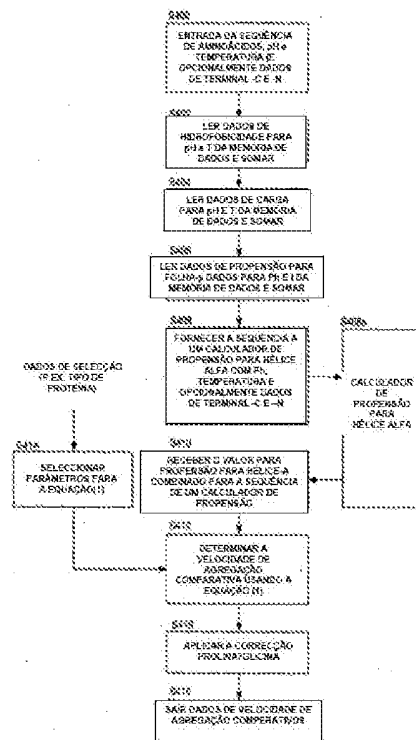


(12) **FASCÍCULO DE PATENTE DE INVENÇÃO**

|   |  |
|---|--|
| (22) Data de pedido: <b>2004.01.19</b>                                      | (73) Titular(es):<br><b>CAMBRIDGE ENTERPRISE LIMITED<br/>THE OLD SCHOOLS TRINITY LANE<br/>CAMBRIDGE CAMBRIDGESHIRE CB2 1TN</b> <b>GB</b> |
| (30) Prioridade(s): <b>2003.01.20 GB 0301235<br/>2003.01.21 US 441914 P</b> |  |
| (43) Data de publicação do pedido: <b>2005.10.19</b>                        | (72) Inventor(es):<br><b>CHRISTOPHER DOBSON</b> <b>GB</b><br><b>FABRIZIO CHITI</b> <b>GB</b><br><b>JESUS ZURDO</b> <b>GB</b>             |
| (45) Data e BPI da concessão: <b>2010.02.25<br/>103/2010</b>                | (74) Mandatário:<br><b>ALBERTO HERMÍNIO MANIQUE CANELAS<br/>RUA VÍCTOR CORDON, 14 1249-103 LISBOA</b> <b>PT</b>                          |

(54) Epígrafe: **MÉTODO COMPUTACIONAL E APARELHO PARA PREVER A AGREGAÇÃO DE POLIPÉPTIDOS OU SOLUBILIDADE**

(57) Resumo:

**RESUMO****"MÉTODO COMPUTACIONAL E APARELHO PARA PREVER A AGREGAÇÃO DE POLIPÉPTIDOS OU SOLIBILIDADE"**

Um método preditivo para prever o efeito de uma modificação de aminoácido na velocidade de agregação (solubilidade) de um polipéptido de referência compreendendo: calcular a diferença de hidrofobicidade ( $\Delta\text{Hidr}$ ) entre o polipéptido de referência e um polipéptido modificado, calcular a diferença na propensão para folha- $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ) entre o polipéptido de referência e o polipéptido modificado, calcular a diferença de carga ( $\Delta\text{Carga}$ ) entre o polipéptido de referência e o polipéptido modificado e calcular:  $[x \cdot \Delta\text{Hidr}] + [y \cdot (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z \cdot \Delta\text{Carga}]$ , em que x, y e z são factores de escala.

## **DESCRIÇÃO**

### **Área Técnica**

Esta invenção relaciona-se com métodos para determinar o efeito de uma modificação de aminoácido na velocidade de agregação de um polipéptido, calculando a propensão de um polipéptido modificado para agregar, *i.e.* a solubilidade de um polipéptido modificado relativamente a um polipéptido de referência. A invenção relaciona-se ainda com um método para design de um polipéptido modificado com uma capacidade especial para agregar, ou seja uma solubilidade especial. A invenção relaciona-se ainda com métodos deste tipo realizados por meio de software de computador e hardware de computador programados para levar a cabo os métodos.

### **Antecedentes da Invenção**

A compreensão dos efeitos de modificações de péptidos e proteínas, tais como substituições de aminoácidos, nas propensões de polipéptidos específicos para (se) agregar é de importância crucial para elucidar a base molecular de doenças de deposição de proteínas, tais como doença de Alzheimer e outras doenças amilóides e para compreender os mecanismos de acção das mutações associadas com formas hereditárias destas doenças.

Em cada uma das várias condições patológicas

associadas com a deposição de proteínas e péptidos, um péptido ou proteína específico que é normalmente solúvel é depositado, ou intacto ou numa forma fragmentada, em agregados insolúveis que se acumulam em um ou mais tipo(s) de tecido. Descobriu-se que numerosas mutações estavam associadas com formas familiares de doenças de deposição de proteínas e mostrou-se que mais de 100 envolvem directamente a sequência do péptido ou proteína responsável pela agregação (Siepen e Westhead, 2002). Muitas destas mutações foram identificadas nos últimos 5 anos, e espera-se que o número aumente de forma dramática no futuro próximo. Investigação dos mecanismos pelos quais mutações naturais resultam em comportamento patológico provou ser de importância fundamental para explorar a base molecular da doença subjacente, mesmo nos casos em que são esporádicos e não de origem familiar (Selkoe, 2001; Volles & Lansbury, 2002).

Descobriu-se que a capacidade de formar agregados altamente organizados, possuindo características estruturais comuns, tais como amilóide, é uma propriedade genérica de polipéptidos, independentemente da sequência ou semelhança estrutural, e não simplesmente uma característica de um pequeno número de proteínas associadas com condições patológicas reconhecidas (Dobson, 2001).

No estado nativo, os resíduos hidrofóbicos estão normalmente encastrados no cerne da proteína, sendo a assim limitada a oportunidade destes resíduos interagirem. No

entanto, as proteínas são dinâmicas e existe um equilíbrio entre a conformação dobrada, estável e os estados parcialmente ou completamente desdobrados, destabilizados. O valor de energia livre ( $\Delta G$ ,  $\text{kJ mol}^{-1}$ ) para uma proteína fornece uma indicação da estabilidade da proteína. Agregação ocorre quando proteínas no seu estado nativo, desnaturam; à medida que a proteína desdobra, ligações intramoleculares são quebradas, permitindo que a cadeia polipeptídica principal (esqueleto) e as cadeias laterais hidrofóbicas sejam expostas. Podem então formar-se ligações de hidrogénio e outras interacções entre moléculas de proteína parcial ou completamente desnaturadas, resultando em associações intermoleculares e formação de agregados.

Nalguns casos pode ser desejável formar agregados, em particular fibrilhas, por exemplo para uso como materiais plásticos, em electrónica, como condutores, para catálise ou como uma forma de libertação lenta do polipéptido, ou quando fibrilhas de polipéptido são fiadas num "fio" polipeptídico para várias aplicações; por exemplo, como descrito em pedidos de patente publicadas WO0017328 (Bobson) e WO0242321 (Dobson &McPhee).

No entanto, noutras circunstâncias a formação de agregados é desvantajosa, por exemplo, quando se deseja usar um péptido em concentrações ou sob condições desejáveis para actividade fisiológica, administração terapêutica ou aplicação industrial. Em particular, o uso de péptidos bioactivos e proteínas como agentes terapêuticos é limitada

quando o péptido ou a proteína tende a formar agregados durante a manufactura, processamento, armazenamento ou administração seguinte. Estas questões são geralmente reconhecidas na indústria biotecnológica e farmacêutica e constitui um problema de vulto e um fardo económico, que pode ser difícil de ultrapassar e pode requerer o uso de técnicas de expressão e de enrolamento sofisticadas, o desenvolvimento de formulações específicas, agentes estabilizadores e excipientes, distribuição em cadeia de frio ou reconstituição imediata antes de uso. Quase todos os produtos terapêuticos polipeptídicos conhecidos apresentam estes problemas, e.g. insulina, interferon- $\gamma$ , BMPs, calcitonina, glucagon, anticorpos.

Sabe-se que vários factores afectam e tendência de um polipéptido para agregar. Alguns destes factores são locais a resíduos de aminoácidos, outros factores são globais e podem afectar a proteína inteira. Por exemplo quando ocorrem mutações num polipéptido, factores locais na região da mutação tais como hidrofobicidade aumentada ou tendência para se converter da conformação de hélice- $\alpha$  à conformação de folha- $\beta$ , resulta numa velocidade de agregação mais elevada do que a da proteína tipo selvagem (wild type) (não-mutada). Modificações "globais" ou totais devido a mutações também podem afectar a velocidade de agregação; por exemplo, uma modificação na carga do péptido mutante tornando-o mais próximo da neutralidade, resulta numa tendência aumentada de um polipéptido para agragar. Mutações que destabilizam o estado nativo do polipéptido também resultam numa agregação facilitada.

Um estudo mutacional detalhado numa proteína modelo, acilfosfatase (AcP) do músculo, demonstraram que a velocidade de agregação de um ensemble de conformações parcialmente desnaturadas pode ser facilmente seguida para AcP usando uma variedade de sondas espectroscópicas. A velocidade de agregação foi determinada para mais de 50 variantes mutacionais desta proteína (Chiti et al., 2002a; 2002 b: 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 (2002 a); 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 (2002 b)). Descobriu-se que muitas destas mutações, particularmente as que envolvem os resíduos 16-31 e 87-98, perturbam a velocidade de agregação de AcP muito significativamente (Chiti et al., 2002 a; 2002 b). Chiti (2002 a) concluiu que as modificações medidas na velocidade de agregação sobre mutação correlacionavam positivamente com as modificações na hidrofobicidade e propensão para folha- $\beta$  das regiões da proteína nas quais estão localizadas as mutações. Chiti (2002 b) examinou mutações na AcP que alteravam o estado de carga da proteína AcP sem afectar significativamente a hidrofobicidade ou propensões da estrutura secundária da cadeia polipeptídica. Foi relatada uma correlação inversa entre a velocidade de agregação de variantes da proteína em condições desnaturantes e a carga total útil da proteína.

Os factores que afectam a velocidade de agregação de uma proteína são diversos. Quando são feitas substituições de aminoácidos numa proteína, estão envolvidos vários factores em extensões diferentes. Uma única mutação pode aumentar a carga útil, desfavorecendo assim a agregação (por exemplo a substituição de Ala por Asp numa proteína carregada positivamente). No entanto a mesma mutação pode aumentar a hidrofobicidade, trazendo assim uma contribuição aceleração para a velocidade de agregação. Finalmente, a mesma mutação também altera as propensões para hélice- $\alpha$  e para formação de folha- $\beta$  de uma cadeia polipeptídica, introduzindo outros factores. A relação entre os factores e a sua importância relativa para a agregação (solubilidade) não estão bem caracterizados.

Assim, não tem sido possível prever com precisão a tendência de uma proteína para formar agregados insolúveis e ordenados; tais como fibrilhas amilóides, nem predizer ou calcular o efeito de modificações de aminoácidos específicas, tais como substituições na agregação/solubilidade. A incapacidade de fazer estes prognósticos ou calculus constitui um problema na concepção e/ou manuseamento de péptidos *in vivo* ou *in vitro*.

A capacidade de prognosticar os efeitos intrínsecos de mutações na agregação de polipéptidos é de importância crucial para elucidar o efeito patogénico de um grande número de mutações associadas doenças de deposição

de proteínas. Seria desejável estabelecer, em casos específicos, se uma dada mutação daria origem à doença como resultado directo do seu efeito no processo de agregação do polipéptido envolvido, ou por outros mecanismos. Além disso, o estabelecimento de princípios gerais na agregação tornaria possível o uso de métodos estatísticos para analisar a relação entre mutação, agregação e doença. A compreensão dos efeitos das substituições de aminoácidos nas propensões de proteínas específicas para se agregar, permitiria o estabelecimento de critérios para modificar racionalmente as propriedades de agregação de péptidos e proteínas naturais ou concebidos (designed) para processos industriais, propósitos de investigação, tratamento médico ou aplicação biotecnológica.

Além disso, métodos da invenção podem ser usados para identificar ou desenhar sequências polipeptídicas com uma propensão para agregação reduzida, polipéptidos reconcebidos poderiam ser administrados por métodos tais como terapia genética para tratar certas desordens, particularmente as associadas com agregação de proteínas. A capacidade de identificar ou desenhar polipéptidos com propriedades de agregação específicas será importante para o desenvolvimento e manufactura de polipéptidos para aplicação na áreas de materiais e dispositivos, tais como os descritos em W00017328 (Dobson) e W00242321 (Dobson &McPhee).

Artigos com o título "Computer-aided design of  $\beta$ -sheet peptides" por Lopez de Paz et al, J Mol Biol (2001)

312, 229-246 XP24466132 descreve testes de validade de métodos de design automático para criar e/ou melhorar péptidos de folha- $\beta$ . Um utensílio de design automático usado é PERLA (protein engineering rotamer library algorithm). PERLA é também usado em "De novo designed peptide-based amyloid fibrils" por Lopez de Paz et al, Proceedings of the national academy of sciences of the United States, vol 99, no 25, 10 Dez 2002, páginas 16052-16057 XP2279769 que discute as interacções que conduzem à formação de folhas- $\beta$  e as forças intermoleculares que levam à formação de fibrilhas amilóides. Também está descrito em EP 1255209 a concepção de proteínas de novo e a concepção de proteínas ajudado por computador.

A formação de fibrilhas amilóides também é discutida no artigo com o título "The fibril one on-line database: Mutations, experimental conditions, and trends associated with amyloid fibril formation" por Siepen et al em Protein Science, vol 11, no 7, Julho 2002, pag. 1862-1866 XP2279770. Os factores que afectam as preferências de aminoácidos para conformações de hélice  $\alpha$ , cadeia- $\beta$  e outros estados conformacionais da cadeia principal em proteínas está discutido em "Amino acid conformational preferences and salvation of polar cackbone atoms in peptides and proteins" por Avbel et al, em Jornal of Molecular Biology, London vol 300, no 5, Julho 2000, pag. 13335-1359 XP4469107 e "Role of main-chain electrostatics, hydrophobic effect and side-chain conformational entropy in determining the secondary structure of proteins" por Avbel

et al, em Journal of Molecular Biology, London vol 279, no 3, 12 Junho 1998, paginas 665-684 XP 4453960.

Seria por isso útil poder prognosticar se um determinado polipéptido formaria agregados insolúveis e prognosticar o efeito que uma determinada modificação ou modificações da sequência de aminoácidos teria nas propriedades de agregação/solubilidade de um polipéptido.

#### Revelação da invenção

A invenção corresponde ao especificado nas reivindicações independentes. Nós descrevemos um método preditivo para prognosticar o efeito de uma modificação de aminoácido na velocidade de agregação (solubilidade) de um polipéptido de referência, cujo método compreende calcular a diferença na hidrofobicidade ( $\Delta\text{Hidr}$ ) entre o polipéptido de referência e um polipéptido modificado, calculando a diferença na propensão para formação de folha- $\beta$  ( $\Delta\Delta_{\text{Gespiral-}\alpha} + \Delta\Delta_{\text{G}\beta\text{-espiral}}$ ) entre o polipéptido de referência e o polipéptido modificado, calculando a diferença de carga ( $\Delta\text{Carga}$ ) entre o polipéptido de referência e o polipéptido modificado e calculando :  $[x*\Delta\text{Hidr}] + [y*(\Delta\Delta_{\text{Gespiral-}\alpha} + \Delta\Delta_{\text{G}\beta\text{-espiral}})] - [z*\Delta\text{Carga}]$ , em que x, y e z são factores de escala. Um valor para  $[x*\Delta\text{Hidr}] + [y*(\Delta\Delta_{\text{Gespiral-}\alpha} + \Delta\Delta_{\text{G}\beta\text{-espiral}})] - [z*\Delta\text{Carga}]$  de um primeiro sinal indica que o polipéptido modificado tem uma maior propensão para agregar relativamente ao polipéptido de referência e um valor para  $[x*\Delta\text{Hidr}] + [y*(\Delta\Delta_{\text{Gespiral-}\alpha} + \Delta\Delta_{\text{G}\beta\text{-espiral}})] - [z*\Delta\text{Carga}]$  de um

sinal oposto ao primeiro sinal indica que o polipéptido modificado tem uma propensão reduzida para agregar relativamente ao polipéptido de referência. Se o efeito da modificação de aminoácido na velocidade de agregação for expressa como  $\ln(v_{\text{mod}}/v_{\text{ref}})$ , um valor positivo para  $\ln(v_{\text{mod}}/v_{\text{ref}})$  indica que o polipéptido modificado tem uma maior propensão para agregar (menor solubilidade) em relação ao polipéptido de referência; e um valor negativo para  $\ln(v_{\text{mod}}/v_{\text{ref}})$  indica que o polipéptido modificado tem uma propensão reduzida para agregar (maior solubilidade em relação ao polipéptido de referência).

Num método preferido o factor de escala x é um valor de 0,59 a 0,64, o valor de escala y é um valor de 0,19 a 0,22 e o factor de escala z é um valor de 0,49 a 0,51. É particularmente preferido que o factor de escala x seja 0,6 , 0,63 ou 0,633 , o factor de escala y seja 0,2 ou 0,198 e o factor de escala z seja 0,5 , 0,49 ou 0,491.

Também descrevemos um método de identificação para identificar uma modificação de aminoácido que reduz a velocidade de agregação (aumenta a solubilidade) de um polipéptido de referência compreendendo o uso de um método da invenção para predizer a modificação na velocidade de agregação para um ou mais polipéptido(s) modificados, tendo o polipéptido modificado uma ou mais modificações de aminoácidos em comparação com o polipéptido de referência, comparar as velocidades de agregação previstas do polipéptido de referência e um ou mais polipéptidos

modificados, e identificar um ou mais polipéptido(s) modificados tendo uma velocidade de agregação prevista reduzida relativamente ao polipéptido de referência. Também descrevemos um método de previsão como descrito aqui anteriormente para uso no método de identificação ou na preparação de um polipéptido modificado.

Também descrevemos um método de identificação para identificar uma modificação de aminoácido que aumenta a velocidade de agregação (diminui a solubilidade) de um polipéptido que compreende o uso de um método da invenção para prever a modificação na velocidade de agregação para um ou mais polipéptido(s) modificado(s), possuindo cada polipéptido modificado uma ou mais modificações de aminoácidos quando comparado com o polipéptido de referência, comparar as velocidades de agregação previstas do polipéptido de referência e um ou mais polipéptidos(s) modificado(s) e identificar um ou mais polipéptidos modificados possuindo uma velocidade de agregação prevista aumentada relativamente ao polipéptido de referência.

Também é descrito um método para produzir um polipéptido possuindo uma velocidade de agregação reduzida (solubilidade aumentada) compreendendo o uso de um método da invenção para identificar uma modificação que se prevê que reduza a velocidade de agregação de um polipéptido e produzir um polipéptido modificado contendo a referida modificação.

É descrito um método para produzir um polipéptido possuindo uma velocidade de agregação aumentada (solubilidade diminuída) que compreende o uso de um método da invenção para identificar uma modificação que se prevê que aumente a velocidade de agregação de um polipéptido e produzir um polipéptido modificado contendo a referida modificação.

Num aspecto da invenção o polipéptido de referência e o(s) polipéptido(s) modificado(s) são estruturalmente relacionados em termos de composição de aminoácidos e sequência. Polipéptidos estruturalmente relacionados têm pelo menos 60%, preferivelmente pelo menos 70%, mais preferivelmente pelo menos 80%, ainda mais preferivelmente pelo menos 90%, e ainda mais preferivelmente pelo menos 95% de homologia de sequência de aminoácidos. Num aspecto alternativo, o polipéptido de referência e o(s) polipéptido(s) modificado são estruturalmente não-relacionados. Preferivelmente o polipéptido de referência é um polipéptido tipo selvagem e o polipéptido modificado é um mutante deste. Um polipéptido modificado é preparável por modificação química e/ou por modificações tais como substituição, deleção e/ou adição de um ou mais aminoácidos da proteína de referência. Um aminoácido substituído ou adicionado à proteína de referência pode ser um aminoácido natural ou um aminoácido quimicamente sintetizado ou quimicamente modificado. Preferivelmente 1 a 20, 1 a 16, 1 a 12, ou 1 a 10 aminoácidos são substituídos, deletados e/ou adicionados; mais preferivelmente 1, 2, 3,

4, 5 ou 6 aminoácidos são substituídos, delectados e/ou adicionados. Preferivelmente a modificação do polipéptido é por substituição de aminoácidos, que pode ser substituição de um ou mais aminoácidos, preferivelmente por substituição de 1 a 20, 1 a 16, 1 a 12, ou 1 a 10 aminoácidos, mais preferivelmente substituição de 1, 2, 3, 4, 5 ou 6 aminoácidos. Alternativamente, a modificação do polipéptido pode ser por delecção de um ou mais aminoácidos, preferivelmente por delecção de 1 a 20, 1 a 16, 1 a 12 ou 1 a 10 aminoácidos; mais preferivelmente por delecção de 1, 2, 3, 4, 5 ou 6 aminoácidos.

O polipéptido de referência pode ser um polipéptido natural de qualquer espécie, ou um polipéptido não-natural "concebido". Modificações propostas do polipéptido de referência podem ser seleccionadas comparando a sequência de aminoácidos de um polipéptido de referência com a sequência de aminoácidos de um polipéptido relacionado ou polipéptidos de espécies diferentes. É preferível que a modificação seja feita em regiões de um polipéptido que sejam polimórficas entre espécies diferentes. Modificações propostas de uma proteína de referência também podem ser seleccionadas comparando a sequência de aminoácidos da proteína de referência com variantes polipeptídicas da proteína de referência que ocorrem naturalmente ou são induzidas, preferivelmente comparando polipéptidos tipo selvagem (wild type) e polipéptidos mutantes.

O polipéptido de referência é preferivelmente um

polipéptido humano. Modificações propostas de um polipéptido de referência humano podem ser seleccionadas comparando um polipéptido de referência humano com um polipéptido relacionado de uma fonte não-humana. A modificação pode ser tal que um aminoácido na forma humana de um polipéptido é modificado para o aminoácido presente naquela posição num polipéptido relacionado de uma fonte não-humana. Alternativamente, o polipéptido de referência pode ser um polipéptido não-humano e a modificação pode ser tal que o polipéptido não-humano é "humanizado".

Para polipéptidos bioactivos, é preferível que a modificação seja tal que uma actividade do polipéptido seja mantida ou melhorada. A modificação pode ser fora do sítio activo do polipéptido ou pode ser no interior do sítio activo do polipéptido.

O termo polipéptido, como usado aqui, abrange proteínas e péptidos.

Usando os métodos da invenção, os efeitos intrínsecos de modificações específicas, tais como mutações, na velocidade de agregação de polipéptidos podem ser racionalizados e previstos a uma medida notável na base de princípios físicos simples: os efeitos que as modificações têm nos parâmetros fundamentais de hidrofobicidade e propensão da estrutura secundária no sítio da modificação e na carga da molécula como um todo. Baseado neste cálculo, podem ser concebidos (designed) polipéptidos modificados

(e.g. mutantes) que são mais ou menos susceptíveis de se agregar (que têm uma menor ou maior solubilidade) do que o polipéptido de referência (e.g. tipo selvagem-wild type), ou que têm uma propensão para agregar no intervalo desejado. Assim é possível avaliar os efeitos que as várias modificações de aminoácidos terão nas propriedades de um polipéptido sem ter que produzir polipéptidos modificados e medir experimentalmente o efeito das modificações. É praticável a concepção de um enorme número de polipéptidos modificados, mesmo para um polipéptido de referência relativamente curto. Isto é importante porque as modificações podem ser seleccionadas também para preencher outros critérios ou restrições, tais como estabilidade da proteína, função, etc.

A alteração da velocidade de agregação como resultado de uma modificação (e.g. mutação) pode ser expressa como  $\ln(v_{\text{mod}}/v_{\text{ref}})$ , e.g.  $(\ln(v_{\text{mut}}/v_{\text{tiposelvagem}}))$ . A velocidade de agregação também pode ser expressa em termos de um tempo de agregação, tal como meio tempo de agregação e a razão de velocidade de agregação é equivalente a uma razão de tempos de agregação tal como adicionalmente discutido mais adiante.

Os parâmetros cinéticos para agregação podem ser velocidades de agregação correspondendo a cinética exponencial de agregação ( $v$ ) ou, quando se usa um parâmetro de tempo ( $\tau$ ),  $(\tau_{\text{mod}}/\tau_{\text{ref}}) = (v_{\text{ref}}/v_{\text{mod}}) = 1/\exp(\ln(v_{\text{mod}}/v_{\text{ref}}))$ , pode incluir ou referir exclusivamente a estádios de

nucleação e/ou "fases de latência" da cinética de agregação (T1) ou ao meio-tempo de agregação do processo global (T2).

A modificação da hidrofobicidade ( $\Delta H_{idr}$ ) é calculada usando  $\Delta H_{idr} = H_{idr_{ref}} - H_{idr_{mod}}$ , em que  $\Delta H_{idr}$  é a modificação da hidrofobicidade que resultaria de uma modificação de aminoácido proposta,  $H_{idr_{ref}}$  e  $H_{idr_{mod}}$  são os valores de hidrofobicidade dos resíduos de aminoácidos de referência e modificados, respectivamente.

Uma escala de hidrofobicidade consensus pode ser usada para atribuir um valor de hidrofobicidade para cada aminoácido. Escalas de hidrofobicidade diferentes podem ser usadas para valores de pH diferentes, por exemplo escalas descritas em Cowan, R. & Whittaker, R.G. (1990) *Peptide Research* 3:75-80) podem ser usadas para calcular a hidrofobicidade de polipéptidos a baixo pH. Pode usar-se uma escala de hidrofobicidade média, que pode ser obtida usando uma combinação de escalas, tais como as disponíveis na literatura (e.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). Numa forma de realização preferida, o valor de hidrofobicidade para cada aminoácido é atribuído usando os valores indicados na Tabela 1 para hidrofobicidade de 20 resíduos de aminoácidos a pH neutro, baseado nos coeficientes de partição de água para octanol; os dados são da coluna 6 da tabela 4.8 em Creighton (1993) (Creighton, T.E. *Em Proteins. Structure and molecular properties. Second edition.* W.H. Freeman & Company (New

York,1993),p.154.).

O factor de escala  $x$  pode ser derivado através do gráfico o  $\ln((v_{\text{mod}} / v_{\text{ref}})$  e.g.  $\ln \ln(v_{\text{mut}}/v_{\text{tiposelvagem}})$  versus  $\Delta H_{\text{idr}}$  para um número de sequências polipeptídicas, que podem ser ou não relacionadas (e.g. uma colecção de mutantes de AcP) e determinando a linha de melhor ajuste, o factor de escala  $x$  é o declive (gradiente) da linha. Para calcular a propensão de converter da forma de hélice- $\alpha$  a estrutura folha  $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ), os valores individuais para  $\Delta\Delta G_{\text{espiral-}\alpha}$  e  $\Delta\Delta G_{\beta\text{-espiral}}$  são calculados.

$\Delta\Delta G_{\text{espiral-}\alpha}$  pode ser calculado usando  $\Delta\Delta G_{\text{espiral-}\alpha} = RT \ln(P_{\alpha}^{\text{ref}}/P_{\alpha}^{\text{mod}})$ , em que  $\Delta\Delta G_{\text{espiral-}\alpha}$  é alteração prevista de energia livre para a transição de hélice- $\alpha$  para espiral aleatória que resulta de modificação;  $R=0,008314$  kJ mol<sup>-1</sup>K<sup>-1</sup>,  $P_{\alpha}^{\text{ref}}$  e  $P_{\alpha}^{\text{mod}}$  são respectivamente as propensões para hélice- $\alpha$  (percentagens de hélice) das sequências de referência (p.ex tipo selvagem) e modificada (e.g.mutante) no sítio de modificação de aminoácido, respectivamente. As propensões de hélice- $\alpha$  previstas podem ser calculadas usando software/algoritmos modelo tais como AGADIR ([www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html](http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html)) Munoz & Serrano (1994) Nature Structural Biol 1, 399-409; Muñoz & Serrano (1994) J Mol Biol 245,297-308; Muñoz & Serrano (1997) Biopolymers 41 495 509 e 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). Quaisquer logaritmos apropriados baseados em bases de dados estruturais, bases de dados de preferência estrutural ou bases de dados de preferência rotamer poderiam ser usadas para este cálculo para estimar as propensões de hélice, por exemplo GOR IV: J.Garnier. J.F. Gibrat e B. Robson em Methods Enzymol., vol266, p 540-533 (1996). Garnier, D.Osguthorpe e B.Robson (J.Mol.Biol.120, 97, 1978). J Mol Biol 1987 Dec 5; 198(3): 425-443 (GOR-III); PHD: Rost B, Sander C. J Mol Biol 1993 Jul 20; 232(2): 584-99. Rost B, Sander C. Proteins 1994 May;19(1): 55-72; PREDATOR Frishman D, Argos P. Protein Eng 1996 Feb; 9(2): 133-142; SIMPA/SIMPA96: Levin JM, Robson B, Garnier J. FEBS Lett 1986 Sep 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, eleage G. Protein Eng 1994 Feb; 7(2): 157-164. Geourjon C, Deleage G. Comput Appl Biosci 1995 Dec; 11(6):681-684.

$\Delta\Delta G_{\beta\text{-espiral}}$  pode ser calculado usando  $\Delta\Delta G_{\beta\text{-espiral}} = 13,64 (P_{\beta}^{\text{ref}} - P_{\beta}^{\text{mod}})$ , em que  $\Delta\Delta G_{\beta\text{-espiral}}$  é a variação de energia livre para a transição de espiral aleatória (random espiral) para folha- $\beta$  que resulta da modificação ( $\Delta G_{\beta\text{-espiral}}$ ), 13,64 é a constante de conversão da escala normalizada para unidades de  $\text{kJ mol}^{-1}$ ;  $P_{\beta}^{\text{ref}}$  e  $P_{\beta}^{\text{mod}}$  são as propensões a formação de folha- $\beta$  normalizadas do resíduo de referência (e.g. tipo selvagem) e modificado (e.g. mutante) respectivamente. Valores de propensão a formação de folha- $\beta$

para todos os 20 aminoácidos pode ser determinada usando uma escala publicada. Uma escala preferida apresenta-se na Tabela 1, que fornece valores de propensão a formação de folha- $\beta$  para 19 resíduos de aminoácidos (todos excepto prolina), estes estão normalizados de 0 (propensão elevada a formação de folha- $\beta$ ) a 1 (baixa propensão a formação de folha- $\beta$ ). Estes dados são da coluna 4 da Tabela 1 de Street e Mayo (1999) (Street, A.G. & Mayo, S.L. Intrinsic  $\beta$ -sheet propensities result from van der Waals interactions between side chains and the local backbone. Proc.Natl.Acad.Sci.USA, 96, 9074-9076 (1999). A propensão para folha- $\beta$  da prolina não está relatada devido à dificuldades na sua determinação experimental. A propensão a formação folha- $\beta$  da glicina é obtida a partir de cálculos teóricos.

O factor de escala  $y$  pode ser derivado através do gráfico  $\ln (v_{\text{mod}}/v_{\text{ref}})$  observado versus  $((\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}))$  para um número de sequências polipeptídicas não-relacionadas ou relacionadas (por exemplo através do gráfico  $\ln (v_{\text{mut}}/v_{\text{ts}})$  versus  $(\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})$  para um número de mutações polipeptídicas de AcP e determinar a linha de melhor ajuste, o factor de escala  $y$  é o declive (gradiente) da linha.

Alteração da carga ( $\Delta$  Carga) do polipéptido é calculada usando:

$\Delta\text{Carga} = |\text{Carga}_{\text{mod}}| - |\text{Carga}_{\text{ref}}|$  em que  $\Delta\text{Carga}$  é a alteração de carga resultante da mutação  $|\text{Carga}_{\text{mod}}|$  e  $|\text{Carga}_{\text{ref}}|$

são os valores absolutos da carga para as sequências modificada (e.g. mutante) e referência (e.g. tipo selvagem) respectivamente (obtidas das somas dos valores de carga para resíduos de aminoácidos (e.g. os valores dados na Tabela 1, que correspondem a pH neutro). O operador de "valor absoluto" é introduzido de modo que um valor negativo de  $\Delta\text{Carga}$  resulta da equação quando a mutação provoca que toda a proteína ou péptido se aproximem da neutralidade, independentemente do sinal inicial da sequência da proteína. Um valor positivo de  $\Delta\text{Carga}$  é obtido quando a mutação provoca que toda a sequência da proteína se desvie mais da neutralidade.

O factor de escala  $z$  pode ser derivado através do gráfico  $\ln(v_{\text{mod}}/v_{\text{ref}})$  observado versus  $\Delta\text{Carga}$  para um número de sequências polipeptídicas não-relacionadas ou relacionadas, p.ex através do gráfico  $\ln(v_{\text{mut}}/v_{\text{ts}})$  observado para um número de mutantes AcP e determinando a linha de melhor ajuste; o factor de escala  $z$  é o declive (gradiente) da linha.

Os Factores de escala  $x$ ,  $y$ , e  $z$  podem ser refinados utilizando, nos gráficos, um maior número de pontos correspondentes a dados para modificações polipeptídicas adicionais. Pode usar-se análise de regressão múltipla para determinar os factores de escala  $x$  para  $\Delta\text{Hidr}$ ,  $y$  para  $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ , e  $z$  para  $\Delta\text{Carga}$ . O método de análise de regressão múltipla consiste em calcular simultâneamente os factores de escala individuais

$x$ ,  $y$ , e  $z$  para um determinado grupo de polipéptidos (e.g. polipéptidos mutantes) para os quais estão disponíveis os valores experimentais de  $\ln(v_{\text{mod}}/v_{\text{ref}})$  (e.g.  $\ln(v_{\text{mut}}/v_{\text{ts}})$ ). Podem ser usados nos cálculos estimativas iniciais para os valores de  $x$ ,  $y$  e  $z$ , tais como os valores aqui descritos, ou valores escolhidos numa base arbitrária. Os valores calculados de  $\ln(v_{\text{mod}}/v_{\text{ref}})$  são comparados com os valores obtidos experimentalmente. Os cálculos podem ser repetidos através de um número de iterações, usando de cada vez uma combinação diferente dos valores  $x$ ,  $y$  e  $z$ . As iterações terminarão quando for encontrado um conjunto de melhores valores para  $x$ ,  $y$  e  $z$ , i.e. quando for encontrada a diferença mínima entre o valor calculado e o valor experimental de  $\ln(v_{\text{mod}}/v_{\text{ref}})$ .

O cálculo pode ser modificado por inclusão de um outro termo ou termos, tal como o factor de estabilidade para corrigir para mutações que afectam a estabilidade da proteína. Podem ser adicionados termos à equação para representar outros factores que afectam a velocidade de agregação, tal como a posição da mutação ao longo da sequência, efeitos não-cumulativos de substituições duplas, distância na sequência entre substituições duplas. Refinamentos da equação por inclusão de factores adicionais pode melhorar a exactidão do método. Em conformidade, os métodos da invenção podem compreender adicionalmente um passo ou passos de cálculo adicional, por exemplo um passo adicional no qual é usado um factor de estabilidade para corrigir para mutações que afectam a estabilidade da proteína.

Também descrevemos um programa de computador codificado para, quando em funcionamento, predizer o efeito de uma modificação de aminoácido na velocidade de agregação de um polipéptido, o código compreendendo código para: introduzir (input) uma modificação de aminoácido que converteria o polipéptido de referência a uma forma modificada do polipéptido; calcular a diferença em hidrofobicidade ( $\Delta\text{Hidr}$ ) entre o polipéptido de referência e o polopéptido modificado; calcular a diferença na propensão para conformação em folha  $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ) entre o polipéptido de referência e o polipéptido modificado; calcular a diferença de carga ( $\Delta$  Carga) entre o polipéptido de referência e o polipéptido modificado; e calcular:  $[x * \Delta\text{Hidr}] + [y * (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z * \Delta$  Carga], em que  $x$ ,  $y$  e  $z$  são factores de escala.

Também descrevemos um programa de computador codificado para, em funcionamento, identificar um polipéptido, compreendendo o código: código para prever a alteração na velocidade de agregação para um ou mais polipéptidos modificados, compreendendo código como descrito acima; e código adicional para identificar um ou mais dos referidos polipéptidos modificados dependendo da referida alteração prevista na velocidade de agregação.

Também descrevemos um sistema informático para determinar a velocidade de agregação de um segundo polipéptido em relação a um polipéptido de referência, pos-

suindo o referido segundo polipéptido e o referido polipéptido de referência cada um uma sequência de aminoácido, compreendendo a sequência de aminoácidos do referido segundo polipéptido uma versão modificada da sequência de aminoácidos do referido polipéptido de referência, compreendendo o sistema de computador um armazenamento de dados para armazenar dados compreendendo dados de hidrofobicidade, dados de propensão a formação de folha- $\beta$  e dados de carga para um conjunto de aminoácidos; uma área de armazenamento de programas que armazena um código implementável por processador; e um processador, acoplado à referida área de armazenamento de programas (program store) e ao referido armazenamento de dados (data store) para implementar o referido código armazenado, compreendendo o código um código para controlar o processador para : introduzir uma sequência de aminoácidos para o referido segundo polipéptido; ler os dados de hidrofobicidade para a sequência de aminoácidos do referido segundo polipéptido do referido armazenamento de dados e determinar um valor de hidrofobicidade para o referido segundo polipéptido; ler a alteração dos dados para a sequência de aminoácidos do referido segundo polipéptido do referido armazenamento de dados e determinar um valor de propensão para formação de folha- $\beta$  para o referido segundo polipéptido; determinar um valor de propensão para formação de hélice- $\alpha$  para o referido segundo polipéptido; obter dados de hidrofobicidade, dados de carga, dados de propensão para formação de folha- $\beta$  e um valor de propensão para formação de hélice- $\alpha$  para o referido polipéptido de referência; e determinar a referida velocidade de agregação relativa usando a referida hidro-

fobicidade, alteração, dados de propensão para formação de folha- $\beta$  e hélice- $\alpha$  para os referidos segundo polipéptido e polipéptido de referência, avaliando  $x\Delta H+y\Delta G-z\Delta C$  em que  $x$ ,  $y$  e  $z$  são factores de escala,  $\Delta H$  representa a diferença de hidrofobicidade entre o referido segundo polipéptido e o referido polipéptido de referência,  $\Delta G$  representa a diferença entre o referido segundo polipéptido e o referido polipéptido de referência na propensão de converter da estrutura de hélice- $\alpha$  para uma estrutura de folha- $\beta$  e  $\Delta C$  representa a diferença de carga entre o referido segundo polipéptido e o referido polipéptido de referência e em que um valor para  $x\Delta H+y\Delta G-z\Delta C$  de um primeiro sinal indica que o segundo polipéptido tem uma maior propensão para agregar relativamente ao polipéptido de referência e um valor de  $x\Delta H+y\Delta G-z\Delta C$  de sinal oposto ao primeiro sinal indica que o segundo polipéptido tem uma propensão reduzida para agregar relativamente ao polipéptido de referência. Os termos "segundo polipéptido" e "polipéptido modificado" são usados de forma permutável.

Numa forma de realização a propensão para formação de folha- $\beta$  pode ser expressa em termos de energia livre. O conjunto de aminoácidos pode compreender, por exemplo, todos os resíduos de aminoácidos naturais. As propensões para formação de hélice- $\alpha$  das sequências polipeptídicas de referência e modificado podem ser determinadas usando um código no interior do sistema de computador ou pode mandado um pedido para um sistema de computador separado, por exemplo numa rede (network), e os valores de propensão para formação de hélice- $\alpha$  recebidos

deste segundo sistema de computadores. Os valores de hidrofobicidade, valores de carga e valores de propensão estrutural para os polipéptidos de referência e modificado podem ser determinados e a partir de escalas de valores para cada aminoácido que foram anteriormente publicados e introduzidos (input) no sistema informático e/ou armazenado por exemplo no armazenamento de dados (data store). A velocidade relativa de agregação determinada pelo aparelho pode simplesmente abranger a determinação de se as velocidades de agregação (ou, noutras formas de realização, a solubilidade) do segundo polipéptido está acima ou abaixo da do polipéptido de referência; alternativamente, pode ser feita uma determinação quantitativa das velocidades de agregação relativas destes polipéptidos. Um ou mais dos factores de escala podem ser seleccionados sensíveis a, por exemplo, o tipo de polipéptido para qual está a ser feita a determinação.

O sistema de computador pode ser usado para avaliar a pluralidade de segundos péptidos, apresentando os resultados, por exemplo como uma lista ordenada ou gráfico e/ou seleccionando uma sequência candidata promissora para a síntese e avaliação adicional.

Também descrevemos um método para determinar um indicador de velocidade de agregação relativa, predizendo o referido indicador da velocidade de agregação relativa, a velocidade de agregação de um segundo polipéptido em comparação com a velocidade de agregação de um polipéptido de referência, compreendendo o referido segundo polipéptido

uma ou mais modificações de aminoácidos quando comparada com o polipéptido de referência, compreendendo o método:

determinar a diferença de hidrofobicidade ( $\Delta H_{idr}$ ) entre o referido segundo polipéptido e o referido polipéptido de referência; determinar uma diferença na diferença na propensão para formação de folha- $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\text{espiral-}\beta}$ ) entre o referido segundo polipéptido e o referido polipéptido de referência; determinar uma diferença na carga ( $\Delta \text{Carga}$ ) entre o referido segundo péptido e o referido polipéptido de referência; formar uma combinação ponderada da referida diferença em hidrofobilidade, referida diferença na propensão da estrutura secundária, referida diferença na carga para determinar o referido indicador de velocidade de agregação relativa calculando  $[x*\Delta H_{idr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\text{espiral-}\beta})] - [z*\Delta \text{Carga}]$ , em que x, y e z são factores de escala e em que um valor para  $[x*\Delta H_{idr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\text{espiral-}\beta})] - [z*\Delta \text{Carga}]$  de um primeiro sinal indica que o segundo polipéptido tem uma maior propensão para agregar (menor solubilidade) relativamente ao polipéptido de referência e um valor para  $[x*\Delta H_{idr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\text{espiral-}\beta})] - [z*\Delta \text{Carga}]$  de um sinal oposto ao primeiro sinal indica que o segundo polipéptido tem uma propensão reduzida para agregar (solubilidade mais elevada) relativamente ao polipéptido de referência.

Também descrevemos código de programa de computador para implementar os métodos acima descritos e aparelhos de computador programados para implementar os métodos. Formas de realização dos métodos podem ser implementadas usando código de programa de computador em aparelhos de síntese de péptidos, preferivelmente aparelhos para sintetizar automaticamente um polipéptido baseado em resultados obtidos por aplicação dos métodos.

O código de programa pode ser fornecido num veiculador ou meio de armazenamento, tal como um disco rígido ou disco floppy, ROM ou CD-ROM, ou num veiculador óptico ou de sinal eléctrico, por exemplo via uma rede de comunicação. O código de controlo do processador pode compreender código do programa em qualquer linguagem convencional de programação por exemplo C ou assembler ou código de máquinas, e formas de realização da invenção podem ser implementadas num sistema informático geral.

### **Breve descrição das figuras**

A Figura 1 mostra o gráfico da alteração na velocidade de agregação de AcP resultante da mutação, versus (a) a alteração prevista da hidrofobicidade, (b) propensão para converter de conformação de hélice- $\alpha$  para folha- $\beta$  e (c) carga.

A Figura 2(a) mostra a variação da velocidade de agregação, no caso de mutação, calculada versus observada

para péptidos pequenos ou proteínas desenroladas na forma nativa, listadas na tabela 2.

A Figura 2(b) mostra a variação, calculada versus observada, da velocidade de agregação no caso de mutação para 27 substituições de aminoácidos de AcP nas duas regiões da sequência que parecem ser relevantes para a agregação e que abrangem os resíduos 16-31 e 87-98.

A Figura 3 mostra um diagrama de bloco de um sistema de computador para implementar um procedimento de determinação da velocidade de agregação de acordo com uma forma de realização de um aspecto da presente invenção.

A Figura 4 mostra um diagrama de fluxo de um procedimento comparativo de determinação da velocidade de agregação, de acordo com uma forma de realização da presente invenção.

A Figura 5 mostra um diagrama de fluxo de um procedimento de determinação candidato, de síntese de proteínas automatizada.

## **Exemplos**

### **Exemplo 1: Trabalho experimental para AcP**

Foram medidas as velocidades de agregação para a proteína AcP tipo selvagem ( $v_{ts}$ ) e para vários mutantes de

AcP (variantes) ( $v_{mut}$ ) após desnaturação em 25% de TFE, em tempos de fluorescência ThT, como descrito por Chiti et al., 2002 a (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 (2002 a)). Todas as medidas de velocidade de agregação foram realizadas sob condições nas quais todas as variantes de proteínas consistem de ensembles de conformações relativamente não estruturadas. A variação da velocidade de agregação como resultado de uma mutação foi expressa em todos os casos como o logaritmo natural da razão das constantes de agregação da proteína mutante e tipo selvagem ( $\ln(v_{mut}/v_{ts})$ ).

Na tabela 1, os valores de hidrofobicidade dos 20 resíduos de aminoácidos a pH neutro, baseiam-se nos coeficientes de partição de água para octanol. Estes valores são da coluna 6 da Tabela 4.8 em Creighton, T.E. *In proteins. Structure and molecular properties. Second edition.* W.H. Freeman & Company (New York, 1993), p.154)). As propensões para formação de folha- $\beta$  de 20 resíduos de aminoácidos estão normalizados de 0 (propensão elevada para formação de folha- $\beta$ ) a 1 (baixa propensão para folha- $\beta$ ). Estes valores são da coluna da Tabela 4 de Street e Mayo (1999) (Street, A.G. & Mayo, S.L. Intrinsic  $\beta$ -sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc.Natl.Acad.Sci.USA*, 96, 9074-9076(1999). A propensão para formação de folha- $\beta$  da prolina não está apresentada devido à dificuldade em determinar experimentalmente. A propensão para formação de

folha- $\beta$  da glicina é obtida a partir de cálculos teóricos. Os valores de carga correspondem a pH neutro. Os valores entre parentesis são a um pH mais baixo do que 6,0, quando o resíduo de histidina está carregado positivamente.

| Tabela 1. Escalas de hidrofobicidade, propensão para folha- $\beta$ e carga para os 20 aminoácidos naturais |  |                               |       |
|---|--|-------------------------------|-------|
| Resíduo de aminoácido   | Hidrofobicidade (kJmol <sup>-1</sup> ) | propensão para folha- $\beta$ | 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     |

Usando os valores na Tabela 1, a variação na hidrofobicidade ( $\Delta\text{Hidr}$ ), propensão para converter da estrutura hélice- $\alpha$  a folha- $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ) e a variação da carga ( $\Delta\text{Carga}$ ) foram quantificadas para AcP

usando os valores tabulados para todos os resíduos de aminoácidos.

A variação na hidrofobicidade ( $\Delta\text{Hidr}$ ) resultante de mutação foi calculada usando  $\Delta\text{Hidr} = \text{Hidr}_{ts} - \text{Hidr}_{mut}$  em que  $\text{Hidr}_{ts}$  e  $\text{Hidr}_{mut}$  são os valores dos resíduos selvagem e mutante respectivamente (os valores de hidrofobicidade para todos os 20 aminoácidos estão listados na Tabela 1).

Para calcular a propensão para converter da estrutura hélice- $\alpha$  a folha- $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ) foi necessário calcular  $\Delta\Delta G_{\text{espiral-}\alpha}$  e  $\Delta\Delta G_{\beta\text{-espiral}}$ .

A variação de energia livre para a transição de espiral aleatória  $\rightarrow$  folha- $\beta$  resultante de mutação ( $\Delta\Delta G_{\beta\text{-espiral}}$ ) foi calculada usando  $\Delta\Delta G_{\beta\text{-espiral}} = 13,64 (P_{\beta}^{ts} - P_{\beta}^{mut})$ .  $P_{\beta}^{ts}$  e  $P_{\beta}^{mut}$  são propensões para folha- $\beta$  normalizadas dos resíduos tipo selvagem e mutante (os valores de propensão para formação de folha- $\beta$  para todos os 20 aminoácidos estão listados na Tabela 1), e 13,64 é o valor da constante de conversão da escala normalizada para unidades de  $\text{kJ mol}^{-1}$ .

A variação prevista de energia livre para a transição de hélice- $\alpha$   $\rightarrow$  espiral aleatória resultante de mutação ( $\Delta\Delta G_{\text{espiral-}\alpha}$ ) foi calculada usando  $\Delta\Delta G_{\text{espiral-}\alpha} = RT \ln(P_{\alpha}^{ts}/P_{\alpha}^{mut})$ .  $P_{\alpha}^{ts}$  e  $P_{\alpha}^{mut}$  são as propensões para hélice- $\alpha$  previstas (percentagens de hélice) do tipo selvagem e sequências mutadas, no sítio da mutação, respectivamente, que foram calculadas usando o algoritmo de AGADIR em [www.embl-heidelberg.de/Services/serrano/agadir/agadir-](http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-)

start.html); R=008314 kJ mol<sup>-1</sup>K<sup>-1</sup>. (ver também Lacroix, E., Viguera AR & Serrano, L. (1998). J.Mol.Biol.284,173-191).

A variação de carga resultante da mutação ( $\Delta$ Carga) foi calculada usando  $\Delta$ Carga = |Carga<sub>mut</sub>| - |Carga<sub>wt</sub>|, em que |Carga<sub>wt</sub>| e |Carga<sub>mut</sub>| são os valores absolutos de carga para as sequências tipo selvagem e mutadas respectivamente (obtidas das somas dos valores de carga de todos os resíduos indicados na Tabela 1).

Foram feitos os gráficos individuais da variação da velocidade de agregação sobre mutação  $\ln((v_{mut}/v_{ts})$  versus  $\Delta$ Hidr, versus ( $\Delta\Delta G_{\text{espiral-}\alpha}$  +  $\Delta\Delta G_{\text{espiral-}\beta}$ ) e versus  $\Delta$ Carga, estes gráficos mostram-se na Figura 1 a, 1 b, e 1 c, respectivamente.

As mutações relatadas nas figuras 1a e 1b descritas anteriormente (Chiti et al., 2002 a, ibid.) não envolvem alterações na carga. As mutações relatadas na Figura 1c, descritas anteriormente (Chiti et al., 2002 b, ibid.), foram concebidas para minimizar a variação de hidrofobicidade e propensões para estrutura secundária. A maioria das substituições de aminoácidos de AcP envolvem resíduos no interior das duas regiões da sequência, que abrangem os resíduos 16-31 e 87-98, que se pensa serem relevantes para a agregação.

As linhas sólidas através dos dados (valores) representam as linhas de melhor ajustamento às funções lineares. Mostram-se, em cada caso os valores de r e p

resultantes de cada correlação e o declive do melhor ajustamento.

Em cada uma das análises, os pontos correspondentes aos dados estão consideravelmente espalhados à volta das linhas que representam as linhas de melhor ajustamento às funções lineares. Esta dispersão pode ser atribuída ao facto de só um parâmetro ser considerado em cada caso, à dificuldade de prever com exactidão modificações na hidrofobicidade e propensões para estrutura secundária e às importâncias relativas variáveis dos diferentes sítios de mutação no processo de agregação. Apesar da dispersão presente em cada gráfico, no entanto, observou-se que a variação da velocidade de agregação no caso de mutação ( $\ln(v_{mut}/v_{ts})$ ) para AcP correlaciona-se significativamente com cada um destes parâmetros individualmente (Figura 1 a, 1 b e 1 c). Foi calculada a dependência média de  $\ln(v_{mut}/v_{ts})$  de cada parâmetro (o declive da linha de melhor ajustamento resultante de cada análise). Os valores encontrados foram:

|  |       |
|--|-------|
| $\Delta H_{idr}$   | 0,633 |
| $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ | 0,198 |
| $\Delta Carga$   | 0,491 |

Seguindo esta análise, a Equação 1 foi elaborada e usada para determinar variação da velocidade de agregação no caso de mutação (o valor de  $\ln(v_{mut}/v_{ts})$ ):

$$\ln(v_{mut}/v_{ts}) = 0,633 * \Delta H_{idr} + 0,198 * (\Delta \Delta G_{\text{espiral-}\alpha} + \Delta \Delta G_{\text{espiral-}\beta}) - 0,491 * \Delta Carga$$

em que os números que precedem os parâmetros  $\Delta H_{idr}$ ,  $(\Delta \Delta G_{\text{espiral-}\alpha} + \Delta \Delta G_{\text{espiral-}\beta})$  e  $\Delta Carga$  são valores para x, y e z respectivamente que correspondem aos declives dos três gráficos indicados na Fig. 1 (i.e. dependências de  $\ln(v_{mut}/v_{ts})$  destes três parâmetros.

**Exemplo 2: Comparação da variação observada versus calculada na velocidade de agregação no caso de mutação de proteína AcP/ velocidades de agregação relativa de proteínas AcP mutantes**

Usando a equação 1, foi calculada a variação da velocidade de agregação para 27 substituições de aminoácidos de AcP no interior das duas regiões da sequência que parecem ser relevantes para a agregação e que abrangem os resíduos 16-31 e 87-98. A variação da velocidade de agregação para cada mutação foi determinada experimentalmente, como descrito no exemplo 1, sob condições nas quais todas as variantes de proteínas consistem em ensembles de conformações relativamente não-estruturadas. Foram feitos os gráficos dos valores de  $\ln(v_{mut}/v_{ts})$  calculados versus experimentais para todas as mutações de AcP, como se mostra na figura 2 b. A correlação observada era altamente significativa ( $r=0,756$  e  $p<0,0001$ ) e o declive era próximo de 1.

**Exemplo 3: Comparação da variação na velocidade de agregação observada versus calculada no caso de mutação para um conjunto de polipéptidos**

A função combinada, equação 1, foi aplicada para calcular a variação na velocidade de agregação no caso de mutação ( $\ln(v_{mut}/v_{ts})$  calculada) para 26 mutações nos polipéptidos amilina, péptidos prião,  $\alpha$ -sinucleína, péptido  $\beta$ -amilóide, tau, repetição rica em leucina e um péptido modelo, como está listado na Tabela 2.

Os valores para  $\Delta H_{idr}$ ,  $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\text{espiral-}\beta}$  e  $\Delta\text{Carga}$  foram calculados para cada mutação polipeptídica usando os métodos descritos no Exemplo 1.

As 26 mutações consideradas incluem tanto mutações fisiologicamente relevantes associadas a formas genéticas de doenças de deposição de proteínas como outras substituições que foram usadas na investigação para abordar questões específicas. Eram todas mutações de proteínas não estruturadas (péptidos), ou polipéptidos que parecem ser desenroladas na forma nativa, tais como o péptido  $\beta$  amilóide, o polipéptido amilóide dos ilhéus,  $\alpha$ -sinucleína, tau, polipéptidos pequenos dissecados da sequência da proteína priônica e outros péptidos modelo. Só foram consideradas na análise mutações de ponto único em péptidos curtos, não estruturados ou proteínas que são desenroladas sob condições próximas das condições fisiológicas. Foram

incluídas todas as mutações para as quais estavam directamente disponíveis os valores experimentais de  $\ln(v_{mut}/v_{ts})$  ou podiam ser determinados a partir de dados na literatura. Foram excluídas mutações que actuavam simplesmente por destabilização do estado nativo da proteína envolvida. Foram considerados dados independentemente das técnicas experimentais empregues pelos diferentes autores para investigar agregação, desde que fosse possível levar a cabo análise quantitativa. Quando não estavam explicitamente indicadas constantes de tempo ou velocidade, os gráficos descrevendo os perfis cinéticos de agregação foram "scanned" e analisados por computador. Este procedimento permitiu que se reconstruissem gráficos com valores numéricos dos pontos e se analisassem para obter valores de constantes de velocidade. Quando eram evidentes fases de latência e fases de crescimento nos perfis cinéticos de agregação, só foi considerada a fase de crescimento. Quando estavam indicados dados em períodos fixos de tempo (por exemplo por meio de gráficos de barras), o valor observado de  $\ln(v_{mut}/v_{ts})$  foi obtido a partir do quociente dos parâmetros de agregação da proteína (péptido) mutada e tipo selvagem, antes que fosse atingido o equilíbrio.

Não foram analisadas as mutações que envolviam resíduos de prolina devido à dificuldade em obter estimativas quantitativas da variação da propensão para folha- $\beta$  como resultado destas mutações (ver tabela 1). Também não foram consideradas mutações para as quais estavam

indicadas, por autores diferentes, discrepâncias substanciais no valor de  $\ln(v_{mut}/v_{ts})$  (quando as discrepâncias eram significativas mas não substanciais, considerámos  $\ln(v_{mut}/v_{ts})$  resultantes de médias dos dados disponíveis).

Tabela 2. Variações na hidrofobicidade, propensão na estrutura secundária, carga e velocidade de agregação como resultado de mutações de ponto único(single point) de péptidos não estruturados ou proteínas desenroladas na forma nativa

| Mutação                                   | $\Delta H_{idr}$ (kJmol <sup>-1</sup> ) | $\Delta\Delta G_{\beta\text{-espiral}}$ (kJmol <sup>-1</sup> ) | $\Delta\Delta G_{\text{espiral-}\alpha}$ (kJmol <sup>-1</sup> ) | $\Delta$ Carga | $\ln(v_{mut}/v_{ts})$ calculado | $\ln(v_{mut}/v_{ts})$ observado | ref |
|---|---|--|---|----------------|---------------------------------|---------------------------------|-----|
| <b>amilina</b>                            |   |  |   |                |                                 |                                 |     |
| 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                                      | 0,30                                    | 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  |
| <b>Péptidos priónicos</b>                 |   |  |   |                |                                 |                                 |     |
| H111A                                     | 3,26                                    | -1,36  | -3,21   | -1             | 1,65                            | 0,60                            | 13  |
| HII-1K                                    | 0,10                                    | 0,41   | -1,72   | 0              | -0,20                           | -0,26                           | 13  |
| A117V                                     | 0,91                                    | 4,63   | 2,37  | 0              | 1,96                            | 1,51                            | 13  |
| V210I                                     | 0,52                                    | 0,41   | -0,97   | 0              | 0,22                            | 0,84                            | 14  |
| <b><math>\alpha</math>-sinucleína</b>     |   |  |   |                |                                 |                                 |     |
| 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  |
| <b>Péptido <math>\beta</math>Amiloide</b> |   |  |   |                |                                 |                                 |     |
| 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  |
| 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  |

(continuação)

| Mutação                          | $\Delta H_{idr}$ (kJmol <sup>-1</sup> ) | $\Delta \Delta G_{\beta\text{-espiral}}$<br>(kJmol <sup>-1</sup> ) | $\Delta \Delta G_{\text{espiral-}\alpha}$<br>(kJmol <sup>-1</sup> ) | $\Delta$ Carga | $\ln(v_{mut}/v_{ts})$<br>calculado | $\ln(v_{mut}/v_{ts})$<br>observado | ref      |
|----------------------------------|---|--|---|----------------|------------------------------------|------------------------------------|----------|
| <b>Tau</b>                       |   |  |   |                |                                    |                                    |          |
| G272                             | 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   |
| <b>Repetição rica em leucina</b> |   |  |   |                |                                    |                                    |          |
| 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       |
| <b>Péptido modelo</b>            |   |  |   |                |                                    |                                    |          |
| D6E                              | 0,90                                    | 5,04   | -2,27   | 0              | 1,12                               | 0,40                               | 25       |
| D6N                              | 1,90                                    | 4,36   | 0,00  | 1              | 1,57                               | 0,52                               | 25       |

Foi obtido o gráfico do valor calculado versus experimental de  $\ln(v_{mut}/v_{ts})$  e mostra-se na figura 2 (a). A correlação altamente significativa ( $r=0,84$ ,  $p<0,0001$ ) e o valor do declive, próximo de 1,0, indicam uma concordância próxima entre os efeitos calculados e experimentais de mutações nas velocidades de agregação deste grupo heterogénio de polipéptidos. As modificações observadas na velocidade de agregação no caso de mutação apresentam-se numa amplitude de ca. 800 vezes, *i.e.* de 15 mais lento a 53 mais rápido do que o polipéptido tipo selvagem correspondente (Figura 2 a e Tabela 2). 84% destas mutações têm valores calculados de  $\ln(v_{mut}/v_{ts})$  que variam de um factor de 3 dos valores observados de  $\ln(v_{mut}/v_{ts})$ . A percentagem aumenta para 92% e 96% se são considerados factores de dispersão de 5 e 10, respectivamente. Exemplos em que há uma concordância próxima entre valores teóricos e experimentais incluem mutações associadas com encefalopatias espongiformes hereditárias, tais como as substituições

A117V e V210I da proteína priónica (Tabela 2). Os valores previstos e experimentais também estão em estreita concordância para a mutação associada com o início precoce de doença de Parkinson e para várias mutações associadas com o péptido  $\beta$  amilóide e responsável por ou início precoce de doença de Alzheimer ou hemorragia cerebral hereditária com amiloidose (Tabela 2).

Se a análise fosse repetida usando um só determinante para calcular os valores de  $\ln(v_{mut}/v_{ts})$ , foram ainda encontrados entre os valores calculados e observados de  $\ln(v_{mut}/v_{ts})$  ( $p=0,003$  usando somente  $\Delta H_{idr}$  para calcular  $\ln(v_{mut}/v_{ts})$ ,  $p=0,036$  usando somente  $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\text{espiral-}\beta}$  e  $p=0,011$  usando somente  $\Delta Carga$ ). No entanto estas correlações são menos notáveis do que as observadas quando se considera uma combinação de todos os três factores e os declives são significativamente menores que 1,0 (0,61, 0,19, e 0,1 usando somente  $\Delta H_{idr}$ , somente  $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\text{espiral-}\beta}$  e somente  $\Delta Carga$  respectivamente. Isto demonstra que a equação na qual estes factores são combinados fornece um método mais exacto para determinar o quociente da velocidade de agregação para polipéptidos modificados (e.g. mutantes) e referência (e.g. tipo selvagem).

A correlação mostrada na figura 2(a) entre efeitos teóricos e experimentais de mutações na agregação considerou-se notável, considerando o grupo heterogénio de proteínas e sistemas peptídicos usados na análise, assim como a variabilidade dos sítios nos quais ocorrem as várias mutações.

**Exemplo 4: Aplicabilidade do algoritmo a modificações que envolvem vários resíduos de aminoácidos e o uso de parâmetros cinéticos sem ser a "velocidades de agregação".**

A equação 1 foi testada em relação a outros sistemas para avaliar a sua aplicabilidade a sistemas mais gerais. Os cálculos usados para derivar a equação 1 baseiam-se na cinética de agregação experimentada por variantes de proteínas e péptidos que diferem num único resíduo em relação à sequência original. As velocidades ( $v_{mut}$  e  $v_{ts}$ ) usadas na expressão correspondem à fase exponencial de agregação para cada um dos péptidos, e não incluem qualquer período de latência ou fase de nucleação que preceda essa fase.

Foram incluídos derivados de duas variações de Calcitonina para testar a validade desta expressão para prever as propensões de agregação de péptidos. A primeira era para avaliar se o efeito de várias substituições poderia ser previsto do mesmo modo que o algoritmo era capaz com mutações pontuais. A segunda era para incluir como parâmetro cinético o quociente relativo de tempos de agregação ( $\tau_{mut}/\tau_{ts}$ ). Incluindo o efeito de uma fase de latência na cinética de agregação exibida pelos péptidos, os tempos de agregação para cada um dos péptidos ( $\tau$ ), poderia ser definido de dois modos diferentes: o primeiro era o tempo de nucleação ou tempo que precede a iniciação da agregação ou o desenvolvimento da turbidez da solu-

ção(T1), e o segundo corresponderia ao meio tempo de agregação ou o tempo no qual variações nas medições usadas para monitorizar a agregação ("light scattering" ou qualquer outro método) atingiam metade do seu valor máximo (T2). Isto pode permitir a aplicação da equação para prever as propensões de agregação para o intervalo mais alargado de moléculas com aspectos de concepção (design) diferentes.

Os cálculos foram feitos em duas variantes de Calcitonina, usando dados disponíveis na literatura (Arvinte, et al. 1993, J Biol Chem 268: 6415-6422) e estudos prévios incluídos noutra pedido de patente por alguns membros do grupo (Zurdo & Dobson, WO 02/083734, PCT/GB02/01778). Os cálculos foram efectuados usando dados divulgados nessas publicações, produzindo os valores indicados na tabela 3. Em ambos os casos o valor para o parâmetro  $\tau_{ts}$  foi obtido independentemente.

| <b>Tabela 3.</b> Variações previstas e experimentais nos tempos de agregação exibidos pelos vários péptidos da calcitonina quando comparados com a sequência humana   |                                 |  |                                    |                                    |
|---|---------------------------------|--|------------------------------------|------------------------------------|
|   | $\ln(v_{mut}/v_{ts})$ calculado | $\ln(v_{mut}/v_{ts})$ observado        | $(\tau_{mut}/\tau_{ts})$ calculado | $(\tau_{mut}/\tau_{ts})$ observado |
| 1Salmão-1   | -10,54                          | -10,31                                 | 37;681,05                          | $\sim 30000^a$                     |
| 2SEQ ID No 14   | -5,60                           | -4,61 <sup>b</sup> /-5,71 <sup>a</sup> |                                    | 100 <sup>b</sup> /300 <sup>a</sup> |
| <sup>1</sup> Dados obtidos de Arvinte et al (1993)J Biol Chem 268, 6415-6422. Salmon calcitoninhas 16 modified positions when compared to the human sequence.<br><sup>2</sup> Sequência revelada em Zurdo&Dobson (WO02/083734,PCT/GB02/01778) e Zurdo&Dobson(observações não publicadas).Sequência IDno14 mostra 6 posições modificadas quando comparadas com a sequência humana.<br><sup>a</sup> Valores para calcular $\tau$ foram obtidas usando T1 como descrito acima.<br><sup>b</sup> Valores para calcular $\tau$ foram obtidas usando T2 como descrito acima. |                                 |  |                                    |                                    |

Os cálculos para variações no tempo de agregação foram efectuados assumindo as seguintes relações com velocidades de agregação descritas pela equação 1.

$$(v_{mut}/v_{ts}) = (v_{ts}/v_{mut}) = 1/\exp(\ln(v_{mut}/v_{ts}))$$

Esta análise mostra que a equação 1 pode ser usada para prever o comportamento de agregação de um dado polipéptido que tem mais do que uma modificação de aminoácidos comparada com a sequência polipeptídica original. Além disso, sugere que, em sistemas em que está presente uma fase de latência, ou a velocidade de agregação pode ser difícil de calcular, parâmetros cinéticos alternativos representados pelos tempos de agregação (T1-tempo de nucleação-ou T2-meio tempo de agregação) podem fornecer valores válidos para comparar com as previsões dadas pela equação 1.

**Exemplo 5: Aplicabilidade do algoritmo para modificações envolvendo adição ou deleção de resíduos de aminoácidos: péptidos A $\beta$  ligados com a doença de Alzheimer**

Os péptidos A $\beta$  (1-40) e A $\beta$ (1-42) que estão associados com a doença de Alzheimer apresentam diferenças nas suas propensões para agregação. Os péptidos diferem na sequência somente em dois resíduos no terminal C. Os métodos da invenção explicam a elevada propensão para agregar da forma de 42 resíduos, relativamente à forma de 40 resíduos do péptido amilóide  $\beta$  associado com a doença de

Alzheimer (Jarrett et al., 1993). De facto, embora a propensão para formação de hélice- $\alpha$  e a carga do péptido como um todo não pareçam ser alterados com a adição do dipéptido Ile-Ala no terminal C, os valores de hidrofobicidade e propensão para formação de folha- $\beta$  dos dois resíduos são mais elevados do que os valores médios calculados para o péptido inteiro.

De um ponto de vista quantitativo, a variação da hidrofobicidade resultante da adição de dois resíduos no terminal-C pode ser calculada como  $\Delta H_{idr} = H_{idr_{ts}} - H_{idr_{mut}}$ , em que  $H_{idr_{ts}}$  é a hidrofobicidade média dos 40 resíduos que formam a forma curta do péptido;  $H_{idr_{mut}}$  é a hidrofobicidade média dos dois resíduos inseridos (Ile-Ala). A variação na propensão para folha- $\beta$  resultante da inserção pode ser calculada de modo semelhante. Isto leva à previsão de que a forma longa se agrega 7 vezes mais depressa do que a forma curta, em boa concordância com o perfil cinético relatado por Jarrett et al., 1993 que observou uma aceleração de 7-8 vezes (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)).

### **Exemplo 6: Sistema informático para operar o método da invenção**

No que se refere à Figura 3, esta mostra um diagrama de bloco de um sistema informático para imple-

mentar uma forma de realização do método acima descrito. Um sistema de computador para uso geral 300 compreende um processador 300a acoplado a memória de computador 300b que armazena código de programa de computador para implementar as formas de realização do método, como descrito mais abaixo, e interfaces 300c tais como écran de computador convencional, teclado, rato e impressora, assim como outras interfaces tais como uma interface de rede, uma interface de controlo para um sintetizador de péptidos e interfaces de software tais como uma interface de base de dados.

O sistema de informático 300 aceita a introdução por um utilizador a partir de um dispositivo de entrada 304 tal como um teclado, ficheiro de dados de entrada, ou uma interface de rede, e fornece um resultado para um dispositivo de saída 308 tal como uma impressora, interface de rede ou dispositivo de armazenamento de dados. O dispositivo de entrada 304 recebe uma entrada (input) que compreende uma sequência de aminoácidos para o péptido modificado (e.g. mutante) assim como valores de pH e temperatura apropriados para um ambiente para o qual é determinada a velocidade de agregação do polipéptido. Também pode ser introduzido um factor de correcção glicina/prolina, tal como uma ponderação para um factor de distorção estrutural que interfere com a formação de folha- $\beta$  intermolecular ou agregação. O dispositivo de saída 308 fornece informação comparativa da velocidade de agregação tal como um log (base 10 ou natural) do quociente de agregação, por exemplo, um quociente de metade para a

agregação de um mutante em comparação com um polipéptido tipo selvagem.

O sistema de computador 300 está acoplado com um armazenamento de dados 302 que armazena dados de hidrofobicidade, dados de propensão para formação de folha- $\beta$  (como valores de propensão *per se* ou em termos de energia livre) e dados de carga. Este dado é armazenado para cada aminoácido (resíduo) e preferivelmente, é armazenado uma pluralidade de conjuntos de cada um destes tipos de dados correspondendo a valores diferentes de pH e temperatura. O sistema de computador, na forma de realização ilustrada, é mostrado com interface para um calculador 306 de propensão para hélice- $\alpha$ . Isto pode ser uma máquina separada, por exemplo acoplada a um sistema de computador 300 através de uma rede, ou pode compreender um programa separado que corre num sistema de computador de uso geral 300, ou noutras formas de realização o código de propensão para formação de hélice- $\alpha$  pode ser armazenado numa área de armazenamento de programa 300b e operar de forma unitária com o código de determinação da velocidade de agregação descrito abaixo. No entanto, qualquer que seja o método empregado, o calculador de propensão para hélice- $\alpha$  recebe os dados de sequência, indirectamente do dispositivo de entrada do utilizador, e fornece o dado de propensão para formação de hélice- $\alpha$  em retorno. Estes dados e os dados no armazenamento de dados 302 podem ser determinados para um aminoácido na base de um aminoácido ou podem ser determinados tendo em conta o contexto de sequência, por exemplo, usando uma janela sobre

a sequência para modificar valores dos dados dependendo dos aminoácidos na vizinhança.

Como ilustrado, o sistema informático 300 também pode fornecer uma saída de controlo de dados 310 a um sintetizador de péptidos automático 312. Os dados de controlo compreendem geralmente uma sequência de aminoácidos de um polipéptido. Deste modo o sistema de computador 300 pode ser programado para comparar automaticamente as propriedades de um número de polipéptidos modificados (e.g. mutantes) e seleccionar um ou mais destes que se prevê que tenham propriedades favoráveis para a síntese automatizada. Um exemplo de um sintetizador deste tipo seria um Sintetizador de péptidos ABI 433 A (Applied Biosystems).

Referindo-nos a seguir à figura 4, esta mostra um procedimento para determinar uma velocidade de agregação comparativa de acordo com o descrito acima. A Figura 4 representa um diagrama de fluxo de uma forma de realização de código a correr na memória do programa 300b da Figura 3.

No passo S400, um utilizador entra uma sequência de aminoácidos, dados de pH e temperatura, opcionalmente com dados do terminal C e N para a sequência. Depois, no passo S402 o sistema de computador lê os dados de hidrofobicidade para a sequência entrada do armazenamento de dados e soma-os para fornecer uma estimativa da hidrofobicidade para o péptido codificado pela sequência. Quando estão disponíveis, como altamente preferido, dados para um intervalo de

valores de pH e temperatura são buscados os dados que mais correspondem ao pH e temperatura desejados. Depois, como passos S404 e S406, o procedimento lê os dados de carga e dados de propensão para formação de folha- $\beta$  do armazenamento de dados, de um modo semelhante, somando os dados de carga, para fornecer uma estimativa da carga para o polipéptido que corresponde à sequência entrada e, de modo semelhante, somando os dados de propensão para formação de folha- $\beta$  (normalmente expressa em termos de energia livre). Com a prolina, não está disponível um dado de propensão para folha- $\beta$  e por isso o resíduo de prolina pode ser saltado quando se soma estes valores, ou pode ser empregue um valor arbitrário de propensão para formação de folha- $\beta$  ou um correspondente a outro aminoácido. Por exemplo, se a propensão para formação de folha- $\beta$  for expressa em termos de energia livre, pode ser usado um valor arbitrário de 1, ou um valor correspondente a outro aminoácido. Opcionalmente os passos S402 e S406 podem empregar uma "janela" (por exemplo 3, 5, 7 ou mais aminoácidos) que incluiriam um factor de correcção para o efeito de resíduos flanco nas propriedades de um aminoácido particular, (*i.e.* para ter em conta vizinhos próximos numa sequência de aminoácidos), em vez de considerar cada aminoácido da sequência individualmente.

No passo S408, o procedimento fornece a sequência entrada a um calculador de propensão para formação de de hélice- $\alpha$ , com os dados de pH e temperatura, e, quando disponíveis os dados dos terminais C e N. Um calculador de

propensão para formação de hélice- $\alpha$  S408a opera nestes dados e retorna dados de volta ao procedimento do passo S410, compreendendo os dados retornados um valor de propensão para hélice- $\alpha$  para a sequência completa. Um código de programação conveniente para um calculador de propensão para hélice- $\alpha$  S408a compreende o código AGADIR disponível em <http://www.embl-heidelberg.de/Services/serrano/agadir/agadir-start.html>, o código GOR4 disponível em [http://npsa-pbil.ibcp.fr/cgi-bin/npsa\\_automat.pl?page=npsa\\_gor4.html](http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=npsa_gor4.html) e outros códigos descritos acima. Uma pessoa especializada reconhecerá que, se desejado, este código ou um código recentemente concebido derivado de dados publicamente acessíveis (descritos na literatura científica) ou dados experimentais adicionais, podem ser incorporados no código, ao implementar o procedimento da Figura 4, em vez de ser implementado como um procedimento separado.

No passo S412 o procedimento determina então a velocidade de agregação comparativa do polipéptido, definido pela sequência de aminácidos importada, em comparação com um polipéptido de referência, usando a equação 1 acima. Pode ver-se da equação 1 que a determinação da velocidade de agregação comparativa requer uma diferença em hidrofobicidade, propensão na estrutura secundária e carga, e os valores para hidrofobicidade, propensão na estrutura secundária e carga para polipéptidos de referência pode ser determinada, ou repetindo os passos S400 a S410 para o polipéptido de referência, ou lendo os valores armazenados destes parâmetros do armazenamento de dados 302, ou de

qualquer outro modo convencional. Se desejado, no passo S412 os parâmetros ou factores de escala na equação 1 que operam nas diferenças em hidrofobicidade, propensão estrutural e carga, podem ser seleccionados de conjuntos de parâmetros convenientes (passo S414) em resposta a dados entrados tais como dados tipo polipéptido. Por exemplo, um polipéptido de espiral aleatória pode usar parâmetros diferentes de um polipéptido parcialmente enrolado ou estruturado. Um polipéptido rico num tipo específico de resíduo, tais como aminoácidos aromáticos ou carregados, pode também requerer parâmetros diferentes.

Após a determinação da velocidade de agregação comparativa, pode ser aplicada uma correcção opcional ao passo S416 para resíduos de prolina e ou glicina para ter em conta preferências conformacionais ou estruturais adicionais que possam impedir a formação de estruturas folha- $\beta$  ou agregadas intermoleculares por um dado péptido e então no passo S418 o sistema faz sair o resultado do cálculo da velocidade de agregação comparativa. Isto pode compreender um simples valor positivo ou negativo indicando se a velocidade de agregação do polipéptido modificado (e.g. mutante) é maior ou menor do que a do polipéptido de referência, mas preferivelmente, compreende dados quantitativos que se relacionam com velocidades de agregação comparativas tais como um log do quociente de velocidades de agregação.

A figura 5 mostra um diagrama de fluxo de uma

forma de realização vantajosa para o procedimento da figura 4. Em particular a Figura 5 mostra um método para triagem de polipéptidos modificados (e.g. mutações) para seleccionar candidatos com propriedades promissoras para investigação adicional e, opcionalmente, síntese. Assim no passo S500 uma sequência de aminoácidos para um polipéptido de referência é introduzida juntamente com dados que identificam uma ou mais posições modificadas (e.g. mutantes). Opcionalmente o procedimento também pode permitir uma modificação ou intervalo de modificações a ser especificadas, por exemplo em termos de um conjunto pré-determinado ou selecção de aminoácidos.

A seguir à inicialização, no passo S502 o procedimento gera uma sequência modificada que representa uma das permutações possíveis definidas pelos dados entrados e a seguir, no passo S504, determina uma velocidade de agregação comparativa para o polipéptido modificado em comparação com o polipéptido de referência, por exemplo usando o procedimento na Figura 4. A seguir, no passo S506, o procedimento verifica se há mais algumas permutações para as quais se execute o cálculo, e, em caso afirmativo, volta de novo ao passo S502 até que tenha sido gerado um conjunto completo de permutações possíveis. Em seguida, no passo S508, o conjunto de dados de velocidade de agregação comparativa para cada polipéptido modificado (em comparação com a proteína de referência) é exportado, por exemplo como uma autolista, gráfico ou de qualquer outro modo conveniente. Estes dados podem então ser usados,

por exemplo para identificar candidatos para a síntese e/ou para comparação com outros dados tais como imunogenicidade/antigenicidade. Em particular, um ou mais dos "melhores" polipéptidos modificados, por exemplo mutantes com uma velocidade de agregação particularmente elevada ou baixa, podem ser coleccionados e os dados de sequência para estes polipéptidos modificados exportados para um sintetizador automático de péptidos tal como o sintetizador 312 da Figura 3 para produzir automaticamente as proteínas mutantes para , por exemplo, investigação adicional.

### **Referências**

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).

Volles, M. J. & Lansbury, P. T. Jr. Vesicle permeabilization by protofibrillar  $\alpha$ -synuclein is sensitive to Parkinson's disease-linked mutations and occurs by a pore-like mechanism. *Biochemistry*, 41, 4595-4602 (2002).

Para as referências seguintes os números são os usados na tabela 2:

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 wildtype amylin. *Am. J. Pathol.* 157, 2101-2109 (2000).
13. Salmona, 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).
14. Thompson, A. J., Barnham, 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.

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  $\alpha$ -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.

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., Stenh, C., Luthman, J., Teplov, D. B., Younkin, S. G., Naslund, J. & Lannfelt, L. (2001). The 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced Ab protofibril formation. *Nature Neurosci.* 4, 887-893.

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  $\beta$ -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).

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).

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  $\beta$ -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).

**REIVINDICAÇÕES**

1. Um método de previsão implementado por computador para prever o efeito de uma modificação de aminoácido na velocidade de agregação (solubilidade) de um polipéptido de referência que compreende:

calcular a diferença de hidrofobicidade ( $\Delta\text{Hidr}$ ) entre o polipéptido de referência e o polipéptido modificado;

calcular a diferença na propensão para formação de folha- $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ) entre o polipéptido de referência e o polipéptido modificado;

calcular a diferença de carga ( $\Delta$  Carga) entre o polipéptido de referência e o polipéptido modificado;

calcular  $[x*\Delta\text{Hidr}] + [y*(\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga], em que  $x$ ,  $y$  e  $z$  são factores de escala; e o referido efeito previsto da referida modificação de aminoácido na referida velocidade de agregação (solubilidade) de  $[x*\Delta\text{Hidr}] + [y*(\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga], em que um valor para  $[x*\Delta\text{Hidr}] + [y*(\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga] de um primeiro sinal indica que o polipéptido modificado tem uma maior propensão para agregar (menor solubilidade) relativamente ao polipéptido de referência e um valor para  $[x*\Delta\text{Hidr}] + [y*(\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga] de um sinal oposto ao primeiro sinal indica que o polipéptido modificado tem uma propensão reduzida

para agregar (solubilidade mais elevada) relativamente ao polipéptido de referência.

2. Um método de acordo com a reivindicação 1 em que o efeito da modificação de aminoácido na velocidade de agregação é expresso como  $\ln(v_{\text{mod}}/v_{\text{ref}})$ , em que um valor positivo para  $\ln(v_{\text{mod}}/v_{\text{ref}})$ , indica que o polipéptido modificado tem uma maior propensão para agregar (menor solubilidade) relativamente ao polipéptido de referência; e um valor negativo para  $\ln(v_{\text{mod}}/v_{\text{ref}})$ , indica que o polipéptido modificado tem uma propensão reduzida para agregar (maior solubilidade) relativamente ao polipéptido de referência.

3. Um método de acordo com qualquer reivindicação 1 ou 2, em que o factor de escala x é um valor de 0,59 a 0,64, o factor de escala y é um valor de 0,19 a 0,22 e o factor de escala z é um valor de 0,49 a 0,51.

4. Um método de acordo com a reivindicação 1, em que x é 0,6, y é 0,2 e z é 0,5.

5. Um método de acordo com a reivindicação 1, em que x é 0,63, y é 0,20 e z é 0,49.

6. Um método implementado por computador para identificar uma modificação de aminoácido que reduz a velocidade de agregação (aumenta a solubilidade) de um polipéptido de referência que compreende:

usar um método de acordo com qualquer das reivindicações 1 a 5 para prever a variação na velocidade de agregação para um ou mais polipéptido(s) modificado(s), tendo cada polipéptido modificado uma ou mais modificações de aminoácido(s) em comparação com o polipéptido de referência,

comparar as velocidades de agregação previstas do polipéptido de referência e referidos um ou mais polipéptidos modificados e identificar um ou mais polipéptido(s) modificado(s) que têm uma velocidade de agregação prevista reduzida relativamente ao polipéptido de referência.

7. Um método implementado por computador para identificar uma modificação de aminoácido que aumenta a velocidade de agregação (diminui a solubilidade) de um polipéptido de referência que compreende:

usar um método de acordo com qualquer das reivindicações 1 a 5 para prever a variação na velocidade de agregação para um ou mais polipéptido(s) modificado(s), tendo cada polipéptido modificado uma ou mais modificações de aminoácidos em comparação com o polipéptido de referência,

comparar as velocidades de agregação previstas do polipéptido de referência e referidos um ou mais polipéptidos modificados e identificar um ou mais polipéptido(s) modificado(s) que têm uma velocidade de

agregação aumentada prevista relativamente ao polipéptido de referência.

8. Um método implementado por computador para produzir um polipéptido com velocidade de agregação reduzida, que compreende:

usar o método da reivindicação 6 para identificar uma modificação que se prevê que reduza a velocidade de agregação do polipéptido; e  
produzir um polipéptido com velocidade de agregação reduzida com a referida modificação.

9. Um método implementado por computador para produzir um polipéptido com velocidade de agregação aumentada, que compreende:

usar o método da reivindicação 7 para identificar uma modificação que se prevê que aumente a velocidade de agregação do polipéptido; e  
produzir um polipéptido com velocidade de agregação aumentada com a referida modificação.

10. Um método de acordo com qualquer das reivindicações precedentes, em que a modificação compreende uma ou mais substituições, deleções e/ou adições de aminoácidos.

11. Um programa de computador que compreende um

código de software para, quando em funcionamento, prever o efeito de uma modificação de aminoácido na velocidade de agregação de um polipéptido, em que o código compreende código para:

entrar uma modificação de aminoácido que converteria o polipéptido a uma forma modificada do polipéptido;

calcular a diferença de hidrofobicidade ( $\Delta\text{Hidr}$ ) entre o polipéptido e o polipéptido modificado;

calcular a diferença na propensão para formação de folha- $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ) entre o polipéptido e o polipéptido modificado;

calcular a diferença de carga ( $\Delta$  Carga) entre o polipéptido e o polipéptido modificado.;

calcular  $[x*\Delta\text{Hidr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga], em que  $x$ ,  $y$  e  $z$  são factores de escala; e o referido efeito previsto da referida modificação de aminoácido na referida velocidade de agregação (solubilidade) de  $[x*\Delta\text{Hidr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga], em que um valor para  $[x*\Delta\text{Hidr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga] de um primeiro sinal indica que o polipéptido modificado tem uma maior propensão para agregar (menor solubilidade) relativamente ao polipéptido de referência e um valor para  $[x*\Delta\text{Hidr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga] de um sinal oposto ao primeiro sinal indica que o polipéptido modificado tem uma propensão reduzida para agregar (solubilidade mais elevada) relativamente ao polipéptido de referência.

12. Programa de computador que compreende software codificado para, quando em funcionamento, identificar um polipéptido modificado, compreendendo o código:

código para prever uma modificação na velocidade de agregação para um ou mais polipéptidos modificados, tendo cada polipéptido modificado uma ou mais modificações de aminoácidos quando comparado com o polipéptido de referência, sendo a variação na velocidade uma modificação de uma velocidade de agregação do referido polipéptido de referência, em que o código compreende o código como reivindicado na reivindicação 11; e

código para identificar um ou mais dos referidos polipéptidos modificados dependendo da referida variação prevista na velocidade de agregação.

13. Um suporte de dados que transporta o programa das reivindicações 11 ou 12.

14. Um sistema de computador para determinar a velocidade de agregação de um segundo polipéptido em relação a um polipéptido de referência, tendo o referido segundo polipéptido e o referido polipéptido de referência uma sequência de aminoácidos, compreendendo a sequência de aminoácidos do referido segundo polipéptido uma versão modificada da sequência de aminoácidos do referido polipéptido de referência, compreendendo o sistema de computador (informático):

um armazenamento de dados para armazenar dados compreendendo dados de hidrofobicidade, dados de propensão para formação de folha- $\beta$  e dados de carga para um conjunto de aminoácidos;

uma área de armazenamento de programas que armazena um código implementado por um processador; e

um processador, acoplado à referida área de armazenamento de programas e ao referido armazenamento de dados para implementar o referido código armazenado, compreendendo o código um código para controlar o processador para:

introduzir uma sequência de aminoácidos para o referido segundo polipéptido;

ler os dados de hidrofobicidade para a sequência de aminoácidos do referido segundo polipéptido a partir da referido armazenamento de dados e determinar um valor de hidrofobicidade para o referido segundo polipéptido;

ler os dados de carga para a sequência de aminoácidos do referido segundo polipéptido do referido armazenamento de dados e determinar um valor de propensão para formação de folha- $\beta$  para o referido segundo polipéptido;

determinar um valor de propensão para formação de hélice- $\alpha$  para o referido segundo polipéptido;

obter os dados de hidrofobicidade, os dados de carga, os dados de propensão para formação de

folha- $\beta$  e um valor de propensão para formação de hélice- $\alpha$  para o referido polipéptido de referência; e

determinar a referida velocidade de agregação relativa, usando a referida hidrofobicidade, carga, dados de propensão para formação de folha- $\beta$  e propensão para hélice- $\alpha$  para os referidos polipéptidos segundo e de referência avaliando  $x\Delta H + y\Delta G - z\Delta C$  em que  $x$ ,  $y$  e  $z$  são factores de escala,  $\Delta H$  representa a diferença em hidrofobicidade entre o referido segundo polipéptido e o referido polipéptido de referência,  $\Delta G$  representa a diferença entre o referido segundo polipéptido e o referido polipéptido de referência na propensão para converter de uma estrutura de hélice- $\alpha$  para uma estrutura de folha- $\beta$  e  $\Delta C$  representa a diferença de carga entre o referido segundo polipéptido e o referido polipéptido de referência e em que um valor para  $x\Delta H + y\Delta G - z\Delta C$  de um primeiro sinal indica que o segundo polipéptido tem uma maior propensão para agregar relativamente ao polipéptido de referência e um valor de  $x\Delta H + y\Delta G - z\Delta C$  de um sinal oposto ao primeiro sinal indica que o segundo polipéptido tem uma propensão reduzida para agregar relativamente ao polipéptido de referência.

15. Um sistema informático como reivindicado na reivindicação 14, em que o referido código para determinar

a referida velocidade de agregação relativa compreende um código para determinar um valor aproximado para um logaritmo de um quociente entre uma velocidade de agregação do referido segundo polipéptido e a velocidade de agregação do referido polipéptido de referência.

16. Um sistema informático como reivindicado nas reivindicações 14 ou 15 em que o referido código para determinar um valor de hidrofobicidade para o referido segundo polipéptido é configurado para combinar os valores de hidrofobicidade para aminoácidos da sequência do segundo polipéptido, tendo em conta a vizinhança de um aminoácido na referida sequência.

17. Sistema informático como reivindicado nas reivindicações 14, 15 ou 16 em que o referido código para determinar um valor de propensão para formação de folha- $\beta$  para o referido segundo péptido é configurado para combinar valores de propensão para folha- $\beta$  para aminoácidos da sequência para o segundo polipéptido, tendo em conta a vizinhança de um aminoácido na referida sequência.

18. Sistema informático como reivindicado em qualquer das reivindicações 14 a 17 em que o referido código compreende adicionalmente um código para introduzir um factor de correcção para Glicina e/ou Prolina, e em que o referido código para determinar a referida velocidade relativa de agregação compreende adicionalmente um código para aplicar o referido factor de correcção.

19. Um sistema informático como reivindicado em qualquer das reivindicações 14 a 18 em que o referido armazenamento de dados é configurada para armazenar os referidos dados de hidrofobicidade, dados de propensão para folha- $\beta$  e dados de carga para uma pluralidade de temperaturas e/ou valores de pH, em que o referido código compreende adicionalmente um código para entrada de dados de pH e/ou temperatura para o referido segundo péptido e em que o referido código para ler os referidos dados de hidrofobicidade, dados de propensão para formação de folha- $\beta$  e dados de carga, está configurado para seleccionar dados a ler a partir do referido armazenamento de dados dependendo dos referidos dados de entrada de pH e/ou temperatura.

20. Um sistema informático como reivindicado na reivindicação 19 em que o referido código compreende adicionalmente um código para entrada de dados de selecção e um código para seleccionar um referido factor de escala, do referido armazenamento de dados, em resposta aos referidos dados de selecção.

21. Um Sistema informático tal como reivindicado em qualquer das reivindicações 14 a 20, em que o referido código compreende adicionalmente um código para gerar sequências para um ou mais segundos polipéptidos, um código para determinar a referida velocidade de agregação relativa de cada uma das referidas sequências dos referidos um ou mais segundos polipéptidos e um código para saída de dados

correspondentes às referidas velocidades de agregação relativas, determinadas.

22. Um Sistema informático tal como reivindicado em qualquer das reivindicações 14 a 20 em que o referido código compreende adicionalmente um código para gerar sequências para um ou mais dos referidos segundos polipéptidos, um código para determinar a referida velocidade de agregação relativa de cada uma das referidas sequências dos referidos um ou mais segundos polipéptidos, um código para seleccionar a referida sequência gerada e um código para saída das referidas sequências seleccionadas.

23. Um programa de computador que compreende um código de software de qualquer das reivindicações 14 a 22.

24. Um suporte de dados legível por computador que compreende o programa da reivindicação 23.

25. Um método implementado por computador para determinar um indicador da velocidade de agregação relativa, predizendo o referido indicador de velocidade de agregação relativa, a velocidade de agregação de um segundo polipéptido em comparação com a velocidade de agregação de um polipéptido de referência, compreendendo o referido segundo polipéptido uma ou mais modificações de aminoácidos em comparação com o polipéptido de referência, compreendendo o método:

determinar a diferença de hidrofobicidade ( $\Delta\text{Hidr}$ ) entre o referido segundo polipéptido e o referido polipéptido de referência;

determinar a diferença na diferença na propensão para formação de folha- $\beta$  ( $\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}}$ ) entre o referido segundo polipéptido e o referido polipéptido de referência;

determinar a diferença de carga ( $\Delta$  Carga) entre o referido segundo polipéptido e o referido polipéptido de referência;

formar uma combinação ponderada da referida diferença de hidrofobicidade, a referida diferença de propensão na estrutura secundária, a referida diferença de carga, para determinar o referido indicador da velocidade de agregação relativa calculando  $[x*\Delta\text{Hidr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga], em que x, y e z são factores de escala e em que um valor para  $[x*\Delta\text{Hidr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga] de um primeiro sinal indica que o segundo polipéptido tem uma maior propensão para agregar (menor solubilidade) relativamente ao polipéptido de referência e um valor para  $[x*\Delta\text{Hidr}] + [y* (\Delta\Delta G_{\text{espiral-}\alpha} + \Delta\Delta G_{\beta\text{-espiral}})] - [z*\Delta$  Carga] de um sinal oposto ao primeiro sinal indica que o segundo polipéptido tem uma propensão reduzida para agregar (solubilidade mais elevada) relativamente ao polipéptido de referência.

26. Um método como reivindicado na reivindicação 25 em que o referido indicador da velocidade de agregação

relativa compreende um logaritmo de um quociente de velocidades de agregação do referido segundo polipéptido e do referido polipéptido de referência.

27. Um método tal como reivindicado nas reivindicações 25 ou 26 compreendendo adicionalmente determinar os pesos para a referida combinação ponderada, utilizando velocidades de agregação conhecidas para mutações do referido polipéptido de referência.

28. Um método tal como reivindicado nas reivindicações 25, 26 ou 27 compreendendo adicionalmente determinar os pesos para a referida combinação ponderada usando velocidades de agregação conhecidas para mutações de um segundo polipéptido de referência.

29. Um método implementado por computador para seleccionar um polipéptido para síntese que compreende:

criar um conjunto de polipéptidos candidatos a partir de uma sequência polipeptídica de referência;  
determinar uma velocidade de agregação relativa para cada um dos referidos candidatos usando o método de qualquer das reivindicações 25 a 28; e  
seleccionar um referido candidato para a síntese dependendo do resultado da referida determinação das velocidades de agregação relativas.

30. Programa de computador adaptado para executar o método de qualquer das reivindicações 25 a 29.

31. Programa de computador adaptado para executar o método de selecção de um polipéptido para síntese, compreendendo os passos:

criar um conjunto de polipéptidos candidatos a partir de uma sequência polipéptídica de referência;  
determinar a velocidade de agregação relativa para cada um dos referidos candidatos, usando o método de qualquer das reivindicações 25 a 29; e  
seleccionar um referido candidato para a síntese dependendo dos resultados da referida determinação da referida velocidade de agregação relativa.

32. Um suporte de dados legível por computador compreendendo o programa de computador das reivindicações 30 ou 31.

33. Um sistema informático que compreende meios para executar o método das reivindicações 25-29.

Lisboa, 21 de Maio de 2010

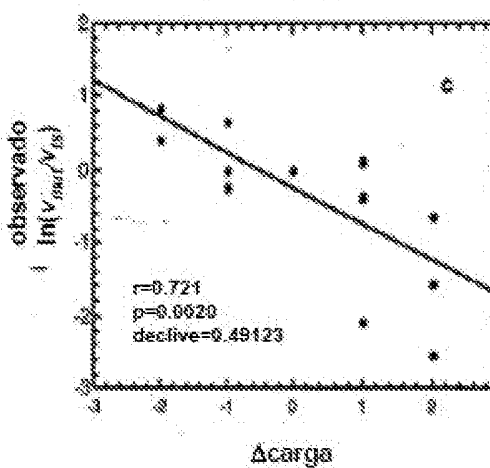
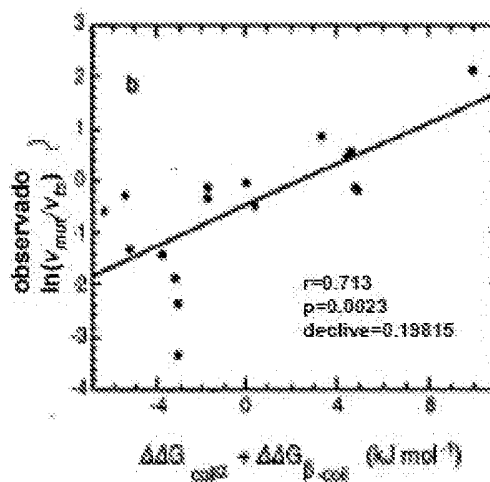
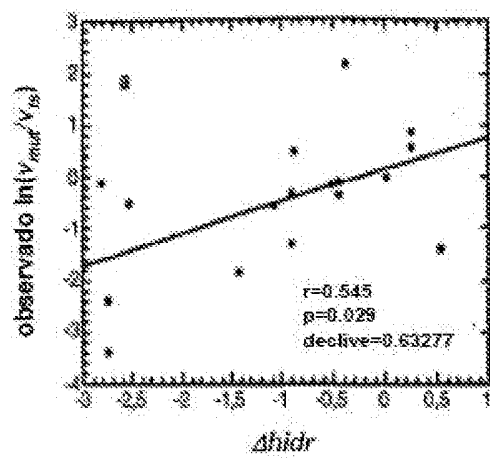


Figura 1

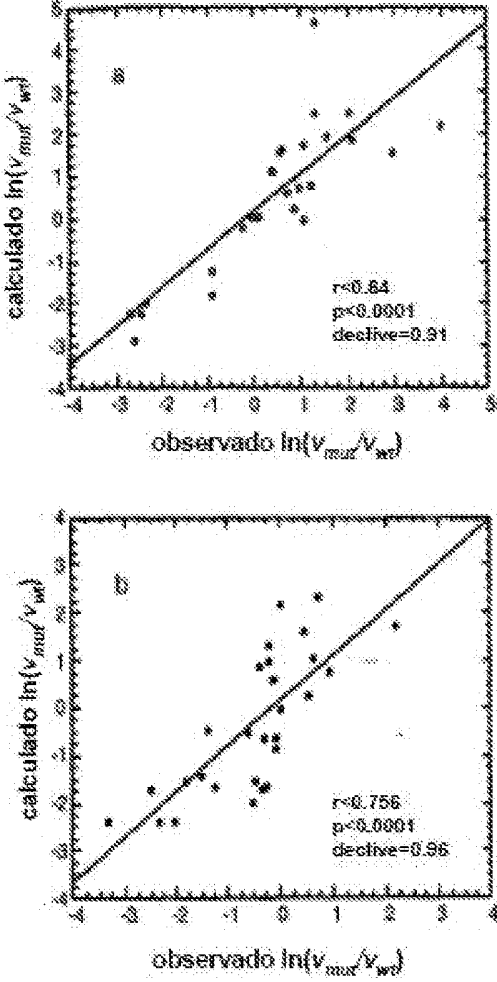


Figura 2

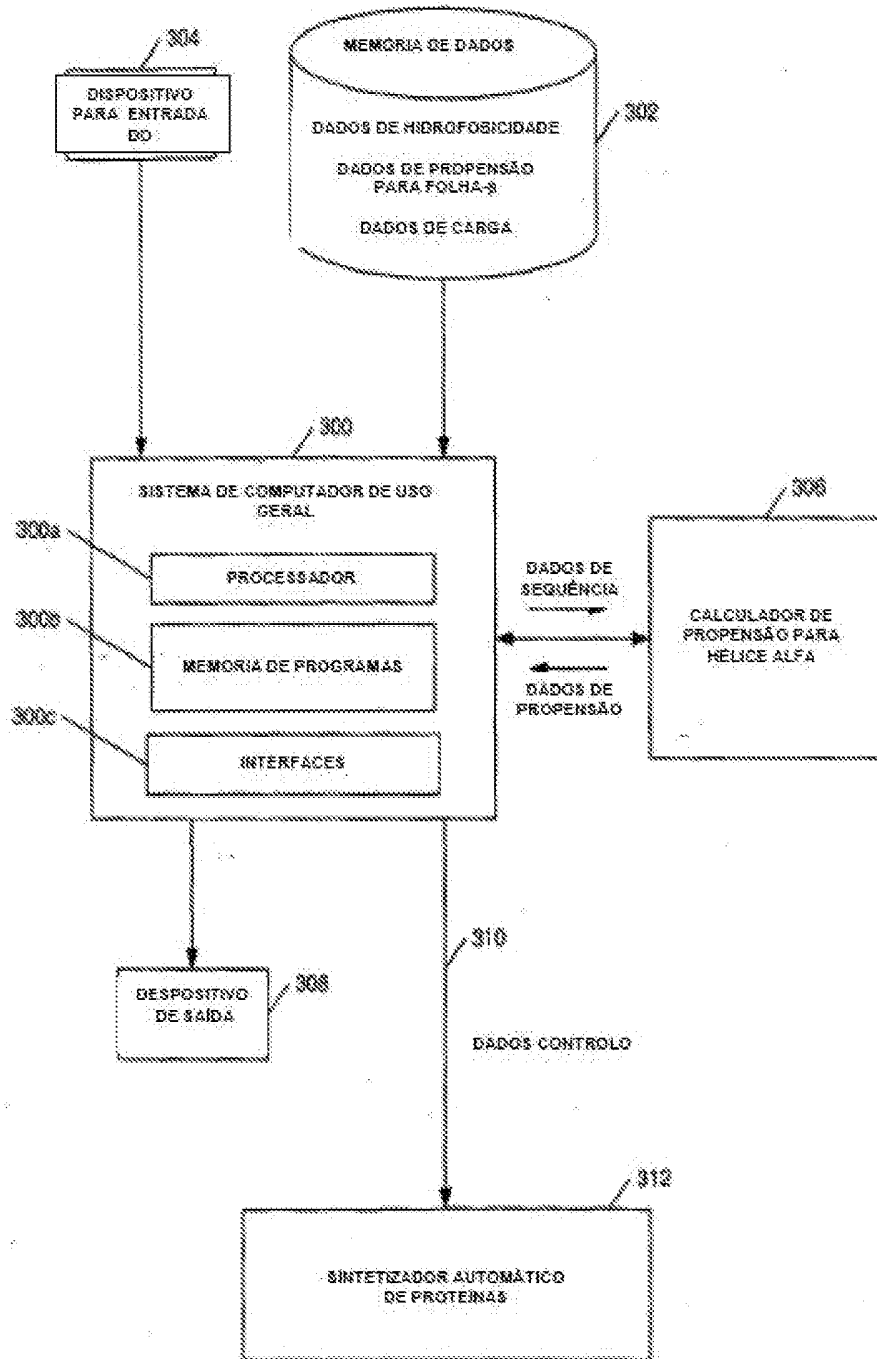


Figura 3

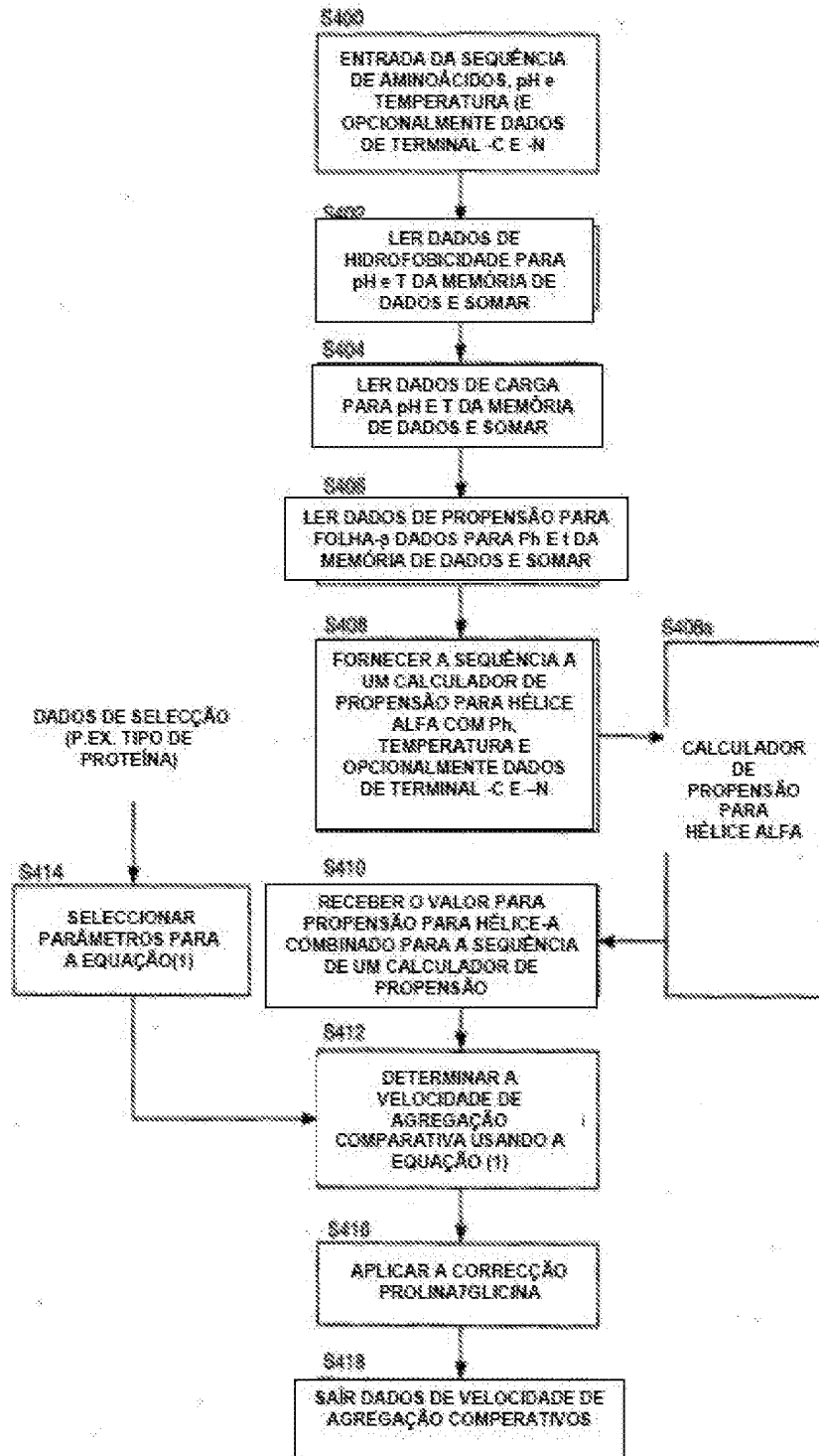


Figura 4

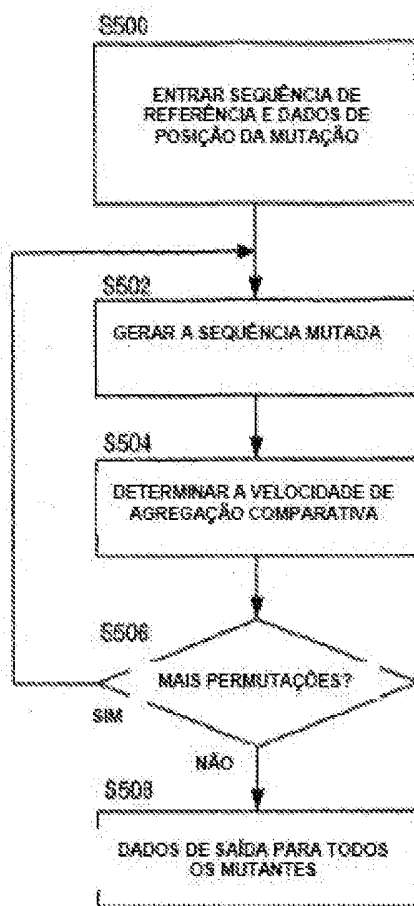


Figura 5

## REFERÊNCIAS CITADAS NA DESCRIÇÃO

Esta lista de referências citadas pelo requerente é apenas para conveniência do leitor. A mesma não faz parte do documento da patente Europeia. Ainda que tenha sido tomado o devido cuidado ao compilar as referências, podem não estar excluídos erros ou omissões e o IEP declina quaisquer responsabilidades a esse respeito.

### Documentos de patentes citadas na Descrição

- WO 0017326 A, Dobson
- WO 0242321 A, Dobson & McPhee
- EP 1255209 A
- WO 02063734 A, Zurdo & Dobson
- GB 0201778 W

### Literatura que não é de patentes citada na Descrição

- Chiti, F.; Taddei, N.; Baroni, F.; Capanni, C.; Stefani, M.; Ramponi, G.; Dobson, C. M. et al. Kinetic partitioning of protein folding and aggregation. *Nature Struct. Biol.*, 2002, vol. 9, 137-143
- 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*, 2002, vol. 99, 16419-16426
- Lopez de Paz et al. Computer-aided design of  $\beta$ -sheet peptides. *J Mol Biol*, 2001, vol. 312, 229-246
- Lopez de Paz et al. De novo designed peptide-based amyloid fibrils. *Proceedings of the national academy of sciences of the United States*, 10 December 2002, vol. 99 (25), 16052-16057
- Siepen et al. The fibril one on-line database: Mutations, experimental conditions, and trends associated with amyloid fibril formation. *Protein Science*, July 2002, vol. 11 (7), 1862-1866
- Avbel et al. Amino acid conformational preferences and salvation of polar backbone atoms in peptides and proteins. *Journal of Molecular Biology, London*, 28 July 2000, vol. 300 (5), 1335-1359
- Avbel et al. Role of main-chain electrostatics, hydrophobic effect and side-chain conformational entropy in determining the secondary structure of proteins. *Journal of Molecular Biology, London*, 12 June 1996, vol. 279 (3), 665-684
- Cowan, R.; Whittaker, R.G. *Peptide Research*, 1990, vol. 3, 75-80
- Fauchere J.-L.; Pliska V.E. *Eur. J. Med. Chem.*, 1983, vol. 18, 369-375
- Kyte J.; Doolittle R.F. *J. Mol. Biol.*, 1982, vol. 157, 105-132
- Creighton, T. E. *Proteins. Structure and molecular properties*. W. H. Freeman & Company, 1983, 154
- Munoz ; Serrano. *Nature Structural Biol*, 1994, vol. 1, 393-409
- Muñoz ; Serrano. *J Mol Biol*, 1994, vol. 245, 297-308
- Muñoz ; Serrano. *Biopolymers*, 1997, vol. 41, 495-509
- Lacroix et al. *J Mol Biol*, 1998, vol. 284, 173-191
- Rost, B. et al. *J Mol Biol*, 1993, vol. 232, 584-589
- Rost, B. et al. *Methods Enzymol*, 1996, vol. 266, 525-539
- Garnier J et al. *J Mol Biol*, 1978, vol. 120, 97-120
- Garnier J et al. *Methods Enzymol*, 1996, vol. 266, 540-553
- J. Garnier ; J.F. Gibrat ; B. Robson. *Methods Enzymol*, 1996, vol. 266, 540-553
- J. Garnier ; D. Osguthorpe ; B. Robson. *J. Mol. Biol.*, 1978, vol. 120, 97
- *J Mol Biol*, 05 December 1987, vol. 198 (3), 425-443
- Rost B ; Sander C. *J Mol Biol*, 20 July 1993, vol. 232 (2), 584-599
- Rost B ; Sander C. *Proteins*, May 1994, vol. 19 (1), 55-72
- Frishman D ; Argos P. *Protein Eng*, February 1996, vol. 9 (2), 133-142
- Levin JM ; Robson B ; Garnier J. *FEBS Lett*, 15 September 1986, vol. 205 (2), 303-308
- J. LEVIN ; J. GARNIER. *Biochim. Biophys. Acta*, 1986, vol. 955, 283-295
- Levin JM. *Protein Eng.*, 1997, vol. 7, 771-776
- Geourjon C ; Deleage G. *Protein Eng*, February 1994, vol. 7 (2), 157-164
- Geourjon C ; Deleage G. *Comput Appl Biosci*, December 1995, vol. 11 (6), 661-684
- Street, A. G. ; Mayo, S. L. Intrinsic  $\beta$ -sheet propensities result from van der Waals interactions between side chains and the local backbone. *Proc. Natl. Acad. Sci. USA*, 1999, vol. 96, 9074-9078
- 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.*, 2002, vol. 9, 137-143

- Lacroix, E.; Viguera AR; Serrano, L. *J. Mol. Biol.*, 1998, vol. 284, 173-191
- Arvinte et al. *J Biol Chem*, 1993, vol. 268, 6415-6422
- 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*, 1993, vol. 32, 4693-4697
- Dobson, C. M. Protein folding and its links with human disease. *Biochem. Soc. Symp.*, 2001, vol. 68, 1-28
- 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*, vol. 32, 4693-4697
- Selkoe, D. J. Alzheimer's disease: genes, proteins, and therapy. *Physiol. Rev.*, 2001, vol. 81, 741-769
- Siepen, J. A.; Westhead, D. R. The fibril\_one on-line database: Mutations, experimental conditions, and trends associated with amyloid fibril formation. *Protein Sci.*, 2002, vol. 11, 1862-1866
- Volles, M. J.; Lansbury, P. T. Jr. Vesicle permeabilization by protofibrillar  $\alpha$ -synuclein is sensitive to Parkinson's disease-linked mutations and occurs by a pore-like mechanism. *Biochemistry*, 2002, vol. 41, 4595-4602
- 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.*, 2001, vol. 276, 34158-34161
- Sakagashira, S.; Hiddinga, H. J.; Tateishi, K.; Sanke, T.; Hanabusa, T.; Nanjo, K.; Eberhardt, N. L. 520G mutant amylin exhibits increased in vitro amyloidogenicity and increased intracellular cytotoxicity compared to wild-type amylin. *Am. J. Pathol.*, 2000, vol. 157, 2101-2109
- Salmona, M.; Malesani, P.; De Gioia, L.; Gorla, S.; Bruschi, M.; Molinari, A.; Della Vedova, F.; Pedrotti, B.; Marrari, M. A.; Awan, T. Molecular determinants of the physicochemical properties of a critical prion protein region comprising residues 106-126. *Biochem J.*, 1999, vol. 342, 207-214
- Thompson, A. J.; Barnham, 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.*, 2001, vol. 1544, 242-254
- Conway, K. A.; Lee, S. J.; Rochet, J. C.; Ding, T. T.; Williamson, R. E.; Lansbury, P. T. Jr. 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*, 2000, vol. 97, 571-576
- Giasson, B. I.; Murray, I. V.; Trojanowski, J. Q.; Lee, V. M. A hydrophobic stretch of 12 amino acid residues in the middle of  $\alpha$ -synuclein is essential for filament assembly. *J. Biol. Chem.*, 2001, vol. 276, 2360-2366
- Van Nostrand, W. E.; Melchor, J. P.; Cho, H. S.; Greenberg, S. M.; Rebeck, G. W. Pathogenic effects of D23N Iowa mutant amyloid beta -protein. *J. Biol. Chem.*, 2001, vol. 276, 32863-32866 [6089]
- Miravalle L; Tokuda T; Chiarle R; Giaccone G; Bugiani O; Tagliavini F; Frangione B; Ghiso J. *J Biol Chem*, 01 September 2000, vol. 275 (35), 27110-6
- Nilssonberth, C.; Westlind-Danielsson, A.; Eckman, C. B.; Condon, M. M.; Axelman, K.; Forsell, C.; Sten, C.; Luthman, J.; Teplow, D. B.; Younkin, S. G. 'Arctic' APP mutation (E693G) causes Alzheimer's disease by enhanced A $\beta$  protofibril formation. *Nature Neurosci.*, 2001, vol. 4, 687-693
- 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  $\beta$ -amyloid congener disrupts peptide folding and abolishes plaque competence. *Biochemistry*, 1996, vol. 35, 13914-13921
- Barghorn, S.; Zheng-Fischhofer, Q.; Ackmann, M.; Biernat, J.; von Bergen, M.; Mandelkow, E. M.; Mandelkow, E. Structure, microtubule interactions, and paired helical filament aggregation by tau mutants of frontotemporal dementias. *Biochemistry*, 2000, vol. 39, 11714-11721
- Gambiin, 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*, 2000, vol. 39, 6136-6144
- 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.*, 1999, vol. 447, 195-199
- 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.*, 2002

- 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  $\beta$ -protein. *FEBS Lett.*, 1997, vol. 412, 397-403
- Orpiszewski, J. ; Benson, M. B. Induction of beta-sheet structure in amyloidogenic peptides by neutralization of aspartate: a model for amyloid nucleation. *J. Mol. Biol.*, 1999, vol. 289, 413-428