

SUPERINTENDENCIA DE INDUSTRIA Y COMERCIO



No. 13-274153- -00008-0000

Fecha: 2015-06-11 16:41:29 Dep. 2020 DIR.NUEVASCR
Tra. 11 PATENTEIYII Eve: 378 EASENACIONALI
Act. 523 RESPUESTA Folios: 16 -17

Señor

SUPERINTENDENTE DE INDUSTRIA Y COMERCIO

DIVISIÓN DE NUEVAS CREACIONES

E. S. D.

Trámite: 11
Evento: 378
Actuación: 523

Re: PI.-THE GOVERNMENT OF THE UNITED STATES OF AMERICA AS
REPRESENTED BY THE SECRETARY OF THE DEPARTMENT OF HEALTH
AND HUMAN SERVICES.-Solicitud de Patente de Invención en Fase Nacional
PCT para "INMUNOTOXINA RECOMBINANTE DIRIGIDA A LA
MESOTELINA".-Clase A61K 39/00.-Expediente número 13-274.153.-

1. Me refiero al auto dictado por ese Despacho notificado por fijación en lista del 17 de marzo de 2015.

2. Con el fin de responder las objeciones expuestas por el Señor Examinador en su Concepto Técnico correspondiente al Oficio No. **2583**, en relación con las Excepciones a la Patentabilidad, los Usos, la Claridad del Capítulo Reivindicatorio y del Capítulo Descriptivo, la Unidad de Invención y el Nivel Inventivo de la presente solicitud, me permito dirigir su atención a las siguientes modificaciones y argumentos técnicos por medio de los cuales no sólo se superan completamente dichas objeciones sino que además se establece de manera incontrovertible la patentabilidad de la solicitud en estudio.

3. CAPÍTULO REIVINDICATORIO.

3.1. MODIFICACIONES AL CAPÍTULO REIVINDICATORIO.

Inicialmente, me permito manifestar que es deseo del solicitante presentar un **capítulo reivindicatorio modificado**, que reemplaza al actualmente presentado, esto con el fin de superar las objeciones presentadas como parte del concepto técnico realizado por ese Despacho. Tales modificaciones se exponen a continuación:

Nueva Reivindicación 1. Corresponde a la reivindicación 1 actualmente presentada, sin embargo ahora se ha modificado de la siguiente manera:

- ✓ se ha reemplazado la sigla PE por su significado, es decir: "exotoxina de *Pseudomona*";

- ✓ se han incluido las limitaciones de la reivindicación 4 actualmente presentada, es decir que ahora se especifican claramente las 6 CDRs que definen el anticuerpo correspondiente a la fracción M;
- ✓ se ha especificado que L1 une el carboxilo terminal del anticuerpo con el N-terminal del FCS; y
- ✓ con relación al dominio funcional III de la exotoxina de *Pseudomonas*, se han definido determinadas sustituciones opcionales de aminoácidos, lo cual encuentra pleno soporte en la página 37 de la descripción.

Nueva Reivindicación 2. Corresponde a la reivindicación 2 actualmente presentada.

Nueva Reivindicación 3. Corresponde a la reivindicación 3 actualmente presentada, sin embargo se ha modificado para especificar las secuencias de aminoácidos de cadena pesada y ligera de conformidad con la reivindicación 14 actualmente presentada.

Nuevas Reivindicaciones 4 a 13. Corresponden respectivamente a las reivindicaciones 6 a 13, 15 y 16 actualmente presentadas, sin embargo, su dependencia ha sido adaptada a la nueva numeración y algunas de ellas incluyen modificaciones menores como por ejemplo la aclaración de la sigla PE mediante la frase "exotoxina de *Pseudomona*", entre otras.

Nueva Reivindicación 14. Corresponde a la reivindicación 17 actualmente presentada, sin embargo se ha modificado para especificar la proporción de la molécula reclamada dentro de la composición, lo cual encuentra pleno sustento en la página 54 de la descripción.

Nueva Reivindicación 15. Corresponde a la reivindicación 23 actualmente presentada, sin embargo ahora con relación al dominio funcional III de la exotoxina de *Pseudomonas*, se han definido determinadas sustituciones opcionales de aminoácidos, lo cual encuentra pleno soporte en la página 37 de la descripción.

Las reivindicaciones 4, 5, 14, 18-22 y 24 se han eliminado del capítulo reivindicatorio.

De esta forma, el **capítulo reivindicatorio modificado** está constituido por **15 reivindicaciones** que se refieren de manera más clara y concisa a las características técnicas esenciales de la molécula quimérica y el polipéptido de fusión que son objeto de la presente solicitud, de acuerdo con las enseñanzas de la descripción.

3.2. CLARIDAD DEL CAPÍTULO REIVINDICATORIO.

Como parte de su concepto técnico, el señor Examinador objetó la claridad del capítulo reivindicatorio actualmente presentado, resaltando diferentes observaciones para determinadas reivindicaciones. En tal sentido, a continuación me permito responder a tales observaciones sobre cada una de las reivindicaciones objetadas:

Reivindicaciones 1 y 23:

El señor Examinador señala que la molécula quimérica de la reivindicación 1 no está correctamente caracterizada, ya que ésta debe reclamarse por su secuencia de aminoácidos completa con un número de identificación y no a partir de una caracterización general de cada uno de sus componentes.

Frente a lo anterior, llamo la atención del señor Examinador sobre el capítulo reivindicatorio modificado que se presenta con este memorial, en donde ahora se define detalladamente cada uno de los componentes de la fórmula M-L1-FCS-FL-dominio funcional III de la PE. Para una mayor claridad, a continuación me permito referirme a cada uno de los componentes de la mencionada fórmula M-L1-FCS-FL-dominio funcional III de la PE:

(i) Anticuerpo que se une a mesotelina (M):

De acuerdo con el señor Examinador, la fracción M, que corresponde a un anticuerpo, sólo se define en términos de su función y considera que tal anticuerpo M se debe definir en términos de sus cadenas variables ligera y pesada o a partir de sus seis CDRs. Adicionalmente indica que no es claro en cual parte del anticuerpo se enlaza L1-FCS-FL-dominio funcional III de la PE.

Tomando en cuenta el punto de vista del señor Examinador, me permito resaltar que la reivindicación 1 ha sido modificada con el fin de especificar las CDRs del anticuerpo, las cuales corresponden a las descritas en la reivindicación 4 actualmente presentada.

Ahora bien, respecto a la unión del anticuerpo con la fracción L1-FCS-FL-dominio funcional III de la PE, cabe señalar que la nueva reivindicación 1 también se ha modificado para especificar que L1 une el carboxilo terminal del anticuerpo con el N-terminal del FCS, lo cual encuentra pleno sustento en las páginas 45 y 46 de la descripción.

(ii) L1:

En opinión del señor Examinador, no están definidos los 10 residuos de aminoácidos que corresponden a L1, además, indica que según la descripción, sólo existiría sustento para las SEQ ID NO. 19 o SEQ ID NO: 68.

Frente a tal afirmación, respetuosamente me permito señalar que L1 es un enlazador o "linker" que no tiene una secuencia específica, y que une el anticuerpo M al FCS de la molécula reivindicada. En este sentido, queda claro que el enlazador L1 no tiene una función atribuida a una secuencia específica, más bien, la función del enlazador L1 es exclusivamente unir el anticuerpo a la secuencia FCS.

Efectivamente, la función de unir el anticuerpo a la secuencia FCS puede llevarse a cabo mediante cualquier secuencia de 1 a 10 residuos de aminoácidos continuos, como bien se reivindica. En ese orden de ideas, la reivindicación 1 define claramente que L1 (i) une el carboxilo terminal del anticuerpo con el amino terminal del FCS y (ii) que consiste de 1 a 10 residuos de aminoácidos continuos.

Por consiguiente, una persona versada en la materia puede determinar fácilmente el alcance del linker L1 aquí reclamado. Por lo tanto, la objeción de en este sentido ha sido superada.

(iii) FCS:

De otra parte, el señor Examinador señala que la fracción FCS, tampoco se encuentra caracterizada por sus aminoácidos.

Con relación a lo anterior, muy respetuosamente me permito indicar que los múltiples sitios de corte de furina ya se conocían en el estado de la técnica al momento de presentación de esta solicitud, como se evidencia, por ejemplo, en el documento de Duckert et al¹, el cual se adjunta al presente memorial; y por ende la persona versada en la materia ya tenía conocimiento de este tipo de sitios de corte.

Adicionalmente, cabe resaltar que el capítulo descriptivo proporciona una descripción detallada de los sitios de corte de furina, por ejemplo, en las páginas 33 a 36.

Ahora bien, como se explica en la página 34 de la descripción, el documento de Duckert et al. resume la información disponible sobre las secuencias de 38 proteínas que en la literatura se informa tienen sitios de corte por furina.

Es así como, tanto la descripción como el documento de Duckert et al. explican que habitualmente, el sitio de corte mínimo es R-X-X-R, ocurriendo el corte después de la

¹ Duckert et al. "Protein Engineering, Design & Selection 17(1):107-112 (2004).

segunda "R" (Ver la página 34 de la descripción y la pág. 109 del documento de Duckert et al).

Con referencia a la secuencia P6...P5' que se menciona en la página 34 de la descripción, en la presente solicitud y en el documento de Duckert se explica que una arginina menos favorable en la posición P4 puede ser compensada por los residuos arginina o lisina en las posiciones P2 y P6 (Ver la pág. 109 del documento de Duckert et al).

El capítulo descriptivo también proporciona 22 ejemplos de secuencias FCS como las SEQ ID NOS: 17 y 29-49 (en las páginas 34 a 36 de la descripción), así como las modificaciones específicas que se pueden incorporar a dichas secuencias sin perder la función. Las secuencias FCS que son adecuados también se describen en los párrafos [0032] - [0038] del documento WO 2009/032954, el cual se adjunta al presente escrito.

De igual forma, la descripción de la presente solicitud explica que mediante métodos conocidos en la técnica se puede determinar si cualquier secuencia particular puede o no ser cortada por furina. En este sentido, la descripción menciona, por ejemplo, que incubando la secuencia con la furina en un tampón de furina (NaOAc 0,2 M (pH 5,5), CaCl₂ 5 mM) en una relación molar 1:10 de enzima/sustrato a 25°C por 16 horas, se puede probar si una secuencia puede o no ser cortada por furina (ver la página 36 de la descripción). Además, la descripción también indica que se puede probar fácilmente cualquier secuencia particular que puede ser cortada por furina mediante su inclusión en una inmunotoxina y sometiendo a prueba la inmunotoxina resultante *in vitro* en una línea celular de mesotelina⁺ (ver la página 36 de la descripción).

Consecuentemente, en vista de la descripción detallada de los sitios de corte de furina en la descripción y en las referencias citadas (Duckert et al. y WO2009/032954), una persona versada en la materia comprendería fácilmente cuales secuencias pueden ser cortadas por furina. Por lo tanto, el término "sitio de corte de furina" que es empleado en las reivindicaciones, es claro y bien conocido para la persona versada en la materia, por lo que la objeción en este sentido ha sido superada.

(iv) FL:

Asimismo, el señor Examinador afirma que la fracción FL, no corresponde a cualquier secuencia de aminoácidos seleccionados independiente de serinas y glicinas, puesto que en la descripción sólo existe soporte para las secuencias SEQ ID NO: 18, 50, 69 y 70. Además, el señor Examinador añade que la fracción FL en la reivindicación 23 no corresponde a cualquier secuencia de aminoácidos seleccionados independiente de serinas y glicinas, puesto que en la descripción sólo existe soporte para las secuencias SEQ ID NO: 18 y 71 a 73.

Frente a la posición del señor Examinador, amablemente me permito afirmar que el alcance de las reivindicaciones con respecto a FL es claro. Primeramente, cabe resaltar que la longitud de la secuencia FL que se menciona en las reivindicaciones es corta, es decir, solamente consta de 3 a 8 residuos de aminoácidos. Y en segundo lugar, las reivindicaciones mencionan que sólo hay dos opciones para cada residuo de aminoácido de FL, es decir, glicina y serina. En consecuencia, el número total de secuencias comprendidas por la fracción FL que se reivindica de ninguna manera puede considerarse ilimitado.

Por otro lado, debe tenerse en cuenta que la descripción de la presente solicitud, proporciona un total de 11 ejemplos de secuencias FL, no sólo siete, como menciona el señor Examinador. En este sentido, sumado a las secuencias SEC ID N°s: 18, 50, 69 y 70 a 73, la descripción también proporciona soporte para los siguientes cuatro ejemplos de secuencias FL:

1. GSG (reivindicación 10 originalmente presentada);
2. GGG (reivindicación 10 originalmente presentada);
3. GGS (pág. 46, segundo párrafo); y
4. GSGGGS (En la Figura 1B, la porción subrayada al lado derecho de la secuencia correspondiente a SS1-LR/GGSx2, y en la SEQ ID NO: 22).

A la luz de lo anterior, es evidente que el alcance de las reivindicaciones modificadas 1 y 23 (Nuevas reivindicaciones 1 y 15) con respecto a FL es claro, y por lo tanto la objeción al respecto ha sido superada.

(v) Dominio funcional III de la exotoxina de Pseudomonas:

Ahora bien, con respecto al Dominio funcional III de la PE, el señor Examinador afirma que las sustituciones comprendidas en tal dominio deben caracterizarse y no pueden mencionarse de manera general. El señor Examinador también indica que la expresión "... mantienen la inmunogenicidad de un epítipo..." es un resultado a alcanzar, que no debe ser mencionado en las reivindicaciones.

Tomando en cuenta el punto de vista del señor Examinador, me permito señalar que las reivindicaciones 1 y 23 han sido modificadas con el fin de especificar la sustitución de residuos de aminoácidos específicos de la SEQ ID NO: 1, así como la sustitución con Ala, Gly, Ser o Glu, y también con el fin de eliminar la expresión "... mantienen la inmunogenicidad de un epítipo...". De esta manera las reivindicaciones 1 y 23 (Nuevas reivindicaciones 1 y 15) ya no mencionan un dominio III de la PE que comprende sustituciones indefinidas.

De igual forma, es importante destacar que en las reivindicaciones 1, 8, 9, 13 y 23 (nuevas reivindicaciones 1, 6, 7, 11 y 15) se ha remplazado la frase "dominio funcional III de la PE" por " *Dominio funcional III de la exotoxina de Pseudomonas*", como lo sugirió el señor Examinador. Por lo tanto, las objeciones realizadas al respecto no proceden en vista de las modificaciones realizadas.

Reivindicación 2:

El señor Examinador objetó la reivindicación 2 dado que, en su opinión no existe sustento para todas las realizaciones allí reclamadas e indica que sólo existe sustento para un fragmento de proteína Fv estabilizado por disulfuro (dsFv).

Frente a tal objeción, me permito resaltar que la reivindicación 1 se ha modificado para indicar las seis CDRs del anticuerpo que se une específicamente a mesotelina. En este sentido, cabe señalar que todos los anticuerpos y fragmentos de anticuerpo mencionados en la reivindicación 2 tienen las CDRs que se definen en la nueva reivindicación 1, de la cual depende la nueva reivindicación 2.

Como es bien conocido por una persona versada en la materia, las regiones CDR determinan la especificidad de los anticuerpos y fragmentos de anticuerpos. Por consiguiente, comprendería fácilmente que un minicuerpo, un diacuerpo, un triacuerpo, un fragmento Fab', un fragmento F(ab)'₂ y una proteína Fv de cadena sencilla ("scFv") que tienen las mismas CDRs que el fragmento de proteína Fv estabilizado por disulfuro ("dsFv") tendrían la misma especificidad anti-mesotelina que el dsFv que se reclama y para el cual el señor Examinador ha reconocido pleno sustento.

Así, la persona versada en la materia también esperaría que una molécula química que comprende un minicuerpo, un diacuerpo, un triacuerpo, un fragmento Fab, un fragmento F(ab)'₂ y un scFv que tienen las mismas CDRs que el dsFv de reivindicación 2, también proporcionen la misma citotoxicidad contra las células objetivo.

Dada las anteriores razones, solicito amablemente que la objeción realizada sea retirada.

Reivindicación 3:

El señor Examinador señala que la reivindicación 3 se refiere de forma general a las regiones variables de inmunoglobulina comprendidas en el anticuerpo, sin describir su secuencia.

Tomando en cuenta la posición del señor Examinador, se ha modificado la reivindicación 3 con el fin de indicar que la región variable de la cadena ligera comprende

la secuencia de aminoácidos SEQ ID NO: 6 y que la región variable de la cadena pesada comprende la secuencia de aminoácidos SEQ ID NO: 7. Por lo tanto la objeción realizada ha sido superada.

Reivindicación 5:

El señor Examinador señala que la reivindicación 5 es redundante puesto que la CDR3 de secuencia SEQ ID NO: 57 ya había sido reclamada en la reivindicación 4.

Frente a lo anterior, me permito manifestar que la reivindicación 5 no hace parte del capítulo reivindicatorio modificado que se presenta junto con este memorial. En consecuencia la anterior objeción ha sido superada.

Reivindicaciones 14 y 15:

El Examinador indica que las reivindicaciones 14 y 15 no deben referirse a nombres o nomenclatura propia (SS1) para describir a los anticuerpos y que tales nombre deben reemplazarse por el número de identificación de las secuencias.

En conexión con tal señalamiento, cabe resaltar que la reivindicación 14 ha sido eliminada del capítulo reivindicatorio modificado que se presenta junto con este memorial y que por su parte el nombre "SS1" se ha eliminado de la reivindicación 15 (nueva reivindicación 12).

En este sentido, también debe tenerse presente que las cadenas pesada y ligera se encuentran claramente definidas por medio de sus SEQ ID NOs en la reivindicación 15 modificada (nueva reivindicación 12). En consecuencia la anterior objeción ha sido superada.

Reivindicación 17:

Respecto a la reivindicación 17 (nueva reivindicación 14), el señor Examinador indica que una composición debe caracterizarse a partir de sus ingredientes (principio activo y excipientes) y las proporciones de los mismos.

Frente al señalamiento anterior, amablemente me permito resaltar que la reivindicación 17 ya indicaba que la composición comprende la molécula de cualquiera de las reivindicaciones 1 a 16. En consecuencia, la reivindicación 17 en efecto indicaba claramente el ingrediente activo de la composición. No obstante, la reivindicación objetada se ha modificado con el fin de especificar que la composición comprende de 0,1% a 75% del ingrediente activo que en este caso corresponde a la molécula reivindicada.

Ahora bien, la descripción de la presente solicitud, también contiene una discusión detallada de los excipientes farmacéuticamente aceptables que pueden incluirse en la composición reivindicada, particularmente en las páginas 53 a 58. En la descripción igualmente se describen las diversas vías de administración, así como ejemplos de excipientes farmacéuticamente aceptables adecuados.

Por lo tanto, las objeciones realizadas al respecto no proceden en vista de las modificaciones realizadas.

Reivindicaciones 18, 19 y 24:

El señor Examinador señala que la molécula de ácido nucleico de las reivindicaciones 18 y 24 debe caracterizarse en términos de la secuencia de nucleótidos y no por el polipéptido que codifica. Además, indica que el vector de la reivindicación 19 debe ser reclamado con una identificación y una estructura definida, indicando al menos el tipo de vector y la secuencia del ácido nucleico que contiene.

Frente a lo anterior, me permito manifestar que el Solicitante ha decidido eliminar las reivindicaciones 18, 19, y 24 y por lo tanto éstas no hacen parte del capítulo reivindicatorio modificado que se presenta junto con este memorial. En consecuencia la anterior objeción ha sido superada.

En vista de todo lo discutido anteriormente, resulta evidente que el capítulo reivindicatorio modificado supera completamente las objeciones presentadas por el Señor Examinador y se enmarca completamente dentro de lo preceptuado en el Artículo 30 de la Decisión Andina 486 de 2000.

4. EXCEPCIÓN A LA PATENTABILIDAD.

De otra parte, el señor Examinador indica que el objeto de la reivindicación 20 actualmente presentada no es patentable dado que, en su opinión, estaría haciendo referencia a una célula que no está definida en términos de una línea celular.

Sobre este particular me permito señalar que tal y como se expuso en el numeral 3.1 del presente memorial, la reivindicación 20 no hace parte del capítulo reivindicatorio modificado.

Por lo anterior, dicho capítulo reivindicatorio modificado no está de ninguna manera dirigido a materia no patentable y en consecuencia, se ajusta completamente a los requerimientos del Artículo 20 de la Decisión Andina 486 de 2000.

5. REIVINDICACIONES RELATIVAS A USOS.

Asimismo, el señor Examinador señala que el objeto de las reivindicaciones 21 y 22 actualmente presentadas no es patentable, debido a que éstas se relacionan con una molécula para ser usada como un medicamento para tratar un cáncer.

En este sentido me permito resaltar que tal y como se indica en el numeral 3.1., las reivindicaciones 21 y 22 fueron eliminadas del capítulo reivindicatorio.

Por lo tanto, el mencionado capítulo reivindicatorio modificado no se refiere de ninguna manera a usos y se enmarca así dentro de lo establecido en los Artículos 14 y 21 de la Decisión Andina 486 de 2000.

6. CAPÍTULO DESCRIPTIVO.

De otro lado, el señor Examinador considera que existe una inconsistencia entre la materia reclamada y la materia divulgada en la descripción, dado que en su opinión, en los ejemplos de la memoria descriptiva sólo se presentan resultados para una realización de molécula química denominada SS1-LR/GGS/8M, pero en las reivindicaciones se reclama una molécula química con una estructura indefinida y que no está limitada a la modalidad que fue probada.

Frente a la anterior objeción, en primer lugar me permito resaltar que contrario a lo que menciona el señor Examinador, la molécula química que se reclama en el capítulo reivindicatorio modificado posee una estructura bien definida, dado que la nueva reivindicación independiente 1 indica que el objeto de la invención corresponde a una molécula química que comprende un polipéptido de fusión que tiene la fórmula M-L1-FCS-FL-dominio funcional III de la exotoxina de Pseudomonas, en donde cada una de las partes, tales como M, L1, FCS, FL y el dominio funcional III de la exotoxina de Pseudomonas, se encuentra claramente definida en la misma reivindicación 1, y a lo largo del capítulo reivindicatorio. Vale la pena destacar que la materia reclamada se encuentra plenamente en concordancia con las divulgaciones de la descripción.

En segundo lugar, amablemente me permito manifestar que de ninguna manera las reivindicaciones deben estar limitadas a las modalidades específicas ensayadas en los ejemplos. En realidad, el solicitante tiene derecho a reclamar un alcance razonable en las reivindicaciones, el cual no está limitado únicamente a las realizaciones específicas analizadas en los ejemplos.

Cabe resaltar que, con base en la información contenida en la presente solicitud, una persona versada en la materia esperaría que otras modalidades que caen dentro del alcance del capítulo reivindicatorio, también proporcionen el efecto técnico de mejorar la citotoxicidad.

Además, teniendo en cuenta que el Solicitante desarrolló el linker corto "FL" que aquí se reclama y evidenció su efecto técnico sorprendente, es claro que también tiene derecho a reclamar, no sólo las realizaciones exactas probadas en los ejemplos, sino también un alcance razonable de moléculas adicionales, que se relacionan con aquellas realizaciones exactas ensayadas, las cuales ahora se espera de manera razonable que sean consistentes con el desarrollo del Solicitante.

Por lo tanto, es evidente que el capítulo descriptivo se ajusta totalmente a los requerimientos del Artículo 28 de la Decisión Andina 486 de 2000.

7. **UNIDAD DE INVENCION.**

Adicionalmente, el Señor Examinador considera que el capítulo reivindicatorio actualmente presentado carecería de unidad de invención, señalando que éste comprendería al menos 4 grupos inventivos.

Al respecto, amablemente me permito señalar que, de acuerdo con el Art. 25 de la D. 486 de 2000: *"una solicitud de patente sólo podrá comprender una invención o un grupo de invenciones relacionadas entre sí, de manera que conformen un único concepto inventivo"*. (Resaltado fuera de texto).

En este sentido, también es apropiado mencionar que la Guía para Examen de Solicitudes de Patente de Invención y Modelo de Utilidad de la SIC, en el numeral 2.11.1, menciona que: *"Se debe entender por "único concepto inventivo" al **conjunto de características (o elementos) técnicas nuevas e inventivas que son comunes a todas las invenciones**"* (Resaltado fuera de texto).

De acuerdo con lo anterior, cabe resaltar que para el caso que no ocupa, el "único concepto inventivo" común que unifica todas las reivindicaciones modificadas se puede definir como la inserción de un linker peptídico corto y flexible ("FL") de 3 a 8 aminoácidos de longitud y consistente en residuos de glicina y/o serina, en secuencia entre el sitio de corte de furina (FCS) y el dominio funcional III de la exotoxina de *Pseudomonas*, tal como se reivindica.

En este sentido, cabe resaltar que la molécula química reivindicada se caracteriza por tener una citotoxicidad mejorada para las células que expresan la mesotelina, y el polipéptido de fusión reivindicado también se caracteriza por tener una citotoxicidad mejorada. Asimismo, las moléculas de exotoxina de *Pseudomonas* reivindicadas retienen una alta actividad citotóxica con la eliminación de los epítopes de células B. De tal manera que, sorprendentemente, la inclusión del linker flexible corto mejora la citotoxicidad de la molécula sin alterar en forma sustancial el corte de la molécula por la furina (Ver la página 4 de la descripción).

Como se explica detalladamente en el numeral 6 del presente memorial, el concepto inventivo común descrito anteriormente no se enseña o sugiere en ninguna de las referencias citadas. En consecuencia, el capítulo reivindicatorio modificado que se presenta con este memorial se encuentra cobijado bajo el mismo concepto inventivo y por lo tanto dicho capítulo cumple a cabalidad con lo exigido por el Artículo 25 de la Decisión Andina 486 de 2000.

8. **NIVEL INVENTIVO.**

Finalmente, el Señor Examinador considera que la presente solicitud carecería de altura inventiva con respecto a las enseñanzas de los documentos **WO2009032954 (D1) y US07592426 (D2).**

A este respecto el Señor Examinador asegura que D1 sugiere la conjugación del sitio de corte de furina FCS y el dominio II de PE, así como sustituciones del dominio III de PE y mecanismos para incrementar la actividad citotóxica (citando los párrafos [0020] - [0022] y [0029] del documento D1). En este sentido, el señor Examinador concluye que la persona versada en la materia habría llevado a cabo la incorporación de un enlazador flexible en la molécula química del documento D1 con lo cual habría sido posible llegar al objeto reclamado en la reivindicación 1.

Sobre este particular me permito señalar que la combinación de las enseñanzas de los documentos D1 y D2 no sugiere la molécula química y tampoco el polipéptido de fusión, tal y como se reclaman dentro del capítulo reivindicatorio modificado que se presenta junto con este memorial.

El documento D1 divulga de manera general que un enlazador denominado R² que tiene de 1 a 10 residuos de aminoácidos puede unir FCS con el dominio funcional III de la exotoxina de *Pseudomonas* PE (D1, párrafo [0009]). El documento D1 también enseña que la citotoxicidad de la exotoxina de *Pseudomonas* se puede aumentar modificando los residuos en el dominio III de la secuencia nativa (D1, párrafo [0029]).

Sin embargo, contrario a las afirmaciones del señor Examinador, D1 no enseña y mucho menos sugiere que un tipo específico de enlazador que une FCS y el dominio funcional III de la exotoxina de *Pseudomonas* (PE) puede incrementar la citotoxicidad de la PE, como sucede con las moléculas aquí reivindicadas. Por el contrario, el documento D1 se dirige en general a una PE en la cual todo el dominio II, excepto por FCS, ha sido eliminado (D1, párrafo [0021]).

De hecho, los párrafos [0020] - [0022] del documento D1 que fueron citados por el señor Examinador, enseñan que la PE de D1, la cual carece de todo el dominio II a excepción de FCS, "retiene... la actividad citotóxica" y proporciona inmunogenicidad reducida". Esa reducción en la inmunogenicidad permite una dosis más alta que se dará con un "aumento concomitante en la actividad anti-tumoral" (D1, párrafo [0020]).

De manera que si bien el documento D1 enseña que la modificación de residuos en el dominio III puede aumentar la toxicidad (D1, párrafo [0029]), tal anterioridad no menciona que cualquier tipo particular de enlazador o "linker" que une FCS con el dominio funcional III de la exotoxina de *Pseudomonas* (PE) puede incrementar la citotoxicidad de la molécula, como bien se reclama en la presente solicitud. En consecuencia, es evidente el nivel inventivo que posee la invención reclamada.

Por otra parte, y según el señor Examinador, los enlazadores flexibles de glicina y serina son conocidos en el arte previo por su capacidad para facilitar la interacción entre los dominios y optimizar la actividad intrínseca de las moléculas fusionadas (citando el link http://parts.igem.org/Protein_domains/Linker. Sitio web Protein domains/linker).

No obstante, aunque que el sitio web Protein domains/linker menciona que los enlazadores a menudo se componen de residuos flexibles como glicina y serina, de modo que los dominios de proteína adyacentes son libres de moverse uno respecto al otro, tal sitio web, de ninguna manera enseña que los enlazadores o "linkers" flexibles de glicina y serina son conocidos por su capacidad de optimizar la actividad intrínseca de moléculas fusionadas, como quiere hacer notar el señor Examinador.

Realmente, no existe ninguna divulgación en el mencionado sitio web Protein domains/linker que enseñe o sugiera que la citotoxicidad de una exotoxina de *Pseudomonas* (PE) puede mejorarse empleando enlazadores flexibles de glicina y serina, como se reivindica en la presente solicitud. Esta es una razón más, por la cual es evidente el carácter inventivo de la presente invención.

Como se explicó anteriormente y como se evidencia en la descripción de la presente solicitud (páginas 3 y 4), de manera sorprendente, la inclusión del enlazador o "linker" flexible corto mejora la citotoxicidad de la molécula sin alterar en forma sustancial el corte de la molécula por la furina.

Se tiene entonces que el documento D1 enseña moléculas de PE, el documento D2 enseña anticuerpos anti-mesotelina, y el sitio web Protein domains/linker enseña enlazadores de glicina-serina. Sin embargo, los documentos D1 y D2 y el sitio web Protein domains/linker no enseñan un FL que consiste de 3 a 8 residuos de aminoácidos continuos seleccionados independientemente de la glicina y la serina conectando el FCS de la exotoxina de Pseudomonas (PE) al dominio funcional III de la PE, como se reclama en la presente invención.

En este sentido, es evidente que no existe ninguna enseñanza en los documentos D1 y D2 y tampoco en el sitio web Protein domains/linker, que sugiera que la inclusión del enlazador flexible (FL) que consiste en "3 a 8 residuos de aminoácidos continuos seleccionados independientemente entre glicina y serina", para conectar el FCS con el dominio funcional III de la exotoxina de Pseudomonas, como aquí se reclama, proporcionaría el efecto sorprendente de mejorar la citotoxicidad de la molécula sin alterar sustancialmente el corte de la molécula por la furina.

En ese orden de ideas, es claro que el efecto técnico sorprendente de la molécula reivindicada no se puede predecir a partir de las enseñanzas de los documentos D1, D2 y el sitio web Protein domains/linker y de las técnicas conocidas en el arte previo.

En vista de todo lo anterior, es indiscutible que la combinación de los **documentos D1 y D2 no conduciría a la persona versada en la materia a obtener la molécula quimérica y el polipéptido de fusión de la presente solicitud.**

Es evidente que en la técnica anterior no había ninguna enseñanza que sugiera que el efecto técnico sorprendente que consiste en la mejora de la citotoxicidad, fuese proporcionado por el FL reivindicado. De hecho, el señor Examinador no cita ninguna referencia que divulgue o sugiera el efecto técnico sorprendente conseguido en la presente invención.

En este sentido, muy respetuosamente me permito señalar que la única motivación que tendría la persona versada en la materia para llegar a la molécula quimérica y al polipéptido de fusión reclamados, sería la revelación hecha por la presente invención, por lo que el análisis de nivel inventivo realizado por el Señor Examinador está enmarcado dentro de lo que en el ámbito de las patentes se conoce como "hindsight", o análisis "Ex post facto", toda vez que conociendo de antemano la invención el Señor Examinador está seleccionado partes específicas de las enseñanzas contenidas dentro de D1 y D2, y modificándolas de manera que permitan construir la materia reclamada por la presente solicitud, pasos que de ninguna manera podrían ser evidentes para la persona versada en la materia en el momento en el cual se desarrolló la invención.

En conexión con lo anterior, vale la pena resaltar que al tener conocimiento de la invención es de esperarse que dichas modificaciones particulares de las enseñanzas de D1 y D2 parezcan obvias, toda vez que ya se sabe de antemano que efectivamente estas modificaciones resuelven el problema planteado; sin embargo, esta información no era conocida por el inventor en el momento de desarrollar la invención, por lo que no podría tener certeza de la utilidad de realizar estos pasos.

Dentro de este marco de ideas, me permito referirme al señalamiento que hace el Tribunal de Justicia de la Comunidad Andina con respecto a las definiciones de "obvio" y "evidente" en la evaluación del Nivel Inventivo de una solicitud de patente²:

"Resulta pertinente, a los efectos del artículo 4 de la Decisión 344, definir qué se entiende por "obvio" y por "evidente" a fin de determinar qué se debe entender por nivel inventivo.

Lo obvio es aquello que salta a la vista de manera clara y directa; lo "que se encuentra o pone delante de los ojos, muy claro o que no tiene dificultad". Mientras que lo evidente consiste en la "certeza clara y manifiesta de la que no se puede dudar; cierto, claro, patente y sin la menor duda". De forma que algo que resulte obvio no es necesariamente evidente; empero lo que es evidente, es también obvio.

En este orden de ideas, se puede concluir que una invención goza de nivel inventivo cuando a los ojos de un experto medio en el asunto de que se trate, se necesita algo más que la simple aplicación de los conocimientos técnicos en la materia para llegar a ella, es decir, que de conformidad con el estado de la técnica el invento no sea consecuencia clara y directa de dicho estado."

En vista de este señalamiento, y a la luz de lo discutido en el presente memorial, me permito resaltar que al no "saltar a la vista de manera clara y directa" a partir de los documentos D1 y D2 la molécula quimérica de la presente invención y al no tenerse "certeza clara y manifiesta" de que dicha molécula poseerían una citotoxicidad mejorada sin presentar alteraciones en el corte de la molécula por la furina, es claro que los documentos D1 y D2 no pueden de ninguna manera afectar el Nivel Inventivo de la solicitud en estudio.

9. **CONCLUSIÓN.**

Para finalizar me permito resaltar que el capítulo reivindicatorio modificado presentado junto con este memorial es totalmente claro y conciso al reclamar la molécula quimérica y el polipéptido de fusión de la invención, cumpliendo así con lo exigido por la mencionada Decisión Andina 486 de 2000.

² Tribunal de Justicia de la Comunidad Andina, PROCESO 192-IP-2005

Por otro lado, en virtud de los argumentos discutidos con respecto a los documentos citados, se estableció de manera incontrovertible que la molécula química y el polipéptido de fusión que son objeto de la solicitud en estudio, de ninguna manera resultarían evidentes para la persona versada en la materia a partir de las enseñanzas de los documentos D1 y D2, demostrando irrefutablemente la patentabilidad de la presente solicitud.

10. Anexo me permito presentar:

- (i) Copia de las páginas **94 a 98** que corresponde al **capítulo reivindicatorio modificado**, tal como se discutió en el presente memorial, constituido por **15** reivindicaciones;
- (ii) Copia del documento: Duckert et al. "Protein Engineering, Design & Selection 17(1):107-112 (2004); y
- (iii) Copia de la primera página correspondiente al documento de publicación internacional WO 2009/032954.

11. En vista de lo anterior, ruego a usted proceder a estudiar nuevamente la solicitud a la luz de las modificaciones y los argumentos presentados y conceder el privilegio de patente solicitado.

Señor Superintendente,



ALICIA LLOREDA RICAURTE

T. P. 53.215

REIVINDICACIONES

1. Una molécula quimérica que comprende un polipéptido de fusión que tiene la fórmula M-L1-FCS-FL-dominio funcional III de la exotoxina de *Pseudomonas*, CARACTERIZADA porque

M es un anticuerpo que se une específicamente a mesotelina, o a un fragmento del mismo que se une específicamente a mesotelina, en donde el anticuerpo o fragmento de anticuerpo comprende:

una cadena V_H que comprende una secuencia de aminoácidos CDR1 de GYTMN (SEQ ID NO: 51), una secuencia de aminoácidos CDR2 de LITPYNGASSYNQKFRG (SEQ ID NO: 52), y una secuencia de aminoácidos CDR3 de GGYDGRGFDY (SEQ ID NO: 53), y

una cadena V_L que comprende una secuencia de aminoácidos CDR1 de SASSSVSYMH (SEQ ID NO: 54), una secuencia de aminoácidos CDR2 de DTSKLAS (SEQ ID NO: 55), y una secuencia de aminoácidos CDR3 seleccionada del grupo compuesto por QQWSGYPLT (SEQ ID NO: 56), QQWSKHPLT (SEQ ID NO: 57), QQWSGHPLT (SEQ ID NO: 58), QQWSAHPLT (SEQ ID NO: 59), QQWSQIPLT (SEQ ID NO: 60), QQWGFNPLT (SEQ ID NO: 61), QQWGTNPLT (SEQ ID NO: 62), QQWGSHPPLT (SEQ ID NO: 63), QQWGDFPLT (SEQ ID NO: 64), QQWGDHPLT (SEQ ID NO: 65), QQWSAHPLT (SEQ ID NO: 66), y QQWSGYPTT (SEQ ID NO: 67);

L1 consiste de 1 a 10 residuos de aminoácidos continuos, en donde L1 une el carboxilo terminal del anticuerpo con el N-terminal del FCS;

FCS es un sitio de corte de furina que consiste de residuos de aminoácidos continuos capaces de ser cortados por furina;

FL consiste de 3 a 8 residuos de aminoácidos continuos seleccionados independientemente de la glicina y la serina;

el dominio funcional III de la exotoxina de *Pseudomonas* consiste de residuos de aminoácidos continuos idénticos en secuencia a SEQ ID NO:1 de la posición 395 a 613, que comprende opcionalmente (i) una sustitución de uno o más residuos correspondientes a 609-613 de la SEQ ID NO:1, (ii) una sustitución de glicina, alanina, valina, leucina o isoleucina por arginina en una posición correspondiente a la posición 490 de la SEQ ID NO:1, (iii) una sustitución de, independientemente, alanina, glicina, serina o glutamina en lugar de uno más residuos correspondientes a los residuos D403, D406, R412, R427, E431, E432, R458, D461, R467, R490, R505, R513, E522, R538, E548, R551, R576, K590, Q592 y L597 de la SEQ ID NO: 1, o (iv) una combinación de cualquiera de (i)-(iii).

2. La molécula de la reivindicación 1, CARACTERIZADA porque el anticuerpo o fragmento de anticuerpo es un minicuerpo, un diacuerpo, un triacuerpo, un fragmento Fab', un fragmento F(ab)'₂, una proteína Fv de cadena sencilla ("scFv") o un fragmento de proteína Fv estabilizado por disulfuro ("dsFv").

3. La molécula de la reivindicación 1 o 2, CARACTERIZADA porque el anticuerpo o fragmento de anticuerpo comprende regiones variables de inmunoglobulina de cadena ligera y pesada estabilizadas por disulfuro fusionadas con el resto de la molécula a través de la unión de la cadena pesada a L1, en donde la región variable de cadena ligera comprende la secuencia de aminoácidos de SEQ ID NO: 6 y la región variable de cadena pesada comprende la secuencia de aminoácidos de SEQ ID NO: 7.

4. La molécula de cualquiera de las reivindicaciones 1 a 3, CARACTERIZADA porque L1 tiene 3 a 5 aminoácidos de longitud.

5. La molécula de cualquiera de las reivindicaciones 1 a 4, CARACTERIZADA porque L1 es ASGG (SEQ ID NO: 19).

6. La molécula de cualquiera de las reivindicaciones 1 a 5, CARACTERIZADA porque el dominio funcional III de la exotoxina de *Pseudomonas* tiene una sustitución de, independientemente, alanina, glicina o serina en lugar de al menos un residuo de aminoácidos correspondiente a un residuo de aminoácidos de la SEQ ID NO: 1, seleccionado del grupo que consiste de D403, R412, R427, E431, R458, D461, R505, E522, R538, R551, R576 y L597.

7. La molécula de cualquiera de las reivindicaciones 1 a 6, CARACTERIZADA porque el dominio funcional III de la exotoxina de *Pseudomonas* tiene una sustitución de, independientemente, alanina, glicina o serina en lugar de al menos un residuo de aminoácidos correspondiente a un residuo de aminoácidos de la SEQ ID NO: 1, seleccionado del grupo que consiste de D406, R432, R467, R490, R513, E548, K590 y Q592.

8. La molécula de cualquiera de las reivindicaciones 1 a 7, CARACTERIZADA porque FL es GGS, (GGS)₂ (SEQ ID NO: 18), GSGG (SEQ ID NO: 71), GGSGG (SEQ ID NO: 72), GGSG (SEQ ID NO: 73), GSG, o GGG.

9. La molécula de cualquiera de las reivindicaciones 1 a 8, CARACTERIZADA porque FL es GGS.

10. La molécula de cualquiera de las reivindicaciones 1 a 9, CARACTERIZADA porque la secuencia de corte de furina es RHRQPRGWEQL (SEQ ID NO: 17).

11. La molécula de cualquiera de las reivindicaciones 1 a 10, CARACTERIZADA porque el dominio funcional III de la exotoxina de *Pseudomonas* es el dominio funcional III de la exotoxina de *Pseudomonas* de LR que tiene la secuencia de aminoácidos de la SEQ ID NO: 4 de la posición 20 a 237 o de la SEQ ID NO: 5 de la posición 20 a 237.

12. La molécula de cualquiera de las reivindicaciones 1 a 11, CARACTERIZADA porque la molécula comprende la cadena ligera variable de la SEQ ID NO: 6 y el polipéptido de fusión de PE de la cadena pesada variable de la SEQ ID NO: 8, donde las cadenas variables ligera y pesada forman un anticuerpo estabilizado por disulfuro.

13. La molécula de cualquiera de las reivindicaciones 1-12, CARACTERIZADA porque el FCS es idéntico en secuencia a la SEQ ID NO: 1 de la posición 274 a 284 o es la secuencia de aminoácidos RHRSKRGWEQL (SEQ ID NO: 29).

14. Una composición farmacéutica CARACTERIZADA porque comprende (a) 0,1% a 75% de la molécula de cualquiera de las reivindicaciones 1 a 13 y (b) un excipiente farmacéuticamente aceptable.

15. Un polipéptido de fusión que tiene la fórmula FCS-FL-Dominio funcional III de la exotoxina de *Pseudomonas*, CARACTERIZADO porque:

FCS consiste de residuos de aminoácidos continuos idénticos en secuencia a la SEQ ID NO: 1 de la posición 274 a 284 o de la secuencia de aminoácidos RHRSKRGWEQL (SEQ ID NO: 17);

FL consiste de 3 a 8 residuos de aminoácidos continuos seleccionados de manera independiente de la glicina y serina;

el Dominio funcional III de la exotoxina de *Pseudomonas* consiste de residuos de aminoácidos continuos idénticos en secuencia a la SEQ ID NO: 1 de la posición 395 a 613, que opcionalmente comprende (i) una sustitución de uno o más residuos correspondientes a 609-613 de la SEQ ID NO: 1, (ii) una sustitución de glicina, alanina, valina, leucina o isoleucina por arginina en una posición correspondiente a la posición 490 de la SEQ ID NO: 1, (iii) una sustitución de, independientemente, alanina, glicina, serina o glutamina en lugar de uno o más residuos correspondientes a los residuos D403, D406, R412, R427, E431, E432, R458, D461, R467, R490, R505, R513, E522, R538, E548, R551, R576, K590, Q592 y L597 de la SEQ ID NO: 1, o (iv) una combinación de cualquiera de (i)-(iii).

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
12 March 2009 (12.03.2009)

PCT

(10) International Publication Number
WO 2009/032954 A1

(51) International Patent Classification:

C07K 14/21 (2006.01) C12N 15/11 (2006.01)
C07K 19/00 (2006.01) A61K 38/16 (2006.01)
C12N 15/62 (2006.01)

FITZGERALD, David (US/US); 37 Convent Drive,
Room 5116, Bethesda, MD 20892 (US).

(21) International Application Number:

PCT/US2008/075296

(74) Agents: HYMAN, Laurence, J. et al.; Townsend and
Townsend and Crew LLP, Two Embarcadero Center, 8th
Floor, San Francisco, CA 94111-3834 (US).

(22) International Filing Date:

4 September 2008 (04.09.2008)

(25) Filing Language:

English

(26) Publication Language:

English

(30) Priority Data:

60/969,929 4 September 2007 (04.09.2007) US
61/018,853 3 January 2008 (03.01.2008) US

(81) Designated States (unless otherwise indicated, for every
kind of national protection available): AE, AG, AI, AM,
AO, AT, AU, AZ, BA, BB, BG, BH, BR, BW, BY, BZ, CA,
CH, CN, CO, CR, CU, CZ, DE, DK, DM, DO, DZ, EC, EE,
EG, ES, FI, GB, GD, GE, GH, GM, GT, HN, HR, HU, ID,
IL, IN, IS, JP, KE, KG, KM, KN, KP, KR, KZ, LA, LC, LK,
LR, LS, LT, LU, LY, MA, MD, ME, MG, MK, MN, MW,
MX, MY, MZ, NA, NG, NI, NO, NZ, OM, PG, PH, PL, PT,
RO, RS, RU, SC, SD, SE, SG, SK, SL, SM, ST, SV, SY, TJ,
TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, ZA, ZM,
ZW.

(71) Applicant (for all designated States except US): THE
GOVERNMENT OF THE UNITED STATES OF
AMERICA, as represented by THE SECRETARY OF
THE DEPARTMENT OF HEALTH AND HUMAN
SERVICES [US/US]; NIH Office of Technology Trans-
fer, 6011 Executive Boulevard, Suite 325, Rockville, MD
20852-3804 (US).

(84) Designated States (unless otherwise indicated, for every
kind of regional protection available): ARIPO (BW, GH,
GM, KE, LS, MW, MZ, NA, SD, SI, SZ, TZ, UG, ZM,
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,
FR, GB, GR, HR, HU, IE, IS, IT, LT, LU, LV, MC, MT, NL,
NO, PL, PT, RO, SE, SI, SK, TR), OAPI (BF, BJ, CF, CG,
CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (for US only): PASTAN, Ira,
H. [US/US]; 37 Convent Drive, Room 5116, Bethesda,
MD 20892 (US). WELDON, John [US/US]; 37 Con-
vent Drive, Room 5116, Bethesda, MD 20892 (US).

Published:

- with international search report.
- with sequence listing part of description published sepa-
rately in electronic form and available upon request from
the International Bureau

WO 2009/032954 A1

(54) Title: DELETIONS IN DOMAIN II OF PSEUDOMONAS EXOTOXIN A THAT REDUCE NON-SPECIFIC TOXICITY

(57) Abstract: The invention provides mutated, cytotoxic forms of Pseudomonas exotoxin A (PE) comprising a furin cleavage sequence conjugated or fused directly to residues 395-613 of PE or variants of that sequence. These minimal forms of PE are smaller than previous cytotoxic forms of PE, reduce non-specific toxicity, and reduce immunogenicity due to domain I or domain Ib of PE. The invention further provides nucleic acids encoding said PEs, chimeric molecules containing them, and methods of use thereof.

Prediction of proprotein convertase cleavage sites

Peter Duckert, Søren Brunak and Nikolaj Blom¹

Center for Biological Sequence Analysis, BioCentrum-DTU, Technical University of Denmark, Building 208, DK-2800 Lyngby, Denmark

¹To whom correspondence should be addressed.
E-mail: nikob@cbs.dtu.dk

Many secretory proteins and peptides are synthesized as inactive precursors that in addition to signal peptide cleavage undergo post-translational processing to become biologically active polypeptides. Precursors are usually cleaved at sites composed of single or paired basic amino acid residues by members of the subtilisin/kexin-like proprotein convertase (PC) family. In mammals, seven members have been identified, with furin being the one first discovered and best characterized. Recently, the involvement of furin in diseases ranging from Alzheimer's disease and cancer to anthrax and Ebola fever has created additional focus on proprotein processing. We have developed a method for prediction of cleavage sites for PCs based on artificial neural networks. Two different types of neural networks have been constructed: a furin-specific network based on experimental results derived from the literature, and a general PC-specific network trained on data from the Swiss-Prot protein database. The method predicts cleavage sites in independent sequences with a sensitivity of 95% for the furin neural network and 62% for the general PC network. The ProP method is made publicly available at <http://www.cbs.dtu.dk/services/ProP>.

Keywords: furin/neural network/propeptide/protease specificity/secretory precursor

Introduction

Post-translational processing by limited proteolysis of inactive secretory precursors to produce active proteins and peptides is an ancient mechanism that enables cells to regulate the level of specific bioactive polypeptides and to generate diverse products from precursor molecules. Many biologically active proteins and peptides are initially synthesized as larger, inactive precursors, usually in the form of pre-pro-proteins, which are post-translationally modified to generate the mature molecule. The N-terminal pre-regions are signal peptides, which direct the precursors to the appropriate cellular compartment, while the prodomains can have several functions. Some prodomains act as intramolecular chaperones that mediate correct folding of the newly synthesized proteins, while other prodomains are only indirectly involved in folding and have other functions such as transport and localization, oligomerization, regulation of activity (Shinde and Inouye, 2000) and quality control of folding (Bauskin *et al.*, 2000). Precursor cleavage frequently occurs at motifs containing multiple basic residues, arginine or lysine, by limited endoproteolysis of the corresponding precursor proteins immediately

C-terminally of the basic amino acid motifs (Seidah and Chretien, 1999).

Most of the enzymes responsible for this type of processing belong to a family of evolutionary conserved dibasic- and monobasic-specific Ca^{2+} -dependent serine proteases called subtilisin/kexin-like proprotein convertases (PCs) (Seidah *et al.*, 1998). Since the identification of the yeast enzyme Kex2 (kexin), seven mammalian kexin-related proprotein convertases have been identified: PC1, PC2, furin, PC4, PC5, PACE4 and PC7. Furin was the first PC to be discovered; a database search identified furin as the first mammalian kexin homolog (Fuller *et al.*, 1989), and furin was soon shown to correctly process the precursors of von Willebrand factor and beta-nerve growth factor (Bresnahan *et al.*, 1990; van de Ven *et al.*, 1990).

The PCs are the major endoproteolytic processing enzymes of the secretory pathway in mammals (Steiner, 1998). These enzymes process precursors at sites, which usually contain the consensus sequence $[\text{R/K}]\text{-X}_n\text{-}[\text{R/K}]\downarrow$, where X indicates any amino acid residue, n , the number of spacer amino acid residues, is 0, 2, 4 or 6, $[\text{R/K}]$ indicates either an arginine or a lysine residue, and the arrow ($\downarrow$) indicates the site of cleavage (Seidah and Chretien, 1999). Furin has a more stringent substrate specificity and preferentially recognizes sites that contain the sequence motif $\text{R-X-}[\text{R/K}]\text{-R}\downarrow$ (Nakayama, 1997).

Recently, the first reports on the crystal structure of members of the PC family, namely furin and Kex2, were published (Henrich *et al.*, 2003; Holyoak *et al.*, 2003). The detailed analysis of the proteolytic domains explains the preference for basic residues, in particular arginine, at the substrate cleavage site P1. In addition, the preference observed for other positions, in particular for P4 to P2 in furin, fits well with the observed enzyme–substrate crystal structure.

The PCs are involved in the activation of a large variety of proteins like peptide hormones, neuropeptides, growth and differentiation factors, adhesion factors, receptors, blood coagulation factors, plasma proteins, extracellular matrix proteins, proteases, exogenous proteins such as coat glycoproteins from infectious viruses (e.g. HIV-1 and influenza virus) and bacterial toxins (e.g. diphtheria and anthrax toxin). PCs not only catalyze removal of prodomains, but are also involved in processing of multifunctional precursors like POMC and proglucagon. Both undergo differential processing, dependent on their sites of production (Steiner, 2002). Therefore, PCs play an essential role in many vital biological processes like embryonic development and neural function, and in viral and bacterial pathogenesis. In addition, PCs are implicated in a number of pathologies such as cancer and neurodegenerative diseases (Thomas, 2002).

From a medical and biotechnological viewpoint it can be of interest to control the production of peptides involved in various diseases by specifically inhibiting the enzymatic

activity of the PCs (Bergeron *et al.*, 2000; Thomas, 2002). Several naturally occurring sequences are known to inhibit PCs, including the prodomains of the proteases themselves and the neuroendocrine proteins 7B2 and proSAAS (Cameron *et al.*, 2002; Rockwell *et al.*, 2002).

We present here work on the characterization and prediction of PC cleavage sites. Since not all single or paired basic amino acid residues are potential cleavage sites of PCs, we examined the sequence patterns characteristic of experimentally verified sites and describe a neural network based method for predicting whether a given site is a potential cleavage site for the PC enzymes. Not all furin cleavage sites contain the furin consensus sequence and not all sites containing the consensus sequence are cleavage sites for furin (Nakayama, 1997). Furin also has a preference for basic amino acid residues at P3, P5 and P6 (even at P7 and P8), while other specific requirements exist at P1' and P2'. For instance, lysine residues are not accepted at the latter positions (Henrich *et al.*, 2003). Favorable residues at P2 and P6 can compensate for less favorable ones at P1 and P4 (Molloy and Thomas, 2002). Together, these facts indicate that the furin recognition motif is non-linear and more complicated than previously believed. Therefore, a method capable of classifying complex patterns, such as a neural network, may be more suitable than a simple consensus sequence, which is not able to include correlated effects often termed 'exceptions'. In the case of non-furin cleavage sites, that are primarily characterized by a dibasic motif, the problem of false predictions by using a simple dibasic consensus pattern becomes evident and necessitates the use of a more sophisticated classifier. The neural network based method described in this paper can be used to predict PC cleavage sites in single proteins, but may also be used when large databases are scanned for novel growth factors, hormones and secreted peptides containing propeptides or for automated annotation of large protein sets.

Materials and methods

Two data sets were collected: (i) furin cleavage sites and (ii) general PC cleavage sites. The latter should ideally represent cleavage sites for all PCs. The furin cleavage data were taken from the literature (Nakayama, 1997; Lehmann *et al.*, 1998) and consisted of 38 proteins, which contained reported furin sites, including proteins from mammals and from pathogenic bacteria and viruses.

The general proprotein convertase data set was based on Swiss-Prot annotations (release 39.0, 05/00; Bairoch and Apweiler, 2000) and also included the furin cleavage data described above, totaling 235 sites in 227 proteins. The following search criteria were used in the extraction of data from the Swiss-Prot database: signal peptide present, experimentally determined propeptide region not having comments like 'PROBABLE', 'BY SIMILARITY', or 'POTENTIAL' (Junker *et al.*, 1999), and a basic residue (K or R) at P1, i.e. the first position N-terminal to the cleavage site. Proteins were from eukaryotes and viruses only, and redundancy in the data set was reduced by removal of doublet sequences, which had identical 13-mer amino acid sequences centered around the cleavage site.

Sequence logos of precursor proteins, aligned by their cleavage sites, were used for displaying the position-specific features of multiple sequence alignments (Schneider and Stephens, 1990).

In order to avoid training and testing on homologous sequences, each data set was divided into four partitions of approximately equal size based on phylogenetic trees, which were constructed using multiple alignments and the neighbour-joining algorithm of the ClustalX program (Thompson *et al.*, 1997). This further eliminates the redundancy issue, and the risk that the predictive performance is overestimated owing to training and test set similarities.

The neural network architecture was feed-forward fully connected, with zero or one layer of two, four, eight or 16 hidden units (Baldi and Brunak, 2001). The neural networks were trained by back-propagation (Rumelhart *et al.*, 1986), and the sequence data were presented to the network using sparsely encoded moving windows (Qian and Sejnowski, 1988; Brunak *et al.*, 1991). Only positions with lysine or arginine were used during training and testing. Symmetric input windows with size varying from five to 23 positions were tested. The learning rate was 0.005, and the decision threshold value for the output unit was 0.5 for all networks.

A correlation coefficient (Matthews, 1975) was calculated from the numbers of correctly and incorrectly predicted positive and negative sequence windows generated with the selected threshold value. The correlation coefficients of both the training and test sets were monitored during training, and the performance at the training cycle with the maximal test set correlation was recorded for each training run. The test performances were calculated by 4-fold cross-validation: every network run was carried out with one part as the test data and the other three parts as the training data. For each of the four combinations, one neural network architecture was chosen based on the test set correlation coefficients. If more than one network architecture gave rise to approximately the same correlation coefficient, we chose the smallest network with respect to input window size and the number of hidden units. The combined performance for each of the four subsets when used as test sets was then calculated. The cross-validated Matthews correlation coefficient for the combined ensemble was calculated from the total numbers of true positives, true negatives, annotated cleavage sites and annotated non-cleavage sites.

The trained networks provide scores between zero and one for each arginine (R) or lysine (L) residue in an amino acid sequence. The combined prediction scores were then calculated as an average over the scores from the four networks. By default, for scores above 0.5 we predict peptide bond cleavage at the C-terminal side of the amino acid residue in question.

To illustrate the prediction performance for varying thresholds and prediction sensitivities, we present receiver operating characteristic (ROC) curves, plotting sensitivity on the *x*-axis and false-positive rate on the *y*-axis.

Results

Sequence logos

The sequence pattern at the cleavage sites is shown as sequence information logos in Figure 1. Sequence logos were generated for each of the two data sets, with P1 as the central position and seven flanking amino acid residues on each side. The sequence logos emphasize amino acid residues that are frequently found at the propeptide cleavage sites, but cannot reveal correlated effects. Note that the data sets were redundancy reduced prior to the logo analysis.

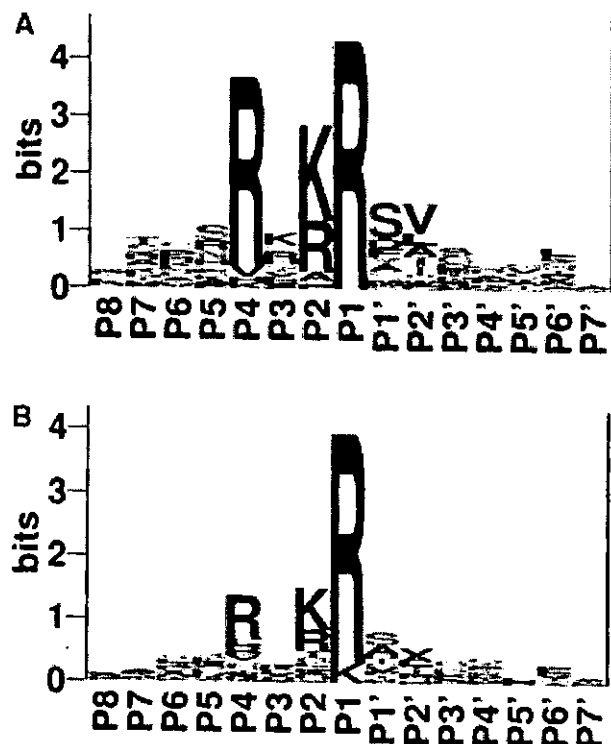


Fig. 1. Sequence logos of aligned propeptide cleavage sites centered at P1, where cleavage takes place between P1 and P1'. Shannon information is shown in units of bits: (A) 38 experimentally determined furin sites found in the literature; (B) 234 general PC sites extracted from the Swiss-Prot database.

Furin sequence logo

Furin preferentially recognizes the cleavage site sequence R-X-[R/K]-R↓ (Nakayama, 1997). This consensus cleavage sequence is clearly visible from the logo, and all furin cleavage sites in the data set had an arginine residue at position P1. The frequency of arginine at position P4 was 89%, while only one of the cleavage motifs (3%) contained a lysine residue at P4. Of the remaining amino acids only histidine (3%) and valine (5%) were found at this position. At P2, R and K were the most frequent, 32 and 58%, respectively, and the remaining amino acids found at this position were alanine (5%), glycine (3%) and proline (3%). Serine was the most frequent amino acid residue at P1' (42%), while leucine, isoleucine and valine did not occur at this position, but these amino acids were frequent at P2', 55% in total. Both phenylalanine and tyrosine were found at P1', four (11%) and three (8%), respectively, but not tryptophan, arginine, lysine, asparagine, proline or threonine. The acidic amino acids glutamic acid and aspartic acid occurred at P1' in one (3%) and four (11%) of the cleavage motifs, and histidine was found in one (3%) of the cleavage motifs. At P2', arginine was absent, and only one lysine residue was found at this position.

Of the 38 cleavage motifs in the furin data set, 31 (81%) had the R-X-[R/K]-R consensus sequence. Eleven (29%) of these were R-X-R-R and 20 (52%) were R-X-K-R. The P2 basic amino acid residue is not essential, but can greatly enhance the processing efficiency (Thomas, 2002). Therefore, the minimal furin cleavage sequence is R-X-X-R, but a less favorable amino acid residue at P4 can be compensated for by arginine or

lysine residues at P2 and P6. Three (8%) of the cleavage motifs contained only the minimal furin cleavage sequence, while the remaining four (11%) did not have an arginine residue at P4. In earlier studies, the following general rules were proposed for furin specificity: (i) an arginine residue is essential at the P1 position; (ii) in addition to the P1 arginine residue, at least two out of the three residues at P2, P4 and P6 are required to be basic for efficient cleavage; (iii) at the P1' position, an amino acid residue with a hydrophobic aliphatic side chain (i.e. leucine, isoleucine or valine) is not suitable (Nakayama, 1997).

These rules are consistent with most of the furin cleavage sites in the data set: arginine was always present at P1 and leucine, isoleucine or valine were never observed at P1'. In contrast, two of the cleavage sites did not follow the rule of at least two basic amino acid residues at P2, P4 and P6, namely human insulin-like growth factor IA precursor (LKPAKSAR↓) and shiga toxin A-chain precursor (HHASRVAR↓). In these cases, a lysine at P7 or two histidines at P7 and P8, respectively, may have compensatory effects.

General PC sequence logo

The general PC sequence logo is based on 234 of the 235 cleavage sites, which were selected as having R or K at P1. One of the 235 cleavage sites is located only three residues from the C-terminus of the protein and therefore, for that single site, it was not possible to extract seven residues C-terminally of the cleavage site as required for generating the sequence logo in Figure 1B.

Arginine was by far the most frequent amino acid residue at P1, corresponding to 92%. At P2 the frequencies of R and K were 22 and 43%, respectively, while the frequency of R was 50% at P4. Only six lysine residues were found at this position, corresponding to 3%. At the P1' position the frequency of serine was 24%, while the frequency of the hydrophobic, aliphatic amino acids leucine, isoleucine and valine was 17% in total. The furin consensus sequence R-X-[R/K]-R can also be recognized in this logo, but to a much lesser extent than in the furin logo (Figure 1A). The furin consensus sequence was found in 104 (44%) of the cleavage sites, and the minimal furin consensus sequence R-X-X-R appeared in 117 (50%) of the cleavage sites. These observations indicate that up to 50% of the cleavage sites in the general data set may be substrates of non-furin PCs.

Performance of the neural networks

The furin network was trained on a data set containing 38 experimentally verified cleavage sites out of a total number of 3004 sites containing R or K at the P1 position, of which 1589 had an R at P1. Using a 4-fold cross-validated training approach, the values for the optimal symmetric window size/hidden units were 13/2, 19/4, 17/8 and 11/2, respectively, for the four networks. Thus, optimal window size range from 11 to 19 residues, which correspond well to the contact area of other known protein-protein interactions, e.g. kinase-substrate contact (10–11 residues; Blom *et al.*, 1999). The fact that there were at least two hidden units in the networks indicates that the PC cleavage site prediction problem is indeed non-linearly separable only owing to the correlations and compensatory effects in the cleavage motifs (Baldi and Brunak, 2001). The number of true positives predicted reached 94.7% with a specificity of 83.7%, equivalent to a Matthews correlation coefficient of 0.89.

From a practical point of view, the most important aspect of a prediction method is its ability to make correct predictions.

As prediction methods are never perfect, one is always faced with the dilemma of choosing between making few false-positive predictions and having a high sensitivity, i.e. correctly identifying as many positive examples as possible. This trade-off can be visualized as what is known as the ROC curve in which the rate of false positives is plotted as a function of the sensitivity by varying the score threshold used for making positive predictions. Figure 2A shows the ROC curve for the furin site predictor, where the standard threshold of 0.5 corresponds to a false-positive rate of 0.2% at a sensitivity of 94.7%. When the sensitivity is increased further, the false-positive rate rises steeply and reaches ~30% at a sensitivity of 100%, i.e. when all cleavage sites are predicted correctly. For neural networks without hidden units, the highest correlation coefficient was 0.83, and the specificity of the corresponding network was 76.1% at the default threshold of 0.5.

Four of the known cleavage sites in the furin data set lacked the arginine residue at the P4 position characteristic of the furin consensus sequence. Two of these sites were correctly learned by the neural network and classified correctly, even when being part of an independent test set: human vitamin-K-dependent protein C precursor (KKRSHLKR¹⁹⁹↓DTED; score 0.797) and human parathyroid hormone precursor (DGKSVKKR³¹↓SVSE; score 0.851). These findings indicate that the neural network has picked up sequence correlations, which indicate furin cleavage in non-consensus cleavage sites. This also may explain the fact that a hidden layer of neurons is required for optimal performance of the neural network.

Two of the known cleavage sites in the furin data set were not predicted correctly, i.e. they were false negatives when being part of an independent test set: human insulin-like growth factor IA precursor (LKPAKSAR¹¹⁹↓SVRA) and human serum albumin precursor (YSRGVFR²⁴↓DAHK). The cleavage scores produced by the furin-specific neural network were low, 0.100 and 0.128, respectively. In the case of human insulin-like growth factor IA precursor cleavage has been observed in furin-deficient cell types indicating that another processing enzyme may be acting *in vivo* (Duguay *et al.*, 1995). For human serum albumin precursor, PCs other than furin, namely PACE4 and PC7, have been suggested to be involved in processing (Mori *et al.*, 1999), and this may explain the low cleavage site score of the furin-specific network. In this sense, the network may be more correct than the statistics on the experimental data indicate.

Another caveat when using the furin-type cleavage site method is that prediction of processing by furin does not mean that the substrate is actually cleaved by furin *in vivo*. Three other proprotein convertases, PACE4, PC5 and PC7, have sequence specificities similar to that of furin (Nakayama, 1997). Since there seems to be a considerable overlap in substrate specificity between the different PC members, predictions for a given PC, e.g. furin, should be interpreted with this in mind.

After implementation of the neural network prediction method for furin-type cleavage sites, we scanned the published literature for recent reports on furin-mediated cleavage in proteins, which were not used in our training data. We found three examples reported as furin cleavage sites and correctly predicted by our furin-type network. All reported sites matched the R-X-[R/K]-R consensus: human ectodysplasin-A RVRNRK¹⁵⁹↓SK (score 0.819) (Chen *et al.*, 2001), Marburg virus glycoprotein VFRRKR⁴³⁵↓SI (score 0.712) (Volchikov *et al.*, 2000) and mouse ZP3 glycoprotein

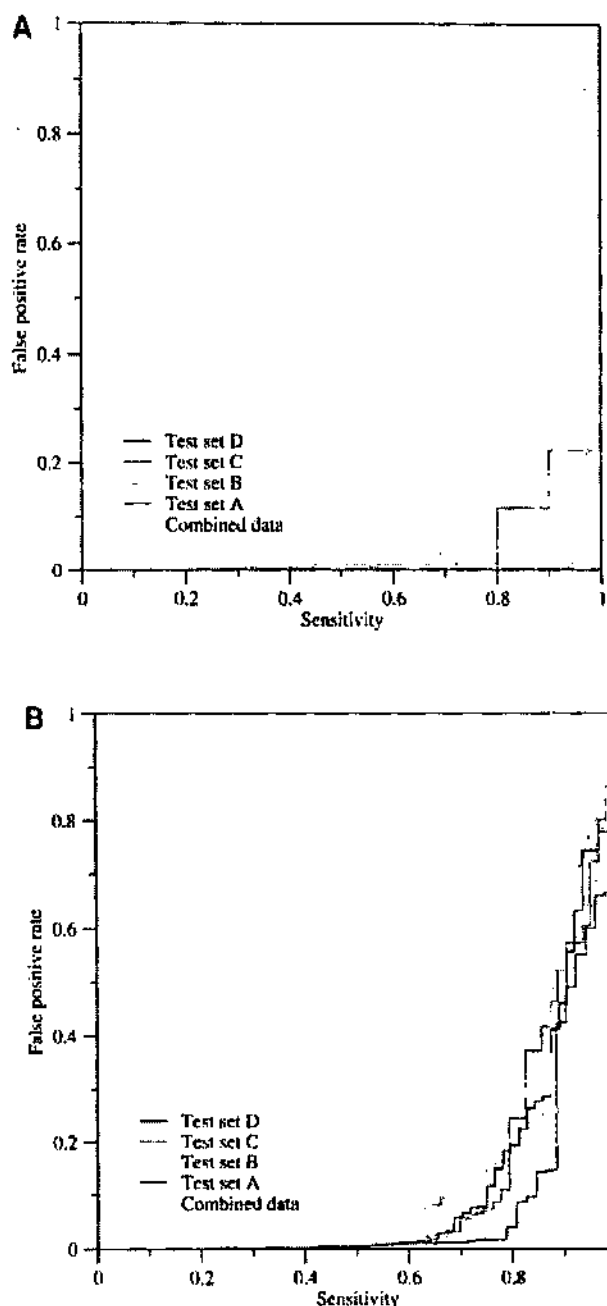


Fig. 2. The predictive performance shown as sensitivity versus false-positive rate in a ROC diagram. The plot was constructed from results obtained for the independent test sets, and corresponds to the expected performance for novel, unrelated proteins. (A) The furin network performance; (B) the general PC neural network. For a given category, e.g. the combined data for the general PC neural network, a sensitivity of 80% can be achieved with a false-positive rate of 20%, corresponding to 80% correct prediction on both categories, i.e. true positives and true negatives. Random performance would correspond to a line along the diagonal.

LVSRNR³⁵³↓HV (score 0.762) (Williams and Wassarman, 2001). More interestingly, our neural network method was able to correctly classify consensus-matching non-cleavage sites in two of these proteins, namely human ectodysplasin-A ESRRVR¹⁵⁶↓NK (score 0.482) and Marburg virus glycoprotein LVCRLR⁵⁶¹↓LA (score 0.446). Again, these findings indicate that the neural network algorithm is able to incorporate

sequence features other than consensus sequence alone and may be well suited for rejecting putative furin cleavage sites which match the consensus sequence perfectly.

It was recently claimed, based on the crystal structure of Kex2, that substrates with a glutamate residue at P3 might be selected against (Holyoak *et al.*, 2003). In our data set of negative furin sites we found an example of a site that strongly matches the furin consensus site but nevertheless is predicted as being non-cleaved. In human integrin α -3 precursor (ITA3_HUMAN) the site RDRR⁵²⁵↓PP (score 0.207) contains a glutamate (D) at P3, which may, in addition to a non-favored proline at P1', result in the correct, non-cleaved classification.

The general PC network was trained on 235 cleavage sites out of a total of 11 399 R or K residues. Of these, 372 were false positives eliminated during the learning process in order to enhance performance (339 R and 33 K), leaving 10 792 negative sites with R or K at position P1. The elimination of potentially false positives, which had a detrimental effect on the predictive performance, was done by monitoring in which order the network did learn the positives and negatives (Blom *et al.*, 1999). A 4-fold cross-validated training gave the following optimal network architecture (window size versus number of hidden units): 13/16, 13/4, 13/4 and 11/8, respectively, for the four networks. At the standard threshold of 0.5, the sensitivity was 61.7%, the specificity was 59.7%, and the Matthews correlation coefficient was 0.60 for the best ensemble of networks. The ROC curve (Figure 2B) showed that the 0.5 threshold was a good choice, since the false-positive rate was 0.9% at a sensitivity of 61.7%. The rise in false-positive rate is moderate until a sensitivity of ~75%, where the false-positive rate is ~7%, but this is still a high number because the number of true negative sites is much larger than the number of true positive sites. The importance of hidden units is again evidenced by the fact that the highest correlation coefficient for a network without hidden units was 0.41, and the specificity of this network was only 28.9%.

We estimate that furin-like sites are frequently represented in the general PC data set, since the furin consensus sequence, R-X-[R/K]-R, is present in 44% of the cleavage sites. This may explain the difficulties when predicting PC1 and PC2 sites, as can be seen from the following examples. Comparing our prediction performance on PC1 and PC2 proprotein processing sites, which were reported late in our study, we observed that not all reported sites could be correctly predicted (Cameron *et al.*, 2002). In particular, of the 12 PC1/PC2 sites in rat proenkephalin and six sites in rat proglucagon, our method correctly predicts five sites and one site, respectively. Almost all sites contain a dibasic motif at P2 and P1, but lack a basic residue on P4. Thus, assuming that our general data set is dominated by furin-type cleavage sites, we do not expect very high prediction performance on bona fide PC1/PC2 cleavage sites.

Discussion

Proteases are highly relevant both as biotechnological tools and as pharmaceutical targets. In both aspects, a deeper understanding of protease specificity and the ability to rapidly scan thousands of protein sequences for potential candidates are desirable.

One of the obstacles when mammalian cells are designed for overproduction of secreted bioactive proteins and peptides is

the relatively low level of endogenous proprotein convertases (Seidah and Chretien, 1997). Recombinant furin has been used successfully to increase the yield of over-expressed secreted bioactive proteins in engineered eukaryotic cells (Seidah and Chretien, 1997; Himmelsbach *et al.*, 2000). Design of recombinant proteins containing highly efficient furin cleavage sites, while still retaining native sequence after cleavage, may be facilitated by the fast methods described in this paper.

Understanding substrate specificity may in particular be useful for designing specific PC inhibitors. PC inhibitors may be designed for treatment of diseases dependent on normal or aberrant PC function, such as anthrax, Ebola virus infections or cancer (Bergeron *et al.*, 2000; Rockwell *et al.*, 2002; Thomas, 2002). Being able to predict and rapidly screen large sets of synthetic cleavage substrates of certain proteases such as the PCs may substantially speed up the selection of highly efficient cleavable peptides. The sequence and deduced structure of these peptides may be used as lead structures for the development of peptidomimetic protease inhibitors.

In addition, the recently published structures on related PC enzymes will allow for a more focused search for lead structures for potential PC inhibitors (Henrich *et al.*, 2003; Holyoak *et al.*, 2003). A deeper understanding of the structural requirements of PC substrates may also be used to improve on the predictions of potential substrates, e.g. by penalizing side chains which are incompatible with the protease structure.

Ideally, one would like to train neural networks for each of the PC members. However, the amount of experimentally verified sites, where the bona fide PC is known, is still limited. This stems from the fact that it is much easier to identify the site of cleavage in a protein, e.g. by N-terminal sequencing, than to identify the nature of the physiologically active PC. Since we were able to extract only a limited number of verified cleavage sites for PCs other than furin, we decided to include all basic proprotein cleavage sites in our general approach. This situation is in many ways similar to our earlier experiences in characterizing phosphorylation sites at a time when few kinase-specific data were available (Blom *et al.*, 1999).

The performance values obtained for the general type PC network compared with the furin-specific network confirmed our expectations. Although a basic cleavage site preference is observed for all PCs, slight differences in substrate specificity imply that a single method is not able to predict all PC sites with the same performance. In particular, we expect that the performance of the general network on PC1 and PC2 sites may be sub-optimal owing to an under-representation of these sites in the data set. A natural development from this will be to design PC1 and PC2 data sets, individual or combined, and train specific neural networks on these data.

Three other PCs, PC5, PACE4 and PC7, have substrate specificities which are similar to furin (Nakayama, 1997) and may be quite difficult to identify *in vitro* owing to cross-specificity and redundancy. Sites predicted by the furin network may therefore, in some cases, be physiological substrates of a furin-like PC.

The prediction of protease cleavage sites may also be used as functional fingerprints, for example, as an additional functional feature input to systems biology approaches attempting to classify the function of orphan proteins (Jensen *et al.*, 2002). Also, in systematic screenings for novel secreted factors, the presence of a strong predicted processing site may be used as a search criterion.

Acknowledgements

We thank Kristoffer Rapacki for assistance with database extraction, Lars Juhl Jensen for profitable discussions and Neil Taylor for helpful advice. The work was supported by the National Danish Research Foundation and a grant from NsGene A/S.

References

- Bairoch, A. and Apweiler, R. (2000) *Nucleic Acids Res.*, **28**, 45–48.
- Baldi, P. and Brunak, S. (2001) *Bioinformatics: The Machine Learning Approach*, 2nd edn. MIT Press, Cambridge, MA.
- Bauskin, A.R., Zhang, H.P., Fairlie, W.D., He, X.Y., Russell, P.K., Moore, A.G., Brown, D.A., Stanley, K.K. and Breit, S.N. (2000) *EMBO J.*, **19**, 2212–2220.
- Bergeron, F., Leduc, R. and Day, R. (2000) *J. Mol. Endocrinol.*, **24**, 1–22.
- Blom, N., Gammeltoft, S. and Brunak, S. (1999) *J. Mol. Biol.*, **294**, 1351–1362.
- Bresnahan, P.A., Leduc, R., Thomas, L., Thorne, J., Gibson, H.L., Brake, A.J., Barr, P.J. and Thomas, G. (1990) *J. Cell Biol.*, **111**, 2851–2859.
- Brunak, S., Engelbrecht, J. and Knudsen, S. (1991) *J. Mol. Biol.*, **220**, 49–65.
- Cameron, A., Apletalina, E. and Lindberg, I. (2002) In Dalbey, R.E. and Sigman, D.S. (eds), *The Enzymes*, 3rd edn. Academic Press, San Diego, CA, Vol. 22, pp. 291–332.
- Chen, Y., Molloy, S.S., Thomas, L., Gambee, J., Bachinger, H.P., Ferguson, B., Zomana, J., Thomas, G. and Morris, N.P. (2001) *Proc. Natl Acad. Sci. USA*, **98**, 7218–7223.
- Duguay, S.J., Lai-Zhang, J. and Steiner, D.F. (1995) *J. Biol. Chem.*, **270**, 17566–17574.
- Fuller, R.S., Brake, A.J. and Thorne, J. (1989) *Science*, **246**, 482–486.
- Henrich, S., Cameron, A., Bourenkov, G.P., Kiefersauer, R., Huber, R., Lindberg, I., Bode, W. and Than, M.E. (2003) *Nat. Struct. Biol.*, **10**, 520–526.
- Himmelspach, M., Pfeiderer, M., Fischer, B.E., Plaimauer, B., Antoine, G., Falkner, F.G., Dörner, F. and Schlokot, U. (2000) *Thromb. Res.*, **97**, 51–67.
- Holyoak, T., Wilson, M.A., Fenn, T.D., Kettner, C.A., Petsko, G.A., Fuller, R.S. and Ringe, D. (2003) *Biochemistry*, **42**, 6709–6718.
- Jensen, L.J. et al. (2002) *J. Mol. Biol.*, **319**, 1257–1265.
- Junker, V.L., Apweiler, R. and Bairoch, A. (1999) *Bioinformatics*, **15**, 1066–1067.
- Lehmann, M., Andre, F., Bellan, C., Remacle-Bonnet, M., Garrouste, F., Parat, F., Lissitsky, J.C., Marvaldi, J. and Pommier, G. (1998) *Endocrinology*, **139**, 3763–3771.
- Matthews, B.W. (1975) *Biochim. Biophys. Acta*, **405**, 442–451.
- Molloy, S.S. and Thomas, G. (2002) In Dalbey, R.E. and Sigman, D.S. (eds), *The Enzymes*, 3rd edn. Academic Press, San Diego, CA, Vol. 22, pp. 199–235.
- Mori, K., Imamaki, A., Nagata, K., Yonetomi, Y., Kiyokage-Yoshimoto, R., Martin, T.J., Gillespie, M.T., Nagahama, M., Tsuji, A. and Matsuda, Y. (1999) *J. Biochem. (Tokyo)*, **125**, 627–633.
- Nakayama, K. (1997) *Biochem. J.*, **327**, 625–635.
- Qian, Y. and Sejnowski, T.J. (1988) *J. Mol. Biol.*, **202**, 865–884.
- Rockwell, N.C., Krysan, D.J., Komiyama, T. and Fuller, R.S. (2002) *Chem. Rev.*, **102**, 4525–4548.
- Rumelhart, D.E., Hinton, G.E. and Williams, R.J. (1986) In Rumelhart, D.E., McClelland, J.L. and the PDP Research Group (eds), *Parallel Distributed Processing: Explorations in the Microstructure of Cognition*. MIT Press, Cambridge, MA, Vol. 1, pp. 318–362.
- Schneider, T.D. and Stephens, R.M. (1990) *Nucleic Acids Res.*, **18**, 6097–6100.
- Seidah, N.G. and Chretien, M. (1997) *Curr. Opin. Biotechnol.*, **8**, 602–607.
- Seidah, N.G. and Chretien, M. (1999) *Brain Res.*, **848**, 45–62.
- Seidah, N.G., Day, R., Marcinkiewicz, M. and Chretien, M. (1998) *Ann. N. Y. Acad. Sci.*, **839**, 9–24.
- Shinde, U. and Inouye, M. (2000) *Semin. Cell Dev. Biol.*, **11**, 35–44.
- Steiner, D.F. (1998) *Curr. Opin. Chem. Biol.*, **2**, 31–39.
- Steiner, D.F. (2002) In Dalbey, R.E. and Sigman, D.S. (eds), *The Enzymes*, 3rd edn. Academic Press, San Diego, CA, Vol. 22, pp. 163–198.
- Thomas, G. (2002) *Nat. Rev. Mol. Cell Biol.*, **3**, 753–766.
- Thompson, J.D., Gibson, T.J., Plewniak, F., Jeanmougin, F. and Higgins, D.G. (1997) *Nucleic Acids Res.*, **25**, 4876–4882.
- van de Ven, W.J., Voorberg, J., Fontijn, R., Pannekoek, H., van den Ouweland, A.M., van Duijnhoven, H.L., Roebroek, A.J. and Siezen, R.J. (1990) *Mol. Biol. Rep.*, **14**, 265–275.
- Volchkov, V.E., Volchkova, V.A., Stroher, U., Becker, S., Dolnik, O., Cieplik, M., Garten, W., Klenk, H.D. and Feldmann, H. (2000) *Virology*, **268**, 1–6.
- Williams, Z. and Wassarman, P.M. (2001) *Biochemistry*, **40**, 929–937.

Received July 31, 2003; revised October 28, 2003; accepted October 28, 2003

Edited by Valerie Daggett