UNIVERSIDADE DE SÃO PAULO
INSTITUTO DE FÍSICA DE SÃO CARLOS

GABRIEL BELEM DE ANDRADE

Estudos estruturais de dockerinas e cohesinas em *Ruminococcus flavefaciens* e sua aplicação no desenvolvimento de matrizes auto montáveis de proteínas

São Carlos
2017

GABRIEL BELEM DE ANDRADE

Estudos estruturais de dockerinas e cohesinas em *Ruminococcus flavefaciens* e sua aplicação no desenvolvimento de matrizes auto montáveis de proteínas

Dissertação apresentada ao Programa de Pós-Graduação em Física do Instituto de Física de São Carlos da Universidade de São Paulo, para obtenção do título de Mestre em Ciências.

Área de concentração: Física Aplicada
Opção: Física Biomolecular

Orientador: Prof. Dr. Marcos Vicente de Albuquerque Salles Navarro

Versão Corrigida

(versão original disponível na Unidade que aloja o Programa)

São Carlos
2017

AUTORIZO A REPRODUÇÃO E DIVULGAÇÃO TOTAL OU PARCIAL DESTE TRABALHO, POR QUALQUER MEIO CONVENCIONAL OU ELETRÔNICO PARA FINS DE ESTUDO E PESQUISA, DESDE QUE CITADA A FONTE.

Ficha catalográfica revisada pelo Serviço de Biblioteca e Informação do IFSC,
com os dados fornecidos pelo(a) autor(a)

Andrade, Gabriel Belem de
Estudos estruturais de dockerinas e cohesinas em
Ruminococcus flavefaciens e sua aplicação no
desenvolvimento de matrizes auto montáveis de
proteínas / Gabriel Belem de Andrade; orientador
Marcos Vicente de Albuquerque Salles Navarro -
versão corrigida -- São Carlos, 2017.
122 p.

Dissertação (Mestrado - Programa de Pós-Graduação em
Física Biomolecular) -- Instituto de Física de São
Carlos, Universidade de São Paulo, 2017.

1. Dockerina. 2. Cohesina. 3. *Ruminococcus*
flavefaciens. 4. Redes automontáveis. 5.
Nanomaterial. I. Navarro, Marcos Vicente de
Albuquerque Salles, orient. II. Título.

AGRADECIMENTOS

Ao Professor Dr. Marcos V. A. S. Navarro, pela orientação, ensinamentos e oportunidades durante esse mestrado.

Aos meus pais, por nunca pouparem esforços por mim, nunca serei grato na mesma proporção.

Ao meu irmão, Tulio, pela parceria e diversão constantes.

À minha namorada Lygia, pelo constante apoio, compreensão e companheirismo.

Aos grandes amigos Rafael, Bruno, Leo, Wesley, Tripa, Darlan e Dany pela amizade e risadas certas.

Aos amigos da pós-graduação Rapha, Ton, Nai, Carla, Helô, Ana, Sina, pela ajuda no dia a dia do laboratório e pela divertida companhia nos almoços e cafés.

Aos funcionários do laboratório da biofísica, Rafael, Fernando, Bel e Andressa, e aos funcionários do IFSC, por todo o tipo de auxílio.

A todos os professores que contribuíram para minha formação durante esse mestrado.

Ao Instituto de Física de São Carlos, e à Universidade de São Paulo, por toda a estrutura e excelência no ensino e pesquisa.

À CAPES, pela bolsa concedida.

-Podes dizer-me, por favor, que caminho devo seguir para sair daqui?
-Isso depende muito de para onde queres ir - respondeu o gato.
-Pouco me importa aonde ir - disse Alice.
-Nesse caso, pouco importa o caminho que sigas - replicou o gato.

Lewis Carroll

RESUMO

ANDRADE, G. B. **Estudos estruturais de dockerinas e cohesinas em *Ruminococcus flavefaciens* e sua aplicação no desenvolvimento de matrizes auto montáveis de proteínas**. 2017. 122 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

O celulosomo é um complexo multienzimático extracelular utilizado por bactérias anaeróbias para a degradação de biomassa vegetal. Ele é composto por escafoldinas, estruturas alongadas que abrigam diversos módulos cohesina, às quais se ligam dockerinas, seus parceiros de interação específica de alta afinidade, fusionados às enzimas celulolíticas. Os módulos cohesina e dockerina compõem o elemento central da interação entre todos os componentes que integram o celulosomo. Esses módulos são divididos em tipos, de acordo com sua sequência primária. Essa divisão reflete efeitos funcionais distintos, sendo o tipo I responsável pela ligação de enzimas às escafoldinas, enquanto o tipo II medeia a ligação de escafoldinas à célula. O mais complexo celulosomo conhecido é o de *Ruminococcus flavefaciens*, e na classificação por tipos, suas sequências divergem, formando o tipo III, que foi posteriormente subdividido em 6 grupos para significância funcional. Nesse sistema, o principal responsável pela integração de enzimas ao sistema é a escafoldina primária ScaA, a qual interage com escafoldina adaptadora ScaB. A especificidade dessa ligação - dockerina de ScaA (Rf-DocA) com cohesinas de ScaB (Rf-CohB1-7) - é classificada como único membro do grupo 5, na divisão de grupos que compõem o tipo III. Assim, essa interação é de suma importância para a organização do celulosomo desse organismo, tendo sido estudada por meio de experimentos biofísicos e bioquímicos. Porém a falta de uma estrutura cristalina resolvida desses componentes limita a compreensão que podemos ter sobre a interação.¹⁻² Nesse trabalho, apresentamos as estruturas cristalográficas de Rf-DocA, em complexo com a Rf-CohB4, além da estrutura dessa cohesina isolada, e ainda, a Rf-CohB1, e alguns de seus mutantes pontuais. Com isso, esclarecemos aspectos estruturais desses módulos, como a presença de dois sítios funcionais de ligação a cálcio em Rf-DocA. Também é observável pelos modelos gerados, detalhes da ligação entre eles, como os resíduos participantes da interação. Estudos de afinidade entre esses módulos foram conduzidos para elucidar algumas propriedades da ligação entre esses módulos, de forma que descobrimos que ela ocorre de uma única maneira, e que há um loop na cohesina cuja flexibilidade afeta a afinidade da ligação. Isso sugere um mecanismo de alteração conformacional que regula a ligação à dockerina. Adicionalmente, buscamos o emprego desses módulos em uma aplicação

tecnológica, desenhando redes automontáveis de proteínas, visando a construção de um nanomaterial. Essas redes são formadas por características intrínsecas das proteínas que os compõem, sendo o principal fator considerado sua simetria rotacional.³ Nesse sentido, as dockerinas e cohesinas foram utilizadas para ligação entre proteínas de diferentes simetrias. Utilizamos proteínas de simetrias C3, C4 e C6 com fusão a dockerinas, que se conectam às cohesinas fusionadas a proteínas de simetria C2, as quais formam o elemento linear da ligação entre os diferentes módulos. Esse desenho experimental permite expressão e purificação independentes dos componentes, o que facilita a obtenção das redes, a partir da mistura dos dois componentes. Através de análises preliminares por microscopia eletrônica de transmissão, observamos a formação de filmes bidimensionais extensos e nanotubos com a construção testada.

Palavras-chave: Dockerina. Cohesina. *Ruminococcus flavefaciens*. Redes automontáveis. Nanomaterial.

ABSTRACT

ANDRADE, G. B. **Structural studies of dockerins and cohesins of *Ruminococcus flavefaciens* and their application in self-assembling arrays of proteins**. 2017. 122 p. Dissertação (Mestrado em Ciências) – Instituto de Física de São Carlos, Universidade de São Paulo, São Carlos, 2017.

The cellulosome is an intricate multienzyme extracellular complexes evolved by anaerobic bacteria for degradation of cellulosic biomass. It is composed of scaffoldins, elongated structures, which bare numerous cohesin modules, which bind to dockerin modules, their high affinity and specificity partners, borne by cellulolytic enzymes. The cohesin and dockerina modules constitute the central element of the interaction between every component of the cellulosome. These modules are categorized in types, according to their primary sequence. That distribution reflects distinct functions, in which the type I is responsible for integration of enzymes to scaffoldins, while type II mediates anchoring of scaffoldins to the cell wall. The cellulosome of *Ruminococcus flavefaciens* is the most intricate known to date, which is categorized into a third type of cohesins and dockerins, due to sequence diversion. The type III was further divided into 6 groups to impart functional significance. In that system, the main enzyme integrating component is the primary scaffoldin ScaA, which interacts to the adaptor scaffoldin ScaB. The specificity of this interaction - dockerina of ScaA (Rf-DocA) to ScaB cohesins (Rf-CohB1-7) - is sorted as a single member of group 5, in the subtypes of type III. Thus, this interaction is essential for cellulosome organization, having been studied by biophysical and biochemical experiments. However, the lack of a solved crystalline structure of these components narrows our understanding on this interaction. In the present study, we present the structures of Rf-DocA, complexed to Rf-CohB4, besides the structure of this isolated cohesin, and also Rf-CohB1 and its point mutants. Due to these data, we clarify structural aspects of these modules, such as the occurrence of two functioning calcium binding sites in Rf-DocA. We also identified details of their binding, such as the interacting residues. Through binding affinity studies, we concluded that the interaction between these modules occurs in a single mode, and that there is a loop in the cohesin module whose flexibility has direct effects on the binding affinity to dockerin. Additionally, we sought to utilize these modules in a downstream application, by designing self-assembling arrays of proteins, aiming for the construction of a nanomaterial. These arrays are constructed from the intrinsic properties of its constituent proteins, in which the main factor is rotational symmetry. In this context, dockerina and cohesin modules were used of binding different symmetry proteins. We utilized C3, C4 and C6 point symmetry proteins fused to dockerin modules,

which bind to the cohesin modules fused to C2 point symmetry proteins, which establish the linear connection between the distinct components. This experimental design allows for the independent expression and purification of the components, which facilitates the achievement of the arrays, by simple mixture of the two components. Through preliminary analysis by transmission electron microscopy, we observed the construction of two-dimensional films and nanotubes.

Keywords: Dockerin. Cohesin. *Ruminococcus flavefaciens*. Self-assembling arrays. Nanomaterial.

LISTA DE FIGURAS

- Figura 1 - Micrografia Eletrônica de bactérias produtoras de celulosomo em uma célula vegetal. Os celulosomos são identificados pelos distintos pontos pretos anexos à superfície celular. 23
- Figura 2 – Desenho representativo de um celulosomo e seus componentes, tomando como exemplo a escafoldina CipA, de *Clostridium thermocellum*. A escafoldina, em amarelo, possui cohesinas do tipo I, para integração de enzimas através da ligação com dockerinas de tipo I, em vermelho, que estas carregam. Possui ainda um CBM, módulo de ligação a celulose, que ancora o complexo ao substrato. Sua dockerina C-terminal de tipo II, em verde, que porta um módulo X, em roxo, que se liga a uma cohesina tipo II, incorporada na superfície celular, o que ancora todo o complexo à célula..... 26
- Figura 3 - Árvore filogenética a partir da sequência primária de cohesinas. A relação filogenética entre cohesinas determina sua separação em diferentes tipos. Os três tipos conhecidos estão agrupados, como indicado pelos arcos na figura. A barra da escala indica a porcentagem (0.1) de substituição de aminoácidos. 28
- Figura 4 - Estrutura das dockerinas e cohesinas de tipos I e II. Todos os módulos retratados são de *Clostridium thermocellum*. Em vermelho, a dockerina do tipo I de Xilanase I, em ciano a cohesina 2 de CipA (PDB ID: 1OHZ). Em laranja, dockerina do tipo II, de CipB, que possui um módulo-X, em roxo; em amarelo, a cohesina 2 de Orf2p (PDB ID:5K39). As esferas em amarelo representam íons de cálcio. As cohesinas estão orientadas de forma que sua face frontal corresponde à que contata a dockerina. 29
- Figura 5 - Estruturas de complexos do Tipo III. (A) Díade Módulo-X:Doc de CttA e a Cohesina da escafoldina ScaE (PDB ID: 4IU2). (B) Uma dockerina do grupo 3 e a cohesina da escafoldina ScaC (PDB ID:5LXV). (C) cohesina de ScaG e uma díade módulo-X:dockerina (PDB ID:4WKZ)..... 31
- Figura 6 - Arquitetura do celulosomo de *Ruminococcus flavefaciens*, linhagem 17, um dos mais complexos conhecidos atualmente. Nesse sistema, há uma proteína específica para ligação ao substrato, CttA, um grande número de enzimas e escafoldinas adaptadoras. 33
- Figura 7 - Modelo da organização do celulosomo de *R. flavefaciens*, linhagem FD-1. As cores de fundo destacam as subdivisões do reconhecimento dockerina-cohesina. Em A, proteínas celulosomais. Em B, Proteínas que se ancoram à célula. Em C, dockerinas incompletas, do grupo 2. Em D, a proteína CttA, que medeia o anexo do sistema ao substrato. Os diferentes grupos estão destacados por cores, sendo azul (grupo 1), amarelo (grupo 2), verde (grupos 2 e 4), e vermelho (grupo 5). 36
- Figura 8 – Diferentes construções para os materiais automontáveis a partir de proteínas simétricas. Em a e b, observa-se a estratégia de nanoedros,

ligando proteínas simétricas com um *linker* rígido. Em c, a ligação é feita com *linker* flexível, o que não permite o controle da geometria.41

Figura 9 - Estratégia de construção de redes auto montáveis de proteínas utilizando proteínas de simetria C3, C4 e C6 como vértices da rede fusionadas a dockerinas e proteínas de simetria C2 fusionado a cohesinas. (A) Módulos selecionados para compor os vértices (proteínas de simetria C3, C4 e C6), fusionados a dockerinas e proteínas de simetria C2 selecionadas para compor o elemento de conexão linear entre os vértices, fusionadas a cohesinas. (B) A fusão de diferentes módulos origina cada elemento da rede, os quais são expressos e purificados independentemente. (C) A associação entre os componentes construídos da fusão das proteínas precursoras dá origem às redes automontáveis, através da ligação da dockerina com a cohesina, presentes em módulos distintos.53

Figura 10 - Representação da clonagem pela técnica LIC mostrando uma das junções entre o vetor e o inserto de interesse. Em vermelho, estão destacados os nucleotídeos retirados pela ação da T4 DNA Polimerase, criando extremidades de fita simples de DNA. As sequências são desenhadas para apresentar complementaridade, e incorporada nos *primers* (em verde). Os dois DNAs são então hibridizados e usados para transformar bactérias, que reparam a desconexão das fitas.55

Figura 11 - Detalhamento das extremidades desenhadas para ligação entre cada componente na clonagem. Em verde, estão exibidas as sequências incorporadas aos *primers*. Em vermelho, os nucleotídeos que serão extraídos no tratamento com a T4 DNA Polimerase.....56

Figura 12 - Mapa do vetor pET_SUMO utilizado na construção de nossas clonagens para expressão de proteínas componentes das matrizes auto montáveis. As setas roxas, externas ao vetor, indicam os locais em que os *primers* iniciam a amplificação do vetor.57

Figura 13 – Estratégias de inserção da fusão genética utilizando LIC e SLIC. Na parte superior, a estratégia LIC, e na parte inferior, SLIC. Em ambas partes da figura, mostramos a fusão da proteína trimérica de PDB ID:3C19 com uma dockerina de tipo II como exemplo. As setas vermelhas indicam a ação exonuclease da enzima T4 DNA Polimerase. Os círculos em amarelo indicam os pontos em que ocorre a fosforilação, através dos *primers*, para permitir a ligação *in vitro*.59

Figura 14 - Análise da purificação de Rf-CohB1 em TALON por SDS-PAGE. (M) Marcador de massa molecular, com a massa de cada banda indicada à sua esquerda. (1-8) Primeira purificação em TALON. (1) Fração insolúvel. (2) Eluato não ligado à coluna. (3-5) Lavagem com Tampão A (Tris 25 mM pH8,0, NaCl 300 mM, CaCl₂ 2 mM) adicionado de 5mM Imidazol. (6) Lavagem com Tampão A. (7-8) Lavagem com Tampão A adicionado de 300mM Imidazol. (10-14) Segunda purificação em TALON. (10) Eluato não ligado à coluna. (11-12) Lavagem com tampão A. (13-14) Lavagem com tampão A adicionado de 500mM Imidazol. O poço (9) não

foi utilizado. (B) Análise da cromatografia de exclusão molecular de Rf-CohB1. (M) Marcador de massa molecular, com a massa de cada banda indicada à sua esquerda. (1-5) Amostras de frações coletadas da cromatografia.	66
Figura 15 - Estrutura da Rf-DocA e seus dois sítios de ligação a cálcio. (A) A dockerina Rf-DocA colorida em padrão arco-íris, com N-terminal em azul, e C-terminal vermelho. Os íons de cálcio estão representados por esferas amarelas. (B) O primeiro sítio de ligação a cálcio adjacente à hélice 1. (C) O segundo sítio de ligação a cálcio, adjacente à hélice 3. Em (B) e (C), os íons de cálcio estão representados por esferas verdes.	69
Figura 16 - Sobreposição do segundo sítio de ligação a cálcio de Rf-DocA e da dockerina CttA, mostrando a inserção que ocorre no loop entre as hélices 2 e 3, semelhante nas duas proteínas. Em verde, Rf-DocA e o íon de cálcio coordenado. Em bege, a dockerina de CttA e o íon de cálcio coordenado.	70
Figura 17 - Interações que contribuem para estabilidade da estrutura da Rf-DocA. A figura mostra, em laranja, o grampo intramolecular ¹⁶ entre W802 e P878, nas regiões N-terminal e C-terminal, respectivamente. Adicionalmente, observamos a interação da região estendida do N-terminal (em rosa) com a hélice 2 (região em amarelo) e com a inserção do loop de ligação ao cálcio (em azul).....	70
Figura 18 - Sobreposição da Rf-DocA com os 3 diferentes tipos de dockerina. (A) Sobreposição da Rf-DocA com a dockerina tipo I de PDB 4FL4 (ciano). (B) Sobreposição da Rf-DocA com a dockerina tipo II de PDB 2Y3N (amarelo). (C) Sobreposição da Rf-DocA com a dockerina tipo III de PDB 4WKZ (laranja). (D) Sobreposição da Rf-DocA com a dockerina tipo III de PDB 5LXV (cinza). Rf-DocA representada em verde (A-D).....	72
Figura 19 - Estrutura da Rf-CohB4 colorida em padrão arco-íris, com N-terminal em azul, e C-terminal em vermelho.....	73
Figura 20 - Conformações do loop 6-7 da cohesina Rf-CohB4 e β -flap8, na cohesina Rf-CohE, na face de ligação de suas respectivas dockerinas. (A) Diferentes conformações do Loop 6-7 de Rf-CohB4, em ciano, fechado, em verde, aberto. (B) Diferentes conformações do β -flap8 de Rf-CohE, em cinza, fechado, em rosa, aberto.	74
Figura 21 - Sobreposição da Rf-CohB4 com os 3 diferentes tipos de cohesina. (A) Sobreposição da Rf-CohB4 com a cohesina de tipo I de PDB 3UL4 (laranja). (B) Sobreposição da Rf-CohB4 com a cohesina de tipo II de PDB 2Y3N (vermelho). (C) Sobreposição da Rf-CohB4 com a cohesina de tipo III de PDB 5AOZ (amarelo). Rf-CohB4 está representada em verde (A-C).	74
Figura 22 - Estrutura do complexo Rf-DocA/Rf-CohB4 e interações de sua interface. A dockerina (ciano) se liga à cohesina (em verde) por meio de uma rede de interações hidrofóbicas, no interior da interface (B) e por contatos	

periféricos compostos de ligações polares, como ligações de hidrogênio e pontes salinas (C e D). A interação se dá, principalmente, pelo contato entre a superfície plana das hélices 1 e 3 da dockerina com a superfície da folha da cohesina composta pelas fitas β 8, 3, 6 e 5, além dos loops que conectam as fitas 5-6, 6-7 e 8-9. Em (D), está evidenciado o contato do resíduo E873, da dockerina, com os resíduos K536 e H538, da cohesina, mostrando que esse resíduo contribui significativamente para a interação.76

Figura 23 - Assimetria entre as hélices de Rf-DocA. A Rf-DocA em ciano corresponde à presente no complexo observado em nosso modelo cristalográfico. Já em verde, é representada a Rf-DocA₁₈₀, ou seja, dockerina virtualmente rotacionada, para simulação do modo alternativo de ligação. Estão destacados os resíduos V813, V817, L868 e V872, que fazem interação com a cohesina, e seriam substituídos, respectivamente, por V861, E865, N820 e N824 na rotação da dockerina.78

Figura 24 - Alinhamento das cohesinas da escafoldina ScaB de *Ruminococcus flavefaciens*. Em verde estão destacados os principais resíduos que fazem contato com a Rf-DocA. A numeração dos resíduos, acima das seqüências, segue a numeração da Rf-CohB4 como referência.79

Figura 25 - Ensaios de calorimetria de titulação isotérmica da Rf-CohB4. (A) Titulação da Rf-CohB4 em Rf-DocA_{WT}. (B) Titulação da Rf-CohB4 em Rf-DocA_{V813E}. (C) Titulação da Rf-CohB4 em Rf-DocA_{V861E} (D) Titulação da Rf-CohB4 em Rf-DocA_{V817E}. O mutante Rf-DocA_{V861E} apresentou uma interação muito semelhante à de Rf-DocA_{WT}, enquanto Rf-DocA_{V817E} resultou em uma dinâmica de interação bastante distinta. O mutante Rf-DocA_{V813E} não apresentou interação com Rf-CohB4.83

Figura 26 - Ensaios de calorimetria de titulação isotérmica da Rf-CohB6. (A) Titulação da Rf-CohB4 em Rf-DocA_{WT}. (B) Titulação da Rf-CohB4 em Rf-DocA_{V813E}. (C) Titulação da Rf-CohB4 em Rf-DocA_{V861E} (D) Titulação da Rf-CohB4 em Rf-DocA_{V817E}. De forma semelhante às interações com Rf-CohB4, o mutante Rf-DocA_{V861E} apresentou uma interação muito semelhante à de Rf-DocA_{WT}, enquanto Rf-DocA_{V817E} resultou em uma dinâmica de interação bastante distinta. O mutante Rf-DocA_{V813E} não apresentou interação com Rf-CohB6.85

Figura 27 - Estrutura da Rf-CohB1. (A) A cohesina acima, colorida em padrão arco-íris, com N-terminal em azul, e C-terminal vermelho. (B) Vista frontal do loop que conecta as fitas β 6-7, com a medição das distâncias relativas ao resíduo G169, de posição estática entre as estruturas (em verde, a conformação fechada, em magenta, a conformação aberta).87

Figura 28 - Comparação das conformações adotadas pelo loop entre as fitas 6-7 das cohesinas Rf-CohB1 e Rf-CohB4. (A) Sobreposição das estruturas Rf-CohB1 (verde) e Rf-CohB4 (ciano), ambas nas conformações fechadas do loop 6-7. (B) Comparação das orientações dos loops de Rf-CohB1 (magenta), e Rf-CohB4 (ciano) na conformação fechada (C) Comparação

das orientações dos loops de Rf-CohB1 (magenta), e Rf-CohB4 (ciano) na conformação aberta	88
Figura 29 - Sobreposição da Rf-CohB1 com os três tipos de cohesinas. (A) Sobreposição da Rf-CohB1 com a cohesina de tipo I de PDB 4UMS (amarelo). (B) Sobreposição da Rf-CohB1 com a cohesina de tipo II de PDB 2OZN (rosa). (C) Sobreposição da Rf-CohB1 com a cohesina de tipo III de PDB 5AOZ (magenta). Rf-CohB1 está representada em verde (A-C).....	89
Figura 30 - Estrutura da cohesina Rf-CohB1 _{R129A} comparada com a conformação aberta das cohesinas Rf-CohB1 _{WT} e Rf-CohB4. A conformação do loop 6-7 em Rf-CohB1 _{R129A} é mais similar à observada em Rf-CohB4. Em magenta, Rf-CohB1 _{WT} , em ciano, Rf-CohB4, em verde e amarelo, as duas moléculas de Rf-CohB1 _{R129A} presentes no modelo.	93
Figura 31 - Sobreposição do loop 6-7 na comparação entre Rf-CohB1 _{R129A} , Rf-CohB1 _{WT} e Rf-CohB4. Observa-se uma grande divergência no ângulo adotado pelo resíduo G124 (que equivale a A542 em Rf-CohB4), ocasionado pelo contato cristalino nessa região. O resíduo R129 é um dos principais participantes desse contato, tal que sua mutação para alanina reduz o efeito desse artefato e a G124 pode adotar uma conformação mais favorável, semelhante à que ocupa A542 em Rf-CohB4. Em amarelo, Rf-CohB1 _{R129A} , em magenta, Rf-CohB1 _{WT} e em ciano, Rf-CohB4.....	93
Figura 32 - Estrutura da Rf-CohB1 _{G128A} comparada à Rf-CohB1 _{WT} nas duas conformações que observamos e à Rf-CohB4 (A) Sobreposição das quatro cadeias de Rf-CohB1 _{G128A} presentes no modelo estrutural que resolvemos (em ciano, vermelho, verde e laranja) às duas conformações observadas para a cohesina Rf-CohB1 _{WT} (em amarelo, fechada, em magenta, aberta) (B) Conformação do loop 6-7 da Rf-CohB1 _{G128A} (em amarelo), comparada às duas conformações observadas para Rf-CohB1 _{WT} , e à conformação fechada de Rf-CohB4, observada na estrutura em que foi cristalizada isolada. A orientação da G124 na Rf-CohB1 _{G128A} é mais semelhante às observadas nas duas formas da Rf-CohB1 _{WT}	94
Figura 33 - Curvas de ligação entre Rf-DocA e variantes da Rf-CohB1, em ensaios de termoforese em microescala. As escalas de todas as curvas foram normalizadas de acordo com as frações ligada e não ligada em cada experimento.....	97
Figura 34 - Clonagem através da técnica LIC, mostrando como exemplo a clonagem entre os fragmentos 3C19 e a Dockerina tipo II (PDB ID:2Y3N). (a) Gel de agarose mostrando a amplificação dos fragmentos 3C19 (1) e Dock2 (2) e amplificação após sua fusão (3). (M) Marcador de massa molecular. (b) Resultado do sequenciamento mostrando em pontilhado a duplicação na região em que deveria haver ocorrido o anelamento devido a complementaridade (LIC). O tracejado destaca o ponto onde houve a ligação entre os fragmentos. A sequência esperada para a clonagem correta dos fragmentos está indicada acima do resultado obtido. As	

sequências terminais dos dois fragmentos (3C19 e Dock2) foram incluídas como referência, indicados por seu nome na figura.99

Figura 35 - Clonagem por SLIC. (A) Amplificação dos DNAs que codificam os componentes das fusões que originam os módulos das redes proteicas bidimensionais. (M) Marcador de massa molecular. (1) 3C19 (2) 2Y3M (3) 3BUT (4) 2Y3N. Todos esses fragmentos de DNA foram amplificados com *primers* que codificam o *linker* 1 (L1). São visíveis bandas espúrias nesse gel, motivo pelo qual o DNA é recortado do gel e tratado com DpnI para degradação de DNA parental. (B) PCR para verificação de transformantes positivos para a clonagem de 3C19-2Y3M-L1. (1-3) Diferentes colônias que apresentaram contaminação na reação de amplificação, sendo descartados. (4-5) Transformantes positivos para a clonagem de 3C19-2Y3M-L1101

Figura 36 - Confirmação de transformantes por PCR. A clonagem dos módulos 2Y3N-1UDV-L2 (1), 2Y3N-1UDV-L1 (2), 1BKB-2Y3M-L2 (3), 1LJO-2Y3M-L2 (4), 2Y3N-3BUT-L1 (5) e 3C19-2Y3M-L1 (6).101

Figura 37 - Purificação da fusão proteica 3C19-2Y3N-L1. (A) Primeira purificação em TALON (Clontech) (M) Marcador de massa molecular. (1) Fração insolúvel (2) Fração solúvel (3) Eluato não ligado à coluna (4) Lavagem com tampão sem imidazol (5) Lavagem com 20 mM de imidazol (6) Lavagem com 50 mM de Imidazol (7) Lavagem com 100 mM de Imidazol (8) Lavagem com 300 mM de Imidazol (9) Lavagem com 500 mM de Imidazol (B) Segunda purificação em TALON (Clontech), para retirada de SUMO e Cromatografia de exclusão molecular (SEC) (M) Marcador de massa molecular. (1) Eluato não ligado à coluna (2) Lavagem com tampão sem imidazol (3-14) Frações coletadas da SEC, correspondendo à proteína de interesse (4-9) e SUMO (12-14).103

Figura 38 - Purificação da fusão proteica 2Y3N-3BUT-L1. (A) Primeira purificação em TALON (Clontech) (M) Marcador de massa molecular. (1) Fração insolúvel (2) Fração solúvel (3) Eluato não ligado à coluna (4) Lavagem com tampão sem imidazol (5) Lavagem com 20 mM de imidazol (6) Lavagem com 50 mM de Imidazol (7) Lavagem com 100 mM de Imidazol (8) Lavagem com 300 mM de Imidazol (9) Lavagem com 500 mM de Imidazol (B) Segunda purificação em TALON (Clontech), para retirada de SUMO e Cromatografia de exclusão molecular (SEC) (M) Marcador de massa molecular. (1) Eluato não ligado à coluna (2) Lavagem com tampão sem imidazol (3) Segunda lavagem com tampão sem Imidazol (4) Lavagem com 500 mM Imidazol (5) Segunda lavagem com 500 mM Imidazol (6-16) Frações coletadas da SEC, correspondendo à proteína de interesse (6-12) e SUMO (14-16).104

Figura 39 - Micrografias eletrônicas dos primeiros testes de formação de redes bidimensionais automontáveis, a partir das proteínas 3C19-2Y3M-L1 e 2Y3N-3BUT-L1, fusionadas por *linker* rígido de 5 resíduos. (A) Nanotubos formados com incubação de 30 minutos no grid. (B) Filmes bidimensionais formados com incubação de 30 minutos no grid. (C) Filmes bidimensionais sobrepostos formados com incubação de 4 horas

	diretamente no grid. (D) Nanotubos formados com incubação de 30 minutos em solução. (E) Nanotubos formados por incubação de 4 horas em solução. (F) Filme bidimensional formado com incubação de 4 horas em solução.	105
Figura 40 -	Micrografias eletrônicas do segundo teste de formação de redes bidimensionais automontáveis, com concentração menor das proteínas 3C19-2Y3M e 2Y3N-3BUT. A diminuição da concentração de proteínas inviabilizou a formação de filmes bidimensionais, sendo visível a nucleação destes. (A) Pontos de nucleação de filmes bidimensionais. (B) Pontos de nucleação de filmes bidimensionais em maior aumento.....	106
Figura 41 -	Micrografias eletrônicas do terceiro teste de formação de redes bidimensionais automontáveis, com grids carregados positivamente e negativamente. Nos grids positivamente carregados (A), a formação de filmes foi dificultada, sendo observados apenas em pequenas extensões, próximas das grades de cobre que compõem o <i>grid</i> . Nos <i>grids</i> carregados negativamente (B) e (C), os filmes bidimensionais foram observados em toda sua extensão, sendo presentes em múltiplas camadas em alguns pontos (C).....	107
Figura 42 -	Rede bidimensional automontável visualizada em alta resolução. (A) Componentes da rede, em verde, a proteína tetramérica ALAD, de <i>E.coli</i> , em amarelo, a streptavidina em complexo com streptagI. (B) Representação do planejamento das redes formadas pela mistura da proteína tetramérica ALAD fusionada à StreptagI, com Streptavidina dimérica. (C) Micrografia de alta resolução mostrando o padrão simétrico formado na rede automontada dessas proteínas.....	108

LISTA DE TABELAS

Tabela 1 -	Parâmetros estatísticos da coleta, processamento e refinamento dos dados cristalográficos do complexo Rf-DocA/Rf-CohB4 e Rf-CohB4 isolada. Os dados em parênteses correspondem faixa de maior resolução.....	67
Tabela 2 -	Classificação de similaridade estrutural à Rf-DocA. O ranking foi editado para retirada de duplicações, quando uma mesma entrada possui duas cadeias com similaridade à proteína consultada. (n) indica a posição do resultado no ranking.	71
Tabela 3 -	Classificação de similaridade estrutural à Rf-CohB4. O ranking foi editado para retirada de duplicações, quando uma mesma entrada possui duas cadeias com similaridade à proteína consultada. (n) indica a posição do resultado no ranking. * A cohesina Rf-CohB1 foi incluída, com maior similaridade, mas não tem número do ranking pois não está depositada no PDB.	75
Tabela 4 -	Interações intermoleculares entre Rf-DocA e Rf-CohB4. O tipo de interação está indicado na terceira coluna: Interação Hidrofóbica (HF), Ligação de Hidrogênio (LH), Ponte Salina (PS).....	77
Tabela 5 -	Análise de resíduos que fariam interação entre as dockerinas Rf-DocA e Rf-DocA ₁₈₀ , e as cohesinas Rf-CohB4 e Rf-CohB6. O levantamento foi feito com base nos resíduos de interação entre Rf-DocA e Rf-CohB4, observados na estrutura do complexo, e no alinhamento entre as cohesinas de ScaB. Os motivos da cohesina destacados com (*) indicam que não é observada interação nessa posição, mas é a vizinhança que acomoda o resíduo V817, na estrutura Rf-DocA/Rf-CohB4, e apresenta uma divergência significativa para a cohesina Rf-CohB6.....	79
Tabela 6 -	Parâmetros termodinâmicos da interação entre as cohesinas Rf-CohB4 e Rf-CohB6 e as dockerinas Rf-DocA _{WT} e mutantes Rf-DocA _{V813E} , Rf-DocA _{V817E} e Rf-DocA _{V861E}	83
Tabela 7 -	Parâmetros estatísticos da coleta, processamento e refinamento dos dados cristalográficos de ambas as formas cristalinas de Rf-CohB1. Os dados em parênteses correspondem faixa de maior resolução.	86
Tabela 8 -	Classificação de similaridade estrutural à Rf-CohB1. O ranking foi editado para retirada de duplicações, quando uma mesma entrada possui duas cadeias com similaridade à proteína consultada. (n) indica a posição do resultado no ranking. *A cohesina Rf-CohB4 foi incluída, com maior similaridade, mas não tem número do ranking pois não está depositada no PDB.	89
Tabela 9 -	Parâmetros estatísticos da coleta, processamento e refinamento dos dados cristalográficos dos mutantes Rf-CohB1 _{G128A} e Rf-	

CohB1 _{R129A} . Os dados em parênteses correspondem faixa de maior resolução.....	92
Tabela 10 - Constantes de dissociação entre Rf-DocA e variantes da Rf-CohB1 obtidas através de ensaios de termoforese em microescala.	96
Tabela 11- Detalhes das construções dos componentes das redes automontáveis de proteínas.....	102
Tabela 12 - Relação das proteínas sintetizadas por GenScript. Os tipos de proteínas (C2, C3, C4 e C6) indicam sua simetria interna. As referências correspondem ao PDB ID de cada, quando disponível. Referências indicadas com asterisco (*) referem-se a pseudônimos utilizados em nossos documentos para diferenciar componentes de uma mesma entrada PDB. As dockerinas e cohesinas de <i>Ruminococcus flavefaciens</i> não requereram síntese, pois já se encontravam em nosso laboratório, sendo incluídas nessa tabela apenas para apresentação de sua sequência. Assim, foram pedidos os <i>primers</i> para amplificá-las, apenas.	115
Tabela 13 - Relação de <i>primers</i> sintetizados para amplificação dos insertos componentes das construções de redes proteicas auto montáveis. As sequências estão agrupadas pelo tipo de proteína, simetrias C2, C3, C4 e C6, Dockerinas e Cohesinas, além dos <i>primers</i> utilizados para amplificação do vetor. Os sufixos indicam o tipo de <i>linker</i> composto pelo primer, sendo: SGEEARRGS (L1), GSEAARREAEARGS (L2) GSGGSG (L3).	119

Sumário

1	Introdução.....	23
1.1	Celulossomo – uma nanomáquina celulolítica.....	23
1.1.1	Arquitetura e função dos celulosossomos	23
1.1.2	Cohesinas e dockerinas: os componentes centrais do celulosomo	26
1.1.3	Celulosomo de <i>Ruminococcus flavefaciens</i>	32
1.1.4	Aplicações envolvendo Dockerinas e Cohesinas	37
1.2	Redes automontáveis de proteínas	38
2	Objetivos	43
3	Materiais e Métodos	45
3.1	Sub-clonagem dos componentes celulosossomais de <i>Ruminococcus flavefaciens</i> em vetor de expressão	45
3.2	Transformação de células BL21(DE3) para expressão dos módulos cohesina e dockerina de <i>R. flavefaciens</i>	45
3.3	Expressão e purificação das cohesinas e dockerinas de <i>R. flavefaciens</i>	46
3.4	Cristalização das cohesinas Rf-CohB1 e Rf-CohB4 de <i>R. flavefaciens</i> e complexo Rf-DocA/Rf-CohB4	47
3.5	Difração e resolução da estrutura das cohesinas 1 e 4 de <i>R. flavefaciens</i> e complexo Rf-DocA/Rf-CohB4	48
3.6	Cristalização dos mutantes da cohesina Rf-CohB1	49
3.7	Difração e resolução da estrutura dos mutantes de Rf-CohB1	49
3.8	Calorimetria de Titulação Isotérmica	49
3.9	Termoforese em Microescala	50
3.10	Planejamento da clonagem para produção de redes auto montáveis de proteínas	52
3.11	Clonagem por LIC	60
3.12	Clonagem por SLIC.....	61
3.13	Transformação de células DH5α ultra competentes.....	62
3.14	Expressão e purificação das fusões	62
3.15	Microscopia Eletrônica de Transmissão com coloração negativa.....	63
4	Resultados	65
4.1	Expressão e purificação de cohesinas e dockerinas de <i>R. flavefaciens</i>	65
4.2	Determinação estrutural do complexo Rf-DocA/Rf-CohB4 e Rf-CohB4 isolada	66

4.2.1 Análise estrutural da dockerina de ScaA presente no complexo Rf-CohB4/Rf-DocA	68
4.2.2 Análise estrutural do quarto módulo cohesina de ScaB presente no complexo Rf-CohB4/Rf-DocA	72
4.2.3 Estrutura do complexo Rf-DocA/Rf-CohB4.....	75
4.3 Calorimetria de Titulação Isotérmica para estudar o modo de ligação de Rf-DocA com cohesinas de ScaB	81
4.4 Determinação estrutural de Rf-CohB1 isolada.....	85
4.4.1 Estrutura da Rf-CohB1.....	86
4.5 Cristalização e resolução estrutural de mutantes pontuais de Rf-CohB1	91
4.6 Análise da influência do loop 6-7 na afinidade da Rf-CohB1 à Rf-DocA por termoforese em microescala	95
4.7 Clonagem por LIC.....	98
4.8 Clonagem por SLIC	99
4.9 Expressão e purificação dos módulos componentes das redes automontáveis	102
4.10 Análise de formação das redes proteicas automontáveis por microscopia eletrônica de transmissão	104
5 Conclusão.....	109
Referências.....	111
Apêndice A – Tabela de sequências sintetizadas para clonagem	115
Apêndice B – Tabela de Primers.....	119

1 Introdução

1.1 Celulossomo – uma nanomáquina celulolítica

1.1.1 Arquitetura e função dos celulossomos

Celulossomos são complexos multienzimáticos extracelulares desenvolvidos por bactérias anaeróbias para superar a limitação energética da produção de enzimas para a degradação de parede celular vegetal (Figura 1). Na montagem desse sistema, a conexão das subunidades enzimáticas com a estrutura central não-catalítica, escafoldina, é mediada por Dockerinas e Cohesinas, módulos de alta especificidade e afinidade e, portanto, centrais para a arquitetura e composição dos celulossomos. As subunidades enzimáticas dotadas de dockerinas se ligam aos módulos cohesina contidos na escafoldina, de modo que as enzimas atuam em sinergia. A natureza modular desses sistemas os torna organizados e eficientes, possibilitando a reciclagem enzimática e a direta assimilação de produtos metabólicos.²⁻⁴

Figura 1 - Micrografia Eletrônica de bactérias produtoras de celulossomo em uma célula vegetal. Os celulossomos são identificados pelos distintos pontos pretos anexos à superfície celular.

Fonte: Adaptada de GÜLLERT.⁵

Celulossomos foram observados em bactérias mesofílicas e termofílicas, englobando um grande número de espécies do gênero *Clostridium*, *Acetivibrio cellulolyticus*, *Bacteroides cellulosolvens*, além de algumas espécies do gênero *Ruminococcus*. Celulossomos simples, até hoje somente observados em bactérias mesofílicas, possuem apenas uma escafoldina que incorpora tantas enzimas quanto há cohesinas em sua composição. O modo de ancoragem dessas escafoldinas à parede celular bacteriana é desconhecido. Essas escafoldinas são classificadas como Escafoldinas Primárias, mais expressas, dotadas de muitos domínios cohesina, capazes de interação com as enzimas e, portanto, mais importantes no sistema. Geralmente contém um CBM, módulo de ligação a carboidrato, que direciona o complexo ao substrato. As escafoldinas de ancoragem, presentes nos sistemas mais elaborados, são responsáveis por anexar o complexo à superfície celular. São dotadas de um variado número de cohesinas e são incorporados na parede celular através de domínios SLH, ou covalentemente, através de motivos *Sortase*. Ainda nos sistemas mais elaborados, são encontradas as escafoldinas adaptadoras, que conectam duas escafoldinas, geralmente intermediando os dois tipos anteriores, tendo principalmente o papel de amplificar o complexo de forma a comportar mais enzimas. Alternativamente, podem intermediar a interação entre uma escafoldina e uma enzima (quando possuem uma única cohesina), o que tem um papel regulatório na organização do celulossomo, podendo alterar a afinidade de uma enzima que inicialmente não se encaixaria na escafoldina. Esse evento tem grande impacto na adaptabilidade das bactérias a diferentes substratos.²

A comunicação entre esses diferentes tipos de escafoldina e enzimas ocorre de forma organizada, em decorrência do reconhecimento específico entre os diferentes tipos de dockerinas e cohesinas. As dockerinas e cohesinas do tipo-I tem o papel de conectar enzimas às escafoldinas primárias. Já as de tipo-II medeiam a ligação entre escafoldinas, geralmente entre as de ancoragem e as primárias ou adaptadoras. Esse segundo tipo de dockerinas é acompanhado de um módulo X, hidrofóbico, com o qual faz interações que a estabilizam estruturalmente, além de promover um distanciamento da superfície celular. Em *Bacteroides cellulosolvens*, curiosamente, as funções dos tipos são invertidas, de modo que as enzimas possuem dockerinas de tipo-II. Há ainda o tipo-III, em *Ruminococcus flavefaciens*, que foi assim classificado devido a divergências significativas em sua sequência. As particularidades de cada tipo serão discutidas em mais detalhes abaixo.²

O celulossomo de *Clostridium thermocellum* foi o primeiro a ser descoberto, em 1983, se tornando o mais bem caracterizado e, portanto, considerado modelo desse sistema. Nele, a

escaffoldina primária CipA possui 9 cohesinas capazes de se ligar a enzimas através de dockerinas do tipo-I, um Módulo de ligação a carboidratos (CBM), e em seu C-terminal, um Módulo X seguido de uma dockerina tipo-II, que pode se ligar a três diferentes scaffoldinas de ancoragem, OlpB, Orf2p ou SdbA, que possuem 7, 2 e 1 cohesinas, respectivamente (Figura 2). O caráter multimodular dos celulosossomos torna difícil a cristalização de partes grandes do sistema, de forma que estruturas são obtidas de pequenos complexos entre módulos de forma isolada. Uma combinação entre outros métodos, como espalhamento de raios-X a baixo ângulo (SAXS), crio-microscopia eletrônica e outras análises computacionais fornecem conhecimento sobre fragmentos maiores. Com isso, foi observado que *linkers* - que não tem efeito sobre a interação dockerina-cohesina - são intrinsicamente flexíveis, mas adotam uma conformação compactada na região central de CipA, alternando a orientação em que direcionam as cohesinas, de modo que as enzimas são intercaladas evitando impedimento estérico. Outra forma que as dockerinas de tipo-I possuem de evitar impedimento estérico é através dos dois modos de ligação possíveis à cohesina, que será discutido em mais detalhes abaixo.⁴

A interação entre dockerina e cohesina é uma das mais fortes conhecidas na natureza, sendo a força necessária para dissociação desses componentes cerca de metade da necessária para a quebra de ligações covalentes. Isso se justifica pela necessidade de manter a estabilidade do sistema no ambiente em que esse sistema atua, exposto ao ambiente extracelular, com condições adversas e variedade de substratos e organismos. Nesse cenário, há muitos organismos não celulolíticos que se aproveitam dos produtos da degradação da parede celular vegetal pelos celulosossomos, que os fornecem substratos consumíveis. Nesse sentido, o celulosossomo pode ser também uma ferramenta de estabelecimento de relações ecológicas. Inclusive, foram identificadas diversas proteínas de diferentes funções associadas a dockerinas, em organismos não celulolíticos, em alguns casos sem a presença de cohesinas em seu genoma, além de muitas vezes, as sequências serem divergentes o suficiente para não se encaixar em nenhum dos três tipos conhecidos de dockerinas e cohesinas. Isso indica que a montagem dos celulosossomos é apenas uma das funções conhecidas das dockerinas e cohesinas.²⁻⁴

Figura 2 – Desenho representativo de um celulosomo e seus componentes, tomando como exemplo a escafoldina CipA, de *Clostridium thermocellum*. A escafoldina, em amarelo, possui cohesinas do tipo I, para integração de enzimas através da ligação com dockerinas de tipo I, em vermelho, que estas carregam. Possui ainda um CBM, módulo de ligação a celulose, que ancora o complexo ao substrato. Sua dockerina C-terminal de tipo II, em verde, que porta um módulo X, em roxo, que se liga a uma cohesina tipo II, incorporada na superfície celular, o que ancora todo o complexo à célula.

Fonte: Adaptada de SMITH.⁴

1.1.2 Cohesinas e dockerinas: os componentes centrais do celulosomo

Entender a relação entre os módulos cohesina e dockerina, além da relação entre cada módulo e os elementos adjacentes nos permite entender melhor a arquitetura e o comportamento do celulosomo, além de permitir usar os pares dockerina-cohesina para aplicações biotecnológicas e nanotecnológicas. A especificidade de cada tipo de dockerina-cohesina é preservada de modo a não haver reações cruzadas entre módulos de tipos diferentes, o que se deve às divergências características de cada um dos tipos. No entanto, há

alguns exemplos de reconhecimento interespecífico, em que dockerinas de um organismo reconhecem cohesinas de um outro, sendo mais comum nos sistemas simples de organismos mesofílicos. No tipo II, a promiscuidade é ainda maior, por exemplo, com a cohesina de SdbA de *Clostridium thermocellum*, que reconheceu dockerinas de tipo II de sua própria espécie, de *Bacteroides cellulosolvens* e *Acetivibrio cellulolyticus*, sendo cogitada a possibilidade de se tratar de uma cohesina universal para tipo II.⁶ A ideia de que há, em cada dockerina, um par de resíduos de reconhecimento à cohesina é bastante difundida, sobretudo em dockerinas do tipo I, em que a simetria entre as hélices permite verificar o duplo modo de ligação através da mutação desse par de resíduos. Há inclusive estudos que invertem a especificidade entre dockerinas por meio da inversão do par de reconhecimento entre elas.⁷ Outra constatação interessante com relação à simetria na dockerina é a possibilidade de inverter a ordem de seus segmentos simétricos, ou mesmo ter um deles duplicado, sem prejuízo para a interação.⁸

Outras regiões foram verificadas importantes para a ligação entre cohesina e a dockerina. O N-terminal da dockerina é essencial para a manutenção da estrutura, sendo visto que sem esses resíduos, a estruturação das hélices diminui, resultando num aumento de estruturas desordenadas. Os íons de cálcio também são imprescindíveis para a ligação, apesar de sua retirada não alterar a estrutura da dockerina. O uso de EDTA pode inibir a ligação, e inclusive dissociar dockerinas e cohesinas ligadas, quando são expressas sem fusões. Curiosamente, a presença de *tags* ou outros módulos fusionados a dockerina, como enzimas, e a cohesina, como *linkers*, módulos-X, ou CBMs, não interferem na ligação.⁸ Há estudos que exploraram essa propriedade, convertendo enzimas livres, de organismos não secretores de celulosomos, em componentes de celulosomos, sem prejuízo para a funcionalidade de nenhum dos módulos.⁹ A fusão desses módulos, no entanto, pode proteger a interação dockerina-cohesina da ação do EDTA, provavelmente bloqueando seu acesso ao cálcio. A interação, como já citado, é muito estável e de afinidade alta, com Kd tipicamente na ordem de nanomolar. Por esse motivo, foi observado que a injeção adicional de cohesina em experimentos com os módulos já em complexo não foi capaz de dissociá-los. Além disso, fatores como variação de pH de 5 a 9 e temperatura até 50^o C tampouco afetaram significativamente a interação.⁸

A divisão de cohesinas em tipos I e II origina da divergência encontrada em *C. thermocellum* e foi feita de acordo com suas sequências primárias (Figura 3). Consequentemente, as dockerinas com as quais faziam interação se encaixaram no mesmo tipo. Assim, as 9 cohesinas tipo-I presentes na escafoldina primária CipA interagem com dockerinas portadas por enzimas, enquanto as cohesinas da escafoldinas de ancoragem, que

interagem com a dockerina C-terminal de CipA compõem o tipo-II. No momento de suas descobertas, novas cohesinas são classificadas de acordo com a similaridade às de *C. thermocellum*. Ao serem incluídas nessa análise, as sequências de *Ruminococcus flavefaciens* apresentaram uma maior divergência filogenética, sendo separadas no tipo III (Figura 3). Uma única cohesina, a de ScaC, apresentou similaridade com tipo-I. No entanto, a significância biológica dessa separação é questionável, uma vez que não traduz as diferentes funcionalidades de seus integrantes. Essa divergência interna do tipo-III causou sua adicional separação em grupos, que serão discutidos abaixo, esclarecendo razoavelmente essa questão. Outras cohesinas, ainda, são dispostas em ramos separados da árvore filogenética, não se encaixando em nenhum dos tipos conhecidos, como é o caso da escafoldina de *Clostridium acetobutylicum* e as de *Archaeoglobus fulgidus*.⁶

Figura 3 - Árvore filogenética a partir da sequência primária de cohesinas. A relação filogenética entre cohesinas determina sua separação em diferentes tipos. Os três tipos conhecidos estão agrupados, como indicado pelos arcos na figura. A barra da escala indica a porcentagem (0.1) de substituição de aminoácidos.

Fonte: Adaptada de HAIMOVITZ.⁶

Apesar das particularidades dos tipos de cohesinas e dockerinas, há uma estrutura comum a cada módulo. As cohesinas são compostas de cerca de 150 resíduos de aminoácidos, dispostos em 9 fitas- β que formam um sanduíche, com topologia de rocambole (*jellyroll*), em que as fitas são organizadas em antiparalelo, com exceção das fitas 1 e 9 que são paralelas entre si. As fitas- β 9, 1, 2, 7 e 4 formam uma das faces do sanduíche, enquanto a outra é formada pelas fitas- β 8, 3, 6 e 5, à qual a dockerina se liga (Figura 4).¹⁰

As dockerinas são formadas de cerca de 70 resíduos de aminoácidos, dois segmentos de 22 aminoácidos que incluem duas α -hélices, formando uma superfície planar que possibilita a interação com a cohesina. Cada um desses segmentos contém em seu N-terminal um domínio *F-hand* de 12 resíduos, similar ao *EF-hand*, que liga cálcio, o qual induz o enovelamento do módulo (Figura 4). Inicialmente foi sugerido que as dockerinas apresentavam uma estrutura fracamente ordenada e adquiriam sua conformação ao interagir com cohesinas. Posteriormente, estruturas geradas por difração de raios-X e outros experimentos como NMR verificaram que a dockerina apresentava sua conformação apta à interação com cohesina de forma inerente e dependente de cálcio.¹⁰

Figura 4 - Estrutura das dockerinas e cohesinas de tipos I e II. Todos os módulos retratados são de *Clostridium thermocellum*. Em vermelho, a dockerina do tipo I de Xilanase I, em ciano a cohesina 2 de CipA (PDB ID: 1OHZ). Em laranja, dockerina do tipo II, de CipB, que possui um módulo-X, em roxo; em amarelo, a cohesina 2 de Orf2p (PDB ID:5K39). As esferas em amarelo representam íons de cálcio. As cohesinas estão orientadas de forma que sua face frontal corresponde à que contata a dockerina.

Fonte: Elaborada pelo autor.

As dockerinas de tipo I apresentam alta simetria entre as hélices, porém, sua interação com cohesinas ocorre de maneira assimétrica, ou seja, uma das hélices domina a interação, tendo uma proximidade maior com a cohesina. No geral, a hélice C-terminal é dominante na interação, fazendo contato com a cohesina por toda sua extensão, enquanto a hélice N-terminal contata a cohesina apenas em sua porção C-terminal. Nessa interação prevalecem ligações de hidrogênio, mas também há a contribuição de interações hidrofóbicas. A grande similaridade entre as hélices possibilita que o acoplamento à cohesina ocorra em dois modos distintos, ou seja, caso a dockerina seja rotacionada em 180^0 , a hélice N-terminal passa a ser dominante na interação, e o restante dos pontos de contato entre os módulos permanece homólogo. Este fenômeno, conhecido como duplo modo de ligação, já foi verificado em estruturas e ensaios de afinidade, em que a hélice C-terminal é mutada de maneira que a orientação alternativa se torna preferencial, e a afinidade é conservada. Acredita-se que essa propriedade possa, também, prevenir impedimento estérico na montagem do celulosomo, além de potencialmente prover flexibilidade para auxiliar acesso da enzima a seu substrato. Apesar de ser característico de tipo-I, não está presente em todos seus integrantes, e portanto, não é determinante para a classificação da dockerina.^{2,10}

O tipo II não apresenta simetria interna, salvo exceções, o que sugere baixa probabilidade de ocorrência do duplo modo de ligação, o qual ainda não foi observado nesse grupo. O mesmo ocorre para tipo III. Estruturalmente, tipo-II difere do tipo-I por apresentar uma α -hélice entre as fitas- β 6 e 7, na “coroa” da cohesina, além de 2 β -flaps que interrompem o curso normal das fitas- β 4 e 8. Seu modo de ligação à dockerina é através de toda a extensão de ambas as hélices, e a orientação da dockerina no complexo é girada de 20^0 no sentido anti-horário, em relação ao observado em tipo-I. Nessa ligação, há uma contribuição muito maior de interações hidrofóbicas. O módulo-X, já mencionado, acompanha a dockerina do tipo-II, com a qual faz interações hidrofóbicas que auxiliam na estabilidade estrutural, mantém um afastamento entre a célula e o celulosomo, além de aumentar a afinidade entre as dockerinas e suas cohesinas parceiras.^{2,10-11}

O Tipo-III inclui uma grande diversidade de dockerinas e cohesinas que apresentam divergências dos tipos I e II em sua sequência primária, porém, alguns elementos estruturais e funcionais em comum a cada tipo. No momento da redação desse trabalho, apenas três estruturas contendo complexos entre cohesinas e dockerinas do tipo-III estão disponíveis (Figura 5), sendo a primeira entre a díade Módulo-X:Doc de CttA e a Cohesina da escafoldina ScaE, a segunda entre uma dockerina do grupo 3 e a cohesina da escafoldina ScaC e a terceira, muito semelhante à primeira, envolve a cohesina de ScaG e uma díade módulo-

X:dockerina. A cohesina de ScaE (PDB:4IU2) apresenta algumas semelhanças com as de tipo-II, como β -flaps 4 e 8, porém não apresenta a hélice ligando as fitas- β 6 e 7. O elemento mais distinguível dessa cohesina é uma hélice proeminente de 13 resíduos, seguida de um loop de 17 resíduos contendo uma segunda hélice de 6 resíduos ligando as fitas- β 8 e 9, além de um extenso N-terminal de 24 resíduos que envolve essas estruturas fazendo numerosas interações de hidrogênio que conferem estabilidade. A dockerina é mais divergente das demais conhecidas, contendo três inserções que culminam num total de 5 α -hélices. As hélices 1 e 3 são antiparalelas e formam a superfície planar que interage com a cohesina, como ocorre nas demais dockerinas, porém, na hélice 2, ocorre uma inserção que provoca uma alteração no motivo *F-hand*, ocasionando que o segundo sítio de cálcio desse módulo seja considerado atípico. É presente ainda um módulo-X, bastante alongado, composto de 5 fitas- β e 2 hélices conectando as fitas 1-2 e 3-4, que faz interações hidrofóbicas com as inserções e hélices adicionais da dockerina. A interação entre os dois módulos ocorre de maneira semelhante ao tipo-II, ou seja, através da extensão de ambas as hélices, porém com a diferença de ser rotacionada em 190° comparado ao tipo-I. Assim, a orientação é mais semelhante a estruturas que mostram o modo de ligação alternativo de tipo-I, que tem a hélice 1 como dominante da interação. Há uma outra entrada no PDB, de referência 2ZF9, que traz apenas a cohesina de ScaE da linhagem 17 de *R. flavefaciens*, cuja maior diferença para esta, da linhagem FD-1, é a disposição do β -flap 8, $10\text{\AA}$ mais recolhido ao centro da cohesina. Isso sugere uma mudança conformacional na cohesina para acomodar a ligação à dockerina.^{2,11-12}

Figura 5 - Estruturas de complexos do Tipo III. (A) Díade Módulo-X:Doc de CttA e a Cohesina da escafoldina ScaE de PDB 4IU2. (B) Uma dockerina do grupo 3 e a cohesina da escafoldina ScaC de PDB 5LXV. (C) cohesina de ScaG e uma díade módulo-X:dockerina de PDB 4WKZ.

Fonte: Elaborada pelo autor.

A cohesina de ScaC (PDB: 5LXV) tem um β -flap 8 mais extenso que o comum, com 17 resíduos, e a hélice entre fitas- β 6 e 7, que são características semelhantes às cohesinas de tipo-II. No entanto, há ainda uma hélice entre as fitas- β 4 e 5, um aspecto distinguível dessa cohesina. A dockerina, membro do grupo 3 e contida numa proteína multimodular de função desconhecida, apresenta a estrutura básica conhecida para outras dockerinas, sem peculiaridades. A interação entre os módulos ocorre também com ambas as hélices, tendo a hélice entre as fitas- β 4 e 5 e o β -flap 8 da cohesina como margens da região de interação. Apesar disso, a deleção do β -flap afeta na afinidade entre os módulos. Curiosamente, a cohesina de ScaC se situa mais próxima a cohesinas de tipo-I nas árvores de filogenia, a despeito de suas semelhanças funcionais e estruturais com o tipo-II.¹³

Existem ainda algumas estruturas disponíveis de cohesinas de tipo-III desacompanhadas de dockerinas, com outras peculiaridades, que corroboram a inexistência de um consenso com relação a características do tipo III, visto que há uma diferença grande de similaridade entre seus membros. Nas 223 dockerinas de FD-1 essa similaridade varia de 20 a 98%. Essa ampla variação motivou a subsequente divisão das dockerinas de tipo-III em seis grupos e onze subgrupos, que serão discutidos em maiores detalhes abaixo.

1.1.3 Celulossomo de *Ruminococcus flavefaciens*

Ruminococcus flavefaciens é uma espécie de bactéria Gram positiva, anaeróbia, e a única do rúmen bovino a apresentar um celulossomo definido. Seu celulossomo é o mais complexo conhecido atualmente, contendo mais de 220 sequências portadoras de dockerina, sendo notável o número de sequências de hemicelulases e o de proteínas não relacionadas a celulossomo, como peptidases, serpinas e transglutaminases, além de numerosas sequências de módulos sem função conhecida, não similares a qualquer outra proteína. Todas as suas linhagens possuem o cluster gênico *sca*, que contém as sequências codificadoras para escafoldinas ScaC, ScaA, ScaB, ScaE e CttA e, portanto, são capazes de expressar um celulossomo. A proteína CttA é uma proteína que possui dois CBM fusionados a uma dockerina, o que revela uma particularidade do celulossomo desse organismo, já que o modelo conhecido em outros organismos é possuir um domínio CBM em uma de suas principais escafoldinas, entre as sequências de cohesina. Essa proteína reconhece exclusivamente a cohesina de ScaE, ancorada covalentemente à célula através de um domínio canônico reconhecido por *sortases*. A escafoldina ScaB, que na linhagem 17 é adaptadora e

na FD-1 acumula também a função de primária, também tem como única afinidade a cohesina de ScaE, de modo que essas duas dockerinas apresentam um módulo-X e inserções em sua sequência, as quais adicionam hélices que promovem maiores interações com o módulo-X, para dar maior suporte ao complexo que carregam (Figura 6).^{1,6,14}

A divisão das dockerinas em grupos foi feita levando-se em conta sequência, elementos da estrutura secundária, resíduos presentes em regiões específicas, resíduos de ligação a Ca^{2+} e a cohesinas. Os grupos têm características únicas, compartilhadas por seus membros, porém, essas não refletem diferenças no reconhecimento de cohesinas, sendo cogitado que ocasionem diferenças na estabilidade da interação com as cohesinas, ou com os módulos funcionais que carregam. A maior diferença entre as linhagens de *R. flavefaciens* está no número e especificidade das cohesinas de ScaB. Em FD-1 há 9 cohesinas, sendo as quatro primeiras similares às duas de ScaA, agregando módulos funcionais ao celulosossomo, enquanto as outras 5 reconhecem exclusivamente a dockerina de ScaA. Apesar disso, há homólogos entre essas linhagens com diferenças que os colocam em diferentes grupos nessa análise.^{1,14}

Figura 6 - Arquitetura do celulosossomo de *Ruminococcus flavefaciens*, linhagem 17, um dos mais complexos conhecidos atualmente. Nesse sistema, há uma proteína específica para ligação ao substrato, CttA, um grande número de enzimas e escafoldinas adaptadoras.

Fonte: Adaptada de BAYER.¹⁵

O grupo 1 é o mais numeroso, subdividido em quatro subgrupos (a-d), contendo 96 dockerinas, que não apresentam similaridade com outras dockerinas conhecidas. Possuem na hélice do primeiro sítio de cálcio o motivo conservado VxIMQxxxNP, e na do segundo, ALxIQKxxL. Possuem a maior região *linker* entre os dois sítios de cálcio, tendo de 37 a 45

resíduos, sendo o subgrupo 1b o maior, devido a uma inserção conservada. O segundo sítio de cálcio tem um número atípico de resíduos nesse grupo, sendo a norma 12 resíduos, enquanto os subgrupos 1a, 1b e 1c tem 13 resíduos e 1d, 11. Reconhecem cohesinas de ScaA e suas similares em ScaB, na linhagem FD-1. O grupo 1b contém ScaC, uma escafoldina adaptadora importante na adaptabilidade do celulosomo a diferentes substratos, a qual possui duas inserções únicas em sua sequência. A maioria das escafoldinas, no entanto, foi classificada em grupos contendo poucas enzimas. Um grande número de glicosil hidrolases, carboidrato esterases e CBMs estão presentes nos subgrupos 1a, 1b e também grupo 3, os quais incluem a maioria das proteínas multidomínio observadas, geralmente associadas a xilanases ou hemicelulases. Por esse motivo, há um grande número de peptidases, CBM e algumas proteínas que contém cohesinas nesses três grupos, e ainda algumas no grupo 6. A maioria das celulases confirmadas se encontra no subgrupo 1a e um menor número em 1b, assim como carboidrato esterases e polissacarídeo liases. Deste último tipo de enzimas, nenhuma foi encontrada nos subgrupos 1c e 1d, nem nos grupos 2 e 4.^{1,14}

O grupo 2 também contém dockerinas sem similaridade com outras conhecidas, porém, com a peculiaridade de apresentarem apenas um sítio de cálcio, semelhante ao C-terminal das dockerinas de grupo 4, aparentando ser versões truncadas destas.^{1,14}

O grupo 3 tem dockerinas homólogas às clostridiais, também homólogas à proteína CE3B, conhecida ligante de ScaC. Apresentam uma variação incomum na hélice do primeiro sítio de cálcio e na região *linker* entre os dois sítios. A maior parte das enzimas nesse grupo, bem como o grupo 6 constitui hemicelulases.^{1,14}

O grupo 4 contém dockerinas exclusivas a FD-1, sendo as únicas a apresentar simetria interna. Os membros do subgrupo 4a são similares às dockerinas de ScaB e CttA, conhecidas por reconhecer cohesina de ScaE, ancorada à célula, mesmo não contendo as inserções apresentadas por estas, nem o módulo-X. As do subgrupo 4b se diferenciam do 4a por ter um *linker* entre os sítios de cálcio menor, um padrão diferente de resíduos conservados nas duas hélices dos seus sítios de cálcio e 15 resíduos em seu primeiro sítio de cálcio. Os 6 membros desse subgrupo integrados a proteínas que possuem domínios semelhantes a transglutaminases, peptídeo sinal e C-terminal *upstream* à dockerina conservados. Dessas, 5 proteínas estão justapostas cabeça-a-cauda, sendo potencialmente controladas por um operon. Cerca de metade das dockerinas associadas a proteases encontra-se no grupo 4.^{1,14}

O grupo 5 possui apenas a dockerina de ScaA, escafoldina primária que possui características de sequência únicas e desempenha um papel central nos celulosomos das

diferentes linhagens de *R. flavefaciens*. O N-terminal dessa dockerina é bastante conservado em todas essas linhagens.^{1,14}

No grupo 6, temos homólogas de algumas dockerinas de *Clostridium* e *Ruminococcus*. No subgrupo 6a, há um resíduo conservado em cada sítio de cálcio, na primeira hélice e no *linker* entre os sítios de cálcio. O subgrupo 6b apresenta o menor *linker*, com 32 resíduos.^{1,14}

Outras 8 sequências não se encaixaram em nenhum grupo, por baixa similaridade. Há ainda 118 módulos de função desconhecida, que não apresentaram similaridade com proteínas conhecidas. A maioria das sequências abordadas nesse estudo apresentava peptídeo sinal N-terminal, 156 tinham a dockerina em seu C-terminal, 32 eram internas e 19 no N-terminal, após o peptídeo sinal.^{1,14}

A organização em grupos das dockerinas de *R. flavefaciens* refletia a similaridade encontrada entre suas sequências, mas sua significância funcional só foi entendida recentemente, e os padrões de seletividade estabelecidos entre cada grupo nos permite entender um pouco melhor a arquitetura desse celulosomo (Figura 7). Os membros do grupo 1 contém predominantemente enzimas, CBMs e algumas escafoldinas, que reconhecem as cohesinas de ScaA e suas semelhantes em ScaB de FD-1. Um dos membros desse grupo A é a escafoldina ScaC, uma adaptadora monovalente, que assume notada importância na adaptabilidade da bactéria a diferentes substratos, visto que permite a ligação de enzimas que não tem afinidade diretamente com as cohesinas de ScaA e similares. A cohesina de ScaC é semelhante a cohesinas do tipo 1, em sua sequência primária, e é reconhecida pelas dockerinas dos grupos 3 e 6, em sua maioria hemicelulases. Estas, no entanto, apresentam assimetria em suas hélices, sugerindo que a ligação à cohesina é feita de um único modo. Alinhamentos com esses grupos revelam uma conservação maior dos supostos resíduos de reconhecimento à cohesina quando a sequência de um dos grupos está invertida, equivalente à rotação de 180° na estrutura da dockerina. Isso sugere que essas enzimas adotam um único modo de ligação à cohesina, mas opostos entre si, facilitando a sua integração na escafoldina.^{1,14}

As dockerinas do grupo 2 se assemelham com uma versão truncada das do grupo 4, como já mencionado. Os membros de ambos os grupos reconhecem a cohesina de ScaH e também a de ScaE, ancorada à superfície celular, mesmo não possuindo as significativas inserções e hélices ou módulos-X, como as dockerinas de CttA e ScaB. O grupo 4 apresenta simetria interna, o que é atípico para dockerinas do tipo-III, comprovadamente refletindo em duplo modo de ligação a cohesinas. As dockerinas de CttA e ScaB fazem parte desse grupo, sendo a exceção no caso da simetria, devido justamente às inserções já citadas. Outras

dockerinas de escafoldinas integram o grupo, como ScaF, ScaI e ScaH, a qual curiosamente reconhece sua própria cohesina.^{1,14}

A dockerina de ScaA, membro único do grupo 5, reconhece as 7 cohesinas de ScaB na linhagem 17 e as cohesinas 5-9 na linhagem FD-1. Por se tratar de uma escafoldina primária, e uma das principais do sistema pela incorporação de enzimas ao complexo, sobretudo na linhagem 17, foi alvo de diversos estudos 16–18 bioquímicos e biofísicos, na tentativa de elucidar seu comportamento na ausência de uma estrutura cristalina. Algumas conclusões, no entanto, dependem da observação do modelo estrutural, o que foi obtido no presente trabalho.

Figura 7 - Modelo da organização do celulosomo de *R. flavefaciens*, linhagem FD-1. As cores de fundo destacam as subdivisões do reconhecimento dockerina-cohesina. Em A, proteínas celulosomais. Em B, Proteínas que se ancoram à célula. Em C, dockerinas incompletas, do grupo 2. Em D, a proteína CttA, que medeia o anexo do sistema ao substrato. Os diferentes grupos estão destacados por cores, sendo azul (grupo 1), amarelo (grupo 2), verde (grupos 2 e 4), e vermelho (grupo 5).

Fonte: Adaptada de RUIFY.¹

1.1.4 Aplicações envolvendo Dockerinas e Cohesinas

O celulosomo tem sido inspiração para diversas aplicações, que na maioria, almejam a degradação de biomassa vegetal para obtenção de etanol. Nesse sentido, mini celulosomos foram projetados tanto para melhorar a compreensão sobre esse sistema quanto para ser extensões dos celulosomos existentes. No último caso, a incorporação de enzimas externas auxilia na degradação de substratos, resultando no melhoramento da atividade como um todo. A incorporação de novas escafoldinas adaptadoras a sistemas existentes permite amplificar o número de enzimas a eles integradas. Outra abordagem usada é a introdução de celulosomos em organismos que não os produzem naturalmente, aumentando sua capacidade de degradação de substratos vegetais. Dessa forma, os componentes do celulosomo podem ser incorporados a organismos mais bem conhecidos e já utilizados para a degradação de substratos de origem vegetal, otimizando o processo. Há também a tentativa de inserção de enzimas que façam a conversão de substratos fermentescíveis a biocombustíveis, sobretudo em organismos termofílicos, pois são atrativos aos processos industriais. O custo do cultivo anaeróbio e a inibição de atividades enzimáticas pelos combustíveis gerados no processo limitam essa aplicação. Outro uso dos minicelulosomos sintéticos é a produção de outros tipos de moléculas de valor agregado, através do desvio de rotas metabólicas. A sinergia característica desse sistema pode ainda ser aproveitada para tornar vias metabólicas mais eficientes, através do tunelamento de substratos, em que reações que ocorrem em cascata sejam favorecidas.²

A arquitetura do celulosomo também inspirou o desenho de outros tipos de complexos, buscando aumentar o número de enzimas associadas. Assim, foram criados complexos com enzimas ligadas a nanopartículas, a estreptavidinas, a escafoldinas de DNA e a proteínas de alta simetria interna, de forma que ao se organizarem, geravam complexos de 12 ou 18 enzimas. Outros parâmetros importantes como a flexibilidade estrutural, recrutamento de substrato e controle do conteúdo enzimático não podem ser obtido nesses complexos.² Outras aplicações foram propostas para a utilização de dockerinas e cohesinas, porém poucas foram de fato exploradas. Um biossensor que utiliza dockerinas e cohesinas de tipos I e II para amplificação de sinal foi construído, além de sistemas de purificação de proteínas utilizando um dos módulos como caudas fusionadas a proteínas e o seu módulo parceiro imobilizado a resina para retenção da fusão de interesse.^{2,19-20}

No presente trabalho, os módulos dockerina e cohesina foram utilizados, devido a sua afinidade, para a montagem de redes altamente ordenadas a partir de proteínas de simetria

rotacional. A fusão desses módulos aos monômeros dessas proteínas ocasiona a automontagem dessas redes, quando ambas as fusões proteicas são misturadas, como será detalhado abaixo.

1.2 Redes automontáveis de proteínas

Um dos objetivos centrais da nanotecnologia é o desenvolvimento de materiais organizados na escala nanométrica que possam ser funcionalizados de forma a exceder as funcionalidades iniciais de seus precursores, e assim, ser úteis para as mais diversas áreas, como a biomedicina, processos bioquímicos, microeletrônica e a cristalização de biomoléculas. A propriedade de automontagem ocorrente em vários sistemas naturais, como em DNA e proteínas, é hoje o meio mais viável de manipulação dos blocos de construção utilizados para essa finalidade, e tem sido crescentemente explorada nas tentativas de obtenção desses materiais. A utilização de DNA como bloco de construção para esses sistemas já foi demonstrada, e sua manipulação é bem caracterizada, devido à relativa simplicidade no reconhecimento entre as bases nitrogenadas que o compõem. Essa técnica, denominada DNA origami, foi aplicada no desenvolvimento de máquinas moleculares, encapsulamento de moléculas e montagem de diversos formatos. No entanto, a síntese dessas moléculas é complexa e custosa. O uso de proteínas não é tão bem caracterizado quanto o de DNA, devido à complexidade envolvida na relação entre os aminoácidos. Contudo, as propriedades particulares das proteínas envolvidas e mesmo da natureza química dos aminoácidos podem conferir, aos materiais, funcionalidades como reconhecimento molecular, ligação a outras moléculas orgânicas, ou inorgânicas, atividade catalítica, além das estruturas mais complexas que podem arquitetar. As proteínas ainda apresentam a versatilidade de poder ser conjugadas com muitos outros tipos de moléculas, permitindo importar suas propriedades.

²¹ Essas características são ainda mais interessantes quando pensadas no contexto de materiais inteligentes, ou seja, responsivos a diferentes estímulos, como por exemplo, mudanças de temperatura, pH, ou reconhecimento de moléculas ou tecidos específicos. A atividade mecânica de algumas proteínas tem sido explorada, de tal forma que vamos nos aproximando do desenvolvimento de nano robôs. Assim, o desenvolvimento de materiais compatíveis deve avançar juntamente. ²¹⁻²² As limitações intrínsecas de algumas proteínas, como instabilidade estrutural, podem, também, conferir obstáculos para a construção de tais materiais, por isso é crucial a seleção cuidadosa de proteínas estáveis para essas aplicações. ²¹

Uma das mais importantes propriedades que torna as proteínas os candidatos ideais para que esses materiais sejam factíveis é a capacidade de formar complexos através de sua simetria. A obtenção de um alto grau de ordenação é chave para a automontagem, e a ocorrência de simetria é o caminho mais simples para alcançá-lo, devido à minimização de diferentes contatos entre as subunidades utilizadas na construção de um padrão. Por esse motivo, também, é tão frequente na natureza, uma vez que a evolução se apropria dos caminhos mais simples possíveis. Assim, a simetria guia as tentativas de obtenção de nanomateriais por automontagem, seja utilizando simetrias naturais ou projetadas artificialmente.²³⁻²⁴ A primeira estratégia explorada faz uso de proteínas de simetria conhecida fusionadas de modo rígido, de forma a controlar a geometria obtível. São selecionadas proteínas com simetria rotacional, como dímeros e trímeros, cujos C ou N- terminais são compostos por hélices e assim se desenha uma hélice que possa ser estendida, compreendendo ambas as proteínas a serem fusionadas. Assim, o primeiro exemplo que se tem de um material desse tipo é uma jaula de geometria tetraédrica contendo 12 subunidades, formadas da fusão de dímeros e trímeros. A estratégia, conhecida como “nanoedros”, foi utilizada para formar outros tipos de simetria e validada por cristalografia (Figura 8).²⁵ Mais simples que essa estratégia, é o uso de estruturas que naturalmente produzem encapsulamento, como os capsídeos virais, desprovidos de material genético, encapsulinas bacterianas, a ribonucleoproteína citoplasmática *vault* e a ferritina.²⁶ Utilidades dessas estruturas foram demonstradas utilizando-se tanto do lúmen quanto de suas faces externas. No primeiro caso, pode ser explorado o encapsulamento de moléculas para transporte até locais ou tecidos específicos. Já a superfície externa pode ser aproveitada como apresentação de moléculas de sinalização e de antígenos, estabelecendo vacinas sintéticas.²⁶⁻²⁷

Os *coiled coils* são motivos estruturais anfifílicos formados por um entrelaçamento de hélices, associadas por interações intra- e intermoleculares e, portanto, são um tipo de interação utilizável no contexto da automontagem de proteínas. A compreensão sobre a natureza dessa interação permite que interações sejam precisamente desenhadas em variadas conformações, como fibras, tubos, jaulas e hidro géis. O desenho desses contatos é facilitado por se tratar de elementos peptídicos pequenos e alguns autores afirmam que essa característica traz um avanço na sofisticação dos nano materiais automontáveis. Outro avanço no desenvolvimento desses materiais é o desenho computacional de novas interfaces, artificiais, otimizadas para a automontagem, aplicável para uma, duas ou três dimensões. Essa prática permite que não se dependa da ocorrência natural de simetria nas proteínas, ou peptídeos, como *coiled coils* e que sejam criadas superfícies que requerem menor energia.²¹⁻²³

Metais são frequentemente encontrados na superfície de proteínas estabilizando estruturas, atuando em atividades catalíticas ou mediando interações entre proteínas. Assim, a coordenação de metais tem sido explorada como mediador na montagem desses materiais. Ao posicionar adequadamente resíduos que coordenam metais na superfície de proteínas, podem ser criadas interfaces que podem ser acrescidas a outras existentes na proteína, resultando em arranjos periódicos bem determinados. Uma propriedade interessante decorrente dessa estratégia é a capacidade de regulação da automontagem através de um estímulo químico externo, ou seja, a introdução do metal no sistema.^{3,23,28} Dessa forma, por exemplo, foi possível criar nano tubos helicoidais e matrizes bidimensionais através da coordenação de zinco na superfície de citocromo cb-562. A funcionalização de um resíduo em outra superfície com uma rodamina permitiu o empilhamento das matrizes 2D para a criação de uma matriz 3D.²⁸ Os usos propostos para esses tipos de arquitetura incluem seu uso como material microporoso para separação de moléculas, além de suporte para reações de tunelamento de substrato, um grande objetivo da engenharia metabólica. Porém, talvez o mais promissor papel dessas arquiteturas seja para a cristalografia, atuando potencialmente como cristais artificiais, onde proteínas de difícil cristalização poderiam ser inseridas, facilitando o processo de resolução de estruturas.²¹⁻²³

Figura 8 – Diferentes construções para os materiais automontáveis a partir de proteínas simétricas. Em a e b, observa-se a estratégia de nanoedros, ligando proteínas simétricas com um *linker* rígido. Em c, a ligação é feita com *linker* flexível, o que não permite o controle da geometria.

Fonte: Adaptada de SINCLAIR.³

Há estratégias dedicadas à obtenção de camadas 2D e matrizes 3D, denominadas “*crystalins*”, em que a fusão de elementos de simetria ocorre de forma flexível, ao contrário dos nanoedros, em que há rigidez entre os componentes. Com isso, a versatilidade de formas diminui, uma vez que não se controla a geometria do sistema. Essas estruturas são o foco de aplicação buscado nesse trabalho, as quais possuem algumas diretrizes bem caracterizadas que foram incorporadas em nosso planejamento experimental e serão exploradas em detalhes. Em primeiro lugar, é imprescindível o uso de elementos simétricos, uma vez que esses sistemas dependem muito da ordenação resultante da simetria, sobretudo quando pensados para utilização como cristais artificiais. Estratégias que combinam diferentes simetrias em um único elemento de fusão trazem dificuldades na purificação, exigindo cuidados para a dissociação e posterior reconstituição da interação. Dessa forma, é mais simples trabalhar com sistemas em que se purifica os componentes separadamente. Os principais fatores a serem considerados no desenho dessas estruturas são a força de interação entre os componentes, a flexibilidade entre eles e a capacidade de reparo de ligações desfavoráveis. O ajuste entre esses fatores é determinante para que a montagem do material consiga explorar diferentes conformações de modo a alcançar um mínimo energético global, que irá lhe conferir a maior estabilidade. Assim, é necessário que a ligação entre os componentes seja forte e específica, mas reversível, para permitir o reparo de erros. A falta dessa propriedade pode fornecer

regiões em que a organização é obtida, mas não é estendida a todo o sistema. De forma semelhante, o *linker* entre os elementos deve ser curto, mas suficientemente flexível para permitir o crescimento da estrutura sem perda da cristalinidade. Dessa forma, a menor imposição sobre a geometria local é compensada pela média global, que faz a conformação do sistema como um todo atingir um mínimo energético. Alguns trabalhos ressaltam que a rigidez do *linker* e mesmo seu tamanho pode ser determinante para a obtenção de redes automontáveis, havendo relatos de diferentes resultados com 1 resíduo de aminoácido de diferença entre os linkers utilizados. Ainda, é desejável que o sistema apresente modularidade, ou seja, a capacidade de controle externo sobre a oligomerização, o que pode ser feito com estímulo químico ou alterações ambientais, como variação de pH, temperatura ou força iônica.^{3,23,29-30}

Neste trabalho, buscamos utilizar as características da interação dockerina-cohesina, alta afinidade e especificidade, como meio de desenhar racionalmente redes automontáveis em duas dimensões.

2 Objetivos

O objetivo primeiro de nossa pesquisa foi compreender a interação entre a escafoldina primária ScaA com a principal escafoldina adaptadora, ScaB, de *Ruminococcus flavefaciens*, a qual é central para a montagem do celulosomo nesse organismo. Entender a montagem e arquitetura do celulosomo é fundamental para propor aplicações na nanotecnologia e biotecnologia. Esse conhecimento passa pelo estudo estrutural e biofísico da interação entre os componentes primordiais dessas proteínas, as cohesinas e dockerinas. Revisitamos resultados prévios obtidos no laboratório e conduzimos novas análises, contemplando:

- Cristalização e determinação estrutural de diferentes módulos cohesina de ScaB.
- Cristalização e determinação estrutural do complexo entre a dockerina de ScaA e módulos cohesina de ScaB.
- Estudos de interação molecular entre a dockerina de ScaA e cohesinas de ScaB por calorimetria isotérmica de titulação (ITC) e termoforese em microescala (MST).

Adicionalmente, planejamos empregar os módulos cohesina e dockerina em uma aplicação nanotecnológica, através da construção de um nanomaterial a partir da construção de uma rede 2D cristalina automontável. Essa rede conta com a ligação entre cohesinas e dockerinas para conexão interna da estrutura. Para isso, selecionamos, baseados em estruturas depositadas no banco de dados PDB, proteínas de simetria rotacional de ordens 2, 3, 4 e 6 (grupos pontuais C₂, C₃, C₄ e C₆) e pares dockerina-cohesina de diferentes tipos e organismos. As proteínas de ordens 3, 4 e 6, constituem o vértice dessas redes, as quais foram fusionadas às dockerinas, e proteínas de simetria rotacional de ordem 2 foram fusionadas às cohesinas para compor a conexão entre os vértices desse sistema. Essa estratégia permite purificação independente dos componentes, facilitando a busca de condições ótimas para montagem das redes. Para isso, realizamos:

- Construção das fusões genéticas que compõem os módulos da rede, expressão e purificação das proteínas por elas codificadas.
- Testes de diferentes condições para montagem das redes *in vitro* e caracterização inicial da formação das redes por microscopia eletrônica.

3 Materiais e Métodos

3.1 Sub-clonagem dos componentes celulosossomais de *Ruminococcus flavefaciens* em vetor de expressão

As sequências codificadoras da dockerina C-terminal de ScaA (Rf-DocA, resíduos 795-879), da primeira cohesina da escafoldina ScaB (Rf-CohB1, resíduos 25-157) e da quarta cohesina da escafoldina ScaB (Rf-CohB4, resíduos 472 – 610) e da sexta cohesina de ScaB (Rf-CohB6, resíduos 775-948) de *R. flavefaciens*, linhagem 17 (número de acesso GenBank: AJ278969) foram sintetizadas com códons otimizados para expressão em *E. coli* (Epoch Biolabs) e subclonados independentemente no vetor pET-SUMO, através dos sítios de restrição BamHI/XhoI. Esse vetor promove a expressão da proteína recombinante fusionada a SUMO (Small Ubiquitin-like Modifier) no N-terminal, incluindo um sítio para reconhecimento da SUMO protease (Ulp1) para posterior remoção dessa proteína de fusão. Adicionalmente, esse vetor promove maior solubilidade e rendimento à proteína recombinante expressa e contém uma cauda de 6 resíduos histidina em seu N-terminal, para permitir a purificação por afinidade a metal.

Os mutantes pontuais de Rf-DocA e Rf-CohB1 foram gerados utilizando o *kit* de mutagênese sítio-dirigida QuikChange (Stratagene), de acordo com as instruções do fabricante. Todas as mutações foram verificadas por sequenciamento. As clonagens haviam sido previamente executadas pela Dr^a Juliana Cheleski Wiggers, durante seu pós-doutorado em nosso laboratório, e assim apenas a expressão das proteínas fez-se necessária para novos ensaios.

3.2 Transformação de células BL21(DE3) para expressão dos módulos cohesina e dockerina de *R. flavefaciens*

Cerca de 100ng de cada DNA plasmidial foi utilizado para transformar células de BL21 (DE3) por choque térmico. O DNA foi adicionado às células, seguido de incubação em gelo por 5 minutos. O choque térmico foi feito a 42^o C por 1 minuto, seguido de incubação em gelo por 2 minutos. Foram adicionados 800 µL de meio LB (10 g/L de Triptona; 5 g/L de Extrato de levedura; 10 g/L de NaCl) às células, e essa mistura foi então incubada a 37^o C por

1h, com agitação de 250 rpm. Ao final, a suspensão bacteriana foi espalhada em placas de Petri com meio LB suplementado com 50 µg/mL de canamicina, as quais foram incubadas *overnight* a 37 °C.

3.3 Expressão e purificação das cohesinas e dockerinas de *R. flavefaciens*

Uma colônia de BL21 transformante crescida na placa de Petri foi inoculada em 5 mL de meio LB suplementado com 50 µg/mL de canamicina e incubada *overnight* a 37 °C. Essa cultura inicial foi usada no preparo de um inóculo de 500 mL de meio LB suplementado com 50 µg/mL de canamicina, o qual foi incubado a 37 °C e 250 rpm até que a densidade óptica a 600 nm atingisse um valor entre 0,5 e 0,6. A expressão então foi induzida pela adição de IPTG a uma concentração final de 100 µM, e incubada *overnight* a 18°C e 250 rpm. Após a expressão, cultura foi centrifugada por 25 min a 4000 x g e 4 °C, e o sobrenadante descartado. O *pellet* remanescente foi ressuspensionado em 20 mL de Tampão A (Tris 25 mM pH8,0, NaCl 300 mM, CaCl₂ 2 mM) contendo 10 mM de Imidazol e lisado por sonicação com 15 pulsos de 30 segundos, intercalados por 30 segundos de descanso, com amplitude de 20% (Sonic Dismembrator Mode 500 – Fisher Scientific). O lisado foi centrifugado a 12000 x g por 30 min a 4 °C, para a recuperação da fração solúvel.

A purificação da proteína recombinante foi realizada por cromatografia de afinidade em coluna de cobalto (IMAC, resina TALON - Clontech). A coluna foi equilibrada com 10 VC (volumes de coluna, ou seja, 1 VC corresponde ao volume de resina presente na coluna) de Tampão A contendo 10 mM de Imidazol. Após a passagem da fração solúvel do lisado pela coluna, foi feita sua lavagem com 5 VC de Tampão A contendo 10 mM de Imidazol, seguido da lavagem com 5 VC do mesmo tampão, porém sem imidazol. A proteína foi eluída com a passagem do Tampão A contendo concentrações crescentes de Imidazol (concentrações 10 mM, 20 mM, 50 mM, 100 mM e 300 mM). As frações contendo a proteína de interesse foram identificadas através da análise por SDS-PAGE, agrupadas e adicionadas a 5 µM da protease Ulp1, com incubação *overnight* a 4 °C.

Após a clivagem, foi realizada uma cromatografia em coluna HiPrep 26/10 (GE Healthcare) para retirada de Imidazol. Em seguida, uma segunda IMAC foi realizada para remoção da SUMO clivada, a interage com a resina enquanto a proteína de interesse, livre de fusão, é coletada na fração não ligada (*flowthrough*) e na lavagem com 5 VC tampão A sem Imidazol.

A proteína foi então purificada por cromatografia de exclusão molecular, através da coluna Superdex 75 10/300 GL (GE Healthcare), utilizando tampão A. A pureza das amostras foi acessada por SDS-PAGE. Todas as proteínas purificadas foram concentradas por centrifugação utilizando filtros *Millipore* 10 KDa, quantificadas através das seus coeficientes de absorvidade molar (calculados pelo ProtParam³¹), aliquotadas, congeladas em nitrogênio líquido e estocadas em ultrafreezer.

3.4 Cristalização das cohesinas Rf-CohB1 e Rf-CohB4 de *R. flavefaciens* e complexo Rf-DocA/Rf-CohB4

A cristalografia de proteínas é uma das técnicas que permite mais alta resolução na determinação da estrutura de proteínas. Nessa técnica, um padrão da difração de raios-X por um cristal de proteína é utilizado para calcular a densidade eletrônica dessa proteína. A obtenção dos cristais é essencial para a realização da técnica, uma vez que sua periodicidade faz com que seja possível tratá-lo como uma rede de difração em 3D. A formação dos cristais consiste na gradual concentração da proteína em solução, ocasionando o empacotamento dessa macromolécula de forma ordenada e simétrica. As técnicas utilizadas para sua obtenção baseiam-se na difusão de vapor entre uma solução contendo a proteína e outra solução, sem a proteína, e mais concentrada em sais e/ou precipitantes. Assim, a proteína passa pela supersaturação, e então a nucleação dá início à formação dos cristais.³²⁻³³

Triagens das condições de cristalização de Rf-CohB1_{WT} (nativa), Rf-CohB4 e o complexo Rf-DocA/Rf-CohB4 foram realizadas a partir de kits comerciais (Hampton Research e Qiagen), pelo método de gota sentada. As condições iniciais foram otimizadas, com variação de pH, concentração de sais e precipitantes, melhorando o crescimento ordenado dos cristais. Esses novos ensaios foram realizados pelo método de gota suspensa, misturando a solução de proteína com a solução do poço em volumes iguais. Os cristais do complexo Rf-DocA/Rf-CohB4 cresceram em gotas contendo 27 mg/mL de proteína (concentração inicial) em tampão 100 mM BIS-TRIS pH 5.5, 25% polietileno glicol 3350. Os cristais de Rf-CohB4 cresceram em gotas com 35 mg/mL de proteína (concentração inicial) em tampão citrato 100 mM pH4, 2,2 M sulfato de amônio. A cohesina Rf-CohB1_{WT} cristalizou em gotas com 30 mg/mL de proteína (concentração inicial), em tampão 2 M sulfato de amônio, 0.1 M Bis Tris (pH 6.5), sendo obtidas duas formas cristalinas distintas. Todos cristais cresceram em incubação a 18 °C.

3.5 Difração e resolução da estrutura das cohesinas 1 e 4 de *R. flavefaciens* e complexo Rf-DocA/Rf-CohB4

Cada onda difratada pelo cristal é caracterizada por sua amplitude e fase, porém, no experimento de difração de raios-X, apenas a intensidade de cada onda difratada é mensurável, havendo a perda da informação sobre a fase da mesma. Essa perda de informação é conhecida como o problema das fases. Diferentes métodos são utilizados para conseguir essa informação, como a utilização de fases presentes na estrutura de uma proteína homóloga (método da Substituição Molecular) e a incorporação de átomos pesados no cristal, e a partir da sua localização calculada, determinar a estrutura da proteína. Nessa última categoria encontram-se os métodos de substituição isomórfica simples ou múltipla, podendo utilizar conjuntamente o sinal anômalo da difração (SIR – Substituição Isomórfica Simples; MIR – Substituição Isomórfica Múltipla; SAD - Difração Anômala Única; MAD - Difração Anômala Múltipla; SIRAS - Substituição Isomórfica Simples com Sinal Anômalo; MIRAS – Substituição Isomórfica Múltipla com Sinal Anômalo).³⁴

As estruturas do complexo Rf-DocA/Rf-CohB4 e Rf-CohB1_{WT} (forma cristalina I) foram resolvidas por SAD, utilizando o sinal anômalo do iodo rapidamente incorporado nos cristais (*quick-soaking*) através da incubação rápida (aproximadamente 10 s) dos cristais em solução crioprotetora adicionada de 500 mM NaI. As estruturas de Rf-CohB4 isolada e Rf-CohB1_{WT} (forma cristalina II) foram resolvidas por substituição molecular, utilizando, respectivamente, a estrutura da Rf-CohB4 presente no complexo e a estrutura de Rf-CohB1_{WT}, previamente resolvidas por SAD, como modelos de busca.

Previamente às coletas de dados de difração de raios X, os cristais foram mergulhados rapidamente em soluções crioprotetoras, de mesma composição do poço em que cresceram, suplementadas com 20% (v/v) de etileno glicol. Os dados de difração foram coletados na linha MX-2 do Laboratório Nacional de Luz Síncrotron, CNPEM, Campinas-SP (Rf-DocA/Rf-CohB4 e Rf-CohB4) ou no anodo rotatório do IFSC, São Carlos-SP. (Rf-CohB1_{WT}). Todos os dados de difração coletados foram processados no pacote de programas XDS35. Para determinação das posições dos íons de iodo nos cristais derivados, foi utilizado o pacote de programas SHELX C/D36. As estruturas foram subsequentemente resolvidas por SAD com o programa PHENIX37. No caso das estruturas resolvidas pelo método de substituição

molecular, foi utilizado o programa PHASER38. O refinamento dos modelos foi feito através de ciclos iterativos com programas COOT39 e PHENIX37.

Tanto a cristalização quanto a resolução dessas estruturas haviam sido realizadas previamente, pelos Drs. Wesley Cardoso Generoso e Juliana Cheleski Wiggers, de modo que ao revisitar seus dados, foi necessário apenas o refinamento das estruturas para obtenção do modelo final.

3.6 Cristalização dos mutantes da cohesina Rf-CohB1

A cristalização dos mutantes de Rf-CohB1 (Rf-CohB1_{G124A}, Rf-CohB1_{Q125A}, Rf-CohB1_{G126A}, Rf-CohB1_{G128A} e Rf-CohB1_{R129A}) foi realizada em torno das condições previamente estabelecidas para a cohesina nativa. Semelhantemente ao processo de otimização de condições, o ensaio de cristalização foi realizado por gota sentada em tampão 0,1 M BIS-Tris pH 6,5, variando Sulfato de Amônio em 0,1 M, abrangendo 2,5 M - 1,4 M. As proteínas foram preparadas em três concentrações diferentes de 10-30 mg/mL e misturadas com a solução do poço (1 µL de cada solução). Os cristais cresceram em incubação a 18 °C e depois de coletados, foram mergulhados rapidamente em suas respectivas soluções crioprotetoras e congelados, conforme já descrito.

3.7 Difração e resolução da estrutura dos mutantes de Rf-CohB1

As estruturas dos cristais dos mutantes Rf-CohB1_{G128A} e Rf-CohB1_{R129A} foram resolvidas por substituição molecular, utilizando como modelo a estrutura de Rf-CohB1_{WT}, previamente resolvida. Os dados de difração desses cristais foram coletados na linha MX-2 do Laboratório Nacional de Luz Síncrotron, CNPEM, Campinas-SP, e o seu processamento, resolução estrutural e refinamento foram conduzidos como descrito anteriormente.

3.8 Calorimetria de Titulação Isotérmica

Calorimetria de Titulação Isotérmica (ITC) foi utilizada para quantificar a afinidade da interação proteína-proteína entre as cohesinas Rf-CohB4 e Rf-CohB6 e as dockerinas Rf-

DocA_{WT} (Rf-DocA nativa) e mutantes Rf-DocA_{V813E}, Rf-DocA_{V817E} e Rf-DocA_{V865E}, e assim, investigar a ocorrência de duplo modo de ligação entre a dockerina Rf-DocA e as cohesinas Rf-CohB4 e Rf-CohB6. Essa técnica se baseia no princípio de que há troca de calor com o ambiente em qualquer interação molecular. No equipamento em que a medida ocorre, há duas células, uma referência e outra de amostra, que são mantidas a uma temperatura constante. Quando se titula um ligante a uma proteína, há liberação ou absorção de calor, o que altera a temperatura da célula de amostra. A diferença entre a temperatura de cada célula é compensada pelo equipamento, que fornece energia às células. O aumento ou diminuição de energia aplicada à célula de amostra, em reações endotérmicas ou exotérmicas, respectivamente, é registrado a cada titulação. À medida que as titulações são feitas, vão se saturando os sítios de ligação das proteínas, de modo que a energia aportada para manutenção da temperatura vai diminuindo a cada injeção. O conjunto das medidas consiste na energia aplicada em função do tempo e forma uma curva de ligação, isoterma, e representa o calor da reação a cada injeção, como função da relação entre concentração de ligante e concentração da proteína. O ajuste dessa curva a um modelo de ligação permite calcular os parâmetros termodinâmicos da ligação, como a constante de dissociação, mudanças na energia livre de Gibbs e entropia.⁴⁰

O equipamento utilizado nos experimentos foi um calorímetro VP-ITC Microcal (Malvern). As medidas foram conduzidas a 25 °C (298 K) e as proteínas foram preparadas no mesmo tampão utilizado para a purificação (25 mM Tris pH 8.0, 300 mM NaCl, 2 mM CaCl₂). À célula de amostra, de 1,4 mL, foi colocada cada variante da dockerina, a 10 µM. As cohesinas Rf-CohB4 e Rf-CohB6 foram tituladas, na concentração de 100 µM, na célula contendo dockerina a cada experimento, em 30 injeções de 8 µL, separadas a intervalos de 240 segundos, com agitação de 300 rpm. O calor de diluição, associado à titulação das cohesinas em tampão, foi subtraído para correção da linha de base. Os dados foram analisados no Microcal ORIGIN versão 7.0 (Microcal Software, Northampton, MA) e as constantes de dissociação (K_d) e estequiometria da ligação foram estimados com o ajuste das curvas com o modelo de sítio único de ligação.

3.9 Termoforese em Microescala

A termoforese em microescala (MST) é uma técnica que permite a quantificação de interações entre proteínas ou mesmo entre proteínas e pequenas moléculas. Essa técnica se

baseia no movimento das moléculas quando em um gradiente de temperatura. Uma variedade de propriedades, como tamanho, carga, camada de solvatação e conformação têm influência sobre seu movimento. Durante o experimento, um laser infravermelho incide sobre a amostra, inserida em um capilar, gerando o gradiente de temperatura. O movimento da molécula estudada é monitorado através de fluorescência, tipicamente através da prévia marcação de uma das moléculas participantes da interação com um fluoróforo, ou através de fluorescência intrínseca. Uma das proteínas é diluída serialmente na outra (marcada), de forma que há uma curva decrescente na relação entre a concentração de cada uma. Essas diferentes concentrações geram perfis de movimento no gradiente, correspondentes ao efeito da interação na mobilidade da proteína marcada. A comparação entre a fluorescência inicial e após a incidência do laser infravermelho gera os pontos correspondentes a cada diluição, originando uma curva. O conjunto desses dados é ajustado a um modelo de ligação que permite o cálculo dos parâmetros da interação. Assim, essa técnica é sensível mesmo a pequenas mudanças, permite a quantificação de interações de alta afinidade e exige uma mínima quantidade de amostra.

A Dockerina Rf-DocA foi marcada com fluoróforo NT-647-NHS, com uso do *kit Monolith NT™ Protein Labeling kit* (NanoTemper Technologies), conforme instruções do fabricante. Esse *kit* de marcação utiliza um éster de N-Hidroxisuccinimida (NHS) conjugado a um fluoróforo, que reage com aminas primárias presentes na proteína (grupos α -amino, no N-terminal da proteína e ϵ -amino, presentes em resíduos de lisina), formando ligações amida estáveis. O fluoróforo utilizado tem seu pico de excitação em aproximadamente 650 nm e o de emissão em 670 nm. Devido à natureza da reação, foi necessário fazer a troca de tampão da proteína, uma vez que este continha Tris. Essa etapa foi realizada com os reagentes presentes no mesmo *kit*, conforme instruções do fabricante. Assim, após a troca do tampão, misturamos 100 μ L de Rf-DocA a 10 μ M a 100 μ L de sonda a 30 μ M e incubamos essa mistura por 30 minutos à temperatura ambiente e protegida de luz. Para remover o excesso de sonda, não incorporada à Rf-DocA, foi realizada uma purificação, com coluna de dessalinização fornecida no *kit*, conforme as instruções do fabricante. A proteína marcada foi eluída com tampão MST (50 mM Tris-HCl pH 7.4, 150 mM NaCl, 10 mM MgCl₂, 0.05 % Tween-20) (NanoTemper Technologies) e coletada em frações.

Para realização dos experimentos de MST, a Rf-DocA marcada foi diluída a uma concentração aproximada de 2 nM em tampão MST suplementado com 0,4 mg/mL de soro albumina bovina, sendo consequentemente distribuída em alíquotas de 20 μ L, para realização da diluição seriada da cohesina. A cohesina Rf-CohB1_{WT} foi preparada em concentração

inicial de 8 μM em tampão MST suplementado com 0,4 mg/mL de soro albumina bovina. Um volume de 20 μL dessa solução de cohesina foi misturado à primeira alíquota de Rf-DocA, 20 μL dessa mistura foram misturados à alíquota seguinte de Rf-DocA e assim sucessivamente. Descartamos 20 μL ao fim da última diluição, obtendo um volume final de 20 μL em todas as frações da diluição seriada. As duas primeiras diluições de cada experimento foram descartadas, pois apresentavam uma queda na fluorescência, devido ao efeito de diluição do tampão na qual se encontravam as cohesinas inicialmente.

As concentrações iniciais das cohesinas foram ajustadas conforme a necessidade, de modo a se obter número suficiente de pontos para ajuste adequado do modelo sigmoidal de ligação. Assim, as concentrações iniciais otimizadas para cada experimento foram: Rf-CohB1_{WT} 8 μM ; Rf-CohB1_{G124A} 1 μM ; Rf-CohB1_{Q125A} 16 μM ; Rf-CohB1_{G126A} 8 μM ; Rf-CohB1_{G128A} 64 μM ; Rf-CohB1_{R129A} 8 μM . A cada experimento, as 16 diluições eram monitoradas por 5 s sem incidência de laser IR, 20 s com incidência de laser IR e 5 s após desligamento do laser. A potência do laser IR foi ajustada para 60%, e o laser LED, que monitora a fluorescência, para 95%, com leituras de fluorescência próximas a 200 u.a., conforme orientação do manual para sistemas de alta afinidade. Os experimentos foram realizados pelo menos em triplicata, usando capilares com tratamento padrão (NT. 115) e o equipamento Monolith NT. 115 (NanoTemper Technologies).

3.10 Planejamento da clonagem para produção de redes auto montáveis de proteínas

Como mencionado anteriormente, neste trabalho buscamos uma aplicação para as dockerinas e cohesinas utilizando-as na automontagem de matrizes proteicas de duas dimensões, com a possibilidade de expandi-las para três dimensões. Para tanto, selecionamos no Protein Data Bank (PDB) 41 proteínas de simetria rotacional C3, C4 e C6 para compor os vértices condutores da simetria do sistema, além de simetria C2 para compor o elemento linear de ligação a cada vértice (Figura 9). As proteínas foram filtradas pelos organismos das quais provinham, dando preferência para proteínas de Archea, por se tratar de proteínas mais estáveis, e compatíveis com expressão em *Escherichia coli*. Análise das estruturas individuais das candidatas foi feita para selecionar aquelas passíveis de fusão em seu N-terminal, no caso das C2, ou C-terminal, no caso das simetrias de maior ordem, tendo em vista a fusão dos módulos cohesina e dockerina, respectivamente. Dessa forma, duas proteínas de cada simetria foram escolhidas para compor o rol das candidatas ao sistema que criaríamos.

As dockerinas e cohesinas foram selecionadas com base em um estudo⁶ que analisou a especificidade de ligação entre diversos pares dockerina/cohesina, dos quais selecionamos os mais específicos, sendo 3 pares de tipo I e 1 de tipo II. Ainda, buscamos um par adicional de cada tipo com estrutura resolvida no PDB. Esses módulos, juntamente com as proteínas de simetria foram sintetizados por GenScript, que otimizaram as sequências para expressão em *E. coli*. Tínhamos ainda, outras dockerinas e cohesinas disponíveis no laboratório, provenientes de *Ruminococcus flavefaciens*, as quais requereram apenas o desenho de *primers* adequados. As sequências utilizadas estão disponíveis no Apêndice A.

Figura 9 - Estratégia de construção de redes auto montáveis de proteínas utilizando proteínas de simetria C3, C4 e C6 como vértices da rede fusionadas a dockerinas e proteínas de simetria C2 fusionado a cohesinas. (A) Módulos selecionados para compor os vértices (proteínas de simetria C3, C4 e C6), fusionados a dockerinas e proteínas de simetria C2 selecionadas para compor o elemento de conexão linear entre os vértices, fusionados a cohesinas. (B) A fusão de diferentes módulos origina cada elemento da rede, os quais são expressos e purificados independentemente. (C) A associação entre os componentes construídos da fusão das proteínas precursoras dá origem às redes automontáveis, através da ligação da dockerina com a cohesina, presentes em módulos distintos.

Fonte: Elaborada pelo autor.

Para ligação entre as proteínas simétricas e cada módulo do par de ligação, foram desenhados três *linkers*, de variados tamanhos e flexibilidades. Dois eram rígidos, formando hélices de 5 e 10 resíduos (sequências: SGEEARRGS e GSEAARREAEARGs) e um era flexível, com 6 resíduos (sequência: GSGGSG). Essas sequências foram desenhadas posicionando estrategicamente resíduos de ácido glutâmico e arginina, de forma a otimizar a estabilização da hélice projetada, com base em estudo 42 específico. O *linker* flexível foi desenhado com base em outro estudo que demonstrou que aminoácidos de cadeia lateral curta e polares como serina e glicina são adequados para fornecer flexibilidade e ao mesmo tempo, estabilidade em ambiente aquoso.⁴³

Tendo em vista o grande número de combinações entre os componentes selecionados, fez-se necessária uma estratégia de clonagem que permitisse o escalonamento do processo, o que seria dificultoso através da tradicional técnica de restrição e ligação de fragmentos de DNA. Por esse motivo, no primeiro momento, selecionamos a estratégia LIC 44 – Clonagem Independente de Ligação, na sigla em inglês – a qual era conhecida em nosso laboratório. Nessa técnica, fragmentos de DNA são anelados por extremidades coesivas de 12 a 20 nucleotídeos, mais extensas que as obtidas através do uso de enzimas de restrição, de tipicamente 4 nucleotídeos. Esses segmentos de fita simples são criados por enzimas que possuem atividade exonuclease, como é o caso da T4 DNA Polimerase, que na ausência de nucleotídeos apresenta atividade 3'-exonuclease. Essa atividade é controlável, uma vez que se pode prover um dos nucleotídeos no meio reacional, fazendo com que ao encontrar o mesmo nucleotídeo na sequência, suas atividades exonuclease e polimerase entrem em equilíbrio, determinando a parada da excisão de nucleotídeos. Dessa forma, ao desenhar sequências em que os terminais sejam deficientes de um dos 4 nucleotídeos, é possível delimitar a criação de uma extremidade coesiva de tamanho desejado (Figura 10). A hibridização dessas extremidades complementares estabiliza o DNA de forma que é possível usá-lo para transformar células bacterianas que repararão a interrupção em ambas as fitas, gerando um DNA recombinante.

Figura 10 - Representação da clonagem pela técnica LIC mostrando uma das junções entre o vetor e o inserto de interesse. Em vermelho, estão destacados os nucleotídeos retirados pela ação da T4 DNA Polimerase, criando extremidades de fita simples de DNA. As sequências são desenhadas para apresentar complementaridade, e incorporada nos *primers* (em verde). Os dois DNAs são então hibridizados e usados para transformar bactérias, que reparam a desconexão das fitas.

Fonte: Elaborada pelo autor.

A sequência dos *linkers* foi então otimizada para a criação dessas extremidades complementares e incorporadas nos *primers* das demais sequências. Assim, cada sequência codificadora de proteína tinha em uma das extremidades três diferentes *primers*, contendo a sequência para cada um dos *linkers*, além de um outro primer desenhado para permitir a ligação ao vetor pelo mesmo método (Figura 11). As sequências dos *primers* estão disponíveis no Apêndice B

Ponto de ligação entre vetor e inserto, à jusante da proteína de fusão SUMO	
5' - ... GTT <u>T</u> ACCAGGAGCAAACGGGAGG - 3'	5' - ACCAGGAGCAAACGGGAGGTGAGACC... - 3'
3' - ... CAAATGGTCTCTGTTGCCCTCC - 5'	3' - TGGTCTCTGTTTGGCCTCCACTCTGG... - 5'
Vetor	Inserto de clonagem
Ponto de ligação entre inserto e vetor, à jusante do inserto	
5' - ... TCCTGAAGCGGTGGCGGTCTG - 3'	5' - GCGGTGGCGGTCTGAACGACAT... - 3'
3' - ... AGGACTTCGCCACCGCCAGAC - 5'	3' - CGCCACCGCCAGACITGCTGTA... - 5'
Inserto de clonagem	Vetor
Ligação entre insertos, compondo o linker SGEEARRGS	
5' - ... AGCGGTGAAGAAGCGCGCCGCGG- 3'	5' - GAAGAAGCGCGCCGCGGTAGC... - 3'
3' - ... TCGCCACTTCTTCGCGCGGCGCC- 5'	3' - CTTCTTCGCGCGGCGCCATCG... - 5'
Inserto 1	Inserto 2
Ligação entre insertos, compondo o linker GSEAARREAEARGS	
	5' - CGCGAAGCGGAAGCGCGTGGTAGC... - 3'
	3' - GCGCTTCGCCTTCGCGCA ^u CCATCG... - 5'
	Inserto 2
5' - ... GGTAGCGAAGCGGCGCGTTCGCGAAGCGGAAGCGCG- 3'	
3' - ... CCATCGCTTCGCCGCGCAGCGCTTCGCCTTCGCGC- 5'	
Inserto 1	
Ligação entre insertos, compondo o linker GSGGSG	
5' - ... GGTAGCGGCGGCGAGCGG- 3'	5' - AGCGGCGGCGAGCGGT... - 3'
3' - ... CCATCGCCGCGGTCGCG- 5'	3' - TCGCCGCGGTCGCGCA... - 5'
Inserto 1	Inserto 2

Figura 11 - Detalhamento das extremidades desenhadas para ligação entre cada componente na clonagem. Em verde, estão exibidas as sequências incorporadas aos *primers*. Em vermelho, os nucleotídeos que serão extraídos no tratamento com a T4 DNA Polimerase.

Fonte: Elaborada pelo autor.

O vetor utilizado para as clonagens foi o pET-SUMO-LIC (Figura 12), semelhante ao utilizado para expressão das cohesinas e dockerinas de *R. flavefaciens*, mas adaptado a LIC, através de sequências apropriadas incorporadas aos *primers* utilizados em sua amplificação. Todos os *primers* foram otimizados, tendo temperatura de melting próximas a 60⁰ C, e sintetizados por Sigma Aldrich.

Figura 12 - Mapa do vetor pET_SUMO utilizado na construção de nossas clonagens para expressão de proteínas componentes das matrizes auto montáveis. As setas roxas, externas ao vetor, indicam os locais em que os *primers* iniciam a amplificação do vetor.

Fonte: Elaborada pelo autor.

A estratégia de LIC é muito versátil e permite a clonagem de múltiplos fragmentos, desde que em números pares. Quando se trabalha com um número ímpar de fragmentos, adaptações e outras técnicas precisam ser aplicadas para promover outras extremidades complementares entre todos os componentes. Essa limitação ocorre devido à complementaridade das sequências na clonagem. Como consequência, cada par de fragmentos que será fusionado na clonagem deverá ser tratado com T4 DNA Polimerase na presença de um deoxirribonucleotídeo complementar ao do fragmento correspondente, por exemplo, tratando o vetor com dTTP e o inserto com dATP. Em nosso caso, tivemos que unir as sequências das duas proteínas componentes da fusão e então inserir esse fragmento no vetor. Dessa forma, um dos fragmentos foi tratado com T4 DNA polimerase na presença de dTTP para poder se ligar ao fragmento tratado na presença de dATP. Esse inserto unificado tinha uma extremidade previamente tratada com dATP, estando apta a inserir-se no vetor tratado com dTTP, enquanto sua outra extremidade precisou de um segundo tratamento, com dATP para se adequar à clonagem. Para possibilitar esse procedimento, foi introduzida uma etapa de ligação *in vitro* entre os dois fragmentos, antes desse segundo tratamento com T4 DNA

Polimerase, para assegurar a ação da enzima no local designado (Figura 13). Para tanto, os *primers* utilizados na amplificação desses fragmentos, que davam origem aos *linkers* entre eles, tiveram que ser fosforilados *in vitro*. Devido às diluições envolvidas nas etapas de todos esses processos, optamos por amplificar esse inserto unificado, com polimerase de alta fidelidade Phusion (New England Biolabs), antes do tratamento para excisão de nucleotídeos.

Alternativamente, testamos técnica de SLIC⁴⁵ – Clonagem independente de sequência e ligação, na sigla em inglês –, bastante semelhante à LIC, porém, independente da sequência presente nas extremidades. A parada da excisão de nucleotídeos é determinada por tempo de reação, portanto, a composição das sequências é irrelevante, desde que haja complementaridade entre as extremidades. Por esse motivo é possível trabalhar com números ímpares ou pares de fragmentos. A reação de excisão de nucleotídeos pela T4 DNA Polimerase é feita na ausência de quaisquer deoxirribonucleotídeos, sendo interrompida, após o tempo adequado, pela adição de qualquer um dos 4 deoxirribonucleotídeos ao meio reacional, seguido de incubação em gelo. As sequências que desenhamos para LIC eram compatíveis com essa técnica, por apresentar a complementaridade necessária, o que nos permitiu testá-la sem incorrer em preparos adicionais. A vantagem da utilização dessa técnica, em nosso planejamento, foi a possibilidade de hibridizar os três fragmentos após um único tratamento com exonuclease (Figura 13).

Figura 13 – Estratégias de inserção da fusão genética utilizando LIC e SLIC. Na parte superior, a estratégia LIC, e na parte inferior, SLIC. Em ambas partes da figura, mostramos a fusão da proteína trimérica de PDB ID:3C19 com uma dockerina de tipo II como exemplo. As setas vermelhas indicam a ação exonuclease da enzima T4 DNA Polimerase. Os círculos em amarelo indicam os pontos em que ocorre a fosforilação, através dos *primers*, para permitir a ligação *in vitro*.

Fonte: Elaborada pelo autor.

3.11 Clonagem por LIC

Os *primers* que originaram os *linkers* entre os fragmentos clonados foram fosforilados utilizando a enzima T4 Polinucleotideo Kinase (PNK) (New England Biolabs), em reação contendo, em concentrações finais: tampão específico da enzima 1X, ATP 1 mM, Primer 5 μ M, 8 U de T4 PNK, e água, para volume final de 20 μ L. a reação foi incubada a 37 °C por 1 h, e inativada por incubação a 65 °C por 20 minutos.

Os insertos foram amplificados por PCR com Phusion DNA Polymerase (New England Biolabs), em reação contendo tampão HF 1X, 200 μ M dNTPs, 0,1 μ M de cada primer, 100 ng de vetor contendo os genes sintéticos, 3% de DMSO, 1 Unidade de Phusion DNA Polymerase, em concentrações finais, e água, num volume reacional de 50 μ L. Após desnaturação inicial de 3 minutos a 98 °C, a reação foi conduzida por 35 ciclos de desnaturação a 98 °C por 20 segundos, hibridização dos *primers* a 52 °C por 30 segundos e extensão a 72 °C por 20 segundos, além de extensão final por 5 minutos adicionais.

A amplificação do vetor pET-SUMO-LIC seguiu o mesmo protocolo, alterando-se o DNA molde para 25 ng do vetor pET-SUMO fechado, a temperatura para hibridização dos *primers* para 60°C e o tempo de extensão de cada ciclo para 3 min. Após confirmação por gel de agarose, os produtos amplificados foram purificados com o *kit* Wizard® SV Gel and PCR Clean-Up System (Promega). Foram então concentrados em Speedvac (Thermo Savant) e tratados com DpnI (ThermoScientific), para degradação de DNA parental, em reação contendo concentração final 1X de tampão Fast Digest, 10 Unidades de DpnI e água, para volume final de 10 μ L. Essa mistura foi incubada por 1h30 a 37 °C e então a 80 °C por 5 minutos para inativação da enzima.

O tratamento com T4 DNA Polimerase para criação de segmentos simples fita, as extremidades coesivas, foi feito utilizando o volume total da reação anterior, acrescido de tampão Fast Digest para concentração final 1X, 0,5 mM de dATP (para as sequências codificadoras de proteínas de simetria C2 e dockerinas) ou de dTTP (para as sequências codificadoras de proteínas de simetrias C3, C4, C6, cohesinas, e o vetor pET-SUMO), DTT (ditiotreitól) 5 mM, 0,5 Unidade de T4 DNA Polimerase (ThermoScientific) (previamente diluída a 0,5 U/ μ L em tampão Fast Digest 1X) e água para um volume final de 20 μ L. A mistura foi incubada a 22 °C por 30 minutos e então inativada por 20 minutos a 65 °C.

Os fragmentos que compõem o inserto da clonagem foram ligados *in vitro* pela reação contendo 10 μ L de cada fragmento tratado com T4 DNA Polimerase, 2,5 mM ATP, 600

Unidades de T4 DNA Ligase (New England Biolabs), tampão CutSmart 1X e água, num volume final de 30 μ L. A reação foi incubada a 22 °C por 3 h, e inativada a 65 °C por 10 minutos.

Foi então feita a amplificação do inserto ligado, utilizando os mesmos parâmetros já descritos, empregando o *primer forward* do primeiro inserto (cohesina, ou proteínas de simetrias C3, C4, ou C6) e o *reverse* do segundo (dockerinas ou proteínas de simetria C2). O produto da amplificação foi purificado por gel de agarose, através do *kit* Wizard® SV Gel and PCR Clean-Up System (Promega). Esse inserto amplificado foi tratado com T4 DNA Polimerase, na presença de dATP, nas mesmas condições descritas anteriormente.

Para hibridização dos DNAs, foi feita a mistura na razão 3:1 (inserto:vetor), em molaridade. A mistura foi feita num volume final de 10 μ L, contendo 150ng do vetor, e a quantidade equivalente do inserto, calculadas considerando 660pg/pmol como massa molecular média de um par de nucleotídeos.

A quantificação precisa dos DNAs envolvidos no processo é um passo crucial para o sucesso da técnica. No entanto, essa quantificação deve ser feita antes da adição de deoxinucleotídeo, uma vez que há interferência deste na leitura espectroscópica do DNA. Dessa forma, a leitura foi feita assim que purificado cada produto de amplificação e a concentração final de DNA foi calculada levando em conta as diluições das etapas subsequentes. A mistura foi incubada a 37°C por 15 minutos e então usada para transformar 40 μ L de células ultra competentes de DH5 α , previamente preparadas de acordo com o método de Inoue.⁴⁶

3.12 Clonagem por SLIC

A clonagem por SLIC se deu pela amplificação dos insertos e vetor, como descrito anteriormente, utilizando *primers* não forforilados, uma vez que não foi necessária a subsequente ligação *in vitro*. Após purificação com o *kit* Wizard® SV Gel and PCR Clean-Up System (Promega) e tratamento com *DpnI*, os fragmentos foram tratados com T4 DNA Polimerase como descrito anteriormente, porém na ausência de quaisquer deoxinucleotídeos, num volume final de 15 μ L. A mistura foi incubada a 22 °C por 30 minutos e então adicionamos 1 μ L de dATP a 10 mM à mistura, incubando-a em gelo imediatamente.

Para hibridização dos DNAs, foi feita a mistura equimolar dos três elementos necessários. A mistura foi feita num volume final de 10 μ L, contendo 150 ng do vetor, e a

quantidade equivalente de cada inserto, calculadas considerando 660 pg/pmol como massa molecular média de um par de nucleotídeos, resultando em 0,04 μM de cada componente. O DNA foi quantificado ao final da purificação de todas as amplificações, conforme descrito anteriormente. A mistura foi incubada a 37 °C por 15 minutos e então usada para transformar 40 μL de células ultra competentes de DH5 α .

3.13 Transformação de células DH5 α ultra competentes

A transformação por choque térmico de células DH5 α ultra competentes previamente preparadas de acordo com o método de Inoue⁴⁶ seguiu o mesmo protocolo descrito para transformação de células BL21(DE3) com os DNAs codificadores das cohesinas e dockerinas de *R. flavefaciens*, descrito acima, com a diferença da incubação inicial de 15 minutos em gelo da mistura de DNAs hibridizados proveniente da clonagem por SLIC adicionada às células.

Colônias crescidas nessas placas foram inoculadas em 5 mL de meio LB suplementado com 50 $\mu\text{g/mL}$ de canamicina e incubadas *overnight* a 37 °C. O seu DNA plasmidial foi então extraído com o *kit* Wizard® Plus SV Minipreps DNA Purification System (Promega) e testado, por PCR para confirmar recombinantes. Essa reação foi feita com *kit* MyTaq (Bioline) e continha tampão 1X (1 mM dNTPs, 3 mM MgCl_2), 0,1 μM de cada primer T7 promoter e T7 terminator, 1 μL do DNA plasmidial previamente diluído 5 vezes, 2,5 Unidades de MyTaq DNA Polimerase e água para volume final de 25 μL . O resultado foi analisado por gel de agarose, e plasmídios confirmados como recombinantes tiveram as sequencias confirmadas por sequenciamento através do método de terminação de cadeia, de Sanger e foram utilizados para transformar células de BL21(DE3) quimiocompetentes, do mesmo modo que descrito anteriormente.

3.14 Expressão e purificação das fusões

A expressão e purificação dessas proteínas seguiram os mesmos passos descritos para as cohesinas e dockerinas de *R. flavefaciens*, descrita acima, com a diferença da composição do Tampão utilizado (Tris 25 mM pH 8,0, NaCl 300 mM).

3.15 Microscopia Eletrônica de Transmissão com coloração negativa

A Microscopia eletrônica de transmissão é uma técnica de microscopia de alta resolução, que se deve à utilização de feixes de elétrons como fonte de energia para visualização da amostra, os quais possuem um comprimento de onda muito baixo. O feixe passa através da amostra, tipicamente um filme fino (<100nm) ou uma suspensão num *grid*, de forma que sua interação com o material gera o contraste que forma a imagem, que é o resultado dos elétrons transmitidos. Amostras biológicas são formadas de átomos leves, que não interferem intensamente com elétrons transmitidos, resultando em um contraste baixo na aquisição de imagens. Nesses casos, costuma-se utilizar a técnica de coloração negativa, em que um sal de elemento pesado (por exemplo Urânio, Tungstênio, Molibdênio, Vanádio) é usado para permear e recobrir a amostra, gerando um espalhamento diferente de elétrons entre a amostra biológica, relativamente transparente aos elétrons, e o corante negativo, opaco aos elétrons. Assim, consegue-se um contraste suficiente, que representa o envelope molecular da partícula analisada.⁴⁷

Foram realizados testes preliminares para algumas condições de obtenção das redes auto montáveis a partir das proteínas 3C19-2Y3M e 2Y3N-3BUT. Os primeiros testes foram realizados no LCE - Laboratório de Caracterização Estrutural, DEMa / UFSCar, São Carlos-SP, utilizando um microscópio FEI TECNAI G² F20 HRTEM. As amostras foram montadas em *grids Cflat 400 mesh R2*.

Para testar as condições de formação das redes proteicas, misturamos as proteínas a concentrações finais de 5 µM, em tampão 20 mM Tris pH 8,0, 300 mM NaCl, 2,5 mM CaCl₂, variando o tempo de incubação (30 minutos ou 4 horas), a qual foi realizada em solução ou diretamente no *grid*. As proteínas incubadas em solução foram transferidas para o *grid*, onde ficaram por 1 minuto antes da secagem. Os *grids* foram então lavados 2 vezes com tampão HEPES 20 mM pH 7,4 e então recobertos com acetato de uranila 1%, por 1 minuto.

No segundo teste, alteramos a concentração final das proteínas para 2,5 µM, deixando-as incubadas em solução por 4 horas. O acetato de uranila foi depositado sobre a amostra por 20 segundos. O restante do tratamento foi igual ao descrito anteriormente.

Testes também foram realizados no LNNano, no Centro Nacional de Pesquisa em Energia e Materiais (CNPEM), Campinas-SP, utilizando um microscópio TEM JEOL 1400 PLUS. As amostras foram preparadas utilizando *grids Ultrathin Carbon film holey carbon 400 mesh*. Testamos aplicar cargas positiva e negativa de 25 mA por 25 segundos aos *grids*,

testando a incubação das proteínas a concentrações finais de 5 μ M com 2,5 mM CaCl_2 , como nos primeiros preparos. As proteínas foram incubadas em solução por 4 horas e depositadas nos grids por 1 minuto e lavadas com água. O acetato de uranila 2% foi depositado em duas etapas de 30 segundos.

4 Resultados

4.1 Expressão e purificação de cohesinas e dockerinas de *R. flavefaciens*

A pesquisa com os componentes do celulosossomo de *R. flavefaciens* já estava em desenvolvimento em nosso laboratório, tendo sido iniciada pelos Drs. Juliana Cheleski Wiggers e Wesley Cardoso Generoso. Numerosos módulos cohesina e dockerina de escafoldinas se encontravam clonados em vetor de expressão, e alguns estudos, sobretudo estruturais, haviam sido realizados, como já mencionado. A cristalização e resolução estrutural das cohesinas Rf-CohB1 e Rf-CohB4, bem como do complexo Rf-DocA/Rf-CohB4 foram executadas, de forma que foi necessário finalizar o refinamento para obtenção do modelo final. Ainda, havia estudos preliminares de afinidade entre a Rf-DocA e algumas das cohesinas, que foram revisados para o planejamento de novos ensaios. Para a conclusão desses trabalhos, visando a compreensão das bases da montagem do celulosossomo de *R. flavefaciens*, realizamos novos testes de afinidade entre Rf-DocA e as cohesinas Rf-CohB1 (MST), Rf-CohB4 e Rf-CohB6 (ITC), incluindo diversos mutantes de Rf-CohB1 (Rf-CohB1_{G124A}, Rf-CohB1_{Q125A}, Rf-CohB1_{G126A}, Rf-CohB1_{G128A} e Rf-CohB1_{R129A}) e de Rf-DocA (Rf-DocA_{V813E}, Rf-DocA_{V817E} e Rf-DocA_{V865E}), também previamente preparados. A cristalização de mutantes da Rf-CohB1 foi realizada para resolução estrutural, visando esclarecer o mecanismo de interação à Rf-DocA e esses resultados, juntamente com as estruturas previamente determinadas, foram contextualizadas à literatura, preenchendo uma lacuna do conhecimento que carecia de estudos estruturais. O conjunto de experimentos e análises foi realizado com a contribuição dos doutorandos Everton Silva e Raphael Meneghello. Para a realização desses ensaios, foi necessário expressar e purificar os módulos citados, através da transformação de células BL21 (DE3) com os plasmídeos previamente preparados.

Após o cultivo de transformantes e indução da expressão, cada proteína foi purificada por IMAC em coluna TALON (Clontech), tratada com a protease Ulp1 para clivagem de SUMO, proteína de fusão que foi em seguida removida com uma segunda IMAC em TALON (Clontech). As proteínas clivadas foram então purificadas por cromatografia de exclusão molecular, através da coluna Superdex 75 10/300 GL (GE Healthcare). A pureza das amostras foi testada por SDS-PAGE em cada etapa de purificação, como mostrado na Figura 14, no exemplo, a purificação da Rf-CohB1. O mesmo procedimento foi dado às proteínas Rf-DocA,

Rf-CohB4, Rf-CohB6 e todos os mutantes envolvidos, resultando em expressão de proteínas homogêneas e puras, em rendimento suficiente para realização de todos ensaios planejados.

Figura 14 - Análise da purificação de Rf-CohB1 em TALON por SDS-PAGE. (M) Marcador de massa molecular, com a massa de cada banda indicada à sua esquerda. (1-8) Primeira purificação em TALON. (1) Fração insolúvel. (2) Eluato não ligado à coluna. (3-5) Lavagem com Tampão A (Tris 25 mM pH8,0, NaCl 300 mM, CaCl₂ 2 mM) adicionado de 5mM Imidazol. (6) Lavagem com Tampão A. (7-8) Lavagem com Tampão A adicionado de 300mM Imidazol. (10-14) Segunda purificação em TALON. (10) Eluato não ligado à coluna. (11-12) Lavagem com tampão A. (13-14) Lavagem com tampão A adicionado de 500mM Imidazol. O poço (9) não foi utilizado. (B) Análise da cromatografia de exclusão molecular de Rf-CohB1. (M) Marcador de massa molecular, com a massa de cada banda indicada à sua esquerda. (1-5) Amostras de frações coletadas da cromatografia.

Fonte: Elaborada pelo autor.

4.2 Determinação estrutural do complexo Rf-DocA/Rf-CohB4 e Rf-CohB4 isolada

Os dados de difração dos cristais de Rf-DocA/Rf-CohB4 e Rf-CohB4 isolada foram coletados na linha MX-2 do Laboratório Nacional de Luz Síncrotron, CNPEM, Campinas-SP. O complexo Rf-DocA/Rf-CohB4 cristalizou no grupo espacial P4₁2₁2, com um heterodímero na unidade assimétrica e difratou raios-X a uma resolução máxima de 1,6 Å. A resolução da estrutura foi realizada por SAD a partir do sinal anômalo da difração de átomos de iodo, através do quick-soaking do cristal no tampão da condição de cristalização suplementado de 500 mM de NaI e 20% etilenoglicol. O cristal incorporou 17 átomos de iodo na unidade assimétrica.

A Rf-CohB4 isolada cristalizou no grupo espacial C222₁, com um monômero na unidade assimétrica, e difratou raios-X a uma resolução máxima 1,5 Å. A estrutura foi resolvida por substituição molecular utilizando como modelo de busca a estrutura da cohesina obtida da construção do modelo do complexo Rf-DocA/Rf-CohB4 resolvido por SAD. Ambas

as estruturas foram refinadas a valores de $R_{\text{factor}}/R_{\text{free}}$ de 13,6/15,1 (para Rf-CohB4 isolada) e 16,0/18,0 (para o complexo Rf-DocA/Rf-CohB4), apresentando excelentes parâmetros geométricos. O resumo dos dados cristalográficos da coleta e refinamento pode ser encontrado na Tabela 1.

Tabela 1 - Parâmetros estatísticos da coleta, processamento e refinamento dos dados cristalográficos do complexo Rf-DocA/Rf-CohB4 e Rf-CohB4 isolada. Os dados em parênteses correspondem faixa de maior resolução.

	Complexo Rf-CohB4/Rf-DocA	Rf-CohB4
Dados da Coleta		
Comprimento de onda (Å)	1,4590	1,4590
Temperatura (K)	100	100
Grupo Espacial	P4 ₁ 2 ₁ 2	C222 ₁
Célula Unitária (Å)	58,3; 58,3; 174,8	59,3; 87,9; 60,5
Resolução (Å)	55,3 - 1,6 (1,7 - 1,6)	49,2 - 1,5 (1,6 - 1,5)
Reflexões Únicas	76178 (12169)	25440 (2844)
Reflexões Totais	779522 (119961)	92874 (6466)
Redundância	10,2 (9,8)	3,6 (2,3)
Completeza (%)	99,7 (98,4)	93,9 (65,7)
I/σ (I)	20,8 (5,4)	20,9 (8,8)
R-meas (%)	8,3 (47,0)	5,3 (11,0)
Refinamento da Estrutura		
R-work (%)	16,0	13,6
R-free (%)	18,0	15,1
Número de Átomos		
Água	247	209
Iodo	17	-
Proteínas	1712	1083
B-fator Médio		
Água	29,9	28,2
Iodo	33,2	-
Proteínas	14,8	9,3
RMSD		
Ligações (Å)	0,017	0,019
Ângulos (°)	1,692	1,811
Gráfico de Ramachandran		
(%)		
Favoráveis	99,1	99,3
Permitidas	0,9	0,7
Não Permitidas	0,0	0,0

Fonte: Elaborada pelo autor.

4.2.1 Análise estrutural da dockerina de ScaA presente no complexo Rf-CohB4/Rf-DocA

A escafoldina ScaA de *Ruminococcus flavefaciens*, é um dos principais componentes do celulosomo dessa espécie bacteriana, sendo responsável pela integração de enzimas à escafoldina ScaB, e consequentemente, ao sistema como um todo. Por esse motivo, a interação da dockerina de ScaA com as cohesinas de ScaB é central para a montagem do complexo celulosomo. Inclusive, essa interação é classificada como única integrante do grupo 5, na subdivisão das cohesinas e dockerinas do tipo III. Alguns estudos sobre essa interação foram conduzidos anteriormente,¹⁶⁻¹⁸ gerando evidências bioquímicas e biofísicas que elucidaram parcialmente a sua natureza. No entanto, alguns aspectos da interação só podem ser investigados plenamente com a obtenção do modelo cristalográfico de sua estrutura, o que não estava disponível até o momento. Assim, nosso trabalho mostra a primeira estrutura da dockerina Rf-DocA, permitindo a elucidação completa de aspectos estruturais da dockerina e da interação a cohesinas de ScaB.

A estrutura da Rf-DocA observada no cristal do complexo com Rf-CohB4 é composta de três α -hélices, como observado nas estruturas de dockerinas de tipos I e II, onde as hélices 1 e 3 são antiparalelas, formando uma superfície planar que interage com as cohesinas de ScaB (Figura 15A). Comum a todas as dockerinas já descritas, cada uma dessas hélices em Rf-DocA contém em seu N-terminal um sítio de ligação a cálcio, ao contrário do proposto pelo modelo de Karpol¹⁸, o qual considerou que apenas o primeiro sítio, canônico, seria funcional para ligação de cálcio. Esse sítio é composto do motivo loop-hélice que forma o EF-hand clássico de 12 resíduos que coordenam o cálcio numa configuração bipirâmide pentagonal. Os resíduos localizados nas posições 1 (D804), 3 (N806) e 5 (D808) fornecem o oxigênio do carboxilato de suas cadeias laterais como ligantes para o cálcio, enquanto o resíduo da posição 7 (V810) fornece o oxigênio da carbonila da cadeia principal. Uma molécula de água faz a ponte com o resíduo da posição 9 (N812) e o resíduo na posição 12 (D815) fecha o loop de coordenação ao cálcio através de uma interação bidentada (Figura 15B).

O segundo sítio de cálcio (Figura 15C) é semelhante ao da dockerina de CttA de *R. Flavefaciens* (PDB ID:4IU2)¹², sendo que em ambas ocorre uma inserção atípica (de 14 resíduos em Rf-DocA e 13 resíduos na dockerina de CttA) que interrompe o loop, deslocando as posições de coordenação do cálcio normalmente observadas nos motivos *EF-hand* (Figura 16). Mesmo com essa inserção, esse loop provê as mesmas interações observadas em um motivo EF-hand clássico. Nesse caso, o íon cálcio também é coordenado com configuração

bipirâmide pentagonal. Oxigênios da carbonila da cadeia principal são providos como ligantes nos resíduos das posições 3 (V842), 21 (K858) (posição 7 no motivo EF-hand clássico) e 4 (D843), sendo neste último, coordenado por intermédio de uma molécula de água. O resíduo da posição 23 (T860) (posição 9 no motivo EF-hand clássico) interage com o íon cálcio através de uma molécula de água, de modo similar ao do loop clássico. Resíduos convencionais, que fornecem oxigênios da carboxila das cadeias laterais são encontrados nas posições 1 (D840) e 26 (D863) (posição 12 no motivo EF-hand clássico), o qual é o ligante bidentado (Figura 15C).

Figura 15 - Estrutura da Rf-DocA e seus dois sítios de ligação a cálcio. (A) A dockerina Rf-DocA colorida em padrão arco-íris, com N-terminal em azul, e C-terminal vermelho. Os íons de cálcio estão representados por esferas amarelas. (B) O primeiro sítio de ligação a cálcio adjacente à hélice 1. (C) O segundo sítio de ligação a cálcio, adjacente à hélice 3. Em (B) e (C), os íons de cálcio estão representados por esferas verdes.

Fonte: Elaborada pelo autor.

Havia um modelo proposto previamente¹⁶ para Rf-DocA, no qual os terminais N e C formavam um fecho intramolecular, através da interação entre W802 e P878. Essa interação estabilizaria a dockerina, permitindo a ligação a suas cohesinas correspondentes. Nosso modelo cristalográfico revela que, de fato, esses dois resíduos apresentam proximidade.

Porém, também estão próximos do segmento horizontal da já mencionada inserção de 14 resíduos. Este loop parece ser estabilizado por uma porção extensa do N-terminal, pela hélice 2 e o C-terminal. Isso sugere uma rede de interações ainda mais extensa que contribui para a estabilidade do módulo (Figura 17).

Figura 16 - Sobreposição do segundo sítio de ligação a cálcio de Rf-DocA e da dockerina CttA, mostrando a inserção que ocorre no loop entre as hélices 2 e 3, semelhante nas duas proteínas. Em verde, Rf-DocA e o íon de cálcio coordenado. Em bege, a dockerina de CttA e o íon de cálcio coordenado.

Fonte: Elaborada pelo autor.

Figura 17 - Interações que contribuem para estabilidade da estrutura da Rf-DocA. A figura mostra, em laranja, o grampo intramolecular 16 entre W802 e P878, nas regiões N-terminal e C-terminal, respectivamente. Adicionalmente, observamos a interação da região estendida do N-terminal (em rosa) com a hélice 2 (região em amarelo) e com a inserção do loop de ligação ao cálcio (em azul).

Fonte: Elaborada pelo autor.

A divergência entre o modo de ligação das dockerinas Rf-DocA e a dockerina de CttA mostra um padrão de interações similar ao dos tipos I e II, que converge com as funções desempenhadas, embora ambas sejam classificadas como grupo III. A dockerina de CttA possui módulo-X acoplado, e sua ligação à cohesina de ScaE é dada por ambas as hélices da dockerina. Essa interação é responsável pela ancoragem da proteína à superfície celular. Já Rf-DocA é ligada à Rf-CohB4 numa interação que se assemelha à de tipo-I. A semelhança entre ambas é a orientação da dockerina no complexo, que se aproximam mais do modo alternativo de ligação possível para dockerinas do tipo-I, rotacionadas de 180°. O alinhamento tridimensional da estrutura de Rf-DocA contra a base de dados do PDB,⁴¹ através do DALI Server⁴⁸ revelou uma semelhança notavelmente maior a dockerinas de tipo-I (Tabela 2).

Tabela 2 - Classificação de similaridade estrutural à Rf-DocA. O ranking foi editado para retirada de duplicações, quando uma mesma entrada possui duas cadeias com similaridade à proteína consultada. (n) indica a posição do resultado no ranking.

n	Referência (PDB ID)	Tipo	Z-score	rmsd (Å)	Resíduos alinhados	Total resíduos	Identidade sequencial (%)	Organismo
2	4FL4	I	10,7	1,7	66	68	29	<i>C. thermocellum</i>
5	4UYP	I	10,2	2,0	68	73	19	<i>A. cellulolyticus</i>
9	2CCL	I	9,7	1,5	62	62	29	<i>C. thermocellum</i>
12	3UL4	I	9,5	1,7	63	63	32	<i>C. thermocellum</i>
15	2MTE	I	9,2	1,8	65	69	32	<i>C. thermocellum</i>
16	4DH2	I	9,0	2,1	67	70	25	<i>C. thermocellum</i>
18	2VN6	I	8,2	1,9	61	64	25	<i>C. cellulolyticum</i>
19	5LXV	III	8,0	1,6	59	65	22	<i>R. flavefaciens</i>
21	4IU3	III	7,8	2,9	81	236	16	<i>R. flavefaciens</i>
22	1OHZ	I	7,8	1,6	56	56	29	<i>C. thermocellum</i>
26	4WKZ	III	6,8	2,4	77	231	17	<i>Flavefaciens</i>
27	2OZN	II	4,4	2,2	49	131	16	<i>C. perfringens</i>
28	2Y3N	II	4,0	2,4	47	60	30	<i>B. cellulosolvens</i>

Fonte: Elaborada pelo autor.

A dockerina à qual Rf-DocA mais se assemelha é a de dockerina Cel9D, de tipo-I de *Clostridium thermocellum* (PDB ID:4FL4), com z-score 10.7, r.m.s.d. 1.7 Å, com 66 resíduos alinhados de seus 68, e 29% de identidade sequencial (Figura 18A). Essa dockerina, no entanto, não possui a inserção no sítio de cálcio, tem uma hélice 2 mais longa, e o loop que conecta as hélices 1 e 2 não se alinham completamente. Essas observações foram comuns quando comparando outras dockerinas de tipos I e III, enquanto as de tipo-II (PDB Ids: 2OZN, 2Y3N) (Figura 18B) foram menos similares a Rf-DocA (z-scores 4,4 e 4,0), já que além das diferenças notadas anteriormente, essas dockerinas não possuem a hélice-2 em sua estrutura. As dockerinas de tipo III de CttA e a de uma enzima não identificada complexada com a cohesina de ScaG (PDB Ids: 4IU3, 4WKZ - z-scores: 7.8 e 6.8) foram as únicas a

apresentar a inserção de 14 resíduos em suas estruturas, nas quais apareciam próximas ao módulo-X que apresentam, levando ao alinhamento incompleto desse loop em relação ao de Rf-DocA (Figura 18C). Essas duas dockerinas, no entanto, apresentam uma estrutura muito mais complexa, com hélices e loops mais longos, além de duas hélices adicionais que interagem com seu módulo-X. A dockerina de tipo-III de uma proteína não identificada, complexada com a cohesina de ScaC (Figura 18D) em sua estrutura (PDB ID: 5LXV - z-score: 8.0) não apresenta a inserção de 14 resíduos e tem um loop que conecta as hélices 1-2 mais curto, que aparece orientado mais ao centro da dockerina, como efeito das interações que faz com a cohesina, através do β -flap protuberante de sua fita 8. As características distintas da Rf-DocA são seu N-terminal alongado e a inserção de 14 resíduos no segundo sítio de cálcio, que o fazem ser atípico.

Figura 18 - Sobreposição da Rf-DocA com os 3 diferentes tipos de dockerina. (A) Sobreposição da Rf-DocA com a dockerina tipo I de PDB 4FL4 (ciano). (B) Sobreposição da Rf-DocA com a dockerina tipo II de PDB 2Y3N (amarelo). (C) Sobreposição da Rf-DocA com a dockerina tipo III de PDB 4WKZ (laranja). (D) Sobreposição da Rf-DocA com a dockerina tipo III de PDB 5LXV (cinza). Rf-DocA representada em verde (A-D)

Fonte: Elaborada pelo autor.

4.2.2 Análise estrutural do quarto módulo cohesina de ScaB presente no complexo Rf-CohB4/Rf-DocA

A cohesina Rf-CohB4 forma um β -sanduíche de 9 fitas com a clássica topologia de rocambole. As fitas 8,3, 6 e 5 compõem a face de ligação à dockerina, enquanto as fitas 9, 1, 2, 7, e 4 formam a face oposta do sanduíche (Figura 19). As fitas são antiparalelas entre si, exceto pelas fitas paralelas 1 e 9. A fita 8 é alongada o suficiente para conseguir se conectar à fita 3, por seu C-terminal, e à fita 9, por seu N-terminal, interligando ambas as faces do

sanduíche. Esse enovelamento é compartilhado com as cohesinas dos tipos I e II, bem como com a cohesina de ScaE de *R. flavefaciens* (tipo III) (Figura 19).

Figura 19 - Estrutura da Rf-CohB4 colorida em padrão arco-íris, com N-terminal em azul, e C-terminal em vermelho.

Fonte: Elaborada pelo autor.

Na cohesina de ScaE, a fita 8 é interrompida por um β -flap (Figura 20B), que ocupa posições diferentes quando há interação com a dockerina de CttA e quando está isolada (PDB IDs:4IU3 e 2ZF9), indicando que esse β -flap possa exercer um papel regulatório na ligação à dockerina.¹² Apesar da ausência de tal β -flap na Rf-CohB4, o loop que conecta as fitas- β 6-7 parece ter uma função análoga na interação com Rf-DocA (Figura 20A). A comparação entre as estruturas da Rf-CohB4 isolada e em complexo com Rf-DocA mostram uma mudança de posição nessa região, que sugere uma mudança conformacional para acomodação da dockerina. O mesmo fato foi observado para a cohesina 1 de ScaB, discutida abaixo, o que sugere que seja um mecanismo comum às interações entre cohesinas de ScaB e a dockerina de ScaA no celulosomo de *R. flavefaciens*. Possivelmente, cada grupo apresente diferentes loops da cohesina como reguladores da interação.

Considerando as estruturas disponíveis no PDB para a busca por similaridade estrutural, através do servidor DALI,⁴⁸ Rf-cohB4 é mais similar à cohesina 3 de ScaB de *R. flavefaciens* linhagem FD-1 (PDB ID:5AOZ) (Figura 21C), com z-score 17,1, r.m.s.d. 1,9Å, com alinhamento de 131 resíduos, dos 141 possíveis e uma identidade sequencial de 27% (Tabela 3). A maior diferença na face de ligação à dockerina é um loop 2 resíduos mais longo ligando as fitas 6-7. A face oposta apresenta uma inclinação em relação à Rf-CohB4, de modo que o N-terminal das fitas 1 e 9 e o C-terminal da fita 2 não se alinham inteiramente. Devido a

serem de linhagens diferentes, essas cohesinas pertencem a grupos diferentes, o que significa que se ligam a dockerinas diferentes. A alta similaridade entre as estruturas sugere que a divergência entre as sequências seja responsável pela especificidade de interações de cada uma, tendo em vista a identidade sequencial de apenas 27% entre elas.

Figura 20 - Conformações do loop 6-7 da cohesina Rf-CohB4 e β -flap8, na cohesina Rf-CohE, na face de ligação de suas respectivas dockerinas. (A) Diferentes conformações do Loop 6-7 de Rf-CohB4, em ciano, fechado, em verde, aberto. (B) Diferentes conformações do β -flap8 de Rf-CohE, em cinza, fechado, em rosa, aberto.

Fonte: Elaborada pelo autor.

Figura 21 - Sobreposição da Rf-CohB4 com os 3 diferentes tipos de cohesina. (A) Sobreposição da Rf-CohB4 com a cohesina de tipo I de PDB 3UL4 (laranja). (B) Sobreposição da Rf-CohB4 com a cohesina de tipo II de PDB 2Y3N (vermelho). (C) Sobreposição da Rf-CohB4 com a cohesina de tipo III de PDB 5AOZ (amarelo). Rf-CohB4 está representada em verde (A-C).

Fonte: Elaborada pelo autor.

Um grande número de cohesinas de tipo-I aparecem em seguida na classificação (Tabela 3) por similaridade, abrangendo z-scores 15.2-13.9, r.m.s.d. 2.1-2.5 Å (Figura 21A). As dockerinas de tipo-III de ScaC (PDB:5LXV) e ScaE (PDB:2ZF9) exibiram z-scores 13.5 e 13.2, com r.m.s.d. 2.3 Å e 2.1 Å. As de tipo-II de *Clostridium perfringens* (PDB:2OZN) e *Bacteroides cellulosolvens* (PDB:2Y3N) (Figura 21B), exibiram z-values 12.2 e 10.5, com r.m.s.d. 3.0 Å e 2.6 Å. Comparando as estruturas em detalhes, observamos fitas 5 e 6 mais

extensas, que aparentam prover uma superfície maior para ligação de dockerinas. Loops mais longos conectando as fitas 4-5 e as fitas 6-7 foram observadas na maioria das estruturas examinadas. A ausência de estruturas típicas de cohesinas de tipo-II, como β -flaps nas fitas 4 e 8, ou hélice conectando as fitas 6-7 e a disposição da lista de classificação de similaridade estrutural corroboram a ideia de que as cohesinas de ScaB da linhagem 17 de *R. Flavefaciens* se assemelham funcionalmente a cohesinas de tipo-I (Figura 19).

Tabela 3 - Classificação de similaridade estrutural à Rf-CohB4. O ranking foi editado para retirada de duplicações, quando uma mesma entrada possui duas cadeias com similaridade à proteína consultada. (n) indica a posição do resultado no ranking. * A cohesina Rf-CohB1 foi incluída, com maior similaridade, mas não tem número do ranking pois não está depositada no PDB.

n	Referência (PDB ID)	Tipo	Z-score	rmsd (Å)	Resíduos alinhados	Total resíduos	Identidade sequencial (%)	Organismo
*	Rf-CohB1	III	21,9	1,3	131	133	44	<i>R. flavefaciens</i>
1	5AOZ	III	17,1	1,9	131	141	27	<i>R. flavefaciens</i>
2	3UL4	I	15,2	2,3	127	142	19	<i>C. thermocellum</i>
3	4UMS	I	15,1	2,1	130	154	21	<i>B. cellulosolvens</i>
4	4UYP	I	15,0	2,1	128	144	16	<i>A. cellulolyticus</i>
7	4DH2	I	14,8	2,2	130	156	12	<i>C. thermocellum</i>
9	1AOH	I	14,6	2,3	128	147	20	<i>C. thermocellum</i>
10	3KCP	I	14,6	2,4	129	303	19	<i>C. cellulolyticum</i>
11	2VN5	I	14,6	2,3	129	142	21	<i>C. thermocellum</i>
13	1G1K	I	14,4	2,5	130	143	21	<i>C. cellulolyticum</i>
14	3P0D	I	14,3	2,5	130	309	18	<i>C. thermocellum</i>
16	4FL4	I	14,3	2,3	128	308	19	<i>C. thermocellum</i>
25	5LXV	III	13,5	2,3	128	173	17	<i>R. flavefaciens</i>
27	1OHZ	I	13,2	2,3	121	140	19	<i>C. thermocellum</i>
28	1ANU	I	13,2	2,4	121	138	19	<i>C. thermocellum</i>
29	2ZF9	III	13,2	2,1	128	180	15	<i>R. flavefaciens</i>
31	2CCL	I	13,0	2,5	125	149	20	<i>C. thermocellum</i>
35	4N2O	III	12,5	2,7	132	191	12	<i>R. flavefaciens</i>
44	4WKZ	III	12,3	2,5	130	184	12	<i>R. flavefaciens</i>
47	2OZN	II	12,2	3,0	125	133	15	<i>C. perfringens</i>
50	4IU2	III	11,6	2,4	130	209	17	<i>R. flavefaciens</i>
52	2Y3N	II	10,5	2,6	126	173	18	<i>B. cellulosolvens</i>
54	1TYJ	II	10,4	2,6	125	170	18	<i>B. cellulosolvens</i>
55	3BWZ	II	10,3	2,9	124	171	17	<i>A. cellulolyticus</i>
56	3L8Q	II	10,3	2,6	125	337	16	<i>A. cellulolyticus</i>

Fonte: Elaborada pelo autor.

4.2.3 Estrutura do complexo Rf-DocA/Rf-CohB4

A interface do complexo Rf-DocA/Rf-CohB4 consiste da face da cohesina formada pelas fitas 8, 3, 6, e 5, os loops que conectam as fitas 5-6, 6-7 e 8-9 juntamente com a hélice 1 de Rf-DocA, o C-terminal da hélice 3 e o loop que conecta as hélices 1-2 (Figura 22). A região do interior da interação é predominantemente hidrofóbica, especialmente no contato

das hélices da dockerina com a folha β da cohesina, enquanto ligações de hidrogênio e interações iônicas são os principais promotores das interações periféricas, que auxiliam na estabilidade do complexo. A Tabela 4 traz a lista dos resíduos participantes da interação. Em um estudo prévio,¹⁷ foi observado que a deleção cumulativa de resíduos do C-terminal da Rf-DocA não afeta significativamente sua afinidade à cohesina Rf-CohB7. Quando deletaram 7 resíduos, incluindo E873, a afinidade foi afetada, ao que se atribuiu o efeito estabilizante que esse resíduo que tem sobre a hélice 3. Adicionalmente, é notável pelos resultados aqui apresentados que E873 é um participante significativo da interação, via ligação de hidrogênio e pontes salinas (Figura 22D). Assim, além de ter um papel na estabilidade da hélice 3, E873 participa diretamente na interface de interação com Rf-CohB4.

Figura 22 - Estrutura do complexo Rf-DocA/Rf-CohB4 e interações de sua interface. A dockerina (ciano) se liga à cohesina (em verde) por meio de uma rede de interações hidrofóbicas, no interior da interface (B) e por contatos periféricos compostos de ligações polares, como ligações de hidrogênio e pontes salinas (C e D). A interação se dá, principalmente, pelo contato entre a superfície plana das hélices 1 e 3 da dockerina com a superfície da folha da cohesina composta pelas fitas β 8, 3, 6 e 5, além dos loops que conectam as fitas 5-6, 6-7 e 8-9. Em (D), está evidenciado o contato do resíduo E873, da dockerina, com os resíduos K536 e H538, da cohesina, mostrando que esse resíduo contribui significativamente para a interação.

Fonte: Elaborada pelo autor.

Tabela 4 - Interações intermoleculares entre Rf-DocA e Rf-CohB4. O tipo de interação está indicado na terceira coluna: interação Hidrofóbica (HF), Ligação de Hidrogênio (LH), Ponte Salina (PS)

Rf-DocA	Rf-CohB4	Tipo de Interação
V813	A505	HF
	Q507	LH
	A540	HF
	F581	HF
A814	F581	HF
V816	A542	HF
V817	A505	HF
N820	Q543	LH
	A544	LH
R821	E585	LH / PS
Y828	N586	LH
N830	N586	LH
L868	V531	HF
	A540	HF
	A542	HF
I871	A540	HF
V872	V531	HF
	N533	LH
	A540	HF
E873	N533	LH
	K536	LH / PS
	H538	LH / PS
L874	V531	HF

Fonte: Elaborada pelo autor.

Embora a ligação da Rf-DocA na Rf-CohB4 se assemelhe a interação do tipo-I, na qual uma única hélice predomina na interação, é atípico que a hélice 1 seja a dominante. Isso resulta numa semelhança maior com dockerinas do tipo-I nas quais a hélice 3 é mutada, promovendo a ligação através da hélice 1. Assim, por exemplo, o complexo Rf-DocA/Rf-CohB4 apresenta maior semelhança com a estrutura de um complexo entre a dockerina de uma xilanase e a segunda cohesina de CipA de *C. Thermocellum* cuja hélice 3 sofreu uma mutação para favorecer essa interação (PDB ID:2CCL), do que com o mesmo complexo envolvendo a dockerina nativa (PDB ID:1OHZ). Como observado anteriormente, as dockerinas de tipo-I apresentam um duplo modo de ligação a cohesinas, o que faz com que possam ser rotacionadas em 180° , mantendo os principais resíduos para a interação nas mesmas posições. Nos tipos II e III, essa propriedade ainda não foi verificada, embora a falta de simetria seja indicativo de sua ausência, já que há baixa conservação de resíduos quando ocorre a rotação da dockerina. Embora Rf-DocA não apresente simetria entre suas hélices,

inspecionamos a possibilidade do duplo modo de ligação, fazendo a rotação virtual da dockerina em nosso modelo (Figura 23), invertendo as posições das hélices para averiguar se os resíduos essenciais à interação permaneceriam compatíveis. Observamos que os resíduos VAD (813-815 e 861-863), no N-terminal de cada hélice são as únicas posições conservadas. O restante das hélices compreende resíduos que resultariam em alteração no reconhecimento da cohesina no evento da rotação. Essa análise indica fortemente que a interação entre Rf-DocA e Rf-CohB4 ocorre via apenas o modo de ligação observado na estrutura cristalina.

Ainda nesse contexto, baseado na divergência sequencial apresentada pelos módulos cohesina de Rf-ScaB (Figura 24), especulamos a possível forma de interação entre Rf-DocA e as demais cohesinas. As cohesinas 1-4 apresentam uma convergência maior para os resíduos de interação à dockerina, enquanto as cohesinas 5-6 tem uma maior divergência e a cohesina 7 apresenta, em diferentes posições, semelhanças com cada um dos grupos. A partir do nosso modelo construído, analisamos as interações possíveis entre as cohesinas 1-4, 5-6 e 7 com as dockerinas Rf-DocA nativa e rotacionada virtualmente em 180° (que denominamos Rf-DocA₁₈₀), para os pontos de interação observados no complexo Rf-DocA/Rf-CohB4. O resumo de todas as interações observadas no cristal (Rf-CohB4) e preditas (Rf-CohB6 e Rf-CohB7) encontram-se na Tabela 5.

Figura 23 - Assimetria entre as hélices de Rf-DocA. A Rf-DocA em ciano corresponde à presente no complexo observado em nosso modelo cristalográfico. Já em verde, é representada a Rf-DocA₁₈₀, ou seja, dockerina virtualmente rotacionada, para simulação do modo alternativo de ligação. Estão destacados os resíduos V813, V817, L868 e V872, que fazem interação com a cohesina, e seriam substituídos, respectivamente, por V861, E865, N820 e N824 na rotação da dockerina.

Fonte: Elaborada pelo autor.

Figura 24 - Alinhamento das cohesinas da escafoldina ScaB de *Ruminococcus flavefaciens*. Em verde estão destacados os principais resíduos que fazem contato com a Rf-DocA. A numeração dos resíduos, acima das sequências, segue a numeração da Rf-CohB4 como referência.

Fonte: Elaborada pelo autor.

Tabela 5 - Análise de resíduos que fariam interação entre as dockerinas Rf-DocA e Rf-DocA₁₈₀, e as cohesinas Rf-CohB4 e Rf-CohB6. O levantamento foi feito com base nos resíduos de interação entre Rf-DocA e Rf-CohB4, observados na estrutura do complexo, e no alinhamento entre as cohesinas de ScaB. Os motivos da cohesina destacados com (*) indicam que não é observada interação nessa posição, mas é a vizinhança que acomoda o resíduo V817, na estrutura Rf-DocA/Rf-CohB4, e apresenta uma divergência significativa para as cohesinas Rf-CohB 5-7.

Rf-DocA	Rf-DocA ₁₈₀	Rf-CohB4	Rf-CohB6	Rf-CohB7
V813	V861	A505	G834	G1250
		A540	D870	V1286
		F581	K913	Y1328
		Q507	Q836	Q1252
V816	S864	A542	A872	A1288
V817	E865	Bolsão(XXFVSD)*	Bolsão+2aa (EGSLKIFD)*	Bolsão+2aa (IKDLYVVD)*
		A505	G834	G1250
N820	L868	Q543	N873	D1289
		A544	A874	G1290
R821	K869	E585	K917	E1332
L868	N820	A540	D870	V1286
		A542	A872	A1288
I871	L823	V531	T861	I1277
		A540	D870	V1286
V872	N824	A540	D870	V1286
		V531	T861	I1277

Fonte: Elaborada pelo autor.

A posição V813 na Rf-DocA interage com os resíduos A505, A540 e F581 de Rf-CohB4, por interações hidrofóbicas. Na Rf-DocA₁₈₀, V861 ocupa a posição equivalente à V813, o que implicaria na conservação das interações no caso da rotação da dockerina. A posição Q507 da Rf-CohB4 é conservada em todas as cohesinas e, portanto, sua interação de hidrogênio com a cadeia principal de V861 da Rf-DocA₁₈₀ seria mantida. Nas cohesinas 5 e 6, resíduos A540 e F581 dão lugar a D870 e K913, respectivamente (a partir daqui a numeração usada refere-se a Rf-CohB6), os quais possuem cadeias laterais eletricamente carregadas e,

portanto, não fariam interações com V861 de Rf-DocA₁₈₀. Pode-se destacar que a posição ocupada por A540 em Rf-CohB4 é chave para interação do complexo, uma vez que interage tanto com a região N-terminal de hélice $\alpha 1$ quanto C-terminal de $\alpha 3$. No entanto, os resíduos N824 e T860 da Rf-DocA₁₈₀ poderiam interagir com esses resíduos polares, compensando a divergência. A505 é substituído por G834 nas cohesinas 5-6, o que não implicaria em alteração da interação com V861 de Rf-DocA₁₈₀. No caso da Rf-CohB7, a posição ocupada por A540 em Rf-CohB4 é conservada.

Em relação ao resíduo R821 de Rf-DocA, que faz uma ponte salina com E585 de Rf-CohB4, sua posição seria ocupada por K869 na Rf-DocA₁₈₀, preservando assim essa interação. No caso das cohesinas 5-6 a posição de E585 é trocada por um resíduo eletricamente positivo, K917, impedindo a formação de ponte salina. Por outro lado, em Rf-CohB7, essa posição é conservada pela E1332.

A posição V816 de Rf-DocA faz interação hidrofóbica com o resíduo A542 de Rf-CohB4, resíduo conservado nas cohesinas 4-7 e substituído por G nas cohesinas 1-3, o que manteria a interação. Na Rf-DocA₁₈₀, a S864 ocupa a posição de V816, rompendo a interação com A542, o que potencialmente teria um efeito negativo em relação à interação dessa orientação de Rf-DocA com todas as cohesinas de Rf-ScaB.

A V817 de Rf-DocA fica localizada, no contato com a Rf-CohB4, próxima a um loop que forma um bolsão hidrofóbico, de motivo XXFVSD, conservado entre as cohesinas 1-4. Na Rf-DocA₁₈₀, V817 é substituído por E865, sendo que as cohesinas 5-6 e 7 exibem respectivamente aa sequência EGSLKIFD e IKDLYVVD no lugar do bolsão, dois resíduos mais longo que nas cohesinas 1-4, o que pode ocasionar um ambiente mais volumoso capaz de acomodar E865 de Rf-DocA₁₈₀. O resíduo V817 de Rf-DocA faz ainda uma interação hidrofóbica com A505 na Rf-CohB4, que é substituída por G nas outras cohesinas, sendo um aminoácido menos volumoso, o que pode, também, contribuir para a acomodação do E865 de Rf-DocA₁₈₀.

O resíduo N820 de Rf-DocA é o doador nas ligações de hidrogênio com Q543 e A544, de Rf-CohB4, interações que seriam rompidas na Rf-DocA₁₈₀, que substitui esse resíduo pela L868. A544 é conservada em todas as cohesinas, e poderia fazer interação hidrofóbica com L868 de Rf-DocA₁₈₀. Por outro lado, L868 faz, na Rf-DocA, interações hidrofóbicas com A540, A542 e V531, compatíveis com N820. Nas cohesinas 5 e 6, A540 e V531 de Rf-CohB4 são substituídos por D870 e T861, respectivamente, que criam um ambiente mais favorável à interação com N820 de Rf-DocA₁₈₀ que com L868 de Rf-DocA. As posições referentes aos

resíduos A540 e V531 em Rf-CohB4 são novamente conservadas em Rf-CohB7 pelos resíduos V1286 e I1277.

O resíduo I871 de Rf-DocA interage com A540 de Rf-CohB4, hidrofobicamente. Seu substituto D870, nas cohesinas 5-6 acomodam mais favoravelmente a L823 que substitui a I871 em Rf-DocA₁₈₀, compensando a ocupação estérica. A V872 de Rf-DocA faz interações hidrofóbicas com A540 e V531 de Rf-cohB4, substituídos por D870 e T861, nas cohesinas 5-6, que são adequados a fazer interações de hidrogênio com N824, posição equivalente da Rf-DocA₁₈₀. Em comparação com a Rf-CohB7, a conservação já mencionada nas posições referentes aos resíduos A540 e V531 de Rf-CohB4 sugere que a interação com Rf-DocA seja favorecida na forma observada no cristal.

Coletivamente, essas análises indicam a possibilidade de interação das cohesinas 5-6 com a Rf-DocA orientada diferentemente do observado na estrutura cristalográfica do complexo. Por outro lado, apesar de uma maior divergência em relação aos agrupamentos 1-4 e 5-6, a cohesina 7 de Rf-ScaB potencialmente interage com a dockerina Rf-DocA da mesma forma observada para Rf-CohB4. Esse achado nos motivou então a investigar o modo de ligação entre Rf-CohB6 e Rf-DocA, o que poderia demonstrar a capacidade da cohesina em compensar a falta de simetria nas hélices de uma dockerina, fornecendo diferentes superfícies para modos alternativos de interação. Para tal, realizamos ensaios de afinidade, por ITC, entre a Rf-DocA (nativa e mutantes V813E, V817E e V861E) e as cohesinas Rf-CohB4 e Rf-CohB6.

4.3 Calorimetria de Titulação Isotérmica para estudar o modo de ligação de Rf-DocA com cohesinas de ScaB

Para avaliar a possibilidade do duplo modo de ligação entre Rf-DocA e cohesinas de ScaB, construímos os mutantes pontuais V813E, V817E e V861E de Rf-DocA (Rf-DocA_{V813E}, Rf-DocA_{V817E} e Rf-DocA_{V861E}), os quais foram testados contra as cohesinas Rf-CohB4 e Rf-CohB6 nativas, por ITC. A seleção dos resíduos foi feita com base na análise da estrutura do complexo Rf-DocA/Rf-CohB4. As Valinas 813 e 817 são chaves na interação, participando de interações em que substituições para resíduos carregados e mais longos, como o ácido glutâmico, seriam desfavoráveis à interação observada na estrutura de Rf-DocA/Rf-CohB4. Embora V861 não participe de interações na orientação observada nesta estrutura, esse resíduo ocupa a posição equivalente à V813, caso ocorra a ligação da Rf-DocA₁₈₀ com as cohesinas (Figura 23). Os resíduos V813 e V861 são dois dos poucos conservados na rotação

da dockerina, localizados no N-terminal das hélices 1 e 3, respectivamente. Por esse motivo, e pela vasta rede de interações envolvendo V813 (Tabela 4), esses mutantes foram desenhados de forma a excluir cada um dos possíveis modos de ligação. Assim, no mutante Rf-DocA_{V861E}, o modo de ligação esperado era o que observamos na estrutura resolvida. Já no mutante Rf-DocA_{V813E}, caso ocorresse interação com a cohesina, seria observado o modo de ligação alternativo, em que a hélice 3 toma o lugar da hélice 1, ou seja, através da rotação da dockerina para o encaixe. Semelhantemente, o resíduo E865 ocuparia, na Rf-DocA₁₈₀, a posição em que se encontra a V817 na Rf-DocA. Essa é uma das posições que apresenta maior discrepância na natureza de aminoácidos, na comparação entre os dois possíveis modos de ligação com as cohesinas de ScaB. Dessa forma, essa mutação torna as hélices simétricas considerando esses quatro pontos (tendo V813, E817, V861, E865), e a ligação da Rf-DocA com essa mutação indicaria a possibilidade de ligação da Rf-DocA₁₈₀, caso ocorresse. (A partir daqui, Rf-DocA_{WT} é a notação utilizada para a Rf-DocA nativa, para diferenci-la dos mutantes)

Ensaio de ITC com mutante Rf-DocA_{V813E} não mostrou interação alguma contra Rf-CohB4, indicando que a posição V861 não compensou a mutação, o que exclui a possibilidade do duplo modo de ligação (Figura 25B). O mutante Rf-DocA_{V861E} confirmou a hipótese de único modo de ligação, uma vez que seus parâmetros termodinâmicos de ligação à Rf-CohB4 foram muito próximos dos observados para a interação da Rf-DocA_{WT} com Rf-CohB4 (Figura 25A e C). Nesses casos, foi observado uma interação de alta afinidade (Kd em torno de 1 nM) (Tabela 6), conforme esperado para a interação dockerina-cohesina. Adicionalmente, esses resultados indicam que os parâmetros da interação nativa não resultam da média dos dois possíveis modos de ligação ocorrendo simultaneamente.

Curiosamente, o mutante Rf-DocA_{V817E} mostrou uma diminuição de 3 ordens de magnitude para o Kd (em torno de 270 nM) de interação com Rf-CohB4, em relação à Rf-DocA_{WT}, com um ΔH positivo. Embora significativamente afetada, a interação *in vitro* não foi erradicada, como no mutante V813E. Isso sugere que a mutação pôde ser tolerada, devido ao ambiente extenso na região da Rf-CohB4 que circunda esse resíduo, o que poderia acomodar essa substituição. Apesar disso, é evidente a alteração na termodinâmica da interação, uma vez que o resíduo de ácido glutâmico é inserido em uma superfície hidrofóbica, o que faz com que a entropia da interação aumente substancialmente (80 J/K) em relação às demais interações (12 J/K). Esse ganho entrópico compensa a entalpia positiva, possibilitando a interação, o que não seria possível no caso do mutante V813E, uma vez que essa posição é localizada na extremidade da hélice, ou seja, exposta ao solvente. No entanto, a significância

biológica de uma diminuição de afinidade nessa proporção equivale à não ocorrência de tal interação. Essas observações reforçam que o contato entre o N-terminal da hélice 1 e o C-terminal da hélice 3 da Rf-DocA com a Rf-CohB4 são determinantes para a interação, enquanto o C-terminal da hélice 1 aparenta admitir certo grau de flexibilidade. O resumo dos dados referentes à interação pode ser consultado na Tabela 6.

Figura 25 - Ensaios de calorimetria de titulação isotérmica da Rf-CohB4. (A) Titulação da Rf-CohB4 em Rf-DocA_{WT}. (B) Titulação da Rf-CohB4 em Rf-DocA_{V813E}. (C) Titulação da Rf-CohB4 em Rf-DocA_{V861E} (D) Titulação da Rf-CohB4 em Rf-DocA_{V817E}. O mutante Rf-DocA_{V861E} apresentou uma interação muito semelhante à de Rf-DocA_{WT}, enquanto Rf-DocA_{V817E} resultou em uma dinâmica de interação bastante distinta. O mutante Rf-DocA_{V813E} não apresentou interação com Rf-CohB4.

Fonte: Elaborada pelo autor.

Tabela 6 - Parâmetros termodinâmicos da interação entre as cohesinas Rf-CohB4 e Rf-CohB6 e as dockerinas Rf-DocA_{WT} e mutantes Rf-DocA_{V813E}, Rf-DocA_{V817E} e Rf-DocA_{V861E}.

	Rf-DocA _{WT}	Rf-DocA _{V813E}	Rf-DocA _{V861E}	Rf-DocA _{V817E}
Rf-CohB4	Kd = 0,7 ± 0,9 nM		Kd = 1 ± 1 nM	Kd = 270 ± 70 nM
	ΔH = -9100 ± 100 J	-	ΔH = -8700 ± 200 J	ΔH = 15800 ± 500 J
	ΔS = 12 J/K	-	ΔS = 12 J/K	ΔS = 80 J/K
	ΔG = -12500 J		ΔG = -12200 J	ΔG = -8900 J
Rf-CohB6	Kd = 50 ± 30 nM		Kd = 60 ± 40 nM	Kd = 16000 ± 6000 nM
	ΔH = -2390 ± 80 J	-	ΔH = -2200 ± 100 J	ΔH = 19000 ± 6000 J
	ΔS = 25 J/K	-	ΔS = 17 J/K	ΔS = 90 J/K
	ΔG = -9800 J		ΔG = -7300 J	ΔG = -7800 J

Fonte: Elaborada pelo autor.

Para testar a hipótese levantada anteriormente, sobre a possibilidade de haver um modo diferente de ligação entre Rf-DocA_{WT} e as cohesinas 5 e 6 de ScaB, reproduzimos o ensaio do ITC, testando a interação entre Rf-DocA_{WT} e Rf-CohB6. Para observar o segundo modo de ligação, era esperado que o mutante Rf-DocA_{V813E} apresentasse interação comparável à da Rf-DocA_{WT}, e que o mutante Rf-DocA_{V861E} não apresentasse afinidade pela

Rf-CohB6. Ou seja, caso fosse visto o modo alternativo de ligação, veríamos o resultado oposto do observado para Rf-CohB4. Já para a mutação Rf-DocA_{V817E}, supusemos que o aumento do loop da cohesina Rf-CohB6 que circunda esse resíduo da Rf-DocA_{WT} acomodaria mais favoravelmente essa mutação. Assim, era esperada uma afinidade também comparável com a da Rf-DocA_{WT}, uma vez que essa mutação torna as hélices simétricas, como já discutido.

Os resultados, no entanto, foram qualitativamente muito semelhantes aos observados para Rf-CohB4, indicando que a ligação da dockerina ocorre do mesmo modo para ambas as cohesinas. O mutante Rf-DocA_{V813E} não apresentou interação com Rf-CohB6, enquanto o mutante Rf-DocA_{V861E} apresentou interação comparável à Rf-DocA_{WT} (Figura 26). O mutante Rf-DocA_{V817E} demonstrou afinidade à Rf-CohB6 diminuída em 3 ordens de grandeza (em torno de 16 μ M) e ΔH positivo, também como observado para a interação com a Rf-CohB4 (Figura 26D). É interessante destacar que a afinidade de Rf-DocA_{WT} por Rf-CohB6 (K_d em torno de 50 nM) é uma ordem de grandeza menor que sua afinidade com a Rf-CohB4, corroborando a análise estrutural que mostra uma posição chave de interação ocupado por D870 em Rf-CohB6 no lugar de A540 em Rf-CohB4. Em Rf-CohB7, essa mesma posição é conservada (V1286) e medidas de K_d em torno de 1 nM¹⁸, semelhantes a Rf-CohB4, corroboram a importância dessa posição para a afinidade da interação. O resumo dos dados referentes à interação consta na Tabela 6.

Figura 26 - Ensaios de calorimetria de titulação isotérmica da Rf-CohB6. (A) Titulação da Rf-CohB4 em Rf-DocA_{WT}. (B) Titulação da Rf-CohB4 em Rf-DocA_{V813E}. (C) Titulação da Rf-CohB4 em Rf-DocA_{V861E}. (D) Titulação da Rf-CohB4 em Rf-DocA_{V817E}. De forma semelhante às interações com Rf-CohB4, o mutante Rf-DocA_{V861E} apresentou uma interação muito semelhante à de Rf-DocA_{WT}, enquanto Rf-DocA_{V817E} resultou em uma dinâmica de interação bastante distinta. O mutante Rf-DocA_{V813E} não apresentou interação com Rf-CohB6.

Fonte: Elaborada pelo autor.

A diferença de afinidade a cada uma das 7 cohesinas de ScaB pode indicar como a escafoldina ScaA se acomoda na escafoldina ScaB. Essa interação é responsável por grande parte da organização do celulosomo em *R. flavefaciens*, como já destacado. Vale ressaltar que a comparação está sendo feita com medidas *in vitro*, com proteínas purificadas, e que há outros fatores que poderiam impactar as afinidades desses módulos *in vivo*.

4.4 Determinação estrutural de Rf-CohB1 isolada

Os dados de difração dos cristais de Rf-CohB1 foram coletados no ânodo rotatório do laboratório de cristalografia do IFSC/USP São Carlos. Ambos os cristais difratados cristalizaram na mesma condição, sendo obtidos dois empacotamentos cristalinos distintos. A forma cristalina I contém duas moléculas na unidade assimétrica e dados de difração foram coletados à uma resolução máxima de 2,3 Å. A determinação de sua estrutura foi realizada por SAD, conforme já descrito para o cristal do complexo Rf-DocA/Rf-CohB4. O cristal incorporou 9 átomos de iodo na unidade assimétrica.

A forma cristalina II difratou à resolução máxima de 1,5 Å e contém uma molécula na unidade assimétrica. A estrutura dessa forma cristalina foi resolvida por substituição molecular utilizando uma das estruturas da Rf-CohB1, obtida da construção do modelo da

forma cristalina I, como modelo de busca. O resumo dos dados cristalográficos da coleta e resolução estrutural de ambas as estruturas pode ser encontrado na Tabela 7.

Tabela 7 - Parâmetros estatísticos da coleta, processamento e refinamento dos dados cristalográficos de ambas as formas cristalinas de Rf-CohB1. Os dados em parênteses correspondem faixa de maior resolução.

	Rf-CohB-1 cristal I	Rf-CohB-1 cristal II
Dados da Coleta		
Comprimento de onda (Å)	1,5418	1,5418
Temperatura (K)	100	100
Grupo Espacial	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Célula Unitária (Å)	59,8; 60,0; 100,6	28,3; 60,3; 72,2
Resolução (Å)	32,3-2,3 (2,4-2,3)	40,0-1,5 (1,6-1,5)
Reflexões Únicas	31447 (4714)	21138 (3063)
Reflexões Totais	222892 (30661)	87147(11505)
Redundância	7,1 (6,5)	4,1 (3,8)
Completeza (%)	98,6 (91,7)	93,6 (85,7)
I/σ (I)	17,8 (2,8)	27,4 (7,2)
R-meas (%)	8,7 (72,2)	3,5 (20,8)
Refinamento da Estrutura		
R-work (%)	21,8	16,2
R-free (%)	26,2	17,9
Número de Átomos		
Água	148	158
Iodo	9	-
Proteínas	1928	999
B-fator Médio		
Água	37,141	27,469
Iodo	87,624	-
Proteínas	33,046	11,587
RMSD		
Ligações (Å)	0,008	0,006
Ângulos (°)	1,085	1,126
Gráfico de Ramachandran		
(%)		
Favoráveis	96,2	99,3
Permitidas	3,4	0,7
Não Permitidas	0,4	0,0

Fonte: Elaborada pelo autor.

4.4.1 Estrutura da Rf-CohB1

Todas as moléculas nas unidades assimétricas das duas formas cristalinas (3 no total) foram refinadas independentemente. A cohesina Rf-CohB1 apresenta o mesmo enovelamento e a topologia já descrita para Rf-CohB4, ou seja, o clássico sanduiche β de 9 fitas antiparalelas (com exceção das fitas 1 e 9), com topologia de rocambole (*jellyroll*), comum às

cohesinas dos três tipos (Figura 27A). As faces são formadas pelas fitas 8, 3, 6 e 5 (face de ligação à dockerina) e fitas 9, 1, 2, 7, e 4. A sobreposição das três moléculas refinadas, feita a cada par de moléculas, resultou em um r.m.s.d entre 0,3-0,8 Å, indicando que são virtualmente idênticas. Análise dessa sobreposição revelou que essa diferença se dá, predominantemente, pela flexibilidade do *loop* que conecta as fitas 6 e 7 (Figura 27B), que será discutida em mais detalhes posteriormente.

Semelhantemente a Rf-CohB4, a estrutura de Rf-CohB1 não apresenta β -flaps nas fitas 4 ou 8, nem hélice conectando as fitas 6-7, estruturas típicas de cohesinas do tipo II. A sobreposição entre Rf-CohB1 e Rf-CohB4 resulta em um notável z-score de 21,9, com r.m.s.d. de 1,3 Å, alinhando 131 de 134 possíveis resíduos, com identidade sequencial de 44% (Tabela 8). A análise dessa sobreposição (Figura 28A) revela como elementos mais distintos o loop entre as fitas 5-6, 2 resíduos mais longo na Rf-CohB4 e a orientação do loop 6-7, mais distante da fita 6 (próximo à “coroa” da cohesina) na Rf-CohB4. Essa observação se mantém mesmo comparando-se os loops de ambas as cohesinas na conformação fechada ou na conformação aberta, entre si. O principal fator que altera essa conformação é a orientação do resíduo G124 em Rf-CohB1 (correspondente a A542 em Rf-CohB4), que ocupa uma região não permitida a aminoácidos que apresentam cadeia lateral (Figura 27B)

Figura 27 - Estrutura da Rf-CohB1. (A) A cohesina acima, colorida em padrão arco-íris, com N-terminal em azul, e C-terminal vermelho. (B) Vista frontal do loop que conecta as fitas β 6-7, com a medição das distâncias relativas ao resíduo G169, de posição estática entre as estruturas (em verde, a conformação fechada, em magenta, a conformação aberta).

Fonte: Elaborada pelo autor.

Figura 28 - Comparação das conformações adotadas pelo loop entre as fitas 6-7 das cohesinas Rf-CohB1 e Rf-CohB4. (A) Sobreposição das estruturas Rf-CohB1 (verde) e Rf-CohB4 (ciano), ambas nas conformações fechadas do loop 6-7. (B) Comparação das orientações dos loops de Rf-CohB1 (magenta), e Rf-CohB4 (ciano) na conformação fechada (C) Comparação das orientações dos loops de Rf-CohB1 (magenta), e Rf-CohB4 (ciano) na conformação aberta

Fonte: Elaborada pelo autor.

Essa grande similaridade entre Rf-CohB1 e Rf-CohB4 deve-se ao fato de serem irmãs, presentes na mesma escafoldina, e membros do mesmo grupo 5, na subdivisão do tipo III, ou seja, ambas têm afinidade pela mesma dockerina, Rf-DocA. Devido a essa alta similaridade, as mesmas conclusões observadas para as sobreposições das estruturas com a Rf-CohB4 foram encontradas aqui. Há uma similaridade maior com a cohesina 3 de ScaB de *R. flavefaciens* linhagem FD-1 (PDB ID:5AOZ) (Figura 29C), com z-score 16.1, r.m.s.d. 1.8 Å, 127 resíduos alinhados dos 141 possíveis e 24% em identidade sequencial (Tabela 8). Essa cohesina é classificada em um grupo diferente e tem afinidade por um grupo diferente de dockerinas. Em seguida na classificação de similaridade estrutural (Tabela 8), diversas cohesinas do tipo I e algumas do tipo III seguem, e algumas do tipo II aparecem em menor pontuação, com z-scores 10.6-10.2 e r.m.s.d. 2.5-2.7 Å. No geral, são observadas fitas 5 e 6 mais longas e loops conectando fitas 4-5 e 6-7 mais longos nas estruturas comparadas (Figura 29). Ainda como visto na Rf-CohB4, a ausência β -flaps nas fitas 4 e 8, ou hélice conectando as fitas 6-7 e a disposição da lista de classificação de similaridade estrutural corroboram a

ideia de que as cohesinas de ScaB da linhagem 17 de *R. flavefaciens* se assemelham funcionalmente a cohesinas de tipo-I.

Figura 29 - Sobreposição da Rf-CohB1 com os três tipos de cohesinas. (A) Sobreposição da Rf-CohB1 com a cohesina de tipo I de PDB 4UMS (amarelo). (B) Sobreposição da Rf-CohB1 com a cohesina de tipo II de PDB 2OZN (rosa). (C) Sobreposição da Rf-CohB1 com a cohesina de tipo III de PDB 5AOZ (magenta). Rf-CohB1 está representada em verde (A-C).

Fonte: Elaborada pelo autor.

Tabela 8 - Classificação de similaridade estrutural à Rf-CohB1. O ranking foi editado para retirada de duplicações, quando uma mesma entrada possui duas cadeias com similaridade à proteína consultada. (n) indica a posição do resultado no ranking. *A cohesina Rf-CohB4 foi incluída, com maior similaridade, mas não tem número do ranking pois não está depositada no PDB.

n	Referência (PDB ID)	Tipo	Z-score	rmsd (Å)	Resíduos alinhados	Total resíduos	Identidade sequencial (%)	Organismo
*	Rf-CohB4	III	21,9	1,3	131	134	44	<i>R. flavefaciens</i>
1	5AOZ	III	16,1	1,8	127	141	24	<i>R. flavefaciens</i>
2	4UMS	I	15,3	2,2	128	154	18	<i>B. cellulosolvens</i>
3	4UYP	I	15,1	2,1	128	146	16	<i>A. cellulolyticus</i>
4	3UL4	I	15,1	2,2	126	142	21	<i>C. thermocellum</i>
5	4DH2	I	15,1	2,1	128	156	16	<i>C. thermocellum</i>
8	1G1K	I	15,1	2,2	127	143	20	<i>C. cellulolyticum</i>
11	3KCP	I	14,9	2,0	124	303	19	<i>C. thermocellum</i>
12	1AOH	I	14,9	2,2	124	143	20	<i>C. thermocellum</i>
13	2VN5	I	14,8	2,3	126	142	20	<i>C. cellulolyticum</i>
16	3P0D	I	14,6	2,2	126	309	18	<i>C. thermocellum</i>
17	4FL4	I	14,6	2,2	126	308	18	<i>C. thermocellum</i>
25	5LXV	III	13,8	2,1	127	172	17	<i>R. flavefaciens</i>
27	1OHZ	I	13,5	2,3	118	140	20	<i>C. thermocellum</i>
28	1ANU	I	13,5	2,3	119	138	20	<i>C. thermocellum</i>
29	2ZF9	III	13,3	2,2	128	180	20	<i>R. flavefaciens</i>
33	2CCL	I	12,9	2,3	117	149	21	<i>C. thermocellum</i>
35	4N2O	III	12,8	2,1	127	190	15	<i>R. flavefaciens</i>
36	4WKZ	III	12,8	2,1	127	184	15	<i>R. flavefaciens</i>
48	2OZN	II	11,9	2,6	121	133	9	<i>C. perfringens</i>
50	4IU2	III	11,7	2,3	126	209	23	<i>R. flavefaciens</i>
52	2Y3N	II	10,7	2,5	124	173	15	<i>B. cellulosolvens</i>
54	1TYJ	II	10,6	2,6	126	170	15	<i>B. cellulosolvens</i>
55	2B59	II	10,3	2,5	122	166	13	<i>C. thermocellum</i>
57	5K39	II	10,2	2,5	119	167	18	<i>C. thermocellum</i>
58	3BWZ	II	10,2	2,7	123	171	19	<i>A. cellulolyticus</i>
59	3L8Q	II	10,2	2,5	121	335	13	<i>A. cellulolyticus</i>

Fonte: Elaborada pelo autor.

A principal diferença em relação às diferentes moléculas de Rf-CohB1 presentes em ambos os modelos aqui apresentados é a orientação do loop que conecta as fitas β 6-7, que compreende os resíduos 124-129 (GQGAGR). Ele aparece próximo ao loop que conecta as fitas β 8-9, em conformações diferentes entre as duas moléculas da forma cristalina I. Numa delas, é observada a conformação que denominamos “fechada”, enquanto na outra, aparece 6 Å mais afastado, em uma conformação “aberta” (calculamos as distâncias entre os carbonos α dos resíduos G126 e G169 de cada conformação, considerando que a posição de G169 é a mesma nas duas conformações) (Figura 27B). Na molécula presente na forma cristalina II, esse loop está na conformação “fechada”. A diferença entre as conformações aberta e fechada nas cohesinas Rf-CohB1 e Rf-CohB4 é discutida em mais detalhes abaixo.

É notável na sequência desse loop que há três resíduos de glicina intercalados, o que contribui para sua flexibilidade. Isso é corroborado pelas diferentes conformações observadas nas estruturas. É necessário ressaltar que essas conformações provavelmente não representam a conformação adotada pelo loop no complexo dessa cohesina (Rf-CohB1) com a Rf-DocA, visto que em todos os modelos esse loop está engajado em contatos cristalinos (principalmente através de R129), o que provavelmente impõe alguma restrição conformacional como artefato cristalino. Pode-se inferir, então, que na interação a conformação do loop se ajuste de forma favorável, sendo provavelmente mais semelhante à posição que ocupa o mesmo loop da Rf-CohB4, em complexo com a Rf-DocA.

Esse padrão de conservação de três glicinas intercaladas não é observado para nenhuma outra das cohesinas de ScaB. A posição ocupada pela G124 é conservada nas cohesinas 1-3, enquanto que uma alanina aparece nos outros módulos. G126 de Rf-CohB1 é conservado somente em Rf-CohB7, enquanto que a posição G128 é estritamente conservada em todos os módulos cohesina de ScaB (Figura 24, acima). Isso indica que o loop 6-7 apresenta flexibilidade em todas as cohesinas de ScaB, apesar de ser, provavelmente, em menor grau que o observado em Rf-CohB1. De fato, comparações dessa região entre estruturas da Rf-CohB4 isolada e em complexo com Rf-DocA mostra certa flexibilidade, sendo que a orientação do loop diverge somente a partir de G544 (correspondente a G126 em Rf-CohB1), uma vez que o resíduo A542 (G124 na Rf-CohB1) fixa a conformação do loop nesse ponto (Figura 28, acima).

Dependendo da conformação observada desse loop nas diferentes formas, algumas dessas glicinas apresentam ângulos torcionais incompatíveis com resíduos que tenham cadeia lateral, sugerindo que a introdução de alaninas nessas posições deveria introduzir certa rigidez

ao loop, possivelmente favorecendo uma conformação mais semelhante a alguma das que observamos. Com base nessas observações, realizamos mutações pontuais nessa região, substituindo cada resíduo por alanina (Rf-CohB1_{G124A}, Rf-CohB1_{Q125A}, Rf-CohB1_{G126A}, Rf-CohB1_{G128A} e Rf-CohB1_{R129A}) para averiguar se mutações nessas regiões favorecem ou desfavorecem a interação com Rf-DocA.

4.5 Cristalização e resolução estrutural de mutantes pontuais de Rf-CohB1

Os mutantes Rf-CohB1_{G124A}, Rf-CohB1_{Q125A}, Rf-CohB1_{G126A}, Rf-CohB1_{G128A} e Rf-CohB1_{R129A} foram submetidos a ensaios de cristalização em condições próximas às obtidas previamente para a cristalização da Rf-CohB1_{WT}. O ensaio foi realizado pelo método gota sentada em placa de 96 poços, em tampão 0,1 M BIS-Tris pH 6,5, variando Sulfato de Amônio 2,5 M - 1,4 M, em passos de 0,1 M. As proteínas foram preparadas em diferentes concentrações e misturadas com a solução do poço na razão 1:1. A placa foi incubada a 18 °C e os cristais cresceram a partir de 3 dias.

Entre todos testados, somente os mutantes Rf-CohB1_{G124A}, Rf-CohB1_{G128A} e Rf-CohB1_{R129A} renderam cristais, e desses, somente cristais dos últimos dois mutantes apresentaram qualidade adequada para coletas completas de dados de difração. Cristais do mutante Rf-CohB1_{G128A} cresceram em gotas com 17 mg/mL de proteína, na condição de 0,1 M BIS-Tris pH 6,5, 1,8 M Sulfato de Amônio. Cristais do mutante Rf-CohB1_{R129A} cresceram em gotas com 9 mg/mL de proteína, na condição de 0,1 M BIS-Tris pH 6,5, 2 M Sulfato de Amônio. Os conjuntos de dados de difração de raios X foram coletados a uma resolução máxima de 2,0 Å e 1,8 Å, para Rf-CohB1_{G128A} e Rf-CohB1_{R129A} respectivamente. Os dados foram processados em grupos espaciais distintos, P2₁ e P2₁2₁2₁, contendo 4 e 2 moléculas na unidade assimétrica, Rf-CohB1_{G128A} e Rf-CohB1_{R129A}, respectivamente. As estruturas foram resolvidas por substituição molecular utilizando o modelo construído previamente para Rf-CohB1_{WT} como modelo de busca. O resumo dos dados cristalográficos da coleta e refinamento pode ser encontrado na Tabela 9.

Ambas as estruturas são muito similares a Rf-CohB1_{WT}, como esperado, sendo a única diferença entre elas a região do loop que conecta as fitas β 6-7. Na estrutura de Rf-CohB1_{R129A}, o loop se encontra numa conformação bastante diferente de Rf-CohB1, sendo muito mais semelhante à orientação observada no loop 6-7 da Rf-CohB4 em complexo com Rf-DocA (Figura 30). A comparação entre as estruturas da Rf-CohB1_{WT} e a estrutura do

mutante Rf-CohB1_{R129A} mostra que a maior diferença nos ângulos adotados pelos resíduos no loop se dá na posição G124. Foi observado nas estruturas de Rf-CohB1_{WT} que R129 participa intensamente de contatos cristalinos, o que impõe uma orientação ao loop 6-7. Devido a esse contato, o loop fica em uma orientação elevada em relação ao plano formado pela folha β 8, 3, 6 e 5, de forma que G124, bastante flexível, seja torcida para acomodar essa configuração (Figura 31). A substituição de R129 por uma alanina, um resíduo com cadeia lateral curta e apolar, ocasiona que a influência do contato cristalino sobre esse resíduo seja menor, introduzindo certa flexibilidade no loop. Assim, com uma maior liberdade, esse loop adota uma posição mais favorável a G124, a qual é, provavelmente, também favorável à interação com Rf-DocA.

Tabela 9 - Parâmetros estatísticos da coleta, processamento e refinamento dos dados cristalográficos dos mutantes Rf-CohB1_{G128A} e Rf-CohB1_{R129A}. Os dados em parênteses correspondem faixa de maior resolução.

	Rf-CohB-1 _{G128A}	Rf-CohB-1 _{R129A}
Dados da Coleta		
Comprimento de onda (Å)	1,4590	1,4590
Temperatura (K)	100	100
Grupo Espacial	P2 ₁	P2 ₁ 2 ₁ 2
Célula Unitária (Å)	58,80;99,30;59,10	59,20;59,30;100,50
Resolução (Å)	43,8 - 2,0 (2,1 - 2,0)	42,2 - 1,8 (1,9 - 1,8)
Reflexões Únicas	43310 (6924)	35067(5684)
Reflexões Totais	188191 (32315)	272461(37322)
Redundância	2,7 (4,7)	7,8 (5,6)
Completeza (%)	94,8 (93,9)	96,1 (98,1)
I/ σ (I)	20,2 (6,0)	23,6 (2,7)
R-meas (%)	6,0 (29,0)	5,5 (58,0)
Refinamento da Estrutura		
R-work (%)	25,1	19,9
R-free (%)	30,8	22,9
Número de Átomos		
Água	382	360
Proteínas	3856	1944
B-fator Médio		
Água	33,2	41,2
Proteínas	30,6	29,4
RMSD		
Ligações (Å)	0,008	0,003
Ângulos (°)	0,949	0,664
Gráfico de Ramachandran (%)		
Favoráveis	97,9	99,6
Permitidas	1,9	0,4
Não Permitidas	0,2	0,0

Fonte: Elaborada pelo autor.

Figura 30 - Estrutura da cohesina Rf-CohB1_{R129A} comparada com a conformação aberta das cohesinas Rf-CohB1_{WT} e Rf-CohB4. A conformação do loop 6-7 em Rf-CohB1_{R129A} é mais similar à observada em Rf-CohB4. Em magenta, Rf-CohB1_{WT}, em ciano, Rf-CohB4, em verde e amarelo, as duas moléculas de Rf-CohB1_{R129A} presentes no modelo.

Fonte: Elaborada pelo autor.

Figura 31 - Sobreposição do loop 6-7 na comparação entre Rf-CohB1_{R129A}, Rf-CohB1_{WT} e Rf-CohB4. Observa-se uma grande divergência no ângulo adotado pelo resíduo G124 (que equivale a A542 em Rf-CohB4), ocasionado pelo contato cristalino nessa região. O resíduo R129 é um dos principais participantes desse contato, tal que sua mutação para alanina reduz o efeito desse artefato e a G124 pode adotar uma conformação mais favorável, semelhante à que ocupa A542 em Rf-CohB4. Em amarelo, Rf-CohB1_{R129A}, em magenta, Rf-CohB1_{WT} e em ciano, Rf-CohB4.

Fonte: Elaborada pelo autor.

Na estrutura do mutante Rf-CohB1_{G128A}, que possui a R129, além da mutação G128A, observamos a conformação de G124 semelhante à observada em Rf-CohB1_{WT}, ou seja, o resíduo G124 também é orientado para cima do plano da folha β . No entanto, a região central do loop tem sua flexibilidade afetada pela mutação G128A, de forma que o loop ocupa uma posição intermediária às duas conformações observadas para o loop nas diferentes estruturas

de de Rf-CohB1_{WT} (Figura 32). Isso sugere que a flexibilidade do loop seja afetada pela mutação da glicina 128, favorecendo a orientação do loop na conformação “fechada”, embora a estrutura observada no cristal não deva ser considerada definitiva para essa conclusão devido ao artefato introduzido pela R129 na orientação do loop. Assim, para a visualização confiável da conformação do loop ocasionada pela mutação R128A, seria necessário produzir um duplo mutante, contendo as mutações G128A e R129A e cristalizá-lo. No entanto, a conservação dessa glicina em todas as 7 cohesinas de ScaB, leva a crer que a presença desse resíduo, o qual faz parte de um loop intrinsecamente flexível, é essencial para a interação.

Figura 32 - Estrutura da Rf-CohB1_{G128A} comparada à Rf-CohB1_{WT} nas duas conformações que observamos e à Rf-CohB4 (A) Sobreposição das quatro cadeias de Rf-CohB1_{G128A} presentes no modelo estrutural que resolvemos (em ciano, vermelho, verde e laranja) às duas conformações observadas para a cohesina Rf-CohB1_{WT} (em amarelo, fechada, em magenta, aberta) (B) Conformação do loop 6-7 da Rf-CohB1_{G128A} (em amarelo), comparada às duas conformações observadas para Rf-CohB1_{WT}, e à conformação fechada de Rf-CohB4, observada na estrutura em que foi cristalizada isolada. A orientação da G124 na Rf-CohB1_{G128A} é mais semelhante às observadas nas duas formas da Rf-CohB1_{WT}.

Fonte: Elaborada pelo autor.

Como conclusão, identificamos que a interação de Rf-DocA com Rf-cohB1 se dá provavelmente em conformações do loop semelhantes ao observado para as estruturas de Rf-CohB1_{R129A} e Rf-CohB4. A mutação G128A provavelmente afeta negativamente a flexibilidade de Rf-CohB1. Para a confirmação do impacto que a mutação da glicina 128 e os demais resíduos do loop têm sobre sua flexibilidade, conduzimos experimentos de MST para medir a finidade entre Rf-DocA e os mutantes Rf-CohB1_{G124A}, Rf-CohB1_{Q125A}, Rf-CohB1_{G126A}, Rf-CohB1_{G128A} e Rf-CohB1_{R129A}.

4.6 Análise da influência do loop 6-7 na afinidade da Rf-CohB1 à Rf-DocA por termoforese em microescala

Para analisar o efeito da flexibilidade do loop que conecta as fitas β 6-7, na Rf-CohB1, sobre a afinidade de ligação à Rf-DocA, foi realizado um *Alanine scan*, sobre os resíduos que compõem esse loop. A Rf-CohB1_{WT} e mutantes foram testados, por ensaios de termoforese em microescala, contra a Rf-DocA. Era esperado que a mutação de resíduos de glicina 124, 126 e 128 reduzissem a flexibilidade do loop, favorecendo uma das conformações (aberto/fechado). Já para os mutantes Q125A e R129A, presumimos que poderia haver um aumento de flexibilidade, devido à troca da cadeia lateral pela mais curta e apolar da alanina.

O mutante Rf-CohB1_{G124A} apresentou uma afinidade (K_d de aproximadamente 5nM) aumentada à Rf-DocA, comparada à apresentada por Rf-CohB1_{WT} (K_d de aproximadamente 87 nM). Esse resultado indica que a flexibilidade foi afetada e a mutação induz a orientação do loop para uma conformação aberta, mais favorável à interação com a Rf-DocA. A introdução de cadeia lateral na posição 124 da Rf-CohB1 provavelmente impede que a cadeia se eleve em relação ao plano da folha β a partir dessa posição. As cohesinas 4-7 de ScaB tem alanina naturalmente nessa posição, o que poderia indicar maior afinidade da Rf-DocA para essas cohesinas (Figura 24, acima). A constante de dissociação de cada variante da Rf-CohB1 é apresentada na Tabela 10 e o conjunto das curvas de ligação, na Figura 33.

Já o mutante Rf-CohB1_{G128A} apresentou uma diminuição de cerca de 25 vezes na afinidade (K_d de aproximadamente 2,2 μ M), quando comparada à de Rf-CohB1_{WT}. Isso indica um efeito significativo sobre a flexibilidade, porém favorecendo a orientação fechada do loop. Observarmos pelo alinhamento das cohesinas de ScaB (Figura 24, acima), que esse

resíduo é conservado em todas as cohesinas de ScaB, sendo essencial para a interação com Rf-DocA.

Os mutantes Rf-CohB1_{R129A}, Rf-CohB1_{G126A} e Rf-CohB1_{Q125A} (Kds de 48 nM, 90 nM e 130 nM, respectivamente), apresentaram afinidades próximas à Rf-CohB1_{WT} (Kds de 90 nM). Esses resultados apontam um baixo efeito dessas mutações sobre a flexibilidade do loop. O resíduo Q125 é conservado nas cohesinas 3-6, G126 é conservado nas cohesinas 4-7, enquanto R129 só é conservado em Rf-CohB3. Em conjunto, essas observações indicam que esses resíduos não são cruciais para a interação com Rf-DocA, tolerando diferentes resíduos nessas posições.

Esses resultados indicam que a flexibilidade do loop tem influência direta na afinidade que a Rf-CohB1 tem à Rf-DocA, sendo um fator que regula o reconhecimento entre os dois módulos. Uma dinâmica semelhante havia sido proposta para a cohesina da escafoldina de ancoragem ScaE.¹² Foram observadas orientações diferentes para o β -flap da fita 8, comparando a estrutura da cohesina isolada (PDBID:2ZF9) com a de seu complexo com a dockerina de CttA (PDBID:4IU2). No entanto, estas são as primeiras evidências experimentais de que mudanças conformacionais na cohesina influenciam na afinidade à dockerina.

O modelo inicial para interação entre dockerina e cohesina dizia que a dockerina seria um módulo desestruturado em solução, e que sua conformação era ajustada à cohesina no momento da interação. Posteriormente, foi observado que a dockerina é estruturada independentemente da cohesina². Nossos resultados sugerem que a cohesina pode apresentar alterações conformacionais que confirmam adaptabilidade à ligação com a dockerina e que esse é um fator que influencia a afinidade entre esses módulos.

Tabela 10 - Constantes de dissociação entre Rf-DocA e variantes da Rf-CohB1 obtidas através de ensaios de termoforese em microescala.

Módulo Cohesina	Kd (nM)
Rf-CohB1 _{G124A}	5.1 $\pm$ 0.6
Rf-CohB1 _{R129A}	48 $\pm$ 7
Rf-CohB1 _{WT}	90 $\pm$ 20
Rf-CohB1 _{G126A}	90 $\pm$ 10
Rf-CohB1 _{Q125A}	130 $\pm$ 20
Rf-CohB1 _{G128A}	2200 $\pm$ 400

Fonte: Elaborada pelo autor.

Figura 33 - Curvas de ligação entre Rf-DocA e variantes da Rf-CohB1, em ensaios de termoforese em microescala. As escalas de todas as curvas foram normalizadas de acordo com as frações ligada e não ligada em cada experimento.

Fonte: Elaborada pelo autor.

De certa forma, a afinidade medida por MST entre Rf-CohB1 e Rf-DocA não condiz com o esperado através das análises baseadas nas estruturas e identidade sequencial dos módulos cohesina. Esperava-se que o K_d medido fosse comparável ao da interação com Rf-CohB4 (em torno de 1 nM), e não mais próximo ao medido para a interação com Rf-CohB6 (em torno de 50 nM), os quais foram medidos por ITC. No entanto, em ambos os casos os valores medidos sofrem influência das limitações das técnicas, sendo difícil uma comparação direta de afinidades aferidas por diferentes técnicas (ITC e MST). Enquanto que no experimento de MST, uma das proteínas estava marcada com um fluoróforo, que pode exercer influência sobre a interação entre as proteínas, no experimento de ITC o ajuste do modelo nos pontos experimentais é afetado pela alta afinidade da interação (próxima ao limite de detecção da técnica). Possivelmente, a marcação da Rf-DocA para os experimentos de MST impacta negativamente a afinidade com Rf-CohB1, o que não invalida a conclusão a respeito da flexibilidade do loop entre as fitas β 6 e 7, uma vez que todos experimentos de MST foram conduzidos com o mesmo lote de Rf-DocA marcada.

4.7 Clonagem por LIC

Buscamos utilizar um método de clonagem que facilitasse o escalonamento do processo, tendo em vista o grande número de clonagens que planejamos fazer. A cada proteína de simetria, haveria 7 módulos do par dockerina/cohesina a se fundir, utilizando 3 linkers diferentes, totalizando 21 possíveis construções para cada proteína (C2, C3, C4, C6). Tendo 2 proteínas de cada simetria, seria possível construir 168 diferentes fusões, que poderiam ser combinadas de 756 maneiras para a formação das redes automontáveis. Restringindo que a cada combinação de rede seria testado apenas o mesmo *linker* para ambas as fusões (entre cohesinas e proteínas de simetria C2 e entre dockerinas e proteínas de simetrias C3, C4 e C6), teríamos 252 maneiras distintas de combinar esses componentes.

Para a clonagem através da técnica LIC, os *primers* que constituem o *linker* entre os fragmentos clonados foram fosforilados, por ação da enzima T4 PNK, e usados na reação de amplificação de cada inserto, por PCR a partir do DNA sintético (GenScript) (Figura 34). Em seguida, esses fragmentos, purificados, foram tratados, separadamente, com T4 DNA Polimerase na presença de apenas um dos quatro desoxiribonucleotídeos para formar as extremidades de fita simples de DNA, cuja complementaridade permite a hibridização entre ambos. Esse DNA unificado foi então ligado covalentemente *in vitro*, pela T4 DNA Ligase, e amplificado (Figura 34), sendo então tratado com T4 DNA Polimerase na presença de dATP, para posterior hibridização com o vetor amplificado e tratado com T4 DNA Polimerase na presença de dTTP.

Nossa estratégia de introduzir uma ligação *in vitro* entre os insertos de clonagem constitui uma adaptação à técnica, para a ligação de três fragmentos de DNA. Ela se mostrou plausível, uma vez que conseguimos amplificar o inserto unificado, purificá-lo a partir de gel de agarose e então tratá-lo com a T4 DNA Polimerase. No entanto, após inserção desse produto no vetor, foi observado, por sequenciamento, que a etapa de ligação *in vitro* propiciava a ligação de fragmentos que não tinham extremidades coesivas, ou seja, eram ligados fragmentos de DNA com extremidades cegas, ocasionando a duplicação da sequência desenhada para complementaridade para LIC (Figura 34).

Buscamos otimizar as reações de modo a obter maior eficiência na excisão de nucleotídeos por ação da T4 DNA Polimerase, aumentando a quantidade de enzima adicionada à reação, bem como o aumento do tempo de incubação. Procuramos também otimizar parâmetros que favorecessem a ligação de fragmentos complementares ao invés da ligação entre extremidades cegas, como diminuição da concentração da enzima T4 DNA

Ligase, diminuição do tempo de incubação e aumento da temperatura de incubação. No entanto, essas tentativas não foram capazes de contornar o problema.

Figura 34 - Clonagem através da técnica LIC, mostrando como exemplo a clonagem entre os fragmentos 3C19 e a Dockerina tipo II (PDB ID:2Y3N). (a) Gel de agarose mostrando a amplificação dos fragmentos 3C19 (1) e Dock2 (2) e amplificação após sua fusão (3). (M) Marcador de massa molecular. (b) Resultado do sequenciamento mostrando em pontilhado a duplicação na região em que deveria haver ocorrido o anelamento devido a complementaridade (LIC). O tracejado destaca o ponto onde houve a ligação entre os fragmentos. A sequência esperada para a clonagem correta dos fragmentos está indicada acima do resultado obtido. As sequências terminais dos dois fragmentos (3C19 e Dock2) foram incluídas como referência, indicados por seu nome na figura.

Fonte: Elaborada pelo autor.

4.8 Clonagem por SLIC

Devido ao insucesso da técnica LIC em nossas clonagens, buscamos como alternativa a técnica SLIC, mais simples que a primeira, uma vez que não exige restrições no desenho das sequências. O tratamento dos fragmentos é regulado por tempo de reação, com uma quantidade menor de enzima, gerando extremidades compatíveis em todos os fragmentos envolvidos, a partir de um único tratamento com a atividade exonuclease da enzima T4 DNA Polimerase. Dessa forma, após a amplificação dos fragmentos de interesse (Figura 35A), o DNA parental foi degradado com DpnI, e então foi feito o tratamento com a T4 DNA Polimerase. Em seguida, os DNAs foram misturados para hibridização e usados para

transformação de células de DH5 α , que reparam as fitas desconexas de DNA. As colônias crescidas em placa de Petri foram testadas por PCR para amplificação do inserto de interesse, utilizando vetores que hibridizam nas regiões flanqueadoras da clonagem, nos sítios do promotor e terminador T7 (Figura 35B) (é necessário ressaltar que com a utilização desses *primers*, há o acréscimo de cerca de 600 pb nos amplicons do vetor utilizado). Os plasmídeos contendo insertos com tamanho correspondente ao esperado foram confirmados por sequenciamento de DNA e então utilizados para transformar células de BL21(DE3), para expressão da proteína recombinante.

A Figura 35, abaixo, mostra as etapas desse processo, envolvendo as amplificações dos DNAs que codificam para as proteínas 3C19, 2Y3M, 3BUT e 2Y3N. Esses elementos estão nomeados de acordo com seu PDB ID, exceto 2Y3M, que é a dockerina presente na estrutura do complexo 2Y3N (essa nomenclatura foi adotada em nossos experimentos para a distinção entre os dois módulos). As proteínas 3C19 e 3BUT, possuem, respectivamente, simetrias rotacionais 3 e 2 e, portanto, são clonadas com dockerina e cohesina. Detalhes de todas os elementos presentes no desenho experimental estão presentes no Apêndice A. Os *primers* utilizados para amplificar cada um desses DNAs contém as sequências dos linkers que unem cada par desses módulos, sendo todos os exibidos na Figura 35 amplificados com o *linker* 1 (L1 - SGEEARRGS). Assim, cada módulo é denominado por pelos dois componentes que o formam, seguido do *linker* que os une, por exemplo 3C19-2Y3M-L1 é formado pela fusão da proteína trímética 3C19, a dockerina 2Y3M, através do *linker* 1. O mesmo procedimento foi adotado para todas as clonagens, sendo que obtivemos as construções dos módulos das redes: 3C19-2Y3M-L1, 2Y3N-3BUT-L1, 2Y3N-1UDV-L1, 1BKB-2Y3M-L2, 1LJO-2Y3M-L2, 2Y3N-1UDV-L2 (Figura 36). A Tabela 11 mostra detalhes desses módulos clonados pela técnica SLIC.

Figura 35 - Clonagem por SLIC. (A) Amplificação dos DNAs que codificam os componentes das fusões que originam os módulos das redes proteicas bidimensionais. (M) Marcador de massa molecular. (1) 3C19 (2) 2Y3M (3) 3BUT (4) 2Y3N. Todos esses fragmentos de DNA foram amplificados com *primers* que codificam o *linker* 1 (L1). São visíveis bandas espúrias nesse gel, motivo pelo qual o DNA é recortado do gel e tratado com DpnI para degradação de DNA parental. (B) PCR para verificação de transformantes positivos para a clonagem de 3C19-2Y3M-L1. (1-3) Diferentes colônias que apresentaram contaminação na reação de amplificação, sendo descartados. (4-5) Transformantes positivos para a clonagem de 3C19-2Y3M-L1

Fonte: Elaborada pelo autor.

Figura 36 - Confirmação de transformantes por PCR. A clonagem dos módulos 2Y3N-1UDV-L2 (1), 2Y3N-1UDV-L1 (2), 1BKB-2Y3M-L2 (3), 1LJO-2Y3M-L2 (4), 2Y3N-3BUT-L1 (5) e 3C19-2Y3M-L1 (6).

Fonte: Elaborada pelo autor.

Tabela 11- Detalhes das construções dos componentes das redes automontáveis de proteínas.

Fusão	Tamanho do amplicon (pb)	Tamanho do monômero (kDa)	Tamanho do monômero fusionado a SUMO (kDa)
3C19-2Y3M-L1	747	27.2	39.2
2Y3N-3BUT-L1	924	33.9	45.9
2Y3N-1UDV-L1	810	30	42
1BKB-2Y3M-L2	639	23.2	35.2
1LJO-2Y3M-L2	456	16.4	28.4
2Y3N-1UDV-L2	825	30.5	42.5

Fonte: Elaborada pelo autor.

4.9 Expressão e purificação dos módulos componentes das redes automontáveis

A expressão dos módulos componentes das redes automontáveis de proteínas foi feita de modo semelhante ao descrito para as cohesinas e dockerinas de *R. flavefaciens*. Assim, após a indução da expressão de 3C19-2Y3N-L1 *overnight*, a proteína foi purificada por IMAC em TALON (Clontech) (Figura 37A), sendo eluída com concentrações crescentes de imidazol adicionado no tampão 20 mM Tris pH 8,0, 300 mM NaCl. Posteriormente, foi tratada com a protease Ulp1 para clivagem de SUMO, que é em seguida removida com uma segunda IMAC em TALON (Clontech) (Figura 37B). Finalmente, essa proteína foi purificada por cromatografia de exclusão molecular, através da coluna Superdex 75 10/300 GL (GE Healthcare) (Figura 37B). A pureza das amostras é testada por SDS-PAGE em cada etapa de purificação.

A expressão dos elementos clonados seguiu esse mesmo protocolo, gerando proteínas solúveis e homogêneas para todas as construções que continham a cohesina 2Y3N (Tabela 11). No entanto, observamos que as construções contendo a dockerina 2Y3M eram instáveis em solução, formando agregados na última etapa de purificação. A única construção estável que obtivemos para esse módulo foi a 3C19-2Y3M-L1, utilizada conjuntamente com a 2Y3N-3BUT-L1 para os primeiros testes de formação de redes protéicas automontáveis. A purificação desta segunda construção está mostrada na Figura 38.

Figura 37 - Purificação da fusão proteica 3C19-2Y3N-L1. (A) Primeira purificação em TALON (Clontech) (M) Marcador de massa molecular. (1) Fração insolúvel (2) Fração solúvel (3) Eluato não ligado à coluna (4) Lavagem com tampão sem imidazol (5) Lavagem com 20 mM de imidazol (6) Lavagem com 50 mM de Imidazol (7) Lavagem com 100 mM de Imidazol (8) Lavagem com 300 mM de Imidazol (9) Lavagem com 500 mM de Imidazol (B) Segunda purificação em TALON (Clontech), para retirada de SUMO e Cromatografia de exclusão molecular (SEC) (M) Marcador de massa molecular. (1) Eluato não ligado à coluna (2) Lavagem com tampão sem imidazol (3-14) Frações coletadas da SEC, correspondendo à proteína de interesse (4-9) e SUMO (12-14).

Fonte: Elaborada pelo autor.

Figura 38 - Purificação da fusão proteica 2Y3N-3BUT-L1. (A) Primeira purificação em TALON (Clontech) (M) Marcador de massa molecular. (1) Fração insolúvel (2) Fração solúvel (3) Eluato não ligado à coluna (4) Lavagem com tampão sem imidazol (5) Lavagem com 20 mM de imidazol (6) Lavagem com 50 mM de Imidazol (7) Lavagem com 100 mM de Imidazol (8) Lavagem com 300 mM de Imidazol (9) Lavagem com 500 mM de Imidazol (B) Segunda purificação em TALON (Clontech), para retirada de SUMO e Cromatografia de exclusão molecular (SEC) (M) Marcador de massa molecular. (1) Eluato não ligado à coluna (2) Lavagem com tampão sem imidazol (3) Segunda lavagem com tampão sem Imidazol (4) Lavagem com 500 mM Imidazol (5) Segunda lavagem com 500 mM Imidazol (6-16) Frações coletadas da SEC, correspondendo à proteína de interesse (6-12) e SUMO (14-16).

Fonte: Elaborada pelo autor.

4.10 Análise de formação das redes proteicas automontáveis por microscopia eletrônica de transmissão

Realizamos testes preliminares visando otimizar condições da obtenção de redes auto montáveis de proteínas, utilizando as proteínas 3C19-2Y3M-L1 e 2Y3N-3BUT-L1, fusionadas por *linker* rígido de 5 resíduos. O primeiro teste comparou a formação da rede dessas proteínas (concentração final de 5 μ M) quando incubadas em solução ou diretamente aplicadas no *grid*. Em ambas as condições, foram testados 30 minutos ou 4 horas de

incubação. A incubação direta nos *grids* ocasionou a evaporação de parte da água da solução em 30 minutos, e na completa secagem da amostra em 4 horas. No primeiro caso, foi observada a formação de nanotubos (Figura 39A) e de filmes bidimensionais (Figura 39B). Já com 4 horas de incubação, apenas filmes bidimensionais com espessuras diferentes foram observados, provavelmente devido a secagem da amostra, que acarretou na sobreposição de camadas como artefato (Figura 39C). Na incubação em solução, observamos apenas a formação de nanotubos em 30 minutos (Figura 39D) e formação tanto de nanotubos (Figura 39E) quanto filmes bidimensionais (Figura 39F), em 4 horas.

Figura 39 - Micrografias eletrônicas dos primeiros testes de formação de redes bidimensionais automontáveis, a partir das proteínas 3C19-2Y3M-L1 e 2Y3N-3BUT-L1, fusionadas por *linker* rígido de 5 resíduos. (A) Nanotubos formados com incubação de 30 minutos no grid. (B) Filmes bidimensionais formados com incubação de 30 minutos no grid. (C) Filmes bidimensionais sobrepostos formados com incubação de 4 horas diretamente no grid. (D) Nanotubos formados com incubação de 30 minutos em solução. (E) Nanotubos formados por incubação de 4 horas em solução. (F) Filme bidimensional formado com incubação de 4 horas em solução.

Fonte: Elaborada pelo autor.

Num segundo teste, buscamos a diminuição da concentração de proteínas, na tentativa de observar camadas únicas de filme, que seriam utilizadas para difração de elétrons, com a qual poderíamos verificar a simetria da rede. A concentração de proteínas foi reduzida para

2,5 μM , e a mistura foi incubada em solução por 4 horas. Como resultado, a concentração utilizada foi insuficiente para a formação significativa de filmes. Observamos a formação de pequenos pontos discretos, o que pode corresponder à fase de nucleação da formação de filmes bidimensionais (Figura 40 A-B).

Figura 40 - Micrografias eletrônicas do segundo teste de formação de redes bidimensionais automontáveis, com concentração menor das proteínas 3C19-2Y3M e 2Y3N-3BUT. A diminuição da concentração de proteínas inviabilizou a formação de filmes bidimensionais, sendo visível a nucleação destes. (A) Pontos de nucleação de filmes bidimensionais. (B) Pontos de nucleação de filmes bidimensionais em maior aumento.

Fonte: Elaborada pelo autor.

O terceiro teste executado foi utilizando grids carregados positivamente e negativamente. Utilizamos 5 μM de cada proteína, incubando a mistura em solução por 4 horas. O grid carregado positivamente ocasionou dificuldade na formação de filmes, sendo observados apenas em pequenas áreas próximas da grade de cobre que forma os quadrantes do grid (Figura 41A). Já no grid carregado negativamente, observamos a formação de filme por todo o grid, sendo visível, em alguns pontos, a formação de múltiplas camadas (Figura 41 B-C).

Nossa estratégia de formação de redes auto montáveis de proteínas teve êxito, uma vez que observamos a formação de filmes bidimensionais e nanotubos, a partir da mistura dos componentes, como planejado. A ocorrência de nanotubos era esperada, tendo sido reportada em outros trabalhos^{28,49}. No estudo publicado por Brodin e colaboradores,²⁸ o desenho de redes automontáveis foi feito a partir do desenho de sítios de coordenação a zinco estrategicamente posicionados na superfície da proteína, de forma a gerar dímeros não naturais que se intercalavam, dando origem tanto a filmes bidimensionais como nanotubos. O fator determinante para formação de cada tipo de estrutura foi a velocidade de nucleação,

controlada nas condições do experimento. Assim, quando as proteínas eram colocadas em pH alcalino (8,5), ou em concentrações elevadas de zinco (100:1 Zinco:Proteína, em razão molar), a nucleação ocorria de forma mais acelerada, e dava origem a pequenos fragmentos bidimensionais que, ao se unir, se enovelavam em hélice, gerando nanotubos. As condições opostas, de pH mais ácido (5,5) e concentração de zinco mais próxima da concentração da proteína (3:1 Zinco:Proteína, em razão molar), ocasionavam nucleação mais lenta, em que os fragmentos bidimensionais gerados cresciam mais lentamente, criando filmes mais extensos. Esses pesquisadores conseguiram inclusive, o empilhamento ordenado desses filmes, gerando matrizes tridimensionais nessas condições. Ainda, verificaram que era possível gerar matrizes 2D e 3D a partir de nanotubos, variando o pH ou concentração de zinco, mas não era possível realizar o caminho inverso. Essa capacidade de modulação das estruturas formadas é de grande interesse nesses materiais, sobretudo no contexto da geração de materiais responsivos.²²

Figura 41 - Micrografias eletrônicas do terceiro teste de formação de redes bidimensionais automontáveis, com grids carregados positivamente e negativamente. Nos grids positivamente carregados (A), a formação de filmes foi dificultada, sendo observados apenas em pequenas extensões, próximas das grades de cobre que compõem o *grid*. Nos *grids* carregados negativamente (B) e (C), os filmes bidimensionais foram observados em toda sua extensão, sendo presentes em múltiplas camadas em alguns pontos (C). Em (B) é observado um filme monocamada, próxima a uma região da malha formada pelo carbono espesso do grid, ou seja, não foram observados nanotubos nesse caso.

Fonte: Elaborada pelo autor.

Em nossos experimentos, embora preliminares, observamos a formação de nanotubos e filmes bidimensionais, tanto concomitantemente quanto isoladamente. Os nanotubos foram observados em diferentes extensões, com espessuras variáveis e apresentando ramificações (Figura 39 D - E, acima). Nos filmes bidimensionais, observamos que eram bastante extensos, recobrando regiões amplas dos grids, e por vezes, formavam camadas sobrepostas. Planejávamos otimizar a condição de obtenção dessas estruturas, bem como o controle desses diferentes aspectos estruturais, de modo semelhante às variações de condições experimentais

citadas acima. A variação da concentração de cálcio adicionado à mistura de proteínas é um fator que poderia influenciar a velocidade de formação das redes bidimensionais, uma vez que a interação entre dockerina e cohesina é dependente de cálcio. Assim, desejávamos remover o cálcio presente nas proteínas que continham dockerina em sua fusão e posteriormente adicioná-lo à mistura de proteínas, verificando tanto se há a possibilidade de estímulo externo para formação das redes, quanto se a concentração de cálcio afeta a velocidade de formação dessas redes.

A interdependência dos componentes para a formação das redes torna provável que a relação entre suas concentrações influencie na velocidade em que se unem para formar as redes. Assim, outro teste que pretendíamos fazer era o de variar as concentrações de um ou ambos os componentes, de modo a obter diferentes razões entre essas variáveis, e verificar se haveria favorecimento da formação de filmes ou nanotubos em cada condição. Ainda, a composição do tampão, com variação de pH ou concentração de sais poderia exercer influência sobre a interação das proteínas, em especial a sobreposição de camadas observada para os filmes bidimensionais. Com esses testes, buscávamos otimizar o preparo das redes automontáveis, chegando idealmente em altas resoluções, como observado no trabalho de Sinclair e colaboradores⁵⁰. Ao fusionar domínios de uma proteína tetramérica ao peptídeo StreptagI, e utilizando streptavidina como conector entre esses tetrâmeros, esses pesquisadores construíram redes bidimensionais que visualizaram com alta resolução, sendo possível observar a simetria do desenho das redes (Figura 42).

Figura 42 - Rede bidimensional automontável visualizada em alta resolução. (A) Componentes da rede, em verde, a proteína tetramérica ALAD, de *E.coli*, em amarelo, a streptavidina em complexo com streptagI. (B) Representação do planejamento das redes formadas pela mistura da proteína tetramérica ALAD fusionada à StreptagI, com Streptavidina dimérica. (C) Micrografia de alta resolução mostrando o padrão simétrico formado na rede automontada dessas proteínas.

Fonte: Adaptada de SINCLAIR.⁵⁰

5 Conclusão

Em nosso trabalho, nos propusemos a elucidar a interação entre a escafoldina primária ScaA e as cohesinas da escafoldina ScaB, de *Ruminococcus flavefaciens*. Essa interação é dada pelo contato dos módulos dockerina, da ScaA com as cohesinas de ScaB, e é o componente central na organização do celulosomo desse organismo. Através da resolução estrutural do complexo entre esses módulos, que resultou na primeira estrutura de Rf-DocA reportada, pudemos compreender a estrutura de cada módulo, e quais os pontos de contato entre eles. Esclarecemos que o contato da dockerina com a cohesina é dado predominantemente por uma das hélices, tendo um interior hidrofóbico, e contatos periféricos polares, com ligações de hidrogênio e pontes salinas. Confirmamos uma questão pendente na literatura sobre a ligação de cálcio no módulo dockerina, concluindo que de fato, seus dois sítios de ligação a cálcio são funcionais, apesar de uma inserção atípica no segundo sítio.

Análises comparativas de todas as estruturas de Rf-CohB1 e Rf-CohB4 determinadas nesse trabalho, permitiram a identificação de um loop altamente flexível entre as fitas β 6 e 7, exclusivo das cohesinas desse grupo e essencial para interação com a dockerina parceira Rf-DocA. Estudos de interação de mutantes pontuais (*Alanine scanning*) na região G124-R129 confirmam que a glicina altamente conservada na posição 128 (Rf-CohB1) é essencial para a ligação e a presença de um resíduo com cadeia lateral na posição 124 (mutante G124A) aumenta a afinidade por Rf-DocA, provavelmente restringindo a flexibilidade desse loop em uma conformação mais favorável para interação. Esse trabalho mostra pela primeira vez que ocorre uma adaptação conformacional (*induced fitting*) de cohesinas no evento de interação com a dockerina parceira.

Através de experimentos de ITC baseados nas análises estruturais, verificamos a interação de mutantes de Rf-DocA desenhados para favorecer apenas um dos possíveis modos de interação com os módulos cohesina de ScaB, confirmando que a interação de todos os pares DocA-CohB na arquitetura do celulosomo de *R. flavefaciens* é feita via um único modo de ligação. Adicionalmente, o mapeamento dos resíduos essenciais para formação do complexo Rf-DocA/Rf-CohB4 e comparação sequencial com os outros módulos cohesina de ScaB permitiu agrupar os módulos cohesina de acordo com a afinidade pela Rf-DocA, onde os módulos 5-6 apresentariam uma menor afinidade em relação aos módulos 1-4 e 7. Experimentos de ITC confirmaram essa hipótese. A arquitetura e dinâmica de montagem dos componentes de celulosomos são essenciais para seu funcionamento e eficiência. Nesse trabalho demonstramos que a escafoldina primária do celulosomo de *R. flavefaciens*, (ScaA),

responsável pelo recrutamento das enzimas no complexo, liga-se aos módulos coesina de ScaB via apenas um modo e com afinidades distintas.

Buscando utilizar os módulos cohesina e dockerina em uma aplicação tecnológica inovadora, planejamos a confecção de redes proteicas automontáveis em que esses módulos foram empregados devido a sua alta afinidade, unindo proteínas de diferentes simetrias rotacionais, permitindo o livre crescimento da estrutura. Conseguimos obter, em resultados preliminares, nanotubos e filmes bidimensionais, que indicam o êxito do nosso planejamento racional de formação de redes automontáveis, embora não tenha sido possível determinar a simetria apresentada pela rede que os forma, como era pretendido. Não obstante, os resultados preliminares serão considerados no refinamento de condições, visando alcançar a visualização de um padrão na formação de redes automontáveis de proteínas, e futuramente, redes tridimensionais.

Referências

- 1 ISRAELI-RUIMY, V. et al. Complexity of the ruminococcus flavefaciens FD-1 cellulosome reflects an expansion of family-related protein-protein interactions. **Scientific Reports**, v. 7, n. 2017, p. 1–15, 2017.
- 2 ARTZI, L.; BAYER, E. A.; MORAÏS, S. Cellulosomes: bacterial nanomachines for dismantling plant polysaccharides. **Nature Reviews Microbiology**, v. 15, p. 83-95, 2016. doi:10.1038/nrmicro.2016.164.
- 3 SINCLAIR, J. C. Constructing arrays of proteins. **Current Opinion in Chemical Biology**, v. 17, n. 6, p. 946–951, 2013.
- 4 SMITH, S. P.; BAYER, E. A. Insights into cellulosome assembly and dynamics: from dissection to reconstruction of the supramolecular enzyme complex. **Current Opinion in Structural Biology**, v. 23, n. 5, p. 686–694, 2013.
- 5 GÜLLERT, S. et al. Deep metagenome and metatranscriptome analyses of microbial communities affiliated with an industrial biogas fermenter, a cow rumen, and elephant feces reveal major differences in carbohydrate hydrolysis strategies. **Biotechnology for Biofuels**, v. 9, n. 1, p. 121, 2016.
- 6 HAIMOVITZ, R. et al. Cohesin-dockerin microarray: diverse specificities between two complementary families of interacting protein modules. **Proteomics**, v. 8, n. 5, p. 968–979, 2008.
- 7 PINHEIRO, B. A. et al. The clostridium cellulolyticum dockerin displays a dual binding mode for its cohesin partner. **Journal of Biological Chemistry**, v. 283, n. 26, p. 18422–18430, 2008.
- 8 FIEROBE, H. P. et al. Cellulosome from clostridium cellulolyticum: molecular study of the dockerin/cohesin interaction. **Biochemistry**, v. 38, n. 39, p. 12822–12832, 1999.
- 9 CASPI, J. et al. Effect of linker length and dockerin position on conversion of a Thermobifida fusca endoglucanase to the cellulosomal mode. **Applied and Environmental Microbiology**, v. 75, n. 23, p. 7335–7342, 2009.
- 10 NASH, M. A. et al. Single versus dual-binding conformations in cellulosomal cohesin-dockerin complexes. **Current Opinion in Structural Biology**, v. 40, p. 89–96, 2016. doi: 10.1016/j.sbi.2016.08.002
- 11 ALBER, O. et al. Cohesin diversity revealed by the crystal structure of the anchoring cohesin from Ruminococcus flavefaciens. **Proteins: structure, function and bioinformatics**, v. 77, n. 3, p. 699–709, 2009.
- 12 SALAMA-ALBER, O. et al. Atypical cohesin-dockerin complex responsible for cell surface attachment of cellulosomal components: binding fidelity, promiscuity, and structural buttresses. **Journal of Biological Chemistry**, v. 288, n. 23, p. 16827–16838, 2013.
- 13 BULE, P. et al. Single-binding mode integration of hemicellulose degrading enzymes via adaptor scaffoldins in Ruminococcus flavefaciens cellulosome. **Journal of Biological Chemistry**, v. 291, n. 52, 2016. doi: 10.1074/jbc.M116.761643.

- 14 RINCON, M. T. et al. Abundance and diversity of dockerin-containing proteins in the fiber-degrading rumen bacterium *Ruminococcus flavefaciens* FD-1. **PLoS ONE**, v. 5, n. 8, p. e12476, 2010.
- 15 BAYER, E. A. **Ed Bayer's Group**. Disponível em: <http://www.weizmann.ac.il/Biomolecular_Sciences/Bayer/>. Acesso em: 23 abr. 2017.
- 16 SLUTZKI, M. et al. Intramolecular clasp of the cellulosomal *Ruminococcus flavefaciens* ScaA dockerin module confers structural stability. **FEBS Open Bio**, v. 3, p. 398–405, 2013. doi: 10.1016/j.fob.2013.09.006.
- 17 WEINSTEIN, J. Y. et al. Insights into a type III cohesin-dockerin recognition interface from the cellulose-degrading bacterium *Ruminococcus flavefaciens*. **Journal of Molecular Recognition**, v. 28, n. 3, p. 148–154, 2015.
- 18 KARPOL, A. et al. Structural and functional characterization of a novel type-III dockerin from *Ruminococcus flavefaciens*. **FEBS Letters**, v. 587, n. 1, p. 30–36, 2013.
- 19 HYEON, J. E.; KANG, D. H.; HAN, S. O. Signal amplification by a self-assembled biosensor system designed on the principle of dockerin-cohesin interactions in a cellulosome complex. **Analyst**, v. 139, n. 19, p. 4790–4793, 2014.
- 20 CRAIG, S. J.; FOONG, F. C.; NORDON, R. Engineered proteins containing the cohesin and dockerin domains from *Clostridium thermocellum* provides a reversible, high affinity interaction for biotechnology applications. **Journal of Biotechnology**, v. 121, n. 2, p. 165–173, 2006.
- 21 GLOVER, D. J.; CLARK, D. S. Protein calligraphy: a new concept begins to take shape. **ACS Central Science**, v. 2, n. 7, p. 438–444, 2016.
- 22 HOLLINGSHEAD, S.; LIN, C.; LIU, J. C. Designing smart materials with recombinant proteins. **Macromolecular Bioscience**, 2017. doi: 10.1002/mabi.201600554.
- 23 KOBAYASHI, N.; ARAI, R. Design and construction of self-assembling supramolecular protein complexes using artificial and fusion proteins as nanoscale building blocks. **Current Opinion in Biotechnology**, v. 46, p. 57–65, 2017. doi: 10.1016/j.copbio.2017.01.001.
- 24 YEATES, T. O.; LIU, Y.; LANIADO, J. The design of symmetric protein nanomaterials comes of age in theory and practice. **Current Opinion in Structural Biology**, v. 39, p. 134–143, 2016. doi: 10.1016/j.sbi.2016.07.003.
- 25 PADILLA, J. E.; COLOVOS, C.; YEATES, T. O. Nanohedra: using symmetry to design self assembling protein cages, layers, crystals, and filaments. **Proceedings of the National Academy of Sciences of the United States of America**, v. 98, n. 5, p. 2217–2221, 2001.
- 26 REHM, B. H. A. Bioengineering towards self-assembly of particulate vaccines. **Current Opinion in Biotechnology**, v. 48, p. 42–53, 2017. doi: 10.1016/j.copbio.2017.03.018.
- 27 HEDDLE, J. G.; CHAKRABORTI, S.; IWASAKI, K. Natural and artificial protein cages: design, structure and therapeutic applications. **Current Opinion in Structural Biology**, v. 43, p. 148–155, 2017. doi: 10.1016/j.sbi.2017.03.007.

- 28 BRODIN, J. D. et al. Metal-directed, chemically tunable assembly of one-, two- and three-dimensional crystalline protein arrays. **Nature Chemistry**, v. 4, n. 5, p. 375–82, 2012.
- 29 RAJAGOPAL, K.; SCHNEIDER, J. P. Self-assembling peptides and proteins for nanotechnological applications. **Current Opinion in Structural Biology**, v. 14, n. 4, p. 480–486, 2004.
- 30 SINCLAIR, J. C. et al. Generation of protein lattices by fusing proteins with matching rotational symmetry. **Nature Nanotechnology**, v. 6, n. 9, p. 558–562, 2011.
- 31 GASTEIGER, E. et al. Protein identification and analysis tools on the ExPASy server. In: WALKER, J. M. (Ed.). **The proteomics protocols handbook**. Totowa: Humana Press, 2005. p. 571–607.
- 32 WLODAWER, A. et al. Protein crystallography for non-crystallographers, or how to get the best (but not more) from published macromolecular structures. **FEBS Journal**, v. 275, n. 1, p. 1–21, 2008.
- 33 BRADER, M. L. et al. Using X-ray crystallography to simplify and accelerate biologics drug development. **Journal of Pharmaceutical Sciences**, v. 106, n. 2, p. 477–494, 2016.
- 34 HENDRICKSON, W. A. Anomalous diffraction in crystallographic phase evaluation. **Quarterly Reviews of Biophysics** v. 47, n. 1, p. 49–93, 2014.
- 35 KABSCH, W. Xds. **Acta Crystallographica Section D: biological crystallography**, v. 66, n. 2, p. 125–132, 2010.
- 36 SHELDRICK, G. M. Experimental phasing with SHELXC/D/E: combining chain tracing with density modification. **Acta Crystallographica Section D: biological crystallography**, v. 66, n. 4, p. 479–485, 2010.
- 37 ADAMS, P. D. et al. PHENIX: a comprehensive python-based system for macromolecular structure solution. **Acta Crystallographica Section D: biological crystallography**, v. 66, n. 2, p. 213–221, 2010.
- 38 MCCOY, A. J. et al. Phaser crystallographic software. **Journal of Applied Crystallography**, v. 40, n. 4, p. 658–674, 2007.
- 39 EMSLEY, P. et al. Features and development of Coot. **Acta Crystallographica Section D: biological crystallography**, v. 66, n. 4, p. 486–501, 2010.
- 40 DU, X. et al. Insights into protein-ligand interactions: mechanisms, models, and methods. **International Journal of Molecular Sciences**, v. 17, n. 2, p. 144, 2016.
- 41 BERMAN, H. M. et al. The protein data bank. **Nucleic Acids Research**, v. 28, n. 1, p. 235–242, 2000.
- 42 BAKER, E. G. et al. Local and macroscopic electrostatic interactions in single α -helices. **Nature Chemical Biology**, v. 11, p. 221–228, 2015. doi:10.1038/nchembio.1739.
- 43 CHEN, X.; ZARO, J. L.; SHEN, W.-C. Fusion protein linkers: property, design and functionality. **Advanced Drug Delivery Reviews**, v. 65, n. 10, p. 1357–1369, 2013.

- 44 ASLANIDIS, C.; JONG, P. J. DE. Ligation-independent cloning of PCR products (LIC-PCR). **Nucleic Acids Research**, v. 18, n. 20, p. 6069–6074, 1990.
- 45 LI, M. Z.; ELLEDGE, S. J. SLIC: a method for sequence- and ligation - independent cloning. In: PECCOUD, J. (Ed.). **Gene synthesis**. Totowa, NJ: Humana Press, 2012. p. 51-59. (Methods in molecular biology, v. 852)
- 46 SAMBROOK, J.; RUSSELL, D. W. The inoue method for preparation and transformation of competent e. coli: “ultra-competent” cells. In: GREEN, M. R.; SAMBROOK, J. (Ed.). **Molecular cloning** : a laboratory manual. 4th ed. New York: Cold Spring Harbor Laboratory Press, 2012. p. 1.112-1.115.
- 47 HARRIS, J. R. Negative staining of thinly spread biological samples. **Methods in Molecular Biology**, v. 369, p. 107–142, 2007. doi: 10.1007/978-1-59745-294-6_7.
- 48 HOLM, L.; ROSENSTRÖM, P. Dali server: conservation mapping in 3D. **Nucleic Acids Research**, v. 38, Supplement 2, p. 545–549, 2010.
- 49 THIRUVENGADATHAN, R. et al. Nanomaterial processing using self-assembly-bottom-up chemical and biological approaches. **Reports on Progress in Physics**, v. 76, n. 6, p. 66501, 2013.
- 50 SINCLAIR, J. C. et al. Generation of protein lattices by fusing proteins with matching rotational symmetry. **Nature Nanotechnology**, v. 6, n. 9, p. 558–562, 2011.

Apêndice A – Tabela de sequências sintetizadas para clonagem

Tabela 12 - Relação das proteínas sintetizadas por GenScript. Os tipos de proteínas (C2, C3, C4 e C6) indicam sua simetria interna. As referências correspondem ao PDB ID de cada, quando disponível. Referências indicadas com asterisco (*) referem-se a pseudônimos utilizados em nossos documentos para diferenciar componentes de uma mesma entrada PDB. As dockerinas e cohesinas de *Ruminococcus flavefaciens* não requereram síntese, pois já se encontravam em nosso laboratório, sendo incluídas nessa tabela apenas para apresentação de sua sequência. Assim, foram pedidos os *primers* para amplificá-las, apenas.

Tipo de proteína	Referência	Organismo	Sequência da proteína
C4	1BKB	<i>Pyrobaculum aerophilum</i>	KWVMSTKYVEAGELKEGSYVVIDGEPCRVVEIEKS KTGKHGSAKARIVAVGVFDGGKRTLSPVDAQVEV PIIEKFTAQILSVSGDVIQLMDMRDYKTIEVPMKYV EEEAKGRLAPGAEEVWQILDRIYKIRVKG
C6	1LJO	<i>Archaeoglobus fulgidus</i>	GAMVLPNQMVKSMVKGKIRVEMKGEENQLVGKL EGVDDYMNLYLTNAMECKGEEKVRSLSGEIVLRGN NVVLIQPQ
C2	1UDV	<i>Sulfolobus solfataricus</i>	TEKLNEIVVRKTKNVEDHVLDVIVLFNQGIDEVILKG TGREISKAVDVYNSLKDRLGDGVQLVNVQTGSEVR DRRRISYILLRLKRVY
C3	1UMJ	<i>Pyrococcus horikoshii</i>	IIVYTTFPDWESAIEKVVKTLLKERLIACANLREHRAF YWWEGKIEEDKEVGAILKTREDLWEEKERIKELHP YDVPAAIRIDVDDVNEDYLLKWLIEETK
C3	2DT4	<i>Pyrococcus horikoshii</i>	VTGMFSLGRTYLFRVPEGEELLTYIKNFCKKEGIETA IINGIGTLKNPKIGYFLEEKKEYKVIPLKGSYELISLIGN VSLKDGEFPVHAHVSLGNEEGIVFGGHLVEGEVFFV AEIFLQELKGEKIERKPTKYGLALWEEKL
C2	2OSD	<i>Methanocaldococcus jannaschii</i>	EKIFPDILEAIRNEEIIKESKKIPMPYFGLFALVIFDKV KELGSETSLYEIGEEFGKMLSPKNIEELKKIFKLMNF GDLEIDENKILLKNPPYKIKLSNPPYQWVSKEPIHD FIAGILAGCLEEIFYFVVNEVECVSQGKDKCVFEV KE
C6	2QTX	<i>Methanocaldococcus jannaschii</i>	NFEYARRLNGKKVKIFLRNGEVLDAEVTGVSNYEI MVKVGDRNLLVFKHAIDYIEY
C4(D2)	2YX5	<i>Methanocaldococcus jannaschii</i>	YKATVIIKLLKGVLPNPEGRTIQRALNFLGFNNVKEV QTYKMIDIIMEGENEEKVKEEVEEMCKKLLANPVI HDYIEIKVEKIE
C6	2ZTS	<i>Pyrococcus horikoshii</i>	PVRRVKSIGPGFDELIEGGFPEGTTVLLTGGTGTGK TTFAAQFIYKGAEYGEFVFTLEERARDLRREM ASFGWDFEKEYEKGKIAIVDGVSSVGLPSEEKFL EDRFNVNDFLRYYRVVKAINAKRLVIDSIPSIALRLE EERKIREVLLKNTILLEMGVTTILTTEAPDPQHGKL SRYGIEEFIARGVIVLDLQEKNIELKRYVLIRKMRETR HSMKKYPFEIGPNGIVVYP
C2	3BUT	<i>Archaeoglobus fulgidus</i>	SLESVKAMWGVVTDSTQTEIVALAKVRNEDVVPVIV SGYHYTIEMNGVKVADGYENSPVTVPASATTLKF SLRLNNSFLREWWVTHIANGKTKIRVAIKPTIEIG GRDVEVPVFLRESEFTTKLL

(continua)

(continuação)

Tabela 12 - Relação das proteínas sintetizadas por GenScript. Os tipos de proteínas (C2, C3, C4 e C6) indicam sua simetria interna. As referências correspondem ao PDB ID de cada, quando disponível. Referências indicadas com asterisco (*) referem-se a pseudônimos utilizados em nossos documentos para diferenciar componentes de uma mesma entrada PDB. As dockerinas e cohesinas de *Ruminococcus flavefaciens* não requereram síntese, pois já se encontravam em nosso laboratório, sendo incluídas nessa tabela apenas para apresentação de sua sequência. Assim, foram pedidos os primers para amplificá-las, apenas.

C3	3C19	<i>Methanopyrus kandleri</i>	SLGLDYEPSHLMFLVTVLDDRDGEVLGDAlQKLiE REEVLACHAVPCVTkKNRPGHVLVLVDGGEDP DRVAEDVARDIMVLTGSTGVDRFDADGVYSVPS RFEDVRVVYGEREWRVSVKIAETEEGEVTVKAE FDECREIGEETGIPPREVKAMVEAAARVGGWVD LKEREIKVQ
C4	3N1U	<i>Legionella pneumophila</i>	KKIKCLICDVGVLSDGLLHIDNHGNElKSfHVQD GMGLKLLMAAGIQVAiITTAQNAVVDHRMEQL GITHYYKGQVDKRSAYQHlKKTlGLNDDEFAYIG DDLPLLIQVQVGLGVAVSNAVPQVLEFADWRT ERTGGRGAVRELCDLiLNAQNKAELAITGYLKQ
C6	4MML	<i>Pseudomonas aeruginosa</i>	GHSLQDPYLNTLRKERVPSiYLVNGIKLQGQIESF AQFVILLKNTVSQMvYKHAISTVVPSRPVRL
Cohesina (II)	2Y3N	<i>Bacteroides celulosolvens</i>	SGSVLTAIDNDKVAVGDKVTLTiNVDKITNfSGYQ FNIKYNTTYLQPWDTiADEAYTDSTMPDYGTLLQ GRFNATDMSKHNLsQGVLNfGRLYMNLSAYRAS GKPESTGAVAKVTFKVIKEiPAEGIKLATFENGSS MNNAVDGTMLFDWDGNMYSSsAYKVvQPLI YPKLE
Dockerina (II)	2Y3M*	<i>Bacteroides celulosolvens</i>	FVKlKGDLNGDGVINMADVMiLAQsFGKAIGNP GVNEKADLNNDGVINSDDAiILAQYFGKTK
Cohesina	4UYP	<i>Acetivibrio celulolyticum</i>	LQVDIGSTSGKAGSVVSPiTFTNVPKSGiYALSFR TNFDPQKVTVASIDAGSLiENASDFTTYNNENGf ASMTFEAPVDRARIIDSDGVfATINfKVSDSAKVG ELYNITTNSAYTSfFYSGTDEIKNVVYNDGKiEVIA L
Dockerina	4UYR*	<i>Acetivibrio celulolyticum</i>	KFiYGDVDGNGSVRSIDAVLiRDYVLGKiNEFPYey GMLAADVDGNGSiKINDAVLVRDYVLGKiFLFPV EEK
Cohesina	Escafoldina CipC	<i>Clostridium cellulovorans</i>	IPGDSLKVTVGTANGKPGDVTVPVTFADVAKM KNVGTCNFYLGyDASLLEVVSVDAGPiVKNAAVN FSSASNGTISFLFDNTITDELITADGVfANIKFKL KSVTAKTTTPVTFKDGGAFGDGTMSKiASVTkTN GSVTIDPGT

(continua)

(continuação)

Tabela 12 - Relação das proteínas sintetizadas por GenScript. Os tipos de proteínas (C2, C3, C4 e C6) indicam sua simetria interna. As referências correspondem ao PDB ID de cada, quando disponível. Referências indicadas com asterisco (*) referem-se a pseudônimos utilizados em nossos documentos para diferenciar componentes de uma mesma entrada PDB. As dockerinas e cohesinas de *Ruminococcus flavefaciens* não requereram síntese, pois já se encontravam em nosso laboratório, sendo incluídas nessa tabela apenas para apresentação de sua sequência. Assim, foram pedidos os primers para amplificá-las, apenas.

Cohesina	Escafoldina CipA	<i>Clostridium thermocellum</i>	NVPSDGVVVEIGKVTGSVGTTVEIPVYFRGVPSK GIANCDFVFRYDPNVLEIIGIDPGDIIVDPNPTKSF DTAIYPDRKIIVFLFAEDSGTGAYAITKDGVF AKIR ATVKSSAPGYITFDEVGGFADNDLVEQKVSFIDG GVNVGNATPT
Cohesina	Escafoldina ScaB	<i>Bacteroides celulosolvens</i>	KSSPGNKMKIQIGDVKANQGDTVIVPITFNEVPV MGVNNCNFTLAYDKNIMEFISADAGDIVTLPMA NYSYNMPSDGLVKFLYNDQAQGAMSIEDGTFA NVKFKIKQSAAFGKYSVGIIKAIGSISALSNSKLIPIES IFKDG SITVTNK
Cohesina (II)	Escafoldina OlpB	<i>Clostridium thermocellum</i>	SELPDSYVIMELDKTKVKVGDITATIKIENMKNFA GYQLNIKYDPTMLEAIELETGSAIAKRTWPVTGG TVLQSDNYGKTTAVANDVGAGIINF AEAYSNLTK YRETGVAAETGIIGKIGFRVLKAGSTAIRFEDTTA MPGAIEGTYMFDWYGENIKGYSVVQPGEIVAEG EE
Dockerina	Celulase familia 48s	<i>Clostridium thermocellum</i>	TKLYGDVNDDGKVNSTDAVALKRYVLRSGISINT DNADLNEDGRVNSTD LGILKRYILKEIDTLPYKN
Dockerina	Celulase família 5a	<i>Clostridium cellulovorans</i>	DPVIVYGDYNNDGNVDALDFAGLK KYIMAADH AYVKNLDVNLDNEVNSTDLAILKKYLLGMVSKLP SN
Dockerina	Escafoldina ScaA	<i>Bacteroides celulosolvens</i>	GTATVLYGDVDNDGNVSDDDYAYMRQWLIGM IADFPGGDIGLANADV DGDGNVSDDDYAYMRQ WLIGMISEFP AEQK
Dockerina (II)	Escafoldina CipA	<i>Clostridium thermocellum</i>	IEGYKVS GYILPDFSFDATVAPLVKAGFKVEIVGTE LYAVTDANGYFEITGV PANASGYTLKISRATYLD R VIANVVVTGDTSVSTSQA PIMMWVGDIVKD NSI NLLDVAE VIRC FNATKGSANYVEELDINRNGAIN MQDIMIVHKHFGATSSDYDAQ
Dockerina (III)	Escafoldina ScaB	<i>Ruminococcus flavefaciens</i>	ANYDHSYAVIETQVGYFFSHDDGVRANGQVGG FNKGQVVTLKIVDVYKDGSEVERPEVNMDLIN F NNNTPEKAYNSRTHINKQTTITDFKYDIPVYYGD MPLVDKSGKALT V TAYIGVKGDITLSNMVDSND ASVALSYT LISTGSNTPANTICQITEEGLQVSSAL DELDNLAAFLGDVNTNEWDPDNWKT T KMGRKI DSNDASQILAFYSKRTSHVYDDVSSNDIWN EILG DARFGK

(continua)

(conclusão)

Tabela 12 - Relação das proteínas sintetizadas por GenScript. Os tipos de proteínas (C2, C3, C4 e C6) indicam sua simetria interna. As referências correspondem ao PDB ID de cada, quando disponível. Referências indicadas com asterisco (*) referem-se a pseudônimos utilizados em nossos documentos para diferenciar componentes de uma mesma entrada PDB. As dockerinas e cohesinas de *Ruminococcus flavefaciens* não requereram síntese, pois já se encontravam em nosso laboratório, sendo incluídas nessa tabela apenas para apresentação de sua sequência. Assim, foram pedidos os primers para amplificá-las, apenas.

Cohesina (III)	Escafoldina ScaE	<i>Ruminococcus flavefaciens</i>	GPAAGQAYDAGNLDVASSPVKPTLSITKKTLTAA EAPNAKVTMELSVEGAADKYAATGLHIQFDPKLL LIPDEGALATAGRAARLLELKKAEADTDNSFFTA TGSSTNNGKDGVLWSFVLQVPADAQPGDKYDV QVAYQSRTTNEDLFTNVKKDEEGLLMQAWTFT QGIEQGYIQVESTTSLE
Dockerina (III)	Escafoldina ScaA	<i>Ruminococcus flavefaciens</i>	PGTKLVPTWGDNTCDGVVNVADVVLNRFLND PTYSNITDQGKVNADVDPQDKSGAAVDPAGV KLTVADSEAILKAIVELITLPQ
Cohesina (III)	Escafoldina ScaB	<i>Ruminococcus flavefaciens</i>	GTVEWVIPKVDADAGDNVTLQVAVKGDADLPV SAAQFDVNGATPLNYQESAGKPYGVDLVVNDTK QHFAFAQSGISKATAKDGDKLFTITFKVPEDCAD GIYPVTWAKQFVSDENGNNITDKVTFTDGWVR VGNPPTTT

Fonte: Elaborada pelo autor.

Apêndice B – Tabela de Primers

Tabela 13 - Relação de *primers* sintetizados para amplificação dos insertos componentes das construções de redes proteicas auto montáveis. As sequências estão agrupadas pelo tipo de proteína, simetrias C2, C3, C4 e C6, Dockerinas e Cohesinas, além dos *primers* utilizados para amplificação do vetor. Os sufixos indicam o tipo de *linker* composto pelo primer, sendo: SGEEARRGS (L1), GSEAAARREAEARGs (L2) GSGGSG (L3).

NOME	SEQUÊNCIA 5'→3'
Simetrias C3	
3C19_Fw	ACCAGGAGCAAACGGGAGGTAGCCTGGGCCTGGATTAC
3C19_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCT TTGCACTTTGATTTACAGTTCT
	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT
3C19_Rv_L2	ACCTTGCACTTTGATTTACAGTTCT
3C19_Rv_L3	CC GCT GCC GCC GCT ACCTTGCACTTTGATTTACAGTTCT
1UMJ_Fw	ACCAGGAGCAAACGGGAGGTATTATTGTCTACACGACCTTCCC
1UMJ_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTTTTGGTTTCTTCAATCAGCCA
	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT
1UMJ_Rv_L2	ACCTTTGGTTTCTTCAATCAGCCA
1UMJ_Rv_L3	CC GCT GCC GCC GCT ACCTTTGGTTTCTTCAATCAGCCA
Simetrias C4	
1BKB_Fw	ACCAGGAGCAAACGGGAGGTAAATGGGTGATGAGTACCAAATATG
1BKB_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTGCCTTTCACACGGATGATT
	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT
1BKB_Rv_L2	ACCGCCTTTCACACGGATGATT
1BKB_Rv_L3	CC GCT GCC GCC GCT ACCGCCTTTCACACGGATGATT
3N1U_Fw	ACCAGGAGCAAACGGGAGGTAAGAAAATTAAATGTCTGATCTGCG
3N1U_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTTTGTTTCAGGTAGCCCGTAATC
	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT
3N1U_Rv_L2	ACCTTGTTTCAGGTAGCCCGTAATC
3N1U_Rv_L3	CC GCT GCC GCC GCT ACCTTGTTTCAGGTAGCCCGTAATC
Simetrias C6	
1LJO_Fw	ACCAGGAGCAAACGGGAGGTGGTGCGATGGTGCTGC
1LJO_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTTTGCGGTTGGATCAGCA

(continua)

(continuação)

Tabela 13 - Relação de primers sintetizados para amplificação dos insertos componentes das construções de redes proteicas auto montáveis. As sequências estão agrupadas pelo tipo de proteína, simetrias C2, C3, C4 e C6, Dockerinas e Cohesinas, além dos primers utilizados para amplificação do vetor. Os sufixos indicam o tipo de linker composto pelo primer, sendo: SGEEARRGS (L1), GSEARREAEARGs (L2) GSGGSG (L3).

1LJO_Rv_L2	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT ACCTTGCGGTTGGATCAGCA
1LJO_Rv_L3	CC GCT GCC GCC GCT ACCTTGCGGTTGGATCAGCA
2QTX_Fw	ACCAGGAGCAAACGGGAGGTAATTTGAATATGCACGTCG
2QTX_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTGTATTCGATGTAGTCGATTGCG
2QTX_Rv_L2	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT ACCGTATTCGATGTAGTCGATTGCG
2QTX_Rv_L3	CC GCT GCC GCC GCT ACCGTATTCGATGTAGTCGATTGCG

Dockerinas

D48sCt_Fw_L1	GAA GAA GCG CGC CGC GGT AGCACCAAACGTATGGCGATGT
D48sCt_Fw_L2	CGC GAA GCG GAA GCG CGT GGT AGCACCAAACGTATGGCGATGT
D48sCt_Fw_L3	AGC GGC GGC AGC GGTACCAAACGTATGGCGATGT
D48sCt_Rv	CAGACCGCCACCGCTTCAATTTTATACGGCAGCGTGTC
D2Y3M_Fw_L1	GAA GAA GCG CGC CGC GGT AGCTTCGTGAAACTGAAAGGCG
D2Y3M_Fw_L2	CGC GAA GCG GAA GCG CGT GGT AGCTTCGTGAAACTGAAAGGCG
D2Y3M_Fw_L3	AGC GGC GGC AGC GGTTCGTGAAACTGAAAGGCG
D2Y3M_Rv	CAGACCGCCACCGCTTCATTTGGTTTTACCGAAATACTGTGC
DxB_Rf_Fw_L1	GAA GAA GCG CGC CGC GGT AGCGCAAACACGATCATTCTTATGCC
DxB_Rf_Fw_L2	CGC GAA GCG GAA GCG CGT GGT AGCGCAAACACGATCATTCTTATGCC
DxB_Rf_Fw_L3	AGC GGC GGC AGC GGTGCAAACACGATCATTCTTATGCC
DxB_Rf_Rv	CAGACCGCCACCGCTTCATTTACCGAAACGTGCATCG
DScaA_Rf_Fw_L1	GAA GAA GCG CGC CGC GGT AGCCCAGGTACTAAACTGGTTCCGA
DScaA_Rf_Fw_L2	CGC GAA GCG GAA GCG CGT GGT AGCCCAGGTACTAAACTGGTTCCGA
DScaA_Rf_Fw_L3	AGC GGC GGC AGC GGTCCAGGTACTAAACTGGTTCCGA
DScaA_Rf_Rv	CAGACCGCCACCGCTTCACTGCGGCAGGGTGATC

(continua)

(continuação)

Tabela 13 - Relação de primers sintetizados para amplificação dos insertos componentes das construções de redes proteicas auto montáveis. As sequências estão agrupadas pelo tipo de proteína, simetrias C2, C3, C4 e C6, Dockerinas e Cohesinas, além dos primers utilizados para amplificação do vetor. Os sufixos indicam o tipo de linker composto pelo primer, sendo: SGEEARRGS (L1), GSEARREAEARGS (L2) GSGGSG (L3).

Cohesinas	
C2Ct_Fw	ACCAGGAGCAAACGGGAGGTAATGTTCCGAGTGATGGCG
C2Ct_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTGGTCGGGGTCGCATTG
C2Ct_Rv_L2	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT ACCGGTCGGGGTCGCATTG
C2Ct_Rv_L3	CC GCT GCC GCC GCT ACCGGTCGGGGTCGCATTG
C2Y3N_Fw	ACCAGGAGCAAACGGGAGGTAGCGGCTCGGTTCTGAC
C2Y3N_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTTCCAGTTTCGGATAGATCAGAC
C2Y3N_Rv_L2	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT ACCTTCCAGTTTCGGATAGATCAGAC
C2Y3N_Rv_L3	CC GCT GCC GCC GCT ACCTTCCAGTTTCGGATAGATCAGAC
CERf_Fw	ACCAGGAGCAAACGGGAGGTGGTCCGGCAGCAGGTC
CERf_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTCTCCAGGCTAGTAGTGGATTCC
CERf_Rv_L2	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT ACCCTCCAGGCTAGTAGTGGATTCC
CERf_Rv_L3	CC GCT GCC GCC GCT ACCCTCCAGGCTAGTAGTGGATTCC
C4BRf_Fw	ACCAGGAGCAAACGGGAGGTGGTACTGTTGAATGGGTTATCCC
C4BRf_Rv_L1	CC GCG GCG CGC TTC TTC ACC GCTAGTAGTGGTTGGCGGGTTG
C4BRf_Rv_L2	CG CGC TTC CGC TTC GCG ACG CGC CGC TTC GCT ACCAGTAGTGGTTGGCGGGTTG
C4BRf_Rv_L3	CC GCT GCC GCC GCT ACCAGTAGTGGTTGGCGGGTTG
Simetrias C2	
1UDV_Fw_L1	GAA GAA GCG CGC CGC GGT AGCACCGAAAACTGAATGAAATCG
1UDV_Fw_L2	CGC GAA GCG GAA GCG CGT GGT AGCACCGAAAACTGAATGAAATCG
1UDV_Fw_L3	AGC GGC GGC AGC GGTACCGAAAACTGAATGAAATCG
1UDV_Rv	CAGACCGCCACCGCTTCAATAGACACGTTTCAGACGCAGC
3BUT_Fw_L1	GAA GAA GCG CGC CGC GGT AGCTCTCTGGAAAGCGTTAAAGCG
3BUT_Fw_L2	CGC GAA GCG GAA GCG CGT GGT AGCTCTCTGGAAAGCGTTAAAGCG

(continua)

(conclusão)

Tabela 13 - Relação de primers sintetizados para amplificação dos insertos componentes das construções de redes proteicas auto montáveis. As sequências estão agrupadas pelo tipo de proteína, simetrias C2, C3, C4 e C6, Dockerinas e Cohesinas, além dos primers utilizados para amplificação do vetor. Os sufixos indicam o tipo de linker composto pelo primer, sendo: SGEEARRGS (L1), GSEARREAEARGS (L2) GSGGSG (L3).

3BUT_Fw_L3 AGC GGC GGC AGC GGTCTCTGGAAAGCGTTAAAGCG

3BUT_Rv CAGACCGCCACCGCTTCACAGCAGTTTGGTCGTAAATTCG

Vetor

pSumoLIC_Fw GCGGTGGCGGTCTGAA

pSumoLIC_Rv CCTCCCGTTTGCTCCTGG

Fonte: Elaborada pelo autor.