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

Adolfo, Larissa Maria Ferreira.
A239i 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-Cα, 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. Frustratrometer. 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(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(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(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(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(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).

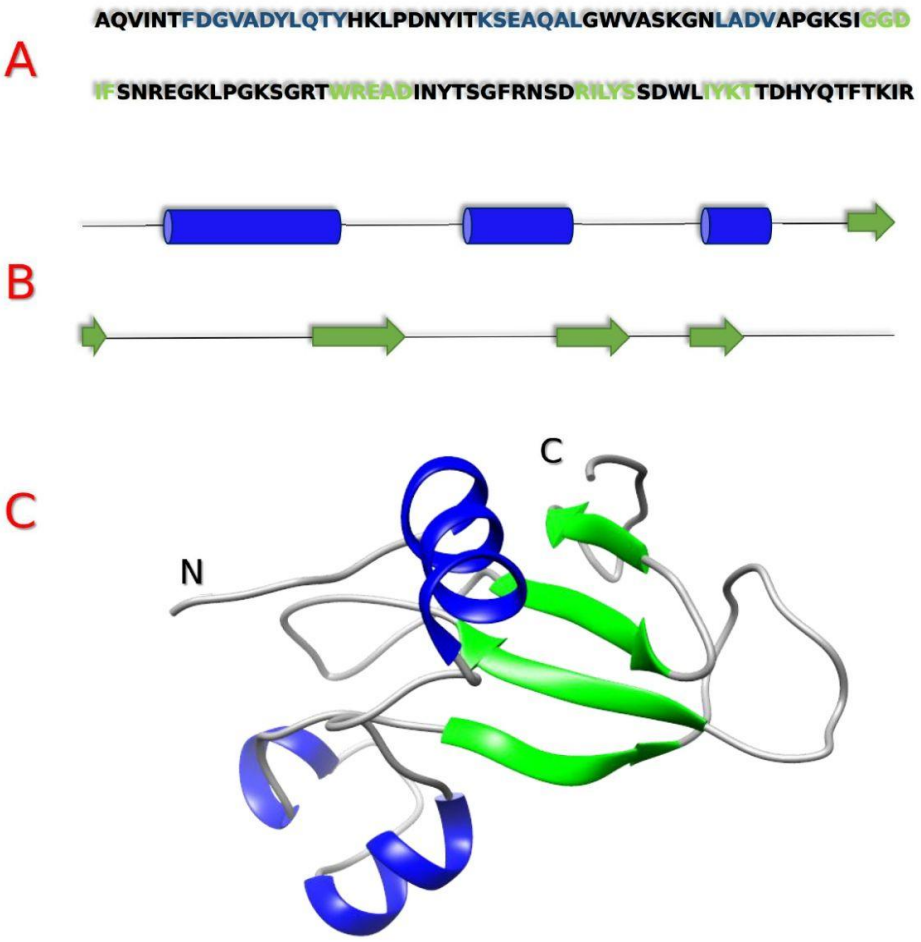
Compreender o enovelamento proteico, de modo a entender quais aminoácidos são essenciais para a formação da estrutura torna-se um importante alvo de estudo, visto que durante processos celulares, muitas vezes o mau funcionamento ou a má formação de proteínas pode acarretar falhas nos mecanismos de controle 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 Gō, 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 (ANFINSEN; 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 dá 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:

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

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.

1.2 Fundamentação Teórica

1.2.1 Modelos Computacionais e Superfície de Energia

Em 1975, Gō e colaboradores, introduziram um modelo minimalista nomeado de 'Modelo de Rede' (TAKETOMI; UEDA; Gō, 1975), que inicialmente consistia em uma rede bidimensional, de modo a simplificar a representação de uma proteína. Nesse modelo de brinquedo (toy model), os monômeros são representados por pontos dessa rede e as ligações peptídicas são representadas por unidades de comprimento da rede. Outros tipos, correspondentes a interações entre esses monômeros foram definidas como contatos que estão próximos espacialmente mas, não estão ligados covalentemente. A partir de simulações utilizando o algoritmo de Metropolis, simulações Monte Carlo, foi proposta a primeira metodologia para enovelar uma proteína com características globais. Estas simulações resultaram no enovelamento completo do estado nativo de uma proteína, ou seja, para uma única conformação a qual representa o estado funcional, sendo este um aspecto marcante desse tipo de sistema.

Ampliando a perspectiva de Gō, Wolynes e seus colaboradores formularam uma teoria baseada na física estatística de forma a elucidar uma visão global de uma superfície energética relacionada com estabilidade termodinâmica e condições mínimas de cinética para o enovelamento de proteínas (BRYNGELSON; WOLYNES, 1987). Essa teoria de enovelamento de proteínas sugere que a energia cinética deve ser alta o suficiente para vencer, sua incapacidade de satisfazer todas as interações favoráveis ou frustração local, para que a proteína seja capaz de se enovelar para sua conformação nativa (BRYNGELSON; WOLYNES, 1989).

Leopold e colaboradores demonstraram a importância das rotas de enovelamento (LEOPOLD; MONTAL; ONUCHIC, 1992), de modo que o processo se torna eficiente

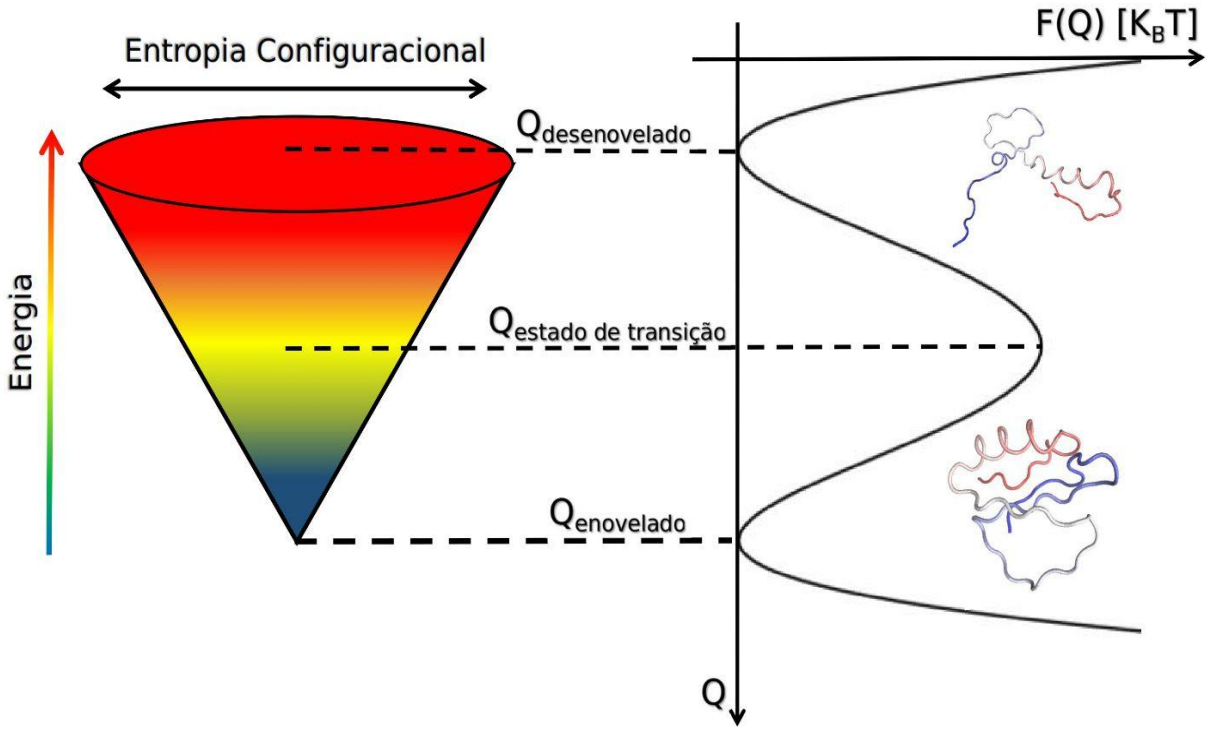
quando há vários caminhos possíveis até o estado nativo. À medida que o processo de enovelamento ocorre, o número de conformações disponíveis, ou seja a entropia que configura a cadeia de aminoácidos, diminui, de tal forma que a barreira de energia livre se torna um balanço entre energia interna e a entropia configuracional, de modo a direcionar a proteína para o seu mínimo de energia livre, assim a superfície energética do enovelamento se assemelha à forma de um funil, como demonstrado na Figura 1.2, de modo que esse processo pode ser entendido como uma reação de dois estados.

A maneira de analisar a superfície energética deste funil se fundamenta na abordagem do problema estudado. Devido à necessidade de acompanhamento do processo e identificação de mudanças conformacionais, escolhe-se uma coordenada de reação, no qual esta é a grandeza que descreve o processo ao longo da superfície de energia. Uma coordenada de reação comumente utilizada é a fração de contatos nativos Q, sendo esses contatos nativos definidos pelo mapa de contato, a qual é a razão entre os contatos nativos formados em uma conformação desenovelada (D) e os contatos nativos presentes na estrutura nativa (N). Devido ao emprego do ensemble canônico (N,V,T), a energia livre de Helmholtz é aqui utilizada. De forma que, o potencial termodinâmico dado pela energia livre F(Q) e o balanço entre a energia interna-entropia construído em função da coordenada de reação é:

$$F(Q) = E(Q) - TS(Q) \tag{1.2}$$

Podemos observar na Figura 1.2 uma relação entre a superfície afunilada e esta coordenada de reação, em que no topo do funil temos representado por uma das diversas estruturas desenoveladas possíveis, de maneira que ao longo do funil, o número de conformações possíveis vão se reduzindo até alcançar sua estrutura nativa, estado de mais baixa energia ou não-degenerado.

Figura 1.2 Representação da superfície de energia em forma de funil e energia livre associada. Representação da superfície de energia em forma de funil (esquerda), em seu topo temos as estruturas da proteína desenovelada, local de mais alta entropia configuracional e alta energia. A direita temos o gráfico de energia livre representando o balanço energia-entropia aqui descrito via coordenada de reação Q, em que $Q_{Desenovelado}$ = conformação desenovelada, $Q_{Estado\ de\ Transição}$ = conformação no estado de transição e Q_{Nativo} = conformação no estado enovelado.



Fonte: Elaborado pelo Autor

Embora o modelo computacional de Gö (Modelo de Rede), tenha sido relevante para entendermos aspectos conceituais do enovelamento de proteínas, sua geometria é altamente restrita e simplista. Então, ONUCHIC e colaboradores, propuseram o modelo Gö fora de rede (CLEMENTI; NYMEYER; ONUCHIC, 2000), mais conhecido como Modelo Baseado em Estrutura (do inglês: *Structure Based Model*, *SBM*), esse modelo permite simulações que varram extensivamente o espaço conformacional conforme modelagem e temperatura com amostragem razoável, de forma a possibilitar o cálculos de propriedades termodinâmicas utilizando fundamentos da física estatística, de forma a conseguir descrever características globais do sistema.

Neste modelo, sua função de energia é construído a partir da conformação da estrutura nativa do sistema em estudo, de forma que, representa transições do processo de enovelamento-desenovelamento, com termos que estruturam seu *backbone* e um termo que caracteriza a interação dos pares que estão em contato na estrutura nativa, e que não representam interações covalentes. Para a definição desses pares de contatos, o *SBM* necessita de um mapa que determina quais são esses pares.

1.2.2 Mapas de Contato

Resíduos que não estão ligados covalentemente na estrutura de uma proteína, mas que estão espacialmente próximos no estado nativo, e complementados por alguma definição de interação específica, são considerados em contato, e o conjunto desses contatos é chamado de mapas de contato. Este mapa é amplamente utilizado como uma representação bidimensional das interações entre resíduos, e apesar de simples, essa representação a partir das distâncias geométricas, define dois resíduos em contato, sem definir suas posições no espaço.

Um mapa de contato pode ser representado por uma matriz binária M, de tal modo que $M(i, j)$ assume o valor um se, e somente se, a interação entre os resíduos i e j da estrutura nativa esteja dentro das definições estabelecidas como contato por um algoritmo utilizado. Como definições mais simples temos o mapa definido pela distância geométrica entre dois contatos aplicando um parâmetro de corte como o mapa de *Cut-Off* (VASSOURA; MARGARA; LENA; MEDRI; FARISELLI; CASADIO, 2008). O segundo mapa aqui estudado, também utiliza-se da distância geométrica com um parâmetro de corte, porém com a introdução de um parâmetro de triagem a fim de remover redundância de informações locais, o Mapa de Contatos de Sombra (NOEL; WHITFORD; ONUCHIC, 2012). O terceiro mapa aqui em estudo utiliza-se da sobreposição das nuvens de van der waals dos resíduos de aminoácidos para definir a interação (SOBOLEV; EDELMAN, 1995). O quarto e último

mapa aqui em estudo foi proposto a partir da metodologia do Frustratrometro (FERREIRO; HEGLER; KOMIVES; WOLYNES, 2007) como uma forma alternativa na definição de interação entre dois resíduos, em que esperávamos a partir dos contatos minimamente frustrados definidos pela metodologia, tivéssemos uma condição de contatos mínimos que um mapa deveria conter, o que não pode ser concluído.

A partir disso, é possível verificar as diferentes definições de interação entre os resíduos de aminoácidos, mas em geral são baseadas em alguma distância de corte entre dois resíduos. Esses mapas têm sido usados para identificação de resíduos essenciais no processo de enovelamento (HERINGA; ARGOS, 1991) quando acompanhados do cálculo de *Phi-Values*. E Auxiliam na estabilidade (CAPRIOTTI; FARISELLI; ROSSI; CASADIO, 2008) e função proteica (SATHYAPRIYA; VIJAYABASKAR; VISHVESHWARA, 2008).

Entendendo a importância dos mapas no estudo do enovelamento de proteínas quando utilizado o modelo baseado em estrutura, faz-se necessário um estudo mais aprofundado sobre esses mapas, visto que temos três metodologias diferentes com amplo emprego na literatura e um quarto mapa aqui proposto, desta forma, com o objetivo de compará-los e entender suas influências.

Com o intuito de compreender como cada mapa define essas interações, apresentamos a seguir os quatro mapas de contato estudados neste trabalho.

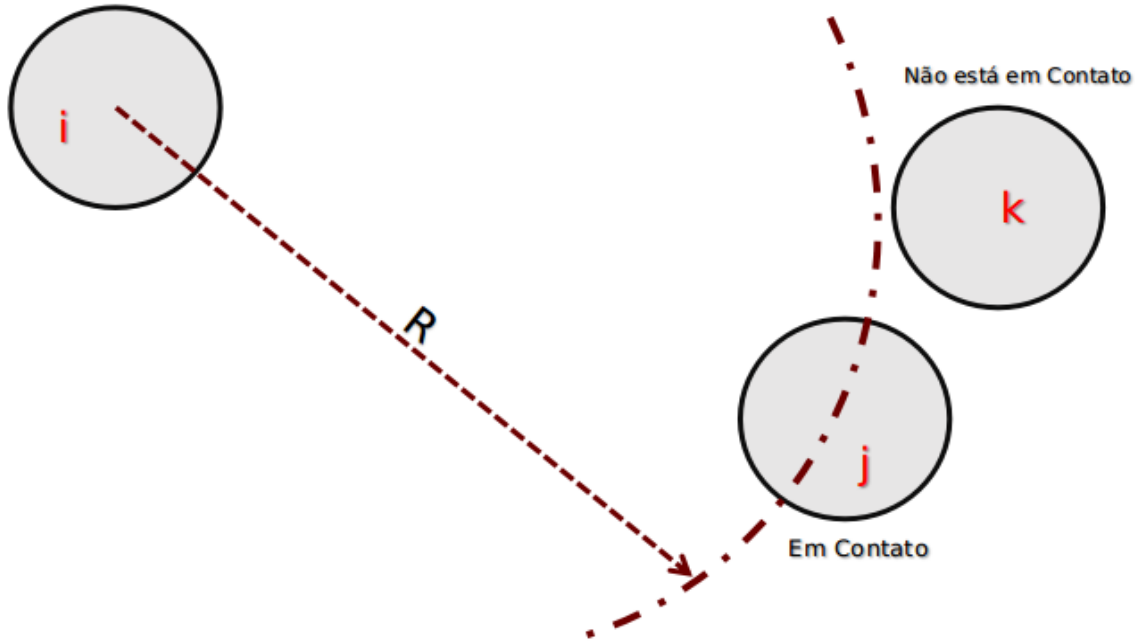
***Cut-Off* (CO)**

O mapa de *Cut-Off* tem a forma mais simples de definir o contato, utilizando de geometria analitica, ou seja, distância geométrica R entre dois resíduos ij que não estão em sequência na cadeia de aminoácidos ($j \geq i + 3$), representação esquemática na Figura 1.3. No qual todos os resíduos que estejam mais próximos que uma distância de corte (em angstrom),

são considerados em contato, ou seja, $S_{ij} = 1$. Segundo Duarte, na região do corte de 8,5 Å os contatos ainda estão na faixa das interações de Van der Waals, e para valores maiores que este, o aumento do contato sobe linearmente com o raio. Quando isso ocorre, há uma sobrecarga de contatos no sistema e consequentemente redundância de informação, desta forma escolhemos um raio de corte de 8Å (DUARTE; SATHYAPRIYA; STEHR; FILIPPIS; LAPPE, 2010). A vantagem deste mapa de contato, se dá pela simplicidade de implementação (ver apêndice A).

$$R = \sqrt{(x_j - x_i)^2 + (y_j - y_i)^2 + (z_j - z_i)^2} \quad (1.3)$$

Figura 1.3 Representação esquemática do Mapa de Contato de Corte (CO). Representação esquemática da definição de contato entre dois resíduos de aminoácidos dentro de uma distância R definida geometricamente. Neste, o contato acontece entre o resíduo i j e o contato não ocorre entre os resíduos de aminoácidos i k.



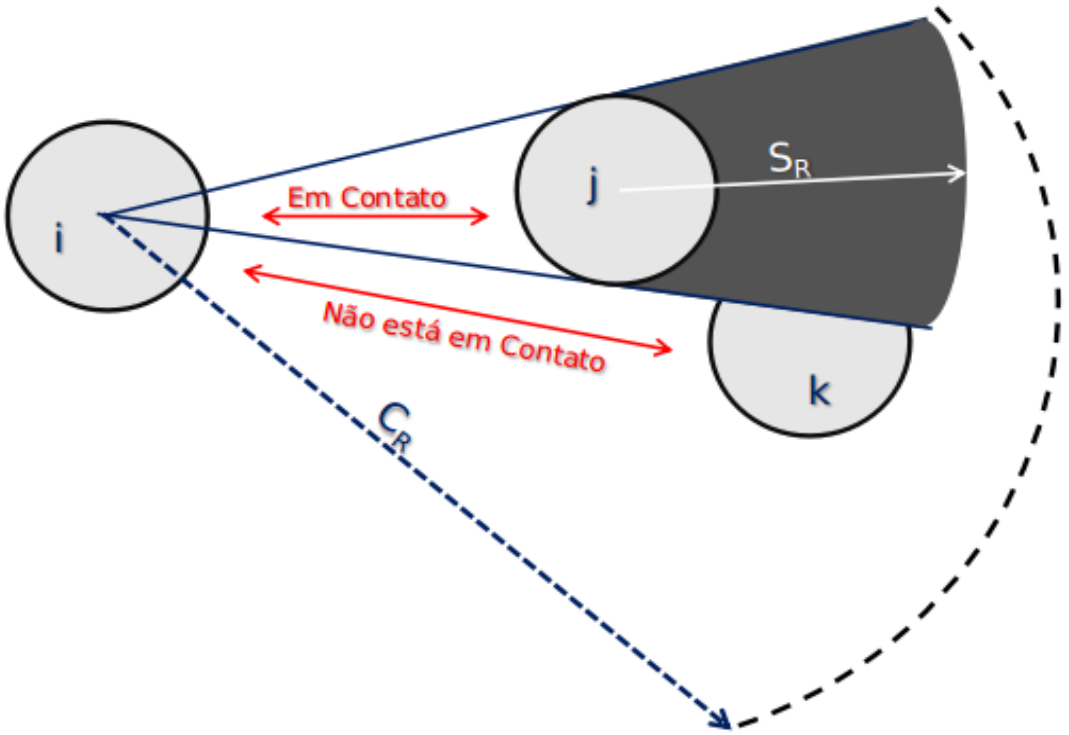
Fonte: Elaborado pelo Autor

Shadow Contacts Map (SCM)

O *Shadow Contacts Map*, utilizando a distância entre dois resíduos e/ou átomos ij não sequenciados na cadeia, com um raio de corte (C_R) pré definido, aplicando um filtro de seleção por meio de efeito de blindagem (S_R) (NOEL; WHITFORD; ONUCHIC, 2012). De

forma a ilustrar a explicação do mapa e explicar a Figura 1.4, considere todos os resíduos de aminoácidos sendo esferas opacas. Agora, considere uma fonte de luz imaginária localizada no centro do resíduo de aminoácido i, em que o resíduo de aminoácido j barra a luz e gera uma sombra de tamanho S_R, desta forma, os resíduos que são sombreados e que estão dentro da distância de corte geométrica, são descartados. Apenas os resíduos de aminoácidos que estão dentro desta distância C_R e geram sombra são considerados contatos, assim o mapa de sombra evita a sobreposição e redundância de informações, $S_R \leq 1$ Å, no qual a matriz pode ser denotada seguinte forma: M_C^S, em que S_R e o tamanho da sombra e C_R o raio de corte. O mapa de contato pode ser acessado via plataforma SMOG, < smog-server.org >.

Figura 1.4 Representação esquemática do Mapa de Contato de Sombra (SCM). Os átomos i e j estão em contato porque estão dentro de C_R e não têm átomo intermediário. Para verificar se os átomos i e k estão em contato, verifica-se se há sombreamento do átomo i no átomo k. Os três átomos são vistos no plano e todos os átomos recebem o mesmo raio de sombreamento S_R referente ao átomo i que causa uma sombra no átomo k, logo os átomos i e k não estão em contato.



Fonte: Elaborado pelo Autor

***Contact of Structural Units* (CSU)**

Com o objetivo de analisar como ocorria a interação entre um cofator com seu sítio de interação, Sobolev define uma nova forma de entender o contato entre dois resíduos de aminoácido não sequenciais, no qual não se utiliza apenas o parâmetro de distância. Inicialmente, ele define o que é superfície acessível ao solvente; essa superfície define-se como uma esfera teste que circunscreve em torno da superfície de van der waals do resíduo de aminoácido, aumentando o valor de seu raio proporcionalmente ao raio da esfera teste, em que o raio desta esfera teste foi definido com $r = 1,4$ Å. Os raios dos aminoácidos foram estudados por Chiwastyk, 2015 e colaboradores.

Considerando a superfície de um resíduo de aminoácido qualquer que denominaremos como A, se aproximar de uma esfera de raio R, em que cada ponto dessa superfície possa estar em contato com esta esfera teste, sem que seja transpassada (SOBOLEV; EDELMAN, 1995), pode ser descrita como:

$$R^{'a} = R_a + R_W \tag{1.4}$$

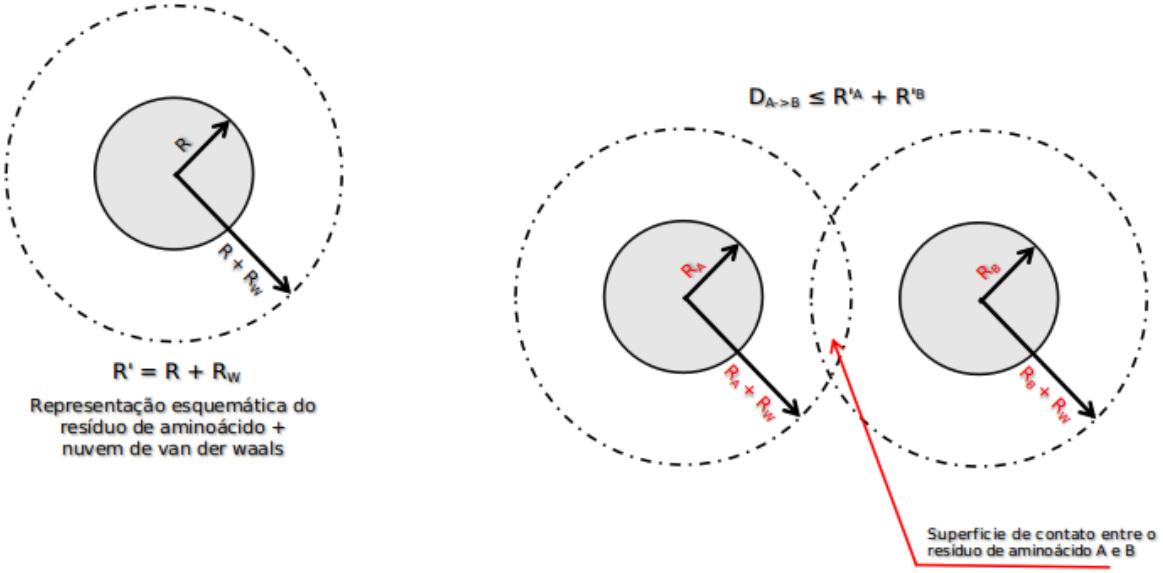
em que R_a é o raio do resíduo de aminoácido A, e R_W é o raio de van der waals deste resíduo de aminoácido A, ou seja, raio adicionado quando a esfera teste percorre a superfície do resíduo de aminoácido. Da mesma forma define-se $R^{'b}$.

Para o cálculo da superfície em contato, divide-se a superfície do átomo A em triângulos de área pequenas e compara-se com a distância de triângulos ínfimos que podem penetrar a área de um resíduos de aminoácidos de raio $R^{'b}$, se:

$$D_{a->b} \geq R_b + R_W \tag{1.5}$$

em que $D_{a->b}$ é a soma de $R^{'a}$ e $R^{'b}$. Então, considera-se como contato, Figura 1.5. Desta forma, Sobolev define como contato entre dois resíduos de aminoácidos, quando a área de superfície do resíduo A estivesse em contato com uma esfera teste e esta mesma estivesse em contato com o resíduo B. Agora, se esta esfera teste penetrar a superfície de vários resíduos ao mesmo tempo, considera-se em contato os dois átomos com centros mais próximos. Portanto, a superfície não depende apenas da distância entre dois resíduos de aminoácidos, mas também do arranjo espacial dos resíduos de aminoácidos determinados pela estrutura da proteína em estudo (SOBOLEV; EDELMAN, 1995).

Figura 1.5 Representação esquemática da determinação de contato pelo método de Contato de Unidades Estruturais (CSU). Área de Superfície acessível ao solvente, linhas tracejadas e relação de contato entre o resíduo de aminoácido A e o resíduo de aminoácido B (linhas cheias).



Fonte: Elaborado pelo Autor

Frustratrômetro (Frt)

A superfície rugosa que pode ser apresentada pela superfície energética do funil de enovelamento se forma pela competição entre as interações favoráveis e sua incapacidade de satisfazer interações locais, no qual podemos chamá-las de frustração. A frustração é um conceito útil para entender o comportamento biológico das proteínas analisando como a

energia é distribuída em sua estrutura e como estas afetam as mutações ou mudanças conformacionais. A alta frustração geralmente indica regiões biologicamente importantes envolvidas na ligação ou na alosteria. Em contraste, ligações minimamente frustradas compreendem um enovelamento estável da molécula (JENIK; PARRA; RADUSKY; TURJANSKI; WOLYNES; FERREIRO, 2012).

O frustratrômetro não é propriamente uma metodologia para construção de mapas de contato, ou seja, não foi construído para este propósito. Este faz parte de uma metodologia chamada *Associative Memory Water-mediated* (AMW) que indica o efeito de memória causado pela hidratação, permitindo caracterizar frustrações energéticas. O seu uso foi adaptado neste trabalho para as propriedades da mínima frustração. Assim, selecionamos apenas as interações de $C\alpha - C\alpha$ e com a energética minimamente frustrada.

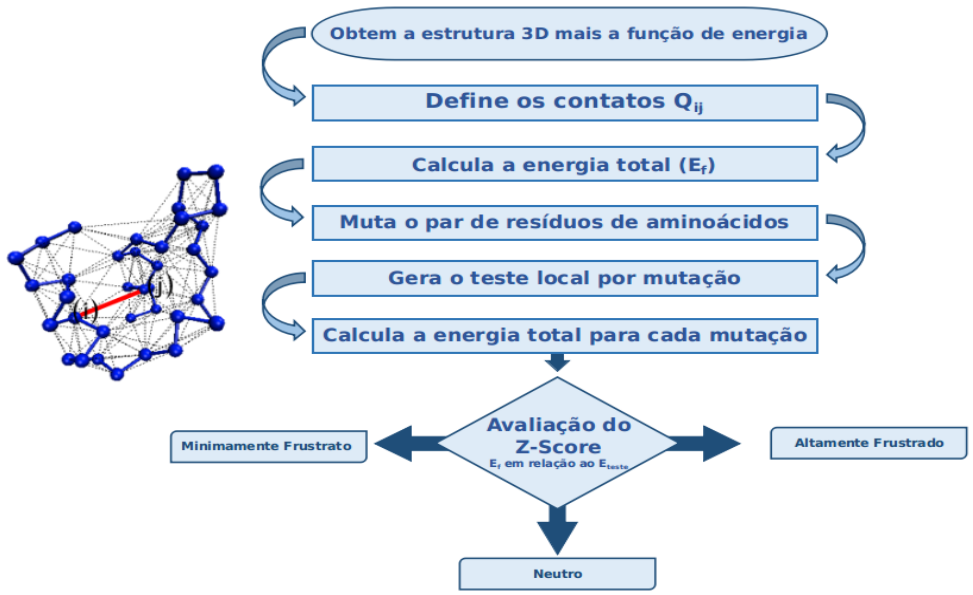
A metodologia do frustratrômetro permite localizar espacialmente e quantificar a frustração energética presente nas estruturas de proteínas nativas, desenvolvendo uma versão local da formulação do critério do princípio da frustração mínima. Essa metodologia compara a contribuição para a energia de estabilização extra atribuída a um determinado par de resíduos de aminoácidos na proteína nativa com as estatísticas das energias pelos aminoácidos que formam um determinado par de contato. Estes pares de contato são alterados, gerando um conjunto teste para os quais a energia total da proteína é recalculada. Um histograma da energia dos testes é comparado à distribuição da energia nativa inicial, H^N. O "índice de frustração" (F_{ij}) para o contato entre os aminoácidos ij é definido como um *Z-score*[1] da energia do par nativo em comparação para os N testes:

$$F_{ij} = (H_{ij}^{N} - < H_{i'j'}^{U} >) / \sqrt{\frac{1}{N}\sum_{k-1}^{n} (H_{i'j'}^{U} - < H_{i'j'}^{U} >)^2} \qquad (1.6)$$

[1] Um z-score é o número de desvios padrão em relação à média de um ponto de informação

em que $H_{i'j'}^{U}$ é a energia do conjunto teste. Este índice de frustração mede o quão favorável uma determinada interação é em relação ao conjunto de todas as interações possíveis naquele local e normalizado, usando a variância da distribuição. Se houver uma estabilização adicional suficiente para um par nativo individual, conforme normatizado pela flutuação de energia típica (de acordo com o critério de Z-score global para frustração mínima), a interação local pode ser chamada de "minimamente frustrada", Z-score de 0,78 ou maior magnitude. Se a estabilização do par nativo estiver no meio da distribuição de alternativas (Z-score menor que 0.78 e maior que -1), a interação pode ser considerada "neutra". Por outro lado, um contato é definido como "altamente" frustrado se a interação estiver na outra extremidade da distribuição com um índice de frustração local menor que -1 (FERREIRO; HEGLER; KOMIVES; WOLYNES, 2007; PAPOIAN; ULANDER; EASTWOOD; LUTHEY-SCHULTEN; WOLYNES, 2004), na Figura 1.6, temos um esquema do processo descrevido no texto.

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

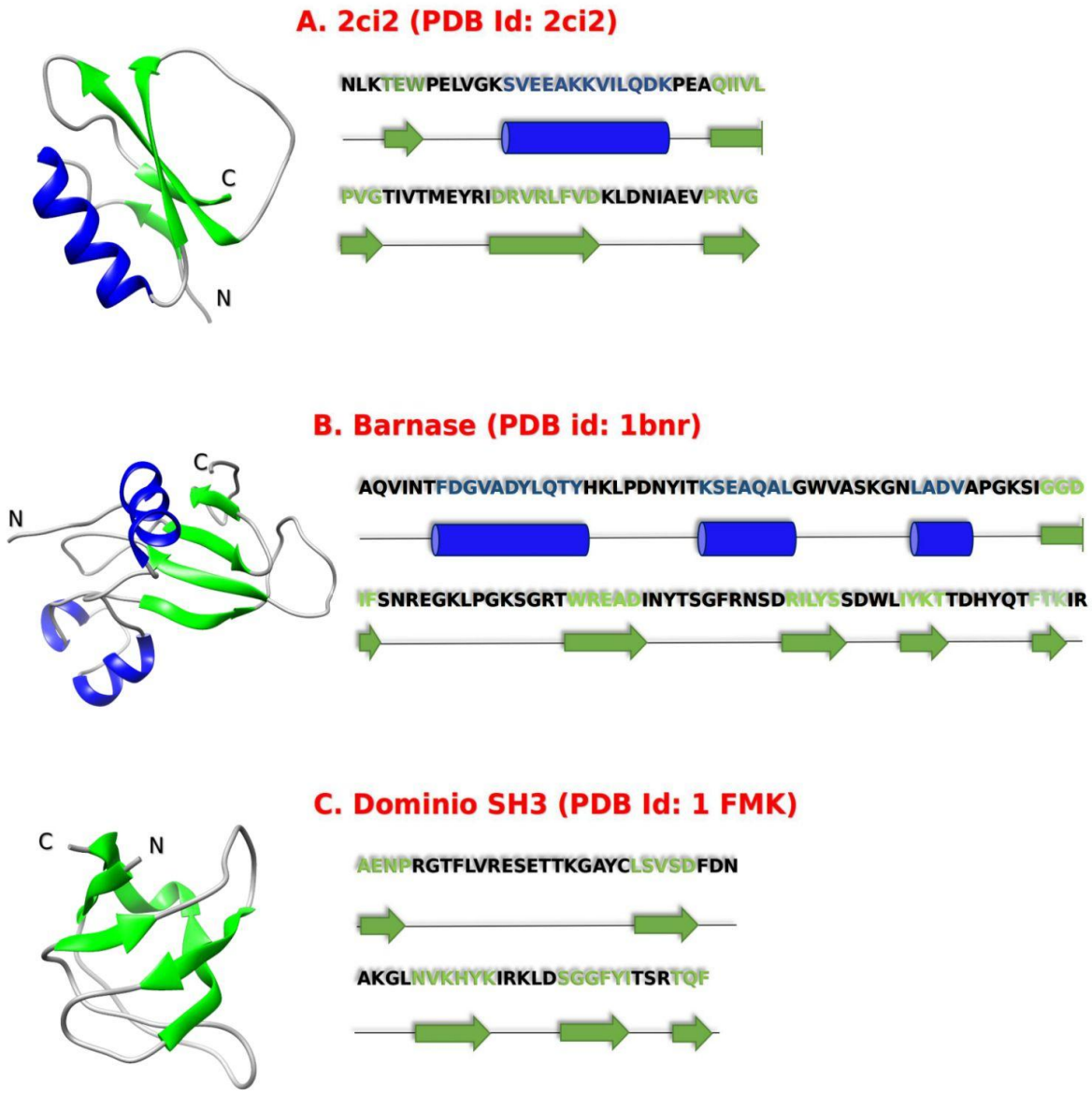


Adaptado de: frustratometer.qb.fcen.uba.ar

1.3 Proteínas Estudadas

Neste estudo, foram empregadas proteínas com diferentes características estruturais, porém amplamente empregadas na literatura e com diferentes características de processo de enovelamento. São essas proteínas a Inibidor de quimotripsina 2 (PDB id: 2ci2), Microbial Ribonuclease, Barnase (PDB id: 1bnr) e o Domínio Src Homology 3 (SH3) da Tirosina Quinase Humana (PDB id: 1 FMK).

Figura 1.7 Representação em cartoon da estrutura cristalográfica e/ou RMN das proteínas de interesse. (A) Chymotrypsin Inhibitor 2 Proteína CI2 (PDB id: 2ci2); (B) Barnase (PDB id: 1bnr) (C) Domínio SH3 (PDB id: 1FMK)



Fonte: Elaborado pelo Autor

A proteína Inibidor de quimotripsina 2 (PDB id: 2ci2), é uma enzima hidrolítica (MCPHALEN; JAMES, 1987), sua estrutura é composta por 65 resíduos de aminoácidos consistindo de quatro fitas β compactadas contra uma hélice α (Figura 1.7 (A)), caracterizando-a como uma proteína α + β, formando um núcleo hidrofóbico. Estudos experimentais definiram que seu enovelamento pode ser modelado por uma cinética de dois estados (JACKSON; ELMASRY; FERSHT, 1993), ou seja, evidenciando a presença de um estado de transição (ET), onde são formadas metade das interações nativas (JACKSON; ELMASRY; FERSHT, 1993). As análises dos valores *phi*, demonstram que as mutações não alteram significamente as estruturas do estado enovelado (JACKSON; ELMASRY; FERSHT, 1993), o que torna esta uma proteína característica para testes computacionais.

Microbial Ribonuclease, Barnase (PDB id: 1 bnr) é uma RNase monomérica de cadeia simples (Figura 1.7 (B)) com 110 resíduos de aminoácidos, sua estrutura em solução foi determinada por experimentos de Ressonância Magnética Nuclear (RMN)[2]. A estrutura da Barnase é composta por três hélices α e cinco fitas β, caracterizando-a como uma proteína α + β. Ela pode sofrer desenovelamento reversível induzido por calor, uréia ou cloreto de guanidínio, que se ajusta a um processo constituído de um estado intermediário, em que fornece um paradigma interessante para estudar o seu enovelamento (FERSHT, 1993), o que torna-a alvo de estudo para processos de enovelamento protéico e estudos computacionais.

Domínio SH3 da tirosina quinase humana é um fragmento de 57 resíduos de aminoácidos (PDB id: 1 FMK), partindo do resíduo T84 até o S140 (Figura 1.7 (C)), formando um barril de fitas-β, formando um arranjo antiparalelo dessas 5 fitas, resolvido por cristalografia de raio-X. Medidas experimentais mostraram que o domínio SH3 tem um enovelamento rápido com mecanismos de dois estados (CLEMENTI; NYMEYER;

[2] Técnica que pode ser utilizada para estudar a cinética de processos visto que a proteína é móvel na solução.

ONUCHIC, 2000). Estudos feitos por simulação mostraram que este possui um caminho de enovelamento específico, determinada por suas geometrias (BORREGUERO; DING; BULDYREV; STANLEY; DOKHOLYAN, 2004; SERRANO; MARTÍNEZ, 1999; LAMMERT; NOEL; ONUCHIC, 2012).

É possível verificar que cada uma destas proteínas tem uma característica marcante em seu processo de enovelamento, de modo que a Barnase tem um estado intermediário presente na sua curva de energia livre, marcado pela presença da primeira fita beta 1 e a hélice alfa 2 e 3, de modo que o que caracteriza este estado intermediário e sua cadeia lateral. O domínio SH3 demonstra um caminho de formação bem definido, de modo que experimentalmente tem-se a formação do *loop* que liga a β_2 as fitas β_3 e β_4 e o *harpin* que encaixa essas duas fitas. A proteína 2ci2 por seu enovelamento robusto, caracteriza-se com uma proteína teste.

Devido a esses aspectos, e a utilização de um modelo simplificado como SBM-Cα, essas proteínas torna-se alvo desse estudo para que consigamos a partir da aplicação de diferentes mapas, verificarmos suas influências.

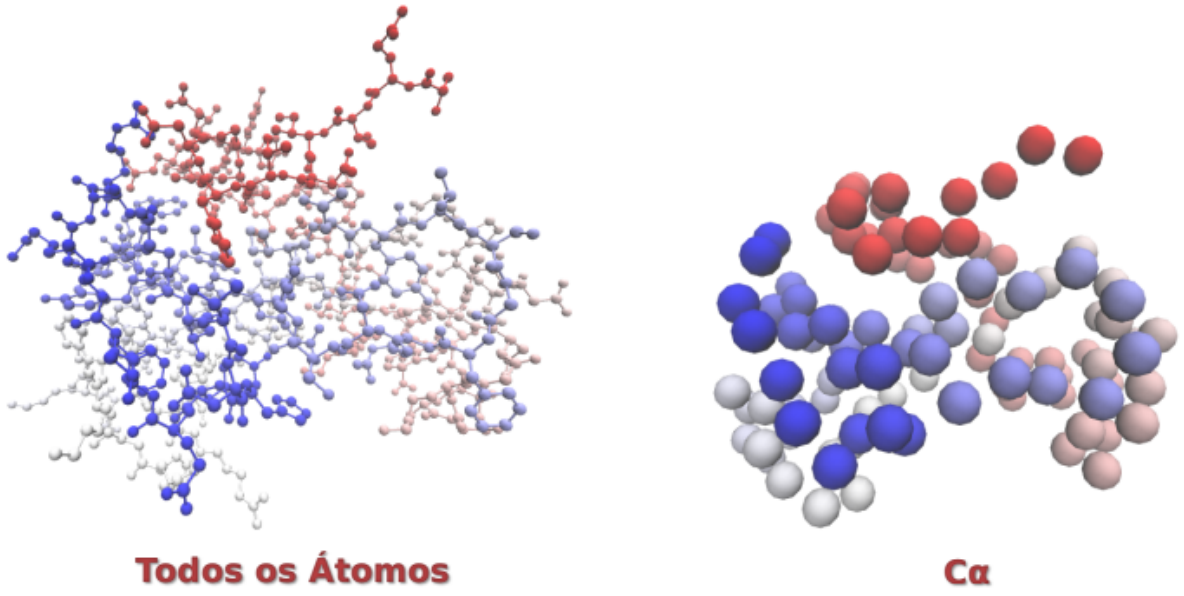
2 OBJETIVO E MOTIVAÇÃO

Uma abordagem para investigar as proteínas associadas a essas doenças envolve o uso de modelos baseados em estrutura, conhecidos como modelos Gō, que se fundamentam na teoria da superfície energética e no princípio da frustração mínima. Esses modelos demonstram sucesso na elucidação de questões fundamentais como o enovelamento de proteínas. Além disso, eles permitem simulações complexas a baixo custo computacional. No contexto do modelo baseado em estrutura, sua eficácia está fortemente relacionada a um mapa de contatos, que define as interações intrínsecas da proteína. Essas interações desempenham um papel crucial no processo de enovelamento. Logo, o objetivo principal deste trabalho é estudar o enovelamento das três proteínas apresentadas, utilizando cada um dos mapas de contato aqui em estudo, de forma a identificar e analisar a captura dos mecanismos de enovelamento e/ou mudanças conformacionais. De modo a entender, quando a escolha do mapa de contato se torna influente para entender o processo de enovelamento de cada uma das três proteínas.

3 MÉTODO

O Modelo Baseado em Estrutura (*SBM*) define-se como uma construção de potenciais em que a energia do sistema é caracterizada pelas propriedades geométricas das conformações fornecidas. Dentre as várias representações possíveis, será empregado apenas com os carbonos α (Cα). Esta representação tem como vantagem ser um modelo simplificado, com baixo custo computacional, de forma a possibilitar a captura de propriedades globais. Porém perde informação de diferenciação entre os resíduos, o que torna necessário o estudo e avaliação dos mapas de contato, a fim de que estes consigam complementar informações impostas pela limitação do modelo.

Figura 3.1 - Representação de todos os átomos e *C*α, por esperas de van der waals da proteína *Barnase.* Estrutura somente com os carbonos alfa (Cα) à direita e estrutura com todos-os-átomos à esquerda, em azul temos representado as folhas β, em vermelho temos representado as hélices α e em cinza o random coil. (PDB Id: 1 bnr).



Fonte: Elaborado Pelo Autor

As informações da proteína para implementação do potencial são encontradas no banco de dados de proteínas (do inglês: *Protein Data Bank*, *PDB)*. As estruturas presentes no *PDB* foram determinadas por técnicas experimentais[3].

[3] Exemplo de técnicas são difração de raio-x e ressonância magnética nuclear.

A energia potencial das atuais conformações (Γ) serão determinadas em relação a sua conformação nativa (Γ_0), na qual todos os parâmetros com índice zero são referente ao estado nativo, dada por:

$$V(\Gamma,\Gamma_0) = \sum_{Covalentes} \frac{\varepsilon_r}{2}(r - r_0)^2 + \sum_{Angulares} \frac{\varepsilon_\theta}{2}(\theta - \theta_0)^2$$

$$+ \sum_{Diedrais} \varepsilon_\varphi[1 - cos(\varphi - \varphi_0) + \frac{1}{2}\{1 - cos(3(\phi - \phi_0))\}] \quad (3.1)$$

$$+ \sum_{Contatos} \varepsilon_{LJ}\left[5\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - 6\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}\right] + \sum_{Não-contatos} \varepsilon_{NC}\left(\frac{4Å}{r_{ij}}\right)^{12}$$

O termo de *covalentes* refere-se às ligações entre dois $C\alpha$ consecutivos, o termo subsequente, *angulares* se refere ao potencial harmônico angular formado por três $C\alpha$ consecutivos, em que θ é o ângulo formado pelos resíduos i, $i + 1$ e $i + 2$ da conformação nativa. O termo *diedrais* modela a estrutura secundária da cadeia, este utiliza quatro $C\alpha$ consecutivos. Os dois termos subsequentes têm relação com o mapa de contato. O termo *contatos* representa a interação entre os $C\alpha$ i e j não ligados covalentemente, mas que realizam um contato na estrutura nativa. Para esta informação é atribuído um potencial 10-12, em que o termo $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12}$ descreve a repulsão e o termo $\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{10}$ descreve a atração. Este potencial é usado porque as distâncias de contatos inerentes σ_{ij} são maiores no modelo $C\alpha$. O parâmetro σ_{ij} é definido por meio do mapa de contato. Os carbonos que não realizam um contato nativo interagem por um potencial repulsivo, representado pelo último termo *não-contatos*.

As constantes ε_r, ε_θ, ε_φ e ε_{NC} são pré-fatores energéticos com valores definidos para o estudo como: 200, 40, 1 e 1 respectivamente, e esses são dados em unidade de ε_{LJ}, sendo

este a energia de interação entre um par de contatos, que apresenta o valor de aproximadamente $1\,kJ/mol$. Esses valores são fixos e foram definidos por Clementi e colaboradores, 2000.

3.1 Detalhes da simulação

Pela fácil implementação do mapa de *Cut-Off*, foi escrito um *script* (ver apêndice A) que, a partir do arquivo de *PDB* da proteína, fez o cálculo das distâncias entre os carbonos alfa tendo um *input* dos pares de resíduos que estavam em uma distância de 8Å.

O mapa de *CSU* tem um *script* próprio <info.ifpan.edu.pl/~rcsu/rcsu/index.html>, em que utiliza o arquivo *PDB,* que gera o *output* com o arquivo com os pares de resíduos e um pdb.

O mapa gerado a partir frustratrômetro, precisa que coloque o arquivo *PDB* no seu servidor web <http://frustratometer.qb.fcen.uba.ar/>. Este gera os arquivos relacionados a demonstrar quais pares de átomos são minimamente frustrados, neutros e altamente frustrados. A partir deste arquivo, pré selecionamos apenas os pares de interação entre os carbonos alfa. Em sequência, selecionamos apenas os contatos que são minimamente frustrados. E, a partir destes contatos minimamente frustrados, foi definido o arquivo com os pares de resíduos.

Para esses três mapas, *Cut-Off, CSU e Frustratrometro,* os arquivos de topologia são gerados utilizando a *web tool SMOG: Structure-based Models for Biomolecule*s <smog-server.org>. Foi utilizado o arquivo *PDB* e o arquivo dos pares de interação que foi gerado por cada mapa.

O mapa de *Shadow foi* obtido a partir da *web tool SMOG*. Para sua obtenção foi adicionado apenas o arquivo *PDB* da proteína apenas o *backbone* ou os carbonos alpha. De forma que foi expostos na plataforma e modificou-se o paramentro de distância de corte de $6\,\text{Å}$ para $8\,\text{Å}$ e manteve-se o parâmetro de triagem, $S_R \leq 1\,\text{Å}$, que foi explicado previsamente.

As simulações foram realizadas utilizando o *software GROMACS* (ABRAHAM; MURTOLA; SCHULZ; PÁLL; SMITH; HESS; LINDAHL, 2015) versão 4.5.7 com integrador estocástico. Todas as simulações realizadas tiveram a mesma quantidade de passos, ou seja, $2\,x\,10^9$ passos, com um intervalo $dt = 0.0005\,ps$, totalizando $1\,\mu s$, em que as trajetórias, velocidades e energias foram salvas a cada 1000 passos, com acoplamento térmico de Berendsen. As simulações ocorreram em um ensemble canônico (N, V, T). Para cada uma das proteínas aqui estudadas, o intervalo de temperatura inicialmente foi analisado em uma gama de temperaturas igualmente espaçadas e depois restringindo as temperaturas para o sistema de interesse.

3.2 Formas de análise

Dois resíduos de aminoácido em contatos nativos, i e j, são formados quando a distância entre eles é menor que $1,2\,d_{ij}$, com d_{ij} sendo a distância entre os dois Cα em contato definidos pelo mapa de contato, sendo que $j \geq i + 3$. O cálculo das propriedades termodinâmicas foi realizado por meio do conjunto de várias simulações realizadas sob diferentes temperaturas, obtendo, assim, a energia livre, F, em função da coordenada de reação, por meio do Método dos Múltiplos Histogramas (do inglês Weighted Histograms Analysis Method - WHAM), que será explicado a seguir.

WHAM - (Weighted Histograms Analysis Method)

Para sistemas mais complexos, não é possível varrer todo o espaço conformacional com uma única temperatura, logo foi desenvolvido o *WHAM - Weighted Histograms Analysis Method* (KUMAR; ROSENBERG; BOUZIDA; SWENDSEN; KOLLMAN, 1992), ou método de pesagem dos múltiplos histogramas. O algoritmo do *WHAM* mede essas distribuições de energia em várias temperaturas, lançando mão de métodos matemáticos com grande aplicação à física estatística a fim de alcançar a degenerescência em função de seu parâmetro de ordem $\Omega(E,Q)$. Para um cálculo deste, a função probabilidade para as temperaturas T e T' são dadas da seguinte forma:

$$P_{\beta}(E,Q) = \frac{\Omega(E,Q)e^{-\beta E}}{\sum_{E}\Omega(E,Q)e^{-\beta E}} \tag{3.2}$$

$$P_{\beta'}(E,Q) = \frac{\Omega(E,Q)e^{-\beta' E}}{\sum_{E}\Omega(E,Q)e^{-\beta' E}} \tag{3.3}$$

no qual $\beta = \frac{1}{k_B T}$, $\beta' = \frac{1}{k_B T'}$ e $\Omega(E,Q)$ é a densidade de estados obtidos através das simulações. Se fizermos a divisão da equação (3.2) pela (3.3), temos:

$$\frac{P_{\beta'}(E,Q)}{P_{\beta}(E,Q)} = \frac{\frac{\Omega(E,Q)e^{-\beta' E}}{\Sigma E\Omega(E,Q)e^{-\beta' E}}}{\frac{\Omega(E,Q)e^{-\beta E}}{\Sigma E\Omega(E,Q)e^{-\beta E}}} \Rightarrow P_{\beta'}(E,Q) = \frac{P_{\beta}(E,Q)e^{-E(\beta'-\beta)}}{Z_{\beta'}/Z_{\beta}} \tag{3.4}$$

visto que $Z_{\beta} = \sum_{E}\Omega(E,Q)e^{-\beta E}$ e $Z_{\beta'} = \sum_{E}\Omega(E,Q)e^{-\beta' E}$ são as função de partição para as temperaturas *T* e *T'* respectivamente, a equação (3.4) torna-se a estimativa a temperatura *T'* a partir de uma simulação a temperatura *T*. Agora, considerando que temos um conjunto de i

simulações, o conjunto dessa geração de vários histogramas pode ser descrito como $N_i(E,Q)$, assim podemos denotar a densidade de estados como:

$$\Omega_i(E,Q) = N_i(E,Q)e^{-\beta_i(F\beta_i - E)} \tag{3.5}$$

em que $F = -K_B T\, ln(Z_\beta)$ ou $F = -ln(Z_\beta)/\beta$ que é a energia livre de Helmholtz do sistema. A densidade de estados $\Omega(E,Q)$ é definida por uma média ponderada das densidades de estados $\Omega_i(E,Q)$ para $i = 1, 2, 3,..., R$ que é obtido nas simulações, assim podemos escrever:

$$\Omega(E,Q) = \sum_{i=1}^{R} p_i(E,Q) \cdot \Omega_i(E,Q) \tag{3.6}$$

$$\sum_{i=1}^{R} p_i(E,Q) = 1 \tag{3.7}$$

Assim, o método mostra uma autoconsistência no qual é estabelecido uma tolerância para a variação $\delta^2\Omega(E,Q)$. A probabilidade do sistema assumir $P(E,Q)$, em uma determinada temperatura T, pode ser escrita como:

$$P_\beta(E,Q) = \frac{\sum_{i=1}^{R} N_i(E,Q)e^{\beta_i E}}{\sum_{i=1}^{R} n_i(E,Q)e^{-\beta_i(F\beta_i - E)}} \tag{3.8}$$

em que n_i representa o número de conformações para a i-ésima simulação para uma dada temperatura T_i. Como o método autoconsistente, no qual os valores vão sendo atribuídos Energia Livre (F) e a partir deste obtém-se d, então o processo é repetido diversas vezes até a convergência. E com essa probabilidade definida, as grandezas termodinâmicas do sistema podem ser calculadas, como por exemplo a energia livre:

$$e^{-F_i} = \sum_E P(E, Q) \tag{3.9}$$

Coordenada de Reação

A coordenada de reação ou parâmetro de ordem, são variáveis utilizadas para acompanhar a evolução e/ou mudanças conformacionais durante o enovelamento protéico em uma simulação computacional e/ou experimental do sistema em estudo. Elas podem evidenciar suas diferentes conformações, permitindo visualizar propriedades termodinâmicas que ocorrem durante as transições. Neste estudo utilizamos três coordenadas de reação: A primeira refere-se a fração de contatos nativos (Q), definida pelo mapa de contato. Em que é a razão entre os contatos nativos formados em uma conformação corrente e o total contatos possíveis, ou seja, total de contatos presentes no mapa de contatos nativo. A segunda é o RMSD, que mede a diferença entre duas conformações durante o processo de enovelamento. E a terceira refere-se aos contatos quaisquer (Z).

A segunda e terceira coordenada de reação utilizada, serão explicadas de forma mais detalhada a seguir.

RMSD - (*Root Mean Square Deviation*)

O *Root Mean Square Deviation* ou Desvio da Raiz Quadrada Média é um método que fornece uma maneira de medir a diferença dos valores observados, neste caso entre os resíduos de aminoácido, de uma estrutura f e sua estrutura de referência i, que estão em um mesmo sistema de coordenadas, no qual pode ser expressado da seguinte forma:

$$RMSD = \sqrt[2]{\frac{1}{N}\sum_{i=1}^{n}(x_f - x_i)^2 + (y_f - y_i)^2 + (z_f - z_i)^2} \tag{3.10}$$

no qual N é o número total de resíduos do sistema. x, y e z referem-se às coordenadas desses resíduos, visto que os índices i e f relacionam-se às posições iniciais e finais de cada conformação respectivamente. Também pode ser utilizado como coordenada de reação para avaliação de mudanças conformacionais e nuances no processo de enovelamento, pois pode ser usado para comparações de estruturas.

R(Q) - Rota de enovelamento

O cálculo da rota de enovelamento permite a análise da especificidade do caminho durante o processo de enovelamento ao longo do funil de energia (PLOTKIN; ONUCHIC, 2000). Essa rota indica a especificidade de um caminho dentre todos os outros caminhos possíveis que podem levar uma proteína para o seu estado nativo. Desse modo, temos que a rota $R(Q)$ é definido:

$$R(Q) = \sum_{i=1}^{M} \frac{(<Q_i> - Q)^2}{M \cdot Q\,(1-Q)} \tag{3.11}$$

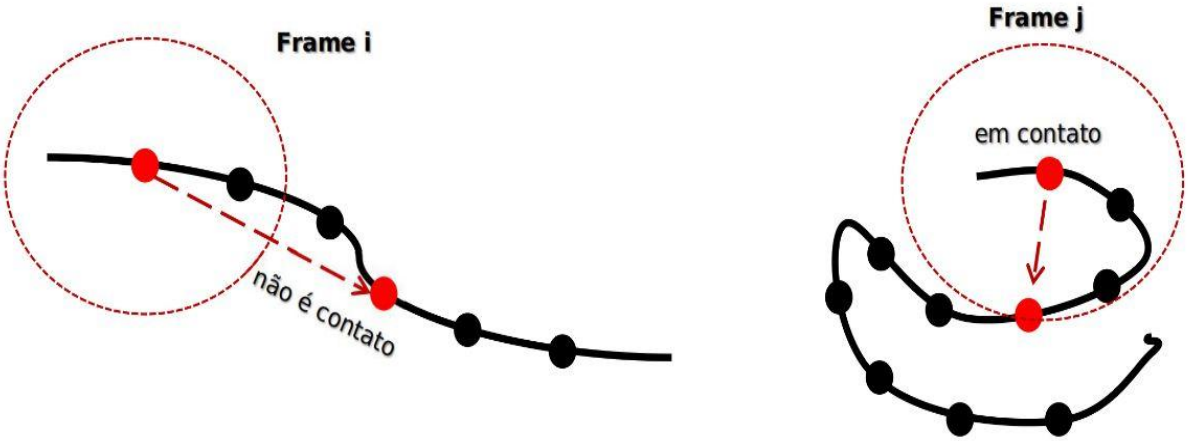
em que $< Q_i >$ específica a formação do contato i em uma conformação, sobre a média de todas as configurações de Q possíveis naquela conformação, com o total M de contatos formados. Portanto, R(Q) apresentará valores entre zero e um, no qual para valores próximos de 1, $R(Q)$ indica alta especificidade na formação de contatos nativos, porém para valores próximos de zero evidenciam a heterogeneidade entre Q_i e Q.

Z - Contatos Quaisquer

Os contatos quaisquer definiu-se como, todos os resíduos que estivessem dentro de uma certa distância de corte (R_Z) definida geometricamente (equação 1.3), dentro de cada frame da simulação considerou-se em contato, representação esquemática Figura 3.2. Assim, o *script* soma todos os contatos presentes no frame que gera um arquivo de contatos por

frame. A distância de corte (R_Z) foi definida como 8.5 Å, pois segundo Duarte (2010), essa é a distância máxima para que não ocorra sobrecarga e/ou redundancia de informação no sistema.

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



Fonte: Elaborado pelo Autor

P_{ij} - Probabilidade de formação de Contatos

A probabilidade de formação de contatos nativos, é calculada a partir do processo do enovelamento da proteína, neste caso em sua temperatura de enovelamento, por apresentar estruturas no estado desenovelado e nativo. Logo, P_{ij} é um parâmetro estatístico que pode ser calculado a partir do arquivo de trajetória e pares de contatos nativos.

Assim, inicialmente ele calcula a distância entre dois contatos possíveis dados pelo mapa de contato, em um conjunto de conformações pré definido, ou seja, a probabilidade de quais contatos estarem formados quando temos um conjunto de conformações que tem entre 10-20% dos contatos formados. Posteriormente verifica-se a distância entre os dois resíduos de aminoácidos que estão 1.2 vezes a distância entre os dois resíduos de aminoácidos determinada pelo mapa de contato em seu estado nativo.

Com isso, ele soma a probabilidade desse par de contato estar formado em todo o conjunto de conformações e divide pela quantidade de conformações.Desta forma temos um arquivo que contém resíduo o, resíduo j e a probabilidade P_{ij} de formação entre os dois, escalonada para variar entre 0 e 1, em que 0 significa que não estiveram em contato naquele conjunto de conformações e 1 que estiveram em contato em todo conjunto de conformações pré definidas.

ELViM - Energy Landscape Visualization Method

O *software* intitulado Método de Visualização de Superfície de Energia, (do inglês: Energy Landscape Visualization Method - ELViM), define uma métrica utilizando as posições dos resíduos de aminoácidos, de forma que as coordenadas internas caracterizam as diferentes conformações do sistema(OLIVEIRA; YANG; WHITFORD; LEITE, 2019). Essa medida é calculada para todos os pares de estruturas k e l de uma dada trajetória, resultando em uma matriz simétrica $n \times n$, em que n é o número total de conformações analisadas. Assim, o cálculo para a medida de comparação entre duas conformações (k, l) pode ser escrita como:

$$q^{k,l} = \frac{1}{N}\sum_{N} exp\left(-\frac{(r_{i,j}^{\ k} - r_{r,j}^{\ l})^2}{\sigma_{i,j}^{\ 2}}\right) \tag{3.12}$$

em que N remete a normalização ao número total de pares de resíduos de aminoácidos s, $r_{i,j}^{\ k}$ ou $r_{i,j}^{\ l}$ refere-se à distância entre o resíduo de aminoácido i e j da conformação k ou l, e $\sigma_{i,j} = \sigma_0 |i - j|^{\epsilon}$, em que $\sigma_0 = 1$ e $\epsilon = 0.15$. O que determina o cálculo da matriz de dissimilaridade.

Posteriormente, o método faz uma redução de ordem de modo a colocar os objetos da matriz n dimensional em posições iniciais aleatórias em um espaço 2D utilizando *Force*

Scheme(TEJADA; MINGHIM; NONATO, 2003) de modo a aplicar um passo de minimização para preservar as distâncias na matriz de dissimilaridade. O processo cessa quando a diferença nas distâncias entre as iterações consecutivas estiver abaixo de um determinado corte ou ao atingir o número máximo de iterações desejadas. Por fim, no resultado final da projeção, cada ponto corresponde a uma estrutura mostrada na dinâmica, em que os eixos não possuem significado, somente a distância entre os pontos se torna relevante, assim a proximidade dos pontos indica a semelhança entre as estruturas representadas, e pontos distantes representam estruturas com pouca relação.

Nas etapas de utilização do *software* ELViM foram realizadas com o auxílio do aluno Murilo Sanches do grupo computacional do professor Doutor Vitor Leite, logo um detalhamento maior sobre não ocorreu, pois utilizou-se como uma alternativa de visualização da superfície de energia.

4 RESULTADOS E DISCUSSÕES

Na fundamentação teórica, foram demonstradas quatro formas diferentes de obter as interações para a formação do mapa de contato. Com o intuito de avaliá-los, foi realizada suas aplicações para cada uma das proteínas propostas. Nas figuras, cada mapa recebe uma cor para representar os seus respectivos resultados, tem-se então, a cor preta para o mapa de *Cut-Off* (CO), azul para o *Shadow Contacts Map* (SCM), vermelho para o mapa de *Contacts of Structural Units* (CSU) e verde para o mapa retirado do *Frustratrômeter* (Frt). Como comumente encontramos na literatura da área, iremos demonstrar inicialmente, os resultados pela coordenada de reação Q, fração de contatos nativos.

O número de interações obtidas para cada mapa de contato e suas respectivas temperaturas de enovelamento (T_f), assim como sua energia livre, F(Q) pela coordenada de reação, Q, para as proteínas 2ci2, Barnase e para o domínio SH3 da proteína 1 FMK, estão apresentados na Tabela 4.1. Todo o tratamento estatístico do enovelamento requer o conhecimento das variáveis termodinâmicas em função da coordenada de reação escolhida, sendo esta definida pelo mapa de contato, de forma a verificar sua influência no processo de enovelamento, pois Q varia tanto em quantidade, verificar Tabela 4.1, quanto em quais resíduos de aminoácidos são considerados em contato para cada mapa aqui em estudo.

Desta forma, podemos a partir dos arquivos de trajetória e energia das simulações a temperatura constante, utilizar o método dos múltiplos histogramas (WHAM), para demonstrar o perfil de energia livre, pela coordenada de reação Q, dado por cada mapa, para cada proteínas aqui apresentadas. Na Figura 4.1, 4.2, 4.3 e 4.4 temos em A, C e E, F(Q) por Q, e em em B, D e F temos R(Q) por Q, para as proteínas 2ci2, Barnase e para o domínio SH3, respectivamente.

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)

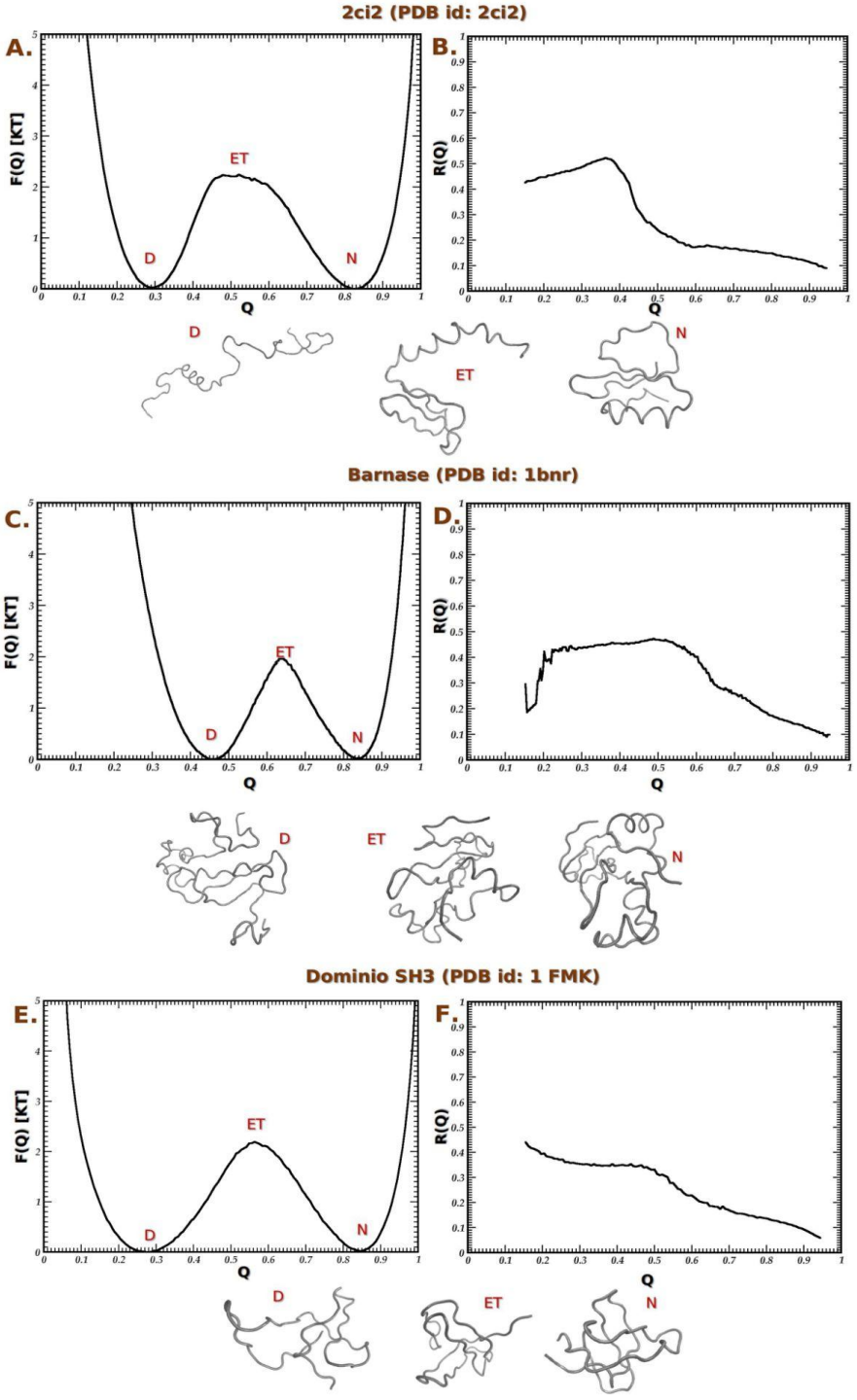
Chymotrypsin Inhibitor 2 Ci2 (PDB id: 2ci2)			
Mapa de Contato	**Quantidade de Contatos (Q)**	**Temperatura de Enovelamento (T_f)**	**F(Q) [KT]**
Cut-Off (CO)	146	122	2,3
Shadow Contacts Map (SCM)	124	117	1,8
Contacts Structures Units (CSU)	136	121	3,1
Frustratrometer (Frt)	95	92	1,5
Barnase (PDB id: 1 bnr)			
Mapa de Contato	**Quantidade de Contatos (Q)**	**Temperatura de Enovelamento (T_f)**	**F(Q) [KT]**
Cut-Off (CO)	287	123	2,0
Shadow Contacts Map (SCM)	241	116	1,7
Contacts Structures Units (CSU)	277	132	4,2
Frustratrometer (Frt)	167	92	1,1
Domínio SH3 da tirosina quinase humana (PDB id: 1 FMK)			
Mapa de Contato	**Quantidade de Contatos (Q)**	**Temperatura de Enovelamento (T_f)**	**F(Q) [KT]**
Cut-Off (CO)	142	125	2,2
Shadow Contacts Map (SCM)	130	123	2,0
Contacts Structures Units (CSU)	140	132	3,7
Frustratrometer (Frt)	86	94	0,7

Fonte: Elaborado pelo Autor

As interações entre os resíduos de aminoácidos determinam como a perda de entropía configuracional, influenciando na altura da barreira de energia livre do sistema. Desse modo,

podemos verificar que para cada mapa, a diferença na fração de Q, varia a altura da barreira de energia livre, de forma a influenciar na estabilização do estado nativo para cada mapa.

Figura 4.1 Gráfico de Energia Livre, F(Q) e Rota, R(Q), pela coordenada de reação, Q, para o mapa de Cut-Off (CO). A, C e E têm F(Q) por Q. Em B, D e F temos R(Q) por Q para as proteínas 2ci2 (PDB id: 2ci2), Barnase (PDB id: 1 bnr) e para o domínio SH3 (PDB id: 1 FMK).



Fonte: Elaborado pelo Autor

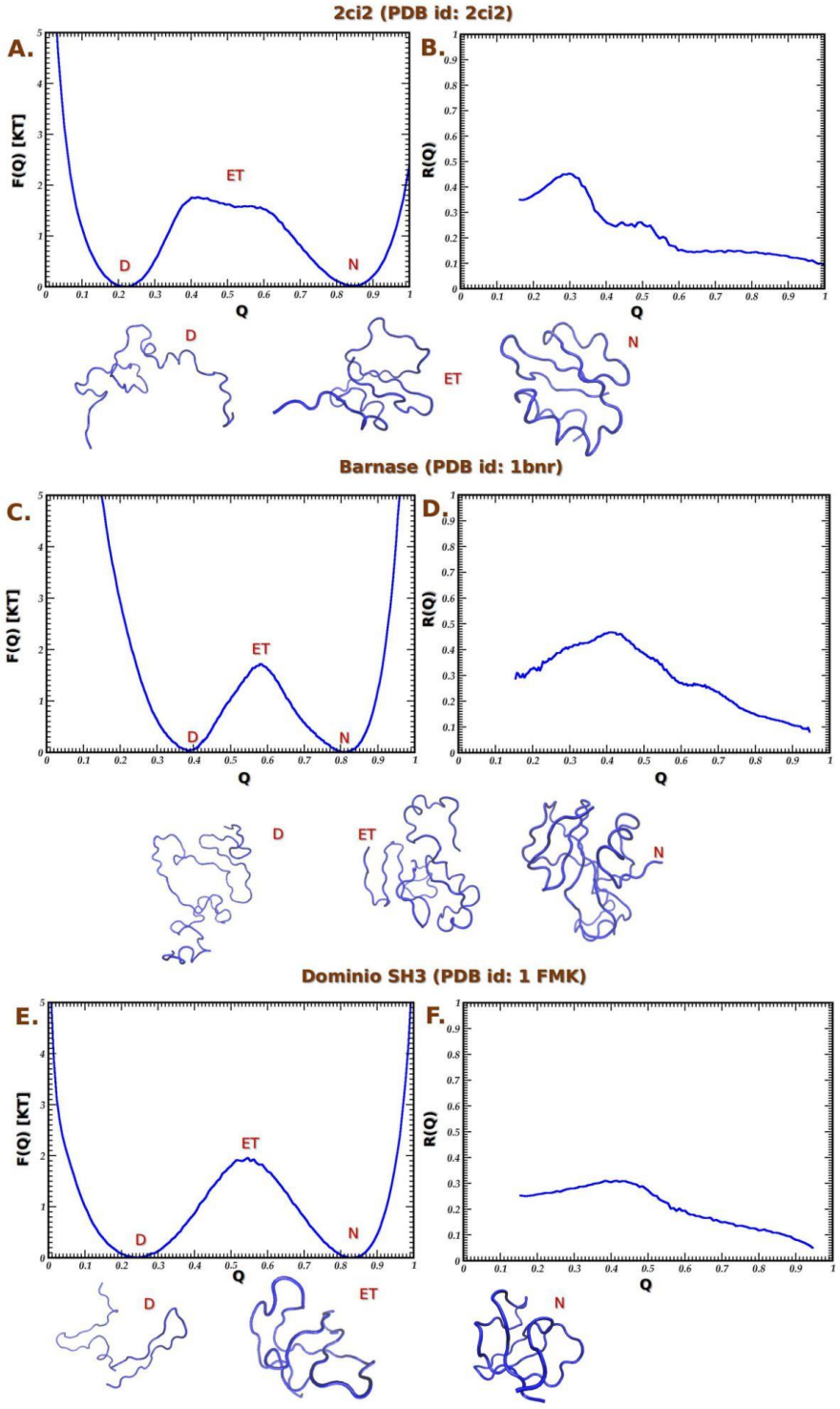
Entendendo que o sistema é modelado por uma reação de dois estados, desenovelado (D) e nativo (N) e que o estado de transição (ET) ocorre para um conjunto de temperaturas em torno de T_f. Temos na Figura 4.1 as curvas de energia livre pela coordenada de reação Q e a rota de enovelamento para as três proteínas estudadas para o mapa de Cut-Off (CO). Para este mapa, temos a fração de contatos em D mínima de 0.3 e uma fração acima de 0.8 para o estado N. Para as três proteínas o mapa demonstra uma reação de dois estados.

Para a proteína 2ci2, este demonstra em B, um aumento de especificidade de rota até atingir a barreira de energia livre (Figura 4.1 A e B). Para a proteína Barnase, em que experimentalmente apresenta um estado intermediário (EI) demonstrado por Matouschek, Serrano e Fersht (1992) em sua curva de energia livre (Figura 4.1 C e D), este mapa não consegue demonstrá-lo, de forma que sua fração de Q aumenta para 0.5, demonstrando uma alta fração de Q formados para que ocorra o processo de enovelamento, o que podemos verificar em sua rota de enovelamento uma especificidade constante até 0.5, de forma a concordar com a curva de energia livre. Para o domínio SH3, o mapa demonstra um declínio constante da especificidade da rota (Figura 4.1 E e F) tendo um leve aumento no início do estado de transição, indicando a necessidade de formação de contatos específica para que ocorra o ET.

Na Figura 4.2, temos a energia livre pela coordenada de reação Q e a rota de enovelamento para as três proteínas, com aplicação do SCM, temos fração de Q acima de 0.2 para o estado D, e 0.8 para o estado N, de forma que o mapa consegue demonstrar uma reação de dois estados para todas as proteínas aqui estudadas. De forma que para a proteína 2ci2 (Figura 4.2 A e B), temos um aumento na especificidade de rota até que ocorra o estado de transição. Para a Barnase (Figura 4.3 C e D), o mapa de SCM não demonstra o ET, de forma que tem a fração de Q em 0.4 em rota de enovelamento crescente até seu estado de transição.

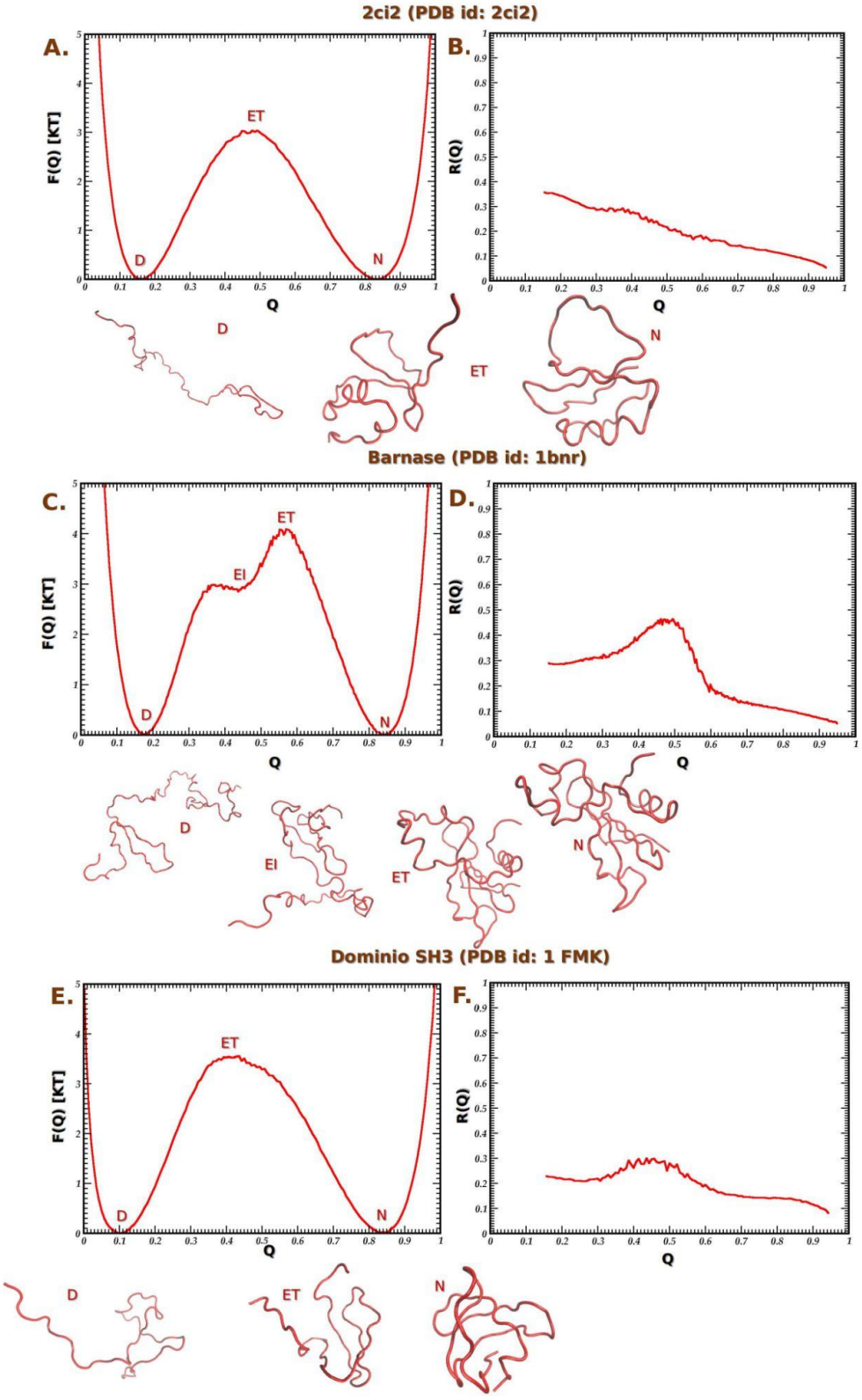
Para o domínio SH3 (Figura 4.3 E e F), há um aumento de rota pouco significativo em seu ET.

Figura 4.2 Gráfico de Energia Livre, F(Q) e Rota, R(Q), pela coordenada de reação, Q, para o mapa de SCM. A, C e F têm F(Q) por Q. Em B, D e F temos R(Q) por Q, para as proteínas 2ci2 (PDB id: 2ci2), Barnase (PDB id: 1 bnr) e para o domínio SH3 (PDB id: 1 FMK).



Fonte: Elaborado pelo Autor

Figura 4.3 Gráfico de Energia Livre, F(Q) e Rota, R(Q), pela coordenada de reação, Q, para o mapa de CSU. A, C e E têm F(Q) por Q. Em B, D e F temos R(Q) por Q, para as proteínas 2ci2 (PDB id: 2ci2), Barnase (PDB id: 1 bnr) e para o domínio SH3 (PDB id: 1 FMK).



Fonte: Elaborado pelo Autor

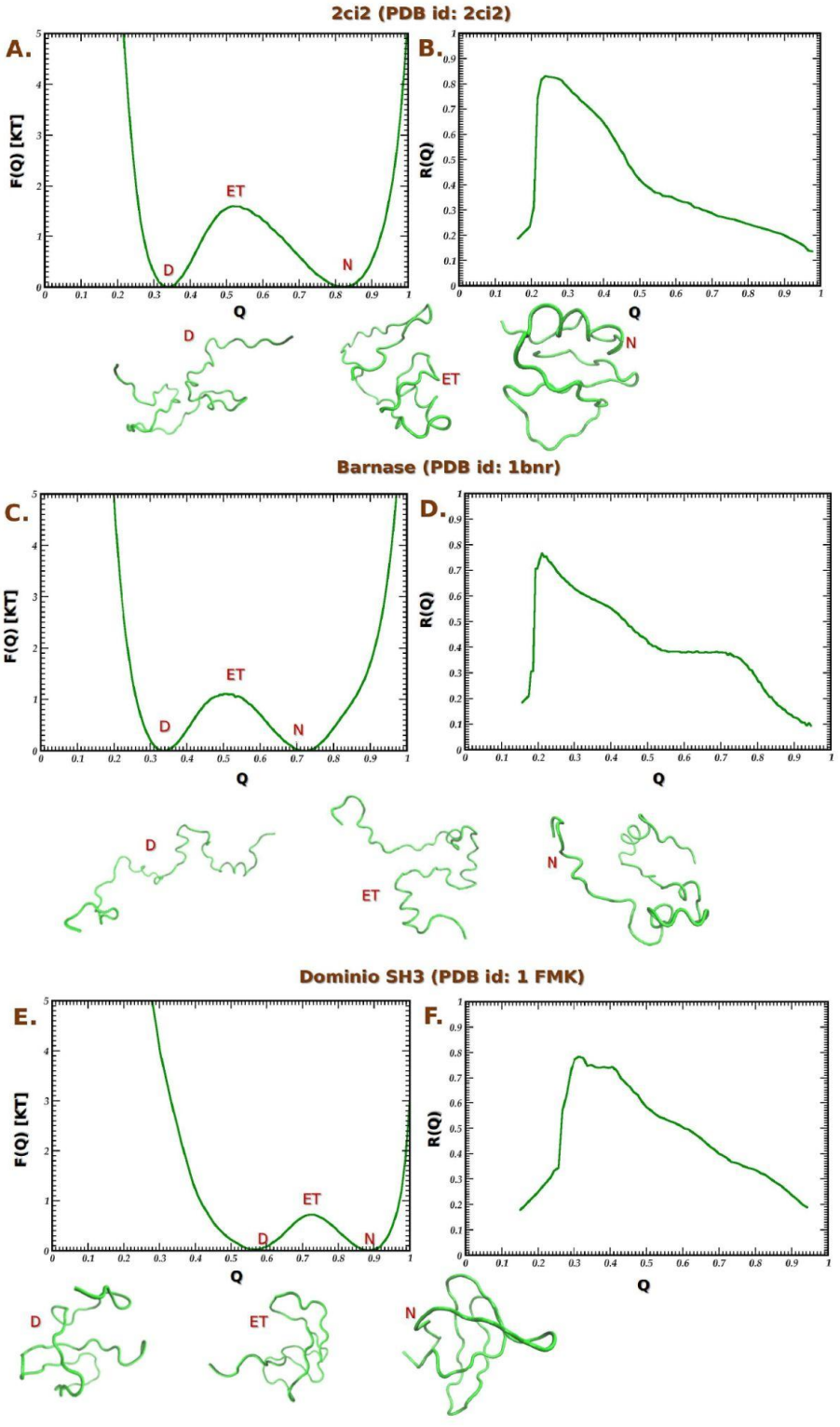
Na Figura 4.3, temos as curvas de energia livre pela coordenada de reação Q e rota de enovelamento para o mapa de CSU, em que para as três proteínas a reação de dois estados consegue ser demonstrada, e sua fração de Q e demonstra em ~ 0.1 no estado D e maior que 0.8 para o estado N. Para a proteína 2ci2, temos (Figura 4.3 A e B), temos um declínio constante da rota de enovelamento, demonstrando pouca especificidade de rota. Para a proteína Barnase (Figura 4.3 C e D), o mapa de CSU apresenta um estado intermediário para seu processo de enovelamento, o que demonstra um pico de especificidade de rota na mesma fração de Q ~ 0.5, de forma que a demonstração deste estado intermediário concorde com resultados experimentais. Para o domínio SH3, (Figura 4.3 E e F), temos uma maior especificidade de rota em seu estado de transição, dado que esta proteína tem um caminho de formação bem definido.

Na Figura 4.4, temos a curva de energia livre e rota de enovelamento para o mapa de contato construído a partir do Frustratrômetro (Frt), em que para as três proteínas a reação de dois estados consegue ser demonstrada. É possível verificar que tem-se a fração de Q ~ 0.3 ou acima para o estado D, e ~ 0.7 ou acima para o estado N, para todas as proteínas a rota de enovelamento é altamente específica, o que deve ser acometido pela baixa quantidade de contatos. Também é possível verificar que este não demonstra o estado intermediário que a proteína Barnase apresenta (Figura 4.4 C), assim como uma alta quantidade de contatos formados para que ocorra o processo de enovelamento do domínio SH3.

A partir das Figuras 4.1, 4.2, 4.3 e 4.4, é possível verificar as diferenças nas curvas de energia livre pela coordenada de reação Q. Além de seguir a modelagem determinada pelo potencial 10-12 capaz de demonstrar a reação de dois estados, desenovelado e nativo para cada proteína. O que varia é como o mapa de contato compreende os dois estados, e como a distribuição desses contatos varia na altura da barreira de energia livre, e como essa

distribuição afeta na especificidade na rota de enovelamento, em especial, no estado de transição das proteínas.

Figura 4.4 Gráfico de Energia Livre, F(Q) e Rota, R(Q), pela coordenada de reação, Q, para o mapa de Frt. A, C e E têm F(Q) por Q. Em B, D e F temos R(Q) por Q, para as proteínas 2ci2 (PDB id: 2ci2), Barnase (PDB id: 1 bnr) e para o domínio SH3 (PDB id: 1 FMK).



Fonte: Elaborado pelo Autor

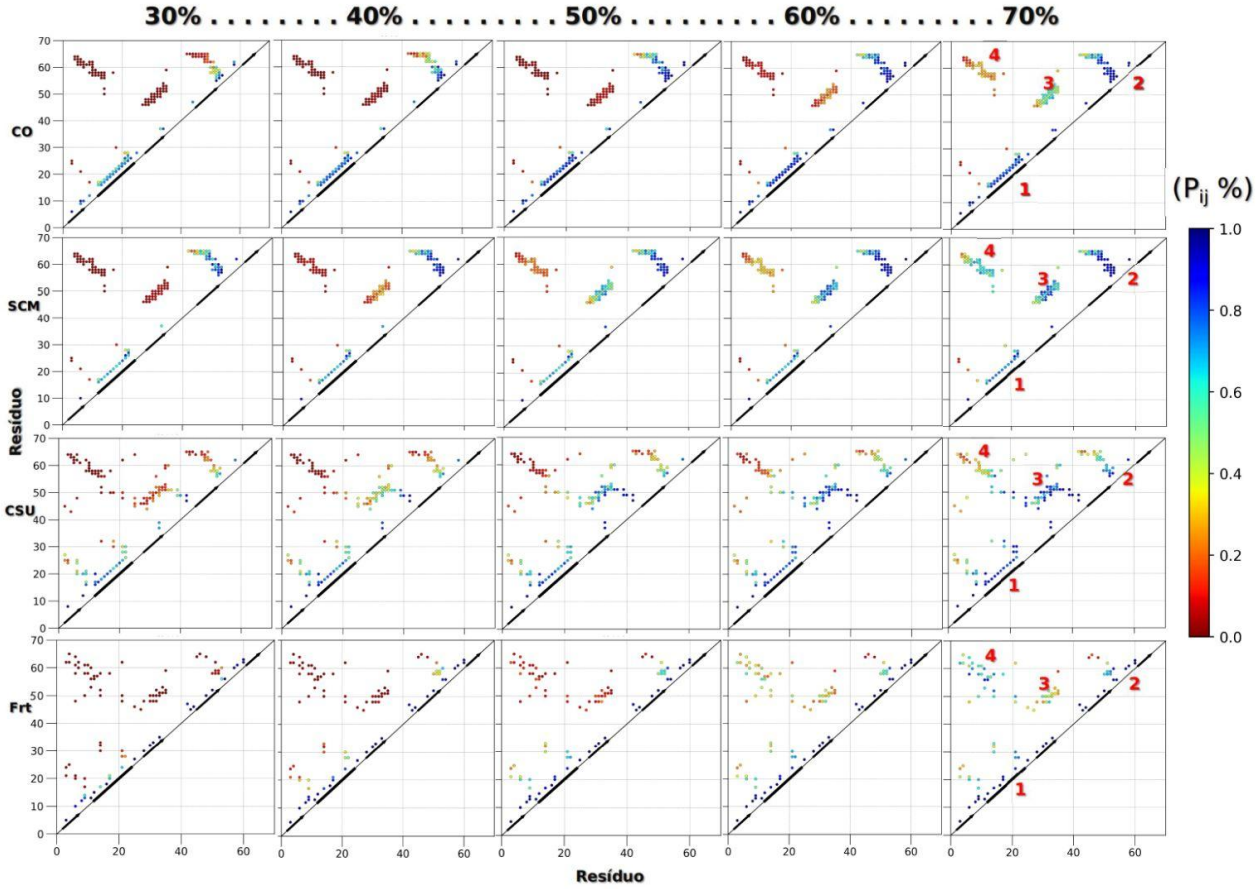
Podemos investigar como esse processo de enovelamento ocorre, a partir da probabilidade de formação de contatos (P_{ij}) no estado de transição para cada uma das proteínas em cada mapa proposto. A fim de verificar como é manifestada esse estado de transição para cada uma das proteínas, esse resultado nos dá a probabilidade de formação dos contatos, como veremos nas Figuras 4.5, 4.6 e 4.7. Esse estado de transição foi definido como a variação da fração de contatos nativos, Q, de forma a demonstrar a transição entre os estados desenovelado ao nativo. Em que as figuras demonstram um conjunto de conformações com uma dada fração de contatos formados, no qual demonstra quais resíduos de aminoácidos e quais estruturas foram formados no respectivo conjunto de conformações.

Essas três figuras, demonstram o caminho de formação durante o processo de enovelamento, assim, os números demonstrados no último quadro (70%+ ou Nativo), mostra qual a ordem de formação das estruturas conforme o aumento de Q, o que podemos comparar com a ordem de formação descrita experimentalmente.

Quanto a P_{ij}, este tem uma escala de cor que varia de vinho a azul marinho, em que vinho demonstra contatos não formados a azul marinho que remete a contatos formados.

Na Figura 4.5 temos a probabilidade de formação de contatos para a proteína 2ci2, em que podemos verificar a probabilidade do caminho de formação dos contatos que ocorre durante o estado de transição. Para esta proteína, é possível verificar que independente do mapa de contato, o caminho de formação não se altera, o que condiz com sua formação característica, de tal modo que para este caso o modelo baseado em estrutura, demonstra ser robusto, pois o mapa de contato não influencia no caminho de formação.

Figura 4.5 Probabilidade de formação de contatos para a proteína 2ci2 (PDB id: 2ci2). De cima para baixo temos a probabilidade de formação do mapa de *Cut-Off* (CO), seguido pelo *Shadow Contacts Map* (SCM), o mapa de *CSU* e o Frustratrômetro (Frt). Temos as probabilidades de 40%, 50% e 60% dos contatos formados, contemplando o estado de transição. As setas indicam o caminho de formação das estruturas. Contatos em vinho, indicam a não formação, contatos indicados em azul, indicam que aqueles contatos estão formados.



Fonte: Elaborado pelo Autor

Na Figura 4.6 temos a probabilidade do caminho de formação dos contatos para a proteína Barnase (PDB id: 1 bnr). No artigo de Matouschek, Serrano e Fersht (1992), há uma descrição do processo de enovelamento da proteína Barnase no qual indica, que a região da folha β é altamente estruturada no estado intermediário, pois é a região do núcleo 3 que consiste no empacotamento do loop 3, que une as fitas β_1 e β_2, e do loop 5, que une as fitas β_4 e β_5, que ligam-se de forma antiparalela. Além disso, o núcleo 2 ou central, formada pelos

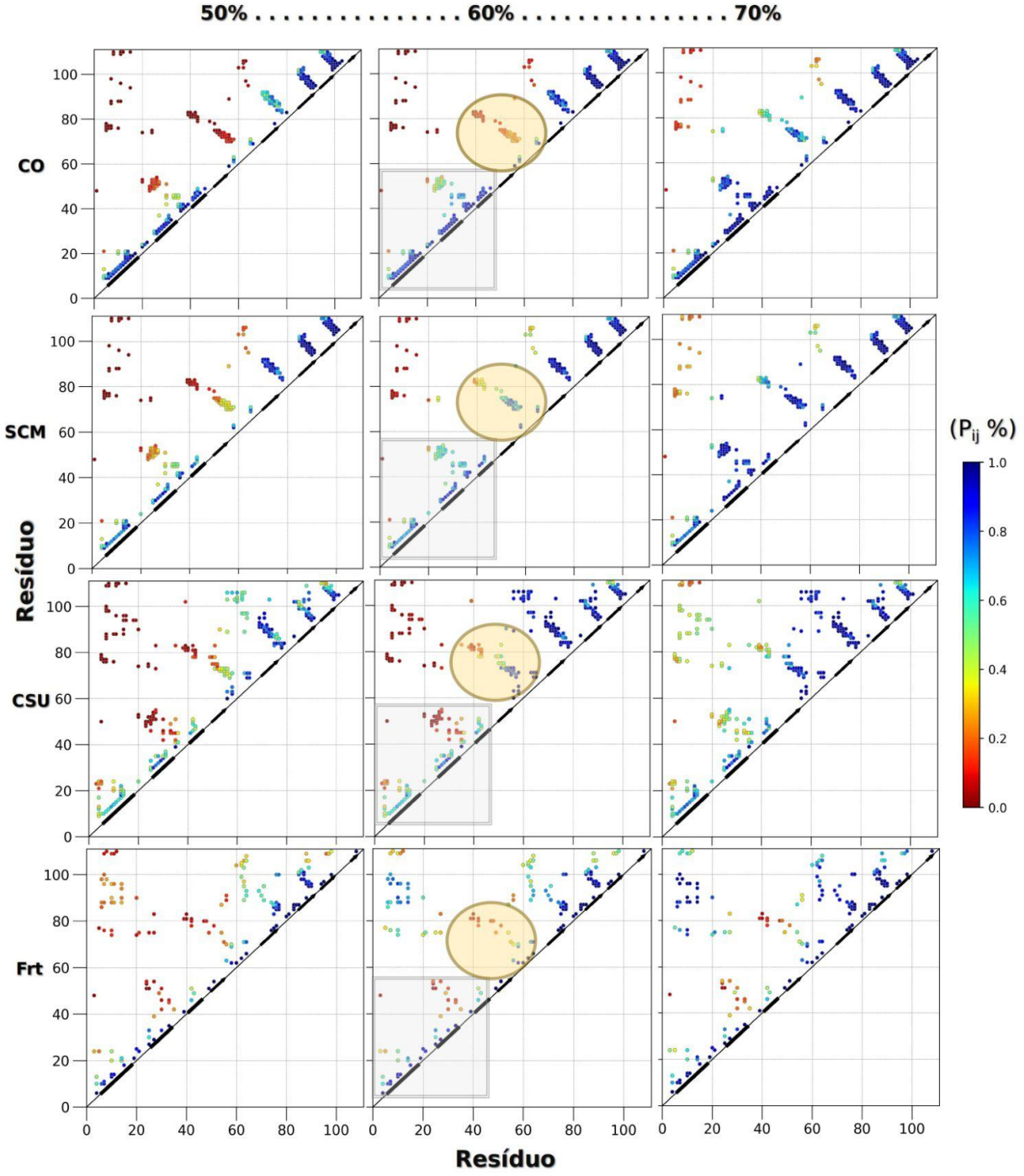
resíduos hidrofóbicos da hélice 2, hélice 3, a fita β_1 e os dois primeiros *loops*, são apenas fracamente formados no estado intermediário e de transição.

De acordo com o artigo, a região que primeiro se forma é a da folha β, que foi bem capturada e tem concordância por todos os mapas. Porém a formação da fita β_1 e a formação parcial das hélices é realizada no estado de transição. Estas regiões foram marcadas por um quadro cinza nas hélices e um círculo ocre na fita β_1. De forma que os mapas de *CO* e *Frt* apresentam uma formação estruturadas das hélices no estado de transição, porém sem a formação da fita β_1. O *SCM*, apresenta uma formação quase que completa da região das hélices e da fita β_1, enquanto que o mapa de *CSU* apresenta uma formação parcial das hélices e da fita β_1.

Era de se esperar a formação precoce da α-hélice devido a fatores energéticos. Porém, a formação demonstra a estruturação da folha β muito bem estabelecida, pois o núcleo e/ou região central, estão altamente relacionados à formação do estado intermediário. De forma que o mapa de CSU foi o que mais se aproximou com os resultados experimentais de Fersht e colaboradores (1992).

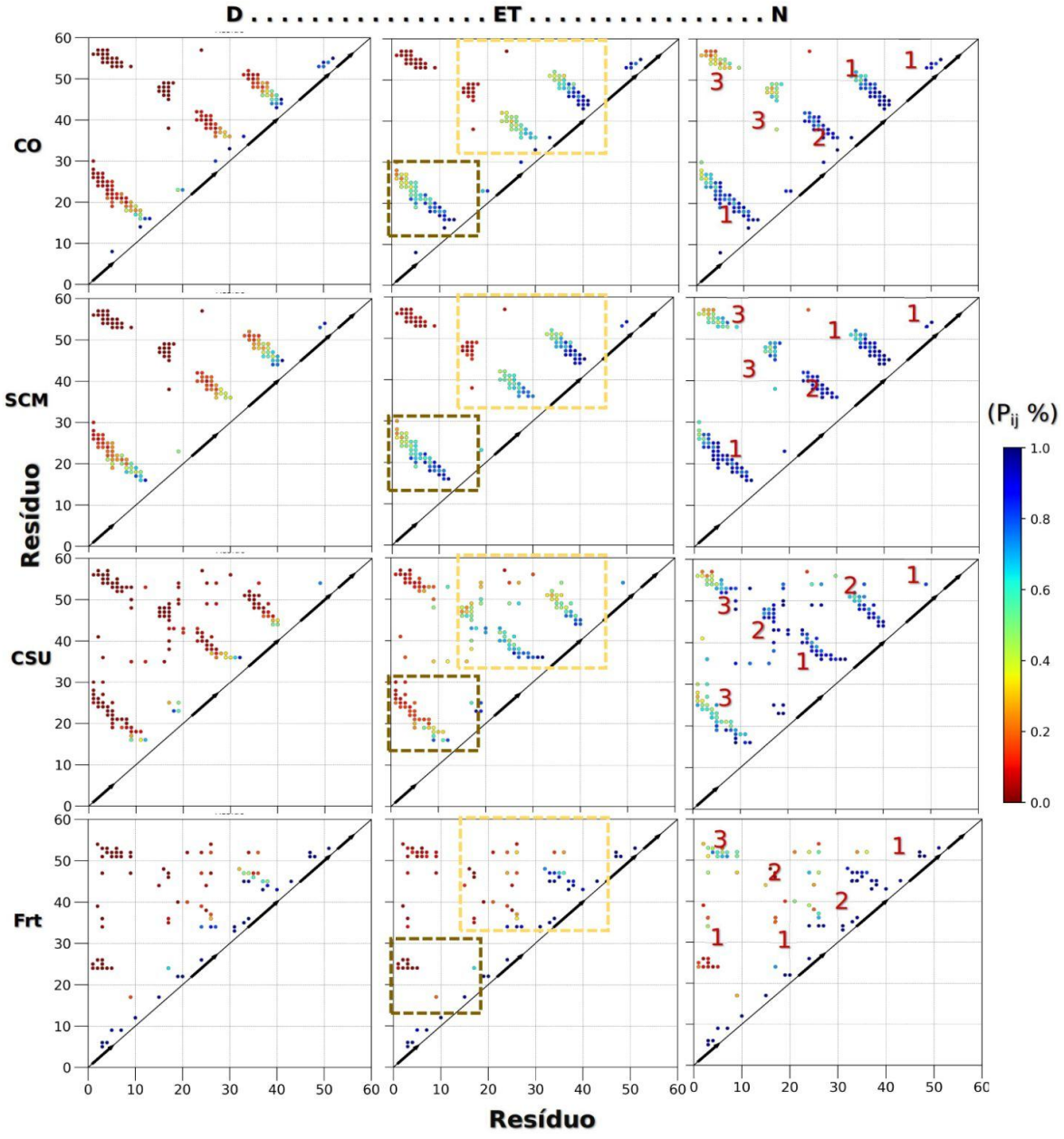
Na Figura 4.7 temos a probabilidade de formação dos contatos durante o processo de enovelamento do domínio SH3 da proteína 1 FMK (PDB id: 1 FMK). Nesse, mostrou-se apenas o estado nativo (N), estado de transição (ET) o estado desenovelado (D), visto que a curva de energia livre (Figura 4.1 C), mostra uma discrepância nos três estados.

Figura 4.6. Probabilidade de formação de contatos para a proteína Barnase (PDB id: 1BNR). De cima para baixo temos a probabilidade de formação do mapa de *Cut-Off* (CO), seguido pelo *Shadow Contacts Map* (SCM), o mapa de *CSU* e o Frustratrômetro (Frt). Temos as probabilidades de 50%, 60% e 70% dos contatos formados, contemplando o estado de transição. Até $P_{ij} = 50\%$, o caminho de formação é igual para todos os mapas. Contatos em vinho, indicam a não formação, contatos indicados em azul, indicam que aqueles contatos estão formados.



Fonte: Elaborado pelo Autor

Figura 4.7. Probabilidade de formação de contatos para o domínio SH3 (PDB id: 1 FMK). De cima para baixo temos a probabilidade de formação do mapa de *Cut-Off* (CO), seguido pelo *Shadow Contacts Map* (SCM), o mapa de *CSU* e o Frustratrômetro (Frt). Temos as probabilidades do estado desenovelado (D), estado de transição (ET) e estado nativo (N). Os números indicam a ordem de formação das estruturas. Contatos em vinho, indicam a não formação, contatos indicados em azul, indicam que aqueles contatos estão formados.



Fonte: Elaborado pelo Autor

Nos estudos de Grantcharova, Riddle e Baker (2000), foi demonstrado que o β-*hairpin* distal se destaca como o elemento estrutural melhor formado no estado de transição. A descoberta de que a reticulação[4] distal do β-*hairpin* aumenta a taxa de enovelamento e não

[4] Processo que ocorre quando cadeias poliméricas lineares ou ramificadas são interligadas por ligações covalentes

tem efeito sobre a taxa de desenovelamento, o que sugere que as duas fitas β ficam muito próximas na base do *hairpin* de forma a se estruturar e encaixar no estado de transição.

Essa região de formação inicial foi marcada em amarelo na Figura 4.4 e a região que tem uma formação posterior marcada em ocre, de modo a deixar explícito o caminho de formação para cada mapa aqui em estudo. A região de formação do estado de transição, descrita experimentalmente, contempla a formação do *loop* que liga a β_2 as fitas β_3 e β_4 e o *harpin* que encaixa essas duas fitas. Assim é possível verificar a partir desta Figura 4.4, que apenas o mapa de *CSU* apresenta caminho de formação que concorda com o resultado experimental.

No estudo de Clemente, Nymeyer e Onuchic (2000) em que foi empregado um estudo computacional as três proteínas aqui empregadas e de Lammert, Noel e Onuchic (2012) em que estuda a rota de enovelamento do domínio SH3 (PDB id: 1 FMK), corrobora com os resultados aqui apresentados. O que nos leva a uma próxima pergunta de porque o mapa de *CSU* consegue contemplar o estado de transição presente na proteína Barnase e consegue reproduzir o caminho de enovelamento definido experimentalmente pelo domínio SH3? É possível que por sua metodologia de complementaridade de superfície, aplicando um raio de van der waals, consiga diferenciar um resíduo do outro, introduzindo um mínimo de heterogeneidade aos resíduos de aminoácidos. Foi verificado a partir disso, se o mapa continha um potencial eletrostático, em que verificamos entre todos os contatos presentes no mapa, se as interações entre os pares de resíduos de aminoácidos, eram apenas atrativas. E por meio de um *script*, verificou-se que há interações atrativas, repulsivas e neutras. Também foi verificada a distância entre os contatos presentes em cada mapa e notou-se que este contempla contatos de mais longa distância e a partir desta fizemos uma aplicação de corte.

Aplicação de raio de corte no mapa de CSU

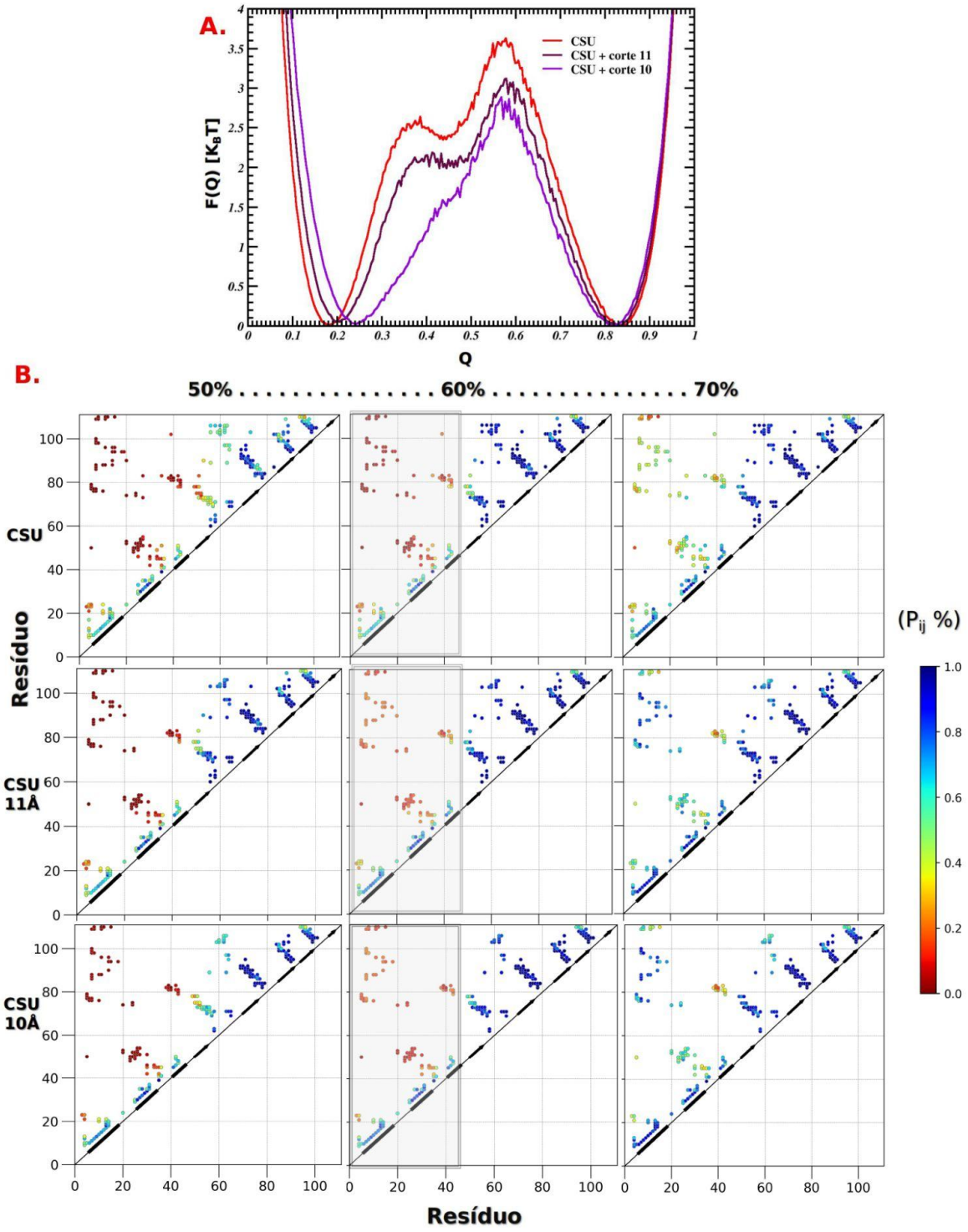
Sabendo que os mapas de CO e SCM apresentam uma distância máxima de 8 Å entre os resíduos com interações, aplicou-se um corte na distância das interações presentes no mapa de *CSU* começando de 8 Å e fomos aumentando de 1 Å, até conseguir replicar os resultados anteriormente. Esse corte no mapa de CSU foi aplicado apenas para a proteína Barnase e o domínio SH3, visto que para a proteína 2ci2 por ter um enovelamento robusto, o mapa de contato não influencia. Os resultados a seguir são apresentados sem a padronização indicada no início deste capítulo, pois nosso interesse não é comparação de mapa e sim verificar se há a necessidade de contatos de mais longa distância.

Na Figura 4.8 apresenta-se a aplicação de corte do mapa de CSU para a proteína Barnase, nesta o resultado inicia-se em 10 Å, para que não tenha redundância de informações. Com um aplicacao de corte de distancia de 10Å na cor violeta, a aplicação do corte na distância de 11Å na cor marrom e o CSU sem aplicação de corte em vermelho.

O mapa de Frt apesar de conter resíduos de longa distância, suas interações são minimamente frustradas, no qual filtra muitos contatos locais, desestabilizando a estrutura. Com os resultados obtidos por esta proteína, verificou-se a necessidade de conter mais interações de resíduos locais e/ou mais curto alcance, e de mais longa distância, a fim de introduzir heterogeneidade ao mapa de contato e consequentemente ao SBM. A fim de confirmar a necessidade de se ter heterogeneidade no mapa de contato, replicamos os resultados para o domínio SH3 da proteína PDB id: 1 FMK que veremos na Figura 4.9 a seguir.

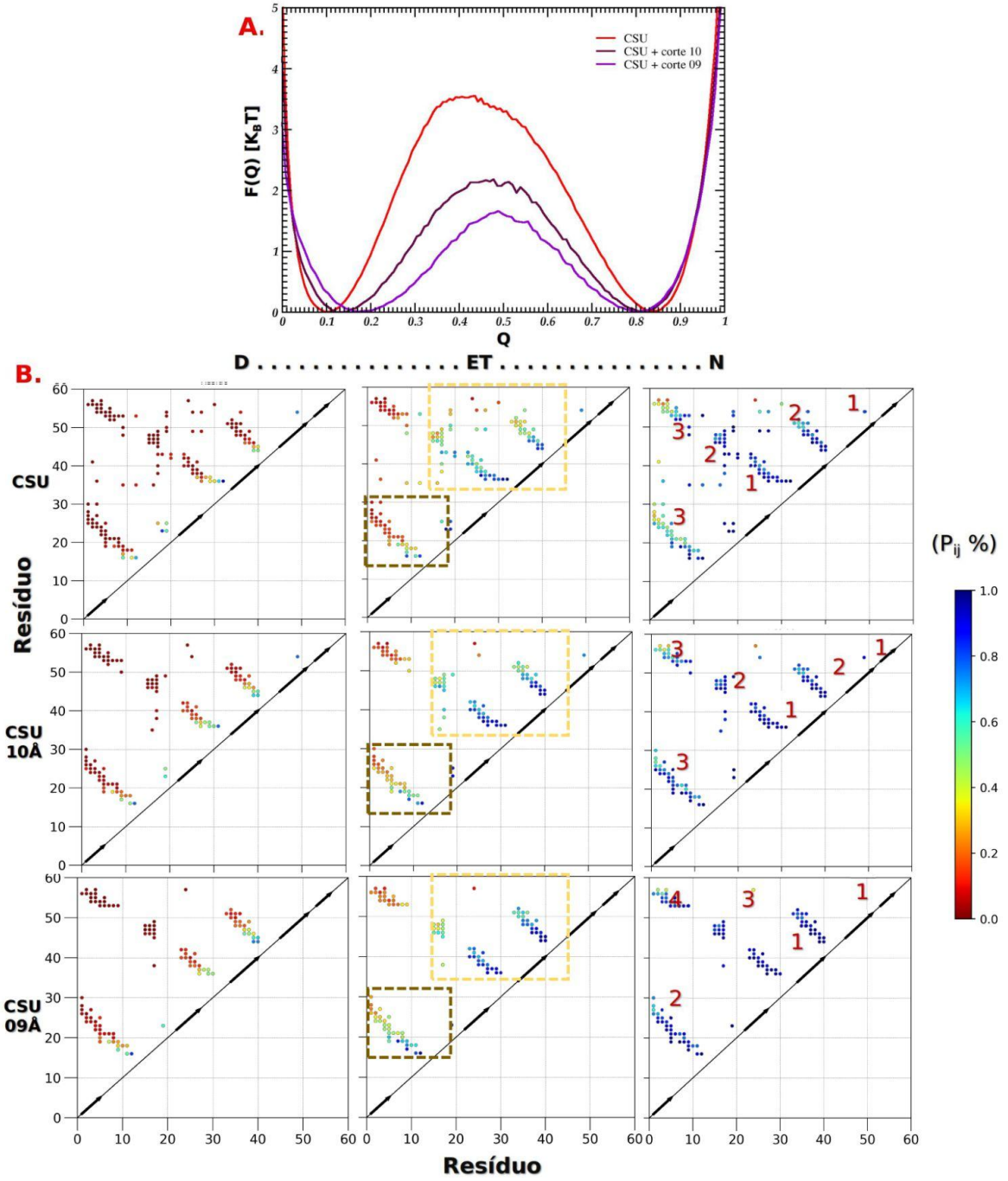
A partir da Figura 4.9, podemos verificar a mudança no caminho de formação, o que confirma a introdução de heterogeneidade ao mapa de contato, aplicada ao mapa de CSU.

Figura 4.8 Aplicação de corte de distância no mapa de CSU, para a proteína Barnase (PDB id: 1 bnr). Em A temos o gráfico de energia livre pela coordenada de reação Q, em que a curva em vermelho o mapa de CSU original, a curva em marrom, o gráfico de CSU com corte de distancia de 11 Å e a curva em lilás o gráfico de CSU com corte de distância de 10 Å. Em B temos o gráfico de probabilidade de formação de contatos (P_{ij}), para o estado de transição da Barnase, para o CSU original e as duas aplicações de corte. Contatos em vinho, indicam a não formação, contatos indicados em azul, indicam que aqueles contatos estão formados.



Fonte: Elaborado pelo Autor

Figura 4.9 Aplicação de corte de distância no mapa de CSU, para o domínio SH3 (PDB id: 1 FMK). Em A temos o gráfico de energia livre pela coordenada de reação Q, em que a curva em vermelho o mapa de CSU original, a curva em marrom, o gráfico de CSU com corte de distancia de 10 Å e a curva em lilás o gráfico de CSU com corte de distância de 09 Å. Em B temos o gráfico de probabilidade de formação de contatos (P_{ij}), para o estado desenovelado, de transição e nativo do domínio SH3, para o CSU original e as duas aplicações de corte. Contatos em vinho, indicam a não formação, contatos indicados em azul, indicam que aqueles contatos estão formados.



Fonte: Elaborado pelo Autor

Mudança na coordenadas de reação

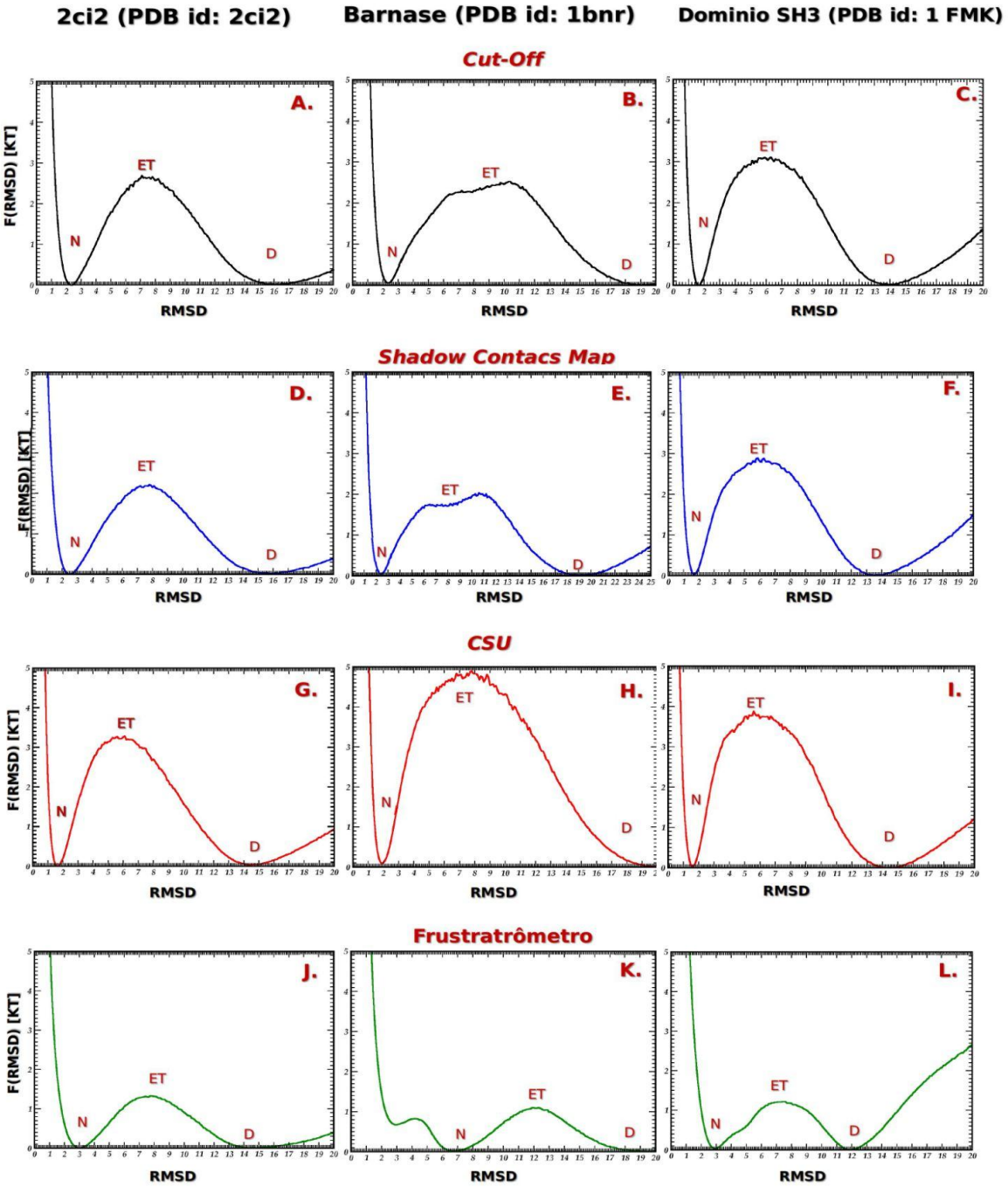
Até aqui, verificamos os resultados presentes em cada mapa utilizando uma única coordenada de reação como é comumente utilizada na literatura, a fração de contatos nativos Q, e mostramos os resultados que cada mapa nos apresenta para cada proteína, investigando o porquê o CSU consegue demonstrar o estado intermediário pela coordenada Q.

Agora, iremos mostrar o sistema por outra coordenada de reação, a raiz quadrada do desvio quadrático médio (RMSD), no qual está nos dá uma diferença estrutural entre duas conformações. Em que Figura 4.10, no qual se verifica quanto a energia livre varia entre a média de distância dos carbonos alfa durante as simulações. Neste, o estado nativo encontra-se do lado esquerdo do gráfico, pois quando a proteína está em seu estado enovelado, seu RMSD é baixo e quando em seu estado desenovelado, seu RMSD é alto. Exceto o gráfico de energia livre para o mapa de frustratrometro para a proteína Barnase (Figura 4.10 K), todos demonstraram uma reação de dois estados prevista pela modelagem do sistema. Além de verificar que para a proteína Barnase, nenhum dos mapas apresentou o estado intermediário presente experimentalmente.

Resolvemos então correlacionar as duas coordenadas de reação Q e RMSD, nos eixos XY e colorir em relação a energia livre. Na Figura 4.11 temos na coluna da esquerda temos os resultados para a proteína 2ci2, na coluna central temos os resultados para a proteína Barnase e na coluna a direita temos os resultados para o domínio SH3. Vemos nessa que o mapa de CSU consegue diferenciar de forma mais explícita o estado desenovelado do estado nativo para todas as proteínas. Também verificamos que para a proteína barnase, os mapas de CO (Figura 4.11 B) e SCM (Figura 4.11 E) não demonstram esse estado intermediário, e o Frt (Figura 4.11 K), mostra um enovelamento incompleto, como se a quantidade de contatos presentes não tivesse força suficiente para se enovelar completamente. O mapa de CSU

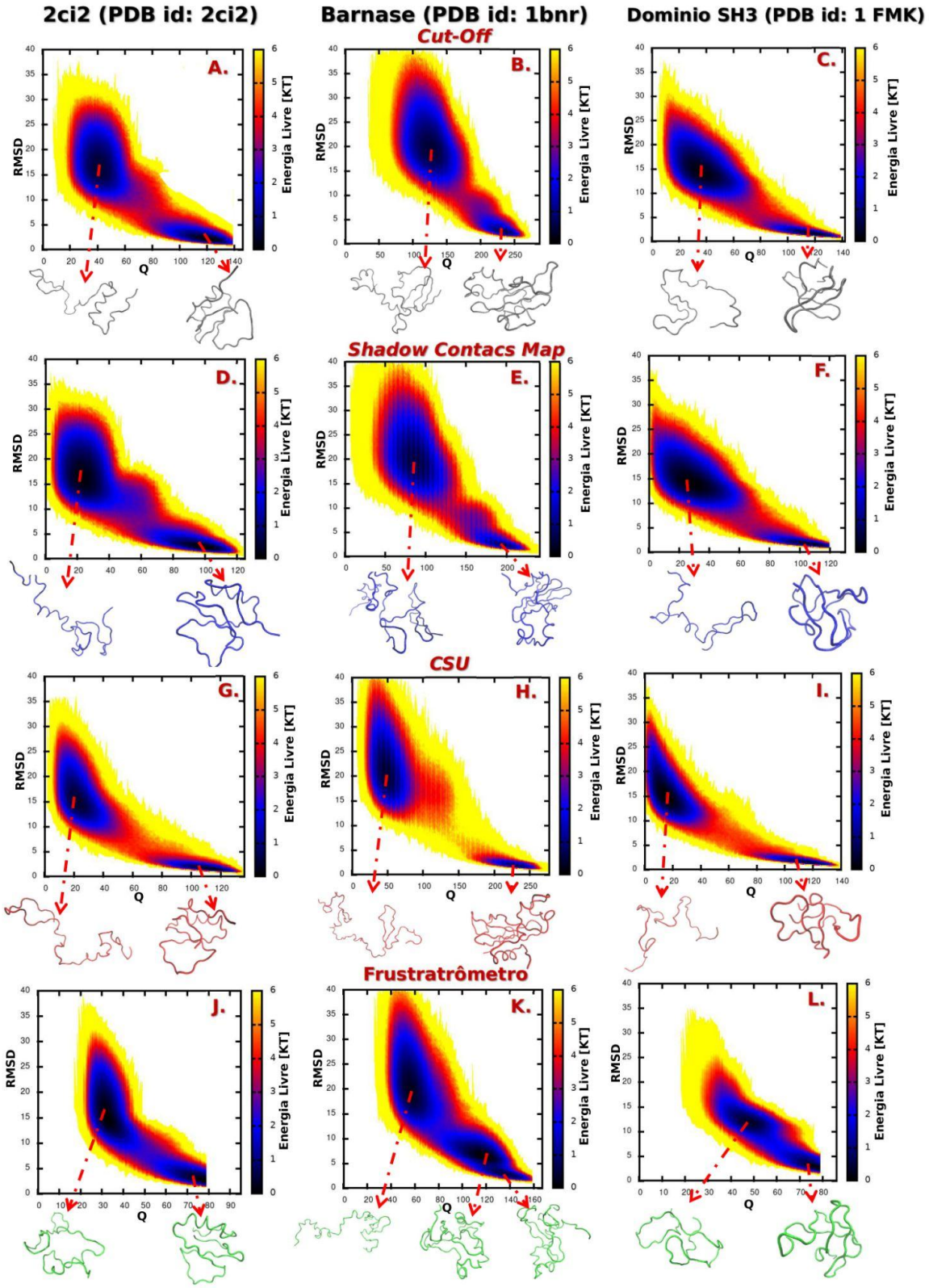
(Figura 4.10 H), demonstra energeticamente um possível estado intermediário em Q ~ 120, porém não há uma diferença estrutural, de modo a definir propriamente um estado intermediário.

Figura 4.10 Energia Livre F(RMSD) pela coordenada de reação RMSD. A coluna da esquerda mostra os resultados para a proteína 2ci2 (PDB id: 2ci2), a coluna central mostra os resultados para a proteína Barnase (PDB id: 1 bnr) e a coluna da direita mostra os resultados para o domínio SH3 (PDB id: 1 FMK). As imagens A, B e C referem-se ao mapa de CO. D, E e F referem-se ao mapa de SCM. G, H e I referem-se ao mapa de CSU e por fim, as imagens J, K e L referem-se ao mapa de Frt.



Fonte: Elaborado pelo Autor

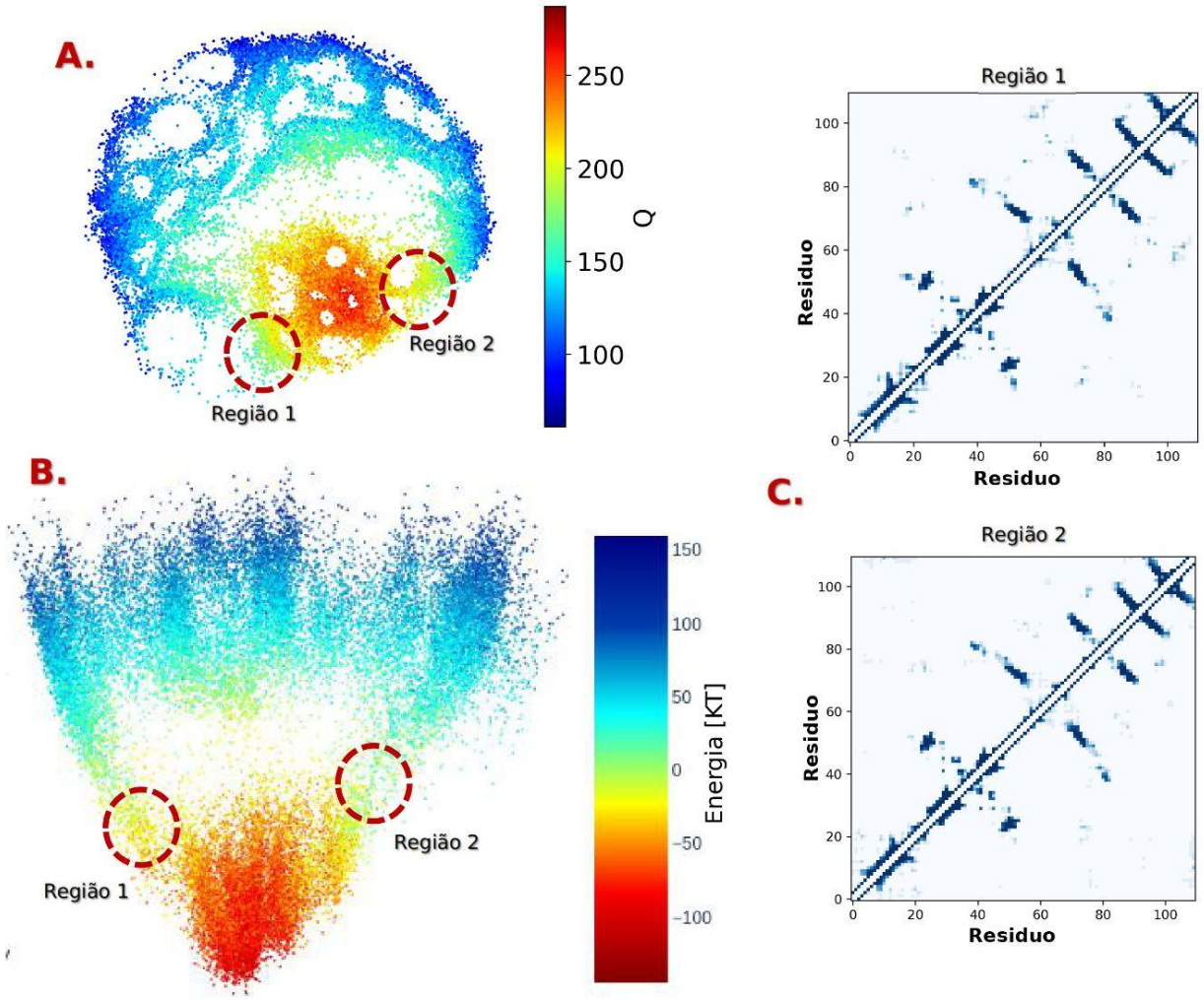
Figura 4.11 Correlação entre as coordenadas de reação Q e RMSD, colorida a partir da energia livre de cada estado. A coluna da esquerda mostra os resultados para a proteína 2ci2 (PDB id: 2ci2), a coluna central mostra os resultados para a proteína Barnase (PDB id: 1 bnr) e a coluna da direita mostra os resultados para o domínio SH3 (PDB id: 1 FMK). As imagens A, B e C referem-se ao mapa de CO. D, E e F referem-se ao mapa de SCM. G, H e I referem-se ao mapa de CSU e por fim, as imagens J, K e L referem-se ao mapa de Frt.



Fonte: Elaborado pelo Autor

Resolvemos agora, investigar esse sistema utilizando a metodologia do ELViM, que utiliza-se da coordenada de $q^{k,l}$(equação 3.13) para construir de forma 2D e 3D a superfície de energia para a proteína Barnase e verificar para todos os mapas o estado de transição e a formação da hélice 2, hélice 3 e a fita β_1, que são apenas parcialmente formados no estado intermediário e de transição demonstrado no artigo de Matouschek, Serrano e Fersht (1992).

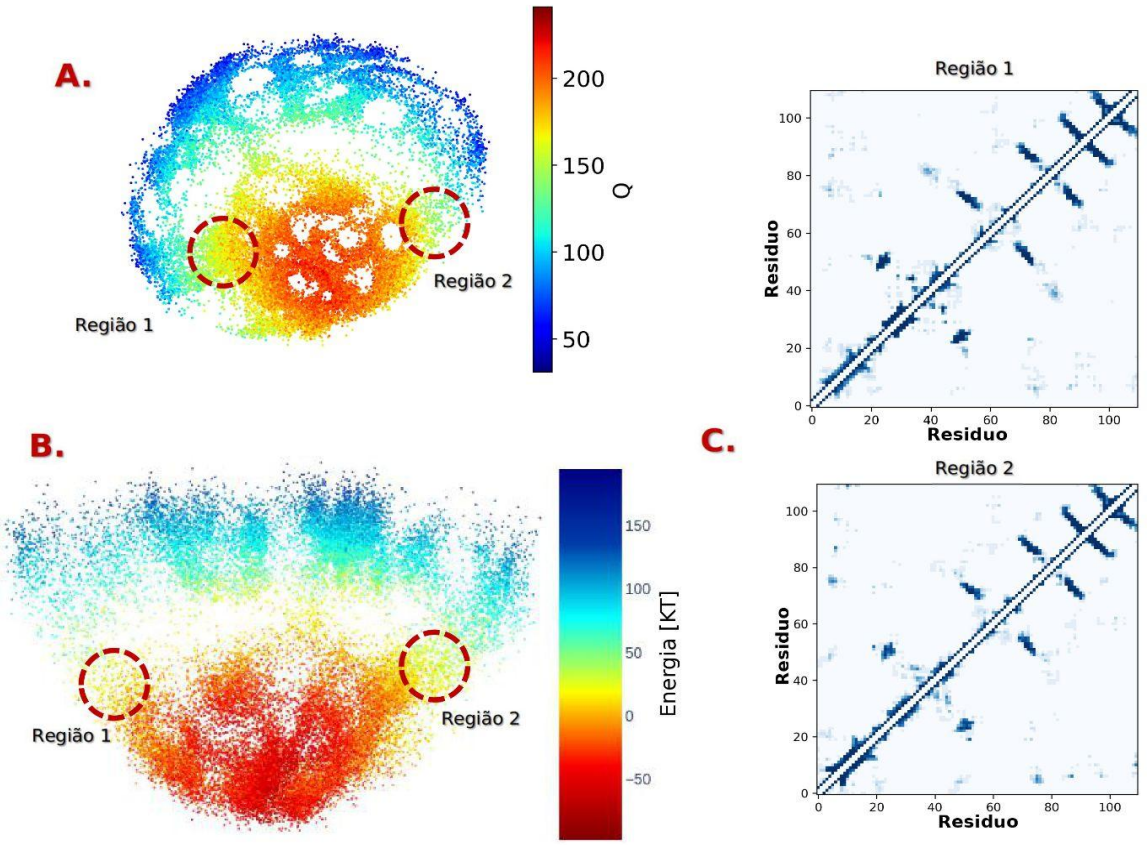
Figura 4.12 Espaço de fase conformacional e funil de estruturas visualizado através do ELVIM e gráfico de formação de estrutura para o mapa de Cut-Off para a proteína Barnase (PDB id: 1bnr). Em A, tem-se o espaço de configurações construído em 2D a partir da coordenada de reação Q em que na cor azul marinho encontram-se, estruturas desenoveladas, variando até a cor vermelho, em que encontram-se estruturas próximas ao estado nativo. Em B, tem-se funil de estruturas em 3D construídas a partir da dinâmica de enovelamento em T_f colorido pela energia livre do sistema, de modo que azul marinho representa energia alta, vermelho representa energia baixa. Em C, tem-se a média dos contatos formados pelas estruturas das regiões marcadas em A e B.



Fonte: Elaborado pelo Autor

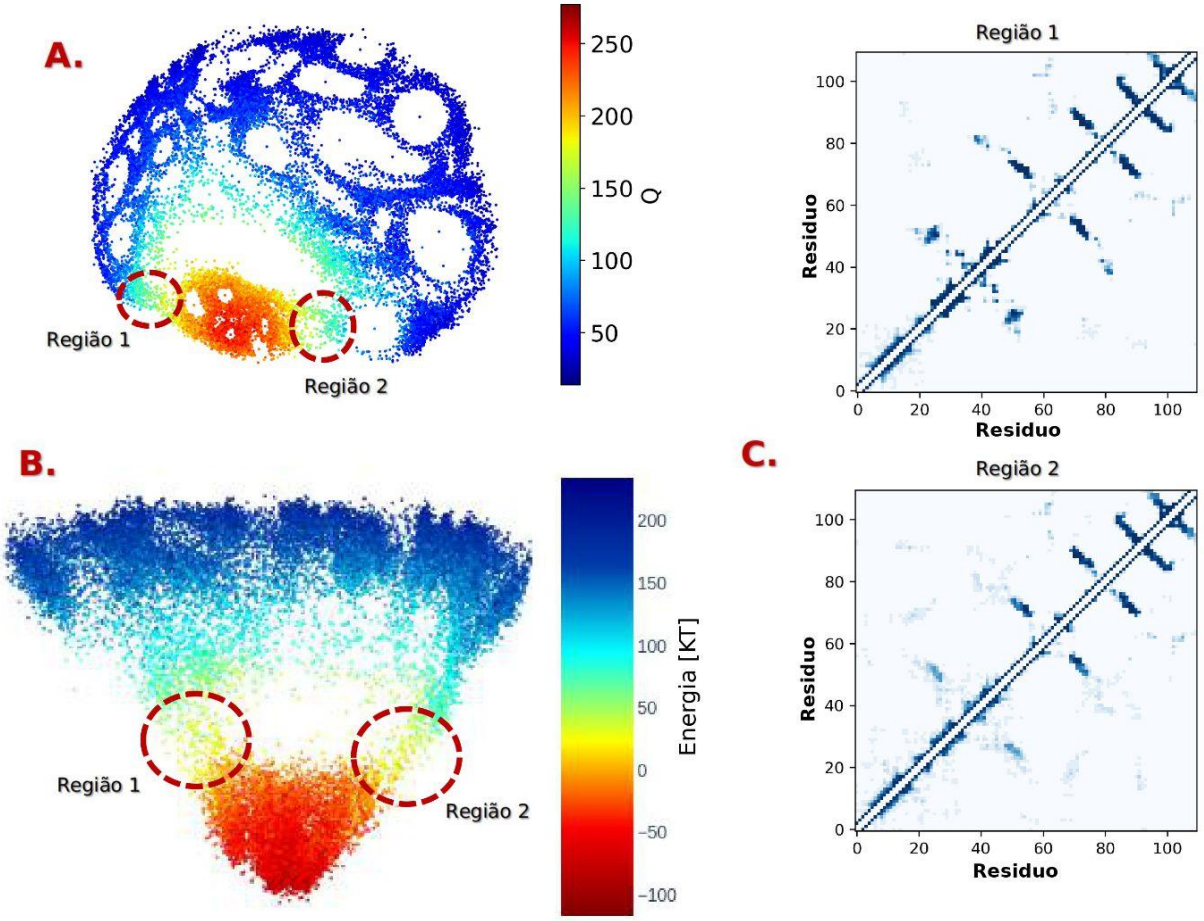
Para as Figuras 4.12, 4.13, 4.14 e 4.15, demonstrou-se em A o espaço de configurações construído em 2D a partir da coordenada de reação Q, fração de contatos nativos. variando da cor de azul marinho para contatos fracamente formados até vermelho, para contatos próximos ao estado nativo. Em B temos uma construção 3D da superfície energética do funil de enovelamento em que cores próximas do vermelho indicam a proteína enovelada variando as cores próximas ao azul marinho, indicando a estrutura em estados desenovelados. Em A e B circularam-se duas regiões em que determinaram o estado de transição, conectando os estados D e N. De modo que em C temos demonstrado a média dos contatos formados nessas regiões.

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). Em A, tem-se o espaço de configurações construído em 2D a partir da coordenada de reação Q em que na cor azul marinho encontram-se, estruturas desenoveladas, variando até a cor vermelho, em que encontram-se estruturas próximas ao estado nativo. Em B, tem-se funil de estruturas em 3D construídas a partir da dinâmica de enovelamento em T_f colorido pela energia livre do sistema, de modo que azul marinho representa energia alta, vermelho representa energia baixa. Em C, tem-se a média dos contatos formados pelas estruturas das regiões marcadas em A e B.



Fonte: Elaborado pelo Autor

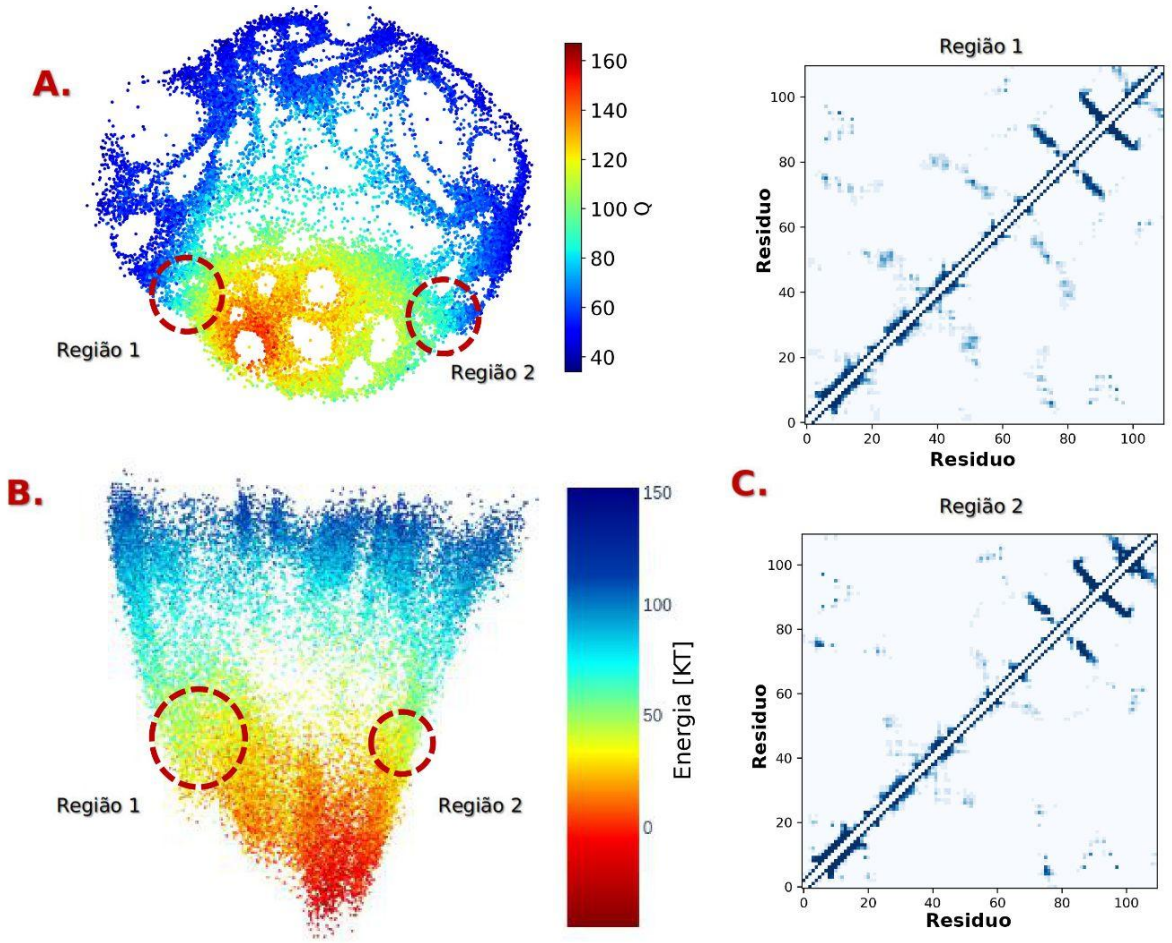
Figura 4.14 Espaço de fase conformacional e funil de estruturas visualizado através do ELVIM e gráfico de formação de estrutura para o mapa de CSU para a proteína Barnase (PDB id: 1bnr). Em A, tem-se o espaço de configurações construído em 2D a partir da coordenada de reação Q em que na cor azul marinho encontram-se, estruturas desenoveladas, variando até a cor vermelho, em que encontram-se estruturas próximas ao estado nativo. Em B, tem-se funil de estruturas em 3D construídas a partir da dinâmica de enovelamento em T_f colorido pela energia livre do sistema, de modo que azul marinho representa energia alta, vermelho representa energia baixa. Em C, tem-se a média dos contatos formados pelas estruturas das regiões marcadas em A e B.



Fonte: Elaborado pelo Autor

As regiões 1 e 2 indicam médias de formação parecida para todos os mapas, de modo que a separação em duas ou mais regiões neste caso, se dá pelas nuances estruturais, definidas pela metodologia do ELViM, pois foram pegas locais com mesmo quantidade de Q, de modo que as energias devam ser semelhantes. De modo que essas pequenas nuances não demonstram diferenças significativas no caminho de formação do sistema, indicando uma robustez do SBM com a troca dos mapas de contato.

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). Em A, tem-se o espaço de configurações construído em 2D a partir da coordenada de reação Q em que na cor azul marinho encontram-se, estruturas desenoveladas, variando até a cor vermelho, em que encontram-se estruturas próximas ao estado nativo. Em B, tem-se funil de estruturas em 3D construídas a partir da dinâmica de enovelamento em T_f colorido pela energia livre do sistema, de modo que azul marinho representa energia alta, vermelho representa energia baixa. Em C, tem-se a média dos contatos formados pelas estruturas das regiões marcadas em A e B.



Fonte: Elaborado pelo Autor

Fizemos essa análise apenas para a proteína Barnase de modo a verificar a existência de um estado intermediário. Para a proteína 2ci2 e para o domínio SH3 essas análises estão no apêndice B, imagens complementares.

Resolvemos utilizar uma quarta coordenada de reação, chamada de contatos quaisquer, Z, em que esta não depende do mapa de contato e sim apenas da trajetória do sistema, em que considera em contato todos os resíduos que estiverem dentro de um raio de corte

pré-estabelecido, em cada frame, sendo esses resíduos considerados em contato ou não pelo mapa de contato, de forma que podemos comparar as trajetórias dos mapas diretamente. Primeiro, verificamos o comportamento de colapso da cadeia pela temperatura, em que:

$$C = \langle Z^2 \rangle - \langle Z \rangle^2 \quad 4.1$$

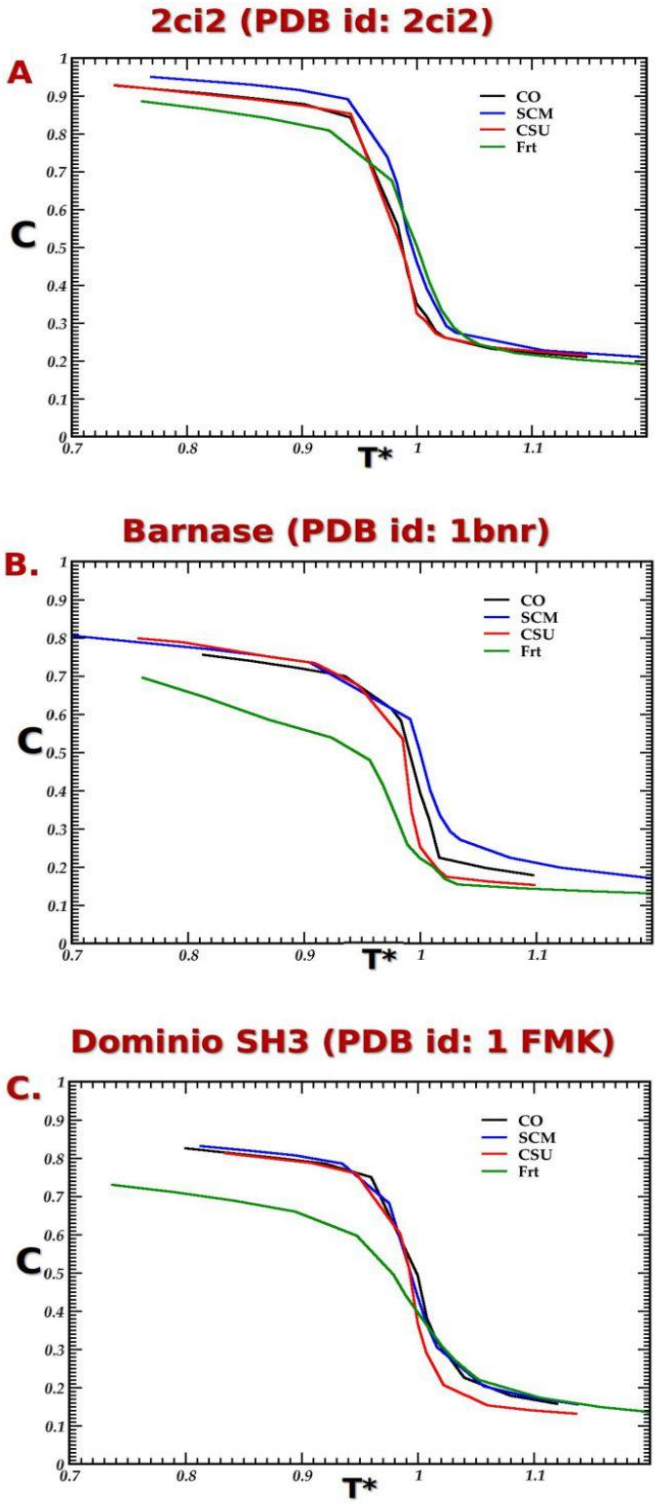
De modo a conseguir verificar se a cadeia se enovela instantaneamente, ou de forma cooperativa, ou seja, se o enovelamento ocorre de forma mais lenta. As coordenadas foram normalizadas de forma a poder compará-las, dividimos o eixo das abscissas pela temperatura de enovelamento, T_f de cada mapa respectivamente e o eixo das ordenadas pela altura do pico máximo.

Podemos ver na Figura 4.16 A, o colapso para a proteína 2ci2, em que podemos verificar uma curva quase que vertical da cadeia mostrando um enovelamento instantâneo da cadeia polipeptídica, demonstrando uma reação de dois estados. Na Figura 4.16 B temos o colapso pela temperatura, para a proteína Barnase no qual, temos um enovelamento quase que instantâneo para os mapas de CO, SCM e CSU e um enovelamento "lento" para o mapa de Frt. Também, por a Barnase apresentar um estado intermediário, esperava-se ver uma representação deste no mapa de CSU, porém também não foi visto. Na Figura 4.16 C temos o colapso da cadeia pela temperatura para o domínio SH3, de modo que para os mapas de CO, SCM e CSU, acontece um enovelamento rápido, enquanto que para o mapa de Frt, a curva demonstra um enovelamento mais lento, de modo a mostrar um não colapso da cadeia, ou enovelamento cooperativo.

Com este resultado verificamos que o estado intermediário previsto para a proteína Barnase, no gráfico de energia livre pela coordenada de reação, não consegue novamente ser

demonstrado. E para o domínio SH3 e a proteína Barnase, o mapa de Frt demonstra não conter contatos suficientes para um enovelamento colapsado da cadeia de aminoácidos.

Figura 4.16 Colapso da cadeia a partir da coordenada de contatos quaisquer Z, pela temperatura. Em A, B e C temos colapso (C) por T* para as proteínas 2ci2 (PDB id 2ci2), para a proteína Barnase (PDB id: 1 bnr) e domínio SH3 (PDB id: 1 FMK) respectivamente. Os mapas foram definidos por cores, em que preto refere-se ao mapa de Cut-Off, azul ao mapa de SCM, vermelho ao mapa de CSU e verde ao mapa de Frt.



Fonte: Elaborado pelo Autor

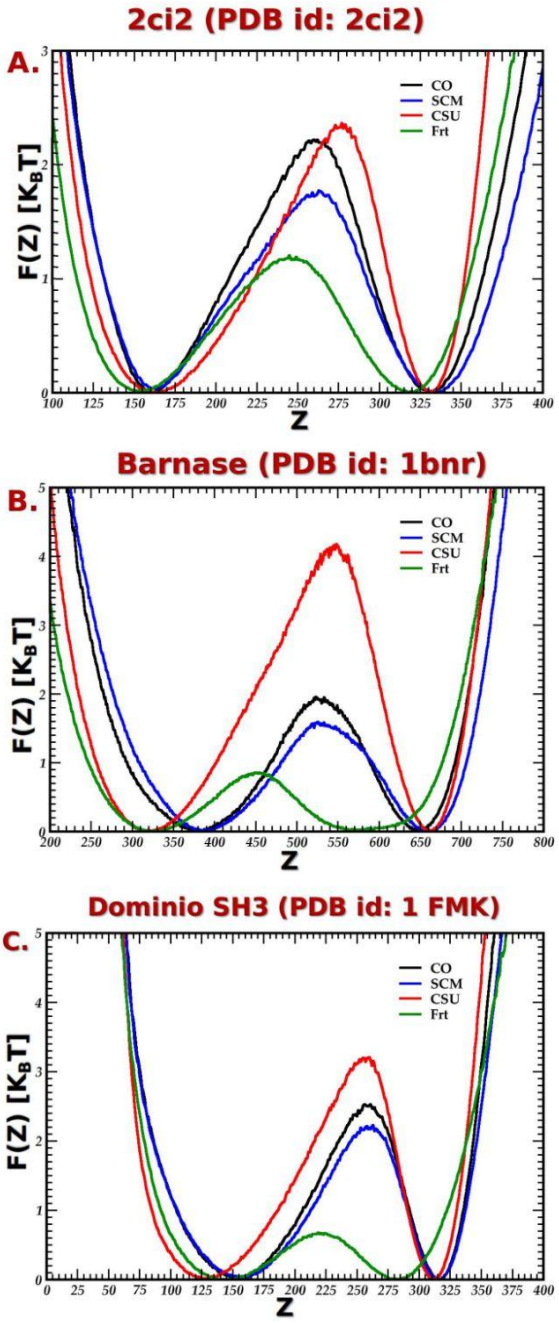
De modo a fazer uma comparação direta entre os mapas de contatos, utilizamos Z como coordenada de reação por não estar diretamente ligado aos mapas de contato, de modo que é definido a partir das trajetórias. Da mesma forma utilizamos a relação das energias de diedros como forma de comparação direta entre as energias do sistema, pois existem combinações de ângulos típicos para as estruturas secundárias, e a conformação dos péptidos é definida pela atribuição de valores em φ e ϕ de modo a definir a quiralidade da cadeia, este utiliza quatro $C\alpha$ consecutivos, no qual também não depende do mapa de contato, apenas do próprio estrutura PDB.

A partir destas, podemos fazer uma comparação direta entre as curvas de energias livres para os quatro mapas propostos neste estudo, referente às três proteínas estudadas, ver Figura 4.17. Para todas as proteínas, os quatro mapas conseguem demonstrar a reação de dois estados, desenovelado e nativo, de modo como o sistema é modelado.

Verificando a variação na altura das curvas da energia livre, de modo que a barreira mais alta energia livre pertence ao mapa de CSU (curva em vermelho), dá-se pela capacidade de introduzir contatos de mais longa distancia ao sistema, de modo a ter uma perda maior de entropia configuracional pela formação desses contatos de longa distância. O mapa de Frt (curva em verde), apesar de conter alguns contatos de longa distância, demonstra uma menor barreira de energia livre, o que se deve à pouca quantidade de interações presentes no mapa, necessitando de uma maior quantidade de contatos formados para se enovelar. Os mapas de CO (curva em preto) e SCM (curva em azul), apresentam resultados semelhantes. Apesar da altura da barreira do SCM ser um pouco menor que a do CO, isso se dá pela quantidade de contatos a menos que o SCM tem, pois suas metodologias são parecidas apesar do parâmetro de filtragem proposto pelo SCM.

Também, esperava-se que pela utilização da energia determinada pelos diedros, energia esta que determina formação da estrutura secundária da proteína, pudesse demonstrar a formação do estado intermediário determinado experimentalmente para a proteína Barnase, seria evidenciado, o que não ocorreu.

Figura 4.17 Energia Livre F(Z) pela coordenada de reação Z, contatos quaisquer. Em A, B e C temos a energia livre F(Z) por Z proteínas 2ci2 (PDB id 2ci2), para a proteína Barnase (PDB id: 1 bnr) e domínio SH3 (PDB id: 1 FMK) respectivamente. Os mapas foram definidos por cores, em que preto refere-se ao mapa de CO, azul ao mapa de SCM, vermelho ao mapa de CSU e verde ao mapa de Frt.



Fonte: Elaborado pelo Autor

Assim, para o modelos simplificado como o SBM-$C\alpha$, a superfície de energia é lisa o suficiente de modo que uma mudança na coordenada de reação não reflita características essenciais do processo de enovelamento. De modo que o mapa de contato deve introduzir a heterogeneidade entre os resíduos de aminoácidos.

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 - Cα, 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 Cα 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. Softwarex, [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 l. 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.

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.

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.

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, Djamal; 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.

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.; VISHVESHWARA, 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.

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 theclostridium thermocellumcellulosome. 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

APÊNDICE A - *Scripts*

Script para mapa de contato *Cut-Off*

Este script tem como objetivo definir os contatos de uma proteína por meio de um corte de distância, a partir da linguagem de programação *python*.

1. Pegar o pdb original e deixar apenas os $C\alpha$;

```
cat file.pdb | grep ATOM | grep CA > protein_CA.pdb
```

2. Segundo o artigo *doi: 10.1186/1471-2105-11-283*, o corte de distância com maior ganho de informação e que não cause redundância no sistema está por volta de 8.5 angstrom;

3. Início do programa;

```
line = []
line2 = []
CO = 8
```

4. Abre e lê o *protein_CA.pdb* por linha*;*

```
with open("protein_CA.pdb", "r") as f:
   fline = f.read().split('\n')[:-1]
```

5. Abre o arquivo que será enviado os pares de contato;

```
arquivo = open("map.mq", "w")
```

6. O primeiro laço lê a linha e separa o aminoácido com suas posicao x, y, e z

```
cont = 0
for i in range(0, len(fline)):
   line = fline[i].split()
   aai = float(line[1])
   xi = float(line[6])
   yi = float(line[7])
   zi = float(line[8])
```

7. O segundo laço repete processo 6, porém com j = i+3 e faz a média geométrica da distância entre o resíduo i e j ;

```
    for j in range(i+3, len(fline)):
      line2 = fline[j].split()
      aaj = float(line2[1])
      xj = float(line2[6])
      yj = float(line2[7])
      zj = float(line2[8])
      r = ((xi-xj)**2 + (yi-yj)**2 + (zi-zj)**2)**(1/2)
```

8. Condição de corte estabelecida, em que temos o *print: c*adeias 1 resíduo i - cadeira 1 resíduo j;

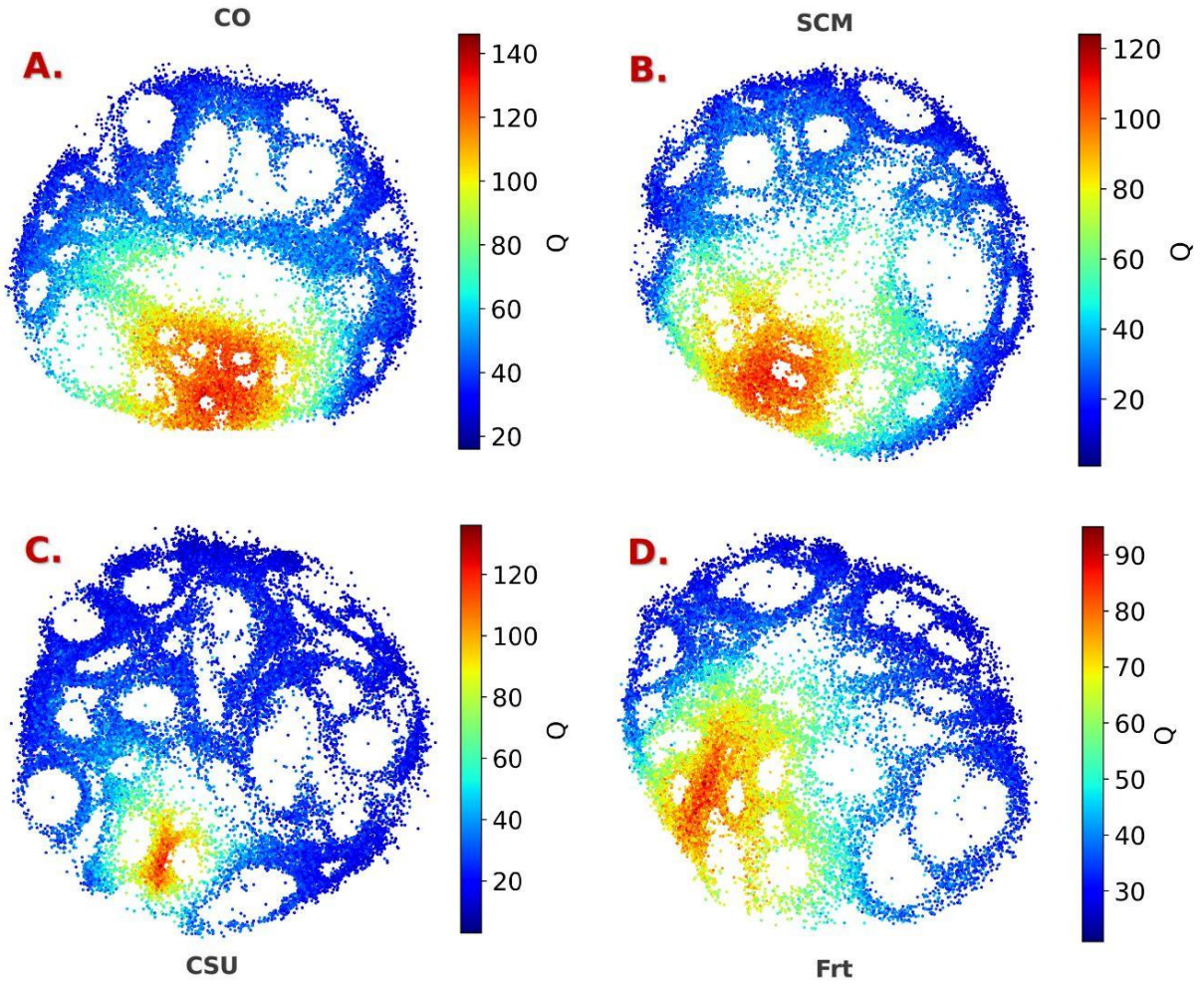
```
      if r <= CO:
        a = float(line[1])
                print(f'1 {aai:.0f} 1  {aaj:.0f}', file=arquivo)
        cont +=1
```

9. Fecha o arquivo *map.mq*.

```
arquivo.close()
print(f'Há {cont} contatos.')
```

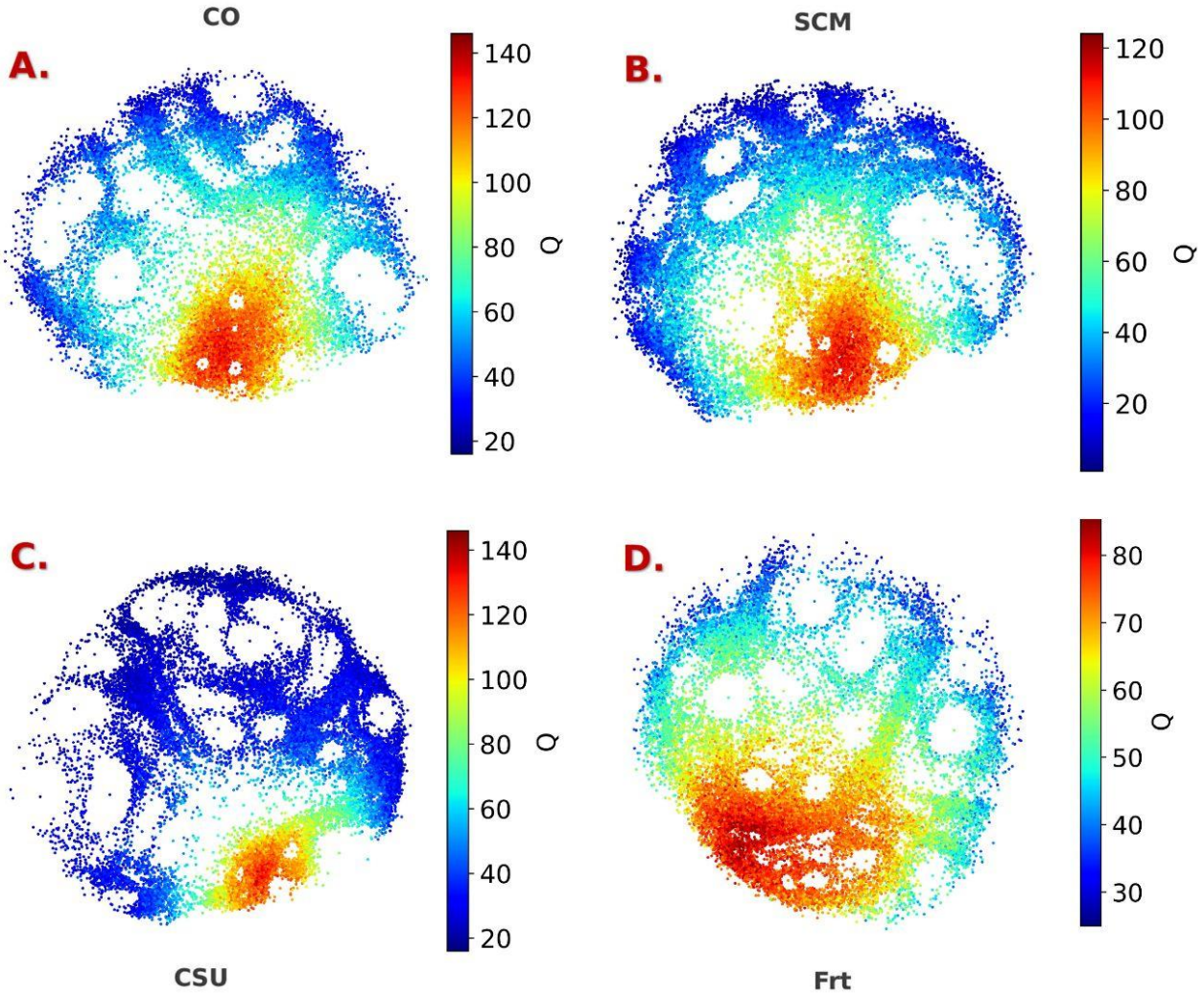
APÊNDICE B - Imagens complementares

Figura B.1 Espaço de fase conformacional visualizado através do ELVIM para a proteína 2ci2 (PDB id: 2ci2). Tem-se o espaço de configurações construído em 2D a partir da coordenada de reação Q em que na cor azul marinho encontram-se, estruturas desenoveladas, variando até a cor vermelho, em que encontram-se estruturas próximas ao estado nativo. Temos então o espaço conformacional para o mapa A. Cut-Off, em B. SCM, em C. CSU e em D. Frt.



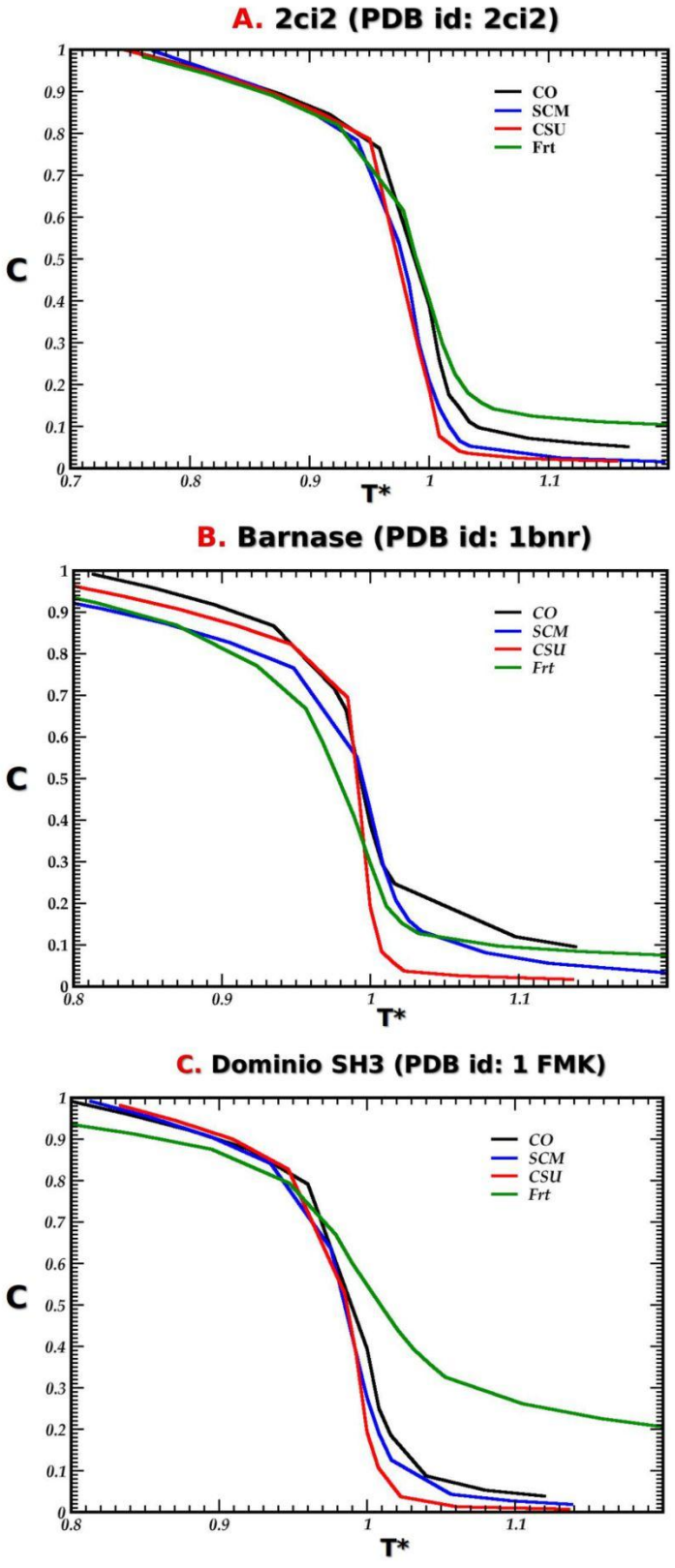
Fonte: Elaborado pelo autor

Figura B.2 Espaço de fase conformacional visualizado através do ELVIM para o domínio SH3 (PDB id: 1 FMK). Tem-se o espaço de configurações construído em 2D a partir da coordenada de reação Q em que na cor azul marinho encontram-se, estruturas desenoveladas, variando até a cor vermelho, em que encontram-se estruturas próximas ao estado nativo. Temos então o espaço conformacional para o mapa A. Cut-Off, em B. SCM, em C. CSU e em D. Frt.



Fonte: Elaborado pelo autor

Figura B.3 Colapso da cadeia a partir da coordenada de fração de contatos nativos, Q, pela temperatura. Em A, B e C temos colapso (C) por T* para as proteínas 2ci2 (PDB id 2ci2), para a proteína Barnase (PDB id: 1 bnr) e domínio SH3 (PDB id: 1 FMK) respectivamente. Os mapas foram definidos por cores, em que preto refere-se ao mapa de Cut-Off, azul ao mapa de SCM, vermelho ao mapa de CSU e verde ao mapa de Frt.



Fonte: Elaborado pelo autor