

RESSALVA

Atendendo solicitação da autora, o texto completo desta dissertação será disponibilizado somente a partir de 04/02/2025.



UNIVERSIDADE ESTADUAL PAULISTA
"JÚLIO DE MESQUITA FILHO"
Câmpus de São José do Rio Preto

Larissa Maria Ferreira Adolfo

**A influência da escolha do mapa no estudo do enovelamento de proteínas
utilização do modelos baseados em estrutura**

**São José do Rio Preto
2023**

Larissa Maria Ferreira Adolfo

**A influência da escolha do mapa no estudo do enovelamento de proteínas
utilização do modelos baseados em estrutura**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES

Orientador: Prof. Dr. Leandro Cristante de Oliveira

São José do Rio Preto

2023

A239i Adolfo, Larissa Maria Ferreira.
A influência da escolha do mapa no estudo do enovelamento de
proteínas utilizando modelo baseado em estrutura / Larissa Maria
Ferreira Adolfo -- São José do Rio Preto, 2023
86 f.

Dissertação (mestrado) - Universidade Estadual Paulista (Unesp),
Instituto de Biociências Letras e Ciências Exatas, São José do Rio Preto
Orientador: Leandro Cristante de Oliveira

1. Mapa de Contato. 2. Enovelamento de Proteínas. 3. Biofísica. I.
Título.

Sistema de geração automática de fichas catalográficas da Unesp. Biblioteca do Instituto de
Biociências, Letras e Ciências Exatas, São José do Rio Preto. Dados fornecidos pelo autor(a).

Essa ficha não pode ser modificada.

Larissa Maria Ferreira Adolfo

**A influência da escolha do mapa no estudo do enovelamento de proteínas
utilização do modelos baseados em estrutura**

Dissertação apresentada como parte dos requisitos para obtenção do título de Mestre em Biofísica Molecular, junto ao Programa de Pós-Graduação em Biofísica Molecular, do Instituto de Biociências, Letras e Ciências Exatas da Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de São José do Rio Preto.

Financiadora: CAPES:

Comissão Examinadora

Prof. Dr. Leandro Cristante de Oliveira
UNESP – Câmpus de São José do Rio Preto
Orientador

Prof. Dr. Laurent Emmanuel Dardenne
LNCC – Laboratório Nacional de Computação Científica

Prof. Dr. Rui Travasso
CFisUC – Centro de Física da Universidade de Coimbra

São José do Rio Preto

04 de Agosto 2023

AGRADECIMENTOS

Inicialmente agradeço a mim por todo esforço e dedicação. Aos meus pais, irmã, sobrinha e minha namorada Jennifer por não deixarem que eu esqueça dos meus objetivos de vida e nem que eu desanime.

Às meninas da república Tilangas (2022/2023), por me fazerem sair do "casulo" e aos pessoal da sala 42 (2022/2023) por todo apoio, discussões e cafés. Também à minha parceira de sala, ou o "Lari-Verso". À Carol, por ser minha amiga desde a graduação e por sempre me apoiar. Ao meu psicólogo e ao departamento de Física e a UNESP.

Ao meu grupo de pesquisa '*DoPreDin*': A Carol e Raphael, por toda a discussão e apoio, sem vocês eu teria surtado. E principalmente ao meu orientador Leandro, por toda a paciência, ensinamento, confiança e amizade gerada nesse tempo.

À banca examinadora, pela disposição e contribuição para este trabalho.

Esta pesquisa tornou-se possível também graças aos recursos computacionais disponibilizados pelo Núcleo de Computação Científica (NCC/GridUNESP) da Universidade Estadual Paulista (UNESP).

O presente trabalho também foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 (88882.461721/2019-01), à qual agradeço.

“A energia que move o mundo,
é a que deixa o caixa eletrônico ligado”

- **Canal “O Algoritmo da Imagem”**

RESUMO

Uma das formas de estudar o enovelamento proteico é utilizando o Modelo Baseado em Estrutura (do inglês: *Structure Based Model*, SBM). Sua energia potencial é determinada pelas propriedades geométricas de uma dada estrutura nativa e um mapa que define os contatos nativos. Esses mapas de contato nativo são responsáveis pela representação das interações de resíduos ij que estabelecem interações atrativas, estabilizando o estado nativo e fornecendo interações de curto alcance. Para tanto, foram estudados três mapas de contato com amplo emprego na literatura: o mapa de corte para átomos pesados (do Inglês: *Heavy Atom Cut-Offs*, CO), o Mapa de Contatos de Sombra (do inglês: *Shadow Contacts Map*, SCM), o mapa de Contatos Estruturais (do inglês: *Contact of Structural Units*, CSU) e adaptamos o Frustratômetro (do inglês: *Frustratometer*, Frt) a fim de obter um quarto mapa. Realizou-se, então, a aplicação desses, para três proteínas diferentes, a Inibidor de quimotripsina 2 (PDB id: 2ci2) uma proteína com enovelamento robusta e característico para testes computacionais. A Barnase (PDB id: 1bnr) que contém um processo de enovelamento constituído de um estado intermediário, e o domínio SH3 (PDB id: 1 FMK) que possui um caminho de enovelamento determinado experimentalmente. Como resultados, temos que o SBM demonstra robustez com a troca de mapa de contato. Analisando unicamente pela coordenada de reação Q como comumente encontrado na literatura, temos que somente o mapa de CSU consegue capturar o estado intermediário presente na curva de energia livre para a proteína Barnase e o caminho de formação demonstrado experimentalmente para o domínio SH3. Porém utilizando diferentes coordenadas de reação nenhum dos mapas atesta o estado intermediário previsto para a Barnase, indicando que para a simplificação SBM-C α , ocorrem perdas significativas de informação. Então, a partir deste verificamos que os mapas de contato devem introduzir a heterogeneidade entre os resíduos de aminoácidos e interações

de mais longa distância como demonstrado pelo mapa do CSU, de forma a corrigir perdas de informações causadas pela simplificação do modelo.

Palavras-chave: *Mapa de Corte para átomos pesados. Mapa de Contato de Sombra. Mapa de Contato por Unidades Estruturais. Frustratrômetro. Modelo Baseado em Estrutura. Enovelamento de Proteínas.*

ABSTRACT

One way of studying protein folding is using the Structure Based Model (SBM). Its potential energy is determined by the geometric properties of a given native structure and a map that defines the native contacts. These contact maps are responsible for representing the interactions of ij residues that establish attractive interactions, stabilizing the native state and providing short-range interactions. For this purpose, three contact maps widely used in the literature were studied: the Heavy Atom Cut-Offs (CO), the Shadow Contacts Map (SCM), the Contact of Structural Units (CSU) map and we adapted the Frustratometer (Frt) in order to obtain a fourth map. These were then applied to three different proteins, the Chymotrypsin 2 Inhibitor (PDB id: 2ci2), a robust and characteristic protein for computational tests. The Barnase (PDB id: 1bnr) which contains a folding process consisting of an intermediate state, and the SH3 domain (PDB id: 1 FMK) which has an experimentally determined folding path. As a result, we have that the SBM demonstrates robustness with the exchange of contact maps. Analyzing only by the Q reaction coordinate as commonly found in the literature, we have that only the CSU map manages to capture the intermediate state present in the free energy curve for the Barnase protein and the experimentally demonstrated formation path for the SH3 domain. However, using different reaction coordinates, they demonstrated that none of the maps attests to the intermediate state predicted for Barnase, indicating that for simplification with only alpha carbons, significant losses of information occur, so that the contact map should bring heterogeneity to the SBMs, to be able to correct these losses of information.

Keywords: *Heavy Atom Cut-Offs. Shadow Contacts Map. Contact of Structural Units. Frustratometer. Structure Based Model. Protein Folding.*

LISTA DE ILUSTRAÇÕES

Figura 1.1 Representação estrutural da proteína PDB id: 1bnr em sua estrutura primária, secundária e terciária.

Figura 1.2 Representação da superfície de energia em forma de funil e energia livre associada.

Figura 1.3 Representação esquemática do Mapa de Contato de Corte (CO).

Figura 1.4 Representação esquemática do Mapa de Contato de Sombra (SCM)

Figura 1.5 Representação esquemática da determinação de contato pelo método de Contato de Unidades Estruturais (CSU).

Figura 1.6 Representação esquemática do processo de interação dos contatos do mapa de Frustratrometro.

Figura 1.7 Representação em cartoon da estrutura cristalográfica e/ou RMN das proteínas de interesse.

Figura 3.1 Representação de todos os átomos e $C\alpha$, por esferas de Van der Waals da proteína *Barnase*.

Figura 3.2 Representação de contato formado pela metodologia de contatos quaisquer (Z).

Figura 4.1 Gráfico de Energia Livre, $F(Q)$ e Rota, $R(Q)$, pela coordenada de reação, Q , e gráfico de Energia Livre, $F(\text{RMSD})$ pela coordenada de reação RMSD, para o mapa de Cut-Off (CO).

Figura 4.2 Gráfico de Energia Livre, $F(Q)$ e Rota, $R(Q)$, pela coordenada de reação, Q , e gráfico de Energia Livre, $F(\text{RMSD})$ pela coordenada de reação RMSD, para o mapa de SCM.

Figura 4.3 Gráfico de Energia Livre, $F(Q)$ e Rota, $R(Q)$, pela coordenada de reação, Q , e gráfico de Energia Livre, $F(\text{RMSD})$ pela coordenada de reação RMSD, para o mapa de CSU.

Figura 4.4 Gráfico de Energia Livre, $F(Q)$ e Rota, $R(Q)$, pela coordenada de reação, Q , e gráfico de Energia Livre, $F(\text{RMSD})$ pela coordenada de reação RMSD, para o mapa de Frt.

Figura 4.5 Probabilidade de formação de contatos para a proteína 2ci2 (PDB id: 2ci2).

Figura 4.6 Probabilidade de formação de contatos para a proteína Barnase (PDB id: 1BNR).

Figura 4.7 Probabilidade de formação de contatos para o domínio SH3 (PDB id: 1 FMK).

Figura 4.8 Aplicação de corte de distância no mapa de CSU, para a proteína Barnase (PDB id: 1 bnr).

Figura 4.9 Aplicação de corte de distância no mapa de CSU, para o domínio SH3 (PDB id: 1 FMK).

Figura 4.10 Energia Livre $F(\text{RMSD})$ pela coordenada de reação RMSD.

Figura 4.11 Correlação entre as coordenadas de reação Q e RMSD, colorida a partir da energia livre de cada estado.

Figura 4.12 Espaço de fase conformacional e funil de estruturas visualizado através do ELVIM e gráfico de formação de estrutura para o mapa de Cut-Off para a proteína Barnase (PDB id: 1bnr).

Figura 4.13 Espaço de fase conformacional e funil de estruturas visualizado através do ELVIM e gráfico de formação de estrutura para o mapa de SCM para a proteína Barnase (PDB id: 1bnr).

Figura 4.14 Espaço de fase conformacional e funil de estruturas visualizado através do ELVIM e gráfico de formação de estrutura para o mapa de CSU para a proteína Barnase (PDB id: 1bnr).

Figura 4.15 Espaço de fase conformacional e funil de estruturas visualizado através do ELVIM e gráfico de formação de estrutura para o mapa adaptado do Frt para a proteína Barnase (PDB id: 1bnr).

Figura 4.16 Colapso da cadeia a partir da coordenada de contatos quaisquer Z , pela temperatura.

Figura 4.17 Energia Livre $F(Z)$ pela coordenada de reação Z , contatos quaisquer.

LISTA DE TABELAS

Tabela 4.1. Informações Obtidas a partir das interações dos mapas de contato e suas respectivas temperaturas de enovelamento para as três proteínas propostas, Chymotrypsin Inhibitor 2 Ci2 (PDB id: 2ci2), Barnase (PDB id: 1 bnr) e o domínio SH3 da tirosina quinase humana (PDB id: 1 FMK).

LISTA DE SIGLAS

CO - Cut-Off

CSU - Contact of Structural Units

D - Desenovelado

ES - Estado de Transição

Frt - Frustratrômetro

GROMACS - GRoningen MACHine for Chemical Simulations

N - Nativo

PDB id - Protein Data Bank Identification

PDB - Protein Data Bank

RMSD - Root Mean Square Deviation

SBM - Structure Based Model

SCM - Shadow Contacts Map

WHAM - Weighted Histogram Analysis Method

SUMÁRIO

1	INTRODUÇÃO	15
	1.2 Fundamentação teórica	19
	1.2.1 Modelos computacionais e superfície de energia	19
	1.2.2 Mapas de contato	22
	1.3 Proteínas estudadas	30
2	OBJETIVO	33
3	MÉTODO	34
	3.1 Detalhes da simulação	36
	3.2 Formas de análise	37
4	RESULTADO E DISCUSSÕES	45
5	CONCLUSÃO	74
	REFERÊNCIAS	76

1 INTRODUÇÃO

Proteínas são macromoléculas biológicas, constituídas comumente por 20 diferentes tipos de aminoácidos que são comumente encontrados na natureza. Elas possuem diversas funcionalidades, como por exemplo controle e regulação (LEHNINGER, NELSON, COX, 201). Nos aminoácidos que compõem as proteínas, com exceção da glicina, o carbono α liga-se a quatro grupos diferentes: um grupo carboxila ($COOH$), um grupo amina (NH_2), um átomo de hidrogênio e um grupo R (ou cadeia lateral), o qual diferencia os aminoácidos em tamanho, estrutura e carga elétrica, assim como diferencia os aminoácidos também na propriedade relacionada com o perfil hidrofóbico.

A estrutura das proteínas é descrita por quatro níveis de complexidade estruturais comumente definidos (LEHNINGER, NELSON, COX, 2014). Uma cadeia de resíduos de aminoácidos formada por ligações peptídicas é dita como estrutura primária. Arranjos estáveis de resíduos de aminoácidos, que formam padrões particulares de estruturas como hélices α e/ou folhas β , são referentes à estrutura secundária. O arranjo tridimensional ou estrutural terciário é formado quando estruturas secundárias interagem através de forças eletrostáticas, ligações de hidrogênio e/ou dissulfeto e forças de Van der Waals, forças que atuam na estruturação e enovelamento protéico, ver Figura 1.1. Por fim, duas ou mais subunidades polipeptídicas formam a estrutura quaternária (dímero, trímero, oligômero, etc).

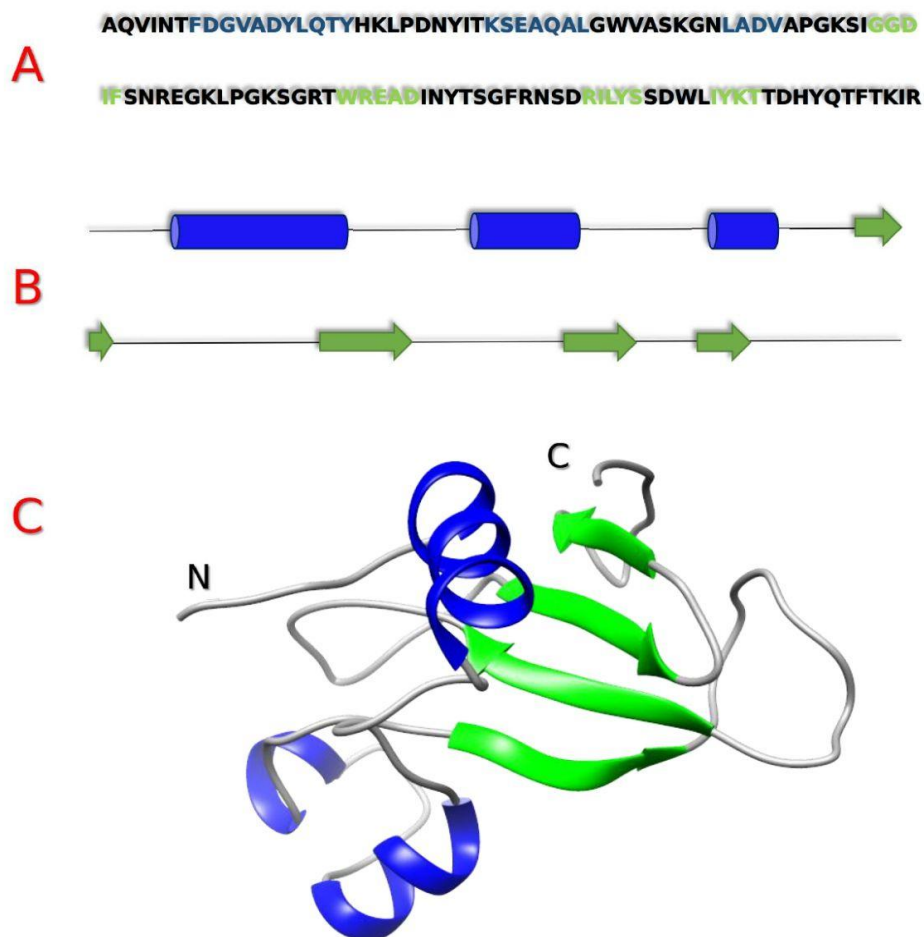
Compreender o enovelamento proteico, de modo a entender quais aminoácidos são essenciais para a formação da estrutura torna-se um importante alvo de estudo, visto que durante processos celulares, muitas vezes o mau funcionamento ou a má formação de proteínas pode acarretar falhas nos mecanismos de controle e regulação (LEHNINGER; NELSON; COX, 2014).

Essas falhas estão diretamente relacionadas ao enovelamento indesejável das proteínas (DOBSON, 2003) e podem causar doenças como fibrose cística (ROSA; DIAS; NOBRE; MORAIS, 2008), doença de Creutzfeldt-Jakob (popularmente conhecida como doença da vaca louca) (AZEVEDO; NASCIMENTO; QUINTELLA; ROSSO; MARANHÃO FILHO; CORRÊA; CHIMELLI; VINCENT; NOVIS, 2001), Alzheimer (SMITH, 1999), diabetes tipo II (MCLELLAN; BARBALHO; CATTALINI; LERARIO, 2007) e outras, esses estudos não serão detalhados aqui.

Uma das formas de estudar as proteínas que estão relacionadas a essas doenças e utilizando os modelos baseados em estrutura, também chamados de modelos do tipo G ϕ , são baseados na teoria da superfície energética e no princípio da frustração mínima (melhor explicado na seção 1.2). São bem-sucedidos em explicar questões fundamentais de, por exemplo, enovelamento de proteínas, enovelamento de RNA e/ou função de proteica (LUTZ; SINNER; BOZIC; KONDOV; SCHUG, 2014). Ao mesmo tempo, são computacionalmente baratos o suficiente para executar simulações complexas. No modelo baseado em estrutura, seu potencial é altamente dependente de um mapa de contato, no qual este define as interações nativas da proteína, interações essas que são essenciais para o processo de enovelamento quando em conjunto com o cálculo de Phi-Values.

Existem três mapas de contato comumente utilizados na literatura e um que aqui foi proposto. Entendendo que o potencial é altamente dependente do mapa, e que há quatro formas possíveis de se compreender as interações que ocorrem entre os resíduos de aminoácidos, se faz necessário um estudo sobre sua influência no processo de enovelamento protéico.

Figura 1.1 Representação estrutural da proteína PDB id 1 bnr em sua estrutura primária, secundária e terciária. As cores e as figuras geométricas indicam a representação unidimensional das estruturas secundárias. As diferentes letras indicam os diferentes tipos de resíduos de aminoácidos que compõem essa proteína. Em (A), temos a representação estrutural primária para a Barnase. Em (B) temos a representação da estrutura secundária. Em (C) temos a representação da estrutura terciária. A informação sobre a estrutura é obtida pelo *Protein Data Bank* - PDB, com o código de identificação (PDB ID: 1bnr). As terminações da proteína com início no N-terminal e final no C-terminal. Em azul é representada a estrutura de hélice α , em verde a estrutura de folhas β e em cinza claro as regiões sem estruturas definidas, ditas como aleatórias ou *random coil*. Figura produzida pelo *software* Chimera alpha versão 1.16.



Fonte: Elaborado pelo Autor

Na década de 1960, Anfinsen, por intermédio de experimentos de físico-química, propôs que a sequência de resíduos de aminoácidos continha a informação necessária para determinar o estado funcional de uma proteína, ou seja, seu estado nativo (ANFINSSEN; HABER; SELA; WHITE, 1961), o que recebeu o nome de hipótese termodinâmica. O experimento consistiu na desnaturação química da Rnase A (ribonuclease pancreática bovina),

uma enzima que tem como função facilitar a interação entre o DNA e o RNA das células pancreáticas dos bovinos. Anfinsen modificou as condições fisiológicas da proteína pela adição de um agente desnaturante, uréia, de forma a conseguir controlar sua desnaturação por meio da contagem das 4 pontes de sulfeto presentes na estrutura. Este experimento também mostrou que ao retornar a condições fisiológicas, i. e., ao retirar o agente desnaturante, a RNase A retornava ao seu estado nativo espontaneamente. A partir disso, ele propôs o que chamou-se de ‘Hipótese Termodinâmica’, que diz que toda informação necessária para encontrar o estado nativo está contido na sequência de aminoácidos que constitui a estrutura primária da proteína, de modo que a estrutura tridimensional de uma proteína nativa em suas condições fisiológicas (solvente, pH, força iônica, etc) está em seu estado de mínima energia livre. Contudo, em 1968, Levinthal, questionou que se a busca pelo estado nativo da proteína pudesse ser feita de maneira aleatória, esses processos poderiam levar a proteína a procurar todos os estados possíveis exaustivamente (ZWANZIG; SZABO; BAGCHI, 1992). Para ilustrar esta afirmação, considere a proteína Rnase A, que contém 102 resíduos de aminoácidos, logo $N = 102$ e que cada um desses N resíduos possa assumir duas 2 conformações, $k = 2$ e cada conformação pode ser acessada a cada 1 picosegundo, então temos:

$$k^N = 2^{102} \approx 10^{30} \approx 40 \text{ milhões de anos} \quad (1.1)$$

Este ficou conhecido como o Paradoxo de Levinthal (FINKELSTEIN, 1997). Motivado por esse paradoxo, novos estudos foram surgindo associados ao problema do enovelamento de proteínas, como por exemplo os modelos de rede cúbica, teoria de *Spin Glass*, teoria da superfície de energia e o modelo baseado em estrutura.

5 CONCLUSÃO

O ponto de partida deste estudo foi entender o modelo baseado em estrutura e como cada um dos quatro mapas de contatos aqui apresentados, de modo a compreender como os respectivos caracterizam as interações entre os resíduos de aminoácidos. Posteriormente, procuramos proteínas com características marcantes em seu processo de enovelamento, e com amplo estudo na literatura da área, de modo a verificar a influência desses mapas em seus processos de enovelamento e/ou capturas marcantes.

O número de interações obtidos para cada um dos mapas de contato para cada uma das três proteínas e seus respectivos resultados termodinâmicos utilizando a coordenada de reação de fração de contatos nativos (Q), coordenada está comumente utilizada na literatura da área, demonstraram temperatura de enovelamento, altura da barreira de energia livre e especificidade da rota de enovelamento diferentes, de modo que apenas o mapa de CSU consegue capturar o estado de transição demonstrado experimentalmente pela proteína Barnase, além de indicar um caminho com melhor concordância também para a proteína Barnase e o domínio SH3.

Investigações sobre o mapa de CSU, de modo que o que diferencia ele dos demais é sua capacidade de contemplar interações de mais longa distância, por sua metodologia contemplar uma diferenciação mínima entre os resíduos de aminoácidos utilizando a superfície de van der Waals. Deste modo que ao utilizar a coordenada de reação Q , o mapa tem uma maior concordância com os resultados experimentais.

Também aumentamos o raio de corte do Shadow Contacts Map, a fim de verificar se esses contatos de mais longa distância sobre uma metodologia diferente, conseguia contemplar esse estado intermediário, o que não ocorreu.

O estado intermediário indicado experimentalmente pela Barnase, assim como o caminho de formação do domínio SH3, é definido pela estruturação da cadeia lateral. Assim, ao utilizarmos SBM - C α , ao utilizar metodologias baseadas apenas na distância geométrica entre os resíduos de aminoácidos, não há diferenciação entre eles. De tal forma que a heterogeneidade entre os resíduos de aminoácidos devem ser introduzida pelo mapa de contato. Na utilização do mapa de CSU, em que a metodologia contempla contatos de mais longa distância, a superfície de energia do sistema utilizando apenas C α ainda é lisa, de tal forma que mudanças na coordenada de reação pode não capturar algumas características essenciais como o estado intermediário. Porém utilizando metodologias de visualização de superfície como o ELVIM, em que nos dá a média dos contatos nativos formados no ET, não evidencia o estado intermediário, mas nos mostra que o caminho de formação para que o processo de enovelamento ocorra, e o mesmo descrito experimentalmente para todos os mapas de contato.

A partir disso temos que o modelo baseado em estrutura é robusto com a troca de mapa de contato, e que o mapa de CSU apresenta maior acordo com os experimentais.

REFERÊNCIAS

- ABRAHAM, Mark James; MURTOLA, Teemu; SCHULZ, Roland; PÁLL, Szilárd; SMITH, Jeremy C.; HESS, Berk; LINDAHL, Erik. GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers. *Software*, [S.L.], v. 1-2, p. 19-25, set. 2015. Elsevier BV. <http://dx.doi.org/10.1016/j.softx.2015.06.001>.
- ANFINSEN, C. B.; HABER, E.; SELA, M.; WHITE, F. H.. THE KINETICS OF FORMATION OF NATIVE RIBONUCLEASE DURING OXIDATION OF THE REDUCED POLYPEPTIDE CHAIN. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 47, n. 9, p. 1309-1314, set. 1961. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.47.9.1309>.
- AZEVEDO, Marlos Fábio Alves de; NASCIMENTO, Fernanda; QUINTELLA, Leonardo; ROSSO, Ana Lucia Zuma de; MARANHÃO FILHO, Péricles A.; CORRÊA, Rosalie B.; CHIMELLI, Leila; VINCENT, Maurice; NOVIS, Sérgio Augusto Pereira. Doença de Creutzfeldt-Jakob: a propósito de um caso com comprometimento medular. *Arquivos de Neuro-Psiquiatria*, [S.L.], v. 59, n. 4, p. 964-967, dez. 2001. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s0004-282x2001000600024>.
- BEST, R. B.; HUMMER, G.; EATON, W. A.. Native contacts determine protein folding mechanisms in atomistic simulations. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 110, n. 44, p. 17874-17879, 15 out. 2013. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.1311599110>.
- BORREGUERO, Jose M.; DING, Feng; BULDYREV, Sergey V.; STANLEY, H. Eugene; DOKHOLYAN, Nikolay V.. Multiple Folding Pathways of the SH3 Domain. *Biophysical Journal*, [S.L.], v. 87, n. 1, p. 521-533, jul. 2004. Elsevier BV. <http://dx.doi.org/10.1529/biophysj.104.039529>
- BRYNGELSON, J D; WOLYNES, P G. Spin glasses and the statistical mechanics of protein folding. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 84, n. 21, p. 7524-7528, nov. 1987. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.84.21.7524>
- BRYNGELSON, Joseph D.; ONUCHIC, José Nelson; SOCCI, Nicholas D.; WOLYNES, Peter G.. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Structure, Function, and Genetics*, [S.L.], v. 21, n. 3, p. 167-195, mar. 1995. Wiley. <http://dx.doi.org/10.1002/prot.340210302>.
- BRYNGELSON, Joseph D.; WOLYNES, Peter G.. Intermediates and barrier crossing in a random energy model (with applications to protein folding). *The Journal Of Physical Chemistry*, [S.L.], v. 93, n. 19, p. 6902-6915, set. 1989. American Chemical Society (ACS). <http://dx.doi.org/10.1021/j100356a007>.
- CAPRIOTTI, Emidio; FARISELLI, Piero; ROSSI, Ivan; CASADIO, Rita. A three-state prediction of single point mutations on protein stability changes. *Bmc Bioinformatics*, [S.L.], v. 9, n. 2, mar. 2008. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1471-2105-9-s2-s6>.
- CLEMENTI, Cecilia; GARCÍA, Angel E.; ONUCHIC, José N.. Interplay Among Tertiary Contacts, Secondary Structure Formation and Side-chain Packing in the Protein Folding Mechanism: all-atom representation study of protein I. *Journal Of Molecular Biology*, [S.L.], v. 326, n. 3, p. 933-954, fev. 2003. Elsevier BV. [http://dx.doi.org/10.1016/s0022-2836\(02\)01379-7](http://dx.doi.org/10.1016/s0022-2836(02)01379-7).
- CLEMENTI, Cecilia; NYMEYER, Hugh; ONUCHIC, José Nelson. Topological and energetic factors: what determines the structural details of the transition state ensemble and "en-route" intermediates for protein folding? an investigation for small globular proteins. *Journal Of Molecular Biology*, [S.L.], v. 298, n. 5, p. 937-953, maio 2000. Elsevier BV. <http://dx.doi.org/10.1006/jmbi.2000.3693>.

Crippen GHHT: Distance Geometry and Molecular Conformation. John Wiley & Sons, New York; 1988. DILL, Ken A.; BROMBERG, Sarina; YUE, Kaizhi; CHAN, Hue Sun; FTEBIG, Klaus M.; YEE, David P.; THOMAS, Paul D.. Principles of protein folding - A perspective from simple exact models. Protein Science, [S.L.], v. 4, n. 4, p. 561-602, 31 dez. 1995. Wiley. <http://dx.doi.org/10.1002/pro.5560040401>.

DOBSON, Christopher M.. Protein folding and misfolding. Nature, [S.L.], v. 426, n. 6968, p. 884-890, dez. 2003. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/nature02261>.

DUARTE, Jose M; SATHYAPRIYA, Rajagopal; STEHR, Henning; FILIPPIS, Ioannis; LAPPE, Michael. Optimal contact definition for reconstruction of Contact Maps. BMC Bioinformatics, [S.L.], v. 11, n. 1, p. 1-11, 27 maio 2010. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1471-2105-11-283>.

EATON, William A.; MUÑOZ, Victor; THOMPSON, Peggy A.; HENRY, Eric R.; HOFRICHTER, James. Kinetics and Dynamics of Loops, α -Helices, β -Hairpins, and Fast-Folding Proteins. Accounts Of Chemical Research, [S.L.], v. 31, n. 11, p. 745-753, 1 nov. 1998. American Chemical Society (ACS). <http://dx.doi.org/10.1021/ar9700825>.

FERREIRO, Diego U.; HEGLER, Joseph A.; KOMIVES, Elizabeth A.; WOLYNES, Peter G.. Localizing frustration in native proteins and protein assemblies. Proceedings Of The National Academy Of Sciences, [S.L.], v. 104, n. 50, p. 19819-19824, 11 dez. 2007. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.0709915104>.

FERSHT, Alan R.. Protein folding and stability: the pathway of folding of barnase. Febs Letters, [S.L.], v. 325, n. 1-2, p. 5-16, 28 jun. 1993. Wiley. [http://dx.doi.org/10.1016/0014-5793\(93\)81405-o](http://dx.doi.org/10.1016/0014-5793(93)81405-o).

FINKELSTEIN, Alexei V; BADRETDINOV, Azat Ya. Rate of protein folding near the point of thermodynamic equilibrium between the coil and the most stable chain fold. Folding And Design, [S.L.], v. 2, n. 2, p. 115-121, abr. 1997. Elsevier BV. [http://dx.doi.org/10.1016/s1359-0278\(97\)00016-3](http://dx.doi.org/10.1016/s1359-0278(97)00016-3).

GIANNI, Stefano. Φ Value Analysis. Encyclopedia Of Biophysics, [S.L.], p. 17-21, 2013. Springer Berlin Heidelberg. http://dx.doi.org/10.1007/978-3-642-16712-6_609.

GO, N; TAKETOMI, H. Respective roles of short- and long-range interactions in protein folding. Proceedings Of The National Academy Of Sciences, [S.L.], v. 75, n. 2, p. 559-563, fev. 1978. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.75.2.559>.

GO, N. Theoretical Studies of Protein Folding. Annual Review Of Biophysics And Bioengineering, [S.L.], v. 12, n. 1, p. 183-210, jun. 1983. Annual Reviews. <http://dx.doi.org/10.1146/annurev.bb.12.060183.001151>.

GRANTCHAROVA, Viara P; RIDDLE, David S.; BAKER, David. Long-range order in the src SH3 folding transition state. **Proceedings Of The National Academy Of Sciences**, [S.L.], v. 97, n. 13, p. 7084-7089, 20 jun. 2000. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.97.13.7084>.

HARRISON, S. C.; DURBIN, R.. Is there a single pathway for the folding of a polypeptide chain? Proceedings Of The National Academy Of Sciences, [S.L.], v. 82, n. 12, p. 4028-4030, 1 jun. 1985. Proceedings of the National Academy of Sciences. <http://dx.doi.org/10.1073/pnas.82.12.4028>.

JACKSON, Sophie E.; ELMASRY, Nadia; FERSHT, Alan R.. Structure of the hydrophobic core in the transition state for folding of chymotrypsin inhibitor 2: a critical test of the protein engineering method of analysis. Biochemistry, [S.L.], v. 32, n. 42, p. 11270-11278, 1 out. 1993. American Chemical Society (ACS). <http://dx.doi.org/10.1021/bi00093a002>.

JENIK, M.; PARRA, R. G.; RADUSKY, L. G.; TURJANSKI, A.; WOLYNES, P. G.; FERREIRO, D. U.. Protein frustratometer: a tool to localize energetic frustration in protein molecules. *Nucleic Acids Research*, [S.L.], v. 40, n. 1, p. 348-351, 29 maio 2012. Oxford University Press (OUP). <http://dx.doi.org/10.1093/nar/gks447>.

KUMAR, Shankar; ROSENBERG, John M.; BOUZIDA, Djamel; SWENDSEN, Robert H.; KOLLMAN, Peter A.. THE weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal Of Computational Chemistry*, [S.L.], v. 13, n. 8, p. 1011-1021, out. 1992. Wiley. <http://dx.doi.org/10.1002/jcc.540130812>.

LAMMERT, Heiko; NOEL, Jeffrey K.; ONUCHIC, José N.. The Dominant Folding Route Minimizes Backbone Distortion in SH3. *Plos Computational Biology*, [S.L.], v. 8, n. 11, 15 nov. 2012. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pcbi.1002776>.

LEHNINGER, T. M., NELSON, D. L. & COX, M. M. *Princípios de Bioquímica*. 6ª Edição, 2014. Ed. Artmed.

LEOPOLD, P. E.; MONTAL, M.; ONUCHIC, J. N.. Protein folding funnels: a kinetic approach to the sequence-structure relationship.. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 89, n. 18, p. 8721-8725, 15 set. 1992. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.89.18.8721>.

LUTZ, Benjamin; SINNER, Claude; BOZIC, Stefan; KONDOV, Ivan; SCHUG, Alexander. Native structure-based modeling and simulation of biomolecular systems per mouse click. *Bmc Bioinformatics*, [S.L.], v. 15, n. 1, p. 1-1, 29 ago. 2014. Springer Science and Business Media LLC. <http://dx.doi.org/10.1186/1471-2105-15-292>

MATOUSCHEK, Andreas; SERRANO, Luis; FERSHT, Alan R.. The folding of an enzyme. **Journal Of Molecular Biology**, [S.L.], v. 224, n. 3, p. 819-835, abr. 1992. Elsevier BV. [http://dx.doi.org/10.1016/0022-2836\(92\)90564-z](http://dx.doi.org/10.1016/0022-2836(92)90564-z).

MCLELLAN, Kátia Cristina Portero; BARBALHO, Sandra Maria; CATTALINI, Marino; LERARIO, Antonio Carlos. Diabetes mellitus do tipo 2, síndrome metabólica e modificação no estilo de vida. *Revista de Nutrição*, [S.L.], v. 20, n. 5, p. 515-524, out. 2007. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1415-52732007000500007>.

MCPHALEN, C. A.; JAMES, M. N. G.. Crystal and molecular structure of the serine proteinase inhibitor CI-2 from barley seeds. *Biochemistry*, [S.L.], v. 26, n. 1, p. 261-269, 13 jan. 1987. American Chemical Society (ACS). <http://dx.doi.org/10.1021/bi00375a036>.

NOEL, Jeffrey K.; WHITFORD, Paul C.; ONUCHIC, José N.. The Shadow Map: a general contact definition for capturing the dynamics of biomolecular folding and function. *The Journal Of Physical Chemistry B*, [S.L.], v. 116, n. 29, p. 8692-8702, 11 maio 2012. American Chemical Society (ACS). <http://dx.doi.org/10.1021/jp300852d>.

OLIVEIRA, Antonio B.; YANG, Huan; WHITFORD, Paul C.; LEITE, Vitor B. P.. Distinguishing Biomolecular Pathways and Metastable States. *Journal Of Chemical Theory And Computation*, [S.L.], v. 15, n. 11, p. 6482-6490, 16 out. 2019. American Chemical Society (ACS). <http://dx.doi.org/10.1021/acs.jctc.9b00704>.

PAPOIAN, Garegin A.; ULANDER, Johan; EASTWOOD, Michael P.; LUTHEY-SCHULTEN, Zaida; WOLYNES, Peter G.. Water in protein structure prediction. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 101, n. 10, p. 3352-3357, 26 fev. 2004. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.0307851100>.

PETTERSEN, Eric F.; GODDARD, Thomas D.; HUANG, Conrad C.; COUCH, Gregory S.; GREENBLATT, Daniel M.; MENG, Elaine C.; FERRIN, Thomas E.. UCSF Chimera?A visualization system for exploratory

research and analysis. *Journal Of Computational Chemistry*, [S.L.], v. 25, n. 13, p. 1605-1612, 2004. Wiley. <http://dx.doi.org/10.1002/jcc.20084>.

PRONK, Sander; PÁLL, Szilárd; SCHULZ, Roland; LARSSON, Per; BJELKMAR, Pär; APOSTOLOV, Rossen; SHIRTS, Michael R.; SMITH, Jeremy C.; KASSON, Peter M.; SPOEL, David van Der. GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics*, [S.L.], v. 29, n. 7, p. 845-854, 13 fev. 2013. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/btt055>.

ROSA, Fernanda Ribeiro; DIAS, Fernanda Gomes; NOBRE, Luciana Neri; MORAIS, Harriman Aley. Fibrose cística: uma abordagem clínica e nutricional. *Revista de Nutrição*, [S.L.], v. 21, n. 6, p. 725-737, dez. 2008. FapUNIFESP (SciELO). <http://dx.doi.org/10.1590/s1415-52732008000600011>.

SATHYAPRIYA, R.; VIJAYABASKAR, M. S.; VISHVESHVARA, Saraswathi. Insights into Protein–DNA Interactions through Structure Network Analysis. *Plos Computational Biology*, [S.L.], v. 4, n. 9, p. 1-10, 5 set. 2008. Public Library of Science (PLoS). <http://dx.doi.org/10.1371/journal.pcbi.1000170>.

SERRANO, Luis; MARTÍNEZ, Jose C.. The folding transition state between SH3 domains is conformationally restricted and evolutionarily conserved. *Nature Structural Biology*, [S.L.], v. 6, n. 11, p. 1010-1016, 1 nov. 1999. Springer Science and Business Media LLC. <http://dx.doi.org/10.1038/14896>.

SOBOLEV, V; A SOROKINE,; PRILUSKY, J; ABOLA, E.; EDELMAN, M. Automated analysis of interatomic contacts in proteins. *Bioinformatics*, [S.L.], v. 15, n. 4, p. 327-332, 1 abr. 1999. Oxford University Press (OUP). <http://dx.doi.org/10.1093/bioinformatics/15.4.327>.

SOBOLEV, Vladimir; EDELMAN, Marvin. Modeling the quinone-B binding site of the photosystem-II reaction center using notions of complementarity and contact-surface between atoms. *Proteins: Structure, Function, and Genetics*, [S.L.], v. 21, n. 3, p. 214-225, mar. 1995. Wiley. <http://dx.doi.org/10.1002/prot.340210304>.

SOCCI, N. D.; ONUCHIC, J. N.; WOLYNES, P. G.. Diffusive dynamics of the reaction coordinate for protein folding funnels. *The Journal Of Chemical Physics*, [S.L.], v. 104, n. 15, p. 5860-5868, 15 abr. 1996. AIP Publishing. <http://dx.doi.org/10.1063/1.471317>.

SPOEL, D. van der et al. *Gromacs User Manual version 4.0*. [S.l.], 2009.

TAKETOMI, Hiroshi; UEDA, Yuzo; Gō, Nobuhiro. STUDIES ON PROTEIN FOLDING, UNFOLDING AND FLUCTUATIONS BY COMPUTER SIMULATION. *International Journal Of Peptide And Protein Research*, [S.L.], v. 7, n. 6, p. 445-459, nov. 1975. Wiley. <http://dx.doi.org/10.1111/j.1399-3011.1975.tb02465.x>.

Taketomi H, Ueda Y, Gō N. Studies on protein folding, unfolding and fluctuations by computer simulation. I. The effect of specific amino acid sequence represented by specific inter-unit interactions. *Int J Pept Protein Res*. 1975;7(6):445-59. PMID: 1201909.

TEJADA, Eduardo; MINGHIM, Rosane; NONATO, Luis Gustavo. On Improved Projection Techniques to Support Visual Exploration of Multi-Dimensional Data Sets. *Information Visualization*, [S.L.], v. 2, n. 4, p. 218-231, dez. 2003. SAGE Publications. <http://dx.doi.org/10.1057/palgrave.ivs.9500054>.

UEDA, Yuzo; TAKETOMI, Hiroshi; Gō on protein folding, unfolding, and fluctuations by computer simulation. II. A. Three-dimensional lattice model of lysozyme. *Biopolymers*, [S.L.], v. 17, n. 6, p. 1531-1548, jun. 1978. Wiley. <http://dx.doi.org/10.1002/bip.1978.360170612>.

VENDRUSCOLO, M.; NAJMANOVICH, R.; DOMANY, E.. Protein Folding in Contact Map Space. *Physical Review Letters*, [S.L.], v. 82, n. 3, p. 656-659, 18 jan. 1999. American Physical Society (APS). <http://dx.doi.org/10.1103/physrevlett.82.656>.

VENDRUSCOLO, Michele; DOMANY, Eytan. Protein folding using contact maps. *Vitamins & Hormones*, [S.L.], p. 171-212, 2000. Elsevier. [http://dx.doi.org/10.1016/s0083-6729\(00\)58025-x](http://dx.doi.org/10.1016/s0083-6729(00)58025-x).

WANG, J.; OLIVEIRA, R. J.; CHU, X.; WHITFORD, P. C.; CHAHINE, J.; HAN, W.; WANG, E.; ONUCHIC, J. N.; LEITE, V. B. P.. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 109, n. 39, p. 15763-15768, 10 set. 2012. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.1212842109>.

WANG, Jin; OLIVEIRA, Ronaldo J.; CHU, Xiakun; WHITFORD, Paul C.; CHAHINE, Jorge; HAN, Wei; WANG, Erkang; ONUCHIC, José N.; LEITE, Vitor B.P.. Topography of funneled landscapes determines the thermodynamics and kinetics of protein folding. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 109, n. 39, p. 15763-15768, 10 set. 2012. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.1212842109>.

WOLYNES, P.; ONUCHIC, J.; THIRUMALAI, D.. Navigating the folding routes. *Science*, [S.L.], v. 267, n. 5204, p. 1619-1620, 17 mar. 1995. American Association for the Advancement of Science (AAAS). <http://dx.doi.org/10.1126/science.7886447>.

ZWANZIG, R.; SZABO, A.; BAGCHI, B.. Levinthal's paradox. *Proceedings Of The National Academy Of Sciences*, [S.L.], v. 89, n. 1, p. 20-22, 1 jan. 1992. *Proceedings of the National Academy of Sciences*. <http://dx.doi.org/10.1073/pnas.89.1.20>.

CHWASTYK, Mateusz; BERNAOLA, Adolfo Poma; CIEPLAK, Marek. Statistical radii associated with amino acids to determine the contact map: fixing the structure of a type i cohesin domain in the clostridium thermocellum cellulosome. *Physical Biology*, [S.L.], v. 12, n. 4, p. 046002, 27 maio 2015. IOP Publishing. <http://dx.doi.org/10.1088/1478-3975/12/4/046002>.