

19



OFICINA ESPAÑOLA DE
PATENTES Y MARCAS

ESPAÑA



11 Número de publicación: **2 376 463**

51 Int. Cl.:
G01N 33/68 (2006.01)
G06F 19/00 (2011.01)

12

TRADUCCIÓN DE PATENTE EUROPEA

T3

- 96 Número de solicitud europea: **04791642 .4**
96 Fecha de presentación: **01.10.2004**
97 Número de publicación de la solicitud: **1680681**
97 Fecha de publicación de la solicitud: **19.07.2006**

54 Título: **MÉTODO Y APARATO PARA EVALUAR LA AGREGACIÓN DE POLIPÉPTIDOS.**

30 Prioridad:
05.11.2003 GB 0325817

45 Fecha de publicación de la mención BOPI:
14.03.2012

45 Fecha de la publicación del folleto de la patente:
14.03.2012

73 Titular/es:
**CAMBRIDGE ENTERPRISE LIMITED
THE OLD SCHOOLS TRINITY LANE
CAMBRIDGE CAMBRIDGESHIRE CB2 1TS, GB**

72 Inventor/es:
**DOBSON, Christopher;
CHITI, Fabrizio;
ZURDO, Jesus;
VENDRUSCOLO, Michele y
DUBAY, Kateri, Hayashi**

74 Agente/Representante:
de Elizaburu Márquez, Alberto

ES 2 376 463 T3

Aviso: En el plazo de nueve meses a contar desde la fecha de publicación en el Boletín europeo de patentes, de la mención de concesión de la patente europea, cualquier persona podrá oponerse ante la Oficina Europea de Patentes a la patente concedida. La oposición deberá formularse por escrito y estar motivada; sólo se considerará como formulada una vez que se haya realizado el pago de la tasa de oposición (art. 99.1 del Convenio sobre concesión de Patentes Europeas).

DESCRIPCIÓN

Método y aparato para evaluar la agregación de polipéptidos

Campo Técnico

5 Esta invención se refiere a métodos, y programas de computadora para determinación de las velocidades de agregación de los polipéptidos.

Antecedentes de la Técnica

10 Una comprensión de las propensiones a polipéptidos específicos a agregarse es de importancia crucial para dilucidar la base molecular de las enfermedades de deposición de proteínas, tales como la enfermedad de Alzheimer y otras enfermedades amiloides, y para comprender los mecanismos de acción de las mutaciones asociadas con formas hereditarias de tales enfermedades.

15 En cada una de las diversas condiciones patológicas asociadas con la deposición de proteínas y péptidos, un péptido o proteína específico que es normalmente soluble se deposita, sea intacto o en forma fragmentada, en agregados insolubles que se acumulan en uno o más tipos de tejido. Se ha encontrado que numerosas mutaciones están asociadas con formas familiares de enfermedades de deposición de proteínas y se ha demostrado que más de 20 100 implican directamente la secuencia del péptido o proteína responsable de la agregación (Siepen y Westhead, 2002). Muchas de estas mutaciones se han identificado a lo largo de los últimos 5 años, y se espera que el número aumente espectacularmente en el futuro próximo. La investigación de los mecanismos por los cuales las mutaciones naturales dan como resultado comportamiento patológico ha demostrado ser de importancia fundamental para explorar la base molecular de la enfermedad subyacente, incluso en aquellos casos en los que las mismas tienen origen esporádico en lugar de familiar (Selkoe, 2001; Volles & Lansbury, 2002).

Se ha encontrado que la capacidad de formar agregados altamente organizados que tienen características estructurales comunes, tales como amiloides, es una propiedad genérica de los polipéptidos, con indiferencia de la semejanza de secuencia o de estructura, y no simplemente una característica de pequeños números de proteínas asociadas con condiciones patológicas reconocidas (Dobson, 2001).

25 En el estado nativo, los residuos hidrófobos se incrustan usualmente en el centro de una proteína, por lo que la oportunidad de que estos residuos interactúen es limitada. Sin embargo, las proteínas son dinámicas y existe un equilibrio entre la conformación estable y plegada, y estados desestabilizados, parcial o completamente desplegados. El valor de la energía libre (ΔG , kJ mol^{-1}) para una proteína proporciona una indicación de la estabilidad de la proteína. La agregación ocurre cuando las proteínas en su estado nativo se desnaturalizan; a 30 medida que la proteína se despliega, se rompen enlaces intramoleculares, permitiendo que queden al descubierto la cadena principal del polipéptido (columna vertebral) y cadenas laterales hidrófobas. Entonces pueden formarse enlaces de hidrógeno y otras interacciones entre las moléculas de proteína parcial o totalmente desnaturalizadas, dando como resultado asociaciones intermoleculares y formación de agregados.

35 En algunos casos, puede ser deseable que se formen agregados, en particular fibrillas, por ejemplo para uso como materiales plásticos, en electrónica, como conductores, para catálisis o como una forma de liberación lenta del polipéptido, o donde las fibrillas de polipéptido deben hilarse en un "hilo" de polipéptido para diversas aplicaciones; por ejemplo, como se describe en las solicitudes de patente publicadas WO 0017328 (Dobson) y WO 0242321 (Dobson & McPhee).

40 Sin embargo, en otras circunstancias la formación de agregados es desventajosa, por ejemplo, cuando se desea utilizar un polipéptido a concentraciones o en condiciones deseables para actividad fisiológica, administración terapéutica o aplicación industrial. En particular, el uso de péptidos y proteínas bioactivos como agentes farmacéuticos es limitado en los casos en que el péptido o la proteína tiende a formar agregados durante la producción, el procesamiento, el almacenamiento o después de su administración. Estas cuestiones están plenamente reconocidas en la industria biotecnológica y farmacéutica y constituyen un problema y una carga 45 económica importantes, que pueden ser difíciles de resolver y pueden requerir el uso de técnicas complejas de expresión y replegamiento, el desarrollo de formulaciones específicas, agentes estabilizadores y excipientes, suministro de cadena fría, o reconstitución inmediata antes de la utilización. Prácticamente la totalidad de los productos terapéuticos polipeptídicos conocidos presentan estos problemas, v.g. insulina, interferón-gamma, HMPs, calcitonina, glucagón, anticuerpos.

50 Se conocen diversos factores que afectan a la propensión a un polipéptido a agregarse. Algunos de estos factores son locales para residuos de aminoácidos, y otros factores son globales y pueden afectar a la proteína completa. Por ejemplo, cuando se producen mutaciones en un polipéptido, los factores locales en la región de la mutación tales como carácter hidrófobo incrementado, o propensión a la conversión de conformación de hélice α a lámina β , dan como resultado una velocidad de agregación mayor que la de la proteína del tipo salvaje (no mutante). Los cambios 55 "globales" o totales debidos a mutaciones pueden afectar también a la velocidad de agregación; por ejemplo, un cambio en la carga neta del polipéptido mutante que lo acerca a la neutralidad da como resultado una propensión

incrementada de un polipéptido a la agregación. Las mutaciones que desestabilizan el estado nativo del polipéptido dan también como resultado una agregación facilitada.

5 Broome, B.H. & Hecht, M.H., Nature disfavors sequences of alternating polar and non-polar amino acids: Implications for Amyloidogenesis. J. Mol. Biol. (2000), 296, 961-968) describen un análisis de 250.514 secuencias de proteínas (79.708.024 residuos) para todos los patrones binarios posibles de residuos de aminoácidos polares y no polares. Los autores encontraron que existen patrones alternantes con una frecuencia significativamente menor que otros patrones con composiciones similares, y que esta infra-representación acoplada con la observación de que tales patrones promueven estructuras semejantes a amiloides en proteínas *de novo* sugiere que las secuencias de aminoácidos alternantes polares y no polares son inherentemente amiloidogénicas y por consiguiente están desfavorecidas por la selección evolutiva.

10 Un estudio detallado de mutaciones en una proteína modelo, la acilfosfatasa muscular (AcP), demostró que la velocidad de agregación para un conjunto de conformaciones parcialmente desnaturalizadas puede seguirse fácilmente para AcP utilizando una diversidad de sondas espectroscópicas. Se determinó la velocidad de agregación para más de 50 variantes de mutación de esta proteína (Chiti et al., 2002a; 2002b: Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. & Dobson, C. M. Kinetic partitioning of protein folding and aggregation. Nature Struct. Biol. 9, 137-143 (2002a); Chiti, F., Calamai, M., Taddei, N., Stefani, M., Ramponi, G. & Dobson, C. M. Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. Proc. Natl. Acad. Sci. USA, 99: 16419-16426 (2002b)). Se encontró que muchas de estas mutaciones, particularmente las que implican los residuos 16-31 y 87-98, perturban la velocidad de agregación de AcP muy significativamente (Chiti et al., 2002a; 2002b). Chiti (2002a) llegó a la conclusión de que los cambios medidos en la velocidad de agregación después de mutación estaban correlacionados positivamente con cambios en la hidrofobicidad y la propensión a lámina β de las regiones de la proteína en las cuales están localizadas las mutaciones. Chiti (2002b) examinó mutaciones de AcP que alteraban el estado de carga de la proteína AcP sin afectar significativamente al carácter hidrófobo o las propensiones a estructura secundaria de la cadena polipeptídica. Se consignó una correlación inversa entre la velocidad de agregación de variantes proteínicas en condiciones desnaturalizantes y la carga neta global de la proteína.

15 Los factores que afectan a la velocidad de agregación de una proteína son diversos. Cuando se realizan sustituciones de aminoácidos en una proteína, se ven implicados varios factores en grados diferentes. Una mutación simple puede aumentar la carga neta, dificultando con ello la agregación (por ejemplo, el reemplazamiento de Ala por Asp en una proteína cargada positivamente). Sin embargo, la misma mutación puede aumentar la hidrofobicidad, aportando con ello una continuación aceleradora a la velocidad de agregación. Finalmente, la misma mutación cambia también las propensiones a hélice α y lámina β de la cadena polipeptídica, introduciendo otros factores. La relación entre estos factores y su importancia relativa para la agregación (solubilidad) no están bien caracterizadas.

20 Por tanto, no ha sido posible predecir con exactitud la propensión a una proteína a formar agregados insolubles y ordenados, tales como fibrillas de amiloide, ni predecir o calcular el efecto de modificaciones específicas de aminoácidos, tales como reemplazamientos, sobre la agregación/solubilidad. La imposibilidad de realizar tales predicciones o cálculos constituye un problema en el diseño y/o la manipulación de los polipéptidos, tanto *in vivo* como *in vitro*.

25 La posibilidad de predecir la agregación de los polipéptidos es de importancia crucial para la dilucidación del efecto patógeno de los grandes números de mutaciones asociados con enfermedades de deposición de las proteínas. El establecimiento de principios generales en la agregación haría posible utilizar métodos estadísticos para analizar las relaciones entre mutación, agregación y enfermedad. Una comprensión de las propensiones a proteínas específicas a agregarse podría permitir el establecimiento de criterios para modificar racionalmente las propiedades de agregación de péptidos y proteínas naturales o de diseño para procesos industriales, propósitos de investigación, tratamiento médico o aplicación biotecnológica. Adicionalmente, los métodos de la invención pueden utilizarse para identificar o diseñar secuencias polipeptídicas con una propensión reducida a la agregación, y podrían administrarse polipéptidos re-diseñados por métodos tales como terapia génica para tratar ciertos trastornos, particularmente los asociados con la agregación de las proteínas. La posibilidad de identificar o diseñar polipéptidos con propiedades de agregación específicas sería importante para el desarrollo y la producción de polipéptidos para aplicaciones en las áreas de materiales y dispositivos, tales como los descritos en WO 0017328 (Dobson) y WO 0242321 (Dobson & MacPhee).

30 Los inventores han descrito previamente la técnica para obtener una velocidad relativa de agregación para dos polipéptidos, un polipéptido de referencia de tipo salvaje y un polipéptido mutante con una secuencia de aminoácidos que tiene una o más modificaciones comparada con la referencia (véase WO 2004/066168 y Chiti F. et al., "Rationalization of the effects of mutations on peptide and protein aggregation rates" Nature, Vol. 424, 14 agosto 2003 (2003-08-14), páginas 805-808; una investigación acerca del efecto de las mutaciones sobre la velocidad de agregación se describe en Chiti Fabrizio et al. "Kinetic partitioning of protein folding and aggregation", Nature Structural Biology, Vol. 9, no. 2, febrero 2002 (2002-02), páginas 137-143. Otro documento elaborado por algunos de los inventores describe secuencias con una propensión elevada a formar lámina β homopolímeras: Lopez de la Paz M et al. "De novo designed peptide-based amyloid fibrils", Proceedings of the National Academy of Sciences of USA, Vol. 99, no. 25, 10 diciembre 2002 (2002-12-10), páginas 16052-16057. La técnica anterior de antecedentes

adicionales puede encontrarse en WO 01/27152, que describe un método de diseño de un polipéptido modificado que tiene una propensión alterada a agregarse, comparado con un polipéptido no modificado. El método compara la propensión a formar la estructura local de un polipéptido modificado con la propensión a formar la estructura local de un polipéptido no modificado.

- 5 Sería útil poder determinar qué partes de una secuencia de aminoácidos promueven la agregación, para determinar si es probable que un polipéptido particular forme agregados insolubles, y poder predecir el efecto que tendrá una modificación o modificaciones particulares de una secuencia de aminoácidos sobre las propiedades de agregación/solubilidad de un polipéptido.

Descripción de la Invención

- 10 La invención se expone en las reivindicaciones independientes.

Se describe un método para identificar una parte de una secuencia de aminoácidos que se predice promueve la agregación de un polipéptido definido por dicha secuencia, comprendiendo el método: determinar las propensiones a la agregación para una pluralidad de partes de dicha secuencia, y comparar dichas propensiones a la agregación a fin de determinar una o más partes de dicha secuencia que se predice promueven la agregación.

- 15 Las realizaciones de este método permiten "la determinación del perfil" una secuencia de aminoácidos para determinar aquellas regiones que es probable que promuevan la agregación. Como en el caso de los métodos de determinación de la velocidad de agregación "absoluta" descritos más adelante, uno o más factores extrínsecos (es decir, factores extrínsecos a la secuencia de aminoácidos) tales como concentración de sales, concentración de proteínas, pH, temperatura y análogos pueden tenerse también en cuenta en la determinación de las partes de la
20 secuencia que se predice promueven la agregación o la "determinación del perfil" de la secuencia. Esto puede hacerse, por ejemplo, por adición de un término dependiente de factores extrínsecos adicionales en los modelos de predicción de la propensión a agregación descritos más adelante. Preferiblemente, la determinación comprende determinar, para cada uno de una pluralidad de aminoácidos de dicha secuencia, un valor de hidrofobicidad, un valor de propensión a hélice α y/o lámina β , un valor de carga, y un valor patrón que representa un patrón de aminoácidos
25 hidrófilos y/o hidrófobos en la proximidad de cada aminoácido, multiplicar cada uno de dichos valores por un factor de escalación, y sumar dichos valores escalados para determinar dichas propensiones a la agregación. El patrón puede comprender un patrón de aminoácidos alternantes hidrófilos e hidrófobos, preferiblemente con una longitud de al menos 5 aminoácidos.

- El método puede comprender adicionalmente la modificación de dicha secuencia de aminoácidos y la repetición de dicha determinación de la propensión relativa a la agregación, a fin de identificar una o más partes de dicha
30 secuencia que se predice promueven la agregación, en particular para cada una de una pluralidad de posiciones en dicha secuencia de aminoácidos, seleccionando cada uno de una pluralidad de aminoácidos alternativos para dicha determinación de la propensión relativa repetida. El método puede incluir la comparación de dichas propensiones a la agregación determinadas repetidamente a fin de identificar una o más partes de dicha secuencia que se predice
35 promueven la agregación.

Se describe también un método para diseñar un polipéptido que incluye predecir una velocidad de agregación para uno o más polipéptidos utilizando el método anterior.

- Se describen también métodos para sintetizar un polipéptido, en particular un método para sintetizar un polipéptido diseñado por el método anterior; y un método para sintetizar un polipéptido que incluye predecir una velocidad de
40 agregación para uno o más polipéptidos. Se proporciona también un polipéptido obtenible u obtenido por un método para sintetizar un polipéptido de acuerdo con la invención.

- Se describe también un código de programa de ordenador para, por ejecución del mismo, identificar una parte de una secuencia de aminoácidos que se predice promueve la agregación de un polipéptido asociado con la secuencia, comprendiendo el código: determinar las propensiones relativas a la agregación para una pluralidad de partes de
45 dicha secuencia; y comparar dichas propensiones relativas a la agregación a fin de determinar una o más partes de dicha secuencia que se predice promueven la agregación.

- Se describe también un sistema de computadora para identificar una parte de una secuencia de aminoácidos que se predice promueven la agregación de un polipéptido asociado con la secuencia, comprendiendo el sistema de computadora: un almacén de datos a almacenar para cada uno de una pluralidad de aminoácidos de dicha
50 secuencia, un valor de hidrofobicidad, un valor de propensión a hélice α y/o lámina β y un valor de carga, un almacén de programas que almacena un código implementable del procesador; y un procesador, acoplado a dicho almacén de programas y a dicho almacén de datos para implementación de dicho código almacenado, comprendiendo el código la codificación para controlar el procesador a fin de: alimentar dicha secuencia de aminoácidos; leer, para cada uno de una pluralidad de aminoácidos de dicha secuencia, uno de dichos valores de
55 hidrofobicidad, un valor de dicha propensión a hélice α y/o lámina β , y un valor de dicha carga, a partir de dicho almacén de datos; determinar los datos de propensión relativa a la agregación para una pluralidad de partes de dicha secuencia con respecto a dicha hidrofobicidad, propensión a hélice α y/o lámina β , y valores de carga y de un valor patrón dependiente de un patrón de aminoácidos hidrófilos y/o hidrófobos en dicha secuencia; y emitir dichos

datos de propensión relativa a la agregación a fin de identificar una parte de dicha secuencia que se predice promueve la agregación de un polipéptido asociado con la secuencia.

5 Se describe también un método de determinación de datos de velocidad de agregación que predicen una velocidad de agregación de un polipéptido definido por una secuencia de aminoácidos, comprendiendo el método: determinar un valor de hidrofobicidad, un valor de carga, y al menos un valor de propensión a forma para dicha secuencia; identificar uno o más patrones que influyen en la agregación dentro de dicha secuencia; determinar un valor patrón para la secuencia que responde a dicha identificación; y determinar dichos datos de velocidad de agregación por determinación de una combinación ponderada de dicho valor de hidrofobicidad, dicho valor de carga, dicho al menos un valor de propensión a forma, dicho valor patrón y al menos un factor extrínseco a dicha secuencia de aminoácidos.

10 Preferiblemente, los datos de velocidad de agregación predicen una velocidad de agregación de dicho polipéptido en una solución, y el al menos un factor extrínseco comprende un factor relativo a la solución, por ejemplo uno o más factores seleccionados de un valor de pH de dicha solución, una fuerza iónica de dicha solución y una medida de una concentración de dicho polipéptido en dicha solución. Factores adicionales que pueden emplearse incluyen temperatura, viscosidad, constante dieléctrica y posiblemente, un factor de ajuste dependiente de si la solución está o no agitada.

15 El al menos un valor de propensión a forma comprende preferiblemente un valor de propensión a hélice α y/o un valor de propensión a lámina β . Preferiblemente, la determinación de la hidrofobicidad, los valores de carga y de propensión a forma de la secuencia comprende sumar los valores de hidrofobicidad, carga y propensión a forma para cada uno de una pluralidad de aminoácidos de la secuencia. Preferiblemente, la velocidad de agregación comprende una velocidad de agregación logarítmica.

20 El patrón puede incluir un patrón de aminoácidos alternantes hidrófobos e hidrófilos, teniendo preferiblemente una longitud de 5 o más aminoácidos. Sin embargo, pueden emplearse patrones adicionales o alternativos tales como un patrón de 3 o más residuos hidrófobos consecutivos. Al igual que los patrones que se cree promueven la agregación, la predicción de la velocidad de agregación puede ser sensible a la identificación de patrones inhibidores de la agregación dentro de la secuencia, tales como cargas consecutivas, prolinas y análogos.

25 Se describe también un método para diseñar un polipéptido, comprendiendo dicho método de diseño un método de acuerdo con la invención. La invención proporciona adicionalmente un método para sintetizar un polipéptido que comprende diseñar un polipéptido utilizando un método de diseño de la invención y sintetizar un polipéptido de acuerdo con dicho diseño. Se proporciona también un polipéptido obtenible u obtenido utilizando un método de síntesis de la invención.

30 Se describe también un polipéptido obtenible u obtenido por determinación de los datos de velocidad de agregación que predicen una velocidad de agregación de un polipéptido definido por una secuencia de aminoácidos y sintetizar un polipéptido con dicha secuencia de aminoácidos. Por ejemplo, los métodos anteriores pueden utilizarse para predecir una propiedad de agregación para un polipéptido (o para muchos polipéptidos, seleccionando entonces uno o más), y posteriormente pueden sintetizarse un polipéptido o polipéptidos con la o las secuencias de aminoácidos definidas. La síntesis de polipéptidos puede realizarse, por ejemplo, por síntesis química, o utilizando métodos de biología molecular. La síntesis de un polipéptido o polipéptidos puede realizarse por un método automatizado.

El término polipéptido, como se utiliza en esta memoria, abarca proteínas y péptidos.

40 Utilizando los métodos de acuerdo con las realizaciones de la invención, las velocidades de agregación de los polipéptidos pueden racionalizarse y predecirse en una extensión notable sobre la base de principios físicos simples: Los efectos que tienen las modificaciones sobre los parámetros fundamentales de hidrofobicidad y propensión a estructura secundaria en el sitio de modificación, y sobre la carga de la molécula como un todo. Basándose en estos métodos, pueden diseñarse polipéptidos modificados (v.g. mutantes) que son más o menos propensos a agregarse (que tienen una solubilidad menor o mayor), o que tienen una propensión a agregación dentro de un intervalo deseado. Así, es posible evaluar los efectos que tendrán diversas modificaciones de aminoácidos sobre las propiedades de un polipéptido sin tener que producir polipéptidos modificados y medir experimentalmente el efecto de los cambios. El diseño de números masivos de polipéptidos modificados es potencialmente factible. Esto es importante dado que las modificaciones pueden seleccionarse también para satisfacer otros criterios o restricciones, tales como estabilidad de las proteínas, función, etc.

45 Puede utilizarse una escala de hidrofobicidad de consenso para asignar un valor de hidrofobicidad para cada aminoácido. Pueden utilizarse diferentes escalas de hidrofobicidad para valores distintos de pH, por ejemplo, las escalas descritas en Cowan, R. & Whittaker, R.G. (1990) Peptide Research 3: 75 - 80) se pueden utilizar para calcular la hidrofobicidad de los polipéptidos a pH bajo. Puede utilizarse una escala de hidrofobicidad ponderada, que se puede obtener utilizando una combinación de escalas, tales como las disponibles en la bibliografía (v.g. Fauchere J.-L & Pliska V.E. (1983) Eur. J. Med. Chem. 18: 369 - 375; Kyte J., Doolittle R.F. (1982) J. Mol. Biol. 157: 105 - 132). En una realización preferida, el valor de hidrofobicidad para cada aminoácido se asigna utilizando los valores dados en la Tabla 1 para hidrofobicidad de los 20 residuos de aminoácidos a pH neutro basados en los

coeficientes de reparto de agua a octanol; los datos proceden de la columna 6 de la Tabla 4.8 en Creighton (1993) (Creighton, T. E. en *Proteins. Structure and molecular properties. Second edition.* W. H. Freeman & Company (Nueva York, 1993), p.154.).

5 Las propensiones predichas para hélice α pueden calcularse utilizando algoritmos de software de determinación de patrones tales como AGADIR (www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html) Muñoz & Serrano (1994) *Nature Structural Biol* 1, 399-409; Muñoz & Serrano (1994) *J Mol Biol* 245, 297 - 308; Muñoz & Serrano (1997) *Biopolymers* 41 495 509 y Lacroix et al (1998) *J Mol Biol* 284 173 -191; PHD (Rost, B. et al. (1993) *J Mol Biol* 232, 584-599); PROF (Rost, B. et al, (1996) *Methods Enzymol* 266, 525-539); GOR4 (Garnier J et al (1978) *J Mol Biol* 120, 97 -120; Garnier J et al (1996) *Methods Enzymol* 266, 540 - 553). Cualesquiera algoritmos adecuados
10 basados en bases de datos estructurales, bases de datos estructurales de preferencia o bases de datos de preferencia de rotámeros podrían utilizarse para este cálculo a fin de estimar las propensiones a hélice, por ejemplo **GOR IV**: J. Garnier, J.F. Gibrat y B.Robson en *Methods Enzymol.*, vol 266, p 540-553 (1996). J.Garnier, D.Osguthorpe y B.Robson (*J.Mol.Biol.*120, 97, 1978). *J Mol Biol* 1987 diciembre 5; 198(3):425-443 (GOR-III); **PHD**: Rost B, Sander C. *J Mol Biol* 1993 julio 20; 232(2):584-99. Rost B, Sander C. *Proteins* 1994 mayo; 19(1):55-72;
15 **PREDATOR** Frishman D, Argos P. *Protein Eng* 1996 febrero; 9(2):133-142; **SIMPA/SIMPA96**: Levin JM, Robson B, Garnier J. *FEBS Lett* 1986 septiembre 15; 205 (2):303-308. J. LEVIN, J. GARNIER. *Biochim. Biophys. Acta*, (1988) 955, 283-295. Levin JM. *Protein Eng.* (1997), 7, 771-776. **SOPM/SOPMA** Geourjon C, Deleage G. *Protein Eng* 1994 febrero; 7(2):157-164. Geourjon C, Deleage G. *Comput Appl Biosci* 1995 diciembre; 11(6):681-684.

20 Los valores de propensión a lámina β para la totalidad de los 20 aminoácidos pueden determinarse utilizando una escala publicada. Una escala preferida se da en la Tabla 1, que proporciona valores de propensión a lámina β para 19 residuos de aminoácidos (todos excepto prolina), valores que están normalizados desde 0 (alta propensión a lámina β) a 1 (baja propensión a lámina β). Estos datos proceden de la columna 4 de la Tabla 1 de Street y Mayo (1999) (Street, A. G. & Mayo, S. L. *Intrinsic β -sheet propensities result from van der Waals interactions between side chains and the local backbone.* *Proc. Natl. Acad Sci. USA*, 96, 9074-9076 (1999)). La propensión a lámina β de la prolina no se consigna debido a la dificultad en la determinación experimental de la misma. La propensión a lámina β
25 de la glicina se obtiene a partir de cálculos teóricos.

Se describe también un sistema de computadora para determinación de los datos de velocidad de agregación a fin de predecir una velocidad de agregación de un polipéptido con una secuencia de aminoácidos definida, comprendiendo el sistema de computadora: un almacén de datos para almacenar datos que comprenden datos de hidrofobicidad, datos de propensión a forma y datos de carga para una serie de aminoácidos; un almacén de programas que almacena un código implementable por un procesador; y un procesador, acoplado a dicho almacén de programas y a dicho almacén de datos para implementar dicho código almacenado, comprendiendo el código un código para controlar el procesador a fin de: introducir una secuencia de aminoácidos para dicho polipéptido y datos relativos a al menos un factor extrínseco para dicha secuencia de aminoácidos; determinar un valor de hidrofobicidad, un valor de carga, y al menos un valor de propensión a forma para dicha secuencia; identificar uno o más patrones que influyen en la agregación dentro de dicha secuencia; determinar un valor patrón para la secuencia que responde a dicha identificación; y determinar dichos datos de velocidad de agregación por determinación de una combinación ponderada de dicho valor de hidrofobicidad, dicho valor de carga, dicho al menos un valor de propensión a forma, dicho valor patrón y dichos datos de factores extrínsecos.

40 El valor de propensión a forma puede comprender propensión a lámina β , expresada por ejemplo en términos de energía libre, y puede comprender además propensión a hélice α , determinada por ejemplo utilizando un código dentro del sistema de computadora o por medio de una solicitud enviada a un sistema de computadora separado, por ejemplo en una red. Puede emplearse una serie de aminoácidos que comprende, por ejemplo, todos los residuos de aminoácidos naturales.

45 Se describe también un código de programa de computadora para, durante su ejecución, determinar datos de velocidad de agregación a fin de predecir una velocidad de agregación de un polipéptido con una secuencia de aminoácidos definida, comprendiendo el código un código para: determinar un valor de hidrofobicidad, un valor de carga y al menos un valor de propensión a forma para dicha secuencia; identificar uno o más patrones que influyen en la agregación dentro de dicha secuencia; determinar un valor patrón para la secuencia que responde a dicha identificación; y determinar dichos datos de velocidad de agregación por determinación de una combinación ponderada de dicho valor de hidrofobicidad, dicho valor de carga, dicho al menos un valor de propensión a forma, dicho valor patrón y al menos un factor extrínseco a dicha secuencia de aminoácidos.

50 El código de programas puede proporcionarse sobre un portador de datos o medio de almacenamiento, tal como un disco duro o flexible, ROM o CD-ROM, o sobre un portador de señal óptica o eléctrica, por ejemplo como una imagen en disco o DLL (biblioteca de enlaces dinámicos) por la vía de una red de comunicaciones. Así, pueden ponerse a disposición realizaciones de la invención, o descargarse, o utilizarse por la vía de un sitio en la web. El código de control del procesador puede comprender un código de programa en cualquier lenguaje convencional de programación, por ejemplo C o Assembler o un código de máquina, y pueden implementarse realizaciones de la invención en un sistema de computadora de uso general o en un aparato de síntesis de péptidos, preferiblemente un aparato para síntesis automática de un polipéptido basado en los resultados obtenidos por aplicación de los métodos. La invención abarca también polipéptidos sintetizados de este modo.

Breve Descripción de los Dibujos

- La Figura 1 muestra el cambio de velocidad de agregación de AcP resultante de una mutación representado gráficamente contra (a) el cambio de hidrofobicidad predicho, (b) la propensión a conversión de una conformación en hélice α a una conformación en lámina β y (c) carga.
- 5 La Figura 2 (a) muestra el cambio calculado frente al observado de la velocidad de agregación después de mutación para los péptidos cortos o proteínas desplegadas naturalmente enumeradas en la Tabla 2.
- La Figura 2 (b) muestra el cambio calculado frente al observado de la velocidad de agregación después de mutación para 27 sustituciones de aminoácidos de AcP dentro de dos regiones de la secuencia que parecen ser relevantes para la agregación y que abarcan los residuos 16-31 y 87-98.
- 10 La Figura 3 muestra un diagrama de bloques de un sistema de computadora para implementación de un primer procedimiento de determinación de la velocidad de agregación.
- La Figura 4 muestra un diagrama de flujo de un procedimiento comparativo de determinación de la velocidad de agregación.
- 15 La Figura 5 muestra un diagrama de flujo de un procedimiento de determinación de candidatos para síntesis automática de proteínas.
- La Figura 6 muestra un diagrama de flujo de un procedimiento para determinación de la propensión relativa intrínseca a la agregación.
- La Figura 7 muestra el perfil de propensión a la agregación de amiloides de PrP.
- La Figura 8 muestra el perfil de agregación de AcP y regiones sensibles.
- 20 La Figura 9 muestra el perfil de agregación de A β 42 y regiones sensibles.
- La Figura 10 muestra perfiles de agregación de PI3 SH3 y α -espectrina SH3 y regiones sensibles.
- La Figura 11 muestra un gráfico de velocidades de agregación absolutas calculadas (logarítmicas) contra las velocidades determinadas experimentalmente.
- 25 Las Figuras 12a y 12b muestran, respectivamente, una distribución de los coeficientes de correlación para la determinación de la velocidad de agregación absoluta, y un gráfico de velocidades de agregación absolutas predichas (logarítmicas) contra las velocidades determinadas experimentalmente.
- La Figura 13 muestra un diagrama de flujo de un procedimiento de acuerdo con una realización de la presente invención para determinación de una estimación de una velocidad de agregación absoluta de un polipéptido.
- 30 En primer lugar se describirán, como antecedentes técnicos útiles para la comprensión de la invención, ejemplos relativos a la predicción de una velocidad de agregación relativa de un polipéptido con relación a un polipéptido de referencia. A continuación se describirán, en el Ejemplo 6, realizaciones de la invención para la determinación del perfil de un polipéptido a fin de identificar regiones sensibles para la agregación y, en el Ejemplo 7, realizaciones afines de la invención para determinar las velocidades de agregación "absolutas" de los polipéptidos.

Ejemplos**35 Ejemplo 1: Trabajo Experimental con AcP (de WO 2004/066168)**

- Se midieron las velocidades de agregación para la proteína AcP de tipo salvaje (v_{wt}) y para diversos mutantes (variantes) de AcP (v_{mut}) después de desnaturalización en 25% TFE, a partir de las evoluciones temporales de la fluorescencia de ThT, como ha sido descrito por Chiti *et al.*, 2002a (Chiti, F., Taddei, N., Baroni, F., Capanni, C., Stefani, M., Ramponi, G. & Dobson, C. M. Kinetic partitioning of protein folding and aggregation. *Nature Struct. Biol.* 9, 137-143 (2002a)). Todas las medidas de velocidad de agregación se realizaron en condiciones en las cuales todas las variantes de proteínas están constituidas por conjuntos de conformaciones relativamente desestructuradas. El cambio de velocidad de agregación como resultado de una mutación se expresó en todos los casos como el logaritmo natural de la ratio de las constantes de velocidad de agregación de la proteína mutante y la de tipo salvaje ($\ln(v_{mut}/v_{wt})$).
- 40
- 45 En la Tabla 1, los valores de hidrofobicidad de los 20 residuos de aminoácidos a pH neutro están basados en los coeficientes de reparto de agua a octanol. Estos datos proceden de la columna 6 de la Tabla 4.8 en Creighton (1993) (Creighton, T. E. *Proteins. Structure and molecular properties*. Segunda edición. W. H. Freeman & Company (Nueva York, 1993), p. 154.). Las propensiones a lámina β de los 20 residuos de aminoácidos están normalizadas desde 0 (alta propensión a lámina β) a 1 (baja propensión a lámina β). Estos datos proceden de la
- 50 columna 1 de la Tabla 4 de Street y Mayo (1999) (Street, A. G. & Mayo, S. L. *Intrinsic R-sheet propensities result*

from van der Waals interactions between side chains and the local backbone. Proc. Natl. Acad. Sci. USA, 96, 9074-9076 (1999)). La propensión a lámina β de la prolina no se consigna debido a la dificultad en su determinación experimental. La propensión a lámina β de la glicina se obtiene a partir de cálculos teóricos. Los valores de carga se dan para pH neutro. Los valores entre paréntesis corresponden a un pH inferior a 6,0, cuando el residuo histidina está cargado positivamente.

5

Tabla 1. Escalas de hidrofobicidad, propensión a lámina β y carga para los 20 aminoácidos naturales			
Residuo de aminoácido	Hidrofobicidad (kJ mol ⁻¹)	Propensión a lámina β	Carga
Arg (R)	3,95	0,35	+1
Lys (K)	2,77	0,34	+1
Asp (D)	3,81	0,72	-1
Glu (E)	2,91	0,35	-1
Asn (N)	1,91	0,40	0
Gln (Q)	1,30	0,34	0
His (H)	0,64 (2,87)	0,37	0 (+1)
Ser (S)	1,24	0,30	0
Thr (T)	1,00	0,06	0
Tyr (Y)	-1,47	0,11	0
Gly (G)	0,00	0,60	0
Pro (P)	-0,99	n.d.	0
cys (C)	-0,25	0,25	0
Ala (A)	-0,39	0,47	0
Trp (W)	-2,13	0,24	0
Met (M)	-0,96	0,26	0
Phe (F)	-2,27	0,13	0
Val (V)	-1,30	0,13	0
Ile (I)	-1,82	0,10	0
Leu (L)	-1,82	0,32	0

Utilizando los datos de la Tabla 1, el cambio de hidrofobicidad ($\Delta Hydr$), la propensión a conversión de estructura en hélice α a estructura en lámina β ($\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$) y el cambio de carga (Δcarga) se cuantificaron para AcP utilizando los valores tabulados para todos los residuos de aminoácidos.

10 El cambio en hidrofobicidad ($\Delta Hydr$) resultante de la mutación se calculó utilizando $\Delta Hydr = Hydr_{wt} - Hydr_{mut}$, donde $Hydr_{wt}$ e $Hydr_{mut}$ son los valores de hidrofobicidad de los residuos de tipo salvaje y mutante, respectivamente (los valores de hidrofobicidad para la totalidad de los 20 aminoácidos se enumeran en la Tabla 1).

Para calcular la propensión a la conversión de estructura de hélice α a estructura de lámina β ($\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$), fue necesario calcular $\Delta\Delta G_{\text{espiral-}\alpha}$ y $\Delta\Delta G_{\beta\text{-espiral}}$.

15 El cambio de energía libre para la transición espiral aleatoria \rightarrow lámina β resultante de la mutación ($\Delta\Delta G_{\beta\text{-espiral}}$) se calculó utilizando $\Delta\Delta G_{\beta\text{-espiral}} = 13,64 (P_{\beta}^{wt} - P_{\beta}^{mut})$. P_{β}^{wt} y P_{β}^{mut} son las propensiones normalizadas a lámina β del residuo de tipo salvaje y el residuo mutante, respectivamente (los valores de propensión a lámina β para la totalidad de los 20 aminoácidos se enumeran en la Tabla 1), y 13,64 es la constante de conversión de la escala normalizada a unidades de kJ mol⁻¹.

20 El cambio predicho de energía libre para la transición hélice $\alpha \rightarrow$ espiral aleatoria resultante de la mutación ($\Delta\Delta G_{\text{espiral-}\alpha}$) se calculó utilizando $\Delta\Delta G_{\text{espiral-}\alpha} = RT \ln(P_{\alpha}^{wt} / P_{\alpha}^{mut})$. P_{α}^{wt} y P_{α}^{mut} son las propensiones a hélice α predichas (porcentajes de hélice) de la secuencia de tipo salvaje y mutadas en el sitio de mutación, respectivamente, que se calcularon utilizando el algoritmo AGADIR en www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html; R = 0,008314 kJ mol⁻¹K⁻¹. (véase también Lacroix, E., Viguera AR & Serrano, L. (1998). J. Mol. Biol. 284,173-191).

El cambio de carga resultante de la mutación ($\Delta Carga$) se calculó utilizando $\Delta Carga = [Carga_{mut}] - [Carga_{wt}]$, donde $[Carga_{wt}]$ y $[Carga_{mut}]$ son los valores absolutos de carga para las secuencias de tipo salvaje y mutada, respectivamente (obtenidas de las sumas de los valores de carga de todos los residuos consignados en la Tabla 1).

5 El cambio de velocidad de agregación después de mutación $\ln(v_{mut}/v_{wt})$ se representó individualmente contra $\Delta Hydr$, contra ($\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$) y contra $\Delta Carga$, representándose estas gráficas en las Figuras 1a, 1b y 1c, respectivamente.

10 Las mutaciones consignadas en las Figuras 1a y 1b, descritas anteriormente (Chiti *et al.*, 2000a, *ibid.*), no implican cambio de carga. Las mutaciones consignadas en la Figura 1c, descrita previamente (Chiti *et al.*, 2002b, *ibid.*), se diseñaron para minimizar el cambio de hidrofobicidad y las propensiones a estructura secundaria. La mayoría de las sustituciones de aminoácidos de AcP implican residuos dentro de las dos regiones de la secuencia, abarcando los residuos 16-31 y 87-98, que se cree son relevantes para la agregación.

Las líneas de trazo continuo a través de los datos representan los ajustes óptimos a funciones lineales. Los valores r y p resultantes de cada correlación y la pendiente de los ajustes óptimos se muestran en cada caso.

15 En cada uno de los análisis, los puntos de datos están dispersados considerablemente alrededor de las líneas que representan los ajustes óptimos a funciones lineales. Esta dispersión puede atribuirse al hecho de que se considera en cada caso un solo parámetro, a la dificultad en la predicción exacta de los cambios en la hidrofobicidad y las propensiones a estructura secundaria, así como a las importancias relativas variables de los diferentes sitios de mutación en el proceso de agregación. Sin embargo, a pesar de la dispersión presente en cada gráfica, se encontró que el cambio de velocidad de agregación después de la mutación ($\ln(v_{mut}/v_{wt})$) para AcP estaba correlacionado significativamente con cada uno de estos parámetros individualmente (Figura 1a, 1b, y 1c). No obstante, a pesar de la dispersión presente en cada gráfica, se encontró que el cambio de velocidad de agregación después de la mutación ($\ln(v_{mut}/v_{wt})$) para AcP estaba correlacionado significativamente con cada uno de estos parámetros individualmente (Figura 1a, 1b, y 1c). Se calculó la dependencia media de $\ln(v_{mut}/v_{wt})$ de cada parámetro (la pendiente de la línea de ajuste óptimo resultante de cada análisis). Se encontró que los valores eran:

$$\begin{aligned} \Delta Hydr & 0,633 \\ \Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}} & 0,198 \\ \Delta Carga & 0,491 \end{aligned}$$

Después de este análisis, se dedujo la Ecuación 1 y se utilizó para determinar el cambio de velocidad de agregación después de la mutación (valor $\ln(v_{mut}/v_{wt})$):

$$30 \quad \ln(v_{mut}/v_{wt}) = 0,633 * \Delta Hydr + 0,198 * (\Delta\Delta G_{\text{espiral-}\alpha} + G_{\beta\text{-espiral}}) - 0,491 * \Delta Carga$$

donde los números que preceden a los parámetros de $\Delta Hydr$, ($\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$) y $\Delta Carga$ son valores para x, y, y z, respectivamente, que corresponden a las pendientes de las tres gráficas consignadas en Fig. 1 (es decir, las dependencias de $\ln(v_{mut}/v_{wt})$ con respecto a los tres parámetros).

35 **Ejemplo 2: Comparación del cambio observado frente al calculado en la velocidad de agregación en la mutación de la proteína AcP/velocidades de agregación relativas de las proteínas AcP mutantes de WO 2004/066168.**

40 Utilizando la Ecuación 1, se calculó el cambio de velocidad de agregación $\ln(v_{mut}/v_{wt})$ para 27 sustituciones de aminoácidos de AcP dentro de las dos regiones de la secuencia que parecen ser relevantes para la agregación y que abarcan los residuos 16-31 y 87-98. El cambio de velocidad de agregación para cada mutación se determinó experimentalmente, como se describe en el Ejemplo 1, en condiciones en las cuales todas las variantes de proteínas están constituidas por conjuntos de conformaciones relativamente desestructuradas. Los valores calculados de $\ln(v_{mut}/v_{wt})$ frente a los experimentales para todas las mutaciones de AcP se representaron gráficamente como se muestra en la Figura 2b. Se encontró que la correlación observada era altamente significativa ($r = 0,756$ y $p < 0,0001$) y la pendiente estaba próxima a 1.

45 **Ejemplo 3: Comparación del cambio observado frente al calculado en la velocidad de agregación en la mutación para una gama de polipéptidos de WO 2004/066168.**

Se aplicó la función combinada, Ecuación 1, para calcular el cambio en velocidad de agregación después de mutación (valor $\ln(v_{mut}/v_{wt})$) calculado para 26 mutaciones en los polipéptidos amilina, péptidos priónicos, α -sinucleína, péptido amiloide β , tau, repetición rica en leucina y un péptido modelo, como se enumeran en la Tabla 2.

50 Se calcularon los valores para $\Delta Hydr$, $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ y $\Delta Carga$ se calcularon para cada mutación de polipéptido utilizando los métodos descritos en el Ejemplo 1.

Las 26 mutaciones consideradas incluían mutaciones fisiológicamente relevantes asociadas con formas genéticas de enfermedades de deposición de proteínas y otras sustituciones que habían sido utilizadas en investigación para

abordar tejidos específicos. Todas ellas eran mutaciones de proteínas desestructuradas (péptidos), o polipéptidos que parecen estar desplegados naturalmente, tales como el péptido amiloide β , el polipéptido de la isleta amiloide, α -sinucleína, tau, péptidos cortos seccionados de la secuencia de la proteína priónica y otros péptidos modelo. Se consideraron únicamente en el análisis mutaciones de un solo punto dentro de péptidos desestructurados o proteínas que están desplegadas en condiciones próximas a las fisiológicas. Se incluyeron todas las mutaciones para las cuales estaban directamente disponibles valores experimentales reales de $\ln(v_{mut}/v_{wt})$ o podían determinarse a partir de datos de la bibliografía. Se excluyeron las mutaciones que actuaban simplemente por desestabilización del estado nativo de la proteína implicada. Se consideraron los datos con indiferencia de las técnicas experimentales empleadas por los diferentes autores para sondar la agregación, con la condición de que pudiera realizarse un análisis cuantitativo. Cuando no se consignaban explícitamente constantes de tiempo o de velocidad, las gráficas que describen los perfiles cinéticos de agregación se escanearon y se analizaron por computadora. Este procedimiento permitió que se reconstruyeran gráficas con valores numéricos de los puntos de datos y se analizaran para obtener valores de la constante de velocidad. Cuando eran evidentes fases de retardo y de crecimiento en los perfiles de agregación cinéticos, únicamente se consideró la fase de crecimiento. Cuando se consignaban datos en periodos de tiempo fijos (por ejemplo por medio de gráficos de barras), el valor para el valor $\ln(v_{mut}/v_{wt})$ observado se obtuvo a partir de la ratio de los parámetros de agregación de la proteína mutada y la proteína (péptido) de tipo salvaje, antes de alcanzarse el equilibrio.

Las mutaciones que implicaban residuos de prolina no se analizaron debido a la dificultad en la obtención de estimaciones cuantitativas del cambio de la propensión a lámina β como resultado de estas mutaciones (véase la Tabla 1). Tampoco se consideraron las mutaciones en las que se consignaban discrepancias sustanciales en el valor $\ln(v_{mut}/v_{wt})$ por diferentes autores (cuando estaban presentes discrepancias significativas pero no sustanciales, se consideraron los valores $\ln(v_{mut}/v_{wt})$ resultantes de las medias aritméticas de los datos disponibles).

Tabla 2. Cambios de hidrofobicidad, propensiones a estructura secundaria, carga y velocidad de agregación como resultado de mutaciones de un solo punto de péptidos desestructurados o proteínas desplegadas naturalmente

Mutación	$\Delta Hydr$ (kJ mol ⁻¹)	$\Delta\Delta G_{\beta\text{-espiral}}$ (kJ mol ⁻¹)	$\Delta\Delta G_{\text{espiral-}\alpha}$ (kJ mol ⁻¹)	$\Delta Carga$	$\ln(v_{mut}/v_{wt})$ calculado	$\ln(v_{mut}/v_{wt})$ observado	Ref.
Amilina							
N22A	2,30	-0,95	-3,36	0	0,60	0,69	11
F23A	-1,88	-4,64	-3,90	0	-2,88	-2,65	11
G24A	0,39	1,77	-2,84	0	0,04	-0,03	11
I26A	-1,43	-5,05	-0,32	0	-1,97	-2,39	11
L27A	-1,43	-2,05	0,36	0	-1,24	-0,93	11
S20G	1,24	-4,09	0,00	0	-0,03	1,01	12
Péptidos priónicos							
H111A	3,26	-1,36	-3,21	-1	1,65	0,60	13
H111K	0,10	0,41	-1,72	0	-0,20	-0,26	13
A1 17V	0,91	4,63	2,37	0	1,96	1,51	13
V21 0I	0,52	0,41	-0,97	0	0,22	0,84	14
α-Sinucleína							
A53T	-1,39	5,59	2,83	0	0,79	1,18	15
A76E	-3,30	1,64	0,00	1	-2,25	-2,72	16
A76R	-4,34	1,64	0,64	-1	-1,80	-0,93	16
Péptido amiloide β							
A21G	-0,39	-1,77	3,27	0	0,05	-0,07	17
E22K	0,14	0,14	-1,72	-2	0,76	0,92	18
E22Q	1,61	0,14	0,00	-1	1,54	2,92	17, 18
E22G	2,91	-3,41	4,30	-1	2,51	2,03	19

Tabla 2. Cambios de hidrofobicidad, propensiones a estructura secundaria, carga y velocidad de agregación como resultado de mutaciones de un solo punto de péptidos desestructurados o proteínas desplegadas naturalmente

Mutación	$\Delta Hydr$ (kJ mol ⁻¹)	$\Delta\Delta G_{\beta\text{-espiral}}$ (kJ mol ⁻¹)	$\Delta\Delta G_{\text{espiral-}\alpha}$ (kJ mol ⁻¹)	$\Delta Carga$	$\ln(v_{mut}/v_{wt})$ calculado	$\ln(v_{mut}/v_{wt})$ observado	Ref.
D23N	1,90	4,36	-1,72	-1	2,22	3,97	17
F19T	-3,27	0,95	-1,76	0	-2,23	-2,48	20
Tau							
G272V	1,30	6,41	-1,71	0	1,75	1,04	21, 22
R406W	6,08	1,50	0,00	-1	4,64	1,25	21, 22, 23
Y310W	0,66	-1,77	0,00	0	0,07	0,05	23bi s
Repetición rica en leucina							
D24N	1,90	4,36	-3,43	-1	1,88	2,08	24
D24Q	2,51	5,18	-3,10	-1	2,49	1,25	24
Péptido modelo							
D6E	0,90	5,04	-2,27	0	1,12	0,40	25

Se describe aquí el modo en que se utilizaron los datos experimentales de la bibliografía para determinar los valores experimentales de $\ln(v_{mut}/v_{wt})$ para cada una de las mutaciones consignadas en la Tabla 2 anterior: **mutaciones de amilina**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ se calcularon todos ellos a partir de Fig. 2b (puntos de datos a 4 min) de la ref. 13; **mutación S20G de amilina**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación S20G de amilina procedían de Fig. 5 de la ref. 14. Los datos se representaron de nuevo gráficamente a fin de obtener constantes de velocidad dentro de las fases de elongación. El valor de $\ln(v_{mut}/v_{wt})$ considerado en el análisis de los inventores es el valor medio de los dos valores $\ln(v_{mut}/v_{wt})$ obtenidos de las dos concentraciones consignadas; **mutaciones H111A, H111K y A117V de un péptido priónico**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en el péptido 106-126 del prión humano eran de Fig. 2 de la ref. 15. Los datos se representaron de nuevo gráficamente para determinar las velocidades iniciales de agotamiento de monómero; **mutación V2101 de un péptido priónico**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en el péptido 198-218 del prión humano eran de Fig. 8 (las velocidades de agregación se tomaron de las pendientes de la gráfica consignada) de la ref. 16; **mutación A53T de α -sinucleína**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación A53T de α -sinucleína eran de la ref. 17: los datos se tomaron de la Figura 1b (tiempo de 14 días), de la Figura 2A (tiempo de 49 días) y de la Figura 3A (tiempo de 66 días). El valor consignado de $\ln(v_{mut}/v_{wt})$ es resultado de un valor medio de los tres valores; **mutaciones A76E y A76R de α -sinucleína**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en las mutaciones A76E y A76R de α -sinucleína eran de Fig. 3 (tiempo de 2 días) de la ref. 12; **mutaciones A21G y D23N de A β** : Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en las mutaciones A21G y D23N de A β eran de Fig. 3 de la ref. 18; los datos se representaron de nuevo gráficamente para obtener la velocidad de agotamiento del rojo Congo. Esto implicaba considerar la velocidad de las 6 primeras horas para el mutante D23N (antes de alcanzar el equilibrio) y entre 0 y 48 horas para los péptidos A21G y de tipo salvaje; **mutación E22K de A β** : Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación E22K de A β eran de Fig. 2 de la ref. 19: Los puntos de datos se representaron de nuevo gráficamente y se ajustaron a una función exponencial simple a fin de obtener los valores de la constante de velocidad; **mutación E22Q de A β** : Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación E22Q de A β eran de (1) Fig. 2 de la ref. 19: los puntos de datos se representaron de nuevo gráficamente y se ajustaron a una función exponencial simple para obtener los valores de la constante de velocidad; (2) Fig. 3 de la ref. 18: los datos se representaron de nuevo gráficamente a fin de obtener la velocidad de agotamiento del rojo Congo; esto implicaba considerar la velocidad de las 6 primeras horas para el mutante E22Q (antes de alcanzar el equilibrio) y entre 0 y 48 horas para el péptido de tipo salvaje. El valor consignado de $\ln(v_{mut}/v_{wt})$ resulta de un valor medio de los dos valores; **mutación E22G de A β** : Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación E22G de A β eran de Fig. 5a, b de la ref. 20: Los datos se representaron de nuevo gráficamente para obtener la velocidad de agotamiento de monómero/dímero. Esto implicaba considerar la velocidad de las 5 primeras horas para el mutante E22G (antes de alcanzar el equilibrio) y entre 0 y 50 h para el péptido de tipo salvaje; **mutación F19T de A β** : Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación F19T de A β eran de Fig. 4 de la ref. 21: se obtuvieron valores similares de $\ln(v_{mut}/v_{wt})$ a diferentes concentraciones de A β ; **mutación R5L de tau**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación R5L de tau eran de Fig. 5 de la ref. 22: los puntos de datos se representaron de nuevo gráficamente y se ajustaron a una función exponencial simple (para la proteína de tipo salvaje) o una función exponencial doble (para el mutante) a fin de obtener valores de la constante de velocidad; los datos experimentales de $\ln(v_{mut}/v_{wt})$ se calcularon para ambas fases observadas para el mutante. El valor consignado de $\ln(v_{mut}/v_{wt})$ es resultado de un valor medio de los dos valores; **mutación G272V de tau**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación G272V de tau eran de (1) la Tabla 1 (constantes de tiempo) de la ref.

24; (2) las Figuras 5 y 6 (velocidades durante las fases de elongación) de la ref. 23. El valor consignado de $\ln(v_{mut}/v_{wt})$ es resultado de un valor medio de los 3 valores; **mutación R406W de tau**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación R406W de tau eran de la Tabla 1 (constantes de tiempo) de la ref. 24, la Figura 5 (velocidades durante las fases de elongación) de la ref. 23 y de Fig. 1 (velocidades durante las fases de elongación) de la ref. 25. El valor consignado de $\ln(v_{mut}/v_{wt})$ es resultado de un valor medio de los 3 valores; **mutación Y310W de tau**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la mutación Y310W de tau eran de la Figura 3A de la ref. 26: se representaron de nuevo gráficamente los puntos de datos en presencia de heparina y se ajustaron a funciones exponenciales simples a fin de obtener valores de la constante de velocidad: **mutaciones del péptido de repetición rico en leucina**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en la repetición rica en leucina eran de Fig. 2 de la ref. 27. Las velocidades de agregación se tomaron de las pendientes de la gráfica consignada; **mutaciones del péptido VTVKVDVAVKTV**: Los datos experimentales de $\ln(v_{mut}/v_{wt})$ en el péptido modelo de 12 residuos eran de Fig. 8 de la ref. 28 (elipticidad media del residuo en el pico dentro de la región de 215-220 nm sustraída por la elipticidad media del residuo para la espiral aleatoria obtenida de Fig. 6A).

Se representó gráficamente el valor calculado de $\ln(v_{mut}/v_{wt})$ frente al valor experimental y se muestra en la Figura 2(a). La correlación altamente significativa ($r = 0,84$, $p < 0,0001$), y el valor de la pendiente que está próximo a 1,0, indican una concordancia estrecha entre los efectos calculados y los experimentales de las mutaciones sobre las velocidades de agregación de este grupo heterogéneo de polipéptidos. Los cambios observados de velocidad de agregación después de la mutación abarcan un intervalo de aprox. 800 veces, a saber desde 15 veces más lenta a 53 veces más rápida que el polipéptido de tipo salvaje correspondiente (Figura 2a y Tabla 2). El 84% de estas mutaciones tienen valores calculados de $\ln(v_{mut}/v_{wt})$ que varían dentro de un factor de 3 respecto a los valores observados de $\ln(v_{mut}/v_{wt})$. El porcentaje aumenta hasta 92% y 96% si se consideran factores de dispersión de 5 y 10, respectivamente. Los ejemplos en los que se encuentra concordancia estrecha entre los valores teóricos y los experimentales incluyen mutaciones asociadas con encefalopatías espongiiformes hereditarias, tales como las sustituciones A117V y V210I de la proteína priónica (Tabla 2). Los valores predichos y experimentales están también en concordancia estrecha para la mutación A53T asociada con la enfermedad de Parkinson de aparición precoz y para diversas mutaciones asociadas con el péptido amiloide β y responsables de la enfermedad de Alzheimer de aparición precoz o de hemorragia cerebral hereditaria con amiloidosis (Tabla 2).

Si se repite el análisis utilizando únicamente un solo determinante para calcular los valores $\ln(v_{mut}/v_{wt})$, se encontraron todavía correlaciones significativas entre los valores calculados y los observados de $\ln(v_{mut}/v_{wt})$ ($p = 0,0003$ utilizando únicamente $\Delta Hydr$ para calcular $\ln(v_{mut}/v_{wt})$, $p = 0,036$ utilizando solamente $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ y $p = 0,011$ utilizando sólo $\Delta Carga$). Sin embargo, estas correlaciones son menos notables que la observada cuando se consideraba una combinación de la totalidad de los 3 factores, y las pendientes son significativamente menores que 1,0 (0,61, 0,19 y 0,10 utilizando sólo $\Delta Hydr$, sólo $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ y sólo $\Delta Carga$, respectivamente). Esto demuestra que la ecuación en la que se combinan estos factores proporciona un método más exacto para determinación de la ratio de velocidad de agregación para los polipéptidos modificado (v.g. mutante) y de referencia (v.g. tipo salvaje).

Se encontró que la correlación que se muestra en la Figura 2 (a) entre los efectos teóricos y experimentales de las mutaciones sobre la agregación es notable, considerando el grupo heterogéneo de sistemas de proteínas y péptidos utilizados en el análisis, así como la variabilidad de sitios en los cuales ocurren las diversas mutaciones.

Ejemplo 4: Aplicabilidad del algoritmo a modificaciones que implican varios residuos de aminoácidos y el uso de parámetros cinéticos distintos de las "velocidades de agregación" de WO 2004/066168

Se testó la Ecuación 1 contra otros sistemas para evaluar su aplicabilidad a sistemas más generales. Los cálculos utilizados para derivar la Ecuación 1 están basados en la cinética de agregación experimentada por variantes de proteínas y péptidos que difieren en un solo residuo de la secuencia original. Las velocidades (v_{mut} y v_{wt}) utilizadas en la expresión corresponden a la fase exponencial de agregación para cada uno de los péptidos, y no incluyen ningún posible periodo de retardo o fase de nucleación precedente a dicha etapa.

Para testar la validez de esta expresión en la predicción de las propensiones a agregación se incluyeron péptidos derivados de dos variaciones de calcitonina. La primera consistió en evaluar si el efecto de varias sustituciones podría predecirse del mismo modo que el algoritmo era capaz de hacer con mutaciones de un solo punto. La segunda consistió en incluir como parámetro cinético la ratio relativa de los tiempos de agregación (T_{mut}/T_{wt}). Por inclusión del efecto de una fase de retardo sobre la cinética de la agregación exhibida por los péptidos, los tiempos de agregación para cada uno de los péptidos (t), pudieron definirse de dos maneras diferentes: la primera era el tiempo de nucleación o tiempo que precede a la iniciación de la agregación o al desarrollo de turbidez en la solución ($T1$), y la segunda podría corresponder al semi-tiempo de agregación o al tiempo en el que las variaciones en las medidas utilizadas para monitorizar la agregación (dispersión de la luz, o cualquier otro método) alcanzaban la mitad de su valor máximo ($T2$). Esto podría permitir la aplicación de la ecuación a la predicción de las propensiones a la agregación para una gama mucho más amplia de moléculas con aspectos de diseño importantes.

Los cálculos se realizaron sobre dos variantes de calcitonina, utilizando datos disponibles en la bibliografía (Arvinte, et al., 1993, J Biol Chem 268:6415-6422), y estudios previos incluidos en otra solicitud de patente por algunos de los miembros del grupo (Zurdo & Dobson, WO 02/083734, PCT/GB02/01778). Los cálculos se efectuaron utilizando

datos descritos en dichas publicaciones, obteniéndose los valores indicados en la Tabla 3. En ambos casos, el valor para el parámetro τ_{wt} se obtuvo independientemente.

Tabla 3. Cambios predichos y experimentales en los tiempos de agregación exhibidos por diversos péptidos de calcitonina cuando se comparan con la secuencia humana				
	$\ln(v_{mut}/v_{wt})$ calculado	$\ln(v_{mut}/v_{wt})$ observado	(τ_{mut}/τ_{wt}) calculado	(τ_{mut}/τ_{wt}) observado
¹ Salmón-1	-10,54	-10,31	37.681,05	~30.000 ^a
² SEQ ID NO 14	-5,60	-4,61 ^b / -5,71 ^a	271,70	100 ^b / 300 ^a

¹ Datos obtenidos de Arvinte et al. (1993) J Biol Chem 268, 6415-6422. La calcitonina de salmón tiene 16 posiciones modificadas cuando se compara con la secuencia humana.

² Sequence comunicada en Zurdo & Dobson (WO 02/083734, PCT/GB02/01778), y Zurdo & Dobson (observaciones no publicadas). La secuencia ID NO 14 exhibe 6 posiciones modificadas cuando se compara con la secuencia humana.

^a Los valores para el cálculo de τ se obtuvieron utilizando $T1$ como se ha descrito arriba.

^b Los valores para el cálculo de τ se obtuvieron utilizando $T2$ como se ha descrito arriba.

Los cálculos para los cambios en el tiempo de agregación se efectuaron suponiendo las relaciones siguientes con las velocidades de agregación descritas por la ecuación 1.

$$5 \quad (\tau_{mut} / \tau_{wt}) = (v_{wt} / v_{mut}) = 1 / \exp(\ln(v_{mut} / v_{wt}))$$

Este análisis demuestra que la Ecuación 1 puede utilizarse para predecir el comportamiento de agregación de un polipéptido dado que tiene más de una modificación de aminoácido comparada con la secuencia del polipéptido original. Además, sugiere que en los sistemas en los que está presente una fase de retardo, o la velocidad de agregación puede ser difícil de calcular, los parámetros cinéticos alternativos representados por los tiempos de agregación ($T1$ - tiempo de nucleación - o $T2$ - semi-tiempo de agregación -) pueden proporcionar valores válidos para comparar con las predicciones dadas por la Ecuación 1.

Ejemplo 5: Aplicabilidad del algoritmo a modificaciones que implican adición o delección de residuos de aminoácidos: péptidos A β relacionados con la enfermedad de Alzheimer

Los péptidos A β (1-40) y A β (1-42) que están asociados con la enfermedad de Alzheimer exhiben diferencias en sus propensiones a la agregación. Los péptidos difieren en secuencia únicamente por dos residuos en el término C. Los métodos de la invención explican la mayor propensión a agregarse de la forma de 42 residuos, con relación a la forma de 40 residuos, del péptido amiloide β asociado con la enfermedad de Alzheimer (Jarrett et al., 1993). De hecho, aunque la propensión a hélice α y la carga del péptido completo parecen mantenerse inalteradas después de la adición del dipéptido Ile-Ala en el término C, los valores de hidrofobicidad y propensión a lámina β de los dos residuos son mayores que los valores medios calculados para el péptido completo.

Desde un punto de vista cuantitativo, el cambio de hidrofobicidad resultante de la adición de los dos residuos en el término C puede calcularse como $\Delta Hydr = Hydr_{wt} - Hydr_{mut}$, donde $Hydr_{wt}$ es la hidrofobicidad media de los 40 residuos que constituyen la forma corta del péptido; $Hydr_{mut}$ es la hidrofobicidad media de los dos residuos insertados (Ile-Ala). El cambio de propensión a lámina β resultante de la inserción puede calcularse análogamente. Esto conduce a la predicción de que la forma larga se agrega siete veces más rápidamente que la forma corta, en concordancia satisfactoria con el perfil cinético consignado por Jarrett et al., 1993, que encontró una aceleración de 7-8 veces (Jarrett JT, Berger EP, Lansbury PT Jr. The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. Biochemistry, 32, 4693-4697 (1993)).

Velocidades de agregación - ejemplo

Una velocidad de agregación puede definirse por una constante de velocidad en una ecuación de agregación, por ejemplo agregación = $A(1 - e^{-kt})$ donde t es el tiempo. La agregación puede medirse, por ejemplo, en términos de un periodo de tiempo para, supongamos, 50% de agregación. En las ecuaciones descritas en esta memoria puede emplearse una velocidad de agregación o un log (preferiblemente log natural) de velocidad de agregación.

En la práctica, puede determinarse un grado de agregación o constante de velocidad de agregación, por ejemplo, por turbidez o dispersión de la luz, o uno de muchos otros medios - por ejemplo a partir de trazas cinéticas obtenidas por los métodos siguientes: fluorescencia de ThT, turbidez, CD, o análisis directos masa/volumen, tales como sedimentación, cromatografía de exclusión por tamaños, y filtración. No obstante, aunque estos métodos detectan aspectos ligeramente diferentes de la agregación, los mismos están estrechamente ligados, y el (log) de la velocidad de agregación medido es aproximadamente independiente de la técnica de medida empleada.

En algunos casos la agregación es un suceso 'espontáneo' precedido por un retardo temporal, y en estos casos la velocidad de agregación puede corresponder al retardo temporal antes del comienzo de la agregación. Una medida de la velocidad de agregación de este tipo parece estar relacionada con la constante de velocidad de agregación mencionada anteriormente, pero puede no corresponder directamente. En algunos casos, los sistemas sembrados y no sembrados dan como resultado velocidades de agregación cuasi-idénticas si se desprecia la fase de retardo en el sistema no sembrado, pero esto no ocurre siempre. Las ecuaciones descritas en esta memoria, dependiendo de los factores de escalación, pueden emplearse para cualquiera de estos tipos de medida de la velocidad de agregación, o ambos.

Debido a dificultades en la cuantificación de la 'agitación' y su influencia sobre la cinética, ésta no se consideró en los ejemplos descritos en esta memoria, aunque los efectos de la agitación podrían incluirse en las ecuaciones aquí descritas.

Ejemplo de implementación por sistema de computadora de los métodos arriba descritos

Haciendo ahora referencia a la Figura 3, ésta muestra un diagrama de bloques de un sistema de computadora para implementar el método arriba descrito. Un sistema de computadora 300 de uso general comprende un procesador 300a acoplado a un código de programas de computadora de almacenamiento de memoria de programas 300b para implementar el método, como se describe adicionalmente más adelante, e interfaces 300c tales como pantalla convencional de computadora, teclado, ratón, e impresora, así como otras interfaces tales como una interfaz de red, una interfaz de control para un sintetizador de péptidos e interfaces de software tales como una interfaz de base de datos. El sistema de computadora 300 acepta la entrada de usuario desde un dispositivo de entrada 304 tal como un teclado, archivo de datos de entrada, o interfaz de red, y proporciona una salida a un dispositivo de salida 308 tal como una impresora, interfaz de red, o dispositivo de almacenamiento de datos. El dispositivo de entrada 304 recibe una entrada que comprende una secuencia de aminoácidos para el péptido modificado (v.g. mutante) así como valores de pH y temperatura apropiados para un entorno para el cual se determina la velocidad de agregación del polipéptido. Puede suministrarse también un factor de corrección glicina/prolina, tal como un peso para un factor de distorsión estructural que interfiera con la formación de lámina β o agregación inter-molecular. El dispositivo de salida 308 proporciona una información comparativa de velocidad de agregación tal como un log (base 10 o natural) de la velocidad de agregación, por ejemplo, una ratio de semi-tiempos para agregación de un mutante en comparación con un polipéptido de tipo salvaje.

El sistema de computadora 300 está acoplado a un almacén de datos 302 que almacena datos de hidrofobicidad, datos de propensión a lámina β (sea como datos de propensión *per se* o en términos de energía libre) y datos de carga. Estos datos se almacenan para cada aminoácido (residuo) y preferiblemente se almacenan una pluralidad de conjuntos de cada uno de estos tipos de datos correspondientes a valores diferentes de pH y temperatura. El sistema de computadora, en el ejemplo ilustrado, se muestra en interfaz con un calculador de propensión a hélice α 306. Éste puede ser una máquina separada, por ejemplo, acoplada al sistema de computadora 300 a través de una red, o puede comprender un programa separado que corre en el sistema de la computadora de uso general 300, o en otros ejemplos un código de propensión a hélice α puede almacenarse con la memoria de programa 300b y operar de modo unitario con el código de determinación de la velocidad de agregación descrito más adelante. Sin embargo, cualquiera que sea el método que se emplee, el calculador de propensión a hélice α recibe datos de secuencia, indirectamente del dispositivo de entrada del usuario, y proporciona datos de propensión a hélice α como retorno. Estos datos y los datos en el almacén de datos 302 pueden determinarse sobre una base de aminoácido por aminoácido, o pueden determinarse teniendo en cuenta el contexto de secuencia, por ejemplo, utilizando una ventana en la secuencia para modificar los valores de datos dependiendo de los aminoácidos vecinos.

Como se ilustra, el sistema de computadora 300 puede proporcionar también una salida de control de datos 310 a un sintetizador automático de péptidos 312. Los datos de control comprenderán generalmente una secuencia de aminoácidos de un polipéptido. De este modo, el sistema de computadora 300 puede programarse para comparar automáticamente las propiedades de cierto número de polipéptidos modificados (v.g. mutantes) y seleccionar uno o más de los que se predice tendrán propiedades favorables para síntesis automática. Un ejemplo de un sintetizador de péptidos automático de este tipo podría ser un ABI 433A Peptide Synthesizer (Applied Biosystems).

Haciendo seguidamente referencia a la Figura 4, ésta muestra un procedimiento para determinar una velocidad comparativa de agregación a tenor de las líneas arriba descritas. La Figura 4 representa un diagrama de flujo de un ejemplo de código que corre en la memoria de programas 300b de la Figura 3.

En el paso S400, un usuario introduce una secuencia de aminoácidos, datos de pH y temperatura, opcionalmente con datos de los términos C y N para la secuencia. A continuación, en el paso S 402 el sistema de la computadora lee los datos de hidrofobicidad para la secuencia introducida a partir del almacén de datos y suma éstos a fin de proporcionar una estimación de hidrofobicidad para el péptido codificado por la secuencia. En los casos en que, como es muy preferible, están disponibles datos para una gama de valores de pH y temperatura, se recuperan los datos que corresponden más estrechamente al pH y la temperatura deseados. Seguidamente, como pasos S304 y S406, el procedimiento lee datos de carga y datos de propensión a lámina β del almacén de datos de manera similar, sumando los datos de carga a fin de proporcionar una estimación de carga para el polipéptido correspondiente a la secuencia introducida y, análogamente, sumando los datos de propensión a lámina β

(expresados normalmente en términos de energía libre). En el caso de la prolina, no está disponible ningún valor de propensión a lámina β y por tanto puede pasarse por alto un residuo prolina cuando se resumen estos valores, o puede emplearse un valor arbitrario de propensión a lámina β o uno correspondiente a otro aminoácido. Por ejemplo, si la propensión a lámina β se expresa en términos de energía libre, puede utilizarse un valor arbitrario de 1, o un valor correspondiente a otro aminoácido. Opcionalmente, los pasos S402 y S406 pueden emplear una "ventana" (por ejemplo de 3, 5, 7, o más aminoácidos) que podría incluir una corrección por el efecto de los residuos flanqueantes sobre las propiedades de un aminoácido particular (es decir, tener en cuenta vecinos próximos dentro de una secuencia de aminoácidos), en lugar de considerar cada aminoácido de la secuencia individualmente.

El paso S408 del procedimiento proporciona la secuencia de entrada a una calculadora de propensión a hélice α , con los datos de pH y temperatura, y, en caso de estar disponibles, con los datos de los términos C y N. Una calculadora de propensión a hélice α S408a opera con estos datos y devuelve los datos nuevamente al procedimiento en el paso S410, comprendiendo los datos devueltos un valor de propensión a hélice α para la secuencia completa. Un código adecuado del programa para un calculador de propensión a hélice α S408a comprende el código AGADIR disponible de <http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html>, el código GOR4 disponible de http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html y otros códigos arriba descritos. La persona experta reconocerá que, en caso deseado, este código o un código de nuevo diseño derivado de datos accesibles al público (descritos en la literatura científica) o datos experimentales adicionales pueden incorporarse dentro del código que implementa el procedimiento de la Figura 4 en lugar de ser implementados como un procedimiento separado.

En el paso S412, el procedimiento determina luego la velocidad comparativa de agregación del polipéptido definido por la secuencia de aminoácidos introducida en comparación con un polipéptido de referencia, utilizando la Ecuación 1 anterior. Puede verse por la Ecuación 1 que una determinación de velocidad comparativa de agregación requiere una diferencia en hidrofobicidad, propensión a estructura secundaria y carga, y valores para hidrofobicidad, propensión a estructura secundaria y carga para los polipéptidos de referencia pueden determinarse, o bien por repetición de los pasos S400 a S410 para el polipéptido de referencia o por lectura de valores almacenados de estos parámetros del almacén de datos 302, o de cualquier otra manera convencional. Si se desea, en el paso S412, los parámetros o factores de escalación en la Ecuación 1 que opera sobre las diferencias en hidrofobicidad, propensión a estructura y carga pueden seleccionarse de conjuntos de parámetros adecuados para (paso S414) en respuesta a datos de entrada tales como datos de tipo de polipéptido. Por ejemplo, un polipéptido en espiral completamente aleatorio puede utilizar parámetros diferentes que un polipéptido parcialmente desplegado o estructurado. Asimismo, un polipéptido rico en un tipo específico de residuo, tal como aminoácidos aromáticos o cargados, puede requerir parámetros diferentes.

Después de la determinación de la velocidad comparativa de agregación, puede aplicarse una corrección opcional en el paso S416 para los residuos prolina y/o glicina a fin de tener en cuenta preferencias adicionales de conformación o estructura que pueden impedir la formación de lámina β inter-molecular o estructuras agregadas por un polipéptido dado y, posteriormente, en el paso S418, el sistema emite como salida el resultado del cálculo de la velocidad comparativa de agregación. Éste puede comprender un valor simple positivo o negativo que indica si la velocidad de agregación del polipéptido modificado (v.g. mutante) es mayor o menor que la del polipéptido de referencia, pero preferiblemente éste comprende datos cuantitativos relativos a las velocidades de agregación comparativas tales como una ratio logarítmica de las velocidades de agregación.

La Figura 5 muestra un diagrama de flujo de una implementación ventajosa del procedimiento de la Figura 4. En particular, la Figura 5 muestra un método de cribado de polipéptidos modificados (v.g. mutaciones) a fin de seleccionar candidatos con propiedades prometedoras para investigación ulterior y, opcionalmente, síntesis. Así, en el paso S500 se introduce una secuencia de aminoácidos para un polipéptido de referencia junto con datos que identifican una o más posiciones modificadas (v.g. mutantes). Opcionalmente, el procedimiento puede permitir también una modificación o una serie de modificaciones a especificar, por ejemplo en términos de un conjunto o selección pre-determinado(a) de aminoácidos.

Después de la inicialización, en el paso S502, el procedimiento genera una secuencia modificada representativa de una de las posibles permutaciones definidas por los datos de entrada y a continuación, en el paso S504, determina una velocidad comparativa de agregación para el polipéptido modificado en comparación con el polipéptido de referencia, utilizando por ejemplo el procedimiento de la Figura 4. Luego, en el paso S506, el procedimiento comprueba si existen cualesquiera permutaciones adicionales para las cuales realizar el cálculo, y en caso afirmativo vuelve al paso S502 hasta que se ha generado un conjunto completo de permutaciones posibles. Posteriormente, en el paso S508 se emite como salida el conjunto de datos de velocidad de agregación comparativos para cada polipéptido modificado (en comparación con la proteína de referencia), por ejemplo como una lista automática, gráfico, o de cualquier otra manera conveniente. Estos datos pueden utilizarse luego, por ejemplo para identificar candidatos para síntesis y/o para comparación con otros datos tales como inmunogenicidad/antigenicidad. En particular, uno o más de los polipéptidos modificados 'óptimos', por ejemplo mutantes con una velocidad de agregación particularmente alta o baja, pueden recogerse y emitirse como salida los datos de secuencia para estos polipéptidos modificados a un sintetizador automático de péptidos tal como el sintetizador 312 de la Figura 3 a fin de producir automáticamente las proteínas mutantes para, supongamos, investigación ulterior.

Ejemplo 6: Propiedades intrínsecas para la formación de amiloide de aminoácidos y secuencias de polipéptidos: Identificación de las regiones sensibles para agregación

5 A continuación se presenta una fórmula para medir la propensión intrínseca a la agregación de amiloide de un polipéptido. A partir de esta fórmula, se identifican los residuos que promueven la formación de amiloide, se comparan las propensiones a amiloide de cierto número de secuencias, y se identifican las regiones de la secuencia que son particularmente importantes para promover la agregación.

Definición de las Propensiones a la Agregación

10 Los factores intrínsecos del algoritmo arriba descrito se utilizaron para definir una nueva ecuación que especifica P_{agg} como la propensión intrínseca a la agregación de una secuencia. El peso para cada factor intrínseco y extrínseco se determinó simultáneamente utilizando técnicas de regresión sobre un conjunto de datos de 83 secuencias, como se expone a continuación en la Tabla 4. Los pesos para los factores intrínsecos se tomaron del algoritmo resultante y se utilizaron para definir una ecuación de P_{agg} ulterior (Ecuación 2).

Tabla 4

Secuencia	Mutantes	pH	Fuerza iónica	[Péptido]	Referencias
AcP	59	5,5	43 mM	0,04 mM	[3,4, 77]
AP40	2	7,4	150 mM	0,25 mM	[66]
AP40	ninguno	7,4	81 mM	0,03 mM	[80]
AP42	ninguno	7,4	81 mM	0,01 mM	[80]
ABri	ninguno	9,0	89 mM	1,31 mM	[102]
Péptido AChE 586 - 599	ninguno	7,0	7,7 mM	0,20 mM	[109]
Amilina 1-37	2	7,2	1,1 mM	20 mM	[67]
Amilina 1-37	ninguno	7,3	1,4 mM	0,14 mM	[89]
Amilina 8-37	ninguno	7,3	1,4 mM	0,14 mM	[89]
Precursor de IAPP	ninguno	5,0	0,1 mM	0,001 mM	[88]
LRR	1	7,8	3,3 mM	0,39 mM	[64]
Péptido PrP 106 - 126	3	5,0	1,2 mM	0,33 mM	[65]
TTR	3	4,4	130 mM	0,014 mM	[68]

15 Las referencias son como sigue: [3]: Chiti, F., et al., Kinetic partitioning of protein folding and aggregation. Nat. Struct Biol, 2002a. 9(2): p. 137-43; [4]: Chiti, F., et al., Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. Proc Natl Acad Sci USA, 2002b. 99 Suppl 4: p. 16419-26; [64]: Symmons, M.F., et al., X-ray diffraction and far-UV CD studies of filaments formed by a leucine-rich repeat peptide: structural similarity to the amyloid fibrils of prions and Alzheimer's disease beta-protein. FEBS Lett, 1997, 412(2): p.397-403; [65]: Salmona, M., et al., Molecular determinants of the physicochemical properties of a critical prion protein region comprising residues 106-126. Biochem J, 1999 342 (Pt 1): p. 207-14; [66]: Miravalle, L., et al., Substitutions at codon 22 of Alzheimer's abeta peptide induce diverse conformational changes and apoptotic effects in human cerebral endothelial cells. J Biol Chem, 2000. 275(35): p. 27110-6; [67]: Azriel, R. and E. Gazit, Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. J Biol Chem, 2001, 276(36): p.34156-61; [68]: Hammarstrom, P., et al., Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity. Proc Natl Acad Sci U S A, 2002. 99 Suppl 4: p. 16427-32; [80]: Fezoui, Y. and D.B. Teplow, Kinetic studies of amyloid beta-protein fibril assembly. Differential effects of alpha-helix stabilization. J Biol Chem, 2002, 277(40): p. 36948-54; [88]: Kaye, R., et al., Conformational transitions of islet amyloid polypeptide (IAPP) in amyloid formation in vitro. J Mol Biol, 1999. 287(4): p. 781-96; [89]: Goldsbury, C., et al., Amyloid fibril formation from full-length and fragments of amylin. J Struct Biol, 2000. 130(2-3): p. 352-62; [102]: El-Agnaf, O.M., et al., Effect of the disulfide bridge and the C-terminal extension on the oligomerization of the amyloid peptide ABri implicated in familial British dementia. Biochemistry, 2001. 40(12): p. 3449-57; [109]: Cottingham, M.G., M.S. Hollinshead, and D.J. Vaux, Amyloid fibril formation by a synthetic peptide from a region of human acetylcholinesterase that is homologous to the Alzheimer's amyloid-beta peptide. Biochemistry. 2002.41(46): p. 13539-47.

35 *Predicción de las Propensiones a Agregación*

40 Las propensiones a agregación de varios péptidos y pequeñas proteínas se calcularon a pH neutro. Se incluyeron los péptidos β de Alzheimer (Ab40 & Ab42), ABri, el péptido acetilcolinesterasa (586-599) (péptido AchE), acilfosfatasa (AcP), el péptido amilina (1-37), el dominio SH3 de α -espectrina, el dominio SH3 de fosfatidilinositol-3-quinasa (PI3 SH3), α -sinucleína, β 2-microglobulina (β 2m), calcitonina, el dominio N-terminal de la proteína procariota HypF (HypF), insulina, repeticiones ricas en leucina (LRR), la proteína priónica (PrP), el péptido PrP (106-126), y transtiretina (TTR).

Propensiones a la Agregación de los Aminoácidos

Se utilizó la Ecuación 2 para calcular el valor P_{agg} para aminoácidos individuales. I^{pat} no se incluye en dicho cálculo, dado que el término patrón para un residuo depende de la posición del residuo en la secuencia.

Perfiles de Propensión a Agregación

5 Los valores P_{agg} de los aminoácidos individuales se calcularon a lo largo de la longitud de una secuencia. Se utilizó la Eq (1) para calcular un valor P_{agg} por residuo, dando el peso total de I^{pat} para cualquier residuo dentro de una secuencia hidrófoba-hidrófila de 5 residuos secuencialmente alternante. Se alisó luego el perfil de P_{agg} promediando los valores resultantes a lo largo de una ventana de deslizamiento de 5 residuos y se representó gráficamente de acuerdo con el número de residuos centrales. Se creó un perfil de muestra P_{agg} para PrP.

10 *Detección de las Regiones Sensibles*

Las regiones de la secuencia que son particularmente propensas a cambiar las velocidades de agregación de amiloide por mutaciones simples se identificaron como sigue. Se calcularon los perfiles P_{agg} para la secuencia wt y para todos los mutantes simples posibles (20 aminoácidos posibles para cada residuo). Se consideraron los valores de estos perfiles en cada residuo, y los valores P_{agg} máximo y mínimo posibles en dicho residuo se representan gráficamente junto con el valor de tipo salvaje sin alisar. Se calcularon los perfiles de región sensible para AcP (pH 5), Afi42 (pH 5), y dos dominios SH3 (pH 2).

Resultados

Definición de las propensiones a agregación intrínsecas:

20 La propensión intrínseca a formar agregados de amiloide, P_{agg} , se define considerando únicamente los factores intrínsecos (I):

$$P_{agg} = -0.08I^{hydr} + 0.96I^{pat} - 0.07I^{\alpha} + 0.08I^{\beta} - 0.47I^{ch} \quad (\text{Ecuación 2})$$

I^{hydr} representa la hidrofobicidad de la secuencia [Roseman, M. A. (1988). "Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds." J Mol Biol 200(3): 513-22; and Cowan, R. and R. G. Whittaker (1990) "Hydrophobicity indices for amino acid residues as determined by high-performance liquid chromatography." Pept Res 3(2): 75-80]; I^{pat} indica la modelización hidrófoba-hidrófila [Broome, B. M. and M. H. Hecht (2000) "Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis." J Mol Biol 296(4): 961-8]; I^{α} mide la propensión a hélice α [Munoz, V. and L. Serrano (1994) "Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales." Proteins 20(4): 301-11,]; I^{β} es la propensión a lámina β [Street, A. G. and S. L. Mayo (1999) "Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone." Proc Natl Acad Sci U S A 96(16): 9074-6,]; e I^{ch} es el valor absoluto de la carga neta de la secuencia. Dado que el pH influye en tres de estos términos (I^{hydr} , I^{pat} , e I^{ch}), debería especificarse preferiblemente el mismo para resolver la Ecuación (2).

35 La Tabla 5 siguiente, da las escalas de hidrofobicidad, propensión a lámina β y carga para los 20 aminoácidos naturales.

Tabla 5

Escalas de hidrofobicidad, propensión a lámina β y carga para los 20 aminoácidos naturales			
Residuo de aminoácido	Hidrofobicidad (kcal mol ⁻¹) ^a	Propensión a lámina β ^b	Carga ^c
Arg (R)	3,95	0,35	+1
Lys (K)	2,77	0,34	+1
Asp (D)	3,81	0,72	-1
Glu (E)	2,91	0,35	-1
Asn (N)	1,91	0-40	0
Gln (Q)	1,30	0,34	0
His (H)	0,64 (2,87) d	0,37	0(+1) d
Ser (S)	1,24	0,30	0
Thr (T)	1,00	0,06	0

Escala de hidrofobicidad, propensión a lámina β y carga para los 20 aminoácidos naturales			
Residuo de aminoácido	Hidrofobicidad (kcal mol ⁻¹) ^a	Propensión a lámina β ^b	Carga ^c
Tyr (Y)	-1,47	0,11	0
Gly (G)	0,00	0,60	0
Pro (P)	-0,99	n.d.	0
Cys (C)	-0,25	0,25	0
Ala (A)	-0,39	0,47	0
Trp (W)	-2,13	0,24	0
Met (M)	-0,96	0,26	0
Phe (F)	-2,27	0,13	0
Val (V)	-1,30	0,13	0
Ile (I)	-1,82	0,10	0
Leu (L)	-1,82	0,32	0
<p>^a hidrofobicidad, valores de los 20 residuos de aminoácidos a pH neutro basados en los coeficientes de reparto de agua a octanol. Los datos son de la columna 6 de la Tabla 4.8 en la ref. 30.</p> <p>^b propensiones a lámina β de los 20 residuos de aminoácidos normalizados desde 0 (propensión alta a lámina β a 1 (propensión baja a lámina β). Los datos son de la columna 4 de la Tabla 1 en la ref. 29.</p> <p>La propensión a lámina β de la prolina no se consigna debido a la dificultad en la determinación experimental de la misma. La propensión a lámina β de la glicina procede de cálculos teóricos:</p>			
^c Los valores de la carga se refieren a pH neutro.			
^d Los valores entre paréntesis corresponden a un pH menor que 6,0, cuando el residuo histidina está cargado positivamente			

La propensión intrínseca a la agregación Pass es un número sin dimensiones, que puede ser escalado de acuerdo con los factores de la ecuación anterior, y que puede seleccionarse convenientemente para dar valores P_{agg} entre -1 y +1 (correspondiendo -1 a agregación reducida y +1 a agregación incrementada).

Predicción de las Propensiones Intrínsecas a la Agregación:

- 5 Dado que la mayoría de los estudios de agregación de amiloide han estado diseñados hasta ahora para detectar la formación de fibrillas en lugar de medir con precisión las velocidades de agregación, las condiciones utilizadas variaban considerablemente en diferentes experimentos y es difícil evaluar a partir de la bibliografía las propensiones intrínsecas de diferentes secuencias a la agregación. La Ecuación 2 proporciona una separación natural entre los factores intrínsecos y extrínsecos que promueven la agregación de amiloide y hace posible por
- 10 tanto comparar las propensiones intrínsecas a la formación de amiloide de diferentes secuencias. Los inventores clasificaron varias secuencias de polipéptidos estudiadas intensamente de acuerdo con su propensión intrínseca a la agregación. La Tabla 6 presenta una lista de secuencias diferentes y sus propensiones a la agregación, calculadas a pH 3 y pH 7.

Tabla 6

Secuencia	pH 3	pH 7
AR40	-0,03	0,79
AP42	0,21	1,03
ABri	-2,01	0,03
Péptido AChE (586 - 599)	-1,42	0,21
AcP	-2,08	3,53
Amilina 1 - 37	-0,27	0,40

Secuencia	pH 3	pH 7
α -espectrina SH3	-2,57	2,49
PI3 SH3	0,34	0,98
α -sinucleína	-1,05	-1,39
β 2 microglobulina	-0,26	7,80
Calcitonina	0,53	1,48
HypF	0,71	6,99
Insulina	3,85	5,09
LRR	-0,55	0,28
PrP	-0,73	10,13
Péptido PrP (106 - 126)	0,87	1,54
TTR	-3,41	1,12

Es importante considerar el pH cuando se calcula la propensión a la agregación de amiloide, dado que el pH influye en los factores intrínsecos f^{hydro} , f^{pat} e f^{ch} . El conjunto de secuencias y datos proporciona resultados interesantes. En primer lugar, está claro que a pH bajo, la mayoría de las secuencias tienen realmente una baja propensión a agregarse. Tanto la propensión intrínseca a la agregación como la estabilidad de las proteínas plegadas decrecen con el pH. Por tanto, pueden obtenerse fibrillas de amiloide a partir de proteínas plegadas por disminución del pH, aun cuando su propensión intrínseca a agregarse se reduce. A pH neutro, PrP, b2m, HypF, e insulina tienen las propensiones más altas a la agregación de amiloide, y se sabe que forman fibrillas con relativa facilidad. AcP tiene también una propensión muy alta a la agregación intrínseca, y de hecho, esta proteína jugaba un papel importante en el establecimiento del principio de que la formación de fibrillas de amiloide es una propiedad genérica de los polímeros de aminoácidos. Como era de esperar, A β 42 tiene una propensión mayor a la agregación que A β 40 tanto a pH 3 como a pH 7.

Propensiones intrínsecas de agregación de los aminoácidos individuales:

La propensión a la agregación de amiloide de cada aminoácido puede calcularse por la Ecuación 2. La escala resultante es útil en el diseño de mutaciones para aumentar o disminuir la agregación de amiloide. La escala a pH neutro se muestra en la Tabla 7 siguiente, enumerándose los aminoácidos en orden decreciente de propensión.

Tabla 7

Trp	0,23
Leu	0,21
Phe	0,20
Gly	0,17
Ile	0,13
Tyr	0,13
Met	0,13
Ala	0,12
Val	0,12
Cys	0,11 (-0,57 si pH > 8,3)
His	0,06 (-0,61 si pH < 6,0)
Ser	-0,01
Gln	-0,03
Asn	-0,03
Pro	-0,10

Thr	-0,12
Lys	-0,62
Glu	-0,64 (0,03 si pH < 4,3)
Asp	-0,63 (0,05 si pH < 3,7)
Arg	-0,72

A pH neutro, triptófano, leucina, fenilalanina y glicina tienen la propensión máxima a amiloide, mientras que ácido aspártico, lisina, ácido glutámico y arginina tienen la mínima. Es interesante que la presente escala de los inventores asigna a la histidina una propensión mucho menor a la agregación de amiloide que los otros residuos aromáticos, especialmente a valores de pH inferiores.

5 Perfiles de Propensión a la Agregación

Puede utilizarse la Ecuación 2 para calcular la suma de los factores intrínsecos (es decir, hidrofobicidad, patrones hidrófobos, propensiones a estructura secundaria, y carga) individualmente para cada residuo en una secuencia polipeptídica. Esta operación da como resultado un 'perfil de propensión a la agregación', que ilustra de qué modo las diferentes regiones de la secuencia de un polipéptido tienen propensiones intrínsecas significativamente diferentes a la agregación.

Se presenta en primer lugar el perfil de propensión para PrP. La Figura 7 muestra el perfil de propensión a PrP a la agregación de amiloide. El perfil de propensión a la agregación amiloide, a pH 7, se muestra junto con la secuencia de PrP, calculada para cada residuo como si fuera su propia secuencia a partir de Eq (2) y promediada a lo largo de una ventana de deslizamiento de 5 residuos.

Los residuos 55-90 muestran una propensión relativamente alta hacia la agregación. Esto es interesante, dado que se sabe que repeticiones adicionales dentro de esta región están ligadas a enfermedades priónicas. La región de los aminoácidos 106-126 tiene una propensión alta a la agregación, y se sabe que forma fibrillas *in vitro*. Sin embargo, la región que ostenta la característica más interesante de este perfil va desde el residuo 180 al 190. Aunque se sabe que algunas mutaciones en esta región son patógenas, sería interesante ver si diferentes variaciones genéticas en esta región que disminuyen el valor máximo ayudan a proteger a sus portadores contra las enfermedades priónicas. Las mutaciones patógenas conocidas corresponden a algunas de las características más interesantes del perfil, agrupándose alrededor de los residuos 105, 180, y 200.

Identificación de las Regiones Sensibles

La Figura 8 muestra el perfil de agregación y las regiones sensibles de AcP. La Figura 8a muestra el perfil de agregación de amiloide a pH 5,5 representado para AcP. El perfil de tipo salvaje es la curva 800, y los valores de propensión máximo y mínimo posibles para cada residuo se representan gráficamente en las curvas 802, todas ellas sin alisado. La Figura 8b muestra el cambio de velocidad como resultado de diversas mutaciones simples en AcP. Los valores positivos indican un aumento en la velocidad de agregación. Los experimentos se realizaron a pH 5,5.

Una de las observaciones más interesantes en los estudios cinéticos recientes de amiloides es que la secuencia de AcP parece contener regiones "sensibles". Las mutaciones simples de aminoácidos en estas regiones pueden cambiar notablemente las velocidades de agregación. Estas regiones de la secuencia parecen influir particularmente en la velocidad de formación de amiloide (véase la Figura 8b). Un análisis de los perfiles de propensión ofrece una nueva comprensión acerca del origen de estas regiones sensibles. Para cualquier posición dada a lo largo de la secuencia, los inventores calcularon los valores de propensión para todos los mutantes posibles de un solo punto, obteniendo así los valores de propensión máximo y mínimo posibles. Por repetición de este cálculo para cada posición a lo largo de la secuencia pueden construirse así dos nuevos perfiles de propensiones máxima y mínima respectivamente. Estos dos perfiles se comparan con el perfil para la secuencia de tipo salvaje a fin de presentar la gama de aumentos y disminuciones disponibles en cada posición.

Los inventores aplicaron este tipo de análisis a tres secuencias de polipéptidos para las cuales se dispone de datos de mutación extensos sobre las velocidades de agregación: AcP, A β 42, y dos dominios de SH3. El pH de cada perfil se seleccionó para permitir la comparación óptima con los datos experimentales disponibles.

AcP: El perfil de propensión a la agregación para AcP se calculó a pH 5,5 para la secuencia de tipo salvaje (Figura 8A, curva 800) y se representó gráficamente con los perfiles de propensiones máxima y mínima (Figura 8A, curvas 802). Los resultados se comparan con un estudio cinético experimental de 55 mutantes de mono-sustitución de AcP, realizado también a pH 5,5 (Figura 8B) [Chiti 2002a, b]. Es interesante que las dos regiones (residuos 16-31 y 87-98 [Chiti 2002a]) de AcP que se han identificado como sensibles tienen regiones de alta propensión en la secuencia de tipo salvaje. Y, lo que es todavía más interesante, pueden observarse aumentos y disminuciones en las velocidades de agregación de la mayoría de las mutaciones de AcP como áreas de alto cambio potencial entre el perfil de agregación del tipo salvaje y los perfiles potenciales máximo o mínimo. La única excepción importante es el aumento

para el mutante A30G. Sin embargo, este mutante está localizado en el centro de una hélice α . Dado que los experimentos se realizaron en 25% de TFE, que estabiliza las hélices, es probable que esta discrepancia sea debida a una estabilidad residual que no está considerada en la fórmula de los inventores, que está diseñada para tratar sólo con polipéptidos desestabilizados.

- 5 La Figura 9 muestra el perfil de agregación de A β 42 y las regiones sensibles. La Figura 9a muestra el perfil de agregación de amiloide a pH 5 representado para A β 42. El perfil de tipo salvaje se representa en la curva 900 y los valores de propensión máximo y mínimo posibles para cada residuo se representan en las curvas 902, todas ellas sin alisado. La Figura 9b muestra la frecuencia de una mutación aleatoria en cada residuo para 36 secuencias de agregación mínima como fueron determinadas por Wurth *et al.* [Wurth, C., N. K. Guimard y M.Hecht. (2002)
10 "Mutations that reduce aggregation of the Alzheimer's Abeta42 peptide: an unbiased search for the sequence determinants of Abeta amyloidogenesis." J Mol Biol 319(5): 1279-90].

A β 42: El perfil de propensión a la agregación para A β 42 se calculó para la secuencia de tipo salvaje (Figura 9a, curva 900) y se representó gráficamente con los perfiles de propensiones máxima y mínima (Figura 9a, curva 902). La primera característica que se observa en el perfil de tipo salvaje de A β 42 es la alta propensión alrededor de los
15 residuos 32-42. Un trabajo reciente realizado por Petkova *et al.* proporciona un modelo estructural de las fibrillas A β al nivel de residuos [Petkova, A. T., Y. Ishii, et al. (2002). "A structural model for Alzheimer's beta -amyloid fibrils based on experimental constraints from solid state NMR." Proc Natl Acad Sci U S A 99(26): 16742-7]. Sus resultados ponen los residuos 28-42 directamente en el núcleo de la lámina β de la fibrilla de amiloide. Adicionalmente, los
20 residuos 15-24 son capaces de formar fibrillas *in vitro*, de nuevo una región de alta propensión a la agregación de tipo salvaje. De hecho, los residuos 22, 23, y 28 son los únicos residuos de 17-42 que no están maximizados para propensión a amiloide. Esto es particularmente interesante, dado que han sido identificadas varias mutaciones patógenas en el residuo 22. Para comparar ulteriormente los resultados de este análisis con el trabajo experimental, los inventores tuvieron en cuenta un reciente estudio *in vivo* sobre A β 42 realizado por Wurth *et al.* [*ibid.*]. Se enlazó
25 A β 42 a la proteína fluorescente verde (GFP), se sometió a mutación aleatoria, y se expresó en colonias de células. La fluorescencia de GFP se extingue después de la agregación de amiloide de A β 42. Se seleccionaron luego las 36 colonias más fluorescentes (es decir las de agregación mínima) para secuenciación de A β 42 [Wurth 2002]. La Figura 9B presenta un histograma del número de veces que cada residuo sufría mutación en los 36 mutantes de agregación mínima. Dado que el pH *in vivo* no puede conocerse con precisión, se calcularon los perfiles de propensión a la agregación a lo largo de un intervalo de pH, desde pH 2 a pH 9. La Figura 9a presenta el resultado
30 del cálculo a pH 5, pero los perfiles son sólo levemente dependientes de los valores de pH en el intervalo de 4,5 a pH 9. Puede verse claramente que las regiones que se calcula tienen el potencial máximo para reducir la propensión a la agregación son de hecho aquellas regiones en las que la mayoría de las mutaciones se producían aleatoriamente en el estudio de Wurth *et al.* Considerando las complicaciones de un sistema *in vivo*, así como los cambios en el valor P_{agg} debidos a la GFP ligada, estos son resultados excelentes.

35 La Figura 10 muestra los perfiles de agregación y regiones sensibles de PI3 SH3 y α -espectrina SH3. Los perfiles de agregación de amiloide de SH3 corresponden a pH 2. Los perfiles de tipo salvaje (wt) se representan gráficamente en las curvas 1000 y 1004, mientras que los valores de propensión máximo y mínimo posibles para cada residuo se representan en las curvas 1002 y 1006, todas ellas sin alisado. La Figura 10a presenta los perfiles PI3, y la Figura 10b los perfiles de α -espectrina.

40 **SH3:** Se calcularon los perfiles de propensión a la agregación para PI3-SH3 (Figura 10A) y α -espectrina SH3 (Figura 4B) para las secuencias de tipo salvaje (curvas 1000, 1004) y se compararon con los perfiles de propensión máxima y mínima (curvas 1002, 1006). El dominio SH3 de PI3 ha sido estudiado como una proteína globular formadora de amiloide no patógena. El estado nativo de SH3 es muy estable y la proteína tiene que desnaturalizarse antes que
45 puedan formarse fibrillas de amiloide. Dado que las condiciones utilizadas son a menudo fuertemente ácidas para contrarrestar la estabilidad, los perfiles se determinaron a pH 2. Mientras que la mayoría de los dominios SH3, con inclusión de α -espectrina SH3, tienen un consenso excelente en los residuos 25 y 26, típicamente Lys 25-Lys 26, PI3-SH3 tiene los residuos atípicos His 25-Leu 26; α -espectrina SH3, que contiene los residuos Lys 25-Lys 26 no forma fibrillas detectables después de una incubación de 30 días, mientras que PI3 SH3 sí lo hace. Las sustituciones de estos dos residuos lisina en PI3 SH3 hacen que su formación de fibrillas sea asimismo indetectable, mientras que
50 la sustitución de los residuos 20-26 de PI3-SH3 en α -espectrina SH3 hace posible la formación de fibrillas en esta escala de tiempo. Estos experimentos se llevaron a cabo todos ellos en condiciones fuertemente ácidas. En la comparación de los perfiles de agregación de las dos secuencias de tipo salvaje a pH bajo (Figura 10), los valores P_{agg} de los residuos 20-26 son significativamente menores para α -espectrina SH3 que para PI3-SH3, correspondiéndose bien con las observaciones experimentales.

55 Se presenta en esta memoria un método para calcular las propensiones intrínsecas a la agregación de amiloide para varias secuencias de interés en la investigación del amiloide. Se presenta también una escala de propensión a amiloide para aminoácidos individuales, que puede utilizarse en el diseño de mutantes con propensiones a la agregación controladas. Adicionalmente, se han calculado los perfiles de propensión para examinar las características amiloidogénicas de 5 polipéptidos, PrP, AcP, A β 42, PI3-SH3, y α -espectrina SH3. Estos perfiles
60 ofrecen una nueva comprensión de las observaciones experimentales acerca de estas secuencias.

Se han empleado las abreviaturas siguientes: A β = péptido β de Alzheimer; AChE = acetilcolinesterasa; AcP = acetilfosfatasa; HypF = dominio N-terminal de la proteína procariota HypF; LRR = repeticiones ricas en leucina; PI3 = fosfatidilinositol-3-quinasa; PrP = proteína priónica; TTR = transtiretina; GFP = proteína verde fluorescente.

Determinación de patrones - información adicional

- 5 Esta sección proporciona información adicional acerca de los patrones utilizables en los métodos de predicción basados en las Ecuaciones 2 (anterior) y 3 (siguiente).

Los patrones alternantes se encuentran entre los patrones de secuencia más comunes en la naturaleza y son capaces de determinar motivos estructurales secundarios en las proteínas de diseño. Los patrones de 5 residuos hidrófilo-hidrófobos alternantes, en los que los residuos con valores de hidrofobicidad $\leq -0,5$ en la escala de Roseman [Roseman, M.A., Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. J Mol Biol, 1988. 200(3): p. 513-22] se consideraron hidrófobos, y aquéllos que tenían valores $\geq 0,5$, hidrófilos. Se seleccionaron patrones de 5 residuos debido a que parece ser el número mínimo de residuos alternantes que pueden diferenciar entre patrones promotores de lámina β ($\bullet\Delta\bullet\Delta\bullet$) y patrones promotores de hélice α ($\bullet\Delta\bullet\Delta\Delta$). Una forma de representar estos patrones es añadir +1 al término del patrón por cada secuencia alternante de 5 residuos encontrada en una secuencia. Esta representación da como resultado un coeficiente de correlación de 0,47 cuando se utiliza sola para predecir las velocidades absolutas de mutantes de región sensible de AcP. Esto puede refinarse mediante 1) adición de contribuciones de menor peso para los patrones alternantes de 4 y 3 residuos, 2) adición de valores negativos de diversos pesos para un patrón de 5 residuos que coincide con el patrón promotor de hélice α , y 3) adición de términos contributivos para residuos hidrófilos consecutivos o residuos hidrófobos consecutivos. Sin embargo, ninguno de los ajustes anteriores parece proporcionar una mejora significativa la representación simple de un valor +1 por cada patrón alternante de 5 residuos encontrado en la secuencia, al menos para la serie de datos estudiada.

Los aproximadamente 20 aminoácidos que pueden clasificarse como hidrófilos (residuo cargado o residuo polar, por ejemplo serina o cistina) o hidrófobos (no polares) de acuerdo con la definición anterior pueden emplearse los términos (y la clasificación) no polar y polar, aunque "hidrófilo" puede incluir cualquiera o ambos de "polar" y "cargado".

Alternativamente, puede emplearse una clasificación como se muestra a continuación:

hidrófobos: ala, val, phe, ile, leu, met, tyr, trp (algunos autores incluyen tyr y trp como polares, pero atendiendo a su carácter general son realmente hidrófobos)

30 **cargados:** asp, glu, lys, arg, his (algunos autores sitúan his como polar)

polares: ser, thr, cys, gln, asn.

glicina: puede ser hidrófoba o puede clasificarse como una glicina independiente que se considera residuo neutro.

35 Se ha determinado experimentalmente que ciertos patrones de aminoácidos, en particular patrones de aminoácidos hidrófilos ("P")/hidrófobos ("NP") dan como resultado una propensión incrementada a agregarse. Más particularmente, los patrones alternantes dan lugar a una propensión incrementada a la agregación, en particular los patrones alternantes que tienen una longitud de 5 o más aminoácidos (aunque algunas secuencias de 3 o más pueden exhibir un pequeña efecto). Así, por ejemplo, NP P NP P NP y P NP P NP P son ejemplos de patrones alternantes de longitud 5 que dan lugar a propensión incrementada a la agregación. Otros patrones pueden inhibir la agregación, por ejemplo una cadena de aminoácidos hidrófilos, o una cadena de algunos aminoácidos particulares tales como prolina.

45 Los efectos de estos patrones se tienen en cuenta en la Ecuación 2 anterior y la Ecuación 3 siguiente en el término f^{pat} , dándose en una realización a f^{pat} un valor de +1 por cada patrón alternante encontrado en la secuencia. Sin embargo, se apreciará que el incremento dado a f^{pat} para cada patrón identificado es esencialmente arbitrario, que es escalado por su factor multiplicador en la ecuación. Los expertos reconocerán que f^{pat} puede ajustarse por un primer valor para un primer patrón y un segundo valor para un segundo patrón, por ejemplo +1 para una cadena alternante de aminoácidos de longitud 5, y +2 para, por ejemplo, una cadena alternante de aminoácidos de longitud 9. Opcionalmente f^{pat} puede ajustarse por un valor negativo, por ejemplo -1, para un patrón inhibidor de la agregación. Se reconocerá una vez más que aunque sólo se ha incluido un término f^{pat} en la Ecuación 2 anterior, puede incluirse más de un término f^{pat} , cada uno con un factor multiplicador separado. (Es interesante que secuencias alternantes como las arriba mencionadas que tienen propensión a agregarse parecen no estar bien representadas en la naturaleza, debido quizás a que son desfavorables y han sido eliminadas durante la evolución.)

Sistema de computadora para implementar las realizaciones de la invención

55 El sistema de computadora 300 puede emplearse también para implementar la Ecuación 2 anterior, por ejemplo cuando se ejecuta un código de programa para implementar el diagrama de flujo de la Figura 6, y también la

Ecuación 3 siguiente, de acuerdo con el diagrama de flujo de la Figura 11 y el Ejemplo 7, que describe realizaciones de la presente invención.

La Figura 6 muestra un diagrama de flujo de un procedimiento para determinación de la propensión relativa intrínseca a la agregación como se ha descrito arriba, utilizando la Ecuación 2. Muchos de los pasos de la Figura 6 son similares a los descritos previamente con referencia a la Figura 4. Así, en el paso s600 se introducen una secuencia de aminoácidos, junto con datos de pH y temperatura, (para determinación de la carga y la propensión a hélice) y posteriormente, en los pasos s602, s604 y s606 el procedimiento determina, para cada aminoácido *l* de la secuencia, una hidrofobicidad, carga, y propensión a lámina β para el aminoácido. En el paso s608, los datos de secuencia se proporcionan también a un calculador de la propensión a hélice α, junto con los valores de pH y temperatura introducidos en el paso s600. En el paso s608a, un calculador de la propensión a hélice α determina un valor de propensión a hélice α para cada aminoácido en la secuencia, y en el paso s610, ésta es recibida por el código de programa para uso subsiguiente en el paso s614. La propensión a hélice α puede ser calculada por el procedimiento consultando simplemente un valor de propensión para cada aminoácido de la secuencia en una tabla de valores de propensión para cada uno de los 20 aminoácidos más o menos. (El método de la tabla de consulta puede utilizarse también con el procedimiento de la Figura 4). Alternativamente, puede utilizarse un programa de cálculo de la propensión a hélice α para determinar un valor de propensión a hélice α para cada aminoácido, como se describe anteriormente con referencia a la Figura 4. De modo preferible (pero no esencialmente), se proporcionan el pH y la temperatura al código de determinación de la propensión a hélice α.

En el paso s612 se determinan los datos de patrón para cada aminoácido de la secuencia. Como apreciarán las personas expertas, hay muchas maneras por las cuales puede hacerse esto, por ejemplo contando el número de alternancias polar/no polar hasta que éste alcanza 5 o más y asignando luego un valor de datos de patrón (*P^{pat}*) de, por ejemplo, +1 para cada aminoácido en la secuencia de alternancias (alternativamente, estos valores podrían normalizarse de tal modo que, supongamos, cada aminoácido en una secuencia alternante de longitud 5 tenga un valor de +0,2).

Opcionalmente, en el paso s616, puede seleccionarse un juego de parámetros para la Ecuación 2 basado en un tipo o grupo de proteínas al que se desea aplicar la ecuación, por ejemplo ACP y análogos.

En el paso s614, están disponibles para cada aminoácido de la secuencia todos los datos para la ecuación 2 y se aplica esta ecuación para determinar un valor de propensión relativa intrínseca a la agregación para cada aminoácido. Estos datos se emiten luego como salida en el paso s618, por ejemplo como un archivo de datos y/o como una matriz (impresa), como un gráfico, y/o de cualquier otro modo. Las figuras ... (Figuras 2-4 del documento) muestran ejemplos de emisiones de salida gráficas; si se desea, puede emplearse la media aritmética de un pequeño número de aminoácidos (por ejemplo, 2 a 10 aminoácidos) para alisar la curva. Opcionalmente, puede emplearse un procesamiento ulterior para identificar regiones sensibles como anteriormente. Así, hablando en términos generales, pueden sumarse las propensiones relativas intrínsecas a la agregación para cada aminoácido de una secuencia de tipo salvaje y puede determinarse luego en cada posición de la secuencia una suma separada para cada una de las 20 posibles mutaciones más o menos de un solo punto a fin de determinar aquellas posiciones en las cuales es potencialmente más probable que una mutación dé como resultado o contribuya a una velocidad de agregación aumentada. Si se desea, los resultados de un procedimiento de este tipo pueden, una vez más, representarse gráficamente (y/o de las otras maneras arriba mencionadas), como se muestra también en las Figuras 7 a 10).

Ejemplo 7: Predicción de las velocidades de agregación de amiloide absolutas de las cadenas de polipéptidos

Se describe aquí una ecuación que acumula y extiende el procedimiento del Ejemplo 6 y utiliza el conocimiento de la secuencia de aminoácidos y de las condiciones experimentales para reproducir, en algunas realizaciones con un coeficiente de correlación de 0,92, las velocidades de agregación *in vitro* de péptidos o proteínas desnaturalizadas. Estos resultados indican que la formación de agregados de amiloide puede racionalizarse en términos de principios físico-químicos simples. La técnica descrita es capaz de predecir, dentro de una extensa gama de condiciones experimentales posibles, las velocidades de agregación de numerosos péptidos desestructurados no homólogos y proteínas desplegadas o parcialmente desplegadas.

Los inventores introducen la fórmula fenomenológica siguiente para describir las velocidades de agregación absolutas de las cadenas de polipéptidos:

$$\ln(k) = \alpha_0 + \alpha_{hydr} I^{hydr} + \alpha_{mut} I^{mut} + \alpha_\alpha I^\alpha + \alpha_\beta I^\beta + \alpha_{ch} I^{ch} + \alpha_{pH} E^{pH} + \alpha_{ionic} E^{ionic} + \alpha_{conc} E^{conc}$$

Ecuación (3)

donde $\ln(k)$ es el logaritmo natural de la velocidad de agregación *k*, en s⁻¹. Los factores intrínsecos para la secuencia de aminoácidos se designan con *I*, mientras que los factores extrínsecos dependientes de las condiciones se designan con *E*. *I^{hydr}* representa la hidrofobicidad de la secuencia, tomada como la suma de las contribuciones

- hidrófobas de cada residuo de la escala de Roseman, utilizando la escala de Cowan a pH 3,4 para estimar los cambios con el pH [Roseman, M.A., Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J Mol Biol*, 1988. 200(3): p. 513-22; Cowan, R. y R.G. Whittaker, Hydrophobicity indices for amino acid residues as determined by high Performance liquid chromatography. *Pept Res*, 1990. 3(2): p. 75-80.]. f^{pat} corresponde a la existencia de patrones de residuos hidrófobo-hidrófilos alternantes; se asignó un factor +1 para cada patrón de 5 residuos hidrófobos e hidrófilos alternantes consecutivos en la secuencia [Broome, B.M. y M.H. Hecht, Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *J Mol Biol*, 2000. 296(4): p. 961-8.]. f^{β} mide la propensión global a hélice α de la secuencia, tomada como la suma de los logaritmos naturales de las propensiones intrínsecas a hélice α de cada residuo [Munoz, V. y L. Serrano, Intrinsic secondary structure propensities of the amino acids, using statistical phi-psi matrices: comparison with experimental scales. *Proteins*, 1994. 20(4): p. 301-11]. f^{β} es la propensión a lámina β , calculada como la suma del logaritmo natural de la propensión intrínseca a lámina β de cada residuo; se asignó un valor de 1% a prolina (destructora de la lámina β), aunque los resultados no se veían afectados cuando se consideraban valores de hasta 20%; se asignó un valor de 50% a glicina (no determinada) [Street, A.G. y S.L. Mayo, Intrinsic beta-sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc Natl Acad Sci USA*, 1999. 96(16): p. 9074-6]. f^{ch} es el valor absoluto de la carga neta de la secuencia. E^{pH} es el pH de la solución en la que ocurre la agregación y E^{ionic} es la fuerza iónica de la solución, dada en unidades milimolares. Finalmente, E_{conc} es la medida de la concentración de polipéptido, C en la solución, tomada en la forma de $\ln(C+1)$, con C en unidades milimolares.
- 20 La serie de datos utilizada para determinar y testar el algoritmo de predicción comprendía a la vez datos del estudio extenso de mutación sobre AcP y datos sobre otros sistemas disponibles en la bibliografía - véase la Tabla 8 siguiente.

Tabla 8

Secuencia	Mutantes	pH	Fuerza iónica	[Péptido]	Referencias
AcP	59	5,5	43 mM	0,04 mM	[27, 31, 50]
AP40	2	7,4	150 mM	0,25 mM	[53]
AP40	ninguno	7,4	81 mM	0,03 mM	[59]
Af342	ninguno	7,4	81 mM	0,01 mM	[59]
ABri	ninguno	9,0	89 mM	1,31 mM	[55]
Péptido AChE 586 - 599	ninguno	7,0	7,7 mM	0,20 mM	[58]
Amilin 1 - 37	2	7,2	1,1 mM	2,0 mM	[51]
Amilin 1 - 37	ninguno	7,3	1,4 mM	0,14 mM	[56]
Amilin 8 - 37	ninguno	7,3	1,4 mM	0,14 mM	[56]
Dominio HypF	ninguno	5,5	40 mM	0,08 mM	[62]
Precursor IAPP	ninguno	5,0	0,1 mM	0,001 mM	[57]
LRR	1	7,8	3,3 mM	0,39 mM	[54]
Péptido PrP 106 - 126	3	5,0	1,2 mM	0,33 mM	[52]
TTR	3	4,4	130 Mm	0,014 Mm	[43]

Secuencia	Mutantes	pH	Fuerza iónica	[Péptido]	Referencias
Las referencias para la tabla anterior son como sigue:					
					27. Chiti, F., et al., Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. Proc Natl Acad Sci U S A, 2002b. 99 Suppl 4: p. 16419-26.
					31. Chiti, F., et al., Kinetic partitioning of protein folding and aggregation. Nat Struct Biol, 2002a. 9(2): p. 137-43.
					43. Hammarstrom, P., et al., Sequence-dependent denaturation energetics: A major determinant in amyloid disease diversity. Proc Natl Acad Sci U S A, 2002. 99 Suppl 4: p. 16427-32.
					50. Calamai, M., et al., Relative Influence of Hydrophobicity and Net Charge in the Aggregation of two Homologous Proteins. Biochemistry, 2003. submitted.
					51. Azriel, R and E. Gazit, Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. J Biol Chem, 2001. 276(36): p. 34156-61.
					52. Salmona, M., et al., Molecular determinants of the physicochemical properties of a critical prion protein region comprising residues 106-126 Biochem J, 1999. 342 (Pt 1): p. 207-14.
					53. Miravalle, L., et al., Substitutions at codon 22 of Alzheimer's obeta peptide induce diverse conformational changes and apoptotic effects in human cerebral endothelial cells. J Biol Chem, 2000. 275(35): p. 27110-6.
					54. Symmons, M.F., et al., X-ray, diffraction and far-UV CD studies of filaments formed by a leucine-rich repeat peptide: structural similarity to the amyloid fibrils of prions and Alzheimer's disease beta-protein. FEBS Lett. 1997. 412(2); p. 397-403.
					55. El-Agnaf, O.M., et al., Effect of the disulfide bridge and the C-terminal extension on the oligomerization of the amyloid peptides ABri implicated in familial British dementia. Biochemistry, 2001. 40(12): p. 3449-57.
					56. Goldsbury, C., et al., Amyloid fibril formation from full-length and fragments of amylin. J Struct Biol, 2000. 130(2-3): p. 352-62.
					57. Kaye, R., et al., Conformational transitions of islet amyloid polypeptide (IAPP) in amyloid formation in vitro. J Mol Biol, 1999. 287(4): p. 781-96.
					58. Cottingham, M.G., M.S. Hollinshead, and D.J. Vaux, Amyloid fibril formation by a synthetic peptide from a region of human acetylcholinesterase that is homologous to the Alzheimer's amyloid-beta peptide. Biochemistry, 2002. 41(46): p. 13539-47.
					59. Fezoui, Y. and D.B. Teplow, Kinetic studies of amyloid beta-protein fibril assembly. Differential effects of alpha-helix stabilization. J Biol Chem, 2002. 277(40): p. 36948-54.
					62. Chiti, F., et al., Solution conditions can promote formation of either amyloid protofilaments or mature fibrils from the HypF N-terminal domain. Protein Sci, 2001. 10(12): p. 2541-7.

Las velocidades de agregación para variantes de AcP y TTR se determinaron en condiciones que promueven el despliegue del estado nativo en un conjunto de conformaciones desplegadas o parcialmente desplegadas. Esto permitió a los inventores examinar factores que favorecen la formación de amiloide con exclusión de cualquier implicación de cambios en la estabilidad del estado nativo que podrían ocurrir como consecuencia de las mutaciones. Dado que las secuencias remanentes son todas ellas péptidos que no se pliegan en una estructura globular definida, pueden utilizarse datos cinéticos de soluciones tamponadas en tanto que se mantiene la confianza en que los cambios en las velocidades de agregación consignadas en la bibliografía no son debidos a modificación en la estructura del estado nativo.

5 Se determinaron primeramente (véase *Métodos*) los coeficientes α dados en Eq (3) por ajuste de los mismos a partir de los valores experimentales de $\ln(k)$ para las proteínas, péptidos y sus mutantes como se consigna en la Tabla 8. Los valores consignados en la Tabla 9 siguiente representan las mejores estimaciones de los inventores para estos parámetros. La Tabla 9 muestra también su significación estadística (valor p).

Tabla 9

	α	Valor p
Ordenada en el origen	-6,2	
Hidrofobicidad	-0,08	0,005
Patrón	0,96	< 0,001
Hélice- α	-0,07	0,060
Lámina- β	0,08	0,031

	α	Valor p
Carga	-0,47	< 0,001
pH	-0,22	0,284
Iónico	-0,03	<0,001
Concentración	3,05	< 0,001

La Figura 11 muestra los resultados del análisis de regresión realizado sobre la serie de datos completa, que compara las velocidades de agregación calculadas y observadas para diversas secuencias. Los valores calculados para $\ln(k)$, determinados utilizando la Ecuación (3) y los coeficientes α consignados en la Tabla 9, se representan gráficamente contra los valores experimentales. Los datos para AcP de tipo salvaje y sus mutantes o variantes se representan en rombos, mientras que los datos para las otras secuencias en la serie de datos se representan en triángulos. La comparación entre las velocidades de agregación predichas y las experimentales para el dominio N-terminal de HypF se representa como un cuadrado. El coeficiente de correlación lineal de los valores calculados y observados para la serie de datos completa es 0,92 ($p < 0,0001$). El error cuadrático medio entre los valores $\ln(k)$ calculados y observados era 0,7; este valor es una estimación del error estadístico en la predicción de $\ln(k)$, consistente con los resultados obtenidos por el test de autogeneración ('bootstrapping') (véase más adelante).

Validación de las predicciones

Con objeto de testar la exactitud y el poder predictivo de la Ecuación (3) para determinar las velocidades de agregación de las cadenas de polipéptidos se utilizaron dos métodos de validación cruzada, un procedimiento de autogeneración ('bootstrapping') [Press, W.H., et al., Modeling of Data. Numerical Recipes in C++, 2002 (Cambridge University Press): p. 696-697], y un método de navaja ('jackknife') [Mardia, K.V., J.T. Kent, y J.M. Bibby, Multivariate Analysis. Academic Press Londres, 1979].

En el *test de autogeneración*, la serie de datos completa se dividió aleatoriamente en dos subconjuntos. El primer conjunto, compuesto de los dos tercios de las secuencias, se utilizó como el conjunto de aprendizaje, a partir del cual se estimaron los coeficientes α . Estos coeficientes se utilizaron luego para predecir las velocidades de agregación de las secuencias restantes, el conjunto de test. El procedimiento se repitió 25 veces, cada vez con una elección aleatoria diferente del conjunto de aprendizaje. La distribución de los coeficientes de correlación entre los valores predichos y los experimentales se representó gráficamente para los conjuntos de aprendizaje y de test.

La Figura 12a muestra los resultados del test de autogeneración para la Ecuación 3. El histograma muestra la distribución de los coeficientes de correlación de los conjuntos de aprendizaje 1200 y los conjuntos de test 1202 para las 25 pruebas. El coeficiente de correlación para el conjunto de aprendizaje estaba comprendido entre 0,98 y 0,94, con un máximo en 0,92. El valor p es menor que 0,0001 en todos los casos. El coeficiente de correlación para el conjunto de test estaba comprendido entre 0,50 y 0,94 con un máximo en 0,84. Se obtuvieron coeficientes de correlación más bajos que 0,70 sólo en cuatro casos. Una inspección de los conjuntos de aprendizaje utilizados en estos casos reveló que la selección aleatoria había excluido un conjunto completo de datos experimentales (datos correspondientes a medidas realizadas en las mismas condiciones experimentales), lo que hacía algo inexacto el ajuste de los factores dependiente exclusivamente de las condiciones experimentales, es decir los parámetros E extrínsecos.

Se adoptó luego el método de validación cruzada de *navaja*, en el cual se predice sucesivamente una velocidad para cada secuencia después de haber dejado de lado dicha secuencia particular (así como cualesquiera secuencias correspondientes a mutantes de dicho polipéptido original) durante la determinación de los coeficientes α óptimos para las secuencias restantes. Se realizó este procedimiento para la totalidad de los polipéptidos de tipo salvaje y mutados consignados en la Tabla 8; las condiciones experimentales para cada velocidad observada se consignan en la Tabla 8. El coeficiente de correlación lineal entre las velocidades predichas y las observadas era en este caso 0,88. Los resultados de este test para las secuencias no homólogas de tipo salvaje en la serie de datos utilizada por los inventores se muestran en la Figura 12b.

La Figura 12b muestra los valores $\ln(k)$ predichos para la totalidad de las secuencias no homólogas de tipo salvaje en la serie de datos de los inventores por medio del análisis de validación cruzada de la *navaja*. Los valores predichos de $\ln(k)$ para cada una de las secuencias de tipo salvaje representadas se calcularon utilizando un análisis de regresión de los datos para todas las secuencias de la serie de datos excepto los datos correspondientes a la secuencia simple de tipo salvaje predicha. La concordancia relativamente satisfactoria entre las velocidades de agregación predichas y las experimentales para las diversas proteínas y péptidos examinados en este estudio muestra la fiabilidad de la fórmula en la determinación de las velocidades de agregación absolutas de los estados desestructurados.

Un test convincente para la fórmula de los inventores es la predicción de la velocidad de agregación del dominio N-terminal de la proteína procarionota globular HypF. Se ha demostrado que esta cadena de polipéptido de 91 residuos forma fibrillas amiloides en condiciones similares a las utilizadas en los estudios con AcP. HypF forma fibrillas

amiloides incluso más rápidamente que AcP, que tiene una de las velocidades de agregación de amiloide más rápidas en la serie de datos utilizada. Utilizando la Ecuación (3) se predice $\ln(k) = -3,8$ para HypF. Un límite experimental para la velocidad de agregación es $\ln(k) \geq 2,5$. La comparación entre las velocidades de agregación predichas y observadas de HypF (véase la Figura 11) muestra que ambos valores son significativamente más rápidos que cualquier otra velocidad en la serie de datos de los inventores.

Influencia de los factores individuales

Los valores de los coeficientes α que fueron determinados por los inventores nos permiten explorar la influencia de diferentes factores sobre la propensión a una secuencia a formar agregados de amiloide.

Factores intrínsecos

10 *Hidrofobicidad.* Se ha sugerido muchas veces que las interacciones hidrófobas juegan un papel significativo en la formación de amiloides. La escala de hidrofobicidad que utilizaron los inventores asigna valores positivos a los residuos hidrófilos y valores negativos a los residuos hidrófobos [Roseman, M.A., Hydrophilicity of polar amino acid side-chains is markedly reduced by flanking peptide bonds. *J Mol Biol*, 1988.200(3): p. 513-22; Creighton, T.E., 4.2.3 Aqueous Solutions, en *Proteins. Structure and molecular properties*. 1993, W.H. Freeman & Co.: Nueva York. p. 15 Tabla 4.8, columna 6]. Dado que los inventores encontraron un coeficiente negativo (-0,08) significativo ($p = 0,005$) para f^{hydr} , el análisis de los inventores confirma la importancia del efecto de la hidrofobicidad sobre la agregación. A medida que aumenta la hidrofobicidad, f^{hydr} se hace más negativo, conduciendo a una contribución positiva a $\ln(k)$, y dando como resultado una velocidad más rápida.

20 *Patrones hidrófobos.* La determinación de patrones hidrófobos es uno de los determinantes más significativos ($p < 0,001$) de las velocidades de agregación en la Ecuación (3). La importancia de los patrones hidrófobo-hidrófilos ha sido estudiada concienzudamente por Hecht y colaboradores [véase, por ejemplo, Wurth, C., N.K. Guimard, y M.H. Hecht, Mutations that reduce aggregation of the Alzheimer's Abeta42 peptide: an unbiased search for the sequence determinants of Abeta amyloidogenesis. *J Mol Biol*, 2002. 319(5): p. 1279-90], y se ha demostrado que patrones alternantes del tipo utilizado por los inventores se encuentran entre los rasgos menos comunes de las secuencias de proteínas naturales [Broome, B.M. y M.H. Hecht, Nature disfavors sequences of alternating polar and non-polar amino acids: implications for amyloidogenesis. *J Mol Biol*, 2000. 296(4): p. 961-8]. Se encontró que una longitud de 5 residuos alternantes hidrófobos e hidrófobos consecutivos daba la correlación más significativa con la cinética de agregación. El valor positivo del coeficiente para los patrones (0,96) indica que cuanto más patrones de este tipo se encuentran en una secuencia dada, tanto más rápida es la velocidad de agregación.

30 *Propensiones a Estructura Secundaria.* La significación y los signos de los coeficientes para las propensiones a hélice α ($p = 0,057$, $\alpha_\alpha = -0,07$) y lámina β ($p = 0,031$, $\alpha_\beta = 0,08$) indican, como era de esperar, que la formación de fibrillas amiloides se ve favorecida por un valor alto de la propensión global a lámina β y por un valor bajo de la propensión global a hélice α en la secuencia del polipéptido.

35 *Carga.* El signo negativo altamente significativo ($p < 0,001$) para el coeficiente de la contribución de la carga ($\alpha = -0,47$) indica que la velocidad de agregación aumenta a medida que disminuye el valor absoluto de la carga neta; se ha observado anteriormente una correlación de este tipo para AcP y sus mutantes [Chiti, F., et al., Studies of the aggregation of mutant proteins in vitro provide insights into the genetics of amyloid diseases. *Proc Natl Acad Sci US A*, 2002b. 99 Suppl 4: p. 16419-26]. Sin embargo, en un estudio se demostró que cargas de ± 1 eran más favorables para la formación de amiloide que cargas netas de 0 o ± 2 [Lopez De La Paz, M., et al., De novo designed peptide-based amyloid fibrils. *Proc Natl Acad Sci USA*, 2002. 99(25): p. 16052-7]. No obstante, modificaciones de la forma funcional de f^{ch} , el término que describe la contribución de la carga a las velocidades de agregación en Eq (3), desde una forma lineal a una forma polinómica con máximos en ± 1 daban un coeficiente de correlación menor. Es probable que las condiciones que favorecen cinéticas de agregación rápida no coincidan necesariamente con las óptimas para la formación de ensamblajes amiloides bien ordenados, como se ha sugerido en algunos sistemas experimentales analizados hasta ahora. De este modo, si examinamos los parámetros importantes para influir en la cinética de agregación de un polipéptido dado cualesquiera que sean las características morfológicas particulares exhibidas por los ensamblajes finales, los descubrimientos de los inventores son consistentes con resultados previos [Chiti *ibid*], lo que sugiere que la acumulación de cargas ejerce un efecto inhibitorio sobre la agregación de los polipéptidos, con indiferencia de cuál sea la estructura final adoptada por el polipéptido.

Factores extrínsecos

55 *pH.* Nuestros resultados indican que el pH es inversamente proporcional a las velocidades de agregación. Esto es consistente con la observación de que la formación de fibrillas de amiloide se encuentra que ocurre a menudo a pH bajo. El pH es menos significativo ($p = 0,28$) que los otros factores en la Ecuación (3), debido muy probablemente a que el mismo está tenido en cuenta ya en gran parte por otros factores, tales como la hidrofobicidad, los patrones hidrófobos y la carga.

Fuerza iónica. Se encontró una correlación muy significativa ($p < 0,001$) en los datos entre fuerzas iónicas mayores y velocidades de agregación menores. Si se excluye del análisis la fuerza iónica, la Ecuación (3) da todavía un coeficiente de correlación de 0,87 en lugar de 0,92 entre las velocidades de agregación calculadas y las observadas.

La fuerza iónica incrementada puede, al menos en algunos casos, reducir las velocidades de agregación en los campos de valores utilizados en la serie de datos de los inventores.

5 *Concentración de Péptido.* De acuerdo con la Ecuación (3), la velocidad de agregación aumenta significativamente ($p < 0,001$) con la concentración de péptidos C. Se testaron varias formas funcionales de E^{conc} , y la dependencia logarítmica, $E^{conc} = \ln(C+1)$ permitida para las mejores predicciones en un amplio intervalo de C. Dado que todos los datos experimentales que se consideraron se obtuvieron por encima de la concentración crítica para la agregación, la extrapolación de los resultados obtenidos con la Ecuación (3) a C baja debería considerarse cuidadosamente.

Factores adicionales a considerar y mejoras futuras

10 El presente análisis está basado en una cantidad relativamente limitada de datos experimentales disponibles hasta la fecha y, como resultado, podrían haber sido pasados por alto elementos relevantes para definir en detalle la agregación de los polipéptidos debido a la falta de datos. Un enfoque alternativo podría utilizar redes neurales para extraer parámetros, sin necesidad de hacer suposiciones sobre la forma funcional (desconocida) [Rumelhart, D. y J. McClelland, *Parallel Distributed Processing: Exploration in the Microstructure of Cognition*. MIT Press, Cambridge, MA., 1986]; sin embargo, este enfoque podría ser menos informativo en términos de comprensión de la importancia relativa de los diferentes elementos en el mecanismo de la agregación de los polipéptidos.

15 Se ha considerado aquí una colección de factores intrínsecos y extrínsecos que contribuyen al comportamiento de agregación de los polipéptidos. Factores adicionales, tales como la estabilidad del estado nativo, la temperatura o la agitación pueden incluirse en el algoritmo de predicción, con tal que se disponga de datos adecuados para permitir una determinación fiable de sus coeficientes. Se sabe que una temperatura incrementada conduce a velocidades de agregación más rápidas en muchos casos. Sin embargo, la falta de variación entre las temperaturas experimentales para las velocidades incluidas en la serie de datos hizo difícil establecer su contribución con exactitud. Otro factor experimental importante que influye en la cinética de la agregación es el grado en que las soluciones se agitan o se 'remueven'. Si se definieran los efectos de la agitación, este factor podría incluirse entonces posiblemente en la Ecuación 3. Como se ha mencionando, el procedimiento arriba descrito no tiene en cuenta la estabilidad del estado nativo, sino que predice más bien las velocidades de agregación a partir de un estado desestabilizado. En principio, sin embargo, la estabilidad del estado nativo podría considerarse también como un factor adicional en la fórmula.

20 Utilizando una combinación de parámetros intrínsecos y extrínsecos como se ha detallado arriba junto con un análisis multivariante de los datos experimentales disponibles, la Ecuación (3) es capaz de predecir velocidades de agregación absolutas para cualquier secuencia de polipéptido. Las velocidades de agregación calculadas utilizando el método de los inventores están correlacionadas con las velocidades observadas experimentalmente con un coeficiente de 0,92 (autogeneración con validación cruzada 0,84, navaja con validación cruzada 0,88) y puede, por tanto, esperarse que produzcan predicciones exactas dentro de los intervalos de condición incluidos en la serie de datos de los inventores, a saber pH 4,4 a 9,0, fuerza iónica de 0,1 a 150 mM, y concentración de péptido de 0,01 a 2 mM. La fórmula derivada en este ejemplo se obtuvo ignorando el hecho de que ciertas regiones de una cadena polipeptídica son más importantes que otras para la determinación de las velocidades de agregación. Esta aproximación es probablemente responsable de la influencia relativamente pequeña de las propensiones a estructura secundaria que se encontraron en el ejemplo. No obstante, se ha encontrado una correlación muy significativa entre las velocidades de agregación predichas y las experimentales. La calidad de la predicción puede mejorarse ulteriormente por combinación de la ecuación (3) con un algoritmo capaz de predecir las regiones sensibles de una cadena polipeptídica, tales como la arriba descrita con referencia al Ejemplo 6. Sin embargo, el hecho de que regiones sensibles importantes para la agregación no necesiten ser conocidas para utilizar esta fórmula mejora notablemente su aplicabilidad general.

25 Así pues, se ha analizado el efecto de una combinación de propiedades intrínsecas de la secuencia y elementos experimentales extrínsecos para predecir con exactitud las velocidades de agregación exhibidas por diferentes polipéptidos de distinto origen. La notable concordancia entre las velocidades de agregación absolutas predichas y los valores obtenidos experimentalmente demuestra que parámetros simples que definen una secuencia de polipéptido y su entorno pueden utilizarse para racionalizar, en gran medida, su propensión a la agregación. La capacidad de predecir la propensión a la agregación exhibida por un péptido o proteína dado(a) con exactitud y precisión que es potencialmente una herramienta poderosa para ayudar a comprender el comportamiento de los polipéptidos naturales y su propensión a agregarse, así como para establecer de qué modo las secuencias han evolucionado en la naturaleza a fin de evitar el plegado erróneo. Además, este método puede aplicarse para comprender mejor y quizás predecir incluso la aparición de amiloidosis y otras enfermedades de deposición, y para ayudar a explicar estrategias terapéuticas eficaces para su tratamiento.

Series de datos

55 Los datos cinéticos acerca de la agregación de AcP y sus mutantes se obtuvieron de la bibliografía como se expone en la Tabla 8; en estos estudios, se utilizó la fluorescencia de ThT (Tioflavina T) para determinar la velocidad de agregación de cada proteína en solución. Los datos de AcP se midieron todos en condiciones idénticas y proporcionaron la mayor serie de datos utilizada en el presente análisis (60 secuencias). La segunda serie de datos incluía las velocidades de agregación de varios péptidos diferentes en distintas condiciones, obtenidas de resultados

publicados (véase la Tabla 8 para referencias). Se condujo inicialmente una investigación bibliográfica utilizando 'cinética' y 'fibrilla' o 'amiloidé' como palabras clave, dando como resultado una lista inicial de más de 800 referencias. Se seleccionó luego una serie de estudios fácilmente disponibles que describían experimentos cinéticos realizados sobre péptidos cortos o proteínas en una solución tampón que formaban fibrillas detectables al microscopio electrónico durante el curso del experimento. Se seleccionaron así 10 referencias que proporcionaron a los inventores datos cinéticos acerca de 23 secuencias en concentraciones salinas diferentes, ocasionalmente con pequeñas cantidades de codisolvente remanente de la solución stock de péptido. Una vez seleccionados utilizando los criterios arriba descritos, no se excluyó del análisis secuencia alguna, ni se añadió ninguna secuencia nueva.

Las velocidades de agregación se determinaron a partir de trazas cinéticas obtenidas por los métodos siguientes: fluorescencia de ThT (transtiretina), turbidez, CD, o estimación directa de la cantidad relativa de material agregado utilizando técnicas tales como sedimentación, cromatografía de exclusión por tamaños, y filtración. Aunque estos métodos detectan aspectos ligeramente diferentes de la agregación, los mismos están ligados estrechamente, y en algunos sistemas en los que se han aplicado dos o más técnicas experimentales, se ha observado un perfil cinético similar. En el análisis de los inventores no se consideraron fases de retardo, debido a que las mismas no se consignaban a menudo o eran difíciles de extraer de los datos publicados. Además, no se dispone todavía de una comprensión global de las fases de retardo en la agregación de las proteínas, y el análisis presente está enfocado en la cinética de agregación después de la fase de retardo, donde se observa generalmente una fase de elongación con comportamiento exponencial simple. Las trazas cinéticas se ajustaban a la ecuación $y = A(1 - e^{-kx})$ donde k es la constante de velocidad en s^{-1} . En la Ecuación (3) se utilizó el logaritmo natural de la constante de velocidad ($\ln(k)$), dado que los valores de $\ln(k)$ se describían mejor por una distribución normal que k propiamente dicha. En algunos sistemas, soluciones sembradas y no sembradas dan como resultado velocidades de agregación prácticamente idénticas si se desprecia la fase de retardo en la solución no sembrada. Se estimó que una inclusión de la fase de retardo podría cambiar las velocidades de agregación en no más que un factor de 5, dando como resultado un error de 1,6 en el logaritmo; este número sería comparable al valor estadístico de 0,7 en las predicciones de los inventores.

Los valores de $\ln(k)$ determinados por diferentes métodos en estos documentos difieren en menos de 0,2 unidades en todos los casos excepto uno, en el que la cinética de turbidez y la cinética de ThT difieren en 1,9 unidades, probablemente como resultado de otras diferencias en el procedimiento experimental. En los estudios experimentales que fueron considerados por los inventores, se utilizaron análisis masa/volumen en ausencia de una técnica independiente para confirmar los resultados. Sin embargo, dado que estos métodos pueden considerarse como el método de observación más directo del crecimiento de agregados físicos, los datos obtenidos exclusivamente por estos métodos se incluyeron en el análisis.

Deducción de la fórmula

La forma funcional en cada factor en la Ecuación (3) se seleccionó después de examinar una diversidad de combinaciones fenomenológicas de los factores que influían probablemente en la propensión a la agregación. Se consideraron dos clases de factores, intrínsecos y extrínsecos. Los factores intrínsecos incluían las propiedades de la secuencia de aminoácidos, tales como hidrofobicidad, patrones hidrófobos, propensiones a estructura secundaria, y carga. Sus formas funcionales se determinaron por examen de un subconjunto de mutantes de AcP para encontrar la representación que se correlacionaba mejor con los cambios en $\ln(k)$ entre los mutantes. Los factores extrínsecos incluían concentración de péptido, fuerza iónica, y pH. Se utilizó una forma logarítmica para el término que describía el efecto de la concentración de péptido a fin de evitar la sobreestimación de las velocidades a concentraciones mayores. Se supuso que los otros términos tenían una forma lineal.

Las regresiones se realizaron utilizando el software estadístico Rweb1.03 [Rweb1.03, www.math.montana.edu The R Development Core Team Version 1.4.1, 2002] para obtener coeficientes α en la Ecuación (3) que minimicen las diferencias entre los valores $\ln(k)$ calculados y experimentales. En la interpretación del significado de las constantes numéricas en la fórmula debería tenerse en cuenta su naturaleza fenomenológica. La fórmula puede incluir el hecho de que se tengan en cuenta dos veces algunos factores (v.g. hidrofobicidad y patrones hidrófobos), pero esto no constituye un problema importante, dado que los coeficientes están ajustados a partir de datos experimentales y no se derivan de principios fundamentales.

Diagrama de flujo para una implementación por computadora de los métodos del Ejemplo 7

La Figura 13 muestra un procedimiento adicional, que puede implementarse una vez más utilizando el sistema de computadora de la Figura 3 corriendo un código apropiado, para implementar la Ecuación 3 anterior a fin de determinar una estimación de una velocidad de agregación absoluta en lugar de las velocidades de agregación relativas predichas por las Ecuaciones 1 y 2.

Muchos de los pasos de la Figura 13 son similares a los de la Figura 6 anterior y, en particular, los pasos s1300-s1312 y s1315 corresponden en términos generales a los pasos s600-s612 y s616 de la Figura 6. Sin embargo, en el paso s1300, además de los parámetros del paso s600, se introducen en el procedimiento parámetros extrínsecos adicionales, en particular un valor de fuerza iónica (de la solución de polipéptido, por ejemplo en unidades milimolares), y un valor de concentración C que es una medida de la concentración de polipéptido, por ejemplo en

unidades milimolares, y que se utiliza para determinar un parámetro de concentración E^{conc} para la Ecuación 3 utilizando $E^{conc} = \ln(C + 1)$. Estos valores extrínsecos adicionales se utilizan en la determinación de la velocidad de agregación absoluta utilizando la Ecuación 3 en el paso s1314. En los pasos s1302 a s1306, los datos de hidrofobicidad, carga y propensión a lámina β se suman (de manera similar al procedimiento de la Figura 4) en lugar de determinarse para cada aminoácido. En el paso s1312, cada patrón alternante de aminoácidos identificado cuando se avanza a lo largo de la secuencia recibe un valor de, por ejemplo, +1 en lugar de asignar un valor de f^{pat} a cada aminoácido particular de la secuencia. En el paso s1318, los datos de velocidad de agregación absoluta se aportan de cualquier manera convencional para uso ulterior, por ejemplo como se ha descrito previamente.

Referencias

- 10 Dobson, C. M. Protein folding and its links with human disease. *Biochem. Soc. Symp.* 68, 1-26 (2001).
- Jarrett, J. T., Berger, E. P. & Lansbury, P. T. Jr. The carboxy terminus of the beta amyloid protein is critical for the seeding of amyloid formation: implications for the pathogenesis of Alzheimer's disease. *Biochemistry* 32, 4693-4697.
- Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.* 81, 741-766 (2001).
- Siepen, J. A. & Westhead, D. R. The fibril_one on-line database: Mutations, experimental conditions, and trends associated with amyloid fibril formation. *Protein Sci.* 11, 1862-1866 (2002).
- 15 Volles, M. J. & Lansbury, P. T. Jr. Vesicle permeabilization by protofibrillar α -synuclein is sensitive to Parkinson's disease-linked mutations and occurs by a pore-like mechanism. *Biochemistry*, 41, 4595-4602 (2002).
- Para las referencias siguientes, los números son los indicados en la tabla 2:
- 20 11. Azriel, R. & Gazit, E. Analysis of the minimal amyloid-forming fragment of the islet amyloid polypeptide. An experimental support for the key role of the phenylalanine residue in amyloid formation. *J. Biol. Chem.* 276, 34156-34161 (2001).
12. Sakagashira, S., Hiddinga, H. J., Tateishi, K., Sanke, T., Hanabusa, T., Nanjo, K. & Eberhardt, N. L. S20G mutant amylin exhibits increased in vitro amyloidogenicity and increased intracellular cytotoxicity compared to wild-type amylin. *Am. J. Pathol.* 157, 2101-2109 (2000).
- 25 13. Salmons, M., Malesani, P., De Gioia, L., Gorla, S., Bruschi, M., Molinari, A., Della Vedova, F., Pedrotti, B., Marrari, M. A., Awan, T., Bugiani, O., Forloni, G., Tagliavini, F. Molecular determinants of the physicochemical properties of a critical prion protein region comprising residues 106-126. *Biochem J.* 342, 207-214 (1999).
- 30 14. Thompson, A. J., Bamham, K. J., Norton, R. S., Barrow, C. J. The Val-210-Ile pathogenic Creutzfeldt-Jakob disease mutation increases both the helical and aggregation propensities of a sequence corresponding to helix-3 of PrP(C). *Biochim. Biophys. Acta.* 1544, 242-254 (2001).
15. Conway, K. A., Lee, S. J., Rochet, J. C., Ding, T. T., Williamson, R. E. & Lansbury, P. T. Jr. (2000). Acceleration of oligomerization, not fibrillization, is a shared property of both alpha-synuclein mutations linked to early-onset Parkinson's disease: implications for pathogenesis and therapy. *Proc. Natl. Acad. Sci. USA*, 97, 571-576.
- 35 16. Giasson, B. I., Murray, I. V., Trojanowski, J. Q. & Lee, V. M. A hydrophobic stretch of 12 amino acid residues in the middle of α -synuclein is essential for filament assembly. *J. Biol. Chem.* 276, 2380-2386 (2001).
17. Van Nostrand, W. E., Melchor, J. P., Cho, H. S., Greenberg, S. M. & Rebeck, G. W. (2001). Pathogenic effects of D23N Iowa mutant amyloid beta -protein. *J. Biol. Chem.* 276, 32860-32866.
- 40 18. Miravalle L, Tokuda T, Chiarle R, Giaccone G, Bugiani O, Tagliavini F, Frangione B, Ghiso J. *J Biol Chem* 2000 Sep 1;275(35):27110-6
19. Nilsberth, C., Westlind-Danielsson, A., Eckman, C. B., Condron, M. M., Axelman, K., Forsell, C., Sten, C., Luthman, J., Teplow, D. B., Younkin, S. G., Naslund, J. & Lannfelt, L. (2001). The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced A β protofibril formation. *Nature Neurosci.* 4, 887-893.
- 45 20. Esler, W. P., Stimson, E. R., Ghilardi, J. R., Lu, Y. A., Felix, A. M., Vinters, H. V., Mantyh, P. W., Lee, J. P. & Maggio, J. E. Point substitution in the central hydrophobic cluster of a human β -amyloid congener disrupts peptide folding and abolishes plaque competence. *Biochemistry*, 35, 13914-13921 (1996).
21. Barghorn, S., Zheng-Fischhofer, Q., Ackmann, M., Biernat, J., von Bergen, M., Mandelkow, E. M. & Mandelkow, E. (2000). Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias. *Biochemistry*, 39, 11714-11721.

22. Gamblin, T. C., King, M. E., Dawson, H., Vitek, M. P., Kuret, J., Berry, R. W., Binder, L. I. In vitro polymerization of tau protein monitored by laser light scattering: method and application to the study of FTDP-17 mutants. *Biochemistry*, 39, 6136-6144 (2000).
- 5 23. Nacharaju, P., Lewis, J., Easson, C., Yen, S., Hackett, J., Hutton, M. & Yen, S. H. Accelerated filament formation from tau protein with specific FTDP-17 missense mutations. *FEBS Lett.* 447, 195-199 (1999).
- 23b.Li, L., Von Bergen, M., Mandelkow, E. M. & Mandelkow, E. Structure, stability, and aggregation of paired helical filaments from tau protein and FTDP-17 mutants probed by tryptophan scanning mutagenesis. *J. Biol. Chem.* in press (2002).
- 10 24. Symmons, M. F., Buchanan, S. G., Clarke, D. T., Jones, G. & Gay, N.J. X-ray diffraction and far-UV CD studies of filaments formed by a leucine-rich repeat peptide: structural similarity to the amyloid fibrils of prions and Alzheimer's disease β -protein. *FEBS Lett.* 412, 397-403 (1997).
25. Orpiszewski, J. & Benson, M. D. Induction of beta-sheet structure in amyloidogenic peptides by neutralization of aspartate: a model for amyloid nucleation. *J. Mol. Biol.* 289, 413-428 (1999).

REIVINDICACIONES

- 1.- Un método de producción de un polipéptido que tiene una secuencia de aminoácidos, **caracterizándose el método por** determinar el perfil de una secuencia de aminoácidos introducida para identificar una parte de dicha secuencia de aminoácidos que se predice promueve la agregación de un polipéptido definido por dicha secuencia, comprendiendo el método:
- 5 determinar una propensión a la agregación para cada aminoácido de dicha secuencia introducida a fin de determinar un perfil de propensión a la agregación para dicha secuencia introducida; y
- comparar dichas propensiones a la agregación de dicho perfil de propensión a la agregación a fin de determinar una o más partes de dicha secuencia introducida que se predice promueven la agregación;
- 10 utilizar dichas una o más partes de dicha secuencia introducida que se predice promueven la agregación para diseñar un polipéptido modificado que es más o menos propenso a agregarse o que tiene una propensión a agregarse dentro de un intervalo deseado; y
- producir luego un polipéptido de acuerdo con el diseño; y
- 15 en donde dicha determinación comprende determinar, para cada aminoácido de dicha secuencia, un valor de hidrofobicidad, y un valor de propensión a hélice α y/o lámina β , un valor de carga, y un valor patrón que representa un patrón de aminoácidos hidrófilos y/o hidrófobos en la proximidad de cada aminoácido, multiplicar cada uno de dichos valores por un factor de escalación, y sumar dichos valores escalados para determinar dicha propensión a la agregación.
- 2.- Un método de acuerdo con la reivindicación 1, en el cual dicho patrón comprende un patrón de aminoácidos hidrófilos e hidrófobos alternantes.
- 20 3.- Un método de acuerdo con la reivindicación 2, en el cual dicho patrón tiene una longitud de al menos 5 aminoácidos.
- 4.- Un método de acuerdo con una cualquiera de las reivindicaciones 1 a 3 que comprende adicionalmente modificar dicha secuencia de aminoácidos y repetir dicha determinación de la propensión a la agregación para identificar una o más partes de dicha secuencia que se predice promueven la agregación.
- 25 5.- Un método de acuerdo con la reivindicación 4 en el que dicha modificación comprende, para cada una de una pluralidad de posiciones en dicha secuencia de aminoácidos, seleccionar cada uno de una pluralidad de aminoácidos alternativos para dicha determinación de la propensión repetida.
- 6.- Un método de acuerdo con la reivindicación 4 ó 5 que comprende adicionalmente comparar dichas propensiones a la agregación determinadas repetidamente a fin de identificar una o más partes de dicha secuencia que se predice promueven la agregación.
- 30 7.- Un portador que lleva un código de programa de computadora para diseñar un polipéptido modificado que tiene una secuencia de aminoácidos por determinación del perfil de una secuencia de aminoácidos introducida a fin de identificar una parte de dicha secuencia de aminoácidos introducida que promueve la agregación de un polipéptido definido por dicha secuencia, comprendiendo el código codificar de tal modo que, cuando se corre en un programa de computadora, haga que el sistema de la computadora:
- 35 determine una propensión a la agregación para cada aminoácido de dicha secuencia introducida a fin de determinar un perfil de propensión a la agregación para dicha secuencia introducida;
- comparar dichas propensiones a la agregación de dicho perfil de propensión a la agregación a fin de determinar una o más partes de dicha secuencia que se predice promueven la agregación; y
- 40 utilizar una o más partes de dicha secuencia que se predice promueven la agregación para diseñar dicho polipéptido modificado, siendo dicho polipéptido modificado más o menos propenso a agregarse o teniendo una propensión a agregarse dentro de un intervalo deseado,
- 45 en donde dicha determinación comprende determinar, para cada aminoácido de dicha secuencia, un valor de hidrofobicidad, un valor de propensión a hélice α y/o lámina β , un valor de carga, y un valor patrón que representa un patrón de aminoácidos hidrófilos y/o hidrófobos en la proximidad de cada uno de dichos aminoácidos, multiplicar cada uno de dichos valores por un factor de escalación, y sumar dichos valores escalados a fin de determinar dicha propensión a la agregación.
- 8.- Un sistema de computadora para determinar el perfil de una secuencia de aminoácidos de un polipéptido a fin de identificar una parte de dicha secuencia de aminoácidos que se predice promueve la agregación de un polipéptido definido por dicha secuencia, comprendiendo el sistema de la computadora:
- 50

un almacén de datos para almacenar para cada aminoácido de dicha secuencia, un valor de hidrofobicidad, un valor de propensión a hélice α y/o lámina β y un valor de carga,

un almacén de programas que comprende el portador de la reivindicación 7; y

5 un procesador, acoplado a dicho almacén de programas y a dicho almacén de datos para implementar dicho código almacenado, comprendiendo el código adicionalmente un código para controlar el procesador a fin de:

introducir dicha secuencia de aminoácidos;

leer, para cada aminoácido de dicha secuencia, un dicho valor de hidrofobicidad, un dicho valor de propensión a hélice α y/o lámina β , y un dicho valor de carga, a partir de dicho almacén de datos;

10 determinar los datos de propensión a la agregación para cada aminoácido de dicha secuencia a partir de dichos valores de hidrofobicidad, propensión a hélice α y/o lámina β , y carga y de dicho valor patrón; y

emitir como salida dichos datos de propensión a la agregación a fin de identificar una parte de dicha secuencia que se predice promueve la agregación de un polipéptido asociado con la secuencia.

15 9.- Un sistema de computadora de acuerdo con la reivindicación 8 que comprende adicionalmente un servidor de la web.

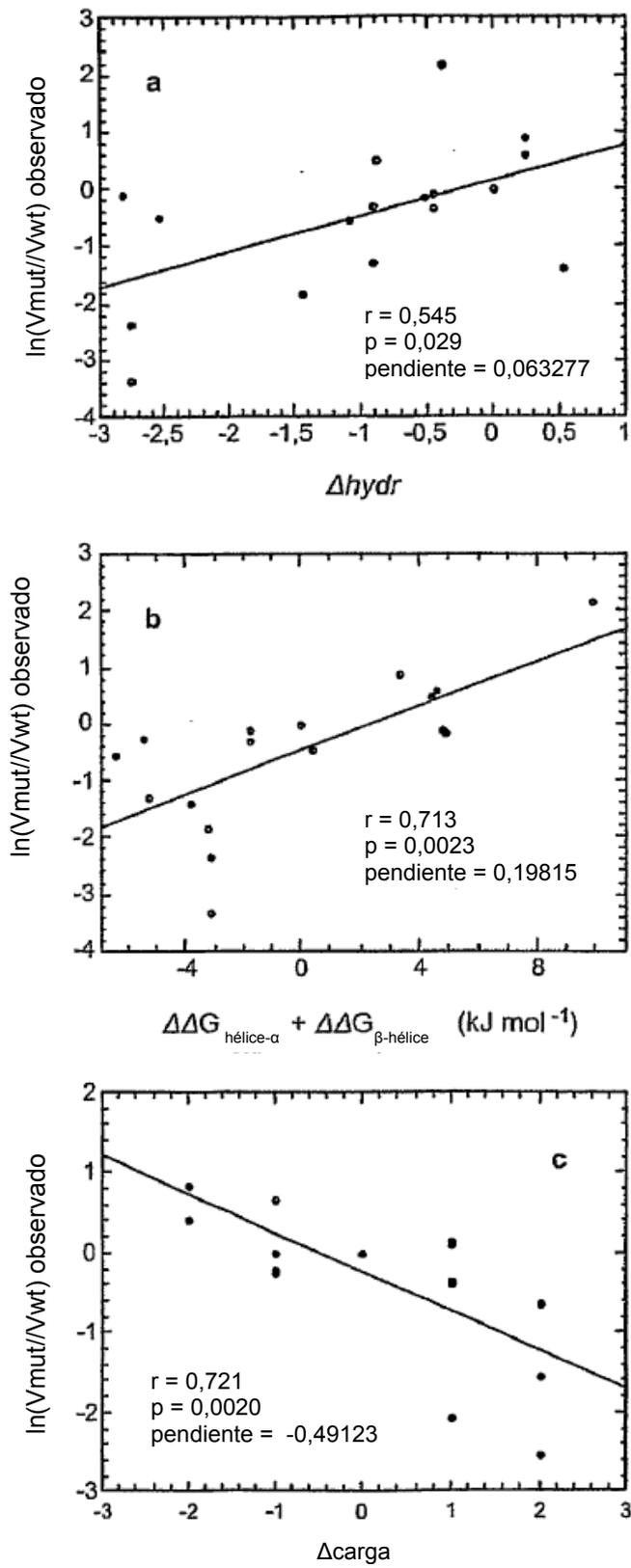


Figura 1

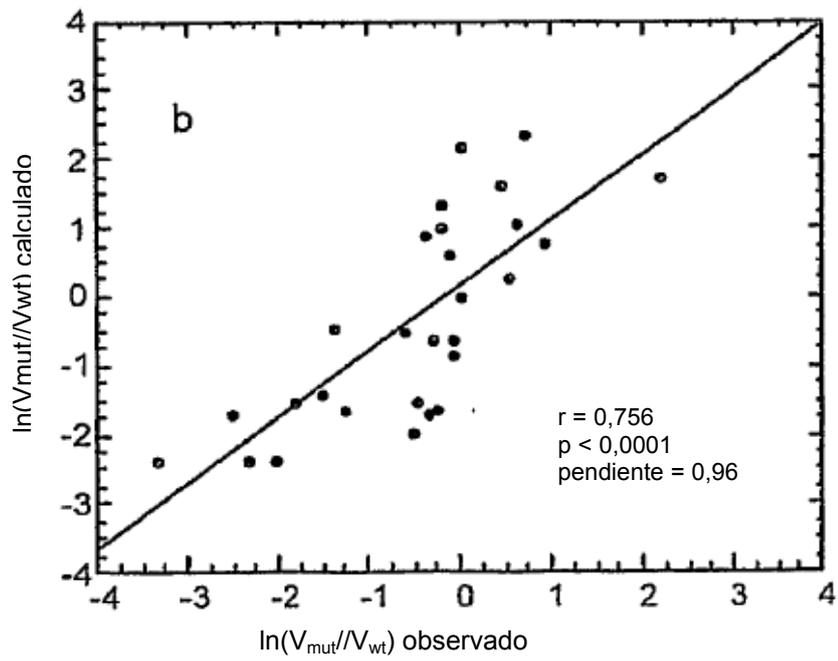
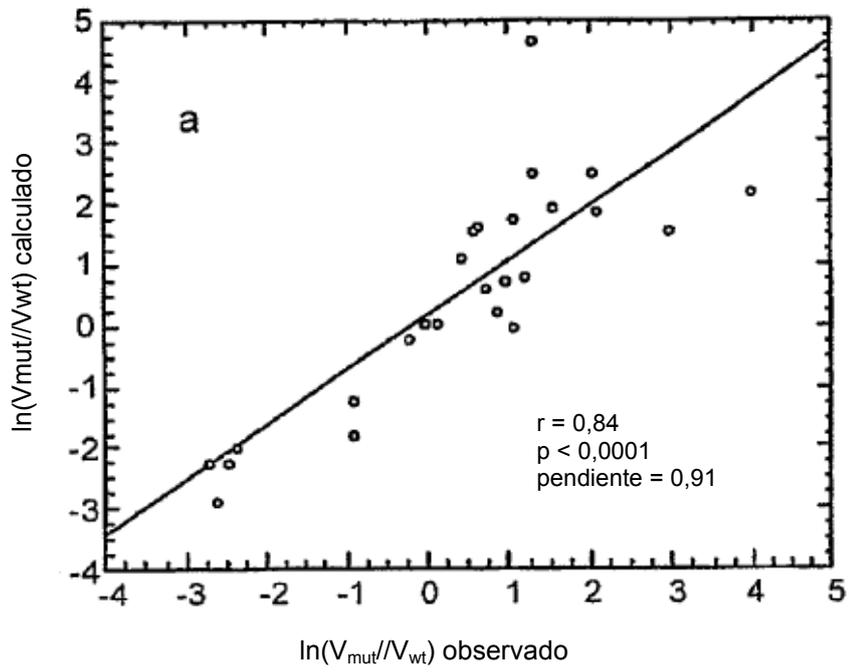


Figura 2

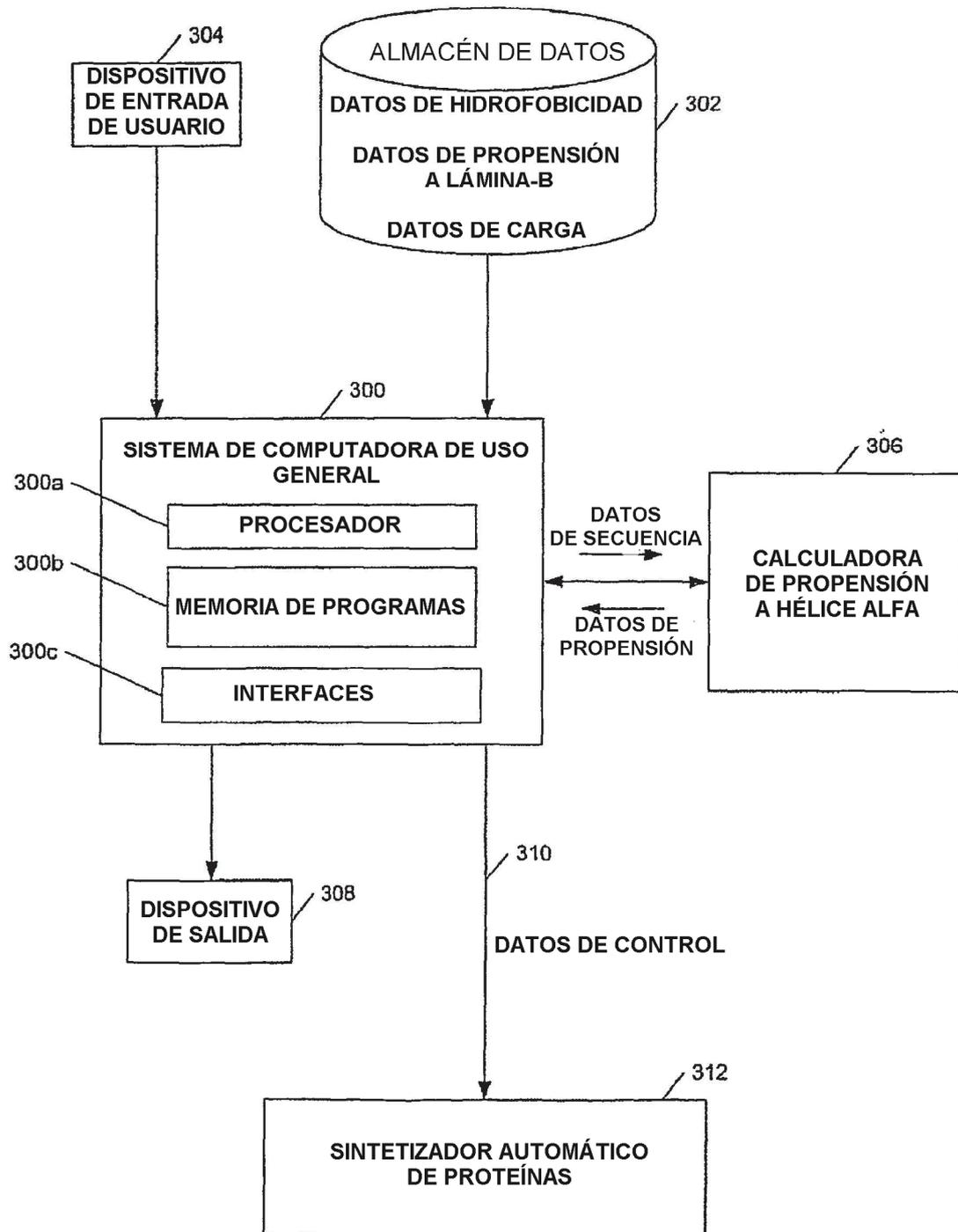


Figura 3

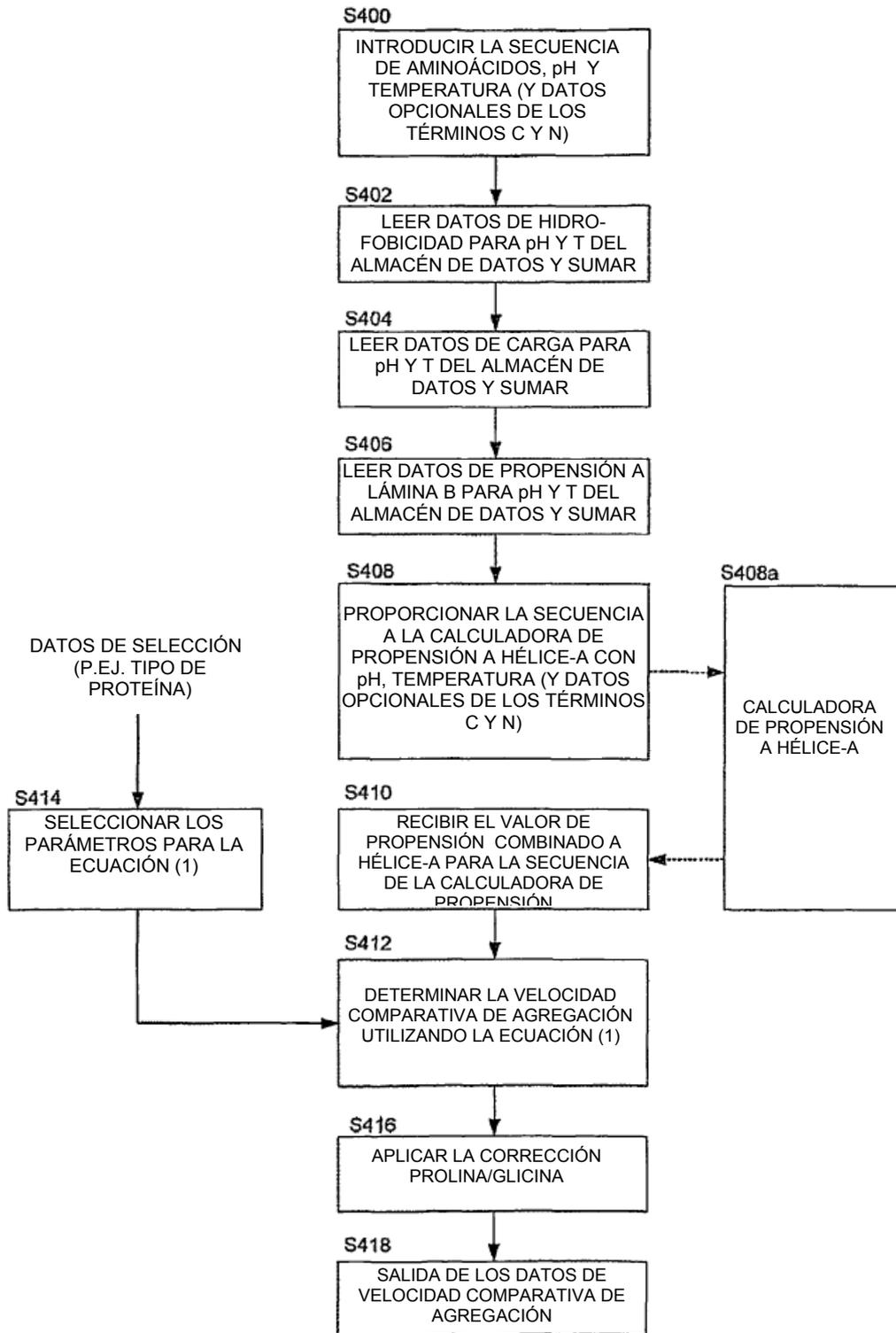


Figura 4

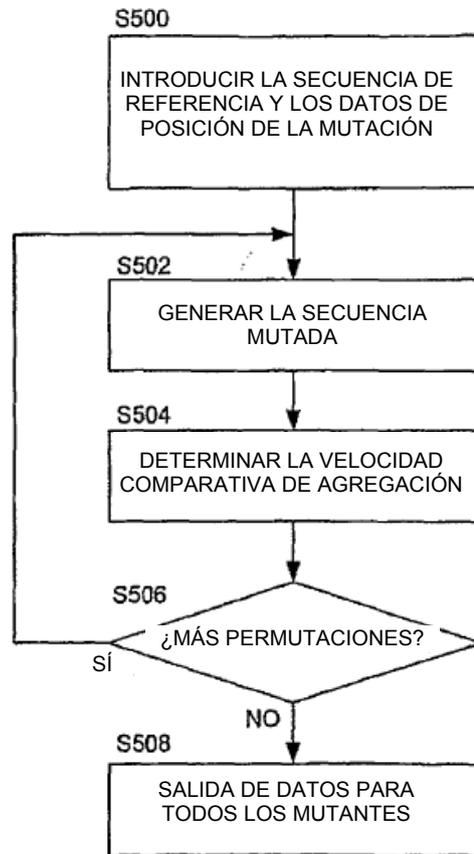


Figura 5

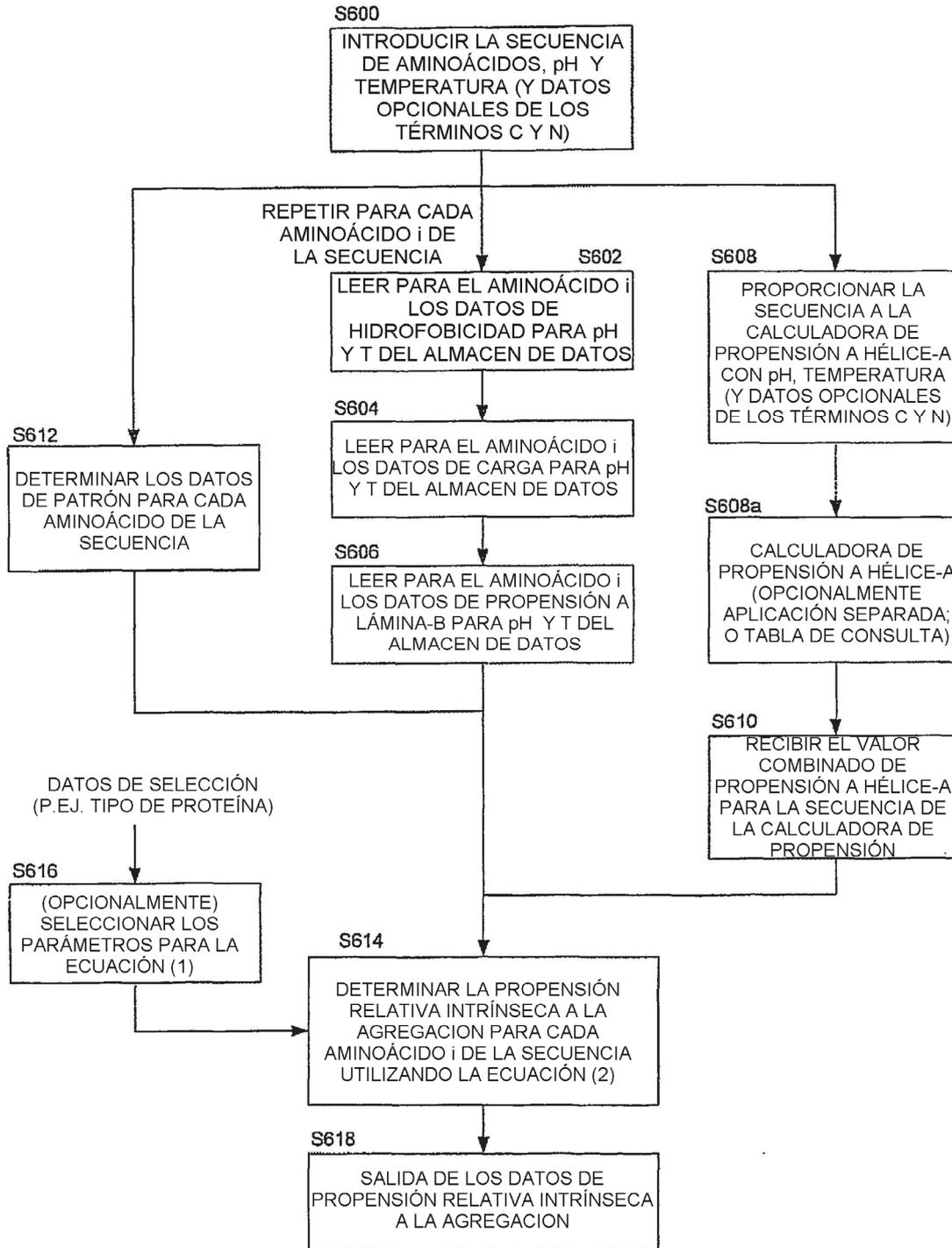


Figura 6

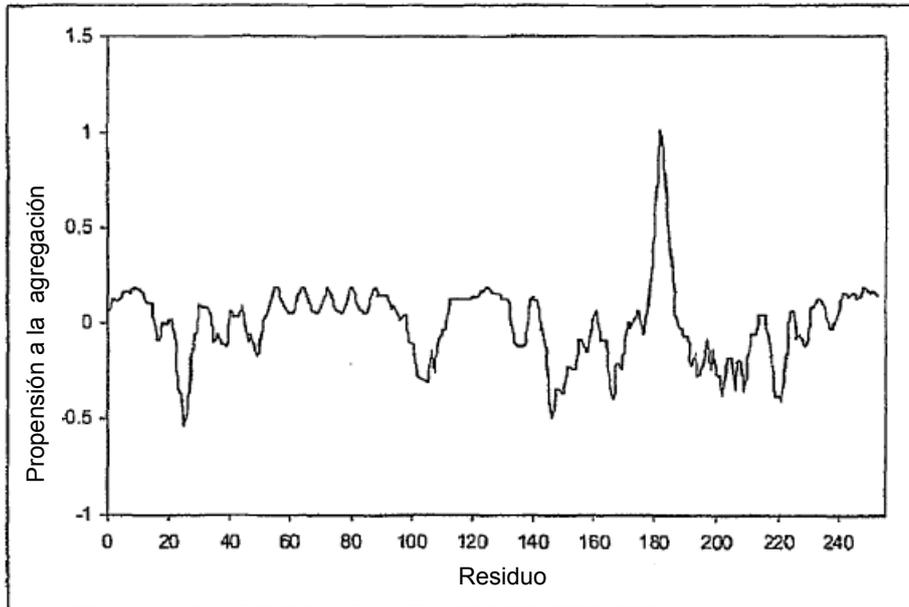


Figura 7

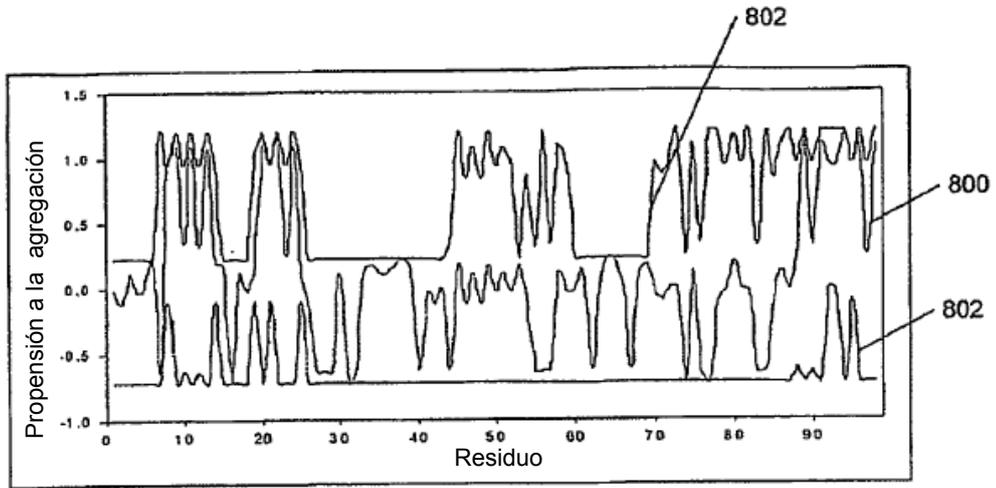


Figura 8a

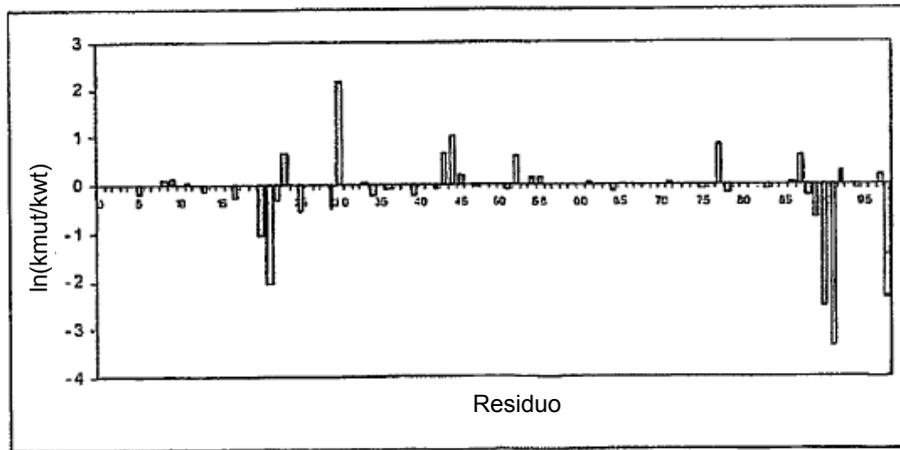


Figura 8b

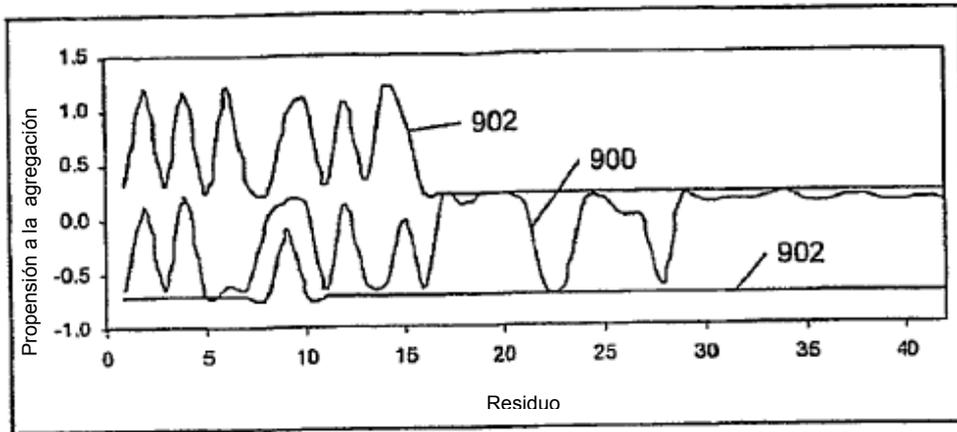


Figura 9a

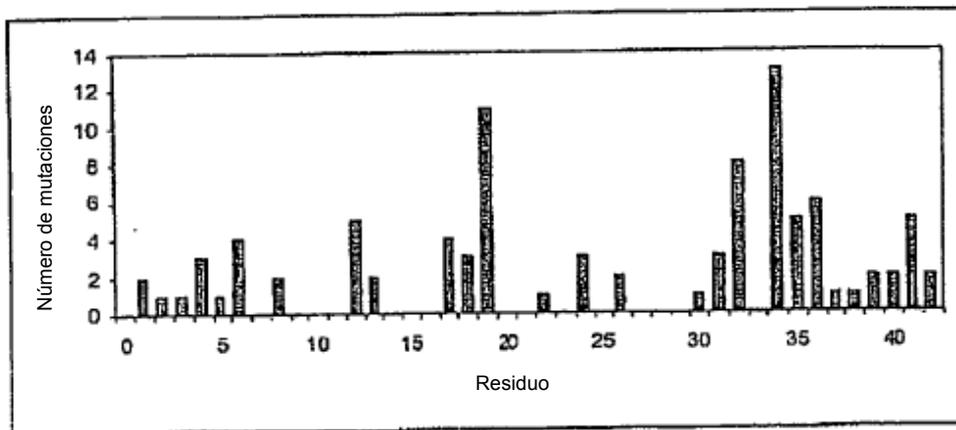


Figura 9b

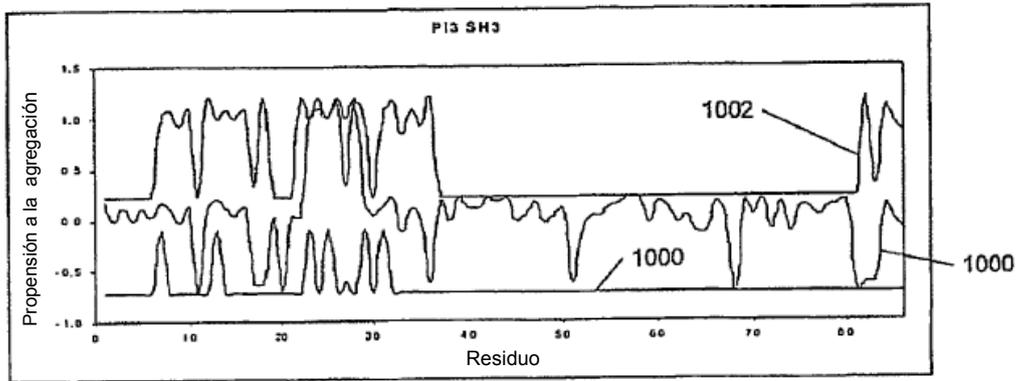


Figura 10a

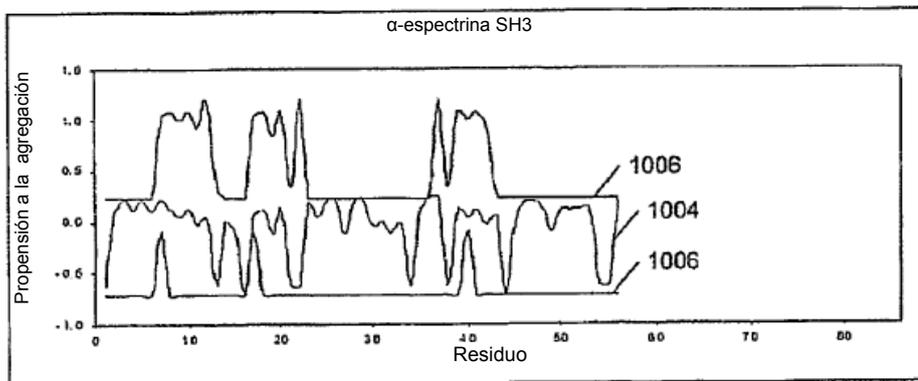


Figura 10b

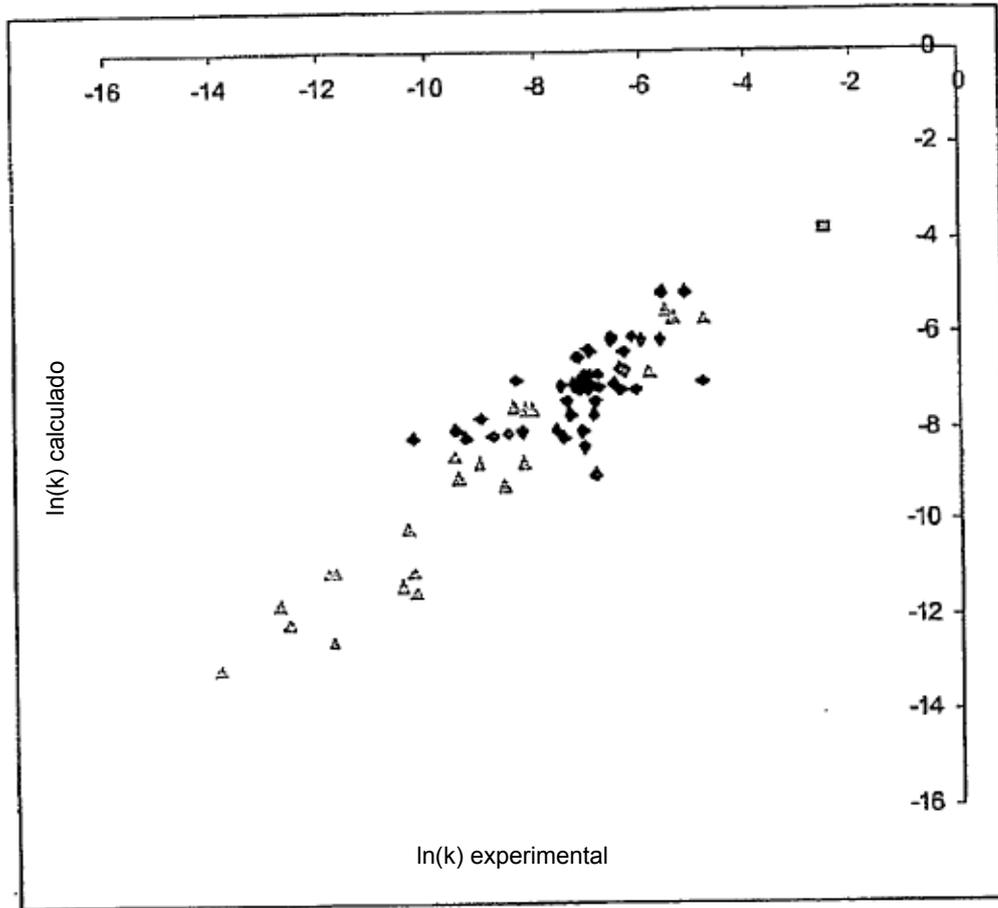


Figura 11

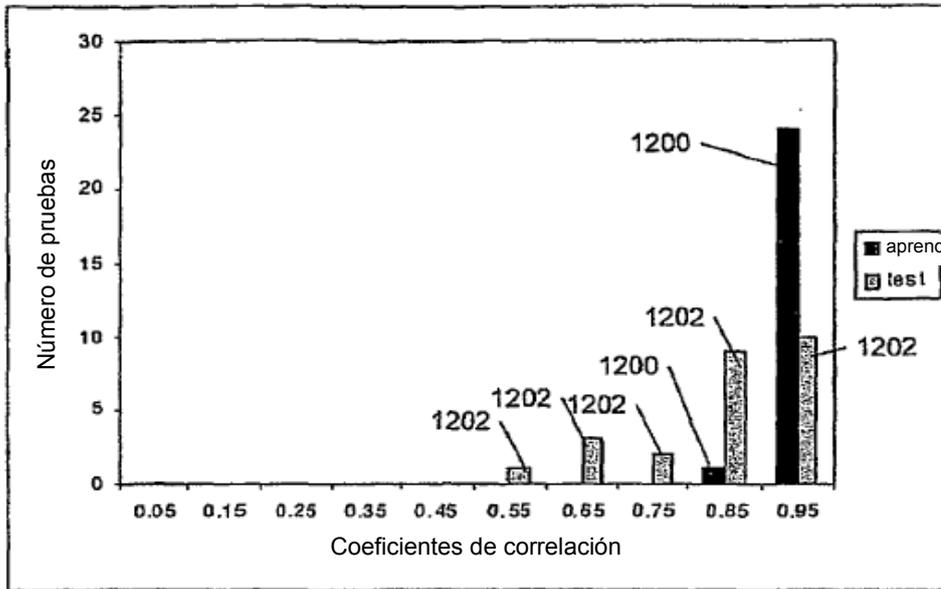


Figura 12a

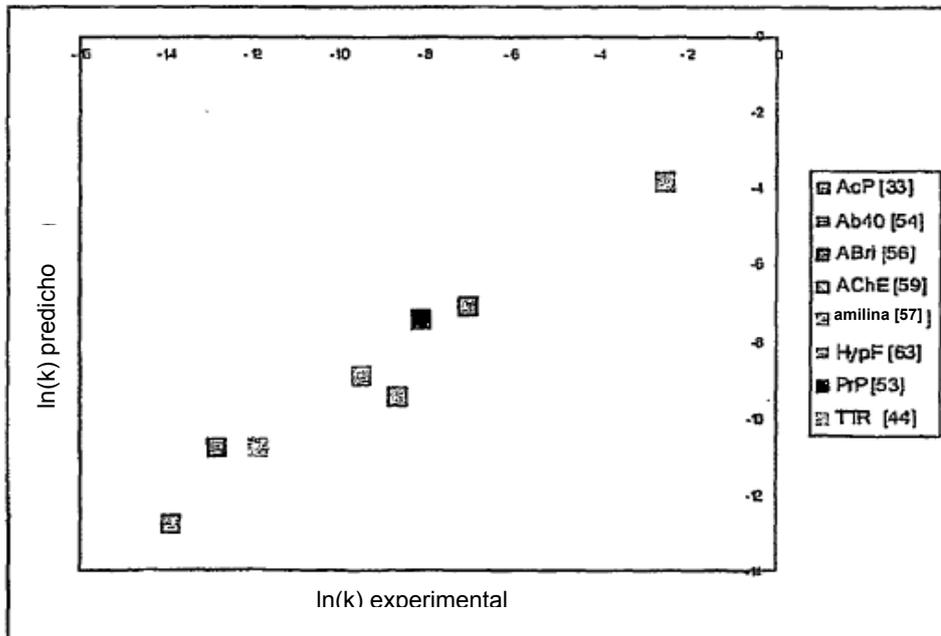


Figura 12b

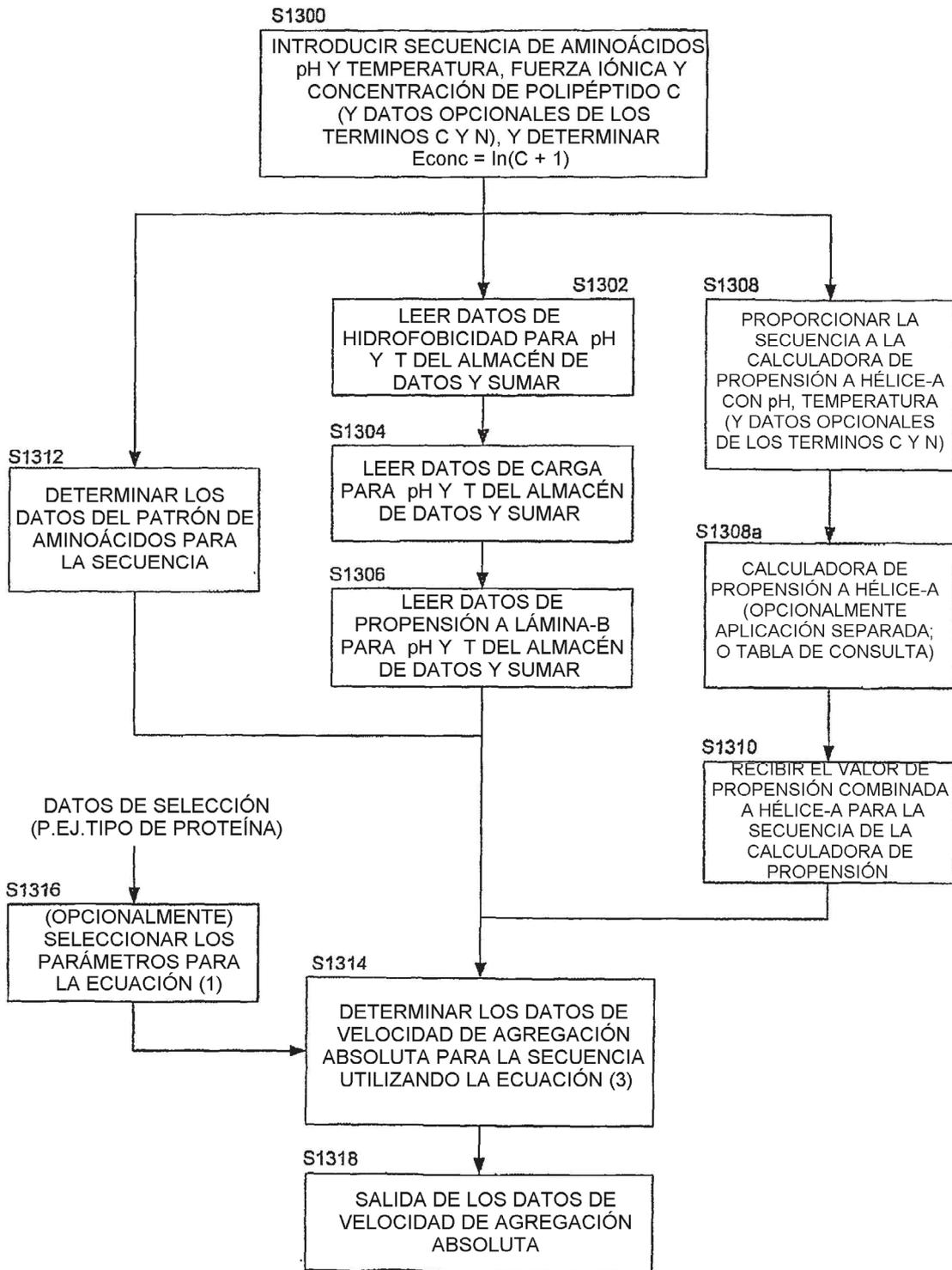


Figura 13