



**UNIVERSIDAD AUTÓNOMA DE MADRID**

**FACULTAD DE CIENCIAS**

**DEPARTAMENTO DE QUÍMICA ORGÁNICA**



**APLICACIONES DE LA RMN AL ESTUDIO DE LA ESTRUCTURA  
DE LECTINAS VEGETALES Y DE SUS INTERACCIONES CON  
OLIGOSACÁRIDOS. UNA VISIÓN 3D DE DOMINIOS DE HEVEÍNA  
MODIFICADOS Y DE UNA LECTINA ESPECÍFICA DEL TOMATE.**

**MEMORIA**

que para optar al grado de

Doctor en Ciencias Químicas

Presenta

**MARÍA ISABEL CHÁVEZ URIBE**

Madrid, 2006



La presente Tesis ha sido realizada en el Centro de Investigaciones Biológicas del Consejo Superior de Investigaciones Científicas bajo la dirección del Prof. Jesús Jiménez Barbero, a quien me es grato agradecer su asesoría así como la oportunidad de trabajar en su grupo de investigación.

También me permito agradecer a los Doctores F. Javier Cañada Vicinay, Nuria Aboitiz Cantalapiedra y Blanca López Méndez, por su dedicación, consejos y ayuda proporcionada para el desarrollo del trabajo de investigación que integra esta Tesis.

Asimismo, me permito agradecer al Dr. Tomás Torres, catedrático de la Universidad Autónoma de Madrid, por haber aceptado fungir como mi tutor para llevar a cabo el programa de doctorado en esta Universidad.

Por otro lado, agradezco al Dr. Gabriel Cuevas, Investigador del Instituto de Química de la UNAM, por fungir como responsable ante la Dirección General de Asuntos del Personal Académico, y sobre todo, su acertada sugerencia para integrarme al grupo del Dr. Jesús Jiménez-Barbero.

Por último, tengo a bien expresar mi agradecimiento a la Universidad Nacional Autónoma de México, que mediante la Dirección General de Asuntos del Personal Académico (DGAPA) y a través del Programa de Apoyo a la Superación del Personal Académico (PASPA), ha proporcionado el apoyo financiero correspondiente para la realización de los estudios de doctorado. Extiendo mi reconocimiento a las autoridades e instancias académicas del Instituto de Química de la UNAM, que otorgaron el aval y apoyo necesarios para la realización de los estudios doctorales. También me permito hacer mención de agradecimiento al Consejo Nacional de Ciencia y Tecnología (México) por el apoyo recibido a través del Sistema Nacional de Investigadores.



*A mis padres:  
Arcadia Uribe Torres<sup>†</sup>  
Eduardo Chávez Muñoz*



## ÍNDICE

<b>INTRODUCCIÓN GENERAL</b>	3
<b>GENERALIDADES</b>	
1. Aplicación de la Resonancia Magnética Nuclear en la Determinación de la Estructura Tridimensional de Polipéptidos y Proteínas	
1.1 Introducción	13
1.2 Métodos de RMN	17
1.3 La Asignación de los Espectros	
1.3.1 Identificación de Diferentes Sistemas de Espín	22
1.3.2 Asignación de NOEs Secuenciales, y de Media y Larga Distancia	23
1.4 El Cálculo de Estructuras	26
1.4.1 Los Métodos basados en Geometría de Distancias	27
1.4.1.1 Programa DYANA	
1.4.1.2 Programa CYANA	29
1.4.2 Minimización Energética Restringida a través de Enfriamiento Simulado	34
1.4.3 Minimización Energética Restringida con Moléculas de Agua Explícitas	37
1.5 Simulación Mediante Dinámica Molecular utilizando Disolvente Explícito	37
2. Aplicaciones de la RMN al Reconocimiento Molecular de Carbohidratos por Proteínas	
2.1 Introducción	39
2.2 Método por Perturbaciones de Desplazamiento Químico del Receptor	40
<b>CAPÍTULO 1: DETERMINACIÓN DE LA ESTRUCTURA 3D DE UN DOMINIO DE HEVEÍNA NO NATURAL (AcAMP2Pff18/Pff20) Y SU INTERACCIÓN CON N, N', N''-TRiacetil-QUITOTRIOSa MEDIANTE RMN</b>	
1. Introducción	45
2. Resultados y Discusión	47
2.1 Determinación de la Estructura Tridimensional del Péptido AcAMP2Pff18/Pff20 unido a N, N', N''-Triacetil-Quitotriosa	
2.1.1 Asignación de los Desplazamientos Químicos de <sup>1</sup> H-RMN del AcAMP2Pff18/Pff20 libre y unido a Quitotriosa	47



2.1.2 Generación de las Estructuras 3D del AcAMP2Pff18/Pff20 unido a Quitotriosa mediante el Programa DYANA	52
2.1.3 Optimización de la Estructura 3D del AcAMP2Pff18/Pff20 mediante Minimización Restringida utilizando Simulaciones de Dinámica Molecular	
2.1.3.1 Parametrización del Aminoácido no Natural <i>p</i> -Fluorofenil alanina (Pff)	55
2.1.3.2 Generación de Estructuras Refinadas del AcAMP2Pff18/Pff20 mediante el uso del Programa AMBER	56
2.2 La Estructura Tridimensional del Complejo AcAMP2Pff18/Pff20-Quitotriosa	
2.2.1 Análisis del Sitio de Unión del Trisacárido	60
2.2.2 Estructura del Complejo. Minimización Energética Mediante el Método de Enfriamiento Simulado (SA)	63
2.2.3 Hacia una Visión Tridimensional de la Estructura de AcAMP2Pff18/Pff20 y de su Complejo con Quitotriosa. Dinámica Molecular en Disolvente Explícito (sMD)	64
2.3 La Asociación de AcAMP2Pff18/Pff20 a Quitotriosa. Determinación de la $K_a$ del Proceso y Análisis Comparativo con Valores de $K_a$ y Parámetros Termodinámicos para otros Péptidos Análogos	70
3. Parte Experimental	
3.1 Experimentos 2D-RMN	73
3.2 Cálculo de Geometría de Distancias (DYANA) y Minimización Energética Restringida (REM)	73
3.3 Dinámica Molecular en Disolvente Explícito (sMD)	74
3.4 Protocolo Experimental por RMN de la Valoración para la Asociación de AcAMP2Pff18/Pff20-Quitotriosa	77
4. Conclusiones	77

**CAPÍTULO 2: LA INTERACCIÓN DE UN MUTANTE DE HEVEÍNA TRUNCADA (Hev32S19D) CON OLIGOSACÁRIDOS DE QUITINA**

1. Introducción	81
2. Resultado y Discusión	85
3. Parte Experimental	91
4. Conclusiones	93

## **CAPÍTULO 3: LA ESTRUCTURA 3D EN DISOLUCIÓN DE UN PÉPTIDO AISLADO DEL TOMATE**

1. Introducción	97
2. Resultados y Discusión	98
2.1 Determinación de la Estructura Tridimensional de un Péptido del Tomate	
2.1.1 Asignación de los Desplazamientos Químicos de <sup>1</sup> H-RMN del Péptido del Tomate	98
2.1.2 Generación de las Estructuras 3D del Péptido del Tomate, mediante el Programa CYANA	104
2.1.3 Refinamiento de la Estructura 3D del Péptido del Tomate mediante el Programa AMBER 8.0	108
2.1.3 Simulación con Dinámica Molecular en Disolvente Explícito (sMD) del Péptido del Tomate	112
2.2 Análisis de la Interacción del Péptido del Tomate con Quitotriosa, Quitohexaosa y Quitina utilizando RMN	113
3. Parte Experimental	
3.1 Experimentos de 2D-RMN	114
3.2 Cálculo de Geometría de Distancias con el Programa CYANA y Minimización Energética Restringida (REM) en Disolvente Explícito	114
4. Conclusiones	116
<b>CONCLUSIONES GENERALES</b>	119
<b>BIBLIOGRAFÍA</b>	123
<b>APÉNDICE</b>	133
<b>PUBLICACIÓN</b>	145



## ***INTRODUCCIÓN GENERAL***



El estudio de la estructura tridimensional de las biomoléculas: proteínas, ácidos nucleicos, lípidos y carbohidratos es esencial para la comprensión, a escala molecular, de su función en los procesos vitales.<sup>1</sup>

Más concretamente, los carbohidratos son el grupo de compuestos de origen biológico más abundante en la naturaleza. Presentan una amplia diversidad estructural y funcional y, de hecho, desempeñan papeles clave como nutrientes y fuentes de reserva y transporte de energía o como componentes estructurales de las paredes celulares de los organismos vivos.<sup>2</sup> Además, más modernamente, se ha comprobado que tanto variaciones estructurales de los carbohidratos como alteraciones sutiles de las diversas actividades enzimáticas relacionadas con su síntesis (glicosiltransferasas y glicosidasas) se asocian a muchas patologías.<sup>3</sup>

Sin embargo, la explosión de las glicociencias a partir del último cuarto del siglo pasado quizá se ha debido a la demostración inequívoca de que los carbohidratos intervienen en una gran variedad de procesos fisiológicos, actuando como marcadores de reconocimiento celular. Así, dentro de estos procesos, donde el reconocimiento molecular de carbohidratos específicos por proteínas tiene un gran protagonismo, es posible mencionar las respuestas inmunes e inflamatorias, la organogénesis, la metástasis, y diversos procesos infecciosos, tanto patogénicos como simbióticos.<sup>4</sup>

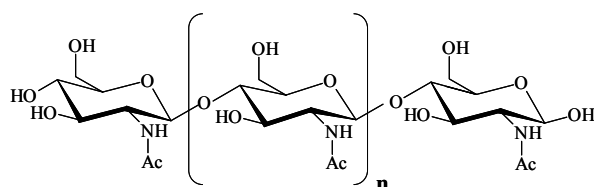
Los receptores de los carbohidratos en muchos de estos eventos biológicos, excluyendo los anticuerpos generados por el sistema inmunitario, son las lectinas, proteínas en muchos casos glicosiladas y sin actividad enzimática, que se encuentran tanto en el mundo animal como vegetal, así como en microorganismos como bacterias, protozoarios y virus.<sup>5</sup> Entre las características que definen a las lectinas, se puede mencionar que algunas tienen la capacidad de aglutinar células (la capacidad de aglutinar eritrocitos fue la primera propiedad con que se definió a las lectinas), aunque actualmente se definen por ser capaces de un reconocimiento específico de carbohidratos, uniéndose reversiblemente a sus ligandos glicosídicos con una exquisita selectividad y sin alterar la estructura covalente de éstos.<sup>6,7</sup>

Las lectinas pueden clasificarse según su estructura molecular, sus propiedades biológicas, y su especificidad de unión a carbohidratos. Su clasificación puede visualizarse en las páginas web donde se definen los módulos de proteínas que

reconocen carbohidratos (carbohydrate binding modules, <http://afmb.cnrs-mrs.fr/CAZY/CBM.htm>) y en la base de datos de lectinas: <http://www.cermav.cnrs.fr/lectines/>). Es interesante mencionar que entre las primeras lectinas que se caracterizaron están las de origen vegetal que, en muchos casos, también se llaman agglutininas o fitohemagglutininas. Por ejemplo, la Concanavalina A (Con A, obtenida a partir de *Canavalia Ensiformes*) fue la primera lectina vegetal analizada y caracterizada mediante difracción de Rayos X, y actúa asociándose específicamente a glucosa y a manosa.<sup>8</sup>

Prestando especial atención a las lectinas vegetales, estas han sido utilizadas ampliamente como herramientas en áreas de investigación biomédica y biológica.<sup>9</sup> Algunas de las lectinas más estudiadas han sido aquellas obtenidas de leguminosas, las lectinas que unen quitina (p. e. dominios de heveína), las que actúan en los ribosomas y aquellas que unen manosa.<sup>10</sup>

Más concretamente, y centrándonos en los estudios realizados en esta Memoria, las lectinas que unen quitina (un polímero de *N*-acetilglucosamina, con enlace glicosídico  $\beta$  (1 $\rightarrow$ 4) (Figura 1) han sido relacionadas con la defensa de plantas frente a insectos y hongos.<sup>11</sup>



**Figura 1.** Estructura química del polímero quitina

Muchas de estas lectinas incluyen un motivo estructural de alrededor de 40 aminoácidos, anudados con 3 ó 4 puentes disulfuro, denominado dominio de heveína.<sup>12</sup> Este nombre procede de la lectina heveína, que es un polipéptido de 43 aminoácidos, que se extrajo por primera vez del látex de la *Hevea brasiliensis*.<sup>13</sup> Independientemente de sus propiedades de defensa en la planta de origen, se ha demostrado que la propia heveína y otras proteínas homólogas presentes en otras plantas están relacionadas con la alergia al látex y con el denominado síndrome cruzado látex-frutas.<sup>14</sup> Los dominios de heveína pueden encontrarse aislados, como en la propia heveína y en la pseudoheveína,

o formando parte de proteínas más grandes e incluso agrupados covalentemente en conjuntos de dos o más dominios. Así, por ejemplo, la lectina de la ortiga (UDA, *Urtica dioica agglutinin*) y la aglutinina del germen de trigo (WGA, *Wheat germ agglutinin*), muestran dos y cuatro dominios de heveína, respectivamente.<sup>15</sup> La actividad aglutinante de la WGA fue una de las primeras propiedades biológicas de estas lectinas que fue demostrada de modo inequívoco y su estructura se caracterizó mediante cristalografía de Rayos X muy tempranamente.<sup>16</sup> Además, fue una de las primeras en las que se estudió su interacción a carbohidratos mediante RMN.<sup>17</sup> Posteriormente, se demostró que la actividad de unión a quitina de estas proteínas es debida a la presencia de uno ó más dominios con gran homología de secuencia y relacionados estructural y funcionalmente con la heveína.<sup>10,18</sup>

En todos los casos descritos anteriormente, las estructuras tridimensionales de ambas entidades, tanto del ligando oligosacárido como del receptor (lectina), juegan un papel primordial. Además, es evidente que el conocimiento de estas estructuras, antes y después de formar el complejo lectina-azúcar, así como de los factores dinámicos y energéticos del proceso de interacción es fundamental para comprender estos fenómenos y poder, en último término, modularlos a voluntad. Por lo tanto, el estudio sistemático de lectinas de pequeño tamaño y de sus interacciones está justificado desde una perspectiva académica, pero también por sus posibilidades de aplicación en el área de la investigación biomédica y biológica.

Investigaciones previas mediante métodos de determinación de estructuras, como cristalografía de Rayos X y espectroscopia de Resonancia Magnética Nuclear o modelado molecular, unidos a estudios de calorimetría y de otras técnicas biofísicas han permitido obtener información detallada acerca de las características estructurales y termodinámicas de una variedad de complejos proteína-carbohidrato.<sup>19-27</sup>

De acuerdo con estos estudios, existen distintos tipos de fuerzas implicadas en el proceso de reconocimiento de carbohidratos por sus receptores proteicos que se diferencian en sus requerimientos de geometría, fuerza y especificidad: enlaces de hidrógeno, fuerzas electrostáticas, de van der Waals e interacciones de apilamiento CH- $\pi$ .<sup>28, 29</sup>

En los sitios de reconocimiento de carbohidratos, aparecen enlaces de hidrógeno cooperativos, bidentados, y formando redes. En general, en un enlace de hidrógeno un



átomo de hidrógeno es compartido entre dos otros átomos, normalmente de marcado carácter electronegativo. En un enlace de hidrógeno, la molécula que posee el hidrógeno que participa en el enlace se denomina donadora, mientras que aquella que porta el átomo electronegativo (oxígeno ó nitrógeno) complementario se denomina aceptora.

Las energías de enlace asociadas con este tipo de enlaces están en el rango de 12 a 30 kJ/mol, por lo que se trata de un enlace relativamente débil. Sin embargo, puede llegar a alcanzar una energía equivalente al 10% de la contenida en los enlaces covalentes C-C, por lo que se le considera el más fuerte de entre los enlaces débiles. Por lo tanto, la presencia de numerosos grupos hidroxilo en los carbohidratos claramente facilita a estos una participación en enlaces intermoleculares de hidrógeno con grupos donadores y aceptores de puentes de hidrógeno de la proteína.<sup>30</sup>

Por otro lado, aunque los azúcares sean marcadamente hidrófilos, las orientaciones específicas de sus grupos no polares C-H pueden crear interacciones hidrófobas, que pueden interactuar recíprocamente con una cavidad, también hidrófoba, del sitio de unión del receptor.<sup>31</sup>

Dentro de las interacciones no polares se encuentran las fuerzas de van der Waals, que son fuerzas no específicas atractivas entre dos átomos o grupos de átomos. Estas fuerzas son más favorables si los átomos se encuentran a 0,2-0,3 nm de distancia mutua, dependiendo de su naturaleza. La fuerza proviene de la atracción transitoria entre el núcleo de un átomo y la nube de electrones del átomo vecino. La energía de enlace de las fuerzas de van der Waals es de aproximadamente 4 kJ/mol. Es decir, estas fuerzas son más débiles y menos específicas que los enlaces de hidrógeno, aunque la suma de muchas interacciones de van der Waals puede apoyar, considerablemente, la formación de enlaces de hidrógeno entre el carbohidrato y las cadenas laterales de la proteína, proporcionando una mayor estabilización del complejo.<sup>30, 32</sup>

Adicionalmente, otras interacciones relevantes para el reconocimiento molecular de carbohidratos por proteínas, y que tienen componentes polares y no polares, son las interacciones entre los grupos C-H de los residuos del carbohidrato con el sistema de electrones  $\pi$  de residuos de aminoácidos aromáticos (triptófano, fenilalanina y tirosina). Estos residuos se localizan frecuentemente en los sitios de unión con carbohidratos. Estas interacciones de apilamiento CH- $\pi$  son determinantes de la conformación y de la

afinidad de los carbohidratos como ligandos, por lo que permiten la formación y modulan la selectividad y estabilidad de los complejos proteína-carbohidrato.<sup>33</sup>

Por otro lado, para entender, de la manera más rigurosa posible, los factores involucrados en el proceso de reconocimiento molecular de carbohidratos por proteínas, es fundamental conocer los parámetros termodinámicos del proceso de asociación.<sup>34</sup>

En este sentido, se ha demostrado que, de modo general, existe un fenómeno de compensación entalpía-entropía, que se ha atribuido a la pérdida de grados de libertad conformacional tanto de la proteína como del carbohidrato al formarse el complejo.<sup>35-37</sup>

En las últimas dos décadas, la utilización y el desarrollo de nuevos métodos de Resonancia Magnética Nuclear ha permitido investigar en profundidad la estructura y características conformacionales de muy diversos carbohidratos en disolución.<sup>38</sup>

Normalmente, los datos experimentales se combinan con simulaciones de mecánica o dinámica molecular (MM y MD) para llevar a cabo la descripción fidedigna de la estructural tridimensional de carbohidratos y proteínas, bien libres, bien asociados, y así comprender, en mayor o menor medida, los factores conformacionales implicados en el proceso de reconocimiento molecular.<sup>39,40</sup>

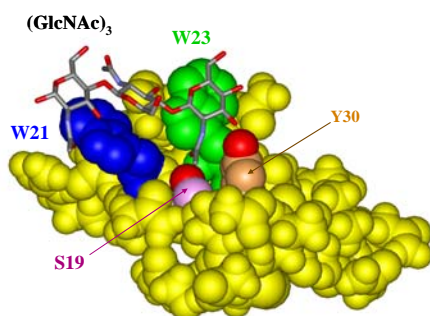
Respecto al tema de estudio de esta Memoria, en los últimos años, se ha determinado la estructura tridimensional de distintos dominios de heveína, tanto mediante cristalografía de Rayos X como por Resonancia Magnética Nuclear, en estado libre o asociado a oligosacáridos (Figura 2). Por ejemplo, se han descrito los casos de la WGA<sup>16,41</sup> y de su dominio B aislado,<sup>42</sup> de la UDA libre<sup>43</sup> y asociada a quitotriosa,<sup>43,44</sup> de la heveína de látex libre,<sup>45</sup> asociada a quitobiosa,<sup>46</sup> quitotriosa<sup>47</sup> y quitopentaosa,<sup>48</sup> de una heveína truncada de 32 aminoácidos,<sup>49</sup> de la heveína de saúco,<sup>50</sup> de la pseudoheveína,<sup>51</sup> de la PL-D2 (péptido aislado de la raíz de la fitolaca o hierba carmín, *Pokeweed*),<sup>52</sup> del péptido antifúngico EAFF2 aislado del árbol de la gutapercha, *Eucommia ulmoides*,<sup>53</sup> o del péptido antifúngico AcAMP2.<sup>54</sup>

## Introducción General

	1	10	20	30	43	
Heveína	EQ	CGRQAGGKLC	PNNLCCSQ	WGWCGSTDE	YCS	PDHNCQSN CKD
Heveína 32	EQ	CGRQAGGKLC	PNNLCCSQ	WGWCGSTDE	YCS	
Pseudoheveína	EQ	CGRQAGGKLC	PNNLCCSQ	YGWCGSSDD	YCS	PSKNCQSN CKGGG
WGA B	KRC	GSQAGGATCP	NNHCCSQ	YGHCGFGAE	YCGAG	CQGGPCRAD
UDA A	QR	CGSQGGGGT	CPALWCC	SIWGWCGDSE	YCGRT	CENK CWSGERSD
PL-D2 A	APE	CGERASGKRC	PNGKCCSQ	WGYCGTTDN	YCGQG	CQSQ CDY
EAFP2	QTC	ASRCP RPC	NAGLCC	SIYGYCGSGAA	YCGAG	NCRCQ CRG
AcAMP2	VG	ECVR	GRC	PSGMCCSQ	FGY	CGKGPKYCGR

**Figura 2.** Alineación de secuencias de aminoácidos de la heveína, la heveína 32 (una heveína truncada) y otros péptidos con dominio de heveína. Se indican en color los aminoácidos que participan en el sitio de unión con el carbohidrato.

En base a los estudios previos de Rayos X y RMN,<sup>41-54</sup> se ha determinado que, en estos dominios polipeptídicos, y considerando la numeración correspondiente a la heveína, los residuos aromáticos en posiciones relativas 21, 23 y 30 juegan un papel clave en la unión con el carbohidrato, permitiendo la estabilización del complejo a través de apilamientos carbohidrato-aromático, mediante interacciones de van der Waals y CH/ $\pi$ . Adicionalmente, el grupo hidroxilo de la serina 19 forma un enlace de hidrógeno con el carbonilo del grupo acetamido de una unidad de *N*-acetilglucosamina, al mismo tiempo que el grupo metilo de este acetamido interactúa con los anillos aromáticos de la cadena lateral de los residuos 21 y 30. Finalmente, existe un enlace de hidrógeno entre el grupo hidroxilo de la tirosina (Tyr) 30 de la heveína con el OH-3 de la misma unidad de azúcar mencionada anteriormente. (Figura 3).



**Figura 3.** Interacciones de apilamiento CH- $\pi$  y enlaces de hidrógeno entre la heveína truncada de 32 aa y la tri-*N*-acetilquitotriosa.<sup>49</sup>

En la Figura 2 se muestra la secuencia de aminoácidos de algunos dominios de heveína, mostrando los residuos que participan en el sitio de unión.

A día de hoy, uno de los intereses y objetivos fundamentales de nuestro grupo de investigación es el investigar y comprender el origen de la especificidad y estabilidad de los complejos proteína-carbohidrato, profundizando en el conocimiento de la importancia relativa que juegan las interacciones de apilamiento carbohidrato-aromático y los enlaces de hidrógeno en el proceso de reconocimiento molecular. Puesto que los dominios de heveína son relativamente pequeños y manejables, ha sido posible, en colaboración con otros grupos de investigación, disponer de heveínas naturales o mutantes mediante métodos de purificación de extractos,<sup>55</sup> de biología molecular<sup>56</sup> o de síntesis química.<sup>57, 58</sup> De este modo, los dominios de heveína se han constituido como sistemas modelo adecuados para llevar a cabo los estudios encaminados a la comprensión y cuantificación de los factores energéticos y estructurales involucrados en los procesos de interacción proteína-carbohidrato.<sup>59, 60</sup>

En concreto, y dentro de estos objetivos generales, en esta Tesis se han abordado estudios de Resonancia Magnética Nuclear (y simulaciones de dinámica molecular) de un dominio de heveína: el polipéptido mutante 4-fluorofenilalanina (Pff)-AcAMP2, sintetizado químicamente por los Dres. Cecilia Andreu y Gregorio Asensio de la Universidad de Valencia. Además, se describe su interacción con N, N', N''-triacetilquitotriosa. En este péptido se ha modificado la naturaleza química de dos de los aminoácidos aromáticos, 18 y 20, (equivalentes a 21 y 23 en heveína) implicados en el sitio de unión con el carbohidrato.

Asimismo, se ha estudiado, también mediante RMN, la asociación de otro péptido mutante de heveína truncada, Hev32S19D, sintetizado con la perspectiva de que mostrase una especificidad diferente de la natural, con N, N', N'', N''', N''''-Pentacetilquitopentaosa y su derivado de tipo quitosano, monodesacetilado en el residuo del extremo no reductor (N', N'', N''', N''''-tetracetilquitopentaosa).

Finalmente, y con el objeto de continuar con el estudio de la estructura 3D de polipéptidos de origen vegetal que uniesen derivados de *N*-acetilglucosamina, se ha abordado la determinación de la estructura tridimensional, así como el análisis de la interacción con quitotriosa, quitohexaosa y con el polímero quitina, mediante RMN de

otro péptido de 39 aminoácidos aislado del fruto del tomate, también con una estructura de 3 puentes disulfuro anudados, pero sin estar clasificado dentro de la familia de los dominios tipo heveína.

## ***GENERALIDADES***



# 1. APLICACIÓN DE LA RESONANCIA MAGNÉTICA NUCLEAR EN LA DETERMINACIÓN DE LA ESTRUCTURA TRIDIMENSIONAL DE POLIPÉPTIDOS Y PROTEÍNAS.

## 1.1 Introducción

Desde 1985 el desarrollo de la espectroscopia de Resonancia Magnética Nuclear de alta resolución, tanto desde el punto de vista metodológico como de la mejora de los equipos accesibles a los investigadores, ha permitido llevar a cabo los estudios estructurales detallados, a escala atómica, de las biomacromoléculas, incluyendo péptidos, proteínas, nucleótidos, ácidos nucleicos y oligo y polisacáridos.<sup>61-65</sup> También mediante RMN ha sido posible estudiar procesos dependientes del tiempo, como son procesos dinámicos intramoleculares, aspectos termodinámicos y cinéticos de fenómenos biológicos relevantes, como son los procesos de reconocimiento molecular o el plegamiento de las proteínas.

La mayoría de estos procesos se desarrollan en disolución similar a la acuosa y, por tanto, el conocimiento de la estructura, dinámica e interacciones de las biomoléculas en disolución es de gran relevancia. En concreto, mediante el uso de RMN es posible conocer, en casos favorables, la estructura tridimensional de proteínas en condiciones de temperatura, pH, y fuerza iónica definidas, similares a las que se encuentran en el medio fisiológico, dando así el primer paso para intentar conocer la relación entre la estructura y la función de estas biomoléculas.

No obstante, es obvio que la aplicación de métodos de RMN en este campo tiene también muchas limitaciones que, normalmente, tienen que ver con la asequibilidad, estabilidad y tamaño de la proteína objetivo. Es bien conocido que la principal limitación de la RMN tiene que ver con la baja sensibilidad intrínseca técnica. Además, en el caso de proteínas de cierto tamaño, la complejidad de los espectros también representa un cuello de botella importante en el proceso de resolución de la estructura. Sin embargo, estos problemas, al menos en parte, se han ido resolviendo gracias al desarrollo de equipos de muy altos campos magnéticos (comercialmente, de hasta 900 MHz en protón, en la actualidad, aunque el imán de 1 GHz está ya prácticamente disponible), unido al acceso a aditamentos que han permitido aumentar la relación



señal/ruido de manera eficaz. Así, a principios de años 90, se popularizó el uso de unidades de gradientes de campo magnético, mientras que después del año 2000, el empleo de sondas criogénicas se ha extendido ampliamente.

Estos desarrollos metodológicos han venido acompañados del empleo de métodos de biología molecular, más o menos accesibles a muchos investigadores no necesariamente expertos en esta área, que han permitido la sobre-expresión, y posterior purificación, de sus proteínas objetivo, en cantidades suficientes, y con un etiquetado isotópico uniforme y selectivo en núcleos de baja abundancia natural, como  $^{15}\text{N}$ ,  $^{13}\text{C}$  y/o  $^2\text{H}$ , lo que ha permitido a que puedan obtenerse, e interpretarse, distintos espectros multidimensionales de muchas proteínas con mejor calidad y resolución y así, tener acceso a la información necesaria para la obtención de su estructura 3D.

En general, aunque siempre depende del problema concreto, empleando únicamente métodos de RMN homonucleares y experimentos bidimensionales (2D-RMN) es posible determinar la estructura de péptidos o proteínas con un peso molecular de hasta 10 kDa y, en casos muy favorables, casi excepcionales, de hasta 15 kDa. Esta es la aproximación que hemos usado en esta Tesis Doctoral. Por otra parte, para realizar estudios estructurales de proteínas de peso molecular de más de 15 kDa, es imprescindible recurrir a métodos heteronucleares multidimensionales, normalmente de 3D-RMN (ó 4D-RMN), utilizando proteínas enriquecidas isotópicamente con  $^{15}\text{N}$  y  $^{13}\text{C}$ .<sup>66,67</sup> Para proteínas de más de 20 kDa, es también necesario emplear muestras en las que los protones de todas, o algunas de, las cadenas laterales de los aminoácidos componentes se han reemplazado por  $^2\text{H}$ .

Aunque una descripción detallada de la RMN está fuera del contexto de esta Tesis, simplemente mencionar que el fenómeno físico, en sí mismo, fue observado por los grupos de Purcell y de Bloch, ambos físicos, de manera independiente, en el año 1945, lo les permitió conseguir el Premio Nobel de Física. No fue hasta principios de los años 1950 cuando la RMN se utilizó por primera vez en la resolución de un problema químico. Desde entonces, con el desarrollo del NOE y después, con la invención de la espectroscopia bidimensional, la RMN se ha ido desarrollando vertiginosamente, a la par que se ha establecido un formalismo, sencillo y riguroso a la par, que permite predecir, de manera muy fidedigna, el resultado final de los distintos experimentos, por

muy complicados que sean. Así, en 1991, Richard Ernst recibió el premio Nobel de Química por sus contribuciones metodológicas a la RMN. El trabajo de Ernst y sus colaboradores fue complementado por el de muchos otros grupos de investigación que han permitido que los métodos de Resonancia Magnética Nuclear sean hoy un apoyo imprescindible a la investigación tanto en los ámbitos académicos como privados, en muy diversas áreas de la química, bioquímica, biomedicina, investigación farmacéutica, ciencia de materiales,..., etc.

En particular, y dentro del ámbito de esta Memoria, es necesario resaltar la contribución del grupo de Kurt Wüthrich (Premio Nobel de Química, 2002) que, a principios de los años 80, puso a punto un protocolo, eficaz y bien definido, para la determinación estructural de biomacromoléculas, especialmente proteínas y ácidos nucleicos. Así, se establecieron estrategias para realizar las asignaciones completas de las señales de RMN pertenecientes a cada uno de los aminoácidos de la cadena polipeptídica (sistemas de espín), seguidas de otras para la identificación de los aminoácidos contiguos en la secuencia de la proteína (*sequential assignment*). Finalmente, el protocolo se dirigía a la identificación de correlaciones entre las señales de  $^1\text{H}$  RMN de los aminoácidos alejados en la secuencia. En un primer paso, esto se consiguió mediante el uso de experimentos bidimensionales homonucleares (COSY, DQF-COSY, RELAY, TOCSY, NOESY) en los que el paso clave era la asignación e identificación inequívoca de la mayor cantidad posible de los picos de cruce del espectro 2D-NOESY. Por integración de estos, las intensidades correspondientes se convertían en restricciones de distancia interprotónicas que, a través de métodos computacionales basados bien en métodos de geometría de distancias, o de movimientos en el espacio multidimensional de todos los ángulos de torsión de la cadena peptídica, permitían generar un conjunto de estructuras 3D de la proteína, compatibles con los datos experimentales de RMN.

En esta Tesis, una vez obtenidos los espectros, asignadas todas las señales y cuantificados los NOES, se ha empleado el programa DYANA (Dynamics Algorithm for NMR Applications), desarrollado por el grupo de Wüthrich,<sup>68</sup> que se detalla en el Esquema 1 ruta A. Este protocolo ha sido mejorado posteriormente por el programa CYANA (Combined assignment and dynamics Algorithm for NMR Applications) merced a la incorporación de módulos iterativos de asignación automática, que permiten automatizar, en gran medida, el cálculo de la estructuras. Un esquema del protocolo

CYANA, también empleado en esta Memoria, implementado por el grupo de P. Güntert,<sup>69-71</sup> se da en el Esquema 1 ruta B.

Finalmente, en todos los casos, resulta conveniente, y así lo hemos realizado, llevar a cabo un refinamiento geométrico, mediante una minimización energética restringida, para encontrar una estructura (mínimo local) con sentido físico. Dentro de los programas que comúnmente se han aplicado para este último paso, CHARMM,<sup>72</sup> DISCOVER,<sup>73</sup> GROMOS,<sup>74</sup> ..., etc, nosotros hemos empleado el campo de fuerzas AMBER.<sup>75-78</sup>

En esta Tesis, adicionalmente, y con el objeto de evaluar la influencia de los datos experimentales de NOE en la estructura 3D final de la proteína, también se han realizado cálculos de dinámica molecular usando moléculas de agua explícitas, con el campo de fuerzas AMBER, y sin la presencia de las restricciones de NOE.



**Esquema 1.** Procedimientos para la determinación estructural de proteínas. Ruta A) Asignación manual de los picos de cruce NOE y uso posterior del Programa DYANA para el cálculo de la estructura. Ruta B) El protocolo incluye la asignación automática iterativa con el subsiguiente cálculo de la estructura (CYANA) a partir de las posiciones e intensidades de los picos de cruce NOESY.

## 1. 2 Métodos de RMN

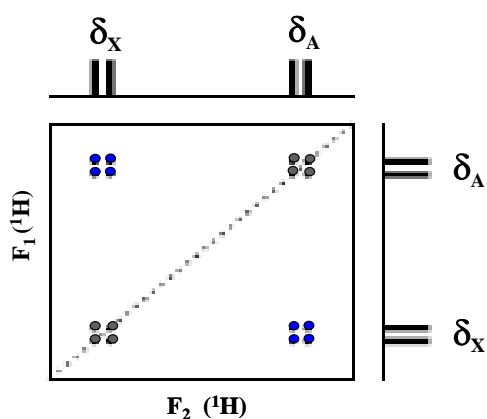
### *UNA BREVE VISIÓN GENERAL DE LOS DISTINTOS EXPERIMENTOS*

En septiembre de 1971, Jean Jeener presentó en la “Ampere Summer School” que tuvo lugar en Basko Polje, en la antigua Yugoslavia, los fundamentos de la Resonancia Magnética Nuclear en dos dimensiones. Posteriormente en 1976, Richard Ernst publicó el primer espectro en dos dimensiones, conocido como COSY (CORrelated SpectroscopY).

Jeener sugirió un experimento consistente en dos pulsos  $\pi/2$ , separados por un tiempo  $t_1$  variable, y de un tiempo  $t_2$  que mediría el tiempo transcurrido después del segundo pulso. Midiendo la respuesta  $s(t_1, t_2)$  de la secuencia de dos pulsos, y haciendo una transformada de Fourier con respecto a las dos variables de tiempo, se produciría un espectro en dos dimensiones  $S(\omega_1, \omega_2)$ .

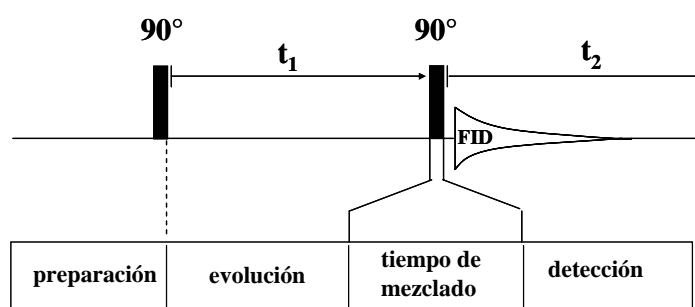
Los dos dominios de frecuencias son el directo,  $F_2$ , (de adquisición) y el indirecto  $F_1$ , (de evolución). Las señales finales son directamente detectadas por el receptor durante el periodo de adquisición,  $t_2$ , aunque están moduladas por las frecuencias operativas en el período de evolución,  $t_1$ . Un espectro 2D es en realidad una combinación de experimentos 1D idénticos, que solamente varían en el valor de  $t_1$  de cada uno de ellos, ya que este se incrementa de manera sistemática (en  $\Delta t_1$ ) desde el primero hasta el último.

Los espectros bidimensionales de correlación homonuclear (figura 1, como COSY, TOCSY, NOESY, etc.) son simétricos con respecto a la diagonal. Las señales fuera de la diagonal, también llamadas picos de cruce, con coordenadas  $(\delta H_A, \delta H_X)$  y  $(\delta H_X, \delta H_A)$ , indican que existe interacción, o correlación entre los núcleos  $H_A$  y  $H_X$ . Dependiendo del tipo de experimento, la correlación puede deberse a la existencia de acoplamiento escalar o a que compartan un mecanismo de relajación dipolar, o de intercambio químico.



**Figura 4.** Espectro COSY, correlación  $^1\text{H}$ - $^1\text{H}$  para un sistema AX.

En general, las secuencias de pulsos de dos dimensiones (2D) contienen siempre los siguientes elementos: preparación, evolución, mezclado y detección. Figura 5



**Figura 5.** Periodos de tiempo que permiten generar un espectro 2D.

Durante el periodo de preparación, al sistema de espines se le permite regresar al equilibrio térmico y, después, mediante los pulsos adecuados, se generan coherencias (magnetización transversal, desde una perspectiva clásica). En el caso del COSY, se emplea un pulso de  $90^\circ$  ( $\pi/2$ ).

El periodo de evolución  $t_1$  permite hacer el etiquetado de las frecuencias en la dimensión de evolución. Es un período crítico, ya que la resolución digital de la segunda dimensión depende del total de valores de  $t_1$ , denominados incrementos, empleados en la adquisición del espectro. Durante este periodo, los espines evolucionan a una frecuencia determinada por su desplazamiento químico (y sus acoplamientos escalares).

Los experimentos de dos dimensiones, difieren en la forma en que la magnetización (coherencia) del núcleo A es transferida al núcleo X. Esta transferencia se lleva a cabo durante el periodo de mezclado, en el que, mediante los pulsos e intervalos de tiempo adecuados, parte (o toda) la coherencia que durante el período  $t_1$ , había evolucionado a la frecuencia de un núcleo A se transfiere a otro núcleo, B.

Como se ha mencionado, la transferencia de información puede darse por alguna de las siguientes interacciones:

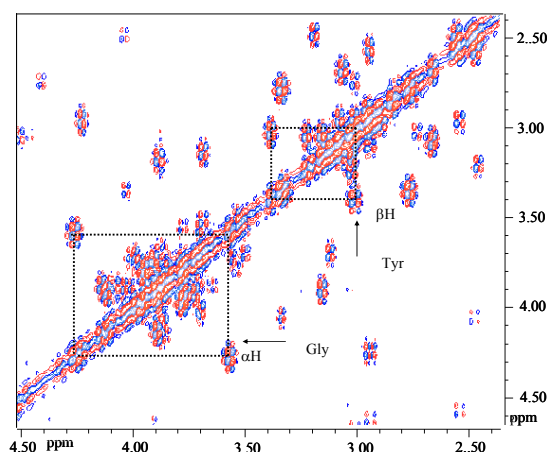
- 1.- Acoplamiento escalar (COSY, DQ-COSY, TOCSY )
- 2.- Relajación Dipolar (NOESY, ROESY).
- 3.- Intercambio químico (EXSY).

Finalmente, en el periodo de detección, cada espectro correspondiente a un valor de  $t_1$ , es adquirido durante el tiempo de adquisición  $t_2$ . Durante este período las señales también contienen la información de las frecuencias a las que habían evolucionado durante  $t_1$ , por lo que se obtiene una matriz de intensidades  $s(t_1, t_2)$ , que después de ser transformada (FT) en ambas dimensiones produce el espectro bidimensional deseado, con dos dimensiones de frecuencia independientes. Las coherencias que no se han transferido totalmente en el periodo de mezclado, han evolucionado con las mismas frecuencias durante ambos periodos de tiempo y dan lugar a los picos diagonales. Aquellas que se han transferido, al menos en parte, han evolucionado con distintas frecuencias en los periodos  $t_1$  y  $t_2$  y, tras la doble FT, dan lugar a los picos de cruce.

### **2D-DQF-COSY (Double Quantum\_Filtered COrelated Spectroscopy)<sup>79</sup>**

La secuencia de pulsos es la siguiente:  **$d1-\pi/2-t_1-\pi/2-\pi/2-t_2$** .

Este experimento identifica las parejas de protones acoplados escalarmente, que pertenecen a un mismo sistema de spin. Las señales provenientes de núcleos no acoplados escalarmente a otros (por ejemplo, un grupo OMe) no aparecen en el espectro. Las características de fase de los picos de este experimento no se van a detallar en esta Memoria, aunque puede mencionarse que los picos de cruce cercanos a la diagonal y adyacentes a señales intensas son fácilmente distinguibles.



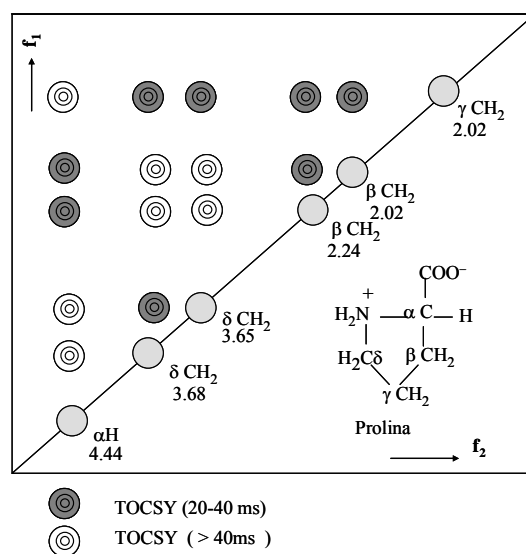
**Figura 6.** Espectro 2D-DQF-COSY de un péptido, se muestran el acoplamiento escalar de los protones  $\alpha$  de la glicina y los  $\beta$  de una tirosina.

### TOCSY (TOtal Correlated SpectroscopY).<sup>80</sup>

La secuencia de pulsos es: **d1- $\pi/2$ -t1-spin-lock-t<sub>2</sub>**.

El pulso  $\pi/2$  crea magnetización en el plano x, y, que evoluciona libremente en el periodo  $t_1$  bajo la influencia de los propios desplazamientos químicos y acoplamientos escalares. A continuación, la magnetización se confina en un lugar de este plano mediante el efecto de un pulso largo (de confinamiento, *spin-lock*, de decenas de ms). El pulso de confinamiento se logra normalmente por medio de una sucesión de pulsos (por ejemplo, MLEV-17), que proporcionan un campo efectivo  $B_1$  que se ajusta a alrededor de 7-12 kHz y que mantiene a la magnetización “confinada”. En estas condiciones de acoplamiento fuerte (los spines pierden la noción de su frecuencia de resonancia), se produce una transferencia de coherencia entre los espines que pertenecen al mismo sistemas de spin, mientras existan spines acoplados a otro común. Cuanto más largo sea el pulso (tiempo de mezclado) más se favorece la transferencia y puede llegar a espines situados relativamente alejados en el sistemas de spines; i. e., a 5 ó 6 enlaces.

Normalmente, en proteínas se obtienen varios espectros TOCSY con diferentes valores de tiempo de *spin lock*. Usando entre 20-40ms se obtienen los picos de cruce de los espines directamente acoplados, mientras que con tiempos mayores pueden identificarse los espines alejados en la cadena. (figura 7)



**Figura 7.** Representación de picos de cruce TOCSY para el aminoácido prolina.

### NOESY (Nuclear-Overhauser-Effect-Spectroscopy)<sup>81</sup>

La secuencia de pulsos es:  $d1-\pi/2-t_1-\pi/2-\tau_m-\pi/2-t_2$ .

El espectro obtenido mediante el experimento NOESY representa la herramienta principal para el análisis conformacional y la determinación estructural de biomoléculas en solución. El NOESY permite obtener información sobre proximidad entre núcleos a través del espacio, siempre y cuando se encuentren a distancias menores de 5-6 Å.

Esta información se encuentra codificada en los picos de cruce que, en un NOESY, son debidos a la interacción dipolar espín-espín. La interacción dipolar da origen al efecto NOE (*Nuclear Overhauser Effect*), mediado por la relajación cruzada entre núcleos próximos. La intensidad del efecto NOE es proporcional a  $1/r^6$  y, por lo tanto, los picos de cruce debidos a la correlación de núcleos que se encuentren muy cercanos serán muy intensos, mientras que los producidos por núcleos, cuyas distancias internucleares sean de alrededor o mayores de 5 Å, serán débiles o inexistentes.

El efecto NOE es generado durante el tiempo de mezclado  $\tau_m$ . Dependiendo de su tamaño, en proteínas resulta conveniente registrar varios espectros NOESY con diferentes valores de  $\tau_m$ , entre 50 ms y 300 ms. A tiempos mayores se propicia el fenómeno conocido como difusión de espín, que permite la transferencia de



magnetización entre dos espines, aunque se encuentren alejados, siempre que tengan uno o varios próximos a ambos, por lo que pueden aparecer picos de cruce entre parejas de protones relativamente alejados entre sí. Este fenómeno depende del tamaño de la molécula y es más eficaz cuanto más grande sea esta (o el campo magnético).

### 1.3 La Asignación de los Espectros

La determinación de la estructura 3D de proteínas de peso molecular de hasta 10 kDa se lleva a cabo mediante los espectros bidimensionales homonucleares H-H: COSY, DQF-COSY, TOCSY y NOESY. Con los tres primeros se realiza la identificación de cada sistema de espín (protones correspondientes para cada aminoácido), y el último se aplica para determinar las conectividades secuenciales, a media y a larga distancia. La cantidad del número de picos de cruce NOE asignados conforman el conjunto de restricciones de distancia que se utilizan en el cálculo para la obtención de la estructura tridimensional. El proceso de asignación se lleva a cabo a través de protocolos establecidos por K. Wüthrich.<sup>62</sup>

#### 1.3.1 Identificación de Diferentes Sistemas de Espín

La identificación de los sistemas de espín se lleva a cabo con el TOCSY y el COSY. En el COSY, solamente se observan las correlaciones entre protones mutuamente acoplados escalaramente; es decir, que se hallan a dos o tres enlaces. Por ejemplo, los mediados por la constante de acoplamiento a tres enlaces que resulta del ángulo de torsión *phi* entre la pareja H $\alpha$ -NH de la cadena principal de cada aminoácido. En el TOCSY, se observan correlaciones entre todos los protones que contiene cada uno de los aminoácidos de la proteína.

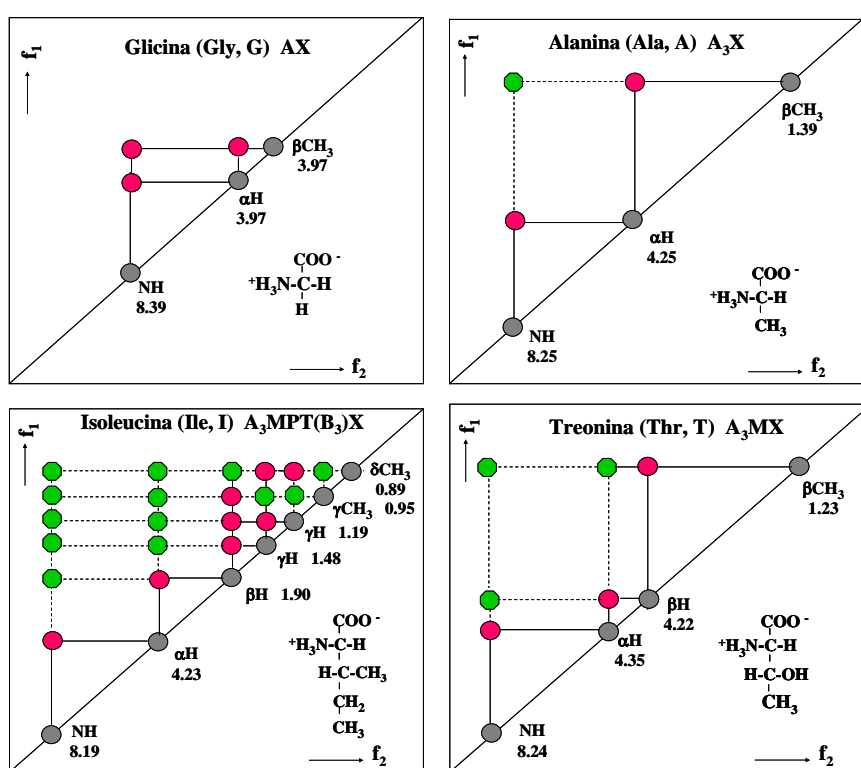
Una vez separados los sistemas de espín, el primer paso es identificar los sistemas de espines de los aminoácidos que contienen patrones característicos (*únicos*), que son:

- Glicina (sistema de espín AX), se localiza fácilmente por contener dos H $\alpha$ ,
- Alanina (A<sub>3</sub>X), por el acoplamiento COSY del H $\alpha$  a único metilo  $\beta$ ,

-Valina ( $A_3B_3MX$ ), Isoleucina ( $A_3MPT(B_3)X$ ) y Leucina ( $A_3B_3MPTX$ ) se reconocen de manera fácil por que a partir de sus correspondientes metilos se puede observar una fila de conectividades TOCSY características

-Treonina ( $A_3MX$ ), por el metilo  $\gamma$  aislado y su acoplamiento de éste con un  $H\beta$ , y de este con un  $H\alpha$ , fácilmente identificables por sus desplazamientos químicos similares.

Figura 7a.

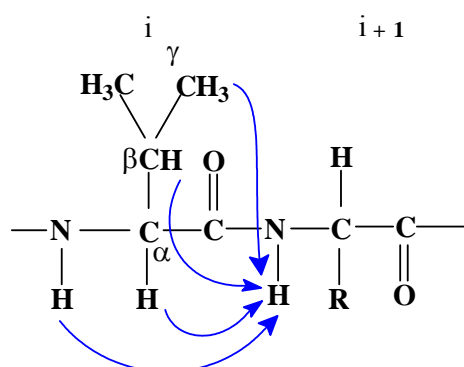


**Figura 7a.** Sistemas de Espín de algunos aminoácidos con patrones característicos. Los círculos de color gris oscuro son picos diagonales, los de color rojo indican correlaciones COSY, y los círculos verdes indican correlaciones TOCSY.

### 1.3.2 Asignación de NOEs Secuenciales, y de Media y Larga Distancia.

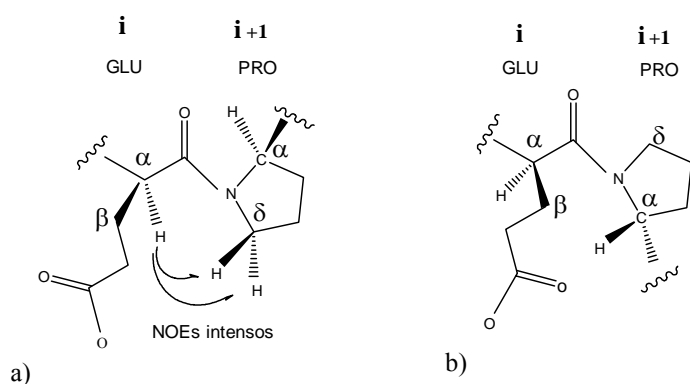
Para llevar a cabo estas asignaciones, se parte de los sistemas de espín de los residuos previamente identificados mediante los espectros TOCSY. En primer lugar, se realiza la determinación secuencial ordenada y una vez realizada esta se concluye la asignación completa de la mayor cantidad posible de los picos de cruce contenidos en el espectro NOESY

La conectividad secuencial resulta de la interacción dipolar, a través del espacio, entre los protones amídicos de la cadena principal HN de un residuo (i) con su homólogo del residuo siguiente HN (i+1) o bien, entre este último protón amídico HN (i+1) de la cadena principal con los protones H $\alpha$ , H $\beta$ , H $\gamma$ , H $\delta$  del residuo anterior (i), como se aprecia en la Figura 8.



**Figura 8.** Se indican en azul las interacciones NOE secuenciales de protones del Residuo (i) con el NH del residuo (i+1)

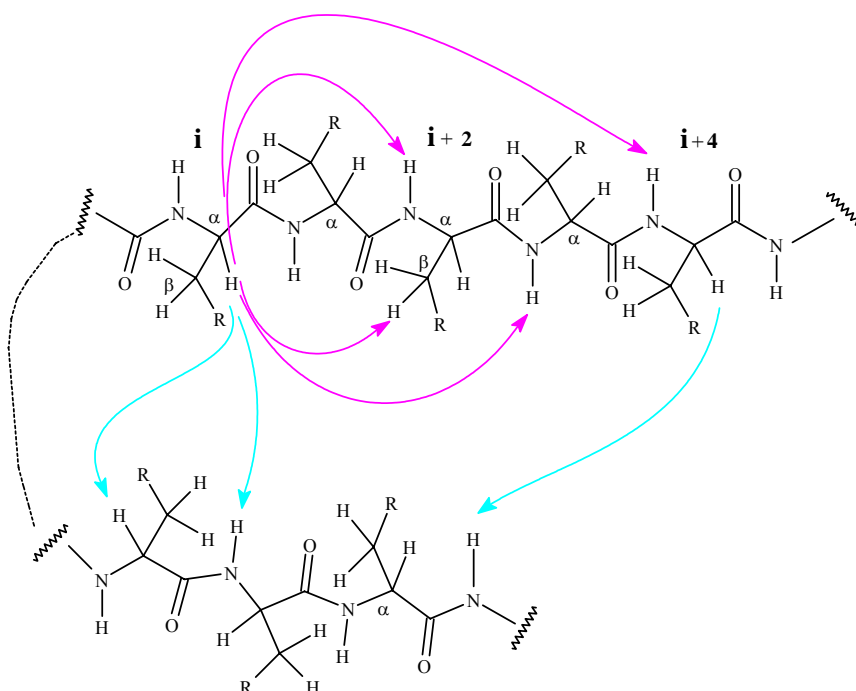
Cabe mencionar que la conectividad secuencial es interrumpida por los residuos de prolina, que no poseen protón amídico. Sin embargo, cuando una prolina en disposición (i+1) adopta una conformación *trans*, se observa interacción NOE entre los H $\delta$  de la misma con el H $\alpha$  del residuo (i). Además de la asignación secuencial, este argumento también permite discernir la presencia de conformación *trans* o *cis* en el enlace peptídico correspondiente, como se observa en la Figura 9.



**Figura 9.** Conformaciones de la prolina a) *trans*, hay interacciones NOE Intensas de los protones  $\delta$  con el aminoácido anterior y b) *cis* se puede considerar interacciones entre el H $\alpha$  y los protones del sistema de espín del aminoácido anterior

Las correlaciones a media distancia (figura 9a) son aquellas interacciones NOE que se observan entre los protones, tanto de cadena principal como de la cadena lateral, de los aminoácidos que se hallan dentro un segmento de cuatro residuos consecutivos.

Finalmente, las correlaciones a larga distancia son aquellas que se originan mediante el acoplamiento dipolar entre los protones de residuos que se encuentran al menos cinco posiciones separados dentro de la secuencia ( $|j-i| \geq 5$ ). Figura 9a



**Figura 9a.** Representación de correlaciones NOE a media y larga distancia. En color rosa, las correlaciones a media distancia y en color azul, las de larga distancia

Las estructuras secundarias de la proteína por RMN se definen por la presencia de patrones característicos de interacciones NOE de corta y media distancia, que involucran a los protones NH, H<sub>α</sub>, y H<sub>β</sub>.

Así, por ejemplo, los valores de distancias claves para una lámina β antiparalela son:  $d_{\alpha\alpha}(i,j)=2.3 \text{ \AA}$ ,  $d_{\alpha N}(i,j)=3.2 \text{ \AA}$  y  $d_{NN}(i,j)=3.3 \text{ \AA}$ . Para una lámina β paralela, los mismos pares de protones son:  $d_{\alpha N}(i,j)=3.0 \text{ \AA}$ ,  $d_{NN}(i,j)=4.0 \text{ \AA}$  y  $d_{\alpha\alpha}(i,j)=4.8 \text{ \AA}$ . Para la hélice-α  $d_{NN}(i,i+1)=2.8$ ,  $d_{\alpha N}(i,i+1)=3.5$ ,  $d_{\alpha N}(i,i+3) = 3.6-4.4 \text{ \AA}$  y  $d_{\alpha N}(i,i+4)=4.4$ .

Puesto que los NOE son inversamente proporcionales a la inversa de la sexta potencia de la distancia entre los núcleos involucrados y resulta realmente difícil cuantificar de forma precisa los picos de cruce del NOESY, normalmente, los NOEs se convierten a restricciones de distancia en Å, con la particularidad de que la distancia concreta se mantiene entre un límite superior y uno inferior. Así, los NOEs se suelen clasificar en fuertes o (intensos), producidos por parejas de protones que se encuentran a una distancia ( $r$ ) < 2.7 Å. Se supone que para los NOEs de intensidad mediana, los protones se encuentran a una distancia,  $r < 3.3$  Å, mientras que los débiles deben encontrarse, normalmente, a una distancia máxima,  $r < 5.0$  Å.<sup>62</sup> En todos los casos, las restricciones inferiores se establecen en base a la suma de los radios de van der Waals de dos protones; es decir, 1.8 Å.

Normalmente, el análisis, la asignación y la interpretación de los espectros se realiza mediante el uso de programas desarrollados específicamente para este fin, como son:

XEASY: <http://www.mol.biol.ethz.ch/wuthrich/software/xeasy/index.html>

NMRview: <http://nmrview.com>

SPARKY: <http://www.cgl.ucsf.edu/home/sparky/>

En esta Tesis se ha utilizado XEASY, en el que los archivos que se generan de la asignación del espectro NOESY son:

- Lista con los desplazamientos químicos de cada uno de los protones de los aminoácidos de la proteína. (**.prot**)
- Lista con el número de picos de cruce asignados, con el valor del volumen (intensidad) correspondiente a cada pico y que, posteriormente, son convertidos a restricciones de distancia (**.peaks**).

#### 1.4 El Cálculo de Estructuras

Para la conseguir el objetivo deseado en este tipo de investigación; esto es, conseguir una estructura lo más real y fidedigna posible, se requiere utilizar, de la manera adecuada, las restricciones de distancia determinadas experimentalmente mediante

NOEs. Esto se consigue mediante el empleo de métodos computacionales que, además de incluir las restricciones de los NOEs, también consideran los parámetros intrínsecos de la estructura covalente del péptido (longitudes de enlace, ángulos de enlace). En todos los casos, es necesario también conocer la secuencia de aminoácidos.

Puesto que en el espacio conformacional multidimensional de los grados de libertad de torsión de una proteína, esta puede adoptar muy variadas conformaciones, es necesario identificar en el espectro NOESY tantas restricciones de distancia como sea posible. Así, es posible restringir el espacio conformacional accesible a la cadena polipeptídica, y definir adecuadamente la estructura tridimensional en disolución.

#### 1.4.1 Los Métodos Basados en Geometría de Distancias

El primer paso para generar la estructura tridimensional es utilizar estas restricciones de distancia experimentales (NOE), mediante un algoritmo basado en un método matemático denominado geometría de distancias.<sup>82, 83</sup> Este protocolo crea una matriz de distancias interatómicas, que genera estructuras consistentes con las restricciones de distancia aplicadas.

Estos métodos pueden funcionar en el espacio de coordenadas cartesianas o en el espacio de los ángulos de torsión, que es el más utilizado.<sup>84, 85</sup> En cualquier caso, el proceso parte de una estructura al azar; por ejemplo, con una cadena peptídica totalmente desplegada.

##### 1.4.1.1 Programa DYANA (**DY**namics **A**lgorithm for **NMR** **A**pplications)

El programa DYANA<sup>68</sup> se aplica para el cálculo de la estructura tridimensional de proteínas, a partir de las restricciones de distancia y ángulos de torsión obtenidos mediante RMN. DYANA es la versión avanzada del programa original, DIANA,<sup>86</sup> y se basa en una combinación de enfriamiento simulado y dinámica molecular en el espacio multidimensional de los ángulos de torsión, mientras que DIANA funcionaba mediante el empleo de una minimización de en una función variable  $T$ , mediante gradientes conjugados, en el espacio de los ángulos de torsión, hasta encontrar la estructura 3D que cumpla las restricciones experimentales. En DYANA, la función variable  $T$  (función diana), toma el papel de la energía potencial  $E_{pot}$  y el sistema se acopla a un baño de alta

temperatura, que posteriormente se enfría lentamente, permitiendo que el sistema cruce las barreras de energía entre los mínimos locales.

La función variable  $T$  se define mediante la ecuación siguiente:

$$T = \sum_{c = u, l, v} w_c \cdot \sum_{\alpha, \beta \in I_c} \left[ \Theta_c \left( \frac{d_{\alpha\beta}^2 - b_{\alpha\beta}^2}{2b_{\alpha\beta}} \right) \right]^2 + w_a \sum_{\alpha \in I_a} \left[ 1 - \frac{1}{2} \left( \frac{\Delta_a}{\Gamma_a} \right)^2 \right] \Delta_a^2$$

donde  $w_c$  son los coeficientes de pesado de las restricciones de distancia, con límites superiores ( $u$ ), límites inferiores ( $l$ ) e interacciones de van der Waals repulsivas ( $v$ ).

$w_a$  = es el coeficiente de pesado de los ángulos de torsión.

$d_{\alpha\beta}$  = es la distancia entre dos átomos

$b_{\alpha\beta}$  = es la distancia límite entre dos átomos

$\Delta_a$  = es la violación de ángulo de torsión

$\Gamma_a$  = es la anchura media del intervalo no permitido del ángulo de torsión

$c$  = contribuciones los límites superiores ( $u$ ), inferiores ( $l$ ) y estéricos ( $v$ ).

Por lo tanto, el valor de la función variable depende de las violaciones generadas a partir de los límites superiores e inferiores, obtenidos experimentalmente, y del valor mínimo asignado para las interacciones de van der Waals.

Los archivos de entrada y salida de DYANA son los siguientes:

- La librería con los 20 aminoácidos estándar (**.lib**).
- La secuencia de aminoácidos (estructura primaria) (**.seq**).
- La lista de restricciones de distancia superiores (**.upl**) e inferiores (**.lol**).
- El cálculo de la estructura se lleva a cabo con el macro **anneal**.
- El macro de inicialización **init** se ejecuta cada vez que se inicia un proceso

DYANA

Dentro de los archivos de salida se incluyen todos los archivos que contienen la información del desarrollo de la minimización y la lista de las restricciones violadas, el resumen (**.ovw**), que incluye la lista completa de números, sumas, y los máximos del valor residual de las restricciones violadas, en cada paso del cálculo de la estructura. Además, se genera una tabla con las violaciones que se dan en todas las estructuras y

otras con los de valores de ángulos diedros (**.ang**) y coordenadas cartesianas de las estructuras (confórmeros) calculados.

Finalmente, el programa también proporciona los valores de desviación cuadrática media, **r. m. s. d** (*root mean square deviation*), que recoge los valores entre todos los pares de estructuras seleccionadas,<sup>87</sup> tanto para los átomos de la cadena principal como para los de la cadena lateral. Es decir, para dos grupos de  $n$  átomos,  $r_1, \dots, r_n$ , y  $q_1, \dots, q_n$ , con  $\sum_i r_i = \sum_i q_i = 0$ , el RMSD es definido por

$$\text{RMSD} = \min_{R \in SO(3)} \sqrt{\frac{1}{n} \sum_{i=1}^n |\hat{r}_i - R\hat{q}_i|^2}$$

Donde  $R$  es la matriz de rotación y  $SO(3)$  es la rotación del grupo.

Un valor de  $\text{RMSD} \leq 1 \text{ \AA}$  indica una buena calidad de la convergencia de las estructuras individuales dentro del conjunto de las mismas.

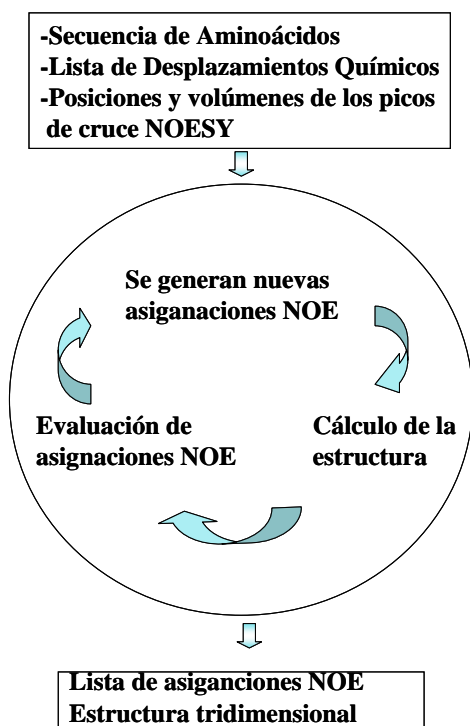
#### 1.4.1.2 Programa CYANA (Combined assignment and dynamics Algorithm for NMR Applications).

En este momento, los métodos automatizados para el cálculo de estructuras mediante RMN están jugando un papel cada vez más prominente. De hecho, es previsible que, en un futuro próximo, probablemente reemplazarán a los protocolos convencionales, manuales, para resolver las estructuras 3D de proteínas en disolución.<sup>70, 71</sup>

En cualquier caso, para la determinación de la estructura 3D, incluyendo estos procedimientos automatizados todavía se requieren las etapas de los protocolos convencionales: preparación de la muestra, obtención de los espectros de RMN, procesamiento de los mismos, localización de los sistemas de espín y asignación secuencial, y, a partir de aquí, automática e iterativamente, asignación de los picos NOE y colección de otras restricciones conformacionales, seguidas del cálculo y refinamiento de la estructura (Figura 10).



Así, una primera aplicación de métodos de automatización se ha dirigido a la asignación de los picos de cruce NOESY, incorporándose a DIANA<sup>86</sup> y DYANA.<sup>68</sup> Entre algunos de estos métodos, cabe citar NOAH,<sup>88</sup> ARIA<sup>89</sup> y CANDID.<sup>69</sup>



**Figura 10.** Esquema general del protocolo para el cálculo de estructura de proteínas mediante asignación automática iterativa del espectro NOESY.<sup>70,71</sup>

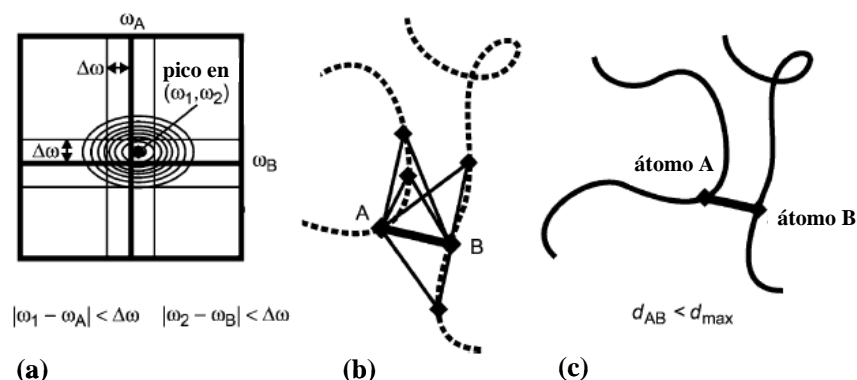
Por ejemplo, NOAH lleva a cabo una asignación múltiple aunque, transitoriamente, no considera los picos de cruce con más de dos posibilidades de asignación. Sin embargo, genera restricciones de distancia individuales para cada una de las posibilidades de asignación, generalmente para los picos de cruce con poca ambigüedad. En el primer ciclo del cálculo se incluyen todos los picos con una o dos posibilidades de asignación. NOAH requiere que todos o la mayoría de los desplazamientos químicos de los protones estén asignados y que las posiciones de los picos de cruce sean las más exactas estén dentro del rango de tolerancia asignado ( $\pm 0.2$  ppm).<sup>70, 88</sup>

ARIA, previamente utilizado con los programas XPLOR<sup>90</sup> y CNS,<sup>91</sup> introduce “restricciones de distancia ambiguas” dentro del proceso de cálculo de la estructura. En

ARIA, el proceso lleva varios ciclos y se inicia a partir de una lista de picos de cruce y de desplazamientos químicos. Normalmente, en el ciclo 0, se genera una estructura inicial que sirve de “plantilla”, mientras que en los ciclos consecutivos se realiza la asignación de los picos NOE, su calibración y un análisis de las violaciones existentes respecto a los promedios de distancia ( $r$ ) interprotónicas calculados para los conformeros de energía más baja del ciclo anterior. Es decir, en cada ciclo se calibran y asignan los espectros NOESY y, finalmente, se calcula un grupo de conformeros con el programa CNS.

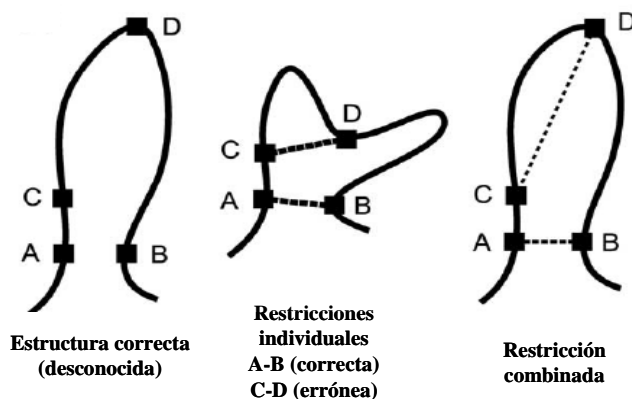
En 2002, el grupo de Wüthrich<sup>69</sup> implementó el módulo CANDID en el programa DYANA, lo que dio origen al programa CYANA (P. Guntert, 2004),<sup>71</sup> que incorpora características de NOAH y de ARIA. CANDID es un método de asignación automática que se basa en el volumen original de un pico NOE. Un ciclo de CANDID comienza generando las asignaciones iniciales para cada pico de cruce, a partir de la lista inicial de picos NOE y la lista de desplazamientos químicos asignados, dentro de los límites de tolerancia. Posteriormente, para cada pico de cruce, estas asignaciones se seleccionan sobre la base de criterios de los valores de desplazamientos químicos y de la posición de los picos de cruce, buscando que haya coherencia dentro una red determinada de NOEs y de que, en su caso, exista compatibilidad con la estructura del ciclo anterior. Por lo tanto, el protocolo está basado en filtros y restricciones de distancia ambiguas, que incluye los conceptos de “*network anchoring*” y de combinación de restricciones (*constraint combination*).

El primer término evalúa que un subgrupo de restricciones asignadas correctamente sea consistente con cualquier red de restricciones de distancia; es decir, que sea suficientemente denso para la determinación de la estructura 3D. El *network anchoring* es un factor importante en la asignación automática mediante CYANA, ya que la estructura que se genera en el primer ciclo tiene que estar definida correctamente, como se indica en la Figura 11.



**Figura 11.** Tres condiciones para la asignación automática válida de picos NOESY mediante CYANA.<sup>70,71</sup> a) Los desplazamientos químicos no deben exceder el límite de la tolerancia. b) “network anchoring” y c) la distancia entre dos átomos (A y B) tiene que ser menor que la distancia máxima.

La combinación de restricciones (*constraint combination*) tiene como objetivo reducir el impacto de restricciones erróneas sobre la estructura resultante. Generalmente, se aplica en los dos primeros ciclos. Genera restricciones de distancia mediante una combinación de asignaciones a partir de de dos restricciones de larga distancia no relacionadas entre sí, como se indica en la Figura 12.



**Figura 12.** Representación esquemática del efecto de una combinación de restricciones correcta para A y B y errónea para C y D. a) El cálculo utiliza estas interacciones como dos restricciones individuales que estarán conformadas para dar el resultado correcto. b) En una conformación deformada. c) Con las restricciones combinadas, una de las distancias es muy corta, lo que permite que se genere una conformación casi sin deformar.<sup>70,71</sup>

En definitiva, CYANA lleva a cabo siete ciclos iterativos de asignación de NOEs, seguidos del consiguiente cálculo de estructura.

Los archivos de entrada que se requieren son:

- La secuencia de aminoácidos (**.seq**)
- La lista de desplazamientos químicos (**.prot**)
- La lista de picos de cruce y sus volúmenes, directamente del espectro NOESY, que pueden ser obtenidos mediante XEASY<sup>92</sup>, NMRview<sup>93</sup> u otros programas.
- La lista de restricciones de distancia inferiores (**.lol**)
- El cálculo de la estructura se lleva a cabo con **calc.cya**
- El macro de inicialización **init.cya** se ejecuta cada vez que se inicia un proceso

En cada uno de los 7 ciclos, CYANA realiza los siguientes pasos:

- 1.- Lee los datos de los archivos de entrada.
- 2.- Calibra las restricciones de distancia, obtenidas de los volúmenes de los NOEs.
- 3.- Genera una lista inicial de asignaciones creando, para cada pico de cruce, una o dos asignaciones basándose en los desplazamientos químicos dentro de los valores de tolerancia dados por el usuario.
- 4.- Filtra las asignaciones iniciales de acuerdo a los criterios mostrados en la Figura 11.
- 5.- Genera restricciones de distancia para todos los picos de cruce con al menos una asignación. Los picos que con una asignación única dan origen a restricciones de distancia no-ambiguas. Aquellos con más de una asignación resultan en restricciones ambiguas.
- 6.- Combina las restricciones de acuerdo con la figura 12
- 7.- Calcula la estructura mediante un proceso dinámico en el espacio de los ángulos de torsión,<sup>68</sup> usando un protocolo de enfriamiento simulado.<sup>94</sup> En este paso se calcula una estructura 3D que se usa como archivo de entrada para el siguiente ciclo.

Es importante mencionar que dentro del proceso del cálculo, en cada ciclo se mejora la precisión de la estructura tridimensional de la proteína.

Respecto a los archivos de salida, en cada ciclo:

- Se genera una lista de asignaciones de los picos de cruce NOESY (**.upl**)
- Se genera información para cada asignación individual (**.noa**)
- Se genera un conjunto de conformeros (**.pdb**)

- Dentro del archivo **.ovw**, se muestra el **r.m.s.d** para los átomos de cadena principal y para los átomos de la cadena lateral.

Finalmente, la estructura CYANA definitiva, dada por los archivos **final.upl** y **final.pdb**, tiene una única lista de restricciones de distancia. Esta estructura posteriormente puede someterse a un proceso posterior de refinamiento, mediante minimización energética con un campo de fuerzas.

#### 1.4.2 Minimización Energética Restringida a través de Enfriamiento Simulado.

Las simulaciones de dinámica molecular (MD) permiten el estudio teórico del comportamiento conformacional y dinámico de biomoléculas. De hecho, pueden proporcionar información detallada sobre las fluctuaciones dependientes del tiempo de todo tipo de moléculas. Actualmente se utilizan de manera rutinaria para investigar la estructura, la dinámica y la termodinámica de las macromoléculas biológicas y sus complejos.

La dinámica molecular está basada en la aplicación de las leyes del movimiento de Newton:

$$F_i(t) = m_i a_i(t)$$

Donde  $F_i$  es la fuerza sobre el átomo  $i$  en el tiempo  $t$ ,  $m_i$  es la masa del átomo  $i$  y  $a_i$  es la aceleración del átomo  $i$  en el tiempo  $t$ . Las fuerzas sobre los átomos se pueden calcular a partir del gradiente de la función de energía potencial. La temperatura  $T$  se puede calcular a partir de las velocidades atómicas que se obtienen de una distribución de Maxwell-Boltzman.

$$3N/2 k_B T = \frac{1}{2} \sum_{i=1}^N m_i v_i^2$$

Donde,  $k_B$  es la constante de Boltzman,  $m_i$  y  $v_i$  son la masa y la velocidad del átomo  $i$  y  $N$  es el número total de átomos ( $3N$  es el número de grados de libertad).

Durante la simulación, la interconversión entre la energía cinética y la potencial hace que la temperatura oscile. Si la temperatura y la presión son constantes, entonces las velocidades y posiciones atómicas pueden calcularse para un tiempo  $(t + \Delta t)$ . Por lo tanto, los átomos cambian su posición, con lo que se genera una nueva conformación. En definitiva, los cambios conformacionales se producen a través de movimientos en una superficie multidimensional, que relaciona la geometría y la energía potencial de las moléculas.

En principio, puede asumirse que la superficie de energía potencial es armónica y permite ajustar las posiciones atómicas en el espacio conformacional correspondiente a un mínimo local en la superficie de energía potencial.

Las estructuras tridimensionales obtenidas mediante los métodos de geometría de distancias o movimientos en el espacio torsional, antes citados, generalmente tienden a dar lugar estructuras alejadas de estos mínimos locales, con niveles energéticos altos, por lo que las estructuras obtenidas directamente de estos métodos son imprecisas e inexactas. Es decir, resulta necesario realizar posteriormente un refinamiento de las estructuras, tal que estas sean sometidas a una minimización energética, incluyendo las restricciones de distancia obtenidas mediante RMN. Es decir, de este modo se consigue relajar la estructura, por medio de movimientos iterativos de los átomos, basados en gradientes de energía potencial, hasta que se obtiene un mínimo local de la energía potencial  $V$ . Normalmente, en biomoléculas, para este fin se emplean simulaciones de dinámica molecular, por lo que al protocolo se le denomina dinámica molecular restringida (rMD)

Existen distintos campos de fuerzas que permiten realizar estas simulaciones. En biomoléculas, los más utilizados son CHARMM,<sup>72</sup> DISCOVER,<sup>73</sup> GROMOS,<sup>74</sup> y AMBER.<sup>75-78</sup>

Normalmente, el campo de fuerzas (*force field*) se define de acuerdo a la suma de las siguientes contribuciones:

$$V_{total} = V_{van\ der\ Waals} + V_{coulomb} + V_{dihedral} + V_{bond} + V_{angle} + V_{NMR}$$

Los dos primeros son los potenciales que se refieren a fuerzas de van der Waals atractivas y repulsivas (potencial de Lennard-Jones para pares de átomos no enlazados

covalentemente) y a las interacciones electrostáticas. Los siguientes tres términos corresponden a los potenciales energéticos necesarios para girar los ángulos de torsión, y para deformar las longitudes y los ángulos de enlace, respectivamente. El último término,  $V_{NMR}$ , corresponde a la penalización que conlleva incumplir las restricciones de distancia experimentales. De hecho, las restricciones de distancia se incorporan de la siguiente manera:

$$K_{NOE} \begin{cases} K_{NOE} (r_{ij} - r_{ij}^u)^2 & \text{si } r_{ij} > r_{ij}^u \\ 0 & \text{si } r_{ij} \leq r_{ij} \leq r_{ij}^u \\ K_{NOE} (r_{ij} - r_{ij}^l)^2 & \text{si } r_{ij} < r_{ij}^l \end{cases}$$

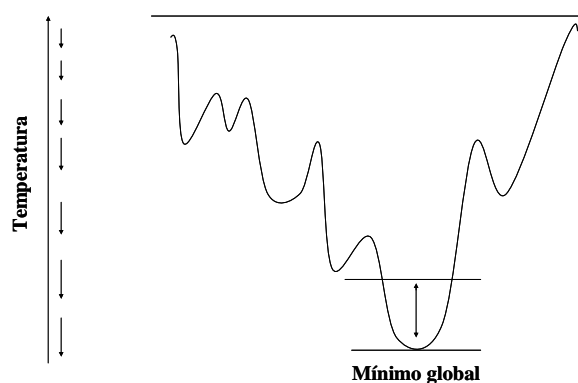
Donde  $r_{ij}^u$  son límites superiores y  $r_{ij}^l$  son los límites inferiores y  $K_{NOE}$  constante de fuerza, del orden de  $1000 \text{ kJ mol}^{-1} \text{ nm}^{-2}$

Puesto que la cantidad de restricciones violadas depende de la estructura inicial, es importante dar valores altos a las constantes de fuerza de estas restricciones, ya que de esta manera, en principio, deben llevar la minimización hacia un mínimo que se ajuste a los datos experimentales.

#### Enfriamiento Simulado (SA) (*Simulated Annealing*)

Mediante la dinámica molecular restringida (**rMD**) los átomos de la estructura inicial están en movimiento hasta que el proceso alcanza un mínimo local. En principio, no existe ninguna garantía de que el mínimo alcanzado sea el mínimo global. Para intentar garantizar que se alcanza este mínimo global, para lo que se requiere salvar las barreras energéticas entre los mínimos locales, primero se simula un calentamiento del sistema a una temperatura alta,  $T$ , que posteriormente se va reduciendo gradualmente (enfriamiento). En todo el proceso se mantiene la influencia del campo de fuerzas lo que permite que se exploren conformaciones diferentes. El resultado de la simulación es una estructura de energía baja, normalmente un mínimo local. El proceso se repite de

manera iterativa y las estructuras obtenidas se usan nuevamente, como iniciales, para otro subsiguiente proceso de enfriamiento simulado (SA). Finalmente, si la convergencia en el espacio conformacional es correcta, se alcanza el mínimo global (Figura 13).



**Figura 13.** Representación del proceso para alcanzar el mínimo global en un proceso de enfriamiento Simulado (SA)

#### 1.4.3 Minimización Energética Restringida con Moléculas de Agua Explícitas.

Más recientemente, se han incluido moléculas explícitas de disolvente (normalmente H<sub>2</sub>O, preferentemente con contraiones) para realizar la determinación de la estructura de proteínas mediante rMD. Este protocolo permite una mejora significativa de la calidad de la estructura.

#### 1.5 Simulación Mediante Dinámica Molecular utilizando Disolvente Explícito

Es evidente que una simulación por dinámica molecular estándar presenta una carencia fundamental, que es la de considerar el disolvente simplemente mediante una constante dieléctrica promedio en el término electrostático del campo de fuerzas. Para mejorar este aspecto, en los últimos años, se han desarrollado métodos computacionales que



permiten llevar a cabo simulaciones en una caja, considerando moléculas de agua explícitas.

En las simulaciones por dinámica molecular, el mayor tiempo de cálculo que se usa es aquel consumido para calcular los términos de no enlace de la función de energía potencial; es decir, las interacciones electrostáticas y las de van der Waals. En principio, es posible reducir considerablemente el cómputo de las interacciones de no enlace, utilizando una distancia de corte (9-12 Å); es decir, asumiendo que las interacciones entre átomos más alejados de ese umbral no tienen importancia. No obstante, existen estudios experimentales que han demostrado la importancia de las interacciones electrostáticas de rango largo por lo que el uso de distancias de corte debe emplearse con cierta precaución.

Normalmente, se emplea el método de sumas Ewald, comúnmente considerado como el adecuado para el tratamiento correcto de las interacciones electrostáticas de rango largo.<sup>95</sup> En concreto, se usa el método denominada *Particle Mesh Ewald* (PME), cuya descripción está fuera del alcance de esta Tesis.

Cuando las moléculas de agua se incluyen explícitamente en la simulación, se emplea lo que se denomina condiciones periódicas de contorno (*periodic boundary conditions*), que crea copias múltiples de la caja de agua, unas a continuación de otras, en todas las dimensiones espaciales, lo que impide que las moléculas de agua se dispersen lejos de la proteína durante la simulación. Además, se utiliza un número limitado de moléculas de agua que mimetice la densidad del medio, de acuerdo con el tamaño de la caja que, a su vez, depende del tamaño de la proteína. Además, la mayoría de las simulaciones se realizan a temperatura y presión constantes.

En las simulaciones de dinámica molecular, se siguen las siguientes etapas:

#### 1.- Iniciación

- a) Minimización energética corta que permite eliminar las interacciones van der Waals muy fuertes (malos contactos), que podrían provocar una distorsión estructural local y, por consiguiente, una simulación inestable.
- b) A continuación, se realiza una dinámica molecular de las moléculas de agua, fijando la proteína, lo que permite que estas moléculas se reajusten a la proteína.

## 2.- Calentamiento del sistema

Se elimina la restricción sobre la proteína y se continúa el proceso de simulación aumentando la temperatura. Al inicio de la simulación, las velocidades atómicas iniciales se generan a una temperatura baja, pero posteriormente se crean nuevas velocidades considerando una temperatura ligeramente cada vez más alta, de manera gradual, hasta que se alcanza la temperatura fijada.

## 3.- Equilibración

Una vez que se alcanza la temperatura deseada, la simulación sigue su proceso, teniendo en cuenta ciertas características clave: la estructura, la presión, la temperatura y la energía. En este punto, de estas características deben permanecer estables en el tiempo.

## 4.- Etapa de producción

La etapa de producción de datos depende del proceso que se desea monitorizar y, por tanto, de la duración de la simulación. Hoy en día, normalmente se acceden a decenas de nanosegundos (ns).

Finalmente, a partir de los datos obtenidos, se puede obtener el promedio de las estructuras y llevar a cabo un análisis conformacional detallado, a escala atómica, también comparado los resultados de la simulación con los experimentales, en su caso.

## 2.- APLICACIONES DE LA RMN AL RECONOCIMIENTO MOLECULAR DE CARBOHIDRATOS POR PROTEÍNAS

### 2.1 Introducción.

Las interacciones proteína-carbohidrato están involucradas en una gran variedad de procesos biológicos, por lo que el conocimiento de la estructura y dinámica de estos complejos es de vital importancia para comprender los fenómenos correspondientes.

En general, desde el punto de vista de la RMN, todos los parámetros espectroscópicos de ligando y receptor se afectan al producirse la asociación ligando (carbohidrato) - receptor (proteína). Es decir, los desplazamientos químicos, constantes escalares y de difusión, velocidades de relajación, NOEs,..., etc se afectan en mayor o menor medida

dependiendo de la afinidad y de la velocidad del intercambio entre los estados libre y asociado.

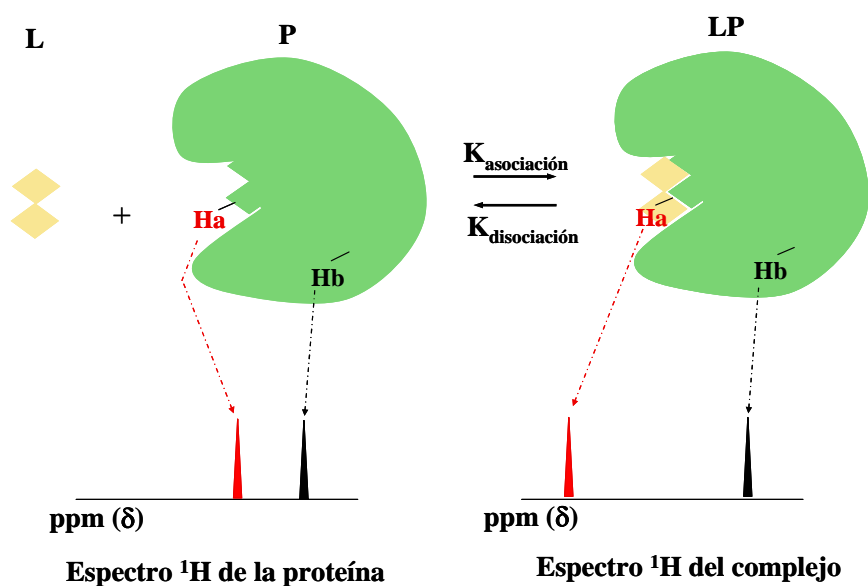
En principio, los métodos de RMN usados para la identificación y estudio de interacciones carbohidrato - proteína pueden dividirse en dos clases:

- a) Siguiendo las variaciones de los parámetros de RMN de las señales del receptor al interaccionar con el ligando (carbohidrato)
- b) Siguiendo los efectos producidos en los parámetros de RMN del ligando al pasar del estado libre a asociado.
- c) La opción entre las dos aproximaciones depende del tamaño de la proteína, de su disponibilidad, y de la  $K_{\text{asociación}}$  del complejo.

En esta Memoria sólo hemos abordado la primera opción; es decir, vamos a centrarnos en la detección de la existencia de interacción siguiendo las señales de la proteína.

## 2.2 Método por Perturbaciones de Desplazamiento Químico del Receptor.

El efecto producido por los diferentes ambientes químicos de los protones de una molécula se traduce en la existencia de distintos desplazamiento químicos. Obviamente, cuando existe asociación ligando-receptor, se generan variaciones en los desplazamientos químicos de ambas especies, específicamente en los núcleos que se encuentran en el sitio de unión y cercanos a éste. La estimación de estas perturbaciones se puede llevar a cabo a través de técnicas de RMN mono y bidimensionales. En los casos en los que se conocen los desplazamientos químicos del receptor (Figura 14).



**Figura 14.** Representación de la asociación ligando-proteína mediante RMN. Se muestra la variación de desplazamiento químico para Ha (rojo), que se localiza en el sitio de reconocimiento. Sin embargo, el desplazamiento químico de Hb (negro) no se afecta.

El seguimiento de estas perturbaciones de desplazamiento químico puede permitir llevar a cabo, mediante titulaciones, la determinación de parámetros termodinámicos ( $K_{\text{asociación}}$ , entalpía de unión, entropía, y la energía libre), usando una aproximación de van't Hoff. Es decir, si el intercambio entre los estados libre y asociado es rápido en escala de tiempos de los desplazamientos químicos, se puede estimar el valor de la constante de asociación. En esta Tesis se han llevado a cabo determinaciones de este tipo mediante metodología estándar.<sup>29, 46, 96, 97</sup>

En estos casos, las variaciones de desplazamiento químico de las señales de la proteína para las distintas concentraciones molares de ligando varían como se describe en la siguiente ecuación:

$$\delta_{\text{obs}} - \delta_{\text{libre}} = (\delta_{\text{comp}} - \delta_{\text{libre}}) \frac{[PL]}{P_T};$$

donde [PL] es la concentración del complejo proteína–ligando,  $P_T$  es la concentración total de proteína,  $\delta_{\text{libre}}$  y  $\delta_{\text{comp}}$  son los desplazamientos químicos de la proteína libre y del complejo, respectivamente;  $\delta_{\text{obs}}$  es el desplazamiento químico observado para una concentración dada de ligando. Puesto que la  $K_{\text{asociación}}$  es directamente proporcional a la concentración de complejo e inversamente proporcional al producto de las concentraciones de la proteína y ligando, la ecuación queda definida:

$$\delta_{\text{obs}} - \delta_{\text{libre}} = (\delta_{\text{comp}} - \delta_{\text{libre}}) \frac{K_{\text{asociación}} [P][L]}{P_T}$$

La concentración de complejo en equilibrio cumple con la siguiente ecuación:

$$[PL] = \frac{1}{2} \left\{ \left( [P_T] + [L_T] + \frac{1}{K_{\text{asociación}}} \right) - \sqrt{\left( \left( [P_T] + [L_T] + \frac{1}{K_{\text{asociación}}} \right)^2 - 4 [P_T][L_T] \right)} \right\}$$

Estos experimentos se realizan manteniendo constante la concentración de la proteína  $P_T$ . Se adquieren los espectros de  $^1\text{H}$  para cada una de las diferentes concentraciones de ligando y la determinación de la  $K_{\text{asociación}}$  se basa en la representación de la curva de valoración  $\delta_{\text{obs}}$  frente a [L], lo que permite realizar un ajuste no lineal a la ecuación, considerando que los parámetros variables son  $K_{\text{asociación}}$  y  $(\delta_{\text{libre}} - \delta_{\text{comp}})$ .



*CAPÍTULO 1*

***DETERMINACIÓN DE LA ESTRUCTURA 3D DE UN DOMINIO DE  
HEVEÍNA NO NATURAL (AcAMP2Pff18/Pff20) Y SU INTERACCIÓN  
CON N, N', N''-TRiacetil-QUITOTRIOSa MEDIANTE RMN.***



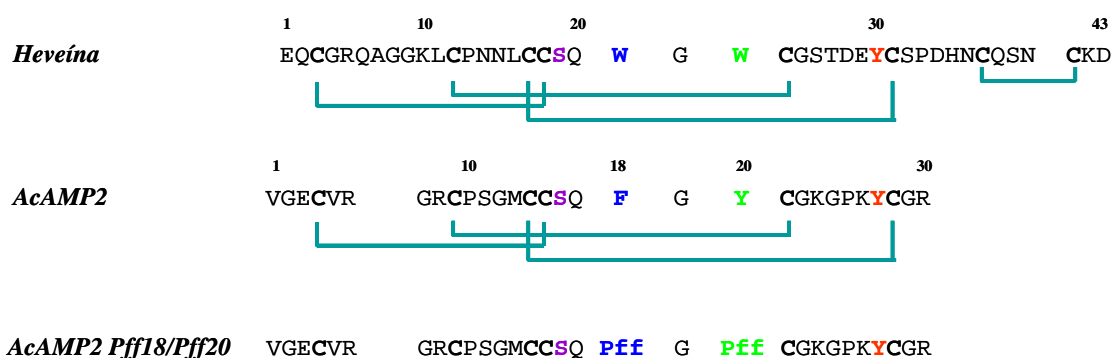


## 1. Introducción

Recientemente, se han incrementado los estudios, mediante Resonancia Magnética Nuclear, de procesos de reconocimiento molecular de carbohidratos por proteínas.<sup>98,99</sup> En nuestro grupo de investigación, venimos trabajando, desde hace unos años, en el estudio de la interacción, a escala atómica, de la interacción de una familia de lectinas, denominadas dominios de heveína, con oligosacáridos de quitina.<sup>42,49,51</sup> De hecho, sus características estructurales intrínsecas, así como su relativa fácil disponibilidad han permitido su utilización como modelos de estudio para investigar el mecanismo, el origen y la participación relativa de los parámetros energéticos y características estructurales involucradas en la formación de los complejos proteína-carbohidrato. En estas investigaciones, hemos podido resolver la estructura molecular de distintos dominios de heveína, bien libres o bien enlazados a carbohidratos. Otros autores también han contribuido a estos estudios. Por ejemplo, el grupo de J.C. Martins, *et al.*,<sup>54,100</sup> estudió el péptido natural AcAMP2 (*Amaranthus caudatus* *AntiMicrobial Peptide*), que está constituido por sólo 30 aminoácidos con tres puentes disulfuro intramoleculares. Este péptido se aísla de las semillas de la planta *Amaranthus caudatus*, conocida por sus propiedades antimicrobianas y antifúngicas.

Como se ha mencionado en la Introducción General de esta Tesis, en la heveína, los residuos que participan en el sitio de unión son la Ser19 y los aminoácidos aromáticos Trp21, Trp23 y Tyr30. Estos mismos aminoácidos, teniendo en cuenta la numeración del AcAMP2, son Ser16, Phe18, Tyr20 y Tyr27 (Figura 1.1).

Los resultados obtenidos hasta la fecha indicaban la importancia de la naturaleza química del anillo aromático para establecer una interacción eficaz con los oligosacáridos. De hecho, en 2004 demostramos que cuánto mayor era la superficie del anillo aromático en la posición 18, mayor era la afinidad y la entalpía de unión de quitotriosa al derivado de AcAMP2 correspondiente.



**Figura 1.1** Alineación de la Secuencia del AcAMP2 y de la heveína, indicando los puentes disulfuro. Además, se muestra la del mutante AcAMP2Pff18/Pff20. Se indican en color los aminoácidos que participan en el sitio de unión.

En este sentido, y con el objeto de continuar estos trabajos, este capítulo está dirigido hacia la resolución de la estructura 3D de un dominio de heveína (realmente de ACAMP2) mutante y el estudio de su afinidad por fragmentos de quitina (N, N', N''-triacetil-quitotriosa) mediante RMN y simulaciones de dinámica molecular. En este péptido (AcAMP2Pff18/Pff20), se ha modificado la naturaleza química de dos de los aminoácidos implicados en el proceso de reconocimiento molecular, la Phe18 y la Tyr20, sustituyendo ambos aminoácidos por uno no natural, la *p*-fluorofenilalanina (Pff). Esta modificación se ha realizado con el objetivo de comprender la importancia de la densidad electrónica de los anillos aromáticos en la interacción con oligosacáridos. Por tanto, además de estudiar la estructura 3D del polipéptido en sus estados libre y asociado, se han determinado los parámetros termodinámicos del proceso de interacción y se han comparado con los descritos para otros dominios de AcAMP2.<sup>60</sup>

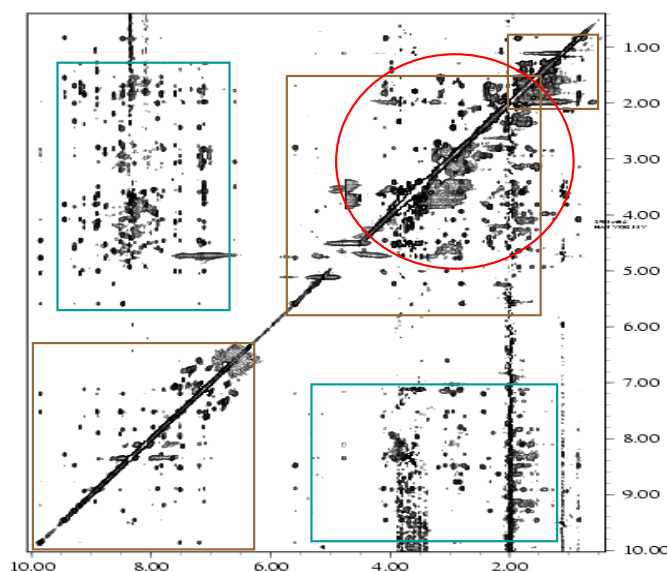
## 2. Resultados y discusión

### 2.1 Determinación de la Estructura Tridimensional del Péptido AcAMP2Pff18/Pff20 Unido a N, N', N''-triacetil-quitotriosa

#### 2.1.1 Asignación de los Desplazamientos Químicos de <sup>1</sup>H RMN del AcAMP2Pff18/Pff20 libre y unido a Quitotriosa

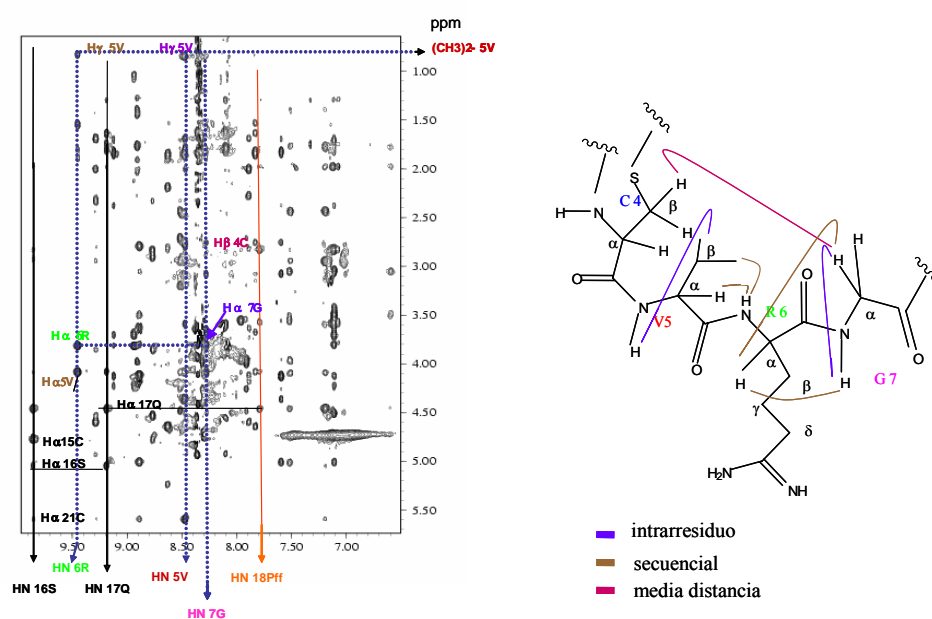
La asignación de los desplazamientos químicos se basó en la combinación de espectros TOCSY y NOESY, de acuerdo a la estrategia y metodología estándar (asignación secuencial, véase el capítulo anterior) establecida por Wüthrich.<sup>62</sup> En nuestro caso, se utilizó el programa XEASY,<sup>92</sup> que permite superponer los distintos espectros, y llevar un registro continuo de las asignaciones. Así, se consiguió la asignación total (Tabla 1.1) del espectro de <sup>1</sup>H RMN del AcAMP2Pff18/Pff20 en ausencia y presencia de quitotriosa.

Como ejemplo, y para mencionar algún detalle del protocolo seguido, se dividió el espectro NOESY en distintas regiones: (Figura 1.2)



**Figura 1.2.** Espectro NOESY (500 MHz, tiempo de mezcla, 300ms, 300 K) de AcAMP2Pff18/Pff20 unido a quitotriosa.  Región de picos de correlación del péptido (-CH<sub>3</sub>, cadena lateral (H $\alpha$ , H $\beta$ , H $\gamma$ , H $\delta$ ), -NH, y Hs aromáticos).  Región de picos de correlación de la quitotriosa.  Péptido: Huella Dactilar

La región de campo alto en donde aparecen principalmente los protones alifáticos y los grupos metilo de Ala, Val, Leu e Ile, así como los de los grupos NHAc del azúcar, en el caso del complejo. La región central, en donde se localizan los protones H $\alpha$  de la cadena principal del polipéptido, algunos H $\beta$  y los que pertenecen a los anillos de piranosa de la quitotriosa. Finalmente, se encuentra la región denominada huella dactilar, donde se localizan los picos de cruce de los NH amídicos, que permiten realizar la asignación secuencial de los protones de la cadena principal del polipéptido: H $\alpha$ , H $\beta$ , H $\gamma$ , H $\delta$  (Figura 1.2). En esta zona también se incluyen los protones aromáticos de los grupos *p*- fluorofenilalanina 18 y 20 y de la Tyr27 y los protones amida de los NHAc del azúcar. Algunos NOEs, correspondientes a interacciones intraresiduales, secuenciales y de media distancia se muestran en la Figura 1.3.



**Figura 1.3** Espectro NOESY mostrando la conectividad de algunos protones amídicos aromáticos. Existen correlaciones intraresiduo, secuencial y media distancia.

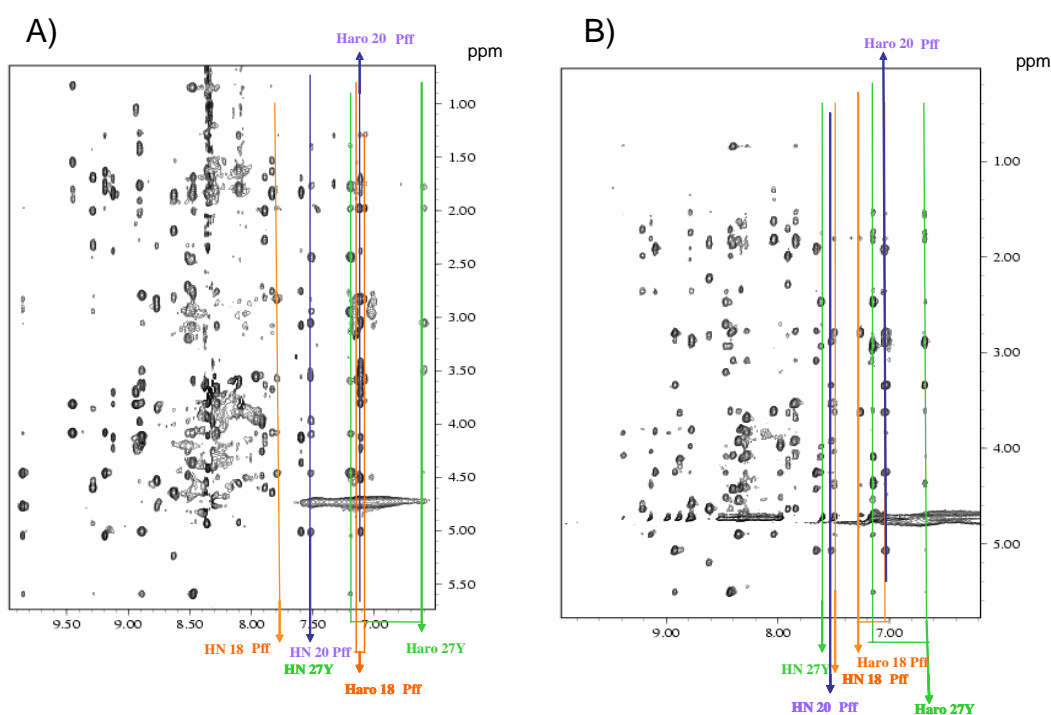
Como ejemplo, considérese las señales aisladas intensas a campo alto que corresponden al grupo metilo  $\gamma$  de la Val5. Si se traza una línea horizontal, se observa una correlación intrarresiduo con el protón del NH (también existente en el TOCSY). Además, este mismo grupo metilo tiene una interacción secuencial con un NH distinto al de su propio aminoácido, que puede identificarse como Arg6. Del TOCSY desde este NH se deduce la posición de su propio H $\alpha$ . En el NOESY, desde el NH de la Arg6, se observa una correlación con otro H $\alpha$ , el de la Gly7. Además, se observa una correlación a media distancia de la interacción con un protón H $\beta$ , que más tarde se identifica como perteneciente a la Cys4. Normalmente, se comienza el proceso de asignación por distintas regiones del espectro. Así, es posible reconocer el NH de la Ser16 como el protón amídico de campo más bajo, típico de dominios de heveína, especialmente cuando la proteína se une a carbohidratos. De la observación de las correlaciones desde este NH, descartando los picos intrarresiduo que se encuentran en el TOCSY, es posible encontrar NOEs interresiduales, que permiten establecer la conectividad con los siguientes residuos, Gln17 y Pff18. En las ampliaciones del NOESY (Figura 1.4), se observan los NOEs de los protones amídicos, y aromáticos de Pff18 y Pff20 y la Tyr27, del AcAMP2Pff18/Pff20, tanto en el complejo con quitotriosa (Fig. 1.4A), como libre (Fig. 1.4B). Por comparación de ambos espectros se pudieron establecer los desplazamientos químicos del péptido en ambos estados (Tabla 1.1)

**Tabla 1.1.** Desplazamientos químicos ( $\delta$ ) de <sup>1</sup>H-RMN del Ac-AMP2Pff18/Pff20 en presencia de quitotriosa en disolución tampón de H<sub>2</sub>O (10% D<sub>2</sub>O) a pH 5.6 y 300 K.

aa	NH	$\alpha$ H	$\beta$ H	$\gamma$ H	Otros
Val1	-	3.813	2.133	1.047, 1.089	
Gly2	8.936	4.223, 3.699			
Glu3	9.121	4.122	1.831,1700	2.380	
Cys4	7.596	4.356	3.076,2.755		
Val5	8.479	4.080	1.859	0.833, 0.853	
Arg6	9.448	3.816	1.807, 1.893	1.547	H $\epsilon$ : 7.145
Gly7	8.280	4.078, 3.670			
Arg8	7.831	4.634	1.841,1.854	1.635,1.533	H $\delta$ : 3.150
Cys9	8.630	5.232	2.816, 2.191		
Pro10	-	4.414	1.679, 2.357	1.968	H $\delta$ : 3.937, 3.416
Ser11	8.354	4.112	3.799		
Gly12	8.888	4.175, 3.633			
Met13	7.891	4.592	2.011, 1.691	2.276	
Cys14	9.284	4.528	2.330, 3.853		
Cys15	8.766	4.768	2.905, 2.833		
Ser16	9.834	5.041	4.455		
Gln17	9.186	4.081	1.755, 1.642	1.811	H $\epsilon$ 21: 6.784, H $\epsilon$ 22: 7.330
Pff18	7.793	4.742	2.830, 3.576		H2,6: 7.119, H3,5: 7.078
Gly19	7.962	3.956, 3.554			
Pff20	7.520	5.004	3.054, 3.498		H2,6(3,5):7.113
Cys21	8.887	5.585	2.794		
Gly22	8.473	3.592, 1.795			
Lys23	8.100	4.926	1.642	1.429, 1.291	
Gly24	8.359	4.539, 3.930			
Pro25	-	4.254	2.287, 1.912	2.017	H $\delta$ : 3.623, 3.828
Lys26	8.909	3.976	1,769, 1.398	1.442, 1.280	H $\delta$ : 1.625
Tyr27	7.512	4.089	2.443, 2.947		H2,6: 7.165, H3,5: 6.602
Cys28	8.516	4.481	3.195, 2.714		
Gly29	7.914	3.974			
Arg30	8.318	4.256	1.816, 1.694	1.599	H $\delta$ : 3.144

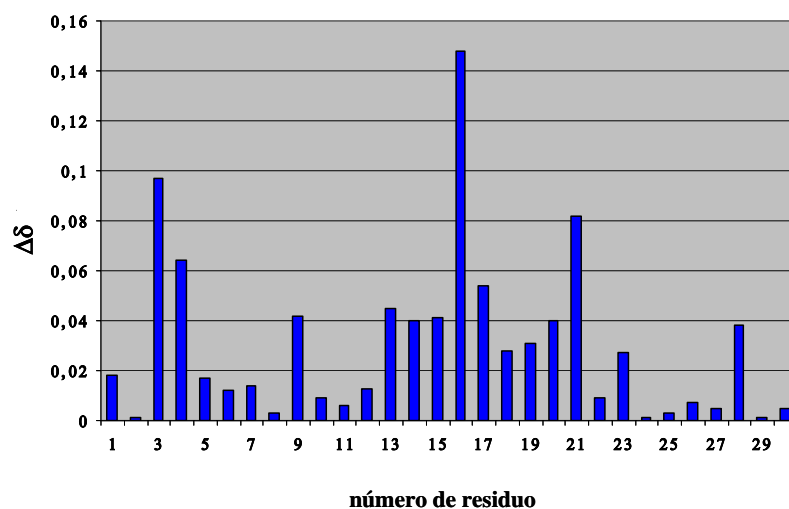
Las perturbaciones de desplazamiento químico observadas en las resonancias del péptido al asociarse al trisacárido permiten establecer, al menos cualitativamente, el

lugar de interacción. Las máximas variaciones de H $\alpha$  (Figura 1.5) tienen lugar en el extremo N-terminal (residuos 3 y 4), entre el residuo 13 y 21, especialmente en 16 y 21, y en el residuo 28. Para los protones aromáticos, la mayor diferencia entre el estado libre y el asociado se observa en los del residuo Pff18.



**Figura 1.4** Región de los protones NHs y aromáticos de los espectros NOESY del AcAMP2Pff18/Pff20 unido a quitotriosa (A) y libre (B).

Cualitativamente, la gran variación de desplazamientos químicos observada en estas regiones permite suponer que el polipéptido modificado AcAMP2Pff18/Pff20 interacciona con la quitotriosa de manera análoga al encontrado para otros dominios de heveína.<sup>41-54</sup> De hecho, en la estructura tridimensional de estos dominios, el extremo N-terminal se halla también en las proximidades del sitio de unión a azúcares.



**Figura 1.5** Representación de las diferencias de desplazamiento químico de los H $\alpha$  ( $\Delta\delta: \delta_{compl} - \delta_{libre}$ ), para el AcAMP2Pff18/Pff20 en estado libre y en su complejo con quitotriosa

### 2.1.2 Generación de las Estructuras 3D del AcAMP2Pff/Pff20 unido a Quitotriosa mediante el Programa DYANA

A partir de anterior análisis del espectro NOESY, se obtuvieron 314 picos NOEs significativos entre los protones del AcAMP2Pff/Pff20, cuando está unido a quitotriosa. Siguiendo el protocolo convencional, y de acuerdo con su intensidad, estos picos de cruce NOEs fueron convertidos en restricciones de distancia. Los picos se clasificaron en 3 grupos, de modo semicuantitativo: picos intensos, de mediana intensidad y débiles a los, que se asignaron valores de 2.5 Å, 3.5 Å y 5.5 Å, respectivamente, como límites superiores de distancia interprotónica.

En el conjunto de picos NOE que se usaron para la determinación de la estructura 3D, no se consideraron las interacciones intrarresiduo a un enlace (ej. HN-H $\alpha$ ), ya que no introducen restricciones geométricas útiles para el cálculo de estructuras. Por lo tanto, quedaron finalmente 248 restricciones, de las que 79 fueron intraresiduo, 99 secuenciales, 24 de rango medio y 47 a larga distancia. Por otro lado, también se incluyeron los puentes disulfuro de forma explícita en el cálculo. Así, se restringieron las distancias S-S del siguiente modo: 2.0 Å < R < 2.1 Å y las C $\beta$ -S como: 3.0 Å < R < 3.1 Å. Las estructuras 3D se generaron usando el programa DYANA<sup>68</sup> (Dynamics



Algorithm for NMR Applications), a partir de las restricciones de NOE antes mencionadas, que se incluyen en el apéndice. El programa DYANA emplea un protocolo de calentamiento-enfriamiento simulado (*simulated annealing* (SA)), combinado con un cálculo en el espacio multidimensional de los ángulos de torsión del péptido. Así, empleando distintas estructuras iniciales, se generaron conjuntos de 400 y 800 estructuras, de las que, en cada caso, se seleccionaron las 50 mejores. El criterio de selección se basó en su función diana, es decir, aquellas estructuras que proporcionan menos violaciones de los datos experimentales.

De la comparación de las 50 estructuras seleccionadas, mediante evaluación de la raíz de la desviación cuadrática media de las coordenadas atómicas, (RMSD), se deduce que la estructura del polipéptido está bien definida tanto en la cadena principal (RMSD < 0.74 Å), como en el conjunto de todos los átomos pesados, incluyendo cadenas laterales (RMSD < 1.5 Å). La orientación de los aminoácidos supuestamente implicados (S16, Pff18, Pff20 y Y27) en el reconocimiento de azúcares también está bien definida (véase más adelante).

A las 248 restricciones utilizadas en el cálculo de DYANA previamente mencionado, se les aplicó el programa GLOMSA<sup>86</sup> (*Global method for obtaining stereospecific assignments*) que permitió establecer la asignación estereoespecífica de los protones proquirales β de los residuos 4, 8, 10, 13, 14, 18, 20, 26, 27 y 30. Posteriormente, se aplicó el módulo (también incluido en DYANA) de modificación de distancias redundantes (*distance modify*), que elimina las restricciones que no tienen valor para el cálculo de geometrías. En este paso, el programa permite un ajuste inicial de los límites de distancia a las intensidades de los NOEs y, además, introduce pseudoátomos cuando no hay asignación estereoespecífica disponible para los picos de correlación NOE correspondientes a protones diastereotópicos. A continuación, se realiza un ajuste final de los valores de restricciones de distancia, usando el programa CALIBA<sup>86</sup> (*Calibration of NOE intensity versus distance constraints*), basándose en las intensidades integradas de los picos de cruce NOESY, realizadas con XEASY.

Así, se obtuvo un conjunto final de 208 restricciones de distancias, que se introdujeron de nuevo en el programa DYANA. Una vez terminado el cálculo, se seleccionaron las 35 estructuras que mejor cumplían las restricciones en términos de la función diana,

obteniendo el conjunto mostrado en la Figura 1.6, que presenta un rmsd de 0.60 Å para los átomos de la cadena principal y de 1.40 Å para todos los átomos.



**Figura 1.6** Superposición de las 35 mejores estructuras DYANA entre los residuos 4..28. RMSD de la cadena principal, 0.60 +/- 0.13- RMSD de los átomos pesados, 1.40 +/- 0.19Å

#### La Estructura Secundaria del AcAMP2Pff18/Pff20.

Una característica típica de los dominios de heveína es la de poseer dos segmentos que forman una horquilla  $\beta$  antiparalela, entre los residuos Met13 y Lys23.<sup>54</sup> El AcAMP2Pff187Pff20, no natural, no es una excepción y presenta el mismo motivo estructural, como se deduce de los NOEs característicos de una horquilla  $\beta$  observados en el espectro NOESY. Los parámetros de RMN característicos de este elemento estructural son:

- NOEs secuenciales:  $\alpha$  N(i, i+1) intensos
- NOEs a larga distancia:  $\alpha$   $\alpha$  (i,j) débiles o de media intensidad

Así, se observan los picos de cruce  $\alpha$ Hi-NHi+1 (Cys15-Ser16, Cys21-Gly22), NHi-NHj (Cys14-Gly22, Ser16-Tyr20) y NHi- $\alpha$ Hj(Cys14-Lys23, Ser16-Cys21). Además, la característica más típica de horquilla  $\beta$  antiparalela se manifiesta principalmente por la presencia de una interacción NOE  $\alpha$ Hi- $\alpha$ Hj (Cys15-Cys21). De hecho, la clasificación

de este segmento como lámina  $\beta$ , de acuerdo con el protocolo de Kabsch-Sander, se corroboró mediante el uso del Programa Insight II.

### 2.1.3 Optimización de la Estructura 3D del AcAMP2Pff18/Pff20 mediante Minimización Restringida utilizando Simulaciones de Dinámica Molecular.

#### 2.1.3.1 Parametrización del Aminoácido no Natural *p*-Fluorofenilalanina (Pff).

Los programas computacionales que se emplean para determinar y refinar la estructura tridimensional de las proteínas incluyen solamente en sus librerías de parámetros estructurales y geométricos de los 20 aminoácidos proteínogénicos estándar. Por lo tanto, para el caso específico del péptido sintético AcAMP2Pff18/Pff20, que contiene dos residuos de un aminoácido no natural, la *p*-fluorofenilalanina (Pff), fue necesario crear e incluir los parámetros para este residuo.

El protocolo que se siguió para crear este aminoácido es el siguiente:

- a) Creación del aminoácido mediante el uso del programa Xleap (módulo de AMBER 7.0 ),<sup>75</sup> generando el correspondiente archivo de coordenadas cartesianas en formato PDB (**P**rotein **D**ata **B**ank).
- b) Cálculo de las cargas deslocalizadas mediante el uso del programa Gaussian98 con un nivel de teoría básico HF/STO-3G, pop=mk iop (6/3302).<sup>101</sup>
- c) A la salida del cálculo anterior se aplicó el módulo RESP (potencial electrostático restringido) del paquete de programas AMBER 7.0, para generar los valores de carga puntuales para cada átomo del residuo de Pff. Posteriormente, se realizó el promedio de las cargas para las posiciones equivalentes sobre la base de las ya establecidas para el resto de los aminoácidos. De esta manera, se obtuvo la carga del átomo de flúor. Siguiendo este protocolo, se crearon los archivos apropiados con los datos correspondientes para este nuevo aminoácido y se introdujeron en las librerías de parámetros de los programas clave para el cálculo de estructuras, DYANA y AMBER.

### 2.1.3.2 Generación de Estructuras Refinadas del AcAMP2Pff18/Pff20 mediante uso del Programa AMBER.

Las estructuras generadas por DYANA (véase anteriormente) fueron refinadas mediante un protocolo de optimización energética usando el campo de fuerzas AMBER versión 5.0.<sup>76</sup> El proceso consistió en:

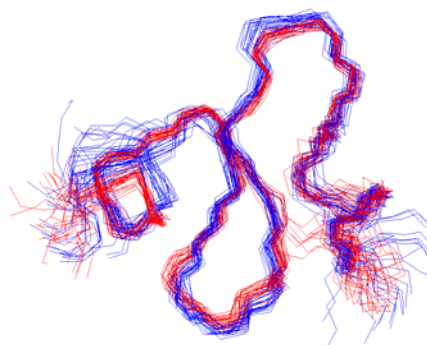
- 1.-Minimización energética restringida (REM), utilizando las restricciones de distancia obtenidas de los NOEs,
- 2.-Simulación de dinámica molecular restringida (RMD), mediante un protocolo de calentamiento y enfriamiento simulado (SA),
- 3.-De nuevo, minimización energética restringida (REM), a baja temperatura.

De todas las estructuras obtenidas mediante este protocolo, se eligieron las 25 mejores, en términos energéticos, que están representadas en la Figura 1.7. El rmsd de los átomos de la cadena principal, entre los residuos 4-28, fue de 0.84 Å y el de todos los átomos, de 1.96 Å.

La comparación de estas 25 estructuras obtenidas por AMBER para el AcMP2Pff18/Pff20 (Figura 1.8) con las 26 estructuras del AcAMP2 natural,<sup>54</sup> [código PDB 1MMC], permitió observar que tanto el plegamiento global como la orientación de las cadenas laterales claves para el reconocimiento molecular de quitooligosacáridos es muy similar en ambos polipéptidos. Por tanto, puede concluirse que la conformación global del péptido modificado AcAMP2Pff18/Pff20, unido a quitotriosa, no se afecta por la mutación de los dos residuos aromáticos involucrados en el sitio de unión a carbohidratos.

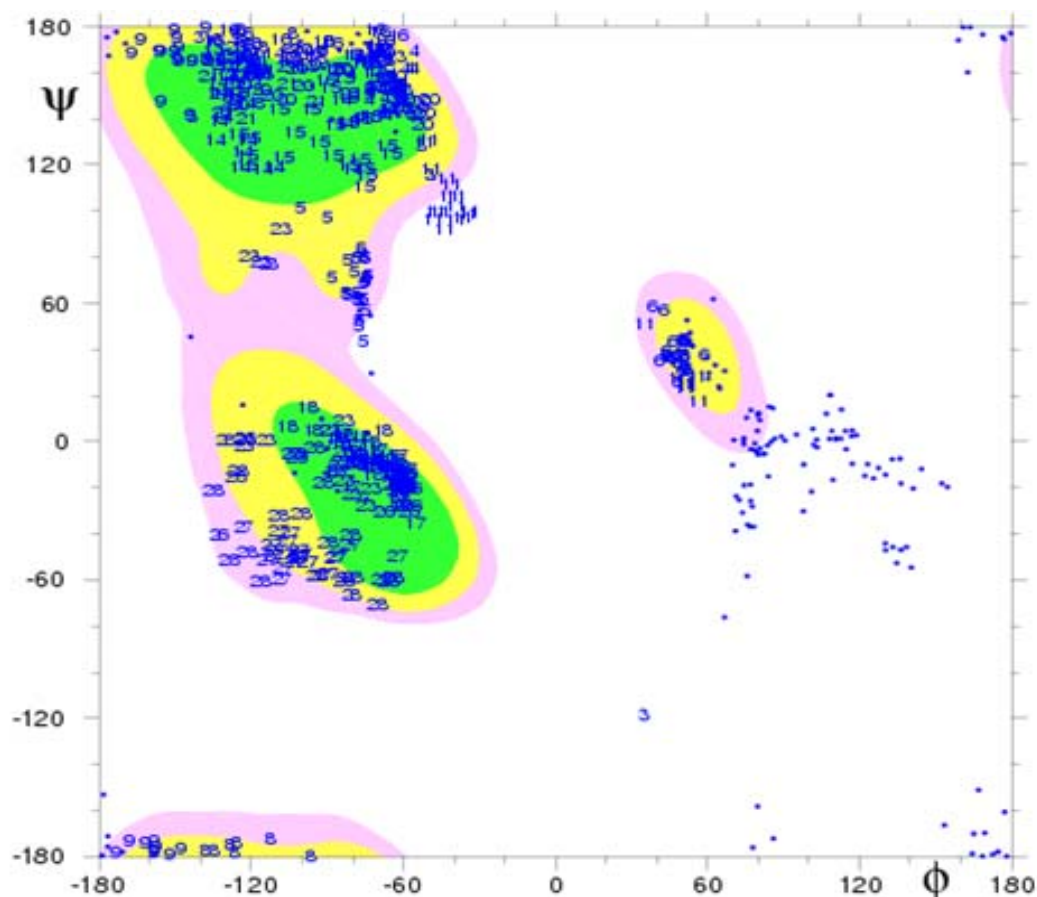


**Figura 1.7** Superposición de la cadena principal (residuos 4-28) de las 25 mejores estructuras AMBER.  
rmsd: cadena principal 0.84 +/- 0.18 Å  
átomos pesados 1.96 +/- 0.37 Å



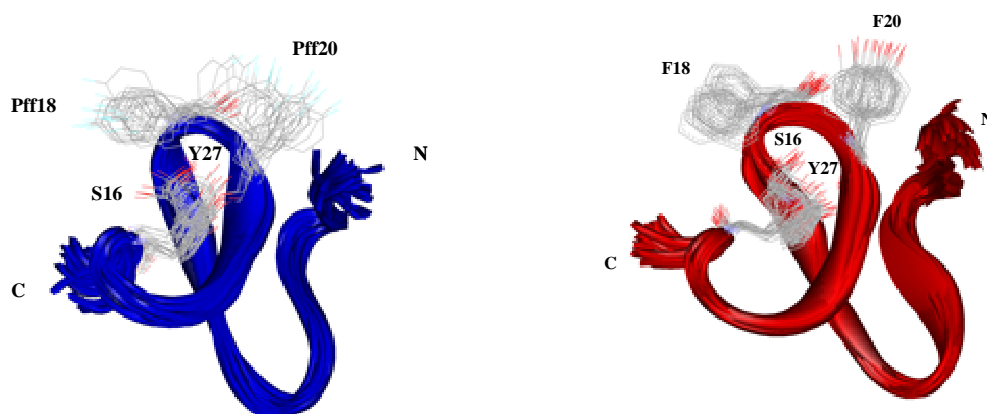
**Figura 1.8** Superposición de la cadena principal (residuos 4-28) de las 25 mejores estructuras de AcAMP2Pff18/Pff20 unido a quitotriosa (azul) con las 26 estructuras reportadas para el AcAMP2 nativo (rojo).

Un modo usual de estimar la bondad de las estructuras obtenidas es el de representar los ángulos  $\phi$  y  $\psi$  del péptido en el mapa de Ramachandran, que denota las áreas  $\phi$ ,  $\psi$  permitidas energéticamente para los distintos elementos estructurales que forman una cadena polipeptídica. El mapa de Ramachandran del AcAMP2Pff18Pff20 se representa en la Figura 1.9. En él, puede observarse que los valores de  $\phi$ ,  $\psi$  para la inmensa mayoría de los residuos se encuentran en áreas permitidas. Los puntos representan los residuos de glicina que, debido a que presentan menor impedimento estérico por la ausencia de sustituyente en  $\beta$ , tienen mayor libertad conformacional, y por tanto, un margen más amplio de conformaciones permitidas.



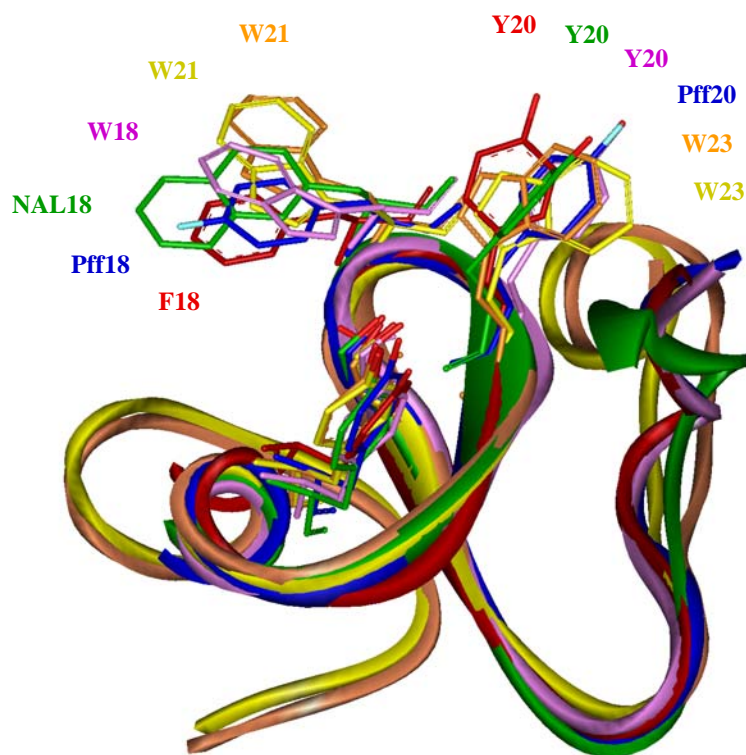
**Figura 1.9** Gráfica Ramachandran de las 25 estructuras del AcAMP2Pff18/Pff20 unido a quitotriosa, calculadas mediante el uso de de AMBER.

Respecto a la orientación de los anillos aromáticos de los residuos que participan en la interacción proteína-carbohidrato, el análisis comparativo de las estructuras del AcAMP2Pff18/Pff20 unido a quitotriosa, con las del AcAMP2 natural, en estado libre, permite deducir la existencia de un ligero cambio en la orientación de los anillos de los residuos Pff18 y Pff20, probablemente debido al apilamiento CH/ $\pi$  con el trisacárido en la forma asociada (Figura 1.10). Este cambio de orientación se ha observado también en una forma truncada de heveína, al pasar del estado libre al asociado.<sup>49</sup>



**Figura 1.10** Representación de la superposición de las 25 estructuras de AcAMP2Pff18/Pff20 asociado a quitotriosa, derivadas a partir de la optimización mediante RMD con AMBER (azul) y de la superposición de las 26 estructuras del AcAMP2 natural (rojo). Se muestra la cadena principal y la orientación de Ser16, Pff18, Pff20, Tyr27, que participan en la interacción con el azúcar.

Asimismo, es interesante mencionar que la topología molecular de la cadena principal del polipéptido modificado aquí estudiado, unido a quitotriosa, es muy similar a las estructuras descritas para el AcAMP2 natural,<sup>54</sup> y a otros dos polipéptidos mutantes de AcAMP2 previamente estudiados en nuestro grupo de investigación, Phe18Trp y Phe18Nal, en los que la Phe18 se ha sustituido por Trp o por naftilalanina, respectivamente.<sup>60</sup> También es muy similar a la estructura de la heveína asociada a quitooligosacáridos<sup>46,47</sup> (Figura 1.11). De hecho, la orientación de los anillos aromáticos de los residuos involucrados en el sitio de unión a quitooligosacáridos no manifiesta cambios significativos, principalmente en lo que se refiere a los anillos de Pff en AcAMP2Pff18/Pff20, con respecto a los Trp de la heveína, e incluso a otros dominios de tipo heveína, como la pseudoheveína<sup>51</sup> y el dominio B de la WGA.<sup>42</sup>



**Figura 1.11** Superposición de las cadenas principales del AcAMP2Pff18/Pff20 unido a quitotriosa (azul), AcAMP2 natural (rojo), AcAMP2W18 (rosa), AcAMP2NAL18 (verde), heveína de RMN (amarillo) y Heveína de Rayos X (naranja). También se muestran los anillos aromáticos implicados en el sitio de unión del carbohidrato.

## 2.2 La Estructura Tridimensional del Complejo AcAMP2Pff18/Pff20- Quitotriosa.

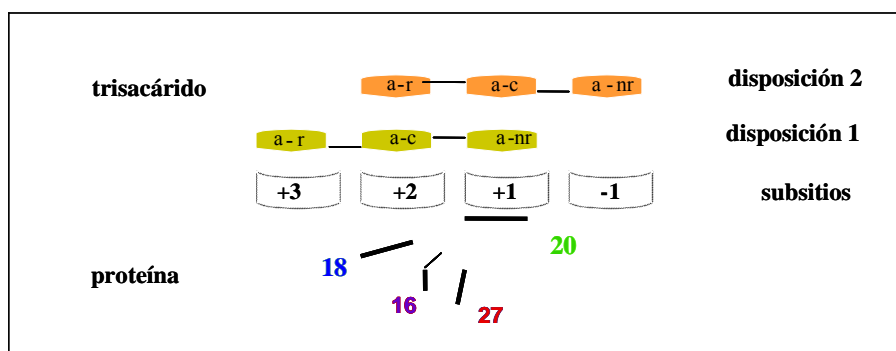
### 2.2.1 Análisis del Sitio de Unión del Trisacárido

El sitio de reconocimiento molecular de carbohidratos en diversos dominios de heveína ha sido estudiado mediante RMN.<sup>41-54</sup> En el caso del AcAMP2Pff18/Pff20, tanto por homología como, experimentalmente, sobre la base sólida de los datos de perturbación de desplazamiento químico y de NOEs intermoleculares proteína-azúcar (véase más adelante), este corresponde a los aminoácidos Ser16, Pff18, Pff20 y Tyr27. En este sentido, es de resaltar la gran diferencia de  $\delta$  observada para el protón  $\alpha$  de la Ser16, así como los de los H $\alpha$  de 21 y 28, inmediatamente consecutivos a los residuos aromáticos clave al pasar del estado libre al asociado. Además, es posible pensar que un cambio en la orientación del anillo del aminoácido Pff20 sea responsable del  $\Delta\delta$  del H $\alpha$  de los



residuos Gly2 y Glu3, también observados (Figura 1.5), debido a la cercanía del extremo N-terminal al residuo 20 en la estructura tridimensional.

Los estudios de la interacción de dominios de heveína con quitooligosacáridos realizados en nuestro grupo de investigación han permitido demostrar que el sitio de unión a azúcares de estos polipéptidos es lo suficientemente extendido como para interactuar con, al menos, 3 unidades de N-acetilglucosamina (GlcNAc). Por lo tanto, para un trisacárido, es posible considerar dos posibles disposiciones en el sitio de unión, tal y como se muestra en la figura 1.12.



**Figura 1.12** Representación de las dos posibilidades de unión proteína-trisacárido. En ellas, **a-r** representa el azúcar reductor, **a-c** es el azúcar central, y **a-nr** es el azúcar no reductor.

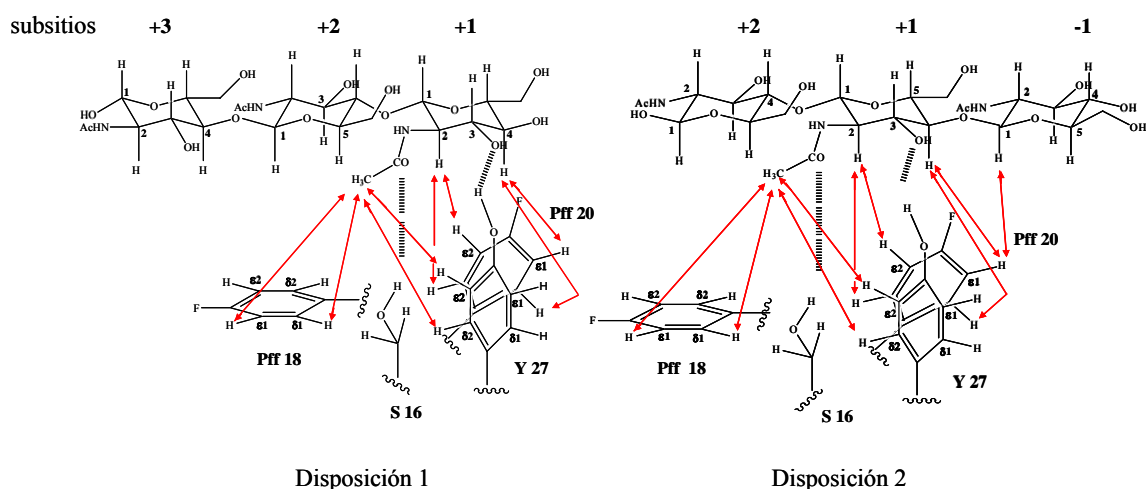
De hecho, las evidencias experimentales derivadas del NOESY permiten proponer la existencia de ambas disposiciones **1** (+3,+2,+1) y **2** (+2,+1,-1) para la unión de la quitotriosa al AcAMP2Pff18/Pff20 (tabla 1.2). Por un lado, existe un pico de cruce entre el metilo del grupo acetamida que pertenece al azúcar no reductor del trisacárido en el subsitio (+1), con los protones aromáticos de los residuos Pff18 y Tyr27. Esta interacción es exclusiva de la disposición **1**. Además, se observa un pico de cruce entre los protones H-2 y H-4 del mismo azúcar no reductor (+1) con los protones aromáticos del residuo Pff20. Por otra parte, también se observan picos de cruce entre los mismos residuos de la proteína, pero ahora con el azúcar central (subsitio +1), indicando la existencia de la disposición **2**. Además, existe un pico de cruce NOE entre el protón

anomérico del azúcar no-reductor con los protones del aminoácido Pff20, que sólo es posible si dicho extremo no reductor se sitúa en el subsitio (-1). Este pico es inambiguo para la disposición 2. Una representación esquemática de estas interacciones se muestra en la Figura 1.13.

**Tabla 1.2** Interacciones NOE intermoleculares AcAMP2Pff18/Pff20-quitosa.

$H_{AcAMP2Pff18/Pff20}$	$\delta H_{AcAMP2Pff18/Pff20}$	$H_{carboh}$	$\delta H_{carboh}$	Disposición 1 subsitios +3,+2,+1	Disposición 2 subsitios +2,+1,-1
18H <sub>δ1,2</sub>	7.12	CH <sub>3</sub>	2.03	+1	+1
18H <sub>ε1,2</sub>	7.08	CH <sub>3</sub>	2.03	+1	+1
20H <sub>δ1,2</sub> , 20H <sub>ε1,2</sub>	7.11	H1	4.52	-	-1
20H <sub>δ1,2</sub> , 20H <sub>ε1,2</sub>	7.11	H2	(3.69, 3.80) <sup>a</sup>	+1	+1
20H <sub>δ1,2</sub> , 20H <sub>ε1,2</sub>	7.11	H4	(3.48, 3.69) <sup>a</sup>	+1	+1
27H <sub>δ1,2</sub>	7.19	CH <sub>3</sub>	2.03	+1	+1
27H <sub>ε1,2</sub>	6.60	CH <sub>3</sub>	2.03	+1	+1
27H <sub>H</sub>	7.46	CH <sub>3</sub>	2.03	+1	+1

<sup>a</sup>Desplazamientos químicos para ambas disposiciones, respectivamente.

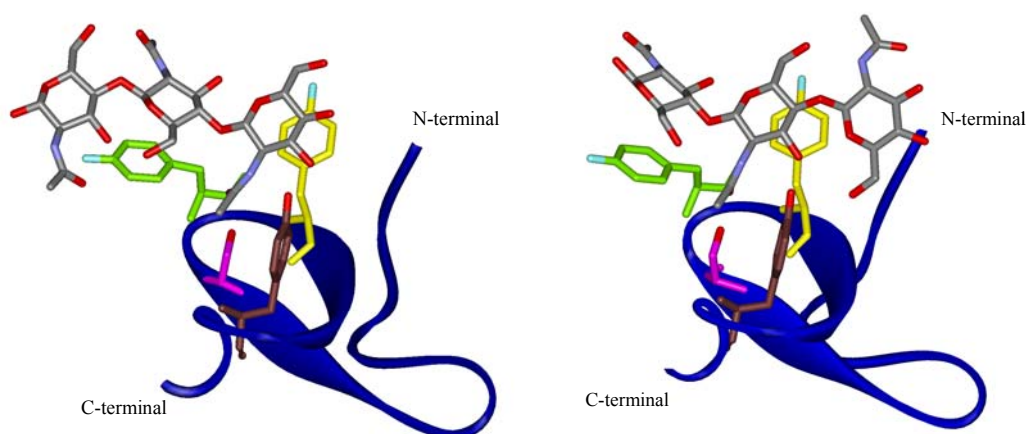


**Figura 1.13** Representación esquemática de los NOEs intermoleculares que evidencian la existencia simultánea de las dos disposiciones de unión del trisacárido al AcAMP2Pff18/Pff20.

### 2.2.2 Estructura del Complejo. Minimización Energética Mediante el Método de Enfriamiento Simulado (SA)

La obtención de la estructura 3D del complejo se realizó mediante el uso de cálculos de dinámica molecular, usando el programa AMBER 5.0. En este caso, además de las restricciones intra-proteína utilizadas anteriormente, también se incluyeron los NOEs intermoleculares entre la proteína y el azúcar, considerando las dos disposiciones descritas en el apartado anterior. Como punto de partida para los cálculos de enfriamiento simulado, se generaron 30 estructuras mediante DYANA, que estaban de acuerdo con todos los datos experimentales. A continuación, se modelaron los dos posibles complejos, utilizando el programa InsightII, adecuando las estructuras a los NOEs intermoleculares y tomando como base el complejo entre una heveína trunca de 32 aminoácidos y la quitotriosa, previamente estudiado en nuestro grupo.<sup>42, 49</sup>

Además de los NOEs intermoleculares, también se incluyeron dos enlaces de hidrógeno entre el trisacárido y los residuos de Ser16 y Tyr27. Estos enlaces de hidrógeno se han descrito en todos los complejos de dominios de heveína descritos hasta el momento. Como resultado de estos cálculos, se obtuvieron las estructuras de los dos posibles complejos mostrados en la Figura 1.14.



**Figura 1.14** Estructura de los dos posibles complejos de AcAMP2Pff18/Pff20 unido a quitotriosa. Las estructuras mostradas se obtuvieron mediante por simulaciones de dinámica molecular, usando restricciones intermoleculares proteína- carbohidrato obtenidas mediante RMN.

### 2.2.3 Hacia una Visión Tridimensional de la Estructura de AcAMP2Pff18/Pff20 y de su Complejo con Quitotriosa. Dinámica Molecular en Disolvente Explícito (sMD).

Las estructuras obtenidas hasta el momento se ha conseguido utilizando restricciones experimentales de RMN (rMD). Es decir, estas estructuras están influidas en cierta medida por los datos de NOE. Para evaluar la influencia de estos datos en las propiedades conformacionales del AcAMP2Pff18/Pff20, se realizaron cálculos de dinámica molecular en disolvente explícito para el polipéptido libre (sMDlibre) y asociado a quitotriosa (sMDcomplejo), eliminando las restricciones de NOE. La bondad de los últimos campos de fuerza desarrollados últimamente permite estimar adecuadamente las propiedades conformacionales y dinámicas de biomoléculas en disolución, así como la influencia de ligandos en las mismas. En primer lugar, se optimizó la geometría del AcAMP2Pff18/Pff20, mediante una simulación de dinámica molecular en una caja incluyendo moléculas de agua explícitas, también usando el campo de fuerzas AMBER versión 5.0.

Como estructura de partida, se eligió una de las estructuras del proceso anterior de minimización restringida, obtenida usando los datos del NOESY como restricciones espaciales. En concreto, la estructura elegida fue aquella que presentaba un menor valor de la energía de penalización, en definitiva, la que mostraba menos violaciones de las restricciones experimentales.

La preparación de la caja de simulación, conteniendo el AcAMP2Pff18/Pff20 y las moléculas de agua se realizó mediante la aplicación del módulo Xleap de AMBER. En este, el comando WATBOX216 8.0 permitió asignar las condiciones de disolvente explícito; esto es, relacionar el tamaño del soluto con el volumen óptimo de la caja. Los límites se extendieron 8 Å más allá del tamaño del polipéptido, proporcionando unas dimensiones espaciales de A x B x C (A=42.557, B=34.187, C=41.908 Å). Se añadieron 1836 moléculas de agua para proporcionar la densidad adecuada. Por otro lado, se incluyeron iones para conseguir que la caja fuera neutra. Dado que el polipéptido está cargado positivamente a pH neutro, en este caso se añadieron cinco iones cloruro. Finalmente, se crearon los archivos de coordenadas y de topología.

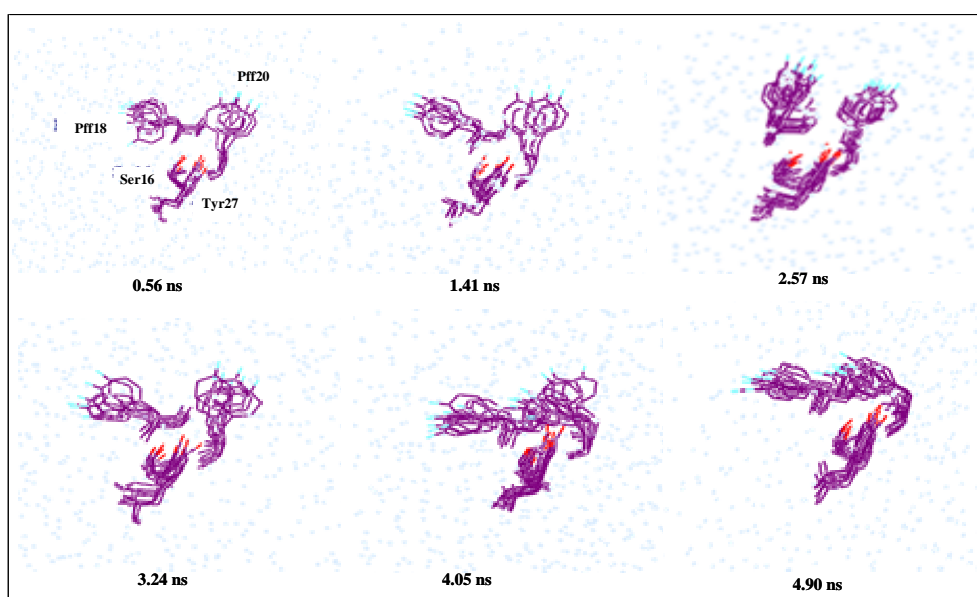
La simulación por dinámica molecular se llevó a cabo mediante el módulo Sander de AMBER. Este proceso requirió un primer paso de minimización corta, utilizando las

restricciones de distancia experimentales. El objetivo de esta minimización es la eliminación de los malos contactos de las moléculas de agua adicionadas entre sí y con el soluto. A continuación, el sistema se equilibró, manteniendo las restricciones, durante 12.5 ps a una temperatura de 300 K y presión de 1 atm, aplicando un corte de 9 Å para el cálculo de las interacciones de no enlace. Así, se consiguió equilibrar la caja con el soluto, las moléculas de agua y los contraiones. Posteriormente, el sistema se trató mediante el método PME (Particle Mesh Ewald) que permite considerar cuantitativamente las interacciones electrostáticas de largo y corto rango. Se mantuvieron las mismas condiciones de temperatura y presión, y el valor de 9 Å para el cálculo de las interacciones de no enlace. Durante los primeros pasos de simulación, las restricciones se fueron eliminando de manera gradual, hasta comenzar el período de simulación, propiamente dicho, sin restricciones. Durante toda la simulación, la temperatura y la presión se mantuvieron constantes, usando el algoritmo propuesto por Berendsen.<sup>102</sup> El paso de integración de las ecuaciones de Newton fue de 2 fs. La trayectoria de la dinámica se guardó en el ordenador usando un muestreo de 0.5 ps.

Con objeto de valorar los parámetros geométricos relacionados con las propiedades conformacionales del AcAMP2Pff18/Pff20, los datos de las trayectorias se tradujeron a formato PDB, mediante el módulo *carnd* de AMBER.

Se prestó atención especial a las variaciones del plegamiento del polipéptido durante el tiempo de simulación, así como a las variaciones de orientación mostradas por los anillos aromáticos de los residuos involucrados en el sitio de unión a carbohidratos.

En la Figura 1.15 se muestran distintos aspectos del proceso de simulación del AcAMP2Pff/Pff20. Es importante señalar la variación de la orientación de los anillos aromáticos de las fluorofenilalaninas 18 y 20 durante toda la simulación, sMDlibre.



**Figura 1.15** Representación de la posición de los residuos involucrados en el reconocimiento molecular de quitooligosacáridos en distintos momentos de la trayectoria de dinámica molecular del AcAMP2Pff18/Pff20, usando disolvente explícito. En cada panel se representan 6, 6, 12, 7, 9 y 10 estructuras, tomadas cada 0.5 ps en torno al tiempo de simulación mostrado abajo.

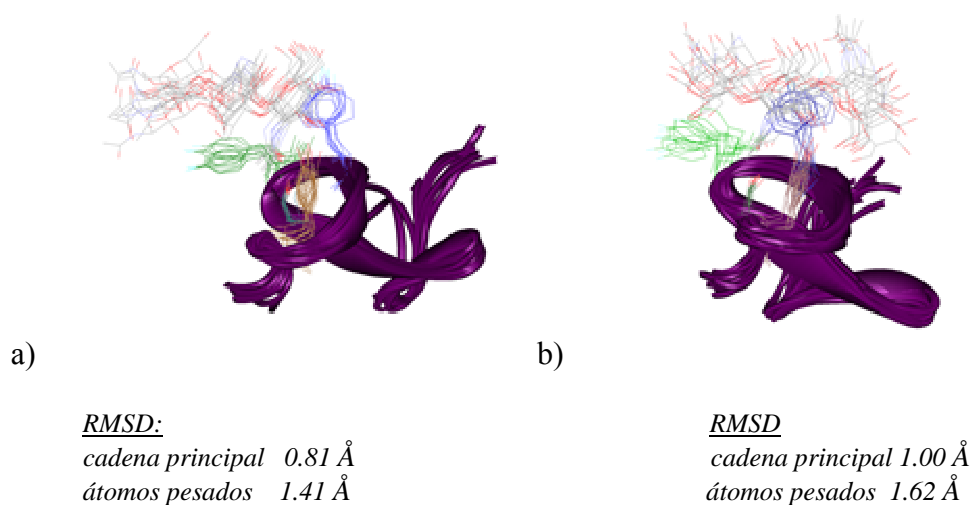
La comparación de un grupo 18 de estructuras de la trayectoria de dinámica molecular en disolvente explícito (sMDlibre) a con aquella obtenida mediante dinámica molecular con restricciones de RMN y con la estructura del AcAMP2 nativo,<sup>54</sup> permite concluir que la mayor variación entre las mismas se da en la orientación de los anillos aromáticos de los aminoácidos 18 y 20 (Figura 1.16). Hay que señalar que la estructura del AcAMP2 nativo se ha obtenido mediante rMD, usando datos de RMN y que nuestra estructura de RMN se ha obtenido en presencia de quitotriosa; es decir, con los anillos de Pff18 y 20 interaccionando con los anillos de GlcNAc del ligando. Es decir, las variaciones entre las estructuras son mínimas y se circunscriben básicamente a la orientación de los residuos Pff18 y Pff20.



**Figura 1.16** Superposición de 18 estructuras de la trayectoria de dinámica molecular de AcAMP2Pff18/Pff20 con disolvente explícito (sMDlibre) (morado), con la estructura obtenida mediante RMN y dinámica molecular restringida rMD, (azul), y la estructura del AcAMP2 nativo<sup>54</sup> (rojo). El RMSD de la cadena principal de AcAMP2Pff18/Pff20 en referencia a la estructura del AcAMP2 nativo es de 0.887 Å.

Una vez estudiada la dinámica del AcAMP2Pff18/Pff20 aislado, se simuló el comportamiento de su complejo con quitotriosa. En concreto, se simularon dos complejos diferentes, empleando disolvente explícito y usando los dos modos de unión del trisacárido al AcAMP2Pff18/Pff20 mencionados en el capítulo anterior. El protocolo que se siguió para llevar a cabo esta dinámica solvatada (sMDcomplejo) fue idéntico al descrito anteriormente para al AcAMP2Pff18/Pff20, usando las geometrías de la dinámica molecular restringida mostradas en el capítulo anterior que mostraron menos violaciones de los datos experimentales.

Los datos de la dinámica indican la interacción del oligosacárido con los residuos clave del polipéptido, que son Ser16, Pff18, Pff20 y Tyr27. Estas interacciones permiten la formación del complejo y su estabilización mediante enlaces de hidrógeno interacciones CH/ $\pi$  y fuerzas de van der Waals, tal y como se muestra la figura 1.17.

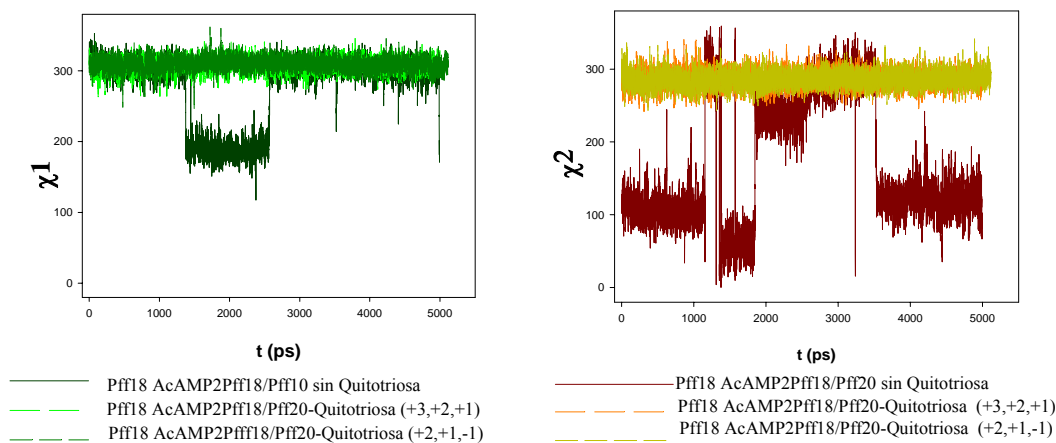


**Figura 1.17** Superposición de 10 estructuras del complejo AcAMP2Pff18/Pff20-quitotriosa obtenidas de la simulación mediante dinámica molecular en disolvente explícito. a) Estructuras explícito. a) Estructuras obtenidas a un tiempo de simulación entre 3.2 y 4.2 ns. El modo de unión es +3,+2,+1. b) Estructuras obtenidas a un tiempo de simulación entre 1.2 y 2.4 ns. El modo de unión es +2,+1,-1.

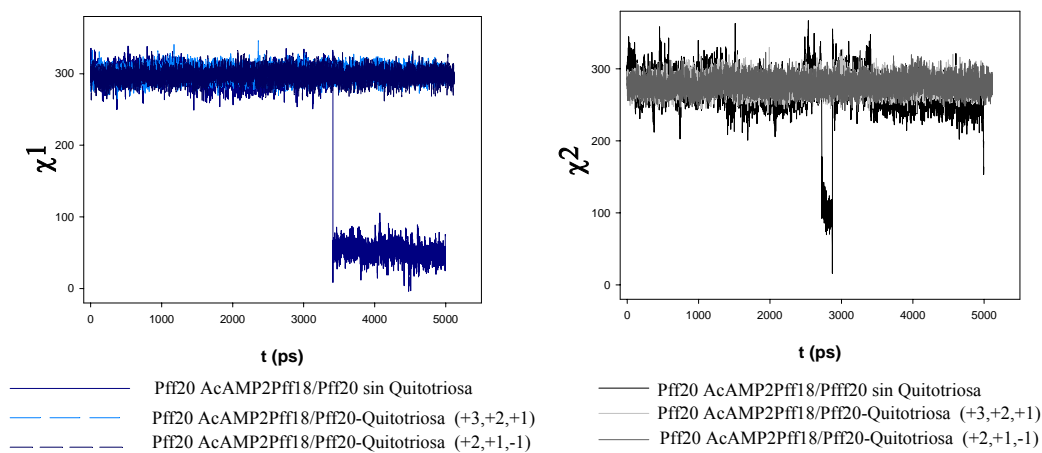
La representación detallada de la trayectoria en función del tiempo (cada 0.5 ps) de la simulación por dinámica molecular en disolvente explícito del AcAMP2F2, en ausencia y presencia de quitotriosa, se muestra en las figuras 1.18 y 1.19, con atención específica a los dos anillos aromáticos, Pff18 y Pff20, que participan en la unión con el trisacárido. La orientación relativa de estos anillos está definida por las torsiones  $\chi_1$ , entre el C $\alpha$  y el C $\beta$ , y  $\chi_2$ , (C $\beta$ -C $\gamma$ ). En las Figuras 1.18 y 1.19 puede observarse la variabilidad de poblaciones dependientes de los valores de estos ángulos de torsión, así como a la influencia del oligosacárido, cuya presencia restringe notablemente la libertad conformacional de estos anillos aromáticos.

Por tanto, se corrobora que la presencia del trisacárido proporciona interacciones estabilizantes con el polipéptido, de apilamiento CH/ $\pi$ , enlaces de hidrógeno y fuerzas de van der Waals, que contribuyen a la estabilización del complejo y que, en este caso, se manifiestan en la orientación homogénea de los anillos aromáticos de Pff 18 y Pff 20.





**Figura 1.18** Representación de la variación de los valores de los ángulos de torsión  $\chi_1$  y  $\chi_2$ , con el tiempo, del anillo aromático de la cadena lateral Pff18 del AcAMP2Pff18/Pff20 en ausencia y presencia de quitotriosa.



**Figura 1.19** Representación de la variación de los valores de los ángulos de torsión  $\chi_1$  y  $\chi_2$ , con el tiempo, del anillo aromático de la cadena lateral Pff20 del AcAMP2Pff18/Pff20, en ausencia y presencia de quitotriosa.

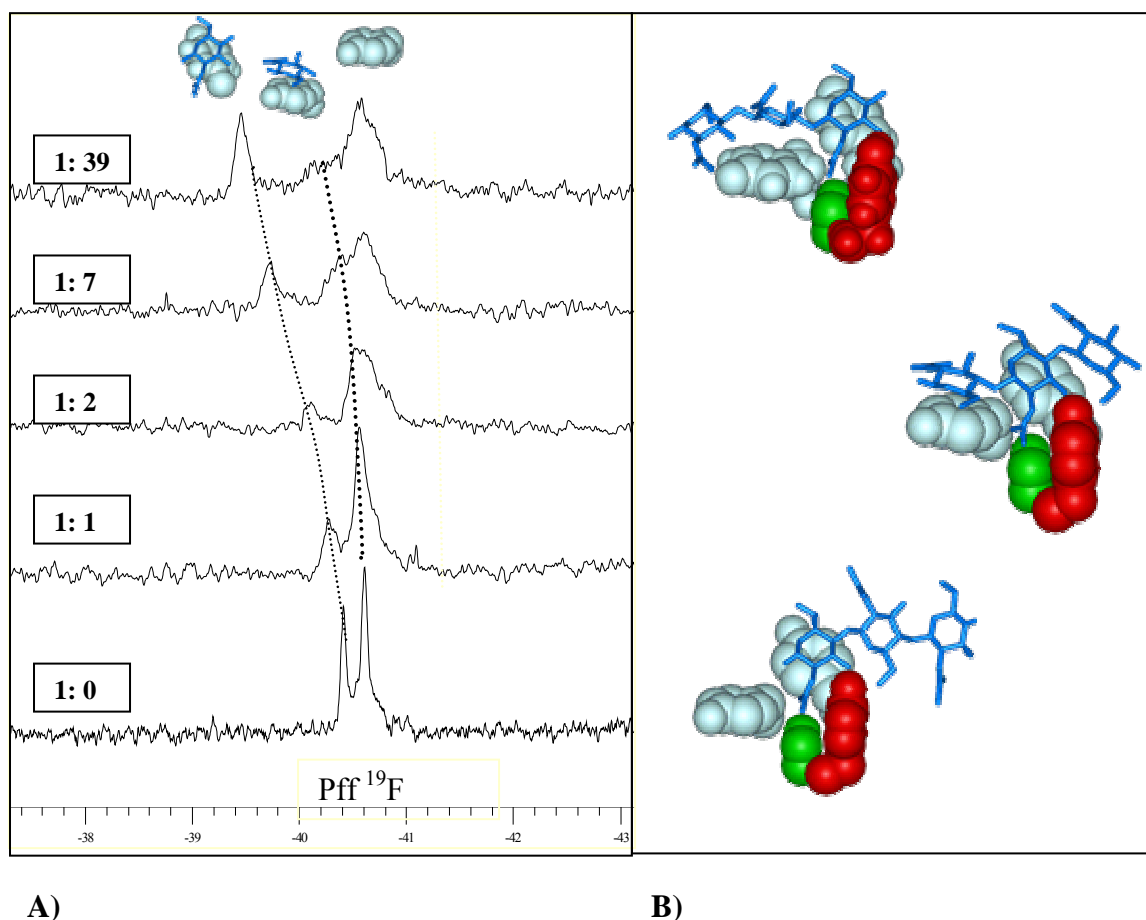
### 2.3 La asociación de AcAMP2Pff18/Pff20 a Quitotriosa. Determinación de la $K_a$ del Proceso y Análisis Comparativo con Valores de $K_a$ y Parámetros Termodinámicos para otros Péptidos Análogos.

La constante de asociación del AcAMP2Pff18/Pff20 a quitotriosa se determinó mediante experimentos de valoración, utilizando RMN de  $^1\text{H}$  y de  $^{19}\text{F}$ , ya que el péptido únicamente presenta dos señales de resonancia de este núcleo, una para el residuo 18 y otra para el 20.

Las valoraciones se realizaron utilizando la metodología estándar de RMN.<sup>29, 46, 96, 97</sup> Para una descripción del protocolo, véase la Parte Experimental.

En el experimento se mantuvo constante la concentración de proteína y se adicionaron concentraciones crecientes de trisacárido. Para cada concentración se llevó a cabo una medición, en la que se siguió la variación del desplazamiento químico de las resonancias de  $^{19}\text{F}$  de la proteína.

En la Figura 1.20 A) se representan los espectros realizados a 298 K en distintos puntos de la valoración. Puede observarse que una de las señales de  $^{19}\text{F}$  se desplaza claramente a campo bajo al aumentar la concentración del carbohidrato. En cambio, la otra se ensancha apreciablemente desdoblándose claramente para una relación molar 1:7, manteniéndose desdoblada para relaciones molares superiores. Esta observación experimental está de acuerdo con la existencia de distintos modos de asociación del oligosacárido al polipéptido, determinada mediante experimentos NOESY, y que se esquematizan nuevamente en la Figura 1.20 B).



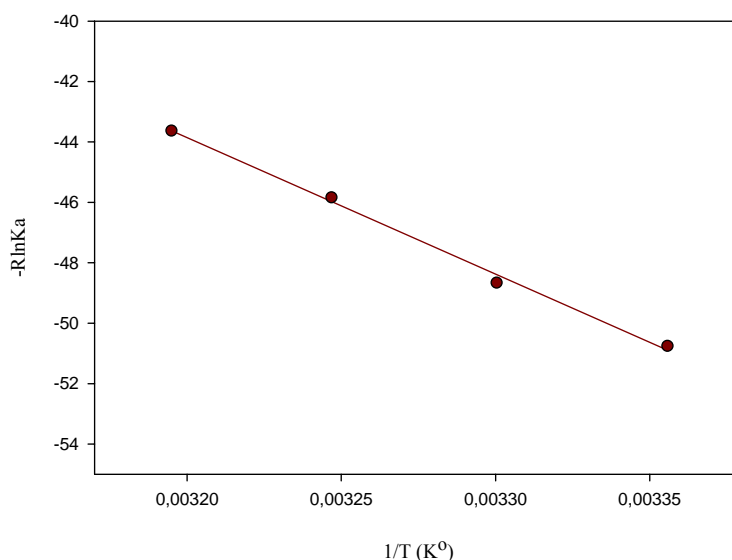
**Figura 1.20** A) Diferentes espectros de <sup>19</sup>F-RMN (298 K, pH 5.6) tomados en cinco puntos de la valoración con (GlcNAc)<sub>3</sub> de una disolución de Ac-AMP2Pff18/Pff20. Puede observarse el desplazamiento sistemático de una de las señales de <sup>19</sup>F en función de la concentración de trisacárido en el tubo de RMN. B) En la imagen derecha se muestran las posibles disposiciones de unión del trisacárido con el péptido sintético AcAMP2Pff18/Pff20.

En la Tabla 1.3 se muestra la comparación de los valores obtenidos para la constante de asociación de AcAMP2Pff18/Pff20 a quitotriosa con las encontradas para otros dominios análogos.<sup>60</sup> En la primera fila se encuentra el valor para el AcAMP2 nativo, mientras que las dos filas siguientes muestran los valores determinados para dos mutantes en los que el residuo 18 se ha mutado por triptófano (fila 2) o naftilalanina (fila 3). Por tanto, puede concluirse que a medida que aumenta el tamaño del anillo aromático de la posición 18, el valor de la asociación ligando-receptor aumenta. Para tamaños análogos, Phe versus Pff, la desactivación del anillo aromático por la introducción de átomos de flúor provoca una disminución de la constante de asociación.

**Tabla 1.3.** Constantes de asociación ( $K_a$ ) y parámetros termodinámicos ( $\Delta H^\circ$  y  $\Delta S^\circ$ ) de la asociación a quitotriosa de AcAMP2Pff18/Pff20 y otros péptidos análogos, determinados mediante valoraciones de RMN.<sup>60</sup>

	$K_a$ [ $M^{-1}$ ]				Parámetros Termodinámicos	
	T=298K	T=303K	T=308K	T=313K	$\Delta H$ [KJ/mol]	$\Delta S$ [J/Kmol]
AcAMP2	1206	721	537	340	-63.4	-154.1
AcAMP2 (Trp18)	1755	1473	916	614	-54.1	-119.0
AcAMP2 (Nal18)	3527	2314	1638	999	-63.9	-146.6
AcAMP2 (Pff18/Pff20)	448	348	255	190	-45.1	-101.0

Los parámetros termodinámicos de la interacción se determinados usando los valores de  $K_a$ , mediante una aproximación de van't Hoff a cuatro temperaturas (298, 303, 308, 313 K, Figura 1.21).



**Figura 1.21** Aproximación de van't Hoff para la interacción AcAMP2Pff18/Pff20 y Quitotriosa

En lo que se refiere a los parámetros termodinámicos, generalmente, estos procesos de reconocimiento molecular proteína-carbohidrato están dirigidos entálpicamente, mientras que la entropía se opone a la asociación. Además, suelen observarse fenómenos de compensación entalpía-entropía.<sup>17,35,103</sup> Como puede observarse en la Tabla 1.3, para la interacción entre quitotriosa y AcAMP2Pff18/Pff20, el valor de  $K_a$  y

de  $\Delta H^\circ$  es sensiblemente inferior al de los otros péptidos, indicando la importancia de la desactivación del anillo aromático en la asociación. El que la entropía se oponga al proceso se debe obviamente a la formación de un complejo a partir de dos entidades moleculares y también a la rigidificación del carbohidrato, de la proteína, o de ambos. El estudio de dinámica molecular antes señalado indica claramente la rigidificación de las cadenas laterales de Pff18 y 20, que evidentemente debe tener lugar con una penalización entrópica. Las variaciones relativas entre los valores de estos parámetros también deben reflejar un distinto comportamiento del disolvente, que obviamente debe solvatar el sitio de unión de distinto modo, dependiendo de la naturaleza química y tamaño de los anillos aromáticos involucrados.

### 3. Parte Experimental

#### 3.1 Experimentos de 2D-RMN

Los espectros fueron obtenidos a 500 MHz y 800 MHz. La muestra de AcAMP2Pff18/Pff20-Quitotriosa se preparó en una solución tampón de (0.6 ml; 1:8 péptido-carbohidrato, 0.3 mM de péptido, en H<sub>2</sub>O/10%D<sub>2</sub>O, 100 mM NaCl, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 5.6). El experimento TOCSY<sup>80</sup> se llevó a cabo utilizando las secuencias estándar a 298 K, con tiempos *spin lock* de 50 y 70 ms. El NOESY-2D,<sup>81</sup> se adquirió con un tiempo de mezclado de 300 ms a 300 K. Los datos de RMN están depositados en la base de datos BMRB, con número de código 6591 para el péptido en complejo con quitotriosa y para el péptido libre 6656.

#### 3.2 Cálculo de geometría de distancias (DYANA) y Minimización Energética

Restringida (REM).

El cálculo de la estructura se inició mediante un protocolo de dinámica en el espacio de los ángulos de torsión con el programa DYANA.<sup>68</sup> Las restricciones de distancia superiores usadas en DYANA fueron obtenidas mediante las intensidades de los picos

de cruce NOESY, analizados a través del programa XEASY.<sup>92</sup> Estas intensidades se clasificaron como fuertes, medias y débiles, y se transformaron en límites superiores de 2.5, 3.5 5.5 Å, respectivamente. El procedimiento CALIBA<sup>86</sup> se usó para llevar a cabo un mejor ajuste de las distancias. Las restricciones de límites inferiores se colocaron de acuerdo a la suma de radios van der Waals entre dos átomos. Las restricciones para los protones diastereotópicos no asignados estereoespecíficamente se definieron de acuerdo a una posición intermedia (pseudoátomo), adicionando a la distancia 2.20 Å.

Los enlaces disulfuro se incluyeron como restricciones de distancia entre S-S ( $2.0 < r < 2.1$  Å) y para C $\beta$ -S ( $3.0 < r < 3.1$  Å). Cabe mencionar que, para el cálculo de la estructura del complejo, se adicionaron al cálculo 10 restricciones de distancia intermolecular péptido-carbohidrato, analizadas en el espectro NOESY.

Como información adicional, el protocolo de GLOMSA<sup>86</sup> permitió la asignación estereoespecífica del 50 % de los  $\beta$ -metilenos proquirales, con subsecuente aplicación del módulo de DYANA, *distance modify*.

Las 35 mejores estructuras en términos de función diana se sometieron a un protocolo de dinámica molecular restringida (RMD), con el campo de fuerzas de AMBER. La etapa inicial fue una minimización energética restringida con 2000 iteraciones de gradiente conjugado. Posteriormente, en el proceso de enfriamiento simulado (SA), las estructuras se equilibraron a una temperatura de 600 K durante 2 ps. Durante este paso, se experimentan varios cambios conformacionales y, en el siguiente paso de 2 ps, se inicia realmente la simulación mediante dinámica molecular restringida (RMD). En un paso subsiguiente, la temperatura se disminuye paulatinamente cada 2 ps hasta llegar a 100 K. Cuando esta última temperatura se alcanza, se llevó a cabo otra RMD de 4 ps. Finalmente, las estructuras se sometieron a una minimización energética restringida (REM), mediante 2000 iteraciones de gradiente conjugado.

La estructura refinada del complejo AcAMP2Pff18/Pff20-Quitotriosa se encuentra depositada en el Protein Data Bank (PDB: 1ZNT).

### 3.3 Dinámica Molecular en Disolvente Explícito (sMD)

La estructura de partida fue la obtenida mediante RMN, mediante los protocolos descritos en el párrafo anterior. Los archivos de topología y coordenadas, y preparación

de la caja con moléculas de agua para esta estructura se generaron a través del módulo Xleap de AMBER. Así, la estructura fue inmersa en una caja de 1836 TIP3P. Para el cálculo de interacciones de no enlace se aplicó un corte de 11 Å. Las simulaciones fueron llevadas a cabo mediante el módulo Sander, utilizando condiciones periódicas de contorno (*periodic boundary conditions*) y la aproximación PME (Particle Mesh Ewald) para introducir efectos electrostáticos de rango largo. Se usó algoritmo SHAKE para fijar los átomos de hidrógeno y el tiempo de integración fue de 2 fs, en cada paso.

Las simulaciones se llevaron a cabo a presión (1 atm) y temperatura (300K) constantes. La equilibración se realizó con las siguientes condiciones: La constante de fuerza para las restricciones fue de 500 kcalM<sup>-1</sup> Å. Manteniendo estas restricciones durante 12.5 ps, a 300K, se consiguió equilibrar la caja en presencia de los iones, usando un corte de 9 Å para el tratamiento de las interacciones electrostáticas. Se usó la aproximación PME para considerar los efectos electrostáticos de rango largo. Posteriormente, el sistema se sometió a varios ciclos de minimización, con 1000 iteraciones en orden descendente, en las que gradualmente fueron eliminándose las restricciones posicionales intra e Inter-moleculares. Es decir, la constante de fuerza fue disminuyendo paulatinamente desde 500 kcalmol<sup>-1</sup>Å hasta un valor de cero. El tiempo de simulación para el AcAMP2Pff18/Pff20 libre y su complejo con Quitotriosa fue de 4.5 ns y las trayectorias de la dinámica molecular se guardaron cada 0.5 ps. Finalmente, las 9000 estructuras guardadas de cada trayectoria se analizaron mediante el módulo *caranal* de AMBER.

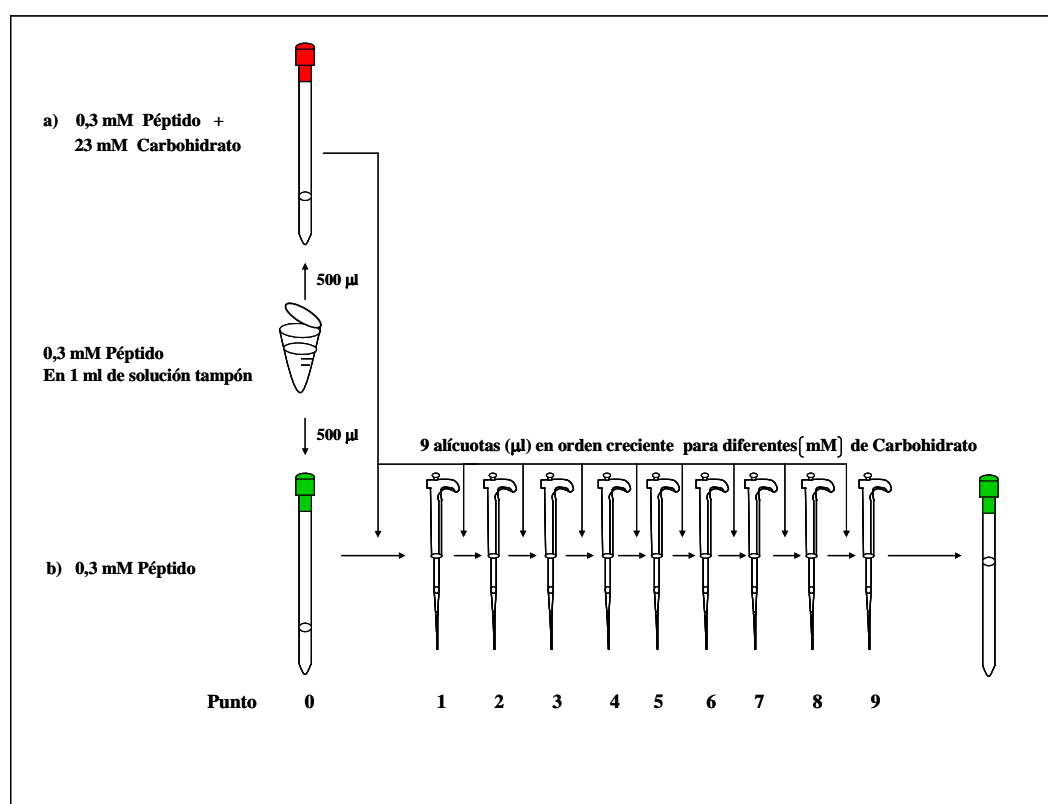
### 3.4 Protocolo Experimental por RMN de la Valoración para la Asociación de AcAMP2Pff18/Pff20- Quitotriosa

La asociación de la N, N', N''-triacetil-glucosamina con AcAMP2Pff18/Pff20 se siguió por Resonancia Magnética Nuclear monodimensional de <sup>19</sup>F, en un equipo de 500 MHz.

El procedimiento consistió de los siguientes puntos (también, véase Figura 1.22):

1.- Preparación de la muestra de proteína a una concentración 0,3 mM en 1 ml de solución tampón (H<sub>2</sub>O/10% D<sub>2</sub>O, 100 mM NaCl, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 5.6).

- 2.- Se dividió en dos partes la muestra del primer punto: **a)** a una alícuota de 500  $\mu$ l de AcAMP2Pff18/Pff20 (0,3 mM), se adicionó la Quitotriosa (23 mM) y **b)** la otra alícuota de 500  $\mu$ l de AcAMP2Pff18/Pff20 se usó como patrón.
- 3.- Se adquirieron espectros de <sup>19</sup>F-RMN de las muestras **(a)** y **(b)**.
- 4.- Posteriormente, para construir la curva de valoración se adicionaron de manera subsiguiente pequeñas alícuotas ( $\mu$ l) de la muestra **(a)** sobre la muestra **(b)**. En este caso se adicionaron 9 alícuotas. Para cada muestra, que contiene diferente concentración de carbohidrato y la misma concentración de polipéptido, se adquirió un espectro de <sup>19</sup>F-RMN, a cuatro diferentes temperaturas (298, 303, 308, 313 K).
- 5.- La constante de afinidad se evaluó para las cuatro temperaturas, mediante un método de regresión no lineal. Finalmente, se calcularon los parámetros termodinámicos de la interacción AcAMP2Pff18/Pff20-Quitotriosa,  $\Delta H^\circ$ ,  $\Delta S^\circ$ , mediante un ajuste de van't Hoff.



**Figura 1.22** Protocolo experimental de valoración mediante Resonancia Magnética Nuclear para la asociación AcAMP2Pff18/Pff20- N, N', N''-triacetil-glucosamina



#### 4. Conclusiones

1. Se ha descrito la estructura tridimensional en disolución de un péptido con dominio heveína, utilizando RMN y simulaciones de dinámica molecular.
2. Este péptido, AcAMP2Pff18/Pff20, que presenta dos residuos aromáticos en su estructura que han sido mutados por dos aminoácidos no naturales de *p*-fluorofenilalanina (*Pff*), mantiene la capacidad de unir oligosacáridos de quitina.
3. El sitio de reconocimiento de oligosacáridos es el mismo que en otros polipéptidos con dominios de heveína; en la familia de los AcAMP2, está constituido por la Ser16 y por los residuos aromáticos en posiciones 18, 20, y 27.
4. Los datos de RMN y de dinámica molecular indican que la estructura tridimensional del AcAMP2Pff18/Pff20 libre o unido a quitotriosa es muy similar a la de la proteína nativa.
5. Los datos de RMN (<sup>19</sup>F y <sup>1</sup>H) indican que, en disolución, existen dos complejos diferentes del AcAMP2Pff18/Pff20 asociado a quitotriosa. Estos complejos se diferencian en el modo de unión del carbohidrato a los residuos clave de AcAMP2Pff18/Pff20. En uno de ellos, el residuo terminal no reductor del trisacárido interacciona directamente con los residuos 20 y 27, mientras que en el otro es el residuo central del trímero es el que establece estas interacciones.
6. La topología del polipéptido libre y asociado es muy similar al encontrado para otros dominios de heveína en estado libre y asociado.
7. Específicamente, mediante el análisis comparativo entre las estructuras del complejo AcAMP2Pff18/Pff20-Quitotriosa y AcAMP2 nativo, únicamente existe una ligera diferencia en la orientación conformacional de los residuos Pff18 y 20, que puede atribuirse a la presencia de la quitotriosa, que restringe la movilidad conformacional de estos residuos. Esto está de acuerdo con los resultados de las simulaciones (sMD), que presentan mayor libertad conformacional, los residuos Pff18 y Pff20, en el estado libre que en el estado enlazado.

8. Se ha determinado la constante de asociación y los datos termodinámicos ( $\Delta H^\circ$  y  $\Delta S^\circ$ ) entre el AcAMP2Pff18/Pff20 y la quitotriosa, empleando valoraciones mediante RMN y mediante la aproximación van't Hoff respectivamente. La constante de asociación y el  $\Delta H^\circ$  encontrados para este péptido, cuyos residuos aromáticos presentan una densidad electrónica disminuida, son ligeramente menores que la encontrada para el péptido nativo y para otros dominios de AcAMP2, con residuos aromáticos con mayor densidad electrónica. Estos datos parecen indicar que la interacción carbohidrato-aromático que tiene lugar en dominios de heveína está modulada por la naturaleza química del anillo y por la densidad electrónica del mismo. El proceso está desfavorecido desde el punto de vista entrópico.

*CAPÍTULO 2*

***LA INTERACCIÓN DE UN MUTANTE DE HEVEÍNA TRUNCADA  
(Hev<sub>32</sub>S19D) CON OLIGOSACÁRIDOS DE QUITINA.***



## 1. Introducción

Como se ha mencionado anteriormente, en los procesos de reconocimiento molecular de carbohidratos por proteínas están involucradas una serie de interacciones clave, como son los enlaces de hidrógeno, las fuerzas de van der Waals, interacciones hidrófobas, y fuerzas de apilamiento, por citar las más evidentes. En definitiva, estos factores energéticos y estructurales modulan la selectividad, estabilidad y afinidad de las interacciones proteína-carbohidrato. Estudios previos, como los denotados en el capítulo 1 de esta memoria, han permitido corroborar que, en las interacciones de van der Waals e hidrófobas, los aminoácidos involucrados son fundamentalmente los aromáticos: Trp, Phe y Tyr, seguidos de Leu, Val, Ala.

Obviamente, para la formación de enlaces de hidrógeno es conocido que fundamentalmente participan los aminoácidos con cadenas laterales polares: Asp, Asn, Glu, Arg, His, Trp, Lys, Tyr, Gln, Ser, y Thr. Estos enlaces de hidrógeno, por sus características de estabilidad, geometría y fuerza de enlace, también contribuyen, de manera significativa, a la afinidad y especificidad de la asociación entre proteínas y carbohidratos. En general, los enlaces de hidrógeno pueden clasificarse en:

a) enlaces de hidrógeno cooperativos, que se pueden describir como:  $(NH)_n \rightarrow OH \rightarrow O$ , donde el OH (donador y aceptor) es un grupo hidroxilo del azúcar, y NH y O son los grupos donadores y aceptores, respectivamente, de los residuos involucrados en el sitio de unión.

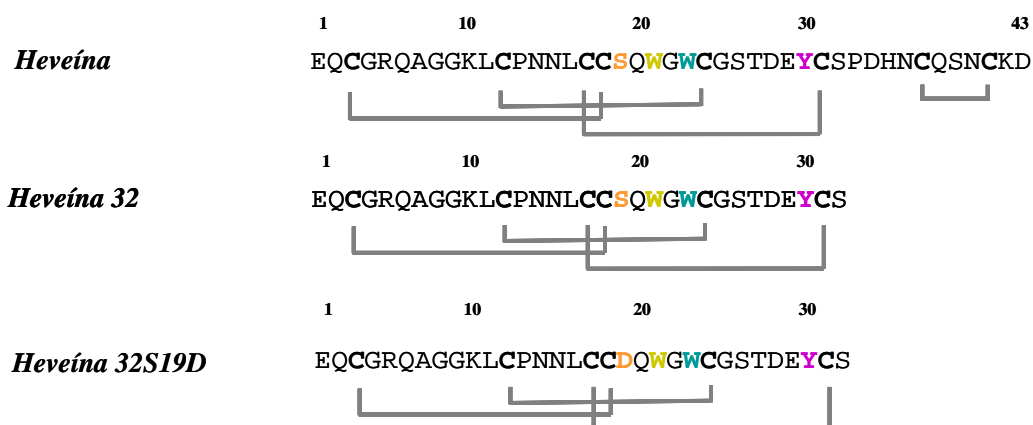
b) enlaces de hidrógeno bidentados, se forman cuando dos grupos OH contiguos de un azúcar interactúan con dos átomos que pertenecen a una cadena lateral (polar y planar) del polipéptido.

c) enlaces de hidrógeno que forman redes, que se forman a partir de la presencia de enlaces de hidrógenos cooperativos y bidentados.<sup>30</sup>

Cabe mencionar que la participación e importancia relativa, tanto de los enlaces de hidrógeno, como de los demás factores que están involucrados en la formación de complejos proteína-carbohidrato, depende de cada proteína en particular.

En el caso que nos ocupa en esta Memoria, el de la heveína y sus análogos estructurales, las características del sitio de unión son tales que estos péptidos son capaces de interactuar, de manera reversible y específica, con oligómeros de *N*-acetilglucosamina.

El conocimiento del modo de interacción de estos dominios de heveína con oligosacáridos de quitina ha permitido diseñar, en colaboración con el grupo del Prof. Dr. David Andreu (Univ. Pompeu Fabra), una heveína truncada, de solamente 32 aminoácidos, que sigue siendo capaz de interactuar, de manera eficaz, con quitoooligosacáridos, perdiendo solamente del orden de un 20% de la afinidad de la heveína nativa frente al trisacárido,<sup>47</sup> (GlcNAc)<sub>3</sub> (Figura 2.1). Al igual que en la heveína nativa, los residuos involucrados en el sitio de unión son aquellos con posiciones relativas 21, 23 (dos anillos de triptófano), que estabilizan el complejo por interacciones de apilamiento CH- $\pi$  y de van der Waals, la Tyr30, que proporciona interacciones de apilamiento con un metilo de uno de los grupos acetamido del azúcar y además, los grupos hidroxilo de los residuos Ser19 y Tyr30 intervienen, a través de enlaces de hidrógeno, con el carbonilo del grupo acetamido que se apila con la Tyr30, y con el OH-3 del residuo correspondiente de GlcNAc, respectivamente (véase Figura 3).

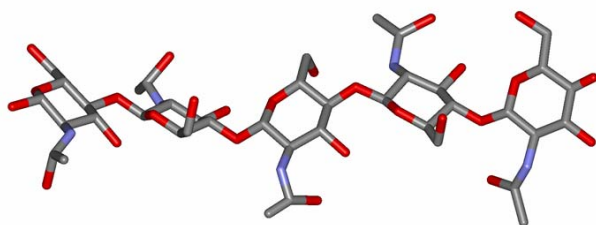


**Figura 2.1** Secuencia de aminoácidos de la heveína truncada32, heveína y el péptido mutante Heveína 32S19D, se indican los puentes disulfuro y en color se muestran los residuos involucrados en el sitio de reconocimiento.

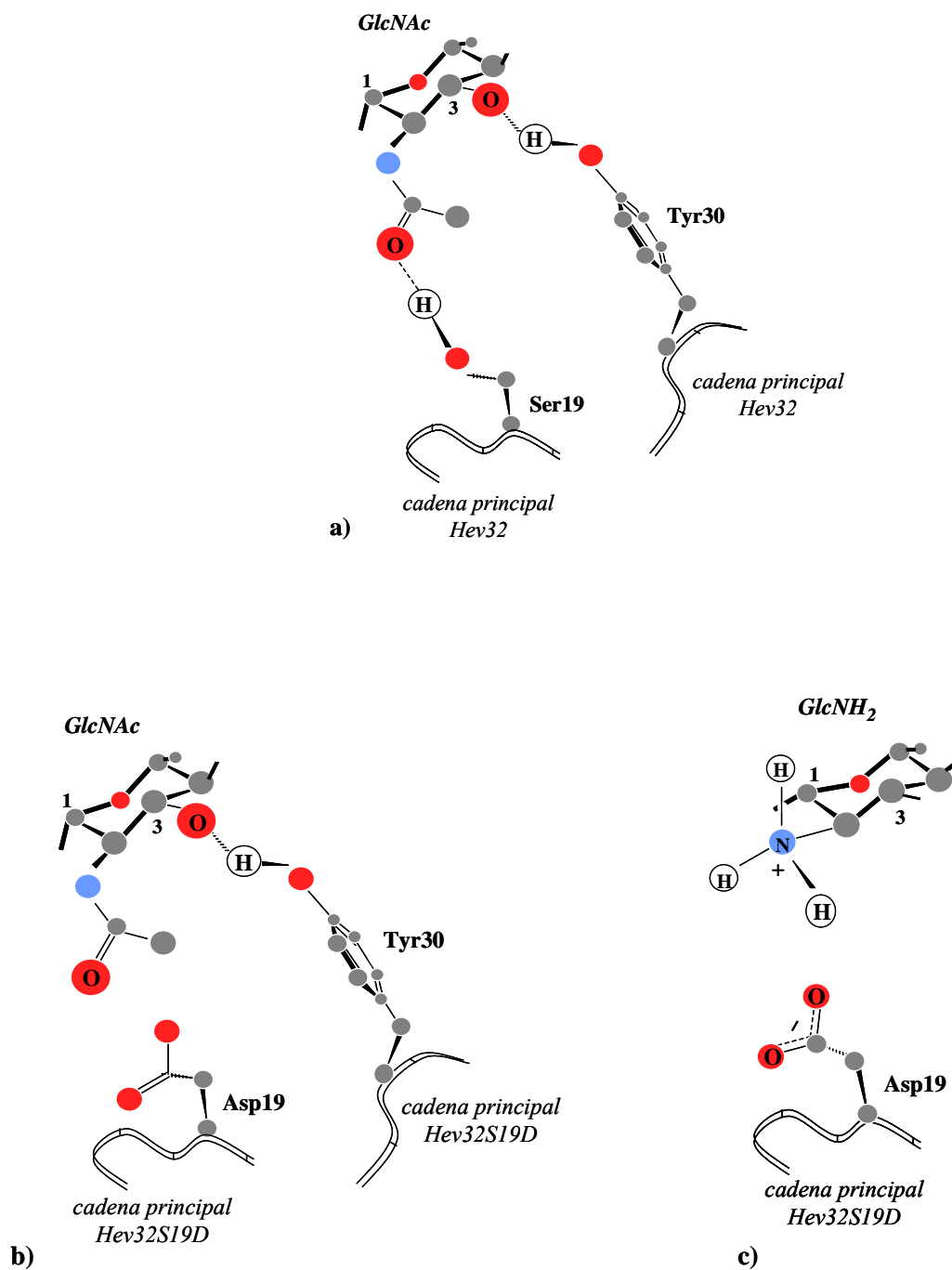
Sobre la base de estos datos estructurales, y observando el modo de interacción con quitooligosacáridos, tanto de la de heveína como de su péptido truncado, Hev32, nos planteamos la posibilidad de modular la especificidad de la heveína, generando un péptido mutante que no solamente fuese capaz de interactuar con quitooligosacáridos, sino también, y de modo preferente, con fragmentos de quitosano (polisacárido derivado de la quitina que está parcialmente N-desacetilado).

Para ello, y en colaboración con el grupo del Prof. Dr. David Andreu, se decidió mutar el residuo de Ser19, que proporciona uno de los enlaces de hidrógeno claves para el reconocimiento de GlcNAc (con el carbonilo del grupo acetamido) por un residuo de ácido aspártico, mutación S19D. El fin último de esta mutación sería mantener un aminoácido con una cadena lateral polar en la posición 19 pero con la posibilidad de establecer una interacción polar fuerte, no solamente con un grupo acetamido (como en la N-acetil glucosamina, GlcNAc), sino también con un grupo amino (como en la glucosamina, GlcNH<sub>2</sub>), componente del quitosano.

Así, en la Universidad Pompeu Fabra, se consiguió la síntesis de este mutante, cuya estructura 3D fue determinada en Madrid, mediante RMN por el Dr. Miquel Vila, del grupo del Prof. Andreu. Este análisis mostró que la mutación S19D no variaba sustancialmente la topología del polipéptido y que la orientación de las cadenas laterales seguía siendo análoga a la existente en el péptido original. Por tanto, se procedió al estudio, mediante RMN, de la interacción de este péptido, Hev32S19D, tanto con N,N',N'',N''',N''''- pentacetilquitopentaosa, (GlcNAc)<sub>5</sub>, como el correspondiente pentasacárido monodesacetilado en el extremo no reductor, GlcNH<sub>2</sub>β(1→4)(GlcNAc)<sub>4</sub>. (Figuras 2.2 y 2.3)



**Figura 2.2** Estructura de la N,N',N'',N''',N''''- pentacetilquitopentaosa, (GlcNAc)<sub>5</sub>.



**Figura 2.3** a) Interacción a través de enlaces de hidrógeno de los OH de la Ser19 y Tyr30 con el grupo acetamido del azúcar y el OH-3, respectivamente. b) Se ha cambiado el aminoácido Ser19 por el Asp. c) Se representa la posibilidad de un puente salino entre el COO<sup>-</sup> del ácido aspártico19 y el NH<sub>2</sub> del extremo no reductor del correspondiente pentasacárido monodesacetilado.

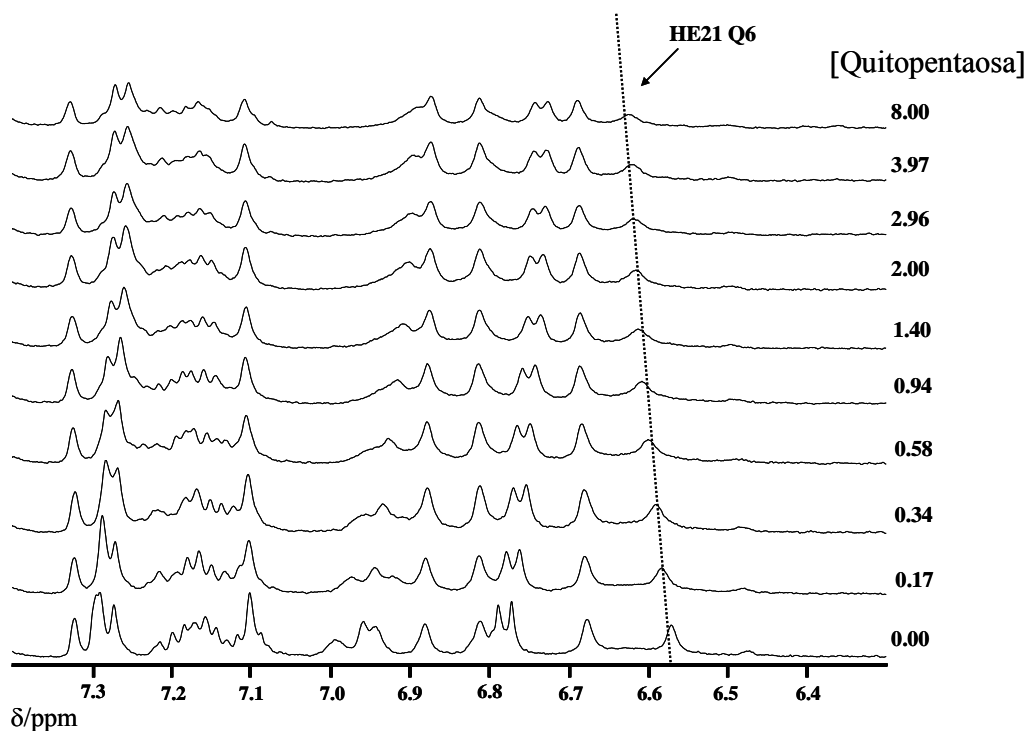


## 2. Resultados y Discusión

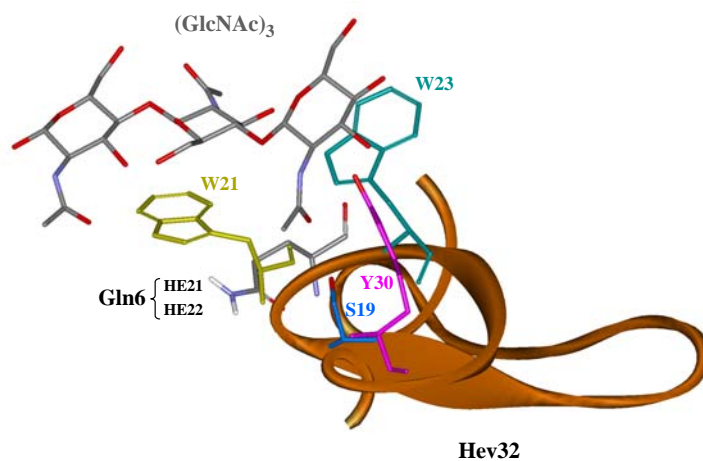
El estudio de la asociación del péptido mutante Hev32S19D con N, N', N'', N''', N''''-penta-acetilquitopentaosa y con su derivado monodesacetilado, en el extremo terminal no reductor, se llevó a cabo mediante experimentos de valoración, usando <sup>1</sup>H RMN. El protocolo se describe en la parte experimental de este capítulo y simplemente, se basa en la realización, registro y análisis de espectros 1D de una muestra que contiene una concentración constante del polipéptido en un tubo de RMN y a la que se le añaden, de manera sistemática, cantidades crecientes del azúcar correspondiente.

Los espectros se realizaron a tres diferentes temperaturas (303, 308 y 313 K), y los cambios observados en los desplazamientos químicos de algunas de las señales del péptido permitieron, por una parte, corroborar la formación de complejo entre el péptido Hev32S19D y los dos carbohidratos y, por otra, estimar sus constantes de asociación correspondientes.

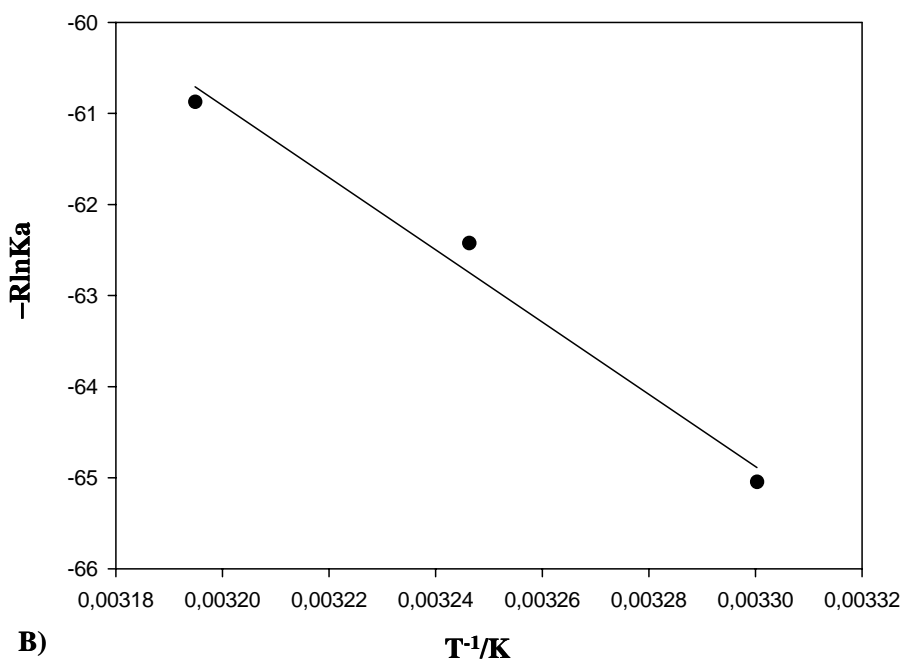
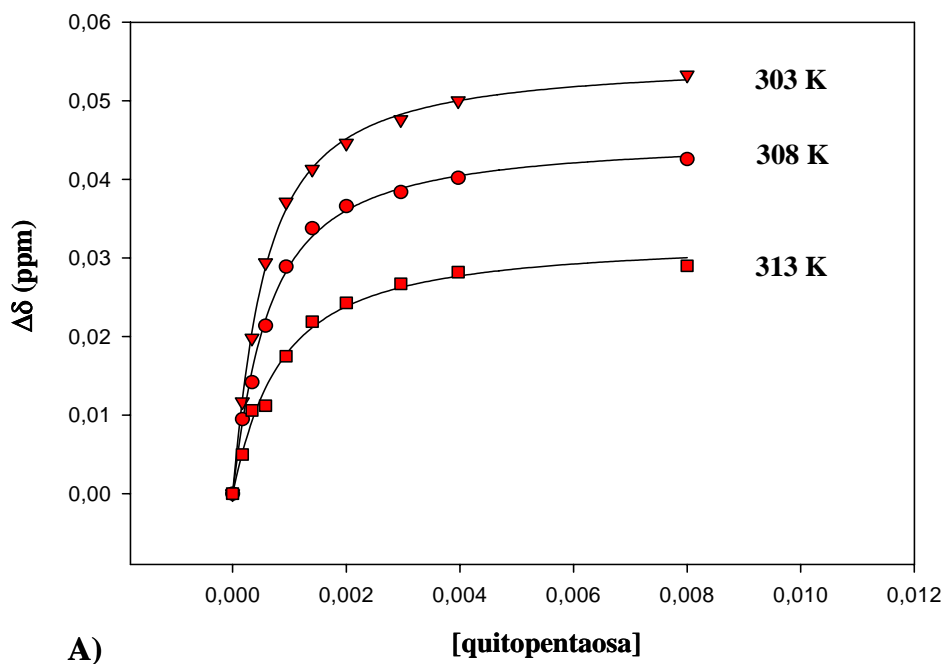
En ambos casos, la escasa variación de anchura de las señales del péptido tras la adición de cantidades crecientes del azúcar permitió deducir que la velocidad del proceso de intercambio rápido entre los estados libre y asociado es rápido en la escala de tiempo de los desplazamientos químicos. Por tanto, aparece una única señal para cada núcleo del polipéptido, cuyo desplazamiento químico observado está promediado entre los estados libre y asociado, de acuerdo con sus fracciones molares correspondientes. Por lo tanto, las constantes de asociación ( $K_a$ ) se determinaron mediante un ajuste no lineal de los desplazamientos químicos observados respecto a las concentraciones de ligando y receptor (Figura 2.6 A)). Se siguió la señal de <sup>1</sup>H RMN correspondiente al protón HE21 del residuo de Gln 6 (Figuras 2.4). Esta señal muestra una variación notable de desplazamiento químico a las distintas concentraciones de ligando y está lo suficientemente aislada para que el error cometido sea mínimo. Desde un punto de vista estructural, y de acuerdo con la conformación del complejo determinada mediante RMN, la cadena lateral del aminoácido Gln 6, en el complejo Hev<sub>32</sub>-quitotriosa<sup>49</sup> está muy cercana al Trp 21, que a su vez está involucrado directamente en la interacción con el carbohidrato (Figura 2.5), por lo que las variaciones de los desplazamientos químicos de este residuo son un reflejo inequívoco del proceso de interacción.



**Figura 2.4** Serie de espectros de <sup>1</sup>H RMN (500 MHz, pH 5.6, 303 K) de la valoración, a diferentes concentraciones (mM) de pentasacárido (GlcNAc)<sub>5</sub>, con el péptido Hev32S19D, a una concentración constante de 0.28 mM. Puede observarse el desplazamiento químico sistemático de la señal HE21 de la Gln 6 en función de la concentración de pentasacárido.



**Figura 2.5** El complejo Hev32-triacetilquitotriosa. Se muestran los residuos del sitio de reconocimiento, y los protones He21 y He22 de la Gln6.



**Figura 2.6** A) Gráfica de la valoración de RMN de la asociación de la penta-acetilquitopentaosa con el mutante *Hev32S19D*. B) Representación de van't Hoff para el complejo *Hev32S19D*-penta-acetilquitopentaosa.

La representación de la variación de los valores de  $K_a$  frente a la inversa de la temperatura permitió obtener los parámetros termodinámicos,  $\Delta H^\circ$  y  $\Delta S^\circ$  (Tabla 2.1), merced a la aproximación de van't Hoff (Figura 2.6 B)).

**Tabla 2.1** Constantes de asociación ( $K_a$ ) a diferentes temperaturas determinadas por RMN ó ITC y parámetros termodinámicos para la heveína y la Hev<sub>32</sub> interaccionando con (GlcNAc)<sub>3</sub> (\*) y (GlcNAc)<sub>5</sub> (\*\*).  $K_a$  para el mutante Hev<sub>32</sub>S19D interaccionando con (GlcNAc)<sub>5</sub> (\*\*) y tetracetilquitopentaosa (\*\*\*)

	$K_a$ [M <sup>-1</sup> ]				Parámetros Termodinámicos		
	T=298 K	T=303 K	T=308 K	T=313 K	$\Delta G$ [kJmol <sup>-1</sup> ]	$\Delta H$ [kJmol <sup>-1</sup> ]	$\Delta S$ [JK <sup>-1</sup> mol <sup>-1</sup> ]
<b>Heveína*</b>	11 500	8700	6900	5700	- 23.1	- 36.4	- 44.5
<b>Heveína**</b>	-	474 000	-	-	-32.6	-40.1	-26.3
<b>Hev32*</b>	7700	4200	3400	2200	- 21.8	- 62.6	-136.0
<b>Hev32S19D**</b>	-	2491	2130	1512	- 19.7	- 39.4	- 64.4
<b>Hev32S19D***</b>	173	118	106	102	- 11.8	- 26.1	- 45.6

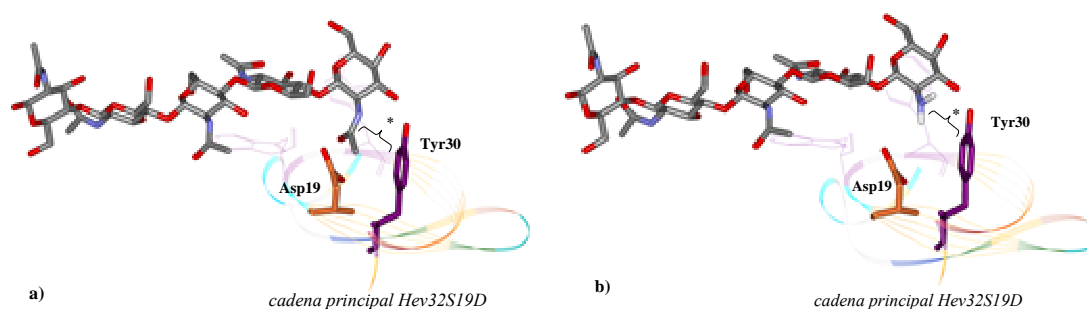
El análisis de los datos de  $K_a$  y de los parámetros termodinámicos mostrados en la Tabla 2.1, obtenidos para el complejo Hev32S19D-(GlcNAc)<sub>5</sub>, permite concluir que, incluso con la variación S19D, el péptido mutante mantiene la capacidad de unirse a oligosacáridos derivados de N-acetilglucosamina con afinidad moderada. Sin embargo, el valor de  $K_a$  de la asociación Hev32S19D-quitopentaosa es más de dos órdenes de magnitud más bajo que el descrito para la interacción de la heveína con el mismo pentasacárido usando microcalorimetría (ITC)<sup>51</sup>. Estudios de ultracentrifugación analítica y de DOSY-RMN realizados en nuestro grupo de trabajo han demostrado que la alta afinidad de la heveína por el pentasacárido de quitina, u oligosacáridos de mayor tamaño, se debe a la existencia de complejos de estequiometría mayor de 1:1, en los que más de un dominio de heveína puede interaccionar con una única cadena de oligosacárido, si este es lo suficientemente largo. En estos casos, es imposible determinar la constante de asociación mediante RMN, puesto que las señales de RMN son muy anchas y el equilibrio es intermedio-lento en la escala de tiempos de desplazamientos químicos. En el caso aquí estudiado, las señales siguen siendo finas y el ajuste se realiza perfectamente asumiendo un modelo de estequiometría 1:1. La afinidad encontrada para la interacción de Hev32S19D con quitopentaosa es algo menor

que la hallada para la interacción de Hev ó Hev32 con el trisacárido de quitina, que también presentan una estequiometría 1:1 (Tabla 2.1).<sup>47, 49</sup>

En principio, y a falta de datos estructurales más fidedignos, el descenso de afinidad para la interacción de Hev32S19D con (GlcNAc)<sub>5</sub>, en comparación con el péptido original, puede deberse a la pérdida (al menos parcial) del enlace de hidrógeno clave entre el residuo en posición 19 con el grupo carbonilo del azúcar y a la pérdida de la interacción de apilamiento entre el metilo del acetamido y la Tyr30. Es decir, un residuo de ácido aspártico, en vez de una serina en posición 19, no puede adoptar la orientación adecuada para interactuar con el residuo de GlcNAc. Asimismo, la eliminación del grupo acetamido del residuo terminal, no reductor, de la cadena del pentasacárido produce una dramática pérdida de afinidad, indicando que, también en este caso, los dominios de heveína están específicamente diseñados para interactuar con quitooligosacáridos y que una pequeña variación estructural, bien en el péptido o bien en el azúcar, produce una pérdida significativa de afinidad.

Con respecto a los parámetros termodinámicos, obtenidos mediante una aproximación de van't Hoff a partir de las  $K_a$  a las distintas temperaturas, los valores negativos de  $\Delta H^\circ$  y  $\Delta S^\circ$  están de acuerdo con un proceso dirigido entálpicamente. Si se comparan con los descritos previamente descritos para la heveína y algunos de sus análogos estructurales (Tabla 2.1), es evidente que los datos manifiestan un fenómeno de compensación entropía-entalpía, típico de los procesos de asociación entre proteínas y carbohidratos. Una mejor entalpía de unión viene acompañada, normalmente, de una mayor pérdida entrópica. Los procesos dirigidos entálpicamente suelen deberse al establecimiento de interacciones de van der Waals y enlaces de hidrógeno en el complejo ligando-receptor, como se ha demostrado en el caso de los dominios de heveína. La pérdida de entropía se debe a la pérdida de grados de libertad traslacionales, rotacionales y de torsión, tanto del ligando como del complejo, al pasar del estado libre al asociado. La comparación específica de los casos concretos del péptido mutante Hev32S19D interactuando con la penta-acetilquitopentaosa y con su derivado monodesacetilado, tetracetilquitopentaosa, indican que, para el complejo Hev32S19D-pentacetilquitopentaosa, la entalpía de asociación es mucho mayor ( $\Delta H^\circ = -39.4 \text{ kJmol}^{-1}$ ) que para aquel con tetracetilquitopentaosa ( $\Delta H^\circ = -26.13 \text{ kJmol}^{-1}$ ). Obviamente, esta diferencia debe atribuirse a la ausencia del grupo acetamido que está implicado en un

enlace de hidrógeno (hidroxilo-carbonilo), que podría reemplazarse por otro, carboxilo-amino, en el caso del derivado de quitosano y, muy probablemente, a la pérdida de una interacción de apilamiento entre el grupo metilo del acetamido y el anillo aromático de la Tyr30 (Figura 2.7). La penalización entrópica ( $\Delta S^\circ = -64.4 \text{ J K}^{-1} \text{ mol}^{-1}$ ) que se incrementa en el complejo *Hev32S19D*-penta-acetilquitopentaosa, sugiere que el acomodo del pentasacárido para dar las interacciones estabilizantes con el péptido viene acompañado de una rigidificación de sus movimiento torsionales, así como de las cadenas laterales correspondientes.



**Figura 2.7** a) Representación del complejo *Hev32S19D-(GlcNAc)<sub>5</sub>*, (\*) se muestra la interacción de apilamiento entre el grupo metilo del grupo acetamido y el anillo aromático de la Tyr30. b) Representación del complejo *Hev32S19D-(GlcNH<sub>2</sub>-(GlcNAc)<sub>4</sub>)* (\*) ausencia del apilamiento del  $\text{NHCOCH}_3$ -Aromático y posibilidad de formación de otro tipo interacción carboxilo (*Asp19*)-amino (*GlcNH<sub>2</sub>*).

### 3. Parte Experimental

Protocolo experimental de la valoración (ver Figura 1.22), mediante <sup>1</sup>H RMN, de la asociación del péptido mutante Hev32S19D con N, N', N'',N''',N''''-penta-acetilquitopentaosa y su derivado monodesacetilado en el extremo terminal no reductor. Ambas valoraciones se realizaron exactamente mediante el mismo protocolo, que consistió de los siguientes pasos:

1.- Se preparó la muestra del péptido Hev32S19D, con una concentración 0,28 mM, en 1 ml de solución tampón (90% H<sub>2</sub>O/10% D<sub>2</sub>O, 100 mM NaCl, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 5.6).

2.- Esta muestra se dividió en dos partes:

**a)** A una alícuota de 500 µl de Hev32S19D (0,28 mM), se añadió el pentasacárido correspondiente (8 mM) **b)** la otra alícuota de 500 µl de Hev32S19D se usó como patrón (punto 0 de la valoración).

3.- Se adquirieron los espectros de <sup>1</sup>H RMN de las muestras **(a)** y **(b)** en un equipo de 500 MHz, a distintas temperaturas.

4.- Posteriormente, para construir la curva de valoración, se adicionaron, de manera sistemática, pequeñas alícuotas (µl) de la muestra **(a)** sobre la muestra **(b)**. En este caso se adicionaron 9 alícuotas. Para cada muestra, que contenía diferentes concentraciones de pentasacárido y la misma concentración de polipéptido, se adquirió un espectro de <sup>1</sup>H RMN, a tres ó cuatro temperaturas diferentes (298, 303, 308 y 313 K). Se tabularon los datos obtenidos de las valoraciones, a cada temperatura, para la asociación entre Hev32S19D y ambos pentasacáridos y, en la Tabla 2.2, se muestran los obtenidos para la valoración con (GlcNAc)<sub>5</sub>, ya que las variaciones observadas para el derivado monodesacetilado fueron mucho menores.

**Tabla 2.2** Datos de los desplazamientos químicos de la señal HE21 de la Gln6, obtenidos a partir de las valoraciones del péptido mutante Hev32S19D con penta-acetilquitopentaosa. Se indican los valores obtenidos para tres temperaturas.

Punto Valoración	Péptido Hev32S19D	Pentasacárido	$\delta_{\text{obs}}$ 303 K	$\delta_{\text{obs}}$ 308 K	$\delta_{\text{obs}}$ 313 K
0	0.28 mM	8.00 mM	6.570	6.609	6.651
1	0.28 mM	3.97 mM	6.582	6.619	6.656
2	0.28 mM	2.96 mM	6.590	6.623	6.661
3	0.28 mM	2.00 mM	6.599	6.631	6.662
4	0.28 mM	1.40 mM	6.607	6.638	6.668
5	0.28 mM	0.94 mM	6.612	6.643	6.672
6	0.28 mM	0.58 mM	6.615	6.646	6.675
7	0.28 mM	0.34 mM	6.618	6.648	6.677
8	0.28 mM	0.17 mM	6.620	6.649	6.679
9	0.28 mM	0.00 mM	6.624	6.652	6.680

5.- Finalmente, para cada concentración, la diferencia de desplazamiento químico de la señal HE21 de la glutamina 6, con el correspondiente a la muestra del punto 0, se usó para estimar la constante de afinidad a las distintas temperaturas, mediante un método de regresión no lineal. A continuación, se calcularon los parámetros termodinámicos  $\Delta H^\circ$  y  $\Delta S^\circ$ , de la interacción Hev32S19D con los pentasacáridos, mediante una aproximación de van't Hoff.



#### 4. Conclusiones

1.- Se ha estudiado la interacción de un péptido mutante con dominio heveína, Hev32S19D, con N, N',N'',N''',N''''-penta-acetilquitopentaosa y un derivado monodesacetilado, mediante valoraciones de <sup>1</sup>H RMN. Las perturbaciones de los desplazamientos químicos del péptido en presencia de concentraciones crecientes del azúcar correspondiente, han permitido obtener los valores de K<sub>a</sub>, y, mediante una aproximación de van't Hoff se han estimado los valores de ΔH° y ΔS°. Los resultados obtenidos permiten concluir que el mutante mantiene la capacidad de unir oligómeros de quitina.

2.- No obstante, se observa una disminución significativa de los valores de K<sub>a</sub> de este péptido mutante Hev32S19D frente a la penta-acetilquitopentaosa con respecto a los determinados previamente para la heveína y la heveína truncada 32. Es decir, se demuestra inequívocamente la aportación clave de la Ser19 para establecer un enlace de hidrógeno con el carbonilo del grupo acetamido del carbohidrato, que contribuye decisivamente a la estabilidad y especificidad del proceso de reconocimiento molecular proteína-carbohidrato.

3.- Sin embargo, el péptido mutante no es capaz de reconocer, de manera significativa, el análogo de quitosano (un pentasacárido monodesacetilado en la posición terminal no reductora). La K<sub>a</sub> disminuye drásticamente, lo que indica que 1) la presencia de grupos acetamido es esencial para el reconocimiento molecular de azúcares, 2) el enlace de hidrógeno entre el hidroxilo de la Ser19 y el acetamido del azúcar no parece que pueda mimetizarse, al menos en este caso, por un puente salino entre el D19 y un residuo de glucosamina y 3) la interacción de apilamiento entre el grupo metilo del acetamido del azúcar y el anillo aromático de la Tyr30. De hecho, los valores termodinámicos de la asociación de Hev32S19D con penta-acetilquitopentaosa y con tetra-acetilquitopentaosa indican que la ausencia del grupo acetamido en el extremo no reductor del pentasacárido provoca una disminución drástica de la entalpía de asociación.

4.- Los datos termodinámicos indican que el proceso de reconocimiento del péptido mutante Hev32S19D por penta-acetilquitopentaosa está dirigido por la entalpía,

mientras que está desfavorecido desde el punto de vista entrópico. Los valores negativos de entalpía están de acuerdo con que las fuerzas de van der Waals y los enlaces de hidrógeno proporcionan estabilidad al complejo

*CAPÍTULO 3*

*LA ESTRUCTURA 3D EN DISOLUCIÓN DE UN PÉPTIDO  
AISLADO DEL TOMATE.*



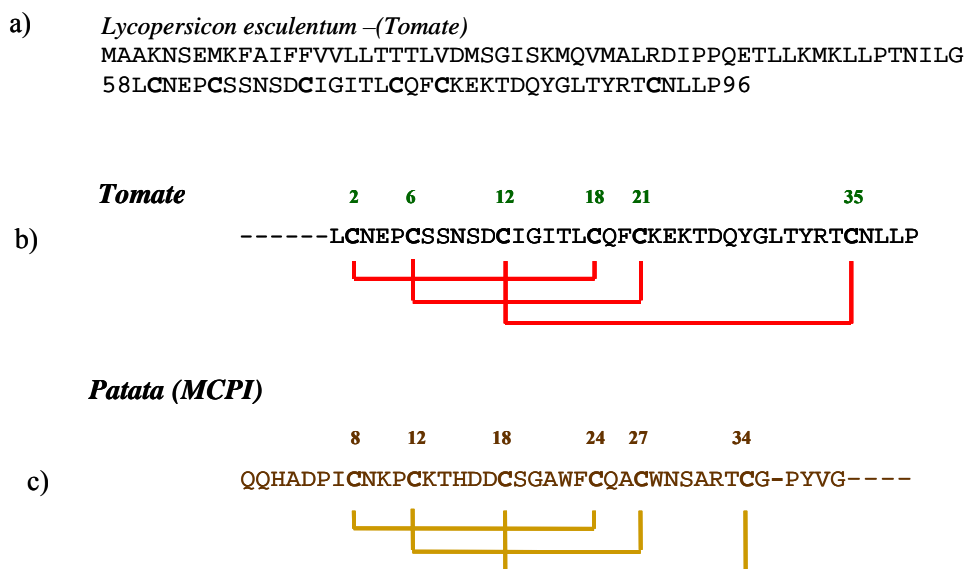
## 1. Introducción.

Se han identificado lectinas en numerosas plantas de la familia de leguminosas, y en particular, se han caracterizado algunas en gramíneas (cereales, como el germen de trigo) y solanáceas (patata y tomate) que unen quitina. Como se mencionó anteriormente, las propiedades antifúngicas y antimicrobianas características de estas plantas se atribuyen a la presencia de dominios de heveína.

En este sentido, y continuando con nuestros estudios sobre proteínas vegetales, en colaboración de la Dra. van Damme, de la Universidad de Gante, nos propusimos estudiar, mediante Resonancia Magnética Nuclear, la estructura tridimensional de un péptido de 39 aminoácidos, aislado del fruto del tomate en esta Universidad y que, de acuerdo con estudios iniciales, parecía ser un dominio de heveína. Este péptido provenía de una proteína más grande, de 96 aminoácidos (Figura 3.1), específica del fruto del tomate (*Lycopersicon esculentum*) y fue purificado mediante una cromatografía de afinidad usando una columna de quitina. Por tanto, además, se procedió a realizar el análisis de su interacción con quitotriosa, quitohexaosa, y con el polímero quitina. Como se verá a continuación, el péptido resultó no ser una heveína y tampoco se detectó ninguna afinidad por quito oligosacáridos de distintos tamaños.

De hecho, buscando en bases de datos de proteínas, se encontró homología con el Inhibidor de Metalloproteína (MCPI), que es un también un péptido de 39 aminoácidos, en este caso aislado de la patata (*Solanum tuberosum*, figura 3.1), cuya estructura había sido previamente caracterizada mediante RMN.<sup>104</sup> No obstante, solamente 12 de los 39 aminoácidos son idénticos.

Por tanto, en este capítulo de la Tesis se describe el protocolo usado para determinar la estructura tridimensional de un péptido del tomate (39 aa), y los intentos fallidos de los estudios de su interacción con oligo y polisacáridos de *N*-acetilglucosamina.



**Figura 3.1** a) Secuencia del producto de traducción primaria de una proteína de 96 aa aislada del tomate (*Lycopersicon esculentum*). b) Secuencia de aminoácidos del péptido (39 aa) en estudio. c) Secuencia de los 39 aminoácidos del Inhibidor de Metalloproteasa (MCPI) aislado de la patata (*Solanum tuberosum*). Se indican los puentes disulfuros de las dos últimas secuencias.

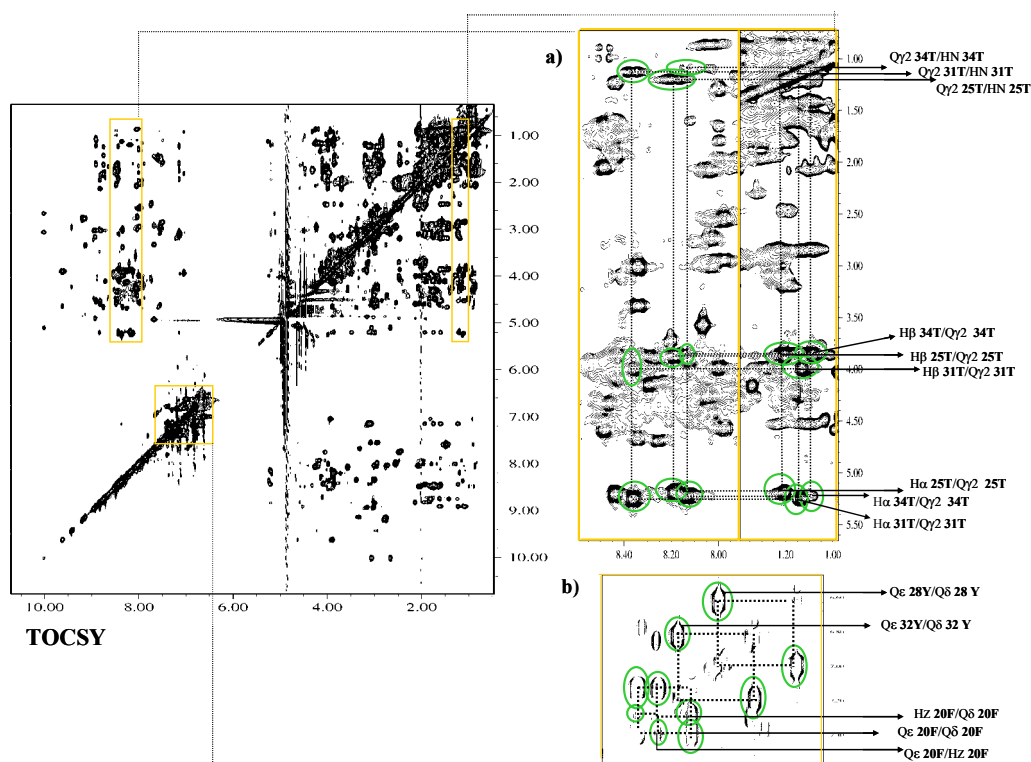
## 2. Resultados y Discusión

### 2.1 Determinación de la Estructura Tridimensional de un Péptido del Tomate

#### 2.1.1 Asignación de los Desplazamientos Químicos de <sup>1</sup>H-RMN del Péptido del Tomate.

Mediante los espectros COSY, TOCSY, y usando el programa XEASY y la metodología estándar de Wüthrich, se procedió a la asignación secuencial de los protones de cada aminoácido que constituyen el péptido. En concreto, el proceso se inició con la identificación de los sistemas de espín con patrones característicos. Así, las treoninas 25, 31 y 34 (Figura 3.2 a), se identificaron fácilmente en el espectro TOCSY, ya que sus H $\beta$  muestran un desplazamiento químico muy similar al de los protones H $\alpha$ .

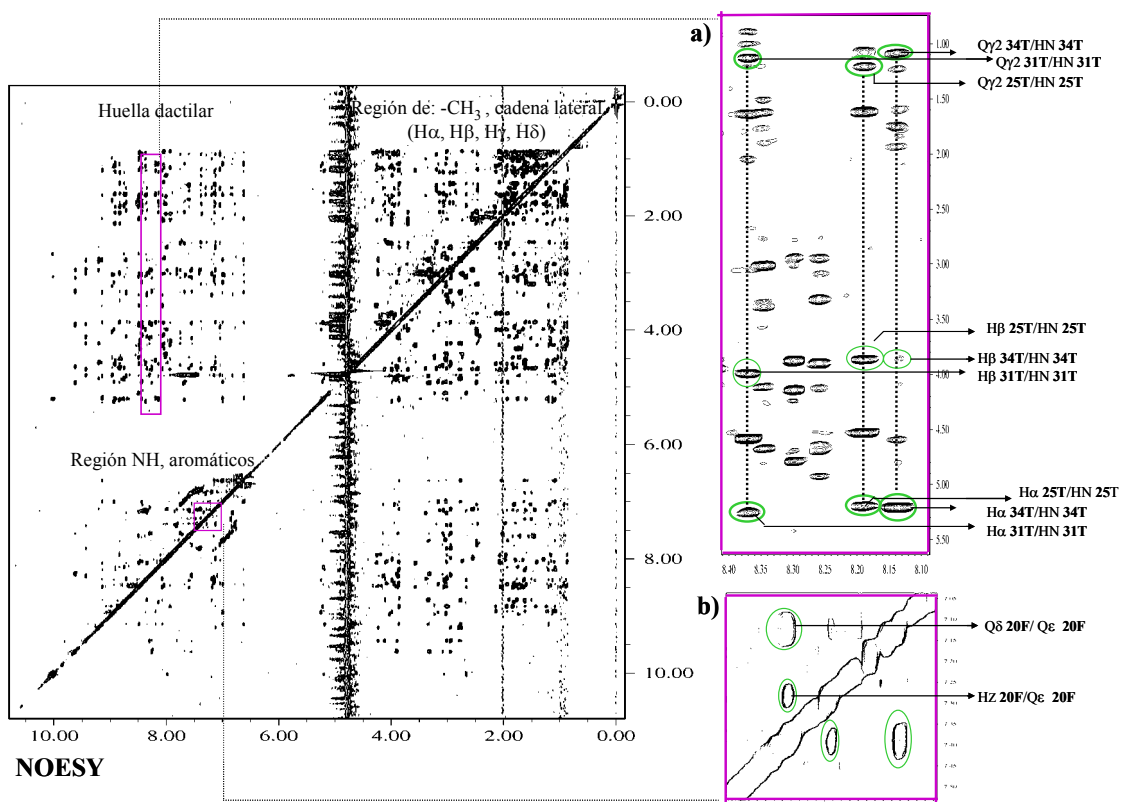
El acoplamiento con los metilos  $\gamma$ , a campo alto, indica específicamente el sistema de espín de Thr. Desde una perspectiva estructural, y para realizar el cálculo de la estructura 3D en base a NOEs, a los grupos metilo se les asigna un pseudoátomo (Q), en el punto intermedio entre los 3 átomos de hidrógeno equivalentes.



**Figura 3.2** Espectro TOCSY (500 MHz, tiempo de mezcla 60 ms, 298K), y las ampliaciones: a) Se muestran en círculos de color verde los sistemas de espín de las Treoninas 25, 31 y 34. b) Se indican en círculos de color verde los sistemas de espín aromáticos, AA'XX'M para la Phe20 y los sistemas AA'XX de las Tyr 28 y Tyr 32

Otros sistemas de espín fácilmente identificables fueron los de la Phe 20 (AA'XX'M), Tyr 28 y Tyr32 (AA'XX'), en la región de los protones aromáticos (Figura 3.2 b).

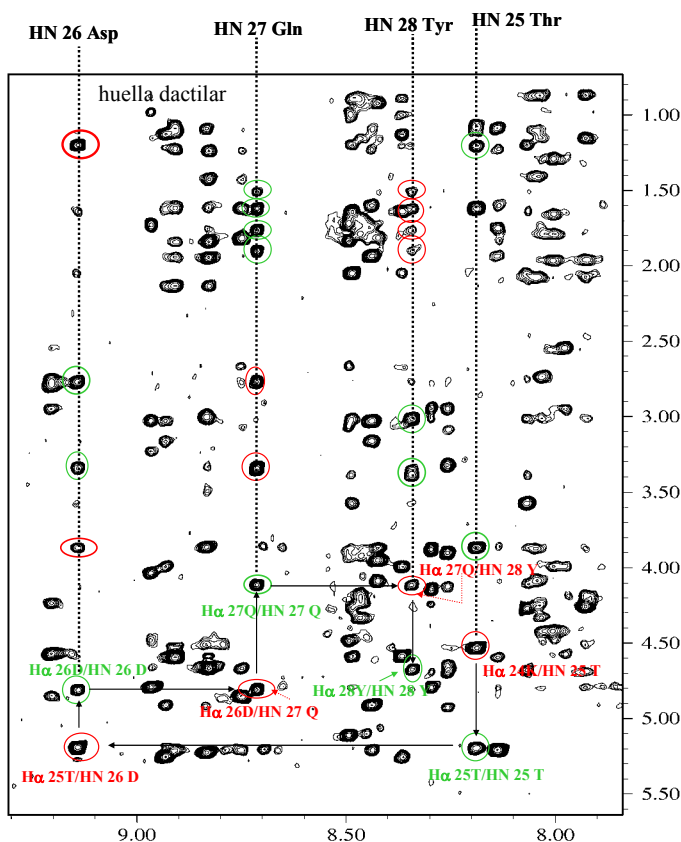
Una vez identificados algunos sistemas de espín, se pasaron al espectro NOESY, Figura 3.3, y así, mediante una continua combinación de análisis de los espectros COSY, TOCSY y NOESY se completaron todos los sistemas de espín de los aminoácidos, así como sus conectividades secuenciales. A continuación, se identificaron los picos de cruce NOE, imprescindibles para llevar a cabo el cálculo de la estructura tridimensional.



**Figura 3.3** Espectro NOESY (800 MHz, tiempo de mezcla 150 ms, 298K) se muestran las ampliaciones: a) Región de la huella dactilar, con los picos de cruce claves de NH con los H $\beta$  y H $\alpha$  de cada una de las Thr 25, 31 y 34. b) Picos de cruce del residuo Phe20.

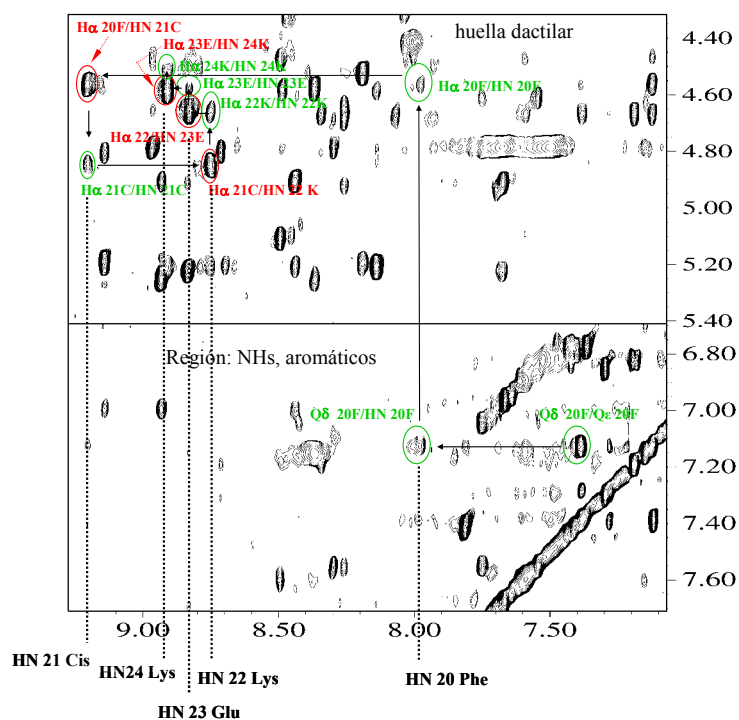
Un ejemplo de conectividad secuencial en el espectro NOESY (y apoyado en el espectro TOCSY) es el del tipo H $\alpha$ <sub>i</sub>-NH<sub>i+1</sub> (Figura 3.4). Aquí, se muestra la correlación NOE del H $\alpha$  de la Thr 25 con el NH del Asp 26, y sobre la línea vertical es este, se localiza el H $\alpha$  de D26, que tiene NOE con el NH de la Gln 27. El H $\alpha$  de éste último residuo muestra NOE (línea vertical) con el NH de la Tyr 28, cuyos protones aromáticos habían sido previamente encontrados en el espectro TOCSY.





**Figura 3.4** Ampliación (huella dactilar) del espectro NOESY, indicando la conectividad secuencial, mediante la interacción NOE ( $H\alpha_i-NH_{i+1}$ ), desde el residuo Thr 25, hasta los residuos Asp26, Gln 27, y Tyr 28. Se muestran en círculos de color verde las interacciones TOCSY y en círculos rojos, las interacciones NOE.

Por otro lado, el fragmento de los residuos 25-28 se conectó con otro fragmento secuencial, identificado de manera análoga, entre los residuos 20-24. Así, en los protones aromáticos  $H\delta$  de la Phe 20 (figura 3.5), se observó una correlación intraresidual con su NH, y sobre la vertical de éste se localizó su  $H\alpha$ , que muestra NOE con el NH del residuo Cys 21. El  $H\alpha$  de 21, localizado en la vertical, muestra NOE con el NH de la Lys 22, cuyo  $H\alpha$  presenta NOE con el NH de Glu 23. El  $H\alpha$  de 23 presenta otro NOE con el NH de la Lys 24.



**Figura 3.5** Ampliaciones (huella, dactilar, y región de aromáticos) del espectro NOESY, mostrando la conectividad secuencial a través de la interacción ( $H\alpha_i-NH_{i+1}$ ) de los residuos 20, 21, 22, 23, 24. Se indican en círculos de color verde los acoplamientos TOCSY y en círculos de color rojo las interacciones NOE.

De esta manera se logró asignar el 98 % de los protones de la secuencia de 39 aminoácidos que constituyen el péptido (véase Tabla 3.1).

**Tabla 3.1.** Desplazamientos químicos ( $\delta$ ) de  $^1\text{H}$ -RMN del Péptido del Tomate de 39 aa.

aa	NH	$\alpha\text{H}$	$\beta\text{H}$	$\gamma\text{H}$	$\delta\text{H}$	Otros
Leu1	-	4.032	1.714, 1.745	1.158	0.983	
Cys2	8.965	4.474	3.025, 3.228			
Asn3	8.828	4.508	3.034		6.964, 7.667	
Glu4	7.929	4.686	1.959, 2.067	2.470, 2.561		
Pro5	-	4.912	1.932, 2.482	2.050, 2.140	3.705, 3.994	
Cys6	7.675	4.922	3.317, 3.343			
Ser7	9.617	4.639	3.888, 3.920			
Ser8	8.259	4.688	3.898, 4.121			
Asn9	9.429	4.235	2.945, 3.014		7.413, 7.814	
Ser10	8.298	4.146	3.875			
Asp11	7.548	4.613	2.945, 3.084			
Cys12	7.751	4.045	2.480			
Ile13	7.045	3.947	1.815,	0.926, 1.815, 1.521	0.870	
Gly14	8.425	3.869, 4.083				
Ile15	8.046	4.256	2.080	0.884, 1.104, 1.409	0.865	
Thr16	7.930	4.227	4.184	1.162		
Leu17	8.465	4.327	1.738, 1.789	1.689	0.901, 0.948	
Cys18	8.035	4.419	2.736, 3.014			
Gln19	8.007	3.986	1.295, 1.657	1.798, 1.969		$\epsilon\text{H}21$ 6.780, 7.367
Phe20	7.980	4.568	2.568, 2.549		7.126	$\epsilon\text{H}7$ 3.86, $\text{ZH}7$ 2.85
Cys21	9.200	4.849	2.755, 2.794			
Lys22	8.748	4.664	1.823	1.243, 1.621,	1.435, 1.769	$\epsilon\text{H}2$ 8.96, 2.956
Glu23	8.827	4.593	1.845, 1.948	2.138		
Lys24	8.908	4.528	1.075, 1.622	1.233, 1.117	1.581,	$\epsilon\text{H}2$ 8.45
Thr25	8.188	5.188	3.867	1.207		
Asp26	9.137	4.807	2.770, 3.341			
Gln27	8.712	4.111	1.768, 1.908	1.512, 1.908		$\epsilon\text{H}6$ 8.56, 7.300
Tyr28	8.344	4.672	3.015, 3.387		7.192	$\epsilon\text{H}6$ 8.21
Gly29	8.070	3.571, 4.265				
Leu30	8.486	4.582	2.054, 1.647	0.893, 1.002		
Thr31	8.366	5.249, 3.993	1.133			
Tyr32	8.930	4.906	3.029, 3.162		6.996	$\epsilon\text{H}6$ 6.23
Arg33	8.440	5.202	1.725, 1.940	1.232, 1.605	3.076, 3.139	$\epsilon\text{H}7$ 1.50
Thr34	8.139	5.223	3.861	1.090		
Cys35	8.832	5.208	3.005			
Asn36	10.012	5.104	2.669, 3.063		7.153, 8.366	
Leu37	8.494	4.478	1.724, 1.841		0.978, 1.037	
Leu38	7.593	4.668	1.451, 1.490	1.564	0.905, 0.927	
Pro39	-	3.863	1.453, 1.619	1.189, 1.545	3.195, 2.464	

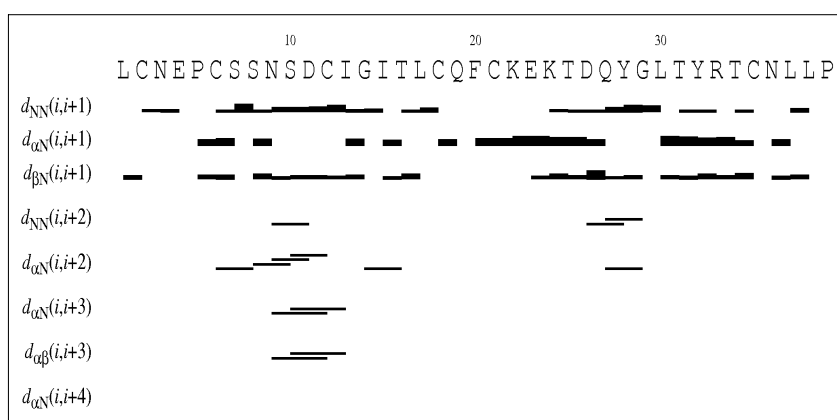
### 2.1.2 Generación de las Estructuras 3D del Péptido del Tomate, mediante el Programa CYANA (versión 2.1).

Una vez realizada la asignación de desplazamientos químicos en el espectro NOESY, se seleccionó, manualmente, la posición de todos los picos de cruce. A continuación, y mediante el empleo del módulo externo *peakint* del programa XEASY, se realizó la integración automática de los picos de cruce seleccionados (1314 picos de cruce, tabla 3.2). Los archivos con las listas de desplazamientos químicos, y de los picos de cruce con sus respectivos volúmenes, se usaron para la determinación de la estructura tridimensional, mediante el uso del programa CYANA,<sup>71</sup> que está basado en DYANA y CANDID. Como se ha mencionado, CYANA lleva a cabo un protocolo combinado de enfriamiento simulado y dinámica molecular en el espacio multidimensional de los ángulos de torsión. Además, permite confirmar la asignación de los picos NOE identificados y realiza una asignación automática de aquellos que no están todavía identificados de manera no ambigua. En nuestro caso, también se incluyeron restricciones superiores e inferiores de distancias para los puentes disulfuro S-S, tal que  $2.0 \text{ \AA} < R_{S-S} < 2.1 \text{ \AA}$  y  $3.0 \text{ \AA} < r_{C\beta-S} < 3.1 \text{ \AA}$ . En primer lugar, los pares S-S se establecieron de acuerdo con los existentes para el inhibidor de carboxipeptidasa de la patata.<sup>104</sup> El protocolo de CYANA realiza, de manera iterativa, siete ciclos de ajuste. En cada uno genera 100 estructuras iniciales que, de acuerdo con sus criterios de evaluación, basados en filtros y en restricciones de distancia ambiguas (*network anchoring*) y en combinaciones de restricciones (*constraint combination*), produce 20 estructuras finales. En cada cálculo con CYANA se genera una tabla (*cyanatable*) que incluye un resumen con la cantidad y el tipo de restricciones utilizadas en cada ciclo. En la tabla 3.2 se indica el caso específico del cálculo de la estructura tridimensional del péptido aislado del fruto del tomate.

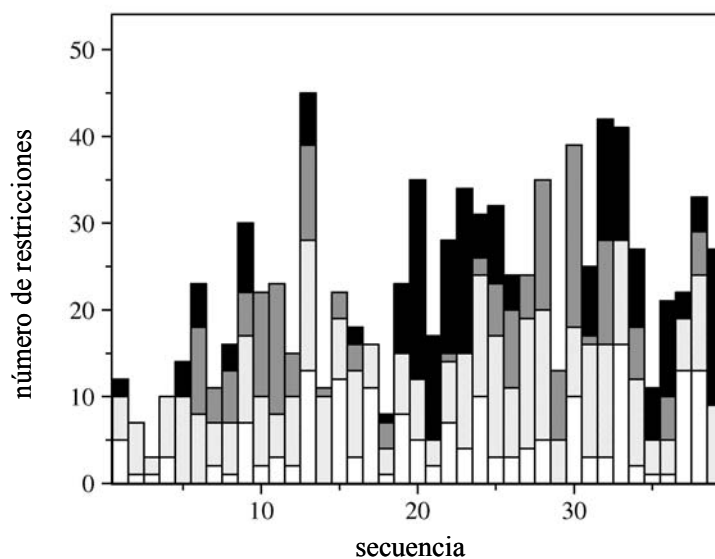
Ciclo	:	1	2	3	4	5	6	7
picos de cruce:								
seleccionados	:	1314	1314	1314	1314	1314	1314	1314
asignados	:	1290	1293	1276	1283	1285	1284	1286
no-asignados	:	24	21	38	31	29	30	28
asignaciones de la diagonal	:	0	0	0	0	0	0	0
picos de cruce :								
asignados fuera de la diagonal	:	1290	1293	1276	1283	1285	1284	1286
asignaciones únicas	:	626	1039	1114	1128	1144	1172	1182
asignaciones a corto rango $ i-j  \leq 1$	:	1052	1048	1042	1041	1039	1033	1033
asignaciones a medio rango $1 <  i-j  < 5$	:	136	119	115	116	119	124	124
asignaciones a largo rango $ i-j  \geq 5$	:	102	126	119	126	127	127	129
comparación con las asignaciones de referencia:								
picos de cruce asig. de referencia	:	850	850	850	850	850	850	850
idénticos con las asig.de referencia	:	472	695	738	748	760	773	777
compatibles con las asig. de referencia	:	837	835	830	832	833	828	828
incompatibles con las asig. de referencia	:	0	3	3	7	7	12	12
asign. adicionales de referencia	:	13	12	17	11	10	10	10
asign. adicionales	:	453	455	443	444	445	444	446
límites superiores de distancia:								
total	:	611	533	514	515	517	516	521 535
corto-rango, $ i-j  \leq 1$	:	429	371	360	354	349	347	335 343
medio-rango, $1 <  i-j  < 5$	:	149	112	73	73	77	78	77 77
largo-rango, $ i-j  \geq 5$	:	33	50	81	88	91	91	109 115
promedio de asig. /restricciones	:	2.92	1.63	1.17	1.16	1.14	1.11	1.00 1.00
valor de función diana promedio	:	4.54	2.68	1.56	0.18	0.08	0.07	0.05 0.02
RMSD (residuos 1..39):								
RMSD de cadena principal	:	3.44	3.85	1.30	0.91	0.95	0.99	0.83 0.85
RMSD de átomos pesados	:	3.85	4.23	1.66	1.35	1.35	1.38	1.20 1.24

**Tabla 3.1** Archivo de salida del cálculo automatizado con CYANA (“cyanatable”). La tabla indica, de manera resumida, el tipo y número de restricciones utilizadas en cada ciclo, en este caso para el cálculo de la estructura 3D del péptido del tomate.

En la figura 3.6 se muestran, esquemáticamente, las restricciones utilizadas para generar el conjunto definitivo de los 20 conformeros del péptido. Por un lado, en el inciso A, se representan de las conectividades secuenciales utilizadas para el cálculo de la estructura 3D. Aquí se observa que las interacciones  $d_{NN}(i,i+1)$  se interrumpen en la prolina 5. Además, se observan las conexiones  $d_{\alpha N}(i,i+1)$  del segmento 20-27. Asimismo, en el inciso B, se muestra gráficamente la distribución de las 535 restricciones utilizadas. De ellas, 343 son secuenciales, y 77 y 115, de media y larga distancia, respectivamente (tabla 3.2).



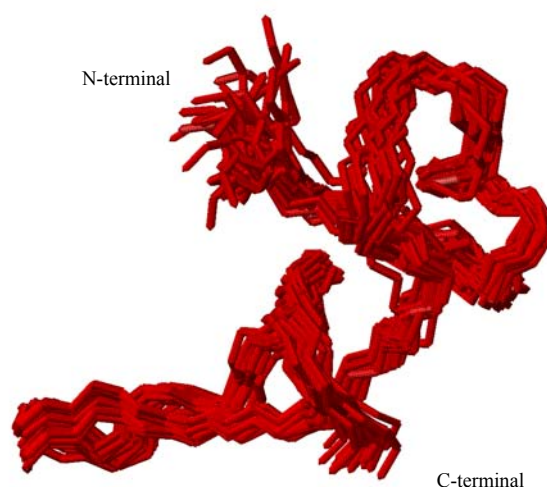
A)



B)

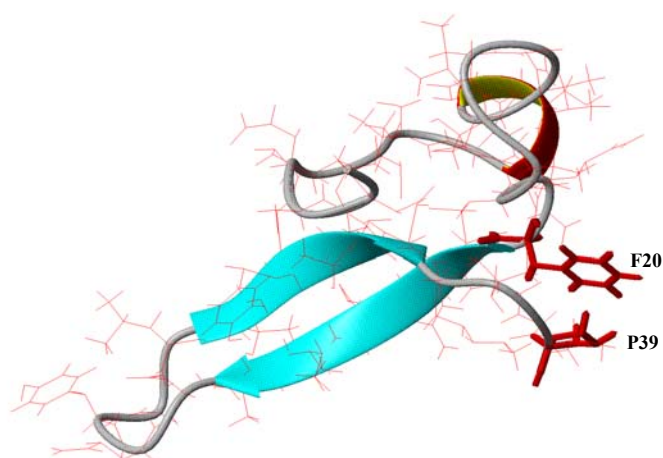
**Figura 3.6** A) Conectividad secuencial para el péptido del tomate, de acuerdo con el tipo de NOE entre los protones NH, H $\alpha$ , y H $\beta$ . B) Representación gráfica de las restricciones secuenciales, a media y a larga distancia.

La superposición de las 20 estructuras CYANA, entre los residuos 1–39, se muestra en la figura 3.7. El valor de RMSD es de 0.85 Å para la cadena principal y de 1.24 Å para los átomos pesados.



**Figura 3.7** Superposición de la cadena principal de las 20 estructuras de CYANA del péptido del tomate. Se indican los extremos N- y C terminales.

Con relación a la estructura secundaria, figura 3.8, existen dos láminas- $\beta$  antiparalelas definidas por la presencia de NOEs intensos, secuenciales del tipo  $d_{\alpha N}(i, i+1)$  y a larga distancia de media intensidad del tipo  $d_{\alpha N}$  y  $d_{\alpha\alpha}(i, j)$ . Además, existen interacciones de apilamiento CH- $\pi$  entre la Phe20 y la Pro39. Cada una de las dos láminas  $\beta$  está formada por 6 aminoácidos: Phe20-Cys21-Lys22-Glu23-Lys24-Thr25 y Thr31-Tyr32-Arg33-Thr34-Cys35-Asn36. Las interacciones  $d_{\alpha(i)-N(j)}$  son Lys22-Cys35, Glu23-Thr34, Lys24-Arg33 y Thr25-Tyr32. Las  $d_{\alpha(i)-\alpha(j)}$  son Cys21-Cys35 y Glu23-Arg33. También existe una hélice  $\alpha$  en el segmento Asn9-Ser10-Asp11-Cys12. Está definida por los NOEs  $d_{NN}(i, i+1)$  Asp9-Ser10 y otros de tipo  $d_{\alpha N}(i, i+2)$  y  $(i, i+3)$  entre Asn9-Asp11 y Asn9-Cys12, respectivamente.



**Figura 3.8** Representación de las dos láminas- $\beta$  antiparalelas y la hélice  $\alpha$  corta que conforman la estructura secundaria del péptido del tomate. En color rojo, puede observarse la interacción de apilamiento entre el anillo aromático de la Phe20 y los protones de la cara superior de la Pro39.

### 2.1.3 Refinamiento de la Estructura 3D del Péptido del Tomate mediante el Programa AMBER 8.0

El conjunto de 20 estructuras derivadas de CYANA fue sometido a una minimización energética, mediante un proceso de gradientes conjugados con dinámica molecular restringida (rMD) en disolvente explícito. Este protocolo se llevó a cabo con el campo de fuerzas AMBER versión 8.0.<sup>78</sup> Previamente, se construyó una caja con moléculas de agua con el modelo TIP3PBOX, también conteniendo átomos de sodio para neutralizar la carga neta del péptido se adicionó un contraión. La caja se escogió con unas dimensiones de 10.0 Å alrededor de cada estructura del péptido, mediante el uso del módulo Xleap de AMBER. La minimización energética se llevó a cabo con el módulo SANDER de AMBER y consistió de tres pasos:



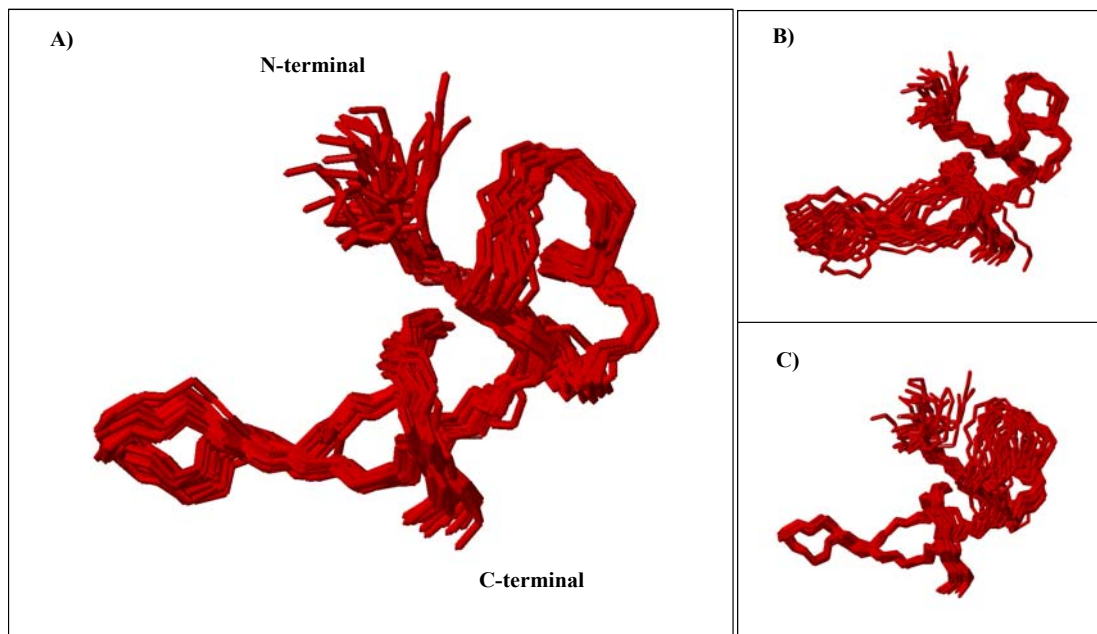
1.- Se eliminaron los malos contactos de van der Waals de las moléculas de agua con el péptido. En este paso, se usó una constante de fuerza de 100 kcal/mol para restringir la posición de los átomos de todos los residuos del péptido.

2.- Minimización restringida de las moléculas de agua: En este paso se relajaron las posiciones de las moléculas de agua, manteniendo restringidas las posiciones de los átomos del péptido.

3.- Minimización restringida de todo el sistema, usando 3000 iteraciones de gradiente conjugado.

Durante todo el proceso se aplicó un corte de 8 Å para computar las interacciones de no enlace, y el volumen de la caja se mantuvo constante.

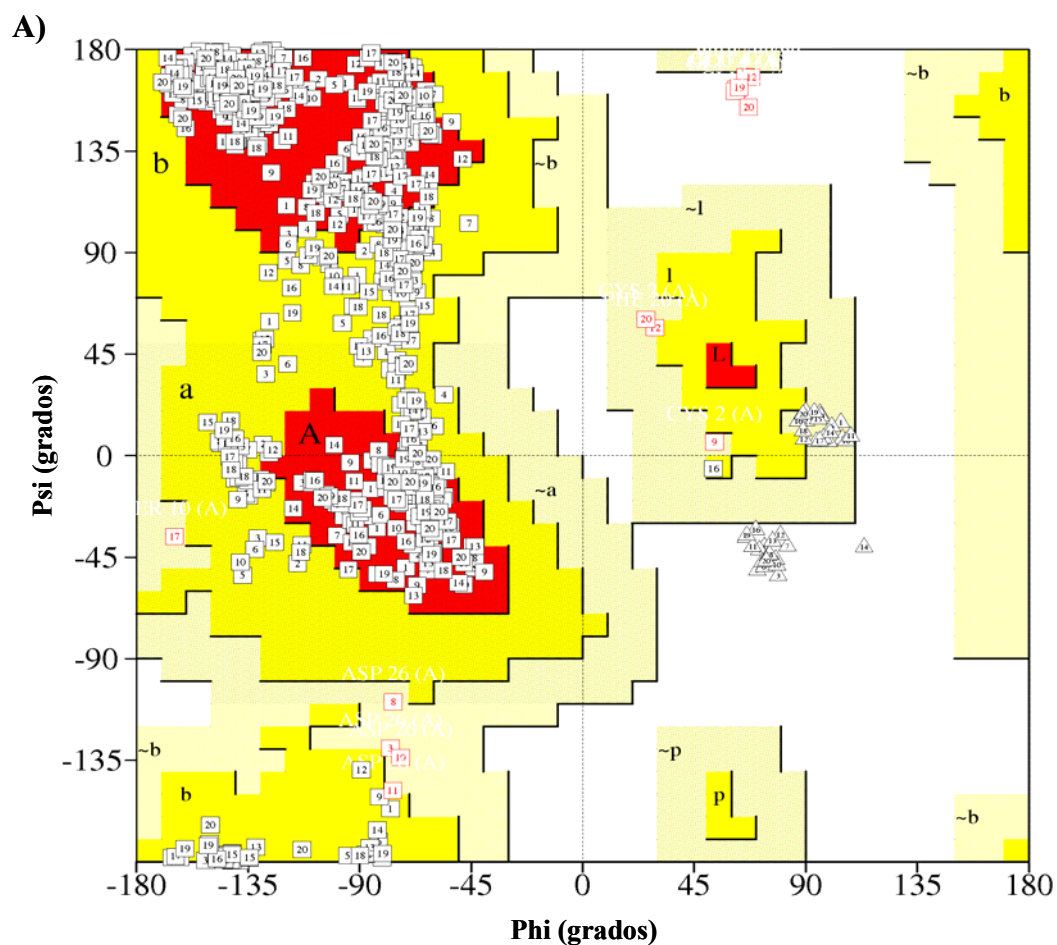
En la figura 3.9 A se representan las 20 estructuras generadas de la minimización energética restringida en disolvente explícito. El RMSD entre los residuos 2-39 de los átomos de la cadena principal fue de 0.69 Å y el de los átomos pesados, de 1.09 Å. Si se comparan estos datos con los correspondientes a la estructuras de CYANA (figura, 3.7), se observa una mejor definición. Cabe mencionar que la cadena principal entre los residuos 20-39 (RMSD 0.44 Å, figura 3.9 B) presenta una mejor convergencia que el segmento comprendido entre los residuos 2-18 (RMSD 0.50 Å, figura 3.9 C). En este sentido, se pensó que la conexión entre ambos segmentos podría ser flexible, por lo que se procedió a realizar una dinámica molecular usando moléculas de agua explícitas, como se describe en el apartado 2.1.4 de este Capítulo.



**Figura 3.9** Superposición de las 20 estructuras obtenidas de AMBER:

A) RMSD entre 2-39: cadena principal,  $0.69 \pm 0.17 \text{ \AA}$ , y átomos pesados,  $1.09 \pm 0.14 \text{ \AA}$ . B) RMSD entre 2-18: cadena principal,  $0.50 \pm 0.12 \text{ \AA}$ , y átomos pesados,  $0.95 \pm 0.13 \text{ \AA}$ . C) RMSD para 20-39: cadena principal,  $0.44 \pm 0.14 \text{ \AA}$  y átomos pesados,  $0.92 \pm 0.11 \text{ \AA}$ .

El mapa de Ramachandran (figura 3.10. A)) permitió evaluar la bondad de las 20 estructuras refinadas. Puede observarse que prácticamente el 100 % de los residuos se localizan en regiones permitidas (Figura 3.10 B)). El mapa y la tabla correspondiente se generaron mediante el método de validación de la base de datos del *Protein Data Bank*: <http://www.rcsb.org>). Solamente las glicinas (triángulos), que no poseen sustituyentes  $\beta$ , presentan mayor libertad conformacional y pueden situarse en diferentes regiones del mapa.

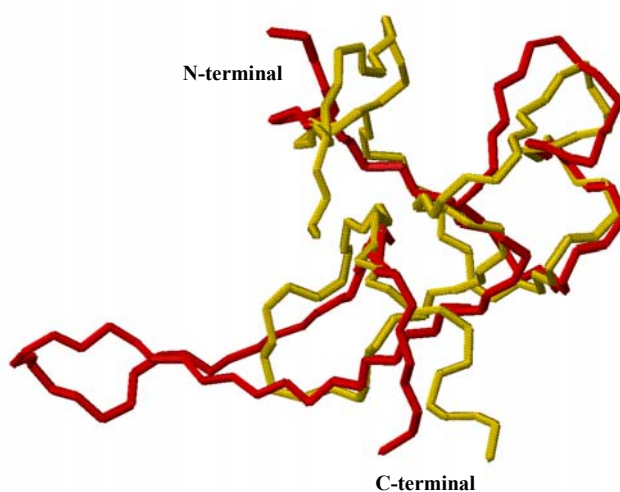


B)

Residuos en regiones más favorecidas	469	69.0 %
Residuos en regiones adicionales permitidas	197	29.0 %
Residuos en regiones generosamente permitidas	8	1.2 %
Residuos en regiones no permitidas	6	0.9 %
Número total de residuos (no-Gly, no-Pro)	680	100 %
Número de residuos finales (excl. Gly y Pro)	20	
Número de residuos glicinas (se muestran con triángulos)	40	
Número de residuos prolina	40	
Número total de residuos	780	

Figura 3.10 A) Mapa de Ramachandran de los residuos de las 20 estructuras refinadas con AMBER. B) Estadísticas de localización en función de los ángulos Phi y Psi.

El análisis comparativo de las estructuras del péptido del tomate derivadas de AMBER con las estructuras del inhibidor de metalocarboxipeptidasa (MCPI) aislado de la patata<sup>104</sup> (código PDB 1H2O, figura 3.10a) permitió corroborar la homología entre ambos péptidos, ya que el plegamiento de ambos péptidos es similar. La similitud con la heveína es nula.



**Figura 3.10a** Superposición de la cadena principal (residuos 2-39) de una de las estructuras del péptido del tomate con una de las estructuras del de la patata (MCPI).

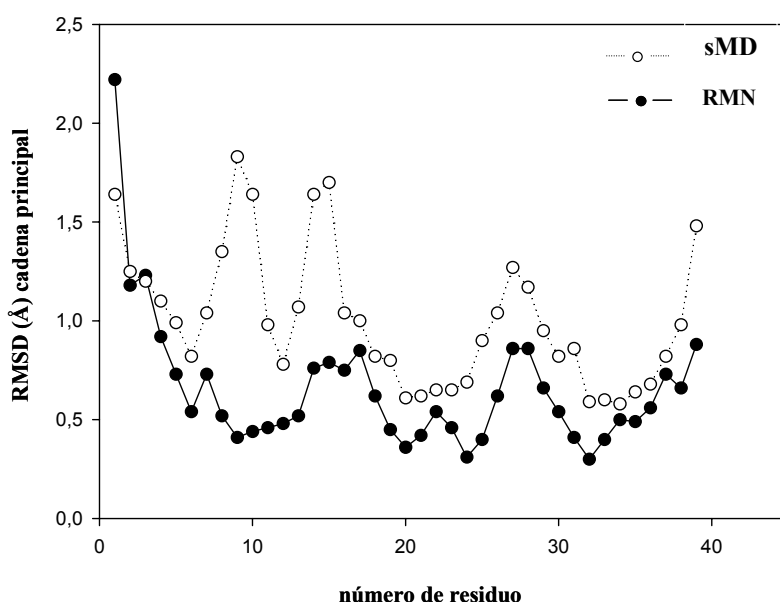
#### 2.1.4 Simulación con Dinámica Molecular en disolvente explícito (sMD) del Péptido del Tomate.

Para evaluar, al menos cualitativamente, la flexibilidad del péptido, se llevaron a cabo 4.5 ns de dinámica molecular solvatada (sMD) con el campo de fuerzas AMBER, usando condiciones equivalentes a las previamente descritas en la parte experimental (apartado 3.3) del capítulo 1 de esta Memoria.

La desviación cuadrática media (RMSD) de la cadena principal de 49 estructuras de la trayectoria de la dinámica molecular solvatada (sMD) se compararon con los RMSD de las 20 estructuras de RMN (figura 3.11). Puede observarse que en la sMD no existe ningún residuo en la zona 16-20 con una desviación significativa que indique una conectividad flexible entre los segmentos N-terminal (2-16) y C-terminal (20-39). Sin

embargo, sí se observa una mayor fluctuación en el extremo N-terminal, que en la región C-terminal. De hecho, en esta se incluyen las dos láminas  $\beta$  antiparalelas y la interacción de apilamiento CH- $\pi$  entre los residuos Phe20 y Pro39, mencionadas anteriormente.

En relación las estructuras 3D de RMN, los residuos del segmento N-terminal (2-18) presentan una convergencia ligeramente inferior que los del segmento C-terminal (20-39). De hecho, en el número de restricciones del N-terminal (figura 3.6 B), se observa una deficiencia de restricciones a larga distancia.



**Figura 3.11** Representación gráfica del RMSD de la cadena principal de las 49 estructuras de la dinámica solvatada (sMD) y de las 20 estructuras de RMN.

## 2.2 Análisis de la Interacción del Péptido del Tomate con Quitotriosa, Quitohexosa y Quitina, utilizando RMN.

Como se mencionó en la introducción de este capítulo, la purificación de la lectina del jugo del tomate se consiguió mediante cromatografía de afinidad sobre quitina cruda. Por lo tanto, se procedió a realizar el análisis de su interacción con quitotriosa (GlcNAc)<sub>3</sub> y con quitohexosa (GlcNAc)<sub>6</sub>. EL protocolo que se siguió fue mediante valoraciones por RMN, como se describió en los capítulos 1 y 2 de esta Memoria. Sin

embargo, no se observó ninguna perturbación en los desplazamientos químicos de las señales del péptido con la presencia de ningún quitoooligosacárido, ni tampoco con el polímero quitina, indicando de nuevo, como se había observado de los datos estructurales que péptido del tomate, aún conteniendo 3 puentes disulfuro anudados, no es un dominio de heveína.

### 3. Parte Experimental

#### 3.1 Experimentos de RMN-2D

El péptido del tomate (2.5 mg) se disolvió en 500  $\mu\text{l}$  de  $\text{H}_2\text{O}$ , 50  $\mu\text{l}$  de  $\text{D}_2\text{O}$ , y 2  $\mu\text{l}$  de una solución de  $\text{H}_2\text{O}/10\%$  HCl. Por lo tanto, las condiciones de la muestra fueron de una concentración de péptido de 1 mM y un pH 4.5.

Los espectros TOCSY y NOESY se obtuvieron en un equipo Bruker Avance 500 MHz, con secuencias estándar y con tiempos *spin lock* de 30 y 60 ms para el TOCSY y de 300 ms para el NOESY, a 298 K.

Además se registró un NOESY en un equipo Bruker 800 MHz, con un tiempo de mezcla de 150 ms a 298 K.

#### 3.2 Cálculo de geometría de distancias con el programa CYANA y Minimización Energética Restringida (REM) en disolvente Explícito.

Las lista de picos de cruce NOE seleccionados y el valor de sus intensidades y de sus desplazamientos químicos se generó usando el programa XEASY. Los archivos con las restricciones de distancia se completaron con otras para los puentes disulfuro S-S ( $2.0 \text{ \AA} < r < 2.1 \text{ \AA}$  y C $\beta$ -S:  $3.0 \text{ \AA} < r < 3.1 \text{ \AA}$ ). Estos fueron los archivos de entrada principales para realizar el cálculo con CYANA. La tolerancia de las coordenadas de los picos de cruce en la dimensión indirecta fue de 0.020 ppm, y de 0.015 ppm para la dimensión directa. El número de pasos durante el cálculo fue de 10000. Las estructuras iniciales para cada ciclo fue de 100, generándose 20 finales.

Posteriormente, las 20 estructuras CYANA se sometieron a un protocolo de minimización restringida en una caja con moléculas de agua explícitas, usando el método de gradiente conjugado, con el campo de fuerzas AMBER versión 8.0.

Con el modulo Xleap de AMBER se preparó la caja de disolvente con el comando TPI3PBOX, con una dimensión de 10 Å sobre el tamaño del péptido, neutralizándose el sistema con un contraiones Na<sup>+</sup>. Posteriormente, cada una de las 20 estructuras se sometió a un protocolo de minimización energética con el campo de fuerzas de AMBER 8.0. En este caso, se usó el parámetro *imin*=1, que procede de la siguiente manera:

Inicialmente, con el objeto de eliminar los malos contactos existentes entre las moléculas de agua y el péptido, se llevó a cabo una minimización (parámetro *nmropt* = 0), que indica que se mantiene fija la posición de los átomos del péptido, utilizando una constante de fuerza de 100 kcal/mol, un volumen constante y una constante dieléctrica de 1.0. Ésta minimización se llevó a cabo con 500 pasos (*maxcyc*=500), con gradientes conjugados.

Posteriormente, se procedió a otra minimización con las mismas condiciones que la anterior, aunque en este caso se permitió la relajación de la posición de las moléculas de agua alrededor del soluto (parámetro *nmropt* = 1) y se consideraron las restricciones de distancia experimentales.

Finalmente, se realizó la minimización energética restringida (*nmropt* = 1), considerando las moléculas del disolvente y el péptido, usando un mayor número de pasos durante la minimización (*maxcyc* = 3000).

#### 4. Conclusiones

1.- Se ha determinado la estructura tridimensional de un péptido aislado del fruto del tomate, utilizando RMN y cálculos de dinámica y mecánica molecular.

2.- Este péptido del tomate tiene homología y similitud estructural con el inhibidor de carboxipeptidasa aislado de la patata (MCPI). El análisis comparativo de las estructuras de ambos péptidos permitió observar que la topología conformacional en las regiones homólogas es similar.

3.- Los resultados de las valoraciones, por RMN, de la asociación del péptido, con (GlcNAc)<sub>3</sub>, (GlcNAc)<sub>6</sub>, y el polímero quitina, indicaron que el péptido en estudio no tiene la capacidad de reconocer derivados de N-acetilglucosamina. Por lo tanto el péptido del tomate no puede clasificarse como un dominio de heveína.



## ***CONCLUSIONES GENERALES***



En esta memoria se han descrito los resultados de los estudios sobre el rol que juegan tanto las interacciones entre residuos aromáticos como los enlaces de hidrógeno en el reconocimiento molecular carbohidrato-proteína. Nuestros estudios se han llevado a cabo mediante la aplicación de la Resonancia Magnética Nuclear, utilizando como modelos un dominio de heveína no natural AcAMP2 Pff18/Pff20 y una heveína truncada mutante Hev32S19D.

En el primer caso, que estructuralmente es un análogo del péptido natural AcAMP2, se han sustituido los aminoácidos Phe18 y Tyr20 por el aminoácido no natural *p*-fluorofenilalanina. El análisis de la estructura tridimensional del complejo con quitotriosa, obtenida mediante RMN, permitió concluir que la mutación de los dos residuos no altera el plegamiento y que la presencia del trisacárido produce cambios en la orientación y libertad conformacional de los residuos aromáticos Pff18 y Pff20, involucrados en el sitio de unión.

Asimismo, la estimación de los valores de la constante de afinidad y parámetros termodinámicos ( $\Delta H^{\circ}$  y  $\Delta S^{\circ}$ ) mediante valoraciones con RMN, y su comparación con los obtenidos para otros dominios de heveína y AcAMP2 natural, permitieron concluir que la presencia del flúor sobre el anillo aromático del residuo del sitio de unión, disminuye la densidad electrónica del mismo y reduce la afinidad y la entalpía de unión con el carbohidrato.

En el péptido de heveína truncada Hev32S19D se ha reemplazado el residuo Ser19 por uno de Asp, con la finalidad de analizar el valor energético y estructural que ejerce la formación del enlace de hidrógeno de la Ser19 con un hidroxilo del carbohidrato. El estudio se llevó mediante valoraciones de RMN, que proporcionaron los valores de la constante de afinidad, y los valores de entalpía y entropía de unión. Comparativamente con los datos obtenidos para los complejos de heveína naturales con Ser19, la heveína truncada muestra una afinidad y una entalpía menor para la interacción con el pentasacárido (GlcNAc)<sub>5</sub>. En todos los casos, en los procesos de reconocimiento molecular se observa un fenómeno de compensación entalpía-entropía.

Por otro lado, la asociación de un pentasacárido de quitina, mono-desacetilado en su extremo no reductor, con el péptido mutante Hev32S19D es mucho menor que para el

derivado penta-N-acetilado, lo que demuestra la importancia de la presencia del grupo acetamido para estabilizar el complejo.

Por lo tanto, puede concluirse que las interacciones de apilamiento CH- $\pi$  son gran valor y relevancia para la interacción proteína-carbohidrato. Asimismo, los enlaces de hidrógeno y las fuerzas de van der Waals contribuyen a la estabilidad y selectividad en los procesos de reconocimiento molecular de carbohidratos por proteínas.

Adicionalmente en esta memoria, mediante la aplicación de un método reciente de asignación automática de datos de RMN con CYANA, y el empleo subsiguiente del campo de fuerzas AMBER, se ha determinado la estructura tridimensional de un péptido (39 aa) aislado del fruto del tomate. Su plegamiento es homólogo al del inhibidor de metalocarboxipeptidasa (MPCI), aislado de la patata. El análisis de afinidad con oligómeros de *N*-acetilglucosamina permitió deducir que éste péptido no es un dominio de heveína.

## ***BIBLIOGRAFÍA***



1. Gabius, H. J., Siebert, H. C., André, S., Jiménez-Barbero, J., Rüdiger, H. *Chem. BioChem.* **2004**, 5, 740-764.
2. Dwek, R.A. *Chem. Rev.* **1996**, 96, 683-720.
3. Montreuil, J., Vliegthart, J. & Schachter, H. *Glycoproteins*. Elsevier, Amsterdam. eds. **1995**.
4. Solís, D., Jiménez-Barbero, J., Kaltner, H., Romero, A., Siebert, H. C., von der Lieth, C. W., Gabius, H. J. *Cells Tissues Organs.* **2001**, 168, 5-23.
5. Protein-Carbohydrate Interactions and Infectious Diseases, The Royal Society of Chemistry **2005**, Ed. Carole Bewley.
6. Lis H, Sharon N. *Chem. Rev.* **1998**; 98, 637-674.
7. Kaltner, H. & Stierstorfer, B. *Acta Anat.* **1998**, 161, 162-179.
8. Ko TP, Ng JD, McPherson, A. *Plant Physiol.* **1993**, 101, 729-44
9. van Damme, E. J. M., Peumans, W. J., Pustazi, A., Bardocz, S. *Handbook of Plant Lectins: Properties and Biomedical Applications*, 1998, Ed. Wiley-VCH.
10. Van Damme, J. M., Peumans, W. J., Barre, A., Rougé, P. *Critical Reviews in Plant Sciences* **1998**, 17, 575-692.
11. Raikhel, N.V., Lee, H.I. & Broekaert, W.F. *Annu. Rev. Plant Physiol. Plant Mol. Biol.* **1993**, 44, 591-615.
12. Beintema J.J. *FEBS Lett.* **1994**, 350, 159-163.
13. Rodríguez-Romero A, Ravichandran K. G, Soriano-Garcia M. *FEBS Lett.* **1991**, 291, 307-309.
14. Wagner, S., Breitneder, H. *Biochem. Soc. Trans.* **2002**, 30, 935-940.
15. Loris, R., Hamelryck, T., Bouckaert, J. & Wyns, L. *Biochim Biophys. Acta* **1998**, 1383, 9-36.
16. Wright, C.S. *J. Mol. Biol.* **1984**, 178, 91-104.
17. Kronis, K.A. & Carver, J.P. *Biochemistry* **1985**, 24, 834-840.
18. Gidrol, X., Chrestin, H., Tan H-L., Kush A. *J. Biol. Chem.* **1994**, 209, 9278-9283.
19. Dam, T. K., Brewer, F. *Chem. Rev.* **2002**, 102, 387-429
20. von der Lieth, C., Siebert, H., Kozar, T., Burchet, M., Frank, M., Gilleron, M., Kaltner, H., Kayser, G, Tajkhorshid, E., Bovin, N. V., Vliegthart, J. F., Gabius,

- H. J. *Acta Anat. (Basel)*. **1998**, 161, 91-109
21. Woods, R. *Glycoconj. J.* **1998**, 15, 209-216
22. Jimenez-Barbero, J., Asensio, J. L., Cañada, F. J., Poveda, A., *Curr. Opin. Struct. Biol.* **1999**, 9, 549-555
23. Imberty, A., Perez, S. *Chem. Rev.* **2000**, 100, 4567-4588
24. Bewley, C. A. *Structure Fol. Des.* **2001**, 9, 931-940
25. Wolmald, M. R., Petrescu, A. J., Pao, Y. L., Gilthero, A., Elliott, T., Dwek, R. A., *Chem. Rev.* **2002**, 102, 371-386
26. Neuman, D., Lehr, C-M., Lenhof, H-P., Kohlbacher, O. *Advanced Drug Delivery Reviews* **2004**, 56, 437-457
27. Puri, K. D., Surolia, A. *Pure & Appl. Chem.* **1994**, 66, 497-502
28. Kogelberg, H., Solís, D., Jiménez-Barbero, J. *Curr Opin Struct Biol.* **2003**, 13, 646-653.
29. Poveda, A., Jiménez-Barbero, J. *Chem. Soc. Rev.* **1998**, 27, 133-143.
30. a) Quioco, F. A. *Pure & Appl. Chem.* **1989**, 61, 1293-1306.  
b) Vyas, N. K. *Curr. Opin. Struct. Biol.* **1991** 1, 732-740
31. Muraki, M. *Protein Pept. Lett.* **2002**, 9, 195-209
32. Toone, E. J. *Curr. Opin. Struct. Biol.* **1994**, 4, 719-728
33. a) Nishio, M., Umezama, Y., Hirota, M., Takeuchi, Y. *Tetrahedron* **1995**, 51, 8665-8671. b) Thomas, A. Meurisse, R., Charloteaux, B. Brausser, R. *Proteins: Structure, Function, and Genetics* **2002**, 48, 628-634
34. Siebert, H. C., Jiménez-Barbero, J., Gabius, H. J. *Methods Enzymol.* **2003**, 362, 417- 434.
35. Lemieux, R. U. *Acc. Chem. Res.* **1996**, 29, 373-380
36. Carver, J. P. *Pure Appl. Chem.* **1993**, 65, 763-770
37. Carver, J. P., Michnik, S. W., Imberty, A., Cumming, D. A. *Ciba Found. Symp.* **1989**, 145, 6-26
38. Jiménez-Barbero, J. & Peters, T. NMR of Glycoconjugates. **2002**, Eds. Weinheim, Wiley-VCH
39. Ford, M. G., Weimar, T., Kohli, T. Woods, R.J. *Proteins: Structure, Function Genetics* **2003**, 53, 229-240.



40. Sagui, C. Darden, T.A. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, 28, 155-179
41. Wright, C. S. *J. Biol. Chem.* **1992**, 267, 14345-14352
42. Espinosa, J. F, Asensio, J. L, Garcia, J. L, Laynez, J., Bruix, M., Wright, C., Siebert, H.C., Gabius, H. J., Canada, F. J, Jimenez-Barbero, J. *Eur J Biochem.* **2000**; 267, 3965-3978.
43. Saul, F. A, Rovira, P., Boulot, G., Damme, E. J., Peumans, W. J., Truffa-Bachi, P. Bentley G. A. *Structure Fold Des.* **2000**, 8, 593-603.
44. Harata, K., Muraki, M. *J. Mol Biol.* **2000**, 297, 673-681.
45. Andersen, N. H., Cao, B., Rodriguez-Romero, A., Arreguin, B. *Biochemistry.* **1993**, 32, 1407-1422.
46. Asensio, J. L., Canada, F. J., Bruix, M., Rodriguez-Romero, A. Jimenez-Barbero, J. *Eur J Biochem.* **1995**, 230, 621-633.
47. Asensio, J.L., Canada, F.J., Bruix, M., Gonzalez, C., Khair, N., Rodriguez-Romero, A. Jimenez-Barbero, J. *Glycobiology* **1998**, 8, 569-577.
48. Asensio, J. L., Canada, F. J., Siebert, H. C., Laynez, J., Poveda, A., Nieto, P. M., Soedjanaamadja, U. M., Gabius, H.J, Jimenez-Barbero, J. *J. Chem Biol.* **2000**, 7, 529-543.
49. Aboitiz, N. Vila-Perelló, M., Groves, P., Asensio, J. L., Andreu, D., Cañada, F. J., Jiménez-Barbero, J. *ChemBioChem* **2004**, 5, 1-12
50. Camelia Mihai, Tesis Doctoral, Vrije Univetsität Bruselas, **2004**, dirigida por J. Martins.
51. Asensio, J. L., Siebert, H. C., von Der Lieth, C. W., Laynez, J., Bruix, M., Soedjanaamadja, U. M., Beintema, J. J., Canada, F. J., Gabius, H. J., Jimenez-Barbero, J. *Proteins.* **2000**, 40, 218-236.
52. Hayashida, M., Fujii, T., Hamasu, M., Ishiguro, M., Hata, Y. *J. Mol. Biol.* **2003**, 334, 551-565.
53. Huang, R. H., Xiang, Y., Tu, G. Z., Zhang, Y. Wang, D. C. *Biochemistry* **2004**, 43, 6005-6012
54. Martins, J.C, Maes, D., Loris, R., Pepermans, H. A, Wyns, L., Willem, R., Verheyden, P. N. *J Mol Biol.* **1996**, 258, 322-333.

55. Archer, B. L. *Biochem. J.* **1960**, 75, 236-240
56. Karisola, P., Mikkola, J., Kalkkinen, N., Airenne, K. J., Laitinen, O. H., Repo, S., Pentikainen, O. T., Reunaka, T., Turjanmaa, K., Johnson, M.S., Palouso, T., Kulomaa, M. S., Alenius, H. *J. Immunol.* **2004**, 172, 2621-2628
57. Muraki, M. Morii, H., Harata, K. *Protein Pept. Lett.* **1998**, 5, 193-198
58. Muraki, M. Morii, H., Harata, K. *Protein Eng.* **2000**, 13, 385-389
59. Siebert, H. C., André, S., Asensio, J. L., Cañada, F. J., Dong, X., Espinosa, J. F., Frank, M., Gilleron, M., Kaltner, H., Kozár, T., Bovin, N. V., von der Lieth, C. W., Vliegthart, J. F., Jiménez-Barbero, J., Gabius, H. J. *Chembiochem.* **2000**, 1, 181-195
60. Chávez, M.I., Andreu, C., Vidal, P., Aboitiz, N., Freire, F., Groves, P., Asensio, J. L. Asensio, G., Muraki, M., Cañada, F. J., Jiménez-Barbero, J. *Chem. Eur. J.* **2005**, 11, 7060-7074
61. Wemmer, D. E., Reid, B. R. *Annu. Rev. Phys. Chem.* 1985, 36, 105-137
62. Wuthrich, K. *NMR of Proteins and Nucleic Acids* **1986** John Wiley&Sons. New York
63. Bax, A. and Grzeziek, S. *Accounts Chem. Res.* **1993**, 26, 131-138
64. Guntert, P. *Q. Rev. Biophys* 1998, 31, 145-237
65. Herrmann, T., Guntert, P. Wuthrich, K. *J. Biomol. NMR* **2002**, 24, 171-189
66. Wagner, G. *J. Biomol. NMR* **1993**, 3, 375-385
67. Kay, L. E., Gardner, K. H. *Curr. Opin. Struct. Biol.* **1997**, 7, 564-570
68. Guntert, P., Mumenthaler, C., Wuthrich, K. *J. Mol. Biol.* **1997**, 273, 283-298
69. Herrmann, T., Güntert, P. & Wüthrich, K. *J. Mol. Biol.* **2002**, 319, 209-227
70. Güntert, P. *Prog. NMR Spectrosc.* **2003**, 43, 105-125
71. Güntert, P. *Meth. Mol. Biol.* **2004**, 278, 353-378
72. a) Brooks, B. R., Karplus, M. *J. Chem. Phys.* **1983**, 79, 6312, b) Brooks, B. R., Burccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S., Karplus, M. *J. Comput. Chem.* **1983**, 4, 187
73. Maple, J. R. Hwang, M. J., Stockfish, T. P. Dinur, U., Waldham, M., Ewing, C. S., Hagler, A. T. *J. Comput. Chem.* **1994**, 15, 162
74. The GROMOS software package (by van Gunsteren, W. F. and Berendsen H. J. C.)

75. Pearlman, D. A., Case, D. A., Caldwell, J. W., Ross, Cheatham, T. E III, Debolt, S., Ferguson, D., Seibel, G. And Kollman, P. *Comp. Phys. Commun.* **1995**, 91, 1-41.
76. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. Weiner, P.J. *J. Am. Chem. Soc.* **1984** 106, 765-76.
77. Ponder J. W., Case, D. A. *Adv. Prot. Chem.* **2003**, 66, 27-85
78. Cheatham III, T. Darden, H. Gohlke, R. Luo, K.M. Merz, Jr., A. Onufriev, C. Simmerling, B. Wang and R. Woods. *J. Computat. Chem.* **2005**, 26, 1668-1688
79. Piantini, U., Sorensen, O. W., Ernst, R. R. *J. Am. Chem. Soc.* **1982**, 104, 6800-6801
80. a) Levitt, M., Freeman, R., Frenkiel, T. *J. Magn. Reson.* **1982**, 47, 238. b) Bax, A., Davis, D. *J. Magn. Reson.* **1985**, 65, 355.
81. a) Jeener, J., Meier, B. H. , Bachmann, P., Ernst, R. R. *J. Chem. Phys.* 1979, 71, 4546-4553. b) Wagner, R., Berger, S. *J. Magn. Reson.* **1996**, 123A, 119
82. Brunger, A. T. & Nigles, M. *Quart. Rev. Biophys.* **1993**, 26, 49-125
83. Havel, T. F. & Wüthrich, K. *Bull. Math. Biol.* 1984, 46, 673-698
84. Nilges, M., Core, G. M. & Gronenborn, A. M. *FEBS Letters* **1988**, 229, 317-324
85. Nilges, M., Kuszewski, J. & Brunger, A. T. Computational Aspects of the Study of Biological Macromolecules by Nuclear Magnetic Resonance Spectroscopy (Hoch, J. C. Poulsen, F. M. & Redfield, C., eds), **1991**, pp. 451-455, Plenum Press, New York
86. Guntert, P., Braun, W. & Wüthrich, K. *J. Mol. Biol.* **1991**, 217, 517-530
87. Evans, J. N. S. *Biomolecular NMR Spectroscopy* **1995** Oxford University Press
88. Mumenthaler, C., Braun, W. *J. Mol. Biol.* **1995**, 254, 465
89. Nilges, M., Macias, M. J., O'Donoghue, S. I., Oschkinat, H. *J. Mol. Biol.* **1997**, 269, 408
90. Brunger, A. T. X-PLOR versión 3.1 1993. A system for X-ray crystallography and NMR, Yale University Press
91. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nigles, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T., Warren, G. L. *Acta Crystallogr. D.* **1998**, 54, 905

92. Bartels, C. H., Guntert, P., Wüthrich, K. *J. Biomolecular NMR* **1995**, 5, 1-10
93. Johnson, B. A., Blevins, R. A. *Journal of Biomolecular NMR* **1994**, 4, 603-614
94. Kirpatrick, S., Gelatt Jr. C. D., Vecchi, M. P. *Science* **1983**, 220, 671-680
95. Darden, T., York, D., Perderson, L. *J. Chem. Phys.* **1993**, 98 10089-10092
96. Fielding, L. *Tetrahedron* **2000**, 56, 6151-170
97. a) Feeney, J., Batchelor, J. G., Albrand, J. P., Roberts, G. C. K. *J. Magn. Reson.* **1979**, 33, 519-529. b) Kronis, K. A., Carver, J. P. *Biochemistry* **1985**, 24, 834-840
98. Peters, P., Meyer, B. *Angew. Chem. Int. Ed.* **2003**, 43, 864-890
99. Bernardi, A., Arosio, D., Potenza, D., Sánchez-Medina, I., Mari, S., Cañada, F. J., Jiménez-Barbero, J. *Chem. Eur. J.* **2004**, 10, 4395-4406
100. Verheyden, P., Pletinckx, J., Maes, D., Pepermans, H. A.M., Wyns, L., Willem, R., Martins, J. C. *FEBS Letters*, **1995**, 370, 245-249
101. Frisch, M. J. et. al. Gaussian 98 (Revision A. 11.3), **1998** Gaussian, Inc., Pittsburgh. P. A.
102. Berendsen, H. J. C., Grigera, J. R., Straatsma, T. P. *J. Phys. Chem.* **1987**, 91, 6269-6271
103. Searle, M. S., Williams, D. H. J. *J. Am. Chem. Soc.* **1992**, 114, 10690-10697
104. González, C., Neira, J. L. Ventura, S., Bronsoms, S., Rico, M. Avilés., F. X. *Proteins: Structure, Function, and Genetics* **2003**, 50, 410-422

*APÉNDICE*



NOEs ASIGNADOS PARA EL PÉPTIDO NO NATURAL AcAMP2Pff18/Pff20

1VAL	HA	2	GLY	HN	2.50
1VAL	HA	3	GLU	HN	5.00
1VAL	HB	2	GLY	HN	5.50
1VAL	QG1	2	GLY	HN	7.70
1VAL	QG2	2	GLY	HN	7.70
2GLY	HN	3	GLU	HN	3.50
3GLU	HN	4	CYSS	HN	5.00
3GLU	HN	20	TYR	QD	7.70
3GLU	HA	4	CYSS	HN	3.00
3GLU	HA	20	TYR	HA	3.50
3GLU	HA	20	TYR	QD	5.70
3GLU	HB2	20	TYR	QD	7.70
3GLU	HB3	20	TYR	QD	7.70
4CYSS	HN	4	CYSS	HB2	3.50
4CYSS	HN	20	TYR	HA	3.50
4CYSS	HN	21	CYSS	HN	5.00
4CYSS	HA	5	VAL	HN	2.50
4CYSS	HB2	5	VAL	HN	3.50
4CYSS	HB2	7	GLY	HN	3.50
4CYSS	HB2	8	ARG	HN	5.00
4CYSS	HB3	5	VAL	HN	3.50
4CYSS	HB3	7	GLY	HN	5.00
4CYSS	HB3	8	ARG	HN	5.00
5VAL	HN	5	VAL	HB	2.50
5VAL	HN	6	ARG	HN	5.00
5VAL	HN	8	ARG	HN	3.00
5VAL	HA	6	ARG	HN	2.50
6ARG	HN	7	GLY	HN	2.50
6ARG	HA	7	GLY	HN	2.50
7GLY	HN	8	ARG	HB2	2.50
8ARG	HN	8	ARG	HB2	2.50
8ARG	HN	8	ARG	HB3	3.00
8ARG	HN	8	ARG	HG2	5.00
8ARG	HN	8	ARG	HG3	5.00
8ARG	HN	9	CYSS	HN	5.00
8ARG	HA	9	CYSS	HN	2.50
8ARG	HB3	9	CYSS	HN	3.00
8ARG	HG2	9	CYSS	HN	5.00
8ARG	HG3	9	CYSS	HN	5.00
9CYSS	HN	9	CYSS	HB2	3.00
9CYSS	HN	9	CYSS	HB3	3.50
9CYSS	HB2	10	PRO	HD2	5.00
9CYSS	HB2	10	PRO	HD3	5.50
9CYSS	HB3	10	PRO	HD2	5.50
9CYSS	HB3	10	PRO	HD3	5.50
10PRO	HA	11	SER	HN	2.50
10PRO	HB2	11	SER	HN	3.00
10PRO	HB3	11	SER	HN	3.50
11SER	HN	12	GLY	HN	5.50
11SER	HA	12	GLY	HN	2.50
11SER	HA	13	MET	HN	5.00

12GLY	HN	13	MET	HN	5.00
13MET	HN	13	MET	HB2	2.50
13MET	HA	14	CYSS	HN	2.50
13MET	HB2	14	CYSS	HN	3.50
13MET	HB2	22	GLY	HN	5.00
13MET	HB3	14	CYSS	HN	3.00
13MET	HB3	21	CYSS	HA	5.50
14CYSS	HN	14	CYSS	HB2	3.50
14CYSS	HN	21	CYSS	HA	3.50
14CYSS	HN	22	GLY	HN	3.00
14CYSS	HN	23	LYS	HA	5.00
14CYSS	HN	27	TYR	HB3	5.50
14CYSS	HA	15	CYSS	HN	2.50
14CYSS	HB2	15	CYSS	HN	4.50
14CYSS	HB2	22	GLY	HN	5.50
14CYSS	HB2	2	TYR	HB2	5.00
14CYSS	HB2	27	TYR	HB3	5.50
14CYSS	HB2	28	CYSS	HN	5.50
14CYSS	HB2	28	CYSS	HA	3.00
14CYSS	HB3	15	CYSS	HN	3.50
14CYSS	HB3	27	TYR	HB3	4.50
15CYSS	HN	15	CYSS	HB2	3.00
15CYSS	HN	15	CYSS	HB3	3.50
15CYSS	HN	16	SER	HN	5.50
15CYSS	HA	16	SER	HN	2.50
15CYSS	HA	21	CYSS	HA	3.50
15CYSS	HB2	16	SER	HN	5.00
15CYSS	HB3	16	SER	HN	5.50
16SER	HN	17	GLN	HN	5.00
16SER	HN	17	GLN	HA	5.00
16SER	HN	20	TYR	HB2	4.50
16SER	HN	21	CYSS	HA	4.50
16SER	HN	27	TYR	QD	7.00
16SER	HA	17	GLN	HN	3.00
16SER	HA	27	TYR	HA	5.50
16SER	HA	27	TYR	HB3	5.50
16SER	HA	27	TYR	QD	7.70
17GLN	HN	17	GLN	HB2	3.50
17GLN	HN	17	GLN	HB3	3.50
17GLN	HN	18	TYR	HN	4.50
17GLN	HA	19	GLY	HN	5.00
18TYR	HN	18	TYR	HB2	3.00
18TYR	HN	19	GLY	HN	3.00
18TYR	HN	20	TYR	HN	5.00
18TYR	HB2	19	GLY	HN	5.00
18TYR	HB2	20	TYR	QD	5.70
18TYR	HB2	27	TYR	QD	7.00
18TYR	HB3	20	TYR	QD	5.70
18TYR	QD	19	GLY	HN	7.70
19GLY	HN	20	TYR	HN	3.50
19GLY	HA1	20	TYR	HN	3.50



19GLY	HA2	20	TYR	HN	3.50
20TYR	HN	20	TYR	HB2	3.50
20TYR	HN	20	TYR	HB3	3.50
20TYR	HA	21	CYSS	HN	2.50
20TYR	HB2	21	CYSS	HN	5.00
20TYR	HB2	27	TYR	QD	7.70
20TYR	HB2	27	TYR	QE	5.70
20TYR	HB3	21	CYSS	HN	3.50
20TYR	HB3	27	TYR	QE	7.00
20TYR	QD	21	CYSS	HN	7.70
21CYSS	HA	22	GLY	HN	2.50
21CYSS	HA	27	TYR	QD	7.70
22GLY	HN	27	TYR	HB2	3.50
22GLY	HN	27	TYR	HB3	3.50
22GLY	HN	27	TYR	QD	7.70
22GLY	HA1	23	LYS	HN	3.50
22GLY	HA2	23	LYS	HN	3.00
23LYS	HN	23	LYS	HG2	5.00
23LYS	HN	23	LYS	HG3	5.00
23LYS	HA	24	GLY	HN	3.00
26LYS	HN	26	LYS	HB2	2.50
26LYS	HN	26	LYS	HB3	3.00
26LYS	HN	27	TYR	HN	3.50
26LYS	HA	28	CYSS	HN	5.00
26LYS	HB2	27	TYR	HN	5.50
26LYS	HB2	27	TYR	QD	5.70
26LYS	HB2	27	TYR	QE	7.70
26LYS	HB3	27	TYR	HN	5.50
26LYS	HB3	27	TYR	QD	7.70
27TYR	HN	27	TYR	HB2	3.00
27TYR	HN	28	CYSS	HN	2.50
27TYR	HB2	28	CYSS	HN	2.40
27TYR	HB2	28	CYSS	HA	5.00
27TYR	HB3	28	CYSS	HN	3.00
28CYSS	HN	28	CYSS	HB2	3.00
28CYSS	HN	28	CYSS	HB3	2.50
28CYSS	HN	29	GLY	HN	3.50
28CYSS	HN	30	ARG	HA	5.00
28CYSS	HB2	29	GLY	HN	5.50
28CYSS	HB3	29	GLY	HN	5.00
29GLY	HN	30	ARG	HA	5.00
30ARG	HN	30	ARG	HB2	5.00
4CYSS	SG	15	CYSS	SG	2.10
4CYSS	SG	15	CYSS	CB	3.10
4CYSS	CB	15	CYSS	CG	3.10
9CYSS	SG	21	CYSS	SG	2.10
9CYSS	SG	21	CYSS	CB	3.10

9CYSS	CB	21	CYSS	SG	3.10
14CYSS	SG	28	CYSS	SG	2.10
14CYSS	SG	28	CYSS	CB	3.10
14CYSS	CB	28	CYSS	SG	3.10

**Pseudoatomos que se introdujeron por el módulo (*distance modify*) de DYANA**

1VAL	QQG	2	GLY	HN	6.59
1VAL	QQG	20	TYR	QD	11.16
2GLY	QA	20	TYR	QD	8.28
3GLU	HN	3	GLU	QB	3.15
3GLU	HN	3	GLU	QG	6.18
3GLU	QB	4	CYSS	HN	4.18
3GLU	QB	20	TYR	HA	6.18
3GLU	QG	4	CYSS	HN	6.18
3GLU	QG	20	TYR	QD	8.28
4CYSS	HN	19	GLY	QA	6.18
4CYSS	HA	5	VAL	QQG	8.96
4CYSS	HB2	7	GLY	QA	6.08
4CYSS	HB3	7	GLY	QA	6.08
6ARG	HN	6	ARG	QG	4.08
8ARG	HN	8	ARG	QD	6.08
8ARG	QD	9	CYSS	HN	6.08
9CYSS	HA	10	PRO	QD	3.39
9CYSS	QB	10	PRO	QD	4.48
9CYSS	QB	13	MET	HN	6.08
9CYSS	QB	13	MET	HA	6.08
9CYSS	QB	13	MET	HB 2	3.08
9CYSS	QB	13	MET	QG	5.00
11SER	HN	11	SER	QB	3.04
11SER	QB	12	GLY	HN	4.08
13MET	HN	13	MET	QG	4.08
14CYSS	HN	21	CYSS	QB	6.08
16SER	HN	16	SER	QB	3.24
16SER	QB	17	GLN	HN	3.08
16SER	QB	18	TYR	HN	4.08
16SER	QB	18	TYR	QD	8.28
16SER	QB	27	TYR	HN	6.08
16SER	QB	27	TYR	HA	4.08
16SER	QB	27	TYR	HB3	6.08
16SER	QB	27	TYR	QD	5.28
16SER	QB	27	TYR	QE	8.28
17GLN	HN	17	GLN	QG	4.08
17GLN	HA	17	GLN	QG	3.51
17GLN	QG	18	TYR	HN	6.08
17GLN	QG	18	TYR	QD	8.28
20TYR	HA	21	CYSS	QB	6.08
21CYSS	QB	22	GLY	HN	4.08

22GLY	QA	23	LYS	HN	2.61
22GLY	QA	27	TYR	QD	8.28
23LYS	HN	23	LYS	QB	3.24
23LYS	HN	24	GLY	QA	6.08
23LYS	HA	23	LYS	QG	3.50
24GLY	QA	26	LYS	HN	6.08
26LYS	HN	26	LYS	QG	4.08
26LYS	HN	26	LYS	QD	6.08
26LYS	HA	26	LYS	QG	3.50
26LYS	HA	26	LYS	QD	6.08
26LYS	HB3	26	LYS	QD	3.52
28CYSS	HN	28	CYSS	QB	2.61
30ARG	HN	30	ARG	QG	6.08
30ARG	HA	30	ARG	QD	6.08

NOEs OBTENIDOS MEDIANTE ASIGNACIÓN AUTOMÁTICA CON CYANA  
 PARA EL CÁLCULO DE LA ESTRUCTURA 3D DEL PÉPTIDO DEL TOMATE

1 LEU QOD	2 CYSS H	4.82
1 LEU HB2	2 CYSS H	4.33
35 CYSS H	35 CYSS QB	3.67
6 CYSS HA	11 ASP HB3	4.20
9 ASN H	9 ASN HB2	3.96
8 SER HB3	9 ASN H	3.95
8 SER HB2	9 ASN H	3.95
9 ASN H	9 ASN HB3	3.96
5 PRO HB2	6 CYSS H	4.15
5 PRO HB3	6 CYSS H	4.15
23 GLU QG	32 TYR HA	5.50
5 PRO HA	6 CYSS H	3.35
6 CYSS HA	7 SER H	3.11
7 SER H	7 SER HB3	3.66
6 CYSS HB2	7 SER H	3.66
7 SER H	11 ASP HB3	4.70
7 SER H	11 ASP HB2	4.70
7 SER H	7 SER HB2	3.66
6 CYSS HA	8 SER H	4.50
6 CYSS HB2	8 SER H	3.67
7 SER H	8 SER H	2.90
8 SER HA	9 ASN H	3.17
9 ASN H	21 CYSS HB3	4.89
8 SER H	9 ASN H	5.50
9 ASN H	10 SER H	3.83
9 ASN H	9 ASN HD21	5.50
9 ASN H	9 ASN HD22	5.50
10 SER H	10 SER QB	3.15
8 SER HA	10 SER H	5.50
9 ASN HB3	10 SER H	5.05
10 SER H	11 ASP H	3.82
37 LEU H	38 LEU H	4.12
10 SER QB	11 ASP H	4.23
11 ASP H	11 ASP HB2	3.95
11 ASP H	11 ASP HB3	3.95
11 ASP H	12 CYSS H	3.60
9 ASN HA	11 ASP H	5.50
12 CYSS H	12 CYSS QB	3.61
9 ASN HA	12 CYSS QB	5.17
10 SER HA	13 ILE H	4.75
10 SER HA	12 CYSS H	5.24
20 PHE QD	39 PRO HA	4.20
12 CYSS H	13 ILE QG2	5.50
12 CYSS H	13 ILE HG12	5.50
12 CYSS H	13 ILE HG13	5.50
9 ASN HA	12 CYSS H	4.94
13 ILE H	13 ILE HB	3.17
13 ILE H	13 ILE HG13	3.76
13 ILE H	13 ILE HG12	3.76
13 ILE H	13 ILE QD1	4.58
13 ILE H	13 ILE QG2	3.87
12 CYSS QB	13 ILE H	4.57
10 SER QB	13 ILE H	5.50
13 ILE H	14 GLY H	4.65
13 ILE QG2	14 GLY H	3.77
13 ILE HB	14 GLY H	4.09
13 ILE HG13	14 GLY H	5.50
14 GLY H	15 ILE H	4.36
15 ILE H	15 ILE HB	3.79
15 ILE H	15 ILE HG12	4.18
15 ILE H	15 ILE QG2	4.23
15 ILE QD1	16 THR H	4.19
15 ILE QG2	16 THR H	4.42
15 ILE HA	16 THR H	3.53

15 ILE H	15 ILE QD1	3.95
15 ILE H	15 ILE HG13	4.18
16 THR H	16 THR HB	3.86
16 THR HB	17 LEU H	3.69
17 LEU H	18 CYSS H	3.93
16 THR H	17 LEU H	5.50
16 THR QG2	17 LEU H	4.49
17 LEU H	17 LEU QD1	4.50
17 LEU H	17 LEU QD2	4.50
17 LEU H	17 LEU HB3	3.68
17 LEU H	17 LEU HG	3.54
37 LEU H	37 LEU QD1	4.53
37 LEU H	37 LEU QD2	4.53
30 LEU H	30 LEU QD1	3.84
30 LEU H	30 LEU QD2	4.37
36 ASN HB2	37 LEU H	4.84
36 ASN HB3	37 LEU H	3.88
37 LEU H	37 LEU HB2	3.24
30 LEU H	30 LEU HB3	3.59
30 LEU H	30 LEU HB2	3.16
28 TYR H	30 LEU QD1	5.09
30 LEU HA	31 THR H	2.74
31 THR H	31 THR QG2	3.81
30 LEU HA	30 LEU QD1	3.47
30 LEU HA	30 LEU QD2	4.60
23 GLU HA	31 THR QG2	4.79
17 LEU HA	17 LEU QD2	4.92
37 LEU HA	37 LEU QD2	4.01
37 LEU HA	37 LEU QD1	4.01
17 LEU HA	17 LEU QD1	4.92
24 LYS HA	25 THR H	2.88
18 CYSS HA	19 GLN H	3.37
19 GLN H	19 GLN HG3	3.96
19 GLN H	19 GLN HB3	4.11
19 GLN H	19 GLN HB2	4.11
19 GLN H	19 GLN HG2	3.96
12 CYSS H	13 ILE H	3.16
25 THR QG2	30 LEU H	4.70
20 PHE H	20 PHE QD	3.64
29 GLY H	30 LEU H	3.25
28 TYR H	29 GLY H	3.13
20 PHE QD	39 PRO HB2	4.32
20 PHE QD	21 CYSS H	5.06
20 PHE HA	21 CYSS H	3.06
9 ASN HA	21 CYSS H	3.89
21 CYSS H	21 CYSS HB2	3.51
21 CYSS H	21 CYSS HB3	3.51
9 ASN HA	21 CYSS HB2	3.99
9 ASN HA	21 CYSS HB3	3.99
21 CYSS HA	22 LYS H	3.07
22 LYS HA	23 GLU H	2.73
22 LYS H	22 LYS QB	3.46
22 LYS H	36 ASN HB2	5.21
22 LYS H	35 CYSS HA	4.20
26 ASP HA	27 GLN H	3.25
22 LYS QB	36 ASN HB2	3.77
23 GLU H	23 GLU HB3	3.29
23 GLU H	23 GLU HB2	3.29
22 LYS HG3	23 GLU H	5.06
34 THR HB	35 CYSS H	3.44
30 LEU HB3	30 LEU QD1	3.26
1 LEU HB2	1 LEU QD	3.08
1 LEU HB3	1 LEU QD	3.08
30 LEU HB3	30 LEU QD2	3.14
30 LEU HB2	30 LEU QD2	3.42
37 LEU HB3	37 LEU QD1	3.37
37 LEU HB3	37 LEU QD2	3.37
37 LEU HB2	37 LEU QD1	3.37

3 ASN	H	4 GLU	H	4.63
4 GLU	HA	5 PRO	HD3	3.15
4 GLU	HA	5 PRO	HD2	3.15
15 ILE	HB	16 THR	H	4.70
10 SER	HA	10 SER	QB	2.95
13 ILE	HA	13 ILE	QD1	4.13
13 ILE	HA	13 ILE	QG2	3.38
13 ILE	HA	13 ILE	HG12	4.24
13 ILE	HA	13 ILE	HG13	4.24
13 ILE	HB	13 ILE	QD1	3.38
15 ILE	QG2	15 ILE	HG13	3.26
15 ILE	QG2	15 ILE	HG12	3.26
15 ILE	HB	15 ILE	QD1	3.16
15 ILE	HA	15 ILE	QG2	3.23
16 THR	HA	16 THR	QG2	3.25
27 GLN	HG3	28 TYR	QD	4.75
23 GLU	H	23 GLU	QG	4.26
22 LYS	HG2	23 GLU	H	5.06
23 GLU	HB2	24 LYS	H	4.55
23 GLU	HB3	24 LYS	H	4.55
23 GLU	QG	24 LYS	H	3.50
23 GLU	HA	24 LYS	H	2.74
24 LYS	H	24 LYS	HB2	4.10
24 LYS	H	24 LYS	HG2	3.96
24 LYS	H	24 LYS	HB3	4.10
24 LYS	HB2	25 THR	H	3.80
25 THR	H	25 THR	QG2	3.91
24 LYS	HB3	25 THR	H	3.80
24 LYS	H	25 THR	H	5.31
25 THR	H	32 TYR	H	5.50
25 THR	HA	25 THR	QG2	3.38
25 THR	H	26 ASP	H	5.50
25 THR	HA	26 ASP	H	2.88
26 ASP	H	26 ASP	HB3	4.12
26 ASP	H	26 ASP	HB2	4.12
25 THR	HB	26 ASP	H	4.33
25 THR	QG2	26 ASP	H	3.73
26 ASP	H	27 GLN	H	5.50
27 GLN	H	27 GLN	HB3	3.63
27 GLN	H	27 GLN	HB2	3.63
27 GLN	H	28 TYR	H	3.75
28 TYR	H	28 TYR	HB3	3.77
28 TYR	H	28 TYR	HB2	3.77
27 GLN	HB2	28 TYR	H	5.18
27 GLN	HB3	28 TYR	H	5.18
25 THR	QG2	29 GLY	H	4.05
29 GLY	H	30 LEU	HB2	4.27
30 LEU	QD1	31 THR	H	5.08
30 LEU	HB2	31 THR	H	4.66
31 THR	H	31 THR	HB	3.37
31 THR	HA	31 THR	QG2	3.44
31 THR	HA	32 TYR	H	2.78
31 THR	H	32 TYR	H	5.50
31 THR	QG2	32 TYR	H	3.36
32 TYR	H	32 TYR	HB2	4.19
32 TYR	HA	33 ARG	H	2.95
32 TYR	H	33 ARG	H	5.50
33 ARG	H	33 ARG	HB2	3.66
33 ARG	H	33 ARG	HB3	3.66
33 ARG	HB2	34 THR	H	4.52
33 ARG	HA	33 ARG	HG3	4.02
33 ARG	HA	33 ARG	HG2	4.02
23 GLU	QG	33 ARG	HG2	4.31
33 ARG	HB3	34 THR	H	4.52
34 THR	H	34 THR	QG2	3.71
23 GLU	HA	34 THR	H	4.44
34 THR	HA	34 THR	QG2	3.38
24 LYS	HG2	24 LYS	QE	3.75

24	LYS	HG3	25	THR	HA	5.06
25	THR	HA	31	THR	QG2	5.50
23	GLU	H	34	THR	H	5.50
34	THR	H	35	CYSS	H	5.50
32	TYR	QD	33	ARG	H	4.17
25	THR	HA	32	TYR	QD	4.16
32	TYR	QD	33	ARG	HA	4.69
33	ARG	HB3	33	ARG	HE	4.72
33	ARG	HB2	33	ARG	HE	4.72
27	GLN	HG2	28	TYR	QD	4.75
33	ARG	HB3	33	ARG	HD2	3.97
22	LYS	HG2	36	ASN	HB3	4.01
33	ARG	HB3	33	ARG	HD3	3.97
33	ARG	HB2	33	ARG	HD2	3.97
33	ARG	HB2	33	ARG	HD3	3.97
20	PHE	QD	39	PRO	HB3	4.32
33	ARG	H	33	ARG	HE	4.67
26	ASP	H	32	TYR	QE	4.48
26	ASP	H	32	TYR	QD	4.35
25	THR	HA	32	TYR	QE	4.61
12	CYSS	HA	12	CYSS	QB	2.90
32	TYR	HA	32	TYR	QD	3.94
37	LEU	H	37	LEU	HB3	3.24
30	LEU	HB2	30	LEU	QD1	3.32
38	LEU	H	38	LEU	HB2	3.72
38	LEU	H	38	LEU	HB3	3.87
38	LEU	HB3	38	LEU	QD1	3.31
38	LEU	HB2	38	LEU	QD1	3.26
38	LEU	QD2	39	PRO	HD2	4.31
20	PHE	HB2	39	PRO	HA	4.17
6	CYSS	H	34	THR	QG2	4.62
38	LEU	HB3	38	LEU	QD2	3.31
38	LEU	HB2	38	LEU	QD2	3.26
32	TYR	QE	34	THR	QG2	3.39
38	LEU	HB2	39	PRO	HD2	3.64
26	ASP	H	30	LEU	HB3	5.11
26	ASP	H	30	LEU	HB2	5.47
24	LYS	HG3	26	ASP	H	5.50
6	CYSS	H	35	CYSS	H	5.50
6	CYSS	HB3	35	CYSS	H	5.50
24	LYS	H	24	LYS	HG3	3.52
24	LYS	HG3	25	THR	H	4.11
16	THR	QG2	18	CYSS	H	4.43
34	THR	QG2	35	CYSS	H	4.15
33	ARG	H	34	THR	QG2	4.87
20	PHE	QD	39	PRO	HD3	4.51
20	PHE	QE	39	PRO	HD3	4.03
20	PHE	HZ	39	PRO	HD3	5.30
20	PHE	QD	39	PRO	HD2	5.27
20	PHE	QE	39	PRO	HD2	4.97
22	LYS	H	36	ASN	HB3	5.50
23	GLU	QG	31	THR	QG2	3.22
9	ASN	H	11	ASP	H	5.50
26	ASP	H	29	GLY	H	5.16
26	ASP	H	28	TYR	H	5.50
26	ASP	H	30	LEU	H	3.86
22	LYS	H	34	THR	H	4.37
27	GLN	H	29	GLY	H	5.50
2	CYSS	H	3	ASN	H	5.04
25	THR	QG2	31	THR	HA	4.30
22	LYS	QB	24	LYS	QE	5.50
32	TYR	H	32	TYR	QD	3.59
27	GLN	H	32	TYR	QE	5.50
28	TYR	H	28	TYR	QD	3.75
28	TYR	QD	30	LEU	H	4.61
6	CYSS	H	7	SER	H	5.50
27	GLN	H	28	TYR	QD	5.41
25	THR	HB	32	TYR	QE	5.50

32 TYR	QE	34 THR	HB	5.50
25 THR	HB	32 TYR	QD	5.50
32 TYR	QD	34 THR	HB	5.50
20 PHE	QD	38 LEU	HA	3.75
20 PHE	HA	20 PHE	QD	3.54
20 PHE	QD	37 LEU	HA	5.07
28 TYR	HA	28 TYR	QD	3.42
24 LYS	H	33 ARG	HA	3.72
25 THR	HA	32 TYR	H	5.50
6 CYSS	H	34 THR	HA	3.91
34 THR	HA	35 CYSS	H	3.55
26 ASP	H	31 THR	HA	5.50
33 ARG	HA	34 THR	H	2.88
24 LYS	HG3	24 LYS	QE	3.94
31 THR	HA	32 TYR	QD	4.31
6 CYSS	HB2	11 ASP	H	4.66
5 PRO	HA	34 THR	HA	4.12
21 CYSS	HA	35 CYSS	HA	4.85
23 GLU	HA	33 ARG	HA	4.30
12 CYSS	QB	15 ILE	QD1	5.02
20 PHE	HB3	39 PRO	HA	4.17
6 CYSS	HB2	11 ASP	HB3	3.53
24 LYS	QD	32 TYR	QE	4.11
6 CYSS	HB2	11 ASP	HB2	3.53
31 THR	HB	32 TYR	H	4.34
30 LEU	HB2	32 TYR	QE	4.15
30 LEU	HB3	32 TYR	QE	4.39
27 GLN	HA	29 GLY	H	4.29
13 ILE	HA	14 GLY	H	3.15
23 GLU	HA	23 GLU	QG	3.56
38 LEU	HA	39 PRO	HD3	2.97
17 LEU	HA	17 LEU	HG	4.04
10 SER	HA	13 ILE	QG2	4.47
10 SER	HA	13 ILE	HB	5.01
10 SER	QB	13 ILE	QD1	4.10
12 CYSS	HA	15 ILE	QG2	5.00
1 LEU	HA	1 LEU	QQD	3.34
1 LEU	HA	1 LEU	HG	4.17
10 SER	QB	13 ILE	HB	5.29
13 ILE	HA	19 GLN	HB3	5.50
15 ILE	HB	16 THR	HB	5.50
32 TYR	HB3	33 ARG	H	4.03
24 LYS	HG3	32 TYR	QE	5.34
31 THR	QG2	32 TYR	QD	5.38
16 THR	H	16 THR	QG2	4.14
28 TYR	QD	30 LEU	QD2	4.98
30 LEU	QD1	32 TYR	QE	5.50
24 LYS	HG2	32 TYR	QE	5.50
32 TYR	QD	34 THR	QG2	3.12
30 LEU	HB3	32 TYR	QD	4.04
30 LEU	HB2	32 TYR	QD	5.19
23 GLU	QG	33 ARG	HE	5.13
28 TYR	QD	30 LEU	HB2	5.50
24 LYS	HA	25 THR	HB	4.31
30 LEU	HA	31 THR	HB	5.08
38 LEU	H	38 LEU	HG	4.55
20 PHE	H	20 PHE	QE	5.48
10 SER	HA	13 ILE	HG13	4.21
22 LYS	QB	36 ASN	HB3	3.98
32 TYR	HB2	34 THR	QG2	3.95
32 TYR	HB3	34 THR	QG2	5.27
25 THR	QG2	29 GLY	HA2	3.33
25 THR	QG2	29 GLY	HA3	4.58
23 GLU	HA	24 LYS	HG2	4.79
23 GLU	HA	33 ARG	HG2	5.06
24 LYS	HA	24 LYS	HG2	4.15
31 THR	QG2	32 TYR	HA	5.50
5 PRO	HA	34 THR	QG2	3.57



6	CYSS	HA	11	ASP	HB2	4.20
38	LEU	HA	39	PRO	HD2	3.19
9	ASN	H	21	CYSS	HB2	4.89
8	SER	HA	21	CYSS	HB2	4.86
8	SER	HA	21	CYSS	HB3	4.86
36	ASN	HB3	38	LEU	HB2	5.50
38	LEU	QD1	39	PRO	HD2	4.31
28	TYR	HB3	30	LEU	QD2	4.76
28	TYR	QE	30	LEU	QD2	5.50
15	ILE	QD1	16	THR	HB	3.78
1	LEU	QQD	16	THR	HB	5.50
1	LEU	QQD	16	THR	HA	5.50
28	TYR	H	29	GLY	HA2	5.50
26	ASP	H	29	GLY	HA2	5.50
30	LEU	HB3	31	THR	H	3.68
12	CYSS	H	13	ILE	HB	5.50
36	ASN	HA	37	LEU	H	3.25
20	PHE	QE	38	LEU	HA	4.74
23	GLU	HB3	31	THR	QG2	4.73
23	GLU	HB2	31	THR	QG2	4.73
16	THR	QG2	17	LEU	HG	4.18
5	PRO	HA	35	CYSS	H	5.36
32	TYR	HB2	33	ARG	H	3.27
6	CYSS	HB3	7	SER	H	4.21
20	PHE	QE	39	PRO	HA	5.50
23	GLU	QG	33	ARG	HG3	4.31
19	GLN	H	37	LEU	HA	5.50
38	LEU	HA	38	LEU	HG	3.95
23	GLU	HA	33	ARG	HG3	5.06
24	LYS	HA	25	THR	QG2	4.36
22	LYS	HG3	36	ASN	HB3	4.01
10	SER	HA	13	ILE	HG12	4.21
10	SER	HA	13	ILE	QD1	3.82
1	LEU	HB3	2	CYSS	H	4.33
15	ILE	HA	15	ILE	QD1	4.21
25	THR	QG2	28	TYR	H	5.50
25	THR	QG2	31	THR	H	5.50
30	LEU	QD2	31	THR	H	5.07
8	SER	H	11	ASP	H	4.29
25	THR	H	25	THR	HB	3.35
27	GLN	HG2	28	TYR	QE	5.22
27	GLN	HG3	28	TYR	QE	5.22
38	LEU	H	38	LEU	QD1	5.50
38	LEU	H	38	LEU	QD2	5.50
7	SER	H	11	ASP	H	5.50
28	TYR	HB2	30	LEU	QD2	4.76
9	ASN	HB2	10	SER	H	5.05
6	CYSS	HB3	8	SER	H	4.84
17	LEU	H	17	LEU	HB2	3.68
38	LEU	HB3	39	PRO	HD3	4.24
38	LEU	HB3	39	PRO	HD2	3.37
36	ASN	HB3	38	LEU	H	4.88
24	LYS	HG2	34	THR	H	5.50
19	GLN	H	20	PHE	QD	4.36
20	PHE	HZ	39	PRO	HD2	5.50
13	ILE	HG12	14	GLY	H	5.50
37	LEU	HB2	37	LEU	QD2	3.37
13	ILE	QG2	19	GLN	HA	5.40
13	ILE	HA	19	GLN	HB2	5.50
25	THR	QG2	27	GLN	HA	5.50
20	PHE	HA	20	PHE	QE	5.50
1	LEU	QB	1	LEU	QQD	2.59
1	LEU	QB	2	CYSS	H	3.80
1	LEU	QQD	2	CYSS	QB	5.03
2	CYSS	H	2	CYSS	QB	3.58
3	ASN	QB	3	ASN	QD2	3.26
4	GLU	H	4	GLU	QB	3.56
4	GLU	H	4	GLU	QG	5.34

4	GLU	HA	4	GLU	QG	3.67
4	GLU	HA	5	PRO	QG	4.95
4	GLU	HA	5	PRO	QD	2.68
4	GLU	QB	5	PRO	QD	3.27
4	GLU	QG	5	PRO	QD	4.03
5	PRO	QB	6	CYSS	H	3.64
5	PRO	QB	34	THR	QG2	5.34
6	CYSS	H	33	ARG	QB	5.34
6	CYSS	HA	11	ASP	QB	3.62
6	CYSS	HB2	11	ASP	QB	2.97
7	SER	H	11	ASP	QB	3.83
8	SER	H	8	SER	QB	3.23
8	SER	H	11	ASP	QB	3.90
8	SER	HA	21	CYSS	QB	4.01
8	SER	QB	9	ASN	H	3.15
9	ASN	H	9	ASN	QB	3.16
9	ASN	H	9	ASN	QD2	4.66
9	ASN	HA	21	CYSS	QB	3.23
9	ASN	QB	9	ASN	QD2	2.94
9	ASN	QB	10	SER	H	4.29
9	ASN	QB	10	SER	QB	5.33
9	ASN	QB	13	ILE	QD1	4.47
9	ASN	QB	19	GLN	QB	4.03
9	ASN	QB	21	CYSS	H	4.35
10	SER	H	11	ASP	QB	4.93
10	SER	HA	13	ILE	QG1	3.68
11	ASP	H	11	ASP	QB	3.15
11	ASP	QB	12	CYSS	H	4.29
13	ILE	H	13	ILE	QG1	3.23
13	ILE	HA	13	ILE	QG1	3.66
13	ILE	HA	19	GLN	QG	5.34
13	ILE	HA	19	GLN	QE2	5.34
13	ILE	HB	14	GLY	QA	5.34
13	ILE	QG2	13	ILE	QG1	3.03
13	ILE	QG2	14	GLY	QA	3.99
13	ILE	QG2	19	GLN	QE2	5.34
13	ILE	QG1	14	GLY	H	4.78
14	GLY	QA	16	THR	H	5.34
15	ILE	H	15	ILE	QG1	3.56
15	ILE	QG2	15	ILE	QG1	2.83
15	ILE	QD1	18	CYSS	QB	4.91
16	THR	H	18	CYSS	QB	5.34
17	LEU	H	17	LEU	QQD	3.72
17	LEU	HA	17	LEU	QQD	3.67
17	LEU	QB	17	LEU	QQD	2.33
18	CYSS	H	18	CYSS	QB	3.54
18	CYSS	H	19	GLN	QB	5.34
18	CYSS	QB	37	LEU	HA	4.43
19	GLN	H	19	GLN	QB	3.38
19	GLN	H	19	GLN	QG	3.46
19	GLN	HA	19	GLN	QG	3.67
19	GLN	QB	19	GLN	QE2	4.59
19	GLN	QB	20	PHE	QD	5.34
19	GLN	QB	20	PHE	QE	5.34
19	GLN	QG	20	PHE	QD	4.23
19	GLN	QG	20	PHE	QE	4.50
20	PHE	H	20	PHE	QB	3.41
20	PHE	QB	39	PRO	HA	3.31
20	PHE	QB	39	PRO	QB	4.91
20	PHE	QD	38	LEU	QQD	4.68
20	PHE	QD	39	PRO	QB	3.56
20	PHE	QD	39	PRO	QG	4.63
20	PHE	QE	38	LEU	QQD	4.72
20	PHE	QE	39	PRO	QB	4.66
20	PHE	QE	39	PRO	QG	4.35
21	CYSS	QB	35	CYSS	HA	5.34
22	LYS	H	22	LYS	QG	3.60
22	LYS	H	22	LYS	QD	5.17

22	LYS	HA	22	LYS	QD	3.96
22	LYS	QB	22	LYS	QG	2.41
22	LYS	QB	22	LYS	QD	3.27
22	LYS	QB	36	ASN	QD2	5.34
22	LYS	QG	22	LYS	QE	3.19
22	LYS	QG	23	GLU	H	4.34
22	LYS	QG	36	ASN	HB2	3.91
22	LYS	QG	36	ASN	HB3	3.33
22	LYS	QD	23	GLU	H	3.80
22	LYS	QD	23	GLU	HA	5.34
22	LYS	QD	36	ASN	HB2	5.27
22	LYS	QD	36	ASN	HB3	5.34
23	GLU	HA	33	ARG	QB	5.34
23	GLU	HA	33	ARG	QG	4.26
23	GLU	QB	31	THR	QG2	4.03
23	GLU	QB	33	ARG	QG	3.86
23	GLU	QB	33	ARG	HE	5.34
23	GLU	QG	33	ARG	QG	3.69
24	LYS	H	24	LYS	QB	3.33
24	LYS	QB	24	LYS	QD	3.47
24	LYS	QB	24	LYS	QE	3.90
24	LYS	QB	25	THR	H	3.10
25	THR	HA	26	ASP	QB	5.34
26	ASP	H	26	ASP	QB	3.33
26	ASP	QB	27	GLN	H	2.84
26	ASP	QB	28	TYR	H	5.34
26	ASP	QB	32	TYR	QD	4.29
26	ASP	QB	32	TYR	QE	3.57
27	GLN	H	27	GLN	QB	3.11
27	GLN	H	27	GLN	QG	3.42
27	GLN	QG	28	TYR	H	3.85
27	GLN	QG	28	TYR	HA	4.77
27	GLN	QG	28	TYR	QD	3.97
27	GLN	QG	28	TYR	QE	4.35
27	GLN	QG	29	GLY	H	5.15
28	TYR	H	28	TYR	QB	3.28
28	TYR	QB	29	GLY	H	4.31
28	TYR	QB	30	LEU	H	4.17
28	TYR	QB	30	LEU	HB2	4.45
28	TYR	QB	30	LEU	HB3	5.06
28	TYR	QB	30	LEU	QD1	3.88
28	TYR	QB	30	LEU	QD2	4.13
33	ARG	H	33	ARG	QB	3.10
33	ARG	H	33	ARG	QG	3.67
33	ARG	HA	33	ARG	QD	5.34
33	ARG	QB	33	ARG	QD	2.98
33	ARG	QB	33	ARG	HE	4.12
33	ARG	QB	34	THR	H	3.71
33	ARG	QG	34	THR	H	4.39
36	ASN	HB3	36	ASN	QD2	3.59
36	ASN	QD2	37	LEU	QB	5.18
36	ASN	QD2	38	LEU	H	3.77
36	ASN	QD2	38	LEU	HB2	4.23
36	ASN	QD2	38	LEU	QQD	5.27
37	LEU	H	37	LEU	QQD	3.97
37	LEU	HA	37	LEU	QQD	3.36
37	LEU	QB	37	LEU	QQD	2.38
37	LEU	QB	38	LEU	H	3.96
38	LEU	HA	38	LEU	QQD	3.16
38	LEU	HB2	38	LEU	QQD	2.80
38	LEU	HB3	38	LEU	QQD	2.74
38	LEU	QQD	39	PRO	HD2	3.48
38	LEU	QQD	39	PRO	HD3	3.84



*Los Resultados del Capítulo 1 de esta Tesis generaron la Publicación que se anexa:*



VIP

# On the Importance of Carbohydrate–Aromatic Interactions for the Molecular Recognition of Oligosaccharides by Proteins: NMR Studies of the Structure and Binding Affinity of AcAMP2-like Peptides with Non-Natural Naphthyl and Fluoroaromatic Residues

M. Isabel Chávez,<sup>[a, b]</sup> Cecilia Andreu,<sup>[c]</sup> Paloma Vidal,<sup>[a]</sup> Nuria Aboitiz,<sup>[a]</sup> Felix Freire,<sup>[a, d]</sup> Patrick Groves,<sup>[a]</sup> Juan Luis Asensio,<sup>[a]</sup> Gregorio Asensio,<sup>[c]</sup> Michiro Muraki,<sup>[e]</sup> Francisco Javier Cañada,<sup>\*[a]</sup> and Jesús Jiménez-Barbero<sup>\*[a]</sup>

*Dedicated to Professor Manuel Martín-Lomas on the occasion of his 64th birthday*

**Abstract:** The specific interaction of a variety of modified hevein domains to chitoooligosaccharides has been studied by NMR spectroscopy in order to assess the importance of aromatic–carbohydrate interactions for the molecular recognition of neutral sugars. These mutant AcAMP2-like peptides, which have 4-fluoro-phenylalanine, tryptophan, or 2-naphthylalanine at the key interacting positions, have been prepared by solid-phase synthesis. Their three-dimensional structures, when bound to the chitin-derived trisaccha-

ride, have been deduced by NMR spectroscopy. By using DYANA and restrained molecular dynamics simulations with the AMBER 5.0 force field, the three-dimensional structures of the protein–sugar complexes have been obtained. The thermodynamic analysis of the interactions that occur upon com-

plex formation have also been carried out. Regarding binding affinity, the obtained data have permitted the deduction that the larger the aromatic group, the higher the association constant and the binding enthalpy. In all cases, entropy opposes binding. In contrast, deactivation of the aromatic rings by attaching fluorine atoms decreases the binding affinity, with a concomitant decrease in enthalpy. The role of the chemical nature of the aromatic ring for establishing sugar contacts has been thus evaluated.

**Keywords:** carbohydrate binding • hevein • molecular dynamics • molecular recognition • NMR spectroscopy • peptides

[a] M. I. Chávez,<sup>\*</sup> P. Vidal,<sup>\*</sup> Dr. N. Aboitiz, F. Freire, Dr. P. Groves, Dr. J. L. Asensio, Dr. F. J. Cañada, Prof. Dr. J. Jiménez-Barbero  
Department of Protein Structure and Function  
Centro de Investigaciones Biológicas, CSIC  
Ramiro de Maeztu 9, 28040 Madrid (Spain)  
Fax: (+34)91-553-1706  
E-mail: jcanada@cib.csic.es  
jjbarbero@cib.csic.es

[b] M. I. Chávez<sup>\*</sup>  
Instituto de Química, UNAM  
Ciudad Universitaria, Coyoacán (México DF)

[c] Dr. C. Andreu,<sup>\*</sup> G. Asensio  
Departamento de Química Orgánica  
Facultad de Farmacia  
Universidad de Valencia  
46100 Burjassot (Spain)

[d] F. Freire  
Departamento de Química Orgánica  
Univ. de Santiago de Compostela y  
Unidad Asociada al CSIC de RMN de Biomoléculas  
Santiago de Compostela (Spain)

[e] M. Muraki  
Biological Information Research Center  
National Institute of Advanced Industrial Science and Technology  
Central 6, 1-1-1 Higashi, Tsukuba, Ibaraki 305-8566 (Japan)

[\*] These authors contributed equally to this work.

Supporting information for this article is available on the WWW under <http://www.chemeurj.org/> or from the author. Schematic view of the NMR signals of the peptides that are significantly affected by the addition of ligand (S1), chemical shifts assignments (S2), sections of NMR spectra for the three polypeptides (S3), list of the intramolecular protein–protein NOEs unambiguously assigned and converted into relevant distance constraints for the peptides (intraresidue, sequential, medium, and long range) (S4), Ramachandran-type plots of the NMR structures of the three polypeptides (S5), titrations and chemical shift perturbation data of the NMR signals of the mutant peptides in the presence of (GlcNAc)<sub>3</sub> (S6), schematic view of the two possible arrangements of the trisaccharide and the intermolecular NOE data at the binding site (S7), NMR/AMBER-based structures of the complex for the Phe18Nal peptide (S8), description of the results of the solvated molecular dynamics simulation for Phe18Trp (S9), and temperature dependence of titration plots of the mutant peptides in the presence of (GlcNAc)<sub>3</sub> (S10).

## Introduction

Carbohydrate–protein interactions play an important role in a wide variety of biological processes, including immunological and inflammatory responses, organogenesis, metastasis, and diverse infections.<sup>[1]</sup> In these processes, defense mechanisms against pathogen microorganism invasions such as those produced by bacteria, viruses, and parasites, are included, since on many occasions they essentially depend on the molecular recognition of specific carbohydrates by proteins.<sup>[2]</sup>

In this context, the determination of the structural and conformational factors, which govern the molecular recognition of these biomolecules, as well as the knowledge of the physicochemical features of these processes, is of paramount importance.<sup>[3]</sup>

Previous investigations by X-ray crystallography, NMR spectroscopy, molecular modeling, and calorimetric studies have allowed information to be obtained on the structural and thermodynamic characteristics of a variety of protein–carbohydrate complexes.<sup>[4]</sup> Due to the amphipathic character of oligosaccharides, different kinds of forces may be involved in the recognition process by a given protein. The presence of the oxygen atoms of the hydroxyl groups evidently provides a potential involvement in intermolecular hydrogen bonds with the side chains of polar amino acids within the polypeptide chain. Nevertheless, not only polar forces are involved in carbohydrate recognition. Depending on the stereochemistry of the monomer constituents of the oligosaccharide chain, the presence of a number of rather apolar C–H groups constitute patches that provide van der Waals, CH– $\pi$ , and hydrophobic interactions.<sup>[5]</sup>

NMR spectroscopy has been used to determine the conformations of diverse carbohydrates in the binding site of proteins, generally by means of transferred NOE experiments. Conformations are usually determined in combination with molecular dynamics calculations, thus leading to an understanding, to varying degrees, of the factors involved in molecular recognition processes.<sup>[6]</sup>

Among the processes in which carbohydrates are recognized by proteins, there are some small proteins of plant origin that bind chitin, a polymer of *N*-acetylglucosamine with  $\beta(1\rightarrow4)$  glycosidic linkages.<sup>[7]</sup> These protein domains have been related with the defense of plants against insects and fungi,<sup>[8]</sup> and usually include a structural motif (sequence) of around 30 to 45 amino acids linked with three or four disulfide bonds, denominated the hevein domain.<sup>[9]</sup>

In the last years, the three-dimensional structures of several hevein domains have been determined, both in the free and in the sugar-associated state, by using X-ray crystallographic analysis and/or NMR spectroscopy. As leading examples, the cases of wheat germ agglutinin<sup>[10,11]</sup> and its isolated B domain,<sup>[12]</sup> *Urtica dioica* agglutinin,<sup>[13,14]</sup> and hevein itself isolated from latex<sup>[15–20]</sup> have been reported. Additional studies for other domains have also been carried out,<sup>[21,22]</sup> including one on the smaller antimicrobial peptide AcAMP2.<sup>[23,24]</sup>

According to these X-ray and NMR studies, the aromatic residues at relative positions 21, 23, and 30 in hevein domains (Phe18, Tyr20, Tyr27 in Ac-AMP2) play an important role in carbohydrate binding by stabilizing the complexes by means of interactions between the aromatic and sugar rings through van der Waals contacts and CH– $\pi$  stacking. Additionally, Ser19 of hevein (Ser16 in Ac-AMP2) is involved in a hydrogen-bonding interaction with the carbonyl group of the acetamide moiety of one GlcNAc residue, and at the same time the methyl group of this acetamide interacts with the side-chain aromatic rings of residues 23 and 30 (Tyr20 and Tyr27 in Ac-AMP2). As a final key interaction, the hydroxyl group of hevein Tyr30 (AcAMP2 Tyr27) also provides additional hydrogen bonding with OH-3 of the same sugar moiety. The same kind of interactions take place in other proteins with one or more hevein domains on the same polypeptide chain, including pseudohevein, *Urtica Dioica* agglutinin (UDA), pokeweed agglutinin, and wheat germ agglutinin (WGA).<sup>[15–22]</sup> No detailed analysis of the binding mode to the antifungal AcAMP2 peptide has been reported to date.<sup>[23]</sup>

The origin of the specificity and stability of protein–carbohydrate complexes has become a topic of major discussion,<sup>[5,25–28]</sup> and, at this point, it seems relevant to verify the relative roles of carbohydrate–aromatic stacking and hydrogen-bonding interactions for the recognition of sugars. The small size of hevein-like peptides, as well as the availability of their three-dimensional structures, make them a suitable system for the study of the structural features and energetics involved in their sugar interaction processes.<sup>[29,30]</sup>

On this basis, we are currently engaged in a multidisciplinary project devoted to the modification of the key interacting residues of hevein domains in order to understand the origin of protein–carbohydrate interactions. Thus, synthetic peptides have been prepared by using solid-phase synthesis with the aim of modifying the chemical nature of the key aromatic amino acids involved in sugar binding. One of us has previously reported on the chemical synthesis of mutated AcAMP2 peptides.<sup>[31,32]</sup> Thus, the aromatic amino acid at relative position 18 (Phe in AcAMP2) was mutated to alanine (non-aromatic residue) and to residues either with larger electron density and aromatic surface such as tryptophan (Trp, Phe18Trp) and naphthylalanine (2-Nal, Phe18Nal), or with the electron-deficient aromatic rings pentafluorophenylalanine and nitrophenylalanine.<sup>[5,31,32]</sup> Herein, we have studied the three-dimensional structure and thermodynamics of the synthetic tryptophan-, 2-naphthylalanine-, and 4-fluorophenylalanine (Pff)-AcAMP-2 mutants bound to chitooligosaccharides. This last variation has also been combined with the simultaneous mutation of Tyr20 to a second 4-fluorophenylalanine residue (Pff). Thus, it is our aim in this work to deduce the influence of the electronic density of the aromatic rings of the lectin in the energetics of binding to the sugars. The three-dimensional structure of the modified lectins in their free and (GlcNAc)<sub>3</sub>-bound state has been studied in aqueous solution by NMR spectroscopy and modeling methods, including molecular dynamics. Fur-



thermore, the thermodynamics of the binding processes with the artificial peptides has been characterized both by fluorescence (for Phe18Trp), <sup>1</sup>H NMR, and <sup>19</sup>F NMR spectroscopy and compared with a chemically synthesized sample of AcAMP-2 that is identical to the natural sequence. Previous qualitative data on the affinity of the Phe18Trp and Phe18Nal towards large polymeric chitin have been reported.<sup>[32]</sup>

A schematic view of the sequences is given in Figure 1.

## Results

**Synthesis:** The synthesis of Phe18Trp and Phe18Nal have been described elsewhere.<sup>[31,32]</sup> As with AcAMP2, Phe18Pff/Tyr20Pff is a polypeptide that consists of 30 amino acid residues with three intramolecular disulfide bonds in its structure and was obtained as an amide peptide. In this modified AcAMP2 analogue, Phe18 and Tyr20 of the native AcAMP2 were changed to 4-fluorophenylalanine (Pff).

The amino acids were manually assembled by solid-phase synthesis by using Fluoren-9-yl-methoxycarbonyl (Fmoc) chemistry according to standard protocols, with DCC and 1-hydroxybenzotriazole hydrate (HOBt) as coupling reagents.

The side chains of the six Cys residues of the sequence were protected with the same protecting group (trityl), which were removed by trifluoroacetic acid (TFA) treatment for resin cleavage. After testing several experimental conditions<sup>[33]</sup> (redox pairs, temperatures, denaturants, pH, and buffers), the fully reduced peptide was subjected to oxidative folding in NH<sub>4</sub>Ac buffer (0.1 M, pH 7.8) in high dilution conditions (peptide concentration 3.1 × 10<sup>-5</sup> M) in the presence of reduced and oxidized glutathione (GSH/GSSG) as redox reagents for three days at room temperature (molar ratio peptide:GSH:GSSG 1:100:10). After folding, the reac-

tion was stopped by acidification to pH 2.5 with TFA and lyophilized. The crude material was analyzed by RP-HPLC (linear gradient from 10% to 90% of acetonitrile in water containing TFA 0.09%, over 30 min at flow rate of 1 mL min<sup>-1</sup>) showing the presence of several different isomers (retention times between 10.4 and 13.5 min), but clearly the first peak represented the major product (retention time 10.6 min). Isolation and purification by using semipreparative RP-HPLC afforded the mutant AcAMP2 peptide in 30% yield from the starting material.

The synthesis and full characterization by HPLC and ESMS is detailed in the Experimental Section.

### Thermodynamic analysis of chitoooligosaccharide binding to AcAMP2, and to the Phe18Trp, Phe18Nal, and Phe18Pff/Tyr20Pff mutants:

The equilibrium association constants (*K<sub>a</sub>*) were first obtained by one-dimensional <sup>1</sup>H NMR titrations. Thus, the binding constants of (GlcNAc)<sub>3</sub> to Phe18Trp, Phe18Nal, and Phe18Pff/Tyr20Pff were obtained (Table 1) by utilising one-dimensional <sup>1</sup>H NMR spectra of a series of

Table 1. Affinity data from the NMR analysis (and fluorescence for Phe18Trp) for the binding of *N,N,N'*-tri-acetylchitotriose to natural AcAMP2 and to the other peptides mutated at positions 18 and 20. Estimated errors amount to ± 15%. The thermodynamic parameters estimated from a van't Hoff analysis are also given. An enthalpy–entropy compensation phenomenon is evident.

	<i>K<sub>a</sub></i> [M <sup>-1</sup> ]				Thermodynamic parameters	
	<i>T</i> = 298 K	<i>T</i> = 303 K	<i>T</i> = 308 K	<i>T</i> = 313 K	Δ <i>H</i> [kJ mol <sup>-1</sup> ]	Δ <i>S</i> [J K <sup>-1</sup> mol <sup>-1</sup> ]
AcAMP2	1206	721	537	340	-63.4	-154.1
Phe18Trp	1755	1473	916	614	-54.1	-119.0
Phe18Nal	3527	2314	1638	999	-63.9	-146.6
Phe18Pff/Tyr20Pff	448	348	255	190	-45.1	-101.0

peptide samples with increasing trisaccharide concentrations.<sup>[34]</sup> Also, similar experiments were performed for the parent AcAMP2 for testing purposes. For the Phe18Pff/Tyr20Pff peptide, additional spectra were recorded using <sup>19</sup>F NMR spectroscopy. For Phe18Trp, additional binding data were obtained by fluorescence titration. In this titration, as observed for hevein, hevein-32,<sup>[34]</sup> and wheat germ agglutinin,<sup>[35]</sup> a blue-shift was observed for the tryptophan emission of Phe18Trp with the peak maximum shifting from

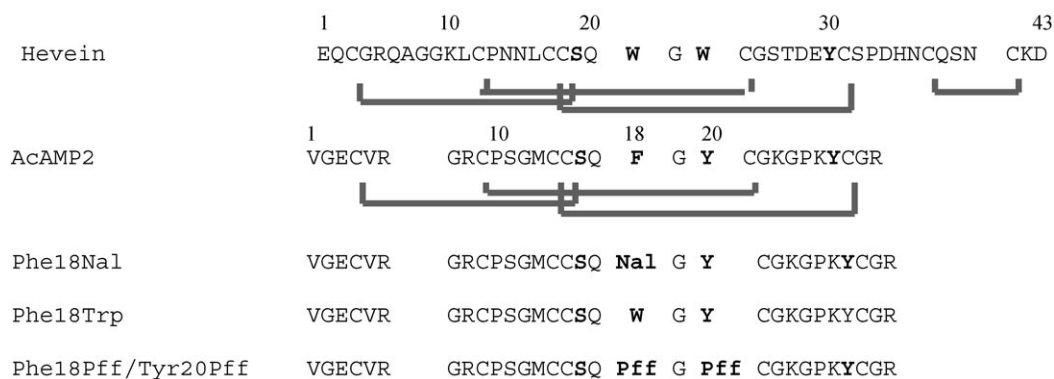


Figure 1. Amino acid sequence alignment of AcAMP2 and hevein (indicating the disulfide bridge pattern) with the three mutant ACAMP2 polypeptides used in this study. The residues involved in the binding site are highlighted.

357 nm in the protein-free form to 352 nm in the protein-bound form (Figure 2). This shift in peak maximum is consistent with a solvent exposed tryptophan of the protein becoming shielded from water molecules upon complexation with the ligand. A 46% increase in fluorescence intensity at 340 nm falls between the 28% increase reported for hevein and 89% increase reported for hevein-32.<sup>[34]</sup> In the NMR ti-

tration, a variety of peptide NMR signals are significantly affected by the addition of ligand (see Supporting Information, S1). The perturbations in chemical shifts allowed the determination of the equilibrium association constants.<sup>[36]</sup> For AcAMP2, and Phe18Trp, Phe18Nal and Phe18Pff/Tyr20Pff, the backbone NH protons of either Ser16 or other NH resonance signals were followed as a function of the

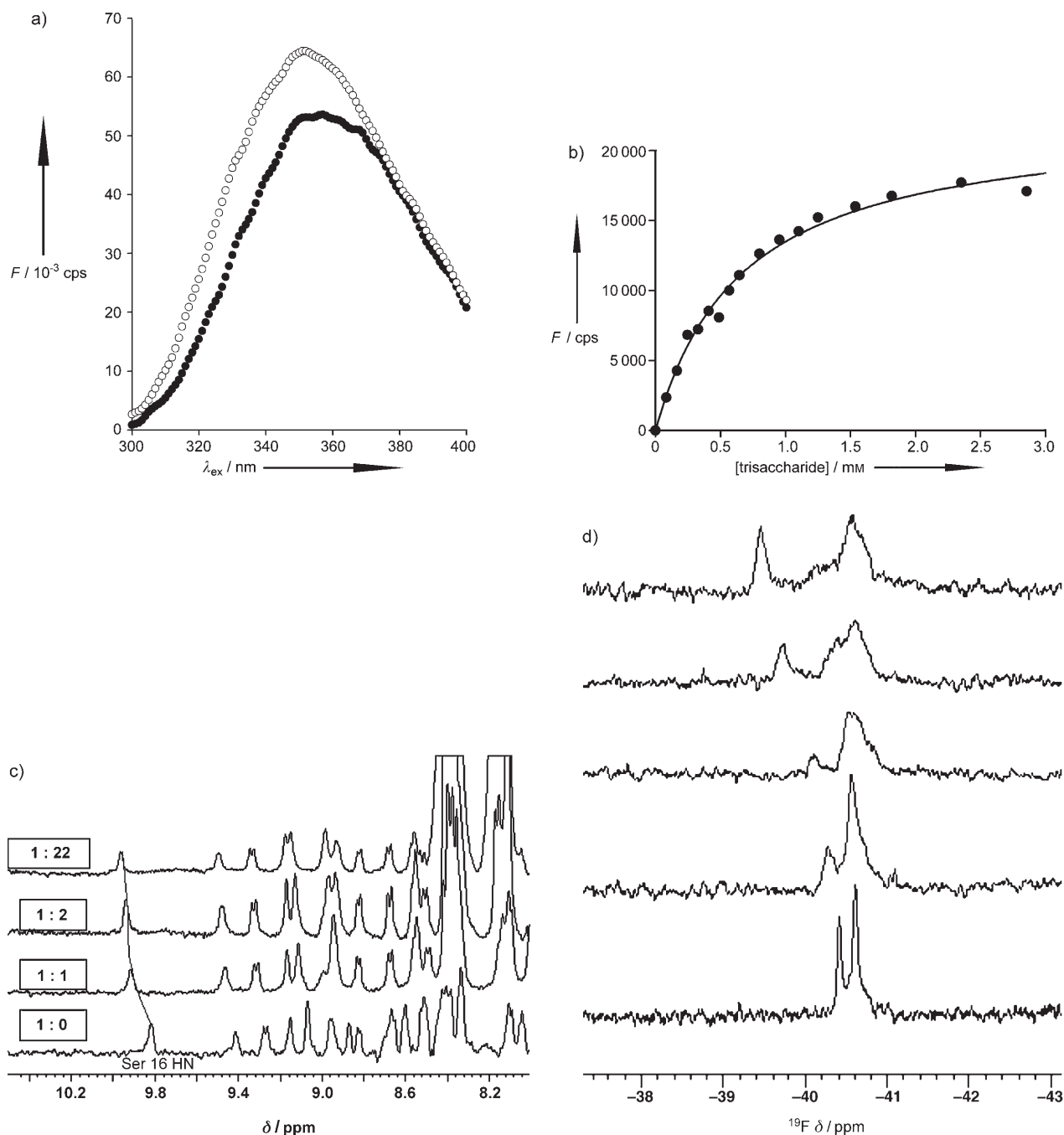


Figure 2. Fluorescence spectroscopy of Phe18Trp. a) Phe18Trp free (filled circles) and trisaccharide-bound (open circles) fluorescence spectra. b) The change in fluorescence of Phe18Trp at 340 nm was monitored as a function of trisaccharide concentration. The data was fitted in SigmaPlot to a first-order binding equation to obtain the binding parameters. c)  $^1\text{H}$  NMR titrations of Phe18Nal (0.7 mM) with increasing amounts of  $N,N,N'$ -triacetylchitotriose at 298 K and pH 5.6. The variations of Ser16 HN are highlighted. d)  $^{19}\text{F}$  NMR titrations of Phe18Pff/Tyr20Pff with  $N,N,N'$ -triacetylchitotriose. The variations of the  $^{19}\text{F}$  signals with the ligand to protein molar ratio are shown at 298 K. The presence of a dynamic process is evident from the broadening of the signals. More than one sugar-lectin complex is present in solution.

concentration of added chitotriose to assess the binding constants (Figure 2C, Table 1). For Phe18Pff/Tyr20Pff, the  $^{19}\text{F}$  NMR spectra permitted a better evaluation of the titration data.

A van't Hoff plot of NMR-determined  $K_a$  values as a function of temperature was used to estimate the equilibrium thermodynamic parameters,  $\Delta H^\circ$  and  $\Delta S^\circ$  (Figure 3, Table 1).

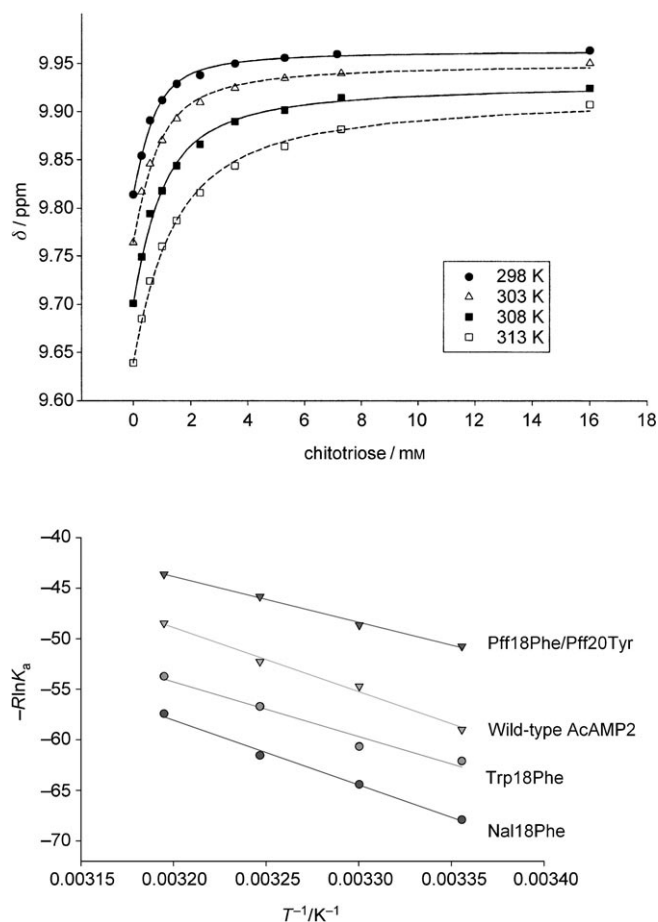


Figure 3. Thermodynamic parameters of  $(\text{GlcNAc})_3$  binding to AcAMP2 and its mutant peptides, as determined by NMR spectroscopy. Top: The binding curves derived from the NMR titrations for the association of  $(\text{GlcNAc})_3$  to Phe18Nal are shown. Bottom: The van't Hoff approximation for the binding of  $(\text{GlcNAc})_3$  to AcAMP2 and the three mutant peptides. In all cases, the binding process is enthalpy driven.

As reported for other hevein domains, the interaction with chitotriose is enthalpy driven and entropy opposes binding.<sup>[17–22]</sup> It should be noted that although the linear assumption in the van't Hoff plot is only approximated, our previous studies with hevein domains have demonstrated that NMR values differ by <10% from those obtained by isothermal titration calorimetry.<sup>[17–22]</sup>

$^{19}\text{F}$  NMR titration experiments for chitotriose binding to Phe18Pff/Tyr20Pff at 298 K resulted in one of the signals being shifted downfield when the concentration of the sugar increases. In contrast, the other peak is remarkable in being

clearly broadened and split at a 1:7 molar ratio, remaining split at larger molar ratios. This experimental observation suggests the presence of different association modes of the trisaccharide to the peptide (see Figure 2 and below). The peptides studied herein only differ in the chemical nature of the aromatic residue at position 18 (and 20 in one case), thus allowing a direct comparison of the influence of this ring (see Discussion).

### The three-dimensional NMR structure of the AcAMP2 mutants bound to $(\text{GlcNAc})_3$

*The protein:* The assignments of the  $^1\text{H}$  NMR spectra of the free and bound Phe18Trp, Phe18Nal, and Phe18Pff/Tyr20Pff in aqueous solution were facilitated by the reported data for the natural peptide (see Supporting Information for chemical shifts assignments, S2). (The NMR data have been deposited in the BMRB database, with accession numbers 6639 and 6647 for free and bound Phe18Trp, 6657 and 6637 for free and bound Phe18Nal, and 6656 and 6591, for free and bound Phe18Pff/Tyr20Pff.) Indeed, the assignments of the three polypeptides were obtained by using well-established protocols through the assignment of regular two-dimensional TOCSY and NOESY spectra. For the complexes, according to the measured binding constant,  $K_a$  (Table 1), and experimental conditions (peptide concentration 0.3–1.0 mM and  $(\text{GlcNAc})_3$  at a 8:1 molar ratio), the different polypeptides are more than 60% bound at 298 K. For both the free and bound peptides, different sets of intramolecular protein–protein NOEs were unambiguously assigned and converted into relevant distance constraints (intraresidue, sequential, medium, and long range) as depicted in Table 2

Table 2. Number and type of conformationally significant NOEs for the complexes of the three mutant peptides to  $(\text{GlcNAc})_3$ .

	Total assigned NOEs	Intra-residue	Sequential	Medium range	Long range
Phe18Nal	348	64	140	52	99
Phe18Trp	331	60	135	59	77
Phe18Pff/Tyr20Pff	248 <sup>[a]</sup>	79	98	24	47

[a] After “distance modify” and CALIBA routines.

(and Supporting Information S3, S4). For Phe18Pff/Tyr20Pff, the low availability of the samples did not allow the use of higher protein concentrations and the number of detected significant NOEs is lower. Thus, starting from 400, 500, and 800 randomized conformations and applying the DYANA protocol,<sup>[37]</sup> a group of 50 structures with low target-function values were obtained (Table 3). The 25 best DYANA structures of Phe18Trp and Phe18Nal, as well as the 35 best structures of Phe18Pff/Tyr20Pff were subjected to further optimization through a simulated annealing protocol by restricted molecular dynamics (rMD) with the AMBER force field. The average backbone root mean square deviations (rmsd) of the refined AMBER structures was below 0.9 Å,

Table 3. Statistics from DYANA and AMBER restrained MD calculations. The data from the DYANA analysis and AMBER restrained MD calculations are given. The limits of variation interval are in parentheses.

Peptide	Method	Number of structures	Range	Backbone RMSD [Å]	Heavy atom RMSD [Å]	DYANA target function
Phe18Nal	DYANA	25	4–28	0.61 ± 0.15 (0.24–1.05)	1.17 ± 0.17 (0.63–1.68)	0.026 (0.012–0.036)
	AMBER	30	4–28	0.81 ± 0.18 (0.28–1.42)	1.45 ± 0.27 (0.77–2.25)	
Phe18Trp	DYANA	25	4–28	0.73 ± 0.23 (0.25–1.56)	1.53 ± 0.30 (0.76–2.44)	0.43 (0.33–0.50)
	AMBER	26	4–28	0.80 ± 0.23	1.86 ± 0.41	
Phe18Pff/Tyr20Pff	DYANA	35	4–28	0.60 ± 0.13 (0.15–0.90)	1.40 ± 0.19 (0.74 ± 2.02)	0.071 (0.044–0.086)
	AMBER	25	4–28	0.84 ± 0.18 (0.00–1.36)	1.96 ± 0.37 (0.00–2.68)	

while the heavy atoms were below 2.0 Å, considering residues 4–28 in all cases (see Table 3 and Supporting Information S5). Moreover, the obtained structures have very small deviations from ideal geometry and there are no nonbonded contacts (the coordinates of the three three-dimensional structures of the mutant peptides have been deposited in the protein data bank, with accession codes 1ZUV, 1ZWU, and 1ZNT, for Phe18Trp, Phe18Nal, and Phe18Pff/Tyr20Pff, respectively). Schematic representations of the superimposition of the backbone and key side chains of the three mutant peptides, as derived from the NMR data, by using restricted molecular dynamics are depicted in Figure 4. The two segments located between residues Met13 and Lys23 build a two strand antiparallel  $\beta$ -sheet, which is also present in the natural AcAMP2<sup>[23]</sup> and hevein.<sup>[15–20]</sup> Indeed, NOEs characteristic for the  $\beta$ -sheet (Cys15HA-Cys21HA, Cys14HN-Lys23HA, and Cys15HN-Gly21HA) were identified for all the mutant polypeptides.

*The complex:* While the structure of the protein was well defined by the NMR data, for all three modified AcAMP2 peptides there was also a number of intermolecular ligand–protein NOEs that supported the experimental structures for the complexes (see Figure 5, Table 4, and the Supporting Information S7). In spite of the strong overlapping of the chemical shifts of the sugar residues protons, the comparison with previous chitotriose complexes of hevein allowed the unambiguous assignment of some ligand protons<sup>[20,21]</sup> and their intermolecular cross-peaks. Regarding the saccharide, the solution conformation of (GlcNAc)<sub>3</sub> in the free state<sup>[38]</sup> has been studied. The conformers are in agreement with the occurrence of the exo-anomeric effect and correspond to dihedral angles of  $\Phi = 50 \pm 20^\circ$  and  $\Psi = 0 \pm 20^\circ$  for the glycosidic linkages.<sup>[20,21]</sup> These angles also agree with data obtained from standard NMR methods, molecular mechanics, and dynamics calculations for chitoooligosaccharides bound to hevein<sup>[16–20]</sup> and hevamine.<sup>[38]</sup> The sugar-binding site of diverse hevein domains have been studied by NMR spectroscopy,<sup>[15–21]</sup> and the contact points correspond to Ser16, Phe18, Tyr20, and Tyr27 in wild-type AcAMP2. In the case of the three modified AcAMP2 peptides under study, the presence of a similar binding arrangement was deduced

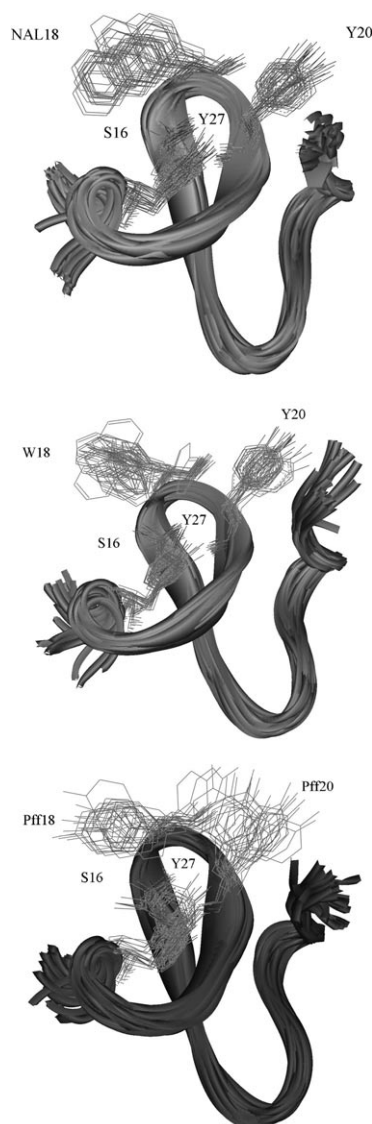


Figure 4. Ribbon representation of the superimposition of the backbone of the NMR-derived structures of the AcAMP2 mutants when bound to (GlcNAc)<sub>3</sub>. In this representation, the sugar has been omitted for clarity. The structures of Phe18Nal (top), Phe18Trp (middle), and Phe18Pff/Tyr20Pff (bottom) are shown. The location of the side chains of the key amino acids for sugar binding is also highlighted. The N and C termini of the peptides are marked in each panel.

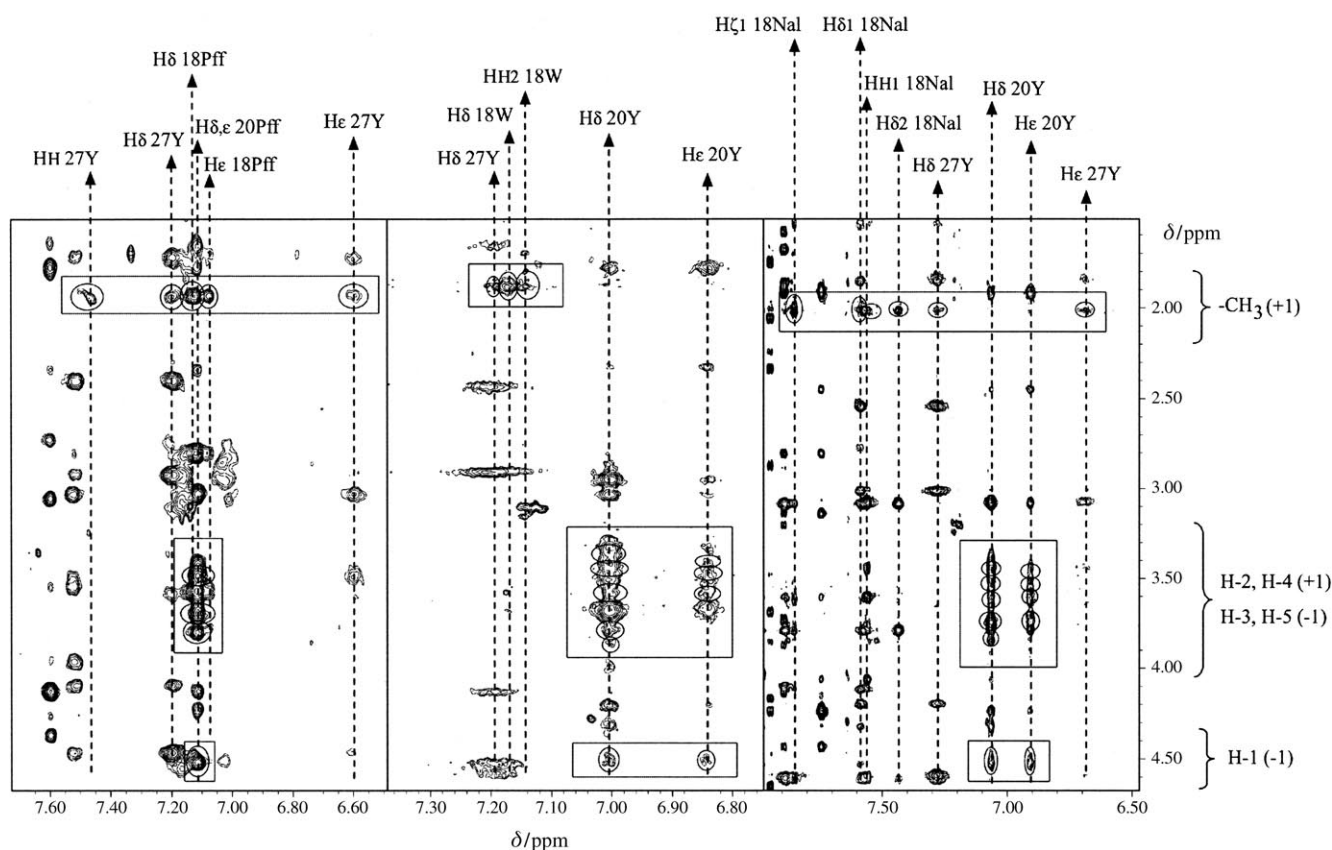


Figure 5. Sections of the 800 MHz NOESY spectra (298 K, 200 ms mixing time), recorded for the complexes of the three polypeptides with (GlcNAc)<sub>3</sub>, showing the assigned sugar–lectin intermolecular NOEs. Phe18Pff/Tyr20Pff (left), Phe18Trp (middle), and Phe18Nal (right).

from the chemical shift perturbation data of the NMR signals of the mutant peptides (free versus bound, see Supporting Information S1 and S6). In particular, important chemical shift differences between the free and bound species were mainly observed for Ser16HA, as well as for Cys21HA and Cys28HA, located immediately after two of the key aromatic residues.

Moreover, the position of the binding site was corroborated by the existence of key intermolecular interactions involving these particular residues in the NOESY spectra (see below and Figure 5). Previous studies on the interaction of hevein domains with small chitooligosaccharides (up to the trisaccharide in length) have also shown that the lectin binding site is sufficiently extended for interacting with at least two *N*-acetylglucosamine units, placed on the two primary subsites arranged with the side chains of residues 16, 18, 20, and 27. Therefore, for a trisaccharide, it is possible to consider two possible arrangements at the binding site (see Figure 6 and Supporting Information S7). Additional evidence for the existence of two binding modes for chitotriose also came from the <sup>19</sup>F NMR titrations of Phe18Pff/Tyr20Pff, as mentioned above.

In fact, the observed intermolecular saccharide–lectin NOESY cross-peaks are fittingly explained when considering the simultaneous existence of two distinct binding modes (Table 4). For example, the intermolecular NOE data for the complex of (GlcNAc)<sub>3</sub> to Phe18Pff/Tyr20Pff will be

detailed, while those for the other two mutants basically follow the same trends (Figure 6 and Supporting Information S7).

In the NOESY spectra of the bound Phe18Pff/Tyr20Pff, NOE cross-peaks are observed between the acetamide methyl group of a nonreducing sugar unit with the aromatic protons of the residues Pff18 and Tyr27; this places the corresponding sugar unit over the aromatic side chain of residue Pff20 and defines the subsite +1 (Figure 6 top). This binding mode may also explain the observed interactions between the H-2 and the H-4 protons from the nonreducing end sugar<sup>[21]</sup> with the aromatic protons of Pff20. These cross-peaks are specific to the so-called binding mode 1.

In contrast, the NOE cross-peaks between the anomeric proton of the nonreducing unit of the trisaccharide with the aromatic protons of Pff20 are only possible if this proton belongs to the nonreducing end sugar unit when it is located at subsite (–1). The existence of this NOE indicates the presence of so-called binding mode 2. In this manner, a number of intermolecular NOEs were assigned (Figures 5 and 6, and Supporting Information S7) and related to either of the two possible binding modes. Then, the three-dimensional structures of the two possible complexes were obtained through a simulated annealing protocol from the AMBER force field program, considering the two binding modes previously described. The peptide–peptide (as described above) and the peptide–trisaccharide restrictions, included as distance



Table 4. Intermolecular peptide–carbohydrate NOE interactions for AcAMP2 Phe18Nal, AcAMP2 Phe18Trp, and AcAMP2 Phe18Pff/Tyr20Pff with chitotriose.

H <sub>peptide</sub>	$\delta H_{\text{Phe18Nal}}$	$\delta H_{\text{Phe18Trp}}$	$\delta H_{\text{Phe18Pff/Tyr20Pff}}$	H <sub>carboh</sub>	$\delta H_{\text{carboh}}$	Binding mode 1 <sup>[a]</sup> subsites +3, +2, +1	Binding mode 2 <sup>[a]</sup> subsites +2, +1, -1
18H $\delta$ 1	7.57			CH <sub>3</sub>	2.01	+1	+1
18H $\delta$ 1		7.17		CH <sub>3</sub>	1.96	+1	+1
18H $\delta$ 2	7.43			CH <sub>3</sub>	2.01	+1	+1
18HZ1	7.85			CH <sub>3</sub>	2.01	+1	+1
18HH1	7.55			CH <sub>3</sub>	2.01	+1	+1
18HH2		7.15		CH <sub>3</sub>	1.96	+1	+1
18H $\delta$ 1,2			7.12	CH <sub>3</sub>	2.03	+1	+1
18H $\epsilon$ 1,2			7.08	CH <sub>3</sub>	2.03	+1	+1
20H $\delta$ 1,2	7.06			H1	4.51	–	–1
20H $\delta$ 1,2	7.06			H2	(3.73, 3.82) <sup>[b]</sup>	+1	+1
20H $\delta$ 1,2	7.06			H3	3.57	–	–1
20H $\delta$ 1,2	7.06			H4	(3.48, 3.62) <sup>[b]</sup>	+1	+1
20H $\delta$ 1,2	7.06			H5	3.50	–	–1
20H $\delta$ 1,2		7.01		H1	4.48	–	–1
20H $\delta$ 1,2		7.01		H2	3.78	+1	+1
20H $\delta$ 1,2		7.01		H3	3.57	–	–1
20H $\delta$ 1,2		7.01		H4	3.43	+1	–
20H $\delta$ 1,2		7.01		H5	3.48	–	–1
20H $\delta$ 1,2, H $\epsilon$ 1,2			7.11	H1	4.52	–	–1
20H $\delta$ 1,2, H $\epsilon$ 1,2			7.11	H2	(3.69, 3.80) <sup>[b]</sup>	+1	+1
20H $\delta$ 1,2, H $\epsilon$ 1,2			7.11	H4	(3.48, 3.69) <sup>[b]</sup>	+1	+1
20H $\epsilon$ 1,2	6.91			H1	4.51	–	–1
20H $\epsilon$ 1,2	6.91			H2	(3.73, 3.82) <sup>[b]</sup>	+1	+1
20H $\epsilon$ 1,2	6.91			H3	3.57	–	–1
20H $\epsilon$ 1,2	6.91			H4	(3.48, 3.62) <sup>[b]</sup>	+1	+1
20H $\epsilon$ 1,2	6.91			H5	3.50	–	–1
20H $\epsilon$ 1,2		6.84		H1	4.48	–	–1
20H $\epsilon$ 1,2		6.84		H3	3.57	–	–1
20H $\epsilon$ 1,2		6.84		H4	3.43	+1	–
20H $\epsilon$ 1,2		6.84		H5	3.48	–	–1
27H $\delta$ 1,2	7.28			CH <sub>3</sub>	2.01	+1	+1
27H $\delta$ 1,2		7.20		CH <sub>3</sub>	1.96	+1	+1
27H $\delta$ 1,2			7.19	CH <sub>3</sub>	2.03	+1	+1
27H $\epsilon$ 1,2	6.69			CH <sub>3</sub>	2.01	+1	+1
27H $\epsilon$ 1,2			6.60	CH <sub>3</sub>	2.03	+1	+1
27HH			7.46	CH <sub>3</sub>	2.03	+1	+1

[a] In binding mode 1, +1 corresponds to the reducing end sugar residue and “–” corresponds to the middle sugar residue. In binding mode 2, when –1 corresponds to the reducing end sugar residue, +1 corresponds to the middle one sugar residue, and “–” corresponds to the sugar residue at the nonreducing end. [b] Chemical shifts for both binding modes, respectively.

constraints, derived from the NOESY spectra were used as input to generate the three-dimensional structures of the complexes.

Besides the intermolecular NOEs, two intermolecular hydrogen bonds were also included between the trisaccharide and the hydroxyl groups of Ser16 and Tyr27. These hydrogen bonds have been described in all the hevein-domain complexes reported to date.<sup>[15–20]</sup> The NMR/AMBER-based structures of the two complexes for Phe18Pff/Tyr20Pff, which just differ in the position of the trisaccharide in the binding site, are shown in Figure 6 (bottom), while that for Phe18Nal is given in the Supporting Information S8.

**Solvated molecular dynamics simulations:** In a parallel manner, a series of solvated molecular dynamics simulations starting from the experimental NMR derived structures was performed in order to obtain, by a semi-independent manner,<sup>[3,29,39]</sup> a complete view of the three-dimensional structures and possible motions of the peptide–chito-

oligosaccharide complexes, especially to complement Phe18Pff/Tyr20Pff, for which the number of experimental restraints was smaller due to the low availability of the sample. Therefore, to evaluate the influence of the NOE-based restrictions on the conformational properties of the mutant peptides in solution, solvated MD simulations (sMD) were performed for the free (sMD<sub>free</sub>) and (GlcNAc)<sub>3</sub>-bound (sMD<sub>complex</sub>) peptides, without NOE restrictions. Again, a description of the results for Phe18Pff/Tyr20Pff will be presented, the results for Phe18Trp and Phe18Nal being very similar (see Supporting Information, S8, S9). Since within the context of the present paper the main aim of the sMD simulations was to support the NMR data for Phe18Pff/Tyr20Pff, a deeper discussion of the results of all the solvated MD simulations will be presented separately. In the case of the complexes, the calculations were performed for the two different binding modes.

In all cases, the solvated MD simulations (sMD) started from the corresponding restricted rMD<sub>free</sub> and rMD<sub>complex</sub>

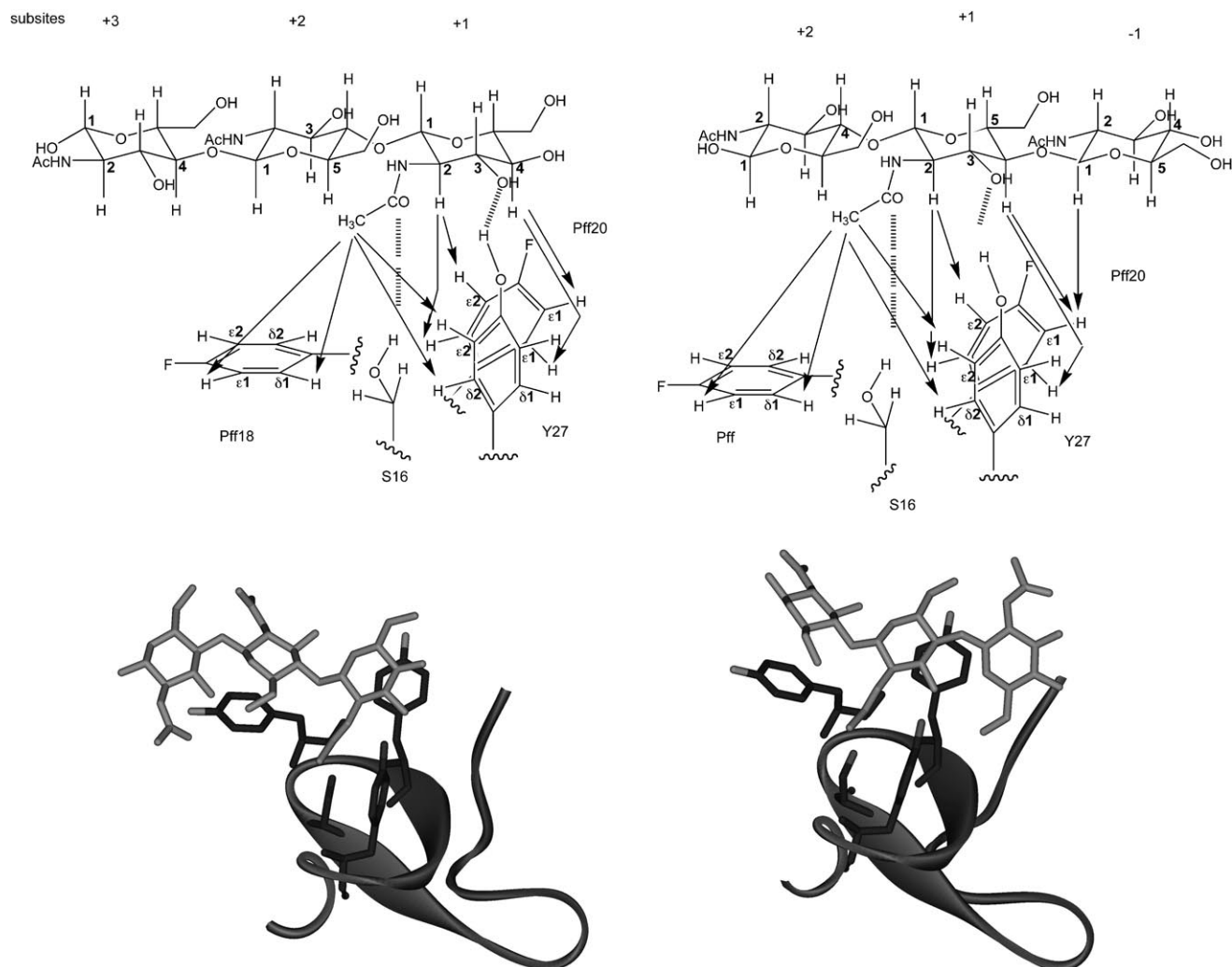


Figure 6. Top: Schematic views of the intermolecular NOEs detected for the complex between  $(\text{GlcNAc})_3$  and Phe18Pff/Tyr20Pff. The two possible binding modes are shown. Bottom: Representations of the corresponding three-dimensional structures, as obtained by restrained MD calculations, using the Amber 5.0 force field for bound Phe18Pff/Tyr20Pff. The non-ambiguous intraprotein and intermolecular protein–trisaccharide NOEs were used as restraints. Two hydrogen bonds between the trisaccharide and the residues Ser16 and Tyr27 were also added.

structures described above. No NMR constraints were applied to retain the starting structure of the proteins or to maintain  $(\text{GlcNAc})_3$  in contact with the protein. It is worth noting that the trisaccharide remained at the binding site and orientation during all simulation times in those calculations carried out for the complexes.

From the methodological viewpoint, the sMD were carried out using AMBER 5.0 with explicit water molecules, counterions, and periodic boundary conditions, while Ewald sums were used for the treatment of electrostatic interactions.<sup>[40]</sup>

A detailed representation of the trajectories of the key torsions that define the orientation of the two aromatic rings of Phe18Pff/Tyr20Pff in the absence and presence of chitotriose are shown in Figure 7.

## Discussion

### Thermodynamic analysis of peptide binding to $(\text{GlcNAc})_3$ :

Previous studies have shown that the  $\Delta G^\circ$  values for hevein binding to  $(\text{GlcNAc})_3$ , based on fluorescence ( $-22.9 \text{ kJ mol}^{-1}$ ) and NMR ( $-23.1 \text{ kJ mol}^{-1}$ ) data, are very similar to  $\Delta G$  measured by isothermal titration calorimetry ( $-22.6 \text{ kJ mol}^{-1}$ ).<sup>[20,34]</sup> Interestingly, the  $\Delta G$  of binding of a truncated analogue of hevein, with the same size as the AcAMP2 analogues and dubbed HEV32, for  $(\text{GlcNAc})_3$  was shown to be only slightly weaker (1.0 (NMR) and 2.2  $\text{kJ mol}^{-1}$  (fluorescence))<sup>[34]</sup> than the corresponding value for hevein (Table 1). These data suggested that the C terminus of hevein does not appear to make a significant contribution to the overall binding affinity of chitooligosaccharide ligands.

The binding data from fluorescence and NMR for Phe18Trp are similar. As mentioned above, a 5 nm blue-

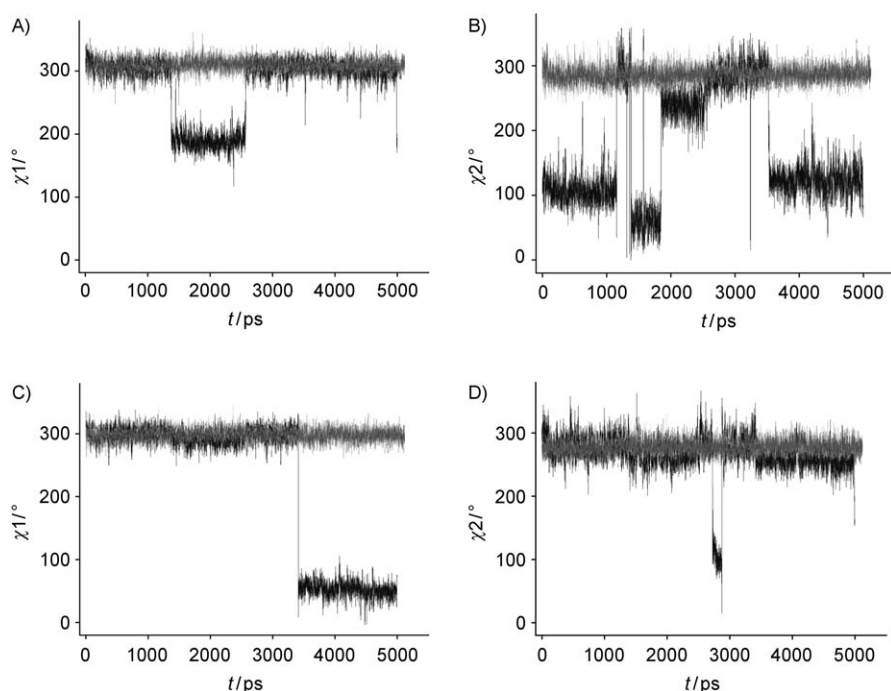


Figure 7. Schematic representation of the trajectory of torsion angles  $\chi_1$  (A, C) and  $\chi_2$  (B, D), for the side chains at position 18 (A, B) and position 20 (C, D) of Phe18Pff/Tyr20Pff during the solvated MD simulation. A) Free peptide (black), (GlcNAc)<sub>3</sub>-bound peptide (gray). B) Free peptide (black), (GlcNAc)<sub>3</sub>-bound peptide (gray). C) Free peptide (black), (GlcNAc)<sub>3</sub>-bound peptide (gray). D) Free peptide (black), (GlcNAc)<sub>3</sub>-bound peptide (gray). Both binding modes 1 and 2 (see text) basically gave the same monotonic trajectory. The importance of the presence or absence of the trisaccharide is remarkable.

shift was observed for the Trp emission of Phe18Trp. The observed 46% increase in fluorescence intensity at 340 nm is in between the values reported for hevein and hevein-32.<sup>[39]</sup> This fluorescence increase is related to changes in fluorescence energy transfer (FRET) between tyrosine and tryptophan residues in the saccharide binding site of the protein. A direct comparison cannot be made as the aromatic residues of Phe18Tyr (2Tyr, 1Trp) are different to those in hevein and hevein-32 (1Tyr, 2Trp). Nonetheless, the increase of 46% in fluorescence unambiguously indicates that changes in the distance and/or relative orientation between Tyr20 and Tyr27 with respect to Trp18 in Phe18Trp occur during the ligand-binding process.

The trisaccharide-induced changes in the tryptophan fluorescence spectrum of Phe18Trp allowed the binding affinity to be measured from a titration experiment. At 25 °C, the binding constant,  $K_a$  was measured as  $1120 \pm 200 \text{ M}^{-1}$ , on the same order of magnitude than the NMR data for AcAMP2 and the three modified analogues (see Table 1). Despite the fact that the measured  $K_a$  values for all the mutants are about one order of magnitude lower than for full-length hevein, the van't Hoff data for the AcAMP2 peptides, analogously to the truncated HEV32, show larger  $\Delta H^\circ$  values than that of hevein, but compensated with an increased  $\Delta S^\circ$  that highly opposes association (Table 1). Although the origin of this enthalpy–entropy compensation phenomenon remains an open question,<sup>[25,26]</sup> it has been reported,<sup>[41]</sup> at this magnitude and sign of  $\Delta S$  and  $\Delta H$ , that hydrogen bond-

ing and van der Waals forces should be the most important factors that stabilize the complex. Part of the observed negative entropy of binding could arise from rigidification of the carbohydrate and/or protein lateral chains,<sup>[25,26,42]</sup> or by reorganization of the water structure. As deduced in earlier studies,<sup>[15–21]</sup> the maximum loss of conformational entropy by freezing of the (GlcNAc)<sub>3</sub> ligand upon binding to lectin domains can reach  $17 \text{ kJ mol}^{-1}$  at 25 °C (the conformational entropy was estimated from the conformational distribution map of (GlcNAc)<sub>3</sub>),<sup>[15–20,38]</sup> much less than the entropy penalty determined for the trisaccharide binding to the mutants, which is between 30 and  $46 \text{ kJ mol}^{-1}$  (Table 1). Thus, apart of the rigidification of the sugar, an additional loss of entropy is taking place. It seems that these shorter hevein domains must pay an entropic

penalty with some reduction of the lateral chains flexibility upon binding, in order to better accommodate the trisaccharide and provide a more favorable binding enthalpy.

The inverse of  $K_a$  values (dissociation constants) for (GlcNAc)<sub>3</sub> binding are in the millimolar range. However, there is a significant trend in the obtained values that is above experimental error and correlates with the qualitative data of chitin binding obtained previously. The obtained binding constant  $K_a$  values for native AcAMP2 and the modified peptides indicated that the larger the aromatic surface of the residue at position 18, the higher the binding affinity. In contrast, the deactivation of the aromatic rings of residues 18 and 20 in Phe18Pff/Tyr20Pff provokes a weaker association constant value in comparison to the other peptides. It has to be considered that this latter peptide (Phe18Pff/Tyr20Pff) involves a double mutation (not only residue 18, but also residue 20). Inspection of the three-dimensional structures of the complexes for the other two peptides (with a Tyr at position 20) indicated that the hydroxyl group of this Tyr residue is not involved in either intra- or intermolecular hydrogen bonding, and thus the measured decrease of affinity for Phe18Pff/Tyr20Pff is not due to the lack of the hydrogen-bond donor character of the hydroxyl group present in Tyr20, but to deactivation of the key aromatic rings.

As mentioned above, the obtained enthalpy and entropy values for all the peptides confirmed that all the molecular recognition processes are enthalpy driven. The entropy of



binding,  $\Delta S^\circ$ , was found to be negative, as also observed for a variety of chitooligosaccharides interacting with hevein itself, pseudohevein, WGA, and UDA.<sup>[11, 15–21, 43, 44]</sup>

The binding enthalpies oscillate between  $-45$  and  $-64$  kJ mol<sup>-1</sup>. In all cases, there is an important loss of entropy that lowers the interaction process, thus increasing the dissociation constant into the millimolar range. The higher enthalpy value is observed for Phe18Nal and the lowest for Phe18Pff/Tyr20Pff. Care should be taken when quantitatively analyzing these thermodynamic data; this observation suggests that the carbohydrate–aromatic interaction (CH/ $\pi$ -type interaction), typical of the hevein domains, is indeed modulated by the chemical nature and the electronic density of the aromatic ring. Although these data should not be overemphasized, some sort of enthalpy–entropy compensation also seems to occur, since the smaller observed binding enthalpy is also accompanied by a lower entropy loss.

**Comparative structural analysis of the three mutant polypeptides with AcAMP2 and with hevein:** The analysis of the NMR data leads to a three-dimensional perspective of the protein–saccharide complexes for the AcAMP2 mutants. The global folding, as well as the orientation of the key side chains of the AcAMP2 mutants, wild-type AcAMP2, and hevein, when bound to the (GlcNAc)<sub>3</sub> are shown in Figures 4, 6, and 8–10. When comparing the free and bound structures, the conformation of the peptide backbone is maintained when the sugar is added and is indeed similar to that deduced for the free peptides, with some slight changes in the orientation of the amino acids involved in the binding site. Indeed, the three-dimensional structures of the complexes of the three mutant peptides were very similar although the orientation of the aromatic ring in Phe18Nal is slightly different to those in the other derivatives.

As mentioned above, the comparison of the refined three-dimensional NMR structures of the mutant peptides with those structures previously reported for wild-type free AcAMP2 by NMR spectroscopy,<sup>[23]</sup> for hevein by X-ray analysis,<sup>[16]</sup> and for hevein as determined by NMR spectroscopy in our research group,<sup>[17–20]</sup> allowed the observation that the molecular topology is very similar. In fact, not only is the polypeptide backbone remarkably similar, but also the orientation of the binding site residues does not reveal significant changes, except for a restriction to motion in the bound state as depicted in Figures 8–10. According to the sMD simulations, the orientations of the key aromatic rings are different in the free and bound states. The orientations of these rings may be defined by the torsions  $\chi_1$  (C $\alpha$ –C $\beta$ ) and  $\chi_2$  (C $\beta$ –C $\gamma$ ). Thus, Figure 7 shows that the aromatic residues at positions 18 and 20 may adopt a variety of values in the free state, while the presence of the sugar strongly influences the available conformational space. In fact, the trisaccharide remarkably restricts the flexibility of these aromatic rings, especially for Pff18 and for Pff20 to a lesser extent. The asymmetry of the aromatic rings in Phe18Trp and Phe18Nal already restrict the motion of these side chains (see Supporting Information S7–S9).

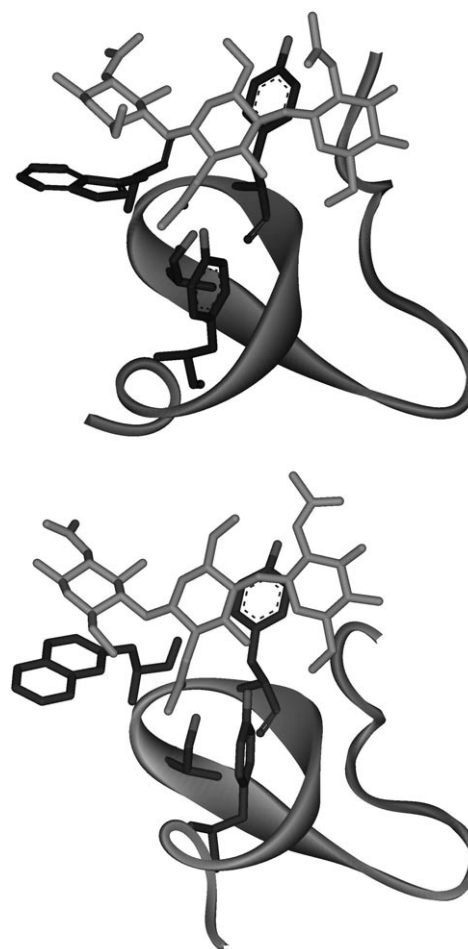


Figure 8. Schematic ribbon representation of the NMR-based structures of Phe18Trp (top) and Phe18Nal (bottom). The key aromatic and Ser16 side chains are depicted. The figure clearly shows the stacking interactions that take place for the molecular recognition of chitotriose by these AcAMP2 mutants. The interaction modes are in total agreement with the detected intermolecular peptide–sugar NOEs. The structures are indeed similar to those depicted for binding mode 2 in Figure 6, for the Phe18Pff/Tyr20Pff double mutant.

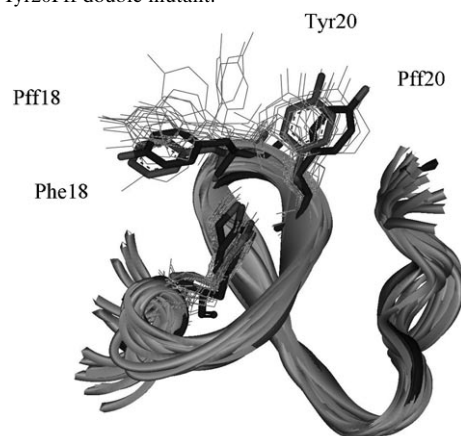


Figure 9. Ribbon representation of the superimposition of the backbone of 18 structures of free Phe18Pff/Tyr20Pff derived from molecular dynamics in explicit solvent (sMDfree). The structure obtained herein by NMR spectroscopy is also shown (gray), as well as that described by Martins et al.<sup>[24]</sup> for wild-type AcAMP2 (black). The similarity is evident. The N and C termini of the peptide chains are marked for clarity.

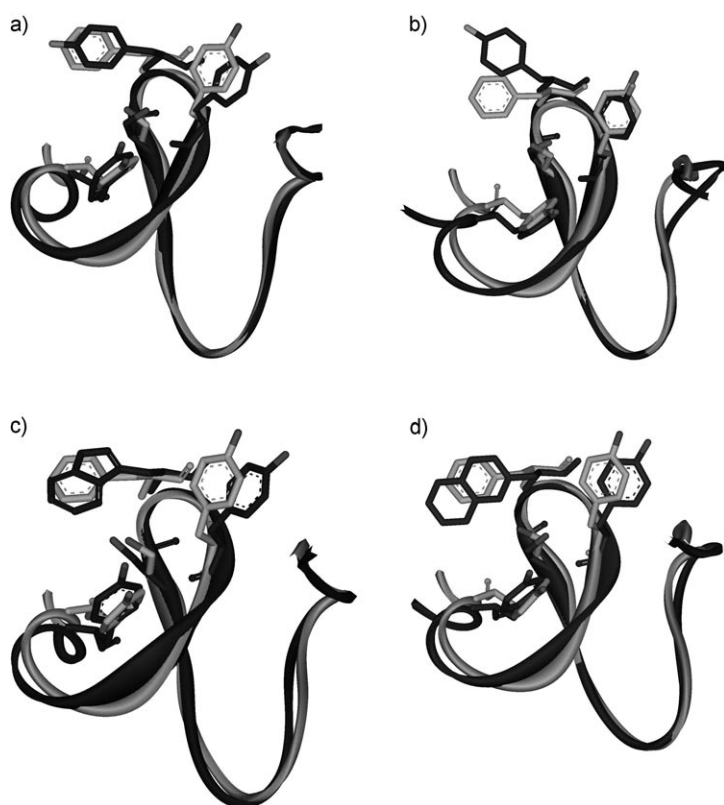


Figure 10. Schematic ribbon representation of the pairwise backbone superimposition of the NMR spectroscopic and solvated MD structures of the three mutant peptides described herein with the NMR structure described by Martins et al.<sup>[24]</sup> for wild-type AcAMP2. a) Phe18Pff/Tyr20Pff by NMR spectroscopy (black) with wild-type AcAMP2 (gray); b) Phe18Pff/Tyr20Pff by solvated MD (black) with wild-type AcAMP2 (gray); c) Phe18Trp by NMR spectroscopy (black) with wild-type AcAMP2 (gray); d) Phe18Nal by NMR spectroscopy (black) with wild-type AcAMP2 (gray).

The simulations confirm that the trisaccharide provides stabilizing interactions with the polypeptide. A combination of CH/ $\pi$  stacking-type interactions, accompanied by hydrogen bonding and van der Waals contacts, contribute to the maintenance of the complexes (Figures 8 and 9), which, for the double mutant, are expressed in the homogeneous orientation of the aromatic rings of the Pff18 and Pff20 moieties (Figure 9).

Regarding similarities, the superimposition of 18 structures taken from the solvated molecular dynamic trajectory (sMD<sub>free</sub>, with a 4.9 ns simulation time) of Phe18Pff/Tyr20Pff with the NMR structure deduced above (and also with the NMR-based wild-type AcAMP2 structure) showed a maximum backbone rmsd of only 0.887 Å. Some differences between the sMD<sub>free</sub> and the NMR structures were observed for the orientation (flexibility) of the aromatic rings of residues 18 and 20, which evidently depend on the presence or absence of the sugar (Figure 9).

When the mutant polypeptides are compared to their parent AcAMP2 structure, the pairwise backbone rmsd are small. In fact, the average NMR structures of Phe18Trp,

Phe18Nal, and Phe18Pff/Tyr20Pff have rmsds with AcAMP2 of 0.984, 0.925, and 0.835 Å, respectively. A pairwise superimposition is given in Figure 10.

### Concluding remarks

The interaction of AcAMP2 and the three mutants with (GlcNAc)<sub>3</sub> has been described in structural terms, making use of a NMR-derived three-dimensional structure and modeling procedures. We have shown that the binding process is enthalpically driven and that both hydrogen bonds and van der Waals forces contribute to the stability of the complexes in aqueous solution. The key role of carbohydrate–aromatic interactions is highlighted. We have demonstrated that the chemical nature of the aromatic residues involved in the protein–carbohydrate interaction strongly influences the binding affinity. Besides, the orientation and mobility of the key aromatic amino acid residues of the lectins implicated in binding depends on the presence of the trisaccharide, experiencing a significant decrease in their conformational freedom upon complexation. Therefore, the nature of the environment of the binding site is significantly important to predict the impact of changes in their constitution for binding affinity. Our data support the physiological relevance and binding ability of small native hevein domains (such as the wild-type AcAMP2) or their mutants. Chitin recognition by a large number of AcAMP2-like molecules would be expected to have a net effect on the dynamic behavior of the polymer, modifying its physical properties. We might speculate that through this mechanism, plant defense proteins, such as AcAMP2, could block the normal development of pathogen organisms, such as fungus, by interfering with the chitin biological function of protection. The role of the C terminus, absent in AcAMP2 with regard to hevein, might be related to other factors, for example, selectivity or as a functional spacer between hevein and other domains of chitinases.

### Experimental Section

**Source of lectins and ligand:** Oligosaccharides were purchased from Toronto Research Chemicals (Canada). The synthesis of the Phe18Trp and Phe18Nal mutants have been described elsewhere.<sup>[31]</sup>

**Synthesis of Phe18Pff/Tyr20Pff—materials and instruments:** All Fluoren-9-ylmethoxycarbonyl (Fmoc) amino acids derivatives, including that of 4-fluorophenylalanine (4-Fphe), and the resin TentaGel S-Ram were purchased from Fluka. The side-chain protection scheme was trityl for all the 6Cys (S-Trt) and Gln (*N*<sub>δ</sub>-Trt); *tert*-butyl for Glu (5-*t*Bu), Ser (O-*t*Bu), and Tyr (O-*t*Bu); *tert*-butoxycarbonyl for Lys (*N*<sub>ε</sub>-Boc), and 2,2,4,6,7-pentamethylidihydrobenzofuransulfonyl for Arg (*N*<sub>δ</sub>-Pbf).

Dicyclohexylcarbodiimide (DCC), 1-hydroxybenzotriazole hydrate (HOBt) and other reagents for peptide synthesis were obtained from Aldrich, and the solvents for the HPLC analysis from Merck.

Reversed-phase high-performance liquid chromatography (RP-HPLC) was performed on a Merck Hitachi Lachrom system. For the analytical work a LiChrospher 100 RP-18 (5  $\mu$ m) 250–4 column was used, with a linear gradient from 10% to 90% of acetonitrile in water (containing

0.09% TFA) over 30 min, and at flow rate of 1 mL min<sup>-1</sup>. The wavelength for the detection was fixed at 214 nm.

The semipreparative separation was performed with a LiChrospher 100 RP-18 (10 μm) 250–10 column, using a gradient of 10% to 37% of acetonitrile in water (containing 0.09% TFA) over 20 min at a flow rate of 2.5 mL min<sup>-1</sup>.

Mass spectrometry was performed using an ion trap mass spectrometer (Esquire 3000 Plus, Bruker) coupled to a liquid chromatograph (Agilent LC 1100 Chemstation). The ionization method was electrospray with positive ion polarity (ESI+). Peptide samples were dissolved in acetonitrile/water (2:3) containing 0.5% formic acid to a concentration of 0.3 mg mL<sup>-1</sup> and infused into the ion spray using a syringe pump at a flow rate of 4 μL min<sup>-1</sup> (capillary voltage -3.8 kV, dry gas temperature 200 °C).

**Solid-phase synthesis:** The AcAMP2 mutant was synthesized as the amide peptide by using a TentaGel S-Ram resin (0.24 mmol g<sup>-1</sup>). The coupling processes were carried out manually with three equivalents of Fmoc amino acids, three equivalents of HOBt, and three equivalents of DCC in DMF during, at least, 90 min until completion, and checked with Kaiser's test. The Fmoc group in each cycle was removed by washing twice with 25% piperidine in DMF for 5 and 12.5 min, respectively. After synthesis, the resin was washed with methylene chloride and diethyl ether and dried under vacuum.

For resin cleavage, the dried resin was suspended in a fresh mixture of trifluoroacetic acid, water, phenol, and triisopropylsilane (88:5:5:2, 10 mL g<sup>-1</sup> of resin) for 5–6 h. Then, the cleavage cocktail containing the peptide was filtered into chilled diethyl ether (10 volumes in relation to the cleavage cocktail), and the peptide precipitated was stored at -80 °C overnight. The crude precipitated peptide was separated by centrifugation (3500 rpm × 15 min), washed twice with cold diethyl ether and air dried. The reduced peptide was directly analyzed by HPLC in the conditions described above (retention time 12.6 min) and characterized by ESMS ([M+2H]<sup>2+</sup> 1604.7, [M+3H]<sup>3+</sup> 1070.6, [M+4H]<sup>4+</sup> 804.0; *M<sub>r</sub>* found: 3208.87; *M<sub>r</sub>* calcd: 3208.74).

**Disulfide formation and purification:** The polythiol peptide was subjected to an oxidative folding reaction under equilibration conditions. The reduced peptide (60 mg) was dissolved in NH<sub>4</sub>OAc buffer (0.1 M, pH 7.8, 600 mL) in the presence of reduced (650 mg) and oxidized (128 mg) glutathione (molar ratio of peptide/GSH/GSSG, 1:100:10). After three days of air oxidation at room temperature, the reaction was stopped by acidification of the mixture with TFA to pH 2.5. The total volume of the reaction was lyophilized to 30 mL, and then directly desalted and purified by semi-preparative RP-HPLC under the conditions described above. The isomer mixture of the different folded peptides were eluted under semi-preparative conditions in the interval of 18 to 25 min, but only the main peak was collected (retention time 18.6 min). This fraction was dried by lyophilization to yield 18 mg of a product that was analyzed by HPLC in the analytical conditions observing a single homogeneous polypeptide peak (retention time 10.6 min) and characterized by ESMS ([M+2H]<sup>2+</sup> 1602.2, [M+3H]<sup>3+</sup> 1069.3, [M+4H]<sup>4+</sup> 803.3, [M+5H]<sup>5+</sup> 643.5; *M<sub>r</sub>* found: 3202.33; *M<sub>r</sub>* calcd: 3202.69). The yield of the oxidative folding reaction was 30% over the starting material.

**Titration experiments—fluorescence:** Fluorescence experiments were performed with 7 × 10<sup>-6</sup> M Phe18Trp. The stock chitotrisaccharide (Toronto Research Chemicals) concentration was calculated based on weight. Experiments were performed at 25 °C, as detailed in Aboitiz et al.<sup>[34]</sup>

**Titration experiments—NMR spectroscopy:** The binding of the carbohydrate to the lectins was monitored by recording one-dimensional 500 MHz <sup>1</sup>H NMR spectra of a series of samples with increasing ligand concentration (ten different concentrations) as previously described.<sup>[17]</sup> The concentration of the protein during the experiments was kept constant (ca. 0.7 mM for AcAMP2; Phe18Trp; Phe18Nal; 0.3 mM for Phe18Pff/Tyr20Pff). The samples were prepared by dissolving the lyophilized protein in buffer (1.0 mL; 85:15, <sup>1</sup>H<sub>2</sub>O/<sup>2</sup>H<sub>2</sub>O, 100 mM NaCl, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 5.6). The one-dimensional NMR spectrum for the sample with the highest ligand/protein ratio was recorded by dissolving the corresponding oligosaccharide (16 mM; 23 mM for Phe18Pff/Tyr20Pff) in the

lectin-containing solution (0.5 mL) described above. The titration curve was established by adding small aliquots of the highest ligand/protein ratio sample to the ligand-free protein sample as previously described. Identical experiments were carried out for Phe18Pff/Tyr20Pff using a series of one-dimensional <sup>19</sup>F NMR spectra. Thermodynamic equilibrium parameters, Δ*S* and Δ*H*, for the lectin-(GlcNAc)<sub>3</sub> interaction were determined from van't Hoff plots in which the affinity constants were assessed at 25, 30, 35, and 40 °C.

**NMR two-dimensional experiments:** The spectra were recorded at 500 and 800 MHz. The samples for the three lectins free (2.5 mM protein) or bound to (GlcNAc)<sub>3</sub> complex [1:8 molar ratio, (ca. 0.7 mM for AcAMP2; Phe18Trp; Phe18Nal; 0.3 mM for Phe18Pff/Tyr20Pff)] were prepared in buffer (0.6 mL; 85:15 H<sub>2</sub>O/D<sub>2</sub>O, 100 mM NaCl, 20 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 5.6) and degassed by passing argon. TOCSY<sup>[45]</sup> (50 ms mixing time) was performed using standard sequences at 298 K. Two-dimensional NOESY experiments<sup>[46]</sup> were performed with mixing times of 200 and 300 ms at 500 and 800 MHz, and 298 K. The NMR data have been deposited in the BMRB database (codes 6639, 6647 (Phe18Trp free and bound, respectively), 6657, 6637 (Phe18Nal free and bound, respectively), 6656 and 6591 (Phe18Pff/Tyr20Pff free and bound, respectively)).

**Structure calculations:** The torsion angle dynamics protocol as implemented in DYANA<sup>[37]</sup> package was followed. Upper limits for proton-proton distances were obtained from NOESY cross-peak intensities at two mixing times (200 and 300 ms). Cross-peak intensities were analyzed with the XEASY software and initially classified as strong, medium, and weak corresponding to upper limits of 2.6, 3.5, and 5.5 Å, respectively. The 800 MHz cross-peaks that were not present at 500 MHz (except as overlapped) were included as 5.5 Å upper limit constraints. The CALIBA protocol was also used to obtain a better definition of the distances. The lower limit for proton-proton distances was set as the sum of the van der Waals radii of the protons. Constraints involving diastereoisomeric atoms were defined to an intermediate position (pseudoatom) and assigned an additional distance of 2.20 Å. Disulfide linkages were included as distance constraints between S-S (2.0 < *r* < 2.1 Å) and between Cβ-S (3.0 < *r* < 3.1 Å). The calculations were performed on a Linux PC computer using DYANA. A set of distance constraints derived from protein-protein NOEs (Table 2) and between 10 and 15 intermolecular distance restraints relating the lectin to the ligand were used in the final round of calculations. The standard protocol with GLOMSA was used,<sup>[34]</sup> which allowed the stereospecific assignments of 50% of the prochiral β-methylens, followed by the application of the "distance modify" and CALIBA modules.

Between 25 to 35 best DYANA structures in terms of target function were subjected to restrained molecular dynamics<sup>[47]</sup> with the AMBER<sup>[48]</sup> force field. After an initial restrained energy minimization (REM) with 2000 conjugate gradient iterations, the structures were equilibrated at 600 K for 2 ps and, at this temperature, their conformational behavior for the next 2 ps was simulated by restrained molecular dynamics (RMD). In the next step, the structures were subjected to a cooling regime, in which the temperature was decreased by 100 K every 2 ps until a temperature of 100 K was reached. At this temperature, 4 ps of RMD calculations were carried out. The final structures were energy-minimized (REM) by using 2000 conjugate gradient iterations. The refined structures can be found in the Protein DataBank (PDB ID: 1ZUV, 1ZWU, and 1ZNT, for the Phe18Trp, Phe18Nal, and Phe18Pff/Tyr20Pff, respectively).

**Molecular mechanics and dynamics calculations:** ϕ is defined as H-1'-C-1''-O-C4'/H-1'-C-1'-O-C4 and ψ as C1''-O-C4'-H4'/C1'-O-C4-H4 for the nonreducing-middle and middle-reducing disaccharide entities, respectively. The MD calculations for the complexes were performed using the AMBER 5.0 package.<sup>[49]</sup>

For Phe18Nal and Phe18Pff/Tyr20Pff, it was necessary to build up the corresponding library for the non-natural naphthylalanine and Pff residues. As a first step, the corresponding structures were generated and minimized with the MM2 force field. Then, the charges were estimated using Hartree Fock 6-31G calculations with Gaussian 98.<sup>[50]</sup> Then, the MOLMOL program<sup>[51]</sup> was used to generate the DYANA and XEASY libraries and the X-LEAP<sup>[52]</sup> module to build up the library for AMBER.

The starting glycosidic torsion angles were taken from previously described MM3\* calculations performed for (GlcNAc)<sub>2</sub> and (GlcNAc)<sub>3</sub>.<sup>[20,21]</sup> The global minimum for the trisaccharide was used as starting structure, since it has been shown that it is very similar to its hevein-bound conformation ( $\Phi/\Psi$  55:5 for both glycosidic linkages).<sup>[20,21]</sup> For the protein coordinates, the input files were prepared from the NMR-derived structures by using the X-LEAP module of the AMBER package. The obtained initial structures were immersed into a box of 3017 TIP3P water molecules in order to obtain accurate solvation. Cut-offs for nonbonding interactions was set to 11.0 Å. The molecular dynamics simulations were carried out by using the Sander module and were performed using periodic boundary conditions and the particle-mesh Ewald approach to introduce long-range electrostatic effects. The SHAKE algorithm for hydrogen atoms, which used a 2 fs time step, was employed.

All solvated simulations were performed at constant pressure and temperature, using the Berendsen coupling algorithm for the latter. Equilibration of the system was carried out as follows; as a first step, a short minimization with positional restraints on solute atoms was run to remove any potentially bad contacts. The force constant for the positional constraints was 500 kcal mol<sup>-1</sup> Å. Next a 12.5 ps molecular dynamics calculation was run at 300 K maintaining positional restraints on the ligand in order to equilibrate the water box and ions. A 9 Å cut-off was used for the treatment of the electrostatic interactions. The system was further equilibrated (12.5 ps run at 300 K) by using the mesh Ewald approach for long-range electrostatic effects. Then, the system was subjected to several minimization cycles (each using 1000 steepest descent iterations) gradually reducing positional restraints on the chitoooligosaccharides from 500 kcal mol<sup>-1</sup> Å to 0. Finally, MD trajectories at constant pressure (1 atm) and temperature (300 K) were collected and analyzed using the Carnal module of AMBER. Structures were recorded every 0.5 ps, for a total calculation time around 4.5 ns in all cases. A deeper discussion of the results of the SMD results will be presented separately.

## Acknowledgments

Funding from the European Union (SACCSIGNET, HPRN-CT2002-00251), Ministry of Education and Science of Spain (grant BQU2003-3550-C03), and Comunidad de Madrid (grant 08.2/0035.1/2000) is gratefully acknowledged. N.A. also thanks Comunidad de Madrid for financial support. We also thank the CAI of NMR of Universidad Complutense of Madrid and the Parc Científic of Barcelona for access to the NMR spectrometers. I.C. thanks UNAM of Mexico for a leave of absence. C.A. and G.A. thank Generalitat Valenciana (Grupos 03/120) for funding. M.M. thanks Ministry of Economy, Trade and Industry of Japan for financial support.

- [1] H.-J. Gabius, H.-C. Siebert, S. Andre, J. Jimenez-Barbero, H. Rudiger, *ChemBioChem* **2004**, *5*, 740–764.
- [2] H. Rudiger, H.-C. Siebert, D. Solis, J. Jimenez-Barbero, A. Romero, C. W. von der Lieth, T. Diaz-Marino, H.-J. Gabius, *Curr. Med. Chem.* **2000**, *7*, 389–416.
- [3] a) A. Imberty, S. Perez, *Chem. Rev.* **2000**, *100*, 4567–4588; b) H. Lis, N. Sharon, *Chem. Rev.* **1998**, *98*, 637–674.
- [4] a) J. Jimenez-Barbero, J. L. Asensio, F. J. Canada, A. Poveda, *Curr. Opin. Struct. Biol.* **1999**, *9*, 549–555; b) H. Kogelberg, D. Solis, J. Jimenez-Barbero, *Curr. Opin. Struct. Biol.* **2003**, *13*, 646–653.
- [5] M. Muraki, *Protein Pept. Lett.* **2002**, *9*, 195–209.
- [6] T. Weimar, R. J. Woods, in *NMR Spectroscopy of Glycoconjugates* (Eds.: J. Jimenez-Barbero, T. Peters), Wiley-VCH, Weinheim, **2002**, pp. 111–114.
- [7] E. J. M. Van Damme, W. J. Peumans, A. Barre, P. Rouge, *Crit. Rev. Plant Sci.* **1998**, *17*, 575–692.
- [8] a) W. J. Peumans, E. J. M. van Damme, *Histochem. J.* **1995**, *27*, 253–271; b) F. Garcia-Olmedo, A. Molina, J. M. Alamillo, P. Rodriguez-Palenzuela, *Biopolymers* **1998**, *47*, 479–491.
- [9] J. J. Beintema, *FEBS Lett.* **1994**, *350*, 159–163.
- [10] a) C. S. Wright, *J. Biol. Chem.* **1992**, *267*, 14345–14352; b) C. S. Wright, G. E. Kellogg, *Protein Sci.* **1996**, *5*, 1466–1476.
- [11] G. Bains, R. T. Lee, Y. C. Lee, E. Freire, *Biochemistry* **1992**, *31*, 12624–12628.
- [12] J. F. Espinosa, J. L. Asensio, J. L. Garcia, J. Laynez, M. Bruix, C. Wright, H. C. Siebert, H. J. Gabius, F. J. Canada, J. Jimenez-Barbero, *Eur. J. Biochem.* **2000**, *267*, 3965–3978.
- [13] F. A. Saul, P. Rovira, G. Boulot, E. J. Damme, W. J. Peumans, P. Truffa-Bachi, G. A. Bentley, *Structure* **2000**, *8*, 593–603.
- [14] K. Harata, M. Muraki, *J. Mol. Biol.* **2000**, *297*, 673–681.
- [15] A. Rodriguez-Romero, K. G. Ravichandran, M. Soriano-Garcia, *FEBS Lett.* **1991**, *291*, 307–339.
- [16] C. A. Reyes-Lopez, A. Hernandez-Santoyo, M. Pedraza-Escalona, G. Mendoza, A. Hernandez-Arana, A. Rodriguez-Romero, *Biochem. Biophys. Res. Commun.* **2004**, *314*, 123–130.
- [17] N. H. Andersen, B. Cao, A. Rodriguez-Romero, B. Arreguin, *Biochemistry* **1993**, *32*, 1407–1422.
- [18] J. L. Asensio, F. J. Canada, M. Bruix, A. Rodriguez-Romero, J. Jimenez-Barbero, *Eur. J. Biochem.* **1995**, *230*, 621–633.
- [19] a) J. L. Asensio, F. J. Canada, M. Bruix, C. Gonzalez, N. Khiar, A. Rodriguez-Romero, J. Jimenez-Barbero, *Glycobiology* **1998**, *8*, 569–577; b) H. C. Siebert, S. Andre, J. L. Asensio, F. J. Canada, X. Dong, J. F. Espinosa, M. Frank, M. Gilleron, H. Kaltner, T. Kozar, N. V. Bovin, C. W. von Der Lieth, J. F. Vliegthart, J. Jimenez-Barbero, H. J. Gabius, *ChemBioChem* **2000**, *1*, 181–195.
- [20] J. L. Asensio, F. J. Canada, H. C. Siebert, J. Laynez, A. Poveda, P. M. Nieto, U. M. Soedjanaamadja, H. J. Gabius, J. Jimenez-Barbero, *Chem. Biol.* **2000**, *7*, 529–543.
- [21] J. L. Asensio, H. C. Siebert, C. W. von Der Lieth, J. Laynez, M. Bruix, U. M. Soedjanaamadja, J. J. Beintema, F. J. Canada, H. J. Gabius, J. Jimenez-Barbero, *Proteins* **2000**, *40*, 218–236.
- [22] P. Karisola, J. Mikkola, N. Kalkkinen, K. J. Airene, O. H. Laitinen, S. Repo, O. T. Pentikainen, T. Reunala, K. Turjanmaa, M. S. Johnson, T. Palosuo, M. S. Kulomaa, H. Alenius, *J. Immunol.* **2004**, *172*, 2621–2628.
- [23] J. C. Martins, D. Maes, R. Loris, H. A. Pepermans, L. Wyns, R. Willem, P. Verheyden, *J. Mol. Biol.* **1996**, *258*, 322–333.
- [24] P. Verheyden, J. Pletinckx, D. Maes, H. A. Pepermans, L. Wyns, R. Willem, J. C. Martins, *FEBS Lett.* **1995**, *370*, 245–249.
- [25] R. U. Lemieux, *Acc. Chem. Res.* **1996**, *29*, 373–380.
- [26] J. P. Carver, *Pure Appl. Chem.* **1993**, *65*, 763–770.
- [27] T. K. Dam, F. Brewer, *Chem. Rev.* **2002**, *102*, 387–429.
- [28] M. R. Wormald, A. J. Petrescu, Y. L. Pao, A. Glithero, T. Elliott, R. A. Dwek, *Chem. Rev.* **2002**, *102*, 371–386.
- [29] R. Woods, *Glycoconj. J.* **1998**, *15*, 209–216.
- [30] a) H.-J. Gabius, *Pharm. Res.* **1998**, *15*, 23–30; b) C. von der Lieth, H. Siebert, T. Kozar, M. Burchert, M. Frank, M. Gilleron, H. Kaltner, G. Kayser, E. Tajkhorshid, N. V. Bovin, J. F. Vliegthart, H. J. Gabius, *Acta Anat.* **1998**, *161*, 91–109.
- [31] M. Muraki, H. Morii, K. Harata, *Protein Pept. Lett.* **1998**, *5*, 193–198.
- [32] M. Muraki, H. Morii, K. Harata, *Protein Eng.* **2000**, *13*, 385–389.
- [33] a) D. Andreu, E. Nicolas, in *Solid-Phase Synthesis* (Eds.: S. A. Kates, F. Albericio), Marcel Dekker, New York, **2000**, pp. 365–375; b) M. Vila-Perello, A. Sanchez-Vallet, F. Garcia-Olmedo, A. Molina, D. Andreu, *FEBS Lett.* **2003**, *536*, 215–219.
- [34] N. Aboitiz, M. Vila-Perello, P. Groves, J. L. Asensio, D. Andreu, F. J. Canada, J. Jimenez-Barbero, *Chembiochem*, **2004**, *5*, 1245–1255.
- [35] J. P. Privat, F. Delmotte, G. Mialonier, P. Bouchard, M. Monsigny, *Eur. J. Biochem.* **1974**, *47*, 5–14.
- [36] L. Fielding, *Curr. Top. Med. Chem.* **2003**, *3*, 39–53.
- [37] a) P. Guntert, W. Braun, K. Wuthrich, *J. Mol. Biol.* **1991**, *217*, 517–530; b) P. Guntert, C. Mumenthaler, K. Wuthrich, *J. Mol. Biol.* **1997**, *273*, 283–298.
- [38] A. Germer, C. Mugge, M. G. Peter, A. Rottmann, E. Kleinpeter, *Chem. Eur. J.* **2003**, *9*, 1964–1973.
- [39] M. G. Ford, T. Weimar, T. Kohli, R. J. Woods, *Proteins* **2003**, *53*, 229–240.

- [40] C. S. A. T. Darden, *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 155.
- [41] K. A. Kronis, J. P. Carver, *Biochemistry* **1985**, *24*, 834–840.
- [42] a) J. P. Carver, S. W. Michnik, A. Imberty, D. A. Cumming, *Ciba Found. Symp.* **1989**, *145*, 6–26; b) F. A. Quioco, *Pure Appl. Chem.* **1989**, *61*, 1293–1306.
- [43] E. Garcia-Hernandez, R. A. Zubillaga, A. Rojo-Dominguez, A. Rodriguez-Romero, A. Hernandez-Arana, *Proteins* **1997**, *29*, 467–477.
- [44] R. T. Lee, H. J. Gabius, Y. C. Lee, *Glycoconj. J.* **1998**, *15*, 649–655.
- [45] A. Bax, D. G. Davis, *J. Magn. Reson.* **1985**, *65*, 355–360.
- [46] A. Kumar, R. R. Ernst, K. Wuthrich, *Biochem. Biophys. Res. Commun.* **1980**, *95*, 1–6.
- [47] R. M. Scheek, W. F. van Gunsteren, R. Kaptein, *Methods Enzymol.* **1989**, *177*, 204–218.
- [48] S. J. Weiner, P. A. Kollman, D. A. Case, U. C. Singh, C. Ghio, G. Alagona, S. Profeta, P. Weiner, *J. Am. Chem. Soc.* **1984**, *106*, 765–784.
- [49] D. A. Pearlman, D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, S. Debolt, D. Ferguson, G. Seibel, P. Kollman, *Comput. Phys. Commun.* **1995**, *91*, 1–41.
- [50] Gaussian98 (Revision A.11.3), M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, V. G. Zakrzewski, J. A. Montgomery, R. E. Stratmann, J. C. Burant, S. Dapprich, J. M. Millam, A. D. Daniels, K. N. Kudin, M. C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Pomelli, C. Adamo, S. Clifford, J. Ochterski, G. A. Petersson, P. Y. Ayala, Q. Cui, K. Morokuma, D. K. Malick, A. D. Rabuck, K. Raghavachari, J. B. Foresman, J. Cioslowski, J. V. Ortiz, B. B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R. L. Martin, D. J. Fox, T. Keith, M. A. Al-Laham, C. Y. Peng, A. Nanayakkara, C. Gonzalez, M. Challacombe, P. M. W. Gill, B. G. Johnson, W. Chen, M. W. Wong, J. L. Andres, M. Head-Gordon, E. S. Replogle, J. A. Pople, Gaussian, Inc., Pittsburgh, PA, **1998**.
- [51] R. Koradi, M. Billeter, K. Wüthrich, *J. Mol. Graphics* **1996**, *14*, 51–55.
- [52] C. E. A. F. Schafmeister, W. F. Ross, V. Romanovsky, University of California, San Francisco, **1995**.

Received: April 1, 2005

Revised: August 19, 2005

Published online: October 12, 2005





