



Universidade Federal do Rio de Janeiro
Centro de Ciências Matemáticas e da Natureza
Instituto de Química

Projeto de Final de Curso (IQWX01)

**Caracterização estrutural e da interação com
DNA do domínio *cold-shock* da proteína AtGRP2
de *Arabidopsis thaliana* por Ressonância
Magnética Nuclear**

Karina de Carvalho Pougy

Rio de Janeiro, 2021

Junho/2021

Caracterização estrutural e da interação com DNA do domínio *cold-shock* da proteína AtGRP2 de *Arabidopsis thaliana* por Ressonância Magnética Nuclear

Karina de Carvalho Pougy

PROJETO DE FINAL DE CURSO SUBMETIDO AO CORPO DOCENTE DO DEPARTAMENTO DE BIOQUÍMICA DO INSTITUTO DE QUÍMICA DA UNIVERSIDADE FEDERAL DO RIO DE JANEIRO COMO PARTE DOS REQUISITOS NECESSÁRIOS PARA A OBTENÇÃO DO GRAU DE QUÍMICO COM ATRIBUIÇÕES TECNOLÓGICAS.

Prof. Anderson de Sá Pinheiro (IQ - UFRJ)

Prof^a. Bianca Cruz Neves (IQ - UFRJ)

Prof. Fábio César Sousa Nogueira (IQ - UFRJ)

RIO DE JANEIRO, RJ – BRASIL
JUNHO/2021

Resumo

AtGRP2 (*Arabidopsis thaliana glycine rich protein 2*) é uma proteína rica em glicina e ligante de RNA que desempenha papéis chave na resposta a estresse abiótico e na regulação do tempo de floração em *Arabidopsis thaliana*. Esta proteína contém um domínio N-terminal *cold-shock* (CSD) e dois domínios C-terminais *zinc finger* do tipo CCHC retrovirais intercalados com regiões ricas em glicinas. Domínios *cold-shock* são conhecidos por se ligarem a ácidos nucleicos. Dada a função de AtGRP2 na tolerância a estresses abióticos em *A. thaliana*, o conhecimento de seu mecanismo de ação se faz interessante dentro do contexto de poluição ambiental e aquecimento global. Assim, o presente projeto teve como objetivo a caracterização estrutural e análise da interação com o oligonucleotídeo T7 do CSD de AtGRP2 de modo a se iniciar o processo de elucidação do mecanismo de ação da proteína a nível atômico. Uma construção do CSD de AtGRP2 (aminoácidos 1-79) foi expressa de forma recombinante em cepas de *E. coli*, BL21 (DE3), e purificada através de uma combinação de técnicas cromatográficas (cromatografia por afinidade a níquel e cromatografia por exclusão molecular). No espectro 2D [¹H,¹⁵N] HSQC de AtGRP2-CSD foi constatada a presença de mais do que o dobro da quantidade de sinais esperados; estes sinais extras são caracterizados por uma baixa dispersão na dimensão do ¹H, encontrados aglomerados em torno de 8,2 ppm, e maior intensidade. Foi concluído que há uma conformação desenovelada em equilíbrio com o estado enovelado da proteína em solução. A partir dos experimentos de RMN multidimensional adquiridos, foi possível assinalar 95% das ressonâncias da forma enovelada de AtGRP2-CSD, com exceção dos 2 artifícios de clonagem e dos resíduos M1 e S41, e 68% da forma desenovelada, sendo os seguintes aminoácidos não identificados: G-1, H0, M1, S2, N5, R10, S14, V15, K16, W17, F18, D19, T20, Q21, K22, G23, F24, D30, D31, H39, Q40, S41, L51, A52 e N65. Este equilíbrio conformacional já foi observado na literatura para as proteínas drkN SH3 e YB-1 CSD. O propósito desta instabilidade parcial e sua função no mecanismo de ação destas proteínas e de AtGRP2-CSD ainda deve ser mais investigado. A análise da interação de AtGRP2-CSD com o oligonucleotídeo T7 foi realizada através de experimentos de titulação por RMN em razões molares crescentes de ligante e a Perturbação do Deslocamento Químico (CSP) foi calculada a partir destes. A superfície de interação proteína-DNA foi identificada a partir de resíduos com valores de CSP maiores que 2 e 3 σ : G25, F26, L36 e V38 (3 σ) e G23, T28, F37, H39, Q40, R49 e L51 (2 σ). Estes fazem parte dos motivos ligantes de ácidos nucleicos RNP-1 e RNP-2 de AtGRP2-CSD e da região de *loop* β 3 β 4, a qual é conhecida na literatura por auxiliar na interação entre CSDs e ácidos nucleicos.

Sumário

Agradecimentos.....	6
Índice de Figuras	7
Índice de Tabelas	8
1. Introdução.....	9
1.1. Regulação pós-transcricional da expressão gênica	9
1.2. Proteínas ricas em glicina (GRPs).....	9
1.3. AtGRP2.....	11
1.4. Domínio <i>cold-shock</i>	14
2. Objetivos.....	16
2.1. Objetivo geral	16
2.2. Objetivos específicos.....	16
3. Material e Métodos	17
3.1. Material	17
3.1.1. Meio mínimo (M9).....	17
3.1.2. Meio Luria-Bertani (LB).....	17
3.1.3. Vetor de expressão.....	17
3.1.4. Cepas Bacterianas	18
3.2. Métodos	18
3.2.1. Transformação Bacteriana.....	18
3.2.2. Expressão de AtGRP2-CSD	19
3.2.3. Eletroforese em gel de poliacrilamida sob condições desnaturantes (SDS-PAGE)	19
3.2.4. Lise Bacteriana	20
3.2.5. Purificação de AtGRP2-CSD	20
3.2.6. Dosagem de Proteína pelo Método Colorimétrico de Pierce	22
3.2.7. Análise da massa molecular por Espectrometria de Massas (MS).....	22
3.2.8. Aquisição dos dados de Ressonância Magnética Nuclear (RMN).....	22
3.2.9. Assinalamento sequencial das ressonâncias	23
3.2.10. Análise da interação com oligonucleotídeo por Perturbação do Deslocamento Químico (CSP).....	23
3.2.11. Modelagem molecular de AtGRP2-CSD	24
4. Resultados e Discussão.....	25
4.1. Resultados	25
4.1.1. Expressão e purificação de AtGRP2-CSD	25
4.1.2. Equilíbrio de conformações enovelado-desenovelado de AtGRP2-CSD....	29
4.1.2.1. 1D ¹ H RMN	29

4.1.2.2. [¹ H, ¹⁵ N] HSQC.....	30
4.1.2.3. Estratégia de assinalamento das ressonâncias.....	32
4.1.2.4. Assinalamento da cadeia principal das formas enovelada e desenovelada de AtGRP2-CSD.....	37
4.1.3. Análise da interação de AtGRP2-CSD com oligonucleotídeo de DNA	39
4.2. Discussão.....	45
4.2.1. Equilíbrio conformacional de AtGRP2-CSD	45
4.2.2. Sítio de interação de AtGRP2-CSD inclui os motivos RNP-1 e RNP-2	48
5. Conclusões	50
6. Referências Bibliográficas	51

Agradecimentos

Primeiramente, gostaria de agradecer ao meu pai, Marcelo Pougy, por me apoiar e acompanhar desde o início da minha decisão de cursar química, por me ouvir e me ajudar a passar pelos momentos difíceis durante o meu caminho e sempre me incentivar a perseverar. À minha tia, Martha Pougy, por estar comigo no começo dessa minha jornada e por me aconselhar nos momentos que precisei, especialmente neste final de curso. Ao meu avô, José Pougy, por sempre me incentivar a seguir em frente com a química e a pesquisa e por acreditar no meu potencial. Também agradeço aos meus outros familiares por todo o apoio ao longo destes anos.

Aos meus amigos do IQ, Guilherme, João, Rafael, Matheus de Assis, Matheus Diniz, Mariana, Marianna e Jéssica, por todo o companheirismo, risadas, conversas filosóficas, ajuda e compreensão nos momentos de surto durante nosso tempo neste curso. A universidade não teria sido a mesma sem a presença de vocês.

Aos professores, por todo o conhecimento transmitido e pelas conversas esclarecedoras fora das aulas. Em especial, agradeço ao professor Sérgio Machado pelos diversos ensinamentos e conselhos no início da minha trilha pela química.

Ao meu orientador, professor Anderson Pinheiro, por ter me apresentado à bioquímica para além do conhecido do ensino médio e ter me ajudado a reviver o meu gosto pela área. Muito obrigada pela oportunidade de fazer este trabalho, por todas as explicações e ensinamentos, pelas advertências necessárias, pelos incentivos e oportunidades de crescimento.

Aos meus colegas de laboratório, alguns que ainda trabalham no LabMol e outros não, Daniel, Igor, Clara, Nathália, Luiza, Carol Lixa, Gustavo, Ana Elisa e Marcelo, pela companhia no laboratório e em reuniões, tornando os dias mais alegres. Em especial, agradeço ao Daniel pela grande ajuda e paciência ao me ouvir e auxiliar em diversas questões no laboratório, e ao Igor, também por sua ajuda e conselhos.

Agradeço também aos professores Bianca Neves e Fábio Nogueira por terem aceitado fazer parte da minha banca examinadora, respondendo prontamente ao meu convite.

Índice de Figuras

Figura 1: Classificação das proteínas ricas em glicina (GRPs).....	11
Figura 2: Representação esquemática dos domínios da proteína AtGRP2.....	12
Figura 3: Fenótipo observado em plantas transgênicas de <i>Arabidopsis thaliana</i>	13
Figura 4: Estrutura tridimensional e conservação na sequência primária de CSDs	15
Figura 5: Representação esquemática da construção Thio ₆ -His ₆ -TEV-AtGRP2-CSD	18
Figura 6: Gel de SDS-PAGE da expressão de AtGRP2-CSD.....	25
Figura 7: Primeira etapa da purificação de AtGRP2-CSD.....	26
Figura 8: Segunda etapa da purificação de AtGRP2-CSD.....	27
Figura 9: Segunda etapa da purificação de AtGRP2-CSD.....	28
Figura 10: Análise da identidade e pureza de AtGRP2-CSD por espectrometria de massas	29
Figura 11: Espectro ¹ H 1D de RMN de AtGRP2-CSD.....	30
Figura 12: Espectro [¹ H, ¹⁵ N] HSQC de AtGRP2-CSD	31
Figura 13: Representação esquemática da transição visual de experimentos de RMN em duas dimensões para três dimensões	33
Figura 14: Representação em uma cadeia polipeptídica da correlação entre sistemas de <i>spin</i> nucleares realizada a partir de experimentos de RMN.....	35
Figura 15: Fatias do espectro 3D HNCACB adquirido de AtGRP2-CSD	37
Figura 16: Espectro [¹ H, ¹⁵ N] HSQC de AtGRP2-CSD 95% assinalado, conformação enovelada	38
Figura 17: Espectro [¹ H, ¹⁵ N] HSQC de AtGRP2-CSD 68% assinalado, conformação desenovelada	39
Figura 18: Sobreposição de espectros [¹ H, ¹⁵ N] HSQC de AtGRP2-CSD em razões molares crescentes de T7	40
Figura 19: Gráfico de Perturbação do Deslocamento Químico (CSP) por resíduo	42
Figura 20: Análise do sítio de interação de AtGRP2-CSD	43
Figura 21: Intensidade dos sinais da conformação enovelada de AtGRP2-CSD em razões molares de T7 diferentes	45

Índice de Tabelas

Tabela 1: Experimentos de RMN em 3D utilizados para o assinalamento de ressonâncias da cadeia principal de proteínas	33
---	----

1. Introdução

1.1. Regulação pós-transcricional da expressão gênica

O número de moléculas de RNAs mensageiros (mRNAs) existentes em uma célula em um dado momento é extremamente grande, variando de 15.000 a 150.000 em leveduras e mamíferos, respectivamente, por exemplo ⁽¹⁾. É de se esperar, portanto, que o destino desses RNAs, como sua localização, transporte, estabilização e degradação ^(1,2), seja finamente controlado de acordo com as necessidades da célula. Este processo ocorre em diversos níveis; um deles é a regulação pós-transcricional da expressão gênica que, tal como seu nome indica, consiste em um conjunto de mecanismos que agem no final do processo de transcrição, ou seja, que atuam sobre a biologia do mRNA. Esta regulação também afeta, conseqüentemente, a síntese proteica.

Em eucariotos, os processos de transcrição e tradução ocorrem em compartimentos diferentes; no núcleo e no citoplasma, respectivamente ⁽³⁾. Isto torna possível a realização de um extenso processamento pós-transcricional de pré-mRNAs, o que inclui os processos de *splicing*, edição, adição de uma cauda de poliadeninas na terminação 3' (cauda poliA) e de um resíduo de 7-metilguanossina na terminação 5' (*cap-5'*) ^(1,3). O destino do mRNA maduro é determinado pelo repertório de biomoléculas que interagem com ele, incluindo proteínas ligantes de RNA (RBP) e RNAs não codificantes, como microRNAs ^(1,3), também chamados de fatores *trans*, os quais se ligam a elementos de sequência, ou elementos *cis*, presentes nas regiões 5' e 3' não-traduzidas do mRNA (UTRs) ^(2,4). Estes elementos podem aparecer como sequências de nucleotídeos repetidas ou estruturas secundárias, e medeiam esses processos de regulação ⁽²⁾. A regulação pós-transcricional oferece, como um todo, a possibilidade de se ter um conjunto ainda mais diverso de mRNAs a partir de um mesmo genoma ⁽³⁾, além de ser fundamental na resposta da célula a mudanças externas a partir do controle rápido da quantidade de mRNAs intracelulares ⁽⁵⁾.

1.2. Proteínas ricas em glicina (GRPs)

As proteínas ricas em glicina (GRPs) formam uma superfamília que possui motivos ricos no aminoácido glicina. Isto implica na formação de domínios cujas estruturas são altamente flexíveis, o que é favorável para interações com outras

proteínas e ácidos nucleicos ⁽⁶⁾. Os genes que codificam estas proteínas são altamente regulados por diversos fatores, como temperatura, ferimentos, salinidade e luz, como é observado no caso de organismos vegetais ⁽⁶⁾.

Atualmente, diversas GRPs são conhecidas, e sabe-se que estas possuem estruturas primárias e domínios funcionais distintos ⁽⁷⁾. Assim, esta superfamília é dividida em cinco classes de acordo com sua estrutura primária (Figura 1). A classe I é formada por GRPs chamadas de clássicas. Estas possuem alto conteúdo de glicina (40 a 70%) e repetições GGX, sendo G é o aminoácido glicina e X é qualquer outro aminoácido. A classe II possui repetições GGXXXGG, e seu domínio C-terminal apresenta diversas cisteínas em posições conservadas. A classe III possui repetições GXGX e domínios de oleosinas ^(7,8).

É válido destacar as GRPs de classe IV, que são divididas em quatro subclasses: IVa-IVd. Esta classe é formada por GRPs que possuem domínios que interagem com ácidos nucleicos, o que implica em estarem localizadas no ambiente intracelular. Isto condiz com o fato de estas proteínas não possuírem peptídeos sinais que as enderecem para endomembranas. Sabe-se que as proteínas desta classe estão envolvidas em mecanismos de defesa de plantas, agindo em diversas etapas do processamento pós-transcricional da expressão gênica ⁽⁸⁾. A subclasse IVa possui um domínio de reconhecimento de RNA (RRM, *RNA recognition motif*). A subclasse IVb possui, para além de um domínio RRM, um domínio *zinc-finger* do tipo CCHC. A subclasse IVc possui um domínio *cold-shock* (CSD) e dois ou mais domínios *zinc-finger*. A subclasse IVd possui dois ou mais domínios RRM ^(7,8). A proteína que é o foco deste trabalho, AtGRP2, é uma GRP que faz parte da classe IV, subclasse c. Por fim, tem-se a classe V, a qual possui repetições GGX/GXGX ^(7,8).

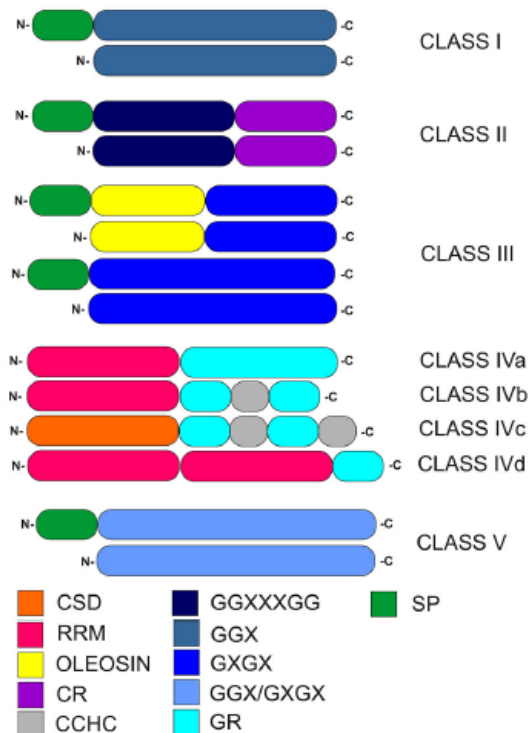


Figura 1: Classificação das proteínas ricas em glicina (GRPs). CSD, domínio *cold-shock*; RRM, domínio de reconhecimento de RNA; *oleosin*, domínio conservado de oleosinas, CR, domínio rico em cisteínas; CCHC, *zinc-finger*; GR, domínio rico em glicinas; GGX, GGXXXGG, GXGX, GGX/GXGX, repetições ricas em glicina, sendo G glicina e X qualquer outro aminoácido; SP, peptídeo sinal. Adaptado de Rurek, M. et al, 2018.

1.3. AtGRP2

A proteína AtGRP2 (do inglês, *Arabidopsis thaliana glycine-rich protein 2*), como já mencionado, é uma GRP da classe IVc. Ela possui 19 kDa, tem teor de glicina de 40% e é composta por um domínio *cold-shock* e dois domínios *zinc-finger* (Figura 2), todos os quais são domínios que se ligam a ácidos nucleicos.

Os domínios *zinc-finger* da AtGRP2 são do tipo CCHC (CX₂CX₄HX₄C, sendo X qualquer aminoácido) retroviral, apresentando grande semelhança com domínios da proteína GAG do HIV⁽⁹⁾. As regiões ricas em glicina desta proteína são encontradas intercaladas com os domínios *zinc-finger*, o que é interessante visto que estas regiões possuem repetições que lembram sequências do tipo RGG-box, as quais já foram encontradas em conjunto com outros domínios que se ligam a ácidos nucleicos em RBPs⁽¹⁰⁾.



Figura 2: Representação esquemática dos domínios da proteína AtGRP2. Esta é composta por um domínio *cold-shock* (CSD) e dois domínios *zinc-finger* (ZKD) intercalados por regiões ricas em glicina.

Esta proteína é encontrada em *Arabidopsis thaliana* como uma dentre quatro GRPs de classe IVc presentes no genoma desta planta (AtGRP1 até AtGRP4) ⁽¹¹⁾, sendo a mais expressa dentre essas ⁽¹²⁾. É codificada pelo gene At4g38680 ⁽¹³⁾. Sua localização é núcleo-citoplasmática, como foi constatado a partir de ensaios de imunolocalização e de sua fusão com GFP ⁽¹⁴⁾, o que é condizente com o esperado para proteínas de sua classe. Além disso, é preferencialmente expressa em tecidos embrionários e com alta taxa de divisão celular, como meristemas, carpelos, anteras e embriões ⁽¹⁴⁻¹⁶⁾.

A partir do silenciamento do gene de AtGRP2 em *Arabidopsis thaliana* mutantes, foram observados os seguintes fenótipos: (i) floração precoce (Figura 3A); (ii) alteração no número de estames; (iii) diminuição da viabilidade das sementes ⁽¹⁴⁾. Em contrapartida, a partir de uma superexpressão ectópica do gene, observou-se: (I) floração tardia (Figura 3B) ⁽¹⁷⁾; (II) retardação na germinação de sementes ⁽¹⁸⁾; (III) diminuição da capacidade de tolerância da planta ao congelamento após aclimatação ao frio ⁽¹⁷⁾. Com as informações destes experimentos, foi concluído que AtGRP2 age como reguladora negativa do tempo de floração de *Arabidopsis thaliana*, estando envolvida em processos de desenvolvimento e atuando na passagem do estado vegetativo para o reprodutor. Ademais, foi confirmado que a atuação de AtGRP2 na resistência ao congelamento em *Arabidopsis thaliana* é como um regulador negativo da expressão de fatores de ligação a repetições ricas em citosina (do inglês, *C-repeat binding factors* – CBFs) ⁽¹⁷⁾. Estes fatores ativam a expressão de genes regulados pelo frio (do inglês, *cold-regulated genes* – COR), os quais geram tolerância ao congelamento ^(19,20). Além disso, plantas mutantes que superexpressavam AtGRP2 apresentaram uma redução na expressão de genes do catabolismo de ácido abscísico (ABA), resultando em níveis significativamente aumentados deste ácido, e genes de biossíntese de giberelina

(18). Foi proposto, então, que AtGRP2 regula negativamente a germinação de sementes a partir de um controle no nível de hormônios, como ABA e giberelina (18).

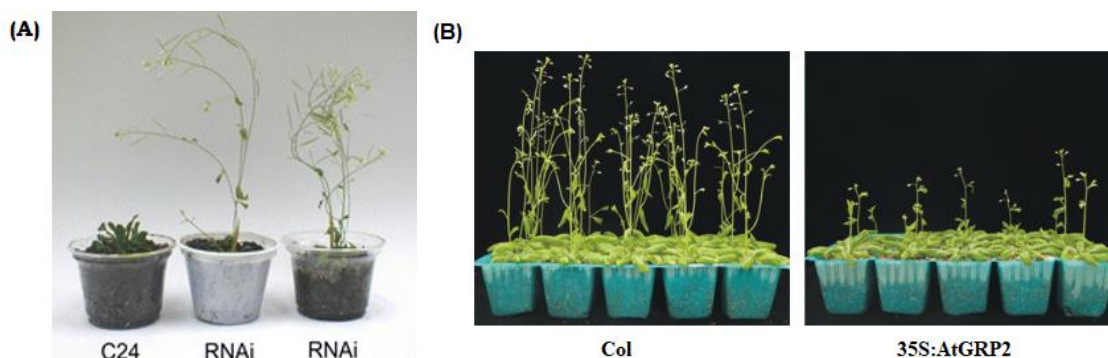


Figura 3: Fenótipo observado em plantas transgênicas de *Arabidopsis thaliana*. (A) Plantas transgênicas onde houve o silenciamento do gene da AtGRP2 (RNAi) comparado com a planta selvagem controle (ecotipo C24). Nota-se uma floração precoce nas plantas transgênicas. Adaptado de Fusaro et al., 2007. **(B)** Plantas transgênicas onde houve a superexpressão ectópica do gene da AtGRP2 (35S:AtGRP2) comparado com a planta selvagem controle (ecotipo Columbia). Nota-se uma floração tardia nas plantas transgênicas. Adaptado de Sasaki et al., 2013.

AtGRP2 é fortemente induzida por baixas temperaturas e estresse salino (14). Neste último caso, a proteína tem efeito negativo sobre a tolerância da planta ao sal, agindo como atenuadora da sinalização celular em resposta a este estresse (21). Ainda, estresse hídrico tem efeito repressor para a proteína (13).

Já foi constatada a afinidade de AtGRP2 por DNA simples fita (ssDNA) e RNA (14), fato também observado para a proteína CspA (*cold-shock protein A*) de *Escherichia coli*, a qual também possui um domínio *cold-shock* e que atua como chaperona de RNA (22). AtGRP2 também promove desnaturação de DNA dupla-fita (23). É esperado, portanto, que AtGRP2 também tenha função de chaperona de RNA.

Esta proteína, como um todo, age como uma reguladora negativa em *Arabidopsis thaliana*, um papel essencial para que haja uma resposta adequada do organismo vegetal a diversos tipos de estresses abióticos, visto que uma superativação de vias de resposta a estes estresses durante o desenvolvimento vegetal poderia ser um empecilho para sua sobrevivência. É interessante, nos dias de hoje, dentro do contexto de escassez de água, aumento da salinidade dos solos, poluição ambiental e aquecimento global, entender o funcionamento da tolerância ao estresse em plantas, principalmente se tratando de respostas a

mudanças no ambiente. Dado este papel de AtGRP2, faz-se interessante o estudo da estrutura tridimensional e o mecanismo de ação desta proteína, ambos os quais são dados ainda não conhecidos.

1.4. Domínio *cold-shock*

Proteínas *cold-shock* recebem este nome por responderem ao estresse pelo frio, e estão dentre algumas das proteínas mais conservadas evolutivamente ⁽²⁴⁾. Foram identificadas pela primeira vez em bactérias, onde uma mudança repentina de temperatura (de 37°C para 10°C) causou um aumento significativo e rápido na expressão da proteína *cold shock A* (CspA) ⁽²⁵⁾. São caracterizadas por possuírem um ou mais domínios *cold-shock* (CSD), o qual é conhecido por se ligar a ácidos nucleicos, o que leva estas proteínas a terem funções relacionadas a regulação da transcrição, tradução e *splicing* ^(25,26).

Domínios *cold-shock* fazem parte da família de proteínas CSDB (do inglês, *cold shock DNA-binding*), a qual, em conjunto com outras 16 famílias, compõem a superfamília NAB (do inglês, *nucleic acid-binding*) ^(26,27). Proteínas e domínios *cold-shock* costumam conter motivos conhecidos de ligação a RNA, RNP-1 e RNP-2 (ribonucleoproteicos 1 e 2) ⁽²⁶⁾, e possuem estrutura tridimensional em forma de um barril- β compacto composto por cinco fitas- β antiparalelas ^(26,27). Exemplos são as proteínas CspA, acima mencionada, e YB-1 CSD (domínio *cold-shock* da proteína ligante de *Y-box*), a qual faz parte de um grupo muito conhecido de proteínas que possuem CSD em humanos, a família *Y-box*. As capacidades de ambas proteínas de se ligarem a ácidos nucleicos já foi relatada ^(28,29). Suas estruturas são apresentadas na Figura 4.

Existe uma conservação na posição de alguns aminoácidos na sequência primária de domínios *cold-shock*, o que é interessante visto que variações no comprimento, na sequência e na conformação, principalmente em regiões de alças e voltas, influenciam a especificidade de ligação com ácidos nucleicos de cada proteína ⁽²⁶⁾. A Figura 4 apresenta a conservação de resíduos por posição na sequência primária de CSDs.

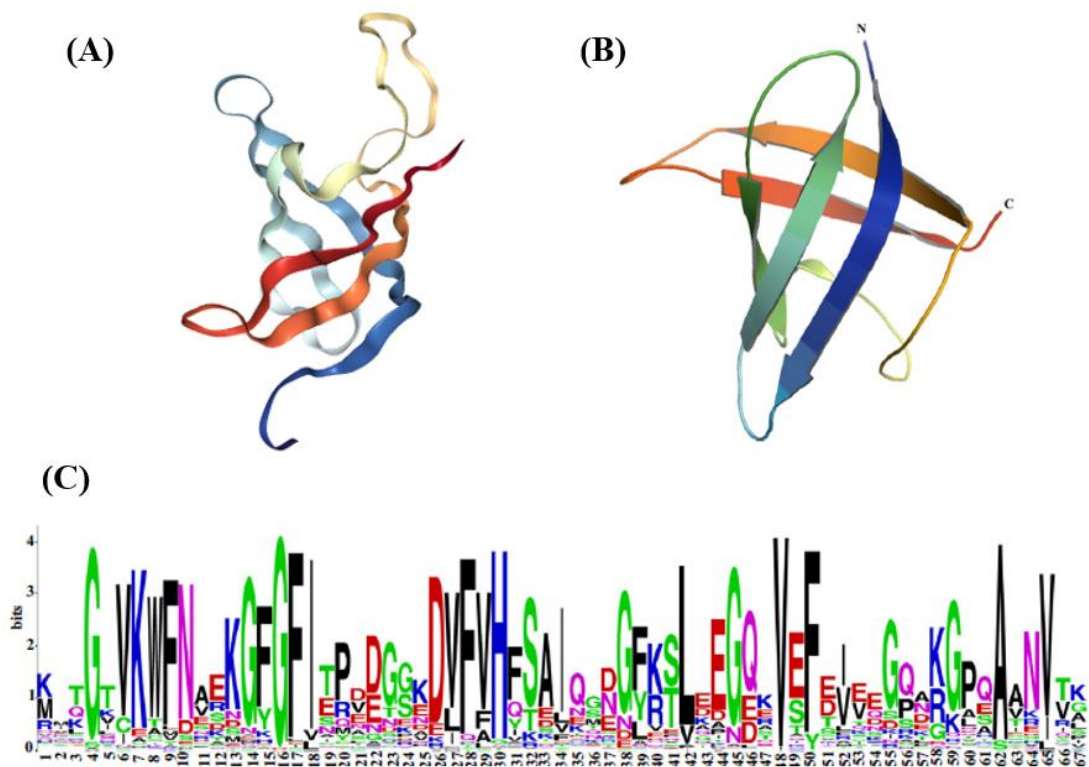


Figura 4: Estrutura tridimensional e conservação na sequência primária de CSDs. Estrutura tridimensional das proteínas, **(A)** YB-1 CSD, código PDB: 1H95 e **(B)** CspA, código PDB: 1MJC. **(C)** Conservação de resíduos por posição na sequência primária de CSDs. A altura de cada letra individual é sua probabilidade de ser a letra daquela posição, enquanto a altura total de cada grupamento vertical de letras representa a quantidade de informação daquela posição em bits. Adaptado de Mohd et al, 2018.

A proteína AtGRP2, como já mencionado anteriormente, possui um domínio *cold-shock*. Seu CSD é composto por 79 aminoácidos (resíduos 1-79) e massa molecular de 8,742 kDa. É esperado que o CSD da AtGRP2 também tenha estrutura em barril- β , apesar de ainda não haverem dados experimentais que confirmem esta informação. Seu PI teórico é de 4,58. Dessa forma, o presente trabalho tem por objetivo a elucidação da estrutura tridimensional do domínio *cold-shock* de AtGRP2 (AtGRP2-CSD) e a caracterização de sua interação com DNA através da Ressonância Magnética Nuclear, a fim de melhor compreender o modo pelo qual esta proteína atua como um regulador negativo de tempo de floração e resposta a estresse abiótico em plantas. Como primeira etapa da estratégia de resolução estrutural por RMN, foi realizado, neste trabalho, o assinalamento sequencial das ressonâncias da cadeia principal de AtGRP2-CSD e a análise de sua interação com o oligonucleotídeo T7.

2. Objetivos

2.1. Objetivo principal

O presente trabalho tem como objetivo principal realizar o assinalamento sequencial das ressonâncias do domínio *cold-shock* da proteína AtGRP2 e caracterizar a interação deste com oligonucleotídeos de DNA por Ressonância Magnética Nuclear.

2.2. Objetivos específicos

- Expressar AtGRP2-CSD recombinante em sistema bacteriano;
- Purificar AtGRP2-CSD por meio de técnicas cromatográficas;
- Coletar experimentos de RMN bi- e tridimensionais;
- Assinalar as ressonâncias da cadeia principal de AtGRP2-CSD;
- Analisar a interação de AtGRP2-CSD com o oligonucleotídeo T7 por meio de RMN;

3. Material e Métodos

3.1. Material

3.1.1. Meio mínimo (M9)

As cepas bacterianas foram cultivadas em meio mínimo de cultura. Este é composto a partir de uma solução de Na₂HPO₄·2H₂O 42,6 g/L, KH₂PO₄ 15 g/L e NaCl 2,5 g/L em pH ajustado para 7,5, e de MgSO₄ 0,25 g/mL, D-glicose 20% m/v, CaCl₂ 0,11 g/mL, Tiamina (Vitamina B1) 10 mg/mL e NH₄Cl 25 g/L. Para produzir amostras isotopicamente enriquecidas com ¹⁵N, o NH₄Cl, única fonte de nitrogênio do M9, foi substituído por ¹⁵NH₄Cl 1 g/L, enquanto que para produção de amostras duplamente enriquecidas com ¹⁵N e ¹³C, também substituiu-se a D-glicose, fonte de carbono do meio, por ¹³C-D-glicose 3 g/L. Todas as soluções utilizadas na composição do meio foram preparadas com água MilliQ e autoclavadas ou esterilizadas com filtro de membrana de poro 0,22 µm. Durante o processo de crescimento, foi acrescentado ao meio canamicina 100 µg/mL, a qual também fora previamente esterilizada.

3.1.2. Meio Luria-Bertani (LB)

O meio Luria Bertani de cultura é composto por 20 g/L do meio Luria-Bertani (LB) (Sigma), o qual contém triptona 10 g/L, extrato de levedura 5 g/L e NaCl 5 g/L. O meio LB sólido, utilizado nas placas de cultivo, é composto acrescentando-se ágar 35 g/L (Sigma).

3.1.3. Vetor de expressão

O vetor de expressão utilizado para expressão do domínio N-terminal da proteína AtGRP2, domínio *cold-shock* (resíduos 1-79), foi o vetor RP1B. Este funde a proteína de interesse a uma cauda N-terminal que contém os primeiros 6 aminoácidos da tioredoxina de *E. coli*, 6 histidinas e um sítio de clivagem para a protease TEV, oriunda do vírus mosaico do tabaco. Além disto, este plasmídeo possui um promotor T7, de origem viral, uma origem de replicação *f1* e o gene que confere resistência à canamicina, *KanR*.

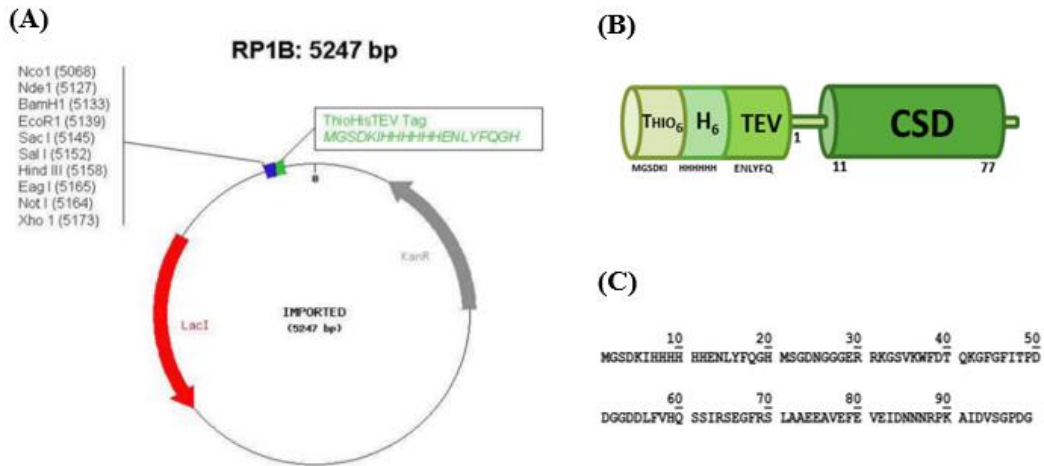


Figura 5: Representação esquemática da construção Thio₆-His₆-TEV-AtGRP2-CSD. (A) Mapa de restrição do vetor RP1B, o qual foi utilizado para expressar o domínio *cold-shock* da AtGRP2. É evidenciado neste o sítio de multi-clonagem, o marcador seletivo e a cauda amino-terminal Thio₆-His₆-TEV. **(B)** Representação esquemática da proteína AtGRP2-CSD com a tag Thio₆-His₆-TEV. **(C)** Sequência primária da proteína AtGRP2-CSD fusionada a cauda Thio₆-His₆-TEV.

A construção da proteína com a cauda N-terminal é composta por 99 aminoácidos e resulta em massa molecular de 10,992 kDa. Após clivagem da cauda N-terminal, a proteína remanesce com 81 aminoácidos e sua massa molecular passa a ser 8,742 kDa.

3.1.4. Cepas bacterianas

Duas cepas bacterianas foram utilizadas neste trabalho. A primeira, *Escherichia coli* BL21 (DE3) [genótipo: F-*ompT hsdSB*(rB-,mB-) *gal dcm* (DE3)], foi utilizada para realizar a expressão da proteína recombinante. Já a segunda, *Escherichia coli* DH5α [genótipo: F-φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17 (rk-,mk+) *phoA supE44 λ-thi-1 gyrA96 relA1*], foi empregada na propagação do plasmídeo recombinante. Ambas as cepas foram gentilmente cedidas pela professora Bianca Cruz Neves do Departamento de Bioquímica do Instituto de Química da Universidade Federal do Rio de Janeiro.

3.2. Métodos

3.2.1. Transformação bacteriana

Para realização da transformação bacteriana, 200 μL de *E. coli* BL21 (DE3) quimicamente competente, previamente congelados a -80°C, foram descongelados em banho de gelo. Foram adicionados 2 μL do estoque de DNA plasmidial recombinante a este volume. As células foram então submetidas a um

choque térmico da seguinte forma: 30 minutos em banho de gelo, incubação de 45 segundos a 42°C e 2 minutos em banho de gelo. Em seguida, 800 µL de meio LB foram acrescentados e nova incubação foi realizada à 37°C e 200 rpm por 1 hora. Decorrido este tempo, as células foram centrifugadas por 10 minutos a 4°C e 8000 rpm por 10 minutos, sendo então 930 µL do sobrenadante descartados. O precipitado de bactérias formado foi ressuspendido no volume remanescente. As células foram plaqueadas em meio LB sólido contendo canamicina 100 mg/mL e incubadas a 37°C por 18 horas. Colônias macroscópicas puderam ser observadas na placa após o término do processo.

3.2.2. Expressão de AtGRP2-CSD

Uma colônia de bactérias transformadas foi transferida para 5 mL de meio LB acrescido de canamicina 100 mg/mL. Uma incubação a 37°C e 200 rpm por 6 horas foi realizada, após a qual transferiu-se 800 µL do cultivo para 80 mL de meio M9 marcado com ¹⁵NH₄Cl ou com ¹⁵NH₄Cl e ¹³C-D-glicose. Este pré-inóculo foi incubado por 18 horas à 37°C e 200 rpm. Decorrido o tempo, 20 mL do cultivo foram transferidos para 300 mL de meio M9, novamente marcado com ¹⁵NH₄Cl ou com ¹⁵NH₄Cl e ¹³C-D-glicose. O inóculo foi realizado em quadruplicata, todos os quais cresceram nas mesmas condições de 37°C e 200 rpm até atingirem densidade ótica a 600 nm (D_{0.600}) de 0,6. Neste ponto, à cada replicata foi acrescentado 300 µL de isopropil-β-D-tiogalactopiranosídeo (IPTG) 1mM para então se realizar uma nova incubação de 30 horas a 18°C e 200 rpm. Por fim, centrifugou-se os cultivos à 4°C e 8000 rpm por 30 minutos, descartou-se o sobrenadante e congelou-se as células à -80°C.

3.2.3. Eletroforese em gel de poliacrilamida sob condições desnaturantes (SDS-PAGE)

A eletroforese SDS-PAGE foi realizada seguindo a técnica descrita por U. K. Laemmli (1970) ⁽³⁰⁾. Para tal, fez-se uso do sistema Mini-PROTEAN Tetra System 20 (BioRad). Foram preparados dois géis: um gel concentrador 4% e um gel fracionador 15%. O primeiro gel, o concentrador, foi composto pela mistura de Tris-HCl 1,0 M (pH 6,8), solução de acrilamida 30% (acrilamida 29%, bis-acrilamida 1%), dodecil sulfato de sódio (SDS) 0,1%, persulfato de amônio (PSA) 0,1% e TEMED 0,01%, sendo estes últimos dois componentes os catalisadores da polimerização e, portanto, necessariamente adicionados após os demais. O

segundo gel, o fracionador, foi preparado pelos mesmos componentes do gel concentrador, fazendo-se uso, contudo, de Tris-HCl 1,5 M (pH 8,8) ao invés de Tris-HCl 1,0 M (pH 6,8). O tampão de corrida adicionado à cuba eletroforética foi composto de Tris-HCl 25 mM (pH 8,3), Glicina 250 mM e SDS 0,1% (p/v). Para o preparo das amostras, acrescentou-se 5 µL de tampão de amostra (Tris-HCl 12,5 mM (pH 6,8), β-mercaptoetanol 25 mM, SDS 0,5% (p/v), azul de bromofenol 0,25%, Glicerol 2,5%) à 15 µL de cada amostra, seguido de fervura por 5 minutos. Foram aplicados 5 µL do padrão de peso molecular (SeeBlue Plus2 Prestained Standard) no primeiro poço de cada gel utilizado e 20 µL das amostras preparadas nos poços seguintes. A corrida foi realizada com 30 mA/gel por um período de 1h e 30 minutos ou até que o *front* da corrida se aproximasse do final do gel. Os géis foram revelados corando-se com Azul de Coomassie R (Vetec).

3.2.4. Lise bacteriana

Os precipitados de bactéria armazenados a -80°C foram descongelados em banho de gelo e ressuspensos em 50 mL de tampão de lise [Tris-HCl 50 mM (pH 8,0), Imidazol 5 mM, NaCl 500 mM e 1 comprimido de coquetel de inibidor de protease (Protease Inhibitor Cocktail Tablets, EDTA-Free (Sigma))]. A lise celular ocorreu através de sonicação realizada em ciclos de 20 segundos com 59 segundos de descanso e 20% de amplitude sob banho de gelo. O lisado obtido foi centrifugado a 8000 rpm à 4°C por 30 minutos, e o sobrenadante filtrado membrana de poro 0,45 nm.

3.2.5. Purificação de AtGRP2-CSD

A purificação da AtGRP2-CSD foi composta por duas etapas de cromatografia de afinidade a níquel seguidas de uma etapa de cromatografia por exclusão molecular. As primeiras duas etapas da purificação foram realizadas no aparelho FPLC Äkta Prime plus (GE Healthcare) com auxílio de uma coluna HisTrap HP (GE Healthcare) acoplada ao sistema. A última etapa fez uso de um aparelho AKTA Explorer (GE Healthcare) e uma coluna Superdex 75 (HiLoad 16/60; GE Healthcare).

Para iniciar o processo de purificação, a coluna cromatográfica foi equilibrada em tampão A [Tris-HCl 50 mM (pH 8,0), Imidazol 5 mM e NaCl 500 mM]. A amostra lisada e filtrada foi injetada na coluna em um fluxo de 1 mL/min de tampão A. As proteínas que não interagiram com a resina de níquel foram

diretamente lavadas e coletadas no *flow through* que perdurou até a estabilização da linha de base da absorbância em 280 nm. Neste momento, o fluxo foi alterado para 3 mL/min para realização da lavagem da coluna. Em seguida, um gradiente de 0 até 100% de tampão B [Tris-HCl 50 mM (pH 8,0), Imidazol 500 mM e NaCl 500 mM], ainda sob mesmo fluxo, foi aplicado a fim de eluir a amostra de interesse. Uma nova lavagem da coluna com tampão Imidazol 2M (pH 8,0) foi realizada até a passagem de cinco vezes o volume da coluna. Cada etapa da purificação foi coletada em tubos cônicos de 50 mL (Falcon), com exceção da eluição com Imidazol, o qual foi coletado em frações de 1 mL, para posterior avaliação do perfil proteico por eletroforese SDS-PAGE.

Após identificação de quais frações do gradiente continham a proteína de interesse, estas foram reunidas em uma membrana de diálise SnakeSkin (3000 Da; Thermo Scientific). A protease TEV (*Tobacco Etch Virus*) foi adicionada a este volume em uma razão 1:10 (TEV:AtGRP2-CSD) em volume. A membrana fechada foi depositada em 2 L de tampão de diálise [Tris-HCl 50 mM (pH 8,0) e NaCl 500 mM] para que houvesse a remoção do Imidazol em excesso na solução e concomitante clivagem da cauda N-terminal Thio₆-His₆-TEV pela protease TEV. Este processo ocorreu a 25°C por 18 horas. Decorrido este tempo, a amostra foi retirada da diálise e filtrada em membrana de poro 0,45 nm.

A segunda etapa da purificação se iniciou com a injeção da amostra à coluna cromatográfica HisTrap HP (GE Healthcare) acoplada ao cromatógrafo e previamente equilibrada em tampão A. Os passos realizados durante esta purificação foram os mesmos descritos para a primeira cromatografia de afinidade a níquel. Todas as etapas foram coletadas e alíquotas de cada etapa foram submetidas a eletroforese SDS-PAGE para investigação do perfil proteico.

As frações contendo a proteína de interesse foram reunidas e concentradas no sistema de filtração CENTRICON (3000 Da; Millipore) até volume final igual ou menor que 5 mL. Os centricons foram centrifugados à 4°C e 3250 rpm em ciclos de 20 minutos, dentre os quais as soluções dentro dos centricons foram ressuspendidas para evitar focos de superconcentração.

Para a última etapa de purificação através da cromatografia por exclusão molecular, a coluna cromatográfica foi previamente equilibrada em tampão [MES 20 mM (pH 6,5), NaCl 50 mM e MgCl₂ 2,5 mM]. A amostra, filtrada em membrana de poro 0,45 nm, foi injetada na coluna em um fluxo de 1 mL/min até a passagem

de 120 mL total. O pico com máximo em 86 mL foi coletado, e uma eletroforese SDS-PAGE foi realizada para verificar o grau de pureza da amostra de AtGRP2-CSD. Por fim, a amostra foi concentrada em centricon até aproximadamente 1 mL, dosada pelo método colorimétrico de Pierce e congelada em -80°C.

3.2.6. Dosagem de proteína pelo Método colorimétrico de Pierce

Foi utilizado o método colorimétrico de Pierce para se determinar a concentração da amostra de proteína. Uma curva padrão com albumina bovina sérica (BSA) de 10 pontos, na qual a concentração de BSA variou uniformemente de 0 µg/mL até 2000 µg/mL, foi utilizada para calibração. Assim, em uma placa de 96 poços, foram adicionados 10 µL de cada ponto da curva e da amostra, e a cada volume foi acrescentado 150 µL do reagente Pierce 660nm Protein Assay (Thermo Scientific). Todos os ensaios foram realizados em triplicata. A placa foi incubada em abrigo da luz por 5 minutos. A leitura das absorvâncias foi realizada a 660 nm em um espectrofotômetro SpectraMax M2e (Thermo Fisher).

3.2.7. Análise por Espectrometria de Massas (MS)

Para verificar o grau de pureza e a massa molecular de AtGRP2-CSD, uma análise por espectrometria de massas foi realizada. Os dados de MS foram adquiridos sob condições desnaturantes em LTQ Velos Orbitrap realizada em uma Orbitrap Elite (Thermo Fisher Scientific) em uma escala de m/z 500 a 2.000 em um valor de 1 milhão de cargas. A proteína foi injetada diretamente. Os padrões de proteína intacta foram analisados com 240.000 de poder de resolução (a m/z 400).

3.2.8. Aquisição dos dados de Ressonância Magnética Nuclear (RMN)

A aquisição dos dados de RMN da proteína AtGRP2-CSD foi realizada em espectrômetro Bruker Avance III HD 900 MHz utilizando uma sonda de tripla ressonância de detecção inversa com gradiente no eixo z TXI HCN (Bruker Biospin). O espectrômetro fica localizado no Centro Nacional de Ressonância Magnética Nuclear (CNRMN; IBqM-UFRJ). O experimento 1D ¹H RMN foi adquirido a 25°C com uma amostra purificada de AtGRP2-CSD com concentração de 100 µM em tampão cacodilato de sódio 20 mM (pH 7,0) e NaCl 100 mM acrescido de 5% D₂O (v/v). Os experimentos de RMN utilizados para o assinalamento sequencial das ressonâncias foram adquiridos a 5°C com uma

amostra purificada de $^{15}\text{N}/^{13}\text{C}$ -AtGRP2-CSD com concentração de 917 μM em tampão MES 20 mM (pH 6,5), NaCl 50 mM e MgCl_2 2,5mM acrescido de 8,3% D_2O (v/v) e 4,2 mM NaN_3 (p/v). A aquisição dos espectros foi feita com o programa Top Spin 3.2 (Bruker Biospin) enquanto o processamento destes foi realizado no programa NMRPipe ⁽³¹⁾.

3.2.9. Assinalamento sequencial das ressonâncias

Para realizar o assinalamento sequencial das ressonâncias da cadeia principal ($^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}^{\alpha}$, $^{13}\text{C}^{\beta}$, $^{13}\text{C}'$) de AtGRP2-CSD, foram utilizados os seguintes experimentos de RMN: 2D [$^1\text{H},^{15}\text{N}$] HSQC (8 acumulações por incremento em t_1 , 2048 pontos coletados na dimensão F_2 e 512 pontos em F_1), 3D HNCA (8 acumulações, 1024 pontos em F_3 , 90 pontos em F_2 e 300 pontos em F_1), 3D HNCO (8 acumulações, 1024 pontos em F_3 , 90 pontos em F_2 e 140 pontos em F_1), CBCA(CO)NH (16 acumulações, 1024 pontos em F_3 , 90 pontos em F_2 e 140 pontos em F_1), 3D HNCACB (32 acumulações, 1024 pontos em F_3 , 90 pontos em F_2 e 140 pontos em F_1) e 3D HNCACO (32 acumulações, 1024 pontos em F_3 , 90 pontos em F_2 e 140 pontos em F_1) ⁽³²⁾. Fez-se uso do programa CARA 1.8.4.2 (<http://www.cara.nmr.ch/doku.php>) para analisar os espectros.

3.2.10. Análise da interação com oligonucleotídeo por Perturbação do Deslocamento Químico (CSP)

Experimentos de titulação por RMN foram realizados com amostras de ^{15}N -AtGRP2-CSD 100 μM em tampão [MES 20 mM (pH 6,5), NaCl 50 mM e MgCl_2 2,5mM acrescido de 10% D_2O (v/v)]. O oligonucleotídeo de DNA sintético utilizado foi T7 (5'-TTTTTTT-3'). Este foi obtido comercialmente da empresa Integrated DNA Technology (IDT, Coraville, EUA) e dissolvido no mesmo tampão da análise. De forma a se obter a superfície de interação entre AtGRP2-CSD e o oligonucleotídeo, espectros [$^1\text{H},^{15}\text{N}$] HSQC foram adquiridos a 25°C em razões molares crescentes de AtGRP2-CSD:T7 (1:0, 1:0,1, 1:0,2, 1:0,3, 1:0,5, 1:0,7, 1:1 e 1:1,5). Estes dados foram obtidos em espectrômetro Bruker Avance III HD 900 MHz com sonda de tripla ressonância de detecção inversa com gradiente no eixo z TXI HCN (Bruker Biospin). O espectrômetro fica localizado no Centro Nacional de Ressonância Magnética Nuclear (CNRMN; IBqM-UFRJ).

Os valores de Perturbação do Deslocamento Químico (do inglês, *Chemical Shift Perturbation* - CSP), o qual indica mudanças no ambiente químico

da proteína por concentração de ligante, foram calculados para as amidas da cadeia principal de acordo com a Equação 1 com auxílio do programa CcpNmr Analysis V2 ⁽³³⁾.

$$\Delta\delta = \left((\Delta\omega_H)^2 + \left(\frac{\Delta\omega_N}{6,66} \right)^2 \right)^{\frac{1}{2}} \quad (1)$$

Onde $\Delta\omega_H$ e $\Delta\omega_N$ se referem às variações no deslocamento químico do $^1\text{H}^N$ e ^{15}N amídico, respectivamente, de um dado sinal.

3.2.11. Modelagem molecular de AtGRP2-CSD

De forma a avaliar a superfície de interação proteína-DNA, um modelo da estrutura tridimensional de AtGRP2-CSD foi obtido a partir de modelagem por homologia estrutural utilizando o programa Phyre2 (Protein Homology/analogy Recognition Engine V2.0). Phyre2 é um servidor online que realiza modelagem, previsão e análise da estrutura, função e mutações de proteínas ⁽³⁴⁾. O modelo da estrutura de AtGRP2-CSD obtido foi analisado com o programa PyMol 2.0 ⁽³⁵⁾.

4. Resultados e Discussão

4.1. Resultados

4.1.1. Expressão e purificação de AtGRP2-CSD

A expressão recombinante de AtGRP2-CSD com a cauda de Thio₆-His₆-TEV foi realizada de acordo com o descrito em 3.2.2. O crescimento dos inóculos foi interrompido ao atingirem densidade ótica a 600 nm (D.O₆₀₀) de 0,6. Neste momento, foi adicionado IPTG, o qual é responsável por induzir a transcrição do gene que codifica a proteína. Este processo foi avaliado por SDS-PAGE e o resultado confirmando a expressão de AtGRP2-CSD pode ser observado na Figura 6.

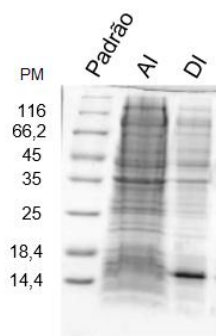


Figura 6: Gel de SDS-PAGE da expressão de AtGRP2-CSD. PM, padrão de massa molecular; AI, antes da indução; DI, depois da indução. É observado o surgimento de uma banda forte entre 14,4 e 18,4 kDa no perfil proteico do extrato celular após a indução.

A purificação de AtGRP2-CSD foi composta por uma combinação de técnicas cromatográficas e uma diálise e clivagem como descrito em 3.2.5. A primeira etapa da purificação consistiu em uma cromatografia por afinidade a níquel; como seu nome indica, esta técnica realiza a separação de substâncias que interagem com o íon Ni²⁺ imobilizado na resina da coluna cromatográfica de substâncias que não interagem. A proteína deste trabalho foi expressa com uma cauda N-terminal que contém seis histidinas, as quais contêm um anel imidazólico, um grupo doador de elétrons, então ocorre coordenação entre estas e o metal de transição. Dessa forma, AtGRP2-CSD foi retida na coluna e separada das demais proteínas no extrato celular. A eluição posterior de AtGRP2-CSD foi dada através de um gradiente contínuo de imidazol, o qual age como um competidor por afinidade, de 5 mM até 500 mM. Os resultados desta cromatografia podem ser observados na Figura 7.

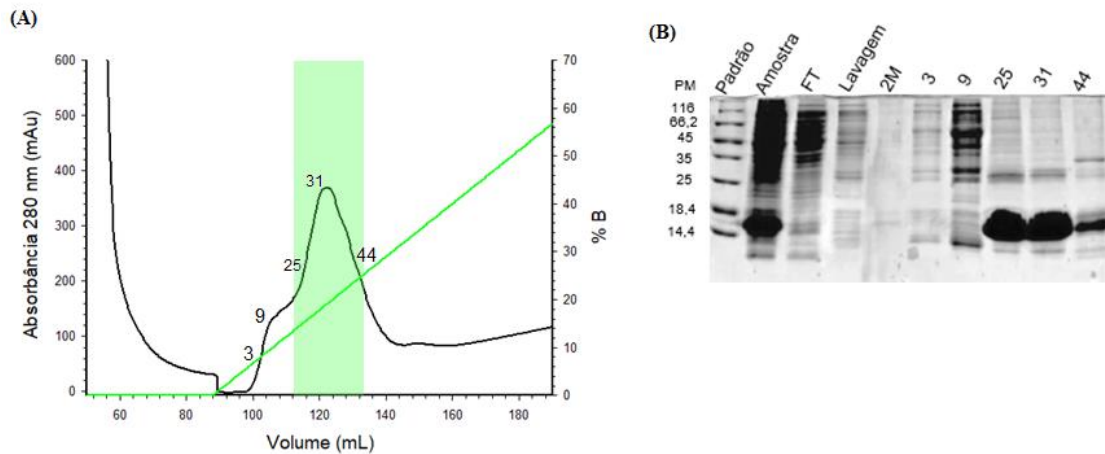


Figura 7: Primeira etapa da purificação de AtGRP2-CSD. (A) Cromatograma da eluição de AtGRP2-CSD, representado pela linha preta, com o gradiente de imidazol, indicado pela linha verde. O pico de eluição da proteína, marcado em verde, ocorreu em ~131 mM Imidazol. **(B)** Gel de SDS-PAGE de frações coletadas de diferentes etapas da cromatografia por afinidade a níquel. PM, padrão de massa molecular (em kDa); FT, *flow through*; Lavagem, lavagem com tampão A; 100%B, final do gradiente, onde se tem 100% do tampão B; 2M, lavagem da coluna com tampão de Imidazol 2M (pH 8,0). Bandas fortes referentes a construção de AtGRP2-CSD podem ser observadas entre 14,4 e 18,4 kDa nos poços Amostra e das frações 25, 31 e 44.

A partir do cromatograma e de uma avaliação do conteúdo proteico por eletroforese, é possível confirmar que a maior parte das proteínas do extrato celular não ficou retida na coluna, como era de se esperar. O pico observado durante o gradiente foi coletado em frações. Verificou-se a presença de uma banda forte entre 14,4 e 18,4 kDa nas frações 25, 31 e 44. Este pico se refere a ~131 mM imidazol. Apesar de a construção Thio₆-His₆-TEV-AtGRP2-CSD possuir 10,991 kDa, esta banda foi entendida como a proteína devido à sua superexpressão comparada com as demais bandas.

Em seguida, foi realizada a clivagem da cauda N-terminal Thio₆-His₆-TEV a partir da incubação da proteína com a protease TEV em razão 1:10 em volume (TEV:AtGRP2-CSD). Esta protease foi escolhida, pois sua atividade não é prejudicada mesmo na presença de diversos inibidores tipicamente adicionados em tampões de purificação e por ser extremamente específica, reconhecendo a sequência EXXYXQ*S/G (sendo X qualquer aminoácido e * a região onde ocorre a clivagem) ⁽³⁶⁾. Esta etapa foi realizada em conjunto com uma diálise para remoção do excesso de imidazol no tampão da amostra. Após a clivagem no sítio específico para TEV na cauda, um resíduo de glicina e um de histidina ainda permanecem com a proteína; por serem somente dois aminoácidos restantes na

região N-terminal da proteína, sua presença não foi um empecilho para as demais análises realizadas com AtGRP2-CSD.

O próximo passo da purificação foi a segunda etapa de cromatografia por afinidade a níquel. Devido à ausência da cauda de poli-histidinas, não ocorre interação significativa entre a proteína e o íon de níquel na resina, o que implica em a proteína passar diretamente pela coluna durante o *flow through*. Nesta etapa, portanto, há a separação da proteína de interesse da cauda Thio₆-His₆-TEV clivada, da protease TEV, a qual é uma construção que possui uma cauda His₆ N-terminal, e de possíveis contaminantes que tenham eluído junto com a proteína durante a primeira cromatografia, visto que estes são todos retidos na coluna. Os resultados desta cromatografia e da clivagem podem ser observados na Figura 8.

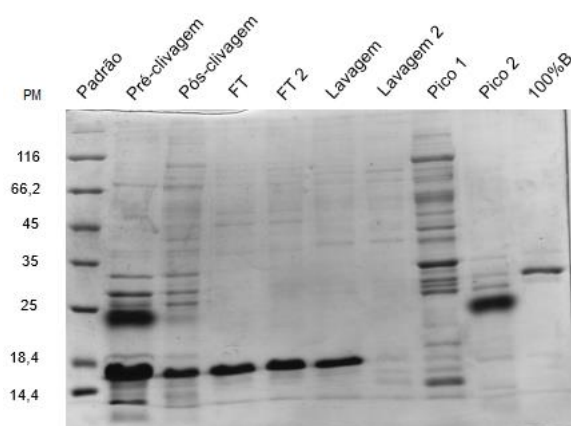


Figura 8: Segunda etapa da purificação de AtGRP2-CSD. Gel de SDS-PAGE de frações coletadas de diferentes etapas da cromatografia por afinidade a níquel. PM, padrão de massa molecular (em kDa); FT 1 e 2, *flow through* 1 e 2; Lavagem 1 e 2, lavagem 1 e 2 com tampão A; 100%B, final do gradiente, onde se tem 100% do tampão B. Bandas fortes referentes a construção de AtGRP2-CSD podem ser observadas entre 14,4 e 18,4 kDa nos poços Pré-clivagem, Pós-clivagem, FT, FT 2 e Lavagem. A banda referente a protease TEV pode ser observada próxima da banda de 25 kDa nos poços Pré-clivagem, Pós-clivagem e Pico 2.

A clivagem se mostra eficiente a partir de uma diminuição na massa molecular da banda referente a proteína no gel de SDS-PAGE devido à perda da cauda N-terminal. Esta variação não é muito aparente no gel devido a elevada concentração da proteína, o que resulta em bandas largas, mas é comprovada com a localização da proteína nas frações coletadas do *flow through* e da lavagem (banda intensa entre 14,4 e 18,4 kDa), além de sua ausência nos picos observados durante o gradiente de imidazol.

A amostra proteica coletada foi concentrada e então submetida à terceira etapa da purificação, uma cromatografia por exclusão molecular, a qual separa substâncias por seu volume molecular. Assim, proteínas maiores eluem antes de proteínas pequenas. Esta etapa de purificação garante um maior grau de pureza da amostra final, além de realizar a troca do tampão da amostra de tampão A (pH 8,0) para o tampão final para as análises com a proteína, tampão MES (pH 6,5). O cromatograma e o gel SDS-PAGE desta etapa podem ser observados na Figura 9.

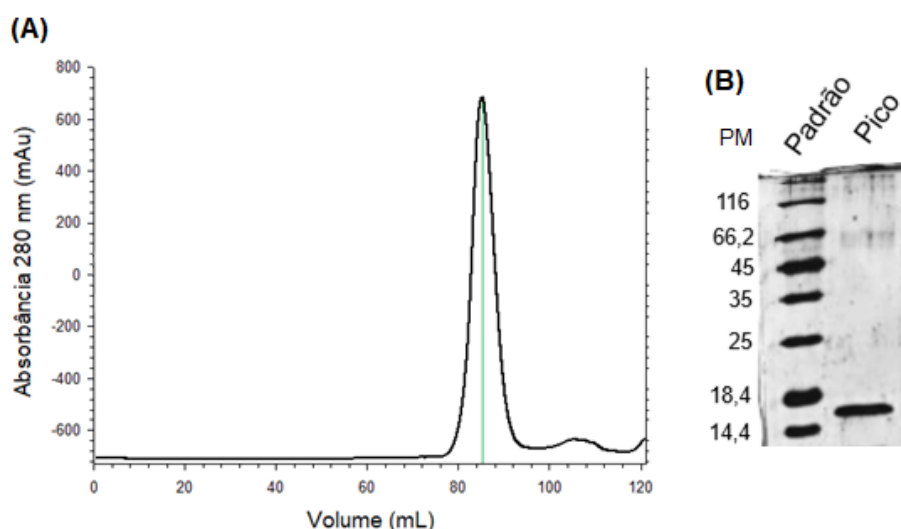


Figura 9: Terceira etapa da purificação de AtGRP2-CSD. (A) Cromatograma de exclusão molecular de AtGRP2-CSD. O pico da proteína, centrado em aproximadamente 86 mL, está marcado em verde. **(B)** Gel de SDS-PAGE do pico observado na cromatografia de exclusão molecular de AtGRP2-CSD. A banda da proteína é observada entre 14,4 e 18,4 kDa. A ausência de outras bandas no gel indica a pureza da amostra final.

Foi observado um único pico de eluição centrado em aproximadamente 86 mL. O gel de eletroforese desta cromatografia apresentou apenas uma banda entre 14,4 e 18,4 kDa. A fim de verificar a identidade e pureza da proteína, um experimento de espectrometria de massas foi realizado (Figura 10). Um único pico correspondente a m/z de 8.741,01 é observado, o que coincide com a massa teórica do domínio *cold-shock* de AtGRP2 de 8.742,40 Da. Desta forma, o protocolo adotado para a expressão e purificação da proteína se mostrou adequado. A discrepância aparente entre a massa da proteína obtida por MS e a indicada pelos géis SDS-PAGE de entre 14,4 e 18,4 kDa, um intervalo correspondente a aproximadamente o dobro de sua massa teórica, é mais explicada adiante.

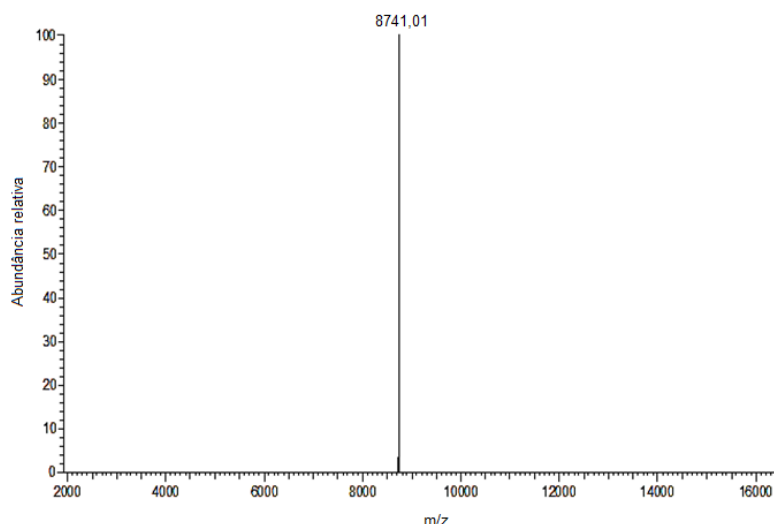


Figura 10: Análise da identidade e pureza de AtGRP2-CSD por espectrometria de massas. O espectro apresenta um único pico relativo a m/z de 8.721,01. Este valor é comparável com a massa molecular teórica do domínio *cold-shock* de AtGRP2 de 8.742,40 Da.

4.1.2. Equilíbrio de conformações enovelado-desenovelado de AtGRP2-CSD

4.1.2.1. 1D ^1H RMN

O espectro unidimensional de RMN de hidrogênio adquirido a partir de uma amostra purificada de AtGRP2-CSD apresenta um perfil compatível com o de uma proteína enovelada (Figura 11), visto que os sinais se apresentam como linhas finas e o deslocamento químico dos hidrogênios amídicos apresenta grande dispersão entre 10-6 ppm. Também é possível observar sinais entre 6-5 ppm, característicos de H^α presentes na estrutura secundária de folha- β , o que indica a existência de uma estruturação na proteína. Um fato importante de se destacar é a presença de dois claros sinais do $\text{H}^{\delta 1}$ indólico do triptofano acima de 10 ppm. Considerando que a sequência primária da AtGRP2-CSD só contém um triptofano, este é um indício de que este aminoácido experimenta mais de uma conformação em solução, as quais encontram-se em troca na escala de tempo de segundos.

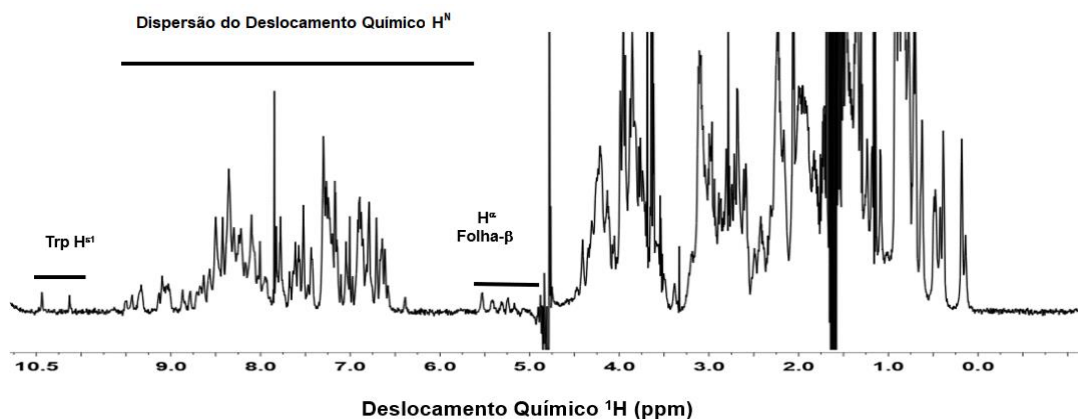


Figura 11: Espectro 1D ^1H de RMN de AtGRP2-CSD. Este foi adquirido a 25°C a partir da proteína purificada com concentração de $100\ \mu\text{M}$ em tampão cacodilato de sódio $20\ \text{mM}$ ($\text{pH}\ 7,0$) e $\text{NaCl}\ 100\text{mM}$ acrescido de 5% D_2O (v/v). Há uma grande dispersão dos deslocamentos químicos bem como a presença de sinais referentes a estrutura secundária de folha- β , indicando que a proteína se encontrava enovelada em solução.

4.1.2.2. [$^1\text{H},^{15}\text{N}$] HSQC

O espectro [$^1\text{H},^{15}\text{N}$] HSQC é um espectro de RMN em 2D de correlação heteronuclear (do inglês, *Heteronuclear Single Quantum Coherence*) onde se observa um pico de correlação para cada hidrogênio covalentemente ligado a nitrogênio na proteína. Assim, tem-se um sinal para cada NH da cadeia principal, ou seja, um sinal para cada aminoácido presente na cadeia polipeptídica com exceção das prolinas, as quais não possuem hidrogênio amídico. Além disso, picos de correlação para os grupos NH_2 presentes nas cadeias laterais também podem ser observados, mas devido à janela espectral na qual esse experimento costuma ser adquirido, somente os sinais referentes aos NH_2 das cadeias laterais de asparaginas e glutaminas, que aparecem em torno de $113\ \text{ppm}$, são de fato visualizados. Este tipo de espectro é único para cada proteína, visto que o deslocamento químico de cada pico de correlação da cadeia principal depende, dentre outros fatores, da estrutura tridimensional da proteína, do ambiente químico ao redor de cada núcleo.

A figura 12 apresenta o espectro [$^1\text{H},^{15}\text{N}$] HSQC em alta resolução de AtGRP2-CSD purificada.

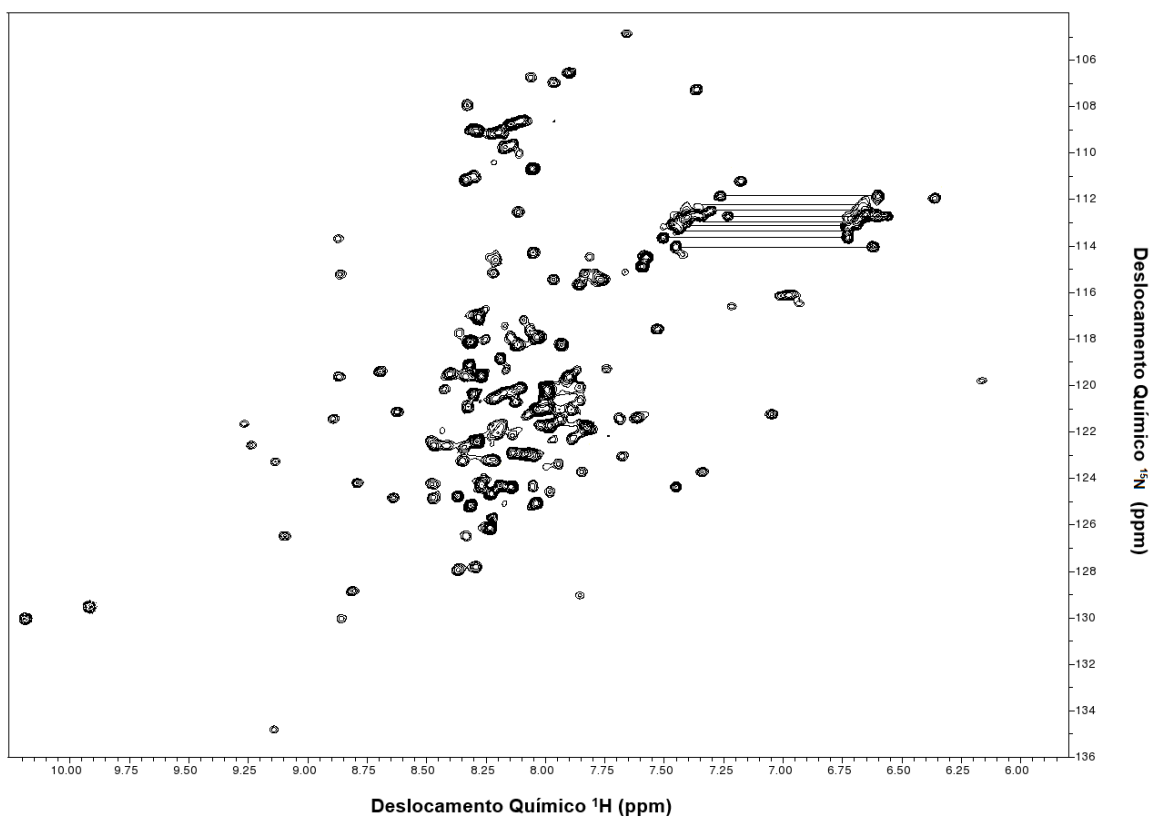


Figura 12: Espectro $[^1\text{H},^{15}\text{N}]$ HSQC de AtGRP2-CSD. Este espectro foi adquirido a 5°C a partir da proteína purificada com concentração de $917\ \mu\text{M}$ em tampão MES $20\ \text{mM}$ (pH 6,5), NaCl $50\ \text{mM}$ e MgCl_2 $2,5\ \text{mM}$ acrescido de $8,3\%$ D_2O (v/v) e $4,2\ \text{mM}$ NaN_3 (p/v). O aglomerado de sinais em torno de $8\ \text{ppm}$ é característico de proteínas desenoveladas, enquanto que os demais sinais bem distribuídos pelo espectro indicam que o domínio da proteína se encontra enovelado, o que implica na existência de duas conformações em equilíbrio de troca. As linhas traçadas indicam os grupamentos NH_2 da cadeia lateral de asparaginas e glutaminas.

O $[^1\text{H},^{15}\text{N}]$ HSQC de AtGRP2-CSD apresenta ressonâncias com linhas finas e boa dispersão no deslocamento químico, características esperadas no espectro de uma proteína bem enovelada em solução. Ademais, algumas informações interessantes podem ser obtidas a partir do espectro desta proteína. Primeiramente, foi identificado um número de sinais muito acima do esperado. Deveriam ser visualizadas 78 ressonâncias amídicas, visto que a construção do domínio *cold-shock* de AtGRP2 possui 79 aminoácidos, sendo 3 destes prolinas, além dos 2 aminoácidos remanescentes da cauda Thio₆-His₆-TEV, mas foram encontrados 166 sinais. A contabilização de mais do que o dobro de sinais esperados sugere que há mais de uma conformação da proteína em equilíbrio em solução. Outro ponto a se destacar é a presença de diversos sinais mais intensos e com baixa dispersão na dimensão do ^1H ; estes estão aglomerados em torno de $8,2\ \text{ppm}$. Sinais com estas características são particulares de

proteínas ou regiões proteicas desenoveladas. Como há uma grande quantidade destes sinais, isto sugere que a outra conformação de AtGRP2-CSD é uma forma desenovelada. É importante ressaltar neste ponto que, havendo duas conformações para a construção inteira da proteína, ainda existem 12 sinais extras que também se encontram aglomerados na região central do espectro. Isto é um indício de que alguns aminoácidos experimentam uma terceira conformação, a qual também é desenovelada. Ademais, é notável a presença de dois sinais na região onde se encontram ressonâncias provenientes de NH indólicos da cadeia lateral de triptofanos, em torno de 10 ppm, sendo que há somente um único triptofano na sequência primária desta proteína. Este fato também foi constatado no espectro 1D ^1H de AtGRP2-CSD.

Em conjunto, estas informações levam a conclusão de que toda a proteína AtGRP2-CSD se encontra em equilíbrio entre conformações, sendo uma enovelada e outra desenovelada, com alguns aminoácidos possuindo uma terceira conformação também desenovelada. Como é possível visualizar este grande grupo de ressonâncias extras no espectro junto com os sinais dispersos da forma enovelada, tem-se que este equilíbrio de troca ocorre em regime lento na escala de tempo do RMN (segundos).

4.1.2.3. Estratégia de assinalamento das ressonâncias

Os espectros 3D de RMN são espectros de tripla ressonância que correlacionam os núcleos ^1H , ^{15}N e ^{13}C . Assim, para além das dimensões de ^1H e ^{15}N observados em um HSQC, há uma terceira dimensão explicitando os deslocamentos químicos do ^{13}C . Considerando-se uma representação de três eixos cartesianos, para cada coordenada no plano $^1\text{H}^{15}\text{N}$ que contém um sinal, tem-se uma ou mais coordenadas correspondentes no eixo ^{13}C . Em outras palavras, para cada sinal NH observado no $[^1\text{H},^{15}\text{N}]$ HSQC, tem-se valores correspondentes ao deslocamento químico do carbono da carbonila, carbono alfa, beta, gama, etc ($^{13}\text{C}'$, $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$, $^{13}\text{C}^\gamma$, etc). O carbono que é detectado no espectro depende do experimento 3D realizado. Os nomes dos experimentos 3D e os sinais de ^{13}C observados neles estão descritos na Tabela 1, enquanto a Figura 13 mostra uma representação esquemática da relação entre as ressonâncias de um $[^1\text{H},^{15}\text{N}]$ HSQC e de um HNCO.

Tabela 1: Experimentos de RMN em 3D utilizados para o assinalamento de ressonâncias da cadeia principal de proteínas. A denominação (i) indica se tratar de uma correlação intra-resíduo, enquanto a (i-1) indica uma correlação inter-resíduo com o aminoácido predecessor àquele sistema.

Experimentos de RMN em 3D	Sinais observados no espectro de RMN
HNCACB	$^{13}\text{C}^\alpha (i-1)$, $^{13}\text{C}^\beta (i-1)$, $^{13}\text{C}^\alpha (i)$ e $^{13}\text{C}^\beta (i)$
HNCA	$^{13}\text{C}^\alpha (i-1)$ e $^{13}\text{C}^\alpha (i)$
CBCA(CO)NH	$^{13}\text{C}^\alpha (i-1)$ e $^{13}\text{C}^\beta (i-1)$.
HNCO	$^{13}\text{C}' (i-1)$.
HNCACO	$^{13}\text{C}' (i-1)$ e $^{13}\text{C}' (i)$
CC(CO)NH	$^{13}\text{C}^\alpha (i-1)$, $^{13}\text{C}^\beta (i-1)$, $^{13}\text{C}^\gamma (i-1)$, etc

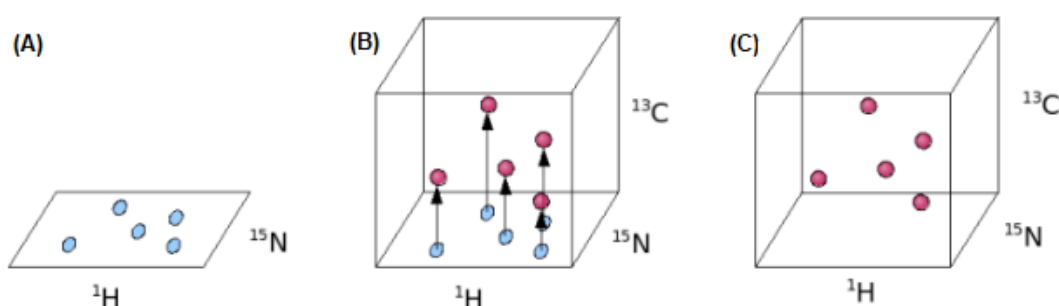


Figura 13: Representação esquemática da transição visual de experimentos de RMN em duas dimensões para três dimensões. (A) Espectro $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ HSQC. (B) Extensão dos sinais do $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ HSQC para incluir a terceira dimensão referente às frequências do ^{13}C . (C) Espectro 3D HNCO, onde se tem as frequências do grupo NH correlacionadas às frequências de $^{13}\text{C}' (i-1)$. Adaptado de Higman (2019).

Em experimentos de RMN, cada aminoácido é reconhecido como um sistema de *spins* nucleares, os quais são conectados entre si por acoplamento escalar e que são associados a um deslocamento químico a partir de experimentos 2D e 3D. Visto que há uma ligação peptídica unindo os aminoácidos de uma proteína em uma sequência, é possível identificar cada sistema a partir de uma correlação através destas ligações. O assinalamento sequencial das ressonâncias da cadeia principal de uma proteína por RMN é, portanto, esta identificação dos aminoácidos de forma ordenada ⁽³⁷⁾, sendo baseado no conjunto de informação obtidas e cruzadas a partir de espectros em 3D (HNCO, HNCACO, HNCA, HNCACB e CBCA(CO)NH, sendo os principais nesta tarefa o HNCACB e CBCA(CO)NH) ⁽³⁸⁾, para além do $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ HSQC ⁽³⁹⁾.

O espectro $[\text{}^1\text{H}, \text{}^{15}\text{N}]$ HSQC oferece diretamente os deslocamentos químicos dos núcleos de hidrogênio ($^1\text{H}^N$) e nitrogênio (^{15}N) amídicos de cada resíduo. Nota-se que neste caso, os sinais observados são somente de uma

correlação intra-resíduo (i). Para espectros 3D, sinais inter-resíduos (i-1) também são observados, sendo sempre uma relação com algum *spin* nuclear do aminoácido anterior àquele sistema na sequência primária. Os deslocamentos químicos dos spins $^{13}\text{C}^\alpha$ e $^{13}\text{C}^\beta$ são identificados a partir dos espectros HNCACB e CBCA(CO)NH, que devem ser analisados juntos. O primeiro apresenta quatro sinais, dois dos quais são referentes ao $^{13}\text{C}^\alpha$ e $^{13}\text{C}^\beta$ do próprio resíduo e são mais intensos, enquanto os outros dois são referentes ao $^{13}\text{C}^{\alpha-1}$ e $^{13}\text{C}^{\beta-1}$, ou seja, $^{13}\text{C}^\alpha$ e $^{13}\text{C}^\beta$ do aminoácido anterior, e são mais fracos. O espectro CBCA(CO)NH apresenta somente dois sinais, $^{13}\text{C}^{\alpha-1}$ e $^{13}\text{C}^{\beta-1}$, os quais têm deslocamento químico igual ao dos sinais de mesmo nome presentes no HNCACB mas os quais aparecem como picos fortes neste espectro, e auxilia na distinção dos sinais do HNCACB. Por vezes, o espectro HNCA também é utilizado. Nele, observam-se sinais para $^{13}\text{C}^\alpha$ e $^{13}\text{C}^{\alpha-1}$, mas por ser um experimento de maior sensibilidade, a diferença de intensidade entre os picos i e i-1 fica mais explícita, sendo outra fonte para resolver possíveis ambiguidades no HNCACB. A Figura 14 mostra como este processo de relacionar um grupo NH aos demais *spins* nucleares funciona a partir de uma cadeia polipeptídica.

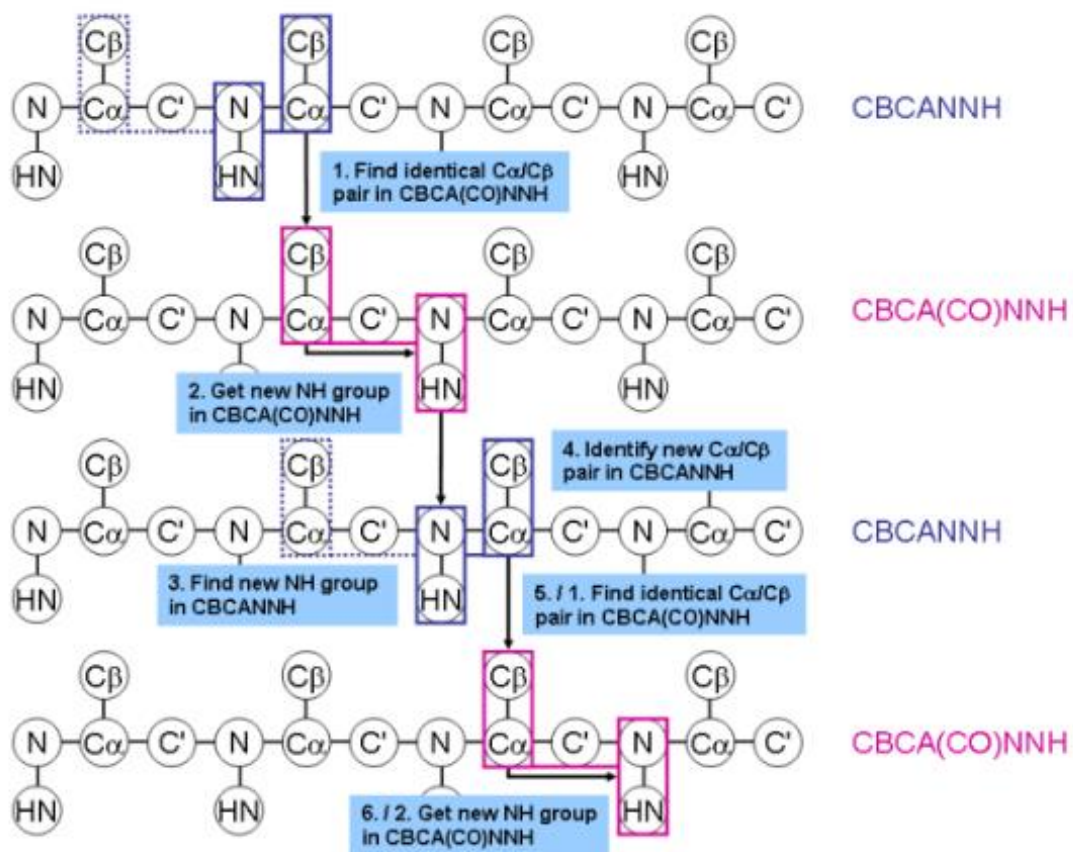


Figura 14: Representação em uma cadeia polipeptídica da correlação entre sistemas de *spin* nucleares realizada a partir de experimentos de RMN. O espectro HNCACB correlaciona um grupo NH com o deslocamento químico do $^{13}\text{C}^\alpha$ e $^{13}\text{C}^\beta$ do próprio resíduo (sinais fortes) e do resíduo anterior (sinais fracos). O espectro CBCA(CO)NH somente correlaciona o grupo NH com o $^{13}\text{C}^\alpha$ e $^{13}\text{C}^\beta$ do resíduo anterior (sinais fortes). Adaptado de Higman (2019).

Os espectros HNC O e HNCAC O também são analisados em conjunto. O HNCAC O apresenta sinais dos *spins* da carbonila do próprio resíduo, $^{13}\text{C}'$, e do resíduo anterior, $^{13}\text{C}'-1$, dentre os quais este segundo é mais fraco. No HNC O , somente um sinal é visualizado, o $^{13}\text{C}'-1$, que aparece como um pico forte no espectro e deve ter o mesmo deslocamento químico do seu sinal correspondente no HNCAC O .

Com todos os deslocamentos químicos das ressonâncias da cadeia principal identificados, é possível começar o assinalamento sequencial. Para tal, faz-se uso dos valores dos deslocamentos químicos de cada *spin* nuclear de cada aminoácido em conformação randômica, os quais estão tabelados. Como alguns aminoácidos possuem deslocamento químico para $^{13}\text{C}^\alpha$ e/ou $^{13}\text{C}^\beta$ muito característicos, esta informação é muito útil na identificação dos resíduos. As

conexões i e $i-1$ obtidas pelos espectros para cada sistema e a sequência primária da proteína também são fundamentais nesta análise.

Alguns exemplos de aminoácidos com deslocamentos químicos distintos são: glicina, alanina, serina e treonina. A glicina mais se destaca, visto que não possui $^{13}\text{C}^\beta$ e seu $^{13}\text{C}^\alpha$ aparece caracteristicamente em torno de 45 ppm. Alaninas possuem $^{13}\text{C}^\beta$ com deslocamento químico muito mais baixo do que os outros aminoácidos, em torno de 18 ppm. Em contrapartida, serinas e treoninas são os únicos que possuem $^{13}\text{C}^\beta$ com deslocamento químico maior que seu $^{13}\text{C}^\alpha$. Estes fatos tornam fácil o reconhecimento de sistemas como algum destes quatro aminoácidos ou como tendo algum deles como predecessor.

Parte-se então de algum sistema que seja distinguível como um destes quatro aminoácidos distintos. O programa CARA 1.8.4.2 foi utilizado para procurar possíveis sistemas de *spin* predecessores àquele escolhido. Para tal, o programa procura por conectividades sequenciais a partir dos deslocamentos químicos do $^{13}\text{C}^{\alpha-1}$, $^{13}\text{C}^{\beta-1}$ e $^{13}\text{C}'-1$ do sistema com os $^{13}\text{C}^\alpha$, $^{13}\text{C}^\beta$ e $^{13}\text{C}'$ dos demais sistemas. Quando há ambiguidade decorrente de valores de $^{13}\text{C}^\alpha$ e $^{13}\text{C}^\beta$ similares, visto que alguns aminoácidos possuem valores parecidos para estes núcleos, a incerteza pode ser resolvida a partir do dado do $^{13}\text{C}'/^{13}\text{C}'-1$, visto que os deslocamentos químicos das carbonilas diferem muito mais entre si. Em alguns casos, e quando disponível, também pode ser feito o uso do espectro CC(CO)NH, o qual apresenta sinais dos carbonos da cadeia lateral $i-1$; como o perfil observado é muito característico para alguns aminoácidos, ele pode ser útil para discernir a relação correta entre sistemas. Os aminoácidos podem então ser assinalados a partir da união das informações da sequência primária da proteína e do padrão das conectividades de diversos sistemas de *spin* consecutivos.

A Figura 15 mostra fatias do espectro HNCACB adquirido para AtGRP2-CSD, onde a conexão entre ressonâncias de sistemas sequenciais e o nome e a posição na sequência primária do aminoácido assinalado para cada sistema estão explicitados.

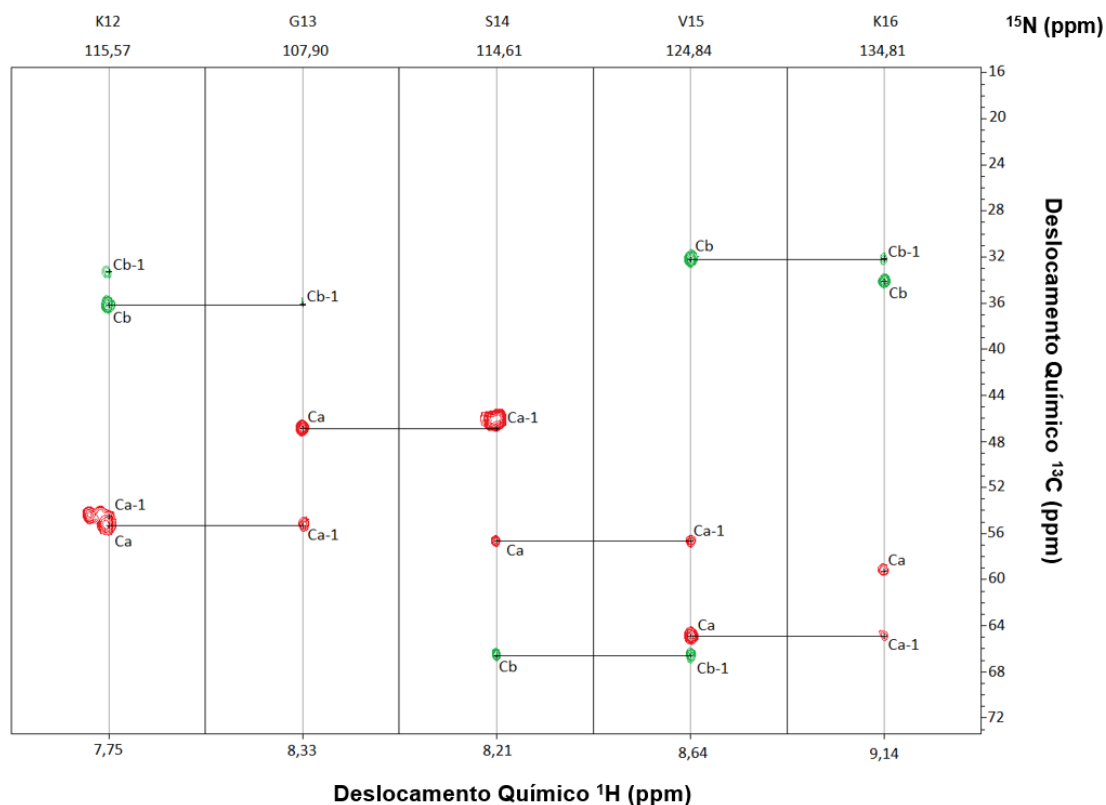


Figura 15: Fatias do espectro 3D HNCACB adquirido de AtGRP2-CSD. As linhas traçadas mostram a conectividade sequencial entre os spins $^{13}\text{C}^{\alpha}/^{13}\text{C}^{\alpha-1}$ e $^{13}\text{C}^{\beta}/^{13}\text{C}^{\beta-1}$ de cada sistema. A abreviatura dos aminoácidos e suas posições na sequência primária estão apresentados no topo da figura acima do sistema para o qual foram assinalados.

4.1.2.4 Assinalamento da cadeia principal das formas enovelada e desenovelada de AtGRP2-CSD

A partir da estratégia de assinalamento sequencial empregada neste trabalho, foi possível assinalar 95% das ressonâncias da forma enovelada de AtGRP2-CSD. Dos 78 aminoácidos da construção do domínio *cold-shock*, foram identificados 74; os aminoácidos não encontrados são os 2 artifícios de clonagem, a M1 e a S41. O GH remanescente da cauda N-terminal e a M1 não puderam ser assinalados devido à grande sobreposição de sinais no espectro $[^1\text{H},^{15}\text{N}]$ HSQC da proteína. O assinalamento da S41 foi particularmente impossibilitado devido ao desaparecimento do seu sinal; este caso é mais esclarecido no próximo tópico. O espectro $[^1\text{H},^{15}\text{N}]$ HSQC com a identificação das ressonâncias da forma enovelada de AtGRP2-CSD está apresentado na Figura 16.

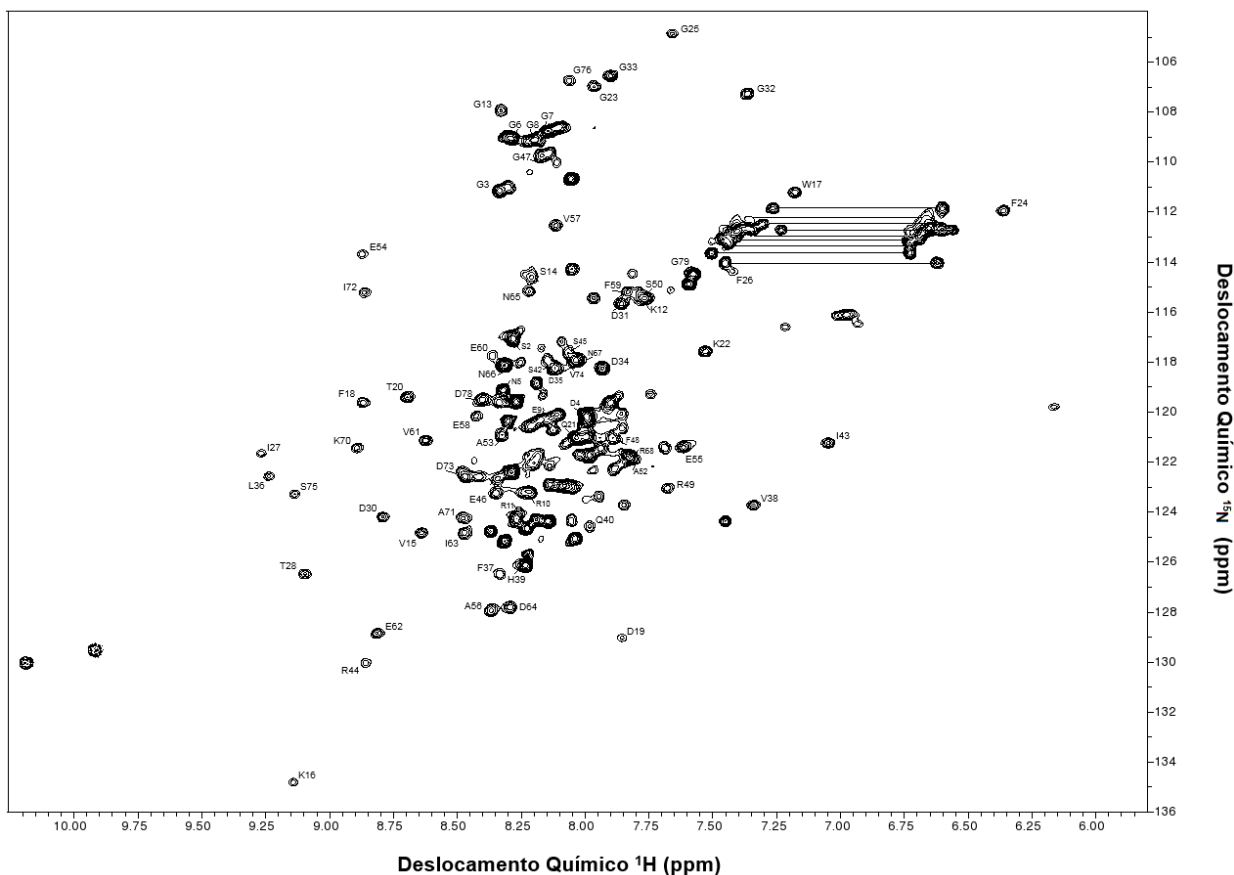


Figura 16: Espectro $[^1\text{H},^{15}\text{N}]$ HSQC de AtGRP2-CSD 95% assinalado, conformação enovelada. Este espectro foi adquirido a 5°C a partir da proteína purificada com concentração de 917 μM em tampão MES 20 mM (pH 6,5), NaCl 50 mM e MgCl_2 2,5mM acrescido de 8,3% D_2O (v/v) e 4,2 mM NaN_3 (p/v). As linhas traçadas indicam os grupamentos NH_2 da cadeia lateral de asparaginas e glutaminas.

Utilizando a mesma estratégia de assinalamento, foi possível assinalar 68% dos sinais da forma desenovelada de AtGRP2-CSD. Dos 78 aminoácidos passíveis de serem encontrados, foi confirmada a localização de 53. Os aminoácidos não encontrados são: G-1, H0, M1, S2, N5, R10, S14, V15, K16, W17, F18, D19, T20, Q21, K22, G23, F24, D30, D31, H39, Q40, S41, L51, A52 e N65. Os sinais da forma desenovelada, como já mencionado anteriormente, são sinais com baixa dispersão na dimensão do ^1H . Por estarem presentes em grande quantidade neste espectro, há uma considerável sobreposição de diversas ressonâncias no centro do espectro. Este fato dificultou o assinalamento não ambíguo destes aminoácidos. O $[^1\text{H},^{15}\text{N}]$ HSQC com os sinais da forma desenovelada de AtGRP2-CSD identificados está apresentado na Figura 17.

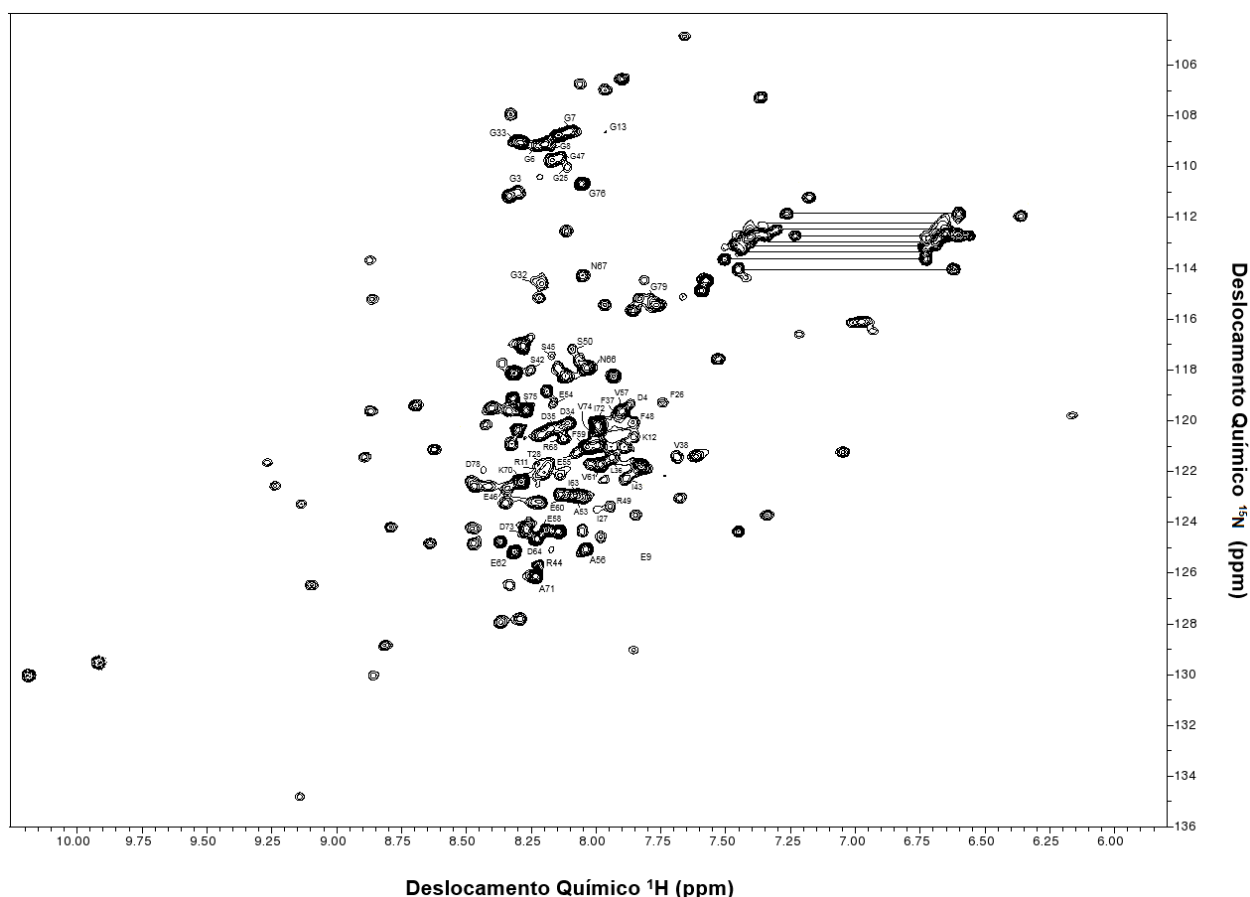


Figura 17: Espectro $[^1\text{H},^{15}\text{N}]$ HSQC de AtGRP2-CSD 68% assinalado, conformação desenovelada. Este espectro foi adquirido a 5°C a partir da proteína purificada com concentração de $917\ \mu\text{M}$ em tampão MES $20\ \text{mM}$ (pH 6,5), NaCl $50\ \text{mM}$ e MgCl_2 $2,5\ \text{mM}$ acrescido de $8,3\%$ D_2O (v/v) e $4,2\ \text{mM}$ NaN_3 (p/v). As linhas traçadas indicam os grupamentos NH_2 da cadeia lateral de asparaginas e glutaminas.

De forma a prosseguir com a resolução da estrutura tridimensional desta proteína, a próxima etapa a ser realizada é o assinalamento das ressonâncias das cadeias laterais dos aminoácidos.

4.1.3. Análise da interação de AtGRP2-CSD com oligonucleotídeo de DNA

Domínios *cold-shock* são conhecidos na literatura por se ligarem a ácidos nucleicos (25,26,28,29,40,41). Experimentos de espectroscopia de fluorescência realizados previamente pelo nosso laboratório confirmaram a interação de AtGRP2-CSD com oligonucleotídeos de DNA e investigaram a sequência com maior especificidade de ligação (dados não mostrados). Dessa forma, este trabalho prosseguiu com a investigação da interação de AtGRP2-CSD com ácido nucleico através da técnica de Ressonância Magnética Nuclear. Para tal,

utilizou-se o oligonucleotídeo T7, o qual apresentou menor constante de dissociação [$K_d = (4,70 \pm 1,90) \mu\text{M}$].

Espectros [$^1\text{H}, ^{15}\text{N}$] HSQC de AtGRP2-CSD foram adquiridos com a proteína livre e na presença do oligonucleotídeo T7 nas seguintes razões molares (proteína:ligante): 1:0, 1:0,1, 1:0,2, 1:0,3, 1:0,5, 1:0,7, 1:1 e 1:1,5. A sobreposição dos espectros está apresentada na Figura 18. De forma a analisar a interação entre a proteína e o ligante, o assinalamento realizado a 5°C foi transferido para o espectro a 25°C através de uma curva de temperatura de 5 até 25°C. A titulação foi realizada até constatação da saturação da interação.

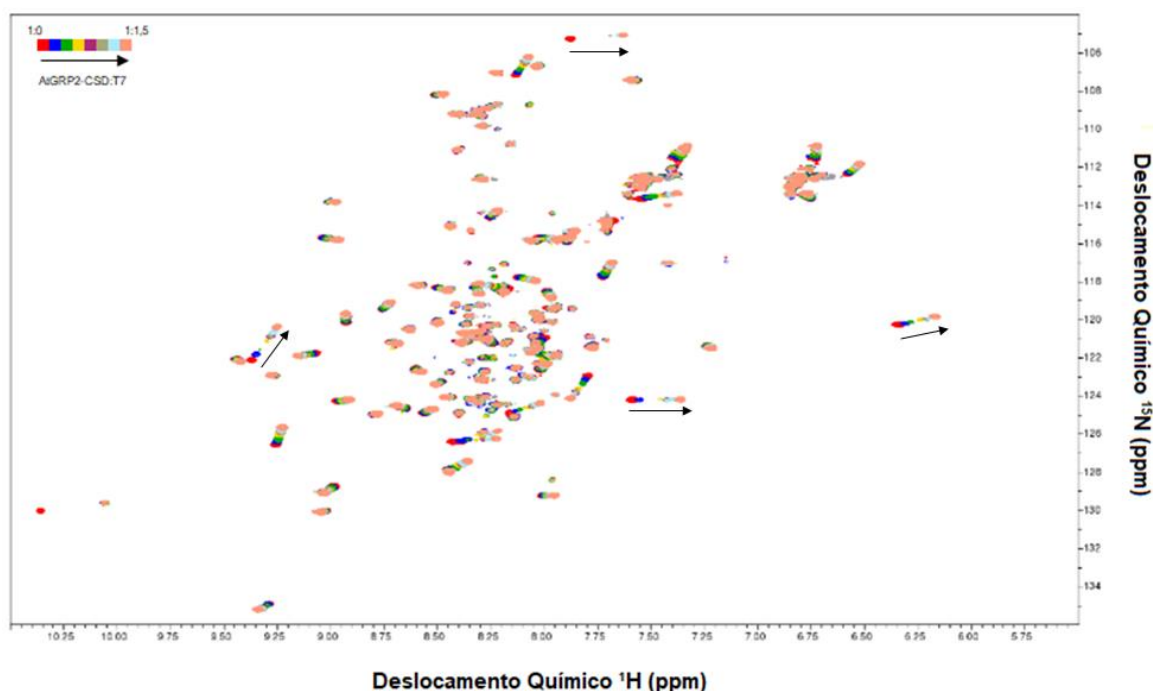


Figura 18: Sobreposição de espectros [$^1\text{H}, ^{15}\text{N}$] HSQC de AtGRP2-CSD em razões molares crescentes de T7. As razões molares de AtGRP2-CSD:T7 utilizadas foram: 1:0 (vermelho), 1:0,1 (azul escuro), 1:0,2 (verde escuro), 1:0,3 (amarelo), 1:0,5 (rosa escuro), 1:0,7 (verde oliva), 1:1 (azul claro) e 1:1,5 (salmão). Os espectros foram adquiridos a 25°C com a proteína purificada com concentração de 100 μM em tampão MES 20 mM (pH 6,5), NaCl 50 mM e MgCl_2 2,5 mM acrescido de D_2O 5% (v/v). As setas em preto indicam o deslocamento de alguns sinais.

Variações nos deslocamentos químicos dos sinais, em uma titulação por RMN, reportam alterações no ambiente químico ao redor de resíduos. Observando a Figura 18, é nítido que há um deslocamento de diversas ressonâncias com o aumento da razão molar entre a proteína e o ligante, o que confirma, como esperado, que há interação entre o domínio *cold-shock* de AtGRP2 e o oligonucleotídeo T7. Este deslocamento contínuo de algumas

ressonâncias com o acréscimo de ligante no meio também indica que a cinética de troca entre a forma livre da proteína e a complexada ocorre em regime rápido na escala do RMN (pico a nano segundos). Neste ponto, é importante destacar o fato de que somente sinais referentes a forma enovelada sofreram alterações, indicando que esta é a conformação responsável pela interação com ácido nucleico.

Foi constatado o desaparecimento e posterior reaparecimento das ressonâncias dos resíduos G25 e V38 ao longo do experimento. No caso de outro resíduo, a L36, houve uma variação acentuada em sua intensidade em alguns pontos da titulação apesar de seu sinal continuar sendo localizado. Interessantemente, o sinal da S41, aminoácido não assinalado, pôde ser identificado no espectro após saturação da ligação. Isto sugere que este resíduo está em regime de troca intermediária (microsegundos na escala do RMN) na forma livre da proteína, o que resulta no desaparecimento do sinal, e que alguma interação ocorrendo na presença do ligante acelerou a cinética da troca, atingindo regime rápido e revelando o sinal. Além disso, a titulação apresentou efeito sobre somente um dos dois sinais referentes ao NH indólico do triptofano (localizado em aproximadamente 10,35 ppm), o qual apresentou deslocamento e desaparecimento com o aumento da razão molar do ligante. Isto sugere que este é o sinal originado da forma enovelada.

A Perturbação do Deslocamento Químico (CSP; do inglês, *Chemical Shift Perturbation*) foi medida em relação à proteína livre para todos os resíduos da forma enovelada em razão molar 1:1,5 AtGRP2-CSD:T7. A superfície de interação foi identificada a partir de valores de CSP maiores que 2σ e 3σ (Figura 19). A modelagem da estrutura tridimensional de AtGRP2-CSD foi realizada através do software Phyre2 de modo a facilitar a análise destes resultados. O modelo foi visualizado com o programa PyMol 2.0.

Os resíduos identificados como presentes no sítio de interação são: G25, F26, L36 e V38 (3σ) e G23, T28, F37, H39, Q40, R49 e L51 (2σ). Os quatro aminoácidos com os maiores valores de CSP compõem o sítio primário de interação com ácido nucleico de AtGRP2-CSD e estão localizados na região da fita β_2 (G25 e F26) e β_3 (L36 e V38). Os demais resíduos com valores de CSP entre 2σ e 3σ podem estar envolvidos em contatos diretos com o ligante ou estarem sofrendo mudanças conformacionais induzidas pela interação com o

oligonucleotídeo. Estes se encontram na fita $\beta 2$ (G23 e T28), fita $\beta 3$ (F37 e H39) e na região do *loop* $\beta 3\beta 4$ (Q40, R49 e L51).

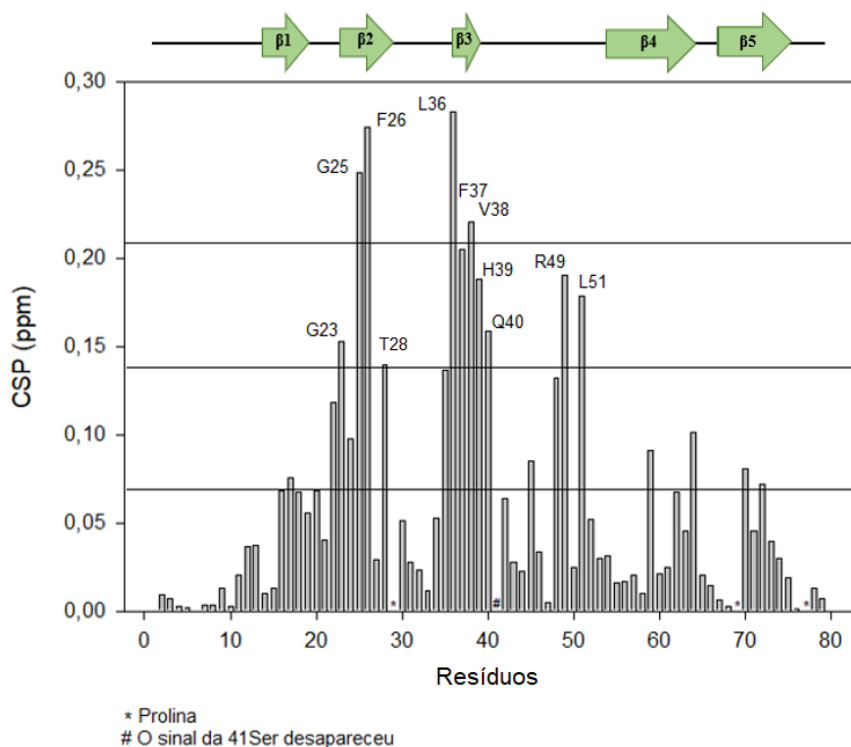


Figura 19: Gráfico de Perturbação do Deslocamento Químico (CSP) por resíduo. As linhas traçadas indicam, de baixo para cima, os valores referentes a σ , 2σ e 3σ , respectivamente. Os resíduos reconhecidos como parte do sítio de interação de AtGRP2-CSD com o oligonucleotídeo T7 estão identificados. Uma representação da estrutura secundária da proteína modelada pelo software Phyre2 está apresentada acima do gráfico. Ela é composta por cinco fitas β ($\beta 1$ a $\beta 5$).

O modelo obtido por modelagem é apresentado nas Figuras 20A e 20B com os aminoácidos que fazem parte do sítio de interação demarcados e identificados. A estrutura obtida apresentou forma de barril- β , sendo composta por cinco fitas- β antiparalelas, como esperado para um CSD. Isto evidencia a posição adjacente entre as fitas $\beta 2$ e $\beta 3$, nas quais estão localizados a maioria dos resíduos com CSPs altos, compondo a superfície de interação. Um ponto interessante que pode ser observado é a orientação da cadeia lateral de resíduos hidrofóbicos (F26 e F37); estes estão voltados para o solvente ao invés de formarem um centro hidrofóbico no interior da proteína, o que corrobora o envolvimento destes aminoácidos na interação com o oligonucleotídeo. Além do perfil hidrofóbico do sítio, também há uma grande presença de resíduos carregados positivamente nesta região (Figura 20C). Todas estas informações levam a conclusão de que a interação entre AtGRP2-CSD e o oligonucleotídeo

de DNA se dá através de um empilhamento das cadeias laterais dos resíduos hidrofóbicos com as bases nucleotídicas do DNA, além de a carga positiva da superfície do sítio favorecer uma interação eletrostática com os fosfatos do ácido nucleico.

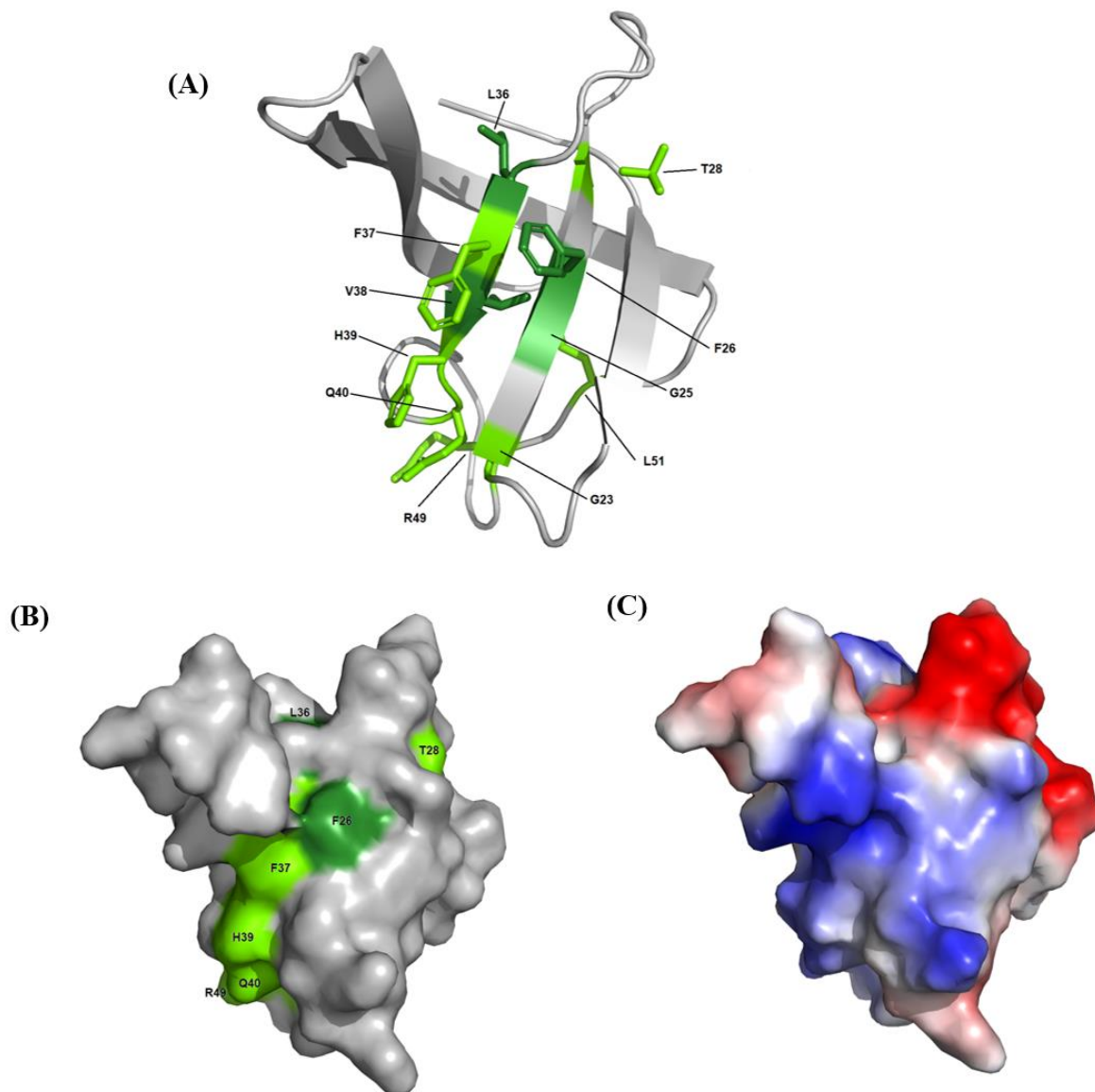


Figura 20: Análise do sítio de interação de AtGRP2-CSD. (A) Diagrama de fita da estrutura tridimensional modelada de AtGRP2-CSD. Os resíduos que fazem parte do sítio de interação com o oligonucleotídeo T7 estão marcados em tons de verde (verde escuro: CSP > 3σ; verde claro: CSP > 2σ). As cadeias laterais destes resíduos estão apresentadas. (B) Representação da superfície do modelo estrutural de AtGRP2-CSD com os resíduos do sítio de interação marcados em tons de verde. (C) Representação da superfície eletrostática de AtGRP2-CSD. Regiões com potencial positivo tem coloração azul, enquanto as áreas com potencial negativo estão em vermelho e com potencial neutro em branco.

Neste ponto, retornamos a outro fato que foi observado durante a titulação de AtGRP2-CSD com o oligonucleotídeo T7. No tópico 4.1.2 foi identificada e

investigada a presença de mais de uma conformação da proteína em equilíbrio em solução e neste tópico foi mencionado que somente as ressonâncias da conformação enovelada sofrem deslocamento com o aumento da razão molar de ligante. Contudo, foi constatada uma diminuição da intensidade dos sinais da forma desenovelada com o aumento da razão molar proteína:ligante. Simultaneamente, foi observado um aumento na intensidade da forma enovelada. Esta mudança está evidenciada na Figura 21, onde as intensidades dos picos da forma enovelada nos pontos de razão 1:0 e 1:1,5 estão apresentados. Os resíduos localizados em regiões de sobreposição de sinal foram retirados do gráfico de modo a não prejudicar a análise. Aproximadamente 80% dos aminoácidos apresentaram aumento na intensidade com o acréscimo de ligante no meio. Este dado indica que houve um deslocamento no equilíbrio de conformações de AtGRP2-CSD para a forma enovelada na presença de ligante, o que sugere um enovelamento da proteína concomitante a sua interação com o oligonucleotídeo. Outro dado que apoia esta conclusão foi obtido através dos experimentos de espectroscopia de fluorescência realizados pelo nosso laboratório (dados não mostrados): para a interação com T7, foi obtido n de Hill = 0,3966. O fator n de Hill representa a cooperatividade na interação; valores menores do que 1 indicam cooperatividade negativa. Um estudo na literatura ⁽⁴²⁾ mostrou que valores de n de Hill < 1 podem ser obtidos para proteínas intrinsecamente desordenadas em equilíbrio com outra conformação ao interagirem com DNA e se enovelarem.

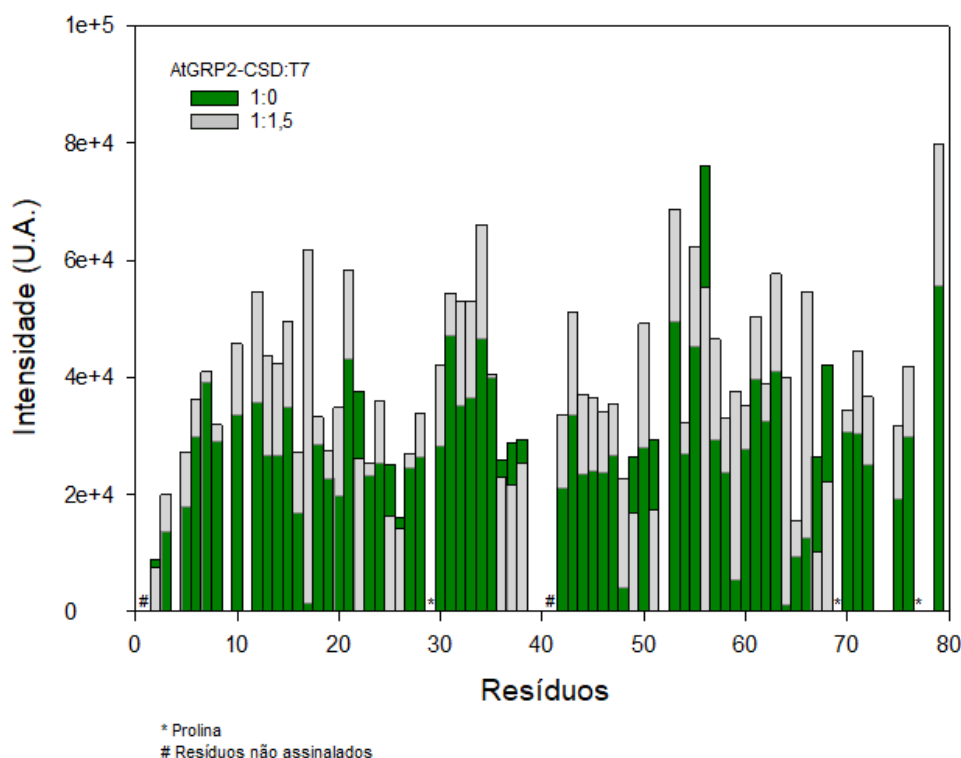


Figura 21: Intensidade dos sinais da conformação enovelada de AtGRP2-CSD em razões molares de T7 diferentes. As razões molares AtGRP2-CSD:T7 apresentadas são: 1:0 (verde escuro) e 1:1,5 (cinza). Os sinais sobrepostos foram retirados da análise. As prolinas e os resíduos não assinalados estão marcados. O gráfico apresenta um aumento majoritário da intensidade dos sinais da forma enovelada de AtGRP2-CSD com o aumento da concentração do ligante.

4.2. Discussão

4.2.1. Equilíbrio conformacional de AtGRP2-CSD

A sequência primária de uma proteína guarda de forma codificada uma vasta quantidade de informações sobre a proteína; sua estrutura tridimensional, a variação desta ao longo do tempo e sua estabilidade termodinâmica são todos derivados e refletidos nesta sequência ⁽⁴³⁾. Conhecer estas informações é, portanto, essencial para se entender a função biológica de uma proteína.

Intrinsecamente relacionados com a estabilidade de uma proteína estão os estados desenovelados. Estes são fundamentais para o entendimento do evento de enovelamento proteico e sua cinética ⁽⁴⁴⁾. Como o enovelamento proteico é um evento rápido, podendo ocorrer na escala de milissegundos ⁽⁴⁵⁾, a caracterização destes estados intermediários se faz difícil.

Neste trabalho, foi constatada a presença de uma conformação desenovelada em equilíbrio em solução com a forma enovelada do domínio *cold-shock* de AtGRP2 em pH próximo do neutro. O resultado anômalo obtido para a

massa molecular desta proteína nos géis de SDS-PAGE, sendo que sua massa foi confirmada por espectrometria de massas, pode ser uma consequência direta deste equilíbrio.

Casos de um equilíbrio como este em proteínas não são comuns, mas alguns outros exemplos já foram observados na literatura; interessante, um destes também é um CSD. O domínio *cold-shock* da proteína YB-1 é o domínio central e ligante de ácidos nucleicos da proteína humana Y-box 1, a qual, tal como outras proteínas da família Y-box, está envolvida em processos de regulação de transcrição e tradução. A presença de uma forma desenovelada em equilíbrio com seu estado enovelado já foi estudada por RMN ⁽⁴⁶⁾. Assim como para AtGRP2-CSD, estas conformações de YB-1 CSD se apresentaram em regime de troca lenta na escala do RMN e ressonâncias extras indicando uma segunda conformação desenovelada para alguns aminoácidos também foram encontradas. Uma razão de 7:3 entre as formas enovelada e desenovelada, respectivamente, foi determinada. Kloks *et al.* (2004) confirmaram em suas análises que o estado desenovelado de YB-1 CSD é diferente da forma desnaturada da proteína e constataram a presença de uma estrutura não randômica no estado desenovelado.

Outro caso mais amplamente estudado na literatura é o do domínio SH3 (*Src Homology 3*) N-terminal da proteína drk (*downstream of receptor kinase*). Esta é uma proteína de *Drosophila* que auxilia em processos de transdução de sinais ⁽⁴⁷⁾. Diferentemente de CSDs, domínios SH3 se ligam a proteínas; em específico, sequências ricas em prolinas ⁽⁴⁷⁾. A cinética de troca entre as conformações enovelada e desenovelada deste domínio também ocorre em regime lento na escala do RMN ⁽⁴⁸⁾. A razão entre seus estados enovelado e desenovelado foi determinada como aproximadamente 1:1 ⁽⁴⁸⁾. Novamente, foi constatado que seu estado desenovelado difere de uma forma desnaturada do domínio ⁽⁴⁹⁾ e que há a presença de uma estrutura residual neste ⁽⁴⁹⁻⁵¹⁾. Marsh *et al* (2007) realizaram uma caracterização estrutural do estado desenovelado de drkN SH3 e obtiveram resultados que sugerem um *ensemble* compacto e a presença de estruturas nativas e não nativas nesta forma da proteína.

As estruturas tridimensionais das formas enoveladas de YB-1 CSD e drkN SH3 já foram determinadas ^(49,52). O que é interessante neste fato é que ambas estruturas obtidas são muito semelhantes às de outros domínios CSD e SH3

conhecidos, mas só existem relatos da existência de uma conformação desenovelada em equilíbrio com a forma nativa para estas proteínas. Isto indica que há uma instabilidade em YB-1 CSD e drkN SH3 não encontrada em seus homólogos que, ao passo em que resulta em uma população significativa de um estado desordenado, não extingue a contribuição da forma enovelada. Mok *et al.* (2001) realizaram um alinhamento entre diversas sequências de domínios SH3 e identificaram uma sequência Gli-Asp ou Gli-Glu altamente conservada em uma volta- β entre as fitas β 3 e β 4 deste domínio. Esta é substituída por Thr-Gln em drkN SH3, uma sequência cuja propensão em formar a volta- β é significativamente menor. Experimentos de mutagênese posteriores confirmaram que a troca deste resíduo de treonina por uma glicina foi capaz de aumentar significativamente a estabilidade da proteína em seu estado enovelado (aumento de 20°C na T_m da proteína) ⁽⁵¹⁾. O grupo também propôs, a partir de seus resultados, que o resíduo de glicina nesta posição desestabiliza interações importantes na manutenção da forma desenovelada visto que é nesta região da volta- β divergente que foi constatada a presença de uma estrutura em hélice não-nativa na forma desenovelada ⁽⁵¹⁾. Estes resultados levantam a possibilidade de um caso semelhante ocorrer com os domínios CSD de AtGRP2 e YB-1.

Interessantemente, só existem relatos deste equilíbrio conformacional para os domínios isolados, o que sugere que a presença do restante da estrutura destas proteínas é capaz de estabilizar os domínios. Mais experimentos devem ser realizados para confirmar este ponto. O papel desta instabilidade que gera o equilíbrio conformacional no mecanismo de ação destas proteínas, as quais possuem funções biológicas variadas, parece ser complexo.

Como um todo, a existência de um estado desenovelado em equilíbrio com a forma enovelada em uma dada condição em que a conformação enovelada se mantém estável cria um excelente modelo para o estudo do enovelamento e estabilidade proteicos e deve ser mais explorado. Além disso, o papel deste equilíbrio de troca conformacional pode ser fundamental para um entendimento mais profundo do mecanismo de ação pelo qual estas proteínas exercem suas funções biológicas.

4.2.2. Sítio de interação de AtGRP2-CSD inclui os motivos RNP-1 e RNP-2

O motivo conservado ribonucleoproteico 1 (RNP-1) foi primeiramente observado na sequência primária da proteína ligante de mRNA poli-A⁽⁵³⁾ e é composto principalmente por resíduos aromáticos e positivos. Posteriormente, o motivo RNP-2, formado por seis aminoácidos e menos conservado, também foi identificado⁽⁵⁴⁾. Apesar de terem sido identificados em domínios RRM (do inglês, *RNA recognition motif*), já foi comprovada a sua presença em CSDs⁽⁵⁵⁾, onde são tipicamente localizados nas fitas $\beta 2$ e $\beta 3$.

Um estudo realizado por Landsman (1992) com diversos CSDs e RBPs mostrou que o motivo RNP-1 possui um centro hidrofóbico nas posições 2-5 com a seguinte composição: [G/A][F/Y][G/A][F/Y]. Este centro é muito conservado, sendo encontrado em mais de 90% dos exemplos analisados. A posição 6 costuma ser preenchida por Ile, Val ou Ala. As posições 1 e 7 são menos conservadas, possuindo um aminoácido básico e um hidrofílico (como Asp, Glu, Asn, Gln) ou uma Thr, respectivamente. Um resíduo hidrofóbico pode ser encontrado na posição 8 ou, mais comumente para CSDs, uma Arg⁽⁵⁶⁾. No caso do RNP-2, a menor conservação resulta em maior variação na sequência observada em CSDs e domínios RRM. Para estes últimos, a sequência consenso é determinada como: [I/V/L][F/Y][I/V/L][X][N][L], sendo X qualquer aminoácido⁽⁵⁷⁾. Para CSDs, o motivo RNP-2 costuma ser descrito com uma composição muito semelhante, sendo por vezes apresentado como uma sequência de quatro aminoácidos, [V/L][F/Y][V/L][H]^(56,58). Os resíduos de Phe de ambos os motivos RNP-1 (posição 5) e RNP-2 já foram descritos como essenciais para a interação de CSDs com ácidos nucleicos⁽⁵⁶⁾.

Baseado nessas informações, as regiões correspondentes aos motivos RNP-1 e RNP-2 de AtGRP2-CSD foram identificadas como sendo:

RNP-1: K22, G23, F24, G25, F26, I27, T28 e P29.

RNP-2: D35, L36, F37, V38, H39 e Q40.

Em verde estão identificados os resíduos pertencentes ao sítio de interação com o oligonucleotídeo de DNA (verde escuro: CSP > 3σ ; verde claro: CSP > 2σ).

Como esperado, há uma grande sobreposição entre os aminoácidos que fazem parte do sítio de interação de AtGRP2-CSD com T7 e estes motivos característicos de interação com ácidos nucleicos em CSDs. Os resíduos do sítio primário de interação estão divididos em ambos os motivos RNP-1 e RNP-2, indicando que ambos são importantes para a interação de ATGRP2-CSD com seu ligante. Interessantemente, apesar de haver uma fenilalanina na posição 3 do motivo RNP-1, seu sinal não apresentou variações grandes de deslocamento químico ao longo da titulação. O resíduo I27, ainda que também condizente com a sequência consenso e presente em uma posição central no motivo, também não apresentou valor alto de CSP.

Além dos resíduos presentes nos motivos RNP-1 e RNP-2 de AtGRP2-CSD, também foram constatados valores altos de CSP para dois aminoácidos presentes no *loop* entre as fitas β 3 e β 4: R49 e L51. A literatura também aponta esta região entre as fitas β 3 e β 4 como sendo importante para a interação com ssDNA ⁽⁵⁹⁾. O estudo de Kleene (2018) apontou a sequência [E][G][F][R/K][T/S][L] como sendo o motivo presente neste *loop* para os CSDs de *B. subtilis* e *X. tropicalis*; esta sequência pode ser encontrada em AtGRP2-CSD: [E46][G47][F48][R49][S50][L51] e contém os dois aminoácidos identificados como parte do sítio de interação com o oligonucleotídeo de DNA. Dessa forma, foi constatado neste trabalho que o sítio de interação do domínio *cold-shock* de AtGRP2 é semelhante ao de outros CSDs e CSPs, sendo composto pelos motivos RNP-1 e RNP-2 e pela região de *loop* entre as fitas β 3 e β 4.

5. Conclusões

- O espectro [$^1\text{H},^{15}\text{N}$] HSQC de AtGRP2-CSD apresenta mais do que o dobro da quantidade de sinais esperados. Isto indica que a proteína se encontra em equilíbrio conformacional, em cinética de troca da ordem de segundos, em solução. Alguns aminoácidos ainda experimentam uma terceira conformação;
- A baixa dispersão na dimensão do ^1H dos sinais extras no espectro [$^1\text{H},^{15}\text{N}$] HSQC indica que a outra conformação de AtGRP2-CSD é uma forma desenovelada;
- A partir de experimentos de tripla ressonância de RMN, foi possível assinalar 95% dos aminoácidos da forma enovelada e 68% da forma desenovelada. Este é o primeiro passo para a determinação da estrutura tridimensional de AtGRP2-CSD;
- AtGRP2-CSD é capaz de se ligar ao oligonucleotídeo T7. A partir dos dados de Perturbação do Deslocamento Químico, foi possível obter a superfície de interação entre a proteína e o ligante. A proteína utiliza uma plataforma de resíduos aromáticos e positivos pertencentes aos motivos RNP-1, RNP-2 e do *loop* entre as fitas $\beta 3$ e $\beta 4$ para realizar esta interação;
- Há um aumento na intensidade dos sinais referentes a forma enovelada de AtGRP2-CSD no espectro [$^1\text{H},^{15}\text{N}$] HSQC com o aumento da concentração de ligante enquanto o oposto ocorre com os sinais da forma desenovelada. Isto sugere um enovelamento da proteína concomitante com sua interação com ácido nucleico.

6. Referências bibliográficas

- 1 - Halbeisen, R.E., Galgano, A., Scherrer, T. *et al.* Post-transcriptional gene regulation: From genome-wide studies to principles. **Cell. Mol. Life Sci**, v. 65, p. 798, 2008.
- 2 - Dassi, E.; Quattrone, A. Tuning the engine, an introduction to resources on posttranscriptional regulation of gene expression. **RNA Biology**, v. 9, n. 10, p. 1224–1232, 2012.
- 3 - Glisovic, T., Bachorik, J.L., Yong, J., Dreyfuss, G. RNA-binding proteins and post-transcriptional gene regulation. **FEBS Letters**, v. 582, p. 1977-1986, 2008.
- 4 - Day, D.A., Tuite, M.F. Post-transcriptional gene regulatory mechanisms in eukaryotes: an overview. **Journal of Endocrinology**, v. 157, p. 361-371, 1998.
- 5 - Tew, S. R., Vasieva, O., Peffers, M. J., Clegg, P. D. Post-transcriptional gene regulation following exposure of osteoarthritic human articular chondrocytes to hyperosmotic conditions. **Osteoarthritis Cartilage**, v. 19, n. 8, p. 1036-1046, 2011.
- 6 – Sachetto-Martins, G., Franco, L.O., de Oliveira, D.E. Plant glycine-rich proteins: a Family or just proteins with a common motif? **Biochimica et Biophysica Acta**, v. 1492, p. 1-14, 2000.
- 7 - Mangeon, A., Junqueira, R.M., Sachetto-Martins, G. Functional diversity of the plant glycine-rich proteins superfamily. **Plant Signaling & Behavior**, v. 5, p. 99-104, 2010.
- 8 - Czolpinska, M., Rurek, M. Plant Glycine-Rich Proteins in Stress Response: An Emerging, Still Prospective Story. **Front. Plant Sci.**, v. 9, p. 302, 2018.
- 9 – Kingsley, P.D., Palis, J. GRP2 proteins contain both CCHC zinc fingers and a cold shock domain. **The Plant Cell**, v. 6, p. 1522-1523, 1994.
- 10 – Gendra, E., Moreno, A., Alba, M.M., Molecular, D.D.G., Fabre, U.P. Interaction of the plant glycine-rich RNA-binding protein MA16 with a novel nucleolar DEAD box RNA helicase protein from *Zea mays*. **The Plant Journal**, v. 38, p. 875-886, 2004.
- 11 – Sasaki, K., Imai, R. Pleiotropic roles of cold shock domain proteins in plants. **Front Plant Sci.**, v. 2, p. 116, 2012.
- 12 – Wang, S., Wang, R., Liang, D., Ma, F., Shu, H. Molecular characterization and expression analysis of a glycine-rich RNA-binding protein gene from *Malushupenhensis*Rehd. **Mol Biol Rep.**, v. 39, p. 4145–415, 2012.
- 13 - Oliveira, D.E., Seurinck, J., Inzé, D., Montagu, M.V., Bottermana, J. Differential Expression of Five Arabidopsis Genes Encoding Glycine-Rich Proteins. **The Plant Cell**, v. 2, p. 427-436, 1990.
- 14 - Fusaro, A.F., Bocca, S.N., Ramos, R.L.B., Barrôco, R.M., Magioli, C., Jorge, V.C., Sachetto-Martins, G. AtGRP2, a cold-induced nucleo-cytoplasmic RNA-binding protein, has a role in flower and seed development. **Planta**, v. 225, n. 6, p. 1339-1351, 2007.
- 15 – Nakaminami, K *et al.* Arabidopsis cold shock domain proteins: relationships to floral and silique development. **J. Exp. Bot.**, v. 60, p. 1047-1062, 2009.
- 16 – Sasaki, K., Kim, M. H., Imai, R. Arabidopsis cold shock domain protein 2 is a RNA chaperone that is regulated by cold and development signals. **Biochem. Biophys. Res. Commun.**, v. 364, p. 633-638, 2007.
- 17 - Sasaki, K., Kim, M.-H., Imai, R. Arabidopsis COLD SHOCK DOMAIN PROTEIN 2 is a negative regulator of cold acclimation. **New Phytologist**, v. 198, p. 95-102, 2013.

- 18 – Sasaki, K., Kim, M. H., Kanno, Y., Seo, M., Kamiya, Y., Imai, R. Arabidopsis cold shock domain protein 2 influences ABA accumulation in seed and negatively regulates germination. **Biochem Biophys Res Commun.**, v. 456, n. 1, p. 380-4, 2015.
- 19 - Jaglo-Ottosen, K. R. Arabidopsis CBF1 overexpression induces COR genes and enhances freezing tolerance. **Science**, v. 280, p. 104-6, 1998.
- 20 – Gilmour, S. J., Sebolt, A. M., Salazar, M. P., Everard, J. D., Thomashow, M. F. Overexpression of the Arabidopsis CBF3 transcriptional activator mimics multiple biochemical changes associated with cold acclimation. **Plant Physiol**, v. 124, p. 1854-65, 2000.
- 21 – Sasaki, K., Liu, Y., Kim, M. H., Imai, R. An RNA chaperone, AtCSP2, negatively regulates salt stress tolerance. **Plant Signal Behav.**, v. 10, n. 8, e1042637, 2015.
- 22 - Khan, F., Daniëls, M.A., Folkers, G.E., Boelens, R., Saqlan Naqvi, S.M., van Ingen, H. Structural basis of nucleic acid binding by *Nicotiana tabacum* glycine-rich RNA-binding protein: implications for its RNA chaperone function. **Nucleic Acid Research**, v. 42, n. 13, p. 8705-8718, 2014.
- 23 – Kim, J. S., Park, S. J., Kwak, K. J., Kim, Y. O., Kim, J. Y., Song, J., Jang, B., Jung, C. H., Kang, H. Cold shock domain proteins and glycine-rich RNA-binding proteins from *Arabidopsis thaliana* can promote the cold adaptation process in *Escherichia coli*. **Nucleic Acids Res.**, v. 35, n. 2, p. 506–516, 2007.
- 24 – Jones PG, Inouye M. The cold-shock response - a hot topic. **Mol Microbiol**, v. 11, n. 5, p. 811-818, 1994.
- 25 – Lindquist, J.A., Mertens, P.R. Cold Shock proteins: from cellular mechanisms to pathophysiology and disease. **Cell communication and Signaling**, v. 16, p. 63, 2018.
- 26 – Mohd, A., Kumar, V., Dohare, R., Islam, A., Ahmad, F., Hassan, I. Sequence, structure and evolutionary analysis of cold shock domain proteins, a member of OB fold family. **Journal of Evolutionary Biology**, v. 31, p. 1903-1917, 2018.
- 27 – M. Amir, V. Kumar, R. Dohare, et al., Investigating architecture and structure-function relationships in cold shock DNA-binding domain family using structural genomics based approach. **International Journal of Biological Macromolecules**, v. 133, p. 484-494, 2019.
- 28 – Rennella, E., Sára, T., Juen, M., Wunderlich, C., Imbert, L., Solyom, Z., Favier, A., Ayala, I., Weinhäupl, K., Schanda, P., Konrat, R., Kreutz, C., Brutscher, B. RNA binding and Chaperone Activity of the E. Coli Cold Shock Protein CspA. **Nucleic Acids Research**, v. 45, n. 7, p. 4255-4268, 2017.
- 29 – Kloks, C.P.A.M., Spronk, C.A.E.M., Lasonder, E., Hoffman, A., Vuister, G.W., Grzesiek, S., Hilbers, W. The Solution Structure and DNA-binding Properties of the Cold-shock Domain of the Human Y-box Protein YB-1. **Journal of Molecular Biology**, v. 316, p. 317-326, 2002.
- 30 – Laemmli UK. Cleavage of Structural Proteins during the Assembly of the Head of Bacteriophage T4. **Nature**, v. 227, p. 680-685, 1970.
- 31 – Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., Bax, A. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. **Journal of Biomolecular NMR**, v. 6, n. 3, p. 277–293, 1995.
- 32 - Sattler, M., Schleucher, J., Griesinger, C. Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulse field gradients. **Progress in NMR Spectroscopy**, v. 34, n. 23, p. 93-158, 1999.

- 33 – Vranken, W. F., Boucher, W., Stevens, T. J., Fogh, R. H., Paion, A., Llinas, M., Ulrich, E. L., Markley, J. L., Ionides, J., Laue, E. The CCPN data model for NMR spectroscopy: development of a software pipeline. **Proteins**, v. 59, n. 4, p. 687-696, 2005.
- 34 – Kelley, L. A., Mezulis, S., Yates, C. M., Wass, M. N., Sternberg, M. J. The Phyre2 web portal for protein modeling, prediction and analysis. **Nature Protocols**, v. 10, p. 845-858, 2015.
- 35 - The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
- 36 - Carrington, J. C., Dougherty, W. G. A viral cleavage site cassette: identification of amino acid sequences required for tobacco etch virus polyprotein processing. **Proceedings of the National Academy of Sciences**, v. 85, n. 10, p. 3391-3395, 1988.
- 37 – Wüthrich, K. NMR of Proteins and Nucleic Acids. New York, Wiley, 1986.
- 38 – Higman, V. Triple Resonance Backbone Assignment. 2019. Disponível em: <<https://www.protein-nmr.org.uk/solution-nmr/assignment-theory/triple-resonance-backbone-assignment/>>. Acesso em: 23 mai. 2020.
- 39 – Grzesiek, S., Bax, A. An efficient experiment for sequential backbone assignment of medium-sized isotopically enriched proteins. **Journal of Magnetic Resonance**, v. 99, p. 201–207, 1992.
- 40 – Karlson, D., Nakaminami, K., Toyomasu, T., Imai, R. A Cold-regulated Nucleic Acid-binding Protein of Winter Wheat Shares a Domain with Bacterial Cold Shock Proteins. **Journal of Biological Chemistry**, v. 277, p. 35248-35256, 2002.
- 41 – Zeeb, M., Balbach, J. Single-stranded DNA-binding of the cold-shock protein CspB from *Bacillus Subtilis*. **Protein Science**, v. 12, p. 112-123, 2003.
- 42 - Munshi, S., Gopi, S., Subramanian, S., Campos, L. A., Naganathan, A. N. Protein plasticity driven by disorder and collapse governs the heterogeneous binding of CytR to DNA. **Nucleic Acids Research**, v. 46, n. 8, p. 4044-4053, 2018.
- 43 - Anfinsen, C. B. Principles that govern the folding of protein chains. **Science**, v. 181, p. 223-230, 1973.
- 44 - Shortle, D. The denatured state (the other half of the folding equation) and its role in protein stability. **FASEB J.**, v. 10, p. 27-34, 1996.
- 45 - Kuwajima, K., Sakuraoka, A., Fueki, S., Yoneyama, M., Sugai, S. **Biochemistry**, v. 27, p. 7419, 1988.
- 46 - Kloks, C.P.A.M., Tessari, M., Vuister, G.W., Hilbers, C.W. Cold Shock Domain of the Human Y-Box Protein YB-1. Backbone Dynamics and Equilibrium between the Native State and a Partially Unfolded State. **Biochemistry**, v. 43, n. 31, p. 10237-10246, 2004.
- 47 - Olivier, J. P., Raabe, T., Henkemeyer, M., Dickson, B., Mbamalu, G., Margolis, B., Schlessinger, J., Hafen, E., Pawson, T. **Cell**, v. 73, p. 179-191, 1993.
- 48 – Zhang, O., Kay, L. E., Olivier, J. P., Forman-Kay, J. D. Backbone ¹H and ¹⁵N resonance assignments of the N-terminal SH3 domain of drk in folded and unfolded states using enhanced-sensitivity pulsed field gradient NMR techniques. **Journal of Biomolecular NMR**, v. 4, p. 845-858, 1994.
- 49 – Zhang, O., Forman-Kay, J. D. Studies of Unfolded States of an SH3 Domain in Aqueous Solution and Denaturing Conditions. **Biochemistry**, v. 36, p. 3959-3970, 1997.
- 50 – Marsh, J. A., Neale, C., Jack, F. E., Choy, W., Lee, A. Y., Crowhurst, K. A., Forman-Kay, J. D. Improved Structural Characterizations of the drkN SH3 Domain Unfolded State Suggest a Compact Ensemble with Native-like and Non-native Structure. **Journal of Molecular Biology**, v. 367, p. 1494-1510, 2007.

- 51 - Mok, Y., Eliseeva, E. L., Davidson, A. R., Forman-Kay, J. D. Dramatic Stabilization of an SH3 Domain by a Single Substitution: Roles of the Folded and Unfolded States. **Journal of Molecular Biology**, v. 307, p. 913-928, 2001.
- 52 - Bezsonova, I., Singer, A., Choy, W.-Y., Tollinger, M., Forman-Kay, J.D. Structural Comparison of the Unstable drkN SH3 Domain and a Stable Mutant. **Biochemistry**, v. 44, p. 15550-15560, 2005
- 53 - Adam SA, Nakagawa T, Swanson MS, Woodruff TK, Dreyfuss G. mRNA polyadenylate-binding protein: gene isolation and sequencing and identification of a ribonucleoprotein consensus sequence. **Mol Cell Biol.**, v. 6, p. 2932–1943, 1986.
- 54 - Dreyfuss G, Swanson MS, Pinol-Roma S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. **Trends Biochem Sci.**, v. 13, p. 86–91, 1988.
- 55 - Maris, C., Dominguez, C., and Allain, F. H. T. The RNA recognition motif, a plastic RNA-binding platform to regulate post-transcriptional gene expression, **FEBS J.**, v. 272, p. 2118-2131, 2005.
- 56 – Landsman, D. RNP-1, an RNA-binding motif is conserved in the DNA-binding cold shock domain. **Nucleic Acids Research**, v. 20, n. 11, p. 2861-2864, 1992.
- 57 - Dreyfuss, G., Swanson, M. S., Pinol-Roma, S. Heterogeneous nuclear ribonucleoprotein particles and the pathway of mRNA formation. **Trends Biochem. Sci.**, v. 13, p. 86–91, 1988.
- 58 – Eliseeva, I. A., Kim, E. R., Guryanov, S. G., Ovchinnikov, L. P., Lyabin, D. N. Y-Box-Binding Protein 1 (YB-1) and It's Functions. **Biochemistry (Moscow)**, v. 76, n. 13, p. 1402-1433, 2011.
- 59 - Kleene, K. C. Y-box proteins combine versatile cold shock domains and arginine-rich motifs (ARMs) for pleiotropic functions in RNA biology. **Biochem J.**, v. 475, p. 2769-2784, 2018.