

PONTIFÍCIA UNIVERSIDADE CATÓLICA DO RIO GRANDE DO SUL
FACULDADE DE BIOCÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM BIOLOGIA CELULAR E MOLECULAR

CARLOS EDUARDO SEQUEIROS BORJA

PREDIÇÃO DA ESTRUTURA 3D DE PROTEÍNAS
MIMETIZANDO O AMBIENTE RIBOSSÔMICO

PORTO ALEGRE
2017

CARLOS EDUARDO SEQUEIROS BORJA

PREDIÇÃO DA ESTRUTURA 3D DE PROTEÍNAS
MIMETIZANDO O AMBIENTE RIBOSSÔMICO

Dissertação apresentada como requisito para a obtenção do título de Mestre em Biologia Celular e Molecular pelo Programa de Pós-Graduação em Biologia Celular e Molecular da Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul.

ORIENTADOR: Prof. Dr. OSMAR NORBERTO DE SOUZA

PORTO ALEGRE
2017

CARLOS EDUARDO SEQUEIROS BORJA

PREDIÇÃO DA ESTRUTURA 3D DE PROTEÍNAS
MIMETIZANDO O AMBIENTE RIBOSSÔMICO

Dissertação apresentada como requisito para a obtenção do título de Mestre em Biologia Celular e Molecular pelo Programa de Pós-Graduação em Biologia Celular e Molecular da Faculdade de Biociências da Pontifícia Universidade Católica do Rio Grande do Sul.

Aprovada em: 23 de fevereiro de 2017.

BANCA EXAMINADORA:

Prof. Dr. Pedro Henrique Monteiro Torres – FIOCRUZ

Prof. Dr. Rafael Andrade Caceres - UFCSPA

Prof. Dr. Luiz Augusto Basso - PUCRS

PORTO ALEGRE
2017

AGRADECIMENTOS

Inicialmente gostaria de agradecer ao Prof. Dr. Osmar Norberto de Souza pela oportunidade e orientação ao longo desses dois anos. Pela confiança num estudante estrangeiro e liberdade para trabalhar no seu laboratório com autonomia. Pela amizade e apoio para me fazer sentir bem vindo num ambiente totalmente novo.

A minha família que encontra-se longe de mim, mas sempre ficaram perto pelas oportunidades e boa educação que me forneceram ao longo da minha vida, eu não seria nada sem sua guia, confiança e amor durante todo o tempo.

Ao meu amigo o Dr. Luís Fernando Saraiva Macedo Timmers por toda a ajuda fornecida para a realização desta dissertação. Pelo apoio moral, acadêmico e pessoal que me ajudou a resolver muitos problemas ocorridos no transcurso do mestrado. Pelas risadas, saídas e amizade que me fez sentir em casa.

Ao meu amigo o Dr. José Fernando Ruggiero Bacheга pela ajuda e ideias para a realização dos experimentos desta pesquisa. Pela amizade e confiança que me deu desde o primeiro dia de me conhecer.

A minha amiga a MSc. Michele dos Santos da Silva pela amizade e ajuda na parte de programação descritas nesta dissertação. Sem sua ajuda eu teria ficado perdido na geração de scripts e automatização das análises.

Ao meu amigo o Dr. Thiago Lipinski-Paes pela amizade e discussão da análise dos resultados e pela guia nos experimentos de REMD desta pesquisa.

A todos os colegas do LABIO e FARMINF, Vanessa, Walter, Gustavo, Rafael, Paula, Isabel e Igor pela amizade e o sentimento que me fez sentir em família.

O presente trabalho foi realizado com apoio do Programa Estudantes-Convênio de Pós-Graduação – PEC-PG, da CAPES/CNPq – Brasil, chamada N° 10/2014 - PEC - PG 2014, Processo 190602/2014-7.

RESUMO

A predição de estrutura 3D de proteínas partindo apenas da sequência de aminoácidos ainda é um grande desafio em bioinformática estrutural. Apesar da dificuldade, a predição precisa de ser acurada e rápida. Nesta dissertação, propõe-se e mesuram-se os efeitos da co-tradução e o uso de um modelo ideal de canal ribossomal na predição da estrutura 3D de proteínas, fazendo uso de dinâmica molecular clássica e dinâmica molecular com intercambio de réplicas. O modelo do canal ribossomal construído foi testado com diferentes velocidades de tradução, e os resultados foram comparados com simulações padrão. Foram testadas diferentes velocidades de tradução para verificar sua influência nas predições, e as velocidades que apresentaram os melhores resultados ficaram na faixa de 80 até 200 ps. A qualidade dos modelos preditos foram boas, apresentando valores de GDT-TS de 1,0, assim como 0,3 Å para RMSD para simulações de apenas 50 ns. No geral, demonstra-se que o uso desta abordagem na predição da estrutura de proteínas, produz satisfatoriamente estruturas nativas ou perto da nativa em três de quatro proteínas testadas, atingindo assim a acurácia e velocidade esperadas. Como conclusão, o uso da co-tradução com um modelo do canal ribossomal é uma abordagem promissora para a predição de estruturas de mini proteínas perto da estrutura nativa. Melhoras na metodologia e no modelo permitirão uma predição de estruturas 3D de proteínas maiores de interesse biológico e biomédico.

Palavras-chave: Modelo do canal ribossomal. Co-tradução. Predição da Estrutura 3D de Proteínas. Dinâmica Molecular. Dinâmica Molecular com Intercambio de Réplicas.

ABSTRACT

Protein structure prediction from just the amino acid sequence continues to be a major challenge in structural bioinformatics. If at all possible, prediction needs to be accurate and fast. In this project, it is proposed and tested the effects of cotranslation within an ideal ribosomal channel model in protein structure prediction using classical molecular dynamics and replica-exchange molecular dynamics simulations. An ideal ribosomal channel model was built, different translation speeds were used and compared the results to control simulations. Different translation speeds were tested to verify their influence on predictions, and the best results were observed at translation speeds between 80 and 200 ps. The quality of the predicted models were as low as 0.3 Å and 1.0 for the RMSDs and GDT-TS parameters, respectively, for simulations of just 50 ns. Overall, the use of this approach to protein structure prediction has successfully produced native and near-native structures in three of the four proteins investigated, thus reaching accuracy and speed as expected. As a conclusion, using cotranslation within an IRCM is a promising approach to predict native-like 3D structures of mini-proteins successfully. Improvements to the methodology should allow the prediction of 3D structures of larger proteins of biological and biomedical interest.

Keywords: Ribosomal Channel Model, Cotranslation, Protein Structure Prediction, Molecular Dynamics, Replica Exchange Molecular Dynamics.

LISTA DE ILUSTRAÇÕES

- Figura 1. Níveis de organização das proteínas. As estruturas primárias contem a sequência de aminoácidos (acima), a estrutura secundária descreve a conformação de partes da cadeia (meio esquerda), a estrutura terciária refere-se às disposições da proteína no espaço 3D (meio direita) e a estrutura quaternária refere-se às disposições das subunidades (embaixo).....13
- Figura 2. Ângulos de torção das proteínas. Apresentam-se os três ângulos de torção presentes em proteínas, o ângulo ômega em roxo (ω), em azul o ângulo psi (ψ) e em laranja o ângulo phi (φ).14
- Figura 3. Modelos do enovelamento de proteínas. A) Visão antiga do caminho de enovelamento onde se tem diferentes estados intermediários e B) Visão moderna de diversas rotas de enovelamento numa área com forma de funil.....17
- Figura 4. Co-tradução e Pós-tradução de proteínas. Sequência de passos da síntese de uma proteína no ribossomo, o extremo N-terminal começa o enovelamento enquanto a proteína é sintetizada até atingir o estado nativo (co-tradução). Uma proteína completamente sintetizada atinge sua estrutura nativa diretamente num passo (pós-tradução).....18
- Figura 5. Representações do canal ribossomal. Canal ribossomal com proteínas associadas (esquerda) e com uma hélice alfa (direita).....21
- Figura 6. Representação gráfica da metodologia REMD. Observam-se quatro réplicas a temperaturas diferentes, depois de um determinado tempo de simulação com DM realiza-se uma tentativa de intercambio de temperaturas entre réplicas adjacentes (setas amarelas). A metodologia REMD permite as configurações da molécula sair dos mínimos locais com o intercâmbio de temperaturas (gráfico de energia versus coordenadas).25

LISTA DE SIGLAS

3D – Tridimensional

Cryo-EM – Crio-microscopia eletrônica

DM – Dinâmica Molecular

DNA – Ácido desoxirribonucleico

EAF – Frequência de tentativas para troca

MC – Monte Carlo

poliA – Polialanina

REMD – Dinâmica Molecular com Intercambio de Réplicas

RMN – Ressonância Magnética Nuclear

RNA_m – Ácido ribonucleico mensageiro

RNA_r – Ácido ribonucleico ribossomal

RMSD – Desvio médio quadrático

SASA – Área acessível ao solvente

GDT-TS – Teste de distância global

LISTA DE SÍMBOLOS

ps – Picosegundo

ns – Nanosegundo

fs – Femtosegundo

K – Temperatura Kelvin

Å – Angström

SUMÁRIO

1.	REFERENCIAL TEÓRICO	12
1.1.	PROTEÍNAS	12
1.2.	ESTRUTURA E HIERARQUIA DE PROTEÍNAS.....	12
1.2.1.	<i>Estrutura primária</i>	14
1.2.2.	<i>Estrutura secundária</i>	14
1.2.2.1.	Hélice- α	15
1.2.2.2.	Folha- β	15
1.2.2.3.	Dobras	15
1.2.3.	<i>Estrutura terciária</i>	15
1.3.	SÍNTESE E ENOVELAMENTO DE PROTEÍNAS	16
1.3.1.	<i>Pós-tradução</i>	18
1.3.2.	<i>Co-tradução</i>	18
1.4.	RIBOSSOMO	20
1.4.1.	<i>Canal ribossomal</i>	20
1.5.	DETERMINAÇÃO E PREDIÇÃO DE ESTRUTURA 3D DE PROTEÍNAS	22
1.5.1.	<i>Métodos experimentais</i>	22
1.5.2.	<i>Métodos computacionais</i>	22
1.5.2.1.	DINÂMICA MOLECULAR (DM)	23
1.5.2.2.	DINÂMICA MOLECULAR COM INTERCAMBIO DE RÉPLICAS (REMD)	24
1.5.3.	<i>Avaliação de estruturas</i>	25
2.	OBJETIVOS E JUSTIFICATIVA	26
2.1.	OBJETIVO GERAL	26
2.2.	OBJETIVOS ESPECÍFICOS	26
2.3.	JUSTIFICATIVA.....	26
3.	CONCLUSÕES	28
4.	CONSIDERAÇÕES FINAIS.....	30
	REFERÊNCIAS.....	31

Introdução

As proteínas são os componentes envolvidos nas mais diversas funções biológicas, abrangendo desde catalisadores moleculares a componentes de nano-máquinas para o seu funcionamento. As funções das proteínas residem na sua estrutura tridimensional (3D), apresentando funções como catalisadores enzimáticos, blocos estruturais e como transportadores ou bombas de íons.

Existem diferentes metodologias experimentais para determinar a estrutura de macromoléculas, entre elas a cristalografia por difração de raios X, ressonância magnética nuclear (RMN) e crio-microscopia eletrônica (Cryo-EM). Ainda que os métodos experimentais sejam os mais confiáveis para a determinação de estruturas de proteínas, estes podem apresentar algumas limitações, como por exemplo o tamanho de algumas proteínas assim como sua flexibilidade. Além das limitações dessas técnicas, existem problemas com os passos prévios, por exemplo a clonagem dos genes em estudo, a purificação, a natureza das próprias moléculas. Além disso, esses métodos têm elevado custo tanto em esforço como em recursos econômicos.

Compreendendo suas deficiências, a utilização de técnicas computacionais ou *in silico* torna-se uma metodologia atraente para tratar esse problema. Diferentes métodos computacionais são empregados para a predição de estruturas como modelagem molecular por homologia, modelagem por *threading* e técnicas *ab initio*. Contudo, assim como os métodos experimentais, as abordagens *in silico* também apresentam desvantagens como a complexidade dos cálculos.

Dentre as técnicas *ab initio* aplicadas à predição de estruturas de proteínas, podemos destacar as metodologias de monte carlo (MC) e dinâmica molecular (DM). A técnica de MC utiliza um algoritmo estocástico e sua trajetória final não é dependente de tempo, enquanto que a DM utiliza um algoritmo determinístico dependente de tempo. De forma simplificada, a DM realiza o cálculo da força que atua em cada átomo de uma proteína, calculando sua posição no espaço 3D ao longo de um determinado tempo e gerando como resultado uma trajetória de movimento (Tramontano, 2006).

As pesquisas que utilizam DM como ferramenta de estudo, podem ser divididas em quatro áreas: i) método de amostragem do espaço conformacional, ii) descrição do sistema em equilíbrio, iii) estudo do comportamento de uma proteína em solução (Karplus &

McCammon, 2002), e iv) predição de estruturas de proteínas (Breda, Santos, Basso, & Souza, 2007; Simmerling, Strockbine, & Roitberg, 2002).

Atualmente as simulações por DM aplicadas à predição de estrutura 3D de proteínas utilizam toda a sequência de aminoácidos desde o início da simulação, isto é, simulam um processo pós-traducional de enovelamento (J. J. Ellis, Huard, Deane, Srivastava, & Wood, 2010). Entretanto, nesta pesquisa, foi desenvolvida uma abordagem diferente para a predição de estruturas de proteínas utilizando a DM. Para isso, foi construído um modelo do canal ribossomal com o intuito de mimetizar a síntese proteica realizada pelo ribossomo, seguindo a proposta de co-tradução.

1. Referencial teórico

1.1. Proteínas

As proteínas são as unidades funcionais para quase todos os processos biológicos, podendo atuar como catalizadores utilizados nas reações químicas dentro dos organismos, abrangendo funções estruturais, de transporte, regulação e sinalização (Lodish et al., 2005). A grande diversidade de funções realizadas pelas proteínas é produto do número de blocos fundamentais dos quais são compostas, os aminoácidos. Cada proteína contém uma sequência única que lhe confere propriedades características (Karp, 2011), embora isto não seja inteiramente certo, pois atualmente o mais aceito é que as propriedades de uma proteína devem-se à sua estrutura tridimensional mais que à sua sequência de aminoácidos, pois proteínas com diferentes sequências de aminoácidos podem ter a mesma função dentro dos seres vivos (Wood, Srivastava, Patton, & Fisher, 2011). Assim, a determinação da estrutura 3D das proteínas ganhou mais interesse que a determinação da sequência de aminoácidos para estudar sua função biológica.

1.2. Estrutura e hierarquia de proteínas

A estrutura de proteínas pode ser classificada em 4 níveis (**Figura 1**), os quais apresentam uma organização hierárquica: i) estrutura primária → sequência linear dos aminoácidos que formam a proteína; ii) estrutura secundária → descreve a conformação de partes da cadeia de aminoácidos como, por exemplo, hélices- α , folha- β e dobras; iii) estrutura terciária → descreve a disposição dos segmentos no espaço tridimensional; iv) estrutura quaternária → descreve como as estruturas terciárias organizam-se em solução (Karp, 2011).

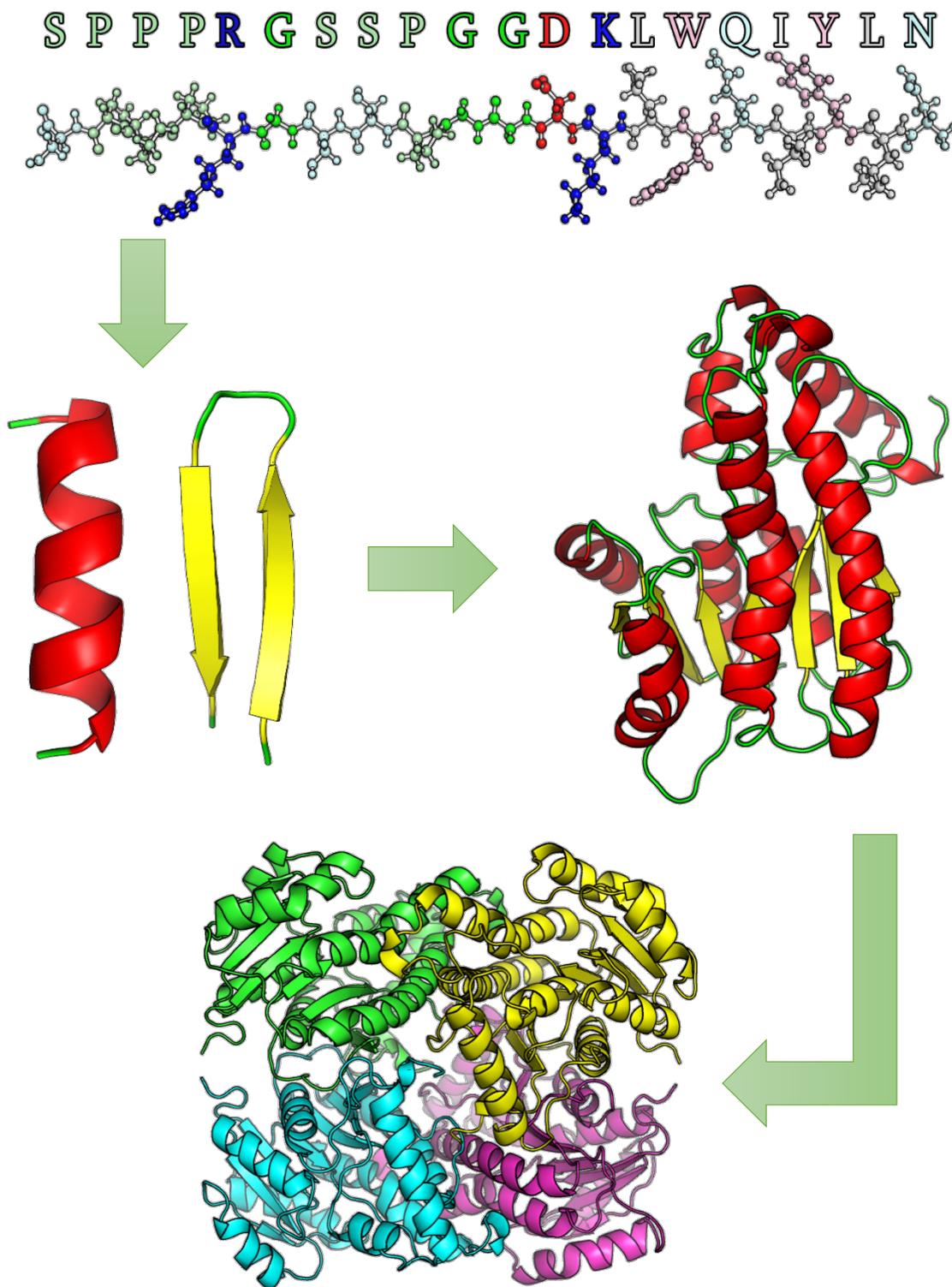


Figura 1. Níveis de organização das proteínas. As estruturas primárias contêm a sequência de aminoácidos (acima), a estrutura secundária descreve a conformação de partes da cadeia (meio esquerda), a estrutura terciária refere-se às disposições da proteína no espaço 3D (meio direita) e a estrutura quaternária refere-se às disposições das subunidades (embaixo).

Fonte: O autor.

1.2.1. Estrutura primária

Os átomos de nitrogênio (N) do grupo amino, carbono α ($C\alpha$) e o grupo carbonila (C=O) formam a cadeia principal dos aminoácidos (N- $C\alpha$ -C=O). Sendo que a união destas cadeias, de forma linear, caracteriza o primeiro nível hierárquico de proteínas. Esta união ocorre por meio de ligações peptídicas, compreendendo o grupamento carboxílico de um aminoácido e o grupamento amino de outro. Além disso, é a partir das torções envolvendo as ligações da cadeia principal que as proteínas começam a assumir seus diferentes tipos de enovelamento. A **Figura 2** apresenta como são representadas as torções.

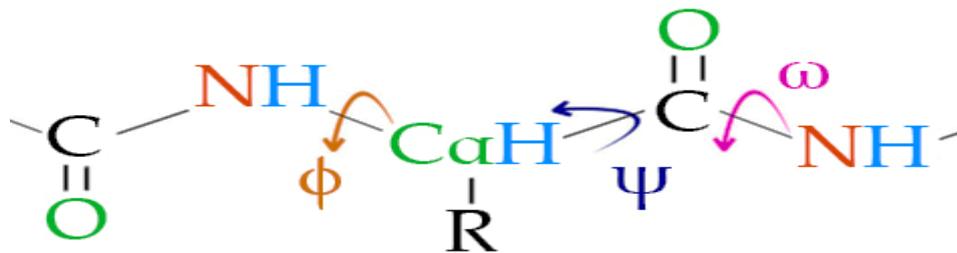


Figura 2. Ângulos de torção das proteínas. Apresentam-se os três ângulos de torção presentes em proteínas, o ângulo ômega em roxo (ω), em azul o ângulo psi (ψ) e em laranja o ângulo phi (ϕ).
Fonte: O autor (2017).

São definidos três ângulos possíveis na cadeia principal, que são: i) o ângulo phi (ϕ), definido entre os átomos N e o $C\alpha$; ii) o ângulo psi (ψ), entre os átomos $C\alpha$ e C do grupamento carbonila, e iii) o ângulo ômega (ω), definido entre os átomos C do grupamento carbonila e o N (Aluru, 2007). Apesar de existirem três ângulos, apenas ϕ e ψ apresentam liberdade de rotação e descrevem como a cadeia polipeptídica pode enovelar. O ângulo ω fica sempre fixo a 180° devido à natureza da ligação peptídica, embora se tem umas poucas exceções onde fica fixo a 0° (Lodish et al., 2005).

1.2.2. Estrutura secundária

A estrutura secundária refere-se aos diversos arranjos espaciais resultantes dos dobramentos de uma cadeia polipeptídica. Estes arranjos são gerados pela presença ou ausência de ligações de hidrogênio. De acordo com o padrão das ligações, o segundo nível hierárquico de proteínas pode ser dividido em três tipos: hélice- α , folha- β e dobras (Lodish et al., 2005).

1.2.2.1. Hélice- α

Neste tipo de estrutura, o átomo de oxigênio do grupo carbonila de cada ligação peptídica forma uma ligação de hidrogênio com o átomo de hidrogênio do grupo amino. Esta interação ocorre a cada quatro resíduos no sentido da extremidade C terminal, permitindo que o esqueleto adquira uma forma de espiral ou hélice. Além disso, as hélices podem apresentar comportamentos hidrofóbicos ou hidrofílicos dependendo da natureza físico-química dos resíduos de aminoácidos (Lesk, 2001; Lodish et al., 2005).

1.2.2.2. Folha- β

A folha- β é uma estrutura composta por duas ou mais fitas- β associadas lateralmente. Cada fita- β é um segmento polipeptídico curto de no mínimo cinco resíduos quase totalmente estendido, no qual as ligações de hidrogênio formam-se entre os átomos da cadeia principal das fitas- β adjacentes. Assim, as folhas- β podem ser paralelas, onde as fitas estão na mesma direção N- e C-terminal, ou antiparalelas, onde as fitas encontram-se em direções opostas (Lesk, 2001; Lodish et al., 2005).

1.2.2.3. Dobras

As voltas são formadas por 3 a 4 resíduos e encontram-se na superfície de uma proteína, formando dobras definidas que redirecionam o esqueleto. Essas estruturas são estabilizadas por ligações de hidrogênio entre os resíduos finais, permitindo o dobramento de proteínas em estruturas altamente compactadas (Lodish et al., 2005).

1.2.3. Estrutura terciária

Se consideramos a estrutura primária como a representação de uma proteína numa só dimensão como, por exemplo, uma linha e a estrutura secundária como o objeto em duas dimensões como um quadrado, a estrutura terciária seria uma figura no espaço 3D, neste caso um cubo ou uma esfera. Diferentemente do que ocorre na estrutura secundária, onde a estabilidade é dada por ligações de hidrogênio, na estrutura terciária a estabilidade da molécula é obtida por diferentes tipos de interações. Dentre estas podemos destacar as interações hidrofóbicas entre grupos

apolares, ligações de hidrogênio entre grupos polares, pontes salinas entre resíduos ácidos e básicos e pontes bissulfeto. Esses tipos de interações mantêm as estruturas terciárias em formas regulares e compactas. Como as interações que estabilizam a estrutura terciária são fracas, essa estrutura não é rígida, conseguindo apresentar diferentes estados conformacionais (Lodish et al., 2005).

A estrutura terciária nos é de especial interesse, pois define a função biológica de uma proteína monomérica, além de ser o objetivo ao qual se deseja chegar com a predição de estruturas 3D partindo de uma sequência de aminoácidos.

2.2.4. Estrutura quaternária

A estrutura quaternária descreve o número e posições relativas das subunidades de proteínas. As proteínas multiméricas podem ser constituídas pela repetição da mesma subunidade (homoméricas) ou por subunidades diferentes (heteroméricas) (Lodish et al., 2005).

1.3. Síntese e enovelamento de proteínas

A síntese de proteínas é um processo pelo qual uma cadeia de ácido desoxirribonucleico (DNA) é traduzida numa proteína. Nesse processo, a sequência de DNA é primeiro transformada numa cadeia de ácido ribonucleico mensageiro (RNAm). Esta cadeia é transportada até o ribossomo, onde a cada 3 nucleotídeos do RNAm podem ser traduzidos em 20 diferentes aminoácidos. Dentro do ribossomo, os aminoácidos são ligados uns com outros por meio de ligações peptídicas numa ordem sequencial, um resíduo por vez (Alberts et al., 1986).

O enovelamento de proteínas não é um processo simples, visto que o número de configurações que uma proteína pode apresentar é proporcional ao número de resíduos da mesma. Com isso, é possível chegar a um número astronômico de configurações possíveis, dando origem ao paradoxo de Levinthal. Segundo o paradoxo, se uma proteína tivesse que atingir sua conformação nativa por meio de uma busca exaustiva aleatória, precisaria um tempo maior que a idade do universo (Kolb, 2001). Desta forma, se tem proposto que as proteínas não necessitam fazer uma busca exaustiva do espaço

conformacional para atingir sua conformação nativa, e sim haveria uma rota alternativa de enovelamento (Englander & Mayne, 2014).

Do ponto de vista termodinâmico, as proteínas no estado nativo apresentam uma maior estabilidade, sendo sua energia livre, a mínima global de todas as estruturas cineticamente acessíveis (Plotkin & Onuchic, 2002). Entretanto, algumas exceções podem ocorrer como no caso da Protease α -lítica, onde o estado nativo da proteína não corresponde ao estado com menor energia livre (Sohl, Jaswal, & Agard, 1998).

A concepção inicial era de que as proteínas atingiam sua conformação nativa através de estados intermediários (**Figura 3A**). No entanto, entende-se que uma cadeia polipeptídica pode apresentar muitos caminhos de enovelamento com diversas conformações (Englander & Mayne, 2014; Plotkin & Onuchic, 2002). Assim, podemos descrever a área energética de uma proteína como um funil no qual tem no fundo a estrutura nativa (**Figura 3B**), apresentando barreiras energéticas e dando origem a mínimos locais (Cabrita, Dobson, & Christodoulou, 2010; Englander & Mayne, 2014).

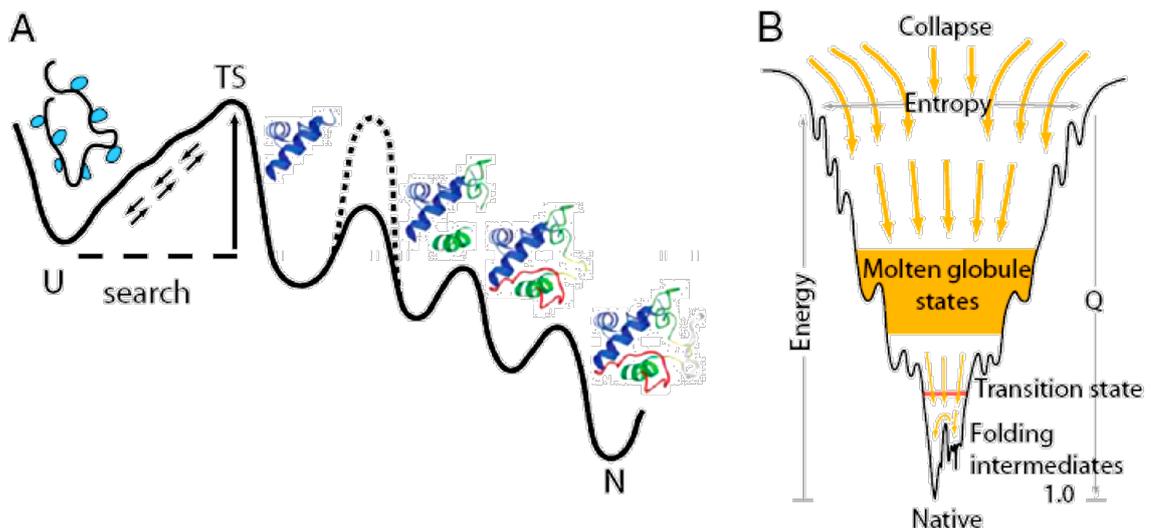


Figura 3. Modelos do enovelamento de proteínas. A) Visão antiga do caminho de enovelamento onde se tem diferentes estados intermediários e B) Visão moderna de diversas rotas de enovelamento numa área com forma de funil.

Fonte: Englander & Mayne 2014.

Dependendo do tempo de enovelamento das proteínas, é possível distinguir dois tipos, o pós-traducional e co-traducional (**Figura 4**), os quais serão descritos nos próximos subtópicos.

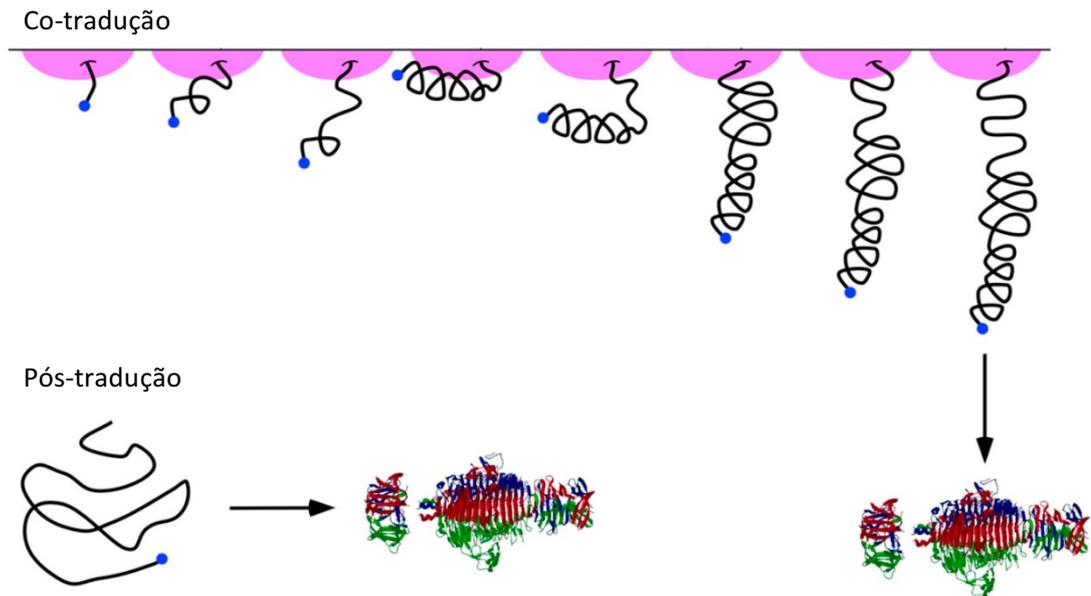


Figura 4. Co-tradução e Pós-tradução de proteínas. Sequência de passos da síntese de uma proteína no ribossomo, o extremo N-terminal começa o enovelamento enquanto a proteína é sintetizada até atingir o estado nativo (co-tradução). Uma proteína completamente sintetizada atinge sua estrutura nativa diretamente num passo (pós-tradução).

Fonte: Evans *et al.*, 2008.

1.3.1. Pós-tradução

O modelo pós-traducional de proteínas foi o primeiro a ser aceito devido aos trabalhos de renaturação de proteínas. Dentre estes trabalhos, podemos destacar a pesquisa de Anfisen, na qual conseguiu obter uma proteína no seu estado nativo após ter realizado a desnaturação por meio de métodos químicos (Anfinsen, 1973). De acordo com o experimento de Anfisen, a cadeia polipeptídica precisa estar completamente sintetizada antes do início do processo de enovelamento. Este princípio é o mesmo utilizado pelos métodos de predição de estruturas *in silico* de proteínas. Entretanto, atualmente é possível encontrar algumas poucas exceções que começam com uma metodologia de co-tradução (Elcock, 2006; J. J. Ellis *et al.*, 2010; Huard, Deane, & Wood, 2006; Morrissey, Ahmed, & Shakhnovich, 2004).

1.3.2. Co-tradução

A co-tradução de proteínas baseia-se numa visão mais temporal da síntese de proteínas, onde uma cadeia polipeptídica é sintetizada um resíduo por vez (Deane, Dong, Huard, Lance, & Wood, 2007; Fedorov & Baldwin, 1997; Huard *et al.*, 2006; Kolb, 2001; O'Brien, Ciryam, Vendruscolo, & Dobson, 2014). Toda proteína começa sua

síntese partindo de seu extremo N-terminal até acabar no extremo C-terminal. Esta direção no processo de enovelamento, faz com que a região N-terminal das proteínas tenha mais tempo para realizar seu enovelamento que a região C-terminal (Clarke IV & Clark, 2008; J. J. Ellis et al., 2010; J. P. Ellis, Bakke, Kirchdoerfer, Jungbauer, & Cavagnero, 2008; Komar, 2009; Wood et al., 2011; G. Zhang, Hubalewska, & Ignatova, 2009). Esta característica, muitas vezes, permite que a região N-terminal atinja sua estrutura nativa estando ainda unida ao ribossomo (Elcock, 2006; Tsai et al., 2008; Wen et al., 2008), e desta forma possa desempenhar seu papel biológico parcial ou total (Nicola, Chen, & Helenius, 1999).

Nos últimos anos se tem estudado com maior profundidade o efeito da co-tradução no enovelamento de proteínas. Sendo os tópicos mais relevantes a velocidade de síntese, a direção e o efeito do ribossomo.

A velocidade de síntese pode ser estudada tanto experimentalmente quanto computacionalmente. Os estudos experimentais baseiam-se principalmente na mutação sinônima de códons (Komar, 2009; Tsai et al., 2008). Ciryam *et al.* realizaram uma análise do proteoma de *Escherichia coli* com a finalidade de encontrar proteínas que apresentam o enovelamento do tipo co-traducional. Nos resultados, foi observado que praticamente um terço do proteoma de *E. coli* enovela-se co-traducionalmente (Ciryam et al., 2013). Em outro estudo realizado por Wang *et al.*, foi analisada a velocidade de síntese. Neste trabalho, foram testadas duas velocidades, e os resultados indicaram que uma maior velocidade favorece a co-tradução (Wang, Wang, Chen, & Xiao, 2015).

Ellis *et al.*, estudaram o efeito da direção de síntese no enovelamento, ou seja, neste trabalho a leitura da proteína foi realizada de N- para C-terminal e também de C- para N-terminal. Os resultados mostraram que quase a totalidade das proteínas testadas seguem uma direção determinada para atingir sua estrutura nativa, que foi a direção de N- para C-terminal (J. J. Ellis et al., 2010). Já nos estudos de Deane *et al.*, foram apresentados dados computacionais de que a co-tradução é mais preponderante em proteínas da classe α/β (Deane et al., 2007).

O principal foco dos estudos sobre o efeito do ribossomo na co-tradução é a utilização desta proteína como uma barreira física, principalmente o canal ribossomal. Em estudos realizados por Elcock, o ribossomo foi simulado a nível atômico para

analisar o comportamento da síntese de três proteínas. Os resultados indicaram que, nas proteínas testadas, o confinamento destas no canal ribossomal não apresenta um efeito significativo no enovelamento, entretanto, foram observados indícios de que a formação de estruturas secundárias das proteínas é sensível dependendo da localização no canal ribossomal (Elcock, 2006).

Em relação ao(s) caminho(s) de enovelamento de proteínas, se tem indícios de que o canal ribossomal possa servir como um ambiente que propicie a formação dos contatos nativos. Desta forma, as proteínas podem seguir um ou dois caminhos de enovelamento (Bui & Hoang, 2016; Chen, Wang, Liu, & Xiao, 2013).

1.4. Ribossomo

A de síntese de proteínas é efetuada pelo ribossomo, o qual é um complexo multienzimático composto por proteínas e ácido ribonucleico ribossomal (RNAr). Os ribossomos eucarióticos e procarióticos são muito similares na sua estrutura e função, cada um deles é composto por uma subunidade maior e uma subunidade menor. Cada subunidade é composta por RNAr e proteínas, porém nos ribossomos eucarióticos, o número de proteínas associadas às subunidades é maior (Alberts et al., 1986; Ban, Nissen, Hansen, Moore, & Steitz, 2000).

O ribossomo não é simplesmente uma “fábrica” de proteínas, mas também o primeiro ambiente que as proteínas em síntese entram em contato, assim como, interagem (Kramer, Boehringer, Ban, & Bukau, 2009). A partir dessa concepção, o interesse em estudar as interações entre o ribossomo e as cadeias polipeptídicas em formação ganhou importância entre os cientistas (Cabrita et al., 2010; J. P. Ellis et al., 2008; Kramer et al., 2009; G. Zhang & Ignatova, 2011). Dentro do ribossomo, é de especial relevância o estudo do canal ribossomal na análise do enovelamento de proteínas, visto que é onde encontram-se as proteínas em formação e onde acontecem as primeiras formações de estruturas secundárias.

1.4.1. Canal ribossomal

O canal ribossomal é uma cavidade que faz parte de um sistema de canais interconectados acessíveis ao solvente. Este sistema é uma rede que permeabiliza o ribossomo e que conecta-se com o exterior por meio de diferentes partes do

complexo, mas a única conexão capaz de acomodar uma cadeia polipeptídica é o canal ribossomal (Voss, Gerstein, Steitz, & Moore, 2006).

Inicialmente acreditava-se que o ribossomo era simplesmente um conduto pelo qual as proteínas eram sintetizadas, mas atualmente se tem conhecimento que para muitas proteínas o canal ribossomal tem um papel mais ativo (J. P. Ellis et al., 2008; Kosolapov & Deutsch, 2009; Kramer et al., 2009; Lu & Deutsch, 2008; Lu, Hua, Kobertz, & Deutsch, 2011; Tu & Deutsch, 2010; Wilson & Beckmann, 2011; G. Zhang & Ignatova, 2011; Y. Zhang, Wölfle, & Rospert, 2013; Ziv, Haran, & Thirumalai, 2005). Esta cavidade apresenta uma forma irregular, mas pode ser descrita como um túnel de 80 a 100 Å de comprimento e de 10 a 20 Å de largura, totalizando um volume de aproximadamente 25.000 Å³ (**Figura 5**) (Voss et al., 2006). As paredes são muito ricas em RNAr, e de acordo os trabalhos de Lu e colaboradores e Voss e colaboradores, pode-se constatar que o canal apresenta uma natureza eletronegativa no interior e quase nula próximo à saída (Lu, Kobertz, & Deutsch, 2007; Voss et al., 2006).

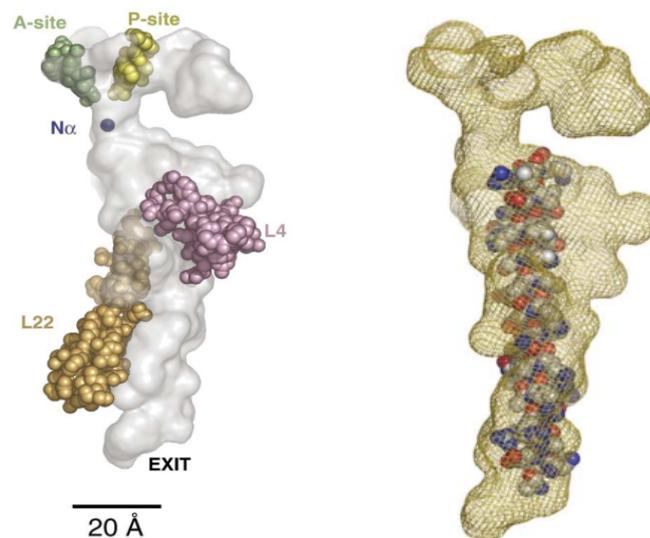


Figura 5. Representações do canal ribossomal. Canal ribossomal com proteínas associadas (esquerda) e com uma hélice alfa (direita).
Fonte: Voss et al., 2006.

Assim, tendo em vista que as proteínas são sintetizadas um resíduo por vez e que o seu primeiro ambiente é o ribossomo, é possível formular a ideia de que o ribossomo e a co-tradução servem como uma rota para as proteínas atingirem seu estado nativo e funcional.

1.5. Determinação e predição de estrutura 3D de proteínas

Existem diferentes procedimentos e metodologias para a determinação da estrutura 3D de proteínas, os quais podem-se dividir em métodos experimentais e computacionais (ou *in silico*). Dentre os métodos experimentais, podemos destacar as abordagens de difração de raios X, a RMN e Cryo-EM. Entre as metodologias computacionais podemos citar a DM e MC, as quais são metodologias determinísticas e estocásticas, respetivamente (Tramontano, 2006).

1.5.1. Métodos experimentais

Como o nome indica, os métodos experimentais baseiam-se na determinação de estruturas experimentalmente, isto é, realizam-se medições reais de preparações proteicas para a determinação de sua estrutura. Esses métodos são os mais confiáveis em termos de determinação de estrutura, mas apresentam algumas limitações, como por exemplo:

- A determinação de uma única proteína tem um elevado custo econômico.
- É necessário de muito tempo para sua realização.
- Existe dificuldade em purificar algumas proteínas.
- Precisa-se de laboratórios especializados para o procedimento.
- No caso da difração de raios X, o resultado é uma única estrutura, não é possível observar o caráter dinâmico das proteínas.

Para superar essas limitações, métodos que permitem a predição de estrutura 3D são desenvolvidos, isto é, obter estruturas 3D sem necessidade de realizar procedimentos experimentais.

1.5.2. Métodos computacionais

As metodologias computacionais ou *in silico* utilizam os cálculos computacionais para a predição de estruturas 3D de proteínas. Os métodos computacionais dividem-se em dois tipos, aqueles que baseiam na homologia de sequências com estrutura tridimensional conhecida, e os que partem somente de uma cadeia de aminoácidos (denominados métodos *ab initio*). Dessas duas metodologias

de predição de estruturas 3D, a modelagem *ab initio* é a mais custosa computacionalmente de realizar.

As metodologias *ab initio* baseiam-se na determinação das posições dos átomos do sistema e dos ângulos de torção *phi* e *psi*, sendo as abordagens mais comuns o MC e a DM. A DM pode ser utilizada em combinação com diferentes metodologias, com o intuito de propiciar uma melhor amostragem do espaço conformacional. Por exemplo, a DM com intercâmbio de réplicas (REMD), faz uso da DM para simular vários sistemas e utiliza o critério de MC para aceitar ou rejeitar as trocas de estruturas, esta metodologia será melhor explicada num próximo subtópico.

1.5.2.1. Dinâmica molecular (DM)

A dinâmica molecular (DM) é a ciência que permite simular os movimentos de um sistema. Esta abordagem tem como elementos essenciais o conhecimento do potencial de interações do sistema (os campos de força que podem ser usados), e as equações de movimento que governam a dinâmica das partículas. Para o comportamento de biomoléculas, as equações de movimento newtonianas clássicas são adequadas (Karplus & Petsko, 1990). Na DM as equações de movimento clássicas para todos os átomos de um sistema são resolvidas simultaneamente num período de tempo adequado, desta forma é possível extrair informações detalhadas sobre o sistema por meio da análise das trajetórias (McCammon, Gelin, & Karplus, 1977).

A DM de biomoléculas baseia-se também em outros parâmetros como, por exemplo, os campos de força, ângulos de torção, tipo de solvente, nível de detalhamento do modelo, temperatura do sistema, tempo de integração, raio de corte, entre outros. O processo de DM clássica pode ser resumido de uma forma simples de acordo com os seguintes passos:

- Definição das posições iniciais de todos os átomos do sistema.
- Sorteio aleatório das velocidades iniciais para cada átomo do sistema dependendo da distribuição de Boltzmann de acordo com a temperatura determinada.
- Determinação das novas posições de cada átomo do sistema usando equações newtonianas e um intervalo na ordem de femtosegundos (fs).
- Determinação do tempo de simulação de acordo com o passo de integração.

- Repetem-se os três últimos passos quantas vezes forem necessárias até acabar a simulação. É muito importante ressaltar que depois do primeiro passo, as velocidades não são novamente sorteadas, estas são calculadas a partir do passo anterior.
- Uma vez finalizada a simulação, tem-se como resultado uma trajetória de todo o sistema em estudo.

Dois atributos das simulações de DM têm desempenhado um papel essencial no seu rápido desenvolvimento e amplo uso. Primeiro, as simulações fornecem movimentos individuais de partículas em função do tempo, assim podem ser testados mais facilmente que os métodos experimentais, permitindo que perguntas específicas sobre as propriedades de um sistema sejam respondidas. Segundo, embora os potenciais usados sejam aproximações, o usuário tem a liberdade de alterá-los, removendo ou modificando contribuições específicas para a simulação do sistema de estudo (Karplus & Petsko, 1990).

1.5.2.2. Dinâmica molecular com intercambio de réplicas (REMD)

O algoritmo REMD envolve várias cópias de um sistema, sendo cada um deles um processo separado. Dependendo do tipo de REMD, são utilizados diferentes parâmetros para as réplicas. A metodologia REMD mais comum utiliza diferentes temperaturas de simulação por sistema. No caso desta dissertação, será explicada somente a REMD com uso de temperaturas.

As cópias ou réplicas são idênticas na totalidade, variando somente a temperatura entre elas. Inicialmente as réplicas são simuladas independentemente entre elas até atingir sua temperatura objetivo mediante acoplamento de temperatura. Depois de um determinado tempo de simulação, cada uma das réplicas tenta intercambiar temperaturas com uma das réplicas adjacentes usando o critério de MC. Se a probabilidade é satisfeita, as réplicas intercambiam temperaturas. Depois do intercambio, as réplicas atingem sua nova temperatura quando ocorre o intercâmbio de temperaturas, ou mantem temperatura constante no caso contrário. Esse ciclo de simulação e tentativa de intercâmbio acontece tantas vezes for necessário para a simulação atingir seu objetivo (García & Sanbonmatsu, 2001; Sanbonmatsu & García, 2002; Sugita & Okamoto, 1999).

Desta forma uma simulação pode superar as barreiras energéticas dos mínimos locais, conseguindo assim uma melhor exploração do espaço conformacional (**Figura 6**).

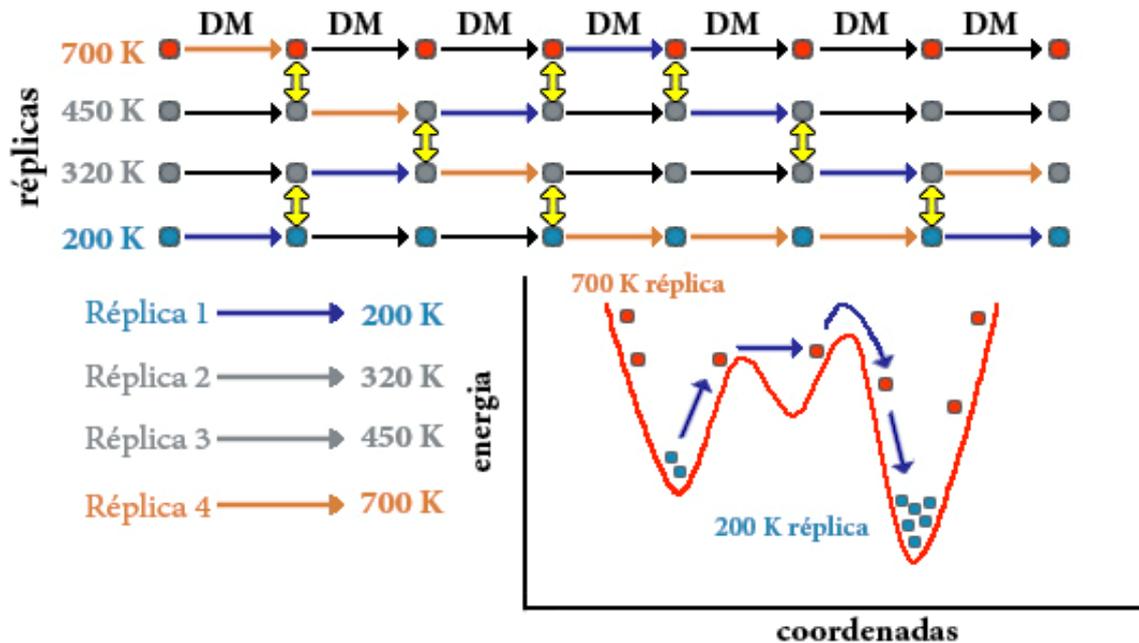


Figura 6. Representação gráfica da metodologia REMD. Observam-se quatro réplicas a temperaturas diferentes, depois de um determinado tempo de simulação com DM realiza-se uma tentativa de intercâmbio de temperaturas entre réplicas adjacentes (setas amarelas). A metodologia REMD permite as configurações da molécula sair dos mínimos locais com o intercâmbio de temperaturas (gráfico de energia versus coordenadas).

Fonte: O autor (2017).

1.5.3. Avaliação de estruturas

A determinação de similaridade de duas estruturas não é uma tarefa simples e existem diferentes abordagens. O desvio quadrático meio (RMSD) é a medida mais utilizada para calcular a similaridade entre duas coordenadas atômicas superpostas. Os valores de RMSD são apresentados em Angstroms (Å) e quanto menor o valor, maior similaridade. A principal desvantagem do RMSD é que esta métrica é muito sensível a valores atípicos.

Uma medida de similaridade que não é muito sensível a valores atípicos é o teste de distância global (GDT-TS). Com o GDT-TS, muitas superposições para duas estruturas são calculadas, enquanto que o RMSD que utiliza apenas a superposição ótima para seu cálculo. O GDT-TS utiliza para o seu cálculo o maior conjunto de C α de resíduos num modelo que está dentro de um raio de corte da sua posição na estrutura experimental. O GDT-TS realiza o cálculo com diferentes raios de corte incrementais. O resultado final do cálculo é um valor entre 0,0 e 1,0, sendo que 1,0 representa a maior similaridade (Kufareva & Abagyan, 2012; Zemla, Venclovass, Moulton, & Fidelis, 1999).

2. Objetivos e Justificativa

2.1. Objetivo geral

- Desenvolver uma nova abordagem de predição da estrutura 3D de proteínas utilizando a Dinâmica Molecular, a co-tradução e um modelo do canal ribossomal, visando obter estruturas mais satisfatórias que os métodos computacionais atuais.

2.2. Objetivos específicos

- Desenvolver um modelo simples do canal ribossomal capaz de apresentar algumas características do canal ribossomal.
- Testar o modelo do canal ribossomal desenvolvido.
- Determinar se a velocidade de síntese tem efeito no resultado de predição.
- Comparar a eficiência da metodologia desenvolvida com a metodologia tradicional.

2.3. Justificativa

Simulações de dinâmica molecular de proteínas a nível atômico tem um elevado custo computacional na predição da estrutura nativa partindo apenas da sequência de aminoácidos. Geralmente estas simulações levam muito tempo para sua finalização, e depois de ter finalizado, os pesquisadores encontram-se com outro problema, o número de estruturas a ser analisadas. Uma simulação de dinâmica molecular simples pode ter como resultado desde Miles até milhões de estruturas.

Diferentes melhoras nas metodologias de predição de estrutura de proteínas tem sido desenvolvidas nos últimos anos, como o uso de restrições nos ângulos de torção (Balaraman, Park, Jain, & Vaidehi, 2011), guiar as simulações usando inferência bayesiana sobre informações heurísticas (Perez, MacCallum, & Dill, 2015), usar modelos *coarse grained* para acelerar as simulações (Ding, Buldyrev, & Dokholyan, 2005), usar *simulated annealing* para obter uma melhor busca no espaço conformacional (Hao, Xu, Yang, & Yang, 2015). Embora estas metodologias obtiveram melhorias na qualidade de estruturas obtidas, as estruturas resultantes são ainda muito numerosas e o custo computacional é elevado. Além disso, todas as metodologias de predição de estrutura tem uma abordagem pós-traducional do enovelamento de proteínas, salvo algumas exceções (J. J. Ellis et al., 2010).

Visto que a co-tradução tem influência no enovelamento de proteínas, uma metodologia utilizando esta abordagem pode oferecer vantagens na predição de estrutura de proteínas. Portanto, nesta pesquisa se propõe desenvolver uma abordagem diferente para a predição de estrutura de proteínas usando a co-tradução e um modelo do canal ribossomal, com o intuito de obter estruturas nativas e uma redução no número de estruturas a serem analisadas.

3. Conclusões

As conclusões obtidas deste trabalho foram:

- O modelo do canal ribossomal desenvolvido nesta pesquisa conseguiu mimetizar algumas das características do canal ribossomal real. A geometria do modelo assemelha-se as dimensões ideais do canal ribossomal, 100 Å de comprimento e 20 Å de diâmetro. O potencial elétrico negativo do canal foi levemente mimetizado usando o íon Cl^- para a construção do modelo.
- O modelo do canal ribossomal desenvolvido foi testado pela capacidade de promover a formação de hélice α numa cadeia de polialanina, apresentando o melhor resultado no modelo feito de íons Cl^- .
- Os resultados de diferentes velocidades de síntese apresentaram diferenças quanto a similaridade das estruturas resultantes, estabilidade da simulação e o número de clusteres gerados. A maioria dos testes apresentaram os melhores resultados no intervalo de 80 a 200 ps de velocidade de síntese.
- Os resultados da DM com o uso do modelo do canal ribossomal melhoraram a predição da proteína Trp-cage quando comparado com o procedimento controle. Entretanto, para a temperatura de 450 K praticamente não existe diferença entre o uso ou não do modelo do canal ribossomal. Para o modulo GA, o uso do modelo do canal ribossomal teve uma leve melhora nos resultados, mas mesmo assim os resultados não foram significativos.
- Nas simulações por REMD do Trp-cage observou-se uma melhoria de todos os tratamentos quando comparados com as simulações por DM padrão. Embora os resultados do Trp-cage sejam aceitáveis, é preciso realizar réplicas das simulações para determinar se existe uma diferença estatística entre os tratamentos e o controle, visto que os resultados não foram tão diferentes quando comparados com a DM padrão. Assim como com o modulo GA, as estruturas aceitáveis da região N-terminal da ubiquitina apresentaram baixa quantidade.
- Neste trabalho foi desenvolvida uma abordagem diferente para a predição de estrutura de proteínas, utilizando a co-tradução e um modelo do canal ribossomal no início do processo de predição. Os resultados indicam que, embora possam ser observadas leves

melhorias em relação a quantidade de estruturas secundárias e uma diminuição dos clusteres gerados, é preciso realizar melhorias no modelo do canal ribossomal.

4. Considerações finais

O presente trabalho apresentou resultados positivos em relação à utilização do modelo do canal ribossomal para a predição de estruturas de proteínas. Entretanto, são necessárias melhorias na descrição do modelo do canal ribossomal para sua utilização com diferentes tipos de proteínas, além disso, também é necessário realizar mais simulações de diferentes proteínas e repetições.

Na predição de estrutura de proteínas é desejável que a amostragem do espaço conformacional de uma proteína seja o mais amplo possível, incrementando assim a probabilidade de encontrar a estrutura nativa. O uso do modelo do canal ribossomal (uma restrição espacial) nos primeiros passos da predição de estruturas tem como resultado uma diminuição do número de clusters gerados, o que significa numa redução da amostragem do espaço conformacional. As duas ideias são contrárias, mas podem ser usadas em conjunto como: (i) inicialmente diminuir o espaço de conformações possíveis mediante uma restrição espacial (o uso do modelo do canal ribossomal), para depois sim (ii) realizar uma ampla busca do subespaço conformacional obtido. Podemos utilizar uma analogia bastante simples como, por exemplo, realizar a busca por uma pessoa num país, se escolhermos os cidadãos de forma aleatória em todo o país pode aumentar a probabilidade de encontrar a pessoa, entretanto o esforço também é maior. No entanto, se inicialmente se reduz a busca a uma cidade somente e depois se faz a escolha de cidadãos, as probabilidades aumentam quando comparado com a primeira tentativa. É preciso ter muito cuidado na hora de usar uma restrição no início do processo de predição, porque nem sempre a redução do espaço conformacional pode ser positiva. Por esta razão, a necessidade de um modelo do canal ribossomal que assemelhe o ambiente inicial das proteínas em síntese resultaria numa aproximação atrativa, mas os resultados do presente trabalho indicam que ainda é necessário melhorar o modelo do canal ribossomal apresentado.

Referências

- Alberts, B., Bray, D., Lewis, J., Raff, M., Roberts, K., & Watson, J. D. (1986). *Biología Molecular de la Célula*. (M. Costa, Ed.). Barcelona: Omega.
- Aluru, S. (2007). *Handbook of Computational Molecular Biology*. (S. Aluru, Ed.) *Briefings in Bioinformatics* (Vol. 8). New York: Chapman & Hall/CRC.
<http://doi.org/10.1093/bib/bbm002>
- Anfinsen, C. B. (1973). Principles that Govern the Folding of Protein Chains. *Science*, *181*(4096), 223–230. <http://doi.org/10.1126/science.181.4096.223>
- Balaraman, G. S., Park, I. H., Jain, A., & Vaidehi, N. (2011). Folding of small proteins using constrained molecular dynamics. *Journal of Physical Chemistry B*, *115*(23), 7588–7596.
<http://doi.org/10.1021/jp200414z>
- Ban, N., Nissen, P., Hansen, J., Moore, P. B., & Steitz, T. a. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, *289*(August), 905–920. <http://doi.org/10.1126/science.289.5481.905>
- Breda, A., Santos, D. S., Basso, L. A., & Souza, O. N. De. (2007). Ab initio 3 - D structure prediction of an artificially designed three - α - helix bundle via all - atom molecular dynamics simulations. *Genetics and Molecular Research*, *6*, 901–910.
- Bui, P. T., & Hoang, T. X. (2016). Folding and escape of nascent proteins at ribosomal exit tunnel. *The Journal of Chemical Physics*, *144*(9), 95102.
<http://doi.org/10.1063/1.4943042>
- Byrne, A., Williams, D. V., Barua, B., Hagen, S. J., Kier, B. L., & Andersen, N. H. (2014). Folding dynamics and pathways of the trp-cage miniproteins. *Biochemistry*, *53*(38), 6011–6021.
<http://doi.org/10.1021/bi501021r>
- Cabrita, L. D., Dobson, C. M., & Christodoulou, J. (2010). Protein folding on the ribosome. *Current Opinion in Structural Biology*, *20*(1), 33–45.
<http://doi.org/10.1016/j.sbi.2010.01.005>
- Case, D. A., Berryman, J. T., Betz, R. M., Cerutti, D. S., Cheatham, T. E., Darden, T. A., ... Kollman, P. A. (2014). AMBER. University of California, San Francisco. Retrieved from <http://ambermd.org/>
- Chen, C., Wang, E., Liu, P., & Xiao, Y. (2013). Simulation study of the role of the ribosomal exit tunnel on protein folding. *Physical Review E - Statistical, Nonlinear, and Soft Matter Physics*, *87*(2). <http://doi.org/10.1103/PhysRevE.87.022701>
- Ciryam, P., Morimoto, R. I., Vendruscolo, M., Dobson, C. M., O'Brien, E. P., & O'Brien, E. P. (2013). In vivo translation rates can substantially delay the cotranslational folding of the Escherichia coli cytosolic proteome. *Proceedings of the National Academy of Sciences of the United States of America*, *110*(2), E132-40.
<http://doi.org/10.1073/pnas.1213624110>
- Clarke IV, T. F., & Clark, P. L. (2008). Rare codons cluster. *PLoS ONE*, *3*(10).
<http://doi.org/10.1371/journal.pone.0003412>
- Daura, X., Gademann, K., Jaun, B., Seebach, D., Van Gunsteren, W. F., & Mark, A. E. (1999). Peptide Folding: When Simulation Meets Experiment. *Angew. Chem. Int. Ed*, *38*(1–2), 236-240.

- Day, R., Paschek, D., & Garcia, A. E. (2010). Microsecond simulations of the folding/unfolding thermodynamics of the Trp-cage miniprotein. *Proteins: Structure, Function and Bioinformatics*, *78*(8), 1889–1899. <http://doi.org/10.1002/prot.22702>
- Deane, C. M., Dong, M., Huard, F. P. E., Lance, B. K., & Wood, G. R. (2007). Cotranslational protein folding - Fact or fiction? *Bioinformatics*, *23*(13), 142–148. <http://doi.org/10.1093/bioinformatics/btm175>
- DeLano, W. L. (2002). The PyMOL Molecular Graphics System. Schrödinger. Retrieved from <http://www.pymol.org>
- Ding, F., Buldyrev, S. V., & Dokholyan, N. V. (2005). Folding Trp-Cage to NMR Resolution Native Structure Using a Coarse-Grained Protein Model. *Biophysical Journal*, *88*(1), 147–155. <http://doi.org/10.1529/biophysj.104.046375>
- Elcock, A. H. (2006). Molecular simulations of cotranslational protein folding: Fragment stabilities, folding cooperativity, and trapping in the ribosome. *PLoS Computational Biology*, *2*(7), 0824–0841. <http://doi.org/10.1371/journal.pcbi.0020098>
- Ellis, J. J., Huard, F. P. E., Deane, C. M., Srivastava, S., & Wood, G. R. (2010). Directionality in protein fold prediction. *BMC Bioinformatics*, *11*, 172. <http://doi.org/10.1186/1471-2105-11-172>
- Ellis, J. P., Bakke, C. K., Kirchdoerfer, R. N., Jungbauer, L. M., & Cavagnero, S. (2008). Chain dynamics of nascent polypeptides emerging from the ribosome. *ACS Chemical Biology*, *3*(9), 555–566. <http://doi.org/10.1021/cb800059u>
- Englander, S. W., & Mayne, L. (2014). The nature of protein folding pathways. *Proceedings of the National Academy of Sciences of the United States of America*, *111*(45), 1–8. <http://doi.org/10.1073/pnas.1411798111>
- Evans, M. S., Sander, I. M., & Clark, P. L. (2008). Cotranslational Folding Promotes β -Helix Formation and Avoids Aggregation In Vivo. *Journal of Molecular Biology*, *383*(3), 683–692. <http://doi.org/10.1016/j.jmb.2008.07.035>
- Fedorov, A. N., & Baldwin, T. O. (1997). Cotranslational Protein Folding. *The Journal of Biological Chemistry*, *272*(52), 32715–32718. <http://doi.org/10.1074/jbc.272.52.32715>
- García, a E., & Sanbonmatsu, K. Y. (2001). Exploring the energy landscape of a beta hairpin in explicit solvent. *Proteins*, *42*(3), 345–54.
- Hao, G.-F., Xu, W.-F., Yang, S.-G., & Yang, G.-F. (2015). Multiple Simulated Annealing-Molecular Dynamics (MSA-MD) for Conformational Space Search of Peptide and Miniprotein. *Scientific Reports*, *5*(October), 15568. <http://doi.org/10.1038/srep15568>
- Hawkins, G. D., Cramer, C. J., & Truhlar, D. G. (1995). Pairwise solute descreening of solute charges from a dielectric medium. *Chemical Physics Letters*, *246*(1–2), 122–129.
- Hawkins, G. D., Cramer, C. J., & Truhlar, D. G. (1996). Parametrized Models of Aqueous Free Energies of Solvation Based on Pairwise Descreening of Solute Atomic Charges from a Dielectric Medium. *The Journal of Physical Chemistry*, *100*(51), 19824–19839. <http://doi.org/10.1021/jp961710n>
- Hornak, V., Abel, R., Okur, A., Strockbine, B., Roitberg, A., & Simmerling, C. (2006). Comparison of multiple amber force fields and development of improved protein backbone parameters. *Proteins: Structure, Function and Genetics*, *65*(3), 712–725. <http://doi.org/10.1002/prot.21123>

- Huard, F. P. E., Deane, C. M., & Wood, G. R. (2006). Modelling sequential protein folding under kinetic control. *Bioinformatics*, *22*(14), 203–210. <http://doi.org/10.1093/bioinformatics/btl248>
- Humphrey, W., Dalke, A., & Schulten, K. (1996). VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, *14*(1), 33–38. [http://doi.org/10.1016/0263-7855\(96\)00018-5](http://doi.org/10.1016/0263-7855(96)00018-5)
- Jain, a. K., Murty, M. N., & Flynn, P. J. (1999). Data clustering: a review. *ACM Computing Surveys*, *31*(3), 264–323. <http://doi.org/10.1145/331499.331504>
- Jamroz, M., & Kolinski, A. (2013). ClusCo: clustering and comparison of protein models. *BMC Bioinformatics*, *14*(1), 62. <http://doi.org/10.1186/1471-2105-14-62>
- Johansson, M. U., de Château, M., Wikström, M., Forsén, S., Drakenberg, T., & Björck, L. (1997). Solution structure of the albumin-binding GA module: a versatile bacterial protein domain. *Journal of Molecular Biology*, *266*(5), 859–65. <http://doi.org/10.1006/jmbi.1996.0856>
- Kannan, S., & Zacharias, M. (2009). Folding of Trp-cage mini protein using temperature and biasing potential replica-exchange molecular dynamics simulations. *International Journal of Molecular Sciences*, *10*(3), 1121–1137. <http://doi.org/10.3390/ijms10031121>
- Karp, G. (2011). *Biología celular y molecular: conceptos y experimentos*. (6 ed.). Mexico: Mc Graw Hill.
- Karplus, M., & McCammon, J. A. (2002). Molecular dynamics simulations of biomolecules. *Nature Structural Biology*, *9*(9), 646–652. <http://doi.org/10.1038/nsb0902-646>
- Kolb, V. A. (2001). Cotranslational Protein Folding. *Molecular Biology*, *35*(4), 584–590. <http://doi.org/10.1023/A:1010579111510>
- Komar, A. A. (2009). A pause for thought along the co-translational folding pathway. *Trends in Biochemical Sciences*, *34*(1), 16–24. <http://doi.org/10.1016/j.tibs.2008.10.002>
- Kosolapov, A., & Deutsch, C. (2009). Tertiary interactions within the ribosomal exit tunnel. *Nature Structural & Molecular Biology*, *16*(4), 405–11. <http://doi.org/10.1038/nsmb.1571>
- Kramer, G., Boehringer, D., Ban, N., & Bukau, B. (2009). The ribosome as a platform for co-translational processing, folding and targeting of newly synthesized proteins. *Nature Structural & Molecular Biology*, *16*(6), 589–597. <http://doi.org/10.1038/nsmb.1614>
- Kufareva, I., & Abagyan, R. (2012). Methods of Protein Structure Comparison. In A. J. W. Orry & R. Abagyan (Eds.), *Homology Modeling: Methods and Protocols* (pp. 231–257). Totowa, NJ: Humana Press. http://doi.org/10.1007/978-1-61779-588-6_10
- Lesk, A. M. (2001). *Introduction to Protein Architecture*. New York: Oxford University Press.
- Lindorff-Larsen, K., Piana, S., Dror, R. O., & Shaw, D. E. (2011). How Fast-Folding Proteins Fold. *Science*, *334*(6055), 517–520. <http://doi.org/10.1126/science.1214018>
- Lodish, H., Berk, A., Matsuidara, P., Kaiser, C. A., Krieger, M., Scott, M. P., ... Darnell, J. (2005). *Biología Celular y Molecular* (5 ed.). Argentina: Editorial Médica Panamericana.
- Lu, J., & Deutsch, C. (2005). Folding zones inside the ribosomal exit tunnel. *Nature Structural & Molecular Biology*, *12*(12), 1123–1129. <http://doi.org/10.1038/nsmb1021>
- Lu, J., & Deutsch, C. (2008). Electrostatics in the Ribosomal Tunnel Modulate Chain

- Elongation Rates. *Journal of Molecular Biology*, 384(1), 73–86.
<http://doi.org/10.1016/j.jmb.2008.08.089>
- Lu, J., Hua, Z., Kobertz, W. R., & Deutsch, C. (2011). Nascent peptide side chains induce rearrangements in distinct locations of the ribosomal tunnel. *Journal of Molecular Biology*, 411(2), 499–510. <http://doi.org/10.1016/j.jmb.2011.05.038>
- Lu, J., Kobertz, W. R., & Deutsch, C. (2007). Mapping the Electrostatic Potential within the Ribosomal Exit Tunnel. *Journal of Molecular Biology*, 371(5), 1378–1391.
<http://doi.org/10.1016/j.jmb.2007.06.038>
- Morrissey, M. P., Ahmed, Z., & Shakhnovich, E. I. (2004). The role of cotranslation in protein folding: A lattice model study. *Polymer*, 45(2), 557–571.
<http://doi.org/10.1016/j.polymer.2003.10.090>
- Neidigh, J. W., Fesinmeyer, R. M., & Andersen, N. H. (2002). Designing a 20-residue protein. *Nature Structural Biology*, 9(6), 425–430. <http://doi.org/10.1038/nsb798>
- Nguyen, H., Maier, J., Huang, H., Perrone, V., & Simmerling, C. (2014). Folding simulations for proteins with diverse topologies are accessible in days with a physics-based force field and implicit solvent. *Journal of the American Chemical Society*, 136(40), 13959–13962.
<http://doi.org/10.1021/ja5032776>
- Nicola, A. V., Chen, W., & Helenius, A. (1999). Co-translational folding of an alphavirus capsid protein in the cytosol of living cells. *Nat. Cell. Biol.*, 1(6), 341–345.
<http://doi.org/10.1038/14032>
- O'Brien, E. P., Ciryam, P., Vendruscolo, M., & Dobson, C. M. (2014). Understanding the influence of codon translation rates on cotranslational protein folding. *Accounts of Chemical Research*, 47(5), 1536–1544. <http://doi.org/10.1021/ar5000117>
- O'Brien, E. P., Stan, G., Thirumalai, D., & Brooks, B. R. (2008). Factors governing helix formation in peptides confined to carbon nanotubes. *Nano Letters*, 8(11), 3702–3708.
<http://doi.org/10.1021/nl8019328>
- O'Brien, E. P., Vendruscolo, M., & Dobson, C. M. (2012). Prediction of variable translation rate effects on cotranslational protein folding. *Nature Communications*, 3(May), 868.
<http://doi.org/10.1038/ncomms1850>
- O'Brien, E. P., Vendruscolo, M., & Dobson, C. M. (2014). Kinetic modelling indicates that fast-translating codons can coordinate cotranslational protein folding by avoiding misfolded intermediates. *Nature Communications*, 5. <http://doi.org/10.1038/ncomms3988>
- Pace, C. N., & Scholtz, J. M. (1998). A helix propensity scale based on experimental studies of peptides and proteins. *Biophysical Journal*, 75(1), 422–427.
[http://doi.org/10.1016/S0006-3495\(98\)77529-0](http://doi.org/10.1016/S0006-3495(98)77529-0)
- Palenčár, P., & Bleha, T. (2010). Folding of polyalanine into helical hairpins. *Macromolecular Theory and Simulations*, 19(8–9), 488–495. <http://doi.org/10.1002/mats.201000034>
- Palenčár, P., & Bleha, T. (2011). Molecular dynamics simulations of the folding of poly(alanine) peptides. *Journal of Molecular Modeling*, 17(9), 2367–2374.
<http://doi.org/10.1007/s00894-011-0997-4>
- Patriksson, A., & van der Spoel, D. (2008). A temperature predictor for parallel tempering simulations. *Physical Chemistry Chemical Physics*, 10(15), 2073.
<http://doi.org/10.1039/b716554d>

- Perez, A., MacCallum, J. L., & Dill, K. A. (2015). Accelerating molecular simulations of proteins using Bayesian inference on weak information. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(38), 11846–51. <http://doi.org/10.1073/pnas.1515561112>
- Plotkin, S. S., & Onuchic, J. N. J. N. (2002). *Understanding protein folding with energy landscape theory. Part I: basic concepts. Quarterly Reviews of Biophysics* (Vol. 35). <http://doi.org/10.1017/S0033583502003785>
- Rasmus, D. M., & Hall, C. K. (1991). Prediction of gas adsorption in 5a zeolites using Monte Carlo simulation. *AIChE Journal*, *37*(5), 769–779. <http://doi.org/10.1002/aic.690370514>
- Roe, D. R., & Cheatham, T. E. (2013). PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *Journal of Chemical Theory and Computation*, *9*(7), 3084–3095. <http://doi.org/10.1021/ct400341p>
- Ryckaert, J. P., Ciccotti, G., & Berendsen, H. J. C. (1977). Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *Journal of Computational Physics*, *23*(3), 327–341. [http://doi.org/10.1016/0021-9991\(77\)90098-5](http://doi.org/10.1016/0021-9991(77)90098-5)
- Sanbonmatsu, K. Y., & García, A. E. (2002). Structure of Met-enkephalin in explicit aqueous solution using replica exchange molecular dynamics. *Proteins*, *46*(2), 225–34. <http://doi.org/10.1002/prot.0000>
- Simmerling, C., Strockbine, B., & Roitberg, A. E. (2002). All-atom structure prediction and folding simulations of a stable protein. *Journal of the American Chemical Society*, *124*(38), 11258–11259. <http://doi.org/10.1021/ja0273851>
- Sohl, J. L., Jaswal, S. S., & Agard, D. A. (1998). Unfolded conformations of alpha-lytic protease are more stable than its native state. *Nature*, *395*(6704), 817–9. <http://doi.org/10.1038/27470>
- Sugita, Y., & Okamoto, Y. (1999). Replica-exchange molecular dynamics method for protein folding. *Chemical Physics Letters*, *314*(1–2), 141–151. [http://doi.org/10.1016/S0009-2614\(99\)01123-9](http://doi.org/10.1016/S0009-2614(99)01123-9)
- Tanaka, T., Hori, N., & Takada, S. (2015). How Co-translational Folding of Multi-domain Protein Is Affected by Elongation Schedule: Molecular Simulations. *PLoS Computational Biology*, *11*(7), 1–20. <http://doi.org/10.1371/journal.pcbi.1004356>
- Tramontano, A. (2006). *Proteins Structure Prediction: Concepts and Applications*. Weinheim: WILEY-VCH Verlag GmbH & Co.
- Tsai, C. J., Sauna, Z. E., Kimchi-Sarfaty, C., Ambudkar, S. V., Gottesman, M. M., & Nussinov, R. (2008). Synonymous Mutations and Ribosome Stalling Can Lead to Altered Folding Pathways and Distinct Minima. *Journal of Molecular Biology*, *383*(2), 281–291. <http://doi.org/10.1016/j.jmb.2008.08.012>
- Tu, L. W., & Deutsch, C. (2010). A Folding Zone in the Ribosomal Exit Tunnel for Kv1.3 Helix Formation. *Journal of Molecular Biology*, *396*(5), 1346–1360. <http://doi.org/10.1016/j.jmb.2009.12.059>
- Voss, N. R., Gerstein, M., Steitz, T. A., & Moore, P. B. (2006). The Geometry of the Ribosomal Polypeptide Exit Tunnel. *Journal of Molecular Biology*, *360*(4), 893–906. <http://doi.org/10.1016/j.jmb.2006.05.023>

- Vujic, B., & Lyubartsev, A. P. (2016). Transferable force-field for modeling of CO₂, N₂, Ar and O₂ in all silica and Na⁺ exchanged zeolite. *Modelling Simul. Mater. Sci. Eng*, *24*, 1–26. <http://doi.org/10.1088/0965-0393/24/4/045002>
- Wang, E., Wang, J., Chen, C., & Xiao, Y. (2015). Computational evidence that fast translation speed can increase the probability of cotranslational protein folding. *Scientific Reports*, *5*(October), 15316. <http://doi.org/10.1038/srep15316>
- Weiser, J., Shenkin, P. S., Still, W. C., Lcpo, O., Shenkin, P. S., & Still, W. C. (1999). Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem*, *20*(2), 217–230.
- Wen, J.-D., Lancaster, L., Hodges, C., Zeri, A.-C., Yoshimura, S. H., Noller, H. F., ... Tinoco, I. (2008). Following translation by single ribosomes one codon at a time. *Nature*, *452*(7187), 598–603. <http://doi.org/10.1038/nature06716>
- Wilson, D. N., & Beckmann, R. (2011). The ribosomal tunnel as a functional environment for nascent polypeptide folding and translational stalling. *Current Opinion in Structural Biology*, *21*(2), 274–282. <http://doi.org/10.1016/j.sbi.2011.01.007>
- Wood, G. R., Srivastava, S., Patton, Y., & Fisher, D. W. (2011). Cotranslational protein folding and terminus hydrophobicity. *Advances in Bioinformatics, 2011*. <http://doi.org/10.1155/2011/176813>
- Yu, C. H., Dang, Y., Zhou, Z., Wu, C., Zhao, F., Sachs, M. S., & Liu, Y. (2015). Codon Usage Influences the Local Rate of Translation Elongation to Regulate Co-translational Protein Folding. *Molecular Cell*, *59*(5), 744–754. <http://doi.org/10.1016/j.molcel.2015.07.018>
- Zemla, A., Venclovas, Č., Moulton, J., & Fidelis, K. (1999). Processing and analysis of CASP3 protein structure predictions. *Proteins: Structure, Function and Genetics*, *37*(SUPPL. 3), 22–29.
- Zerella, R., Chen, P. Y., Evans, P. A., Raine, A., & Williams, D. H. (2000). Structural characterization of a mutant peptide derived from ubiquitin: implications for protein folding. *Protein Science : A Publication of the Protein Society*, *9*(11), 2142–50. <http://doi.org/10.1110/ps.9.11.2142>
- Zhang, G., Hubalewska, M., & Ignatova, Z. (2009). Transient ribosomal attenuation coordinates protein synthesis and co-translational folding. *Nature Structural & Molecular Biology*, *16*(3), 274–280. <http://doi.org/10.1038/nsmb.1554>
- Zhang, G., & Ignatova, Z. (2011). Folding at the birth of the nascent chain: Coordinating translation with co-translational folding. *Current Opinion in Structural Biology*, *21*(1), 25–31. <http://doi.org/10.1016/j.sbi.2010.10.008>
- Zhang, Y., Wölfle, T., & Rospert, S. (2013). Interaction of nascent chains with the ribosomal tunnel proteins Rpl4, Rpl17, and Rpl39 of *saccharomyces cerevisiae*. *Journal of Biological Chemistry*, *288*(47), 33697–33707. <http://doi.org/10.1074/jbc.M113.516641>
- Ziv, G., Haran, G., & Thirumalai, D. (2005). Ribosome exit tunnel can entropically stabilize alpha-helices. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(52), 18956–61. <http://doi.org/10.1073/pnas.0508234102>