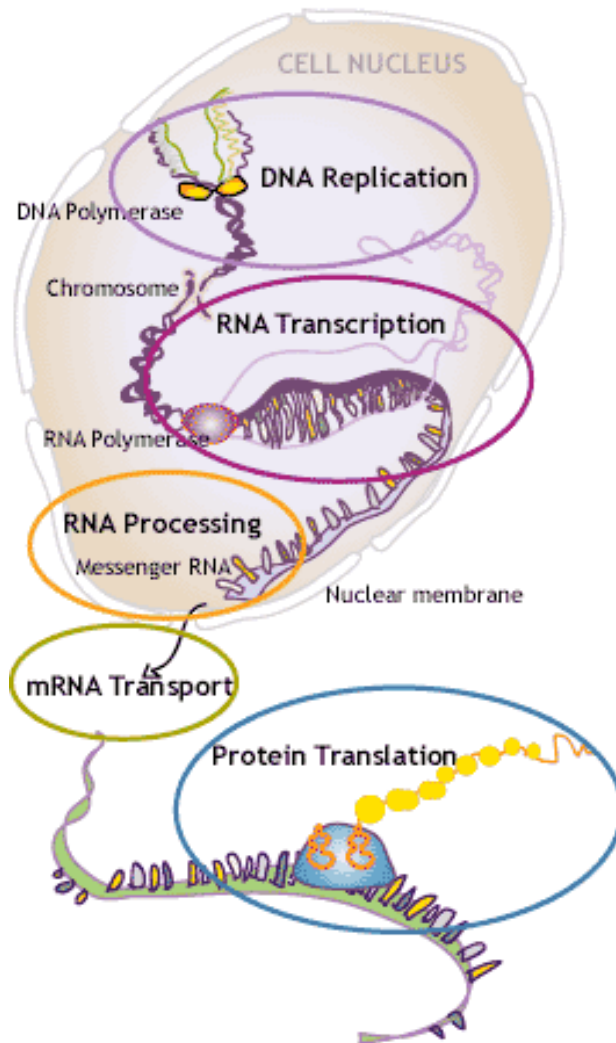


# Folding@Home



Distributed Computing...

# Introdução

O Presente tema surgiu numa aula de Físico-Química, na qual se discutia se o Homem era capaz de reduzir o seu consumo de energia.

Quando entrevi, disse que o Homem não o conseguia fazer, mas que podia aproveitar melhor os recursos utilizados. Então, dei o exemplo da computação distribuída, explicitando em que consistia: uma pessoa que tivesse o seu computador ligado muitas horas por dia a fazer downloads da internet ou simplesmente sem fazer nada, podia participar num projecto de computação distribuída, porque assim o seu computador podia estar a ajudar a descobrir a cura para várias doenças como o cancro, Alzheimer, Parkinson ou BSE.

Não mostrando a professora muito interesse por esta questão, decidi efectuar uma pesquisa na Internet sobre o Folding@Home como forma de fornecer esclarecimentos sobre este assunto.

Neste trabalho vou abordar as seguintes questões:

- O que é o Folding@Home?
- O que é a computação distribuída?
- Enrolamento de proteínas?
- Como posso contribuir?/Como funciona?
- Quem pode foldar?
- O que é a Portugal@Folding?
- Que resultados já existem?
- Que futuro tem a computação distribuída?
- O que é que eu tenho a ver com isso?

Aproveito, para agradecer ao **Metro**, ao **WindWalker**, ao **amjpereira**, ao **Madcaddie** e a todos os outros utilizadores do fórum Techzone que me ajudaram com este trabalho.

## O que é o Folding@Home?

O Folding@Home é um projecto de computação distribuída, dirigido pelo professor de química Vijay Pande da Universidade de Stanford, nos Estados Unidos da América.

É um projecto não lucrativo que se iniciou no dia 1 de Outubro de 2000 e visa conhecer melhor a forma como as proteínas que fazem parte do nosso organismo se enrolam para levar a cabo as suas funções que são muitas e variadas, e as suas respectivas doenças. Estas podem ser Alzheimer, Parkinson, BSE ou mesmo cancro.

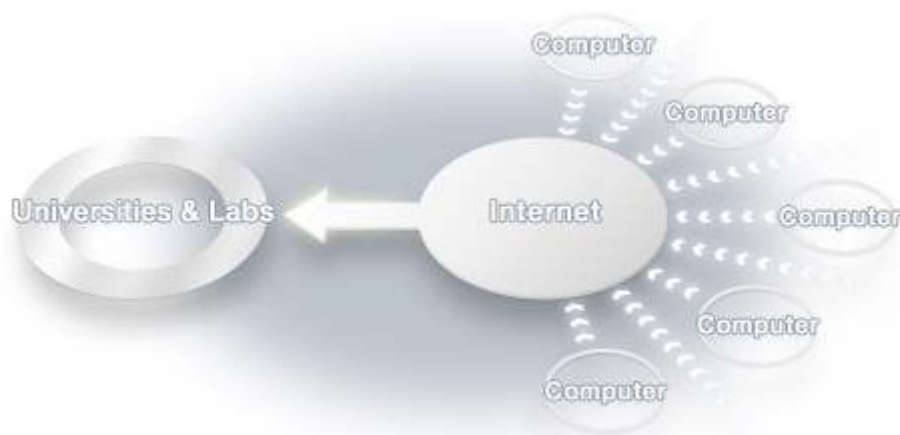
Os resultados deste projecto são depois publicados em revistas da especialidade. É o projecto de computação distribuída com mais resultados publicados e com resultados muito promissores.

## O que é a Computação Distribuída?

A computação distribuída consiste em processar partes diferentes de uma tarefa de um computador usando mais do que um computador. A tarefa é dividida em múltiplas partes: enquanto um computador do Japão processa uma parte, um computador da Argentina faz outra, por exemplo. Mesmo que o utilizador do Japão tenha um Sistema Operativo diferente no seu computador, eles podem dividir o trabalho e terminá-lo em metade do tempo que demoraria num único computador.

É uma abordagem muito inteligente usada por algumas universidades e institutos de pesquisa para usar o imenso poder de computação espalhado à volta do globo, que normalmente não é utilizado (aproximadamente apenas 5 a 10% do poder de computação é usado). A utilização deste método evita ter supercomputadores caríssimos que tinham de fazer o trabalho sozinho, e levariam uma eternidade a realizar esse trabalho.

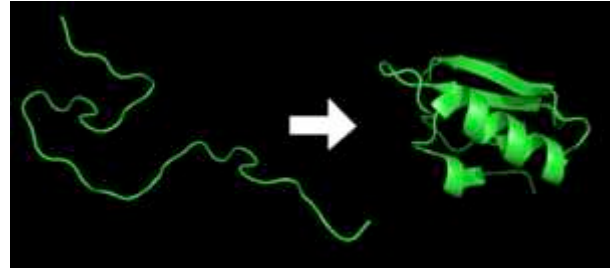
Desta maneira, no fim de cada computador terminar a sua tarefa, os resultados regressam a um servidor, que junta todos os resultados. Retirada do site da Sony, esta imagem explica a computação distribuída de uma maneira muito clara:



Os “pequenos computadores” representam cada voluntário que se juntou a esta causa. Na maioria dos projectos, participa-se realizando o download de um cliente, configurá-lo e deixar o computador fazer o trabalho (computacional).

## Enrolamento de proteínas?

Hoje em dia um dos principais dogmas da biologia juntamente com a teoria da evolução e a teoria celular é este processo: DNA --> RNA --> Proteínas. Este processo é diferente entre Procariotas e Eucariotas. A existência de um núcleo nos Eucariotas obriga a que o produto da transcrição (DNA --> RNA) apresente algumas diferenças em comparação com o produto dos Procariotas porque o RNA precisa de estabilidade suficiente para conseguir atravessar a membrana nuclear para depois ser traduzido em proteína...



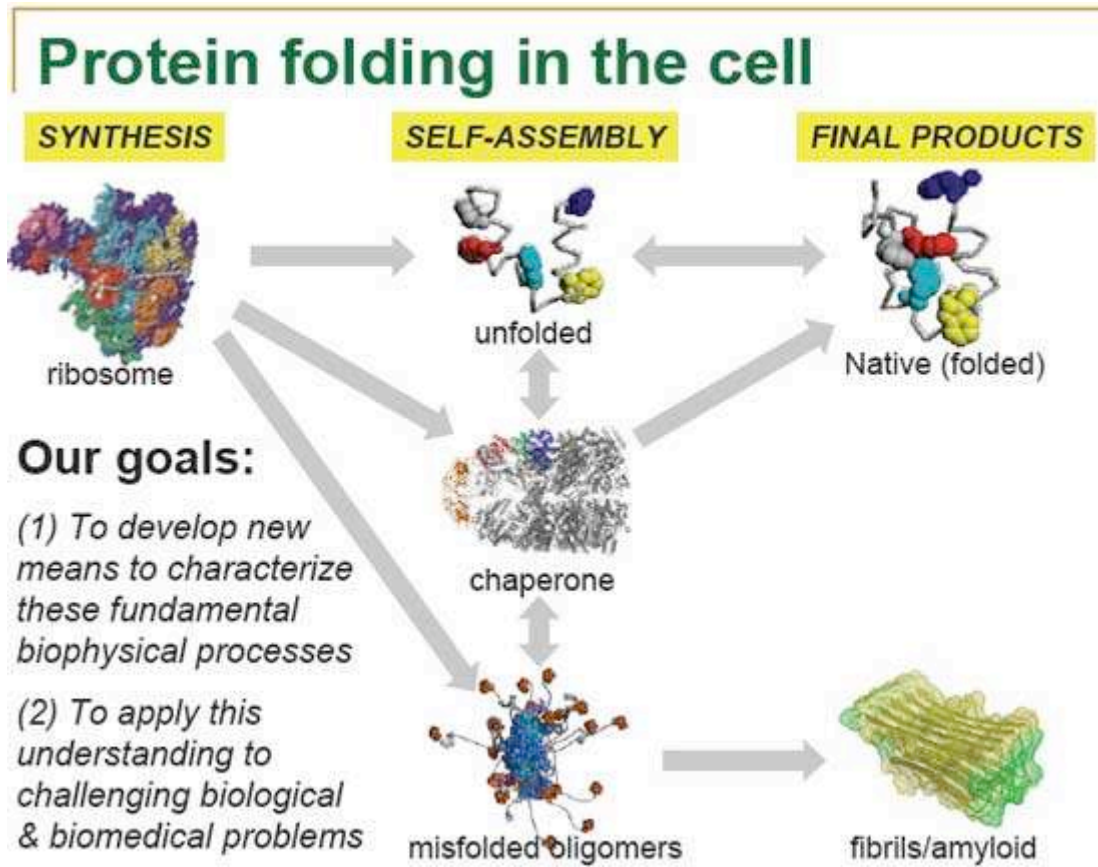
As Proteínas são compostos orgânicos de estrutura complexa e massa molecular elevada (de 5.000 a 1.000.000 ou mais unidades de massa atômica), sintetizadas pelos organismos vivos através da condensação de um grande número de moléculas de  $\alpha$ -aminoácidos, através de ligações denominadas ligações Peptídicas. Uma proteína é um conjunto de 100 ou mais aminoácidos, sendo os conjuntos menores denominados Polipeptídeos.

Nos Eucariotas, o DNA dá origem ao mRNA (RNA mensageiro) dentro do núcleo. Este adquire uma cauda poli-A na extremidade 3' que uma Cap na extremidade 5'. Ele necessita disto para depois sair do núcleo para o citoplasma da célula onde vai ocorrer a tradução desse mRNA. Durante essa tradução ou com ela já finalizada é que vão actuar as Chaperoninas primeiro e depois as Chaperonas no caso da primeira conformação estar errada. Quando adquirem essa conformação diminuem a probabilidade de serem degradadas e facilita o seu transporte para locais específicos da célula.

Por exemplo, os nossos anticorpos (glóbulos brancos) são uma proteína. Para ela conseguir eliminar os corpos estranhos/maliciosos, essa proteína, enrola-se de múltiplas formas consoante o corpo estranho de modo a eliminá-lo. Se esse enrolamento não for correcto, origina-se uma doença.

Resumindo, o Folding calcula energias potenciais dos átomos nas proteínas para prever as suas localizações ao fim de certo intervalo de tempo de interacção. Assim pode-se saber a forma da proteína no fim do seu enrolamento - folding. Sabendo os factores que impedem o correcto enrolamento, pretende-se desenvolver substâncias que o corrijam. Funciona como o modelo da nuvem electrónica, ou seja, probabilidades.

Na imagem seguinte, pode-se observar como funciona o enrolamento das proteínas:



## Como posso contribuir?

Para contribuir para este projecto, basta instalar uma pequena aplicação chamada cliente, e ao fim de uns minutos já estão a participar!

O cliente só está disponível no site oficial do projecto e existem dois tipos de clientes:

- Gráfico: é um cliente que funciona com OpenGL e que consiste num programa onde aparece uma proteína a ser processada e do outro lado informações sobre aquilo que o computador está a processar.
- Texto: consiste numa pequena janela de DOS em que aparece o estado do processo. Para quem costuma jogar, este é o recomendado uma vez que o OpenGL não consegue executar 2 aplicações ao mesmo tempo.

## Como Funciona?

O Folding@Home está organizado num sistema de crédito, tanto individualmente, como colectivamente.

No fim de instalarem o cliente, e o iniciarem pela primeira vez, tem de escolher um nickname e uma equipa. O nickname é a vossa escolha, quanto às equipas, falaremos disso mais à frente.

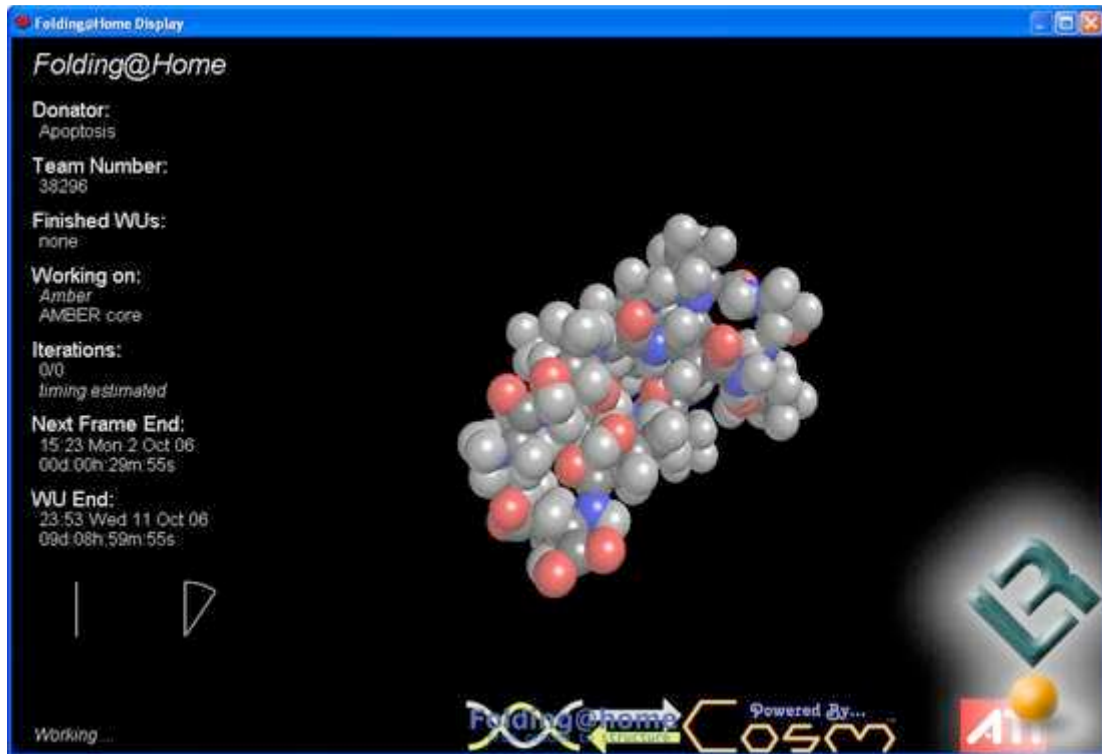
Primeiro que tudo, o cliente funciona em baixa prioridade, ou seja, utiliza recursos que não estão a ser utilizados. Por exemplo, enquanto estão na internet, não estão a usar mais de 10% do vosso processador. Deste modo, o cliente utilizará os 90% restantes. Se entrarem numa aplicação que utilize 70% do processador, o cliente passa automaticamente a usar os restantes 30%. No dia-a-dia, nem se nota que o programa está a correr.

No fim de escreverem o nickname e a equipa para a qual estão a participar, o cliente irá ligar automaticamente ao servidor do projecto e irá fazer o download de uma proteína, ou parte dela. A esta proteína, dá-se o nome de Work Unit (Unidade de Trabalho). No fim de acabarem uma WU, o vosso computador enviará a proteína já processada para o servidor do Folding@Home e receberão outra, e assim sucessivamente.

O tráfego gerado pelo cliente é muito pequeno. Uma WU tem entre 1 a 5Mb de tamanho, sendo que as maiores levam vários dias a ser processadas.

Normalmente, quem participa neste projecto, tem o computador ligado 24 horas por dia, mas ninguém é obrigado a fazê-lo. A única coisa que é preciso ter em atenção é que as WU's têm uma deadline, ou seja, só durante um tempo é que são válidas.

Nesta imagem, é visível a forma como funciona o cliente gráfico para Windows:





## Quem pode foldar?

O número de processadores activos tem vindo a aumentar desde o lançamento do Folding@Home. Dantes, apenas os processadores dos computadores podiam ser usados para processar as WU's. Hoje em dia, alguns dos processadores gráficos, usados em placas gráficas para renderizar gráficos 3D (como aqueles nos jogos) também são suportadas pelo Folding@Home. Mais recentemente, esta possibilidade foi estendida à Playstation 3, uma consola de jogos da Sony. O contributo da Playstation 3 foi tão grande (uma Playstation 3 pode facilmente processar uma WU 30 a 40 vezes mais rápido do que um computador) que, em vez dos poucos números que contribuía para o projecto, os servidores de Stanford tiveram de ser actualizados devido ao excesso de tráfego.

À medida que novos projectos precisam de mais poder computacional, este aumento e extensão do hardware é fundamental.

A terrível verdade na simulação de moléculas dinâmicas é que é preciso muito tempo e poder de computação para simular uma pequena fracção do tempo real. Só para dar uma ideia, é estimado que um 1 ano de simulações é necessário para simular um nanosegundo do enrolamento de uma proteína. Felizmente, ou não, a escala de tempo do enrolamento de proteínas é em nanosegundos.

Sistema Operativo	TFLOPS	CPU's	CPU's
	Actuais	Activos	Totais
Windows	188	197224	2015907
Mac OS X/PowerPC	7	8523	115303
Mac OS X/Intel	25	8024	47912
Linux	54	31793	293583
GPU	101	1459	7372
PLAYSTATION®3	1704	56408	512892
<b>Total</b>	<b>2079</b>	<b>303431</b>	<b>2992969</b>

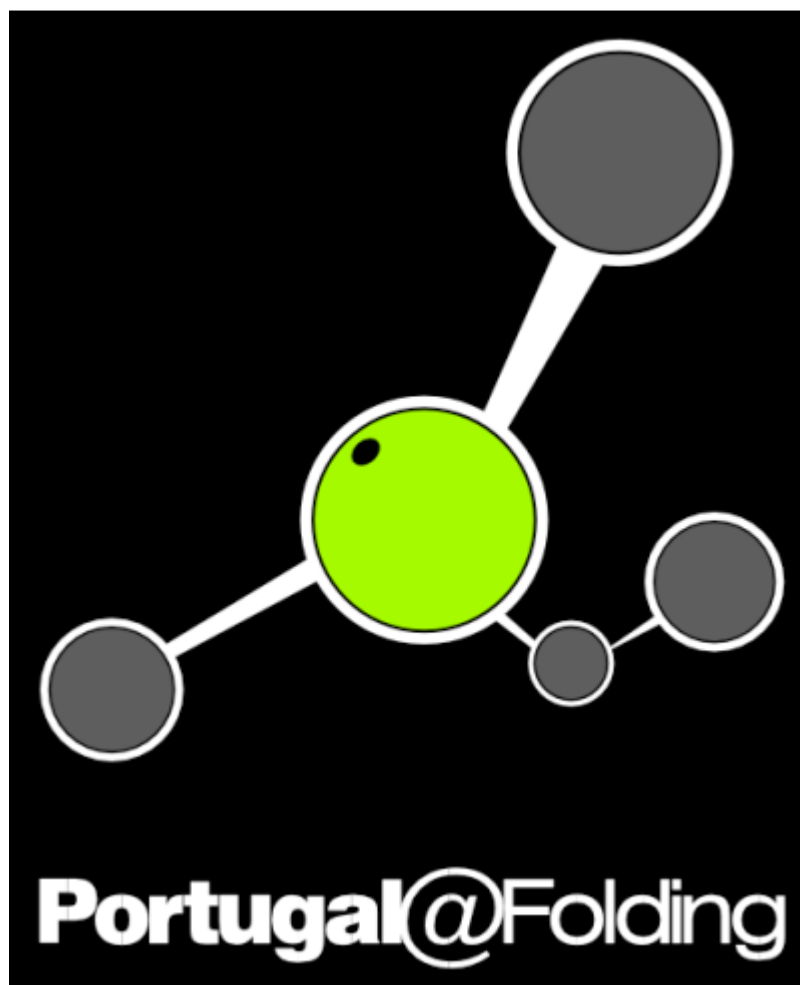
Produção dos vários clientes

## Portugal@Folding

Como disse anteriormente, o folding está organizado em equipas. Portugal@Folding é o nome da maior equipa portuguesa que participa neste projecto. Neste momento está na posição 34 do ranking mundial (já estivemos em 27, mas lentamente, temos vindo a descer).

O número da equipa Portugal@Folding é o 35271. Para contribuir para a Portugal@Folding, quando estão a configurar o cliente, dizem que o número da vossa equipa é o 35271.

A Portugal@Folding conta com cerca de 1500 processadores activos.



# Resultados

Por falar em resultados, não foram muitos até agora. Por um lado, o conceito de computação distribuída é ainda muito recente e normalmente as pessoas são muito relutantes para contribuir (uma vez que envolve tráfego na Internet, as pessoas pensam que é uma maneira de entrar nos seus computadores). Mas pelo outro lado, espera-se que apareçam daqui a uma ou duas décadas de computação. Até ao momento, foram publicados 54 trabalhos científicos, baseado em todo o trabalho efectuado pelos colaboradores. Tal como se pode observar:

- Em 2005, foram publicados os primeiros resultados do estudo do supressor de tumores p53. As mutações do p53 são a causa para quase todos os tipos de cancro. Sabendo como ela se enrola, é mais fácil analisar a doença e as possíveis maneiras de ser curada.
- De 2004 a 2006, novos métodos de ligação de proteínas foram descobertas. São formas muito eficazes na produção de novas drogas.
- Em 2006, resultados concentrados no enrolamento de proteínas em espaços confinados foram publicados. Parece que a água reage de uma forma muito diferente do que era esperado.

## O Futuro

O futuro dos projectos de computação distribuída é ainda indefinido. Existem previsões positivas e negativas. Alguns dizem que nunca irá prevalecer e que só estamos a desperdiçar recursos computacionais.

Porém, a maioria diz que será apenas uma questão de anos para confirmar as suposições: os resultados vão aparecer em breve. Talvez numa década, talvez mais do que isso.

Recentemente, o supervisor do projecto, Vijay Pande deu uma entrevista, referindo o facto que iria levar 10 a 15 anos a encontrar e criar, por exemplo, um inibidor para a doença de Alzheimer. Ele também disse que a sua equipa está limitada pelo poder de computação disponível para o projecto. Citando o Professor Pande “Nós temos planos específicos prontos para por em acção, assim que atingirmos determinados níveis de processadores activos, mais de 1 milhão de processadores. Assim que nós começemos a ter perto de 1 milhão de processadores activos, nós vamo-nos sentar e fazer planos para ir além disso. A nossa pesquisa é muito limitada.”

No entanto, este problema aparenta ter uma solução num futuro próximo. Não são apenas os processadores de computador que estão a evoluir e a ter um maior poder de computação, mas também existe a ajuda das consolas de vídeo de nova geração, que estão a providenciar clientes para estes projectos de computação distribuída, então estes podem ser executados quando a consola não está a executar nenhum jogo. Estes processadores, tem uma arquitectura diferente da dos os computadores de secretária, mas parece que tem uma performance superior quando comparados com os vulgares processadores dos computadores de secretária.

Juntos, estes processadores constituem um imenso poder computacional, poder esse que é requerido pelo professor Vijay Pande e pela sua equipa.

## O que é que eu tenho a ver com isso?

Se têm alguém na família ou mesmo chegado, algum amigo ou conhecido com uma doença que teve a sua origem no enrolamento incorrecto das proteínas, sendo os mais conhecidos, Alzheimer, Parkinson, entre outros, sabem que eles para controlar o "avançar" da doença devem tomar medicação, medicação essa que não caiu do céu por obra e graça do espírito santo.

Foi através de pesquisas, feitas por cientistas e investigadores, e o Folding@home é isso mesmo, o estudo de como criar medicação que ajude a retardar os efeitos negativos de tais doenças e quem sabe daqui a uns anos a cura para a mesma!

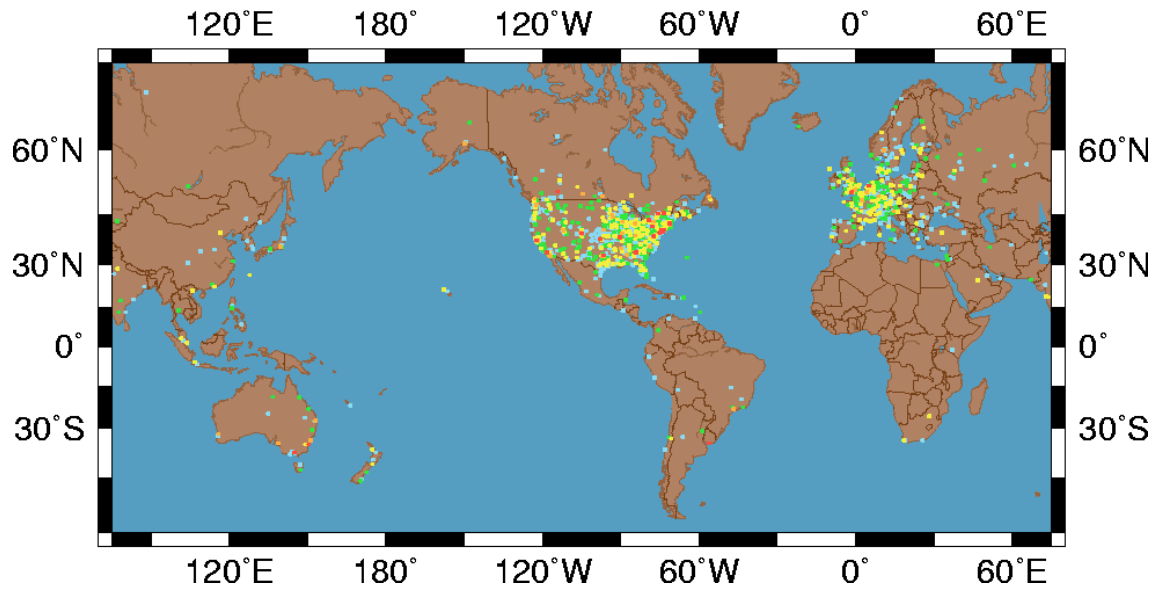
Quando e caso tenham o azar de algum familiar precisar ou mesmo VOCÊS, não gostariam de pensar que contribuíram para que aquela nova medicação "revolucionária" que saiu há uns meses/dias para essa doença, saiu porque vocês, ajudaram?

Porque se todos neste país, ou mesmo no mundo, fossem a pensar como vocês "E o que é que eu tenho a ver com isso?", então não havia qualquer tipo de avanços.

**AJUDEM, NÃO CUSTA NADA!**

## Conclusão

Espero que com este trabalho, vos tenha dado a entender melhor o que é o Folding@Home e a computação distribuída, dos seus benefícios para a humanidade e que existem resultados. Para não acharem que só “uns malucos” é que participam neste projecto, vou deixar-vos um mapa do mundo em, que os pontinhos representam um ou mais foldadores. Quanto mais escura a cor, mais foldadores existem nesse ponto.



# Bibliografia

- Site oficial do Folding@Home  
<http://folding.stanford.edu/>
- Site oficial da Portugal@Folding  
<http://www.portugalfolding.org/>
- Wikipedia  
<http://pt.wikipedia.org/>
- Site oficial do Folding@Home na Playstation 3  
<http://www.scei.co.jp/folding/en/index.html/>

# Índice

<b>Introdução</b> .....	<b>2</b>
<b>O que é o Folding@Home?</b> .....	<b>3</b>
<b>O que é a Computação Distribuída?</b> .....	<b>4</b>
<b>Enrolamento de proteínas?</b> .....	<b>5</b>
<b>Como posso contribuir?</b> .....	<b>7</b>
<b>Como funciona?</b> .....	<b>7</b>
<b>Quem pode foldar?</b> .....	<b>9</b>
<b>Portugal@Folding</b> .....	<b>10</b>
<b>Resultados</b> .....	<b>11</b>
<b>O Futuro</b> .....	<b>12</b>
<b>O que é que eu tenho a ver com isso?</b> .....	<b>13</b>
<b>Conclusão</b> .....	<b>14</b>
<b>Bibliografia</b> .....	<b>15</b>