

Doctor

José Luis Salazar

Director de División de Nuevas Creacion

SUPERINTENDENCIA DE INDUSTRIA Y CC

E. _____ S. _____

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 11-176414- -00006-0000

Fecha: 2013-09-20 16:38:39 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 FASENACIONALI
Act. 523 RESPUESTA Folios: 115

Referencia: Expediente No. 11.176.414

Solicitud de Patente de Invención titulada: "Derivados de suero amiloide P, su preparación y uso"

Solicitante: Promedior Inc.

RESPUESTA AL AUTO NOTIFICADO POR FIJACIÓN EN LISTA

DEL 02 DE JULIO DE 2013, OFICIO NO. 11181 (EXAMEN DE FONDO)

ÁLVARO CORREA ORDÓÑEZ, mayor de edad y vecino de Bogotá D.C., identificado tal como aparece al pie de mi firma, en mi carácter de apoderado de la sociedad en referencia, doy respuesta al auto emanado por ese Despacho dentro del término legal:

1. Objeciones del Examinador

En la acción oficial su Despacho manifiesta lo siguiente:

1.1 Reivindicaciones

1.1.1. El objeto de las reivindicaciones 30 - 35 y 100, las cuales se refieren a un método de tratamiento en humanos, no es patentable de acuerdo con el artículo citado.

1.1.2. Un polipéptido glicosilado debe ser reclamado por su estructura, es decir, por su secuencia de aminoácidos, identificada con un número de identificación SEQ ID N°, y por la estructura tridimensional de la cadena de oligosacáridos (reivindicación 1).

1.1.3. Las expresiones "en donde al menos...", "tiene al menos 50% menos...", "termina con una parte de...", "termina con partes de...", "comprende al menos una...", "está sustancialmente libre...", son imprecisas por lo cual no permiten delimitar la materia que se quiere proteger. Se sugiere ser sustituido por términos concretos (reivindicaciones 1 - 4, 9 - 16).

1.1.4. Se sugiere indicar las modificaciones hechas a la cadena de oligosacáridos del polipéptido glicosilado para que ésta tenga una estructura diferencial bi-, tri-, tetra- o penta-antenaria (reivindicaciones 5 - 8).

1.1.5. En las reivindicaciones 15, 52, 76 y 91, las cuales hacen referencia a un residuo glicosilo modificado de la cadena de oligosacáridos, se debe especificar el tipo de modificación reclamada, para tener claridad sobre el alcance de la invención. De igual manera en las reivindicaciones 16, 53, 77 y 92 se deben especificar los conjugados empleados.

1.1.6. Un polipéptido no debe definirse por porcentajes de identidad ya que esto no permite diferenciar la invención del estado de la técnica (reivindicaciones 18 y 19).

1.1.7. La reivindicación 20, la cual hace referencia al polipéptido como una proteína de fusión que comprende uno o más dominios heterólogos, no es admisible puesto que no se especifican las características esenciales de dichos dominios, y por lo tanto, no se puede tener claridad sobre el real alcance de la invención.

1.1.8. Las reivindicaciones 21, 24, 25, 45, 61, 78, 93, 95, 96 y 99 están expresadas en términos de un resultado a alcanzar, lo cual no es limitativo para la invención, puesto que la solución propuesta incluiría no solo la invención aquí descrita, sino toda aquella realización que llegue a esos mismos resultados. Se sugiere hacer las aclaraciones necesarias, invocando las características esenciales de la invención que le permiten la consecución de dichos resultados.

1.1.9. En la reivindicación 22 se deben especificar las modificaciones y las posiciones en la secuencia de aminoácidos.

1.1.10. Una composición farmacéutica debe estar definida por su estructura en términos de sus elementos esenciales (compuesto activo y excipientes) y sus proporciones (reivindicaciones 27 y 28).

1.1.11. Las reivindicaciones de procedimiento deben referirse a unas actividades organizadas en una serie de pasos de manera determinada. El método para hacer un polipéptido SAP humano, de las reivindicaciones 36, 37 y 94 corresponde a un resultado a alcanzar, puesto que se refieren a la obtención de un polipéptido aislado, sin indicar la técnica para expresar el polipéptido en una célula, las condiciones necesarias para el cultivo de dicha célula, ni los procedimientos para aislar el polipéptido de la célula. Se sugiere caracterizar dicho método a partir de las etapas específicas que posee.

1.1.12. Las reivindicaciones de procedimiento deben referirse a unas actividades organizadas en una serie de pasos de manera determinada. El método para hacer un polipéptido SAP humano, de las reivindicaciones 62 y 79 corresponde a un resultado a alcanzar, puesto que se refieren a la obtención de un polipéptido, sin indicar la forma y las condiciones en las que se va a suministrar un polipéptido SAP, ni los procedimientos para lograr la alteración enzimática o química para producir dicho polipéptido. Se sugiere caracterizar dicho método a partir de las etapas específicas que posee.

1.2 Novedad

El Examinador plantea que, después de realizada la búsqueda de anterioridades, encontró documentos que afectan la novedad de las reivindicaciones 1-5, 9-13, 63-65, 68-70, 74-76-79-80, 83-85 y 89-91 de la presente solicitud.

Particularmente, el Examinador considera que los documentos Pepys *et al.* 1994, (D1) y mantzouranis *et al.* 1985 (D2), afectan la novedad de la presente solicitud por cuanto:

"S Las características esenciales del polipéptido de la reivindicación 1 habían sido divulgadas previamente en el documento D1, el cual da a conocer la glicoestructura del

componente SAP humano que consiste en un oligosacárido N-ligado biantenarico (Resumen) que puede contener dos terminaciones de ácido siálico o ser monosiliada siendo a 2,3 o a 2,6 (Pág. 5603 figura 1, Pág. 5604 resultados).

En consecuencia, la reivindicación 1 carece de novedad, según lo divulgado en el documento D1.

Las reivindicaciones dependientes 3, 5 y 9 - 13 no mencionan alguna característica que haga nueva a la composición de la reivindicación 1, por lo cual se considera que no son nuevas.

De igual manera, las características esenciales de la reivindicación 2, que consiste en un polipéptido glicosilado Amiloide P de suero humano (SAP) que comprende una cadena de oligosacáridos N-ligados, en donde la cadena de oligosacáridos tiene al menos 50% menos de partes de ácido siálico c α 2,6-ligadas, ya habían sido divulgadas en el documento D1.

Por lo tanto, la reivindicación 2 carece de novedad, según lo divulgado en el documento D1.

Las reivindicaciones dependientes 4 - 5 y 9 - 13 no mencionan alguna característica que haga nueva a la composición de la reivindicación 2, por lo cual se considera que no son nuevas.

(...)

Las características esenciales del método de la reivindicación 62 habían sido divulgadas previamente en el documento D1, el cual da a conocer modificaciones de SAP humano que consisten en la remoción de los ácidos siálicos de las terminaciones de las ramas de la cadena de oligosacáridos N-ligados, la desgalactosilación y la resialilación (Pág. 13).

Por lo tanto, la reivindicación 62 carece de novedad, según lo divulgado en el documento D1.

Las reivindicaciones 63 - 65, 68 - 70, 74 - 76, 79 - 80, 83 - 85, 89 o 91 no mencionan alguna característica que haga nueva al método de la reivindicación 62, por lo cual se considera que no son nuevas."

1.3 Nivel Inventivo

Sobre el requisito del nivel inventivo, el Examinador establece que una persona del nivel técnico medio, versada en la materia, ensayaría las soluciones que se proponen en los documentos Pepys et al. (D1) y AU 712819 B2 (D3) y llegaría de manera obvia a la

solución que se propone en la presente solicitud. Particularmente, el Examinador considera que las reivindicaciones 18-19 y 26-2 carecen de nivel inventivo debido a las enseñanzas contenidas en los documentos D1 y D3.

El Examinador establece:

“La secuencia SEQ ID NO 1 de la presente solicitud se encontró 99% idéntica a la secuencia de aminoácidos al componente de suero amiloide P humano aislado de la naturaleza, según reposa en GenBank: AAA60302.1 con fecha de ingreso a la base de datos de 9 de enero de 1995 y divulgado en el documento D2.

Por lo tanto, modificaciones de aminoácidos que resulten en una secuencia casi idéntica a un polipéptido tal como se encuentra en la naturaleza, no representa nivel inventivo, puesto que modificaciones o mutaciones de aminoácidos pueden ocurrir en o durante los procesos moleculares naturales de la célula.

En consecuencia, las reivindicaciones 18 y 19 no cumplen con el requisito de nivel inventivo.

• La reivindicación 26 reclama una preparación farmacéutica para uso en mamíferos que comprende el polipéptido SAB de las reivindicaciones 1 a 25 y un portador farmacéuticamente aceptable.

Las características esenciales de la reivindicación 26 ya habían sido divulgadas previamente en el documento D3, el cual enseña una composición farmacéutica que comprende un componente SAP nativo o recombinante y un vehículo farmacéuticamente aceptable que es formulada para administración nasal (Pág. 1 líneas 5 y 30).

De este modo, la reivindicación 26 carece de nivel inventivo. Las reivindicaciones dependientes 27 - 29 no adicionan características a la composición enseñada en el documento D3, por lo tanto tampoco cumplen con el requisito de nivel inventivo.”

2. Argumentos a favor de la patentabilidad de la solicitud

2.1 Reivindicaciones

En respuesta a las observaciones de su Despacho en relación con las reivindicaciones, el solicitante ha decidido presentar un nuevo pliego reivindicatorio en el cual se responde a todas y cada una de las inquietudes planteadas en la acción oficial.

Por favor note que las reivindicaciones 2, 4-8, 15-22, 24, 25, 27, 28, 30-36, 39, 42-96, 99 y 100 se han cancelado. Se han agregado las nuevas reivindicaciones 2-6.

El soporte para las reivindicaciones modificadas que se encuentran en este escrito aparece a continuación:

Numero de Reivindicación	Soporte en la descripción de la solicitud tal como se presentó
1	Reivindicaciones originales 1, 17 y 19
2	Página 5, reglones 4-11
3	Página 5, reglones 4-11
4	Página 5, reglones 4-11
5	Página 5, reglones 4-11
6	Página 5, reglones 4-11
7	Reivindicación original 3
8	Reivindicación original 9
9	Reivindicación original 10
10	Reivindicación original 11
11	Reivindicación original 12
12	Reivindicación original 13
13	Reivindicación original 14
14	Reivindicación original 23
15	Reivindicación original 26
16	Reivindicación original 29
17	Reivindicaciones originales 37 y 38
18	Reivindicación original 40
19	Reivindicación original 41
20	Reivindicación original 97
21	Reivindicación original 98

En relación con la observación a la reivindicación 1, en referencia con el hecho que un polipéptido debe ser reclamado por su estructura identificándolo con el SEC ID NO el solicitante no esta de acuerdo. Sin embargo, en un esfuerzo por hacer expedito el trámite de la presente solicitud, la reivindicación 1 se ha corregido para especificar que el polipéptido SAP comprende una secuencia de aminoácido que es al menos un 95% idéntica con la SEC ID NO 2:1. De acuerdo con esto, se considera que la corrección a la reivindicación responde a la objeción.

En relación con las observaciones a las reivindicaciones 1-4 y 9-16, sin estar de acuerdo con la observación, las reivindicaciones 2, 15 y 16 se han cancelado. El resto de las

reivindicaciones no cita las frases "comprende al menos uno", "tiene al menos 50% menos", "esta sustancialmente libre", respondiendo a la objeción planteada.

Con respecto a la frase *"en donde todas las ramas de la cadena de oligosacáridos termina con partes de ácido siálico α 2,3-ligadas"*, el solicitante dice de la opinión de su despacho y considera que no se ha apreciado completamente el conocimiento del estado de la técnica con relación a los N-glicanos. Tal como se enseñó en la presente solicitud, y como se evidenció en los anexos A-D presentados junto con este escrito, las estructuras del N-glicanos fueron bien caracterizadas en la técnica en el momento de la presentación. Ver, por ejemplo, la solicitud internacional publicada en las páginas 16, renglón 29, hasta página 17 renglón 20. Por ejemplo, la estructura de un N-glicano SAP se describe claramente en la citada referencia D1. Adicionalmente, si están presentes, las partes de ácido siálico se conocen por estar ligadas a los extremos terminales de las ramas del N-glicano. La estructura del ácido siálico era también bien conocida en la técnica al momento de la presentación. De acuerdo con esto, una persona medianamente versada en la materia al momento de la presentación de esta solicitud hubiera entendido fácilmente los límites de la cadena de oligosacárido (es decir N-glicano) que tiene al menos una rama del oligosacárido que termina con una parte de ácido siálico α 2,3 ligada, de acuerdo con lo anterior el solicitante de manera respetuosa solicita la aceptación de la anterior explicación.

Las reivindicaciones 18 y 19 se objetaron por cuanto definen la secuencia de polipéptido citada en términos de una identidad porcentual (por ejemplo al menos 85% idéntica con la SEC ID 2:1). Sin considerar aceptable tal objeción, las reivindicaciones 18 y 19 se han cancelado, respondiendo de esta manera a su objeción. Con respecto a las reivindicaciones modificadas, la reivindicación 1 esta dirigida a un polipéptido SAP que comprende, en parte, una secuencia de aminoácido que es al menos 95% idéntica con la secuencia de aminoácido de la SEC ID: 1. El solicitante considera que el SAP humano estaba bien definido en el estado de la técnica en el momento de la presentación de la presente solicitud y por lo tanto una persona medianamente versada habría entendido los límites de la invención actualmente reivindicada.

La especificación describe unas especies que caen dentro del alcance del género reivindicado y suministra una variedad de ensayos adecuados para probar otras variantes SAP con el fin de determinar si ellas son funcionalmente activas. Ver, por ejemplo, solicitud

de patente internacional (WO2010/141918 en las páginas 45-52).

Además, la especificación enseña que el SAP pertenece a una familia bien conocida de proteínas de pentaxina que se caracterizan por una estructura pentamérica cíclica. Ver, por ejemplo, la solicitud de patente publicada páginas 12 y 13. La especificación también suministra la secuencia para un promotor SAP humano (es decir, la SEC ID: 1) Así como también las variantes SAP activas de ocurrencia natural encontradas en otras especies. Como se demostró en el anexo E (exhibit E) un promotor SAP humano (es decir la SEC ID: 1) es homólogo al SAP del pollo (es decir, la SEC ID: 1 es 48% idéntica a la SEC ID: 2 de la presente solicitud), el SAP de bovino (es decir, la SEC ID 2:1 es 75% idéntica a la SEC ID: 3 de la presente solicitud), y el SAP de hámster (es decir, la SEC ID: 1 es 67% idéntica a la SEC ID:4 de la presente solicitud). Con base en la amplia variedad de SAP conocidos, el solicitante considera que una persona medianamente versada en la materia hubiera podido alinear tales secuencias y determinar que residuos se conservan entre los homólogos y así la probabilidad de ser importante en la SAP. Además, estos homólogos son ejemplos de los polipéptidos SAP funcionales y por lo tanto suministran especies representativas de las variantes SAP, que tienen una identidad porcentual a un inferior a la SEC ID:1 que está siendo actualmente reivindicada.

Además, el solicitante considera que el SAP era una proteína bien caracterizada al momento de presentarse esta solicitud. En particular, Emsley et al. (Nature, 357:338-345 (1994), presentada como anexo F (exhibit F) fue el primero en reportar la estructura de alta resolución del SAP humano pentamérico. Con base en estos estudios y otros similares (ver, por ejemplo, Hohenester e al., Mol. Biol. 269:570-578 (1997); Srinivasan et al., Structure 2: 1017-1027 (1994); y Brinda et al., Biochem J. 391:1-15 (2005), todos presentados como anexos, G-I), el solicitante considera que los residuos de aminoácido que son importantes para mantener la integridad estructural de cada promotor, así como también los residuos de contacto esencial que median las interacciones promotor - promotor eran bien conocidos en el estado de la técnica. Como primer miembro de la familia en ser resuelto a alta resolución, la estructura del SAP ha sido utilizada en estudios de alineamiento con otra variedad de proteínas de pentaxina evolucionadamente distintas (ver, por ejemplo, anexos H e I). Estos estudios suministran un análisis detallado de los residuos conservados en estas proteínas, particularmente dentro del bolsillo de unión a ligando SAP. Adicionalmente, estos reportes indican claramente importantes regiones estructurales que son necesarias para

la actividad SAP. Por ejemplo, el SAP es conocido por ser una proteína de unión dependiente del calcio que interactúa con una variedad de diferentes ligandos que incluyen, por ejemplo, fibronectina, proteínas que se unen a C4, ADN y glicosaminoglicanos. Tal como se describe en los anexos F-I, el solicitante considera que varios estudio de mapeo de cristalización y computadora han elucidado los residuos de aminoácido claves dentro del SAP que median la unión a estos varios ligandos. De acuerdo con esto, el solicitante asevera que las características estructurales y funcionales de SAP estaban bien caracterizadas en el estado de la técnica al momento de la presentación de esta solicitud.

La biología molecular es un campo científico avanzado y el nivel de un experto en la técnica al momento de la presentación de la solicitud era muy alto. Los métodos para generar variantes de polipéptido que incluyan una o más mutaciones eran bien conocidos en la técnica y así también la rutina. Con base en los ejemplos suministrados en la especificación y en vista del conocimiento en la técnica, la persona medianamente versada habría podido prever otros polipéptidos que cayeran dentro del alcance de las reivindicaciones. De acuerdo con esto, los ejemplos de trabajo ponen al solicitante en posesión de la invención reivindicada. De acuerdo con lo anterior se solicita la aceptación de las reivindicaciones.

Finalmente, el solicitar la definición sin utilizar porcentajes razonables de identidad daría como resultado que la invención se circunvalara de manera muy fácil por algún experto que quisiera utilizar la invención sin violar el derecho. Es decir, que si se reivindica una secuencia específica alguien podría con una homología del 95% obtener los mismos resultados sin que se violara la patente ya que lo que se reivindicó fue una secuencia específica. Esto muestra que reivindicar una secuencia específica lo que haría sería abrirle el campo a un experto para que utilizara los conocimientos aportados y el problema resuelto por esta invención sin violar ningún derecho. De acuerdo con lo anterior se le sugiere a su Despacho ser razonable en el análisis de la presente invención especialmente en lo que se refiere a los porcentajes de identidad.

En relación con las reivindicaciones 36, 37 y 94 se rechazaron por que presuntamente no eran claras. Específicamente, su Despacho asevera que los métodos de las reivindicaciones de manufactura *"deben referirse a unas actividades organizadas en una serie de pasos de una manera determinada...indicando la técnica para expresar el polipéptido en una célula, las condiciones necesarias para cultivar dicha célula, o el proceso para aislar los*

polipéptidos de las células." Sin estar de acuerdo con el rechazo de las reivindicaciones 96 y 94 éstas se han cancelado respondiendo de esta manera a la objeción. Con respecto a la reivindicación 37 rechazada (actual reivindicación 17), de manera respetuosa el solicitante no esta de acuerdo y considera que incluir limitaciones adicionales en las condiciones de cultivo y las técnicas de aislamiento disminuiría de manera indebida el alcance de las reivindicaciones en vista del desarrollo del solicitante.

En primer lugar, al solicitante le gustaría puntualizar que las presentes reivindicaciones no están dirigidas a nuevos métodos para generar y aislar proteínas recombinantes en general. Las presentes reivindicaciones todas están dirigidas a una glicovariante SAP recombinante que comprende una cadena de oligosacárido N-ligada que tiene al menos una rama de la cadena de oligosacárido que termina en una parte de ácido ciálico $\alpha 2,3$ -ligada que se caracteriza por una actividad biológica incrementada, las composiciones farmacéuticas que comprenden las misma y el método para generar tales polipéptidos. Esta glicovariante SAP se expreso primero y se purifico de células CHO. Ver, por ejemplo, la solicitud internacional publicada en las páginas 59 a 62. De acuerdo con esto la actual reivindicación 17 esta dirigida a un método para elaborar un polipéptido SAP humano que comprende de una cadena de oligosacárido, en donde al menos una rama de la cadena de oligosacárido termina con una parte de acido siálico $\alpha 2,3$ -ligada, en donde el método comprende las etapas de: 1) Expresar un polipéptido SAP humano en una célula CHO; y 2) Aislar el polipéptido SAP humano de la célula.

El solicitante considera que la biología molecular es un campo científico avanzado y que el nivel de una persona versada en la técnica al momento de presentar la solicitud era muy alto. Tal como se enseñó en la presente solicitud numerosos métodos para expresar y purificar el recombinante de una proteína de una célula eran bien conocidos en la técnica y en la rutina practicada en las personas versadas. Ver, por ejemplo, solicitud internacional publicada en las páginas 23 a 32. Con base en los ejemplos suministrados en la especificación y en vista del conocimiento del estado de la técnica, la persona experta había podido prever las diferentes condiciones de cultivo y los procesos de aislamiento que pudieran producido la proteína recombinante actualmente mencionada. De acuerdo con esto tal como se evidenció mediante la enseñanza de la presente solicitud, una persona versada en la materia habría entendido fácilmente los limites de la presente invención reivindicada. Las reivindicaciones 5-8, 15, 16, 20-22, 24, 25, 27, 28, 36, 45, 52, 53, 61, 62, 72-79 y 91-96

se rechazaron por ser poco claras. Sin aceptar el rechazo, estas reivindicaciones se han cancelado, respondiendo de esta manera a la objeción planteada.

En relación con la expresión "al menos" el solicitante desea manifestar que dicha expresión no es de ninguna manera ni ambigua ni imprecisa.

A este respecto me permito anotar la observación que aparece en la Guía para el Examen de solicitudes de patentes de invención y modelos de utilidad utilizada por ese Despacho en donde se dice que:

"El examinador presentará una objeción cuando las reivindicaciones incluyan términos imprecisos del tipo "aproximadamente", "alrededor de", "opcionalmente" ya que en ese caso el alcance y ámbito de protección de la reivindicación deja de ser precisa, y no permite una comparación con el estado de la técnica. Se pueden permitir características no esenciales, cuando esté justificado y siempre y cuando permitan distinguir el estado de la técnica, sin ambigüedad.

Por las mismas razones no se admiten términos relativos del tipo "mayor", "delgado", "fuerte, ya que no tienen un significado preciso. En ningún caso estos términos se pueden utilizar para distinguir la invención del estado de la técnica. En estos casos, el examinador requerirá que estas expresiones se sustituyan por términos precisos, o rangos concretos de valores."

La pregunta que sigue es si la expresión o término "al menos" es relativa o imprecisa en vista de que no aparece mencionada en dicha Guía.

Revisemos entonces el significado de la expresión. El Diccionario Prehispánico de dudas de la Real Academia define la expresión "al menos" de la siguiente forma:

"al menos, a lo menos. Locuciones adverbiales que introducen una rectificación o salvedad a lo que se acaba de decir: «No aparecían por allí casi nunca, al menos que yo recuerde» (Mendicutti Palomo [Esp. 1991]); «Eran [las estatuas] de gestos feos y disformes; a lo menos, l[a]s que yo he visto, todas eran así» (DHoy [Ec.] 12.12.96). También significan 'aunque solo sea, como mínimo': «Ahora déjame, al menos, morir en paz» (Gasulla Culminación [Arg. 1975]); «En cada comuna existirá, a lo menos, una Oficina Inscriptora» (Hoy [Chile] 27.10-2.11.97). De las dos, la más usada es al menos. Ambas son equivalentes

de por lo menos.” (resaltado fuera de texto)”

Y en cuanto a la definición de la palabra *"salvedad"* el Diccionario de la real Academia lo define como:

“Razonamiento o advertencia que se emplea como excusa, descargo, limitación o cortapisa de lo que se va a decir o hacer.”

Por su parte el Diccionario de uso del español María Moliner establece lo siguiente en relación con la expresión *“al menos”*:

*“Expresa la restricción a un mínimo de un algo que se pide.” Y luego dice: “Cuando menos, por lo menos → Expresiones *restrictivas.”*

De acuerdo con lo anterior es claro que tenemos que la expresión *“al menos”* es una expresión que introduce una limitación o restricción. ¿Cómo puede entonces una expresión que introduce una limitación o restricción ser imprecisa? Dicho de otra forma ¿cómo algo que esta limitado puede ser indefinido? (La Real academia define la palabra impreciso como indefinido).

Esta claro que hay una abierta y clara contradicción. Esta contradicción lo que muestra de manera categórica es que la expresión *“al menos”* no es de ninguna manera una expresión imprecisa. De hecho hace lo contrario: limita o restringe.

Ésta es una expresión que tiene un muy amplio uso en el lenguaje de patentes a nivel mundial y es aceptada de manera generalizada.

2.2. Novedad

La novedad de una solicitud está legalmente definida en el Artículo 16 de la Decisión 486 así: *“una invención se considerará nueva, cuando no está comprendida en el estado de la técnica. El estado de la técnica comprenderá todo lo que haya sido accesible al público por una descripción escrita u oral, utilización, comercialización o cualquier otro medio antes de la fecha de presentación de la solicitud de patente o, en su caso de la prioridad reconocida...”*

En virtud de la anterior definición, consideramos que los documentos D1 y D2 no constituyen documentos del arte previo que afecten la novedad de la presente solicitud.

Las reivindicaciones 1-13, 62-65, 68-70, 74-76, 79, 80, 83-85, 89-91 fueron rechazadas por cuanto supuestamente la anticipaba el documento Pepys et al. (1994) PNAS, 91: 5602-5606 (denominada en lo sucesivo como "D1"). Sin aceptar el rechazo, las reivindicaciones 2, 4-8, 62-65, 68-70, 74-76, 79, 80 83-85 y 89-91 se han cancelado, respondiendo a esta manera la objeción a su despacho. Con respecto a las reivindicaciones restantes el solicitante de manera respetuosa no esta de acuerdo. Sin embargo, en un esfuerzo para hacer expedito el trámite de la presente solicitud, se ha corregido la reivindicación 1, en parte, para incorporar la materia objeto de la reivindicación 17. De acuerdo con esto, la presente reivindicación esta dirigida a un polipéptido SAP humano recombinante glicosilado. Como la reivindicación 17 no fue rechazada las correcciones de la reivindicación se considera que responden al rechazo. El solicitante por lo tanto requiere, de manera respetuosa, la aceptación de estos argumentos.

2.3. Nivel inventivo

El nivel inventivo de una solicitud se encuentra definido en el Artículo 18 de la Decisión 486 y se establece cuando *"si para una persona del oficio normalmente versada en la materia técnica correspondiente, esa invención no hubiese resultado obvia ni se hubiese derivado de manera evidente del estado de la técnica"*.

Contrario a lo que plantea el Examinador, el documento del arte previo citado no afecta el nivel inventivo de la presente solicitud por cuanto no es posible derivar la presente invención a partir de las enseñanzas presentadas en los documentos D1 y D3.

Las reivindicaciones 18, 19, 26-29 se rechazaron por cuanto se alego la falta de nivel inventivo en vista del documento D1 y D3. Sin estar de acuerdo con el rechazo, las reivindicaciones 18, 19, 27-29 se han cancelado, respondiendo de esta manera la objeción. Con relación al resto de las reivindicaciones objetadas, el solicitante de manera respetuosa manifiesta su desacuerdo. Sin embargo, tal como se indico anteriormente, la reivindicación 1 se ha corregido en parte para incorporar la materia objeto de la reivindicación 17 anterior. De acuerdo con esto la actual reivindicación 1, esta dirigida a un polipéptido SAP humano

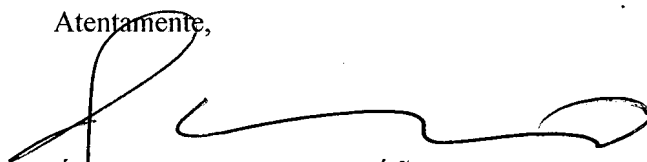
recombinante glicosilado. La reivindicación 29 (reivindicación actual 16) depende de 26 y especifica además rutas alternativas para administrar la preparación farmacéutica (es decir, tópicamente, mediante inyección, mediante inyección intravenosa, mediante inhalación, o mediante deposito continuo o bomba). Tal como se indico anteriormente la reivindicación 17 se reconoció implícitamente como novedosa en vista de la técnica anterior. Además, la reivindicación 17 no fue objeto de rechazo por nivel inventivo. De acuerdo con esto las correcciones de las reivindicaciones han respondido a la objeción planteada es decir nada en la técnica anterior citada enseña o sugiere una preparación farmacéutica que comprenda el polipéptido SAP humano recombinante glicosilado. El solicitante por lo tanto solicita la reconsideración y la aceptación de las actuales reivindicaciones.

Materia objeto no patentable

Las reivindicaciones 30-35 y 100 se rechazaron por cuanto se consideró que estaban dirigidas a materia objeto no reivindicada específicamente, su despacho asevera que los métodos de tratamiento terapéutico que cubren humanos no son reivindicables actualmente en Colombia. Por lo tanto las reivindicaciones 30-35 y 100 se han cancelado.

En virtud de lo anterior, solicito que se acepten las razones expuestas en el presente documento y en consecuencia, se otorgue el privilegio de patente de invención para el invento que se tramita en este expediente.

Atentamente,



ÁLVARO CORREA ORDÓÑEZ

C.C. No. 79.143.366 de Usaquén

T.P. 20.281

Anexos: - Anexos A-D, E, F, G-I
 - Capítulo reivindicatorio modificado
 - Recibo por reivindicaciones adicionales

15

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO

NIT : 800.176.089-2

- / -

 Industria y Comercio SUPERINTENDENCIA	RECIBO DE CAJA	No. 13 - 0095216
		Bogotá D.C., Septiembre 20 de 2013 - 16:26:55

RECIBIDO DE : BAKER	NI 900.532.339
---------------------	----------------

*** Soporte del Pago ***					
TIPO PAGO	BANCO	CUENTA	No. PAGO	FECHA PAGO	VR PAGO
RECIBO DE CA	999999999	95211	20/09/2013		3.475.000.00

*** Conceptos Pagados ***			
CANT. RENTISTICO	CONCEPTO	Vr.UNDITARIO	Vr.CONCEPTO
5 50005-01-01 SOLICITUDES	1607 REIVINDICACION PATENTE UNITARIA ADICIONAL A LAS 10 INIC	30.000.00	150.000.00
			\$150.000.00
=====			

SON: ****CIENTO CINCUENTA MIL PESOS MONEDA CORRIENTE*****

Responsable: _____

Recibo de Caja Aplicado al Expediente No. _____

REIVINDICACIONES

1. Un polipéptido glicosilado Amiloide P de suero humano (SAP) recombinante que comprende una cadena de oligosacáridos N-ligados, en donde una rama de la cadena de oligosacáridos termina con una parte de ácido siálico α 2,3-ligada, y en donde el polipéptido SAP comprende una secuencia de aminoácido que es al menos 95% idéntica con la secuencia de aminoácido de la SEQ ID No.: 1.
2. El polipéptido SAP glicosilado humano de la reivindicación 1, en donde el polipéptido SAP comprende una secuencia de aminoácido que es al menos 96% idéntica con la secuencia de aminoácido de la SEQ ID NO.: 1.
3. El polipéptido SAP glicosilado humano de la reivindicación 1, en donde el polipéptido SAP comprende una secuencia de aminoácido que es al menos 97% idéntica con la secuencia de aminoácido de la SEQ ID NO.: 1.
4. El polipéptido SAP glicosilado humano de la reivindicación 1, en donde el polipéptido SAP comprende una secuencia de aminoácido que es al menos 98% idéntica con la secuencia de aminoácido de la SEQ ID NO.: 1.
5. El polipéptido SAP glicosilado humano de la reivindicación 1, en donde el polipéptido SAP comprende una secuencia de aminoácido que es al menos 99% idéntica con la secuencia de aminoácido de la SEQ ID NO.: 1.

6. El polipéptido SAP glicosilado humano de la reivindicación 1, en donde el polipéptido SAP comprende la secuencia de aminoácido de la SEQ ID No.: 1
7. El polipéptido SAP glicosilado humano de una cualquiera de las reivindicaciones 1-6, en donde todas las ramas de la cadena de oligosacáridos termina con partes de ácido siálico α 2,3-ligadas.
8. El polipéptido SAP glicosilado humano de una cualquiera de las reivindicaciones 1 a 7, en donde la cadena de oligosacáridos N-ligados comprende un núcleo penta sacárido de $\text{Man}[(\alpha 1,6)-(\text{MAN}(\alpha 1,3))-\text{Man}(\beta 1,4)-\text{GlcNAc}(\beta 1,4)-\text{GlcNAc}(\beta 1,N)-\text{Asn}]$.
9. El polipéptido SAP glicosilado humano de una cualquiera de las reivindicaciones 1 a 7, en donde la cadena de oligosacáridos comprende una rama que tiene la estructura $\text{NeuNAc}2\alpha 3\text{Gal}\beta 4\text{GlcNAc}\beta 2\text{Man}\alpha 6$.
10. El polipéptido SAP glicosilado humano de una cualquiera de las reivindicaciones 1 a 9, en donde al menos una rama de la cadena de oligosacáridos está libre de galactosa y N-acetilglucosamina.
11. El polipéptido SAP glicosilado humano de la reivindicación 10, en donde todas las ramas de la cadena de oligosacáridos está libre de galactosa y N-acetilglucosamina.

12. El polipéptido SAP glicosilado humano de una cualquiera de las reivindicaciones 1 a 11, en donde al menos una rama de la cadena de oligosacáridos comprende un residuo de manosa.

13. El polipéptido SAP glicosilado humano de una cualquiera de las reivindicaciones 1 a 12, en donde la cadena de oligosacáridos comprende un residuo de fucosa.

14. El polipéptido SAP glicosilado humano de una cualquiera de las reivindicaciones 1-13, en donde la proteína SAP comprende un aminoácido PEGilado, un aminoácido prenilado, un aminoácido acetilado, un aminoácido biotinilado, ó un aminoácido conjugado con un agente derivatizante orgánico.

15. Una preparación farmacéutica adecuada para uso en mamíferos que comprende el polipéptido SAP humano de una cualquiera de las reivindicaciones 1 a 14 y un portador farmacéuticamente aceptable.

16. La preparación farmacéutica de la reivindicación 15, en donde la preparación farmacéutica es adecuada para la administración a un paciente tópicamente, por inyección, por inyección intravenosa, por inhalación, o por depósito continuo ó bomba.

17. Un método para hacer un polipéptido SAP humano, que comprende una cadena de oligosacárido, en donde al menos una rama de la cadena de oligosacárido termina con una parte de ácido sialico α 2,3-ligada, en donde el método comprende las etapas de:

i) expresar el polipéptido SAP humano en una célula CHO; y

ii) aislar el polipéptido SAP humano de la célula.

18. El método de la reivindicación 17, en donde todas las ramas de la cadena de oligosacárido terminan con partes de ácido siálico α 2,3-ligadas.

19. El método de la reivindicación 17 o 18, en donde la cadena de oligosacáridos está libre de partes de ácido siálico α 2,6-ligadas.

20. Una célula CHO que comprende un polipéptido SAP humano que comprende una cadena de oligosacáridos N-ligados, en donde al menos una rama de la cadena de oligosacáridos termina en una parte de ácido siálico α 2,3-ligada.

21. Una célula CHO que comprende un ácido nucleico que codifica para un polipéptido SAP exógeno.

BM_RAC/BOGDMS-#1689991-V1-P2011_000047-REIVINDICACIONES_MODIFICADAS.DOCX

Structural Aspects of Carbohydrates and the Relation with their Biological Properties

Miloš Hricovíni*

Institute of Chemistry, Slovak Academy of Sciences, 845 38 Bratislava, Slovakia

Abstract: Structural diversity of carbohydrates plays a crucial role in their large variety of roles in biological systems. This paper focuses on aspects of structure and biological functions of three classes of carbohydrates, N-linked oligosaccharides, blood group oligosaccharides and glycosaminoglycans. Conformations and dynamics in solution, as well as structure of protein-carbohydrate complexes are discussed. A short overview also describes theoretical and experimental methodologies that are used in current glycobiological research, particularly high-resolution NMR spectroscopy, X-ray crystallography and methods of computational chemistry.

INTRODUCTION

Carbohydrates and their conjugates belong to essential biomolecules of life. They can be found in all living systems, from plants to humans. Probably the most well-known carbohydrate is the structurally simple polysaccharide cellulose (β -(1-4)-linked glucopyranose), which is also the most abundant biomolecule in nature. Although cellulose and structurally similar compounds, such as xylan or starch are important polysaccharides, current interests of chemists and biologists are focused on saccharides having biological properties, predominantly those that can be used in medical chemistry.

The diversity of biological properties of saccharides are surprisingly large spanning from sources of energy in living systems, to cell adhesion, growth and differentiation, anti-inflammatory and anticoagulant properties, angiogenesis, etc [1]. This large variety of properties originates from the structural arrangement of carbohydrates. Carbohydrates can form linear polymeric forms, like the previously mentioned cellulose, but also very complex branched structures or conjugates with other type of molecules, most importantly, proteins. In glyco-conjugates, e.g. glyco-proteins, the saccharide part can be very small (consisting of even a single monosaccharide unit) up to conjugates where the saccharide part is larger than the protein structure itself.

The structure of monosaccharides, pyranoses or furanoses, allows creation of linkages in various positions. For example, hexopyranoses can be linked with other monosaccharides in positions 1, 2, 3, 4 and 6 (Chart 1), thus a large number of possible glycosidic linkages (one or more linkages at the same time) can be formed among monosaccharide units. Just three different monosaccharide units (hexopyranoses) can form up to nearly 28000 different structures. This considerably differs from peptides where only 6 tripeptides can be formed from three different amino acids. The ability to create complex branched forms and the different linkage types is a unique structural property of carbohydrates with respect to peptides and nucleic acids and,

consequently, result in their functional diversity in biological systems.

Conformation is another feature that has the substantial influence on biological activities. 3D structure of carbohydrates is complicated and originates from their complex constitution. The overall conformation is determined predominantly by the stereochemistry at glycosidic linkages, conformation of pendant groups and the pseudorotation of pyranose and furanose rings.

As carbohydrates are nearly exclusively present in the ring forms, stereo-chemical arrangement of carbohydrates is affected by ring puckering [2]. Pyranoses adopt the 4C_1 (Chart 1) or 1C_4 chair forms in most cases whereas other conformations are relatively rare. However, large substituents can have a pronounced influence on the ring conformation and the boat or the skew conformers can be present in substituted derivatives. Furanose forms are more flexible with respect to pyranose forms and the populations of conformers depend on the structure of neighboring units and substitution. This is also valid for deoxy derivatives, such as 2-deoxy- β -D-erythro-pentose, which is the component of DNA structure. Furthermore, both D and L forms exist in oligo- and polysaccharides, the D form being more abundant. The L form is typical for some residues, e.g. L-fucose in blood group oligosaccharides or L-iduronic acid in glycosaminoglycan compounds.

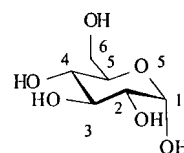


Chart. (1).

Conformation of the hydroxymethyl group is described by the torsion angle ω (ω = O-5-C-5-C-6-O-6). Three stable conformers at the C-5-C-6 linkage are designated as gauche-gauche, *gg*, (ω = -60°), gauche-trans, *gt*, (60°) and trans-gauche, *tg*, (180°) (Chart 2). The first symbol refers to the torsion angle O-5-C-5-C-6-O-6, and the second to C-4-C-5-C-6-O-6. The preference of these conformers depends on the type of monosaccharide and substituents. For example, the gauche effect (the non-bonded interaction between the equatorially oriented hydroxyl group at C-4 with the O-6

*Address correspondence to this author at the Institute of Chemistry, Slovak Academy of Sciences, 845 38 Bratislava, Slovakia; Tel: +4212 5941 0323; Fax: +4212 5941 0222; E-mail: hricovini@savba.sk

oxygen in the *tg* conformer) influences the conformational equilibrium at this linkage and, consequently destabilizes the *tg* conformation in hexopyranoses, such as glucose or mannose [2, 3].

Arrangements of other pendant groups, such as O-acetyl, N-acetyl, carboxylate, N-sulfate, etc. have not been studied in details. Relatively little knowledge about the conformational equilibria of hydroxyl groups though their conformations are crucial for understanding of hydrogen bonds formation (intra- or intermolecular) or recognition processes.

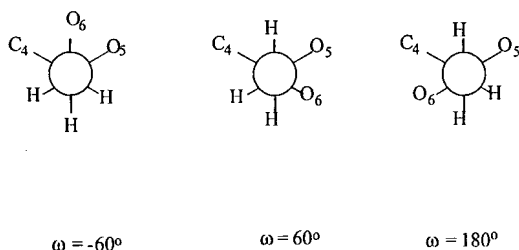


Chart. (2).

Probably the most complicated, and the most discussed, is conformation at the glycosidic linkages. The mutual orientation of neighboring saccharide units is described by the torsion angles ϕ , ψ ($\phi = \text{O}-5'-\text{C}-1'-\text{O}-1'-\text{C}-\text{x}$; $\psi = \text{C}-1'-\text{O}-1'-\text{C}-\text{x}-\text{C}-\text{x}-1$) (Chart 3); the orientation defined by ϕ_{H} , ψ_{H} ($\phi_{\text{H}} = \text{H}-1'-\text{C}-1'-\text{O}-1'-\text{C}-\text{x}$; $\psi_{\text{H}} = \text{C}-1'-\text{O}-1'-\text{C}-\text{x}-\text{H}-\text{x}$) is also often presented in literature and will be used in the present paper. The ϕ , ψ angles can substantially affect the overall shape of oligo- and polysaccharides, therefore, the determination of their values one of the key elements analyzed when dealing with 3D carbohydrate structure. This also comprises an answer for the question whether single or multiple conformers are present at the glycosidic linkages. The above mentioned phenomenon has been thoroughly studied and discussed in terms of "flexibility" and "rigidity" and numerous theoretical and experimental studies [3-5] have been devoted to solve this dilemma.

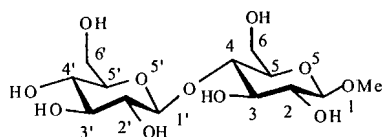


Chart. (3).

In addition to 3D structures, that describe static pictures of compounds or intermolecular complexes in solid state or in solution (theoretical analysis that allows the determination of the lowest energy states of molecules), molecular dynamics provide information on the rates of overall and internal motions as well as the extent of internal motions [6]. Results of several studies have shown that binding of ligand to proteins cannot be fully explained by static pictures [7, 8]. Complete understanding of the biological properties of carbohydrates at the molecular level thus requires the knowledge of their 3D structure and dynamics. The present review deals with these features in three groups of biologically active carbohydrates: N-linked oligosaccharides, blood group oligosaccharides and glycosaminoglycans and their complexes with proteins.

EXPERIMENTAL AND THEORETICAL METHODS

Molecular Mechanics, Molecular Dynamics and Quantum Chemical Methods

Theoretical analysis provides information on structure of carbohydrates in solution, or in solid state, and is mainly used for interpretation of experimental data, such as high-resolution NMR NOEs or coupling constants (both scalar and dipolar) [9-11]. Molecular modeling is also important in analysis of saccharide dynamics and their complexes with proteins.

First theoretical analyses have been based on simple computational schemes, where conformations at the glycosidic linkages were accepted, or rejected, based on the overlap of atoms belonging to different monosaccharide units linked by the glycosidic linkage [12]. This approach has been extended evaluating the exo-anomeric effect [13] and applied in analysis of oligosaccharides related to Lewis blood group determinants [14, 15]. However, it was soon clear that the "rigid-body" approach could not fully explain experimental observations. More recent methodologies have been based mainly on molecular mechanics (MM). In this approach, the molecule is treated as an ensemble of atoms and the interaction among these atoms is described by Newton mechanics. The total energy is then expressed as the sum of contributions of the bond stretching (V_b), the angle bending (V_θ), torsional terms (V_ϕ) and non-bonding interactions ($V_{\text{vanderWaals}} + V_{\text{elstat}}$) [16]:

$$V_{\text{total}} = \sum_{\text{bonds}} V_b + \sum_{\text{angles}} V_\theta + \sum_{\text{torsions}} V_\phi + \sum_{i=1}^N \sum_{j=i+1}^N (V_{\text{vander Waals}} + V_{\text{elstat}})$$

where, e.g. the term for bond stretching, $V_b = \frac{1}{2} k_b (r - r_0)^2$. k_b is the force constant, $r - r_0$ is the difference between interatomic distance and the reference bond length. As the force constants vary in individual terms for different types of molecules (e.g. proteins, carbohydrates), various force-field methods have been presented. A relatively large number of force fields have been applied in the modeling of carbohydrates, for example MM2 [16], MM3 [17], AMBER [18], CFF91 [19], CVFF [20], PEF422 [21], etc. Molecular mechanics (MM) and molecular dynamics (MD) gave reliable results in many cases and could explain satisfactorily the experimental data obtained for saccharides in solution and for protein-carbohydrate complexes [3-5, 22-24]. MM and MD, at present, are prevalently used for modeling of large molecules, such as polysaccharides or protein-carbohydrate complexes where the methods of quantum chemistry are still not applicable.

The last group of theoretical approaches, the quantum chemical methods, has until now been used mainly to model small or medium-sized oligosaccharides. First calculations of carbohydrate structures were based on semi-empirical methods [3, 25]; however, more rigorous methods, *ab initio* and density functional theory (DFT) methods are now largely used [26]. Within the first group of methods, the use of the quality basis sets (e.g. 6-31G**, or 6-311++G) yield energies and geometries that reliably interprets the experimental data. The disadvantage of *ab initio* methods is a demand on computing time, so their applications are still rather limited. DFT methods appear to be better for

calculation of molecular energies and geometries as time requirements are lower than *ab initio* [27]. DFT is at present therefore becoming the popular approach in carbohydrate chemistry [28-30]. In addition, the development of methods, that enable calculation of NMR parameters (chemical shifts and scalar spin-spin coupling constants [31]), have expanded the possibilities to compare theoretical data with experimental observations.

High-Resolution NMR

High-resolution NMR spectroscopy is the main experimental tool used for carbohydrate structural analysis in solution. The 3D structure and dynamics in solution is analyzed from all NMR parameters, chemical shifts, scalar and dipolar coupling constants, NOEs and relaxation times. ^1H and ^{13}C chemical shifts are typically used in primary and secondary structure determination as well as anomeric configuration. Three-bond proton-proton ($^3J_{\text{H-H}}$) and proton-carbon ($^3J_{\text{C-H}}$) coupling constant are often used in 3D structural analysis as they depend on the stereochemical arrangement of bonded atoms. $^3J_{\text{H-H}}$ constants are particularly useful in analysis of pseudorotation of pyranose and furanose rings, hydroxymethyl group conformation. Typical value of $^3J_{\text{H-H}}$ is about 8 Hz for antiperiplanary arranged protons in pyranose rings, whereas $^3J_{\text{H-H}} \approx 3.5$ Hz is the value for the gauche arrangement. Conformation at the glycosidic linkages can be deduced from $^3J_{\text{C-H}}$ values measured for the arrays of H-1'-C-1'-O-1'-C-x atoms (evaluation of the ϕ angle) and the C-1'-O-1'-C-x-H-x atoms (evaluation of the ψ angle). This type of analysis, however, also requires the proper theoretical interpretation of spectral data based on the computed molecular geometry and coupling constants obtained from the parametrized dependences [32, 33].

The magnitude of one-bond proton-carbon coupling constants ($^1J_{\text{C-H}}$) depends on the anomeric configuration. $^1J_{\text{C-H}}$ is ~ 170 Hz in the α -anomer and about 160 Hz in the β -anomer [34, 35]. This difference is often utilized in determination of anomeric configuration by NMR spectroscopy. More recently, theoretical [28, 36-39] and experimental [40, 41] evidences further indicated that $^1J_{\text{C-H}}$ coupling constants can have more general applications as their magnitudes depend also on ϕ and ψ angles at the glycosidic linkages. As oxygen lone pairs orbitals interact with the molecular orbitals localized in the C-H bond, the bond orders change, and consequently, the indirect spin-spin couplings vary. Since these effects are very sensitive to molecular geometry, variations in $^1J_{\text{C-H}}$ magnitudes can be large, up to 15 Hz.

Residual dipolar coupling constants ($^1D_{ij}$) in weakly aligned molecules is another parameter used in conformational analysis [42-44]. Residual dipolar coupling constants (two-bond ^1H - ^1H or one-bond ^1H - ^{13}C) depend on the $\langle r_{ij}^{-3} \rangle$ factor (r is the distance between dipolarly coupled nuclei i and j) as well as the orientation relative to magnetic field. The magnitudes of $^1D_{ij}$ can be thus related with the spatial arrangement of molecules (particularly long-range interactions due to the $\langle r_{ij}^{-3} \rangle$ factor) and complement the results obtained from other NMR methods.

NOEs are probably the most often used NMR parameter in 3D structural analysis of oligo and polysaccharides. Due

to the $\langle r_{ij}^{-6} \rangle$ dependence (r is the distance between the two cross-relaxing protons i and j), NOE magnitudes are sensitive upon the 3D structure. NOEs are frequently used in analyses of conformations at the glycosidic bonds, hydroxymethyl groups or pseudorotation of the pyranose and the furanose rings. However, similarly to $^3J_{\text{C-H}}$ values, proper interpretation of NOE data requires the calculation of theoretical NOEs (computed from the modeled molecular structure) and the comparison with experimental data [45, 46]. NOEs are at present, also often used in studies of protein-carbohydrate complexes through evaluation of transferred NOEs (TR-NOEs) [47-49] and saturation transfer difference (STD) [50, 51] and interpreted with appropriate programs [52, 53].

Dynamics of carbohydrates in solution is analyzed from ^1H or ^{13}C relaxation rates and NOEs [6, 54-56]. In most cases, two-dimensional HSQC-type experiments are carried out at different magnetic fields and T_1 and T_2 data obtained by non-linear fitting to cross-peak intensities. Motional molecular properties are derived from experimental data applying appropriate theoretical models [57, 58].

A large variety of experimental NMR methods, from conventional one-dimensional (1D) to three-dimensional (3D), can be used for determination of chemical shifts, scalar and dipolar coupling constant, NOEs and relaxation times. Two-dimensional (2D) COSY, TOCSY, HSQC and HMBC are the most used methods in the structural analysis of carbohydrates [9, 11]. $^1J_{\text{C-H}}$ (in isotropic liquids) or $^1D_{\text{C-H}}$ (in weakly aligned systems) values can be measured directly from HSQC or E COSY type spectra. As the determination of couplings constants requires high precision, several methodologies, based on the fitting to the intensity of the cross-peaks, have been developed [42].

Although enough precise $^3J_{\text{H-H}}$ values are mostly determined from conventional ^1H spectra, the overlap of ^1H resonances can be severed in carbohydrate spectra and selective 1D methods (e.g. 1D TOCSY) or 2D spectra (COSY or TOCSY) are then easier to apply [59]. $^3J_{\text{C-H}}$ values can be measured by various 2D methods, J-resolved methods [60], constant-time techniques [61] or gradient-enhanced spectroscopy [62]. The precision of these experiments is sufficient to obtain reliable $^3J_{\text{C-H}}$ values.

Most of the currently used experimental schemes, that allow measurements of ^{13}C relaxation times, are based on 2D double INEPT [63]. A series of 2D spectra are measured with various lengths of relaxation intervals. Cross-peaks intensities in 2D spectra vary according to the length of relaxation intervals and relaxation times are determined by fitting of the intensities of the peaks. The interpretation of the T_1 or T_2 values and calculation of motional parameter requires then appropriate models for molecular motions [57].

X-Ray Crystallographic Analysis

X-ray crystallography is the main experimental method for carbohydrate structural determination in solid state. Although only a limited number of carbohydrate structures has been published to date due to the problem of carbohydrate crystallization, bond lengths, bond angles and torsion angles have been obtained together with hydrogen

bonding networks for several biologically interesting molecules [64, 65]. Perhaps even more importantly, current methodologies offer indispensable information on the structure of protein-carbohydrate complexes. Data on 3D structure of both ligand and the protein molecules, hydrogen

bonds, counter ions and solvent molecules in these complexes are essential in drug design. X-ray analysis in solid-state together with NMR data in solution and molecular modeling provide useful basis for the detailed understanding of protein-carbohydrate interactions. Selected

Table 1. Selected Crystal Structures of Protein-Carbohydrate Complexes

Carbohydrate	Protein	PDB code	Ref
Manα(1-3)[Manα(1-6)]Manβ	Concanavalin A	1CNV	66
		1ONA	67
GlcNAcβ(1-2)Manα(1-6)) [GlcNAcβ(1-2)Manα(1-6)] Manα-OMc	Concanavalin A	1TEI	68
(Manα(1-6)[Manα(1-3)])	Agglutinin <i>Galanthus nivalis</i>	1JPC	69
(Manα(1-6)[Manα(1-3)]Man			
(Manα(1-3) Manα(1-4) GlcNAc	Isolectin I <i>Lathyrus ochrus</i>	1LOG	70
MeαMan	Mannose binding protein MBP-A and MBP-C	1KWU	71
Manα(1-3)Man		1KWY	
Manα(1-3)[Manα(1-6)]Man		1KZB	
Man ₆ GlcNAc ₂ Asn		1KX1	
3'-sialyllactose	Wheat germ agglutinin I	1WGC	72
	Wheat germ agglutinin II	2WGC	
3'-sialyl Lewis X	P-, E-selectin	1GIT	73
		1GIR	
		2KMB	74
3'-sulfo Lewis X	Mannose binding protein MBP-A	3KMB	74
4'-sulfo Lewis X		4KMB	
Fucosyllactose	Isolectin II <i>Ulex europaeus</i>	1QOT	75
Fucosylgalactose		1QOT	
Fucosyllactose	<i>Erythrina cristagalli</i> Lectin	1GZ9	76
Fucose	Lectin I <i>Ulex europaeus</i>	1JXN	77
Blood group antigen analogue	Soybeen agglutinin	1G9F	78
Lewis B	Isolectin IV <i>Griffonia Simplicifolia</i>	1LED	79
Lewis Y		1GSL	
Lewis Y	anti-tumor antibody BR96	1CLZ	80
Heparin-pentasaccharide	Antithrombin III	1AZX	81
Heparin-hexasaccharide	FGF-1	2AXM	82
Heparin	FGF-1-FGFR-2	1E0O	83
Heparin-tetrasaccharide	FGF-2	1BFB	84
Heparin-hexasaccharide			
Heparin	FGF-2-FGFR-1	1FQ9	85
Heparin-tetrasaccharide	Annexin V	1G5N	86

recently published carbohydrate-protein crystal structures are listed in Table 1.

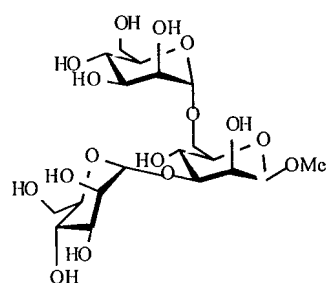


Chart. (4).

3D STRUCTURE AND DYNAMICS OF CARBOHYDRATES AND THE RELATION WITH BIOLOGICAL PROPERTIES

N-linked Oligosaccharides

Asparagine-linked (N-linked) oligosaccharides belong to a class of cell surface carbohydrates that can be divided into three subtypes – oligomannose, complex and hybrid [87]. N-

disaccharides, as well as complex high mannose oligosaccharides [88-91], and were focused on determination of conformational equilibrium on the individual (1-2, 1-3 and 1-6) glycosidic linkages. NOEs and molecular modeling (hard sphere exo-anomeric, HSEA) in model compounds (1-3 and 1-6 linked disaccharides and the core trisaccharide (Chart 4)) indicated that 1-3 linkages are fairly rigid with the (ϕ, ψ) values of ($75^\circ, -135^\circ$). Two approximately equally populated conformers with $\omega = 180^\circ$ and -60° have been observed (values recalculated from those published that were based on the H-5-O-6 definition of the ω torsion angle to values based on the O-5-O-6 definition used presently) for the 1-6 linkages [89]. Narrow torsional excursions within the single energy well ($100^\circ, -90^\circ$) [90] at the 1-3 glycosidic linkage were in qualitative agreement with previous findings. Further evidence of the static concept was found in 500 ps MD simulations performed on 1-2-type linkages in Man₉GlcNAc₂ (Chart 5) in vacuum [91].

On the other hand, both theoretical and experimental studies revealed that this type of oligosaccharides has distinct internal motions [56, 92-96]. Three minima with the comparable energies have been found on energy surfaces computed for the 1-3 linkage in mannosiose in water solutions [92, 96]. The evidence of flexibility was supported

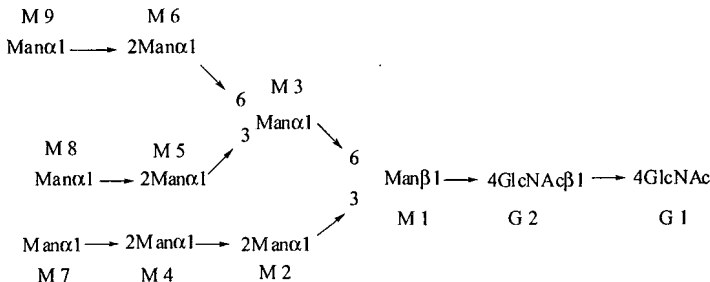


Chart. (5).

linked oligosaccharides constitute the part of many glycoproteins and play significant roles during cell-cell recognition processes. Although there is much known about N-glycans biosynthesis, their 3D structures, dynamics and the roles in glycoprotein folding have been subject of interest during the last two decades.

NMR and molecular modeling studies have been performed on both simple structures, such as mono- or

by the experimental NMR relaxation data. 1D NOEs, collected at various magnetic fields, could be interpreted only considering internal motions at the 1-3 glycosidic linkage in model disaccharide [56]. Molecular modeling studies of N-glycan structures also revealed the occurrence of fluctuations at various glycosidic linkages [97, 98]. Recent detailed studies of the core trimannoside [95, 99] and the complex oligomannoside Man₉GlcNAc₂ [98], (Chart 5) brought further details on internal motions in N-glycans.

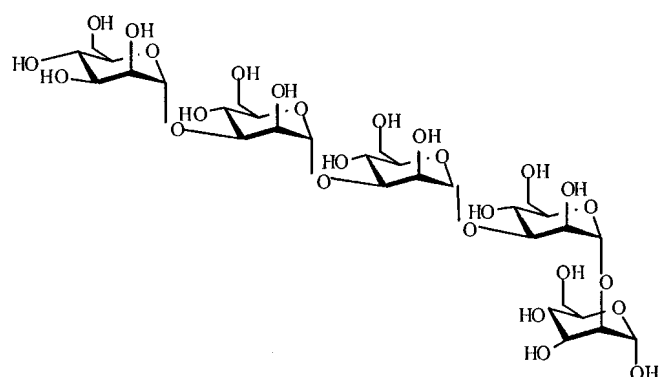


Chart. (6).

Table 2. Thermodynamic Data for Manα-OMe and Selected Mannose-Type Oligosaccharides Complexed with Concanavalin A

Compound	ΔG [kcal. mol ⁻¹]	ΔH [kcal. mol ⁻¹]	TΔS [kcal. mol ⁻¹]	Ref.
Manα-OMe	-5.3	-6.6	-1.3	102
GlcNAcβ(1-2)Man	-5.2	-5.3	-0.1	102
Manα(1-3)Man	-5.7	-10.2	-4.5	103
GlcNAcβ(1-2)Manα(1-3)Manα-OMe	-6.1	-4.6	1.5	104
Cyclic analogue -)6GlcNAcβ(1-2)Manα(1-3)Manα OMe(4-	-5.8	-2.7	3.1	104

Two conformations with different ψ angles existed on 1-2 linkages, namely, $(\phi, \psi) \approx (80^\circ, -60^\circ)$ and $(80^\circ, -180^\circ)$. Interestingly, outer 1-2 linkages adopt a relatively stable stacked conformation with short-lived transitions to the extended conformation. Internal 1-2 linkages showed the larger flexibility though the inter-conversion was relatively slow between the stacked and the extended conformations. The observed flexibility at the 1-2 linkages was not markedly influenced by the short-lived inter-residue hydrogen bonds. In fact, these interactions have not stabilized any of the observed conformations, stacked or extended. Two main conformers at the 1-3 linkage had $(\phi, \psi) \approx (80^\circ, -100^\circ)$ and $(60^\circ, -180^\circ)$ in trimannoside structure; the flexibility at this linkage appeared to be persistent also the complex with mannose-binding protein (MBP) [95, 100]. Similar behavior of the 1-3 linkage, though with slightly restricted flexibility, has been observed in $\text{Man}_9\text{GlcNAc}_2$. In the latter case, small difference in conformational properties of two 1-3 linkages corresponded with differences in the inter-residue hydrogen bonds. Thus, the originally derived structure [88] in this trisaccharide with $(\phi, \psi) (75^\circ, -135^\circ)$ corresponds to the “virtual” conformation resulting from the averaging of two approximately equally populated conformers with $(\phi, \psi) \approx (80^\circ, -100^\circ)$ and $(60^\circ, -180^\circ)$. Detailed analysis of the structurally similar pentasaccharide (Chart 6) also indicated the presence of internal motions as inferred from relaxation data and MD. On the other hand, residual dipolar couplings could be interpreted with a single-well structure [101].

It should be born in mind that the conformational behavior at different glycosidic linkages likely depends on the size of saccharide. As expected, small oligosaccharides tend to be more flexible owing to the reduced steric restrictions of neighboring units. In larger structures, such as the $\text{Man}_9\text{GlcNAc}_2$ both rigid and flexible parts are present in the structure [98]. However, the overall 3D structure of $\text{Man}_9\text{GlcNAc}_2$, seems relatively well preserved despite the local flexibility. This structure does not considerably differ from the conformation adopted in the glyco-conjugate from. It is speculated that the above structural similarities could be due to the compensatory motions at the glycosidic linkages. In addition, correlated motions are likely present as well, influencing the transitions between the glycosidic linkage conformations, such as between the M7-M4 and the M4-M2 residues. The similar type of motions appeared to be present also in the core trisaccharide, $\text{Man}\alpha(1-3)[\text{Man}\alpha(1-6)]\text{Man}\beta\text{-OMe}$ [95, 99]. Perhaps both types of motions, compensatory and correlated, could be the consequence of the complex

hydrogen bond network between N-glycan and solvent. This 3D structure and motional behavior of this N-glycan, together with the hydrogen bond system (partially mediated by water molecules), may be the key to its properties during the binding with proteins. The flexibility at the $\alpha(1-3)$ linkage in the trimannoside structure in the binding site of MBP can compensate, at least in part, the entropic cost resulting from the freezing of motions of the Lys182 side chain in MBP and motions of the trisaccharide hydroxyl groups involved in hydrogen bonds [100]. Consequently, this compensation results in the comparable binding constant values of the core trisaccharide $\text{Man}\alpha(1-3)[\text{Man}\alpha(1-6)]\text{Man}\beta\text{-OMe}$ to MBP with that of monosaccharide, $\alpha\text{-Me}$ mannoside.

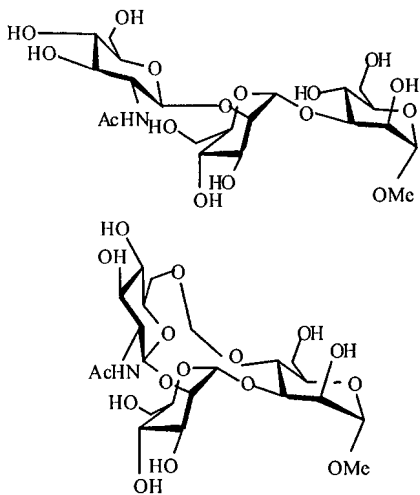


Chart. (7).

Similar evidence has been obtained for various mannose-type oligosaccharides (Table 2). Binding of $\text{GlcNAc}\beta(1-2)\text{Man}\alpha(1-3)\text{Man}\alpha\text{-OMe}$ and its cyclic analog, $-)6\text{GlcNAc}\beta(1-2)\text{Man}\alpha(1-3)\text{Man}\alpha\text{-OMe}(4-$ (Chart 7), with lectin concanavalin A (Con A) showed analogous results [104]. The trisaccharide was found very flexible in solution (with 7 energy minima) and the loss of its flexibility during binding with Con A was accompanied by the less favorable entropy changes ($T\Delta S = 1.5$ kcal/mol) with respect to the cyclic counterpart ($T\Delta S = 3.1$ kcal/mol). 3D structure of this trisaccharide in the complex had the glycosidic linkage conformations $(\phi_{1-2}, \psi_{1-2} \approx 43^\circ, 150^\circ; \phi_{1-3}, \psi_{1-3} \approx 74^\circ, 95^\circ)$. It is noteworthy, that the bound form is comparable with the local minimum structure in the free state $(\phi_{1-2}, \psi_{1-2} \approx 44^\circ, 147^\circ; \phi_{1-3}, \psi_{1-3} \approx 69^\circ, 104^\circ)$ rather than the main

energy minimum arrangement ($\phi_{1-2}, \psi_{1-2} \approx -51^\circ, -172^\circ$; $\phi_{1-3}, \psi_{1-3} \approx 69^\circ, 99^\circ$).

As mentioned in the first paragraph of this section, the influence of saccharide chains on protein part of glycoproteins, and vice versa, is another issue that has to be considered. Several studies pointed out on the significant effect of the protein upon the oligosaccharide conformation of the glycoprotein. NMR data collected in oligosaccharide linked to the Asn78 residue in human chorionic gonadotropin (N-glycosylated at Asn52 and Asn78) indicated that the flexibility of 1-3 and 1-6 linkages in the core

trimannoside fragment ($\text{Man}\alpha(1-3)[\text{Man}\alpha(1-6)]\text{Man}\beta$) differed from that in the aqueous solution [105]. Differences in flexibility between terminal and interior residues were observed also in 21-aminoacid glycopeptide from human serum transferrin [106]. Even relatively small peptides can reduce flexibility of the glycan core what is in agreement with other data [91, 107]. Perhaps more appealing is the observation that N-glycans, such as $\text{Man}_9\text{GlcNAc}_2$ (Scheme 5) can promote protein folding [108] and some complex-type N-glycans could be essential in stabilization of the protein 3D structure [109, 110]. It is argued that the effect of N-glycan is due to its amphiphilic structure, which likely

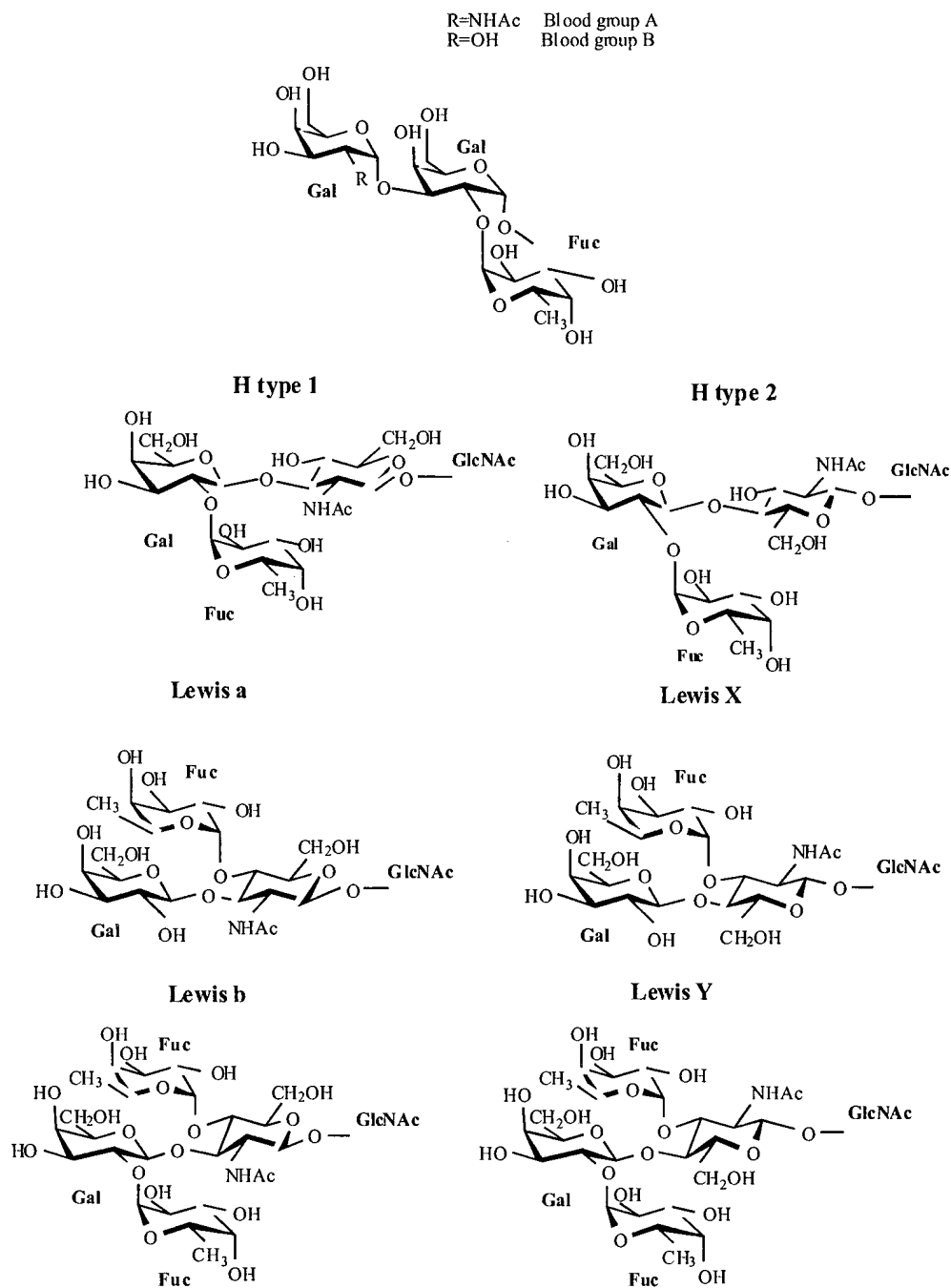


Chart. (8).

contributes to the formation hydrophobic interactions, such as observed in polypeptides in endoplasmatic reticulum [111]. These carbohydrate-protein hydrophobic interactions might play a decisive role in enhancing the protein folding through reducing hydrophobic interactions within the protein part also in other glycoproteins. In accordance, higher N-glycans promoted protein folding better than the truncated one in influenza virus hemagglutinin [112]. Disulfide-dependent dimerization in intestinal mucin also demonstrates the essential role on N-linked oligosaccharides in protein structure [113].

It appears that the main efforts are currently focused on establishing the influences of N-glycans roles in glycoprotein folding, stabilization and dynamics. All these phenomena are crucial in detailed understanding of cell-cell interactions where glycoprotein structures play one of the main roles.

Blood Groups Oligosaccharides

Oligosaccharides related to ABH(O) blood group antigens are synthesized from the type-1, -2, -3 or type-4 glycan precursors by glycosyltransferases. Characteristic for their structure is the $\text{Fuc}\alpha(1-2)\text{Gal}\beta$ - moiety that is substituted in the position 3 of the β -Gal residue by GalNAc and Gal residues, respectively (for A and B determinants, respectively) or linked (1-3 and 1-4) with the GlcNAc residue (H type 1 and H type 2, respectively) (Chart 8). Blood group determinants are covalently bound to proteins and expressed on the red cells surfaces, vascular endothelium as well as other tissues. Due to their biological significance, blood antigen structural properties have been the subject of numerous studies. Initial investigations in early eighties were based on HSEA-type calculations and NOE NMR measurements [14, 114]. The analysis of the data led to the conclusions that these oligosaccharides have well-defined structures. Prevalently rigid structures were inferred also from interglycosidic NOEs between H-1_{Fuc} - $\text{H-5}_{\text{GlcNAc}}$ and H-1_{Fuc} - $\text{H-3}_{\text{GlcNAc}}$ in blood group A oligosaccharides [115] and in H-trisaccharides [116, 117]. A different view on the conformational properties in the blood group A trisaccharide, $\text{GalNAc}\alpha(1-3)[\text{Fuc}\alpha(1-2)]\text{Gal}\beta\text{-O-R}$, was based on the presence of two conformers at the Fuc-Gal linkage. Experimental NOEs, together with their theoretical interpretation, indicated that both conformers (with the ϕ_{1-2} , $\psi_{1-2} \approx 50^\circ$, 25° and ϕ_{1-2} , $\psi_{1-2} \approx 25^\circ$, -48°) have comparable abundances in aqueous solution [118-120]. Satisfactory agreement with experimental data was obtained only when the weighted average from both conformers was taken for calculation of NOEs. Further NMR data of this trisaccharide were collected in the presence of *D. biflorus* lectin [118]. Computed NMR transferred NOEs indicated that only conformer with the ϕ_{1-2} , $\psi_{1-2} \approx 50^\circ$, 25° at the Fuc-Gal linkage was selected by the receptor. Higher flexibility at the Fuc-Gal linkage, rather than the Gal-Gal linkage, was also in agreement with MD simulations [121]. However, Monte Carlo (MC) and solvated MD simulations of T-antigen disaccharide $\text{Gal}\beta(1-3)\text{GalNAc}\alpha\text{OMe}$ also reveal a degree of flexibility at the Gal-Gal linkage [122]. Two energy minima with ϕ , $\psi \approx 53^\circ$, 8° and ϕ , $\psi \approx 171^\circ$, -2° have been obtained by MC for aqueous solution. Computed NOEs, based on the presence of both conformers, agreed

with experimental data. In the latter case, the lack of the fucose residue can, however, results in different conformational properties of this T-antigen disaccharide in comparison with the previously mentioned trisaccharide. It is also noteworthy, that transferred NOEs in $\text{Gal}\beta(1-3)\text{GalNAc}\alpha\text{OMe}$ in the presence of *Maclura pomifera* agglutinin are in agreement with two different 3D structures with ϕ , $\psi \approx 45^\circ$, -65° and ϕ , $\psi \approx -65^\circ$, -18° pointing out on some degree of persistent flexibility in the presence of the protein [123]. Furthermore, both conformers differed (predominantly in the ϕ angle) from the disaccharide structure in crystal state (ϕ , $\psi \approx 39^\circ$, -8°) [124]. This evidence is remarkable and might in part originate from the presence of water molecules that create a bridge between the saccharide molecule and amino acids in the binding site. It appears that the hydrogen bond system disaccharide-water-receptor in solid state can also stabilize other oligosaccharide conformers than those observed in solution in the complex with protein. Consequently, the solution structure of the saccharide bound form may not always correspond to that observed in solid state.

Only a few crystal studies have been published to date dealing with blood groups antigen analogs, such as Me-fucoside, fucosyllactose, fucosylgalactose or penataccharide analog (Table 1). In the complex of structurally related fucosyllactose ($\text{Fuc}\alpha(1-2)\text{Gal}\beta(1-4)\text{Glc}$) with *Ulex europaeus* isolectin II the observed bioactive conformation at the $\alpha\text{Fuc1-2}\beta\text{Gal}$ linkage had the values ϕ_{1-2} , $\psi_{1-2} \approx -3^\circ$, -12° and the $\text{Gal}\beta(1-4)\text{Glc}$ linkage ϕ_{1-4} , $\psi_{1-4} \approx 31^\circ$, -37° [75]. The comparison of stereochemistry of the $\alpha\text{Fuc1-2}\beta\text{Gal}$ linkage in similar systems indicate that this arrangement slightly differs from the second conformer for the blood group A trisaccharide, $\text{GalNAc}\alpha(1-3)[\text{Fuc}\alpha(1-2)]\text{Gal}$ as well as from the geometry of the energy minima found in fucosyllactose in molecular modeling studies [119]. These values appeared different also from that found in a high-resolution crystal structure of *Erythrina cristagalli* lectin in complex with fucosyllactose [76]. In the complex with lectin, ϕ_{1-2} , $\psi_{1-2} \approx 47^\circ$, 47° and ϕ_{1-4} , $\psi_{1-4} \approx 49^\circ$, -20° , and clearly shows that the 3D structure of the saccharide bound form can be quite different in various proteins. The latter discrepancy between the structure of fucosyllactose in *Ulex europaeus* isolectin II and *Erythrina cristagalli* lectin originates from dissimilar structural bases of both proteins binding sites. Although some similarities can be found in the interactions of the fucose residue with amino acids in the binding site, the stereochemistry of the complex appears quite different. The hydroxyl groups 2-OH and 4-OH are involved in hydrogen bonds with Asn133 and Tyr107 in *Erythrina cristagalli* lectin (in *Ulex europaeus* isolectin II with Gly106 and Ser104). In addition, Tyr106 in the lectin is located on the other side of the fucose residue with respect to the Tyr residue (Tyr135) in isolectin II. Experimental results further indicate (increased B-factors) that there could also be an increased flexibility of the monosaccharide binding site, which partly explains the observed, more favorable entropic term in the lectin thermodynamic data. The data also indicate that water molecules play an important role in the lectin binding site.

Well-defined conformations were derived from HSEA calculations and steady-state NOEs in the early studies of structure of Lewis blood group determinants [14,114]. The

overall shapes of Lewis a (Le^a) and Le^x were found similar, the same for Le^b and Le^y . Molecular modeling study using the MM3 force field later found that 3D structures of Le^x and Le^y are more rigid than Le^a and Le^b [119, 125]. In the Le^x trisaccharide, $\text{Gal}\beta(1-4)[\text{Fuc}\alpha(1-3)]\text{GlcNAc}$, the geometry of the energy minimum in solution ($\phi_{1-3}, \psi_{1-3} \approx 39^\circ, -89^\circ$ for the Fuc-GlcNAc linkage; $\phi_{1-4}, \psi_{1-4} \approx 45^\circ, 16^\circ$ for the Gal-GlcNAc linkage) was found comparable with that observed in the solid state [126]. On the other hand, several energy minima have been obtained in theoretical study, and the solvent influenced the conformational equilibria in solutions, [127] suggesting that these oligosaccharides might be more flexible than originally assumed. Experimental NMR relaxation data and MD simulations of Le^x [128] also support these conclusions. The values of generalized order parameters, $S^2 \approx 0.8$, derived from experimental ^{13}C NMR relaxation times, indicate that fluctuations exist at least at the GlcNAc-Gal linkage and the Fuc-Glc linkage. On the other hand, analysis of one-bond residual dipolar coupling constants in Le^a and Le^x were found to be in agreement with well-defined, tightly folded structures of both oligosaccharides [129]. Experimental NMR dipolar couplings and MC simulations using the MM3 force field with continuum solvent model revealed a similar overall shape of both determinants and thus arrived to conclusions as made in earlier studies.

Only a very few studies have examined the 3D structure of Lewis determinants in complexes with proteins. A 2.0 Å resolution analysis of co-crystals of Le^b and Le^y with lectin IV of *Griffonia simplicifolia* indicated that the conformation of both tetrasaccharides is nearly the same [79]. The torsion angles at the glycosidic linkages in the bound state were comparable each other and, in fact, corresponded well also with the results of HSEA calculations. Similar 3D structure at the glycosidic linkages in the complex with anti-tumor antibody BR96 adopted also Le^y derivative, where the largest difference was only 14° in the ϕ torsion angles at the Fuc-GlcNAc linkage [80].

Recent high-resolution NMR study of conformation of Le^x in the complex with two proteins, isolectin A from *Lotus tetragonolobus* and *Aleuria aurantia* agglutinin, revealed that the bioactive conformation at the Fuc-GlcNAc linkage was considerably different compared to that in the free state [130]. It was argued that, as both proteins are fucose-recognizing proteins, the bioactive conformation enabled better contact with the protein surface. In the free solution, the GlcNAc and the Fuc residues showed the stacking interactions that prevent the proper interactions of the Fuc hydrophobic side with the protein receptor. In addition, the diversity in bound conformations at the Fuc-

GlcNAc linkage observed between these two proteins demonstrated that different proteins may select different 3D structures of the same ligand.

Structurally related tetrasaccharide sialyl Le^x (sLe^x , $\text{Neu5Ac}\alpha(2-3)\text{Gal}\beta(1-4)[\text{Fuc}\alpha(1-3)]\text{GlcNAc}$, Chart 9) tetrasaccharide is involved in the adhesion of leukocytes to vascular endothelial cells through the interactions with trans-membrane glyco-proteins. The recognition process of sLe^x by selectins is thus important in understanding such events as the cell adhesion associated with inflammatory processes. The 3D structure of sLe^x cannot be compared directly with that of the related blood group determinant, Le^x . The differences in conformation, brought about by the presence of the αNeuAc residue, are relatively pronounced. The stereochemistry at the $\text{Neu5Ac}\alpha(2-3)\text{Gal}$ linkage is influenced by the presence of the carboxylate group linked to C-2 as well as by the lack of oxygen at C-3. The data based on GESA and MM2 calculations, together with NOEs and coupling constants across the glycosidic linkages, indicated rigidity of the whole molecule, including the $\text{Neu5Ac}\alpha(2-3)\text{Gal}$ linkage [131]. Other NMR and MD investigations confirmed the presence of conformational equilibria at the $\text{Neu5Ac}\alpha(2-3)\text{Gal}$ linkage, [132, 133] though the presence of multiple conformers at other linkages was not clear [133, 134]. However, differences were found in conformational properties at the 2-3 and the 2-6 linkages in $\text{Neu5Ac}\alpha(2-3)\text{Gal}\beta(1-4)\text{GlcNAc-Asn}$ and $\text{Neu5Ac}\alpha(2-6)\text{Gal}\beta(1-4)\text{GlcNAc-Asn}$. Three different conformers were found at the 2-3 linkage, whereas the extended low energy surface for the 2-6 linkage indicated higher flexibility of the latter linkage [135]. A detailed study of conformational properties of $O\text{-sLe}^x$ linked to the octapeptide fragment of the mucin domain MAdCAM-1 further shows [136] that conformation of the carbohydrate moiety remains essentially the same as in the free solution. The influence of the glycosylation upon the peptide conformation has also been studied in this work. The observed differences in five different peptides, caused by the different carbohydrate moieties, further supported other evidence on the crucial role of glycosylation in glycoproteins 3D structure and dynamics, as already discussed within N-glycans [105-113]. GlcNAc and LacNAc residues showed no significant effect on the octapeptide conformation. However, further fucosylation caused changes in the 3D glycopeptide structure, specifically in the N-terminal region. Two proline residues near this region can be in the *cis-trans* forms what results in structural heterogeneity of the peptide. The glycosylation reduced the *cis-pro* conformation whereas the *trans-pro* form further stabilized at the same time.

Structures of the sLe^x bound to E-, P- and L-selectins have been thoroughly studied by NMR, including hydroxyl

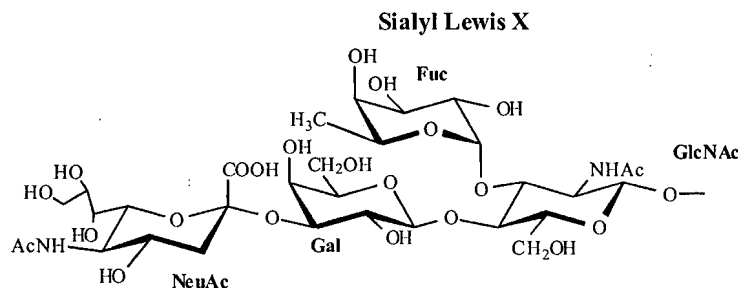


Chart. (9).

Table 3. Conformation of sLe^x in the Complex with E-Selectin as Inferred from NMR Studies and Retrieved from the Crystal Data. The ϕ, ψ Torsions Angles are Defined as $\phi = \text{H-1'-C-1'-O-1'-C-x}$ (C-1'-C-2'-O-2'-C-x in the Case of Neu5Ac); $\psi = \text{C-1'-O-1'-C-x-H-x}$ (C-2'-O-2'-C-x -H-x in the Case of Neu5Ac)

	$\alpha\text{Neu5Ac2-3}\beta\text{Gal}$	$\beta\text{Gal1-4GlcNAc}$	$\alpha\text{Fuc1-3GlcNAc}$	Ref.
	$\phi \psi$	$\phi \psi$	$\phi \psi$	
Solid-state	-65 -11	33 16	44 22	73
Solution NMR	-58 -22	24 34	71 14	133
Solution NMR	-43 -12	45 19	29 41	134
Solution NMR	-76 6	39 12	38 26	137

proton resonances in water solution. Transferred NOEs allowed identification of the bound conformations and the recognition of some of the key interactions between E-selectin and sLe^x tetrasaccharide [133, 134, 137]. The derived bound conformations from transferred NOEs are listed in Table 3 together with the X-ray structure (Fig. (1)) [73].

In general, all structures can be considered as equivalent each other, though small differences exist. The largest variations seem at the Neu5Ac α (2-3)Gal linkage, which is also the most flexible linkage in solution. The largest difference between the crystal (-65°, -11°) and NMR data is about 20°. On the other hand, the small differences (within 6°) between the solution and solid state structures at the other two linkages were obtained in the first study [137].

Furthermore, the transferred NOE data indicate that only relatively minor differences exist between the solution structure and the bound structure. This is not only the case for the Neu5Ac α (2-3)Gal glycosidic linkage, but also for the Fuc α (1-3)GlcNAc linkage (unlike to Le^x). The preserved conformation at the latter linkage is likely due to the interaction of the fucose 3-OH and 4-OH groups with calcium ions in the binding site. This interaction does not require changes in stereochemistry at the linkage as the both OH groups are accessible also in the most abundant conformation.

Analysis of 3D sLe^x structures in complex with other proteins, *Lotus tetragonolobus*, *Aleuria aurantia* agglutinin, P- and L-selectins, indicates that the preservation of the solution structure in the complex is not always the

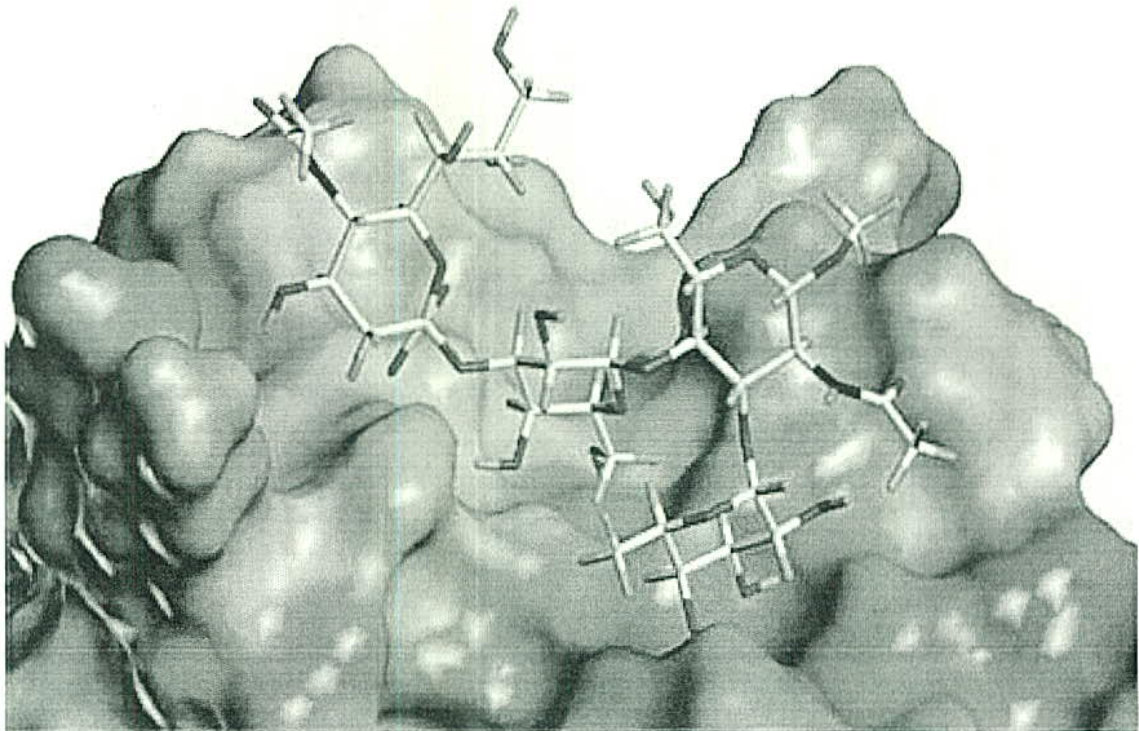


Fig. (1). Structure of the tetrasaccharide sLe^x- E-selectin complex. Torsions angles at the glycosidic linkages are listed in Table 3. Retrieved from the 1G1T pdb file [73].

case. In fact, again similarly to Le^x, the comparison of free and bound conformations reveal variations at the Fuc α (1-3)GlcNAc linkage [130] caused by the proper arrangement of the fucose residue in the binding site. The changes at the Neu5Ac α (2-3)Gal linkage [133] during binding with P- and L-selectins, as well as other similar oligosaccharides in complex with proteins [123], confirmed previous observations that receptor does not necessarily select the global minimum structure of the ligand.

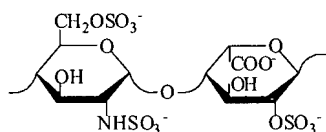


Chart. (10).

Sulfated Saccharides

Although sulfated derivatives belong to a relatively minor group of carbohydrates, they are involved in a large variety of biological functions, such as, blood coagulation, cell growth and differentiation, lipid metabolism, viral invasion, angiogenesis, etc. [138-142]. Within this group of

this polysaccharide as well as in heparin-like oligosaccharides [146-148]. In heparin, the IdoA2S residue adopts two conformations, the skew-boat ²S₀ conformation and the chair ¹C₄ form (Chart 11). The abundances of these two forms are in ratio $\approx 40 : 60$ (²S₀ : ¹C₄) [148]. This equilibrium can be different in other structurally-related compounds. For example, in the heparin-pentasaccharide (Chart 12), which corresponds to the binding site of heparin for antithrombin, the ratio is approximately reversed, thus $\approx 60 : 40$ (²S₀ : ¹C₄). One of the main factors influencing this conformational equilibrium is the 2-O-sulfate group in the IdoA2S residue. The lack of the 2-O-sulfate group results in the preference of the chair form, like is in the heparin-derived tetrasaccharide (Chart 13) in which the ratio ²S₀ : ¹C₄ $\approx 25 : 75$ [149]. However, the conformational equilibria in heparin-like structures are relatively complex and the presence of the third conformer, ⁴C₁, had to be also considered to interpret numerous experimental NMR data [150]. Furthermore, complex inter-conversions among the ¹C₄ chair form, the skew ⁰S₂ and ³S₁ forms as well as the ⁴C₁ form with ²S₀ and ¹S₃ conformations have also been considered [151]. It appears that the pseudorotation among the ²S₀, ¹S₃, ⁰S₂ and ³S₁ forms is energetically less demanding than between the chair and the skewed forms. In

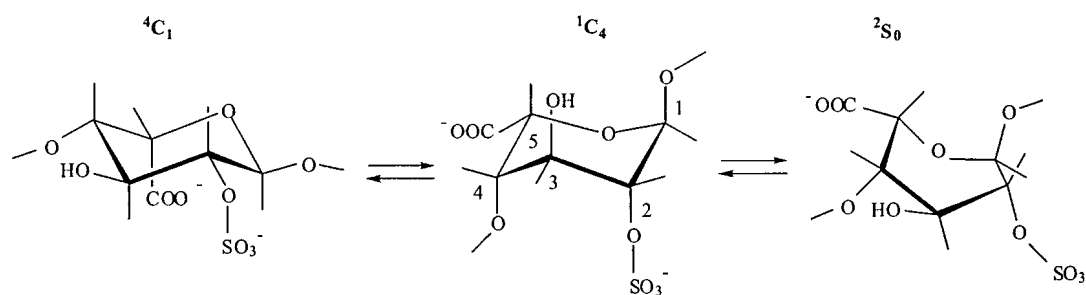


Chart. (11).

saccharides, efforts in structural analyses [143-145] have been focused primarily on glycosaminoglycans (GAGs). GAGs are linear polysaccharides made up by alternating uronic acid (glucuronic acid (GlcA) or L-iduronic acid (IdoA)) and hexosamine (2-amino-2-deoxy-D-glucose (GlcN), 2-amino-2-deoxy-D-galactose (GalN)), some parts being O- or N-sulfated (O-SO₃⁻ (O-S) or N-SO₃⁻ (N-S) groups). The most well-known GAG, heparin (Chart 10, predominantly consisting of -IdoA2S α (1-4)GlcN,6S α (1-4)-) has been the first and the most intensively studied sulfated polysaccharide molecule owing to its variety of biological functions [138-142].

Analyses of high-resolution NMR spin-spin coupling constants and NOEs have shown that GlcN,6S residues adopts rigid ⁴C₁ conformation in all GAGs, whereas IdoA2S residues showed rather complex conformational behavior in

addition, the influence of counter-ions (mainly Na⁺ or Ca²⁺) should also be considered as ions can influence abundances of these conformers [152, 153]. All the above mentioned efforts are prompted by the necessity to understand how the IdoA residue influences the 3D structure and dynamics of GAGs, as well as its role in interactions with proteins.

As GAGs are considerably substituted saccharides, their 3D structure is further influenced by the presence of bulky and negatively charged sulfate and carboxylate groups present also in other residues [148, 150]. Consequently, GAGs adopt extended, partly flexible structures in most cases [154, 155] although the formation of well-defined structures has also been described [156]. The analysis of conformations at the glycosidic linkages has been based predominantly on NOEs (ROEs) [148-150, 156-159], ³J_{C-H} [160] and ¹J_{C-H} [145, 149, 161]. Theoretical NOE data were

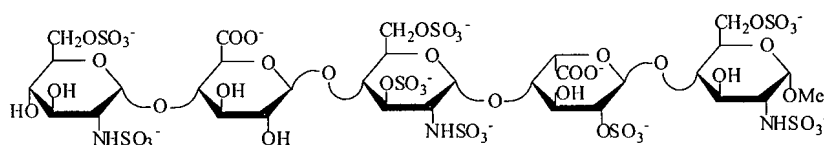


Chart. (12).

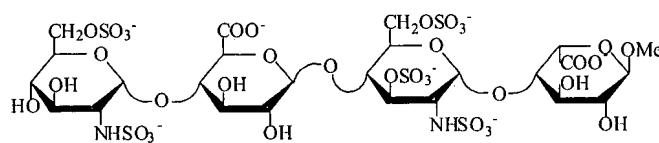


Chart. (13).

compatible with most of the steady-state NOEs in heparin-pentasaccharide, GlcN,6S α (1-4)GlcA β (1-4)GlcN,3,6S α (1-4)IdoA2S α (1-4)GlcN,6S α OMe (Chart 12), and indicated that conformations at the GlcN,6S α (1-4)GlcA, GlcA β (1-4)GlcN,3,6S and GlcN,3,6S α (1-4)IdoA2S glycosidic linkages are comparable for both, nearly equi-energetic, conformers (1C_4 and 2S_0) in the IdoA2S residue [157]. On the other hand, the geometry at the IdoA2S α (1-4)GlcN,6S linkage differed considerably ($\phi_4 = 24^\circ$, $\psi_4 = -48^\circ$ for the 1C_4 conformation and $\phi_4 = 43^\circ$, $\psi_4 = 6^\circ$ for the 2S_0 conformation) for the two iduronate residue forms. In a structurally similar heparin-derived tetrasaccharide GlcN,6S α (1-4)GlcA β (1-4)GlcN,3,6S α (1-4)IdoA α OMe (Chart 13), theoretical ROEs were in a good agreement with experiment ROEs [149]. Small discrepancies between the theoretical and computed data observed for the GlcN,3,6S α (1-4)IdoA linkage originated from the complex internal dynamics at the reducing part of the molecule. It is speculated that the reason of this disagreement is in both pseudorotation of the IdoA residue and flexibility at the GlcN,3,6S α (1-4)IdoA glycosidic linkage.

Interglycosidic NOEs have been analyzed in various sulfated polysaccharides [156, 162, 163]. The decasaccharide model of the heparin molecule was obtained with MM2 force field where both ring forms (1C_4 and 2S_0) were considered for the IdoA residue [156]. The 3D structure of the helix seems to be not significantly affected by

pseudorotation of the IdoA residue, as the conformation of the glycosidic linkages are similar in both the 1C_4 and 2S_0 forms [143, 148, 156, 164]. The overall molecular shape can be characterized as a helix with an overall length of about 50 Å for decasaccharide. In this model the three sulfate groups form a cluster on one side of the chain, while the other three form a similar cluster on the opposite side (Fig. (2)). Such an arrangement of sulfates seems to play the essential role in interaction with proteins.

Analysis of dynamics allowed the determination of motional parameters, internal and overall correlation times, in heparin-oligosaccharides [165, 166] and several polysaccharides [156, 167, 168] using the symmetric top model. In heparin-pentasaccharide [165], the interpretation of ^{13}C T_1 and T_2 relaxation times, collected at different magnetic field strengths, led to the conclusion that the molecule tumbles non-isotropically (with $\tau_{\parallel}$ and $\tau_{\perp} \approx 450$ ps and 1.6 ns, respectively), and that the rate of internal motion is on the picosecond timescale ($\approx 70 - 90$ ps). The order parameters, that characterize the restriction of internal motions, are different for each residue and decrease from the central residue towards both ends of the molecule (from ≈ 0.9 to ≈ 0.7), indicating that the terminal residues are less restricted in their internal motions.

The relaxation data were interpreted in terms of the symmetric top model in polysaccharides such as heparin,

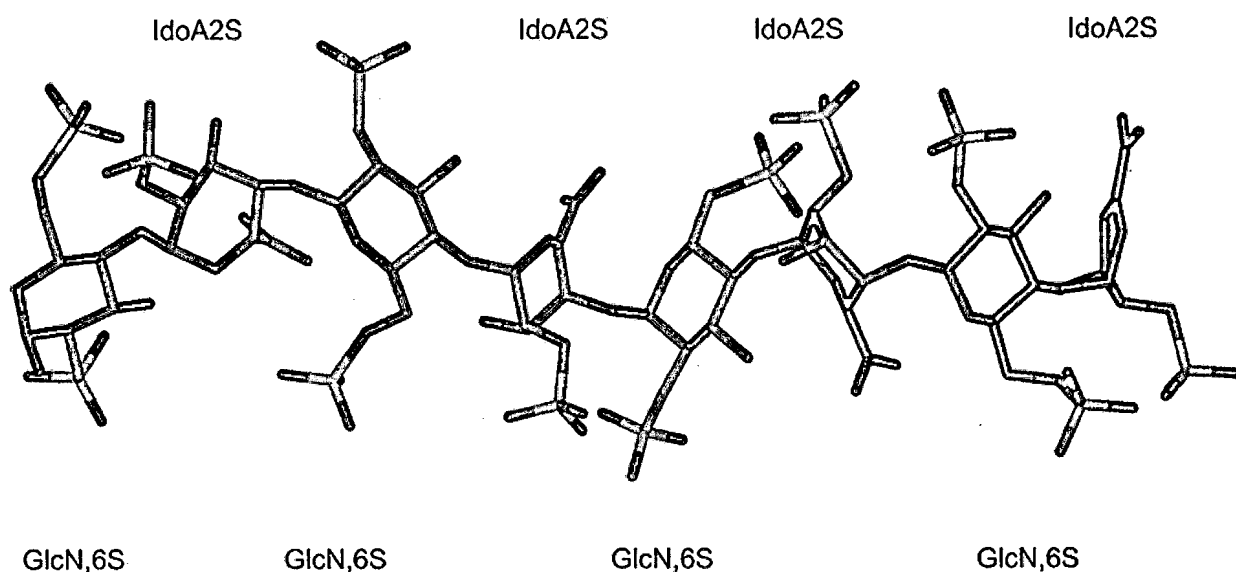


Fig. (2). The overall heparin structure. The N-sulfate and 6-sulfate groups in the glucosamine residues and the 2-sulfate groups in the iduronic acid residues create a cluster at two sites of the polymer chain [156].

Table 4. Values of Overall Correlation Times ($\tau_{||}$, $\tau_{\perp}$, ns/rad), Internal Motion Correlation Times (τ_e , ps/rad) and Order Parameters (S^2_{C-H}) Derived from NMR Relaxation Data for Heparin, Heparin Epoxide and K5 Polysaccharide

	$\tau_{ }$	$\tau_{\perp}$	τ_e	S^2_{C-H}	Ref.
Heparin	0.16	8	a	a	156
Heparin epoxide	0.8	42	0.15	0.65	167
K5	0.35	7.8	0.03	0.79	168

a, not determined

heparin epoxide or K5 [156, 167, 168] and the computed dynamic parameters are given in Table 4. For heparin, the derived correlation times, $\tau_{||} = 0.16$ ns and $\tau_{\perp} = 8$ ns (the axial ratio ($\tau_{\perp} / \tau_{||}$) ≈ 50), indicated a strong motional anisotropy of this polymer. Longer overall correlations times, $\tau_{||}$ and $\tau_{\perp}$, were obtained for heparin epoxide (800 ps and 42 ns, respectively) [167]. The τ_e value (150 ps) suggested that internal motions were slower than in heparin pentasaccharide [165]. The axial ratio was found similar in heparin and heparin epoxide ($\tau_{\perp} / \tau_{||} \approx 50$) and points out on similarity in the molecular shape of these molecules. On the

other hand, the derived values of $\tau_{||}$ and $\tau_{\perp}$ (Table 3) indicated slightly different motional properties of K5 [168]. The axial ratio is smaller (≈ 22) and appears to be in agreement with the smaller size of K5. In addition, τ_e (30 ps) and S^2 (~ 0.8) suggest a different extent of internal motion, compared with heparin epoxide. This difference could originate from the structural differences of these molecules: the presence of the glucuronic acid residue in the K5 leads to a more rigid structure for this polymer.

The interactions between antithrombin and heparin-derived oligosaccharides have been intensively studied [169-

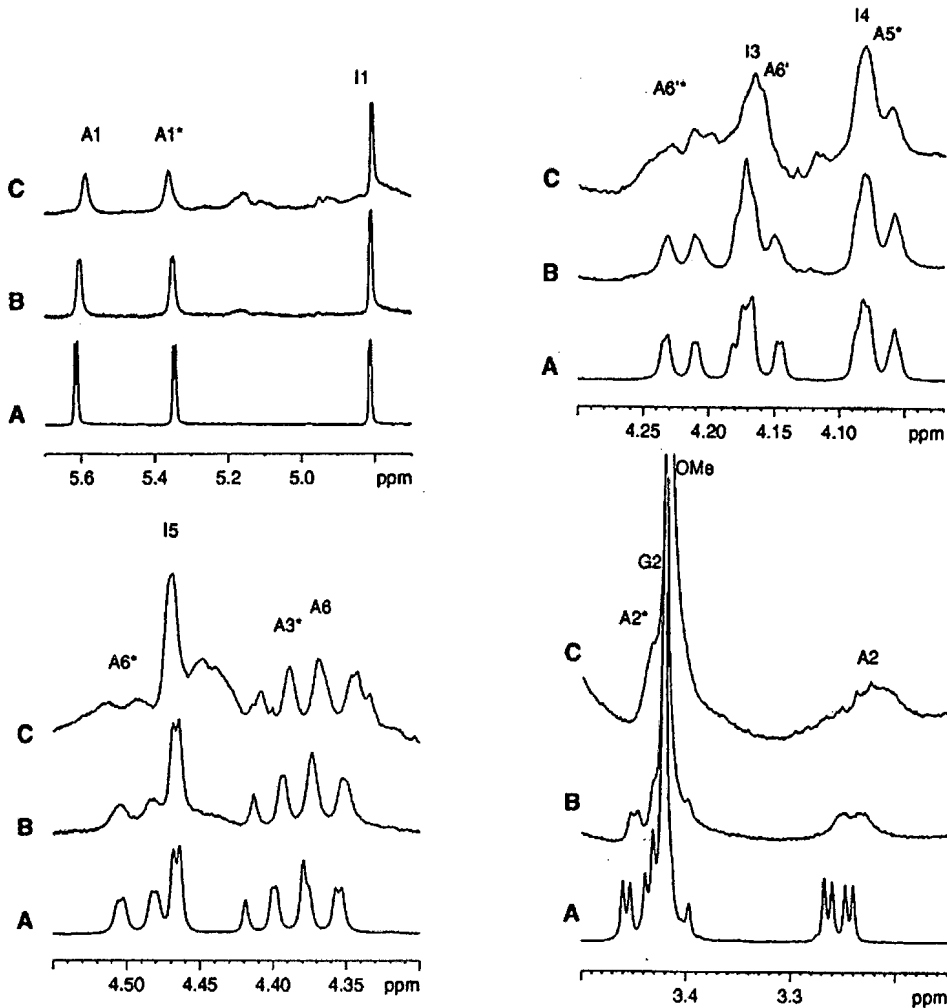


Fig. (3). Regions from ^1H NMR spectra of the heparin-derived tetrasaccharide in aqueous solution (A), and in the complex with antithrombin in molar ratios 20:1 (B) and 10:1 (C).

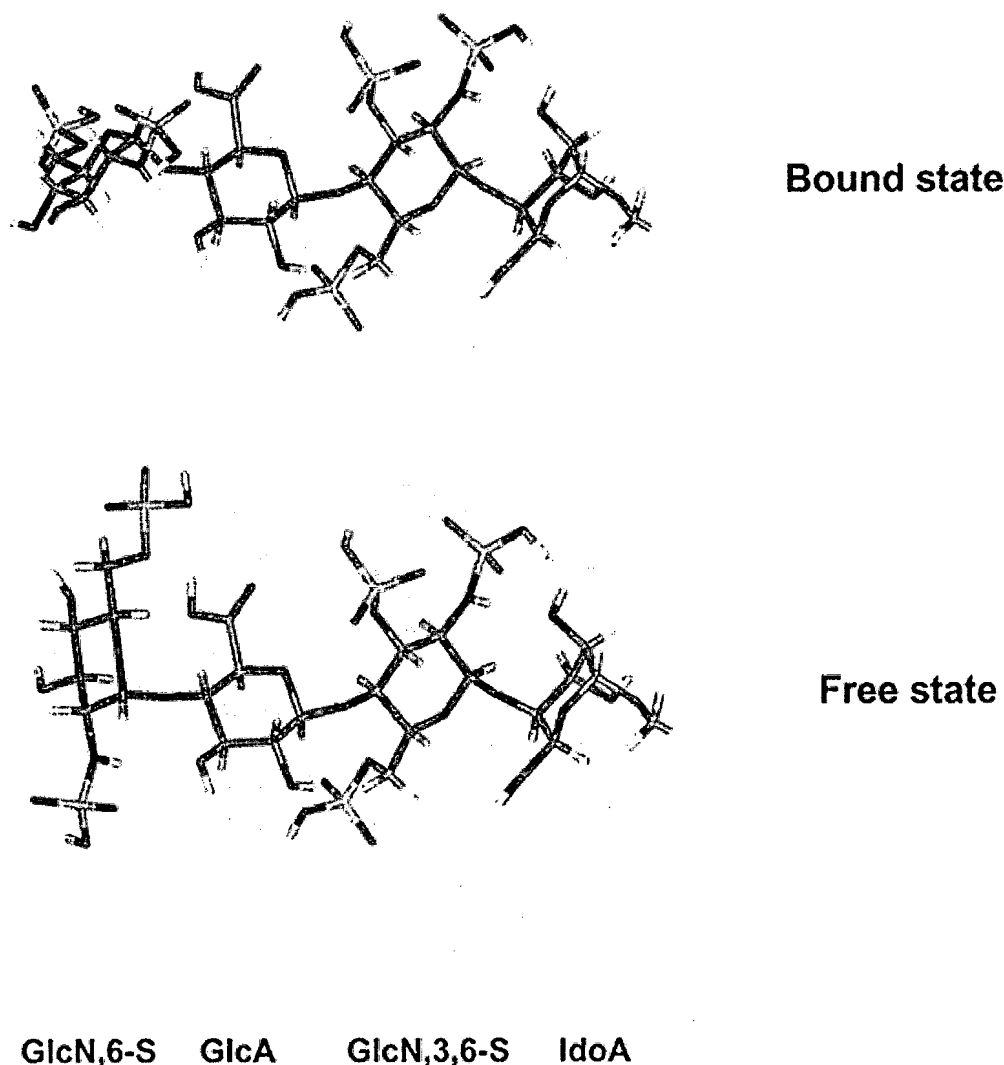


Fig. (4). Heparin-derived tetrasaccharide in the free state and in the complex with plasma protein antithrombin. The corresponding torsion angles at the glycosidic linkages are: ϕ_{1-2}, ψ_{1-2} $-45^\circ, -30^\circ$, (GlcN,6S α (1-4)GlcA), ϕ_{2-3}, ψ_{2-3} $42^\circ, 18^\circ$ (GlcA β (1-4)GlcN,3,6S), ϕ_{3-4}, ψ_{3-4} $-27^\circ, -48^\circ$ (GlcN,3,6S α (1-4)IdoA) for the free state and ϕ_{1-2}, ψ_{1-2} $39^\circ, 12^\circ$, (GlcN,6S α (1-4)GlcA), ϕ_{2-3}, ψ_{2-3} $43^\circ, 14^\circ$ (GlcA β (1-4)GlcN,3,6S), ϕ_{3-4}, ψ_{3-4} $-27^\circ, -48^\circ$ (GlcN,3,6S α (1-4)IdoA) for the bound state.

173] including x-ray crystallography [81, 174], high-resolution NMR [145, 149, 175, 176] and molecular modeling [177-179]. Significant effects of the protein antithrombin on the ^1H and ^{13}C chemical shifts for the heparin-derived tetrasaccharide (Chart 13) have been recently observed (Fig. (3)) in an NMR study of the heparin-tetrasaccharide-antithrombin complex [149]. These large variations are mainly due to strong electrostatic interactions between negatively charged groups in the ligand molecule (sulfates and carboxylates) and positively charged amino acids in the protein binding site. As expected, the most pronounced effects were seen in the central residue (GlcN,3,6S) bearing three sulfate groups where these effects were up to 0.03 ppm for ^1H . Quantitative interpretation of transferred NOEs allowed determination of the tetrasaccharide structure in the bound state. The comparison of both ϕ and ψ torsion angles showed the largest change during the binding with antithrombin was at the GlcN,6S α (1-4)GlcA β linkage (Fig. (4)) [149].

Comparable observations were found in heparin-pentasaccharide-antithrombin complex [175]. Transferred NOEs were interpreted with full relaxation and conformational exchange matrix analysis using the optimized geometry of heparin-pentasaccharide and 8 amino acids in the binding site (Arg 13, Arg 46, Arg 47, Arg 129, Lys 11, Lys 114, Lys 125 and Asn 45). 3D pentasaccharide structures for the free and the bound state (Table 5) revealed that binding is (similarly to tetrasaccharide) accompanied by changes predominantly at the GlcN,6S α (1-4)GlcA β linkage. The small variation at the IdoA2S α (1-4)GlcN,6S linkage has also been observed, whereas the conformation at the remaining linkages, GlcA β (1-4)GlcN,3,6S and the GlcN,3,6S α (1-4)IdoA2S, remains essentially unchanged. The 3D structure at the glycosidic linkages of a synthetic heparin-pentasaccharide analog (bearing extra sulfate and methyl groups) obtained in a recent crystal study [81] was found to be comparable with that of the naturally occurring heparin-pentasaccharide in the free state. Furthermore, the

Table 5. ϕ, ψ Torsion Angles at the Glycosidic Linkages in Heparin-Pentasaccharide $\text{GlcN}, 6\text{S}\alpha(1\text{-}4)\text{GlcA}\beta(1\text{-}4)\text{GlcN}, 3, 6\text{S}\alpha(1\text{-}4)\text{IdoA}2\text{S}\alpha(1\text{-}4)\text{GlcN}, 6\text{S}\alpha\text{OMe}$ Reproducing Structure of the Specific Binding Sequence of Heparin and Heparan Sulphate for Antithrombin in the Heparin-Pentasaccharide-Antithrombin Complex as Derived from NMR Solution Data and Molecular Modeling [175]. ϕ, ψ Values are Listed for Pentasaccharide also in the Free State [157] for Comparison. ϕ, ψ Torsion Angles, Retrieved from the Crystal Structure, for the Pentasaccharide Analogue (Bearing Extra Sulfate and Methyl Groups) [81] are in the Last Two Lines

	ϕ_1, ψ_1	ϕ_2, ψ_2	ϕ_3, ψ_3	ϕ_4, ψ_4
AT-pentasaccharide complex	-56, -60	40, 12	-32, -43	45, 16
pentasaccharide (free state)	-34, -27	47, 6	-35, -42,	43, 6
Pentasaccharide analogue	-31, -35	41, 1	-71, -33	44, 16
(crystal structure) ¹	-31, -22	35, 8	-66, -35	48, 5

¹ corresponds to two different forms of L- and I- antithrombin

analysis of experimental NOEs in the IdoA2S residue showed that this residue adopts the ²S₀ conformation in the complex [175].

Interestingly, both molecular modeling of the heparin-pentasaccharide-antithrombin complex and crystal data of the synthetic analog with antithrombin indicated that the 2-O-

sulfate group was not directly involved in the binding process. In both cases, no interactions have been observed between this group and any amino acid in the binding site. It thus appears that the 2-sulfate group predominantly stabilizes the ²S₀ form of the IdoA2S residue in the complex through electrostatic interactions between the carboxylate

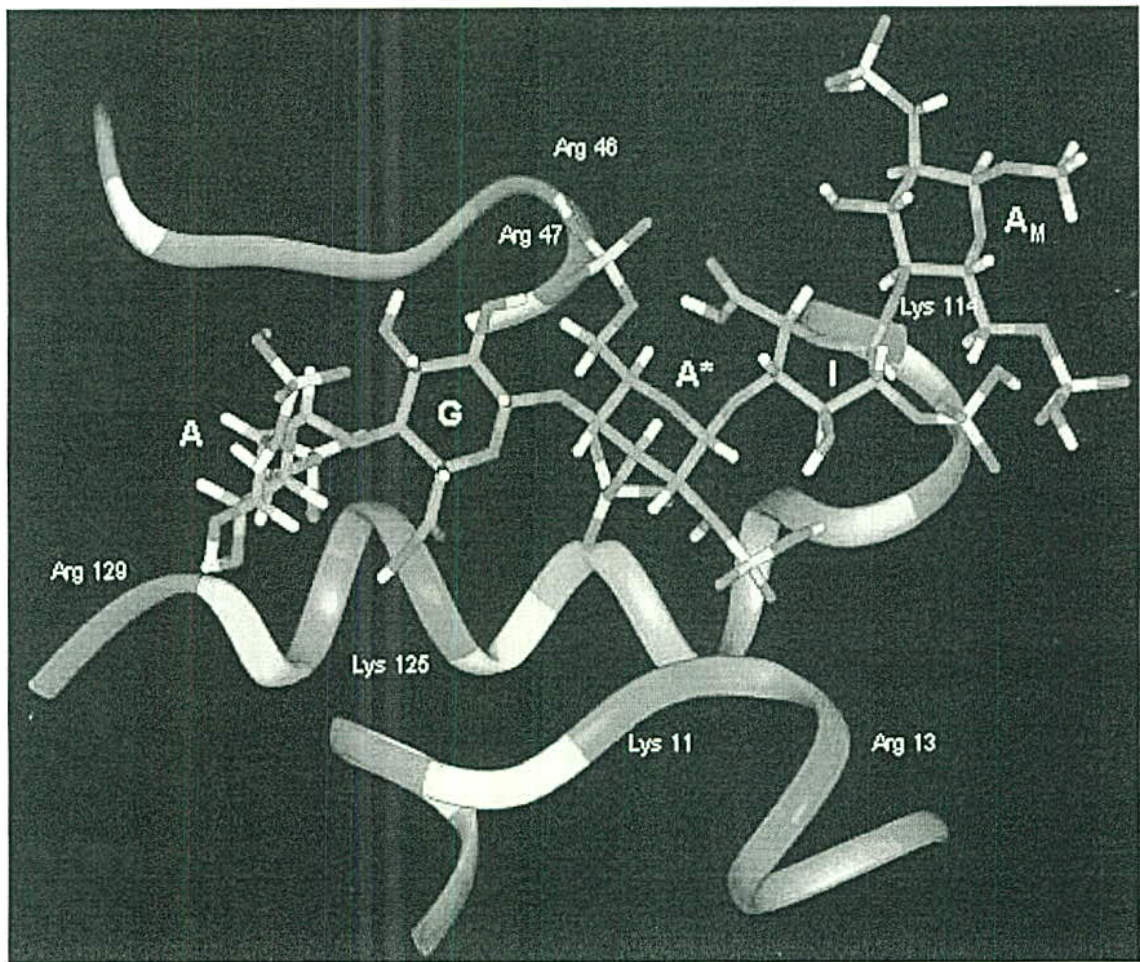


Fig. (5). 3D structure of heparin-pentasaccharide in the complex with plasma protein antithrombin based on high-resolution NMR and molecular modeling. The torsion angles at the glycosidic linkages are listed in Table 5.

group and the Arg 47 and Lys 114 residues. Furthermore, the 2S_0 conformation influences the 3D structure of the reducing end (conformation at the IdoA2S α (1-4)GlcN,6S linkage) enhancing the interactions among the N-sulfate group of the GlcN,6S and Arg 46, Arg 47 and Lys 114 (Fig. (5)). These electrostatic interactions were considerably reduced when the 1C_4 conformation is adopted owing to the different stereochemical arrangement of the IdoA2S α (1-4)GlcN,6S-OMe part. The above evidence indicates that the role of the 2-sulfate group in the IdoA2S residue is essentially affecting the 3D structure of this residue (influencing the chair – skew boat conformational equilibrium) and the part of this pentasaccharide. Such a stereochemical arrangement seems to fit best the 3D structure of the binding site [175, 176] resulting in the high affinity of this pentasaccharide to antithrombin.

The mechanism of the interaction of heparin/heparin sulfate (HS) with fibroblast growth factors (FGFs, FGF-1 and FGF-2) and FGF receptors (FGFR-1 and FGFR-2) have been the subject of thorough studies aimed at the proper understanding of the mechanisms of FGF signal transduction [82-85, 180-185]. The crystal structure of FGF-1 bound to FGFR-2 pointed out on the essential role of HS molecule to cross-link two FGF molecules in the dimer that assembles together two receptor (FGFR-2) molecules creating the complex with 2:2:1 stoichiometry [83]. The analysis of the complex of FGF-2 with FGFR-1 at 3 Å resolution showed that two complexes associate and create the large, positively charged canyon. This cleft represents the HS binding site where two ligand molecules are bound, each occupying one-half of the binding site creating the 2:2:2 stoichiometry complex [85].

One of the key questions in understanding the interaction between HS and FGF involves the role of the chain length of HS as well as the role of N-sulfate and the 6-O-sulfate groups of HS in promoting the assembly of active structures [183, 184]. It has been speculated that 6-O-sulfate groups are important in binding and activation of FGF-1 [186]. Interestingly, 6-O-sulfate groups are probably not crucial for binding of HS with FGF-2, but these groups are important to create the ternary complex with the receptor and to induce the mitogenic activity [85, 187, 188]. In accord with these observations, strong effects of FGF-1 upon shieldings of 1H and ^{13}C resonances in synthetic tetrasaccharides [189] are observed in complexes with FGF-1 [190]. The analysis of the data revealed that the 6-O-sulfate group at the reducing end of the molecule is crucial in interaction with FGF-1. Furthermore, MALDI experiments indicated that tetrasaccharides bearing 6-O-sulfate groups are able to form the 2:1 complexes with FGF-1. Such evidence thus confirmed previous speculations that even relatively short heparin-derived oligosaccharides could create complexes with the *trans* arrangement of two FGF-1 molecules, provided these saccharide derivatives meet proper structural requirements.

High-resolution NMR analysis of structure and dynamics of FGF-2 in the free state and in the presence of heparin-tetrasaccharide and heparin-dodecasaccharide indicated that the presence of tetrasaccharide increased the overall correlation time of the complex. This is compatible with the assumption that tetrasaccharide induces the formation of the dimer of two FGF-2 molecules. In addition,

dodecasaccharide showed even greater effect on the dynamics of the complex due to formation of higher order oligomers. This fact further supports the evidence that the dimer is *trans* oriented in the presence of tetrasaccharide, whereas two *cis* oriented dimers, forming the symmetric tetramer, are assembled in the presence of dodecasaccharide [180].

Recent crystal data of FGF-7 at 3.1 Å resolution provided further insight into the structural basis of growth factors interactions with HS molecules. Crystal data showed that, in addition to N-sulfate and 6-O-sulfate groups, the 3-O-sulfate group might play an important role in interaction with this growth factor [185]. Another requirement for the proper binding appeared the increased length of HS oligosaccharide.

The role of HS-type molecules in cancer biology has also been intensively studied. Cell surface GAGs regulate tumorigenesis, progression and metastasis [191]. Heparin oligosaccharides have shown anti-tumor effects in *in vitro* assays [192]. Although the structural details of antiangiogenic and metastatic properties of GAG molecules are unknown the inhibition of heparanase activity might be the crucial step of the anti-cancer role of heparin-like saccharides [193-196]. Recently, chemically modified heparin showed antiangiogenic activity in *in vivo* models. It was speculated that the increased flexibility in heparin chain is responsible, at least in part, for this effect [197].

There is increasing number of studies dealing with interactions of HS and heparin-derived oligosaccharides with other proteins, such as annexin V [86], heparin-binding growth-associated molecule [198], midkine [199], native vitronectin [200], human immunodeficiency protein model [201], cobra cardiotoxin [202, 203], platelet factor-4 [204], growth-related protein- α [205], selectins [206, 207]. Current efforts are focused particularly on determination of the key functional sulfate groups in heparin-oligosaccharides that are essential for binding with proteins [176, 185, 190, 207]. Other aspects are the lengths of HS chains, variations in 3D structure and dynamics of proteins during the binding process, formation of molecular assemblies. All the above efforts are aimed at detailed description of the structure of protein-carbohydrate complexes and understanding the mechanisms of actions of saccharide molecules in investigated systems in order to design synthetic derivatives with the desired biological activities.

REFERENCES

- [1] Brooks, S.A.; Dwek, M.V.; Schumacher, V. *Functional and Molecular Glycobiology*, BIOS Scientific Publishing: Oxford, 2002.
- [2] Stoddart, J.F. *Stereochemistry of Carbohydrates*, Wiley-Interscience: New York, 1971.
- [3] Computer Modeling of Carbohydrate Molecules, French, A.D.; Brady, J.W. Eds.; ACS series 430: Washington DC, 1990.
- [4] Rao, V.S.R.; Qasba, P.K.; Balaji, P.V.; Chandrasekaran, R. *Conformation of Carbohydrates*, Harwood Academic Publishers: Amsterdam, 1998.
- [5] Imberty, A.; Perez, S.; *Chem. Rev.*, 2000, 100, 4567.
- [6] Daragan, V.A.; Mayo, K.H. *Progr. NMR Spectr.*, 1997, 31, 63.
- [7] Stone, M.J. *Acc. Chem. Res.*, 2001, 34, 379.
- [8] Akerund, T.; Thulin, E.; van Etten, R.L.; Akke, M. *J. Mol. Biol.*, 2002, 322, 137.
- [9] van Halbeek, H. In *Encyclopedia NMR*, Grant, D.M.; Harris, R.K. Eds.; 1996, J. Wiley: New York, p. 1107.

- [10] Duus, J.O.; Gotfredsen, C.H.; Bock, K. *Chem. Rev.*, **2000**, *100*, 4589.
- [11] *NMR Spectroscopy of Glycoconjugates*, Jimenez-Barbero, J.; Peters, T. Eds.; Wiley-VCH: Weinheim, **2002**.
- [12] Rao, V.S.R.; Sundararajan, P.R.; Ramakrishnan, C.; Ramachandran, G.N. *Conformations in Biopolymers*, Academic Press: London, **1963**.
- [13] Lemieux, R.U.; Koto, S.; Voisin, D. In *ACS Symp. Series*; Szarek, W.A.; Horton, D. Eds.; Am. Chem. Soc.: Washington D.C., **1979**; Vol. 87, p. 17.
- [14] Lemieux, R.U.; Bock, K.; Delbaere, T.J.; Koto, S.; Rao, V.S.R. *Can. J. Chem.*, **1980**, *58*, 631.
- [15] Thogersen, T.; Lemieux, R.U.; Bock, K.; Meyer, B. *Can. J. Chem.*, **1982**, *60*, 44.
- [16] Burkert, U.; Allinger, N.L. *Molecular Mechanics*, Am. Chem. Soc. Monograph, **1982**, No. 177.
- [17] Allinger, N.L.; Yuh, Y.H.; Liu, J.H. *J. Am. Chem. Soc.*, **1989**, *111*, 8551.
- [18] Weiner, P.K.; Kollman, P.A. *J. Comput. Chem.*, **1981**, *2*, 287.
- [19] Maple, J.; Dinur, U.; Hagler, A.T. *Proc. Natl. Acad. Sci. USA*, **1988**, *85*, 5350.
- [20] Hagler, A.T.; Lifson, S.; Dauber, P. *J. Am. Chem. Soc.*, **1979**, *101*, 5122.
- [21] Rasmussen, K. *Acta Chem. Scand.*, **1982**, *A36*, 323.
- [22] Tvaroška, I.; Perez, S. *Carbohydr. Res.*, **1986**, *149*, 389.
- [23] Engelsen, S.B.; du Penhoat, C.H.; Perez, S. *J. Phys. Chem.*, **1995**, *99*, 13334.
- [24] Woods, R. In *Rev. Comput. Chem.*, Lipkowitz, K.; Boyd, D.B. Eds.; VCH Publishers: New York, **1996**, Vol. 9, p. 129.
- [25] Tvaroška, I.; Kožár, T. *J. Am. Chem. Soc.*, **1980**, *102*, 6929.
- [26] Parr, R.G.; Yang, W. *International Series of Monographs on Chemistry 16: Density Functional Theory of Atoms and Molecules*, Oxford University Press: New York, **1989**.
- [27] Grant, D.M.; Facelli, J.C.; Alderman, D.W.; Sherwood, M.H. In *Nuclear Magnetic Shieldings and Molecular Structure*; Tossell, J.A. Ed.; Kluwer Academic Publishers: Dordrecht, **1993**, p. 367.
- [28] Hricovini, M.; Malkina, O.L.; Bizik, F.; Turi Nagy, L.; Malkin, V.G. *J. Phys. Chem. A*, **1997**, *101*, 9756.
- [29] Stahl, M.; Schopfer, V.; Frenking, G.; Hoffman, R.W. *J. Org. Chem.*, **1997**, *62*, 3702.
- [30] Bose, B.; Zhao, S.; Stenutz, R.; Cloran, F.; Bondo, P.; Bondo, G.; Hertz, B.; Carmichael, I.; Serianni, A.S. *J. Am. Chem. Soc.*, **1998**, *120*, 11158.
- [31] Malkin, V.G.; Malkina, O.L.; Eriksson, L.A.; Salahub, D.R. In: *Modern Density Functional Theory: A Tool for Chemistry*, Seminario, J.M.; Politzer, P. Eds.; Elsevier: Amsterdam, **1995**, p. 273.
- [32] Tvaroška, I.; Hricovini, M.; Petraková, E. *Carbohydr. Res.*, **1989**, *189*, 359.
- [33] Mulloy, B.; Frenkiel, T.A.; Davies, B.D. *Carbohydr. Res.*, **1988**, *184*, 39.
- [34] Perlin, A.S.; Casu, B. *Tetrahedron Lett.*, **1969**, *34*, 2921.
- [35] Bock, K.; Pedersen, C. *J. C. S. Perkin II*, **1974**, 293.
- [36] Hricovini, M.; Tvaroška, I. *Magn. Res. Chem.*, **1990**, *28*, 862.
- [37] Tvaroška, I.; Taravel, F.R. *J. Biomol NMR*, **1992**, *2*, 421.
- [38] Cloran, F.; Carmichael, I.; Serianni, A.S. *J. Phys. Chem. A*, **1999**, *103*, 3783.
- [39] Malkina, O.L.; Hricovini, M.; Bizik, F.; Malkin, V.G. *J. Phys. Chem. A*, **1997**, *105*, 9188.
- [40] Saito, H. *Magn. Reson. Chem.*, **1986**, *24*, 835.
- [41] Gidley, M.J.; Bocick, S.M. *J. Am. Chem. Soc.*, **1988**, *110*, 3820.
- [42] Bax, A.; Kontaxis, G.; Tjandra, N. *Meth. Enzymol.*, **2001**, *339*, 127.
- [43] Martin-Pastor, M.; Bush, C.A. In *NMR Spectroscopy of Glycoconjugates*, Jimenez-Barbero, J.; Peters, T. Eds.; Wiley-VCH: Weinheim, **2002**, p.23.
- [44] Prestegard, J.H.; Glushka, J. In *NMR Spectroscopy of Glycoconjugates*, Jimenez-Barbero, J.; Peters, T. Eds.; Wiley-VCH: Weinheim, **2002**, p.231.
- [45] Weimar, T.; Woods, R.J. In *NMR Spectroscopy of Glycoconjugates*, Jimenez-Barbero, J.; Peters, T. Eds.; Wiley-VCH: Weinheim, **2002**, p.111.
- [46] Peters, T.; Pinto, B.M. *Curr. Opin. Struct. Biol.*, **1996**, *6*, 710.
- [47] Krishna, N.R.; Moseley, H.N.B. In *Biological Magnetic Resonance*, **1999**, *17*, Kluwer Academic/Plenum Publishers: New York, p. 223.
- [48] Poveda, A.; Jimenez-Barbero, J. *Chem. Soc. Rev.*, **1998**, *27*, 133.
- [49] Jimenez-Barbero, J.; Asensio, J.L.; Canada, F.J.; Poveda, A. *Curr. Opin. Struct. Biol.*, **1999**, *9*, 549.
- [50] Mayer, M.; Meyer, B. *J. Am. Chem. Soc.*, **2001**, *123*, 6108.
- [51] Meyer, B.; Peters, T. *Angew. Chem. Int. Ed.*, **2003**, *42*, 865.
- [52] Moseley, H.N.B.; Curto, E.V.; Krishna N.R. *J. Magn. Reson. B*, **1995**, *108*, 243.
- [53] Jayalakshmi, V.; Krishna N.R. *J. Magn. Reson.*, **2002**, *155*, 106.
- [54] Dais, P. *Carbohydr. Res.*, **1994**, *263*, 13.
- [55] Poppe, L.; van Halbeek, H. *J. Am. Chem. Soc.*, **1992**, *114*, 1092.
- [56] Hricovini, M.; Shah, R.N.; Carver, J.P. *Biochemistry*, **1992**, *31*, 10018.
- [57] Lipari, G.; Szabo, A. *J. Am. Chem. Soc.*, **1982**, *104*, 4546.
- [58] Dayie, K.T.; Wagner, G.; Lefevre, J.F. *Annu. Rev. Phys. Chem.*, **1996**, *47*, 243.
- [59] Brisson, J.R.; Sue, S.C.; Wu, W.; McManus, G.; Nghia, P.T.; Uhrin, D. In *NMR Spectroscopy of Glycoconjugates*, Jimenez-Barbero, J.; Peters, T. Eds.; Wiley-VCH: Weinheim, **2002**, p. 59.
- [60] Hricovini, M.; Tvaroška, I.; Uhrin, D.; Batta, G. *J. Carbohydr. Chem.*, **1989**, *8*, 389.
- [61] Uhrin, D.; Melc, A.; Köver, K.E.; Boyd, J.; Dwek, R.A. *J. Magn. Reson. Ser. A*, **1994**, *108*, 160.
- [62] Uhrin, D.; Barlow, P.N. *J. Magn. Reson.*, **1997**, *126*, 248.
- [63] Kay, L.E.; Nicholson, L.K.; Delaglio, F.; Bax, A.; Torchia, D.A. *J. Magn. Res.*, **1992**, *97*, 359.
- [64] Jeffrey, G.A. *Acta Cryst. B*, **1990**, *46*, 89.
- [65] Rini, J.; Drickamer, K. *Curr. Op. Struct. Biol.*, **1997**, *7*, 615.
- [66] Naismith, J.H.; Field, R.A. *J. Biol. Chem.*, **1996**, *271*, 972.
- [67] Loris, R.; Maccs, D.; Poortmans, F.; Wyns, L.; Bouckaert, J. *J. Biol. Chem.*, **1996**, *271*, 30614.
- [68] Moothoo, D.N.; Naismith, J.H. *Glycobiology*, **1998**, *8*, 173.
- [69] Wright, C.S.; Hester, G. *Structure*, **1996**, *4*, 1339.
- [70] Bourne, Y.; Rouge, P.; Cambillau, C. *J. Biol. Chem.*, **1990**, *265*, 18161.
- [71] Ng, K.K.; Kolatkar, A.R.; Park-Snyder, S.; Feinberg, H.; Clark, D.A.; Drickamer, K.; Weis, W.I. *J. Biol. Chem.*, **2002**, *277*, 16088.
- [72] Wright, C.S. *J. Mol. Biol.*, **1990**, *215*, 635.
- [73] Somers, W.S.; Tang, J.; Shaw, G.D.; Camphausen, R.T. *Cell*, **2000**, *103*, 467.
- [74] Ng, K.K.; Weis, W.I. *Biochemistry*, **1997**, *36*, 979.
- [75] Loris, R.; De Greeve, H.; Dao-Thi, M.H.; Messens, J.; Imberty, A.; Wyns, L. *J. Mol. Biol.*, **2000**, *301*, 987.
- [76] Svensson, C.; Teneberg, S.; Nilsson, C.L.; Kjellberg, A.; Schwarz, F.P.; Sharon, N.; Krengel, U. *J. Mol. Biol.*, **2002**, *321*, 69.
- [77] Audette, G. F.; Olson, D. J. H.; Ross, A. R. S.; Quail, J. W.; Delbaere, L. T. J. *Can. J. Chem.*, **2002**, *80*, 1010.
- [78] Buts, L.; Hamelryck, T. W.; Dao-Thi, M.-H.; Loris, R.; Wyns, L.; Etzler, M. E. J. *J. Mol. Biol.*, **2001**, *309*, 193.
- [79] Delbaere, L.T.; Vandonselaar, M.; Prasad, L.; Quail, J.W.; Wilson, K.S.; Dauter, Z. *J. Mol. Biol.*, **1993**, *230*, 950.
- [80] Jeffrey, P. D.; Bajorath, J.; Chang, C. Y.; Yelton, D.; Hellstrom, I.; Hellstrom, K. E.; Sheriff, S. *Nat. Struct. Biol.*, **1995**, *2*, 466.
- [81] Jin, L.; Abrahams, J.P.; Skinner, R.; Petitou, M.; Pike, R.N.; Carrell, R.W. *Proc. Natl. Acad. Sci. USA*, **1997**, *94*, 14683.
- [82] DiGabriele, A.D.; Lax, I.; Chen, D.I.; Svahn, C.M.; Jac, M.; Schlessinger, J.; Hendrickson, W.A. *Nature*, **1998**, *393*, 812.
- [83] Pellegrini, L.; Burke, D.F.; von Delft, F.; Mulloy, B.; Blundell, T.L. *Nature*, **2000**, *407*, 1029.
- [84] Faham, S.; Hileman, R.E.; Fromm, J.R.; Linhardt, R.J.; Rees, D.C. *Science*, **1996**, *271*, 1116.
- [85] Schlessinger, J.; Plotnikov, A.N.; Ibrahim, O.A.; Eliseenkova, A.V.; Yeh, B.K.; Yanon, A.; Linhardt, R.J.; Mohammadi, M. *Mol. Cell*, **2000**, *6*, 743.
- [86] Capila, I.; Hernaiz, M.J.; Mo, Y.D.; Mealy, T.R.; Campos, B.; Dedman, J.R.; Linhardt, R.J.; Seaton, B.A. *Structure*, **2001**, *9*, 57.
- [87] Kornfeld, R.; Kornfeld, S. *Ann. Rev. Biochem.*, **1985**, *54*, 631.
- [88] Brisson, J.R.; Carver, J.P. *Biochemistry*, **1983**, *22*, 1632.
- [89] Brisson, J.R.; Carver, J.P. *Biochemistry*, **1983**, *22*, 3671.
- [90] Homans, S.W.; Dwek, R.A.; Rademacher, T.W. *Biochemistry*, **1987**, *26*, 6553.
- [91] Edge C.J.; Singh, U.C.; Bazzo, R.; Taylor, G.L.; Dwek, R.A.; Rademacher, T.W. *Biochemistry*, **1990**, *29*, 1971.

- [92] Carver, J.P.; Mandel, D.; Mischnick, S.W.; Imberty, A.; Brady, J.W. In *Computer Modeling of Carbohydrate Molecules*, French, A.D.; Brady, J.W. Eds.; ACS series 430, 1990 p.266.
- [93] Imberty, A.; Perez, S.; Hricovini, M.; Shah, R.N.; Carver, J.P. *Int. J. Biol. Macromol.*, **1993**, *15*, 17.
- [94] Balaji, P.V.; Qasba, P.K. Rao, V.S.R. *Glycobiology*, **1994**, *4*, 495.
- [95] Sayers, E.W.; Prestegard, J.H. *Biophys. J.*, **2000**, *79*, 3313.
- [96] Imberty, A.; Gerber, S.; Tran, V.; Perez, S.; *Glycoconjugate J.*, **1990**, *7*, 27.
- [97] Ruthcrford, T.J.; Homans, S.W. *Biochemistry*, **1994**, *33*, 9606.
- [98] Woods, R.J.; Pathiaseril, A.; Wormald, M.; Edge C.J.; Dwek, R.A. *Eur. J. Biochem.*, **1998**, *258*, 372.
- [99] Tian, F.; Al-Hashimi, H.M.; Crieghead, J.; Prestegard, J.H. *J. Am. Chem. Soc.*, **2001**, *123*, 485.
- [100] Sayers, E.W.; Prestegard, J.H. *Biophys. J.*, **2002**, *82*, 2683.
- [101] Almond, A.; Bunkenborg, J.; Franch, T.; Gottfredsen, C.H.; Duus, J.O. *J. Am. Chem. Soc.*, **2001**, *123*, 4792.
- [102] Williams, B.A.; Chervcnak, M.C.; Toone, E.J. *J. Biol. Chem.*, **1992**, *267*, 22907.
- [103] Mandal, D.K.; Kishore, N.; Brewer, C.F. *Biochemistry*, **1994**, *33*, 1149.
- [104] Navarre, N.; Amiot, N.; van Oigen, A.; Imberty, A.; Poveda, A.; Jimenez-Barbero, J.; Cooper, A.; Nutley, M.A.; Boons, G.J. *Chem. Eur. J.*, **1999**, *5*, 2281.
- [105] Thijssen-van Zuylen, C.W.E.M.; de Beer, T.; Leeftang, b.R.; Boelens, R.; Kaptein, R.; Kamerling, J.P.; Vliegthart, J.F.G. *Biochemistry*, **1998**, *37*, 1933.
- [106] Lu, J.; van Halbeck, H. *Biophys. J.*, **1997**, *72*, 470.
- [107] Stubbs, H.J.; Lih, J.J.; Gustafson, T.L.; Rice, K. *Biochemistry*, **1996**, *35*, 937.
- [108] Kimura, N.; Uchida, M.; Nishimura, S.; Yamaguchi, H. *J. Biochem.*, **1998**, *124*, 857.
- [109] Wyss, D.F.; Choi, J.S.; Li, J.; Knoppers, G. *Science*, **1995**, *269*, 1273.
- [110] Toyoda, T.; Arakawa, T.; Yamaguchi, H. *J. Biochem.*, **2002**, *131*, 511.
- [111] Jitsuhara, Y.; Toyoda, T.; Itai, T.; Yamaguchi, H. *J. Biochem.*, **2002**, *132*, 803.
- [112] Ermonval, M.; Duvet, S.; Zonneveld, D.; Cacan, R.; Buttin, G.; Braakman, I. *Glycobiology*, **2000**, *10*, 77.
- [113] Bell, S.L.; Xu, G.; Khatri, I.A.; Wang, R.; Rahman, R.; Forstner, J.F. *Biochem. J.*, **2003**, *373*, 893.
- [114] Biswas, M.; Rao, V.S.R. *Biopolymers*, **1980**, *19*, 1555.
- [115] Bush, A.C.; Yan, Z.Y.; Rao, N.B.N. *J. Am. Chem. Soc.*, **1986**, *108*, 6168.
- [116] Rao, N.B.N.; Dua, V.K.; Bush, A.C. *Biopolymers*, **1985**, *24*, 2207.
- [117] Cagas, P.; Bush, C.A. *Biopolymers*, **1992**, *32*, 277.
- [118] Casset, F.; Peters, T.; Etzler, M.; Korchagina, E.; Nifantev, N.; Perez, S.; Imberty, A. *Eur. J. Biochem.*, **1996**, *239*, 710.
- [119] Imberty, A.; Mikros, E.; Koca, J.; Mollicone, R.; Oriol, R.; Perez, S.; *Glycoconjugate J.*, **1995**, *12*, 331.
- [120] Koca, J.; Perez, S.; Imberty, A. *J. Comp. Chem.*, **1995**, *16*, 296.
- [121] Yan, S.Y.; Bush, C.A. *Biopolymers*, **1990**, *29*, 799.
- [122] Bukowski, R.; Morris, L.M.; Woods, R.J.; Weimar, T. *Eur. J. Org. Chem.*, **2001**, 2697.
- [123] Weimar, T.; Bukowski, R.; Young, N.M. *J. Biol. Chem.*, **2000**, *275*, 37006.
- [124] Lee, X.; Thompson, A.; Zhang, Z.; Ton-that, H.; Biesterfeldt, J.; Ogata, C.; Xu, L.; Johnston, A.Z.; Young, N.M. *J. Biol. Chem.*, **1998**, *273*, 6312.
- [125] Peters, T.; Perez, S.; Imberty, A. *Theochem. J. Mol. Struct.*, **1997**, *395*, 297.
- [126] Perez, S.; Mouhous-Riou, N.; Tsvetkov, Y.E.; Bachet, B.; Imberty, A. *Glycobiology*, **1996**, *6*, 537.
- [127] Bizik, F.; Tvaroška, I. *Chem. Papers*, **1996**, *50*, 84.
- [128] Poveda, A.; Asensio, J.L.; Martin-Pastor, M.; Jimenez-Barbero, J. *J. Biomol. NMR*, **1997**, *10*, 29.
- [129] Azumendi, H.F.; Pastor, M.M.; Bush, C.A. *Biopolymers*, **2002**, *63*, 89.
- [130] Haselhorst, T.; Weimar, T.; Peters, T. *J. Am. Chem. Soc.*, **2001**, *123*, 10705.
- [131] Ichikawa Y.; Lin, Y.C.; Dumas, D.P.; Shen, G.J.; Garcia-Junceda, E.; Williams, M.A.; Bayer, R.; Ketcham, C.; Walker, L.E.; Paulson, J.C.; Wong, C.H. *J. Am. Chem. Soc.*, **1992**, *114*, 9283.
- [132] Ruthcrford, T.J.; Spackman, D.G.; Simpson, P.J.; Homans, S.W. *Glycobiology*, **1994**, *4*, 59.
- [133] Poppe, L.; Brown, G.S.; Philo, J.S.; Nikrad, P.V.; Shah, B.H. *J. Am. Chem. Soc.*, **1997**, *119*, 1727.
- [134] Harris, R.; Kiddle, G.R.; Field, R.A.; Milton, M.J.; Ernst, B.; Magnani, J.L.; Homans, S.W. *J. Am. Chem. Soc.*, **1999**, *121*, 2546.
- [135] Breg, J.; Kroon-Batenburg, L.M.J.; Strecker, G.; montreuil, J.; Vliegthart, J.F.G. *Eur. J. Biochem.*, **1989**, *178*, 727.
- [136] Wu, W.-G.; Pasternack, L.; Huang, D.-H.; Koeller, K.M.; Lin, C.-C.; Scitz, O.; Wong, C.-H. *J. Am. Chem. Soc.*, **1999**, *121*, 2409.
- [137] Schaffler, K.; Brisson, J.R.; Weisemann, R.; Magnani, J.L.; Wong, W.T.; Ernst, B.; Peters, T. *J. Biomol. NMR*, **1997**, *9*, 423.
- [138] Conrad, H.E. *Heparin-Binding Proteins*. Academic Press, San Diego **1998**.
- [139] Mulloy, B.; Forster, M.J. *Glycobiology*, **2000**, *10*, 1147.
- [140] Sasisekharan, R.; Venkataraman, G. *Curr. Opin. Chem. Biol.*, **2000**, *4*, 626.
- [141] Casu, B.; Lindahl, U. *Adv. Carbohydr. Chem. Biochem.*, **2001**, *57*, 159.
- [142] Capila, I.; Linhardt, R.J. *Angew. Chem. Int. Ed.*, **2002**, *41*, 390.
- [143] Casu, B. *Adv. Carbohydr. Chem. Biochem.*, **1985**, *43*, 51.
- [144] Casu, B.; Naggi, A.; Torri, G. In *Glycoscience- Chemistry and Chemical Biology*, Thiem, J.; Ed., Springer Verlag, **2000**, p.1895.
- [145] Hricovini, M.; Nieto, P.M.; Torri, G. In *NMR Spectroscopy of Glycoconjugates*, Jimenez-Barbero, J. Peters, T.; Eds.; Wiley-VCH, **2002**, p. 189.
- [146] Torri, G.; Casu, B.; Gatti, G.; Petitou, M.; Choay, J.; Jacquinot, J.C.; Sinay, P. *Biochem. Biophys. Res. Com.* **1985**, *128*, 134.
- [147] Casu, B.; Choay, J.; Ferro, D.R.; Gatti, G.; Jacquinot, J.-C.; Petitou, M.; Provasoli, A.; Ragazzi, M.; Sinay, P.; Torri, G. *Nature*, **1986**, *322*, 215.
- [148] Ferro, D.F.; Provasoli, A.; Ragazzi, M.; Torri, G.; Casu, B.; Gatti, G.; Jacquinot, J.-C.; Sinay, P.; Petitou, M.; Choay, J. *J. Am. Chem. Soc.*, **1986**, *108*, 6773.
- [149] Hricovini, M.; Guerrini, M.; Bisio, A. *Eur. J. Biochem.*, **1999**, *261*, 789.
- [150] Ferro, D.R.; Provasoli, A.; Ragazzi, M.; Casu, B.; Torri, G.; Bossennec, V.; Perly, B.; Sinay, P.; Petitou, M.; Choay, J. *Carbohydr. Res.*, **1990**, *195*, 157.
- [151] Ernst, S.; Venkataraman, G.; Sasisekharan, V.; Langer, R.; Cooney, C.L.; Sasisekharan, R. *J. Am. Chem. Soc.*, **1998**, *120*, 2099.
- [152] Rabenstein, D.L.; Robert, J.M.; Peng, J. *Carbohydr. Res.*, **1995**, *278*, 239.
- [153] Angelo, J.; De Paz, J.-L.; Nieto, P.M.; Martin-Lomas, M. *Isr. J. Chem.*, **2000**, *40*, 289.
- [154] Ofosu, F.A.; Danishefsky, I.; Hirsh J. (Eds.): *Heparin and Related Polysaccharides*. New York Acad. Sci.: New York, **1989**.
- [155] Jacques, L.B. *Semin. Thromb. Hemost.*, **1991**, *17*, 1.
- [156] Mulloy, B.; Forster, M.J.; Jones, C.; Davies, D.B. *Biochem. J.*, **1993**, *293*, 849.
- [157] Ragazzi, M.; Ferro, D.F.; Perly, B.; Sinay, P.; Petitou, M.; Choay, J. *Carbohydr. Res.*, **1990**, *195*, 169.
- [158] Mikhailov, D.; Mayo, K.H.; Vlahov, I.R.; Toida, T.; Pervin, A.; Linhardt, R.J. *Biochem. J.*, **1996**, *318*, 93.
- [159] Mikhailov, D.; Linhardt, R.J.; Mayo, K.H. *Biochem. J.*, **1997**, *328*, 51.
- [160] Shen, J.; Lerner, L.E. *Carbohydr. Res.*, **1995**, *273*, 115.
- [161] Yates, E.A.; Santini, F.; De Cristofano, B.; Payre, N.; Cosentino, C.; Guerrini, M.; Naggi, A.; Torri, G.; Hricovini, M. *Carbohydr. Res.*, **2000**, *329*, 239.
- [162] Scott, J.E.; Heatley, F.; Wood, B. *Biochemistry*, **1995**, *34*, 15467.
- [163] Mulloy, B.; Forster, M.J.; Jones, C.; Drake, A.F.; Johnson, E.A.; Davies, D.B. *Carbohydr. Res.*, **1994**, *255*, 1.
- [164] Casu, B.; Petitou, M.; Provasoli, A.; Sinay, P. *Trends Biochem. Sci.*, **1988**, *13*, 221.
- [165] Hricovini, M.; Torri, G. *Carbohydr. Res.*, **1995**, *268*, 159.
- [166] Mikhailov, D.; Mayo, K.H.; Pervin, A.; Linhardt, R.J. *Biochem. J.*, **1996**, *315*, 447.
- [167] Hricovini, M.; Guerrini, M.; Torri, G.; Piani, S.; Ungarelli, F. *Carbohydr. Res.*, **1995**, *277*, 11.
- [168] Hricovini, M. Guerrini, M.; Torri, G.; Casu, B. *Carbohydr. Res.*, **1997**, *300*, 69.
- [169] Atha, D.H.; Lormeau, J.-C.; Petitou, M.; Rosenberg, R. D.; Choay, J. *Biochemistry* **1985**, *24*, 6723.
- [170] Petitou, M.; Lormeau, J.-C.; Choay, J. *Eur. J. Biochem.* **1988**, *176*, 637.
- [171] Desai, U. R.; Petitou, M.; Bjork, I.; Olson, S.T. *J. Biol. Chem.* **1998**, *273*, 7478.

- [172] Arocas, V.; Bock, S.C.; Olson, S.T.; Bjork, I. *Biochemistry* **1999**, *38*, 10196.
- [173] Sissi, C.; Lucatello, L.; Naggi, A.; Torri, G.; Palumbo, M. *Semin. Thromb. Hemost.*, **2002**, *28*, 355.
- [174] Johnson, D.J.D.; Huntington, J.A. *Biochemistry*, **2003**, *42*, 8712.
- [175] Hricovini, M.; Guerrini, M.; Bisio, A.; Torri, G.; Petitou, M.; Casu, B. *Biochem. J.*, **2001**, *359*, 265.
- [176] Hricovini, M.; Guerrini, M.; Bisio, A.; Torri, G.; Naggi, A.; Casu, B. *Semin. Thromb. Hemost.*, **2002**, *28*, 325.
- [177] Bitomsky, W.; Wade, R.C. *J. Am. Chem. Soc.*, **1999**, *121*, 3004.
- [178] Grootenhuis, P.D.J.; Westerduin, P.; Meuleman, D.; Petitou, M.; van Boeckel, C.A.A. *Nat. Struct. Biol.*, **1995**, *2*, 736.
- [179] van Boeckel, C.A.A.; Grootenhuis, P.D.J.; Visser, A. *Nat. Struct. Biol.*, **1994**, *1*, 423.
- [180] Moy, F.J.; Safran, M.; Seddon, A.P.; Kitchen, D.; Böhlen, P.; Aviczor, D.; Yayon, A.; Powers, R. *Biochemistry*, **1997**, *36*, 4782.
- [181] Lam, K.; Rao, V.S.R.; Quasba, P.K. *J. Biomol. Struct. Dyn.*, **1998**, *15*, 1009.
- [182] Plotnikov, A.N.; Schlessinger, J.; Hubbard, S.R.; Mohammadi, M. *Cell*, **1999**, *98*, 641.
- [183] Shriver, Z.; Raman, R.; Venkataraman, G.; Drummond, K.; Turnbull, J.; Toida, T.; Linhardt, R.J.; Biemann, K.; Sasisekharan, R. *Proc. Natl. Acad. Sci. USA*, **2000**, *97*, 10359.
- [184] Pellegrini, L. *Curr. Opin. Struct. Biol.*, **2001**, *11*, 629.
- [185] Ye, S.; Luo, Y.; Lu, W.; Jones, R.B.; Linhardt, R.J.; Capila, I.; Toida, T.; Kan, M.; Pelletier, H.; McKechan, W.L. *Biochemistry*, **2001**, *40*, 14429.
- [186] Pey, D.A.; Vives, R.R.; Hyde, P.; Gallagher, J.T. *Glycobiology*, **2000**, *10*, 1183.
- [187] Guimond, S.; Maccarana, M.; Olwin, B.B.; Lindahl, U.; Rapraeger, A.C. *J. Biol. Chem.*, **1993**, *268*, 23906.
- [188] Presta, M.; Leali, D.; Stabile, H.; Ronca, R.; Camozzi, M.; Coco, L.; Moroni, E.; Liekens, S.; Rusnati, M. *Curr. Pharmaceutical Design*, **2003**, *9*, 553.
- [189] Poletti, L.; Lay, L.; Fleischer, M.; Vogel, C.; Guerrini, M.; Torri, G.; Russo, G.A. *Eur. J. Org. Chem.*, **2001**, *14*, 2727.
- [190] Guerrini, M.; Agulles, T.; Bisio, A.; Hricovini, M.; Lay, L.; Naggi, A.; Poletti, L.; Sturiale, L.; Torri, G.; Casu, B. *Biochem. Biophys. Res. Commun.*, **2002**, *292*, 222.
- [191] Sasisekharan, R.; Shriver, Z.; Venkataraman, G.; Narayanasami, U. *Nature Rev.*, **2002**, *2*, 521.
- [192] Parish, C.; Freeman, C.; Brown, K.; Francis, D.J.; Cowden, W.B. *Cancer Res.*, **1999**, *59*, 3433.
- [193] Vlodavsky, I.; Friedmann, Y.; Elkin, M.; Aingorn, H.; Atzmon, R.; Michaeli, R.I.; Bitan, M.; Pappo, O.; Peretz, T.; Michal, I.; Spector, L.; Pecker, I. *Nature Med.*, **1999**, *5*, 793.
- [194] Hullet, M.D.; Freeman, C.; Hamdorf, B.J.; Baker, R.T.; Harris, M.J.; Parish, C.R. *Nature Med.*, **1999**, *5*, 803.
- [195] Okada, Y.; Yamada, S.; Toyoshima, M.; Dong, J.; Nakajima, M.; Sugahara, K. *J. Biol. Chem.*, **2002**, *277*, 42488.
- [196] Levy-Adam, F.; Miao, H.Q.; Heinrikson, R.L.; Vlodavsky, I.; Ilan, N. *Biochem. Biophys. Res. Commun.*, **2003**, *308*, 885.
- [197] Casu, B.; Guerrini, M.; Naggi, A.; Perez, M.; Torri, G.; Ribatti, D.; Carminati, P.; Giannini, G.; Penco, S.; Pisano, C.; Belleri, M.; Rusnati, M.; Presta, M. *Biochemistry*, **2002**, *41*, 10519.
- [198] Kilpelainen, I.; Kaksonen, M.; Kinnunen, T.; Avikainen, H.; Fat, M.; Linhardt, R.J.; E. Raulo, E.; Rauvala, H. *J. Biol. Chem.* **2000**, *275*, 13564.
- [199] Iwasaki, W.; Nagata, K.; Hatanaka, H.; Inui, T.; Kimura, T.; Muramatsu, T.; Yoshida, K.; Tasumi, M.; Inagaki, F. *EMBO J.* **1997**, *23*, 6936.
- [200] Gibson, A.D.; Lamerdin, J.A.; Zhuang, P.; Baburaj, K.; Serpersu, E.H.; Peterson, C.B. *J. Biol. Chem.* **1999**, *274*, 6432.
- [201] Jeon, K.J.; Katsuraya, K.; Inazu, T.; Kaneko, Y.; Mimura, T.; Uryu, T. *J. Am. Chem. Soc.* **2000**, *122*, 12536.
- [202] Sue, S.C.; Brisson, J.R.; Chang, S.C.; Huang, W.N.; Lee, S.C.; Jarrell, H.C.; Wu, W. *Biochemistry* **2001**, *40*, 10436.
- [203] Sue, S.C.; Chien, K.Y.; Huang, W.N.; Abraham, J.K.; Chen, K.M.; Wu, W. *J. Biol. Chem.*, **2001**, *277*, 2666.
- [204] Mayo, K.H.; Roongta, V.; Ilyina, E.; Dundas, M.; Joseph, J.; Lai, C.K.; Maione, T.; Daly, T.J. *Biochem. J.* **1995**, *312*, 357.
- [205] Mikhailov, D.; Young, H.C.; Linhardt, R.H.; Mayo, K.H. *J. Biol. Chem.* **1999**, *274*, 25317.
- [206] Norgard-Sumnicht, K.E.; Varki, N.M.; Varki, A. *Science*, **1993**, *261*, 480.
- [207] Wang, L.; Brown, J.R.; Varki, A.; Esko, J.D. *J. Clin. Invest.*, **2002**, *110*, 127.



Glycoprotein folding, quality control and ER-associated degradation

Gerardo Z Lederkremer

Nascent N-linked glycoproteins possess a large oligosaccharide precursor, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$, which is later sequentially trimmed. Recent studies help understand the code displayed by each structure produced by this trimming and its decoding by lectins. The calnexin folding cycle targets only monoglucosylated oligosaccharides. N-glycans of misfolded glycoproteins are then more extensively trimmed than once thought, being targeted for degradation by removal of three or four mannose residues. A high local concentration of endoplasmic reticulum (ER) mannosidase I in an ER-derived quality control compartment is mainly responsible for this trimming, with the possible participation of other mannosidases. The shortened chains, $\text{Man}_{5-6}\text{GlcNAc}_2$, are recognized by the ubiquitination machinery-associated lectin OS9 but not by lectins that associate with properly folded glycoproteins en route to the Golgi that bind best to $\text{Man}_{8-9}\text{GlcNAc}_2$.

Address

Department of Cell Research and Immunology, George Wise Faculty of Life Sciences, Tel Aviv University, Tel Aviv 69978, Israel

Corresponding author: Lederkremer, Gerardo Z (gerardo@post.tau.ac.il)

Current Opinion in Structural Biology 2009, 19:515–523

This review comes from a themed issue on
 Carbohydrates and glycoconjugates
 Edited by Nathan Sharon, John Gallagher

Available online 17th July 2009

0959-440X/\$ – see front matter
 © 2009 Elsevier Ltd. All rights reserved.

DOI 10.1016/j.sbi.2009.06.004

Introduction

The vast majority of secretory proteins are N-linked glycoproteins. A puzzling issue, for long, has been why a large oligosaccharide precursor, $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (G3M9, Figure 1) is needed for protein N-glycosylation, if it will later be extensively trimmed. Knowledge has been gained on the importance of oligosaccharide structures, products of stepwise processing in the ER, where N-glycosylation occurs, which determine the recognition and targeting of glycoprotein molecules during their folding and delivery to the Golgi or during their misfolding and targeting to ER-associated degradation (ERAD) [1,2]. It was thought that removal of one mannose from the deglycosylated precursor M9 to yield M8B (Figure 1)

targets a terminally misfolded glycoprotein to ERAD. However, a glycoprotein carrying M8B can still be reglucosylated and remain in the calnexin (CNX) folding cycle and later exit to the Golgi. Recent studies indicate that glycoprotein targeting to ERAD involves extensive trimming of M9 to M6 and M5 (Figure 1) [2,3]. This again could pose a problem, as N-glycans of properly folded glycoproteins are also later trimmed to M5. The solution lies in compartmentalization. Properly folded glycoproteins are trimmed in the Golgi, after the quality control checkpoint and are thus released for further processing and productive maturation. By contrast, recent evidence shows that misfolded glycoproteins are trimmed in the ER-derived quality control compartment (ERQC) and are then trapped by the OS9 lectin and delivered to ERAD.

Early stages of glycoprotein folding and quality control—the calnexin cycle

Calnexin and calreticulin

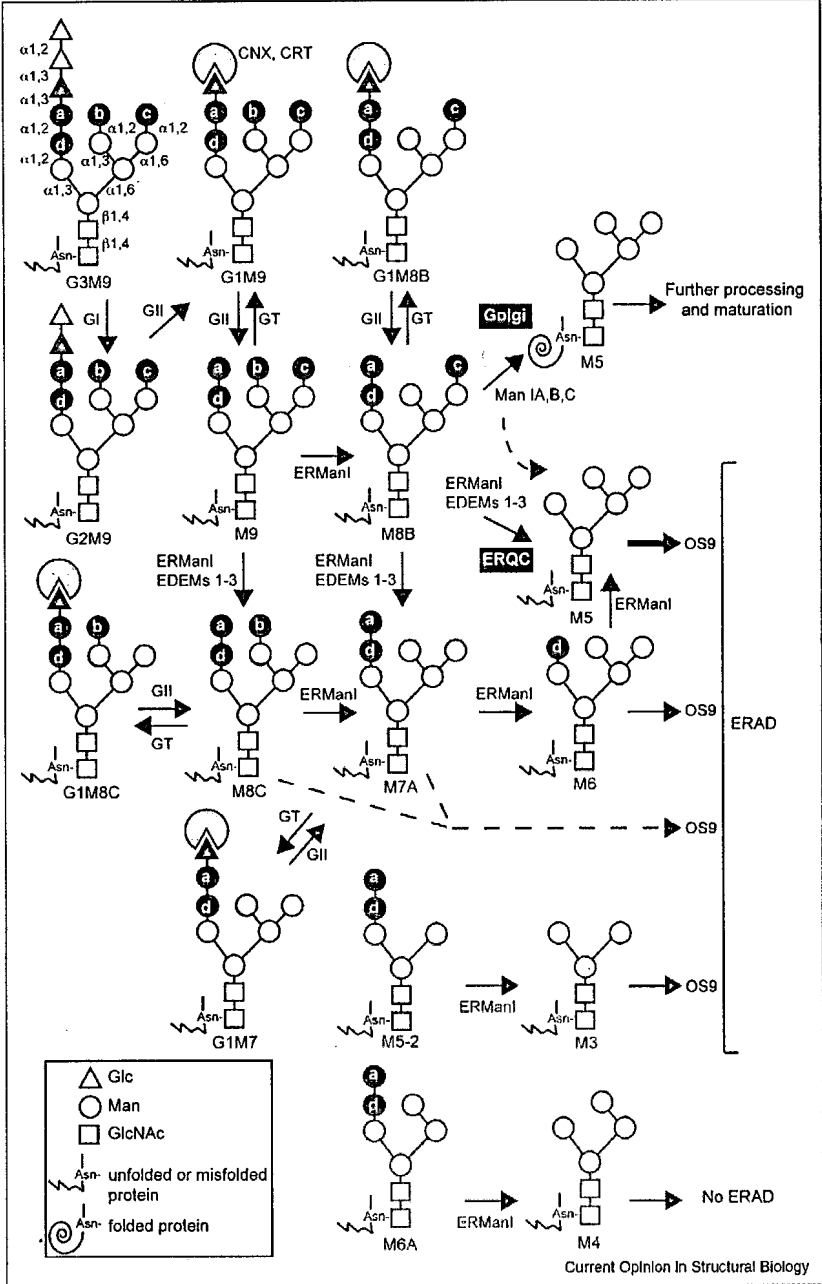
Membrane-bound CNX and its soluble luminal homologue, calreticulin (CRT), are chaperones that associate with nascent glycoproteins after the trimming of two glucose residues from the precursor N-linked oligosaccharide G3M9 by glucosidases I and II (GI and GII), participating in the ER-retention of these glycoproteins until they are properly folded and preventing their aggregation (reviewed in [4]). Dissociation of CNX from the glycoproteins is elicited by removal by glucosidase II of the third, mannose-linked glucose, which can be re-added by UDPGlc:glycoprotein glucosyltransferase (GT). This ‘CNX cycle’ continues until proper folding of the glycoprotein is achieved, which prevents further recognition by the folding sensor GT, allowing exit to the Golgi (Figures 1 and 2). It is important to note that the CNX cycle does not exist in *S. cerevisiae* [4].

Besides their lectin activity, CNX and CRT can bind non-glycosylated proteins and prevent their aggregation *in vitro* [5–7], also shown recently *in vivo* [8*]. Nevertheless, most CNX and CRT binding is carbohydrate dependent [4]. A reasonable model is that the initial binding is generally carbohydrate dependent, leading then to direct peptide binding.

GT and GII

GT recognizes both the core pentasaccharide (M3 in Figure 1) in N-glycans and exposed hydrophobic regions of target glycoproteins [9,10*]. This is consistent with the wide specificity of GT as long as the acceptor mannose

Figure 1



Trimming of the N-glycan precursor during glycoprotein folding and maturation or delivery to ERAD. The precursor oligosaccharide $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ (G3M9), attached to Asn on the nascent polypeptide chain is trimmed by removal of the non-reducing terminal glucose residue by glucosidase I (GI) to G2M9 and of the second glucose by glucosidase II (GII) to G1M9, allowing binding of calnexin (CNX) or calreticulin (CRT). CNX folding cycles then start, where the glycoprotein is deglycosylated by GII and reglucosylated by UDPGlc:glycoprotein glucosyltransferase (GT). During this process mannose-b can be excised from M9 by ER mannosidase I (ERManI). If the glycoprotein is properly folded, it is no longer recognized as a substrate by GT and it exits to the Golgi where it is further trimmed of all its α 1,2 linked mannose residues (blue) by Golgi mannosidases IA, IB and IC to $\text{Man}_5\text{GlcNAc}_2$ (M5). If the glycoprotein cannot fold successfully, it is delivered to the ER-derived quality control compartment (ERQC) and further trimmed to M6 and M5 by ERManI with the assistance of the EDEMs 1-3, and possibly of recycled Golgi α 1,2 mannosidases. This trimming leads to binding of the glycoprotein by the lectin OS9, which commits the glycoprotein to ERAD [21**,26**]. Removal of mannose-c from M9 or from M8B, to yield M8C and M7A, respectively, allows recognition by OS9. However, glycoprotein molecules bearing M8C or M7A can still be rescued from the ERQC and transported to the Golgi by reglucosylation in the ER and proper folding. Trimming of mannose-a from these structures is probably required for efficient ERAD (dashed arrows). The faded lower structures are not naturally present in mammalian cells or yeast. M5-2 is present in wild-type *Trichomonas* [25],

residue (a, see Figure 1) is present. As for GII, recent reports show similar kinetics for removal of Glc from G1M9, G1M8 or G1M7 [11,12] (Figure 1). GT is required for certain slow-folding or misfolded domains in glycoproteins [13*] but not for normal folding of many glycoproteins, implying that they can complete their folding in a single CNX binding event [14*]. Consistent with this, GT was found to be essential in lower eukaryotes only under conditions of high stress [15].

Erp57 and other oxidoreductases

Of the several oxidoreductases and disulfide isomerases that exist in the ER, Erp57 acts specifically on glycoproteins by forming a complex with CNX or CRT [16]. CNX and Erp57 associate with nascent glycoproteins in a complex but can dissociate independently from the bound glycoprotein. Erp57 dissociates rapidly from a misfolded glycoprotein, whereas its association persists to a substrate undergoing productive folding, the opposite being true for CNX [17]. Furthermore, Erp57 does not need to interact with CNX and CRT to act as a chaperone and, for example, promote assembly of MHC class I [18].

Another recently described selenocysteine-containing oxidoreductase, Sep15, is suggested to participate in glycoprotein folding in a complex with GT [19].

Glycoprotein refolding attempts and targeting to ERAD

ER mannosidase I and mannose trimming

A crucial decision is when to abandon refolding attempts through the CNX cycle and to target the glycoprotein to ERAD. Trimming of α 1,2 linked mannose residues is essential for this process and recently the extent of trimming and the required oligosaccharide structures have become more clear [3]. It was initially proposed that removal of a single specific mannose residue (mannose-b in Figure 1) from M9 to generate M8B would mark the glycoprotein for ERAD. However, M8B-bearing glycoproteins remain in the CNX cycle [12] and can be later targeted to the Golgi. In *S. cerevisiae*, lack of the middle mannose-b (as in M8B, Figure 1) is not sufficient for ERAD [20**,21**]. In mammalian cells it is not either, as in mutants that transfer a particular M5 isomer to proteins (M5-2, also devoid of mannose-b, Figure 1), removal of mannose-a residue is still required for ERAD [22]. In wild type mammalian cells it was found that ERAD substrates are processed to M6 and M5, which means that most or all α 1,2 linked mannose residues are removed [23**,24]. A similar determinant for ERAD may also exist in lower eukaryotes. During ERAD in *S. cerevisiae* trimming to an M7 isomer lacking mannose-c (Figure 1) and to a smaller

amount of M5 was observed [20**]. However, this was without inhibiting the proteasomes, which are in charge of degradation in ERAD. Under these conditions most molecules bearing M5 could have been rapidly degraded in the yeast cells like in mammalian cells [24]. In mutants that do not allow exposure of a non-reducing terminal α 1,6-linked mannose, ERAD is blocked (M6A, Figure 1), suggesting that this is an essential determinant, but perhaps also trimming of mannose residues a and d (Figure 1) is necessary in yeast. Consistent with a possible extensive trimming of most or all α 1,2 mannose residues also in *S. cerevisiae*, M5 binds with the highest affinity to the lectin Yos9, whereas M8B binding is negligible (Figure 3) [21**]. Mammalian OS9 and yeast Yos9 are proposed as receptors for misfolded glycoproteins after the mannose trimming step, as described below. In another lower eukaryote, *Trichomonas*, the oligosaccharide transferred to glycoproteins is M5-2 (Figure 1). Mannose residues are removed from this isomer to yield M3 [25*], similar to what was found in mutant mammalian cells [22], suggesting requirement for removal of mannose residues a and d (Figure 1).

Although there might be differences among species, altogether the recent studies suggest that the removal of all or most α 1,2 mannose residues (a-c and perhaps also d) to yield M6 and M5 could be universal in ERAD. Removal of mannose-a takes away the glycoprotein irreversibly from the CNX cycle by eliminating the glucose acceptor and thus precluding CNX binding. Consistently, while M5 binds with high affinity to OS9 and the highest to yeast Yos9, it has no affinity for the lectins VIPL, VIP36 and ERGIC53, involved in ER-Golgi transport (Figures 2 and 3) [26**,27**]. The same M5 structure as in the ER is produced in the Golgi, but in the latter case mannose trimming to M5 would release the glycoprotein from the lectins involved in ER-Golgi transport and allow further maturation through the secretory pathway (Figure 1). By contrast, for misfolded ER-retained glycoproteins, processing to M5 would prevent binding to the ER-Golgi transport lectins and would allow trapping by OS9 and delivery to ERAD. Although mammalian OS9 can also bind M8C and M7A *in vitro* (Figure 3) [26**], it is uncertain that these species would reach OS9 *in vivo* as glycoproteins carrying them will be engaged in the CNX cycle, and probably mannose-a must be removed from them for efficient ERAD, similar to the case of M5-2 discussed above (Figure 1).

An important question is which enzymes are in charge of this crucial decision-making process that targets a misfolded glycoprotein to ERAD. Recent evidence points to

(Figure 1 Legend Continued) in the delta Alg3 *S. cerevisiae* mutant [20**] and in mutant mammalian Mad1A214 cells [22] and glycoproteins carrying it can undergo ERAD without the need of EDEM1 but with the requirement of ERMnl. M5-2 is trimmed to M3, which can be recognized by OS9 [21**]. Glycoproteins carrying the M6A isomer produced in the *S. cerevisiae* delta Alg9 mutant cannot undergo ERAD because they lack a non-reducing terminal α 1,6 linked mannose [20**,21**].

Figure 2

	Protein (1)	Specificity (2)	Localization	Function
Lectins	Calnexin	G1M9-7	ER, ERQC	Folding and quality control
	Calreticulin	G1M9-7	ER, ERQC	Folding and quality control
	Calmeglin	G1M9-7	ER	Folding and quality control? (3)
	Malectin	G2M9	ER	Unknown
	Intracellular MBP (4)	High mannose	ER, ER exit sites	ER to Golgi transport?
	ERGIC-53	High mannose	ERGIC	ER to Golgi transport
	ERGL (5)	Unknown	Unknown	ER to Golgi transport?
	VIP36	M9-M7	ERGIC, cis-Golgi	ER to Golgi transport
	VIPL	M9-M7	ER	ER to Golgi transport
	OS9	M8C-M5 (6)	ER, ERQC? (7)	ERAD
	Fbs1 (8)	High mannose	Cytosol	ubiquitination
	Fbs2 (8)	High mannose	Cytosol	ubiquitination
	EDEM1 (9)	High mannose? (10)	ER, ERQC, vesicles	ERAD
	EDEM2 (9)	High mannose? (10)	ER?	ERAD
	EDEM3 (9)	High mannose? (10)	ER?	ERAD
Glycosidases	ERManI	M9-M6	ERQC	ERAD
	Golgi ManIA	M9-M6	ER, ERGIC, cis-Golgi?	Maturation and ERAD? (11)
	Golgi ManIB	M9-M6	Cis-Golgi	Maturation and ERAD? (11)
	Golgi ManIC	M9-M6	Cis-Golgi	Maturation and ERAD? (11)
	GI	G3M9	ER	Glucose trimming
	GII	G2M9, G1M9-G1M7	ER	Glucose trimming, quality control
Glycosyltransferases	PNGase	Any N-linked	Cytosol	Removal of N-linked glycan from protein
	OST (12)	G3M9	ER	Transfer of G3M9 to proteins
	GT	M9-M7	ER	Transfer of glucose, quality control
	GlcNAc1-phospho-transferase	M8-M7	Cis-Golgi	Targeting to lysosomes

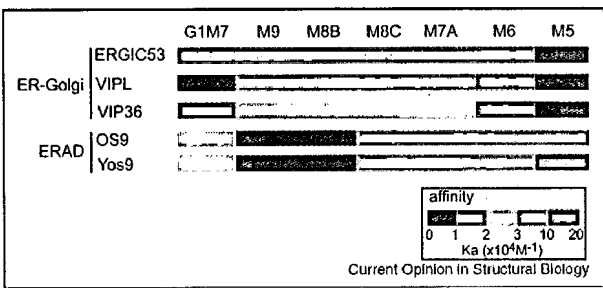
Current Opinion in Structural Biology

Mammalian N-linked glycan-binding proteins of the early secretory pathway: lectins and enzymes. (1) Note that the only proteins in this table with known *S. cerevisiae* homologues are calnexin (although the protein is quite different), OS9, EDEM1, ERManI, GI, GII, PNGase and OST. (2) Specificities refer to naturally occurring ligands on glycoproteins, some synthetic ligands may also be bound, for example, G1M5 to calnexin. (3) Calmeglin is a calnexin homologue expressed exclusively in testis. (4) Intracellular mannan-binding protein. (5) The function and binding specificity of ERGL, (ERGIC53-like protein) have not been characterized yet. (6) *S. cerevisiae* Yos9 has its highest affinity for M5. (7) OS9 is possibly localized in the ERQC like its associated proteins HRD1 and Derlin1 ([55*] and our unpublished results). (8) Fbs1 and Fbs2 are homologous proteins, Fbs1 being neuron-specific and Fbs2 ubiquitous. (9) EDEMs 1-3 are indicated as both lectins and glycosidases as they promote mannose trimming in cells *in vivo* but have not been found yet to have mannosidase activity *in vitro*. (10) Specificity has not been determined yet, but binding is to glycoproteins in the ER, in the high-mannose state. (11) Overexpression of these mannosidases in cells *in vivo* can accelerate ERAD. (12) Oligosaccharyltransferase.

an essential role of ER mannosidase I (ERManI) [23**]. Besides ERManI there is an apparent redundancy of α 1,2 mannosidases and mannosidase-like proteins in mammalian cells, namely the EDEMs 1-3 (described later) and the Golgi α 1,2 mannosidases, Golgi Man IA, IB and IC, all members of the glycosyl-hydrolase 47 family [1,2]. Nevertheless, ERManI may be able to act by itself in trimming M9 to M6-5. Although at low concentrations ERManI removes exclusively mannose-b, at high concentrations it

can trim all α 1,2 mannoses *in vitro* [28] or by overexpression *in vivo* [29]. ERManI is concentrated in the ERQC (Figure 4 and [23**]), a compartment that will be described later, creating a high local concentration capable of removing all four α 1,2 mannoses. Knockdown of ERManI blocks the production of M6-5, even under conditions where cleavage to M8B is not much affected [23**]. Overexpression of Golgi Man IA, IB or IC accelerated ERAD of mutant α 1-antitrypsin, a glycoprotein

Figure 3



Affinities of lectins involved in ER-Golgi transport or in ERAD. The figure illustrates the complementary affinities determined *in vitro* by frontal affinity chromatography of ERGIC53, VIP36 and VIPL, which transport glycoproteins to the Golgi, compared to OS9, which targets glycoproteins to ERAD. Mammalian OS9 has relatively high affinity for M8-5, and even for glucosylated glycans, as long as mannose-c has been removed, but it does not bind M9 or M8B [26**]. Yeast Yos9 has its highest affinity for M5 [21**]. VIP36 and VIPL have their highest affinities for M9, M8 and M7, much lower for M6 and do not bind M5 [27**] and thus they will avoid fully trimmed (misfolded) molecules in the ER and release fully trimmed molecules (processed by Golgi mannosidases) in the Golgi. ERGIC53 shows a rather low affinity and broad specificity but nevertheless it does not bind M5 [27**].

that may exit to the Golgi and be retrieved [30]. However, the ER level of these enzymes may also be increased by their overexpression, leading to glycoprotein trimming and targeting to ERAD even if this is not the normal role of the Golgi mannosidases. Nevertheless, it is possible that these enzymes normally cycle between the ER and the Golgi or that, as in yeast [31], certain ERAD substrates might cycle through the Golgi before delivery to ERAD also in mammalian cells. The ER mannosidase-like proteins EDEMs 1-3 also participate but it is not yet clear whether as mannosidases or as cofactors as discussed below.

EDEMs

EDEM1 was initially thought to be the receptor lectin after the mannose trimming step, but recent data suggest that it participates in the trimming process and that it interacts with its substrate even before the trimming ([32] and our unpublished data). The EDEMs (1-3 in mammalian cells and Htm1 in yeast) may act as cofactors, aiding the activity of ERManI, or may be true mannosidases as was suggested recently [20**,21**,32]. Their mannosidase activity has not been detected yet *in vitro*, except in the case of an EDEM-like protein of *T. cruzi* [25*]. EDEM1 homologues in the yeasts *S. cerevisiae* and *S. pombe* do not trim mannose residues in mutants lacking the ERManI homologue (Mns1) [20**,33], which could mean that initial trimming by ERManI is required or that EDEM1 is a cofactor of ERManI and not a mannosidase *per se*. The complete mannosidase-like domain of EDEM1 cannot even partially replace the whole protein

for mannose trimming in yeast [20**]. On the other hand overexpression of a point mutant that by homology with ERManI should affect the mannosidase activity of EDEM1, could not accelerate mannose trimming of ERAD substrates [20**,32]. However, it could be that slow dissociation of this mutant, interferes with ERManI activity. Acting as a mannosidase or as a cofactor of ERManI, it was shown in yeast that EDEM1 is required for trimming of mannose-c (Figure 1) [20**,21**]. An ERAD substrate carrying M5-2 or M7A (Figure 1) is still degraded in cells lacking a functional EDEM1 [20**,21**]. This suggests that although trimming of mannose residues a and d may still be required, as discussed above, this trimming can be achieved without EDEM1 intervention.

EDEM1 is usually kept at a low level by an autophagy-like degradation process [34]. ERManI is also subjected to lysosomal degradation, perhaps through a similar mechanism. It was suggested that EDEM1, upregulated during the unfolded protein response (UPR) inhibits the degradation of ERManI (unaffected by the UPR [23**]), indirectly accelerating ERAD [35].

At its normal low levels, EDEM1 is located in vesicles [36], and at high levels it is compartmentalized [34] in the ERQC (our unpublished results). Therefore, possible substrate delivery by EDEM1 to ERManI, would be accelerated upon UPR and EDEM1 overexpression.

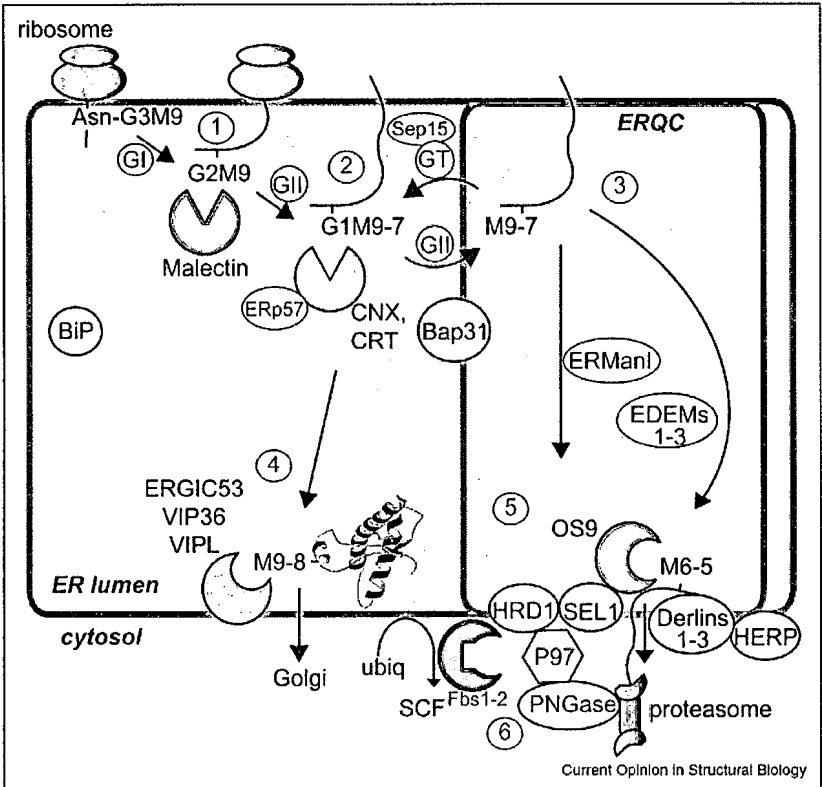
Interestingly, EDEM1 forms a complex with ERdj5, a disulfide reductase required for ERAD [37], possibly replaced by PDI in yeast [38].

OS9 and other lectins in ERAD

At initial quality control stages a novel lectin called mal-ectin binds G2M9 and might modulate the activity of GII [39]. At the other end of the quality control process is OS9. Yeast Yos9 and its mammalian homologues OS9.1, OS9.2 and two forms of XTP3-B interact with the membrane-bound ubiquitination machinery and are postulated ERAD substrate receptors before the ubiquitination and retrotranslocation step, a role once suggested for EDEM1 (reviewed in [40]).

Yos 9 and its mammalian homologues are upregulated by the UPR and are able to bind trimmed sugar chains as discussed above [21**,26**] (Figure 3). Their high glycan specificity suggests their role in a decision-making checkpoint, where glycoproteins are dislocated to the cytosol only if most of their α 1,2 mannose residues have been excised. Nevertheless, they can also bind unglycosylated proteins [40-42]. OS-9 and XTP3-B associate with the HRD1-SEL1L membrane-anchored ubiquitin ligase complex (Figure 4) and with BiP as well as with GRP94, these two chaperones being capable of recruiting non-glycoprotein substrates [43*,44*].

Figure 4



Glycoprotein processing in the ER. *S. cerevisiae* homologues are below in parenthesis whenever they exist. A membrane glycoprotein is translocated and N-glycosylated and the terminal glucose is removed by GI (G1), allowing binding to malectin (step 1). The second glucose is removed by GII (GII) and the CNX folding cycle is initiated (step 2). CNX recruits the oxidoreductase ERp57, whereas GT binds to the reductase Sep15. The glycoprotein is shuttled between the peripheral ER and the ERQC with the aid of Bap31 (steps 2–3). Trimming of up to two α 1,2 mannose residues takes place at this stage. Upon successful folding the glycoprotein is released from the CNX cycle and is transported to ER exit sites and from there to the Golgi with the help of ERGIC53, VIP36, VIPL and perhaps other lectins (step 4). If after several cycles the glycoprotein cannot fold properly it is further trimmed, leading to its release from the CNX cycle because of the absence of the acceptor for glucose transfer by GT. Trimming of most or all α 1,2 mannose residues by ERManI (Mns1) with the aid of the EDEMs 1–3 (Htm1) produces M6 and M5 that are recognized by OS9 (Yos9) (step 5). This initiates the ERAD process, where the glycoprotein is ubiquitinated by the OS9-associated membrane ubiquitin ligase complex of HRD1 (Hrd1) and SEL1 (Hrd3) (or other membrane ubiquitin ligases that are not shown) and encounters the membrane Derlins 1–3 (Der1) and the membrane-associated HERP (Usa1). The glycoprotein is retrotranslocated by an unknown mechanism, is recognized by the cytosolic ubiquitin ligase SCF with the lectin components Fbs1 or 2 and is extracted from the membrane with the aid of the hexameric AAA ATPase P97 (Cdc48) that is associated to HRD1 (step 6). The glycoprotein is deglycosylated by the P97-associated PNGase (Png1) and targeted to degradation by the 26S proteasomes (yeast 26S proteasomes). In the figure, enzymes are in blue text (the EDEMs have not been strictly proven to be mannosidases yet). Lectins are highlighted in colour and with a binding pocket depending on their affinity to glucose (triangle), mannose (circle) or N-acetylglucosamine (square). For simplicity, some accessory components have not been included.

Late stages of ERAD—Fbs1/2

The mechanism of retrotranslocation of the misfolded glycoproteins to the cytosolic proteasomes is still unknown, despite the existence of several candidate polytopic membrane proteins, like subunits of the Sec61 forward translocation machinery, the Derlins 1–3 and a growing number of transmembrane ubiquitin ligases, the best studied being HRD1 (Figure 4) (reviewed in [45]). The cytosolic AAA ATPase P97 (cdc48 in yeast) [46] associates with peptide:N-glycanase (PNGase) for glycoprotein deglycosylation after retrotranslocation and before proteasomal degradation [47]. Deglycosylation takes place in some cases

before complete dislocation [48], but it is not absolutely required that it take place before glycoprotein degradation [49*]. Proteasomes and several other chaperones and cytosolic ubiquitin ligases are recruited to the ER membrane for ERAD (reviewed in [50]). Surprisingly, one of these ligase families contains lectin components, Fbs1 and Fbs2, which recognize glycans on retrotranslocated glycoproteins [51]. Fbs2 is ubiquitous and can bind the glycan while HRD1 binds the polypeptide chain on the same glycoprotein substrate [52]. The similar Fbs1 is neuron-specific and binds to the M3 oligosaccharide core [53*]. Fbs1 and 2 have a broad N-glycan specificity [51] (Figure 2),

suggesting a role in disposal of any glycoprotein that has reached the cytoplasm, in contrast to the high specificity and sorting capability of OS9 [21^{••},26^{••}]

The ER-derived quality control compartment and ERAD

Besides allowing a high local concentration of ERManI for extensive mannose trimming as mentioned above, the microtubule-dependent membrane-enclosed ERQC segregates misfolded proteins that could interfere with ER functions. Upon proteasomal inhibition or UPR induction CNX, CRT and UPR components accumulate in the ERQC but not BiP, PDI, GT or ERp57 [17,54,55[•]]. The ERAD machinery concentrates at the ERQC, which would be a staging ground for ERAD [55[•]] (Figure 4). A transmembrane protein, Bap31 associates with CNX and cycles between the peripheral ER and the ERQC [56] aiding in the delivery of substrates [57]. Non-aggregated ERAD substrates are delivered to the ERQC [54], whereas aggregated glycoproteins like mutant α 1-antitrypsin Z are segregated into CNX-depleted inclusion bodies [58] and delivered to autophagy [59]. Glycoproteins that fold properly and pass the quality control checkpoint are delivered to ER exit sites where they encounter ERGIC53 and other lectins involved in trafficking to the Golgi.

ERGIC53 and other lectins in productive folding

ERGIC53 participates in glycoprotein transport from the ER to the Golgi (Figure 4), although few of its specific substrates have so far been identified [60]. ERGIC53 is upregulated upon UPR [61], which suggests another mechanism to evacuate the ER of accumulated proteins, by accelerating their exit to the Golgi. ERGIC53 has a broad specificity, binding all large high-mannose chains including glucosylated ones, but not M5 (Figure 3) [27^{••}]. Two other lectins, VIPL and VIP36 also appear to participate in glycoprotein transport to the Golgi, but their specificity is more restricted, binding with higher affinity M9, M8B, M8C and with much lower affinity glucosylated or more trimmed sugar chains and, similarly to ERGIC53, they do not bind M5 (Figure 3) [27^{••}]. The specificity of these lectins, complementary to that of Yos9, suggests that they avoid glycoproteins that have been extensively trimmed by the quality control machinery.

Another novel intracellular lectin that might also be involved in ER-Golgi traffic is intracellular mannan-binding protein [62].

Conclusions

Recent reports help clarify the mechanism of glycoprotein sorting to productive folding and exit to the Golgi or to ERAD by compartmentalization, differential N-glycan processing and selective lectin interactions

(Figure 4). Lectins associate at early stages, binding to G2M9 (malectin) and G1M9-8 (CNX and CRT). After deglucosylation by GII, proper folding prevents reglucosylation by GT allowing association with VIPL, VIP36 or ERGIC53, which assist in glycoprotein delivery to the Golgi. Mannose trimming to M5 by Golgi mannosidases IA, IB or IC would prevent reassociation to these lectins and the glycoprotein is released for further processing and delivery to its final destination. Improper folding causes delivery to the ERQC, association with EDEM1-3 and trimming to M6 and M5 by ERManI with the help of the EDEMs and possibly of recycled Golgi α 1,2 mannosidases. M6 and M5 would be recognized by OS9 and its homologues and permit glycoprotein delivery to retrotranslocation. Cytosolic recognition of the sugar chains by Fbs1-2 helps in the ubiquitination process and delivery to the proteasomes for degradation. Basic features of the mechanism seem to exist also in *S. cerevisiae*, with fewer intervening proteins and an absence of the CNX cycle, but apparently with a similar extensive processing of ERAD substrate sugar chains. Although the outline of the overall process is being clarified, there are still several important unresolved issues: (1) The mannosidase/s in charge of creating the signal for ERAD, whether ERManI is the only mannosidase involved or is aided by the Golgi mannosidases, and whether the EDEMs act only as cofactors or as mannosidases as well. (2) The organization and the mechanisms of protein delivery to ER subcompartments. (3) The role of newly discovered lectins that are involved in these processes. (4) The mechanism of protein dislocation to the cytosol in ERAD and the nature of the retrotranslocation channel, which still pose a major challenge for future research.

Acknowledgements

I thank Edward Avezov and Ron Benyair for helpful comments. I apologize to authors whose work I was unable to cite owing to space limitations. Research related to this work is supported by grants from the Israel Science Foundation (1229/07), German-Israeli Project Cooperation (DIP), Israel Cancer Association and Kodesz Institute.

References and recommended reading

Papers of particular interest, published within the period of review, have been highlighted as:

- of special interest
- of outstanding interest

1. Hebert DN, Molinari M: **In and out of the ER: protein folding, quality control, degradation, and related human diseases.** *Physiol. Rev.* 2007, **87**:1377-1408.
2. Molinari M: **N-glycan structure dictates extension of protein folding or onset of disposal.** *Nat Chem Biol* 2007, **3**:313-320.
3. Lederkremer GZ, Glickman MH: **A window of opportunity: timing protein degradation by trimming of sugars and ubiquitins.** *Trends Biochem Sci* 2005, **30**:297-303.
4. Caramelo JJ, Parodi AJ: **Getting in and out from calnexin/calreticulin cycles.** *J Biol Chem* 2008, **283**:10221-10225.
5. Brockmeier A, Brockmeier U, Williams DB: **Distinct contributions of the lectin and arm domains of calnexin to its molecular chaperone function.** *J Biol Chem* 2009, **284**:3433-3444.

6. Michalak M, Groenendyk J, Szabo E, Gold LI, Opas M: **Calreticulin, a multi-process calcium-buffering chaperone of the endoplasmic reticulum.** *Biochem J* 2009, **417**:651-666.
7. Sandhu N, Duus K, Jorgensen CS, Hansen PR, Bruun SW, Pedersen LO, Hojrup P, Houen G: **Peptide binding specificity of the chaperone calreticulin.** *Biochim Biophys Acta* 2007, **1774**:701-713.
8. Ireland BS, Brockmeier U, Howe CM, Elliott T, Williams DB: **Lectin-deficient calreticulin retains full functionality as a chaperone for class I histocompatibility molecules.** *Mol Biol Cell* 2008, **19**:2413-2423.

This paper strengthens the notion that calreticulin can act as a chaperone recognizing peptide domains, independently of its activity as a lectin, in this case in cells *in vivo*.

9. Sousa M, Parodi AJ: **The molecular basis for the recognition of misfolded glycoproteins by the UDP-Glc:glycoprotein glucosyltransferase.** *EMBO J* 1995, **14**:4196-4203.
10. Totani K, Ihara Y, Matsuo I, Tsujimoto T, Ito Y: **The recognition motif of the glycoprotein-folding sensor enzyme UDP-Glc: glycoprotein glucosyltransferase.** *Biochemistry* 2009, **48**:2933-2940.

This paper adds to previous knowledge, showing that GT recognizes the M3 core of N-glycans.

11. Bosis E, Nachliel E, Cohen T, Takeda Y, Ito Y, Bar-Nun S, Gutman M: **Endoplasmic reticulum glucosidase II is inhibited by its end products.** *Biochemistry* 2008, **47**:10970-10980.
12. Totani K, Ihara Y, Matsuo I, Ito Y: **Substrate specificity analysis of endoplasmic reticulum glucosidase II using synthetic high mannose-type glycans.** *J Biol Chem* 2006, **281**:31502-31508.
13. Pearce BR, Gabriel L, Wang N, Hebert DN: **A cell-based reglucosylation assay demonstrates the role of GT1 in the quality control of a maturing glycoprotein.** *J Cell Biol* 2008, **181**:309-320.

This paper shows that GT reglucosylates N-glycans in slow-folding or non-native domains in cells *in vivo*.

14. Solda T, Galli C, Kaufman RJ, Molinari M: **Substrate-specific requirements for UGT1-dependent release from calnexin.** *Mol Cell* 2007, **27**:238-249.

This paper shows that unexpectedly GT deletion does not affect CNX association to many glycoproteins, which must complete their folding in one round of CNX binding.

15. Izquierdo L, Atrih A, Rodrigues JA, Jones DC, Ferguson MA: **Trypanosoma brucei UDP-glucose:glycoprotein glucosyltransferase has unusual substrate specificity and protects the parasite from stress.** *Eukaryot Cell* 2009, **8**:230-240.
16. Jessop CE, Tavender TJ, Watkins RH, Chambers JE, Bulleid NJ: **Substrate specificity of the oxidoreductase ERp57 is determined primarily by its interaction with calnexin and calreticulin.** *J Biol Chem* 2009, **284**:2194-2202.
17. Frenkel Z, Shenkman M, Kondratyev M, Lederkremer GZ: **Separate roles and different routing of calnexin and ERp57 in endoplasmic reticulum quality control revealed by interactions with asialoglycoprotein receptor chains.** *Mol Biol Cell* 2004, **15**:2133-2142.
18. Zhang Y, Kozlov G, Pocanschi CL, Brockmeier U, Ireland BS, Maattanen P, Howe C, Elliott T, Gehring K, Williams DB: **ERp57 does not require interactions with calnexin and calreticulin to promote assembly of class I histocompatibility molecules, and it enhances peptide loading independently of its redox activity.** *J Biol Chem* 2009, **284**:10160-10173.
19. Labunskyy VM, Hatfield DL, Gladyshev VN: **The Sep15 protein family: roles in disulfide bond formation and quality control in the endoplasmic reticulum.** *IUBMB Life* 2007, **59**:1-5.
20. Clerc S, Hirsch C, Oggier DM, Deprez P, Jakob C, Sommer T, Aebi M: **Htm1 protein generates the N-glycan signal for glycoprotein degradation in the endoplasmic reticulum.** *J Cell Biol* 2009, **184**:159-172.

Using *S. cerevisiae* mutants, this paper shows that the exposure of an α 1,6 linked mannose residue, through the activity of Htm1 (EDEMI) is essential for ERAD. It modifies the previous notion that removal of mannose-b from M9 to produce M8B is sufficient for targeting of a glycoprotein to ERAD in yeast.

21. Quan EM, Kamiya Y, Kamiya D, Denic V, Weibezahn J, Kato K, Weissman JS: **Defining the glycan destruction signal for endoplasmic reticulum-associated degradation.** *Mol Cell* 2008, **32**:870-877.

This paper describes the specificity of the *S. cerevisiae* lectin Yos9 *in vitro*, which strengthens its proposed role as a receptor for trimmed N-glycans that targets misfolded glycoproteins to ERAD and suggests that in yeast, as in mammalian cells, there is trimming to M5-6 to target a glycoprotein for ERAD. It also shows that exposure of an α 1,6 linked mannose residue, through the activity of Htm1 (EDEMI) is essential for ERAD modifying the previous notion that removal of mannose-b from M9 to produce M8B is sufficient.

22. Ermonval M, Kitzmuller C, Mir AM, Cacan R, Ivessa NE: **N-glycan structure of a short-lived variant of ribophorin I expressed in the MadIA214 glycosylation-defective cell line reveals the role of a mannosidase that is not ER mannosidase I in the process of glycoprotein degradation.** *Glycobiology* 2001, **11**:565-576.
23. Avezov E, Frenkel Z, Ehrlich M, Herscovics A, Lederkremer GZ: **Endoplasmic reticulum (ER) mannosidase I is compartmentalized and required for N-glycan trimming to Man5-6GlcNAc2 in glycoprotein ER-associated degradation.** *Mol Biol Cell* 2008, **19**:216-225.

This paper shows that the activity of ERMAnI is required for trimming of N-glycans to M5-6 as a signal for targeting glycoproteins to ERAD in mammalian cells. It shows that ERMAnI has a relatively high local concentration in the ER quality control compartment that would allow it to perform this extensive trimming.

24. Frenkel Z, Gregory W, Kornfeld S, Lederkremer GZ: **Endoplasmic reticulum-associated degradation of mammalian glycoproteins involves sugar chain trimming to Man6-5GlcNAc2.** *J Biol Chem* 2003, **278**:34119-34124.
25. Banerjee S, Vishwanath P, Cui J, Kelleher DJ, Gilmore R, Robbins PW, Samuelson J: **The evolution of N-glycan-dependent endoplasmic reticulum quality control factors for glycoprotein folding and degradation.** *Proc Natl Acad Sci U S A* 2007, **104**:11676-11681.

This paper is an extensive study of the quality control machinery of several lower eukaryotes, which shows the existence of homologues of some of the enzymes and chaperones involved in quality control in higher eukaryotes, in relation to the length of their N-glycan precursors, much shorter than G3M9, which is the precursor present in most eukaryotes.

26. Hosokawa N, Kamiya Y, Kamiya D, Kato K, Nagata K: **Human OS-9, a lectin required for glycoprotein ERAD, recognizes mannose-trimmed N-glycans.** *J Biol Chem* 2009, **284**:17061-17068.

This paper determines the specificity of human OS9 *in vitro*, which strengthens its proposed role as a receptor for trimmed N-glycans that targets misfolded glycoproteins to ERAD. It shows that removal of mannose-c and exposure of an α 1,6 linked mannose residue is the feature recognized by OS9. It also presents evidence suggesting that OS9 recognizes N-glycans of the substrate and not of SEL1 in discrepancy with Christianson *et al.* [43].

27. Kamiya Y, Kamiya D, Yamamoto K, Nyfeler B, Hauri HP, Kato K: **Molecular basis of sugar recognition by the human L-type lectins ERGIC-53, VIPL, and VIP36.** *J Biol Chem* 2008, **283**:1857-1861.

This paper describes the specificity of human ERGIC-53, VIPL, and VIP36 *in vitro* by frontal affinity chromatography. The high affinity for M9 and M8B and none for M5 is consistent with the involvement of these lectins in ER-Golgi traffic of properly folded glycoproteins.

28. Herscovics A, Romero PA, Tremblay LO: **The specificity of the yeast and human class I ER alpha 1,2-mannosidases involved in ER quality control is not as strict previously reported.** *Glycobiology* 2002, **12**:14G-15G.
29. Hosokawa N, Tremblay LO, You Z, Herscovics A, Wada I, Nagata K: **Enhancement of endoplasmic reticulum (ER) degradation of misfolded Null Hong Kong alpha1-antitrypsin by human ER mannosidase I.** *J Biol Chem* 2003, **278**:26287-26294.
30. Hosokawa N, You Z, Tremblay LO, Nagata K, Herscovics A: **Stimulation of ERAD of misfolded null Hong Kong alpha1-antitrypsin by Golgi alpha1,2-mannosidases.** *Biochem Biophys Res Commun* 2007, **362**:626-632.
31. Haynes CM, Caldwell S, Cooper AA: **An HRD/DER-independent ER quality control mechanism involves Rsp5p-dependent**

- ubiquitination and ER-Golgi transport. *J Cell Biol* 2002, 158:91-101.
32. Olivari S, Molinari M: **Glycoprotein folding and the role of EDEM1, EDEM2 and EDEM3 in degradation of folding-defective glycoproteins.** *FEBS Lett* 2007, 581:3658-3664.
 33. Movsichoff F, Castro OA, Parodi AJ: **Characterization of Schizosaccharomyces pombe ER alpha-mannosidase: a reevaluation of the role of the enzyme on ER-associated degradation.** *Mol Biol Cell* 2005, 16:4714-4724.
 34. Cali T, Galli C, Olivari S, Molinari M: **Segregation and rapid turnover of EDEM1 by an autophagy-like mechanism modulates standard ERAD and folding activities.** *Biochem Biophys Res Commun* 2008, 371:405-410.
 35. Termine DJ, Moremen KW, Sifers RN: **The mammalian UPR boosts glycoprotein ERAD by suppressing the proteolytic downregulation of ER mannosidase I.** *J Cell Sci* 2009, 122:976-984.
 36. Zuber C, Cormier JH, Guhl B, Santimaria R, Hebert DN, Roth J: **EDEM1 reveals a quality control vesicular transport pathway out of the endoplasmic reticulum not involving the COPII exit sites.** *Proc Natl Acad Sci U S A* 2007, 104:4407-4412.
 37. Ushioda R, Hoseki J, Araki K, Jansen G, Thomas DY, Nagata K: **ERdj5 is required as a disulfide reductase for degradation of misfolded proteins in the ER.** *Science* 2008, 321:569-572.
- This paper describes a mechanism that would participate in protein unfolding in the targeting to retrotranslocation and ERAD.
38. Sakoh-Nakatogawa M, Nishikawa SI, Endo T: **Roles of protein disulfide isomerase-mediated disulfide bond formation of yeast Mnl1p in ER-associated degradation.** *J Biol Chem*. 2009, 284:11815-11825.
 39. Schallus T, Jaechk C, Feher K, Palma AS, Liu Y, Simpson JC, Mackeen M, Stier G, Gibson TJ, Feizi T *et al.*: **Malectin: a novel carbohydrate-binding protein of the endoplasmic reticulum and a candidate player in the early steps of protein N-glycosylation.** *Mol Biol Cell* 2008, 19:3404-3414.
 40. Kanehara K, Kawaguchi S, Ng DT: **The EDEM and Yos9p families of lectin-like ERAD factors.** *Semin Cell Dev Biol* 2007, 18:743-750.
 41. Alcock F, Swanton E: **Mammalian OS-9 is upregulated in response to endoplasmic reticulum stress and facilitates ubiquitination of misfolded glycoproteins.** *J Mol Biol* 2009, 385:1032-1042.
 42. Bernasconi R, Pertel T, Luban J, Molinari M: **A dual task for the Xbp1-responsive OS-9 variants in the mammalian endoplasmic reticulum: inhibiting secretion of misfolded protein conformers and enhancing their disposal.** *J Biol Chem* 2008, 283:16446-16454.
 43. Christianson JC, Shaler TA, Tyler RE, Kopito RR: **OS-9 and GRP94 deliver mutant alpha1-antitrypsin to the Hrd1-SEL1L ubiquitin ligase complex for ERAD.** *Nat Cell Biol* 2008, 10:272-282.
- This paper unveils a role for mammalian OS9 variants in ERAD. It suggests that OS9 binds the glycans of SEL1 and not the substrate, in discrepancy with Hosokawa *et al.* [26*].
44. Hosokawa N, Wada I, Nagasawa K, Moriyama T, Okawa K, Nagata K: **Human XTP3-B forms an endoplasmic reticulum quality control scaffold with the HRD1-SEL1L ubiquitin ligase complex and BiP.** *J Biol Chem* 2008, 283:20914-20924.
- This paper unveils a role for mammalian OS9 variants in ERAD showing some discrepancies with the paper of Christianson *et al.* [43].
45. Hirsch C, Gauss R, Horn SC, Neuber O, Sommer T: **The ubiquitylation machinery of the endoplasmic reticulum.** *Nature* 2009, 458:453-460.
 46. Ballar P, Fang S: **Regulation of ER-associated degradation via p97/VCP-interacting motif.** *Biochem Soc Trans* 2008, 36:818-822.
 47. Zhao G, Zhou X, Wang L, Li G, Schindelin H, Lennarz WJ: **Studies on peptide:N-glycanase-p97 interaction suggest that p97 phosphorylation modulates endoplasmic reticulum-associated degradation.** *Proc Natl Acad Sci U S A* 2007, 104:8785-8790.
 48. Baker BM, Tortorella D: **Dislocation of an endoplasmic reticulum membrane glycoprotein involves the formation of partially dislocated ubiquitinated polypeptides.** *J Biol Chem* 2007, 282:26845-26856.
 49. Kario E, Tirosh B, Ploegh HL, Navon A: **N-linked glycosylation does not impair proteasomal degradation but affects class I major histocompatibility complex presentation.** *J Biol Chem* 2008, 283:244-254.
- This paper shows that unexpectedly glycoproteins can be degraded by the proteasomes without prior deglycosylation.
50. Brodsky JL: **The protective and destructive roles played by molecular chaperones during ERAD (endoplasmic-reticulum-associated degradation).** *Biochem J* 2007, 404:353-363.
 51. Yoshida Y: **F-box proteins that contain sugar-binding domains.** *Biosci Biotechnol Biochem* 2007, 71:2623-2631.
 52. Groisman B, Avezov E, Lederkremer GZ: **The E3 ubiquitin ligases HRD1 and SCF^{Fbs2} recognize the protein moiety and sugar chains respectively of an ER-associated degradation substrate.** *Isr J Chem* 2006, 46:189-196.
 53. Mizushima T, Yoshida Y, Kumanomidou T, Hasegawa Y, Suzuki A, Yamane T, Tanaka K: **Structural basis for the selection of glycosylated substrates by SCF^{Fbs1} ubiquitin ligase.** *Proc Natl Acad Sci U S A* 2007, 104:5777-5781.
- This paper gives structural insight into the broad N-glycan specificity of Fbs1.
54. Kamhi-Nesher S, Shenkman M, Tolchinsky S, Fromm SV, Ehrlich R, Lederkremer GZ: **A novel quality control compartment derived from the endoplasmic reticulum.** *Mol Biol Cell* 2001, 12:1711-1723.
 55. Kondratyev M, Avezov E, Shenkman M, Groisman B, Lederkremer GZ: **PERK-dependent compartmentalization of ERAD and unfolded protein response machineries during ER stress.** *Exp Cell Res* 2007, 313:3395-3407.
- This paper shows recruitment of the ERAD machinery and of UPR sensors to the ER quality control compartment segregating them from BiP.
56. Wakana Y, Takai S, Nakajima K, Tani K, Yamamoto A, Watson P, Stephens DJ, Hauri HP, Tagaya M: **Bap31 is an itinerant protein that moves between the peripheral endoplasmic reticulum (ER) and a juxtanuclear compartment related to ER-associated Degradation.** *Mol Biol Cell* 2008, 19:1825-1836.
 57. Wang B, Heath-Engel H, Zhang D, Nguyen N, Thomas DY, Hanrahan JW, Shore GC: **BAP31 interacts with Sec61 translocons and promotes retrotranslocation of CFTRDeltaF508 via the derlin-1 complex.** *Cell* 2008, 133:1080-1092.
 58. Granell S, Baldini G, Mohammad S, Nicolin V, Narducci P, Storrie B, Baldini G: **Sequestration of mutated alpha1-antitrypsin into inclusion bodies is a cell-protective mechanism to maintain endoplasmic reticulum function.** *Mol Biol Cell* 2008, 19:572-586.
 59. Perlmutter DH: **Autophagic disposal of the aggregation-prone protein that causes liver inflammation and carcinogenesis in alpha-1-antitrypsin deficiency.** *Cell Death Differ* 2009, 16:39-45.
 60. Nyfeler B, Reiterer V, Wendeler MW, Stefan E, Zhang B, Michnick SW, Hauri HP: **Identification of ERGIC-53 as an intracellular transport receptor of alpha1-antitrypsin.** *J Cell Biol* 2008, 180:705-712.
 61. Renna M, Caporaso MG, Bonatti S, Kaufman RJ, Remondelli P: **Regulation of ERGIC-53 gene transcription in response to endoplasmic reticulum stress.** *J Biol Chem* 2007, 282:22499-22512.
 62. Nonaka M, Ma BY, Ohtani M, Yamamoto A, Murata M, Totani K, Ito Y, Miwa K, Nogami W, Kawasaki N *et al.*: **Subcellular localization and physiological significance of intracellular mannan-binding protein.** *J Biol Chem* 2007, 282:17908-17920.



Review

Analysis of *N*-linked oligosaccharides: progress towards the characterisation of glycoprotein-linked carbohydrates

Joanne Charlwood, Duncan Bryant, J. Mark Skehel, Patrick Camilleri *

SmithKline Beecham Pharmaceuticals, New Frontiers Science Park (North), Third Avenue, Harlow, Essex CM 19 5AW UK

Abstract

The covalent attachment of carbohydrate to proteins is a very common co- or post-translational event in the biosynthesis of glycoproteins. The type and heterogeneity of these oligosaccharides can affect a range of physico-chemical and biological properties of a glycoprotein. Thus the development of sensitive, reliable and robust analytical methods for carbohydrate analysis is important in the pharmaceutical industry, especially in the recombinant production of experimental and therapeutic glycoproteins. In this report we have reviewed methodology for the in-gel enzymatic release of *N*-linked oligosaccharides from glycoproteins separated by electrophoresis. These oligosaccharides are derivatised by reductive amination using 3-acetamido-6-aminoacridine (AA-Ac), a novel, highly fluorescent probe. A major advantage of this technique is that glycan derivatives are amenable to analysis by an array of chromatographic and mass spectrometric methods, allowing the resolution and characterisation of a wide variety of glycan structures. It is hoped that in due course the methodology described will be applied to proteomics studies, especially in identifying the role of carbohydrate in protein function and disease. © 2001 Elsevier Science B.V. All rights reserved.

Keywords: Proteomics; Glycoprotein; Carbohydrate; 3-(acetamino)-6-aminoacridine; Matrix-assisted laser desorption ionization time of flight mass spectrometry; Two-dimensional SDS-polyacrylamide gel electrophoresis; Peptide *N*-glycanase F

1. Introduction

In recent years a staggering wealth of information has become available publicly to biochemists, largely due to the large-scale genome sequencing projects. There are now over 38 fully sequenced genomes from a range of species, including the genomes of small organisms like *Haemophilus influenzae* and *Escherichia coli*, to the eukaryotes *Caenorhabditis elegans* and *Saccharomyces cerevisiae* [1,2]. With the announcement earlier this year that 90 percent of the human genome has now been sequenced the term 'post genome era' is fast becoming a reality.

The primary aim of proteomic projects is to identify and characterise the proteins expressed in a tissue, cell or organism at a given time in development and under

defined conditions of growth [3,4]. Current strategies for the identification of proteins are based around the separation of the proteome by two-dimensional SDS-polyacrylamide gel electrophoresis (2-DE). Protein spots are then excised and subjected to in-gel enzymatic digestion with endoproteases such as trypsin. Identification of proteins is then achieved by interrogating sequence databases with mass spectrometric data of the peptides. In the main, protein sequence databases such as those of NCBI (National Centre for Biotechnology Information, USA) and SWISSPROT, are composed of sequences determined from the characterisation of cDNAs. For this reason additional information such as protein function, regulation, location and post-translational modifications cannot be predicted directly.

The advantage of fractionating complex protein mixtures in an array such as 2-DE, is that the technique has an unparalleled resolving capability and in particular when narrow pI range isoelectric focusing gels are employed [5]. Modified forms of the same protein can

* Corresponding author. Tel.: +44-1279-622026; fax: +44-1279-627427.

E-mail address: patrick_camilleri@sbphrd.com (P. Camilleri).

be separated into discrete spots on the gel. For instance, glycosylated proteins give rise to a characteristic train of spots with different pI and apparent molecular weight.

Of all protein post-translational modifications currently known, such as phosphorylation, myristoylation, acetylation, formylation, biotinylation, famesylation and glycosylation, the latter is by far the most common, and is found in both eukaryotes and prokaryotes [6,7]. There are three main classes of linkage of the sugar chain. Carbohydrate may be attached through *O*-glycosyl or *N*-glycosyl linkages. *O*-glycosidic bonds are made between *N*-acetylgalactosamine (GalNAc) and the hydroxyl groups of serine and threonine residues, while *N*-linked carbohydrate chains are formed through bonds between *N*-acetylglucosamine (GlcNAc) and asparagine residues. A third mode of attachment is at the carboxy terminus of some membrane proteins as glycosylphosphatidyl inositol or 'GPI' anchors. In this paper we will consider only *N*-linked glycosylation.

There are three classes of *N*-linked carbohydrates, high mannose, complex and hybrid (Fig. 1). They all contain a common core pentasaccharide consisting of two GlcNAc, and three mannoses (Man). Complex glycans contain no other mannose residues apart from those of the common core. They show the largest amount of variation as they can contain GlcNAc,

galactose and sialic acid residues and can also be multi-antennary. High mannose glycans contain only additional α -mannosyl residues whereas hybrid glycans show characteristics of both high mannose and complex glycans. All three classes of *N*-linked glycans are synthesised by the co-translational addition of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ to the polypeptide chain by oligosaccharyl transferase. This carbohydrate structure then undergoes processing as it passes through the endoplasmic reticulum and Golgi to form the final glycan structure. In the first batch of processing the three glucose residues are removed. The mannose residues are then removed one by one to form high mannose glycans. The complex glycans are synthesised by the addition of GlcNAc, galactose and sialic acid to the core glycan.

Even though there is an enormous complexity of final carbohydrate side chains, there is a single consensus recognition sequence for the initiation of *N*-glycosylation. The sequon Asn-Xaa-Ser/Thr- (where Xaa can theoretically be any amino acid, but is infrequently found to be either proline or aspartate [8]) is necessary for *N*-linked glycosylation (Fig. 1), but not always sufficient. Many possible glycosylation sequons are not glycosylated as they do not possess the necessary three-dimensional conformation.

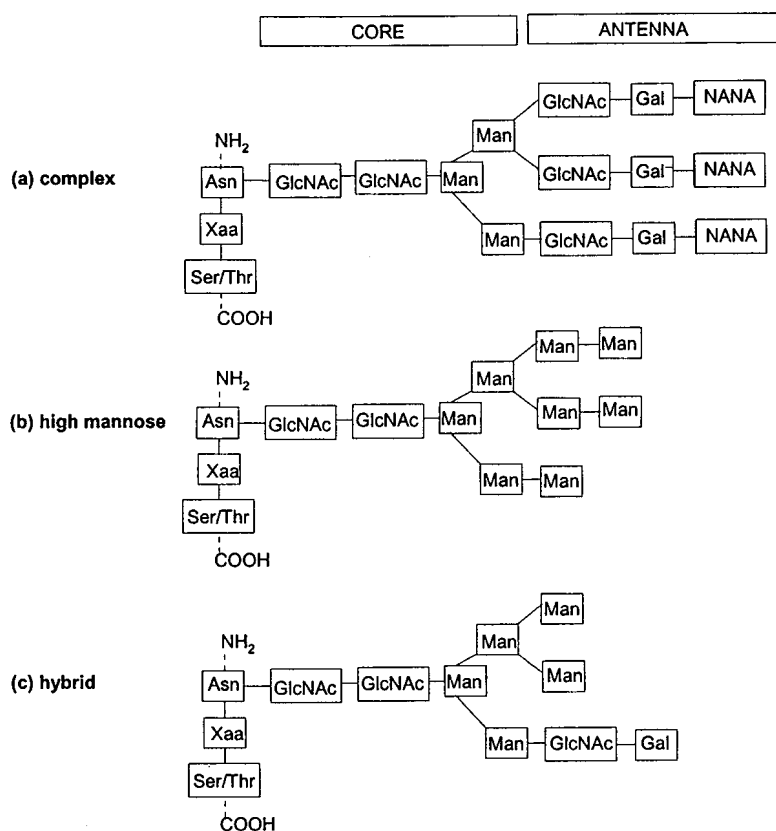


Fig. 1. Three main types of *N*-glycans: (a) tri-antennary complex, (b) high mannose, (c) hybrid.

The carbohydrate moieties of glycoproteins are responsible for diverse biological functions. They are involved in cell adhesion and cell–cell interactions [9,10], while the heterogeneity of the sugar structures encountered are often characteristic of cell types. Indeed, carbohydrates are absolutely required for the correct maturation, function and intracellular sorting of many glycoproteins. For instance, the glycosylation of lysosomal enzymes ensures that they are targeted correctly to lysosomes [11]. Diseases in humans also arise from the failure of glycoproteins to be correctly processed, trafficked and degraded. While certain carbohydrates found on the surface of cells can affect the onset and progression of disease states.

Recombinant DNA techniques are nowadays widely used in the development and manufacture of therapeutic proteins in the treatment of diseases such as cancer, anaemia and hepatitis. The majority of these recombinant therapeutics like erythropoietin (a growth factor) and interferon α -2b (used in the treatment of leukaemia and reversal of acute kidney transplant rejection) are glycoproteins. Carbohydrates linked to these glycoproteins may have a significant influence on their physico-chemical properties, such as their size, conformation, overall charge and solubility. The heterogeneity and the type of carbohydrate in a glycoprotein very often depends on the choice of the host organism and the culture conditions (cell density and age, the presence of glycosidase and glycosyltransferase enzymes, concentration and type of nutrients) used in its biotechnological production, so that it is important to identify the glycosylation pattern that gives optimal clinical benefit. Once this pattern is determined, establishing its reproducibility in the fermentation process of the therapeutic glycoprotein is essential, and is required by the pharmaceutical registration authorities before release for clinical use.

Over the past three to four years, new methods have become available to allow smaller amounts of protein (currently at the fmol level) and glycoprotein oligosaccharides (currently at the pmol level) to be routinely analysed, through major advances in sample preparation, carbohydrate labelling, chromatographic, spectrofluorometric and mass spectrometric techniques. Moreover, the unique resolving power of 2-DE presents us with a powerful electrophoretic purification method, where post-translationally modified proteins can be identified as discrete spots on a gel. To date there are still a limited number of studies where post-translational modifications such as glycosylation have been examined after oligosaccharides have been released from glycoproteins separated by gel electrophoresis, and characterised by a combination of chromatographic and mass spectrometric methods. Some reports [12–16] have used a 1-dimensional SDS-PAGE separation followed by release of the carbohydrate, whereas

there are only a few publications [17–19] involving a 2-DE separation.

This short review is not intended to be a comprehensive study of carbohydrate analysis and as such the examples described have been taken specifically from our current investigations on the analysis of *N*-linked oligosaccharides released from glycoproteins resolved by gel electrophoresis, derivatised with recently developed probes and analysed by chromatographic and mass spectrometric methods.

2. Analysis of *N*-linked glycans

The majority of *N*-linked oligosaccharides do not have a chromophore so that they are difficult to detect at wavelengths above 200 nm. Fortunately the ring structure of reducing glycans is in equilibrium with the 'open' aldehyde form (Fig. 2). The latter functionality allows these molecules to react rapidly with aromatic amines which in turn can be reduced in a 'one-pot' reaction to give stable derivatives. Over the last four years we have developed methodology using two highly fluorescent aromatic probes, namely 2-aminoacridone (2-AMAC) and 3-acetamido-6-aminoacridine (AA-Ac), to derivatise a variety of reducing oligosaccharides [20,21]. The reaction scheme for the reductive amination of GlcNAc (the residue in *N*-glycans reactive to aromatic amines) by AA-Ac is shown in Fig. 2. This derivatisation procedure is quantitative and has the advantage of being carried out in a single reaction vessel, which apart from being practically convenient, reduces possible losses of sample. It can also be miniaturised to microlitre volumes, allowing detection and characterisation of picomole quantities of carbohydrate.

Derivatisation of carbohydrates with 2-AMAC and AA-Ac allows the use of a variety of separation and detection techniques which in combination can provide a great deal of information about the heterogeneity of glycan mixtures. Thus these derivatives are considerably more amenable to chromatographic separation by reverse and normal phase columns. The intense fluorescence of these derivatives at excitation and emission wavelengths in the visible region of the spectrum allows detection at picomole levels. We also found that determination of the molecular weights of these glycan derivatives by mass spectrometric analysis is possible both by matrix assisted laser desorption/ionisation time of flight mass spectrometry (MALDI-Tof-MS) and by electrospray ionization (ESI) mass spectrometry [22–25]. In contrast ionisation of underivatised carbohydrate by MALDI-Tof analysis is not as sensitive as that of these derivatives and requires much higher laser energy. Thus prior reaction of carbohydrates with these aromatic amines significantly enhances the ionisation

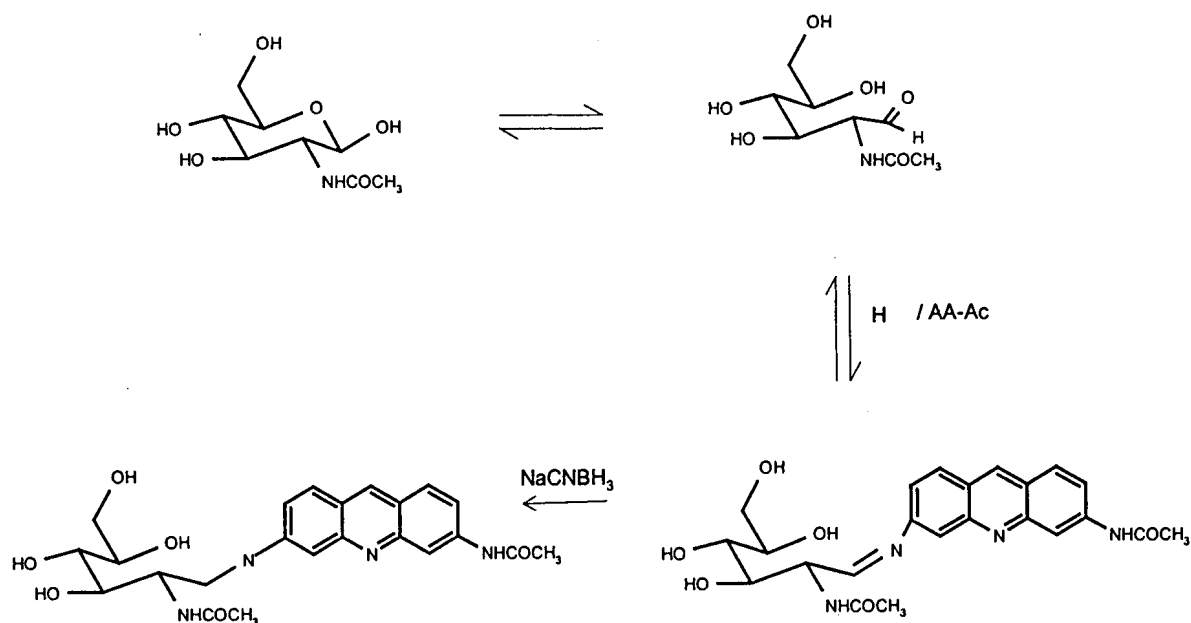


Fig. 2. Reaction scheme for the 'one-pot' derivatisation of reducing glycans with AA-Ac.

properties of the resulting derivatives and so increases the sensitivity of detection.

In a recent comparison of the mass spectrometric properties of a number of probes used in the reductive amination of carbohydrates [26], it was confirmed that the sodium and potassium adducts of 2-AMAC carbohydrate derivatives have excellent ionisation properties when analysis is carried out by MALDI-Tof-MS. However, these adducts are not fragmented easily when analysed by electrospray ionisation mass spectrometry in an MS-MS mode, and therefore are not always suitable to determine the sequence in which glycan residues are attached to one another. To reduce this problem we have specifically developed AA-Ac. This probe besides being approximately three times more fluorescent than 2-AMAC, gives predominantly $[M + H]^+$ carbohydrate adducts in MALDI-Tof-MS analysis. The level at which AA-Ac neutral glycan derivatives are detected by this mass spectrometric method is almost an order of magnitude lower than that possible by 2-AMAC derivatisation. These proton adducts are also amenable to electrospray ionisation MS and MS-MS analysis.

3. Enzymatic release of carbohydrate from glycoproteins separated by gel electrophoresis

The three main strategies for the in situ enzymatic release of *N*-linked glycans from glycoproteins separated by gel electrophoresis are summarised in Fig. 3. All three procedures involve the use of the endo-glycosidase, peptide *N*-glycanase F (PNGase F). In

method 1 the gel is electroblotted onto poly(vinylidene) difluoride (PVDF) membrane [27] and the proteins are visualised by an appropriate staining method such as Coomassie blue or Amido black. Protein spots are then excised from the membrane and the glycans are released from the glycoproteins by incubating with the enzyme. Released glycans are then extracted with water and analysed. This approach was adopted by Packer et al. [17] in an attempt to determine the carbohydrate heterogeneity and so determine what properties influence the migration of glycoforms during 2-DE.

In method 2 the protein bands of interest are excised after staining with Coomassie Blue or Silver. PNGase F is then used to release glycans directly from the proteins immobilised in the gel. Released glycans are extracted for further identification. Procedures such as this were developed by Kuster et al. [15] and employed in the characterisation of the glycans present on recombinant gp120 from HIV. The third method involves digestion of glycoproteins with trypsin, followed by the extraction of the resulting peptides from the gel pieces, essentially following methodology developed by Shevchenko et al. [28]; glycans are finally released from the extracted peptides by digestion in solution with PNGase F.

A comparison of the three methods described has been investigated in our laboratory [19] by releasing glycans from ovomucoid as a model glycoprotein. Oligosaccharides were derivatised with AA-Ac prior to analysis by MALDI-Tof-MS. Data obtained for the derivatised glycans from each of the three release protocols are shown in Fig. 4. The distribution of glycans is essentially the same as that obtained from the digestion of ovomucoid in solution (data not shown) [19].

In comprehensive studies we found that procedure 2, involving the direct in-gel release of glycans using PNGase F (Fig. 4(a)), is the most efficient method. This is not surprising as there are fewer steps involved when the endoglycosidase is used in situ, and this method does not rely on complete extraction of glycopeptides from the gel. In the case where PNGaseF is applied after in situ tryptic digestion (method 3, Fig. 4(b)), mass spectral analysis is further complicated by the appearance of contaminating peptides in the MALDI spectrum. The yield of glycans released from the PVDF blot (Fig. 4(c)) was the lowest (about 10 times less than by in situ PNGase F), suggesting that this method was the least efficient of the three protocols investigated. A reduced recovery of glycans by this method is to be expected because transfer of glycoproteins from the gel to the PVDF membrane is not quantitative and once immobilised, glycans may not be accessible to the PNGase F.

4. Analysis of glycans released from glycoproteins separated by 1D SDS-PAGE

Using protocol 2 we analysed the carbohydrate released from human IgG and a number of therapeutic antibodies [29]. Before electrophoresis by 1D SDS-PAGE, antibodies were reduced and denatured to cleave the disulphide bonds joining the light and heavy chains. This resulted in the appearance of two main bands (Fig. 5(a)) in the electropherogram. Intermediate bands of lower intensity were due to fragments formed due to incomplete cleavage of two or more disulphide bonds. In IgG, carbohydrate is only present on the heavy chain, that is the band of the highest molecular

weight. Release and analysis of this glycan mixture after AA-Ac derivatisation and desialylation gave a number of signals in the MALDI-Tof-MS, as shown in Fig. 5(b). As expected, oligosaccharides released from IgG are of the complex type and contain a fucose residue attached to GlcNAc at their reducing end. Two lower intensity signals (G1FB and G2FB) related to complex bi-antennary glycans containing another GlcNAc residue bisecting the mannose core were also observed. For the structures of these glycans see Table 1.

Although IgG contains a higher proportion of neutral to acidic glycans, it was not possible to identify the latter species using only MALDI-Tof-MS, especially when these are present in small amounts. Difficulty in the analysis of these acidic glycans requires enzymatic desialylation before MALDI-Tof-MS. In the next part we shall demonstrate that an estimation of the ratio of neutral to acidic oligosaccharides in a complex glycan mixture can be obtained when mass spectrometric measurements are combined with chromatographic analysis of the original and desialylated mixtures. This procedure can still be applied to glycans released from as little as 1–10 µg of glycoprotein separated by gel electrophoresis.

The analysis of carbohydrate released from an experimental fusion glycoprotein resolved on a 1D SDS-PAGE gel is illustrated in Fig. 6. The inset gel in Fig. 6(a) shows the separation of 10 µg of this protein which had been purified after expression in a Chinese hamster ovarian cell line. The protein mixture was separated on a 4–20% gradient Laemmli slab gel [30] and visualised by staining with Coomassie brilliant blue R250. The heavily stained band at ≈43 kDa was excised and

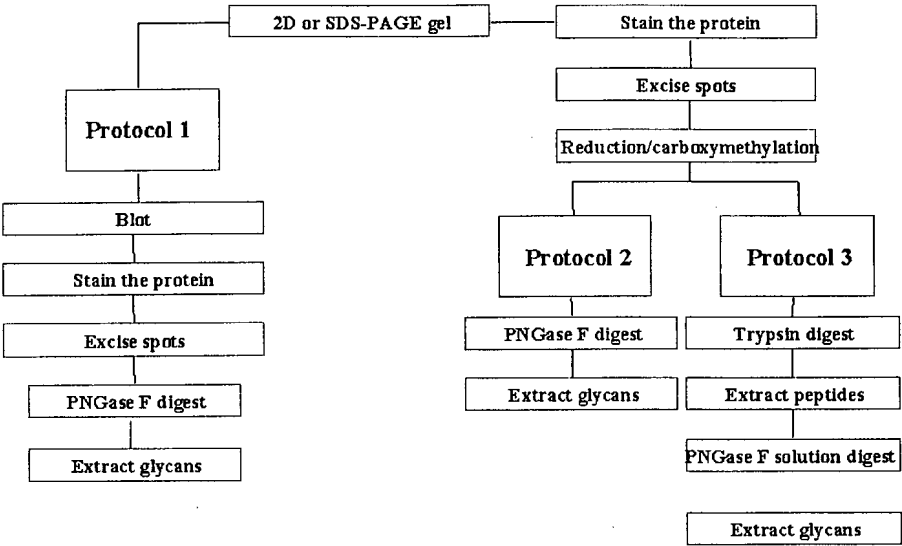


Fig. 3. Flow diagram comparing steps involved in the three methods currently used for the release of glycans following electrophoretic separation.

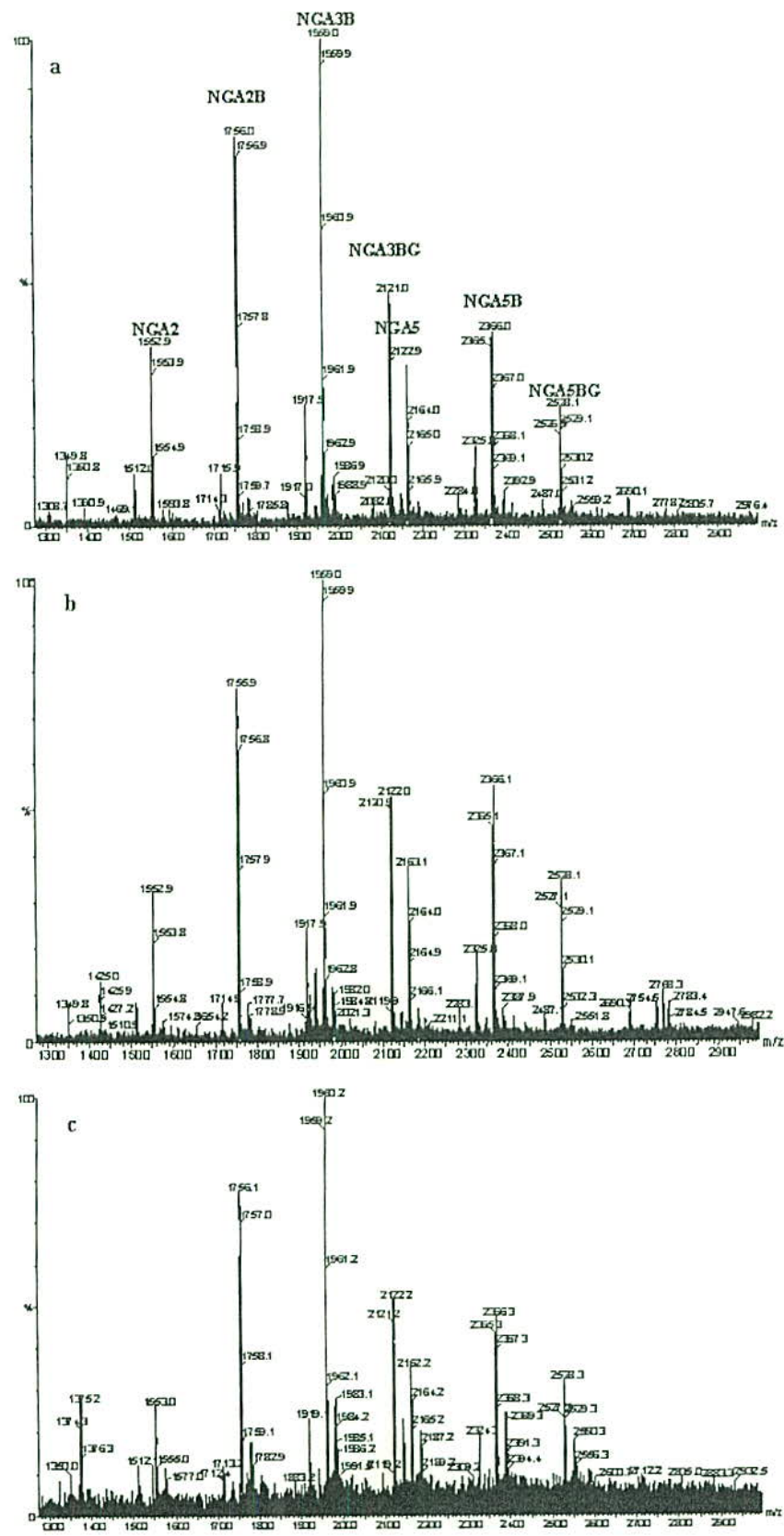


Fig. 4. Glycans released from ovomucoid and analysed by the three different methods. (a) PVDF blot, (b) in situ PNGase F digest, (c) trypsin/PNGase F digest. For explanation of glycan structures see Table 1.

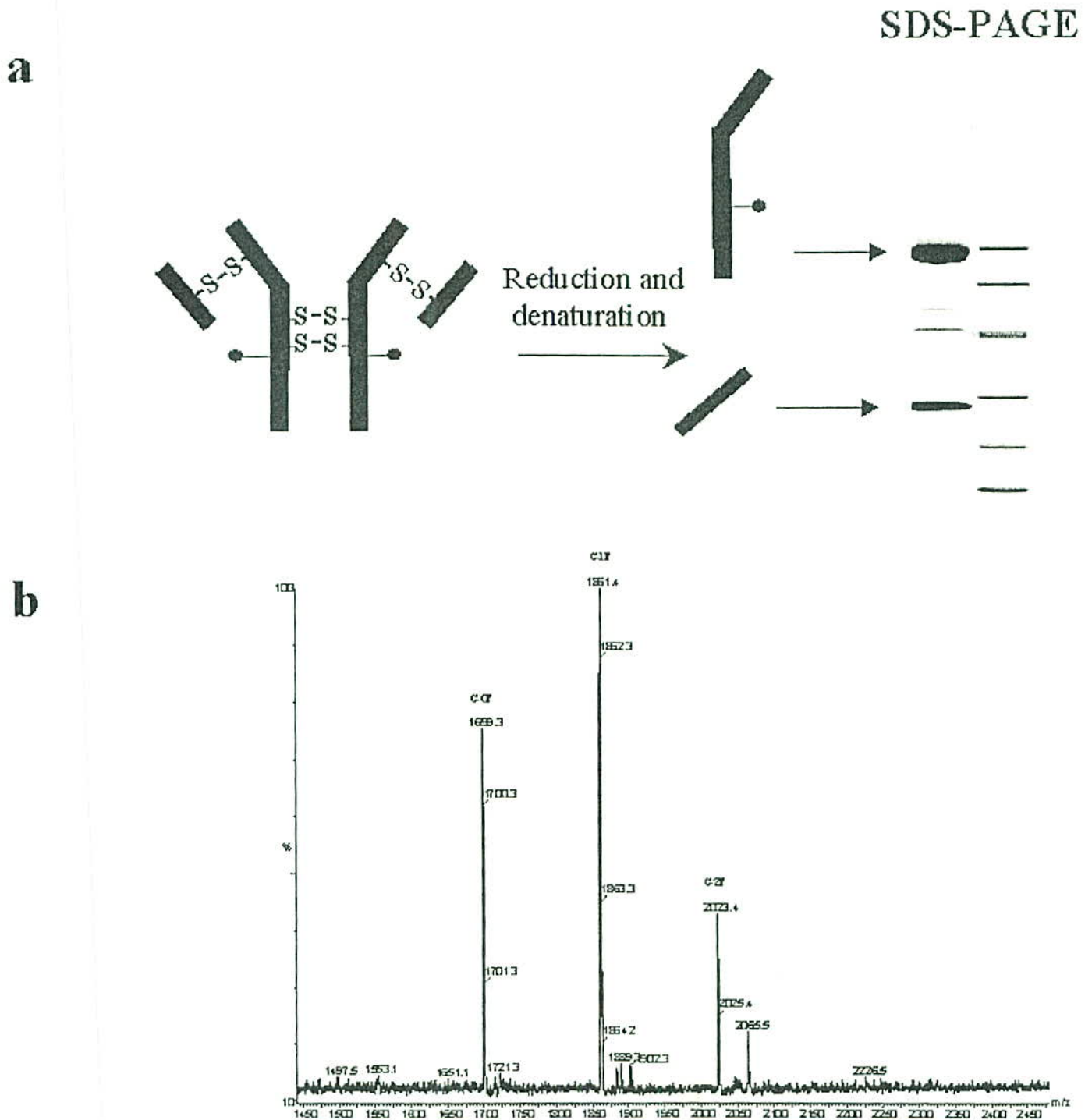


Fig. 5. (a) Separation of IgG on a 4–20% gradient SDS-PAGE gel. The cartoon shows a basic representation of the resolution of the heavy and light chains. (b), MALDI-ToF-MS characterisation of IgG glycans derivatised with AA-Ac after in-gel release from the protein and desialylation.

treated as outlined in Fig. 3, protocol 2. Fig. 6(a) shows the hydrophilic interaction liquid chromatographic (HILC) separation of the released glycans after derivatisation with AA-Ac. In addition to the bi-antennary mono- and di-sialylated species A1F and A2F, a number of multi-antennary sialylated glycans were

present in the chromatogram together with a small amount of the neutral glycan NA2F.

As expected, on treatment with sialidase (Fig. 6(b)), the A1F and A2F glycans lose their sialic acid residues and collapse to a single NA2F peak, while the multi-antennary sialylated glycans give rise to the tri- and

Table 1
Structures of glycans referred to in figures and text

NGA2		NA2	
NGA2F or G0F		NA2F or G2F	
NGA2B		G2FB	
NGA3		NA3	
NGA3B		NA3F	
NGA3BG		NA4F	
NGA5		NA5F	
NGA5B		A1	
NGA5BG		A1F	
G1F		A2	
G1FB		A2F	

The following symbols have been used to identify the various sugar residues: (■), N-acetylglucosamine; (●), mannose; (□), galactose; (○), fructose; (◇), sialic acid. Linkages between sugar residues have been removed for simplicity.

tetra-antennary neutral glycans NA3F and NA4F. A minor component with the longest retention time appeared to be related to a polylactosamine, labelled as NA5F in Fig. 6(b). Assignments of all the other glycans in Fig. 6(b) was made by comparison with retention

times with standard oligosaccharides. This pool of released desialylated glycans was then further characterised by MALDI-Tof-MS (Fig. 6(c)). It is interesting to note that the relative intensities seen in the MALDI-Tof spectrum were similar to those observed in the HILC chromatogram (Fig. 6(b)).

5. Characterisation of glycans released from glycoproteins separated by 2-DE

Two dimensional polyacrylamide gel electrophoresis is now intrinsically associated with proteomic analysis. Fig. 7 illustrates a strategy for the identification and characterisation of both the protein and the post-translational modifications. Fig. 7(a) shows the separation of mouse serum by 2-DE [31]. Mouse serum samples were separated in the first dimension on a pH 3–10 linear IPG isoelectric focusing gel and then in the second dimension on a 4–20% w/v gradient polyacrylamide slab Laemmli gel [30]. The second dimension SDS-PAGE was developed and the proteins visualised with Coomassie Brilliant blue R250. The medium abundant protein indicated in Fig. 7(a) by a circle was chosen for analysis. The loading of sample was such that the protein of interest was present at about 2 µg (≈ 50 pmol). The protein spot was excised and subjected to in-gel enzymatic digestion with trypsin as described elsewhere [27,30]. The peptides generated were desalted on a ZipTip micro reverse phase column [31,32] and analysed by MALDI-Tof-MS Fig. 7(b). The peptide masses obtained were then used to search the NRDB non-redundant protein sequence database (European Bioinformatics Institute, Hinxton, UK), using the database search programme PepSea, part of the Protana Proteomics Software Suite (Protana A/S, Odense, Denmark). For the search 36 peptides were used. The protein α-1-antitrypsin was identified as the first hit, with 12 matching peptides (Fig. 7(c)). The first six proteins identified in the search were all serine protease inhibitors or fragments of the inhibitor. A more detailed evaluation of the data could be carried out using the additional second pass feature of this search programme [33]. Such a search showed that the identified peptides covered 32% of the α-1-antitrypsin protein sequence. The theoretical molecular weight and pI of the protein calculated empirically from the amino acid sequence, of 46 kDa and 5.6, respectively are in general

Fig. 7. Characterisation of a protein from the 2-DE separation of a sample of mouse serum. (a) The circle indicates which protein spots were combined in the analysis. (b) MALDI-Tof mass spectra of the peptide mixture generated by 'in-situ' tryptic digestion. The peptides were first purified on a ZipTip micro reverse phase column before spotting a portion onto the MALDI target as a 1:1 mixture with a solution of the matrix α-cyano-4-hydroxycinnamic acid. (c) A database search using the MALDI-Tof-MS peptide mass map unambiguously identifies the protein with 12 matching peptides. (d) The MALDI-Tof-MS of a AA-Ac derivatised glycan pool released from α1-antitrypsin separated by 2-DE in 7(a). For explanation of glycan structures see Table 1. (e) The total ion chromatogram obtained for the AA-Ac derivatised glycan pool released from α1-antitrypsin. The sialylated glycans were separated by microbore HILC prior to analysis in a Q-Tof hybrid mass spectrometer (Micromass Ltd, Manchester UK).

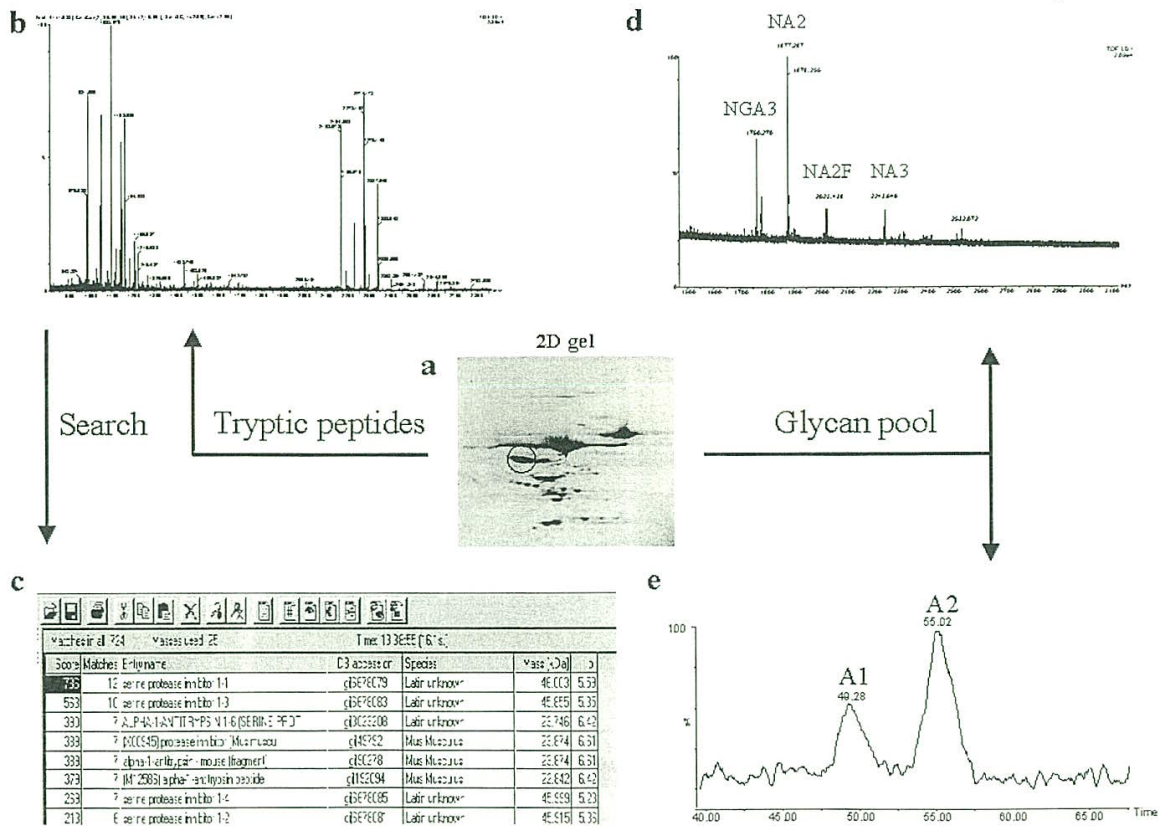


Fig. 7. (Continued)

agreement with the position of the protein on the gel (Fig. 7(a)). However, although the protein can be considered to be unambiguously identified, it is worth noting that of the 36 peptide masses seen in the MALDI-Tof-MS analysis only 30% were matched. The unmatched masses most probably arise from background protein contamination by human keratins. Such contamination could have easily occurred during sample preparation. The MALDI-Tof-MS peptide mass fingerprinting strategy briefly described above, is the preferred method for the identification of proteins in proteomic analysis and has recently been reviewed [33,34].

In order to more fully characterise the glycoproteins, in what could be thought of as a second level of proteome analysis, the train of spots identified as mouse α -1-antitrypsin were combined as shown in Fig. 7(a) and subjected to PNGase F digestion using method 2 to release carbohydrate. For MALDI-Tof-MS analysis (Fig. 7(d)) the released glycan pool was desialylated prior to labelling with AA-Ac. This was necessary as the sialylated glycans show less sensitivity by MALDI-Tof-MS than the neutral glycans. However for analysis by LC-MS the glycan pool released from α -1-antitrypsin was labelled directly with AA-Ac (Fig. 7(e)).

Fig. 7(d) shows that the major oligosaccharides identified from desialylated α 1 -antitrypsin were the bi-antennary glycans NA2 and NA2F, together with small amounts of tri-antennary molecules NGA3, NA3, NA3F and NA3F2 (Table 1). Fig. 7(e) shows the total ion chromatogram obtained from the LC-ESI-MS separation of the sialylated glycan pool of α -1-antitrypsin on a 1 mm GlycoSep N column. The glycans were released from the protein and directly labelled with AA-Ac before analysis. Full details of this methodology can be found in [35]. The major glycans identified were the sialylated complex A1 and A2. It has previously been shown that the analysis of sialylated glycans by MALDI-Tof-MS relies heavily on sample preparation and the acid matrix used [21]. The characterisation of sialylated glycans is affected much more easily using the capillary LC method outlined in this paper, probably because the chromatographic separation overcomes suppression affects of these acidic species.

We have outlined several possible strategies for the analysis of glycans released enzymatically from glycoproteins. In other reports [36,37], intact proteins electroeluted from gels or after electroblotting onto a membrane have been analysed by mass spectrometry. However, these techniques are not generally applicable to proteomic studies, as the amount of protein present in a 2-D gel spot is far too small. Other tools are also being developed to make greater use of the data generated by peptide mass fingerprinting. In a recent publication [38], the programme FindMod was described, for the identification of post-translational modifications.

Here the peptide mass fingerprint obtained after in-gel enzymatic digestion is compared with a theoretical digest of the protein. Any peptides that do not match are further evaluated by calculating the mass differences between the theoretical peptides and seeing if any of these differences is the same as that of a post-translational modification. While this programme is a useful tool for prediction, it relies on the recovery of the majority, if not all of the peptides from the gel piece. After possible identification, the modification still needs to be verified by removal of the moiety chemically or enzymatically followed by mass spectrometric analysis.

6. Conclusions

The derivatisation of *N*-linked glycans with an appropriate probe via reductive amination allows analysis by a range of chromatographic and mass spectrometric techniques. AA-Ac is a highly fluorescent probe, which has proved to offer a number of advantages regarding the characterisation of picomole quantities of complex mixtures of *N*-glycans. Besides being most suitable for the reverse- and normal-phase separation of these molecules, it has proved to be invaluable for identification purposes using highly sensitive mass spectrometric techniques such as MALDI-Tof-MS and ESI-MS. This technology has been routinely applied in our laboratory with a high degree of success to the analysis and characterisation of *N*-glycans released from 1 to 10 μ g of experimental and therapeutic glycoproteins separated by one- and two- dimensional gel electrophoresis. In the future we plan to further miniaturise and automate our derivatisation procedure so that glycans can be released in-situ from glycoproteins in quantities as low as 100 pg (roughly equivalent to 1–5 pmol). This will allow the use of this technology in proteomic studies aimed at relating the glycosylation of proteins to the developmental or disease state of a cell or biological tissue.

References

- [1] Fleischmann RD, Adams MD, White O, Clayton RA, Kirkness EF, Kerlavage AR, et al. *Science* 1995;269:496–512.
- [2] Goffeau A, Barrell BG, Bussey H, Davi RW, Dujon B, Feldmann H, et al. *Science* 1996;274:546–67.
- [3] Wilkins MR, Sanchez JC, Gooley AA, Appel RD, Humphrey-Smith I, Hochstrasser DF, et al. *Biotechnol Genet Eng Rev* 1995;13:19–50.
- [4] Wilkins MR, Pasquali C, Appel RD, Ou K, Golaz O, Sanchez JC, et al. *Biotechnology* 1996;14:61–5.
- [5] Wildgamber R, Harder A, Obermaier C, Boquith G, Weiss W, Fey SJ, et al. *Electrophoresis* 2000;21(13):2610–6.
- [6] Sharon N, Lis H. In: Gabius HJ, Gabius S, editors. *Glycosciences: Status and Perspectives*. Weinheim: Chapman & Hall, 1997:134–62.

- [7] Moens S, Vanderleyden J. *Arch Microbiol* 1997;168:169–75.
- [8] Kornfeld R, Kornfeld S. *Ann Rev Biochem* 1985;54:631–64.
- [9] Lowe JB. In: Fukuda M, Hindsgaul O, editors. *Molecular Biology*. London: Oxford University Press, 1994:163–94.
- [10] Varki A. *Glycobiology* 1993;3:97–130.
- [11] Kornfeld S. *Fed Am Soc Exp Biol J* 1987;1:462–8.
- [12] Anderson JS, Gooch CF, Cooper G, Weitzhandler M. *Glycobiology* 1994;4:459–67.
- [13] Bjellqvist B, Hughes C, Pasquali C, Paquet N, Ravier F, Sanchez J-C, et al. *Electrophoresis* 1993;14:1357–67.
- [14] Oh-eda M, Tominaga E, Nabuchi Y, Matsuura T, Ochi N, Tamura M, et al. *Anal Biochem* 1996;236:369–71.
- [15] Kuster B, Wheeler SF, Hunter AP, Dwek RA, Harvey DJ. *Anal Biochem* 1997;250:82–101.
- [16] Anumula KR, Du P. *Anal Biochem* 1999;275:236–42.
- [17] Packer NH, Lawson MA, Jardine DR, Sanchez J-C, Gooley AA. *Electrophoresis* 1998;19:981–8.
- [18] Packer NH, Wilkins MR, Golaz O, Lawson MA, Gooley AA, Hochstrasser DF, et al. *Biol Technol* 1996;14:66–70.
- [19] Charlwood J, Skehel JM, Camilleri P. *Anal Biochem* 2000;284:49–59.
- [20] Okafo G, Burrow L, Carr SA, Roberts GD, Johnson W, Camilleri P. *Anal Chem* 1996;68:4424–30.
- [21] Charlwood J, Birrell H, Gribble A, Burdes V, Tolson D, Camilleri P. *Anal Chem* 2000;72:1453–61.
- [22] Camilleri P, Tolson D, Birrell H. *Rapid Commun Mass Spectrom* 1998;12:144–8.
- [23] North S, Birrell H, Camilleri P. *Rapid Commun Mass Spectrom* 1998;12:349–56.
- [24] Charlwood J, Langridge J, Tolson D, Birrell H, Camilleri P. *Rapid Commun Mass Spectrom* 1999;13:107–12.
- [25] Charlwood J, Langridge J, Camilleri P. *Rapid Commun Mass Spectrom* 1999;13(14):1522–30.
- [26] Harvey DJ. *Am Soc Mass Spectrom* 2000;11:900–15.
- [27] Pilkington SJ, Skehel JM, Gennis RB, Walker JE. *Biochemistry* 1991;28:866–73.
- [28] Shevchenko A, Jensen ON, Podtelejnikov AV, Sagliocco F, Wilm M, Vorm O, et al. *Proc Natl Sci USA* 1996;93:14440–5.
- [29] Charlwood J, Skehel JM, Camilleri P. *Proteomics* 2001;1:275–84.
- [30] Laemmli UK. *Nature* 1970;227:680–5.
- [31] Skehel JM, Schneider K, Murphy N, Graham A, Benson GM, Cutler P, et al. *Electrophoresis* 2000;21(12):2540–5.
- [32] Pluskal MG. *Nat Biotechnol* 2000;18(1):104–5.
- [33] Jensen ON, Larsen MR, Roepstorff P. *Proteins: Struct Funct Genet* 1998;2:74–89.
- [34] Larsen MR, Roepstorff P. *J Anal Chem* 2000;366:677–90.
- [35] Charlwood J, Birrell H, Bouvier ESP, Langridge J, Camilleri P. *Anal Chem* 2000;72:1469–74.
- [36] Haebel SJC, Anderson SO, Roepstorff P. *Protein Sci* 1995;4:394–404.
- [37] Eckerskorn C, Strupat K, Schleuder D, Hochstrasser D, Sanchez JC, Lottspeich F, et al. *Anal Chem* 1997;69:2888–92.
- [38] Wilkins MR, Gasteiger E, Gooley AA, Herbert BR, Molloy MP, Binz PA, et al. *J Mol Biol* 1999;289:645–57.



Review

Three-dimensional mapping of N-linked oligosaccharides using anion-exchange, hydrophobic and hydrophilic interaction modes of high-performance liquid chromatography

Noriko Takahashi

*GlycoLab, Nakano Vinegar Co. Ltd., 2–6 Nakamura-cho, Handa-City 475, Japan***Abstract**

To determine the structure of N-linked oligosaccharides, a three-dimensional (3-D) sugar mapping technique for pyridylaminated neutral and sialyl oligosaccharides is proposed. The pyridylaminated oligosaccharide mixture is first separated by HPLC on a diethylaminoethyl anion-exchange column and the elution data are placed on the Z-axis. Neutral and mono-, di-, tri- and tetrasialyloligosaccharides are then individually separated on both a hydrophobic octadecylsilylsilica column and a hydrophilic amide-silica column under the same conditions as described previously for neutral oligosaccharides. The validity of the 3-D mapping technique was tested using sialyl pyridylaminated oligosaccharides from human serum glycoproteins.

Contents

1. Introduction	217
2. Outline of procedure, including three-dimensional mapping technique	218
3. Two-dimensional mapping technique for neutral oligosaccharides	218
4. Strategy for the structural determination of an unknown sample on the 2-D map	219
5. Extension to sialyloligosaccharides (3-D mapping technique)	221
6. Chromatograms of oligosaccharide separation leading to a 3-D map	221
7. Conclusion	225
References	225

1. Introduction

Recent developments in elaborate chromatographic techniques allow a new approach to the determination of carbohydrate structure with or without derivatization [1,2]. One of the most promising and popular precolumn derivatization methods is to label the reducing ends of the oligosaccharides with 2-aminopyridine to provide

fluorogenic properties [3,4]. Sometimes liquid chromatography is carried out in two different column-solvent combinations, resulting in “two-dimensional (2-D) mapping” [5,6]. The most extensive mapping method involves the accumulation of HPLC data for 2-pyridylaminated oligosaccharides (PA-oligosaccharides) chromatographed on an octadecylsilica (ODS-silica) column and on an amide-silica column [7]. N-

Linked neutral PA-oligosaccharides are separated by HPLC using the two different columns and their elution positions are expressed in terms of glucose equivalent.

To date more than 220 neutral oligosaccharides have been documented on the 2-D map [8,9]. Many N-linked oligosaccharide structures have been determined by use of this method, e.g., the taste-modifying protein miraculin [10], human IgGs [11], rice α -amylase [12], murine lymphocytes [13], the nicotinic acetylcholine receptor from *Torpedo californica* [14], recombinant erythropoietin [15], human urinary kallidinogenase [16], the sphingolipid activator proteins saposins A, C and D [17] and duck ovomucoid [18].

We have now extended this technique to 3-D mapping [19]. This 3-D mapping method permits the direct analysis of not only neutral oligosaccharides but also of the sialyloligosaccharides [20].

2. Outline of procedure, including three-dimensional mapping technique

As shown in Fig. 1, the asparagine-linked oligosaccharides are enzymatically released after the preparation of the glycopeptides. We always use glycoamidase A from almonds. All N-linked oligosaccharides tested so far have been cleaved from glycopeptides by this enzyme, whether they

were neutral, sialylated, sulfated or phosphorylated. Naturally, all oligosaccharides recorded on the database of the 2-D or 3-D map [7-9,19,20] have been obtained by glycoamidase A digestion. The reducing ends of the oligosaccharide mixture were then derivatized with a fluorescent reagent, 2-aminopyridine. This improved method for PA derivatization works well for sialylated oligosaccharides and for neutral oligosaccharides [4]; also, because of its fluorescent nature, the sensitivity of detection of PA-oligosaccharides is in the picomole range. The PA-oligosaccharide mixtures were separated on three different HPLC columns.

3. Two-dimensional mapping technique for neutral oligosaccharides

We previously proposed the 2-D sugar mapping technique for elucidation of the structures of N-linked oligosaccharides using PA-derivatized neutral oligosaccharides [7]. This is not only a powerful separation technique but it can also establish oligosaccharide structures from the elution positions on two HPLC columns (Fig. 2) [8]. Separation on the amide column (Y-axis) depends mostly on the molecular size of each oligosaccharide. Separation on the ODS column depends on the fine structure of each oligosaccharide. Therefore, oligosaccharides of the same molecular size and which elute at the same position on the amide column can be separated on the ODS column. The elution positions from these two columns provide a unique set of coordinate values for each PA-oligosaccharide, and many sets of such coordinates make up a 2-D map.

The elution time expressed in time or volume units can be variable, depending on the individual column, its age or the batches of buffers used. The introduction of the glucose unit is meant to reduce such variations. For this, we calibrated both the ODS and amide columns with isomaltoligosaccharide mixtures (Fig. 3a and b). The numbers 4, 5, 6, etc., indicate the degree of polymerization of glucose. Next, a sample PA-oligosaccharide was applied to the columns. In this case, the glucose unit of the sample PA-oligosaccharide is measured as 14.7 on the X-axis and 6.7

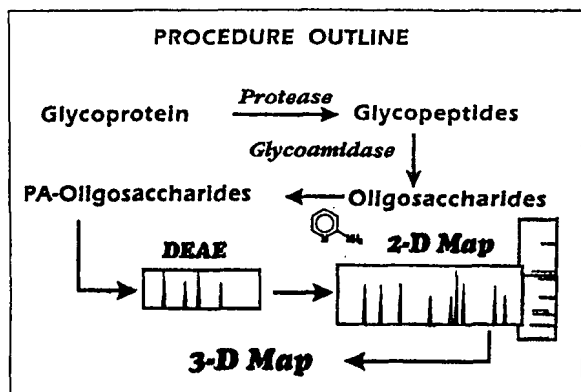


Fig. 1. Scheme for the structural analysis of N-linked neutral and sialyloligosaccharides.

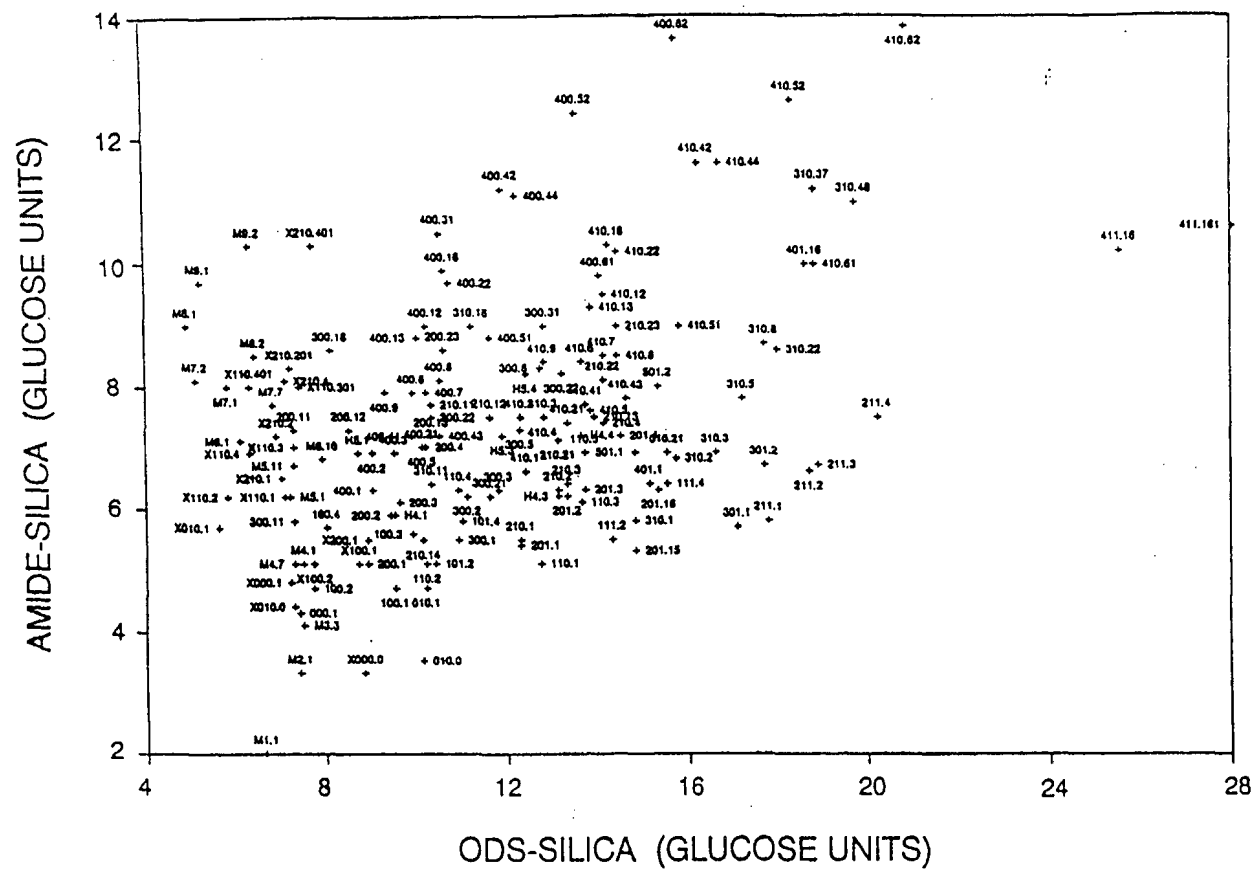


Fig. 2. Two-dimensional sugar map of PA-oligosaccharides expressed as glucose units [8].

on the Y-axis. The next step is to plot the coordinates on the 2-D map (Fig. 3c). The glucose unit, 14.7 on the X-axis and 6.7 on the Y-axis, is plotted on the 2-D map as a coordinate. The structure of an unknown oligosaccharide can be elucidated by comparing its position on the map with the position of the known standards plotted on the 2-D map. Actually, by using a computer search, we choose a couple of candidate PA-oligosaccharides whose coordinates coincide with that of the sample PA-oligosaccharide within an allowable error ($\pm 5\%$).

4. Strategy for the structural determination of an unknown sample on the 2-D map

It is sometimes necessary to go beyond a direct

comparison of the elution positions. For example, if the coordinate 9.5 on the X-axis and 6.0 on the Y-axis of an unknown sample is placed between two known structures 9.6, 6.1 and 9.4, 5.9 (Fig. 4), the strategy for the analysis of the unknown sample is as follows. In many cases, co-injection with a standard PA-oligosaccharide is the most reliable solution. Moreover, a digestion method using several glycosidases is very useful. Although the elution positions of the two candidate oligosaccharides, GlcNAc β 2Man α 6-(Gal β 4GlcNAc β 2Man α 3)Man β 4GlcNAc β 4GlcNAc (9.6, 6.1) and Gal β 4GlcNAc β 2Man α 6-(GlcNAc β 2Man α 3)Man β 4GlcNAc β 4GlcNAc (9.4, 5.9) are very close, after β -N-acetylhexosaminidase digestion, the elution positions of the resultant two oligosaccharides move away as illustrated in Fig. 4. Then, after sequential digestion of β -galactosidase and β -N-acetylhex-

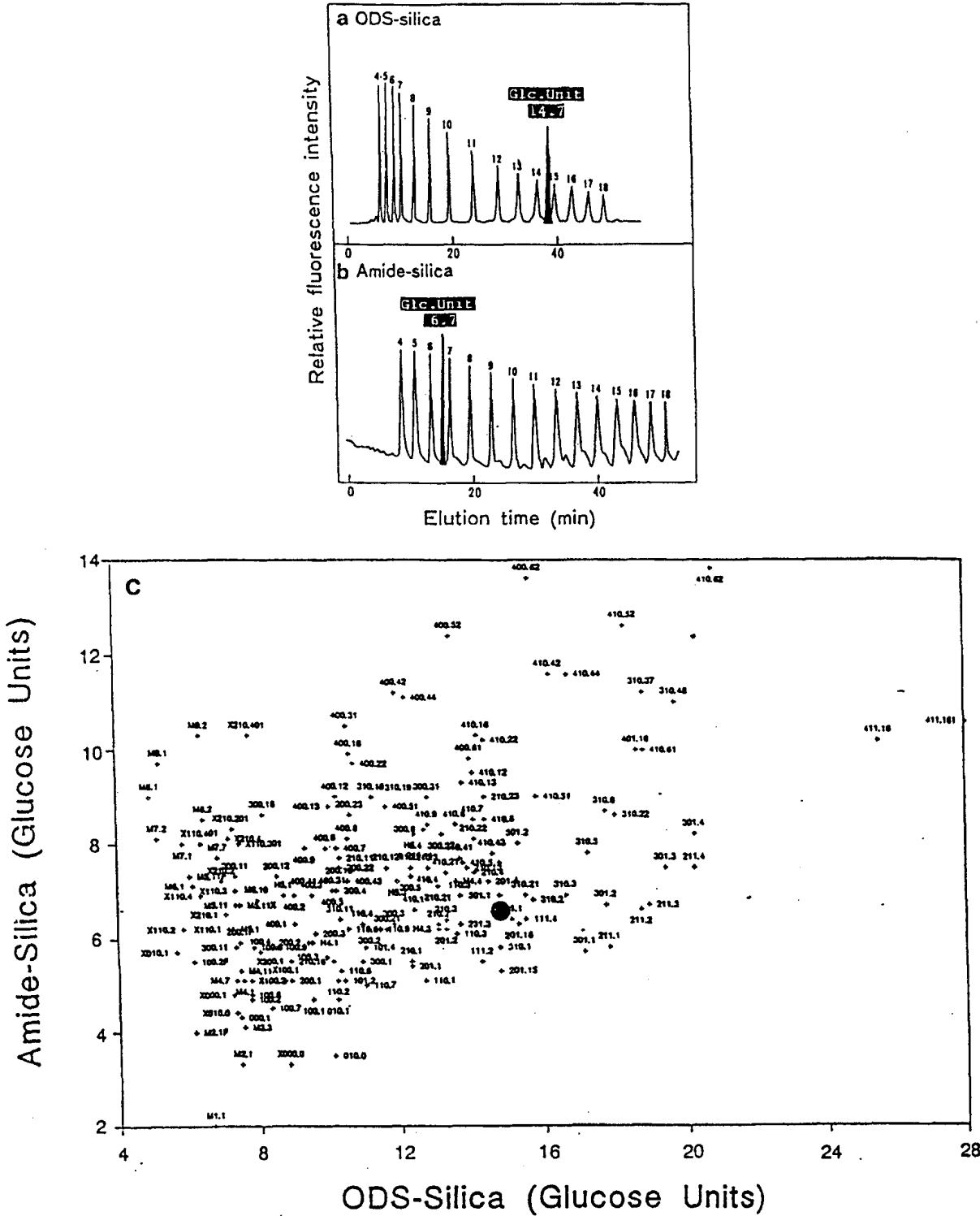


Fig. 3. How to express the elution volume as a glucose unit. Chromatography of PA-derived isomaltooligosaccharide mixture on (a) an ODS-silica and (b) an amide-silica column. Black peaks indicate elution positions of a sample oligosaccharide. (c) The glucose unit 14.7 on the X-axis and 6.7 on the Y-axis is plotted on the 2-D map as a coordinate (black circle).

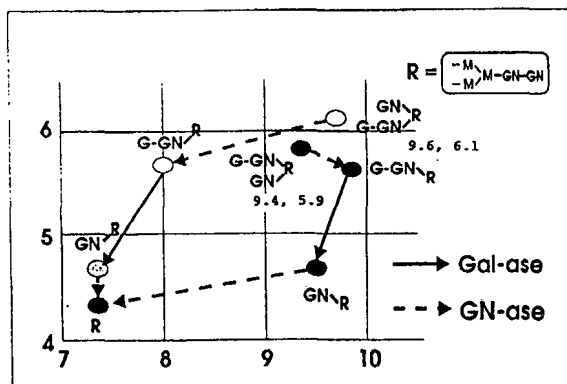


Fig. 4. Strategy for the structural determination of an unknown sample on the 2-D map. Details are described in the text.

osaminidase, the elution position of each standard oligosaccharide finally changes to a common trimannosyl core structure. Therefore, we can establish the originally unknown structure by a series of transformations.

5. Extension to sialyloligosaccharides (3-D mapping technique)

We have more than 220 oligosaccharides on the above-mentioned 2-D map. However, all of these are the neutral oligosaccharides. It is desirable to extend the method to sialyl oligosaccharides.

Fig. 5 shows the steps for constructing a 3-D map for the structural determination of sialyloligosaccharides [19]. First, a PA-oligosaccharide mixture is separated on a diethylaminoethyl (DEAE) column into neutral, mono-, di-, tri- and tetrasialyloligosaccharides according to the sialic acid content of each fraction [20]. Each oligosaccharide separated on the DEAE column is then injected on to an ODS column and the elution volumes are recorded as glucose units, which will form the X-coordinates. Next, each peak from the ODS column is applied to an amide column, and the elution volumes are recorded as glucose units to be plotted on the Y-axis. When all the X- and Y-values are plotted for each sialylation group, a 2-D map is created

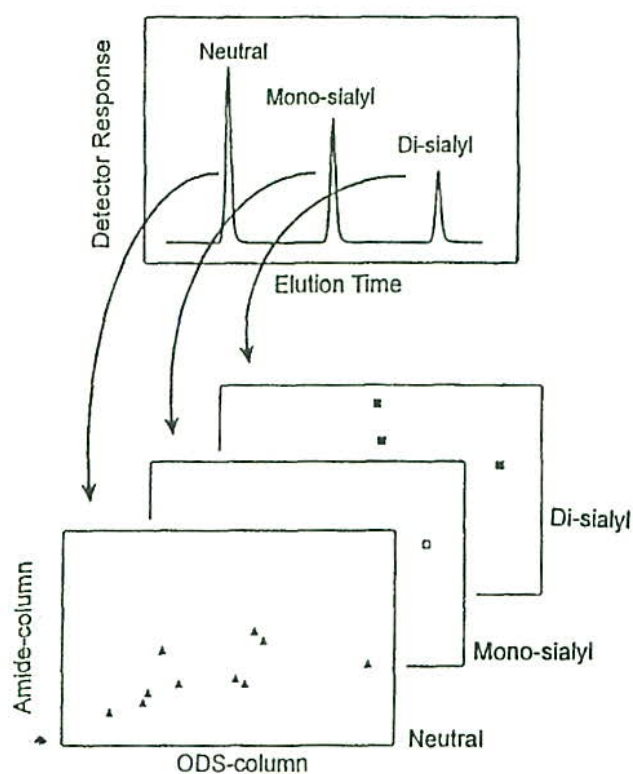
for each of these groups separated by the DEAE column (Fig. 5a).

It is important to point out that for each 2-D map, the HPLC elution conditions used to obtain the X- and Y-coordinates are identical, so the coordinates can be transposed from layer to layer. If we plot all the coordinates from the separated 2-D maps on a single 2-D map, however, the result is slightly confusing because too many coordinates will be crowded on it (Fig. 5b). Three-dimensional mapping is introduced to alleviate such confusion (Fig. 5c). This consists of layers of 2-D maps. Each of these layers lined up on the Z-axis is essentially a 2-D map for a group of oligosaccharides obtained by separation on a DEAE column based on their sialic acid content. The lowest layer is that for the neutral oligosaccharides, and this in fact is the 2-D map we reported previously [7]. The next layer is that for the monosialyloligosaccharides, followed by those for the di-, tri- and tetrasialyloligosaccharides.

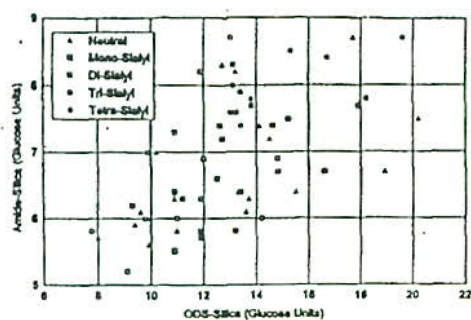
6. Chromatograms of oligosaccharide separation leading to a 3-D map

As an example, we shall briefly describe the structural analysis of N-linked neutral and sialyloligosaccharides derived from human serum glycoproteins [20] (Fig. 6). First, a PA-oligosaccharide mixture is separated on the DEAE column according to the sialic acid content. Neutral and mono-, di- and trisialyloligosaccharides are well separated. The molar ratio of these fractions is about 30, 20, 40 and 10%, respectively. Each of the neutral, mono-, di- and trisialyloligosaccharides is then separated into the many peaks on the ODS column shown in Fig. 6b. The elution volume of each peak is recorded as a glucose unit on the X-axis. Each of the separated oligosaccharides is then applied individually to the amide column, checked for purity and the elution volume of each is recorded as a glucose unit on the Y-axis. In all cases, sample oligosaccharides were found to be homogeneous. The elution volumes from these

a. DEAE-column chromatography



b. Plot on 2-D Map (Overlay)



c. Plot on 3-D Map

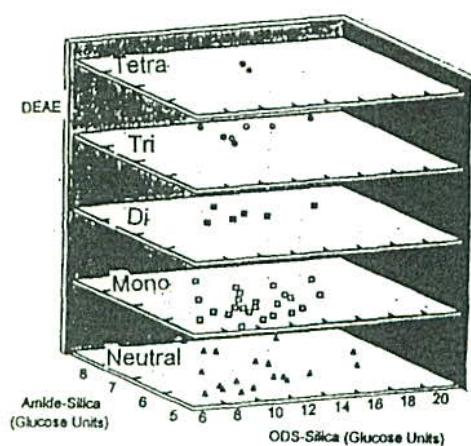


Fig. 5. Transposition of coordinates of PA-oligosaccharides from 2-D to 3-D map [19]. Details are described in the text.

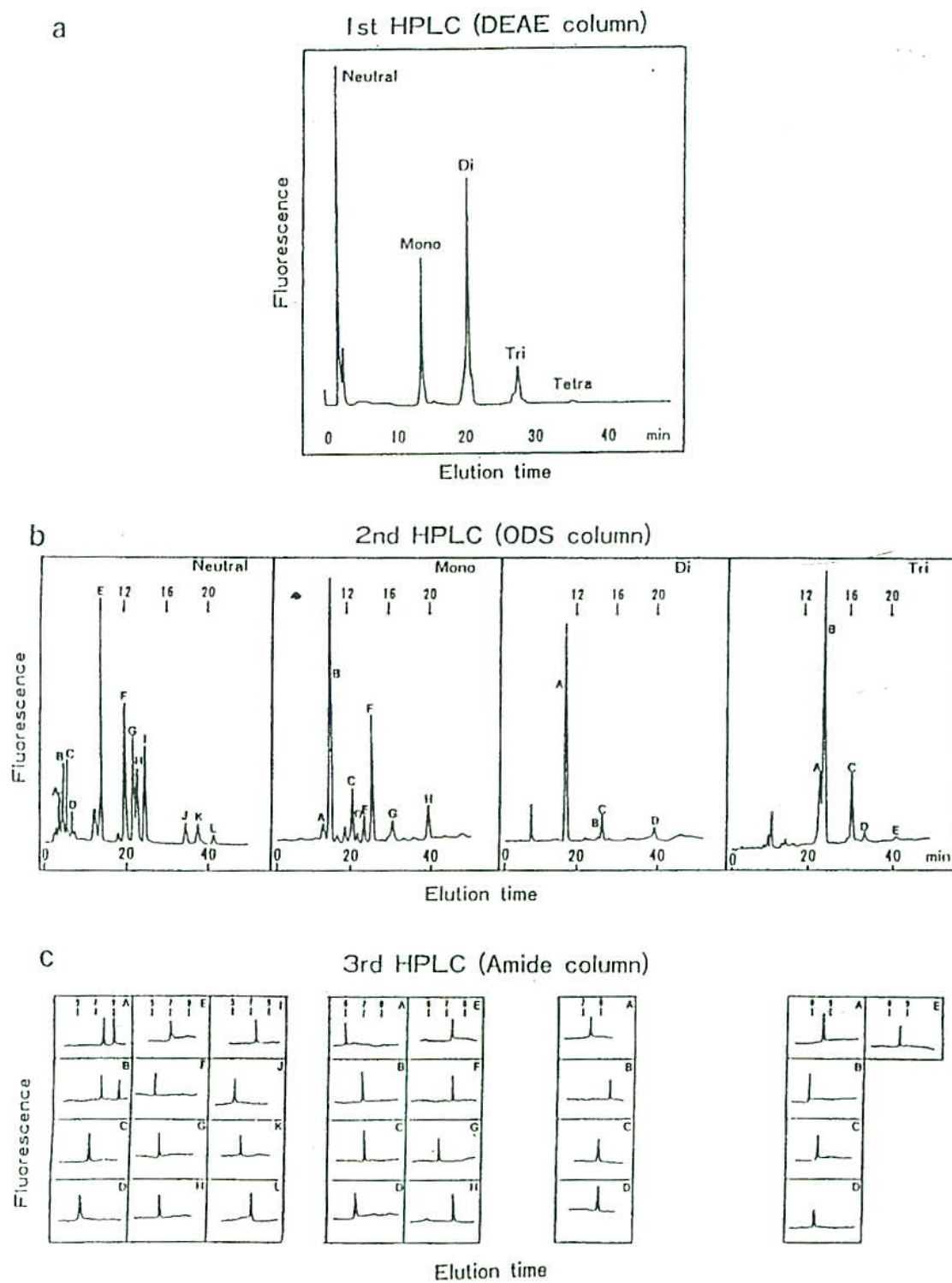


Fig. 6. Elution profile of PA-oligosaccharides from human serum glycoproteins [20]. Details are described in the text.

two columns provide a unique set of coordinates. The coordinates of every oligosaccharide in this sample were coincidental with those of known oligosaccharides, so it was possible to assign a unique structure to each of the unknown oligosaccharides. For example, Table 1 shows the structures of monosialyloligosaccharides from human serum. The main components are bian-

Table 1
Structures and elution positions of PA-derivatized monosialyloligosaccharides obtained from human serum glycoproteins

Code No.	Structure of PA-oligosaccharide	Glc units	
		ODS	Amide
A 1A1-200.3	GlcNAc β 2Man α 6		
	Man β 4GlcNAc β 4GlcNAc	9.3	6.2
	Neu5Ac α 6Gal β 4GlcNAc β 2Man α 3		
B 1A1-200.4	Gal β 4GlcNAc β 2Man α 6		
	Man β 4GlcNAc β 4GlcNAc	9.9	7.0
	Neu5Ac α 6Gal β 4GlcNAc β 2Man α 3		
C 1A2-200.4	Neu5Ac α 6Gal β 4GlcNAc β 2Man α 6		
	Man β 4GlcNAc β 4GlcNAc	12.0	6.9
	Gal β 4GlcNAc β 2Man α 3		
D 1A1-210.3	GlcNAc β 2Man α 6		
	Fuc α 6 Man β 4GlcNAc β 4GlcNAc	12.5	6.6
	Neu5Ac α 6Gal β 4GlcNAc β 2Man α 3		
E 1A1-201.4	Gal β 4GlcNAc β 2Man α 6		
	GlcNAc β 4—Man β 4GlcNAc β 4GlcNAc	12.7	7.2
	Neu5Ac α 6Gal β 4GlcNAc β 2Man α 3		
F 1A1-210.4	Gal β 4GlcNAc β 2Man α 6		
	Fuc α 6 Man β 4GlcNAc β 4GlcNAc	13.2	7.6
	Neu5Ac α 6Gal β 4GlcNAc β 2Man α 3		
G 1A3-210.4	Gal β 4GlcNAc β 2Man α 6		
	Fuc α 6 Man β 4GlcNAc β 4GlcNAc	14.8	6.9
	Neu5Ac α 3Gal β 4GlcNAc β 2Man α 3		
H 1A1-211.4	Gal β 4GlcNAc β 2Man α 6		
	Fuc α 6 GlcNAc β 4—Man β 4GlcNAc β 4GlcNAc	18.3	7.7
	Neu5Ac α 6Gal β 4GlcNAc β 2Man α 3		

tenary with a Neu5Ac α 2 $\rightarrow$ 6 residue on the lower branch with or without a fucose residue on the reducing N-acetylglucosamine. We obtained only one oligosaccharide with a Neu5Ac α 2 $\rightarrow$ 3 residue.

7. Conclusion

Using anion-exchange, hydrophobic and hydrophilic interaction modes of HPLC columns, a mixture of neutral and sialyl oligosaccharides is completely separable. Structural assignment of both neutral and sialyl N-linked oligosaccharides is very easy. Even closely related Neu5Ac α 2 $\rightarrow$ 3- and α 2 $\rightarrow$ 6-containing oligosaccharides can be differentiated on the map. The database on the 3-D map is highly reproducible. Unknown structures whose coordinates do not agree with those of known compounds are easily detected.

References

- [1] S. Honda, *Anal. Biochem.*, 140 (1984) 1.
- [2] K.S. Hicks, *Adv. Carbohydr. Chem. Biochem.*, 46 (1988) 17.
- [3] S. Hase, T. Ikenaka and Y. Matsushima, *Biochem. Biophys. Res. Commun.*, 85 (1978) 257.
- [4] S. Yamamoto, S. Hase, S. Fukuda, O. Sano and T. Ikenaka, *J. Biochem.*, 105 (1989) 547.
- [5] S. Hase, K. Ikenaka, K. Mikoshiba and T. Ikenaka, *J. Chromatogr.*, 434 (1988) 51.
- [6] N.P. Arbatsky, M.D. Martynova, A.O. Zheltova, V.A. Derevitskaya and N.K. Kochetkov, *Carbohydrates*, 187 (1989) 165.
- [7] N. Tomiya, J. Awaya, M. Kurono, S. Endo, Y. Arata and N. Takahashi, *Anal. Biochem.*, 171 (1988) 73.
- [8] N. Takahashi and N. Tomiya, in N. Takahashi and T. Muramatsu (Editors), *Handbook of Endoglycosidases and Glycoamidases*, CRC Press, Boca Raton, FL, 1992, Ch. 8, p. 199.
- [9] N. Takahashi, Y. Wada, J. Awaya, M. Kurono and N. Tomiya, *Anal. Biochem.*, 208 (1993) 96.
- [10] N. Takahashi, H. Hitotsuya, H. Hanzawa, Y. Arata and Y. Kurihara, *J. Biol. Chem.*, 265 (1990) 7793.
- [11] R. Jefferis, H. Mizutani, H. Nakagawa, Y. Kawazoe, Y. Arata and N. Takahashi, *Biochem. J.*, 268 (1990) 529.
- [12] M. Hayashi, A. Tsuru, T. Mitsui, N. Takahashi, H. Hanzawa, Y. Arata and T. Akazawa, *Eur. J. Biochem.*, 191 (1990) 287.
- [13] T. Yoshida, N. Takahashi and I. Nakashima, *Mol. Immunol.*, 28 (1991) 1121.
- [14] H. Shoji, N. Takahashi, M. Ishikawa, I. Shimada, Y. Arata, H. Nomoto and K. Hayashi, *Eur. J. Biochem.*, 207 (1992) 631.
- [15] K.G. Rice, N. Takahashi, Y. Namiki, A.D. Tran, P.J. Lisi and Y.C. Lee, *Anal. Biochem.*, 206 (1992) 278.
- [16] N. Tomiya, J. Awaya, M. Kurono, H. Hanzawa, I. Shimada, Y. Arata, T. Yoshida and N. Takahashi, *J. Biol. Chem.*, 268 (1993) 113.
- [17] K. Ito, N. Takahashi, A. Takahashi, I. Shimada, Y. Arata, J. O'Brien and Y. Kishimoto, *Eur. J. Biochem.*, 215 (1993) 171.
- [18] N. Takahashi, K. Shikami, T. Matsuda, R. Nakamura, I. Shimada and Y. Arata, *Glycoconj. J.*, 10 (1993) 425.
- [19] N. Takahashi, H. Nakagawa, K. Fujikawa, Y. Kawamura and N. Tomiya, *Anal. Biochem.*, 226 (1995) 139.
- [20] H. Nakagawa, Y. Kawamura, K. Kato, I. Shimada, Y. Arata and N. Takahashi, *Anal. Biochem.*, 226 (1995) 130.

Exhibit E

SAP -Homo sapiens (SEQ ID NO:1 of WO 2010/141918)

HTDLSGKVFVFPRESVTDHVNLTITPLEKPLQNFTLCFRAYS DLSRAYS LFSYNTQGRDNELL
VYKERVGEYS LYIGRHKVTSKVIEKF PAPVHICVSWESSSGIAEFWINGTPLVKKGLRQGYF
VEAQPKIVLGQE QDSYGGKFDRSQSFVGEIGDLYMWDSVLPPENILSAYQGTPLPANILDWQ
ALNYEIRGYV IIKPLVWV

SAP - Gallus gallus (SEQ ID NO:2 of WO 2010/141918)

QEDLYRKVFVFREDPSDAYVLLQVQLERPLLNF TVCLRSYTDLT RPHSLFSYATKAQDNEIL
LFKPKPGEYRFYVGGKYVTFRVPENRGWEHVCASWESGSGIAEFWLN GRPWPRKGLQKGYE
VGNEAVVMLGQE QDAYGGGF DVYNSFTGEMADVHLWDAGLSPDKMRSAYLALRLPPAPLAWG
RLRYEAKGDV VVKPRLREALGA

SAP - Bos Taurus (SEQ ID NO:3 of WO 2010/141918)

QTDLRGKVFVFPRESSTDHVT LITKLEKPLKNLTLC LRAYSDLSRGYSLFSYNIH SKDNELL
VFKNIGIGEYS LYIGKTKVTVRATEKF PSPVHICTSWESSTGIAEFWINGKPLV KRGLKQGYA
VGAHPKIVLGQE QDSYGGGF DKNQSF MGEIGDLYMWDSVLSPEEILLVYQGSSSISPTILDW
QALKYEIKGYV IVKPMVWG

SAP - C. migratorius (SEQ ID NO:4 of WO 2010/141918)

QTDLTGKVFVFPRESESDYVKLIPRLEKPLENFTLCFR TYTDLSRPHSLFSYNTKNKDNELL
IYKERMGEYGLYIENVGAIVRGVEEFASPVHFCTSWESSSGIADFWVNGIPWVKKGLKKGYT
VKTQPSIILGQE QDNYGGGF DKSQSFVGEMGDLNMWDSVLTPEEIKSVYEGSWLEPNILDWR
ALNYEMSGYAVIRPRVWH

SAP Alignment (Homo sapiens vs. Gallus gallus)

```
# Aligned_sequences: 2
# 1: Homo_sapiens
# 2: Gallus
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 209
# Identity:      101/209 (48.3%)
# Similarity:    137/209 (65.6%)
# Gaps:          6/209 ( 2.9%)
# Score: 506.0
#
#=====
Homo_sapiens      1 HTDLSGKVFVFPRESVTD-HVNLITPLEKPLQNFTLCFRAYSDDLRSAYSL      49
                  ..||..||||| ||...| :|.|...||:|.|||:|. |:|:|:|. :||
Gallus            1 QEDLYRKVFVF-REDPSDAYVLLQVQLERPLLNFTVCLRSYTDLTRPHSL      49

Homo_sapiens     50 FSYNTQGRDNELLVYKERVGEYSLYIGRHKVTSKVIEKFPAPVHICVSWE      99
                  |||. |: :|||:|: :|:|. |||. |:|...| |:|. |..... |:|. |||
Gallus           50 FSYATKAQDNEILLFKPKPGEYRFYVGGKYVTFRVPENRGWEHVCASWE      99

Homo_sapiens    100 SSSGIAEFWINGTPLVKKGLRQGYFVEAQPKIVLGQEQDSYGGKFDRSQS     149
                  |.|||||||:|:|. |...:| |: :|:|. |.....:| |||||:| |||. ||...|
Gallus          100 SGSGIAEFWLNGRPWPRKGLQKGYEVGNEAVVMLGQEQDAYGGGFVYNS     149

Homo_sapiens    150 FVGEIGDLYMWDSVLPENILSAYQGTPLPANILDWQALNVEIRGYVVIK     199
                  |. ||: |: :|:|. |: :|:|. |||...| |...|. |...|. | |:|. |: :|
Gallus          150 FTGEMADVHLWDAGLSPDKMRSAYLALRLPPAPLAWGRLRYEAKGDVVVK     199

Homo_sapiens    200 PLVWV----      204
                  |.:..
Gallus          200 PRLREALGA      208
```

SAP Alignment (Homo sapiens vs. Bos Taurus)

```
# Aligned_sequences: 2
# 1: Homo_sapiens
# 2: Bos_Taurus
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 205
# Identity:      153/205 (74.6%)
# Similarity:    171/205 (83.4%)
# Gaps:          1/205 ( 0.5%)
# Score: 819.0
#
#
#=====
Homo_sapiens      1 HTDLSGKVFVFPRESVTDHVNLTIPLEKPLQNFTLCFRAYSDSLRSAYSLF      50
                  .|||.|||||||.|||||.|||||.|||:|.|||.|||||||.|||
Bos_Taurus        1 QTDLRGKVFVFPRESSTDHVTLTITKLEKPLKNLTCLRAYSDLSRGYSLF      50
Homo_sapiens      51 SYNTQGRDNELLVYKERVGEYSLYIGRHKVT SKVIEKFPAPVHICVSWES      100
                  |||...:|||||:|...:|||||||:|.|||.|||:|||||.|||
Bos_Taurus        51 SYNIHSDNELLVFKNGIGEYSLYIGKTKVTVRATEKFPSPVHICTSWES      100
Homo_sapiens      101 SSGIAEFWINGTPLVKKGLRQGYFVEAQP KIVLGQE QDSYGGKFDRSQSF      150
                  |:|||||||.|||||:|:|||||.|||.|||||||.|||:|
Bos_Taurus        101 STGIAEFWINGKPLVKRGLKQGYAVGAHPKIVLGQE QDSYGGGF DKNQSF      150
Homo_sapiens      151 VGEIGDLYMWDSVLPPENILSAYQG-TPLPANILDWQALN YEIRGYV IIK      199
                  :|||||||.|||||.|||.||| :...:|||||.|||:||||:|
Bos_Taurus        151 MGEIGDLYMWDSVLSPEEILLVYQGS SSSISPTILDWQALKYEIKGYVIVK      200
Homo_sapiens      200 PLVWV      204
                  |:||.
Bos_Taurus        201 PMVWG      205
```

72

SAP Alignment (Homo sapiens vs. C. migratorius)

```
# Aligned_sequences: 2
# 1: Homo_sapiens
# 2: C_migratorius
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 204
# Identity:      137/204 (67.2%)
# Similarity:    166/204 (81.4%)
# Gaps:          0/204 ( 0.0%)
# Score: 755.0
#
#
#=====
Homo_sapiens      1 HTDLSGKVFVFPRESVTDHVNLTPLEKPLQNFTLCFRAYSDSLRSAYSLF      50
.||||:|||||:|||||.::|:|.|||.||||:|||||.!:||||.:|||
C_migratorius    1 QTDLTGKVFVFPRESESDYVKLIPRLEKPLENFTLCFRTYTDLSRPHSLF      50

Homo_sapiens     51 SYNTQGRDNELLVYKERVGEYSLYIGRHKVT SKVIEKFPAPVHICVSWES     100
||||:..:||||:||||:|.|.||||.....:|:|:|.|.||||
C_migratorius    51 SYNTKNKDNELLIYKERMGEYGLYIENVGAIVRGVEEFASPVHFCTSWES     100

Homo_sapiens     101 SSGIAEFWINGTPLVKKGLRQGYFVEAQPKIVLGQE QDSYGGKFDRSQSF      150
||||:|:|:|.|.||||:|:|.|.|.|.|:||||:|.|.||:||||
C_migratorius    101 SSGIADFVWNGIPWVKKGLKKG YTVKTQPSIILGQE QDNYGGGFDKSQSF      150

Homo_sapiens     151 VGEIGDLYMWDSVLPPENILSAYQGTPLPANILDWQALNYEIRGYVI IKP      200
|||:|.|.|||||.||.|.|.|:|:|.|.||||:||||:|.||:|:|
C_migratorius    151 VGEMGDLNMWDSVLTPEEIKSVYEGSWLEPNILDWRALNYEMSGYAVIRP      200

Homo_sapiens     201 LVWV      204
. ||.
C_migratorius    201 RVWH      204
```

ARTICLES

20. Haslam, C. G. T., Salter, C. J., Stoffel, H. & Wilson, W. E. *Astr. Astrophys. Suppl. Ser.* **47**, 1–143 (1982).
21. Reich, P. & Reich, W. *Astr. Astrophys. Suppl. Ser.* **74**, 7–28 (1988).
22. Davis, R. L., Hodges, H. M., Smoot, G. F., Steinhardt, P. J. & Turner, M. S. *Phys. Rev. Lett.* **69**, 1856–1859 (1992).
23. Salopek, D. S., *Phys. Rev. Lett.* **69**, 3602–3605 (1992).
24. Taylor, A. N. & Rowan-Robinson, M. *Nature* **359**, 396–399 (1992).
25. Maddox, S. J., Elstathiou, G., Sutherland, W. J. & Loveday, J. *Mon. Not. R. astr. Soc.* **242**, 43–47 (1990).

26. Geller, M. J. & Huchra, J. P. *Science* **246**, 897–903 (1990).
27. Sachs, R. K. & Wolfe, A. M. *Astrophys. J.* **147**, 73–90 (1967).
28. Wright, E. L. et al. *Astrophys. J.* **396**, L13–L18 (1992).

ACKNOWLEDGEMENTS. We thank all the staff at NRL and the IAC who have contributed to these results, especially J. Hopkins and R. Hoyland for their technical assistance in the design and operation of the Tenerife instruments. This work has been partly supported by the UK SERC, the EU and the DGICYT Science Programmes. S.H. acknowledges an SERC studentship and a research fellowship at St. John's College, Cambridge University.

Structure of pentameric human serum amyloid P component

Jonas Emsley, Helen E. White, Bernard P. O'Hara*, Glaucius Oliva*, Narayanaswamy Srinivasan, Ian J. Tickle, Tom L. Blundell, Mark B. Pepys† & Steve P. Wood†

Laboratory of Molecular Biology and ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, London WC1E 7HX, UK

† Immunological Medicine Unit, Royal Postgraduate Medical School, Hammersmith Hospital, London W12 0NN, UK

The three-dimensional structure of pentameric human serum amyloid P component at high resolution, the first reported for a pentraxin, reveals that the tertiary fold is remarkably similar to that of the legume lectins. Carboxylate and phosphate compounds bind directly to two calcium ions; interactions with a carboxyethylidene ring are mediated by Asn 59 and Gln 148 ligands of the calcium ions. These X-ray results indicate the probable modes of binding of the biologically important ligands, DNA and amyloid fibrils.

HUMAN serum amyloid P component (SAP) is a decameric plasma glycoprotein composed of identical subunits non-covalently associated in two pentameric rings interacting face-to-face. SAP binds to all forms of amyloid fibril¹ and is universally present in amyloid deposits, including the cerebral amyloid of Alzheimer's disease². Although SAP is not required for amyloid fibrillogenesis *in vitro*, it may protect the fibrils from degradation *in vivo*³. SAP is also the major DNA- and chromatin-binding protein of plasma^{4,5}, and other autologous ligands for SAP include fibronectin, C4-binding protein⁶ and glycosaminoglycans⁷. Binding to all these ligands, and to amyloid fibrils, is absolutely calcium-dependent at physiological pH and ionic strength. In addition, SAP is a calcium-dependent lectin, the best-characterized ligand of which is the 4,6-cyclic pyruvate acetal of β -D-galactose⁸, and this can inhibit all other known calcium-dependent interactions of SAP. Finally, SAP is a normal tissue matrix constituent associated with elastic fibres⁹ and the glomerular basement membrane¹⁰.

Human SAP shows no polymorphism or heterogeneity of its protein or its glycan¹¹ and no occurrence of human SAP deficiency has yet been described, suggesting that the molecule has important functions. Furthermore, SAP and C-reactive protein (CRP), the classical acute-phase protein with which it shares more than 50% sequence identity, belong to the pentraxin family of plasma proteins which have been stably conserved throughout vertebrate evolution^{12,13}. The structure and structure-function relationships of SAP are thus of considerable fundamental interest, in addition to their clinical importance, which arises from

the new information provided by use of radio-labelled SAP as a specific quantitative *in vivo* tracer for amyloid deposits¹⁴.

Recently, we reported the crystallization of SAP in conditions that induce a reversible dissociation of the decamer into a pentameric form¹⁵. We now report the X-ray analysis of the crystals at 2 Å resolution, defining the complete three-dimensional structure of the pentamer. This human plasma protein has a tertiary fold which resembles that of the legume lectins concanavalin A and pea lectin. There are two calcium sites and these are shown to be involved in binding of carbohydrate and the synthetic ligand phosphoethanolamine. However, the calcium sites in SAP and the calcium and manganese sites of concanavalin A differ in their relationship to the common topology. The SAP pentamer shows nearly perfect 5-fold symmetry and sequence comparisons suggest that very similar interactions are retained in CRP pentamers.

Three-dimensional structure

Serum amyloid P component was isolated at greater than 99% purity from pooled human ascites and pleural effusion fluids¹⁶. The crystals obtained at pH 5.5 in a medium containing calcium and acetate ions as described previously¹⁵ were of space group P2₁ and cell dimensions $a=68.9$ Å, $b=99.3$ Å, $c=96.7$ Å and $\beta=95.9^\circ$. The initial model of pentameric SAP was based on X-ray analysis at 2.8 Å resolution using the multiple isomorphous replacement (MIR) technique followed by solvent flattening¹⁷ and 5-fold molecular averaging¹⁸. The refinement statistics are shown in Table 1, the legend of which gives further details of the X-ray analysis. The final model was refined at 2.0 Å resolution to an agreement (R) value of 0.18 to give clear electron density for all residues of the five subunits (see Fig. 1a).

The SAP pentamer consists of five subunits of 204 amino-acid residues, each with a closely similar three-dimensional structure constructed from antiparallel β -strands (A–O) arranged in two sheets (Fig. 2a). The tertiary fold can be envisaged as a jellyroll

* Present addresses: Department of Biochemistry and Molecular Biology, University College, London WC1E 6BT, UK (B.P.O.H.); Instituto de Física e Química de São Carlos, Departamento de Física e Ciência dos Materiais, Universidade de São Paulo, 3560-250 São Carlos SP, Brazil (G.O.).

† To whom correspondence should be addressed.

TABLE 1 Structure determination of serum amyloid P component

Derivative	K ₂ PtI ₆	Na ₃ PO ₄ (WO ₃) ₁₂	Th(NO ₃) ₄	Th(NO ₃) ₄	Ce(SO ₄) ₂	K ₂ AuCl ₄	UO ₂ SO ₄	AgNO ₃	Na ₂ WO ₄
Concentration (mM)	5.0	5.0	4.5	4.5	17.0	2.5	25.0	5.0	11.0
Soaking time (h) at 20 °C	24.0	60.0	28.0	28.0	48.0	1.0	24.0	24.0	19.0 (4 °C)
Collection device	F	F	F	F	F	F	X	I	D
Resolution (Å)	3.5	3.5	4.1	2.8	3.0	3.5	3.2	3.4	6.0
Merging R factor	0.084	0.069	0.062	0.049	0.101	0.110	0.045	0.118	0.075
Isomorphous R factor	0.317	0.237	0.203	0.278	0.222	0.310	0.251	0.165	0.225
% Complete	88.3	80.3	91.7	85.8	76.7	71.1	81.0	93.4	97.1
Number of sites	28	21	17	14	27	22	22	26	—
Phasing power	1.00	1.15	1.43	2.02 (4.1–2.8 Å)	1.02	1.55	1.29	1.04	—

Heavy-atom derivatives used for multiple isomorphous replacement and data collection statistics are shown. Native data were collected from one crystal on a Hilger-Watts 290 4-circle diffractometer with $R_{\text{sym}} = 4.5\%$ for 4,112 independent reflections to 5.6 Å. High-resolution native data were also collected on film at the Synchrotron Radiation Source at Daresbury Laboratory (wavelength 1.468 Å) using three crystals ($R_{\text{sym}} = 7.7\%$ for 78,951 unique reflections to 2.02 Å). The final merged data set comprised low-resolution data from the diffractometer and high-resolution data from the synchrotron (93.6% complete to 2 Å). The major heavy-atom sites were determined from inspection of difference Patterson functions and cross-phase difference Fourier transforms. A multiple isomorphous replacement (MIR) map was calculated, solvent-flattened and averaged. Phases from both maps were used in cross-phase difference Fourier transforms to determine those minor sites related by the 5-fold axis to the major sites. All sites were initially refined with the VECREF program (I.J.T.) to eliminate spurious sites and then with PHARE. A final MIR map was calculated to 2.8 Å and the phasing analysis gave a figure of merit of 0.61. This map was solvent flattened¹⁷, and phases were calculated and recombined with the MIR phases to give a combined figure of merit of 0.85. The solvent-flattened map was then averaged using the PSAVER program (I.J.T.) and the envelope determined in the solvent flattening. The polypeptide of the β -sheet and helix were well defined. Derivative data were also scaled to native data collected from three crystals on a FAST area detector mounted on a CuK α microfocus tube (800W) ($R_{\text{sym}} = 10.0\%$ for 30,996 independent reflections to 2.8 Å). Heavy-atom sites were determined independently from those in Table 1 and were refined using MLPHARE. The resulting electron density map was solvent-flattened and averaged. The map was similar to that used for the refinement. Problems in obtaining an interpretable map arose because not all substituted sites related by the non-crystallographic symmetry were present on early cross-phase difference Fourier transforms because of poor phasing from the derivatives available at the time. Two of the three derivatives with the highest phasing power were the last to be collected. These extra derivatives allowed minor sites related to the major sites by the 5-fold axis to be determined in the early derivatives, as well as multiple occupancy in some sites previously thought to be singly occupied. For example, the dominant thorium nitrate derivative formed a complex with 11 thorium atoms comprising two octahedra, each with edges of ~4 Å and sharing an apex. The direction of the 5-fold axis was determined from a self-rotation function¹⁵, and the approximate position of the molecule on this axis was found by positioning a pentagonal prism in a low-resolution MIR map. A six-dimensional search (three rotations and three translations) for a best correlation in the map for a given rotational operation was performed using LOCROT (I.J.T.). The parameters were then refined with the density correlation programs of Bricogne¹⁸. One averaged subunit of the electron density map obtained from the solvent flattening was displayed in FRODO³⁵ on an Evans and Sutherland PS390. The sequence could be assigned from the position of the disulphide bridge and the putative calcium-binding site for all residues in the sheets and helix. The resolution was extended with simulated annealing using non-crystallographic restraints in XPLOR³⁶ and rebuilding into electron density maps (coefficients $2F_o - F_c$, $F_o - F_c$) calculated from phases that had been combined with the MIR phases until all residues had been inserted. The resolution was then extended to 2 Å and 10 calcium ions, 4 acetate ions and 879 water molecules were added. Least-squares refinement in RESTRAIN³⁷ gave a crystallographic R-factor of 0.18 for all 78,910 reflections in the resolution range 8–2 Å. The r.m.s. deviations from stereochemical ideality are 0.17 Å for bond distances and 3.45° for bond angles. The average isotropic B values are 22.9 Å² for protein and 37.7 Å² for solvent molecules. There were no residues in the disallowed region of a Ramachandran plot³⁸. The CCP4 suite of programs (SERC, Daresbury) has been used for all crystallographic calculations, except the structural refinement.

Merging R factor
$$\frac{\sum_{hkl} \sum_{i=1}^n |I(hkl)_i - I(hkl)|}{\sum_{hkl} \sum_{i=1}^n I(hkl)_i}$$

Isomorphous R factor
$$\frac{\sum_{hkl} |F_{\text{deriv}}(hkl) - F_{\text{nat}}(hkl)|}{\sum_{hkl} |F_{\text{nat}}(hkl)|}$$

Phasing power
$$[\sum_{hkl} F_{\text{heavy}}^2(hkl) / \sum_{hkl} e^2(hkl)]^{1/2}$$

D, Enraf-Nonius CAD-4 diffractometer mounted on a 1,500 W sealed tube; F, Enraf-Nonius FAST area-detector mounted on a CuK α microfocus tube (800 W); X, Siemens Xentronics area-detector mounted on a Siemens XP18 generator; I, MAR research imaging plate mounted on a Siemens XP18 generator.

of strands ABCDKLNO, elaborated by the addition of three further antiparallel strands (EFG and HIJ) forming a β -meander at the same end of each of the sheets to give the topology +11–9+7–1–1–1–3+1+1+5–9–1+12–13. In this arrangement, strands A and M are both hydrogen-bonded to strand L (see Fig. 3). The disulphide between Cys 36 and Cys 95 links the two adjacent strands (L and C) of one β -sheet. A long α -helix between strands L and M is folded on top of this β -sheet. There is also an N-linked oligosaccharide at Asn 32 on this sheet; only one saccharide residue is visible in the electron density map.

The hydrophobic core between the two sheets is comprised mainly of tryptophan, tyrosine, phenylalanine and leucine. The core is closed off by two β -arches between the two sheets; strands joining B to C and J to K are hydrogen-bonded and antiparallel, an arrangement characteristic of proteins derived from a jellyroll motif. One end of the core, formed by 11 residues at the amino-

terminus and those in the N and O strands at the carboxy terminus (Figs 2, 3), has hydrophobic residues accessible to solvent. The other end of the core is involved in interactions with a neighbouring promoter (see below) and so is inaccessible to solvent.

Tertiary fold comparisons

The similarities of the amino-acid sequences of SAP, human CRP, CRP of the arachnid *Limulus polyphemus* (the horseshoe crab) and hamster SAP suggest that they may have similar three-dimensional structures. Comparative modelling¹⁹ shows that these pentraxins can have equivalent antiparallel structures, with insertions and deletions in the loops between the β -strands and α -helices (N.S., H.E.W. and T.L.B., unpublished results). Figure 4 shows that most solvent-inaccessible aromatic side chains are conserved to give compact hydrophobic cores in all members of the family.

ARTICLES

The jellyroll topology of the pentraxins is reminiscent of that of the picornavirus coat proteins, which also have pentameric structures. However, pentraxins most closely resemble legume lectins such as concanavalin A²⁰ and pea lectin²¹ (Fig. 2a). In each case, the arrangement of strands is identical, but the N and C termini are in different positions (Fig. 3). In pea lectin, the N terminus is at strand M and the C terminus at strand L (labelling of the topologically equivalent strands follows those in SAP as shown in Fig. 3). Strands A and M are both hydrogen-bonded to strand L in a similar manner to those in SAP. In the three-dimensional structure of concanavalin A, the N terminus is at strand E and the C terminus at the end of strand D owing to a post-translational cleavage which follows ligation of the true termini between strands L and M²². Pea lectin is additionally cleaved at the loop connecting strands I and J.

Alignment of sequences (Fig. 4) on the basis of topologically equivalent features of the three-dimensional structures²³ shows that helices occupy different positions in the pentraxins and legume lectins and that the amino-acid sequences of the two families have identities of only ~11%. The two main helices in SAP occur before and after strand L, whereas the helices in the legume occur at the C terminus of strand J. There is a long insertion between the end of the helix after strand D and the beginning of strand E in the lectins relative to the pentraxins. Strands G, H and I, together with the type IV β -hairpin between H and I, are identical in both SAP and pea lectin. The so-called pentraxin octapeptide signature sequence, HXCXS/TWXS, is in this region, but this is not conserved in the legume lectins.

Structure of pentamer and decamer

Figure 2b shows the structure of the pentamer. The five subunits are arranged in a ring with a hole 20 Å in diameter and 35 Å deep at the centre. The two layers of β -strands in each subunit are in planes normal to the 5-fold axis. Strands G, I and J of one protomer interact with strand N and loops between strands A and B, C and D, G and H and K and L of an adjacent protomer (Fig. 2b). When SAP is overlaid on the pea lectin, strand J does not give as good a fit as other strands. In SAP this strand has moved to provide inter-protomer contacts, with all residues between Pro 113 and Leu 119 being involved. The subunit interactions consist of hydrogen bonds between main-chain peptide groups, three salt bridges and some hydrophobic contacts, in contrast to other pentameric systems where there are often inter-subunit β -sheets. The surface area of the protomer that is buried on formation of the pentamer is 410.5 Å², comprising 15.4% of the total surface of the protomer. Figure 4 shows the residues involved in these extensive interactions and which account for the considerable stability of the SAP pentamer.

In contrast, the SAP decamer is readily dissociated by reducing the pH to 5.5. The simplest explanation for this is that the decamer is stabilized by ionic interactions involving carboxylate and/or imidazole groups. Electron microscopy has clearly shown that the SAP pentamers are packed face-to-face²⁴ and it seems probable that the faces in contact are those carrying the α -helix, as we show (see below) that the calcium-dependent ligand-binding sites are on the other face and such binding is exhibited by the decamer. Furthermore, adjacent sites are accessible to chymotryptic cleavage in the decamer in the absence of calcium ions. It is not clear which groups are involved in the

decamer stabilization, although Glu 167, positioned on the helix, is a likely candidate. A compact decamer can be modelled if the pentamer 5-fold axes are in line but the subunits are out of step, allowing the helices from one layer to pack between those of the opposite layer. Structure analysis of a crystal form produced at neutral pH with 10 protomers in the asymmetric unit is in progress.

The models of CRP and hamster SAP demonstrate that the pentamers can have very similar inter-subunit interactions. *Limulus* CRP is hexameric²⁵ and we have also constructed this arrangement of protomers by operating with 6-fold axis placed slightly farther from the subunit than the 5-fold axis of SAP.

Calcium-binding site

In serum amyloid P component, we have identified two large spheres of density which are too heavy to be oxygen atoms and are in positions that imply the presence of calcium ions, between 4.0 and 4.3 Å apart in the five subunits and bridged by common side chains. Calcium ion 1 is coordinated to the side chains of Asp 58, Asn 59, Glu 136, Asp 138 and the main-chain carbonyl

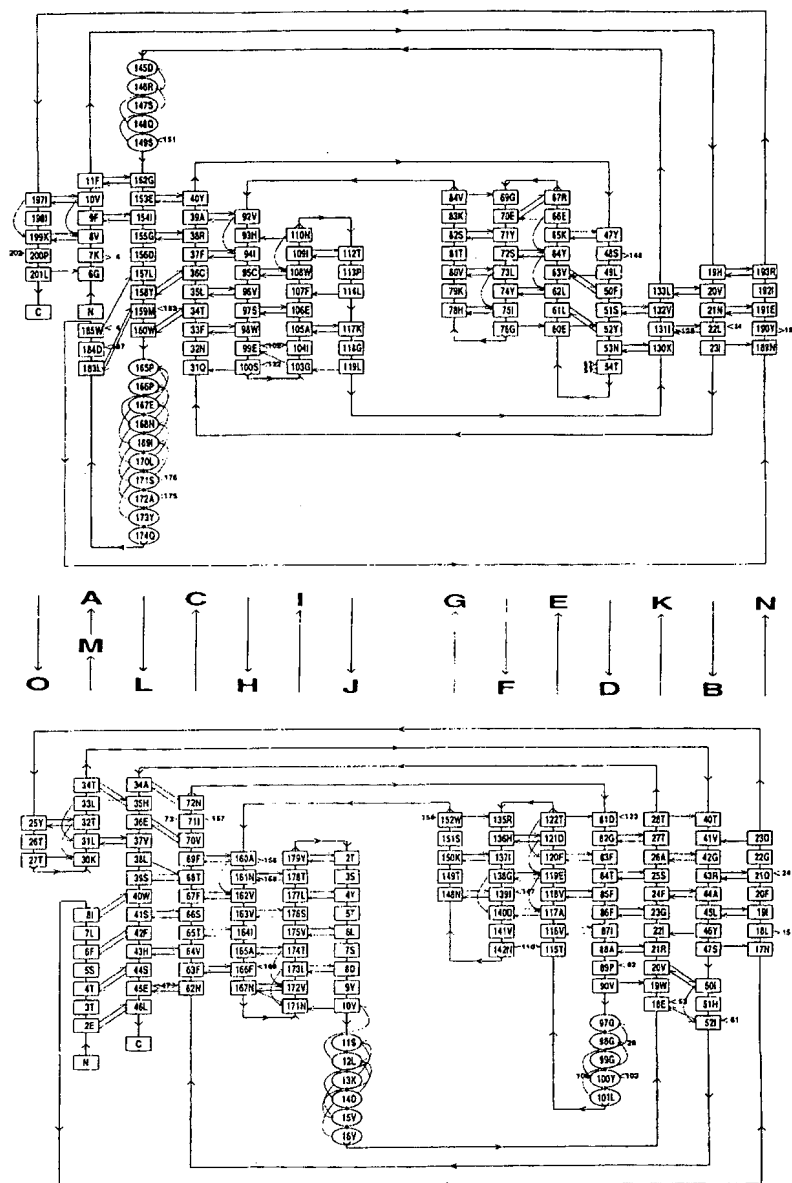


FIG. 3 Hydrogen-bonding diagrams generated by HERA⁴⁰ for SAP (above) and pea lectin (Brookhaven code 2LTN) (below). The secondary structural elements were assigned according to the criteria defined by Kabsch and Sander⁴¹.

of Gln 137 (Fig. 5a). Unlike many calcium-binding sites in proteins, such as parvalbumin, the coordinating residues come from different parts of the sequence. This is achieved by a distortion at the start of strand E carrying residues Asp 58 and Asn 59 and the region containing Glu 136 and Asp 138 looping over towards calcium 1 (Fig. 5a). The calcium ligation is likely to be an important local structural determinant. The seventh coordination site is occupied by a ligand that has the appearance of an acetate ion from the crystallization buffer in protomers 1, 2, 4 and 5,

but in protomer 3 this position is filled by the side chain of Glu 167 of an adjacent molecule in the crystal lattice. Glu 136, Asp 138 and the acetate/lattice contact form a bridge to a second, more loosely bound calcium ion (calcium 2). The coordination of calcium 2 is completed by Gln 148 and two water molecules. In a cross-phase difference Fourier transform of cerium sulphate soaked crystals, we find that calcium 2 is displaced by a cerium ion. Calcium 2 is also removed when the crystals are soaked in calcium-free buffers. These observations are consistent

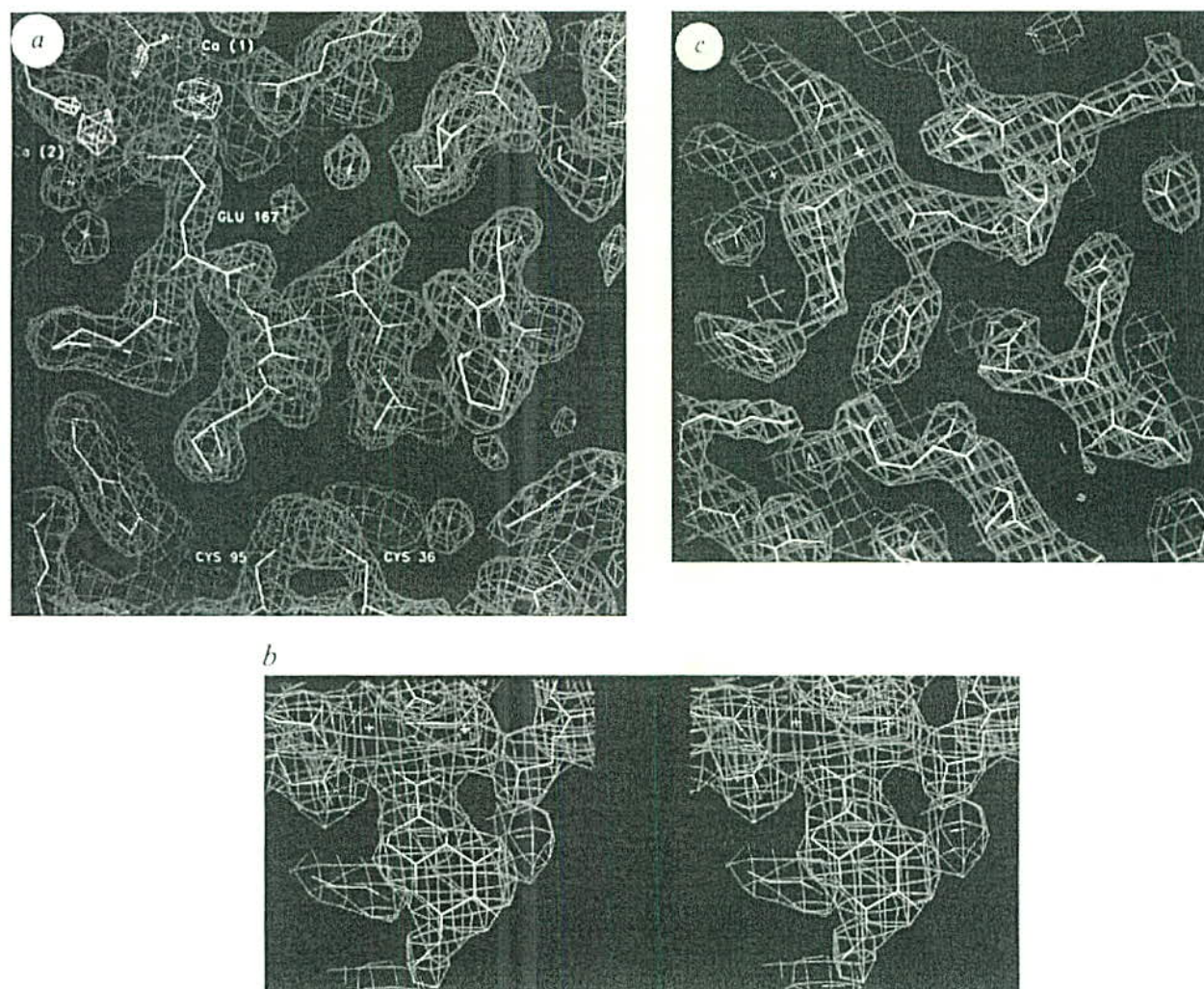


FIG. 1 Final refined electron density of SAP. *a*, Refined electron density map ($2F_o - F_c$) calculated at 2 Å resolution and contoured at 1 r.m.s. This shows the SAP helix packing against the β -pleated sheets in the region of the disulphide bridge Cys 36–Cys 95. The top left hand of the diagram shows a region where the side chain Glu 167 makes contact with the calcium-binding site of a symmetry-related molecule. The region of the calcium ions is contoured at 5 r.m.s., showing that the glutamate bridges the two ions in a similar manner to the acetate ion. *b*, MO/BDG complexed with SAP contoured at 1 r.m.s. electron density level at 2.9 Å resolution. Crystals of SAP prepared by the batch method¹⁵ in the presence of MO/BDG were isomorphous with crystals of native SAP and had cell dimensions $a = 69.06$ Å, $b = 99.3$ Å, $c = 96.75$ Å and $\beta = 95.84^\circ$. Data were collected on a MAR imaging plate system mounted on a Siemens XP-18 rotating anode and a final R_{merge} of 5.3% was obtained. Data were 95.8% complete to 2.9 Å resolution. A final crystallographic R factor of 0.197 for all reflections in the range 10–2.9 Å was obtained after refinement in XPLOR without non-crystallographic restraints. The success of this experiment depended on soaking in an excess of the sugar so as to displace the acetate ion which otherwise occupies the same position at the calcium-binding site. The

MO/BDG is well defined only in promoter 3, shown here, where it is involved in a crystal contact; there is density consistent with the same binding at other subunits. *c*, Phosphoethanolamine (PE) complexed with SAP contoured at 1 r.m.s. electron density level at 2.9 Å resolution. Crystals of SAP prepared by the batch method¹⁵ in the presence of 9 mM PE were not isomorphous with crystals of native SAP and had cell dimensions $a = 67.05$ Å, $b = 103.43$ Å, $c = 102.43$ Å and $\beta = 95.73^\circ$. In space group $P2_1$. Data were collected on a MAR imaging plate system mounted on a Siemens XP-18 rotating anode and a final R_{merge} of 6.4% was obtained. Data were 95.5% complete to 2.9 Å resolution. In the initial refinement, the SAP pentamer was treated as a single rigid body and an R factor of 0.33 was obtained for the 3,513 reflections at 30–6 Å. The five subunits were then treated as individual rigid bodies and refinement produced an R factor of 0.26 for the 11,628 reflections at 18–4 Å. The final crystallographic R factor of 0.20 for all reflections in the range 10–2.9 Å was obtained after refinement in XPLOR; non-crystallographic symmetry restraints were not applied. Coordinates have been deposited with the Protein Data Bank, Chemistry Department, Brookhaven National Laboratory, Upton, New York 11973, USA.

ARTICLES

with the more solvent-accessible position and fewer protein ligands of calcium 2.

Residues that provide ligands to the calcium ions are conserved at the start of strand E carrying residues Asp 58 and Asn 59 and the region containing Glu 136 and Asp 138 looping over towards calcium 1 (Fig. 5a). The calcium ligation is likely to be an important local structural determinant. The seventh coordination site is occupied by a ligand that has the appearance of an acetate ion from the crystallization buffer in protomers 1, 2, 4 and 5, but in protomer 3 this position is filled by the side chain of Glu 167 of an adjacent molecule in the crystal lattice. Glu 136, Asp 138 and the acetate/lattice contact form a bridge to a second, more loosely bound calcium ion (calcium 2). The coordination of calcium 2 is completed by Gln 148 and two water molecules. In a cross-phase difference Fourier transform of cerium sulphate soaked crystals, we find that calcium 2 is displaced by a cerium ion. Calcium 2 is also removed when the crystals are soaked in calcium-free buffers. These observations are consistent with the more solvent-accessible position and fewer protein ligands of calcium 2.

Residues that provide ligands to the calcium ions are conserved in all SAPs, but although Asp 58 is found in hamster SAP, human CRP and *Limulus* CRP, it varies in other CRPs. Nevertheless, the general organization of the site is probably retained. The disposition of these groups on surface loops where sequence differences accumulate could explain the considerable change in calcium affinity between CRP and SAP. In CRP there is evidence that both calciums bind with the same affinity at neutral pH²⁶, whereas in SAP our results show that calcium 2 has fewer protein ligands and can be preferentially unloaded. The tighter calcium binding and equivalence of the sites in CRP could also be due to the substitution of Asp 145 in SAP by glutamate in CRP. The longer side chain in CRP would permit a full complement of protein ligands to calcium site 2.

The existence of two metal ions bridged by protein ligands is reminiscent of concanavalin A. However, although the metal-binding sites in the legume lectins are on the same face of the

protein, they are at a different position, between strands E and F, compared with D and E on SAP. There are no ligands of the two calcium ions that are topologically equivalent to those for the calcium and manganese of the lectins. Thus, although there is evidence for divergent evolution of the protein folds of the plant lectins and the pentraxins, this is not supported by a conservation of similar metal-binding sites.

Ligand binding

SAP binds to methyl 4,6-*O*-(1-carboxyethylidene)- β -D-galactopyranoside (MO β DG), which is not recognized by CRP. Figure 1b shows the electron density for MO β DG complexed with SAP. The sugar derivative binds directly through the acidic group to the two calcium ions in a similar way to the acetate that it replaces. The other interactions include two hydrogen bonds formed between the 4, 6 oxygen atoms of the ring and the amide nitrogen atoms of Gln 148 and Asn 59, respectively, each of which bind to the calcium ions through their amide oxygens (Fig. 5b). Thus, the role of calcium is not to bind the galactopyranoside ring directly, but rather to mediate its binding by orienting side-chain amides in a way that resembles saccharide binding in lectins. There is only one hydrogen bond to the galactopyranoside ring (Fig. 5b). Thus, it is the methyl 4,6-*O*-(1-carboxyethylidene) ring that forms the main interactions with the protein, explaining why neither the non-cyclic acetal of MO β DG nor the simple monosaccharides bind (refs 3, 8 and M.B.P. *et al.*, manuscript in preparation). The *R*-isomer bridge methyl group points into a hydrophobic pocket formed by Leu 62, Tyr 64 and Tyr 74. The differences in the hydrophobic pocket and the ligand distribution at the calciums in CRP may explain the poor binding to MO β DG. It is probable that this site is involved in binding to amyloid fibrils, and its specificity will be discussed elsewhere (M.B.P. *et al.*, manuscript in preparation).

The highest-affinity interaction of human CRP is its calcium-dependent binding to phosphocholine (PC). Studies on human CRP^{27,28} have implicated amino-acid residues 50–60 in the ability

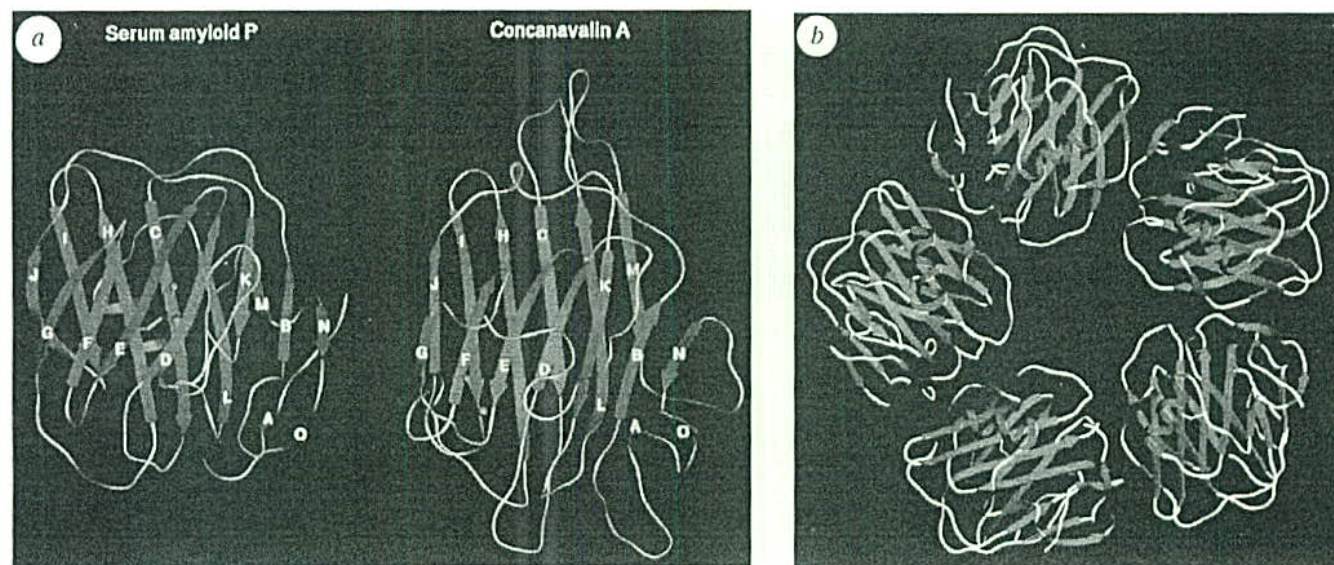


FIG. 2 Ribbon drawings generated with Setor³⁹ with β -strands shown as blue arrows, helices as ribbons and loops as cords. a, The topology of SAP compared with the legume lectin concanavalin A (Con A). The strands are labelled from the N terminus of SAP, ABCDEFGHIJKLMNO, with structurally equivalent strands in Con A labelled similarly. The positions of the N and C termini in Con A are altered with respect to the nascent protein by ligation of the origin N and C termini and proteolytic cleavage between strands D and E. The two structures can be superim-

posed by a least-squares fit using the $C\alpha$ positions of the two sets of antiparallel sheets to produce an r.m.s. deviation of 2.4 Å. There is a greater angle between these sheets in SAP than in Con A. b, Structure of the pentamer of SAP viewed along the non-crystallographic 5-fold axis of symmetry. The pairwise r.m.s. deviations for all main-chain atoms range from 0.19 Å between protomers 1 and 5 to 0.24 Å between protomers 3 and 4.

to bind PC, and more recent mutagenesis experiments²⁹ have identified Lys 55 and Arg 56 as key residues. The native SAP structure shows that Asp 58 and Asn 59 from this loop are involved in coordinating one of the two calcium ions. Human SAP, in contrast, does not bind to PC, although in common with CRP it does bind to phosphoethanolamine (PE). The electron density maps for SAP co-crystallized with PE (Fig. 1c) show a major site in all subunits, which coincides with that which binds acetate ions and MOβDG and indicates a direct interaction between the phosphate group and both calcium ions. In contrast to MOβDG, this interaction with PE displaces Glu 167 from its intermolecular interaction with the calcium ions of protomer 3, explaining the observed disturbance of crystal packing. Binding of MOβDG or PE in the common site probably stabilizes the whole calcium-binding region, including Asn 59 which binds MOβDG and calcium and Asp 58 which binds calcium. Both PE and PC probably bind at the calcium ion in CRP.

SAP aggregates rapidly in neutral solutions in the presence of

calcium ions, presumably through intermolecular interactions involving the surface of the decamer with the exposed calcium-binding site. This is consistent with the observation that this interaction can be inhibited by MOβDG³⁰ and by PE (M.B.P. *et al.*, manuscript in preparation). Similarly, they compete for the calcium site with Glu 167, which is an important lattice contact in the pentamer crystals.

In the presence of calcium, SAP also binds polyanions such as heparan sulphate, dermatan sulphate⁷ and DNA⁴. It has been suggested that the DNA interaction involves a decapeptide around Arg 120, with some resemblance to certain histone sequences, and for which a helical structure was proposed³¹. However, in human SAP this region is not helical and other SAPs do not have arginine at this position. Nevertheless, it does have affinity for multivalent anions, as demonstrated by the binding of phosphotungstate at this site in the heavy-atom isomorphous derivative (see Table 1). There are several basic patches in SAP, for example Arg 120, Arg 77, His 78, Lys 79 and Arg 57, which are on the same surface as the calcium- and

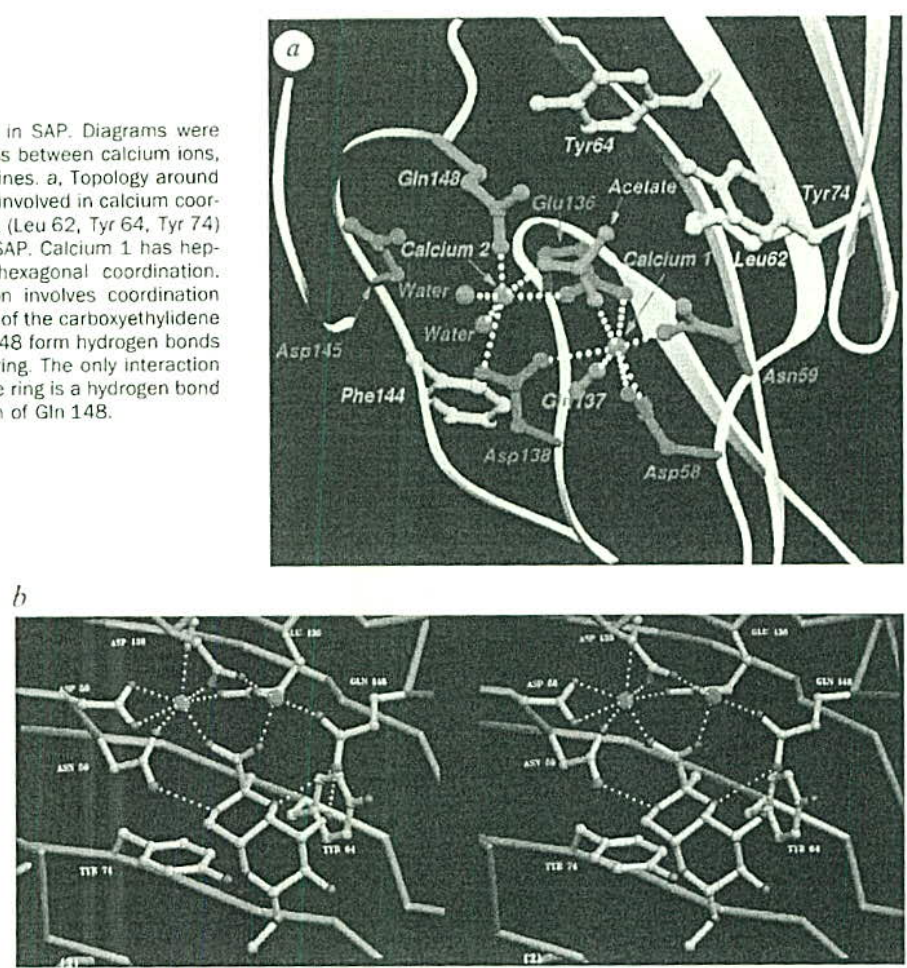


FIG. 4 Alignment of the sequences of the pentraxins (one protomer sapm and one pentamer sapp) of SAP, human C-reactive protein (hcrp), hamster SAP (female hamster protein) (1fhp) and *Limulus* CRP (1lim) with the sequences of legume lectins, concanavalin A (3cna) and pea lectin (2ltu and 1lte). These alignments are based on comparing the three-dimensional structure of SAP with those of the legume lectins²³. Key to characters in the alignments²²: solvent inaccessible, upper case

letter; solvent accessible, lower case letter; positive ϕ , italic letter; cis-peptide, breve over letter; hydrogen bond to other side chain, tilde over letter; hydrogen bond to main-chain amide, bold letter; hydrogen bond to main-chain carbonyl, letter underlined; disulphide bond, cedilla below letter; * metal-binding sites in SAP and Con A; | residues of SAP involved in inter-protomer contacts.

ARTICLES

FIG. 5 The calcium- and ligand-binding sites in SAP. Diagrams were generated using Setor²⁹. Important interactions between calcium ions, ligands and the protein are shown by dotted lines. a, Topology around the calcium-binding site showing the residues involved in calcium coordination and those in the hydrophobic pocket (Leu 62, Tyr 64, Tyr 74) and the cleavage site (Phe 144, Asp 145) in SAP. Calcium 1 has heptagonal coordination, and calcium 2 has hexagonal coordination. b, MO β DG-binding site. The main interaction involves coordination between the calcium ions and the carboxylate of the carboxyethylidene ring. The amide nitrogens of Asn 59 and Gln 148 form hydrogen bonds to the 4, 6 oxygens of the carboxyethylidene ring. The only interaction between the protein and the galactopyranoside ring is a hydrogen bond between the 3 oxygen and the amide nitrogen of Gln 148.



PE-binding sites. It seems more likely that the phosphate backbone of DNA and the sulphated polysaccharides bind both at the calciums and at the basic sites, possibly on more than one subunit simultaneously. These extensive interactions probably account for the ability of SAP to displace from DNA the H1-type histones in chromatin⁵. The sequence differences in the basic regions and in the calcium-binding region could explain the variable affinity for DNA shown by different pentraxins. Human CRP, for instance, binds DNA only at low, non-physiological ionic strength. Low concentrations of both PE and MO β DG enhance the binding of SAP to DNA and both compete with the interaction at high concentration (M.B.P. *et al.*, manuscript in preparation). Similar effects are observed for SAP binding to immobilized PE (M.B.P. *et al.*, manuscript in preparation). As we can find no evidence for a second binding site for PE or MO β DG in the present experimental conditions, these effects may be mediated through inter-subunit interactions.

Ligands for SAP, such as PE in phosphatidylethanolamine and MO β DG-like sugars³², are common in microorganisms. In this respect it resembles CRP¹³ and mannose-binding protein³³, a plasma C-type lectin involved in innate resistance to infection; these recognize widely distributed microbial epitopes, suggesting that SAP, too, may participate in host defence.

Proteinase resistance and amyloidosis

SAP and CRP are both remarkably resistant to proteolytic degradation in the presence of calcium. In contrast, in the absence of calcium both are cleaved by some enzymes, particularly α -chymotrypsin and pronase^{26,34}. Although this cleavage does not cause fragmentation of either the whole molecule or the individual subunits under non-denaturing conditions, it does cause loss of calcium-binding activity by the pentraxins and

abolishes their capacity for calcium-dependent ligand binding. It is therefore of interest that the major site of cleavage of SAP is at residues 144–145 whereas in CRP it is at residues 146–147 (pronase) or 145–146 (Nagarse protease). This is part of a loop held in place by calcium ligation which, in the calcium-free form, may be only loosely associated with the body of the protein and, therefore, susceptible to proteolysis. Most loops of SAP are held close to the body of the pentamer and this makes them less easily accessible to the active sites of proteolytic enzymes.

Resistance to proteinase digestion is likely to be an important aspect of the normal function of SAP, and may also contribute to the persistence of amyloid deposits. The SAP normally associated with the glomerular basement membrane and the surface of elastic fibre microfibrils^{9,10} may protect these components of the extracellular matrix from inappropriate degradation. On the other hand, amyloid fibrils are abnormal extracellular structures which should be recognized and degraded, but which nevertheless persist *in vivo*. In this pathological situation, the binding of SAP to amyloid fibrils may be responsible. Protection could result simply from coating by SAP, which is completely unaltered with respect to its normal circulating form¹¹, and which would therefore not be expected to trigger macrophage activation or phagocytosis. However, the proteinase resistance of SAP itself may be a significant factor. Availability of the complete high-resolution structure of SAP and its ligand-binding site now offer the opportunity for direct modelling of competitive inhibitors of SAP binding and for producing binding-site homologues, either of which could be used as drugs to displace SAP from amyloid deposits *in vivo*. This would open up new avenues for treatment of amyloidosis, enabling the body to mobilize and degrade the fibrils which may otherwise be inappropriately protected by SAP. □

LETTERS TO NATURE

Received 9 August; accepted 9 December 1993.

1. Pepys, M. B., Dyck, R. F., de Beer, F. C., Skinner, M. & Cohen, A. S. *Clin. exp. Immun.* **38**, 284–293 (1979).
2. Coria, F. et al. *Lab. Invest.* **58**, 454–458 (1988).
3. Hind, C. R. K., Collins, P. M., Caspi, D., Baltz, M. L. & Pepys, M. B. *Lancet* **ii**, 376–378 (1984).
4. Pepys, M. B. & Butler, P. J. G. *Biochem. biophys. Res. Commun.* **148**, 308–313 (1987).
5. Butler, P. J. G., Tennent, G. A. & Pepys, M. G. *J. exp. Med.* **172**, 13–18 (1990).
6. de Beer, F. C., Baltz, M. L., Holford, S., Feinstein, A. & Pepys, M. B. *J. exp. Med.* **154**, 1134–1149 (1981).
7. Hamazaki, H. *J. biol. Chem.* **262**, 1456–1460 (1987).
8. Hind, C. R. K. et al. *J. exp. Med.* **159**, 1058–1069 (1984).
9. Breathnach, S. M. et al. *Nature* **293**, 652–654 (1981).
10. Dyck, R. F. et al. *J. exp. Med.* **152**, 1162–1174 (1980).
11. Pepys, M. B. et al. *Proc. 41st ASMS Conf. Mass Spectrometry and Allied Topics*, San Francisco, California (Elsevier, New York, 1993).
12. Pepys, M. B. et al. *Nature* **273**, 168–170 (1978).
13. Pepys, M. B. & Baltz, M. L. *Adv. Immun.* **34**, 141–221 (1983).
14. Hawkins, P. N., Lavender, P. J. & Pepys, M. B. *New Engl. J. Med.* **323**, 508–513 (1990).
15. Wood, S. P. et al. *J. molec. Biol.* **202**, 169–173 (1988).
16. Hawkins, P. N., Tennent, G. A., Woo, P. & Pepys, M. B. *Clin. exp. Immun.* **84**, 308–316 (1991).
17. Wang, B. C. *Meth. Enzym.* **115**, 90–112 (1985).
18. Bricogne, G. *Acta crystallogr.* **A32**, 832–847 (1976).
19. Sali, A., Overington, J. P., Johnson, M. S. & Blundell, T. L. *Trends biochem. Sci.* **15**, 235–240 (1990).
20. Hardman, K. D. & Ainsworth, C. F. *Biochemistry* **11**, 4910–4919 (1972).
21. Einspahr, H., Parks, E. H., Suguna, K., Subramanian, E. & Suddath, F. L. *J. biol. Chem.* **261**, 16518–16527 (1986).
22. Carrington, D. M., Aulfit, A. & Hanke, D. E. *Nature* **313**, 64–67 (1985).

23. Sali, A. & Blundell, T. L. *J. molec. Biol.* **212**, 403–442 (1990).
24. Pepys, M. B. et al. *Lancet* **i**, 1029–1031 (1977).
25. Tennent, G. A. et al. *Eur. J. Biochem.* **214**, 91–97 (1993).
26. Kinoshita, C. M. et al. *Biochemistry* **28**, 9840–9848 (1989).
27. Swanson, S. J. & Mortenson, R. F. *Molec. Immun.* **27**, 679–687 (1992).
28. Liu, T.-Y. et al. *J. Protein Chem.* **6**, 263–272 (1987).
29. Agrawal, A., Xu, Y., Ansari, D., Macon, K. J. & Volanakis, J. E. *J. biol. Chem.* **267**, 25352–25358 (1992).
30. Hind, C. R. K., Collins, P. M. & Pepys, M. B. *Biochim. biophys. Acta* **802**, 148–150 (1984).
31. Turnell, W. G., Satchwell, S. C. & Travers, A. A. *FEBS Lett.* **232**, 263–268 (1988).
32. Hind, C. R. K., Collins, P. M., Baltz, M. L. & Pepys, M. B. *Biochem. J.* **225**, 107–111 (1985).
33. Hartshorn, K. L. et al. *J. clin. Invest.* **91**, 1414–1420 (1993).
34. Kinoshita, C. M. et al. *Protein Sci.* **1**, 700–709 (1992).
35. Jones, T. A. *J. appl. Crystallogr.* **11**, 272–282 (1978).
36. Brunger, A. T., Karpus, M. & Petsko, G. A. *Acta crystallogr.* **A45**, 50–61 (1989).
37. Haneef, I., Moss, D. S., Stanford, M. J. & Borkakoti, N. *Acta crystallogr.* **A41**, 426–433 (1985).
38. Ramachandran, G. N., Ramakrishnan, C. & Sasisekharan, V. *J. molec. Biol.* **1**, 95–99 (1963).
39. Evans, S. V. *J. molec. Graphics* **11**, 134–138 (1993).
40. Hutchinson, E. G. & Thornton, J. M. *Proteins* **8**, 203–212 (1990).
41. Kabsch, W. & Sander, C. *Biopolymers* **22**, 2577–2637 (1983).
42. Overington, J. P., Johnson, M. S., Sali, A. & Blundell, T. L. *Proc. R. Soc. B* **241**, 132–145 (1990).

ACKNOWLEDGEMENTS. This work was supported by project and programme grants from the MRC. We thank D. M. Blow for use of film scanning facilities at Imperial College, R. Pickersgill for data collection on the uranyl derivative, R. Laskowski for running LOCROT, D. Carvin and G. Louie for discussion and J. R. Gallimore and G. A. Tennent for technical assistance. J.E. and H.F.W. are joint first authors.

LETTERS TO NATURE

Transition between spiral and target states in Rayleigh-Bénard convection

Michel Assenheimer & Victor Steinberg

Department of Physics of Complex Systems,
The Weizmann Institute of Science, Rehovot 76100, Israel

RAYLEIGH-BÉNARD convection^{1–4}, which occurs when a shallow fluid layer is heated from below, is commonly regarded as a paradigm for pattern formation under non-equilibrium conditions. The formation of hexagonal arrays of Bénard cells is well known, but more complex patterns such as targets⁵ and spirals^{6–11} have also been reported. Similar patterns have been seen in electrohydrodynamical convection^{12,13}, oscillatory chemical reactions^{14–19} and biological systems^{19,20}. In general, the spiral and target states are found for different experimental conditions. Here we report the observation of a continuous transition between states containing many spirals and many targets, in a fluid undergoing Rayleigh-Bénard convection near the gas-liquid critical point. Whether spirals or targets are observed depends on the Prandtl number, the ratio between the thermal and viscous timescales in the fluid. Neither of these states seems to be predicted by the hydrodynamic equations that describe the fluid motions^{1–4, 21}. The fact that the transformation of one pattern into the other is continuous, and that under some conditions they can coexist, suggests that they may be generated by the same or a similar mechanism.

We studied Rayleigh-Bénard (RB) convection in SF₆ near its thermodynamic critical point (the critical temperature, pressure and density are $T_c = 318.7$ K, $p_c = 37.8$ bar and $\rho_c = 0.73$ g cm⁻³ respectively)⁵. The fluid was enclosed in a stainless steel vessel having a sapphire window 19 mm thick and 102 mm in diameter. A nickel-plated copper bottom-plate (to which a thermofoil heater was attached) sandwiched a mylar spacer. The spacer had an inner diameter $D = 30$ mm and thickness $d = 130$ μ m, giving a radial aspect ratio of $\Gamma = D/2d = 115$. The vessel was coupled by a thin tube to a small hot volume, in which pressure feedback could be performed by heating the fluid. Both vessels were enclosed in a constant-temperature bath in which the long-term stability was better than ± 1.0 mK. Long-term stability of the

bottom temperature was better than ± 0.3 mK. Pressure feedback resulted in long-term stability of ± 0.5 mbar. Shadowgraph techniques, with and without computer enhancement, were used to visualize the patterns. The cell was filled far from the critical point so that the desired density could be obtained with an accuracy of $\pm 0.25\%$ from an equation of state²².

The RB problem has two independent control parameters, namely the Rayleigh number Ra which represents the dimensionless temperature difference across the fluid and is a measure of the external stress applied to the system, and the Prandtl number $Pr^{1-4, 21}$. In the Boussinesq approximation^{1-4, 21} (roughly speaking, temperature-independent fluid properties) the pattern selection is completely determined by the values of Ra , Pr and k , the wavenumber of the pattern^{1-4, 21}. These patterns are summarized in what is known as the Busse balloon (the part of the Ra , Pr , k phase space bounding the stable patterns) and have been verified extensively both experimentally and theoretically¹⁻³. Deviations from the Boussinesq approximation have been shown to influence the pattern selection near onset^{1-4, 21, 23}. Non-Boussinesq behaviour of fluids may be quantified by the parameter Q , a global dimensionless measure of the variations of the relevant fluid properties (density, viscosity, thermal conductivity and isobaric specific heat and thermal expansion) with temperature²¹. Its role for the pattern selection is commonly neglected, especially far from onset.

Recently, we showed that RB convection near T_c has the following advantage because most thermodynamic and kinetic properties of fluids have a universal power law behaviour as a function of the reduced average temperature $\tau = (\bar{T} - T_c)/T_c$: (1) the onset of convection (ΔT_0) decreases as τ is decreased, enabling convection in thin cells and thus large aspect ratios; (2) Pr can be scanned continuously from ~ 1 to practically infinity; (3) Q can be varied independently because it is a function of d , τ and ΔT . Here $\bar{T}$ and ΔT are the average temperature and the temperature difference across the fluid layer respectively⁵.

In Fig. 1a, b and c we present spiral, target and mixed states for values of the Prandtl number of 3.3, 6.6 and 3.6 respectively. For intermediate Prandtl numbers ($3 \leq Pr \leq 4$) as well as in transients, we observed both patterns simultaneously and transitions between them. On the whole, the steady state, comprising many extended patterns, remains time dependent, although single stationary spirals and targets are also present. Spiral-pattern-dominant states exist for $Pr \leq 3.5$, while target pattern dominant states

JMB



Crystal Structure of a Decameric Complex of Human Serum Amyloid P Component with Bound dAMP

Erhard Hohenester^{1*}, Winston L. Hutchinson², Mark B. Pepys²
and Steve P. Wood³

¹Department of
Crystallography, Birkbeck
College, University of London
London, WC1E 7HX, UK

²Immunological Medicine Unit
Royal Postgraduate Medical
School, Hammersmith Hospital,
London, W12 0NN, UK

³School of Biological Sciences
University of Southampton
Southampton, SO16 7PX, UK

Serum amyloid P component (SAP) is a glycoprotein that binds in a calcium-dependent fashion to a variety of ligands including other proteins, glycosaminoglycans and DNA. SAP is universally associated with the amyloid deposits in all forms of amyloidosis including Alzheimer's disease. Small-molecule ligands that displace SAP from amyloid fibrils and thereby expose the fibrils to proteolytic clearance mechanisms hold potential as drugs for the prevention and treatment of amyloidosis. We have carried out a screen for novel SAP ligands and have identified 2'-deoxyadenosine-5'-monophosphate (dAMP) as a ligand. The crystal structure of the SAP-dAMP complex determined at 2.8 Å resolution ($R = 0.232$, $R_{\text{free}} = 0.252$) reveals a decamer in which all interactions between SAP pentamers are mediated by the ligand. The stability of the decamer in solution has been demonstrated by gel filtration chromatography. The two calcium ions of SAP are bridged by the dAMP phosphate group and five hydrogen bonds are formed between the protein and the ligand, including specific interactions made by the adenine base. This mode of dAMP binding is not compatible with the nucleotide being part of double-helical DNA. The SAP-dAMP decamer is stabilized mainly by base-stacking of adjacent ligand molecules and possibly by electrostatic interactions involving the dAMP phosphate groups; decamerization buries 1000 Å² (2.6%) of the pentamer solvent-accessible surface. Ligand-induced decamerization of SAP, which utilizes the high cooperativity of a multiple-site interaction, may be a strategy to overcome the problems for drug design associated with the rather modest affinities of SAP for small-molecule ligands.

© 1997 Academic Press Limited

*Corresponding author

Keywords: pentraxin; amyloidosis; X-ray crystallography; drug design

Introduction

Serum amyloid P component (SAP) is a normal plasma glycoprotein isolated as a decamer of identical subunits in the presence of EDTA (Pinteric *et al.*, 1976; Osmand *et al.*, 1977). Each subunit comprises 204 amino acid residues (molecular mass 25.5 kDa), with one disulphide bond linking cysteine residues 36 and 95 and a single N-linked biantennary oligosaccharide attached to Asn32 (Pepys *et al.*, 1994). SAP and its homologue C-reactive protein (CRP), the classic acute phase protein

in man, belong to the pentraxin family of plasma proteins, which have been highly conserved throughout vertebrate evolution (Osmand *et al.*, 1977; Pepys *et al.*, 1978; Pepys & Baltz, 1983; Gewurz *et al.*, 1995). Although the precise role of SAP has yet to be defined, its ability to form tight complexes with other biological macromolecules is likely to be at the root of its function(s). SAP binds, in a strictly calcium-dependent fashion, to DNA and chromatin (Pepys & Butler, 1987; Butler *et al.*, 1990), glycosaminoglycans (Hamazaki, 1987), agarose (Pepys *et al.*, 1977), fibronectin and C4-binding protein (de Beer *et al.*, 1981) and possibly other proteins. A notable property of SAP is its universal association with the amyloid deposits in systemic amyloidosis, Alzheimer's disease and the transmissible spongiform encephalopathies (reviewed by Pepys, 1994).

Abbreviations used: SAP, serum amyloid P component; CRP, C-reactive protein; MOβDG, methyl-4,6-O-(1-carboxyethylidene)-β-D-galactopyranoside; PE, O-phosphoethanolamine; dAMP, 2'-deoxyadenosine-5'-monophosphate.

The crystal structure of SAP crystallized at pH 5.5 revealed a pentamer, in which the subunits are arranged in a ring with 5-fold rotational symmetry (Emsley *et al.*, 1994). Each subunit has a two-layered β -sheet structure with a jellyroll topology (Emsley *et al.*, 1994). The pentamer has a polarity defined by five double-calcium binding sites on one face (B face) and five α -helices on the other (A) face. Analysis of a complex of SAP with MO β DG (3'-5' cyclic pyruvate acetal of galactose), a monosaccharide constituent of agarose that is bound with modest affinity by SAP (Hind *et al.*, 1984a), has shown that the pyruvate carboxylate group bridges the two resident calcium ions of SAP (Emsley *et al.*, 1994). The direct interaction of an acidic functional group with the calcium ions is likely to be a common feature of other SAP-ligand complexes.

The oligomerization properties of SAP are complex. At neutral pH in the presence of EDTA a stable decameric form exists (Pinteric *et al.*, 1976; Osmand *et al.*, 1977; Wood *et al.*, 1988) and we have speculated that SAP pentamers might interact through contacts between the A faces, thus maintaining accessibility of the calcium-binding sites for ligand recognition (Emsley *et al.*, 1994). In the presence of calcium and absence of ligand, isolated SAP rapidly precipitates (Baltz *et al.*, 1982) and electron microscopy shows long stacks of pentamers (Painter *et al.*, 1982), presumably resulting from a head-to-tail (AB) association. Both small and macromolecular ligands inhibit the calcium-induced aggregation of SAP (Hind *et al.*, 1984a; Hamazaki, 1989; Butler *et al.*, 1990), and this forms the basis of a rapid and efficient screening procedure for prospective ligands.

The principal motivation for the search of novel SAP ligands is the possible involvement of SAP in the pathogenesis of amyloid diseases. Pepys *et al.* have argued that a cloak of SAP covering the amyloid deposits may account for the persistence *in vivo* of these abnormal structures, and that unmasking the fibrils would expose them to the normal proteolytic clearance mechanisms (Hind *et al.*, 1984b; Tennent *et al.*, 1995; Pepys *et al.*, 1996). Millimolar concentrations of MO β DG are required to significantly displace SAP from amyloid fibrils *ex vivo* (Hind *et al.*, 1984b) and in an *in vitro* model system (Tennent *et al.*, 1995), leaving ample scope for the design of ligands with higher affinity for SAP.

Here, we describe the crystal structure at 2.8 Å resolution of a complex of SAP with 2'-deoxyadenosine-5'-monophosphate (dAMP). The compound was selected as an effective inhibitor of calcium-induced SAP precipitation, but also as a ligand of interest with respect to DNA recognition. The nucleotide ligand stabilizes a novel BB decameric form of the protein both in the crystal and in solution, in which the calcium sites are shielded against competing ligands.

Results

Choice of ligand

To identify novel ligands of SAP, we have screened various phosphorylated and sulphated mono- and disaccharides for their ability to inhibit the calcium-induced aggregation of SAP. Compounds were chosen either because of their structural resemblance to the established biological ligands of SAP, DNA and glycosaminoglycans, or because they had been shown to be weak competitors with respect to biological ligands (Loveless *et al.*, 1992). The mononucleotide dAMP was found to be as effective an inhibitor of calcium-induced SAP precipitation as MO β DG and O-phosphoethanolamine (PE), the only two small molecule compounds known to be bound by SAP with appreciable affinity (Pontet *et al.*, 1978; Hind *et al.*, 1984a; Hawkins *et al.*, 1991; Schwalbe *et al.*, 1992). Adenosine 5'-monophosphate is equally effective, demonstrating that a 2'-hydroxyl group is not detrimental. Interestingly, however, the adenine base was found to be essential for high-affinity binding; nucleotides containing guanine or pyrimidine bases are fairly poor ligands.

Decameric structure of the SAP-dAMP complex

Single crystals of SAP in the presence of dAMP and calcium ions could be obtained at pH 8.0 and the crystal structure of the complex was determined by molecular replacement at 2.8 Å resolution. The basic building unit in the crystals is an SAP decamer of point group symmetry 52 with a diameter of 95 Å and a height of 80 Å, in which two pentameric rings interact through their double-calcium ligand binding sites (Figure 1). The asymmetric unit content is one pentamer with subunits labelled A to E and the decamer is generated by a crystallographic dyad. There are surprisingly few contacts across the interface between the two pentamers (the "decamer interface"). The only interaction that can be directly inferred from the electron density map is the stacking of adenine bases of adjacent dAMP molecules (see below). This apart, the interface is conspicuously devoid of stabilizing interactions. Decamer formation results in the burial of only 1000 Å² (2.6%) of the pentamer solvent-accessible surface (calculated with a probe radius of 1.4 Å). There are substantial packing contacts between decamers in the *ab* directions of the tetragonal crystals, whereas packing along *c* involves only the apparently weak decamer interface, resulting in a markedly anisotropic diffraction limit of the crystals (see Materials and Methods).

Given the novel decameric structure of the SAP-dAMP complex in the crystals, we sought to confirm the existence of such a decamer in solution. We decided to use gel filtration chromatography, which has been used successfully in the past to establish the oligomeric state of SAP (Wood *et al.*, 1988). At pH 7.9 in the presence of EDTA, SAP

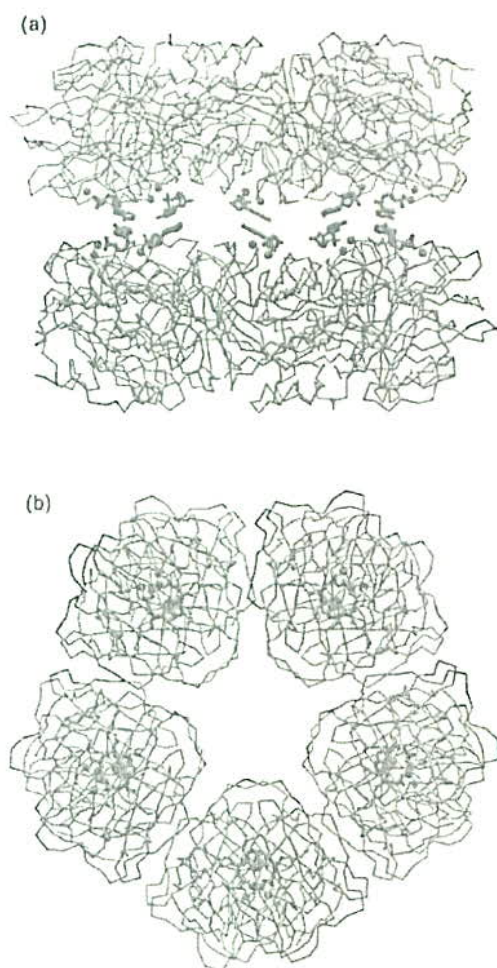


Figure 1. Two orthogonal views of the SAP-dAMP decamer. The two pentamers, shown in red and blue, respectively, are related by a crystallographic dyad. The ligand, dAMP, is shown as a stick model and calcium ions are shown as small spheres. The decamer is viewed along the crystallographic 2-fold axis in (a) and along the non-crystallographic 5-fold axis in (b). The Figure was produced with MOLSCRIPT (Kraulis, 1991).

eluted from a 22.0 ml Superose 12 column (Pharmacia) as a sharp peak at V_e 10.7 ml, corresponding to the decamer known to form under these conditions, whereas the V_e of CRP, a close relative of SAP that always exists as a pentamer of molecu-

lar mass 115 kDa, was 11.9 ml. SAP in the presence of dAMP and calcium ions eluted as a sharp peak at V_e 10.4 ml, demonstrating that a stable decamer exists under these conditions.

Structural details of the SAP-dAMP interaction

The SAP-dAMP complex structure was first refined to convergence without including any ligand molecules in the crystallographic model. Figure 2 shows the resulting difference Fourier map at sites A and E* (the asterisk denotes the crystallographically related subunit), which unambiguously defines the conformation of the dAMP ligand molecules and their interaction with SAP.

The 2'-deoxyribose ring is observed in the envelope C2'-endo conformation, the adenine base is in the anti conformation relative to the sugar moiety, and the torsion angles of the phosphate monoester side-chain are close to values commonly observed in mono- and oligonucleotides (Saenger, 1983). The ligand dAMP is thus bound to SAP in a low-energy conformation corresponding to the major conformer of this nucleotide in solution (Sarma & Dhirga, 1981).

The dAMP phosphate group bridges the two calcium ions of SAP with calcium-oxygen distances of 2.4 to 2.5 Å (Figure 3). One phosphate oxygen atom additionally receives a hydrogen bond from the side-chain of Asn59 ($d = 2.8$ Å). The ester oxygen atom is not involved in any interaction with the protein. The nucleotide sugar moiety interacts with SAP via two hydrogen bonds, one from the side-chain of Gln148 to the 2'-deoxyribose ring oxygen atom ($d = 2.8$ Å) and another from the side-chain of Tyr64, also to the ring oxygen atom ($d = 3.2$ Å). The side-chain of Tyr74, which together with Tyr64 is situated underneath the sugar ring, points towards O3' of dAMP, but the large distance between the tyrosine phenolic oxygen atom and O3' ($d = 3.4$ Å) argues against a hydrogen bond.

The amino group substituent at position 6 of the purine ring donates a hydrogen bond to the side-chain of Asp145 ($d = 3.0$ Å)[†]. Furthermore, the adenine amino group is close to the hydroxyl group of Ser147 ($d = 3.1$ Å), indicating the presence of an additional hydrogen bond. We note that residues Asp145 and Ser147 are contributed by the same loop segment that provides the calcium ligands Glu136, Asp138 and Gln148. The specific interactions made by the adenine ring of dAMP explain the poor affinities of SAP for nucleotides containing other bases. Guanine does not have a hydrogen-bond donor at position 6 and pyrimidines are too small to bridge the distance between the calcium ions and Asp145/Ser147.

Figure 2(b) shows a close-up view of the decamer interface in the vicinity of the ligand-binding sites. The dAMP ligands of adjacent, symmetry-related subunits interact directly through base stacking of their adenine rings, with a stacking separation of 3.5 Å. The phosphate groups of inter-

[†] The side-chain of Asp145 appears to adopt two alternative conformations. In subunits A and D, it hydrogen bonds with the amino group of dAMP, whereas it coordinates to calcium 2 in subunits B and E. The electron density for Asp145 in the crystallographically symmetric C-C* pair is weak, presumably due to static disorder between the two alternative conformations.

acting ligands are brought into close contact with each other (10 Å) and may interact either *via* water molecules or additional metal ions (Figure 2(a)). Apart from the base-stacking interaction, there is no hydrophobic contact stabilizing the decamer. Equally surprising, we do not see any clear indication of hydrogen bonds or salt-bridges linking the pentamers in the decamer. The side-chain of Lys143, if modelled in a fully extended conformation, would be able to donate a hydrogen bond to the carbonyl oxygen atom of Asp138*, but the lack of electron density for Lys143 beyond C γ makes this a speculative proposition. It is equally

likely that Lys143 interacts electrostatically with the dAMP* phosphate group without being fixed in a single conformation.

Discussion

Calcium-dependent binding of dAMP by SAP at physiological pH and ionic strength leads to the face-to-face association of two pentameric rings and the concomitant burial of the ligand-binding sites in the newly formed interface. This interface is unusually small and consists of five contact

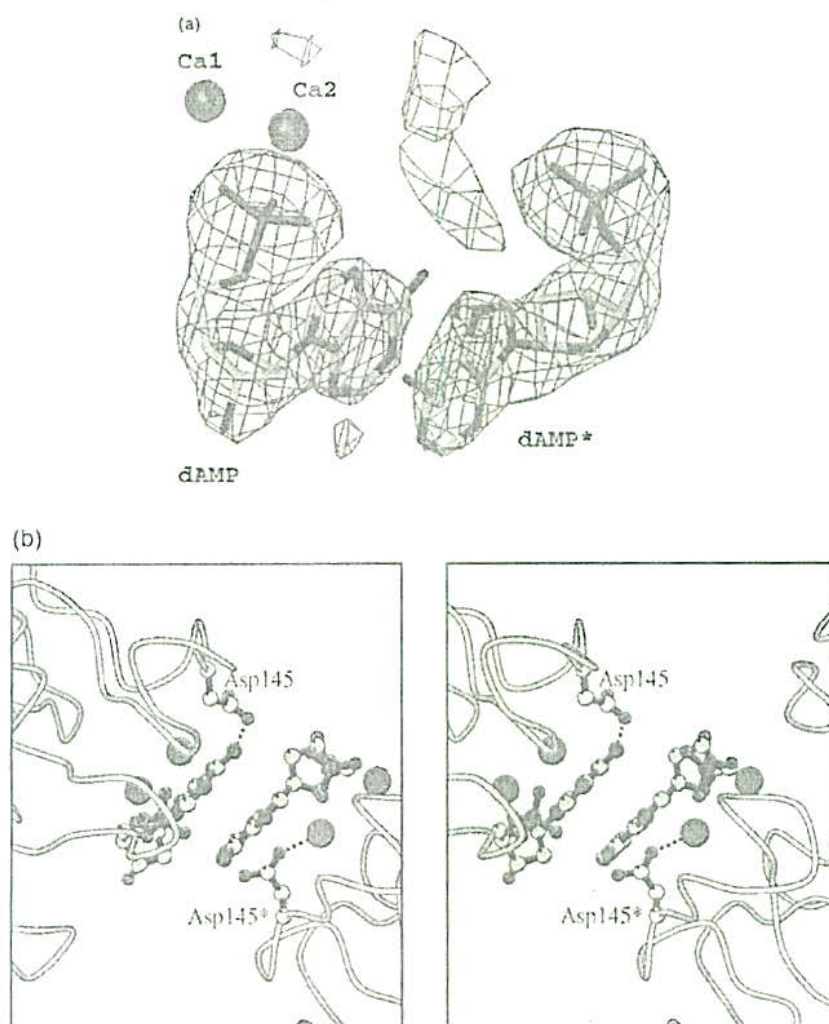


Figure 2. (a) Difference Fourier map at 2.8 Å resolution contoured at 3 σ and calculated prior to the inclusion into the model of the dAMP ligands. The binding sites shown are of subunits A and E*, and similar density is observed for the other sites. Superimposed on the map is the final atomic model, with the calcium ions in subunit A shown as pink spheres. The electron density halfway between the two phosphate groups may correspond to two water molecules or calcium ions. The Figure was produced with O (Jones *et al.*, 1991). (b) Stereoview of the contact region between SAP pentamers. Subunit A is in the upper left corner and subunit E* is in the lower right corner. The two alternative conformations of Asp145 are indicated. The Figure was produced with MOLSCRIPT (Kraulis, 1991).

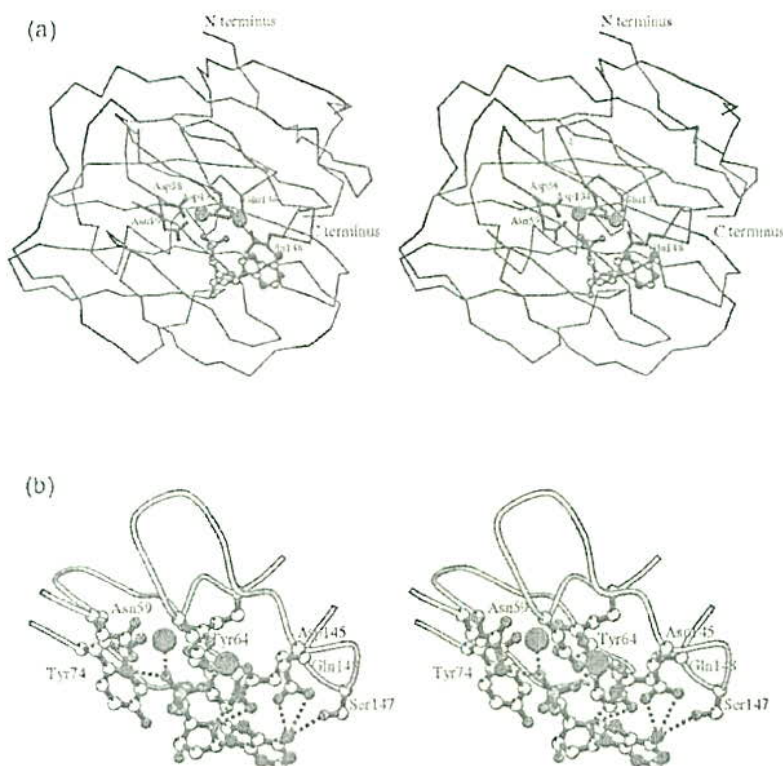


Figure 3. (a) Stereoview of one SAP subunit with bound dAMP. The termini of the polypeptide chain are marked. Side-chains coordinating to the two calcium ions are shown as stick models and have been labelled. The dAMP ligand is shown in atomic detail. (b) Stereoview of the ligand binding site showing the atomic interactions between SAP and dAMP. The two calcium ions are shown as green spheres and residues involved in ligand-binding have been labelled. The Figure was produced with MOLSCRIPT (Kraulis, 1991).

areas of only 200 Å² each. Nevertheless, the SAP-dAMP decamer is a stable entity in solution as demonstrated by gel filtration chromatography and electron microscopy (data not shown).

The stability of the decamer is likely to be the result of a highly cooperative process coupling ligand binding to decamerization. SAP binds monosaccharides with only moderate affinity, like the C-type lectins with which it shares many features of metal and ligand binding (Emsley *et al.*, 1994). High-affinity binding of macromolecular ligands is believed to result from the simultaneous interaction with more than one binding site on the oligomeric lectin (Rini, 1995). Similarly, the isologous association of two SAP pentamers creates a set of five, in this case strictly equivalent contact regions and may be accompanied by a significant gain in free energy even if the individual contributions are modest. The cooperativity of the process should be particularly high, because SAP binding to itself (i.e. decamerization), in contrast to SAP binding to its macromolecular ligands, involves the association of two rigid bodies with an exactly matching oligomeric arrangement.

A detailed analysis of the decamer interface shows that most, possibly even all, contacts between pentamers are mediated by the ligand dAMP, with the stacking of two adenine bases being the predominant interaction. The free energy of stacking two purine bases in aqueous solution has been estimated to be 2 to 6 kcal/mol (Friedman & Honig, 1995; Norberg & Nilsson, 1996). On decamerization, five such interactions are formed simultaneously and it does not seem unreasonable to ascribe much of the decamer stabilization to the stacking of the dAMP purine rings. The specific interactions made by the adenine amino group may aid in this process by fixing the purine ring in a single conformation.

Figure 4 shows the electrostatic properties of the decamer interface. Negative potential, probably originating from the dAMP phosphate groups, is concentrated at the contact sites. In the decamer, the phosphate moieties of interacting dAMP molecules are brought into close contact with each other and with Asp145 of the adjacent subunit (Figure 2(b)). This potentially unfavourable situation would be greatly eased by charge shielding or compensation through intervening water mol-

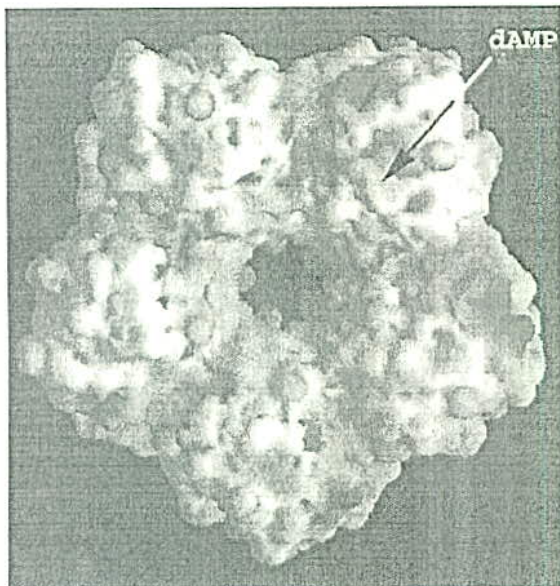


Figure 4. Electrostatic properties of the pentamer face carrying the ligand binding sites (B face) in the SAP-dAMP decamer. Negative and positive potential are indicated by red and blue colouring, respectively. The calcium ions and dAMP phosphate groups were included in the calculation with charges +2 and -2, respectively; the ionic strength was set to 0.15. The SAP pentamer is generally a very dipolar molecule, with a positive B face and a negative A face. In one subunit, an arrow marks the position of the dAMP adenine base, which becomes buried on decamer formation. The Figure was produced with GRASP (Nicholls *et al.*, 1991).

ecules or calcium ions, respectively. Indeed, we see residual difference electron density located between the phosphate groups, suggesting the presence of suitably positioned solvent molecules or metal ions (Figure 2(a)). The side-chain of Lys143, protruding from the ligand-binding site, could also make a contribution to the electrostatic stabilization of the decamer.

The interactions between SAP and the novel ligand dAMP can be thought to be assembled from the binding modes of MO β DG and PE (Emsley *et al.*, 1994). The hydrogen bonds to the sugar ring are shared with the former, while the bridging of two calcium ions by a phosphate group are reminiscent of the latter. In addition, dAMP uniquely interacts with two residues removed from the immediate binding site, Asp145 and Ser147. The calcium-mediated binding of a phosphorylated monosaccharide by SAP has, to our knowledge, no parallel in any other protein-sugar complex of known structure and is distinct from the ligand-binding modes seen in the C-type lectins (Rini, 1995).

Polymeric phosphorylated and sulphated sugars (in the form of DNA and glycosaminoglycans, respectively) are biological ligands of SAP (Pepys &

Butler, 1987; Butler *et al.*, 1990; Hamazaki, 1987). If one assumes that these macromolecular ligands use essentially the same set of interactions at the double calcium site of SAP as small ligands, one can attempt to construct models of DNA and glycosaminoglycan binding. The conformation of dAMP bound to SAP is rather similar to that of a nucleotide unit in B-DNA (Saenger, 1983). In our SAP-dAMP structure, extension at O3' is possible without causing steric clashes and, interestingly, places the phosphate diester group within hydrogen bonding distance of Tyr74 and Lys79, both located at the bottom of the binding site (Figure 3). In contrast, it is not possible to add another nucleotide unit at the dAMP phosphate group without invoking drastic deviations from any of the canonical DNA structures (Saenger, 1983; Dickerson, 1992). This can be easily understood by appreciating that in the SAP-dAMP complex the nucleotide base, which would be roughly normal to the helix axis in double-stranded DNA, lies flat against the SAP subunit. We have to conclude that the present structure offers little structural explanation for the avid binding by SAP of double-stranded nucleosomal DNA (Butler *et al.*, 1990). Binding of DNA may, in fact, involve regions other than the double-calcium site, in particular the basic grooves between subunits in the pentamers (Figure 4). In glycosaminoglycans, where the polymer chain is formed by glycosidic bonds, the sulphate groups are free to bridge the calcium ions at the SAP ligand-binding site and could therefore be bound directly.

As previously noted, the universal association of SAP with amyloid deposits has been recognized as a promising target for drug intervention in a range of human diseases (Tennent *et al.*, 1995; Pepys *et al.*, 1996). An important motivation of the present work was therefore the hope that new SAP ligands might prove valuable in drug design. What can be learned from the structure of the SAP-dAMP decamer? First, the new ligand dAMP utilizes interactions that could not have been deduced from existing structures of SAP-ligand complexes and further design will have to consider this. Second, and more important, the surprising observation of ligand-induced decamerization of SAP suggests a viable strategy to amplify even modest affinities of small ligands into a powerful overall effect by taking advantage of the high cooperativity of the process. Such a strategy appears particularly rewarding as the SAP ligand-binding site is both very shallow and hydrophilic, and does not lend itself easily to structure-based ligand design. Ligand-induced decamerization of SAP efficiently abolishes the availability of the double-calcium ligand-binding sites for other interactions. While this is clearly desirable for the pathological situation of SAP binding to amyloid fibrils, it remains to be seen whether any of the, as yet unknown, physiological function(s) of SAP are adversely affected by drug-induced decamerization.

Materials and Methods

Crystallization

Purified human SAP was obtained as described (Wood *et al.*, 1988). Crystals in the presence of dAMP were grown by vapour diffusion of hanging drops at 4°C: 3 µl of protein solution (20 mg/ml SAP in 10 mM Tris-HCl (pH 8.0), 140 mM NaCl, 20 mM dAMP, 0.02 % (w/v) NaN₃) were mixed with an equal volume of reservoir solution (0.1 M bis-Tris propane-HCl (pH 8.0), 12 to 14 % (w/v) polyethylene glycol 4000, 20 mM CaCl₂) and equilibrated against 1 ml of the latter solution. Crystals appear after two to three days and continue to grow for several weeks. They form as very regular tetragonal bipyramids and frequently attain maximum dimensions of 0.8 mm × 0.8 mm × 0.4 mm.

Data collection

Diffraction data were collected at 4°C with a 30 cm MAR image plate detector on the wiggler beamline 9.6 at the Daresbury synchrotron radiation source (λ = 0.88 Å). The crystals were found to belong to the tetragonal system with unit cell constants *a* = *b* = 190.03 Å, *c* = 119.74 Å, α = β = γ = 90°. The diffraction limit of the crystals is highly anisotropic: diffraction extends to at least 2.8 Å resolution along *a** and *b**, but is weak beyond ≈3.5 Å along *c**. The diffraction data were processed, scaled and merged with MOSFLM (Leslie, 1994) and programs of the CCP4 program suite (Collaborative Computing Project No. 4, 1994). This was first done in space group *P*4 to determine the Laue group as *P*4/*mmm*, and then repeated in space group *P*4₂. A total of 366,263 observations were merged to yield 53,133 independent reflections with an overall *R*_{merge} on intensities of 0.085. The *R*_{merge} in the outermost resolution shell (2.90 to 2.80 Å) is 0.443, which was deemed acceptable considering the almost sevenfold data redundancy and the weakness of reflections with large *l* indices due to the anisotropic diffraction limit. The data set covers 98.0% of all possible reflections between 30 and 2.8 Å resolution. Because no axial reflections with indices 00*l* were recorded, the space group of the crystals could not be unambiguously assigned. Systematic absences along *h*00 were indicative of a 2₁ screw axis, limiting the possible space groups to *P*4₂. Data collection and reduction statistics are summarized in Table 1.

Structure solution and refinement

The asymmetric unit volume can accommodate either one or two SAP pentamers of molecular mass 127.5 kDa, resulting in a calculated solvent content (Matthews, 1968) of 71% or 43%, respectively. Given the relatively poor diffraction limit of the crystals, we favoured the higher solvent content associated with an asymmetric unit containing a single pentamer. The self-rotation function indicated the presence of a 5-fold non-crystallographic symmetry (NCS) axis slightly tilted relative to the tetragonal axis and oriented perpendicular to the crystallographic dyad along [110].

The structure of the new tetragonal crystal form was solved by molecular replacement with the program AMoRe (Navaza, 1994). The search model was the structure of pentameric SAP crystallized at pH 5.5 and solved at 2.0 Å resolution (PDB entry code 1SAC; Emsley *et al.*, 1994). A cross-rotation function calculated with data

Table 1. Data collection and refinement statistics

A. Data Collection	
Space group	<i>P</i> 4 ₂ ,2
Cell constants (Å)	<i>a</i> = <i>b</i> = 190.03, <i>c</i> = 119.74
Number of crystals	3
Resolution range (Å)	30 – 2.8
Number of observations	366,263
Number of independent reflections	53,133
<i>R</i> _{merge} ^a	0.085 (0.443) ^b
(<i>I</i> /σ(<i>I</i>))	6.8 (1.4) ^b
Completeness (%)	98.0 (97.5) ^b
B. Refinement	
Resolution range (Å)	8 – 2.8
Amplitude cut-off	none
Number of reflections in working set	48,167
Number of reflections in test set	2478
Number of atoms contributing to <i>F</i> _{calc}	8253
Number of NCS-restrained atoms	7665
<i>R</i> _{cryst} ^c	0.232
<i>R</i> _{free} ^d	0.252
R.m.s. bonds (Å)	0.009
R.m.s. angles (deg.)	1.6
R.m.s. deviation of NCS-related atoms (Å)	0.07
R.m.s. Δ <i>B</i> of bonded atoms (Å ²)	1.5
R.m.s. Δ <i>B</i> of NCS-related atoms (Å ²)	1.1

^a *R*_{merge} = ΣΣ_{*h*} |(*I*(*h*)) - *I*(*h*)| / ΣΣ_{*h*} *I*(*h*), where *I*(*h*)_{*i*} are the individual observations and (*I*(*h*)) is the mean intensity of reflection *h*.

^b The value in parentheses is for the highest resolution shell (2.90 to 2.80 Å).

^c *R*_{cryst} = Σ |*F*_{obs} - *F*_{calc}| / Σ *F*_{obs}, where *F*_{obs} and *F*_{calc} are the observed and calculated structure factor amplitudes, respectively.

^d *R*_{free} = Σ |*F*_{obs} - *F*_{calc}| / Σ *F*_{obs} evaluated for 5% of the diffraction data not included in the refinement. The test set was chosen in 15 thin shells of resolution.

from 15 to 4 Å resolution and an integration radius of 40 Å yielded five solutions related by the correct NCS and clearly separated from the background (8.0 to 8.5 σ, next highest peak 4.0 σ). Translation functions (8 to 4 Å) were calculated in all four space groups *P*4₂. A solution was only obtained in space group *P*4₂.2 (*R* = 0.370, correlation = 0.662). This solution resulted in a favourable crystal packing, in which an SAP decamer of 52 symmetry is generated from the unique pentamer by the dyad along [110]. A difference Fourier map calculated prior to any refinement shows five strong maxima coinciding with the expected positions of phosphate groups at the double-calcium ligand-binding sites.

Refinement was carried out with X-PLOR (Brünger, 1992a) using the Engh & Huber (1991) parameter set and all data with *F* ≥ 0 between 8 and 2.8 Å resolution. The modest resolution of the data, exacerbated by the poor diffraction along *c**, prompted us to rely heavily on the free *R*-factor (*R*_{free}; Brünger, 1992b) during refinement. The 5-fold NCS was exploited whenever possible, which, together with the high solvent content of the crystals, guaranteed a favourable observation-to-parameter ratio throughout the entire refinement.

After a round of conventional positional refinement with tight NCS restraints on the entire subunit structure, an anisotropic temperature factor correction was applied to the observed amplitudes (*B*₁₁ = *B*₂₂ = 16.2 Å², *B*₃₃ = -32.5 Å²), which greatly improved the appearance of electron density maps. A rebuild of the model showed that several regions involved in crystal packing contacts should not be restrained to obey the NCS, which even-

tually led to the exclusion from the NCS definition of 7% of the subunit atoms. Further refinement, now with tightly NCS-restrained individual B-factors, lowered the R-factor to 0.251 ($R_{\text{free}} = 0.269$) for a model with excellent stereochemistry. A difference Fourier map calculated at this stage allowed the unambiguous placement of five dAMP molecules. A final round of refinement resulted in an R-factor of 0.232 ($R_{\text{free}} = 0.252$). A further reduction to $R = 0.213$ is possible by releasing the NCS restraints. However, as this is accompanied by an increase of R_{free} to 0.260, we decided to retain the more conservative, NCS-restrained model. Refinement statistics are summarized in Table 1. The atomic coordinates of the SAP-dAMP complex and experimental structure factor amplitudes have been deposited with the Brookhaven Protein Data Bank (entry code 1LGN).

Gel filtration chromatography

Gel filtration analysis was carried out using a Pharmacia fast protein liquid chromatography (FPLC) system. Typically, 200 μ l of protein solution (1 to 2 mg/ml SAP or CRP in 10 mM Tris-HCl (pH 8.0), 140 mM NaCl containing either 10 mM EDTA or 5 mM dAMP and 7 mM CaCl_2) were applied to a 22.0 ml Superose 12 column (Pharmacia), previously equilibrated with at least four column volumes of the respective buffer solution. The flow-rate was 0.2 ml/minute. Elution was monitored by measuring absorption at 280 nm. The method yields very reproducible results and the standard errors of quoted elution volumes are less than 0.05 ml, as estimated from experiments carried out in triplicate.

Acknowledgements

This work was supported in part by program and project grants from the Medical Research Council to M. B. P. and S. P. W.. E. H. acknowledges receipt of fellowships from the European Molecular Biology Organization and the Human Frontier Science Program Organization.

References

- Baltz, M. L., de Beer, F. C., Feinstein, A. & Pepys, M. B. (1982). Calcium-dependent aggregation of human serum amyloid P component. *Biochim. Biophys. Acta*, **701**, 229–236.
- Brünger, A. T. (1992a). *X-PLOR Version 3.1: A System for Crystallography and NMR*, Yale University Press, New Haven, CT.
- Brünger, A. T. (1992b). The free R value: a novel statistic quantity for assessing the accuracy of crystal structures. *Nature*, **355**, 472–474.
- Butler, P. J. G., Tennent, G. A. & Pepys, M. B. (1990). Pentraxin-chromatin interactions: serum amyloid P component specifically displaces H1-type histones and solubilizes native long chromatin. *J. Expt. Med.* **172**, 13–18.
- Collaborative Computing Project No. 4, (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
- de Beer, F. C., Baltz, M. L., Holford, S., Feinstein, A. & Pepys, M. B. (1981). Fibronectin and C4-binding protein are selectively bound by aggregated amyloid P component. *J. Expt. Med.* **154**, 1134–1149.
- Dickerson, R. E. (1992). DNA structure from A to Z. *Methods Enzymol.* **211**, 67–111.
- Emsley, J., White, H. E., O'Hara, B. P., Oliva, G., Srinivasan, N., Tickle, I. J., Blundell, T. L., Pepys, M. B. & Wood, S. P. (1994). Structure of pentameric human serum amyloid P component. *Nature*, **367**, 338–345.
- Engh, R. A. & Huber, R. (1991). Accurate bond and angle parameters for X-ray protein structure refinement. *Acta Crystallog. sect. A*, **47**, 392–400.
- Friedman, R. A. & Honig, B. (1995). A free energy analysis of nucleic acid base stacking in aqueous solution. *Biophys. J.* **69**, 1528–1535.
- Gewurz, H., Zhang, X.-H. & Lint, T. F. (1995). Structure and function of the pentraxins. *Curr. Opin. Immunol.* **7**, 54–64.
- Hamazaki, H. (1987). Ca^{2+} -mediated association of human serum amyloid P component with heparan sulfate and dermatan sulfate. *J. Biol. Chem.* **262**, 1456–1460.
- Hamazaki, H. (1989). Calcium-dependent polymerization of human serum amyloid P component is inhibited by heparin and dextran sulfate. *Biochim. Biophys. Acta*, **998**, 231–235.
- Hawkins, P. N., Tennent, G. A., Woo, P. & Pepys, M. B. (1991). Studies *in vivo* and *in vitro* of serum amyloid P component in normals and in a patient with AA amyloidosis. *Clin. Exp. Immunol.* **84**, 308–316.
- Hind, C. R. K., Collins, P. M., Renn, D., Cook, R. B., Caspi, D., Baltz, M. L. & Pepys, M. B. (1984a). Binding specificity of serum amyloid P component for the pyruvate acetal of galactose. *J. Expt. Med.* **159**, 1058–1069.
- Hind, C. R. K., Collins, P. M., Caspi, D., Baltz, M. L. & Pepys, M. B. (1984b). Specific dissociation of fibrillar and non-fibrillar components of amyloid deposits. *Lancet*, **ii**, 376–378.
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallog. sect. A*, **47**, 110–119.
- Kraulis, P. J. (1991). MOLSCRIPT: a program to produce both detailed and schematic plots of protein structures. *J. Appl. Crystallog.* **24**, 946–950.
- Leslie, A. G. W. (1994). *MOSFLM Users Guide*, MRC-LMB, Cambridge.
- Loveless, R. W., Floyd-O'Sullivan, G., Raynes, J. G., Yuen, C.-T. & Feizi, T. (1992). Human serum amyloid P component is a multispecific adhesive protein whose ligands include 6-phosphorylated mannose and the 3-sulphated saccharides galactose, N-acetylglucosamine and glucuronic acid. *EMBO J.* **11**, 813–819.
- Matthews, B. W. (1968). Solvent content of protein crystals. *J. Mol. Biol.* **33**, 491–497.
- Navaza, J. (1994). AMoRe: an automated package for molecular replacement. *Acta Crystallog. sect. A*, **50**, 157–163.
- Nicholls, A., Sharp, K. A. & Honig, B. (1991). Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins: Struct. Funct. Genet.* **11**, 281–296.
- Norberg, J. & Nilsson, L. (1995). Stacking free energy profiles for all 16 natural ribodinucleoside monophosphates in aqueous solution. *J. Am. Chem. Soc.* **117**, 10832–10840.
- Osmand, A. P., Friedenson, B., Gewurz, H., Painter, R. H., Hoffman, T. & Shelton, E. (1977). Characteriz-

- ation of C-reactive protein and the complement sub-component C1t as homologous proteins displaying cyclic pentameric symmetry (pentraxins). *Proc. Natl Acad. Sci. USA*, **74**, 739–743.
- Painter, R. H., de Escallon, I., Massey, A., Pinteric, L. & Stern, S. B. (1982). The structure and binding characteristics of serum amyloid protein (9.55 a1-glycoprotein). *Ann. N.Y. Acad. Sci.* **389**, 199–213.
- Pepys, M. B. (1994). Amyloidosis. In *Santer's Immunological Disease* (Frank, M. M., Austen, K. F., Claman, H. N. & Unanue, E. R., eds), 5th edit., pp. 637–655, Little, Brown and Company, Boston, MA.
- Pepys, M. B. & Baltz, M. L. (1983). Acute phase proteins with special reference to C-reactive protein and related proteins (pentraxins) and serum amyloid P component. *Advan. Immunol.* **34**, 141–212.
- Pepys, M. B. & Butler, P. J. G. (1987). Serum amyloid P component is the major calcium-dependent specific DNA binding protein of the serum. *Biochem. Biophys. Res. Commun.* **148**, 308–313.
- Pepys, M. B., Dash, A. C., Munn, E. A., Feinstein, A., Skinner, M., Cohen, A. S., Gewurz, H., Osmand, A. P. & Painter, R. H. (1977). Isolation of amyloid P component (protein AP) from normal serum as a calcium-dependent binding protein. *Lancet*, **i**, 1029–1031.
- Pepys, M. B., Dash, A. C., Fletcher, T. C., Richardson, N., Munn, E. A. & Feinstein, A. (1978). Analogues in other mammals and in fish of human plasma proteins C-reactive proteins and amyloid P component. *Nature*, **273**, 168–170.
- Pepys, M. B., Rademacher, T. W., Amatayakul-Chantler, S., Williams, P., Noble, G. E., Hutchinson, W. L., Hawkins, P. N., Nelson, S. R., Gallimore, R. J., Herbert, J., Hutton, T. & Dwek, R. A. (1994). Human serum amyloid P component is an invariant constituent of amyloid deposits and has a uniquely homogeneous glycostructure. *Proc. Natl Acad. Sci. USA*, **91**, 5602–5606.
- Pepys, M. B., Tennent, G. A., Booth, D. R., Bellotti, V., Lovat, L. B., Tan, S. Y., Persey, M. R., Hutchinson, W. L., Booth, S. E., Madhoo, S., Soutar, A. K., Hawkins, P. N., Van Zyl-Smit, R., Campistol, J. M., Fraser, P. E., Radford, S. E., Robinson, C. V., Sunde, M., Serpell, L. C. & Blake, C. C. F. (1996). Molecular mechanisms of fibrillogenesis and the protective role of amyloid P component: two possible avenues for therapy. In *The Nature and Origin of Amyloid Fibrils* (Ciba Found. Symp. 199), pp. 73–81, Wiley, Chichester.
- Pinteric, L., Assimeh, S. N., Kells, D. I. C. & Painter, R. H. (1976). The ultrastructure of C1t, a subcomponent of the first component of complement: an e.m. and ultracentrifuge study. *J. Biol. Chem.* **117**, 79–83.
- Pontet, M., Engler, R. & Jayle, M. F. (1978). One-step preparation of both human C-reactive protein and C1t. *Fed. Eur. Biol. Soc. Letters*, **88**, 172–178.
- Rini, J. M. (1995). Lectin structure. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 551–577.
- Sarma, R. H. & Dhingra, M. M. (1981). NMR studies of oligonucleotides in solution. In *Topics in Nucleic Acid Structure* (Neidle, S., ed.), pp. 33–63, Macmillan, London.
- Saenger, W. (1983). *Principles of Nucleic Acid Structure*, Springer, New York.
- Schwalbe, R. A., Dahlback, B., Coe, J. E. & Nelsestuen, G. L. (1992). Pentraxin family of proteins interact specifically with phosphocholine and/or phosphorylethanolamine. *Biochemistry*, **31**, 4907–4915.
- Tennent, G. A., Lovat, L. B. & Pepys, M. B. (1995). Serum amyloid P component prevents proteolysis of the amyloid fibrils of Alzheimer's disease and systemic amyloidosis. *Proc. Natl Acad. Sci. USA*, **92**, 4299–4303.
- Wood, S. P., Oliva, G., O'Hara, B. P., White, H. E., Blundell, T. L., Perkins, S. J., Sardharwalla, I. & Pepys, M. B. (1988). A pentameric form of human serum amyloid P component. Crystallization, X-ray diffraction and neutron scattering studies. *J. Mol. Biol.* **202**, 169–173.

Edited by A. Klug

(Received 18 November 1996; received in revised form 11 March 1997; accepted 4 April 1997)

Comparative analyses of pentraxins: implications for protomer assembly and ligand binding

N Srinivasan¹, Helen E White¹, Jonas Emsley¹, Steve P Wood¹,
Mark B Pepys² and Tom L Blundell^{1*}

¹Laboratory of Molecular Biology and ICRF Unit of Structural Molecular Biology, Department of Crystallography, Birkbeck College, University of London, Malet Street, London WC1E 7HX, UK and ²Immunological Medicine Unit, Department of Medicine, Royal Postgraduate Medical School, Hammersmith Hospital, London W12 0NN, UK

Background: Pentraxins are a family of plasma proteins characterized by their pentameric assembly and calcium-dependent ligand binding. The recent determination of the crystal structure for a member of this family, human serum amyloid P component (SAP), provides a basis for the comparative analysis of the pentraxin family.

Results: We have compared the sequences, tertiary structures and quaternary arrangements of SAP with human C-reactive protein (CRP), Syrian hamster SAP (HSAP) and *Limulus polyphemus* CRP (LIM). These proteins can adopt a β -jelly roll topology and a

hydrophobic core similar to that seen in SAP. Only minor differences are observed in the positions of residues involved in coordinating calcium ions.

Conclusions: Calcium-mediated ligand binding by CRP, HSAP and LIM is similar to that defined by the crystal structure of SAP, but sequence differences in the hydrophobic pocket explain the differential ligand specificities exhibited by the homologous proteins. Differences elsewhere, including insertions and deletions, account for the different (hexameric) quaternary structure of LIM.

Structure 15 November 1994, 2:1017-1027

Key words: acute phase response, C-reactive protein, hamster SAP, *Limulus* CRP, subunit aggregation

Introduction

The pentraxins are a family of oligomeric plasma proteins with the capacity for specific calcium-dependent ligand binding and named for the pentameric arrangement of their protomers [1,2]. They have been conserved throughout vertebrate evolution with retention of ligand specificity, subunit arrangement, and substantial amino acid sequence similarity [3-5]. Their production can be modulated by cytokines, including interleukin (IL)-1, IL-6 and tumour necrosis factor (TNF) [6], and they participate in inflammation and host defence, although they may also have other functions [2,7-9]. The phylogenetically oldest member of the family, the C-reactive protein of an arachnid, *Limulus polyphemus* (the horseshoe crab), has a hexameric rather than a pentameric structure [10].

The human plasma pentraxins are C-reactive protein (CRP), the classical acute phase reactant, and serum amyloid P component (SAP), which is also a normal extracellular matrix protein and a universal constituent of the tissue deposits in the pathological condition of amyloidosis [2]. Human CRP and SAP are clearly products of gene duplication and pentraxins in other vertebrates generally resemble more closely one or other of CRP or SAP. However, there is much overlap in precise ligand specificity, behaviour as acute phase proteins, and deposition in amyloid. Another human protein, PTX3 [11] or TSG-14 [12], with some sequence similarity to the pentraxins has lately been described but has not yet been well characterized, and it is not known whether it shares

the hallmark features of pentameric assembly or calcium-dependent ligand binding.

The calcium-dependent ligand specificity of human CRP for phosphocholine (PC) [13] is of particular interest as it is not shared by human SAP (WL Hutchinson and MB Pepys, unpublished data), although they both bind to phosphoethanolamine (PE) [14,15]. On the other hand, human SAP specifically binds to the 4,6 cyclic pyruvate acetal of β -D-galactose (MOBDG) [16] and also to all types of amyloid fibrils [17], which are not recognized by human CRP. The major pentraxin in Syrian hamsters, formerly known as female protein [18] (here designated HSAP), is the hamster counterpart of human SAP, in that it binds to hamster amyloid [19,20], but it shares with human CRP the capacity to bind to PC as well as PE [18,21].

Recently, the three-dimensional (3D) structure of human SAP has been elucidated at high resolution by X-ray analysis [22]. The protomer has a flattened β -jelly roll topology which is remarkably similar to legume lectins such as concanavalin A [23,24] and pea lectin [25], bacterial 'lectins' such as glucanase [26], endo 1,4- β -xylanase II [27] and cellobiohydrolase I [28], human S-Lac lectin [29] and vertebrate lectin [30]. However, the metal ions and carbohydrate ligand binding sites in SAP and legume lectins are located in different positions on the common fold and the quaternary structures are different (N Srinivasan *et al.*, & TL Blundell, unpublished data). Crystal structures of SAP complexes show

*Corresponding author.

that SAP binding to MO β DG and PE is mediated by two calcium ions [22].

In view of the high degree of sequence similarity between human SAP, human CRP and pentraxins in other species it should be possible to identify structure-function relationships in these other proteins on the basis of the human SAP structure. Using the comparative modelling approach encoded in COMPOSER [31-35] we have built 3D models for the sequences of human CRP [36,37], HSAP [38] and *Limulus* CRP (LIM) [39,40] on the basis of the crystal structure of human SAP. We have constructed pentamers of CRP and HSAP based on the pentamer defined in the crystal structure of SAP and identified the interactions that stabilize subunit aggregation. By operating a six-fold axis, whose position is determined empirically, we have generated a model of the hexamer of LIM and analyzed inter-protomer interactions. Extrapolation of ligand binding modes of SAP to the 3D structure of CRP suggests explanations for the relative affinities of CRP for PC, PE and MO β DG. These comparative analyses of structure and function of the pentraxin family provide a basis for the design of specific ligands of high affinity.

Results and discussion

Comparative analysis of the pentraxin protomers

Fig. 1 shows the alignment of the sequence of SAP with those of CRP, HSAP and LIM; sequence identities with SAP are 52%, 67% and 30%, respectively. Sequences of CRP, HSAP and LIM can be modelled on the fold of SAP with the β -strands and the α -helix conserved. The root mean square deviations (rmsds) between topologically equivalent C α atoms of SAP and the various models are 0.15 Å (CRP), 0.06 Å (HSAP) and 0.65 Å (LIM), with the largest differences occurring as expected in the loop regions. Fig. 2 compares the tertiary structure of SAP defined by X-ray analysis, with those of CRP, HSAP and LIM defined by comparative modelling. Regions with significantly different amino acids relative to SAP are highlighted. Most amino acid substitutions occur in loops, with some in β -sheets, especially in LIM.

CRP versus SAP

The insertion of Ser44A-Thr44B (SAP numbering) in CRP relative to SAP has been modelled by replacing ELSSTR (42-45) with the segment 411-416 of taka amylase (PDB entry code 2TAA). This segment was chosen based on a search for suitable segments in a database of known structures. The insertion region 42-45 of CRP is spatially proximal to another region, Pro89-Val90A which differs from SAP. Pro89 of SAP is in the *cis* conformation, but no equivalent proline is present in CRP. Hence the region Val87-Thr88-Val89 of CRP has been modelled, after a general search of known 3D structures, on 52-54 of subtilisin *Novo* (PDB entry code 2SNI).

Most of the more radical sequence differences in CRP compared with SAP appear to stabilize the structure. For example, the hydrophobic Val16, Phe124 and Leu170, which are surprisingly exposed in SAP, are replaced in CRP by the more hydrophilic residues aspartate, threonine and asparagine, respectively. Thr24, which is buried and hydrogen bonded in SAP, is replaced by alanine in CRP and is in an apolar environment provided by Leu26, Val182 and Leu183. The side chain of Thr90 is hydrogen bonded to that of Asp70, thereby replacing the hydrophobic interaction between Phe90 and Val70 in SAP. Other differences in interactions that may stabilize the structure of the protomer are listed in Table 1.

HSAP versus SAP

There are two regions in HSAP which significantly differ from SAP, one corresponding to the *cis*-proline region of SAP and the other corresponding to the eight-residue carboxy-terminal extension. As in the case of CRP, the former region of HSAP was modelled on a segment from 2SNI. A seven-residue overlap with the preceding conserved region was used to identify the segment, starting from Thr233 of the B-chain of rat mast cell proteinase (PDB entry code 3RP2), as the most suitable carboxy-terminal extension.

Despite high sequence identity between HSAP and SAP, there are some amino acid substitutions that result in differences in interactions between residues (Fig. 1 and Table 1). For example, although Val16 of SAP is replaced by glutamate in HSAP, the side chain of glutamate can form a hydrogen bond to the carbonyl oxygen at Phe144. Glu70 of HSAP is involved in two ion pairs; one with Lys81 and the other with His83 which are threonine and lysine, respectively, in SAP. Phe144 in HSAP packs with the Tyr19, which is a histidine in SAP. Side chains of Asn21 and Arg193 interact in SAP and these residues are changed to lysine and asparagine, respectively, in HSAP.

LIM versus SAP

LIM has many insertions and deletions, reflecting its more distant relationship with SAP. The nine regions with most structural variation from SAP were successfully modelled on the basis of fragments from proteins of known 3D structures in the Brookhaven Protein Data Bank [41] and are listed in Table 2. Amino acid substitutions in LIM relative to SAP occur throughout the molecule, including the β -sheets forming the jelly roll fold. These differences alter the packing and nature of the interactions between the side chains, but they allow the overall 3D structure to be maintained. For example, the interaction between two hydrophobic residues Phe8 and Leu188 in the β -strands in SAP is replaced by a polar-polar interaction between Lys9 and Thr188 in LIM. The interaction between the side chains of Lys65 and Tyr71 in SAP is replaced by a hydrophobic contact between Ile65 and Phe71 in LIM. The interaction between the polar side chains Glu66 and Ser72 in SAP is replaced by the hydrophobic interaction between Leu64 and Leu72 in LIM. In SAP, the side chain of Ser41 is

Fig. 1. Sequence alignment of SAP, CRP, HSAP and LIM. Residues inserted relative to SAP, for example serine and threonine in CRP between residues 44 and 45, are identified as Ser44A and Thr44B. The structural features of SAP are indicated using the program JOY [55]. Key to JOY alignments: solvent inaccessible, UPPER CASE; solvent accessible, lower case; positive ϕ , *italic*; hydrogen bond to other side chain, tilde; hydrogen bond to main chain amide, frame box; hydrogen bond to main chain carbonyl, underline; disulphide bond, cedilla; positive ϕ , ϕ ; β -strand, β ; cis-peptide, ω ; 3_{10} -helix, 3; α -helix, α ; inter-protomer contact residue in SAP, p; calcium binding residue in SAP, c; hydrophobic pocket residue in SAP, h.

Fig. 2. The folds of the pentraxins SAP, CRP, HSAP and LIM. The crystal structure of a subunit of SAP [22] is shown in pink at the top left. Roughly every thirtieth residue is labelled (see Fig. 1 for sequence information). Calcium ions are shown in red and the disulphide bond is in yellow. Models of CRP, HSAP and LIM, are shown with regions of significant residue differences, relative to SAP, shown in green. (Figure generated using SETOR [53].)

The two inserted residues at 44A and 44B of CRP (see legend to Fig. 1 for explanation of nomenclature) are situated in the inter-protomer interface and interact with the loops containing Asp68 and Phe88, resulting in more inter-protomer contacts than in SAP. This region is more polar than in SAP where we find Val68 and Phe88. However, most of the inter-protomer interactions in SAP are conserved in CRP (Table 3). For example, Lys116 and Glu153 form an ion pair across the interface

Table 1. Predicted differences in interactions within a protomer.	
Residue or residue pair in:	Interacting groups in: ^a
SAP	CRP
Ile23-Glu191	Lys23-Glu191
Asn32-Ser82	Ala32-Phe82
Asn53-Tyr140	Ala53-Phe140
Lys65-Val92	Trp65-Val92
Val68-Phe88	Asp68-Thr88 & Thr448
Gly69-Val84	Ile69-Val84
Val70-Phe90	Asp70-Thr90
Val96	Thr96 H-bonded to MC at Ser97
Leu114-Ile109	Arg114-Tyr71
Non-H-bonded Lys143	Asn143-Ser139
SAP	HSAP
Val16-Phe144	Glu16 H-bonded to MC at Phe144
His19-Phe144	Tyr19-Phe144
Asn21-Arg193	Lys21-Asn193
Glu70-Thr81 & Lys83	Glu70-Lys81 & Arg83
Ser82-Leu114	Val82-Trp114
Val96	Thr96 H-bonded to MC at Val95
Leu170	Lys170 H-bonded to MC at Gly111
SAP	LIM
Phe8-Leu188	Lys9-Thr188
Lys65-Tyr71	Ile65-Phe71
Glu66-Ser72	Leu64-Leu72
Phe11-Ile154	Pro11-Leu154
Val196-Val20 & Ile192	Val196-Pro20 & Ala192
Ser41 H-bonded to MC at Val151	Val41-Leu150
Asn53-Asp58	Ala53-Val140
Val63-Ser82	Thr63-Gln82
Ala108-Ile169	Trp108-Ile169
Trp160-Ala172 & Tyr158	Trp160-Leu172 & Tyr173

^aUnless specified the interactions are between side chains. MC = main chain.

Table 2. Modelling of structurally variable regions (SVRs) in LIM.	
SVRs of LIM ^a	Modelled on: ^b
- 2 to - 1	2GDI Ala O 0
11 to 13	2TBV Asp B 217
29 to 30	4TMS Arg 178
42A to 44	1WSY Ala B 269
75A to 78	4APE Gly 21
85 to 91A	1LZ1 Gly 55
114 to 116	1BMV Gly 1 1078
174 to 180	2GDI Asp P 293
204 to 204A	2CRO Gly 62

^aResidue numbering is based on SAP (see Fig. 1). ^bThe protein fragments used for modelling are derived from proteins of known structure. Their PDB codes are given along with the first residue number of the chosen segment. Chain identifiers in the structures with more than one polypeptide chain in the PDB file are specified before the residue number.

in SAP that is retained in CRP by the replacements Lys→Arg and Glu→Asp, respectively.

Calcium binding occurs on one side of the pentamer, on the surface composed of the more concave β-sheets. On the opposite more planar face, each subunit has an

α-helical segment [22]. The existence of SAP in pentameric or decameric form is dependent on pH and the nature and concentration of buffer salts [42] (WL Hutchinson and MB Pepys, unpublished data). The decamer is thought to be formed by packing two pentamers in such a way that the faces containing helices are packed with an angle between the helices from different pentamers; in this arrangement the calcium-containing β-sheets of each pentamer are accessible [22]. The charge distribution at the helix-containing face of the SAP pentamer is shown in Fig. 3a. Both positive and negative charges permit electrostatic interactions between the two pentamers of the decamer. The presence of more negative than positive charges means that not all charges are paired and is indicative of a weak association between the two pentamers. Fig. 3b shows the charge distribution in the helix-containing face of the CRP pentamer. A conspicuous feature is the cluster of strong negative charge at the centre of the pentameric ring. This distribution of charge would lead to repulsion of pentamers in a putative decamer similar to that of SAP. Such repulsion could not be eliminated by the rotation of one of the pentamers with respect to the other as the strong negative charge is evenly spread at the centre. This may explain the observation that CRP does not form a decamer [43].

Most inter-protomer contacts, including ion pairs and hydrophobic contacts, are similar in HSAP and SAP (Table 3). For example, substitution of Ile85 in SAP by valine in HSAP allows retention of the hydrophobic contact with the conserved Phe88. Subunit interactions in HSAP are increased by the presence of a carboxy-terminal extension. Fig. 4 shows the superposition of SAP and HSAP in this inter-protomer contact region.

LIM forms hexamers rather than pentamers [10], which we have modelled as in Fig. 5. The inter-protomer contacts in the hexamer of LIM are listed in Table 3. Two regions where there are deletions of one residue in LIM compared with SAP, 10–12 and 83–89 (see Fig. 1), and the region of two inserted residues in LIM after position 114 of SAP, are in the inter-protomer contact regions and the differences may accommodate hexamer rather than pentamer formation. Although the precise details of inter-protomer interactions might vary depending upon the method of hexamer generation, the aggregation of LIM as a pentamer seems unfavourable. If LIM forms a pentamer similar to SAP, apart from the loss of some of the inter-protomer contacts due to deletions, the insertion of lysine and glycine after position 114 would result in steric clashes with an adjacent protomer. This is also supported by a model of a hypothetical pentamer of LIM which has been constructed in an analogous fashion to the pentameric SAP.

Calcium binding

Calcium coordination is essential for ligand binding by pentraxins [2] and for the resistance of pentraxins to proteolytic degradation [44,45]. There are two calcium

Table 3. Inter-protomer interactions.^a

SAP	CRP	HSAP	LIM
Val10-Ile104	Val10-Ile104	Val10-Ile104	Phe10-Trp101
Pro12(M)-Gly118(M)	Pro12(M)-Ser118(M)	Pro12(M)-Gly118(M)	
Tyr40-Pro113(M)	Tyr40-Pro113(M)	Tyr40-Pro113(M)	Lys40-Cys114(M)
Tyr40-Val115	Tyr40-Val115	Tyr40-Val115	Lys40-Glu104
Asp42-Lys83(M)	Glu42-Glu83(M)	Asp42-Arg83(M)	Asn42-Asn115(M)
Asp42(M)-Lys83	Ser44(M)-Glu83	Asp42(M)-Arg83	
Asp42-Lys117	Glu42-Lys117	Asp42-Lys117	Asn42-Lys114A
	Ser44 & Ser44A-Lys67		
	Thr44B-Asp68		
	Val88A-Pro85		
Ile85-Phe88 & Pro89	Pro85-Val88A	Val85-Phe88 & Ala88A	
			Lys91-Lys114A(M)
Glu99-Lys199	Glu99-Lys199	Glu99-Arg199	Trp101-Val197
Ser102(M)-Tyr195	Ser118-Tyr195	Ser102-Tyr195	Glu102-Ser199
Ile104-Ile197	Ile104-Phe197	Ile104-Val197	Thr106-Glu202
Pro113-Trp203	Pro113-Trp203	Pro113-Val203	His113-Glu202
Val115(M)-Gly152(M)	Val115(M)-Gly152(M)	Val115(M)-Gly152(M)	
Lys116-Glu153	Arg116-Asp 153	Lys116-Glu153	Ala116M-Glu151
			Val121-Val197
Pro166-Val202	Pro166-Leu202		His166-Glu202
		Lys112-Ser204B	

^aInteractions involving the main chain are indicated by 'M' at the residue position.

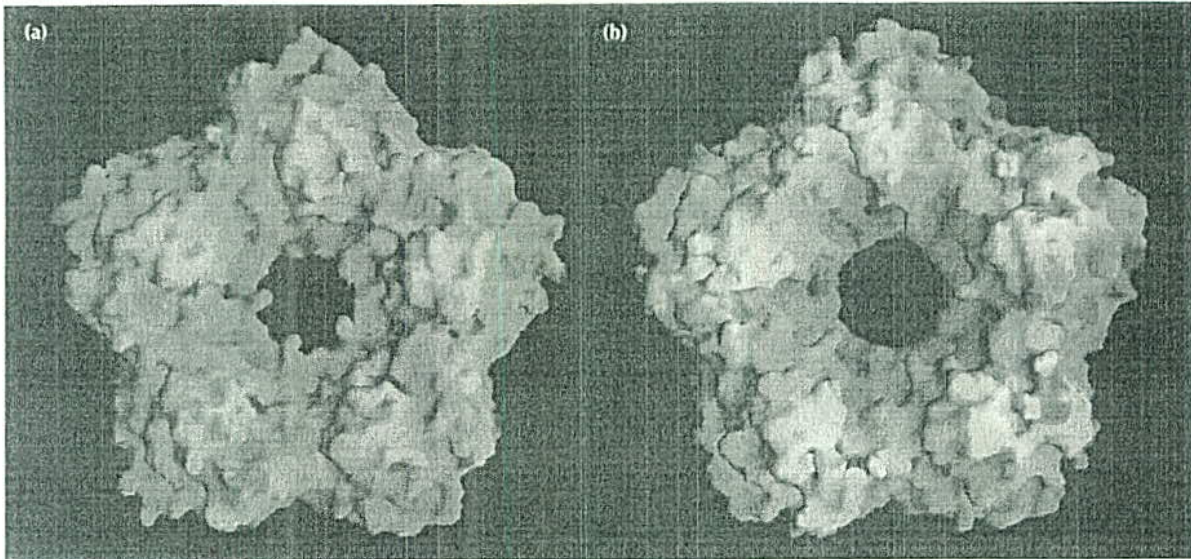


Fig. 3. Representation of the charge distribution calculated using the program GRASP that incorporates the DelPhi algorithm [56]. The pentameric face, containing the helical segments, is shown for (a) SAP and (b) CRP. The charge distribution is represented with the colour code red for negative and purple for positive charges.

ions bound to SAP. Calcium 1 is coordinated to side chains of Asp58 (both oxygens), Asn59, Glu136, Asp138 and the main chain carbonyl of Gln137. The other ligand is an acetate in four of the subunits and the side chain of Glu167 from an adjacent pentamer in the crystal lattice in the fifth subunit. Calcium 2 is less extensively coordinated by SAP, with ligands to Glu136, Asp138, Gln148, and acetate. The coordination is completed by Glu167 and two water molecules [22].

The equivalent residues in CRP, HSAP and LIM are indicated in Fig. 1.

Unlike SAP, CRP binds to the two calcium ions with equal affinity at neutral pH [44]. The backbone at the calcium-binding site of CRP is similar to that in SAP. However, there are some residue changes which could affect the affinity for calcium. In SAP, residue 145 is aspartate, which is not liganded to a calcium ion, but in

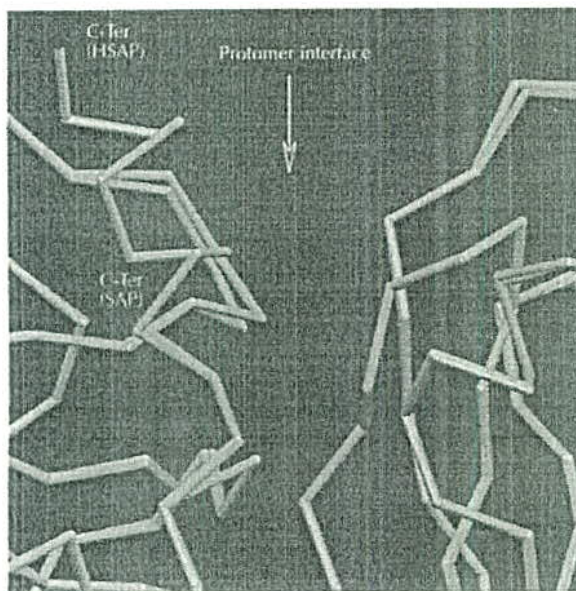


Fig. 4. Inter-protomer region of HSAP (green) superimposed on the inter-protomer region of SAP (pink). The two carboxyl termini are indicated. (Figure generated using SETOR [53].)

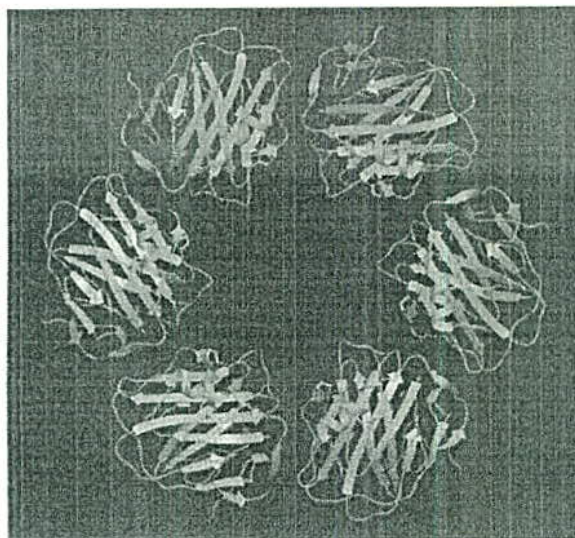


Fig. 5. Hexameric model of LIM. The β -strands and helical segment are represented in pink and blue respectively. Disulphides are shown in yellow. (Figure generated using SETOR [53].)

CRP the larger side chain of glutamate could coordinate the calcium.

In HSAP, the calcium ligands Asn59, Glu136, Gln137, Asp138 and Gln148 are conserved although Asp58 of SAP is replaced by glutamate in HSAP, a change which is unlikely to have much effect on the binding of calcium ions. In a similar way, all calcium-binding residues in LIM are identical to those in SAP with the exception of Glu136 in SAP which is replaced by aspartate.

Calcium-mediated ligand binding

SAP shows calcium-dependent binding of PE but does not bind PC. In the crystal structure of the SAP-PE complex [22], the residues Tyr74, Lys79 and Glu66 are situated close to PE with the side chain hydroxyl of Tyr74 forming a hydrogen bond with the nitrogen of PE. Glu66 and Lys79 form a salt bridge.

PE and PC differ only at the nitrogen where three methyl groups are attached in PC. Modelling of the three methyl groups (PC) resulted in short contacts ranging from 2.1 Å to 2.7 Å between the hydroxyl group of Tyr74 and the methyl carbons. These could not be relieved by rotation around the ethyl nitrogen bond as the side chain hydroxyl of Tyr74 is situated roughly on the rotation axis (Fig. 6) and there is no space for the PC to re-position itself.

Thr74 of CRP is topologically equivalent to Tyr74 of SAP. A shorter side chain at 74 (Tyr→Thr) results in longer distances between the side chain atoms of Thr74 and the three methyl groups of PC (distances range between 3.8 Å and 6.5 Å). Glu66 and Lys79, which form an ion pair in SAP, are replaced by serine and glutamate, respectively. Distances between the side chain atoms of these residues and methyl groups of PC range between 4.9 Å and 8 Å, thus accommodating the methyl groups of PC (Fig. 6).

In the complex of PE with SAP, the hydroxyl of Tyr74 forms a hydrogen bond with the nitrogen of PE. In

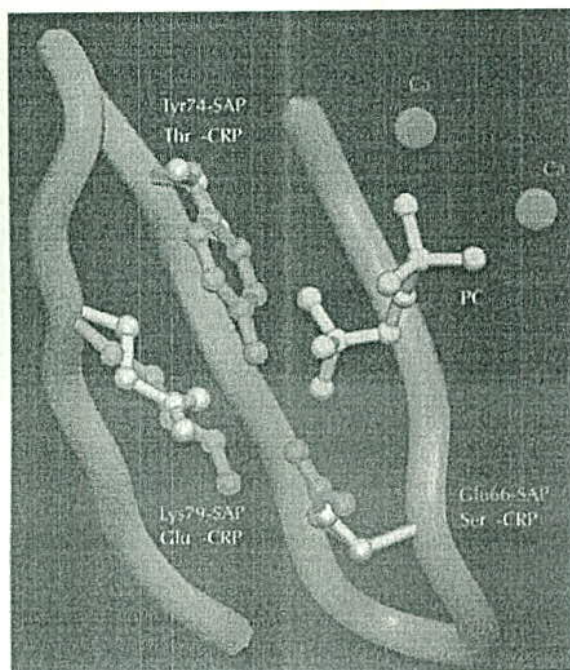


Fig. 6. Docking of phosphocholine. Side chains of SAP (pink) and CRP (blue) are shown superimposed. Phosphocholine is shown in yellow and calcium ions in red. (Figure generated using SETOR [53].)

CRP, Thr74 is too far from the nitrogen of PE to directly stabilize the complex. However, Lys79 in SAP is replaced by a glutamate in CRP, whose side chain could form a hydrogen bond with the nitrogen of PE.

These observations provide an explanation for the binding properties for PE and PC of SAP and CRP. Liu *et al.* [46] and Swanson and Mortensen [47] suggest that the residues 50–60 of CRP are important in binding PC while the mutagenesis work of Agrawal *et al.* [48] implicated Lys55, Arg56 and Trp67 as the key residues. However, it is likely that changes at positions 55, 56 and 67 will have resulted in a structural perturbation at the calcium-binding site.

HSAP resembles CRP in binding both PC and PE [15]. Tyr74 and Glu66, which are close to PC, are conserved, but the third side chain (Lys79 in SAP) that is close to PC is replaced by glycine in HSAP, thus creating more space to accommodate PC. LIM also binds both PC and PE. Tyr74 in SAP, which binds directly with the nitrogen of PE, is replaced by asparagine in LIM. Other changes include Lys79 and Glu66 in SAP which are Leu79 and Asp66 in LIM. Despite these substitutions, which alter the shape/charge of the binding site, both PC and PE can be bound by LIM.

Binding of MO β DG by SAP involves two hydrogen bonds formed between the 4,6 oxygen atoms of the carboxyethylidene ring and the side chains of Gln148 and Asn59 respectively, in addition to the direct interaction between calcium ions and the acidic group of MO β DG [22]. In the crystal structure without an added ligand, acetate from the buffer binds calcium in a similar way. CRP binds MO β DG very weakly [49]. Interestingly, both Gln148 and Asn59 are conserved in CRP (Fig. 1) but there are changes in residues that define the hydrophobic pocket: Tyr64 is a phenylalanine and Tyr74 is a threonine (Fig. 7). These changes alter the size of the pocket as well as the charge distribution and they may account for weaker binding of MO β DG. The higher affinity of CRP for calcium, possibly mediated through Gln145, may also lead to an alteration of the position of calcium 2, and thereby weaken binding of MO β DG.

HSAP binds weakly to MO β DG compared with binding of MO β DG by SAP [21]. The residues in SAP that directly interact with MO β DG (Asn59 and Gln148) and the residues forming the hydrophobic pocket (Leu62, Tyr64 and Tyr74) are all conserved in HSAP (see Fig. 1). However, replacement of Lys79 in SAP by glycine in HSAP may account for the reduced binding of MO β DG.

LIM does not bind to MO β DG (GA Tennent and MB Pepys, unpublished data). Although residues Asn59, Gln148 and Leu62 which interact with MO β DG in SAP, are conserved in LIM, the other hydrophobic residues in the pocket are significantly different in LIM, i.e., Tyr64 and Tyr74 in SAP are Leu64 and Asn74 in LIM (Fig. 1).

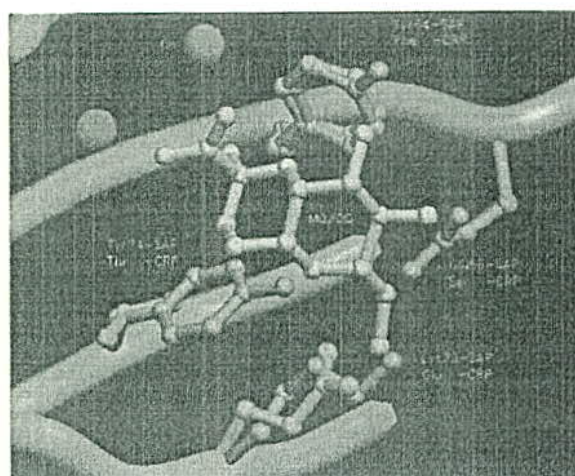


Fig. 7. Superimposition of residues in SAP and CRP in the vicinity of the MO β DG binding site. The backbone is shown in green. The side chains of SAP are shown in pink and those of CRP in blue. MO β DG is represented in yellow and calcium ions in red. (Figure generated using SETOR [53].)

Biological implications

The pentraxins are a family of pentameric vertebrate plasma proteins that have been stably conserved in evolution and share sequence homology, similar subunit assembly and the capacity for specific calcium-dependent ligand binding. The phylogenetically oldest known protein with sequence homology is C-reactive protein (LIM) from the arachnid, *Limulus polyphemus* (the horseshoe crab), which differs from its vertebrate counterparts in being hexameric. Human C-reactive protein (CRP) is the classical cytokine-regulated acute phase protein which is produced in response to inflammation and tissue damage. It is universally measured in clinical practice as an objective marker of disease activity. Human serum amyloid P component (SAP), the other human pentraxin, is a normal constituent of the extracellular matrix as well as a plasma protein. It is also invariably found in the deposits associated with the pathological condition of amyloidosis that underlies such common and important clinical disorders as Alzheimer's disease and diabetes mellitus. SAP probably contributes to persistence of amyloid *in vivo* and radiolabelled SAP is used routinely for diagnosis and monitoring of systematic amyloidosis. Serum amyloid P component from the Syrian hamster (HSAP) is of interest because it shares properties of both human CRP and SAP.

Comparative analyses of the sequences and structures of SAP, CRP, HSAP and LIM, based on the high-resolution crystal structure of SAP provide strong evidence for a highly conserved tertiary fold, despite differences in quaternary structure

and ligand specificity. The multimeric assembly of protomers with their binding sites on one planar face of the pentamer (or hexamer in the case of LIM) enables multivalent attachment. The predominantly β -sheet arrangement of the protomer fold with short, tightly bound joining loops, together with the multiple coordination of calcium ions by residues in solvent-exposed segments, account for the remarkable proteinase resistance of pentraxins. This resistance may be advantageous in the proteinase-rich microenvironment of inflammatory lesions, and it is of interest that CRP is even more proteinase resistant than SAP (GA Tennent and MB Pepys, unpublished data) and that most of the residue differences appear likely to further stabilize the CRP protomer structure.

The models also provide a structure-based explanation of the different ligand specificities among pentraxins, in particular the recognition of phosphocholine by CRP, HSAP and LIM but not by SAP, the recognition of cyclic pyruvate acetals only by SAP, and the recognition of phosphoethanolamine by all of them. Final confirmation will require crystallographic analyses of ligand complexes, but mutagenesis experiments, currently in progress, should provide an experimental test of the proposed interactions. Better understanding of pentraxin-ligand binding may help to elucidate their physiological and pathological functions, and contribute to the development of new diagnostic and therapeutic agents in amyloidosis and inflammation.

Materials and methods

Comparative modelling

The 3D structures of CRP, HSAP and LIM were modelled on the basis of the crystal structure of SAP determined at 2 Å resolution and refined to an R-factor of 0.18 [22]. The suite of programs, encoded in the comparative modelling software COMPOSER [31,32] and incorporated in SYBYL (version 6.0 - Tripos Inc.), was used to obtain initial 3D models. Conserved regions of CRP, HSAP and LIM were modelled by extrapolating from the equivalent regions of SAP; they comprise 15 stretches of β -strands, most of which are also present in lectins, a long α -helical segment at 166-174 which is unique to pentraxins, and some conserved loops. The insertion and deletion regions were modelled by searching a data base of fragments, extracted from known 3D structures, for an appropriate segment satisfying end-to-end inter-C α distances of flanking conserved regions. Steric contact with the rest of the protein was also a consideration in the loop selection. The template matching approach was used in ranking the candidate loops [50]. The selected loop segment was modelled on to the three-residue overlap with flanking SCRs (structurally conserved regions), using an algorithm to relax the structure by 'tweaking' at the ends of SCRs (F Eisenmenger, personal communication). Side chain conformations have been predicted using a set of 1200 rules based

on the conformers in equivalent residues in SAP or from a rotamer library, depending on the nature of substitutions and the type of secondary structure [34].

Energy minimization

The models were energy minimized using the AMBER force-field [51] as incorporated in SYBYL. Hydrogen atoms were also considered in the energy minimization. For every course of energy minimization, 20 cycles of Simplex method and a further 100 cycles of Powell algorithm were employed. During the initial cycles of energy minimization, the backbone atoms were maintained fixed and electrostatic energy terms were not considered. This strategy was employed primarily to relieve short contacts and to correct geometry, particularly at the 'anchor' regions of various loops. In the subsequent cycles, all atoms were allowed to move and electrostatic energy terms were included. A distance-dependent dielectric constant with a cut-off of 9 Å and an ϵ value of 4 was used. Minimization was performed until all inconsistencies in geometry were rectified and all short contacts were relieved.

Modelling of LIM

As LIM is more distantly related to SAP (30% sequence identity) many, usually conservative, amino acid replacements are present, even in the solvent-inaccessible core region. Hence a slightly different strategy, which combines constrained energy minimization at every stage of modelling, was employed to model LIM.

Initially, the structurally conserved regions of LIM, which comprise the β -strands and helix, were built using COMPOSER with SAP as the basis structure. Then the backbone of LIM was kept rigid and the side chains alone were allowed to optimize their positions during energy minimization. Each of the variable regions was built individually and checked for geometry and steric clashes with the 'anchor' regions and rectified by a simulated annealing protocol in SYBYL. Many loop conformations were adjusted manually using interactive graphics and a careful check has been made for the compatibility of conformation at each position [50]. Energy minimization of the (local) loop regions enabled steady improvement in the stereochemical quality of the model. The two disulphides that are present in LIM, in addition to the disulphide in common with SAP, were modelled after using distance constraints on the selection of fragments from the database [52]. Finally, the completed model was energy minimized by allowing all the atoms to move and with an electrostatic term included in the force field.

Generation of quaternary structures

The pentamers of CRP and HSAP were generated by superimposing the model on to each protomer of SAP; special care was taken not to alter the coordinate axes of reference. The generated pentamer model was then energy minimized by moving side chains alone, in order to relieve short contacts at the inter-protomer interfaces. Electrostatic terms were also included in these minimization cycles.

In order to generate the hexamer of LIM, a hypothetical pentamer of LIM was first generated by superposing the LIM model on to each of the protomers of SAP. The centres of C α atomic positions of the protomers define a pentagon

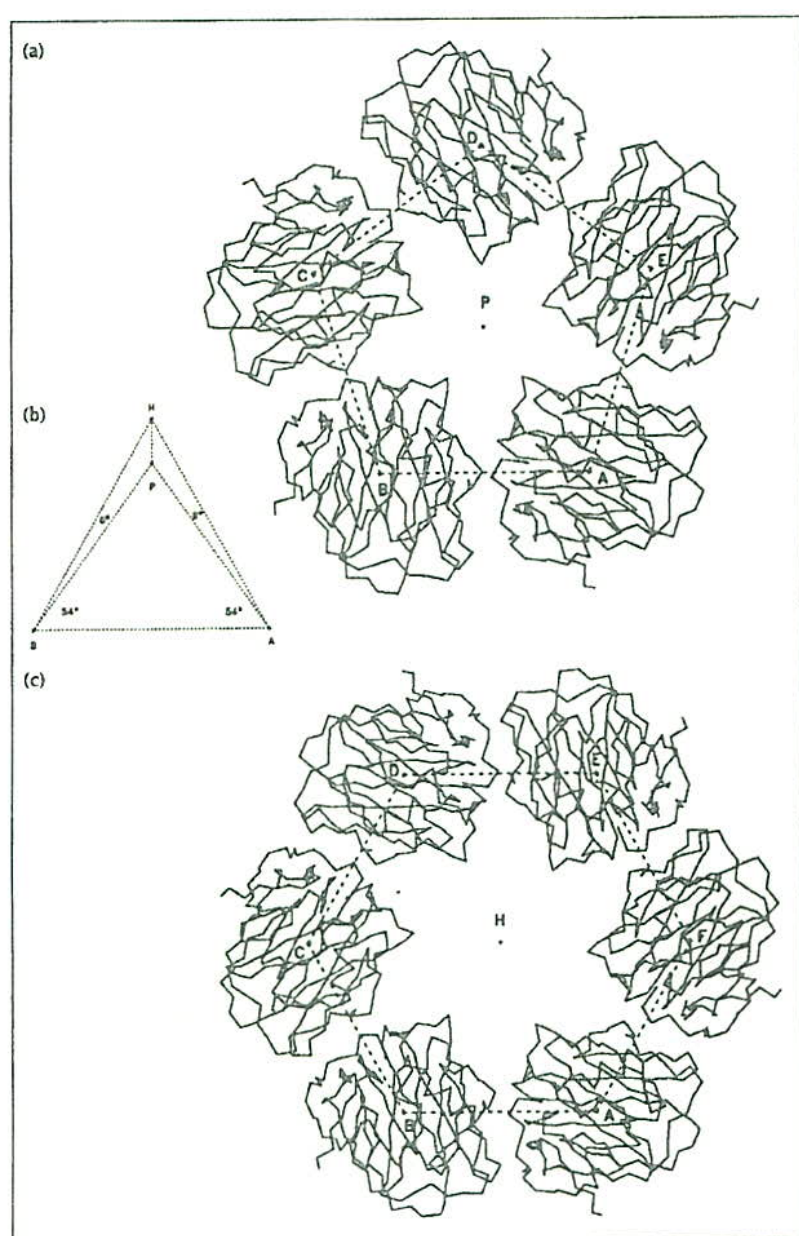


Fig. 8. Generation of a hexameric subunit arrangement for *Limulus* CRP (LIM) based on the pentameric SAP structure. (a) The hypothetical pentamer of LIM with the centre of the C α atoms in each of the protomers marked as A to E. The point P refers to the centre of the pentagon of points A to E and the five-fold axis passes through it. (b) The triangle ABP from the pentamer is used to identify a point H away from the five-fold axis and passing through the six-fold axis. (c) A preliminary model of the hexameric LIM.

whose vertices and centroid are in a plane (points A to E and P in Fig. 8a). The distance from the point P (and hence its coordinates) to a point H in the six-fold axis and in the mean plane formed by the points A to E can be calculated by considering the triangles ABP and ABH where the angles of the triangles and the distances AB, AP and PB are known (Fig. 8b). Using the six-fold axis that passes through H and parallel to the five-fold axis, a preliminary model for a hexamer of LIM was generated (Fig. 8c). The individual protomers were then systematically translated in small steps away from the six-fold axis. For every step of translation, the number of distances, within 10 Å between C β atoms across the two adjacent protomers, was monitored. Translation was terminated when the number of interacting C β atoms was same as in the pentamer of SAP. The hexameric model was then energy minimized by keeping the backbone fixed and allowing side chains to move.

Generation of CRP models bound to PE, PC and MO β DG was accomplished by alterations in side chain positions at the binding site, as guided by the structures of the complexes, SAP-PE and SAP-MO β DG [22].

All computations and development of programs for modelling and analysis were performed on a network of IRIS-SG workstations. Three-dimensional models were analyzed using interactive graphics programs such as SETOR [53] and SYBYL (Tripos Inc., St. Louis) on IRIS workstations and FRODO [54] on an Evans and Sutherland system.

Atomic coordinates of the models are being deposited in the Protein Data Bank [41].

Acknowledgements: This research was supported in part by MRC and ICRF grants to TLB and MRC programme grant G790051 to MBP. NS is supported by Tripos Inc. We thank Ms Beth Sontrop for expert preparation of the manuscript.

References

- Osmand, A.P., Friedenson, B., Gewurz, H., Painter, R.H., Hofmann, T. & Shelton, E. (1977). Characterisation of C-reactive protein and the complement subcomponent C1t as homologous proteins displaying cyclic pentameric symmetry (pentraxins). *Proc. Natl. Acad. Sci. USA* **74**, 739-743.
- Pepys, M.B. & Baltz, M.L. (1983). Acute phase proteins with special reference to C-reactive protein and related proteins (pentraxins) and serum amyloid A protein. *Adv. Immunol.* **34**, 141-212.
- Pepys, M.B., Dash, A.C., Fletcher, T.C., Richardson, N., Munn, E.A. & Feinstein, A. (1978). Analogues in other mammals and in fish of human plasma proteins C-reactive protein and amyloid P component. *Nature* **273**, 168-170.
- Baltz, M.L., et al., & Pepys, M.B. (1982). Phylogenetic aspects of C-reactive protein and related proteins. *Ann. N.Y. Acad. Sci.* **389**, 49-75.
- Pepys, M.B., et al., & Fletcher, T.C. (1982). C-reactive protein and serum amyloid P component in the plaice (*Pleuronectes platessa* L.), a marine teleost, are homologous with their human counterparts. *Biochim. Biophys. Acta* **704**, 123-133.
- Steel, D.M. & Whitehead, A.S. (1994). The major acute phase reactants: C-reactive protein, serum amyloid P component and serum amyloid A protein. *Immunol. Today* **15**, 81-88.
- Pepys, M.B., Rowe, I.F. & Baltz, M.L. (1985). C-reactive protein: binding to lipids and lipoproteins. *Int. Rev. Exp. Pathol.* **27**, 83-111.
- Vigushin, D.M., Pepys, M.B. & Hawkins, P.N. (1993). Metabolic and scintigraphic studies of radioiodinated human C-reactive protein in health and disease. *J. Clin. Invest.* **91**, 1351-1357.
- Tennent, G.A. & Pepys, M.B. (1993). Glycobiology of the pentraxins. *Biochem. Soc. Trans.* **22**, 74-79.
- Tennent, G.A., et al., & Pepys, M.B. (1993). Molecular characterization of *Limulus polyphemus* C-reactive protein. I. Subunit composition. *Eur. J. Biochem.* **214**, 91-97.
- Breviario, F., et al., & Introna, M. (1992). Interleukin-1-inducible genes in endothelial cells. Cloning of a new gene related to C-reactive protein and serum amyloid P component. *J. Biol. Chem.* **267**, 22190-22197.
- Lee, G.W., Lee, T.H. & Vilcek, J. (1993). TSG-14, a tumor necrosis factor- and IL-1-inducible protein, is a novel member of the pentaxin family of acute phase proteins. *J. Immunol.* **150**, 1804-1812.
- Volanakis, J.E. & Kaplan, M.H. (1971). Specificity of C-reactive protein for choline phosphate residues of pneumococcal C-polysaccharide. *Proc. Soc. Exp. Biol. Med.* **136**, 612-614.
- Pontet, M., Engler, R. & Jayle, M.F. (1978). One step preparation of both human C-reactive protein and C1t. *Fed. Eur. Biol. Soc. Lett.* **88**, 172-178.
- Schwalbe, R.A., Dahlback, B., Coe, J.E. & Nelsestuen, G.L. (1992). Pentraxin family of proteins interact specifically with phosphocholine and/or phosphorylethanolamine. *Biochemistry* **31**, 4907-4915.
- Hind, C.R.K., et al., & Pepys, M.B. (1984). Binding specificity of serum amyloid P component for the pyruvate acetal of galactose. *J. Exp. Med.* **159**, 1058-1069.
- Pepys, M.B. (1994). Amyloidosis. In *Samter's Immunological Diseases*. (5th edn), (Frank, M.M., Austen, K.F., Claman, H.N. & Unanue, E.R., eds), pp. 637-655, Little, Brown and Co., Boston.
- Coe, J.E., Margossian, S.S., Slayter, M.S. & Sogn, J.A. (1981). Hamster female protein. A new pentraxin structurally and functionally similar to C-reactive protein and amyloid P component. *J. Exp. Med.* **153**, 977-991.
- Coe, J.E. & Ross, M.J. (1985). Hamster female protein, a sex-limited pentraxin, is a constituent of Syrian hamster amyloid. *J. Clin. Invest.* **76**, 66-74.
- Coe, J.E. & Ross, M.J. (1990). Amyloidosis and female protein in the Syrian hamster. *J. Exp. Med.* **171**, 1257-1267.
- Tennent, G.A., et al., & Pepys, M.B. (1993). Studies of the structure and binding properties of hamster female protein. *Immunology* **80**, 645-651.
- Emsley, J., et al., & Wood, S.P. (1994). Structure of pentameric human serum amyloid P component. *Nature* **367**, 338-345.
- Hardman, K.D. & Ainsworth, C.F. (1972). Structure of concanavalin A at 2.4 Å resolution. *Biochemistry* **11**, 4910-4919.
- Reeke, G.N., Jr, Becker, J.W. & Edelman, G.M. (1975). The covalent and three dimensional structure of concanavalin A. IV. Atomic coordinates, hydrogen bonding and quaternary structure. *J. Biol. Chem.* **250**, 1525-1547.
- Einspahr, H., Parks, E.H., Suguna, K., Subramanian, E. & Suddath, F.L. (1986). The crystal structure of pea lectin at 3.0 Å resolution. *J. Biol. Chem.* **261**, 16518-16527.
- Keitel, T., Simon, O., Borris, R. & Heinemann, U. (1993). Molecular and active-site structure of a *Bacillus* 1,3-1,4-β-glucanase. *Proc. Natl. Acad. Sci. USA* **90**, 5287-5291.
- Torronen, A., Harkki, A. & Rouvinen, J. (1994). Three-dimensional structure of endo-1,4-β-xylanase II from *Trichoderma reesei*: two conformational states in the active site. *EMBO J.* **13**, 2493-2501.
- Divne, C., et al., & Jones, T.A. (1994). The three-dimensional crystal structure of the catalytic core of cellobiohydrolase I from *Trichoderma reesei*. *Science* **265**, 524-527.
- Lobsanov, Y.D., Gitt, M.A., Leffler, H., Barondes, S.H. & Rini, J.M. (1993). X-ray crystal structure of the human dimeric S-lac lectin, L-14-II, in complex with lactose at 2.9 Å resolution. *J. Biol. Chem.* **268**, 27034-27038.
- Liao, D.-I., Kapadia, G., Ahmed, H., Vasta, G.R. & Herzberg, O. (1994). Structure of S-lectin, a developmentally regulated vertebrate β-galactoside-binding protein. *Proc. Natl. Acad. Sci. USA* **91**, 1428-1432.
- Blundell, T.L., Sibanda, B.L., Sternberg, M.J.E. & Thornton, J.M. (1987). Knowledge-based prediction of protein structures and the design of novel molecules. *Nature* **326**, 347-352.
- Blundell, T.L., et al., & Sutcliffe, M. (1988). Knowledge-based protein modelling and design. *Eur. J. Biochem.* **172**, 513-520.
- Sutcliffe, M.J., Haneef, I., Carney, D. & Blundell, T.L. (1987). Knowledge based modelling of homologous proteins. Part I. Three-dimensional frameworks derived from the simultaneous superposition of multiple structures. *Protein Eng.* **1**, 377-384.
- Sutcliffe, M.J., Hayes, F.R.F. & Blundell, T.L. (1987). Knowledge based modelling of homologous proteins. Part II. Rules for the conformations of substituted sidechains. *Protein Eng.* **1**, 385-392.
- Srinivasan, N. & Blundell, T.L. (1993). An evaluation of the performance of an automated procedure for comparative modelling of protein tertiary structure. *Protein Eng.* **6**, 501-512.
- Oliveira, E.B., Gotschlich, E.C. & Liu, T.-Y. (1979). Primary structure of human C-reactive protein. *J. Biol. Chem.* **254**, 489-502.
- Lei, K.-J., Liu, T., Zon, G., Soravia, E., Liu, T.-Y. & Goldman, N.D. (1985). Genomic DNA sequence for human C-reactive protein. *J. Biol. Chem.* **260**, 13377-13383.
- Dowton, S.B., Woods, D.E., Mantzouranis, E.C. & Colten, H.R. (1985). Syrian hamster female protein: analysis of female protein primary structure and gene expression. *Science* **228**, 1206-1208.
- Nguyen, N.Y., Suzuki, A., Cheng, S.M., Zon, G. & Liu, T.-Y. (1986). Isolation and characterization of *Limulus* C-reactive protein genes. *J. Biol. Chem.* **261**, 10450-10455.
- Nguyen, N.Y., Suzuki, A., Boykins, R.A. & Liu, T.-Y. (1986). The amino acid sequence of *Limulus* C-reactive protein. Evidence of polymorphism. *J. Biol. Chem.* **261**, 10456-10465.
- Bernstein, F.C., et al., & Tasumi, M. (1977). The protein data bank: a computer-based archival file for macromolecular structures. *J. Mol. Biol.* **112**, 535-542.
- Wood, S.P., et al., & Pepys, M.B. (1988). A pentameric form of human serum amyloid P component: crystallization, X-ray diffraction and neutron scattering studies. *J. Mol. Biol.* **202**, 169-173.
- Perkins, S.J. & Pepys, M.B. (1986). X-ray and neutron scattering studies on CRP and SAP. In *Protides of the Biological Fluids*. Vol. **34** (Peeters, J., eds), pp. 323-326, Pergamon Press, Oxford.
- Kinoshita, C.M., et al., & Gewurz, H. (1989). Elucidation of a protease-sensitive site involved in the binding of calcium to C-reactive protein. *Biochemistry* **28**, 9840-9848.
- Kinoshita, C.M., et al., & Gewurz, H. (1992). A protease-sensitive site in the proposed Ca²⁺-binding region of human serum amyloid P component and other pentraxins. *Protein Sci.* **1**, 700-709.
- Liu, T.-Y., et al., & Goldman, N. (1987). Comparison of protein structure and genomic structure of human, rabbit and *Limulus* C-reactive proteins: possible implications for function and evolution. *J. Protein Chem.* **6**, 263-272.
- Swanson, S.J. & Mortensen, R.F. (1990). Binding and immunological properties of a synthetic peptide corresponding to the phosphorylcholine-binding region of C-reactive protein. *Mol. Immunol.* **27**, 679-687.
- Agrawal, A., Xu, Y., Ansardi, D., Macon, K.J. & Volanakis, J.E. (1992). Probing the phosphocholine-binding site of human C-reactive protein by site-directed mutagenesis. *J. Biol. Chem.* **267**, 25353-25358.

49. Hind, C.R.K., Collins, P.M., Caspi, D., Baltz, M.L. & Pepys, M.B. (1984). Specific chemical dissociation of fibrillar and non-fibrillar components of amyloid deposits. *Lancet* **ii**, 376-378.
50. Topham, C.M., McLeod, A., Eisenmenger, F., Overington, J.P., Johnson, M.S. & Blundell, T.L. (1993). Fragment ranking in modelling of protein-structure. Conformationally constrained environment amino acid substitution tables. *J. Mol. Biol.* **229**, 194-220.
51. Weiner, S.J., et al., & Weiner, P.K. (1984). A new force-field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* **106**, 765-784.
52. Sowdhamini, R., Srinivasan, N., Shoichet, B., Santi, D.V., Ramakrishnan, C. & Balaram, P. (1989). Stereochemical modelling of disulfide bridges: criteria for introduction into proteins by site-directed mutagenesis. *Protein Eng.* **3**, 95-103.
53. Evans, S.V. (1993). SETOR: hardware-lighted three-dimensional solid model representations of macromolecules. *J. Mol. Graph.* **11**, 134-138.
54. Jones, T.A. (1978). Graphics model building and refinement system for macromolecules. *J. Appl. Crystallogr.* **11**, 268-272.
55. Overington, J.P., Johnson, M.S., Sali, A. & Blundell, T.L. (1990). Tertiary structural constraints on protein evolutionary diversity: templates, key residues and structure prediction. *Proc. R. Soc. Lond.* **241**, 132-145.
56. Nicholls, A. & Honig, B. (1991). A rapid finite difference algorithm, utilizing successive over-relaxation to solve the Poisson-Boltzmann equation. *J. Comput. Chem.* **12**, 435-445.

Received: 15 Aug 1994; revisions requested: 2 Sep 1994;
 revisions received: 13 Sep 1994. Accepted: 14 Sep 1994.

REVIEW ARTICLE

Insights into the quaternary association of proteins through structure graphs: a case study of lectins

K. V. BRINDA, Avadhesha SUROLIA¹ and Sarawathi VISHVESHWARA¹

Molecular Biophysics Unit, Indian Institute of Science, Bangalore, India 560012

The unique three-dimensional structure of both monomeric and oligomeric proteins is encoded in their sequence. The biological functions of proteins are dependent on their tertiary and quaternary structures, and hence it is important to understand the determinants of quaternary association in proteins. Although a large number of investigations have been carried out in this direction, the underlying principles of protein oligomerization are yet to be completely understood. Recently, new insights into this problem have been gained from the analysis of structure graphs of proteins belonging to the legume lectin family. The legume lectins are an interesting family of proteins with very similar tertiary structures but varied quaternary structures. Hence they have become a very good model with which to analyse the role of primary structures in determining the modes of quaternary association. The present review summarizes the results of a legume lectin study as well as those obtained from a similar analysis carried out here on the animal lectins, namely galectins, pentraxins, calnexin, calreticulin and rhesus rotavirus Vp4 sialic-acid-binding domain. The lectin structure graphs have been used to obtain clusters of

non-covalently interacting amino acid residues at the intersubunit interfaces. The present study, performed along with traditional sequence alignment methods, has provided the signature sequence motifs for different kinds of quaternary association seen in lectins. Furthermore, the network representation of the lectin oligomers has enabled us to detect the residues which make extensive interactions ('hubs') across the oligomeric interfaces that can be targeted for interface-destabilizing mutations. The present review also provides an overview of the methodology involved in representing oligomeric protein structures as connected networks of amino acid residues. Further, it illustrates the potential of such a representation in elucidating the structural determinants of protein–protein association in general and will be of significance to protein chemists and structural biologists.

Key words: galectin, graph-spectral method, interface amino acid clusters and hubs, legume lectin, oligomeric-protein structure graph, pentraxin.

INTRODUCTION

Protein–protein association and oligomerization have been found to be extremely important for the functioning of many proteins found in Nature. Understanding the factors determining the nature and states of oligomerization or quaternary associations in proteins has been the aim of numerous studies. Various investigations have been carried out to dissect the structural features of protein interfaces. For instance, the nature and type of amino acid interactions constituting the protein interfaces [1–5], accessible surface area calculations [3–5], conservation of amino acid residues at interfaces [6,7], the geometry and nature of surface patches in monomers constituting the oligomers [3,4,8], conformational entropies of side chains at protein interfaces [9], docking of one monomer on to the other based on empirical methods [10,11], computational design and prediction of protein–protein interactions [12–14], and many more such investigations [5] have all been carried out in the past to understand and elucidate the principles underlying protein associations. Though many of these investigations have provided insights into the factors responsible for protein association, the understanding of the role of both sequence and structure in protein oligomerization is far from complete. The proteins belonging to the lectin family are an excellent model with which to investigate this problem, since they have very similar tertiary structure characterized by the 'jelly-roll' fold, and yet they have very different modes of quaternary

associations. They are known to exist as monomers, dimers and higher oligomers where the dimers and the higher oligomers comprise many different types of protein interfaces and topologically different quaternary associations. The present review provides a detailed perspective on the proteins of the lectin family as an example with which to understand the determinants of protein quaternary association.

THE LECTIN FAMILY

Lectins are carbohydrate-binding proteins that have varied applications in the field of biochemical and biomedical research [15,16]. They are found in almost all organisms, ranging from viruses to vertebrates, and have been implicated in a variety of cellular functions such as cell–cell interactions, cell-surface recognition, the innate immune system etc. [15,16]. The lectin superfamily as classified by SCOP (structural classification of proteins [17]) comprises 15 families that include the legume lectins, β -glucanases, endoglucanases, sialidases, galectins, pentraxins and calnexin/calreticulin, among others. All of them belong to the all- β class and have the ConA (concanavalin-A-like) jelly-roll fold that can be seen in Figure 1. The jelly-roll motif consists of three sets of anti-parallel β -sheets, as can also be seen from the legume lectins shown in Figure 1. There is a six-stranded flat 'back' sheet, a curved seven-stranded 'front' sheet and a short

Abbreviations used: ConA, concanavalin A; CRP, C-reactive protein; DB58 and DBL, *Dolichos biflorus* (horse gram) stem-and-leaf and seed lectins respectively; EcorL, *Erythrina corallodendron* (coral tree) lectin; GS1 and GS4, *Griffonia simplicifolia* (griffonia) lectins 1 and 4; PNA, peanut (*Arachis hypogaea*) agglutinin; SAP, serum amyloid P component.

¹ Correspondence can be addressed to either of these authors (email surolia@mbu.iisc.ernet.in or sv@mbu.iisc.ernet.in).

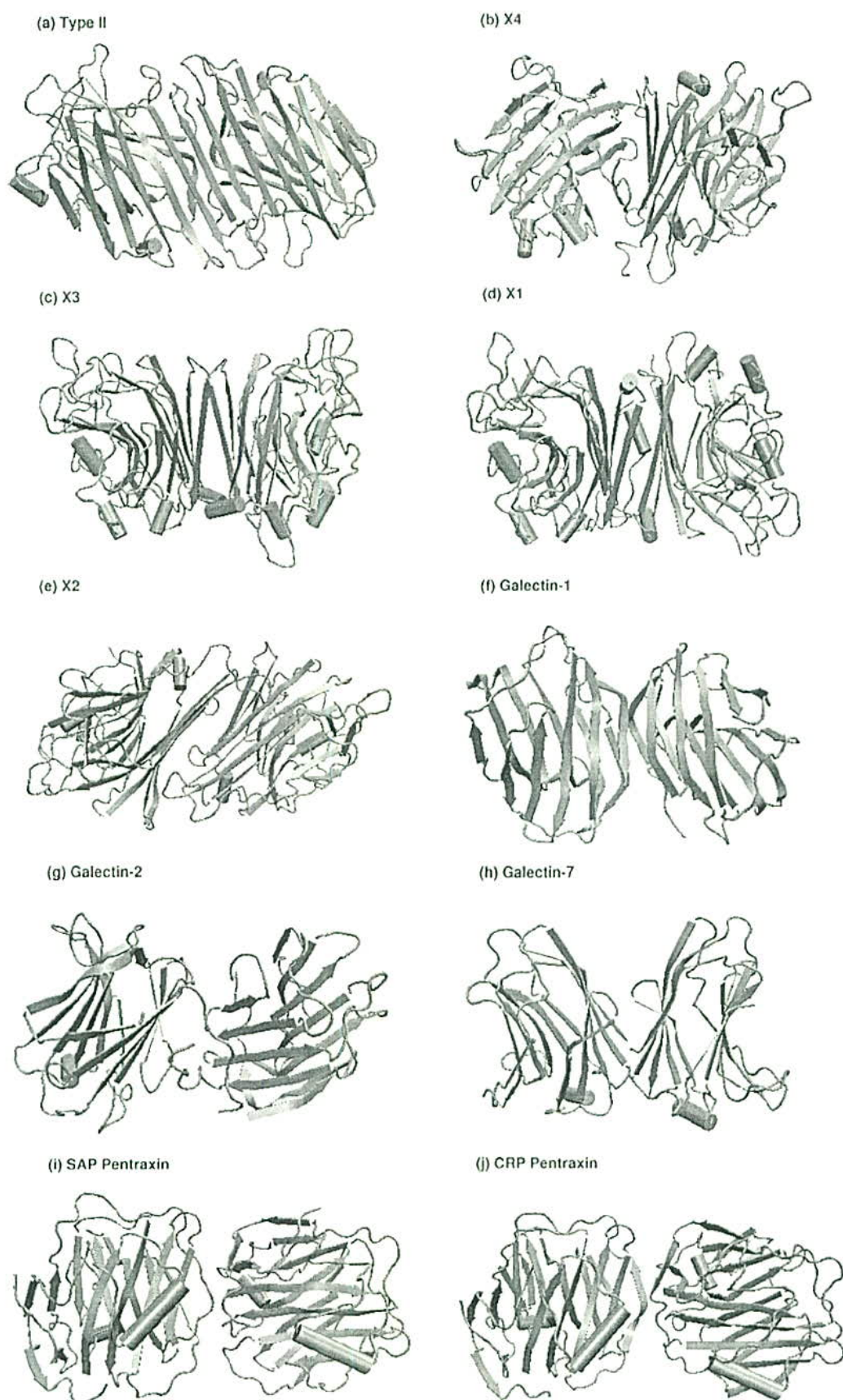


Figure 1 For legend see facing page

five-member sheet at the 'top' of the molecule. The sheets are connected by several loops of various lengths [18–19].

The legume lectins are classified under one family since they have high sequence similarities, very similar tertiary structure and biophysical properties. Further, most of these legume lectins exist as oligomers in nature, and all of them require metal ions for their carbohydrate-binding activities [15,16]. The metal-binding and carbohydrate-binding sites have been identified in various legume lectins, and they are found to be structurally overlapping [15,16]. However, these lectins do differ widely in their carbohydrate specificities and in their quaternary associations [18,19]. The other lectins in this superfamily, such as the galectins, pentraxins, calnexin and calreticulin, have very much less sequence similarity with the legume lectins. However, they do share the same jelly-roll tertiary structure and show different quaternary associations. The reasons for the varied quaternary association in lectins were investigated previously using traditional methods. For example, preliminary analysis of the legume lectin oligomers, including the determination of chemical characteristics of the different legume lectin interfaces [20], correlation with the phylogenetic trees [21], identification of conserved residues from multiple sequence alignment [22] or inspection of pairwise interactions at the intersubunit interface(s) [22], have been carried out. Though these analyses were able to elucidate some of the factors behind the lectin quaternary structures, the exact residues, sequence motifs or structural features responsible for the quaternary association were not deduced from such studies, since the overall similarity in both sequence and tertiary structure is very high in these proteins. Hence, one requires newer methods of analysis to understand protein quaternary association and the choice of the method should be such that it takes a global view of the amino acid interactions at the interfaces and not at the pairwise level. A recent study of the amino acid clusters at the protein interfaces of legume lectin structures using a graph-spectral method has aided significantly in obtaining signature sequences required for a specific type of interface in legume lectins [23]. The present review consolidates these features by detailing similar investigations on several other lectins. Furthermore, all these lectins have been investigated from a network perspective, which takes a global view of the oligomeric structures. The characterization of the amino acid clusters and hubs at the protein interfaces, as elucidated by the network representation of the lectin oligomers, is presented here. The results pertaining to the legume lectin family are taken from [23], whereas the analysis of other lectins, as well as the network perspective of lectins, are previously unpublished results presented here for the first time. Since the concept of protein structure network is relatively new, an overview of the network based methodology and its significance in biology with special emphasis to the protein structure networks is discussed below.

THE NETWORK/GRAPH PERSPECTIVE IN BIOLOGY

In the past decade, the network model has been applied in different aspects of biology to obtain useful insights into protein structure, folding, stability, regulation and evolution, and in genome analysis and comparison [24]. First, a network representation requires consideration of the system under study as a set of nodes (points) and edges (links) that constitute the graph representing the system.

For example, the complete protein interaction network, or the regulatory network, or the metabolic pathway of an organism, say yeast, can be represented as a network graph. In these networks, each protein, ligand or metabolite can be the nodes in the graph and the interaction they make with the other nodes in the graph or the way they biologically regulate the other nodes in the graph can become the edges in these networks. Such networks have recently been constructed and analysed for yeast, yielding valuable insights into the organization of biological networks in genomes [24,25]. Interestingly, all these biological networks are found to have similar overall graphical properties and hence seem to follow some common rules. They belong to a class of real-world networks termed 'scale-free networks', which show a power-law degree distribution of nodes and are characterized by the presence of a small number of highly connected nodes called 'hubs', which yield robustness to these networks [24]. Similarly, the protein domain network has also been constructed by connecting the domains based on structural similarity score, which has been analysed from the domain organization and the evolutionary perspective [26]. An interesting concept relevant to the present review is to consider each protein structure as a network of non-covalent interactions between amino acid residues, and the efforts made in this direction are summarized below.

Networks in protein structures

Studies on protein structure networks have been carried out where the protein structure itself is considered as a network of atoms, amino acids or secondary structural elements according to the requirement, and the edges are constructed based on the interaction between these structural elements. A summary of such investigations is given in Table 1. The studies include the use of graph theory in protein structure comparison [27], where the protein structures to be compared are represented as graphs using their secondary structures as nodes and the interactions between them as edges. The graphs thus generated are compared using standard graph theoretical techniques, such as graph isomorphism, to obtain insights into the structural similarities of the proteins involved (Table 1). The graph representation of protein structures have also been used to understand protein structure and folding at different levels [28–33], dynamics [34,35], comparative modelling [36,37], for the identification of structural domains in proteins [38] and in identifying modular clusters in protein interfaces [39] as given in Table 1. In some of these studies, the amino acid residues in the protein structures or their C^α atoms have been considered as the nodes of the graph. In some others, the secondary structure elements or all the atoms in the protein structure are considered as independent nodes. However, edge-forming criteria vary according to the aim of the study and can vary from sequential proximity to spatial interactions between the nodes.

Another set of analyses that have aided our understanding of protein structure and stability include the identification and analysis of clusters of amino acid residues in protein structures from a protein structure network perspective, using graph-spectral methods [40] (Table 1). Here, the amino acid residues are the nodes in the graph and the edges are determined on the basis of the strength of the non-covalent interactions between them. The spectral parameters of such a graph give the cluster-forming

Figure 1 Different types of dimeric interfaces in plant and animal lectins

The three-dimensional structure of the dimeric interfaces of the legume lectins (II, X4, X3, X1 and X2), galectins (galectin-1, galectin-2 and galectin-7) and pentraxins (SAP and CRP) are shown in cartoon representation with the monomers differentiated using different colours. The jelly-roll fold that characterizes the tertiary structures of these lectin can be clearly seen in the Figure. The monomeric lectins, such as arcelin-5, galectin-3, Charcot-Leyden protein, calnexin and calreticulin, which do have the jelly-roll tertiary structure, are not shown.

Table 1 Summary of protein structure networks

Group and reference(s)	Network	Analysis
Grindley et al. [27] Dokholyan et al. [33]	Protein structure graphs based on secondary-structural elements Protein structure graphs based on C α atoms	Subgraph isomorphism for structure comparison Network properties and hubs relating to protein folding using graphs of transition states and intermediates
Vendruscolo et al. [31,32]	Protein structure graphs based on C α atoms	Network properties and hubs relating to protein folding using graphs of transition states and intermediates
Atilgan et al. [29] Greene and Higman [30] Brinda and Vishveshwara (unpublished work)	Protein structure graphs based on C α atoms Protein structure graphs based on atomic details Residue-based protein structure graphs	Shortest path length and clustering coefficients, network properties Network properties, scale-free nature of protein structure graphs Network properties, scale-free nature, role of hubs in protein structures
Vishveshwara and co-workers [40–45] Sistla et al. [38] Reichman et al. [39] Przytycka et al. [28] Bahar et al. [34,35]	Residue-based protein structure graphs Residue-based protein structure graphs Protein structure graphs based on secondary-structural elements Protein structure graphs based on C α atoms	Graph-spectral parameters for detection of amino acid clusters at protein core, protein/protein and protein/DNA interfaces Identification of structural domains in proteins structures from graph spectra Modular clusters at protein/protein interfaces Rules of protein folding in β -proteins by basic graph manipulations Graph-spectral parameters for elucidating protein dynamics using the Gaussian network model (GNM) and the anisotropic network model (ANM)
Samudrala and Moult [36,37]	Protein graphs based on sterically allowed conformations of protein side chains (rotamers)	Clique-detecting algorithm used for modelling the side chains in proteins with unknown structures

residues in the protein structure [40,41]. This cluster information can be used to identify active-site clusters, folding clusters or intersubunit interface clusters. The main advantage of this method is that it requires a single numeric computation and also involves the global topology of the protein, because the complete protein structure is represented in the form of a connected network [41]. This kind of graph representation has yielded very useful results regarding protein stability [42], protein–protein interactions [43] and protein–DNA interactions [44]. The interface analysis of a set of functional homodimers using this method gave the amino acid clusters and their cluster centres at the interfaces, which aided the identification of interface ‘hot spots’ and also provided a method to predict the interactive surfaces on monomers [43]. It has also helped in the identification of residues important in the formation of the α – α dimer in RNA polymerase, where a single mutation at the α/α interface predicted using this method, led to the de-activation of the enzyme due to the destabilization of the α – α dimer [45]. A similar graph-spectral analysis of multidomain proteins has yielded an elegant and easy method to identify structural domains in proteins and also the residues forming the interface between the domains [38]. Recently, Reichmann et al. [39] used a residue-based interaction graph of proteins to identify modular clusters at protein/protein interfaces aiding our understanding of the architecture of intersubunit interfaces. Thus the concept of representing protein structures as networks and the identification of amino acid clusters in protein structures have, in many ways, enhanced our understanding of protein structures, folding, stability and interactions and hence has the potential to address various other aspects related to protein structural biology. In the following sections, we present the necessary description of this method and its application to the analysis of various kinds of quaternary associations seen in legume lectins, galectins and pentraxins.

METHODOLOGY

Recently, the concept of using protein structure graphs to determine amino acid clusters at protein interfaces has been applied to a set of legume lectin oligomers and the analysis yielded the sequence motifs characteristic of each type of quaternary association seen in them [23]. We have further extended this study to other proteins exhibiting the legume lectin fold, ranging from plants to vertebrates. These analyses underscore the power of such

an algorithm in dissecting protein/protein interfaces across the diverse members of this highly complex family of proteins with strikingly disparate modes of oligomerization. The algorithm consists of representing protein structures as graphs comprising of a set of nodes and edges, where the amino acid residues are nodes and the strength of the non-covalent interactions between them determines the edges as explained below [40]. Such a graph can then be analysed in various ways to obtain information regarding clusters of amino acid residues and highly connected amino acid residues (known as hubs) involved in these protein structure networks as explained below.

Representation of protein graphs

The atomic co-ordinates of the protein structures were obtained from the Protein Data Bank [46]. Each amino acid in the protein structure is represented as a node and the non-covalent interactions between their side-chains are evaluated for edge-formation [40], as follows:

$$I_{ij} = \{n_{ij}/[\min(N_i^0, N_j^0)]\} \times 100 \quad (1)$$

where n_{ij} is the number of distinct side-chain atoms pairs of i and j coming within a distance of 0.45 nm (4.5 Å), evaluated from the crystal structures. N_i^0 and N_j^0 are the normalization values for residue type i and j , which were evaluated previously from a non-redundant set of proteins [40] and are given in Table 2. It can be noted from Table 2 that these normalization values correlate well with the size of the residues.

A cut-off value of interaction, $I_{\min}$, is then considered, and any ij pair that has I_{ij} greater than $I_{\min}$ is connected in the graph. For example, when an $I_{\min}$ of 6% is used, all the connected residues in the graph interact with a value more than 6%. Figure 2(a) shows an example of high interaction (8%) between two aromatic residues. The variable parameter in the construction of a protein structure graph is the interaction cut-off, $I_{\min}$, which can be varied rationally to obtain interesting results, as elucidated below.

Cluster analysis

The protein structure graph obtained as explained above can be represented in the form of a matrix called the ‘Laplacian matrix’, which has the connectivity information among the residues [40]. This matrix is an $N \times N$ matrix, where N is the number of residues

Table 2 Normalization values of amino acids

For further details, see the text.

Amino acid	Normalization value
Ala	55.76
Arg	93.79
Asn	73.41
Asp	75.15
Cys	54.95
Gln	78.13
Glu	78.83
Gly	47.31
His	83.74
Ile	67.95
Leu	72.25
Lys	69.61
Met	69.26
Phe	93.31
Pro	51.33
Ser	61.39
Thr	63.71
Trp	106.70
Tyr	100.72
Val	62.37

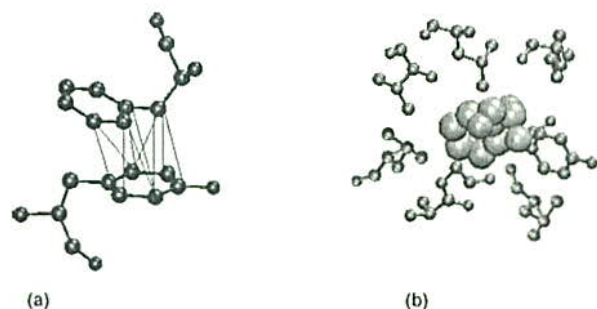


Figure 2 Interaction strength and hubs

(a) Two aromatic residues (shown in ball-and-stick representation) interacting with high value of interaction strength ($I_{ij} = 8\%$). The non-covalent atom-atom contacts occurring within a distance of 0.45 nm (4.5 Å) are indicated by thin lines. (b) Hubs: a phenylalanine residue (shown in van der Waals representation) interacting with seven other residues (shown in ball-and-stick representation), thus acting as a hub in the protein structure.

in the protein structure. From the matrix we obtain the spectra of the graph comprising eigenvalues and vector components, which provide information regarding the residues forming clusters in the protein structures and those that form the centres of these clusters [40,41]. The clusters can be obtained at different interaction cut-offs (I_{min}), which would yield clusters comprising residues interacting at various strengths. A higher I_{min} indicates stronger interaction among the residues forming the cluster, whereas a lower I_{min} indicates weaker interaction among the same. The typical working cut-off that is used for cluster analysis varies from 4 to 12 % [40,41]. The clusters obtained using this clustering method can be further combined with other traditional multiple sequence-alignment methods to obtain information regarding conserved clusters [43]. Although these conserved clusters are a set of residues that are far apart in sequence, they interact in the three-dimensional structure of the protein and are also conserved across different species. The analysis of the conserved clusters at the interfaces of some plant and animal lectin oligomers has been carried out and are discussed in detail below.

The graph-spectral method has the distinct advantage over traditional clustering methods in dealing with weighted graphs,

in identifying the cluster centres and in detecting subclusters in a connected graph [41]. Hence this method has been useful in protein structure analysis where clusters of biological significance, such as active-site clusters, hydrophobic-core-forming clusters, clusters imparting thermal stability to proteins, clusters at the protein/protein and protein/DNA interface [40–45] have been identified. This has aided in the understanding of the role of non-covalent interactions in the folding, stability and interaction of protein structures.

Hubs

Hubs are defined as highly connected nodes in a graph. In a protein structure graph, those nodes with more than four connections (links, edges), are termed the hubs. The presence of hubs is a characteristic feature of many real-world networks, where it has been found that there are a limited number of nodes acting as hubs in these networks, which help in reducing the distance between any two nodes in the network and thus form highly connected, compact networks [24]. It has also been found that these hubs provide robustness to the networks, because random attacks on the non-hubs do not affect the network organization or its stability, whereas targeted attacks on the few hubs present can cause severe damage to the network [24]. Applying the same principle to the protein structure networks, one would expect the hubs (residues) in these residue-based protein structure networks to play an important role in stabilizing the folded structure of the protein, and hence a targeted mutation of the hub residue may destabilize the protein structure. An earlier analysis of hubs in protein structures showed that the hubs identified at different I_{min} values correlated with experimentally available thermodynamic and kinetic parameters (K. V. Brinda and S. Vishveshwara, unpublished work) and also showed a preference for phenylalanine, tyrosine, tryptophan and arginine as hubs at higher I_{min} values, whereas leucine and isoleucine were preferred at lower I_{min} values (K. V. Brinda and S. Vishveshwara, unpublished work).

As explained above, the evaluation of interaction between two residues in a protein structure involves the normalization values of both the residue types. However, for the identification of hubs in a protein structure, it would be accurate to use the normalization value of the hub-forming residue alone. Hence, the interaction equation given in eqn (1) reduces to the following for hub identification:

$$I_{ij} = (n_{ij}/N_i^0) \times 100 \quad (2)$$

where, I_{ij} , n_{ij} and N_i^0 are the same as in eqn (1), with i being the residue whose hub character is being evaluated. An example of a hub-forming residue in the residue-based protein structure graph is shown in Figure 2(b), where a single phenylalanine residue interacts with many other residues. The concept of hubs is applied to some of the lectin oligomers and is presented in the subsection below entitled 'Interface hubs'.

Size of the largest cluster

The size of the largest cluster in a network is one of the important parameters that are generally used to understand the nature and properties of the network [24,26]. In the case of the protein structure graphs described above, most of the residues in the protein exist as a part of the largest cluster at lower I_{min} and the size of the largest cluster decreases with increasing I_{min} . A characteristic profile of the protein structure graphs is observed when the size of the largest cluster is plotted as a function of I_{min} . The analysis of a non-redundant set of monomeric proteins showed that a plot of the

size of the largest cluster versus $I_{\min}$ exhibits a sigmoidal profile with a transition around $I_{\min} = 4\%$ in proteins of all sizes and folds, indicating a universal behaviour in protein structures (K. V. Brinda and S. Vishveshwara, unpublished work). This transition in the size of the largest cluster around $I_{\min} = 4\%$ is found to occur due to the loss of numerous contacts made mainly by the hydrophobic residues such as leucine and isoleucine. In physical terms this change refers to a structural transition that occurs when one large connected cluster (as seen at $I_{\min} = 0\%$) splits into smaller distinct clusters (as seen at $I_{\min}$ greater than 4%). The analysis of the size of the largest cluster in the lectin oligomers is also presented in the subsection 'Size of the largest cluster as a function of interaction cut-off ($I_{\min}$)' below.

INSIGHTS INTO LEGUME LECTIN QUATERNARY ASSOCIATION

The legume lectins are known to have varied types of quaternary associations in spite of very similar tertiary structures. What precisely determines their mode of quaternary association remained unresolved until the protein-structure-graph approach was used [23]. The key results of the analysis are presented here.

Classification of legume lectin interfaces and quaternary structures

The legume lectins can be structurally classified into nine types on the basis of their overall quaternary structure, and these nine quaternary structure types consist of seven known dimeric interface types. The different kinds of dimeric interfaces seen in legume lectin oligomers include types II (canonical), X1 [DB58 (*Dolichos biflorus* stem-and-leaf lectin)-type], X2 (non-canonical interface of ConA), X3 [EcorL (*Erythrina corallodendron* lectin)-type, handshake], X4 [GS4 (*Griffonia simplicifolia* lectin 4)-type, back-to-back] and the unusual interfaces of PNA [peanut (*Arachis hypogaea*) agglutinin] and GS1 (*G. simplicifolia* lectin 1) (Figure 1). The higher oligomers, mainly tetramers, are generally dimers of dimers and hence have combinations of these dimeric interfaces. The nine different kinds of legume lectin quaternary structures made up of these seven dimeric interface types comprise of canonical (II dimer), EcorL-type (X3 dimer), GS4-type (X4 dimer), DB58-type (X1 dimer), DBL (*D. biflorus* seed lectin)-type (II + X1 tetramer), ConA-type (II + X2 tetramer), the open quaternary structure of PNA (II + X4 + unusual tetramer), GS1-type (X4 + unusual tetramer) and arcelin-5-type (monomer) [arcelin-5 is a lectin-like defence protein from *Phaseolus vulgaris* (French bean)]. As can be seen, some of the interface types are found to be present in the dimeric as well as the tetrameric legume lectins (like II, X1 and X4), whereas some of them are seen exclusively in dimeric legume lectins (like X3) and some are seen exclusively in the tetramers (like X2 and the unusual interfaces of PNA and GS1). The legume lectins with known structures are thus classified into different types of quaternary structures, and the higher oligomers are separated into pairs of dimers so as to classify their interface types.

Consensus signature sequence motifs

The method for obtaining the signature sequences determining each of the legume lectin interface types actually involves a combination of the cluster-detecting algorithm and the traditional sequence-alignment methods. The amino acid clusters in the legume lectin structures are first identified using the graph-spectral method explained above. The interface clusters in these oligomers are then identified by the fact that these clusters comprise residues from both the monomers forming the interface. The interaction cut-off ($I_{\min}$) used for cluster identification has been optimized for each lectin so as to obtain distinct interface

clusters differentiated from the bulk of the protein [23]. These interface cluster-forming residues are then mapped on to the multiple sequence alignments (obtained using ClustalW; [48]) of all those legume lectins that belong to a particular interface type. Those residues that are conserved in the multiple sequence alignments and are also present in the interface clusters of all the legume lectins belonging to a particular interface type constitute the consensus signature sequence for each interface type. The consensus patterns thus obtained for five of the prominent legume lectin interface types, namely the type II, X1, X2, X3 and X4 are given in Figure 3. Further, this algorithm also works to an extent even when there is only one known structure of a particular interface type. In such a case the mapping on to multiple sequence alignments cannot be carried out, since there is only one known example. However, the information regarding the cluster-forming residues at the interface can still be obtained to give some idea regarding the residues required for the interface formation, though it might not be conclusive. This was carried out in the two unusual interfaces of PNA and GS1 where there are only single structures available. The unusual interface of GS1 is characterized by residues W10, T26, G28, Q31, T35, F75, Y226 and L228 and that of PNA is characterized by residues L27, Q33, S28, V160, R221, N31, E72, K74 and G158 (for brevity the one-letter amino acid notation is used). The residues present in the interface clusters of the legume lectin interface types II, X1, X2, X3, X4 and the unusual interfaces of PNA and GS1 are shown in Figure 4. It is clear from the Figure 4 that the different types have different clustering patterns, which further lead to their specific consensus signature motif. The analysis thus provides the signature sequence motifs for each legume lectin interface type, and the higher oligomers with multiple interfaces were found to contain the signatures of all the interface types that they comprise.

Comparison of different legume lectin interfaces based on the signature motifs

The comparison of the consensus patterns of all the interface types showed that they were localized in a few sequence regions, namely the N-terminus and the C-terminus and the region near residues 70–80 and 160–200 (Figure 3). The type II interface and the unusual interfaces of PNA and GS1 are characterized mainly by the N-terminal and the C-terminal regions, whereas X1, X2, X3 and X4 have their consensus patterns in the 70–80 and 160–200 regions. Hence, X1, X2, X3 and X4 mutually exclude each other and therefore cannot co-exist with each other in the higher oligomers. Therefore X1, X2, X3 and X4 can only combine with either type II or the unusual interfaces. Theoretically, any of the X1, X2, X3 and X4 dimeric interfaces can combine with the type II or the unusual interfaces of PNA and GS1 to form tetramers. However, not all these combinations are observed in Nature, and there are specific preferences to the combinations seen as explained above. A look at the multiple sequence alignments indicates that these preferences are mainly due to the absence of the residues required for the other interface types, which leads to the exclusion of the other interface types. For example, in the X4-forming legume lectins GS1 and GS4, the type II residues are found to be absent, and for that reason they cannot form type II interfaces. However, in PNA, the II + X4 combination is seen because the residues required for both II and X4 are present in PNA (which also has an unusual interface and an open quaternary structure). Moreover, GS4 exists as an X4 dimer only, whereas GS1 forms the X4 + unusual tetramer. When the sequences of the two were compared, it was found that although six out of the eight residues required for the unusual interface formation of GS1 are present in GS4 also, two of the significant residues are mutated (T26I and F75L), leading to the loss of some important

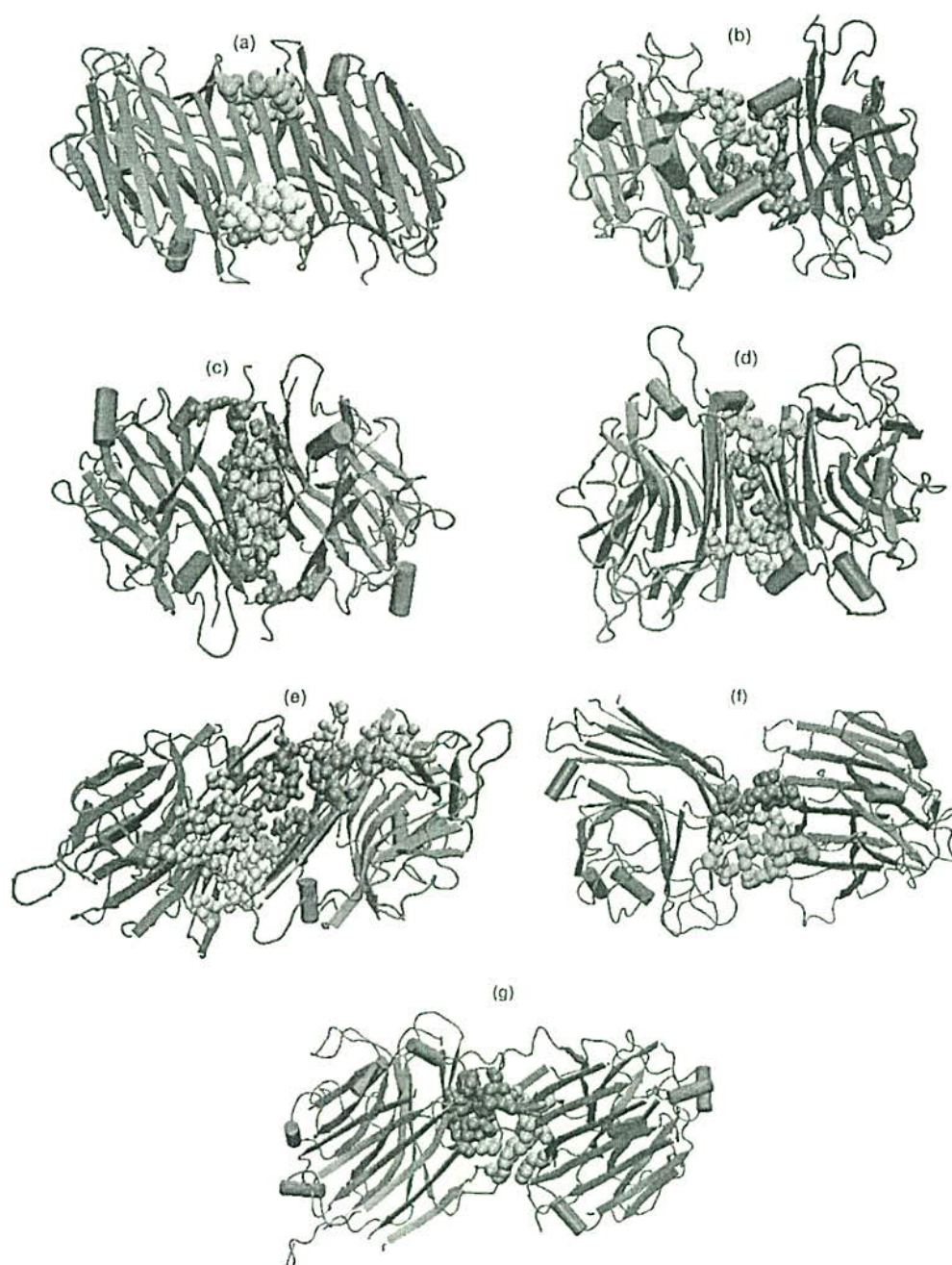


Figure 4 Interface clusters in the seven types of interfaces in legume lectins

(a)–(g) give the cartoon representation of the three-dimensional structure of the legume lectins belonging to these seven types of dimeric interfaces. The monomer chains are represented by blue- and red-coloured cartoons respectively. Although the tertiary structures are very similar in all the seven cases, their quaternary associations are different, as seen in the Figure. The interface cluster-forming residues are represented as van der Waal's spheres. Each cluster is coloured differently to differentiate them in the three-dimensional space. (a) Canonical, type II (1fnyAB at 8% cut-off); (b) EcorL-type, X3 (1axy at 5% cut-off); (c) GS4-type, X4 (1gsl at 6% cut-off); (d) DB58-type, X1 (1qmwAC at 4% cut-off); (e) non-canonical interface of ConA-type, X2 (1dglAC at 4% cut-off); (f) unusual interface of PNA (2pelBD at 6% cut-off); (g) unusual interface of GS1 (1hqIAC at 6% cut-off).

Further, the unusual interfaces are generally found to substitute for the type II interfaces and occur when type II is sterically or sequentially disallowed, as in the case of PNA or GS1. Hence the unusual interfaces involve the same sequence regions as type II. Thus the graph-spectra-based interface cluster analysis of legume lectins has answered most of the questions regarding specificities of interfaces in the legume lectin quaternary structures.

Prediction of quaternary association for lectins with unknown structure

The acid test for the signature sequence motifs of each legume lectin interface type, identified using the conserved interface cluster method, was carried out using the legume lectins with unknown structures. The sequences of six such lectins [from

Onobrychis viciifolia (sainfoin), *Cytisus scoparius* (scotch broom), *Lotus tetragonolobus* (asparagus pea), *Vatairea macrocarpa*, *Bauhinia purpurea* (orchid tree) and *Cicer arietinum* (chickpea) were aligned with the others with known interface types and were checked for the presence of any of the signature motifs so that their nature of oligomerization can be predicted from the presence or absence of these consensus patterns. This was then compared with the information about their state of oligomerization available from biochemical experiments and those predicted from phylogenetic tree analysis [21]. The results of this analysis predicted that *O. viciifolia* lectin should form type II dimers, *Cy. scoparius*, *V. macrocarpa* and *Ci. arietinum* lectins should form type II + X1 tetramers and *B. purpurea* lectin perhaps forms an X4 + unusual tetramer. The results for *L. tetragonolobus* were inconclusive from this analysis. Thus the method could predict the interfaces in five of the six lectins, which also correlated well with the available experimental evidence and results from the phylogenetic tree analysis, proving that this method has a significant predictive value.

Inferences from the legume lectin study

The analysis of the legume lectin quaternary structure using the graph-spectral method has yielded the signature sequences required for each type of quaternary association seen in the legume lectins. It has also aided in characterizing the different interface types and quaternary associations seen in the legume lectins based on the clusters of amino acids seen at these interfaces, in understanding the different factors responsible for the specific oligomerization type of these lectins and in predicting the quaternary association of legume lectins with unknown structures. One of the important points that need to be highlighted is that the application of the clustering algorithm to the oligomeric interfaces of proteins gives a better understanding of the interface nature and composition, as can be seen from these results, because the clusters involve all types of interactions namely, charged, hydrophobic, polar, hydrogen bonds and van der Waal's interactions. There is no discrimination of interactions based on the nature of residues, which helps in obtaining an overall picture of the interactions at the interfaces rather than the pairwise interactions across the interface. Moreover, the algorithm combines the identification of amino acid clusters in protein structures with sequence alignments, thus enabling us to obtain a consensus of both sequentially and structurally conserved clusters of interacting amino acids. Further, the graph-spectral cluster-determining algorithm gives the information regarding the residues which form the centres of clusters, which often coincide with the highly conserved residues in legume lectins. Hence such an analysis has been found to be more successful in handling the complexity of oligomerization in legume lectins and also has the potential to be applied to address various other aspects of protein structure, stability, folding and interactions.

GALECTINS

The representation and analysis of the structure graphs of galectin and pentraxin oligomers are presented here for the first time.

Background

Galectins, though originally discovered in animals, have recently been found in mushrooms as well [49]. They are lectins that bind β -galactose-containing glycoconjugates and do not require bivalent cations to carry out their function, unlike the legume lectins [15,16]. The proposed biological functions of galectins include roles in cell-cell adhesion, cell-matrix adhesion, direct ef-

fects on cell growth and viability, potential intracellular functions in regulating metabolism, induction of apoptosis or programmed cell death, and induction of metabolic changes, such as cellular activation and mitosis [15,16]. Their expression in many types of tumour cells has led to the hypothesis that galectins may also be involved in tumorigenesis and metastasis.

Galectins exist as both monomers and dimers. Galectins 1, 2 and 7 are homodimers, whereas the other galectins are monomers. Interestingly, the tertiary structure of galectins is similar to the jelly-roll fold found in many leguminous plant lectins, although they differ considerably in their primary structures. The crystal structures of galectins 1, 2, 3 and 7, congerin and the Charcot-Leyden protein are available and therefore have been considered in this analysis. Galectin-2 is a fungal galectin (Cgl2 from *Coprinopsis cinerea* [49]). The homodimeric galectin interfaces, namely those of galectins 1, 2 and 7, are topologically different from each other, as can be seen from Figure 1. These are also significantly different from the legume lectin interfaces mentioned above. Congerin forms a galectin-1 type interface and galectin-3 and the Charcot-Leyden protein remain as monomers.

Characterization of galectin interfaces

A careful examination of the sequences of the monomeric and dimeric galectins shows that they do not contain most of the residues present in the signature sequences of the interface types seen in legume lectins, namely II, X1, X2, X3 and X4 (as can be seen from Figure 3). The galectins have their very own signatures that characterize their quaternary association types. The amino acid clusters at the interfaces of the dimeric galectins were identified using the graph-spectral method for those galectins whose structures are available (galectins 1, 2 and 7, shown in Figure 5, and congerin) and were mapped on to the multiple sequence alignment of these galectins obtained along with the monomeric ones (galectin-3 and the Charcot-Leyden protein). These interface cluster-forming residues are highlighted in the multiple sequence alignment shown in Figure 6(a). A comparison of the interface cluster residues characterizing each galectin interface type gave interesting results regarding the factors determining the mode of quaternary association in these lectins. Galectins 1, 2 and 7 have different sequence patterns that determine their oligomerization type, as can be seen from Figure 6(a). The galectin-1 type interface seen in galectin-1 and congerin involves a few residues from the N-terminus and the C-terminus. The T/K + N (Thr/Lys + Asn) pattern observed at the N-terminus in these galectins is partially conserved in galectin-2 and the F + F/L (Phe + Phe/Leu) at the C terminus is partially conserved in the monomers and galectin-7 (Figure 6a). However, none of them have both patterns conserved, which could render them incapable of forming the galectin-1 type interface. The interface clusters in galectin-1 are shown in Figure 5.

The galectin-7 interface is characterized by a number of charged-charged interactions among a set of arginine residues (numbers 14, 20, 22, 133) at the N-terminus and C-terminus, along with His¹⁰⁸, Glu⁸⁷, Asp⁹⁵ and Asp¹⁰³. Phe¹³⁵ and Val¹⁰⁰ also contribute to the cluster (Figure 5). The arginine residues from one chain are neutralized by the aspartate residues from the other, with histidine and glutamate also adding to the stability. On examining the other galectin sequences for the presence of these residues (Figure 6a), we find that all the others do have some of these residues, but none of them have all of them. In fact, those that have the negatively charged residues conserved in those positions lack the complementary positively charged residues and vice versa. Therefore none of them have the complete complement of the arginines, histidine, glutamate and aspartate residues, thus

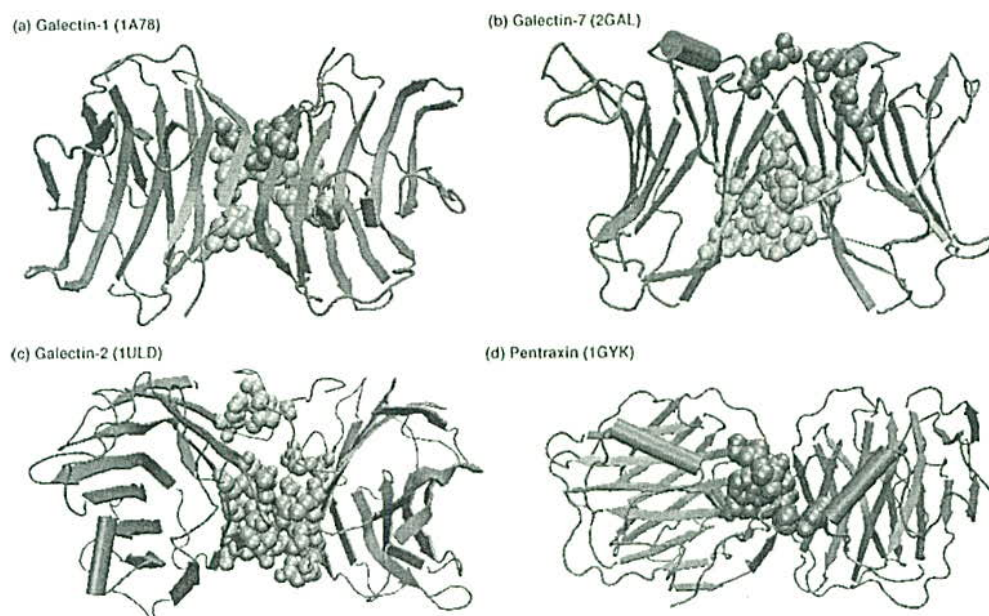


Figure 5 Interface clusters in galectins and pentraxins

The dimeric interfaces of galectins 1, 2 and 7 and pentraxin, along with the amino acid clusters obtained at the interfaces at about $I_{\text{min}} = 6\%$ are shown. The lectins are represented in cartoon diagram and the interface clusters in van der Waals representation. Each monomer and interface clusters are coloured differently.

rendering them incapable of forming the galectin-7 type interface.

The galectin-2 type interface is also characterized by charged interactions across the interface. However, here the positive charge is contributed by Arg⁹⁹, Arg¹⁰³ and His⁹⁶, and the negative charge is contributed by Asp⁹⁸, Gln¹⁰¹ and Glu²⁰ (Figure 6a). In contrast with galectin-7, the positive charges in the interface of galectin-2 come from the residues-90–100 region of the sequence instead of the N-terminus and C-terminus as in galectin-7. The interesting feature of this interface (and of galectin-7 interface type) is that the positively charged residues and negative charged residues that neutralize them always come from two different monomers, thus making the interface extremely strong and stable. Further, Ser¹⁰⁹ and Tyr¹¹¹ from both of the chains add to the stability of the cluster. An additional interface cluster comprising His¹⁴¹, Pro¹⁴⁵ (from same chain) and Leu¹⁴⁷ (from other chain) also stabilize the interface. Figure 5 shows the interface clusters in galectin-2. Although some of these residues are conserved in the other galectins, in none of them is the complete pattern conserved (Figure 6a).

Thus the factors determining the type of quaternary association in galectins can be obtained from this approach. Moreover, the signature sequence motifs that characterize each of them have also become apparent from this analysis.

PENTRAXINS

Background

The earliest described pentraxins, CRP (C-reactive protein) and SAP (serum amyloid P component), are cytokine-inducible acute-phase proteins implicated in innate immunity, whose concentrations in the blood increase dramatically upon infection or trauma [15,16]. The pentraxins are a phylogenetically ancient family of oligomeric plasma proteins, all of which bind Ca^{2+} ions. The binding of Ca^{2+} is necessary for the expression of ligand-binding activities. These proteins have evolved very little

and hence they are highly conserved. They exhibit remarkable conservation of structure and binding specificities. Since the proteins have evolved conservatively and they are present in vertebrates, this suggests an essential function for these proteins in the body. Within vertebrates, there exist two main branches of the pentraxin family, namely the CRP-like proteins and SAP-like proteins. The pentraxins that bind phosphocholine are CRP-like and those that bind carbohydrate moieties are SAP-like pentraxins. All of the pentraxins are oligomers arranged in a discoid-like pentagonal (rarely hexagonal) cyclic symmetry [50], as shown in Figure 7. These proteins were named pentraxins because of their cyclic configuration of five non-covalently bound identical subunits. These proteins consist of a hydrophobic core and have a beta-jelly roll motif as seen in the leguminous-plant lectins.

Characterization of pentraxin interfaces

The pentraxin pentamer consists of five similar dimeric interfaces contributed by five monomers. The dimeric interface of the pentraxin is shown in Figure 1. Each monomer makes two types of interactions with the two other monomers using two different faces coming from different sequential regions. Face 1 of monomer 1 interacts with face 2 of monomer 2 and the face 1 of monomer 2 interacts with face 2 of monomer 3 and so on and so forth till the pentamer is formed (Figure 7). Hence, the dimers themselves are asymmetric, since the interacting monomeric regions at the dimeric interface are different. However, the same pattern is followed subsequently in all five dimeric interfaces, to give rise to a symmetric pentamer. An example each of CRP and SAP, both from humans and whose crystal structures are available, have been considered in this analysis. A comparison of the pentraxins with the consensus patterns of all the legume lectin interface types shows that they lack the signature residues required for all these patterns (Figure 3). It is evident from Figure 3 that the pentraxins lack the motifs required for II, X1, X2, X3 or X4 interface types of the legume lectin family. The pentraxins also

(a) Multiple sequence alignment of Galectins:

```

1a78_A      ----ASAGVAVTNLNLKPGHCVEIKG---SIPPDCKGFAVNLG---EDASNFLHFNAR 49
1c1f_A      -----SGGLQVKNFDFTVGKFLT VGG---FINNSPQRFVS NVG---ESMNSLSLHLDHR 48
1lcl_A      ---MSLLPVVPTAAASLSTGSTVTIKGRPLVCFPLNEPYLQVDPHTEMKEESDIVHFHQC 57
1a3k_A      ----LIVPYNLPLPGGVVPRMLITILG---TVKPNANRIALDFQ---RGNDVAFHFNPR 49
2gal_A      ----SNVPHKSSLPEGIRPGTVLRIRG---LVPPNASRPHVNLCCGEEQGSDAALHFNPR 53
1uld_A      MLYHLFVNQVKLQNDPKPESVAAIRSS--AFNSKGGTTVFNFLS---AGENILLHISIR 55

1a78_A      FDLHGDVNKIVCNS--KEADAWGSEQREEVFPFQQAEMVCFEYQTQKIIKFFSSGDQF 107
1c1f_A      FNYGADQNTIVMNSTLKGDNWETEQRSTNFTLSAGQYFEITLSYDINKFYIDILDGPNL 108
1lcl_A      FGRR-----VVMNS--REYGAWKQVESKNMPPQDQGFELSI SVLPDKYQVMVNGQSSY 110
1a3k_A      FNENN-RRVIVCNT--KLDNNWGREERQSVFPFESGKPKFIQVLVEPDHFKVAVNDAHL 106
2gal_A      LDTS----EVVFN--KEQGSWGREERGPGVPFQRPPEVLIIASDDGFKAVVGDQAYH 107
1uld_A      PGEN----VIVFNS-RLKNGAWGPEERIPYAEKFRPPNPSTITVIDHGRFQIRFDYGTST 110

1a78_A      SFPVRKV-LPSIFFLSLEG-LAFKSITTE----- 134
1c1f_A      EFPNRY-S-KEFLPFLSLAGDARLT LVKLE----- 136
1lcl_A      TFDHRIK-PEAVKMVQVWRDISLT KFNVSYLKR----- 142
1a3k_A      QYNHRVKKLNEISKLGISGDIDLTSASVTMI----- 137
2gal_A      HFRHRLP-LARVRLVEVGGDVQLDSVRIF----- 135
1uld_A      YYNKRIKENAAAIAVNAENSLFSSPVTVDVHGLLPPLPPA 150

```

(b) Multiple sequence alignment of Pentraxins:

```

1B09_A      -----QTDMSRKAFVFPKESDTSYVSLKAPLT KPLKAF TVCLHFYTEL 44
1GYK_A      -----HTDLGSKVFPVPRESVDHVNLT PLEKPLQNFTLCFRAYS DLS 44
sp|Q07203|  MERFALWFI FLAGSLAQEDLVGNVFLFPKPSVTIT YAILKPEVEKPLKNLT VCLRSYTTLT 60
              : * : : * : : * : : * : : * : : * : : * : : * : : * : : * : :
1B09_A      STRGYSIFS YATKR--QDNEILIFNSKDIGYSFTVGGSEILFEVPEVTVA P VHICTSWES 102
1GYK_A      --RAYSLFSYNTQG--RDNELLVYKERVGEYSLYIGRHKVT SKVIEKFPAP VHI CVSWES 100
sp|Q07203|  --RFHSLLSLATS NPLQDN AFLFLSKPPNQCSIYINQEENVF KVDPTAVENKHT CVSWDS 118
              : * : : * : : * : : * : : * : : * : : * : : * : : * : : * : :
1B09_A      ASGIVEFWVDGKPRVRKSLKKGTVGAESI ILGQE QDSFGGNFEGSQSLVGDIGNVNMW 162
1GYK_A      SSGIAEFWINGTPLVKGLRQGYFVEAQPKI VLGQE QDSYGGKFPDRSQSFVGEIGDLYMW 160
sp|Q07203|  VSGVVELWIDGKLYPRTVSKKASSIGFPSSI IQGQE QDSFGGPNIDQSFVGEISDVHWM 178
              * : : * : : * : : * : : * : : * : : * : : * : : * : : * : : * : :
1B09_A      DFVLSPEINTIYLG-GPFSNVNLNRALKYEVQGEVFTKPOLWP----- 206
1GYK_A      DSVLPENILSAVQG-TPLPANILDWQALNYEIRGVII KPLVWV----- 204
sp|Q07203|  DYVLTPDHIQKVL FANMDFNGNI ISWRS LQYELRGQATTQPKRQCKTLEHHYGLFAKCYK 238
              * * : : * : : * : : * : : * : : * : : * : : * : : * : : * : :

```

Figure 6 Sequence alignments of (a) galectins and (b) pentraxins

The galectins include galectin-1 (1a78), congerin (1c1f), galectin-2 (1uld), galectin-3 (1a3k), Charcot-Leyden protein (1lcl) and galectin-7 (2gal). The pentraxins include human CRP (1B09), human SAP (1GYK) and a pentraxin from frog (*X. laevis*) (Swiss-Prot accession code Q07203), which is known to exist as a homodimer. The residues present in the interface clusters of galectins and pentraxins are highlighted and underlined. Galectin-3 (1a3k) and Charcot-Leyden protein are monomers and so have no interface clusters. The structure of the frog pentraxin is not known, and hence we do not have information regarding the interface clusters in this lectin. It is evident from the Figure that galectin-1-like interfaces (including galectin-1 and congerin), galectin-2, galectin-7 and pentraxins all have completely different residues contributing to the interfaces. These can be used for characterizing these interfaces, since they are mutually exclusive.

lack the signature motifs required for all the three galectin-like interfaces (1, 2, and 7; sequence alignments not shown).

The interface clusters obtained in the pentraxins (shown in Figure 5) are as usual mapped on to the sequence alignment of CRP with SAP (Figure 6b), and the consensus pattern of residues required for the oligomerization of the pentraxins are thus obtained (highlighted in Figure 6b). The pentraxin dimeric interface signature is contributed by residues Glu¹⁰⁸ (Glu¹⁰⁶ in SAP), Val¹¹⁷ (Val¹¹⁵ in SAP) and Arg¹¹⁸ (Lys¹¹⁶ in SAP) from one monomer and Tyr⁴⁰ (Tyr⁴⁰ in SAP), Glu⁴² (Asp⁴² in SAP), Pro⁹³ (Pro⁹¹ in SAP), Asp¹⁵⁵ (Glu¹⁵³ in SAP) and Trp²⁰⁵ (Trp²⁰³ in SAP) from the other monomer. One of the CRPs from African clawed frog (*Xenopus laevis*) is known to exist as a homodimer, but its structure is yet unknown (SwissProt accession number Q07203). This sequence was examined for the presence of the residues contributing to the pentraxin interface obtained from the human CRP and SAP analysis (also shown in Figure 6b). It can be seen

that this sequence has only two of the three residues required for one of the monomeric faces of the dimer and two out of the five residues required for the other monomeric face of the dimer. Hence, this pentraxin interface could be considerably destabilized, owing to the mutations of some critical residues leading to its dimeric nature rather than the pentameric form normally seen in the other pentraxins. Thus the study of pentraxins shows that they have a characteristic signature sequence motif different from that of the legume lectins or galectins, which determines their unique pentameric nature of association.

Calnexin/calreticulin/Vp4 sialic-acid-binding domain

Calnexin and calreticulin form part of the quality-control system for glycoproteins in the endoplasmic reticulum [51,52], which bind monoglucosylated N-glycans. They bind to terminal glucose residues on N-linked oligosaccharides and retain misfolded

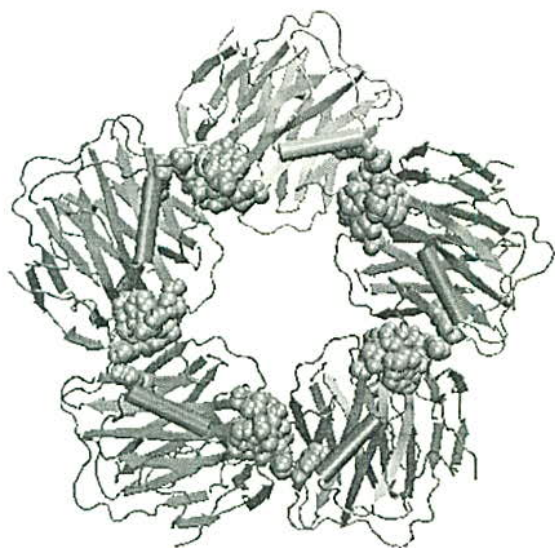


Figure 7 Discoid arrangement of pentameric pentraxin
Each monomer is shown in a differently coloured cartoon representation. The interface clusters at each of the dimeric interfaces are shown as van der Waal's spheres.

glycoproteins in the endoplasmic reticulum [51,52]. They act as chaperones for N-linked glycoproteins so as to keep the folding intermediates in a folding-competent state. Calnexin is a transmembrane protein and calreticulin is a soluble protein retained in the lumen by a C-terminal retention signal. The luminal N-terminal portion of calnexin is very similar to calreticulin, although one of the repeated segments of calnexin is absent from calreticulin. They share the same jelly-roll tertiary structure as the legume lectins. However, they remain monomeric and do not form higher oligomers. Similar to calnexin/calreticulin, the Vp4 sialic-acid-binding domain also has the lectin-like jelly-roll tertiary structure, but remains a monomer. The sequence alignments of calnexin, calreticulin and the Vp4 sialic-acid-binding domain with the legume lectins are shown in Figure 3. When these alignments were analysed for the presence of the signature sequences of the legume lectin interface types, we found that although these

proteins have some of the residues required for a few of the interface types, they lack most of the residues required for all of the legume lectin oligomerization types (II, X1, X2, X3 and X4), indicating why they remain monomeric.

NETWORK MODEL OF LECTIN INTERFACES

Apart from the amino-acid-cluster analysis at the lectin interfaces, we have also analysed the protein structure graphs of the lectin oligomers from the network perspective. This includes the size of the largest cluster in these proteins at different I_{min} values and the residue hubs obtained at their interfaces. These concepts have been presented for the first time in the present review. One representative from each legume lectin interface type, galectin interface type and pentraxin interface type has been chosen for this analysis (Table 3).

Interface hubs

In network terminology, hubs are highly connected nodes in the network. In the protein structure graphs, hubs are those residues that make more than four contacts with other residues and the interface hubs are those which make more than four contacts, with at least one residue belonging to the chain other than its own. Table 3 gives the interface hubs obtained in the ten selected lectins at $I_{min} = 4\%$ or 2% (when there are no interface hubs at $I_{min} = 4\%$). Most of these interface hubs are also present in the signature motifs of their respective interface types (Figures 3 and 6). It can be observed from Table 3 that there are nine aromatic hubs, eleven hydrophobic hubs, nine negatively charged hubs and 21 positively charged hubs. Out of the positively charged ones, 11 are arginine residues, five are histidine residues and five are lysine residues. We see a predominance of charged hubs, especially arginine, in these interfaces. Except X2, unusual GS1 and galectin-1 interface types, all the others have at least one charged hub in their interfaces. The contribution of charged hubs in these interfaces is about 60%. The hydrophobic and aromatic contributions are relatively lesser. This is significantly different from the hub preferences seen in the monomeric protein cores, where there was no clear domination of the charged residues at any I_{min} value, except a meagre excess of arginine at $I_{min} = 4\%$ (K. V. Brinda and S. Vishveshwara, unpublished work). This dominance

Table 3 Interface hubs in various lectin interface types from lectin structure graphs

Abbreviations: DGL, *Dioclea grandiflora* (mucana) lectin; PDB, Protein Data Bank.

Lectin*	PDB code	Interface types in quaternary structure†	Interface hubs (I_{min} in %)‡	I_{min} at which interface cluster is the largest (%)
ConA	2cna	II (canonical) + X2 tetramer	W88A, W88B, Q137A, Q137B (4%)	6
GS4	1gsf	X4 (back-to-back) dimer	R194A, R194B, Y72A, Y72B (4%)	6
ECoRL	1axy	X3 (handshake) dimer	I178A, I178B, H180A, H180B (2%)	4
DBL	1bjq	II + X1 tetramer	H51A, V57A, R60A, V64A, K116A, N55B, R60B (2%)	0
DGL	1dgl	II + X2 tetramer	I187A, I187B (2%)	0
GS1	1hql	X4 + unusual tetramer	W10A, W10B (4%)	0
PNA	2pel	II + X4 + unusual tetramer	N29C, Q33C, E72C, K74C, L219C, R221C, N29D, L219D, R221D (2%)	0
Galectin-1	1a78	Galectin-1 dimer	L121A, I131A, F90B, F128B (2%)	0
Galectin-2	1uld	Galectin-2 dimer	H96A, Q101A, H96B, Q101B (4%)	6
Galectin-7	2gal	Galectin-7 dimer	R20A, R22A, K98A, R20B, R22B (2%)	0
Pentraxin (CRP)	1b09	Pentraxin pentamer	F199A, K201A, L204A, R118B (2%)	4
Pentraxin(SAP)	1gyk	Pentraxin pentamer	K116A (4%)	4

* Galectin-3 (1a3k), Charcot-Leyden protein (1lcl), calreticulin (1gv9), calnexin (1jhn) and arcalin-5 (1ioa) are monomeric types and are not presented here; only one representative from the other dimeric interface types are presented; congerin (1c1f), which has an interface similar to that of galectin-1, is also not presented.

† All the interfaces contributing to the quaternary structure of the lectin are given; the specific interfaces whose hubs are given in the Table are highlighted in bold.

‡ Interface hubs are given at $I_{min} = 4\%$; if no interface hubs were obtained at $I_{min} = 4\%$, those obtained at $I_{min} = 2\%$ are given; the one-letter notation for amino acids is used.

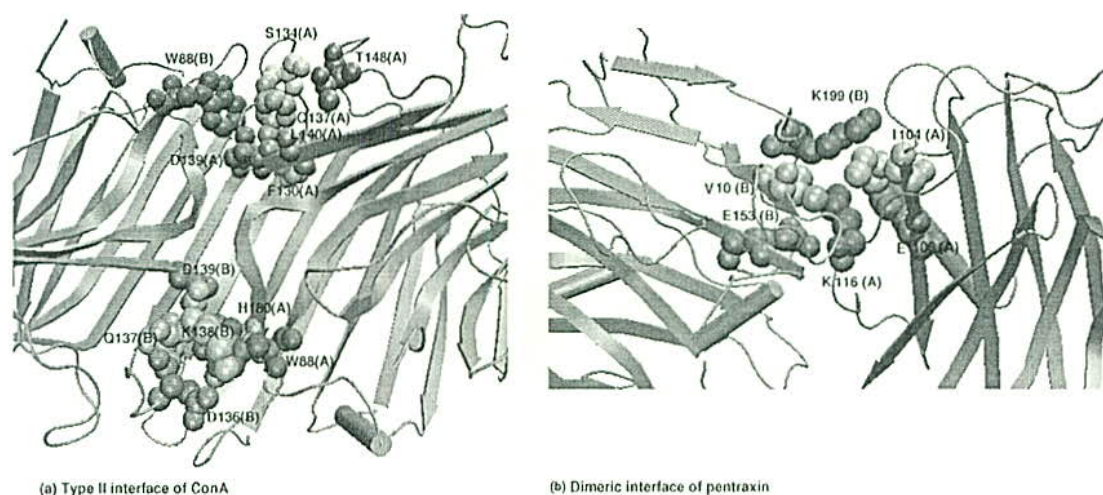


Figure 8 Interface hubs in (a) ConA (type II) interface and (b) SAP pentraxin (dimeric) interface at $I_{\min} = 4\%$

The proteins are shown in cartoon representation, and each monomer in the dimer is coloured differently. The interface hubs and the residues with which they interact are represented as van der Waals spheres with each residue being coloured differently. The residue name and number are indicated in the Figure, and the chain to which each residue belongs is given within parentheses. Using the one-letter amino acid notation, W88 (A) and Q137 (A) in ConA and K116 (A) in SAP pentraxin form the interface hubs. It can be seen from the Figure that these hubs interact with different types of residues from both the chains, thus stabilizing the dimeric interfaces.

of the polar and charged residue hubs at interfaces could be due the fact that these residues can be stabilized by interactions with water when they are exposed on the surface of the monomer and can be neutralized by oppositely charged residues when they get buried at the interface during oligomerization. However, in the monomeric proteins, the charged residues are preferentially exposed on the surface and hence the probability of their forming hubs is generally less. The preference of arginine residues at protein/protein and protein/DNA interfaces have been elucidated previously by a few groups [4,43,44]. The dominance of charged residues as hubs in the interfaces is only an observation based on the present set of lectins, and a detailed analysis of a complete non-redundant set of protein oligomers holds the conclusive results for the hub preferences at protein interfaces.

The interface hubs obtained in the canonical (type II) interface of ConA and SAP pentraxin dimeric interface are shown in Figure 8. The Figure clearly shows that the hubs play a major role in integrating the two monomers structurally through a series of non-covalent interactions with residues belonging to both monomers. Hence, these hubs contribute significantly to the stability of the interface and so a mutation of the hub can severely destabilize the interface. The hubs also provide robustness to the interface interaction network, because a random mutation of a non-hub residue is unlikely to affect the stability of the interface interaction network, whereas the mutation of a hub can have drastic effects on the oligomerization and hence the function of the protein.

Size of the largest cluster as a function of interaction cut-off ($I_{\min}$)

The normalized size-of-the-largest-cluster-versus- $I_{\min}$ plots for the selected lectins (one representative from each interface type) is shown in Figure 9. As the Figure shows, the plots are sigmoidal with a transition (where there is a drastic decrease in size) at about $I_{\min} = 3-4\%$. The largest clusters at different $I_{\min}$ values were obtained for each of these lectins. Whether the largest cluster belongs to the interface or not was examined in each of these lectins at $I_{\min}$ values varying from below the transition (0%) to above the transition (6%). Clearly, at $I_{\min} = 0\%$, the oligomer exists as one big cluster in all the lectins and hence the largest

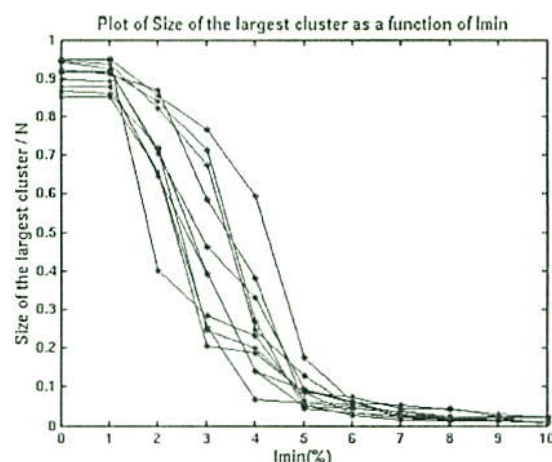


Figure 9 Plot of the size of the largest cluster (normalized with respect to the total number of residues in the protein structure, N) versus $I_{\min}$ for selected lectins

The plot clearly shows that the curve is sigmoidal in nature.

cluster includes the interface in all the proteins at $I_{\min} = 0\%$. An example is shown in Figure 10(a), where, at $I_{\min} = 0\%$, the oligomer exists as one big cluster that includes the interface. However, as the $I_{\min}$ is increased, the size of the largest cluster reduces and after the transition it splits into smaller distinct clusters. Hence, the largest cluster at and above the transition (4 and 6% respectively) may or may not belong to the interface, depending upon the strength of interface of the particular lectin. An example of a lectin with the largest cluster still at the interface at $I_{\min} = 6\%$ is shown in Figure 10(b). In the present set we find that the type II and X4 interfaces of the legume lectins and the galectin-2 type interface contribute to the largest cluster, even at $I_{\min} = 6\%$, the X3 interface of legume lectins and the dimeric interface of the pentraxins contribute to the largest cluster even at $I_{\min} = 4\%$, whereas the largest clusters of X1 and X2 interfaces of legume lectins and that of galectins 1 and 7 do not involve

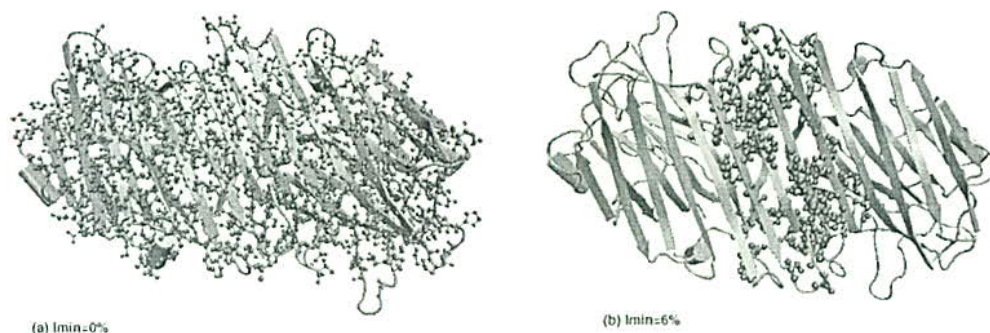


Figure 10 Largest cluster at (a) $I_{min} = 0\%$ and (b) $I_{min} = 6\%$ in ConA (2cna) type II dimer

The protein is shown in cartoon representation with the monomers coloured differently. The cluster-forming residues are shown in a grey ball-and-stick representation. At $I_{min} = 0\%$, the whole dimer forms a single large cluster, as expected. However, at $I_{min} = 6\%$, the single large cluster splits into smaller ones. In this case the largest cluster at $I_{min} = 6\%$ is an interface cluster, as can be seen in the Figure.

their oligomeric interfaces beyond $I_{min} = 0\%$. This shows that the type II, X4, galectin-2, X3 and pentraxin interfaces are as strong as their hydrophobic cores and these are considerably stronger than the X1, X2 and the galectins 1 and 7 interfaces.

Thus the I_{min} value used in the present analysis is a good measure of the strength of the interaction of the interacting residues, which can be used to estimate the strength of the interacting interfaces too. It is noteworthy that, in several cases, the oligomerization has formed by a strong interface that is almost as strong as their monomeric protein cores. Moreover, the identification of interface hubs can be a useful tool for predicting hot spots at protein interfaces, which can be mutated to considerably destabilize the interface. Hence, the interface hub analysis can aid in designing mutants in order to understand protein quaternary association and protein-protein interactions in general.

CONCLUSIONS

The present review provides a comprehensive analysis of the nature and types of quaternary associations found in plant and animal lectins. Specifically, the graph-spectral algorithm has provided the signature sequence motifs characterizing the oligomeric interfaces of legume lectins, galectins and pentraxins, thus giving insights into the factors determining the nature and type of quaternary association in each one of them. It has clearly elucidated why each one of them prefers a specific type of association and why the other known types are excluded in these lectins. The prediction of the oligomerization modes of lectins with unknown structures using the signature motifs established using the present method is also very rational and successful.

The interface hubs identified using the protein structure graphs provide a simple yet novel tool for identifying hot spots at protein interfaces. The graph representation of protein structures based on the strength of the non-covalent interactions among amino acid residues also aids in estimating the strengths of protein interfaces. Thus the graph theoretical algorithm presented in combination with traditional sequence-alignment methods is found to be extremely successful in identifying the signature motifs for a given type of quaternary association, has a very high predictive value and can be used in designing mutants that can destabilize the oligomeric interface. The present study has also aided in assessing the relative strengths of interactions at different types of lectin interfaces.

The representation of protein structures as a network of non-covalently interacting amino acid residues is an interesting way to

analyse both monomeric and oligomeric protein structures, which gives valuable insights into the tertiary and quaternary structures of proteins. This can be a useful tool to understand protein-protein interactions and various other aspects of protein structure.

A.S. thanks the Department of Biotechnology (DBT), India, for funding this project. S.V. acknowledges the computational genomics initiative at the Indian Institute of Science, funded by the Department of Biotechnology, India, for support. K.V.B. thanks the Council of Scientific and Industrial Research, India, for the award of the fellowship.

REFERENCES

- 1 Fernandez, A. and Scheraga, H. A. (2003) Insufficiently dehydrated hydrogen bonds as determinants of protein interactions. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 113–118
- 2 Xu, D., Tsai, C. J. and Nussinov, R. (1997) Hydrogen bonds and salt bridges across protein-protein interfaces. *Protein Eng.* **10**, 999–1012
- 3 Chakrabarti, P. and Janin, J. (2002) Dissecting protein-protein recognition sites. *Proteins* **47**, 334–343
- 4 Bahadur, R. P., Chakrabarti, P., Rodier, F. and Janin, J. (2003) Dissecting subunit interfaces in homodimeric proteins. *Proteins* **53**, 708–719
- 5 Janin, J. and Wodak, S. J. (2002) Protein modules and protein-protein interaction. *Introduction. Adv. Protein Chem.* **61**, 1–8
- 6 Valdar, W. S. and Thornton, J. M. (2001) Protein-protein interfaces: analysis of amino acid conservation in homodimers. *Proteins Struct. Funct. Genet.* **42**, 108–124
- 7 Ma, B., Elkayam, T., Wolfson, H. and Nussinov, R. (2003) Protein-protein interactions: structurally conserved residues distinguish between binding sites and exposed protein surfaces. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 5772–5777
- 8 Jones, S. and Thornton, J. M. (1997) Analysis of protein-protein interaction sites using surface patches. *J. Mol. Biol.* **272**, 121–132
- 9 Smith, G. R. and Sternberg, M. J. (2002) Prediction of protein-protein interactions by docking methods. *Curr. Opin. Struct. Biol.* **12**, 28–35
- 10 Camacho, C. J. and Vajda, S. (2002) Protein-protein association kinetics and protein docking. *Curr. Opin. Struct. Biol.* **12**, 36–40
- 11 Valencia, A. and Pazos, F. (2002) Computational methods for the prediction of protein interactions. *Curr. Opin. Struct. Biol.* **12**, 368–373
- 12 Kortemme, T. and Baker, D. (2004) Computational design of protein-protein interactions. *Curr. Opin. Chem. Biol.* **8**, 91–97
- 13 Cole, C. and Warwicker, J. (2002) Side-chain conformational entropy at protein-protein interfaces. *Protein Sci.* **11**, 2860–2870
- 14 Wodak, S. J. and Mendez, R. (2004) Prediction of protein-protein interactions: the CAPRI experiment, its evaluation and implications. *Curr. Opin. Struct. Biol.* **14**, 242–249
- 15 Liener, I. E., Sharon, N. and Goldstein, I. J. (1986) *The Lectins: Properties, Functions and Applications in Biology and Medicine*, Academic Press, New York
- 16 Varki, A., Cummings, R., Esko, J., Freeze, H., Hart, G. and Marth, J. (eds) (2002) *Essentials in Glycobiology*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY
- 17 Murzin, A. G., Brenner, S. E., Hubbard, T. and Chothia, C. (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* **247**, 536–540
- 18 Loris, R., Hamelryck, T., Bouckaert, J. and Wyns, L. (1999) Legume lectin structure. *Biochim. Biophys. Acta* **1383**, 9–36

- 19 Vijayan, M. and Chandra, N. (1999) Lectins. *Curr. Opin. Struct. Biol.* **9**, 707–714
- 20 Elgavish, S. and Shaanan, B. (2001) Chemical characteristics of dimer interfaces in the legume lectin family. *Protein Sci.* **10**, 753–761
- 21 Manoj, N. and Suguna, K. (2001) Signature of quaternary structure in the sequences of legume lectins. *Protein Eng.* **10**, 735–745
- 22 Srinivas, V. R., Reddy, G. B., Ahmad, N., Swaminathan, C. P., Mitra, N. and Surolia, A. (2001) Legume lectin family, the 'natural mutants of the quaternary state', provide insights into the relationship between protein stability and oligomerization. *Biochim. Biophys. Acta* **1527**, 102–111
- 23 Brinda, K. V., Mitra, N., Surolia, A. and Vishveshwara, S. (2004) Determinants of quaternary association in legume lectins. *Protein Sci.* **13**, 1735–1749
- 24 Barabasi, A. L. (2002) *Linked: the new science of networks*, Perseus Publishing, Cambridge, MA
- 25 Wuchty, S., Ravasz, E. and Barabási, A.-L. (2003) The Architecture of Biological Networks. In *Complex Systems in Biomedicine* (Deisboeck, T. S., Yasha Kresh, J. and Kepler, T. B., eds.), Kluwer Academic Publishing, New York
- 26 Dokholyan, N. V., Shakhnovich, B. and Shakhnovich, E. I. (2002) Expanding protein universe and its origin from the biological Big Bang. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 14132–14136
- 27 Grindley, H. M., Artymiuk, P. J., Rice, D. W. and Willett, P. (1993) Identification of tertiary structure resemblance in proteins using a maximal common subgraph isomorphism algorithm. *J. Mol. Biol.* **229**, 707–721
- 28 Przytycka, T., Srinivasan, R. and Rose, G. D. (2002) Recursive domains in proteins. *Protein Sci.* **11**, 409–417
- 29 Atilgan, A. R., Akan, P. and Baysal, C. (2004) Small-world communication of residues and significance for protein dynamics. *Biophys. J.* **86**, 85–91
- 30 Greene, L. H. and Higman, V. A. (2003) Uncovering network systems within protein structures. *J. Mol. Biol.* **334**, 781–791
- 31 Vendruscolo, M., Paci, E., Dobson, C. M. and Karplus, M. (2001) Three key residues form a critical contact network in a protein folding transition state. *Nature (London)* **409**, 641–645
- 32 Vendruscolo, M., Dokholyan, N. V., Paci, E. and Karplus, M. (2002) Small-world view of the amino acids that play a key role in protein folding. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* **65**, 061910
- 33 Dokholyan, N. V., Li, L., Ding, F. and Shakhnovich, E. I. (2002) Topological determinants of protein folding. *Proc. Natl. Acad. Sci. U.S.A.* **99**, 8637–8641
- 34 Bahar, I., Atilgan, A. R. and Erman, B. (1997) Direct evaluation of thermal fluctuations in proteins using a single-parameter harmonic potential. *Fold. Des.* **2**, 173–181
- 35 Bahar, I., Atilgan, A. R., Demirel, M. C. and Erman, B. (1998) Vibrational dynamics of folded proteins: significance of slow and fast motions in relation to function and stability. *Phys. Rev. Lett.* **80**, 2733–2736
- 36 Samudrala, R. and Moult, J. (1997) Handling context-sensitivity in protein structures using graph theory: *bona fide* prediction. *Proteins: Struct. Funct. Genet.* **1**, 43–49
- 37 Samudrala, R. and Moult, J. (1998) A graph-theoretic algorithm for comparative modeling of protein structure. *J. Mol. Biol.* **279**, 287–302
- 38 Sistla, R. K., Brinda, K. V. and Vishveshwara, S. (2005) Identification of domains and domain interface residues in multidomain proteins from graph spectral method. *Proteins* **59**, 616–626
- 39 Reichmann, D., Rahat, O., Albeck, S., Meged, R., Dym, O. and Schreiber, G. (2005) The modular architecture of protein–protein binding interfaces. *Proc. Natl. Acad. Sci. U.S.A.* **102**, 57–62
- 40 Kannan, N. and Vishveshwara, S. (1999) Identification of side-chain clusters in protein structures by a graph spectral method. *J. Mol. Biol.* **292**, 441–464
- 41 Vishveshwara, S., Brinda, K. V. and Kannan, N. (2002) Protein structure: insights from graph theory. *J. Theor. Comp. Chem.* **1**, 187–211
- 42 Kannan, N. and Vishveshwara, S. (2000) Aromatic clusters: a determinant of thermal stability of thermophilic proteins. *Protein Eng.* **13**, 753–761
- 43 Brinda, K. V., Kannan, N. and Vishveshwara, S. (2002) Analysis of homodimeric protein interfaces by graph-spectral methods. *Protein Eng.* **4**, 265–277
- 44 Sathyapriya, R. and Vishveshwara, S. (2004) Interaction of DNA with clusters of amino acids in proteins. *Nucleic Acids Res.* **32**, 4109–4118
- 45 Kannan, N., Chander, P., Ghosh, P., Vishveshwara, S. and Chatterji, D. (2001) Stabilizing interactions in the dimer interface of α -subunit in *Escherichia coli* RNA polymerase: a graph spectral and point mutation study. *Protein Sci.* **10**, 46–54
- 46 Berman, H. M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T. N., Weissig, H., Shindyalov, I. N. and Bourne, P. E. (2000) The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242
- 47 Reference deleted
- 48 Thompson, J. D., Higgins, D. G. and Gibson, T. J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680
- 49 Walser, P. J., Haebel, P. W., Kunzler, M., Sargent, D., Kues, U., Aebi, M. and Ban, N. (2004) Structure and functional analysis of the fungal galectin CGL2. *Structure* **12**, 689–702
- 50 Kilpatrick, J. M. and Volanakis, J. E. (1991) Molecular genetics, structure, and function of C-reactive protein. *Immunol. Res.* **10**, 43–53
- 51 Trombetta, E. S. and Helenius, A. (1998) Lectins as chaperones in glycoprotein folding. *Curr. Opin. Struct. Biol.* **8**, 587–592
- 52 Parodi, A. J. (2000) Protein glucosylation and its role in protein folding. *Annu. Rev. Biochem.* **69**, 69–93