



**Elisabeth Nelly Pires**

Licenciatura em Química Aplicada

**A organização molecular do Celulossoma:  
caracterização estrutural da proteína CipA e  
estudo da diversidade das interações  
coesina-doquerina**

Dissertação para obtenção do Grau de Mestre em Bioquímica

Orientador: Doutora Benedita Andrade Pinheiro, UCIBIO, FCT/UNL  
Co-orientador: Doutora Ana Luísa Moreira de Carvalho, UCIBIO, FCT/UNL

Júri:

Presidente: Professor Doutor Pedro António de Brito Tavares  
Arguente: Doutora Joana Luís Armada Brás  
Vogais: Doutora Benedita Andrade Pinheiro  
Doutora Joana Luís Armada Brás



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

**Outubro, 2015**





**Elisabeth Nelly Pires**

Licenciatura em Química Aplicada

**A organização molecular do Celulossoma:  
caracterização estrutural da proteína CipA e  
estudo da diversidade das interações  
coesina-doquerina**

Dissertação para obtenção do Grau de Mestre em Bioquímica

Orientador: Doutora Benedita Andrade Pinheiro, UCIBIO, FCT/UNL  
Co-orientador: Doutora Ana Luísa Moreira de Carvalho, UCIBIO, FCT/UNL

Júri:

Presidente: Professor Doutor Pedro António de Brito Tavares  
Arguente: Doutora Joana Luís Armada Brás  
Vogais: Doutora Benedita Andrade Pinheiro  
Doutora Joana Luís Armada Brás



FACULDADE DE  
CIÊNCIAS E TECNOLOGIA  
UNIVERSIDADE NOVA DE LISBOA

**Outubro, 2015**



**“A organização molecular do Celulossoma: caracterização estrutural da proteína CipA e estudo da diversidade das interações coesina-doquerina”**

Copyright © em nome de Elisabeth Nelly Pires, da Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa.

A Faculdade de Ciências e Tecnologia e a Universidade Nova de Lisboa têm o direito, perpétuo e sem limites geográficos, de arquivar e publicar esta dissertação através de exemplares impressos reproduzidos em papel ou de forma digital, ou por qualquer outro meio conhecido ou que venha a ser inventado, e de a divulgar através de repositórios científicos e de admitir a sua cópia e distribuição com objetivos educacionais ou de investigação, não comerciais, desde que seja dado crédito ao autor e editor.



## Agradecimentos

Dedico esta dissertação aos meus pais, aos meus irmãos e ao João. Um Obrigada especial pelo vosso amor! ♡

Ao Pepe e Pipocas, os meus bebes.

Agradeço a minha família.

Obrigada à Professora Maria João Romão pela oportunidade de realizar a minha dissertação no laboratório de Cristalografia, bem como à formação teórica que me proporcionou, essencial para a realização desta tese em Bioquímica.

Obrigada às minha orientadoras, Doutora Benedita Pinheiro e Doutora Ana Luísa Carvalho, por me terem ajudado durante esta etapa, pela disponibilidade e por todos os ensinamentos que me proporcionaram.

Obrigada à Doutora Angelina Palma por toda a ajuda que me facultou no estudo em *Microarrays* e pela sua inteira disponibilidade.

Agradeço a todos os membros do XTAL: Doutora Teresa, Doutor Jorge, Doutora Márcia, Doutor Filipe, Doutora Catarina, Viviana, Cecília, Diana, Rita, Francisco, Hugo e Marino. Obrigada pela recetividade, pela ajuda e por terem contribuído para o meu crescimento.

Agradeço à Raquel, Vanessa e Joana, por todo o apoio, amizade e por estarem sempre comigo desde do primeiro dia nesta instituição. Obrigada por tudo. Adoro-vos!

Agradeço à Ana Inês, Raquel, Cátia, Tiago e Rita por terem estado presentes desde do início deste percurso, pela amizade e por termos partilhado alguns momentos de desespero.

Obrigada a todos os meus amigos e colegas de Licenciatura e Mestrado que me acompanharam durante o meu percurso académico na FCT.

Obrigada a todos!



## Resumo

A celulose, principal componente da parede celular vegetal e o composto orgânico mais abundante da biosfera, possui inúmeras aplicações biotecnológicas. O microrganismo anaeróbio *Clostridium thermocellum* (*C.thermocellum*) tem sido alvo de grande interesse pela sua capacidade em degradar eficientemente a celulose e outros componentes da parede celular vegetal, por meio de um complexo multi-enzimático altamente eficiente, denominado de celulosoma. A montagem deste complexo ocorre através de uma proteína multi-modular denominada CipA. Esta proteína estrutural possui módulos não catalíticos (coesinas tipo I) que se ligam a módulos complementares (doquerinas tipo I) presentes nas enzimas celulolíticas modulares. A CipA possui ainda um módulo doquerina de tipo II que permite a ancoragem deste complexo multi-enzimático à parede celular da bactéria. Na presente dissertação foram utilizadas as metodologias de Cristalografia de Raios-X, para caracterizar a interação coesina-doquerina a nível atômico e molecular, e de *Microarrays*, com o intuito de estudar as possíveis especificidades e afinidades dessas interações. Com base na primeira técnica foram elucidadas as estruturas do módulo coesina C4 da CipA em complexo com a doquerina da enzima modular Xyn10B e do módulo coesina C9 isolado. As estruturas foram comparadas com o complexo do módulo coesina C2-doquerina Xyn10B já publicado. Esta análise encontra-se descrita no capítulo 3. Por último, a técnica de *Microarrays*, associada à eletroforese em gel de poliacrilamida em condições nativas, permitiu a caracterização das diferenças de afinidade e especificidade entre os vários pares coesina-doquerina dos celulosomas de *C. thermocellum* e de *Rumminococcus flavefaciens* (*R. flavefaciens*). As especificidades e afinidades dos módulos doquerina, dos celulosomas mencionados anteriormente estão descritas no capítulo 4.

**Palavras-chaves:** Celulosoma, *Clostridium thermocellum*, *Rumminococcus flavefaciens*, coesina, doquerina, Cristalografia de Raios-X e *Microarrays*



## Abstract

Cellulose, the main component of the cell wall and the most abundant organic compound of the biosphere, has numerous biotechnological applications. The anaerobic microorganism *Clostridium thermocellum* (*C. thermocellum*) has been the subject of great interest for its ability to efficiently degrade cellulose and other components of the plant cell wall, through a highly efficient multi-enzimatic complex, called cellulosome. The assembly of this complex occurs through a multi-modular protein called CipA. This structural protein has non-catalytic modules (cohesins type I), which bind to complementary modules (dockerins type I) that are present in modular cellulolytic enzymes. CipA has a type II dockerin module that allows the anchoring of this multi-enzymatic complex to the cell wall of the bacteria. In this dissertation, X-ray Crystallography methods were used to characterize the cohesin-dockerin interaction at atomic and molecular levels. In addition, the Microarray technology was implemented to study the possible specificities and affinities of these interactions. This work describes the 3D structure solution of the C4 cohesin module from CipA in complex with the dockerin Xyn10B, and of the isolated C9 cohesin module. A structural comparison with the already known and published cohesin-dockerin complex C2-Xyn10B is described in chapter 3. Finally, the Microarray technology was implemented in association with polyacrylamide gel electrophoresis in native conditions, to evaluate the affinity and specificity differences, between the various cohesin-dockerin pairs of the cellulosomes of *C. thermocellum* and *Rumminococcus flavefaciens* (*R. flavefaciens*). The specificities and affinities of the dockerin modules are described in chapter 4.

**Keywords:** Cellulosome, *Clostridium thermocellum*, *Rumminococcus flavefaciens*, cohesin dockerin, X-Ray crystallography and Microarrays



# Índice

Agradecimentos.....	III
Resumo.....	V
Abstract .....	VII
Índice.....	IX
Índice de figuras .....	XIII
Índice de tabelas .....	XIX
Lista de abreviaturas, siglas/acrónimos e símbolos .....	XXI
Capítulo 1 - Introdução .....	23
1. Introdução .....	1
1.1. Parede Celular Vegetal.....	3
1.1.2. Degradação enzimática da parede celular vegetal .....	4
1.2. Descrição do microrganismo <i>Clostridium thermocellum</i> .....	5
1.3. Celulossoma do <i>Clostridium thermocellum</i> : Arquitetura e Função.....	5
1.3.1. Proteínas do celulossoma do <i>Clostridium thermocellum</i> .....	7
1.3.1.1. <i>Cellulosome Integrating Protein</i> (CipA) e proteínas de ancoragem .....	7
1.3.1.2. Interação coesina-doquerina.....	7
1.3.1.3. Módulos de ligação a hidratos de carbono .....	9
1.3.2. Cristalização de proteínas.....	12
1.3.3. Difração de Raios-X e resolução da estrutura .....	14
1.3.4. Refinamento e validação .....	16
1.4. Tecnologia dos <i>Microarrays</i> .....	17
1.4.1 <i>Microarrays</i> de proteína.....	17
1.4.1.1. Desenho do padrão e impressão das proteínas em suporte sólido .....	19
1.4.1.2. <i>Arrayers</i> – manual e robótico.....	19
1.4.1.3. Suportes sólidos .....	20
Capítulo 2 - Objetivos .....	23
2. Objetivos .....	25
Capítulo 3 - Caracterização estrutural da proteína CipA por Cristalografia de Raios-X .....	27
3.1. Introdução .....	29
3.2. Métodos e Materiais .....	29
3.2.1. Transformação de células competentes .....	29
3.2.2. Expressão dos complexos coesina-doquerina .....	30
3.2.3. Purificação dos complexos coesina-doquerina.....	30
3.2.3.1. Cromatografia de afinidade por íons metálicos imobilizados.....	31
3.2.3.2. Cromatografia de exclusão molecular .....	31
3.2.3.3. Cromatografia de permuta iónica .....	32
3.2.4. Ensaios de cristalização dos complexos C1X10B, C4X10B e do módulo coesina C9 .....	32
3.2.5. Recolha de dados de difração de raios-X e resolução de estruturas tridimensionais.....	33
3.3. Apresentação e discussão dos resultados .....	34
3.3.1. Expressão dos complexos coesina-doquerina .....	34

3.3.2. Purificação dos complexos por cromatografia de afinidade por íons metálicos imobilizados .....	37
3.3.3. Purificação dos complexos por cromatografia de exclusão molecular .....	39
3.3.4. Purificação dos complexos por cromatografia de permuta iônica .....	40
3.3.5. Ensaios de cristalização das amostras em estudo, experiência de difração de raios-X e resolução de estruturas tridimensionais .....	41
3.3.5.1. Complexo coesina-doquerina C1X10B .....	42
3.3.5.1.1. Ensaios de cristalização e experiência de difração de raios-X .....	42
3.3.5.2. Complexo coesina-doquerina C4X10B .....	43
3.3.5.2.1. Ensaios de cristalização e experiência de difração de raios-X .....	43
3.3.5.2.2. Resolução da estrutura tridimensional do complexo C4X10B .....	47
3.3.5.2.3. Resolução da estrutura tridimensional do complexo C4X10B .....	47
3.3.5.3. Módulo coesina C9 .....	49
3.3.5.3.1. Ensaios de cristalização e experiência de difração de raios-X .....	49
3.3.5.3.2. Resolução da estrutura tridimensional do módulo coesina C9 .....	53
3.3.5.3.3. Caracterização da estrutura tridimensional do módulo C9 .....	54
3.3.6. Comparação das estruturas C2-Xyn10B, C4X10B e C9 .....	55
3.4. Conclusão .....	58
Capítulo 4 - Análise das interações coesina-doquerina no <i>C.thermocellum</i> e <i>R.flavefaciens</i> usando a tecnologia de <i>microarrays</i> .....	59
4.1. Introdução .....	61
4.2. Materiais e Métodos .....	61
4.2.1. Transformação de células competentes .....	61
4.2.2. Expressão dos módulos coesina e doquerina de tipo I e II .....	62
4.2.3. Purificação dos módulos coesina e doquerina de tipo I e II .....	62
4.2.3.1. Cromatografia de afinidade por íons metálicos imobilizados .....	62
4.2.4. Ensaios de ligação e de competição entre módulos coesina e doquerina .....	63
4.2.4.1. Ensaios de ligação e de competição através da técnica de <i>microarray</i> .....	63
4.2.4.2. Ensaios de ligação e competição em gel de poliacrilamida em condições nativas .....	67
4.3. Apresentação e discussão dos resultados .....	68
4.3.1. Expressão e purificação dos módulos coesina e doquerina de tipo I e II .....	68
4.3.1.1. Expressão e purificação dos módulos coesina de tipo I e II .....	68
4.3.1.2. Expressão e purificação das doquerinas de tipo I e tipo II .....	70
4.3.2. Ensaios de ligação e de competição entre os módulos coesina e doquerina .....	73
4.4. Conclusão .....	82
Capítulo 5 – Conclusão e perspectivas futuras .....	89
5. Conclusão geral e perspectivas futuras .....	85
Capítulo 6 - Bibliografia .....	89
6. Bibliografia .....	89
Capítulo 7 - Anexos .....	93
76A – Anexos referentes ao estudo estruturais da proteína CipA por Cristalografia de Raios-X .....	95
7A.I – Composição dos meios de cultura .....	95
7A.II - Protocolo de purificação por cromatografia de afinidade por íons metálicos imobilizados .....	95
7A.III -Protocolo de purificação por cromatografia de exclusão molecular .....	95

7A.IV - Protocolo de purificação por cromatografia de permuta iónica .....	96
7A.V - Screen de cristalização PEG/ION 1K.....	96
7A.VI - <i>Screen</i> de cristalização PEG/ION 8K .....	98
7A.VII - <i>Screen</i> de cristalização JBS 5 .....	99
7A.VIII – <i>Screen</i> de cristalização JBS 6 .....	100
7A.IX – <i>Screen</i> de cristalização JBS 7.....	100
7A.X – <i>Screen</i> de cristalização JBS 8 .....	101
7A.XI – <i>Screen</i> de cristalização JCSG+.....	102
7A.XII – <i>Screen</i> de cristalização 80! .....	103
7B – Anexos referentes ao estudo das interacções coesina-doquerina no <i>C. thermocellum</i> e <i>R. flavefaciens</i> por <i>microarrays</i> .....	106
7B.I - Informação sobre os módulos coesina e doquerina.....	106
7B.II – Padrão de impressão dos módulos coesina realizado roboticamente .....	108
7B.III – Padrão de impressão dos módulos coesina realizado manualmente .....	109
7B.IV – Protocolo de Microarrays .....	110
7B.V – Gráficos – Intensidades de fluorescência em função da posição do módulo coesina .....	112



## Índice de figuras

<b>Figura 1.1-</b> Esquema representativo da parede celular vegetal. Adaptado do Bioenergy Science Center ( <a href="http://bioenergy.ccruc.uga.edu/">http://bioenergy.ccruc.uga.edu/</a> ) .....	3
<b>Figura 1.2</b> - Estrutura da celulose. Adaptado de Béguin et Lemaire, 1996.....	4
<b>Figura 1.3</b> – Superfície do <i>C.thermocellum</i> . Adaptado de Bayer e Lamed, 1986.....	5
<b>Figura 1.4</b> – Estrutura da superfície celular do <i>C. thermocellum</i> e visualização da mesma por microscopia eletrônica – Microscopia de transmissão eletrônica (TEM). Adaptado de Sholman et al., 1999. ....	6
<b>Figura 1.5</b> – Representação esquemática da interação do celulosoma do <i>C.thermocellum</i> com a celulose, e da interação celulosoma-parede bacteriana. Adaptado de Fontes e Gilbert, 2010. ....	6
<b>Figura 1.6</b> – Representação esquemática da percentagem de identidade entre as coesinas tipo I da CipA, do <i>C.thermocellum</i> . Adaptado Fontes e Gilbert, 2010. ....	7
<b>Figura 1.7</b> – A - Representação esquemática a nível atômico das estruturas determinadas por Cristalografia de Raios-x. Adaptado de Fontes e Gilbert, 2010. B - Estrutura obtida por Cristalografia de Raios-X do complexo C2-Xyn10B e representação da conformação em EF-hand presente na doquerina.....	8
<b>Figura 1.8</b> - Representação esquemática das principais etapas para a obtenção de uma estrutura tridimensional por cristalografia de Raios-X. ....	10
<b>Figura 1.9</b> - Espectro eletromagnético. Os raios-X encontram-se entre os 0,1 e 100Å (1Å=10 <sup>-10</sup> m). ( <a href="http://www.esrf.eu/about/synchrotron-science/synchrotron-light">http://www.esrf.eu/about/synchrotron-science/synchrotron-light</a> ) .....	11
<b>Figura 1.10</b> – Representação esquemática de 3 fontes de raios-X distintas: A – Tubo de <i>Coolidge</i> ; B – Ânodo rotativo; C – Síncrotrão. Adaptado de <a href="http://www.esrf.eu/">http://www.esrf.eu/</a> e Rhodes, 2006. ....	12
<b>Figura 1.11</b> -Representação esquemática da unidade assimétrica, da célula unitária e da rede cristalina. Adaptado de Rhodes, 2006.....	12
<b>Figura 1.12</b> -Representação esquemática da técnica de difusão de vapor. Adaptado de Taylor <i>et al.</i> , 2003 e Hall <i>et al.</i> , 2007. A – método da gota suspensa; B – método da gota assente; C – diagrama de fases representando a variação das concentrações de proteína e agente precipitante durante o processo de cristalização de uma proteína. ....	13
<b>Figura 1.13</b> -Representação da Lei de Bragg. <sup>[38]</sup> .....	15
<b>Figura 1.14</b> -Representação esquemática dos três tipos de <i>microarrays</i> : A – <i>Microarrays</i> Analíticos; B – <i>Microarrays</i> de Fase Reversa; C – <i>Microarrays</i> Funcionais. Adaptado de Sheehan et al, 2005. [43].....	18
<b>Figura 1.15</b> - Representação esquemática das principais etapas envolvidas na técnica de <i>microarrays</i> de proteína. ....	18
<b>Figura 1.16</b> – Fotografias de um arrayer manual (esquerda) e robótico (direita). Adaptado de <a href="http://www.arrayit.com">http://www.arrayit.com</a> . ....	19
<b>Figura 1.17</b> - Ilustração do método de impressão de um slide de 1 pad, a partir do modelo criado na microplaca de 384 poços, e do respectivo arrayer manual utilizado para o efeito. ....	20
<b>Figura 3.1</b> -Constructo utilizado para expressão dos complexos coesina-doquerina. Os termos Doc e Coh correspondem aos módulos doquerina e coesina, respetivamente. Cada um destes componentes é expresso simultaneamente, com promotores independentes. A coesina apresenta no seu C-terminal uma <i>His-tag</i> . ....	29

- Figura 3.2** - Resultados da análise eletroforética em condição desnaturantes das frações eluídas durante a purificação por IMAC dos complexos Coh-Doc. As condições de expressão estão indicadas no topo de cada figura. Indução por IPTG, a 37°C durante 4 horas e 19°C durante 16 horas. E1 – Primeira eluição em tampão 50mM Hepes a pH7,5 com 60mM Imidazol; E2 – Segunda eluição em tampão 50mM Hepes a pH7,5 com 300mM Imidazol. M corresponde ao marcador de pesos moleculares da NZYTech..... 34
- Figura 3.3** - Resultados da análise eletroforética em condições desnaturantes das frações eluídas durante a purificação por IMAC dos complexos Coh-Doc. As concentrações de imidazol utilizadas para eluição das respectivas frações estão apresentadas no topo da figura. VM – Volume morto; 10mM – Solução tampão 50mM Hepes a pH 7,5 com 10mM de imidazol; 50mM – Solução tampão 50mM Hepes a pH 7,5 com 50mM de imidazol; 100mM – Solução tampão 50mM Hepes a pH 7,5 com 100mM de imidazol; 250mM – Solução tampão 50mM Hepes a pH 7,5 com 250mM de imidazol; M corresponde ao marcador de pesos moleculares da NZYTech. .... 36
- Figura 3.4** - Resultados da análise eletroforética em condições nativas da purificação por IMAC dos complexos Coh-Doc. As concentrações de imidazol utilizadas para eluição das respectivas frações estão apresentadas no topo da figura; A BSA é apresentada como referência de corrida. .... 36
- Figura 3.5** - Resultados da purificação dos complexos Coh-Doc por IMAC: Cromatograma e análise eletroforética em condições desnaturantes das frações recolhidas. A - Resultados obtidos para a purificação do complexo C1X10B; B - Resultados obtidos para a purificação do complexo C4X10B; C - Resultados obtidos para a purificação do complexo C9X10B; VM corresponde ao volume morto; M corresponde ao marcador de pesos moleculares da NZYTech. Os círculos azuis mostram as frações escolhidas para posterior protocolo de purificação..... 38
- Figura 3.6** -Resultados da purificação dos complexos Coh-Doc por cromatografia de exclusão molecular: Cromatograma e análise eletroforética em condições desnaturantes e não desnaturantes das frações recolhidas. A - Resultados obtidos para a purificação do complexo C1X10B; B - Resultados obtidos para a purificação do complexo C4X10B; M corresponde ao marcador de pesos moleculares da NZYTech. BSA é apresentada como referência de corrida.Os círculos azuis mostram as frações escolhidas para posterior protocolo de purificação.. 39
- Figura 3.7** - Resultados da purificação dos complexos Coh-Doc por cromatografia de permuta iônica: Cromatograma e análise eletroforética em condições desnaturantes e nativas das frações recolhidas. A - Resultados obtidos para a purificação do complexo C4X10B; B - Resultados obtidos para a purificação do complexo C9X10B; M corresponde ao marcador de pesos moleculares da NZYTech. BSA é apresentada como referência de corrida. Os círculos azuis mostram as frações escolhidas para posteriores ensaios de cristalização. .... 41
- Figura 3.8** - Resultados obtidos para os ensaios de cristalização do complexo C1X10B utilizando o *screen* de 80!.O agente precipitante utilizado para a cristalização bem como os tamanhos dos cristais obtidos compreendidos entre os 0.1 e 0.4 mm estão apresentados na figura. .... 42
- Figura 3.9** -Resultados obtidos para os ensaios de cristalização do complexo C4X10B utilizando o *screen* de 80!, a 20°C. O agente precipitante utilizado para a cristalização, os tamanhos dos cristais (compreendidos entre 0,1 e 0,2 mm) e as concentrações em que foram obtidos encontram-se apresentados na figura. [1] – Concentração de 19,49 mg/mL; [2] – Concentração de 9,75 mg/mL ..... 43
- Figura 3.10** - Resultados obtidos para as primeiras duas semanas de observação da otimização realizada a 20°C, para a condição A6. As proporções de proteína:precipitante, as percentagens de agente precipitante e os tamanhos dos cristais (compreendidos entre 0,4 e 0,5 mm) estão esquematizados na figura. .... 44
- Figura 3.11** - Resultados obtidos para os ensaios de cristalização do complexo C4X10B utilizando o *screen* de 80!, a 20°C. As percentagens de agente precipitante e os tamanhos dos cristais, 0,4 mm, estão esquematizados na figura. [1] – Concentração de 19,49 mg/mL; [2] – Concentração de 9,75 mg/mL. .... 44
- Figura 3.12** - Resultado obtido para a primeira semana de observação da nova otimização da condição A6, a 20°C. A percentagem de agente precipitante e o tamanho dos cristais (0,4 mm) estão esquematizadas. A concentração de proteína utilizada tem o valor de 9,75 mg/mL..... 45
- Figura 3.13** - Padrão de difração do cristal do complexo C4X10B. .... 45

<b>Figura 3.14</b> - A - Representação das duas moléculas de C4X10B presentes na unidade assimétrica. B - Representação da estrutura tridimensional do módulo Xyn10B. Este apresenta um enrolamento em três hélices $\alpha$ (Ser17B-Val27B; Asp35B-Ala41B; Ala51B-Leu61B). As imagens criadas foram conseguidas através da utilização do programa Chimera. ....	48
<b>Figura 3.15</b> - Superfície de potencial electrostático do módulo coesina C4 do complexo C4X10B. O módulo Xyn10B está representado pela sua cadeia principal. ....	49
<b>Figura 3.16</b> - Resultados obtidos para os ensaios de cristalização da coesina C9 utilizando o <i>screen</i> de 80!, a 20°C. O agente precipitante utilizado para a cristalização e os tamanhos dos cristais, compreendidos entre 0,1 e 0,4 mm, encontram-se na figura. ....	50
<b>Figura 3.17</b> - Resultados obtidos para os ensaios de cristalização utilizando o <i>screen</i> de 80!, a 4°C. O agente precipitante utilizado para a cristalização e os tamanhos dos cristais (0,4 mm) encontram-se na figura. ....	50
<b>Figura 3.18</b> - Resultados obtidos para a otimização das condições E4 e F4 realizadas manualmente, para a temperatura de 20°C. As proporções de proteína:precipitante e as percentagens de agente precipitante estão esquematizadas na figura. Os tamanhos dos cristais estão compreendidos entre os 0,1 e 0,4 mm. A concentração de proteína utilizada tem o valor de 11,14mg/mL.....	51
<b>Figura 3.19</b> -Padrão de difração do cristal do módulo C9.....	52
<b>Figura 3.20</b> - Representação das duas moléculas de C9 presentes na unidade assimétrica; B - Representação da estrutura tridimensional do módulo C9; C - Superfície de potencial electrostático do módulo coesina C9. As imagens foram conseguidas através da utilização do programa Chimera. ....	54
<b>Figura 3.21</b> - Alinhamento múltiplo das seqüências dos módulos coesina C2, C4 e C9 pela ferramenta bioinformática MUSCLE, e respetiva árvore filogenética .....	55
<b>Figura 3.22</b> - A- Sobreposição das estruturas C2-Xyn10B (Laranja), C4X10B (Azul) e C9 (verde); B-Superfície de potencial electrostático dos módulos coesina C2, C4 e C9. Nas figuras representantes dos complexos C2-Xyn10B e C4X10B, o módulo Xyn10B está representado pela sua cadeia principal. ....	56
<b>Figura 4.1</b> - Constructos utilizados para expressão dos módulos coesina-doquerina. Os termos Doc e Coh correspondem aos módulos doquerina e coesina, respetivamente. As coesinas apresentam no seu N-terminal um módulo TEV e uma <i>His-tag</i> . As doquerinas possuem um módulo TRX e uma <i>His-tag</i> no seu N-terminal. ....	61
<b>Figura 4.2</b> - Respresentação esquemática da visualização do sinal de fluorescência emitido por um bloco após impressão e após o ensaio. ....	64
<b>Figura 4.3</b> - Representação esquemática dos principais passos de um ensaio de <i>microarrays</i> de proteínas. ....	65
<b>Figura 4.4</b> - Representação esquemática da construção da grelha para cada digitalização do slide, através do programa GenePix Pro da Molecular Devices. ....	66
<b>Figura 4.5</b> - Esquema do <i>template</i> de ensaio de ligação e de competição. Doc: doquerina; Coh1:coesina 1; Coh2: coesina 2; BSA: albumina sérica bovina. ....	67
<b>Figura 4.6</b> - Resultados da análise eletroforética em condições desnaturantes, após dessalinização, dos 30 módulos coesina expressos e purificados. M corresponde ao marcador de pesos moleculares da NZYTech. ....	69
<b>Figura 4.7</b> - Resultados da análise eletroforética, em condições nativas, após protocolo de dessalinização dos 30 módulos coesina expressos. A BSA é apresentada como referência de corrida. ....	69
<b>Figura 4.8</b> - Esquema da otimização dos protocolos de expressão, purificação e de dessalinização das 46 doquerinas em estudo. A figura descreve para cada volume de expressão utilizado, os vários protocolos testados, isto é, que colunas de purificação utilizadas (His GraviTrap <sup>TM</sup> e His SpinTRap <sup>TM</sup> ) e os respetivos volumes de coluna (600 $\mu$ L e 1mL); os volumes de eluição (200 $\mu$ L, 1, 2,5 e 5mL); e as colunas de dessalinização utilizadas (PD-10 e GE-25). Os resultados dos vários protocolos estão assinalados com um certo (a verde) aquando da obtenção de um resultado favorável; com um ponto de interrogação (a laranja) na presença de resultados	

inconclusivos; e com uma cruz (a vermelho) aquando da obtenção de um resultado não favorável à otimização do protocolo. .... 72

**Figura 4.9** - Resultados da análise eletroforética, em condições desnaturantes, após protocolo de dessalinização das 42 doquerinas expressas. M corresponde ao marcador de pesos moleculares da NZYTech ..... 73

**Figura 4.10** - Gráficos obtidos para a quantificação dos ensaios de ligação pela técnica de *microarray* para as quatro concentrações dos módulos coesina (C1, C2, OlpA, OlpC, SbdA, Orf2-1, EAM45267-1 e COH2) e para a concentração de 0,1  $\mu$ M de doquerina (CelA e XDoc2). .... 74

**Figura 4.11** - Gráficos obtidos para a quantificação dos ensaios de ligação pela técnica de *microarray* para as concentrações de 30 e 75  $\mu$ g/mL dos módulos coesina (C1, C2, OlpA, OlpC, SbdA, Orf2-1 e EAM45267-1) e para a concentração de 0,1  $\mu$ M de doquerina (CelA e XDoc2). .... 75

**Figura 4.12** - Resultados dos ensaios de ligação e de competição realizados por *microarrays* e eletroforese em condições nativas (entre os módulos coesina C2 e C9, e o módulo doquerina Xylz). A BSA é apresentada como referência de corrida. .... 78

**Figura 4.13** - Resultados dos ensaios de ligação e de competição realizados por *microarrays* e por eletroforese em condições nativas, para a doquerina CelS (entre os módulos coesinas C2 e C9, C2 e OlpA, e C1 e OlpA). A BSA é apresentada como referência de corrida. .... 78

**Figura 4.14** - Resultados dos ensaios de ligação e de competição realizados por eletroforese em condições nativas, para a doquerina Xyn10B (entre os módulos coesina C2 e C3, C2 e C4, C4 e C9, e C2 e C9). A BSA é apresentada como referência de corrida. .... 79

**Figura 4.15**- Esquema das três experiências realizadas no ensaio de competição entre os módulos coesina C2 e C9 em complexo com o módulo doquerina CelS. .... 81

**Figura 4.16** - Resultados relativos aos ensaios de competição pela técnica de *microarrays* de proteínas para as concentrações de 7,5, 30, 75 e 300  $\mu$ g/mL de coesina imobilizada: A – Gráficos referentes aos resultados para os módulos C2 e C9 imobilizados; B – Gráficos referentes à menor concentração onde se verificou a capacidade de competição entre os módulos em estudo. .... 81

**Figura 7B.1** – Representação esquemática da impressão dos slides de 16 *blocos* dos módulos coesina. Cada módulo foi impresso em quatro concentrações diferentes (7,5, 30, 75 e 300  $\mu$ g/mL). .... 108

**Figura 7B.2** - Representação esquemática da impressão dos slides de 1 *bloco* dos módulos coesina. Cada módulo foi impresso em quatro concentrações diferentes (7,5, 30, 75 e 300  $\mu$ g/mL). .... 109

**Figura 7B.3** - Representação esquemática da impressão dos slides de 2 *blocos* dos módulos coesina C2 e C9. Cada módulo foi impresso em quatro concentrações diferentes (7,5, 30, 75 e 300  $\mu$ g/mL). .... 109

**Figura 7B.4** - Representação esquemática da colocação dos slides de 1 ou 2 *blocos* na tampa de caixa de petri. O uso de algodão húmido proporciona um ambiente húmido, evitando assim a secagem da matriz de nitrocelulose. O parafilm promove a estabilização dos slides durante todo o ensaio. .... 111

**Figura 7B.5** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente. .... 112

**Figura 7B.6** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente. .... 113

**Figura 7B.7** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma

posição no microarray. Uma coesina tem 4 posições no microrrays correspondentes a uma concentração diferente. .... 114

**Figura 7B.8** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no microarray. Uma coesina tem 4 posições no microrrays correspondentes a uma concentração diferente. .... 115

**Figura 7B.9** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no microarray. Uma coesina tem 4 posições no microrrays correspondentes a uma concentração diferente. .... 116

**Figura 7B.10** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo II do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no microarray. Uma coesina tem 4 posições no microrrays correspondentes a uma concentração diferente. .... 116

**Figura 7B.11**- Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do *R. Flavefaciens* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no microarray. Uma coesina tem 4 posições no microrrays correspondentes a uma concentração diferente. .... 117



## Índice de tabelas

<b>Tabela 3.1</b> - Condição de expressão selecionada para cada complexo Coh-Doc. ....	30
<b>Tabela 3.2</b> - Concentrações utilizadas nos ensaios de cristalização .....	32
<b>Tabela 3.3</b> - Resultado mais promissor de expressão para cada complexo em estudo. ....	35
<b>Tabela 3.4</b> - Soluções tampão utilizadas nos vários protocolos de purificação. ....	37
<b>Tabela 3.5</b> - Percentagens de agente precipitante e de glicerol utilizadas para a realização da crio-preservação dos cristais. ....	42
<b>Tabela 3.6</b> -Estatísticas de recolha de dados para o cristal do complexo Coh-Doc C4X10B. ....	46
<b>Tabela 3.7</b> -Condições escolhidas para a experiência de difração e respectivos concentrações de agente precipitante e percentagens de glicerol utilizadas para a realização da crio-preservação dos cristais. ....	51
<b>Tabela 3.8</b> - Tabela resumo dos programas e respetivos resultados obtidos na análise de cada conjunto de dados. ....	53
<b>Tabela 3.9</b> - Ligações por pontes de hidrogénio entre o módulo Xyn10B e os módulos C2 e C4 .....	56
<b>Tabela 3.10</b> - Ligações por pontes de hidrogénio entre os módulos C9. ....	57
<b>Tabela 4.1</b> -Concentrações calculadas após realização dos protocolos (1), (2) e (3). ....	70
<b>Tabela 4.2</b> - Resumo dos resultados obtidos nos ensaios de ligação e de competição entre pares Coh-Doc realizados por eletroforese em, condições nativas. ....	80
<b>Tabela 7A.1</b> - Composição dos meios de expressão utilizados nos testes de expressão em <i>E. coli</i> . Os meios foram preparados com água destilada. Foram de seguida autoclavados a 120°C durante 20 minutos. ....	95
<b>Tabela 7B.1</b> - Informação sobre os módulos coesina em estudo do <i>C. thermocellum</i> . A tabela apresenta o número de resíduos de aminoácidos (aa) presentes na sequência, o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta Expasy .....	106
<b>Tabela 7B.2</b> - Informação sobre os módulos doquerina em estudo do <i>C. thermocellum</i> . A tabela apresenta o número de resíduos de aminoácidos presentes na sequência, o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta Expasy. ....	107
<b>Tabela 7B.3</b> - Informação sobre os módulos coesina em estudo do <i>R. flavefaciens</i> . A tabela apresenta o número de resíduos de aminoácidos presentes na sequência o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta Expasy. ....	108
<b>Tabela 7B.4</b> - Informação sobre os módulos doquerina em estudo do <i>R. flavefaciens</i> . A tabela apresenta o número de resíduos de aminoácidos presentes na sequência, o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta Expasy. ....	108



## Lista de abreviaturas, siglas/acrónimos e símbolos

2xYP - Meio *2x Yeast extract and peptone*

% - Percentagem

% (V/V) – Percentagem volume/volume

Å – Angstrom

Ala – Alanina

Amp – Ampicilina

Asp –Ácido aspártico

Asn - Asparagina

BSA - albumina sérica bovina

*C. thermocellum* – *Clostridium thermocellum*

C1X10B – Módulo coesina C1 da CipA, complexada ao módulo doquerina Xyn10B

C2-Xyn10B - Módulo coesina C2 da CipA, complexada ao módulo doquerina Xyn10B

C3X10B – Módulo coesina C3 da CipA, complexada ao módulo doquerina Xyn10B

C4X10B – Módulo coesina C4 da CipA, complexada ao módulo doquerina Xyn10B

C9X10B – Módulo coesina C9 da CipA, complexada ao módulo doquerina Xyn10B

Ca<sup>2+</sup> - íão cálcio

CaCl<sub>2</sub> – Cloreto de cálcio

CBM – Módulo de ligação a hidratos de carbono

CBM3a – Módulo de ligação à hidratos de carbono da família 3, grupo a

CBP - *Consolidated BioProcessing*

CipA - *Cellulosome Integrating Protein*

Coh – Coesina

Coh1 – Coesina 1

Coh2 – Coesina 2

Coh-Doc – Complexo coesina-doquerina

Crio-EM - Microscopia Electrónica a Temperatura Criogénica

D.O.<sub>600nm</sub> – Densidade ótica a 600 nm

Da – Dalton

Doc – Doquerina

Doc1 – Doquerina 1

Doc2 – Doquerina 2

DNA – Ácido desoxirribonucleico

*E.coli* - *Escherichia coli*

ESRF - *European Synchrotron Radiation Facility*

Glu – Glutamato  
Gly - Glicina  
HEPES - Ácido N-(2-hidroxietil)piperazina-N-etanosulfónico  
HIS - Histidina  
IMAC - Cromatografia de afinidade por iões metálicos imobilizados  
IPTG - Isopropil-β-D-1-tiogalactopiranosido  
K - Potássio  
KAN - Canamicina  
LB – Meio Luria-Bertani  
Leu - Leucina  
MAD – *Multiple Wavelength Anomalous Dispersion*  
MgCl<sub>2</sub> - Cloreto de magnésio  
MIR - *Multiple Isomorphous Replacement*  
Mpa – Megapascal  
MR – *Molecular Replacement*  
(NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> – Sulfato de amónio  
Na - Sódio  
NaCl – Cloreto de sódio  
OlpA - *Outer layer protein A*  
OlpB - *Outer layer protein B*  
OlpC - *Outer layer protein C*  
Orf2p - *Open reading frame protein 2*  
o/n - overnight  
PEG - Polietilenoglicol  
PDB - *Protein Data Bank*  
*R. flavefaciens* – *Ruminococcus flavefaciens*  
RMN - Ressonância Magnética Nuclear  
RMSD - Root-mean-square deviation  
RNA – Ácido ribonucleico  
Rpm – Rotações por minuto  
s/i – Sem informação  
SAD - *Single Wavelength Anomalous Dispersion*  
SdbA - *S-layer glycoprotein A*  
Ser – Serina  
SOLEIL - *Source optimisée de lumière d'énergie intermédiaire du LURE*  
TB – Meio *Terrific Broth*  
TEM – Microscopia de transmissão eletrónica

TEV - Tobacco Etch Virus

Thr – Treonina

TRIS - 2-Amino-2-hydroxymethyl-propane-1,3-diol

TRX - tioredoxina

Tyr - Tirosina

Val - Valina

VM – Volume morto



Capítulo

1

**Introdução**

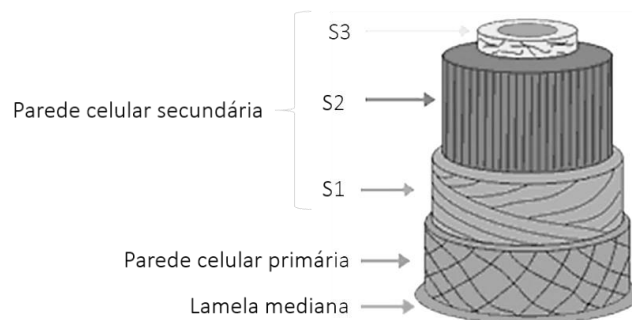


## 1. Introdução

### 1.1. Parede Celular Vegetal

No reino vegetal, a parede celular tem uma variedade de funções. Além de funcionar como exoesqueleto protegendo e conferindo resistência às células vegetais, é também flexível o suficiente para resistir ao rompimento celular e permeável para permitir o transporte intercelular. As células vegetais possuem dois tipos de parede celular com diferente função e composição, a parede celular primária e a secundária (Figura 1.1).

A parede celular primária proporciona resistência mecânica e elasticidade celular, sendo constituída por microfibrilas de celulose cristalina que, por sua vez, se encontram numa matriz formada de polissacáridos não celulósicos, tais como, hemiceluloses, pectinas e glicoproteínas.<sup>[1]</sup> A parede celular primária controla o crescimento da célula e forma a base estrutural da planta.<sup>[2]</sup> Após o crescimento celular a parede torna-se mais espessa pois forma-se, internamente à parede primária, uma segunda parede celular, a chamada parede celular secundária. Estruturalmente, esta rodeia o citoplasma e é composta por três camadas distintas (S1, S2 e S3). As camadas exterior, S1, e interior, S3, são constituídas por microfibrilas de celulose orientadas transversalmente, enquanto que a camada S2, camada intermédia, possui as mesmas microfibrilas de celulose orientadas longitudinalmente. Esta diferente orientação endurece as paredes celulares, prevenindo a degradação bioquímica e os danos físicos de que podem ser alvo.<sup>[3]</sup> Além da celulose, componente principal das paredes vegetais<sup>[4]</sup>, pode ainda ser encontrado na parede secundária, hemicelulose e lenhina. A parede celular vegetal representa assim a mais abundante fonte terrestre de carbono<sup>[5]</sup>.

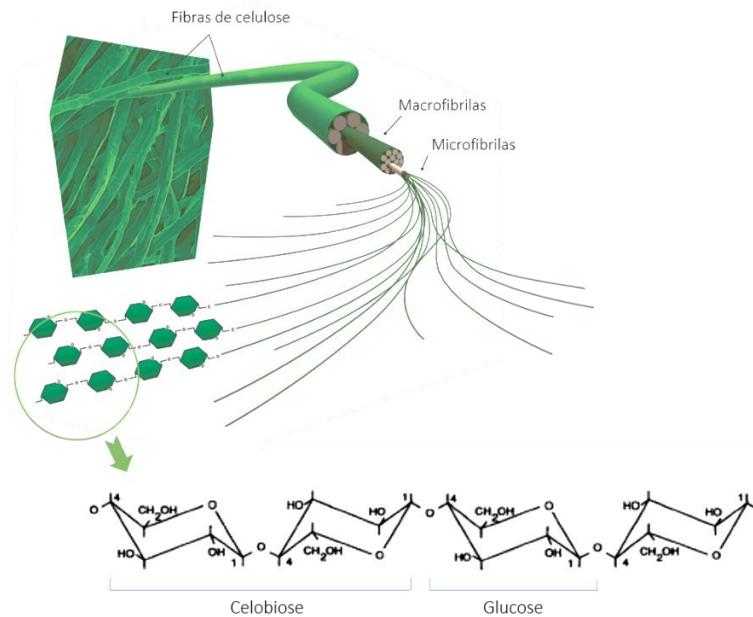


**Figura 1.1-** Esquema representativo da parede celular vegetal. Adaptado do Bioenergy Science Center (<http://bioenergy.crcr.uga.edu/>)

#### 1.1.1. Celulose

A celulose é o composto orgânico mais abundante na biosfera<sup>[6]</sup> e o principal componente das plantas. É constituída por cerca de 7000 a 15000 unidades de D-glucose, unidas entre si por ligações glicosídicas  $\beta$ -1,4.<sup>[7]</sup> É de ter em conta que cada molécula de glucose sofre uma rotação de  $180^\circ$  em relação à sua unidade vizinha, podendo-se também afirmar que a celulose é composta por unidades de celobiose (Figura 1.2).<sup>[8]</sup> Na parede celular, quer primária, quer secundária, as microfibrilas de

celulose são compostas por moléculas de celulose alinhadas paralelamente, orientação esta que é garantida por pontes de hidrogénio e interações de van der Waals. Como referido anteriormente, esta disposição confere uma alta resistência da parede celular vegetal ao ataque enzimático.<sup>[1]</sup> A associação de microfibrilas de celulose dá origem às macrofibrilas, que ao se agruparem formam as fibras de celulose, onde podemos encontrar moléculas de celulose numa forma cristalina, paracristalina ou amorfa. As regiões cristalinas são as que conferem uma forte coesão e estabilidade, tornando as fibras de celulose impermeáveis à água e mais resistentes à degradação química e biológica.<sup>[9]</sup>



**Figura 1.2** - Estrutura da celulose. Adaptado de Béguin et Lemaire, 1996.

### 1.1.2. Degradação enzimática da parede celular vegetal

Tal como referido na secção anterior, a celulose é o principal componente das plantas, proporcionando forma e resistência à parede celular. A complexidade física da parede leva a que a sua acessibilidade por parte de enzimas seja reduzida, tornando assim a sua degradação num processo complexo. Na natureza, existe uma grande diversidade de microrganismos que permitem a degradação dos polissacáridos vegetais. Estes microrganismos têm a particularidade de serem capazes de produzir enzimas com homologia entre si, que vão converter os componentes da parede celular nos seus constituintes mais simples, os açúcares, utilizando estratégias distintas.<sup>[10]</sup>

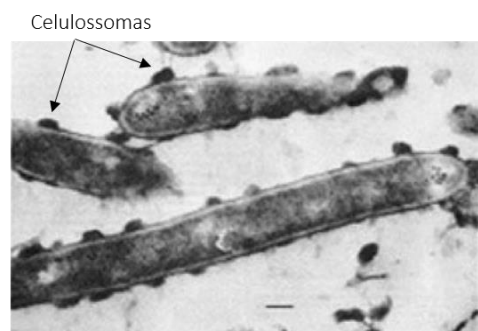
De entre estes microrganismos, encontramos os microrganismos aeróbios: bactérias dos géneros *Bacillus*, *Micromonospora*, *Cellvibrio* e *Pseudomonas* <sup>[11]</sup>, e fungos dos géneros *Trichoderma* e *Aspergillus* <sup>[8][12]</sup>; mas também organismos anaeróbios, tais como, as bactérias dos géneros *Clostridium* <sup>[13]</sup>, *Ruminococcus* <sup>[14]</sup>, *Thermotoga* <sup>[15]</sup> e os fungos dos géneros *Neocallimastix* <sup>[16]</sup> e *Piromyces* <sup>[17]</sup>. Enquanto que os microrganismos anaeróbios possuem sistemas integrados, os microrganismos aeróbios possuem, por sua vez, sistemas não integrados.<sup>[18]</sup> Em género de exemplo, os organismos

aeróbios secretam enzimas para o meio extracelular que vão atuar individualmente na degradação da parede celular, enquanto que organismos anaeróbios utilizam complexos multi-enzimáticos altamente eficientes para esse feito.

Nos últimos anos, o microrganismo anaeróbio *Clostridium thermocellum* (*C.thermocellum*), envolvido na degradação da parede celular vegetal, tem sido alvo de grande interesse pois possui um complexo modular multi-enzimático com alta capacidade em degradar eficientemente a celulose e outros componentes da parede celular vegetal.<sup>[19]</sup> A escassez de combustíveis fósseis tem impulsionado o estudo de outras alternativas, tais como a produção de biocombustíveis. Esta procura deve ter em conta três aspetos, a fim de ser considerado como viável. Deve ter o potencial para satisfazer parte das necessidades energéticas mundiais, deve ser capaz de reduzir o impacto ambiental da escassez dos combustíveis fósseis, devendo, por último, possuir um custo competitivo.<sup>[20]</sup> Uma dessas alternativas tem passado pela utilização da celulose na produção de biocombustíveis e de compostos químicos de alto-valor em processos de *Consolidated BioProcessing* (CBP). Nestes processos, a biomassa linhocelulósica é processada diretamente em produtos como o etanol. Os CBP são processos económicos e viáveis que combinam a produção enzimática, a hidrólise e os passos de fermentação apenas numa etapa, diminuindo assim os custos de produção.

## 1.2. Descrição do microrganismo *Clostridium thermocellum*

O *Clostridium thermocellum* é um organismo anaeróbio, gram-positivo e termófilo com um crescimento ótimo a 55°C, capaz de degradar a celulose com elevada eficiência.<sup>[20]</sup> Em termos morfológicos, tem a forma de bastonete, possuindo na sua superfície um complexo modular multi-enzimático. Este complexo de tamanho *megaDalton* é chamado de celulosoma e a sua organização modular ocorre após o reconhecimento de substratos como a celulose. A presença destes celulosomas na superfície pode ser visualizada por microscopia eletrónica (Figura 1.3.).



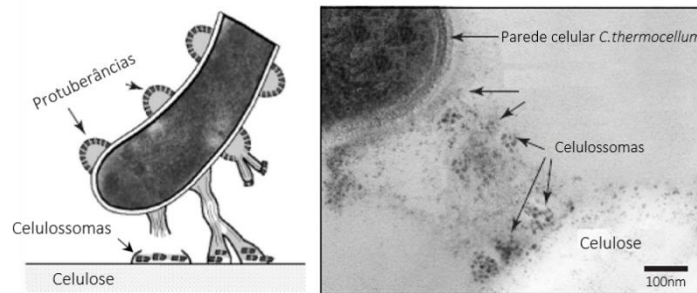
**Figura 1.3** – Superfície do *C.thermocellum*. Adaptado de Bayer e Lamed, 1986.

Em comparação com outros organismos anaeróbios, o *C.thermocellum* apresenta uma das maiores taxas de utilização de celulose, sendo que, por exemplo, a eficiência do seu celulosoma é 50 vezes mais elevada do que o sistema do organismo *Trichoderma*.<sup>[21]</sup>

## 1.3. Celulosoma do *Clostridium thermocellum*: Arquitetura e Função

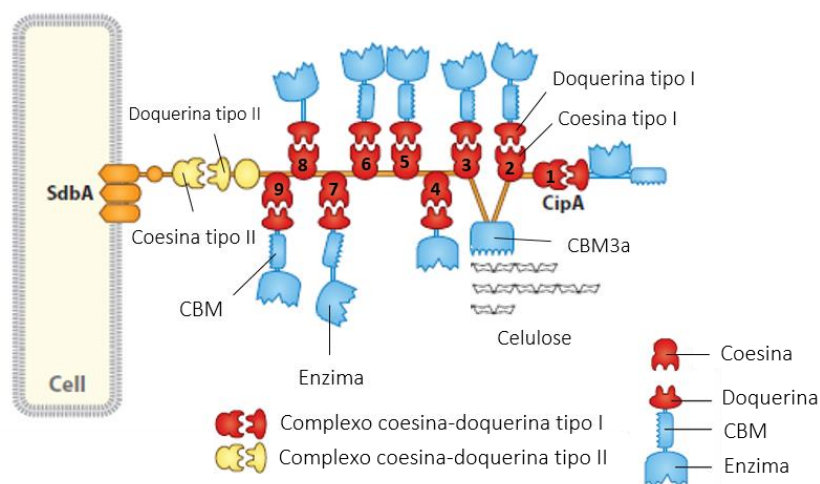
Os celulosomas são complexos multi-enzimáticos com peso molecular superior a  $2 \times 10^6$  Da <sup>[22]</sup> e de alta eficiência, que estão localizados em pequenas protuberâncias da superfície celular do *C.thermocellum*, tendo sido observados pela primeira vez em 1986, por Bayer e Lamed.<sup>[23]</sup> Ligados à superfície da bactéria, possuem flexibilidade suficiente para se ligarem à celulose da parede celular

vegetal. Isto é, após a ligação da celulose aos celulosomas, estes organelos sofrem uma alteração conformacional formando fibras alongadas entre o substrato e a superfície da célula bacteriana (Figura 1.4.).<sup>[24]</sup>



**Figura 1.4** – Estrutura da superfície celular do *C. thermocellum* e visualização da mesma por microscopia eletrônica – Microscopia de transmissão eletrônica (TEM). Adaptado de Sholman et al., 1999.

O principal componente do celulosoma do *C. thermocellum* é denominado de *Cellulosome Integrating Protein* (CipA) ou *Scaffolding protein* e possui uma elevada massa molecular (200 kDa). É responsável pela ligação do celulosoma à celulose, tal como pela ancoragem do mesmo à parede bacteriana, através de um módulo não catalítico chamado doquerina tipo II. De um modo geral, a CipA garante a organização celulosomal.<sup>[25][26]</sup> Além da doquerina tipo II, tem na sua composição outros nove módulos não-catalíticos, chamados coesinas tipo I e um módulo de ligação a hidratos de carbono (CBM3a) - Figura 1.5. A ligação das enzimas necessárias à degradação da celulose, dependentes de  $Ca^{2+}$ , à CipA é realizada por intermédio de proteínas modulares das quais fazem parte os módulos não-catalíticos designados doquerinas tipo I.<sup>[27]</sup> A interação entre os módulos coesina-doquerina potencia a estabilidade e a organização do celulosoma.<sup>[28]</sup>



**Figura 1.5** – Representação esquemática da interação do celulosoma do *C. thermocellum* com a celulose, e da interação celulosoma-parede bacteriana. Adaptado de Fontes e Gilbert, 2010.

### 1.3.1. Proteínas do celulosoma do *Clostridium thermocellum*

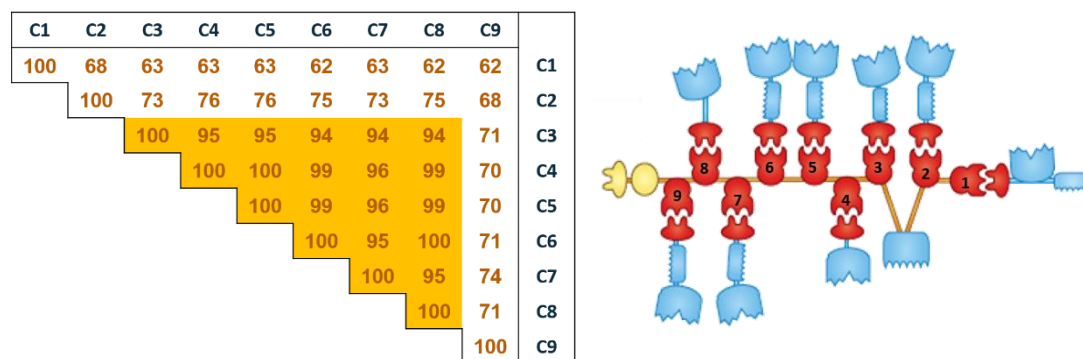
#### 1.3.1.1. Cellulosome Integrating Protein (CipA) e proteínas de ancoragem

A CipA do *C.thermocellum* faz a ancoragem do celulosoma à parede celular deste organismo através da interação coesina-doquerina do tipo II. O módulo doquerina tipo II da CipA interage com o módulo coesina tipo II presente na proteína de ancoragem, promovendo a ligação do celulosoma à superfície bacteriana. Até à data são conhecidos quatro tipos de proteínas de ancoragem com coesinas do tipo II. Cada um destes tipos de proteínas de ancoragem difere no número destes módulos (1, 2 e 7) pelas quais são compostas. Com um módulo temos a proteína SdbA (*Scaffoldin dockerin binding protein A*), com dois módulos temos a Orf2p (*Open reading frame 2 protein*), com sete módulos temos a OlpB (*Outer layer protein B*) e a proteína de ancoragem Cthe\_0736.<sup>[18]</sup> Esta última proteína de ancoragem parece permitir a formação de um complexo livre, sem a ligação à parede bacteriana. É ainda importante mencionar que cada uma destas coesinas tipo II promove a ligação de uma molécula CipA e que este complexo livre pode acomodar até a um número total de sete destas moléculas (CipA).

Estão também identificadas outras duas proteínas de ancoragem, OlpA (*Outer layer protein A*) e OlpC (*Outer layer protein C*), que possuem em vez de módulos coesina tipo II módulos coesina tipo I, permitindo a ligação direta das proteínas modulares à superfície bacteriana através da interação coesina-doquerina do tipo I.<sup>[28][18]</sup>

#### 1.3.1.2. Interação coesina-doquerina

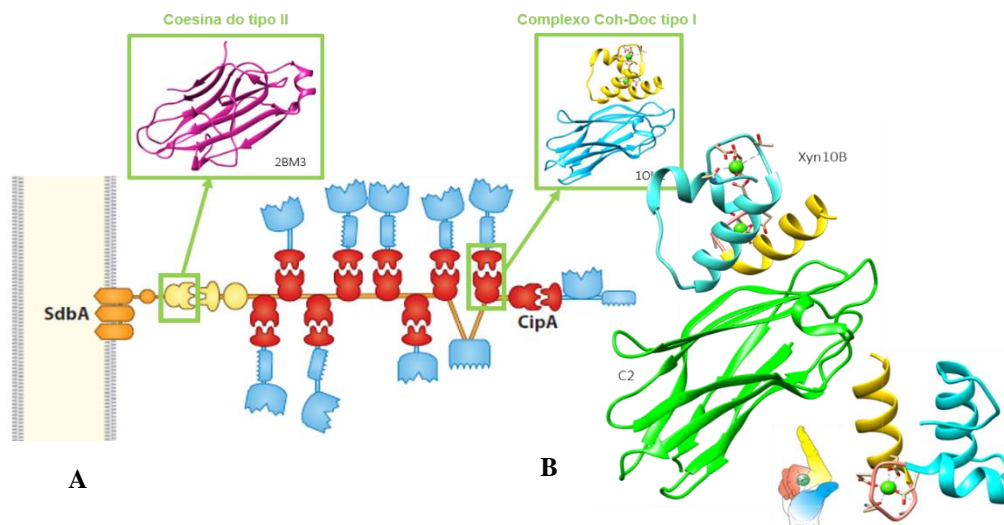
Os nove módulos coesina presentes na CipA, classificadas como coesinas tipo I de 1 a 9, possuem cerca de 140 resíduos de aminoácidos. Na Figura 1.6. estão apresentadas as percentagens de identidade, baseadas na homologia das sequências, entre os nove módulos da *scaffolding protein*. Esta classificação permite demonstrar que os módulos coesina 3 a 8 são aqueles que possuem maior identidade (superior a 94%) e que os módulos coesina 1, 2 e 9 são aqueles que possuem menor identidade com todos os outros, acreditando que a disposição dos vários módulos ao longo da CipA poderá influenciar a estrutura dos mesmos e estar relacionado com a diferença de afinidade que apresentam a certas doquerinas.



**Figura 1.6**– Representação esquemática da percentagem de identidade entre as coesinas tipo I da CipA, do *C.thermocellum*. Adaptado Fontes e Gilbert, 2010.

Estes módulos coesina interagem com outros módulos não-catalíticos denominados de módulos doquerina tipo I. Os módulos doquerinas tipo I são compostos por cerca de 60 a 70 aminoácidos<sup>[8]</sup> sendo que parte da sequência tem uma estrutura semelhante à denominada de *EF-hand*, na qual os resíduos de ácido aspártico (Asp) e asparagina (Asn) se encontram conservados, sendo responsáveis pela coordenação do íon  $\text{Ca}^{2+}$ . Este íon promove o aumento da atividade do celulosoma na degradação dos componentes da parede celular.<sup>[29]</sup> São conhecidos cerca de 70 módulos doquerina pertencentes ao celulosoma do *C. thermocellum*.<sup>[30]</sup>

A interação coesina-doquerina, quer do tipo I, quer do tipo II, é crucial para a montagem e organização do celulosoma e conseqüentemente, para a conversão de biomassa em açúcares simples. Esta interação coesina-doquerina pode dizer-se específica pois, coesinas do tipo I ou tipo II apenas reconhecem doquerinas do mesmo tipo, e vice-versa, sendo que esta interação só acontece entre coesinas e doquerinas da mesma espécie. A elucidação das estruturas tridimensionais para os vários complexos coesina-doquerina do tipo I, da CipA, tem permitido avaliar o mecanismo de montagem do celulosoma. Até à data, a Cristalografia de Raios-X conduziu à determinação de algumas estruturas deste celulosoma. São dois exemplos o complexo C2-Xyn10B – complexo Coh-Doc tipo I - e o módulo coesina do tipo II associada à proteína de ancoragem SdbA (Figura 1.7.).<sup>[31]</sup>



**Figura 1.7**– A - Representação esquemática a nível atômico das estruturas determinadas por Cristalografia de Raios-x. Adaptado de Fontes e Gilbert, 2010. B - Estrutura obtida por Cristalografia de Raios-X do complexo C2-Xyn10B e representação da conformação em EF-hand presente na doquerina.

A estrutura do complexo C2-Xyn10B, exemplo de uma estrutura de um complexo Coh-Doc do tipo I e de grande interesse para este trabalho, permitiu determinar quais as interações entre o módulo coesina C2 da CipA e o módulo doquerina Xyn10B, tal como avaliar a disposição estrutural dos dois módulos. O módulo C2 está organizado numa estrutura em barril  $\beta$  composta por nove cadeias  $\beta$  organizadas em *jellyroll* e de uma hélice  $\alpha$  entre as folhas  $\beta$ 4 e  $\beta$ 5. Já o módulo doquerina Xyn10B possui 2 átomos de  $\text{Ca}^{2+}$ , coordenados pelos resíduos pertencentes à estrutura *EF-hand*, bem como 3

hélices  $\alpha$ . Foi também determinado que a interação coesina-doquerina é realizada por meio de pontes de hidrogénio entre o módulo coesina e uma das hélices  $\alpha$  do módulo doquerina Xyn10B. Estudos de mutagénesse permitiram determinar que a interação do módulo coesina Xyn10B é feita através de uma das duas superfícies de ligação – hélices  $\alpha$  1 e 3, nas quais encontramos os resíduos Ser11-Thr12 e Ser45-Thr46, respetivamente. O fato da interação Coh-Doc poder ser feita por uma destas duas hélices faz com que o módulo Xyn10B possa ter duas orientações equivalentes.<sup>[32]</sup>

Esta primeira elucidação sobre as interações entre os módulos coesina-doquerina da CipA, por Cristalografia de Raios-X, permitiu iniciar a caracterização dos complexos Coh-Doc do tipo I, a nível atómico. A combinação de outras metodologias como os *microarrays* de proteínas poderá permitir a caracterização das diferentes forças de ligação e especificidades existente entre coesinas e doquerinas.

### 1.3.1.3. Módulos de ligação a hidratos de carbono

Tal como referido na secção 1.3., a ligação das enzimas necessárias à degradação da celulose à CipA é realizada por intermédio de módulos chamados doquerinas tipo I. Porém, além destes módulos, as enzimas que atuam na degradação dos componentes da parede celular vegetal, possuem os chamados CBMs – módulos de ligação a hidratos de carbono. Por definição, um CBM é uma sequência contínua de 30 a 200 aminoácidos que está associada a uma enzima que tem por função degradar hidratos de carbono. Estes CBMs reconhecem o substrato a degradar, permitindo a sua ligação à enzima.<sup>[33]</sup>

Até à data estão identificadas 71 famílias de CBM, que podem ser consultadas na base de dados CAZy ([www.cazy.org/carbohydrates-binding-modules.html](http://www.cazy.org/carbohydrates-binding-modules.html), último acesso a 24 de Agosto de 2015). A nomenclatura adotada para a referência de uma família de CBM permite que seja incluída a função e especificidade da mesma. Temos como exemplo o CBM3a, encontrado na CipA (Figura 1.5.).<sup>[34]</sup> Por definição este módulo de ligação a hidratos de carbono pertence à família 3a. O número 3 é referente à família 3 de entre as 71 conhecidas e informa-nos que tem como função o reconhecimento da celulose. Por outro lado, a letra “a” que segue o número 3 permite dizer-nos que estamos perante um CBM que interage com polissacáridos cristalinos insolúveis, caso da celulose. De uma forma geral, os CBMs podem agrupar-se em 3 grupos (a, b ou c), consoante a sua mecânica de interação com polissacáridos-alvo. Os grupos b e c classificam os CBMs por se ligarem a cadeias individuais de polissacáridos solúveis ou insolúveis, e por se ligarem a pequenos açúcares, respetivamente.<sup>[33]</sup>

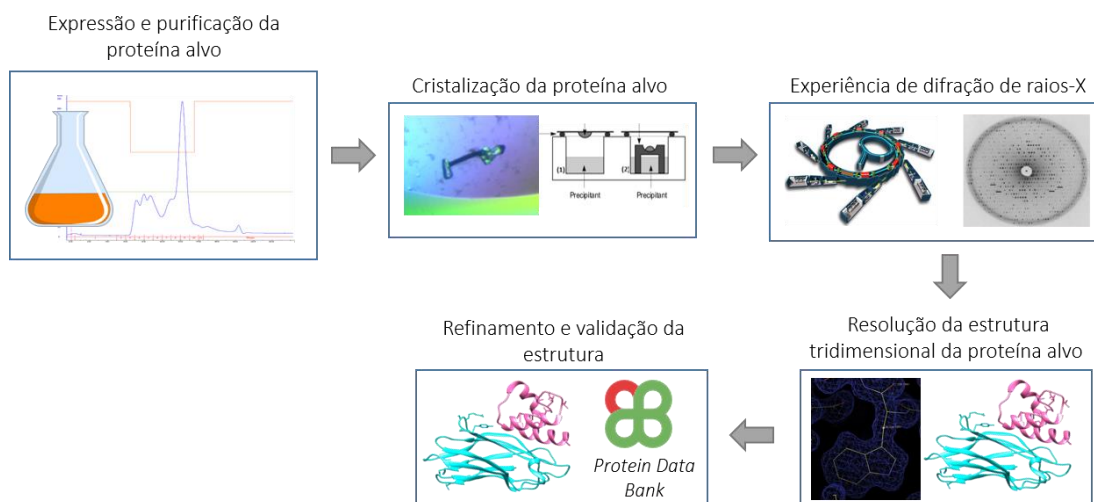
## 1.3. Cristalografia de Raios-X

A resolução da estrutura tridimensional de uma proteína, através da Cristalografia de Raios-X, revela-se essencial na obtenção de informações estruturais, a nível atómico, e na compreensão da sua função biológica. Os avanços científicos têm, no entanto, sugerido outras metodologias que também o permitem, tais como a Ressonância Magnética Nuclear (RMN) e a Microscopia Eletrónica a Temperatura Criogénica (Crio-EM). Muitas vezes de uma forma complementar e integrada, as três

técnicas têm contribuído para o número total de estruturas depositadas na base de dados *Protein Data Bank* (PDB). Até 25 de Agosto de 2015, das 111 385 estruturas depositadas no PDB, 99 220 foram resolvidas por Cristalografia de Raios-X, originando uma percentagem de 89,1% das estruturas totais determinadas, mostrando que é, a metodologia de Cristalografia de Raios-X a mais aplicada (<http://www.ebi.ac.uk/pdbe/>)

A aplicação desta metodologia a moléculas biológicas começou com a resolução das estruturas tridimensionais do colesterol (1937), da vitamina B12 (1945) e da penicilina (1954), por Dorothy Hodgkin. A determinação destas importantes estruturas valeram-lhe o Prémio Nobel da Química, em 1964.<sup>[35]</sup> Em 1958, foram John Kendrew e Max Perutz, que contribuíram para a resolução da estrutura da primeira proteína, a mioglobina de cachalote.<sup>[36]</sup> Já nos últimos anos, o desenvolvimento e aperfeiçoamento dos programas empregues, desde a recolha de dados de difração até à validação de uma estrutura tridimensional, permitiu o aumento da utilização desta metodologia e, consequentemente, o aumento do número de estruturas depositadas do PDB.

A determinação da estrutura tridimensional de uma proteína, por Cristalografia de Raios-X, envolve várias etapas (Figura 1.8): 1) expressão e purificação da proteína alvo; 2) cristalização; 3) experiência de difração de raios-X e resolução da estrutura tridimensional; 4) refinamento e validação da estrutura. Algumas destas etapas serão abordadas em detalhe ao longo das próximas secções da introdução.

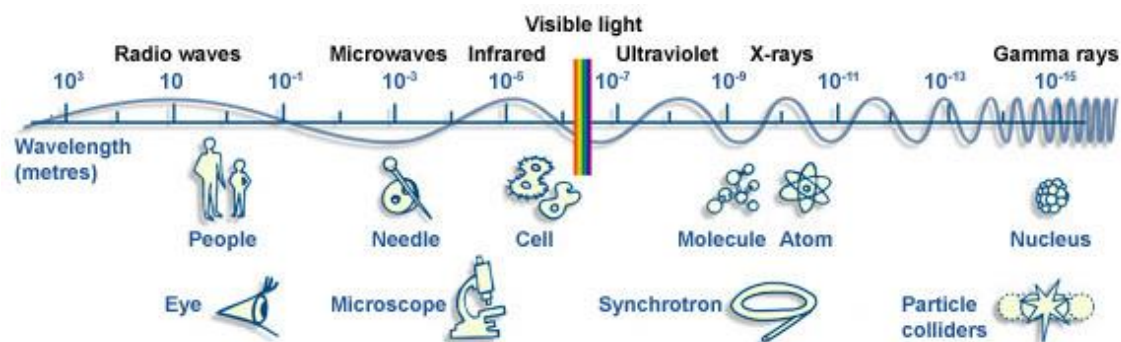


**Figura 1.8-** Representação esquemática das principais etapas para a obtenção de uma estrutura tridimensional por cristalografia de Raios-X.

### 1.3.1. A radiação de Raios-X

Em 1895, o físico alemão Wilhelm Conrad Röntgen fez a descoberta dos raios-X, sendo assim denominados pois a sua natureza era desconhecida. Atualmente esta radiação é caracterizada como sendo radiação eletromagnética, tal como a luz visível, mas de comprimento de onda entre os 0,1 e os

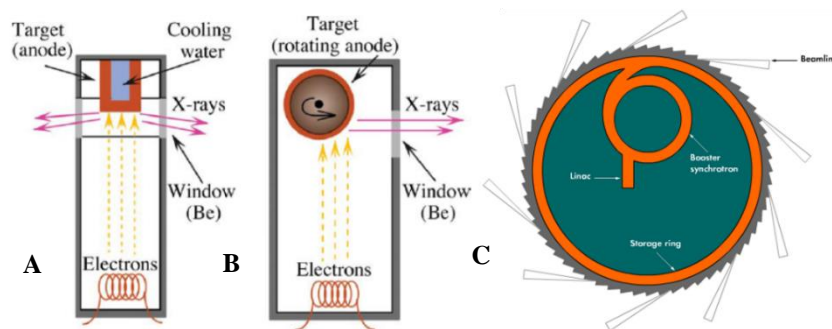
100 Å (Figura 1.9), encontrando-se no espectro eletromagnético entre os raios ultravioletas e os raios gama. Esta é também a zona em que o comprimento de onda da radiação se assemelha aos comprimentos de ligação entre os átomos, o que torna os raios-X adequados para a caracterização estrutural ao nível atômico.



**Figura 1.9** - Espectro eletromagnético. Os raios-X encontram-se entre os 0,1 e 100Å ( $1\text{Å}=10^{-10}$  m). (<http://www.esrf.eu/about/synchrotron-science/synchrotron-light>)

Para fins científicos, os raios-X podem ser produzidos por três fontes distintas: o tubo de *Coolidge* (também conhecido como ampola de raios-X), o ânodo rotativo e as infraestruturas de sincrotrão (Figura 1.10). Os dois primeiros baseiam-se num princípio de funcionamento muito semelhante, sendo o segundo uma versão melhorada do primeiro. Este baseia-se no bombardeamento de um metal (cobre, no caso específico da cristalografia de proteínas) <sup>[37]</sup> com elétrões, produzidos por um filamento aquecido, normalmente de tungsténio ou molibdénio, e acelerados por um campo elétrico. Os elétrões acelerados removem elétrões do metal de um dos níveis energéticos internos, produzindo-se radiação X quando outros elétrões transitam de uma orbital externa para uma orbital interna livre. Os raios-X são emitidos em todas as direções possuindo os comprimentos de onda correspondentes às diferentes transições entre orbitais. A aplicação de um monocromador permite a seleção de um comprimento de onda de interesse. No caso das infraestruturas de radiação de sincrotrão, os elétrões são acelerados até atingirem altas velocidades e injetados num grande anel. Neste movimento circular induzido por magnetos poderosos, os elétrões perdem energia sob a forma de radiação X. Estes raios-X são depois direcionados por sistemas de espelhos e monocromadores, originando feixes de raios-X monocromáticos, com diferentes comprimentos de onda e de alta energia.

[38]



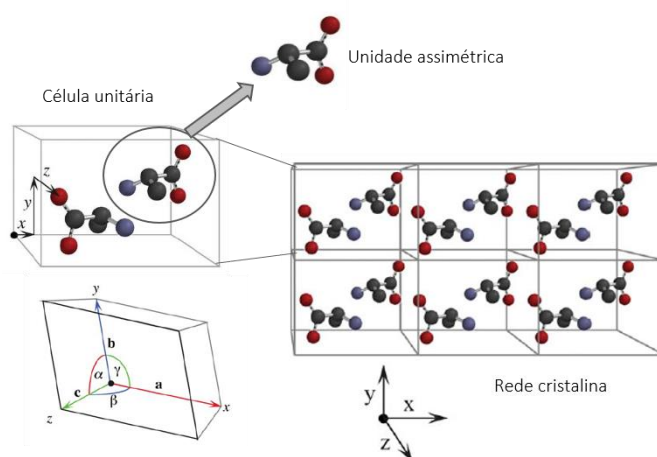
**Figura 1.10** – Representação esquemática de 3 fontes de raios-X distintas: A – Tubo de *Coolidge*; B – Ânodo rotativo; C – Sincrotrão. Adaptado de <http://www.esrf.eu/> e Rhodes, 2006.

### 1.3.2. Cristalização de proteínas

Um cristal de proteína é caracterizado por ser um sólido composto de átomos dispostos num padrão periódico em três dimensões. A célula unitária é o elemento mais simples e mais pequeno, que, por operações de translação, constitui a totalidade do cristal. Cada célula unitária é definida por seis parâmetros, sendo conhecidos como as “constantes da célula”: 3 vectores  $a$ ,  $b$  e  $c$  e 3 ângulos  $\alpha$ ,  $\beta$  e  $\gamma$  (Figura 1.11). Diferentes combinações destes parâmetros originam os chamados sistemas cristalinos.

Associado ao termo “célula unitária” encontra-se o termo “unidade assimétrica”. A unidade assimétrica define-se como a unidade volumétrica mais pequena que contém toda a informação estrutural. A célula unitária pode conter uma ou mais unidades assimétricas, que se relacionam entre si por operações de simetria.

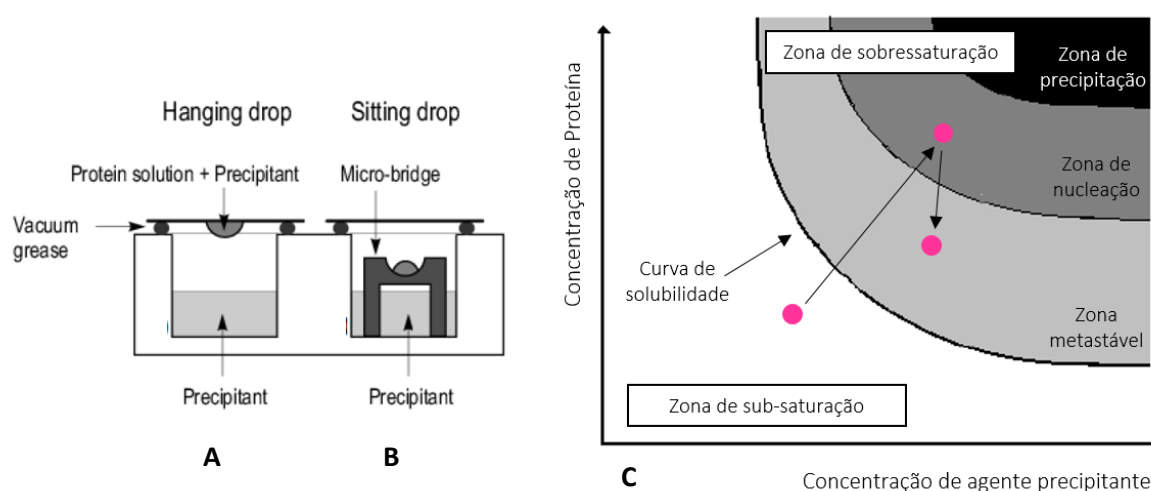
A rede cristalina é mantida por ligações não covalentes, nomeadamente pontes de hidrogénio, estabelecidas entre resíduos superficiais da proteína, podendo ser mediadas por moléculas de água. O teor de solvente deve estar compreendido entre os 30% e os 70% do volume total do cristal.



**Figura 1.11**-Representação esquemática da unidade assimétrica, da célula unitária e da rede cristalina. Adaptado de Rhodes, 2006.

É importante referir que a obtenção de cristais de proteína de qualidade, é fundamental para que os estudos por Cristalografia de Raios-X sejam realizados com sucesso.

Após a obtenção da proteína num elevado grau de pureza, a cristalização é induzida por uma solução que contém um agente precipitante. Esta indução é testada com diferentes agentes precipitantes (nos quais também há variação das suas concentrações e/ou proporções proteína-precipitante), a diferentes temperaturas, valores de pH e utilizando diferentes concentrações de proteína. As condições de cristalização são variáveis consoante a proteína em estudo, sendo por isso necessária a realização de vários ensaios iniciais onde se testam diversas condições experimentais, na tentativa de encontrar uma que nos dê cristais de proteína de qualidade. São conhecidos vários métodos de cristalização, sendo que o mais utilizado é o método de difusão de vapor. Este baseia-se na evaporação e na difusão das moléculas de água entre soluções de diferentes concentrações, como um meio para alcançar a sobressaturação das moléculas proteicas. Por outras palavras, o método de difusão de vapor consiste no estabelecimento de um equilíbrio, em sistema isolado, entre a gota, que contém a proteína em estudo e o agente precipitante, e o reservatório que contém apenas a solução com o agente precipitante. A diferença de concentração do agente precipitante na gota (mais baixa) e no reservatório (mais elevada) permite que ocorra a evaporação das moléculas de água, presentes na gota, para o reservatório. Esta difusão de água, sob a forma de vapor, tem por objetivo aproximar as concentrações de agente precipitante e, simultaneamente, concentrar a proteína na gota de uma forma lenta e controlada (induzindo a formação de cristais). O método de difusão de vapor inclui duas técnicas distintas: a gota suspensa e a gota assente. Na técnica de gota suspensa, a gota é colocada numa lamela silanizada, enquanto que na técnica de gota assente esta é colocada numa microponte, que por sua vez, se encontra dentro do reservatório da placa de cristalização (Figura 1.12-A e B).



**Figura 1.12** -Representação esquemática da técnica de difusão de vapor. Adaptado de Taylor *et al.*, 2003 e Hall *et al.*, 2007. A – método da gota suspensa; B – método da gota assente; C – diagrama de fases representando a variação das concentrações de proteína e agente precipitante durante o processo de cristalização de uma proteína.

A experiência de difusão de vapor pode ser explicada pela análise do conhecido “diagrama de fases da cristalização” (Figura 1.12-C). A cristalização de uma proteína ocorre em duas etapas: nucleação e crescimento. Aquando do início de um ensaio de cristalização, a proteína presente na gota encontra-se num estado não saturado (zona de sub-saturação). A difusão das moléculas de água vai induzir o início da nucleação. Nesta etapa, os vários núcleos cristalinos são formados, permitindo a passagem da proteína para a zona de sobresaturação, onde se verifica a redução da concentração da mesma. O crescimento lento do cristal dá-se na zona metastável, pela adição lenta e ordenada de novas moléculas e formação da rede cristalina.

Uma vez formados cristais únicos, estes podem ser preservados em condições criogénicas, recorrendo-se a azoto líquido, na presença de agentes crio-protetores. Estes impedem a formação de gelo nos canais de solvente, protegendo os cristais, e minimizam a dispersão dos raios-X pelas moléculas de água. [38]

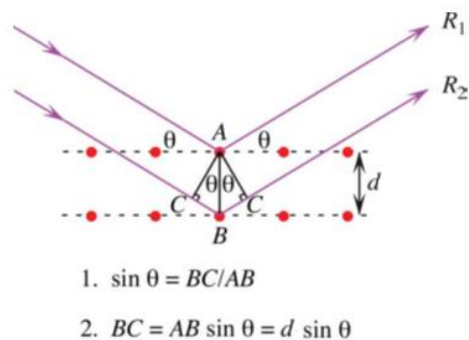
### 1.3.3. Difração de Raios-X e resolução da estrutura

Após a criopreservação dos cristais da proteína em estudo, seguem-se as etapas de difração de raios-X e resolução da estrutura. A experiência de difração consiste em incidir um feixe de fótons de raios-X no cristal obtido, montado num goniómetro acoplado a um equipamento de difração. O cristal difrata os fótons, provenientes da interação dos mesmos com os eletrões das moléculas de proteína, a diferentes ângulos, originando padrões de difração registados num detector. Este padrão é composto pelas reflexões correspondentes às ondas difratadas, dando-nos informações sobre a organização interna da rede cristalina do cristal.

A experiência de difração é normalmente efetuada a temperaturas criogénicas e é explicada através da Lei de Bragg (Equação 1):

$$n \lambda = 2 d \sin \theta \quad (\text{Equação 1})$$

onde,  $\theta$  é o ângulo de incidência,  $d$  é a distância interplanar,  $\lambda$  é o comprimento de onda da radiação incidente e  $n$  é um inteiro. A Lei de Bragg diz-nos que o conjunto de feixes de raios-X incidentes sobre o conjunto de planos da rede cristalina sofre uma interferência construtiva quando o ângulo incidente tem o mesmo valor que o ângulo do feixe refletido. Além disso, para que esta lei seja validada o espaçamento interplanar dos vários planos que fazem parte da rede cristalina deve ser igual a um múltiplo inteiro ( $n$ ) do comprimento de onda. A Figura 1.13 apresenta o diagrama esquemático utilizado para mostrar as relações angulares utilizadas na dedução de Bragg.



**Figura 1.13-**Representação da Lei de Bragg. <sup>[38]</sup>

Na figura acima, os pontos vermelhos representam dois planos paralelos da rede cristalina de um cristal com uma certa distância interplanar ( $d$ ). Os feixes de raios-X incidentes no conjunto de planos da rede cristalina vai dar origem aos feixes  $R_1$  e  $R_2$ , que são refletidos com o mesmo ângulo de incidência ( $\theta$ ). A este fenómeno designamos de interferência construtiva. Isto ocorre quando a distância percorrida pela radiação entre dois planos paralelos é igual a um número inteiro do seu comprimento de onda ( $n\lambda$ ). Por outro lado, aquando da reflexão dos feixes  $R_1$  e  $R_2$ , com ângulo de incidência e de reflexão distintos, a Lei de Bragg não se aplica e dizemos que estamos perante uma interferência destrutiva. Nestes casos não são produzidos reflexões e não se consegue prosseguir com a etapa posterior. É devido a este fenómeno que se deve rodar o cristal aquando desta etapa. <sup>[38][37]</sup>

A análise dos padrões de difração de um cristal permite determinar a intensidade e atribuir um índice a cada reflexão. Esta informação, juntamente com a estimativa dos ângulos de fase da radiação difratada e após cálculos específicos, permite determinar a posição dos átomos da proteína. Ao, contrário da intensidade de cada reflexão, o ângulo da fase não pode ser medido experimentalmente a partir do padrão, levantando-se o chamado “problema da fase” na Cristalografia de Raios-X.

A resolução deste problema pode ser conseguido recorrendo-se a diferentes metodologias que vão permitir estimar o ângulo da fase de cada reflexão. Entre eles encontramos o método de Substituição Molecular, do inglês *Molecular Replacement* (MR), a Substituição Isomórfica Múltipla, do inglês *Multiple Isomorphous Replacement* (MIR) e a Dispersão Anómala Múltipla ou Simples, do inglês, *Multiple/Single Wavelength Anomalous Dispersion* (MAD ou SAD). No âmbito deste trabalho foi utilizado o método de Substituição Molecular, pelo que os outros métodos não serão descritos.

O método de Substituição Molecular é utilizado quando são conhecidas as estruturas de proteínas semelhantes em termos de sequência primária. Uma sequência pode ser considerada semelhante (e um potencial modelo para o método da Substituição Molecular) se a identidade entre as sequências for superior a 25%. <sup>[39]</sup> Esta metodologia vai, numa primeira etapa, posicionar a proteína-modelo na célula unitária da estrutura desconhecida através de operações de rotação e de translação. A segunda etapa passa pelo cálculo dos fatores de estrutura, da molécula conhecida, posicionada na célula unitária da

estrutura desconhecida, a fim de se obterem as fases da molécula desconhecida. Estas fases estimadas são combinadas com as intensidades medidas experimentalmente e um mapa de densidade eletrónica pode ser finalmente calculado. Um mapa interpretável é indicativo da boa resolução da estrutura tridimensional da proteína desconhecida. No PBD, cerca de 85% das estruturas resolvidas por Cristalografia de Raios-X foram determinadas por Substituição Molecular (dado consultado em <http://www.rcsb.org/>). Os outros métodos são utilizados quando não é conhecida nenhuma estrutura potencialmente semelhante.

#### 1.3.4. Refinamento e validação

A última etapa passa pelo refinamento e validação da estrutura. O primeiro modelo de densidade eletrónica obtido deve ser melhorado até dar origem à estrutura final da proteína, por sucessivos processos de refinamento. Matematicamente falando, estes refinamentos promovem a minimização das diferenças entre as amplitudes observadas e calculadas. Em cada refinamento é calculado o fator R (Equação 2). Este fator avalia a qualidade do refinamento e espelha a concordância entre o modelo construído e os dados observados, devendo ser sempre o menor valor possível. Estruturas bem resolvidas apresentam um valor de R inferior a 20%.<sup>[38]</sup>

$$R = \frac{\sum |F_{obs}| - |F_{calc}|}{\sum |F_{obs}|} \quad (\text{Equação 2})$$

O fator R permite aferir a qualidade global da estrutura, porém existem outros fatores que permitem a análise dos dados recolhidos e a avaliação de locais específicos do modelo. A avaliação dos dados recolhidos pode ocorrer através de critérios como: valores elevados de  $I/\sigma(I)$ , *completeness* (superior a 90%), valores baixos de  $R_{merge}$  e multiplicidade elevada. Estes serão explicados na secção 3. Já os outros fatores que permitem avaliar locais específicos do modelo são: o fator B, também conhecido como fator de temperatura; o fator de ocupação; e o diagrama de Ramachandran.

O fator B relaciona a oscilação dinâmica de um átomo numa dada posição. Átomos com valores altos de fator B (em comparação com átomos vizinhos) são preocupantes pois indicam que podem estar numa posição incorreta – densidade eletrónica  $F_{obs}-F_{cal}$  negativa. É ainda de referir que átomos encontrados na superfície da proteína e em zonas com maior mobilidade possuem geralmente valores de B superiores aos encontrados no interior na mesma. Associado a este fator de temperatura encontra-se o fator de ocupação. Este outro fator varia entre 0 e 1 e representa a fração de moléculas no cristal em que um determinado átomo se encontra naquela posição. Aquando de uma posição correta este valor é 1. Se o valor for inferior a 1 poderá indicar que o átomo ocupa aquela posição apenas numa fração das células unitárias do cristal e, possivelmente, o resíduo de aminoácido (onde esse átomo se insere) possui uma conformação alternada. Por último, o diagrama de Ramachandran permite-nos confirmar a correta posição dos átomos. Este é caracterizado pelas suas regiões permitidas e proibidas, que estão relacionadas com os ângulos diedros *phi* ( $\phi$ ) e *psi* ( $\psi$ ) em torno do  $C\alpha$ . A incorreta orientação de um plano peptídico leva a que esse resíduo se encontre numa zona proibida,

dando-nos possibilidade de verificar, refinar e validar o modelo. Uma vez o modelo validado, são concluídos os sucessivos refinamentos e o modelo pode então ser depositado do PDB.

#### **1.4. Tecnologia dos *Microarrays***

No início dos anos oitenta, Augenlicht e os seus colaboradores propuseram, pela primeira vez, a ideia básica da tecnologia de *microarrays*, através da monitorização dos níveis de expressão de genes humanos marcados radioativamente.<sup>[40]</sup> Esta consistia na aplicação direta de uma mistura, contendo a biomolécula alvo, em forma de ponto (*spot*) sobre uma membrana, para posterior identificação, análise e quantificação. Desde dessa altura muitos têm sido os avanços relacionados com esta tecnologia. Nos últimos anos, a determinação da função e da regulação de sistemas celulares complexos, através da tecnologia de *microarrays* tem vindo a tornar-se uma das ferramentas essenciais no mundo da bioquímica e da biologia molecular. O número de estudos envolvendo a utilização de *microarrays* para a identificação de mecanismos celulares tem crescido exponencialmente, pois esta tecnologia permite a investigação de milhares de moléculas alvo em simultâneo.

Pode-se definir os *microarrays* como uma técnica que permite a análise simultânea de inúmeras biomoléculas (DNA, proteína, hidratos de carbono e metabolitos) numa única experiência, tornando esta tecnologia rápida e mais eficiente quando se têm um conjunto alargado de moléculas para estudar. A técnica baseia-se na imobilização ordenada de biomoléculas num suporte sólido (proteína, ligandos ou substratos), para posteriores estudos de interação com amostras biológicas de interesse. No âmbito deste trabalho foi utilizada a técnica de *microarrays* de proteína para a análise da interação proteína-proteína.

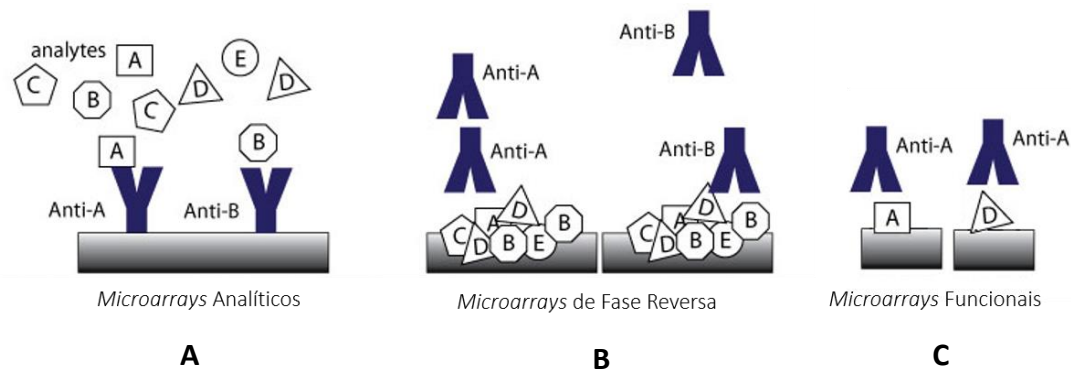
##### **1.4.1 *Microarrays* de proteína**

David Hall e os seus colaboradores<sup>[41]</sup> definem três tipos de *microarrays* de proteína (Figura 1.20):

- *Microarrays* Analíticos – usados para o estudo de uma mistura complexa de proteínas, utilizando um *microarray* de anticorpos. Existem dois tipos: um em que o analito está marcado, outro em que se faz um ensaio *sandwich* e se usa além do anticorpo imobilizado, o anticorpo de captura, um segundo anticorpo, anticorpo repórter, para detetar o analito. Ambos os ensaios são utilizados para observar diferenças em níveis de expressão de proteínas de entre várias amostras, mas com diferenças na especificidade, sensibilidade e reprodutibilidade. (Figura 1.14-A);

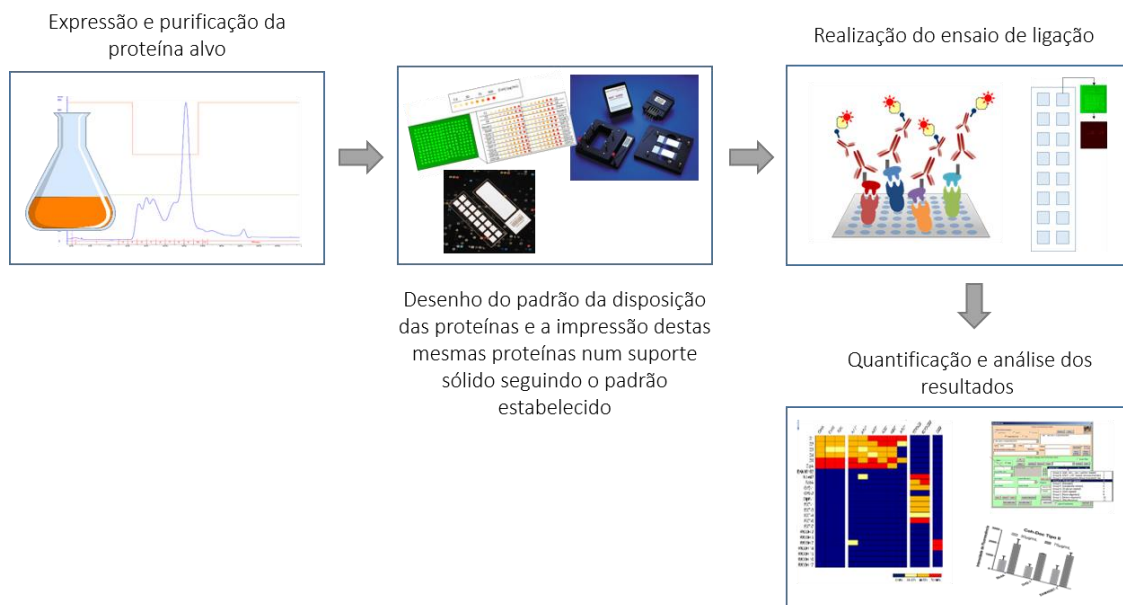
- *Microarrays* de Fase Reversa – usados no estudo de proteínas isoladas a partir da lise celular. O lisado celular é impresso e a deteção é realizada através de anticorpos (Figura 1.14-B).

- *Microarrays* Funcionais – usados no estudo da atividade e especificidade de uma ou mais proteínas numa única experiência, tendo como objetivo estudar várias interações em simultâneo (proteína-proteína; proteína-DNA, proteína-RNA; entre outros), através da detecção por anticorpos específicos (Figura 1.14-C);



**Figura 1.14**-Representação esquemática dos três tipos de *microarrays*: A – *Microarrays* Analíticos; B – *Microarrays* de Fase Reversa; C – *Microarrays* Funcionais. Adaptado de Sheehan et al, 2005. [43]

Os *microarrays* de proteína em estudo nesta dissertação podem ser incluídos no terceiro grupo, o dos *Microarrays Funcionais*, e envolvem 4 etapas gerais (Figura 1.15): 1) Expressão e purificação da(s) proteína(s) alvo; 2) Desenho do padrão da disposição das proteínas e a impressão destas mesmas proteínas num suporte sólido seguindo o padrão estabelecido; 3) Realização do ensaio de ligação; 4) Quantificação e análise dos resultados.



**Figura 1.15** - Representação esquemática das principais etapas envolvidas na técnica de *microarrays* de proteína.

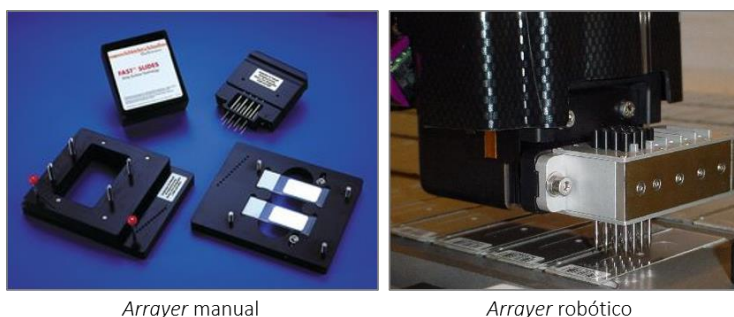
#### 1.4.1.1. Desenho do padrão e impressão das proteínas em suporte sólido

Após expressão e purificação das proteínas do estudo, os *microarrays* são preparados pela imobilização dessas moléculas num suporte sólido. Para isso, as proteínas a serem testadas, são pipetadas em pequenos volumes para o respectivo reservatório de uma microplaca de 384 poços. A disposição das amostras na microplaca é feita de forma a definir uma matriz de colunas e linhas, que constitui o modelo para a impressão de cada slide, e o padrão de cada *microarray*. A imobilização é posteriormente realizada através de um equipamento denominado de *arrayer*.

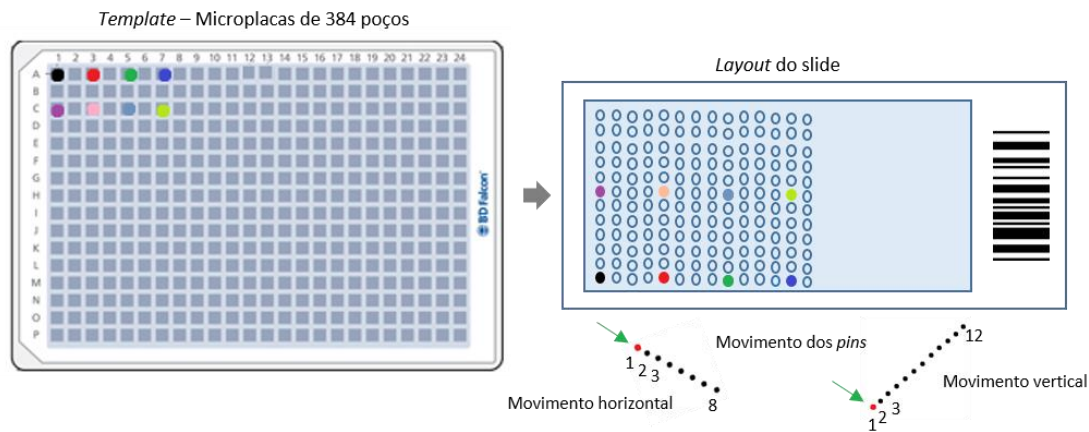
#### 1.4.1.2. Arrayers – manual e robótico

Os *arrayers* possibilitam a imobilização das proteínas em estudo através da impressão em suportes sólidos. Existem diferentes formatos deste equipamento em que os princípios da impressão da amostra (*printing*) variam, por exemplo: *arrayer* manual em que a deposição é por contato e o *arrayer* robótico em que impressão pode ser por contato ou sem contato.

No caso da utilização de um *arrayer* manual, os *pins* presentes no equipamento são mergulhados na microplaca, sendo as proteínas imobilizadas pelo contato desses *pins* com o suporte sólido. Estes *pins* têm a particularidade de serem planos ou terem a ponta côncava e permitem o alojamento de um pequeno volume de amostra a ser depositada no suporte sólido. O *arrayer* manual ilustrado na Figura 1.16 foi o utilizado durante este trabalho. É composto por duas fileiras de quatro *pins*, permitindo um total de 8 imobilizações a cada contato com o suporte. Este fator torna a impressão de um grande número de proteínas muito lenta, podendo ser influenciada pela possível evaporação mais rápida de alguns compostos ou pela perda de parte da fluorescência quando marcados com fluoróforos. É assim indispensável que o ambiente do espaço onde estes ensaios são realizados possua um certo teor de humidade. A deslocação do *arrayer* (movimento horizontal e vertical) sob o suporte sólido é efetuado seguindo a orientação dos dois *pins* mais à esquerda (Figura 1.17). O movimento horizontal é feito seguindo o deslocamento sucessivo de um desses *pins* num total de oito posições. Intercalado com este movimento horizontal, é feito também um movimento vertical onde se segue este mesmo *pin* deslocando-se ao longo de um total de 12 posições. O uso de um *arrayer* manual tem como vantagens a sua fácil implementação e o seu baixo custo.



**Figura 1.16**– Fotografias de um arrayer manual (esquerda) e robótico (direita). Adaptado de <http://www.arrayit.com>.



**Figura 1.17-** Ilustração do método de impressão de um slide de 1 pad, a partir do modelo criado na microplaca de 384 poços, e do respectivo arrayer manual utilizado para o efeito.

No caso da utilização de um *arrayer* robótico em que impressão é feita por não contato, o equipamento enche os *pins*, por capilaridade com as proteínas a imobilizar. Os *pins* mergulham na microplaca e emergem levando uma camada fina de proteína na ponta. O *robot* pressiona ligeiramente cada capilar que liberta uma pequena quantidade de proteína, que fica imobilizada no suporte sólido sem ocorrer o contato dos pins com a superfície. Após a primeira linha de marcas circulares, designadas de “*spots*”, ficar impressa, o *robot* move o seu braço para impressão da segunda linha. E assim sucessivamente. Um exemplo de *arrayer* robótico é ilustrado na Figura 1.16. A impressão utilizando o *robot*, orientada pelo modelo realizado na microplaca de 384 poços, tem a vantagem de originar *spots* com um diâmetro menor aos impressos manualmente, bem como imprimir um maior número de *spots* em simultâneo, e garante a uniformidade na forma dos *spots*. A sua desvantagem passa pelo elevado custo do equipamento.

### 1.4.1.3. Suportes sólidos

Existe uma variedade de superfícies – slides – comercializados que podem ser utilizados. Estas permitem manter a conformação e a funcionalidade da proteína aquando da impressão, mas também permitem que estas mantenham a sua capacidade máxima de ligação.<sup>[42]</sup> A escolha da superfície deve ter em conta o tipo de moléculas a serem imobilizadas. São exemplos as superfícies de níquel - afinidade para proteínas com cauda de histidinas<sup>[42]</sup>, superfícies com estreptavidina – afinidade para proteínas biotinizadas<sup>[43]</sup>; superfícies de vidro funcionalizadas que permitem a imobilização covalente de proteínas através do grupo amina<sup>[44]</sup>; e superfícies de nitrocelulose – para a imobilização de proteínas por adsorção. As diferenças encontradas entre os vários slides comercializados podem fazer com as etapas de imobilização e os protocolos de deteção sejam também diferentes.

Ao longo deste estudo foram utilizados slides com matriz de nitrocelulose. Estes são comercializados com 1, 2, 3, 4, 8, 12, 16, 24 e 64 blocos de nitrocelulose onde são feitas as experiências.

De um modo geral, a técnica de *microarrays* de proteínas permite a identificação de novos alvos para serem estudados por outras técnicas como a Cristalografia de Raios-X. Esta técnica é capaz de analisar várias interações proteína-proteína e proteína-ligando, tornando-a rápida e eficaz no estudo de amostras biológicas complexas como um proteoma ou amostras de pacientes. A contínua otimização dos *microarrays* tem permitido aumentar o número de aplicações que vão desde da determinação de padrões de resposta, de sensibilidade e de toxicidade a determinados compostos, estudo de doenças resultantes de alterações a funções e interações de proteínas até ao desenvolvimento de ferramentas para diagnóstico simultaneamente sensíveis e específicas. Tudo isto utilizando pequenos volumes e/ou pequenas concentrações de amostra.



Capítulo

2

**Objetivos**



## 2. Objetivos

A presente dissertação tem como principal objetivo a elucidação das interações coesina-doquerina na montagem/organização do celulosoma do microrganismo *C. thermocellum*. Assim, como alvo deste estudo selecionou-se a CipA – proteína que pertence ao celulosoma do *C. thermocellum* – dando assim continuidade a estudos realizados anteriormente. As metodologias principais de eleição foram a Cristalografia de Raios-X, que permite a caracterização estrutural a nível atómico dos módulos em estudo, e a metodologia de *Microarrays* de proteínas, que possibilita averiguar a especificidade das interações coesina-doquerina. Este estudo pode assim ser dividido em dois objetivos:

- No primeiro pretende-se a caracterização, através da Cristalografia de Raios-X, dos módulos coesina C1, C3, C4 e C9 da proteína CipA, todos eles complexados ao módulo doquerina Xyn10B, podendo assim elucidar possíveis diferenças nos módulos coesina escolhidos e clarificar as interações com o módulo, já caracterizado estruturalmente, Xyn10B. Para tal, foi necessário, numa primeira etapa, sobre-expressar em *E. coli* e purificar cada um dos complexos. Posteriormente tentou-se determinar e otimizar as condições de cristalização dos quatro complexos, de forma a se resolver as suas estruturas tridimensionais. Todos estes passos foram realizados no laboratório de Cristalografia de proteínas do UCIBIO-REQUIMTE, FCT-UNL. (Capítulo 3)

- No segundo pretende-se elucidar as diferentes especificidades e afinidades, através da técnica de *Microarrays*, das interações coesina-doquerina estabelecidas entre vários módulos coesina e doquerina do tipo I e II dos microrganismos *C. thermocellum* e *R. flavefaciens*, permitindo investigar a interação entre várias espécies. Para dar seguimento a esta análise, foi necessário sobre-expressar em *E. coli* e purificar todos os módulos em estudo. A investigação decorreu nos laboratórios de Cristalografia de proteínas e de Glico-microarrays, ambos do UCIBIO-REQUIMTE, FCT-UNL, bem como no laboratório da Professora Ten Feizi, no Imperial College em Londres, onde se realizou a impressão robótica dos slides de 16 blocos utilizados durante o trabalho desta tese. (Capítulo 4)



Capítulo

3

## **Caracterização estrutural da proteína CipA por Cristalografia de Raios-X**



### 3.1. Introdução

Os módulos coesina e doquerina que contemplam os complexos coesina-doquerina (Coh-Doc) em estudo foram clonados em trabalho anterior pela Doutora Benedita Andrade Pinheiro (UCIBIO-FCT/UNL), no vetor pET28a(+) alterado especialmente para o efeito. Neste vetor os módulos coesina e doquerina que compõem o complexo foram clonados juntos, de modo a serem expressos simultaneamente. O constructo utilizado para a expressão dos complexos em estudo é apresentado na Figura 3.1, no qual verificamos a presença de 6 histidinas (*His-tag*) no C-terminal de cada coesina.



**Figura 3.1** -Constructo utilizado para expressão dos complexos coesina-doquerina. Os termos Doc e Coh correspondem aos módulos doquerina e coesina, respetivamente. Cada um destes componentes é expresso simultaneamente, com promotores independentes. A coesina apresenta no seu C-terminal uma *His-tag*.

O objetivo deste capítulo foi determinar a estrutura tridimensional de quatro complexos Coh-Doc da CipA. Estes complexos são formados pelos módulos coesina C1, C3, C4 e C9, todos eles em complexo com o módulo doquerina Xyn10B, tendo sido denominados de C1X10B, C3X10B, C4X10B e C9X10B, respetivamente. Estudos anteriores por Cristalografia de Raios-X bem como estudos de afinidade entre os módulos coesina e doquerina motivaram a escolha da doquerina Xyn10B.<sup>[31] [32] [46]</sup> O módulo doquerina sofreu uma mutação nos resíduos Ser51-Thr52, os quais foram substituídos por Ala51-Ala52, de maneira a restringir a ligação do módulo doquerina Xyn10B à coesina por um local de ligação único (hélice  $\alpha$  1). A informação correspondente aos vários módulos coesina e doquerina envolvidos neste estudo encontra-se no Anexo 7B.I.

### 3.2. Métodos e Materiais

#### 3.2.1. Transformação de células competentes

A transformação de células competentes da estirpe *Escherichia coli* BL21(DE3) foi realizada à chama, em ambiente estéril. Neste protocolo adicionou-se 3  $\mu$ L de DNA do respetivo complexo em estudo, a 50 $\mu$ L de células competentes, deixando repousar a mistura em gelo durante 30 minutos. A mesma foi posteriormente incubada a 42°C cerca de 40 a 60 segundos. Após um segundo repouso de 2 a 5 minutos em gelo, adicionou-se 1mL de meio de cultura LB e incubou-se a 37°C (Orbital Shaker-Incubator ES-20, da Grant.bio) durante 1 hora. As células foram então centrifugadas durante 5 minutos a 3400x g (MiniSpin® Plus, da Eppendorf), retirando-se o excesso de meio e ressuspendendo-se o *pellet* no restante meio. O produto da transformação foi plaqueado em placas de LB-agar com 50 $\mu$ g/mL canamicina (KAN), para posterior seleção dos transformantes. As placas foram incubadas a 37°C (Orbital Shaker-Incubator ES-20, da Grant.bio) durante 16 horas até ao aparecimento de colónias e posteriormente guardadas a 4°C.

### 3.2.2. Expressão dos complexos coesina-doquerina

De forma a se obter a condição mais favorável à expressão de cada um dos complexos Coh-Doc foi realizado um teste de expressão para cada complexo, testando-se vários meios de cultura (TB, 2xYP e LB) – Anexo 7A.I, diferentes temperaturas de expressão (19 e 37°C), bem como diferentes tempos de expressão (4 e 16 horas) e duas concentrações de IPTG (0,5mM e 1mM). Os resultados foram analisados tendo em conta o nível de expressão de cada complexo e a sua solubilidade. Cada complexo foi assim isolado a partir de células de *E.coli* BL21(DE3), na condição de expressão que melhores resultados obteve (Tabela 3.1).

**Tabela 3.1** - Condição de expressão selecionada para cada complexo Coh-Doc.

Complexo Coh-Doc	Meio de cultura	Temperatura (°C)	[IPTG] (mM)	Tempo de indução (h)
C1X10B	2xYP	37	0,5	4
C3X10B	TB		0,5	
C4X10B	TB		1	
C9X10B	TB		0,5	

Um clone proveniente de cada transformação foi inoculado em meio de cultura LB, condicionado com 50µg/mL do antibiótico KAN. O pré-inóculo foi incubado a 37°C, 200 rpm, o/n (Orbital Shaker-Incubator ES-20, da Grant.bio). No dia seguinte, foi adicionado 10mL do pré-inóculo realizado anteriormente, a 500mL do respetivo meio escolhido (1:50) condicionado com 50µg/mL do antibiótico KAN. Cada cultura foi incubada a 37°C, 210 rpm (Shaker IS-971R da Lab Companion), até se alcançar um D.O.<sub>600nm</sub> de 0,6-0,8. Uma vez atingido este intervalo, cada cultura foi induzida com a concentração adequada de IPTG (Nzytech) e incubada, durante 4 horas, à respetiva temperatura de indução, a 150 rpm (Shaker IS-971R da Lab Companion). Uma vez passadas as 4 horas de indução, as culturas foram centrifugadas durante 15 minutos, a 4°C e a 5000 x g (Rotor JA-10, Avanti J-26 XPI da Beckman Coulter). O sobrenadante foi descartado e o *pellet* guardado a -80°C para posterior protocolo de purificação.

### 3.2.3. Purificação dos complexos coesina-doquerina

Os complexos Coh-Doc foram submetidos a vários processos de purificação com o intuito de se obter um maior grau de pureza. A estratégia de clonagem destes complexos permitiu a utilização da cromatografia de afinidade por iões metálicos como primeiro método de purificação.

Após recolha das células por centrifugação estas foram ressuscitadas em tampão 50mM Hepes a pH 7,5, 1M NaCl e 10mM de imidazol (1g/10mL) juntamente com 1µg/µL de DNase I (Sigma-Aldrich), 300µg/µL de lisozima (Sigma-Aldrich), 10mM MgCl<sub>2</sub> e uma pastilha de inibidores de proteases (*SigmaFAST™ Protease Inhibitor Tablets*). Posteriormente as células foram lisadas por ultrassons em 3 ciclos de 1 minuto (UP50H da Lab Device). A fração solúvel foi recolhida por

centrifugação (Rotor JA-10, Avanti J-26 XPI da Beckman Coulter) a 10 000 rpm, durante 30 minutos a 4°C.

### **3.2.3.1. Cromatografia de afinidade por íões metálicos imobilizados**

A cromatografia de afinidade por íões metálicos imobilizados (IMAC) é conhecida pela sua elevada eficácia na separação de proteínas *His-tagged*. Esta eficácia deve-se à elevada afinidade e seletividade dos resíduos de histidinas com íões de metais bivalentes, presentes na matriz da coluna. Cada complexo foi assim inicialmente submetido a uma IMAC numa coluna de níquel, His-Trap™ de 5 mL, acoplada ao cromatógrafo ÄKTA START, ambos da GE-Healthcare. O método de purificação foi delineado utilizando o programa UNICORN™ start 1.0, tendo sido sempre aplicada uma pressão inferior a 0,5 MPa (Anexo 7A.II).

Neste protocolo foram utilizadas duas soluções tampão 50mM Hepes a pH 7.5, 1M NaCl a duas concentrações diferentes de imidazol (10 e 250 mM). A coluna His-Trap™ foi primeiramente lavada com cerca de 50 mL de água Milli-Q e posteriormente equilibrada com 20 mL do tampão a 10mM imidazol. O extrato celular foi então carregado na coluna. Procedeu-se a um primeiro patamar de lavagem em que foi utilizada 10% do tampão 50mM Hepes a pH 7,5, 1M NaCl, 250mM imidazol. Este patamar tem por objetivo uma lavagem inicial em que ocorre a separação dos constituintes de baixa afinidade para o níquel. Após esta lavagem foi realizado um gradiente de imidazol até à concentração máxima de 250 mM, por cerca de 100mL, com a finalidade de recolhermos o complexo Coh-Doc. É então realizado um segundo patamar com 100% da solução tampão a 250mM imidazol. As frações recolhidas foram posteriormente analisadas por eletroforese em gel de poliacrilamida em condições desnaturantes.

### **3.2.3.2. Cromatografia de exclusão molecular**

A cromatografia de exclusão molecular, também conhecida como filtração em gel, é utilizada para separar constituintes de uma mistura pela sua massa molecular. Cada complexo foi submetido a uma coluna Superdex 75™ 10/300 GL, de 25 mL, acoplada ao cromatógrafo ÄKTAFPLC, ambos da GE-Healthcare. O método de purificação foi delineado utilizando o programa UNICORN™ 5.11, sendo sempre aplicada uma pressão inferior a 1,8 Mpa (Anexo 7A.III).

Neste protocolo foi utilizada a solução tampão 50mM Hepes a pH 7,5, 150mM NaCl e 2mM CaCl<sub>2</sub>. A coluna Superdex 75™ 10/300 GL foi primeiramente lavada com cerca de 50 mL de água Milli-Q e posteriormente equilibrada com outros 50 mL da solução tampão. Para realização deste protocolo, a amostra foi concentrada (*Centricon* Vivaspin® Turbo 15 de 5kDa, da GE-Healthcare) em tampão 50mM Hepes a pH 7,5, 150mM NaCl e 2mM CaCl<sub>2</sub>. Posteriormente, a mesma é aplicada em injeções individuais de 500µL com ajuda de uma seringa, mantendo-se um fluxo da mesma solução tampão na coluna. As frações recolhidas foram posteriormente analisadas por eletroforese em gel de poliacrilamida em condições nativas e desnaturantes.

### 3.2.3.3. Cromatografia de permuta iónica

A cromatografia de permuta iónica baseia-se na interação reversível entre uma proteína carregada e um determinado meio com carga oposta. A cromatografia de permuta iónica foi efetuada com auxílio de uma coluna MonoQ™ 5/50 GL de 1mL, acoplado ao cromatógrafo AKTA HPLC, ambos da GE-Healthcare. O protocolo de purificação foi delineado e aplicado utilizando o programa UNICORN™ 5.31, tendo sido sempre aplicada uma pressão inferior a 4 MPa (Anexo 7A.IV).

Neste protocolo foram utilizados duas soluções tampão 20mM Tris a pH 7,5, 2mM CaCl<sub>2</sub>, com duas concentrações diferentes de NaCl (150mM e 1M). A amostra foi previamente concentrada (*Centricons Vivaspin® Turbo 15* de 5kDa) em tampão 20mM Tris pH 7,5 com 150mM de NaCl. A coluna MonoQ™ 5/50 GL foi primeiramente lavada com cerca de 5 mL de água Milli-Q e posteriormente equilibrada segundo o anexo 7A.IV. A amostra foi aplicada em injeções individuais de 2mL com ajuda de uma seringa, mantendo-se um fluxo na coluna com a mesma solução tampão a 150mM, no intuito de diminuir a ligação de proteínas com menor afinidade à matriz da coluna. A separação das proteínas é conseguida através de um gradiente de NaCl até à concentração máxima de 1M NaCl. As frações recolhidas foram posteriormente analisadas por eletroforese em gel de poliacrilamida em condições nativas e desnaturantes.

As concentrações dos complexos em estudo foram determinadas através da lei de Lambert-Beer, utilizando o Nanodrop SpectraMax 190 da Molecular Devices, acoplado ao programa SoftMax® Pro Softwares.

### 3.2.4. Ensaio de cristalização dos complexos C1X10B, C4X10B e do módulo coesina C9

Os ensaios de cristalização dos dois complexos coesina-doquerina (C1X10B e C4X10B) e do módulo coesina C9 em estudo foram realizados pelo método de difusão de vapor com as concentrações apresentadas na Tabela 3.2., nas soluções tampão descritas, e utilizando as técnicas de gota assente e gota suspensa em condições aeróbias. Para isso recorreu-se ao equipamento de formulação e aplicação de nano-gotas Oryx8 (Douglas Instruments, designado *robot* de cristalização), bem como placas de cristalização de 24 reservatórios.

**Tabela 3.2-** Concentrações utilizadas nos ensaios de cristalização

Amostra	Tampão	Concentração 1 (mg/mL)	Concentração 2 (mg/mL)
C1X10B	50mM Hepes, 150mM NaCl e 2mM CaCl <sub>2</sub> (pH 7,5)	1,8	-
C4X10B	20mM Tris, 2mM CaCl <sub>2</sub> , 150mM NaCl (pH 7,5)	19,48	9,75
C9		11,14	-

Durante o trabalho de pesquisa de condições de cristalização para cada amostra em estudo foram testados vários conjuntos de soluções para pesquisa/varrimento de condições (usualmente designado por *screening*) que apresentamos nos Anexos 7A.V a 7A.XII). Aquando da utilização do *robot* Oryx8, são colocados 50µL da solução de cristalização em cada um dos 96 reservatórios de uma placa de cristalização, e numa ordem pré-estabelecida. A aplicação informática dedicada ao *robot*, WaspRun, permite escolher a proporção proteína:precipitante. Aquando da utilização das placas de cristalização de 24 reservatórios, são colocados 700µL de uma das soluções de cristalização em cada um dos reservatórios. Após escolhermos a proporção de proteína:precipitante a ser utilizada, esta mistura é aplicada sobre uma lamela que, por sua vez, é invertida e colocada sobre o respetivo reservatório, assentando sobre uma fina camada de cera, selando o sistema e estabelecendo o equilíbrio entre gota e reservatório. Os ensaios de cristalização foram testados a 4 e 20°C. A visualização e registo das gotas foi efetuado recorrendo ao microscópio óptico.

### **3.2.5. Recolha de dados de difração de raios-X e resolução de estruturas tridimensionais**

Os cristais obtidos foram posteriormente arrefecidos a temperaturas criogénicas e as suas propriedades de difração de raios-X foram analisadas sob radiação de sincrotrão no *European Synchrotron Radiation Facility* (ESRF), em Grenoble, e no *Source optimisée de lumière d'énergie intermédiaire du LURE* (SOLEIL), em Saint-Aubin, ambos em França. Alguns cristais foram também submetidos a difração *in house* com o difractor D8 Venture (Bruker) usando radiação K $\alpha$  de cobre ( $\lambda=1,5418\text{\AA}$ ). Os dados recolhidos durante a experiência de difração foram processados por vários programas com o objetivo da determinação da estrutura tridimensional das proteínas cristalizadas.

Os programas iMOSFLM e XDS permitiram processar os dados de difração recolhidos, indexando as reflexões, determinando as possíveis soluções para as constantes da célula e grupo espacial, bem como integrando todas as intensidades.<sup>[47] [48]</sup> Os resultados foram automaticamente gravados num ficheiro que contém toda a informação, ficheiro .mtz. Os programas AIMLESS e MATTHEWS foram utilizados para processar todas as reflexões, confirmar as constantes da célula e o grupo espacial e ainda, estimar o conteúdo em solvente da célula unitária, utilizando o peso molecular do complexo em estudo. O programa PHASER foi utilizado para determinar as fases por Substituição Molecular (*Molecular Replacement* - MR), utilizando-se como modelo a estrutura depositada no PDB com o código 2CCL.<sup>[32] [49]</sup> O programa DM foi utilizado com o objetivo de melhorar a densidade eletrónica inicialmente obtida, enquanto que o BUCCANEER permitiu fazer a identificação e construção automática de cadeias de proteínas no mapa de densidade eletrónica obtido anteriormente.<sup>[50] [51]</sup> Por último, a estrutura foi refinada usando o programa REFMAC5.<sup>[52]</sup> Durante os vários ciclos de refinamento a estrutura foi visualizada e ajustada com auxílio do programa COOT versão 0.8.1. As imagens foram geradas no programa Chimera versão 1.9.

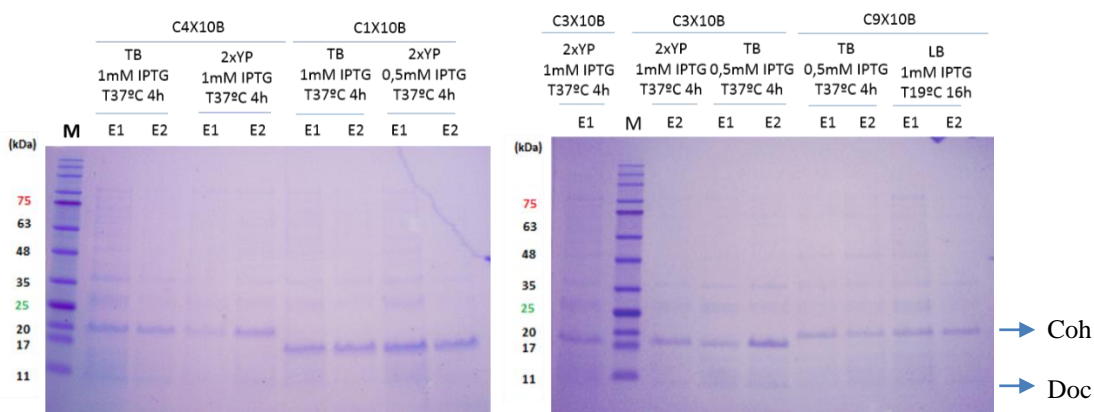
Os programas iMOSFLM, AIMLESS, MATTHEWS, PHASER, DM, BUCCANEER e REFMAC5 são disponibilizados pelo CCP4 (Collaborative Computational Project, Number 4).

### 3.3. Apresentação e discussão dos resultados

Tal como foi referido nas secções anteriores, os estudos estruturais dos complexos Coh-Doc levarão a que, futuramente, a organização/montagem do celulosoma do *C.thermocellum* seja decifrada, informação importante no desenvolvimento de várias aplicações biotecnológicas.

#### 3.3.1. Expressão dos complexos coesina-doquerina

Numa primeira etapa, tal como referido na secção 3.2.2, realizamos vários testes de expressão para cada complexo Coh-Doc testando três meios de cultura (LB, 2xYP e TB), duas temperaturas de expressão (19 e 37°C), duas concentrações de IPTG (0,5 e 1mM) e dois tempos de expressão (4 e 16 horas). A análise eletroforética em condições desnaturantes das frações solúveis e insolúveis de cada condição em estudo permitiu-nos visualizar as condições de expressão para cada complexo, e seleccionar os que melhores resultados obtiveram. Face a estes resultados escolhemos os dois melhores e prosseguimos para um protocolo de purificação por IMAC, em colunas *His SpinTrap*, segundo o protocolo da GE-Healthcare. A análise eletroforética das frações eluídas, apresentada na Figura 3.2, permitiu a escolha da melhor condição de expressão para cada complexo. Os pesos moleculares dos módulos coesina e doquerina em estudo são apresentados no Anexo 7B.I. Eram esperadas duas bandas no gel, uma para cada proteína. Contudo, por ter um peso molecular inferior, a banda do módulo doquerina não é visível em todos eles, contrariamente com o que acontece com a banda referente ao módulo coesina.



**Figura 3.2** - Resultados da análise eletroforética em condição desnaturantes das frações eluídas durante a purificação por IMAC dos complexos Coh-Doc. As condições de expressão estão indicadas no topo de cada figura. Indução por IPTG, a 37°C durante 4 horas e 19°C durante 16 horas. E1 – Primeira eluição em tampão 50mM Hepes a pH7,5 com 60mM Imidazol; E2 – Segunda eluição em tampão 50mM Hepes a pH7,5 com 300mM Imidazol. M corresponde ao marcador de pesos moleculares da NZYTech.

A observação da Figura 3.2. mostra-nos quais as frações solúveis purificadas com maior e menor teor de proteína. No caso do complexo C1X10B, verificamos que as eluições com 60 e 300 mM de imidazol são equiparáveis nos dois meios de expressão escolhidos anteriormente. O mesmo acontece para o complexo C9X10B, escolhendo assim uma das duas condições de expressão. Por outro lado, os complexos C3X10B e C4X10B apresentam uma maior quantidade de proteínas solúvel purificada em meio de indução TB à temperatura de 37°C, com 0,5 e 1mM de IPTG, respetivamente. A Tabela 3.3. apresenta os resultados mais promissores de cada complexo.

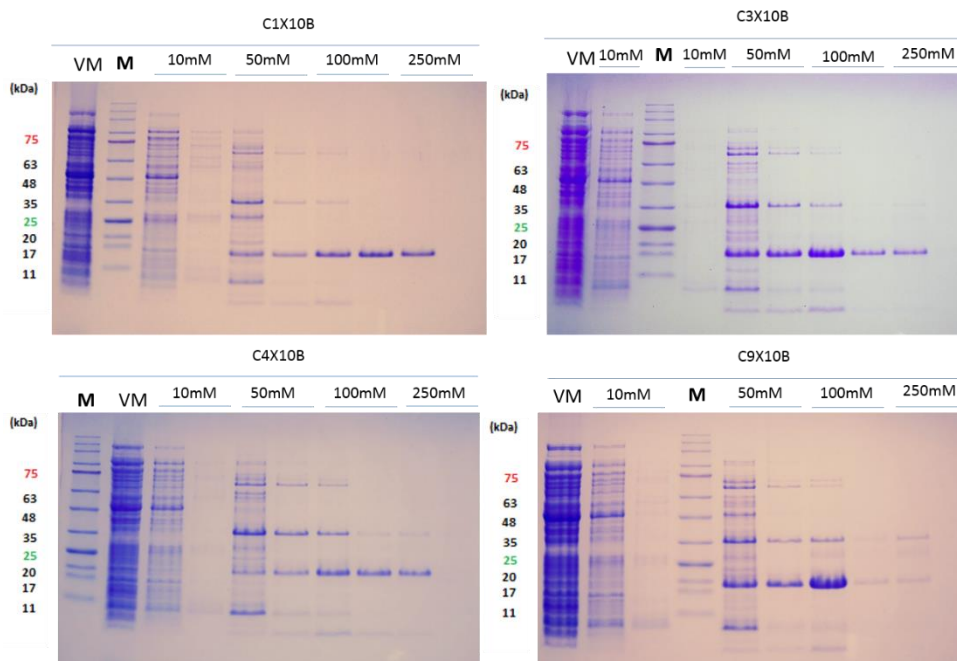
**Tabela 3.3-** Resultado mais promissor de expressão para cada complexo em estudo.

Complexo Coh-Doc	Meio de indução	[IPTG] (mM)	T <sub>indução</sub> (°C)	t <sub>indução</sub> (H)
C1X10B	Meio 2xYP	0,5	37	4
C3X10B	Meio TB	0,5		
C4X10B		1		
C9X10B		0,5		

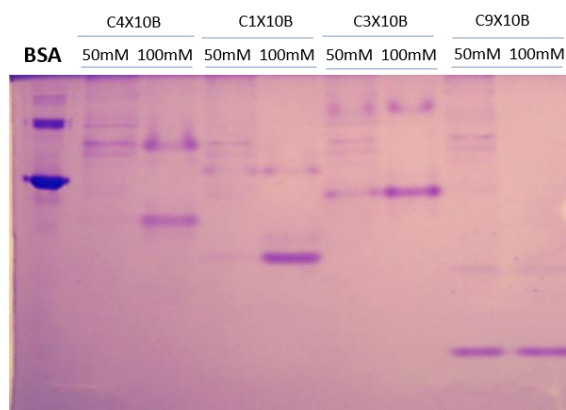
Numa segunda etapa, com o objetivo de verificar se estes resultados eram reprodutíveis, prosseguiu-se para uma expressão e purificação de cada complexo numa escala 50 vezes maior. A purificação foi realizada recorrendo a uma IMAC, em colunas *His GraviTrap*, onde se fizeram várias lavagens, a diferentes concentrações de imidazol (10, 50, 100 e 250mM), de modo a otimizar o protocolo de purificação. Os resultados da análise eletroforética das frações eluídas para os complexos C1X10B, C3X10B, C4X10B e C9X10B, encontram-se na Figura 3.3.

A mesma figura permitiu-nos verificar que a maioria das proteínas de *E.coli* não se liga à coluna saindo logo no início do protocolo (volume morto), tal como nas lavagens com 10mM e 50mM de imidazol. É também nesta última concentração de imidazol mencionada que começamos a verificar a eluição dos complexo Coh-Doc em estudo, contudo ainda bastante contaminados com proteínas do hospedeiro. É ainda possível observar em todos os géis da Figura 3.3, que aos 100mM de imidazol existe a eluição completa do complexo Coh-Doc através da presença das duas bandas, correspondentes aos módulos coesina e doquerina. Por outro lado, nalguns casos verificamos também a eluição de alguma da coesina em excesso que é eluída em maior quantidade aquando da lavagem com 250mM de imidazol.

De modo a confirmarmos a presença de cada complexo Coh-Doc realizou-se a análise eletroforética em condições nativas das frações eluídas a 50 e 100mM de imidazol (Figura 3.4).



**Figura 3.3** - Resultados da análise eletroforética em condições desnaturantes das frações eluídas durante a purificação por IMAC dos complexos Coh-Doc. As concentrações de imidazol utilizadas para eluição das respectivas frações estão apresentadas no topo da figura. VM – Volume morto; 10mM – Solução tampão 50mM Hepes a pH 7,5 com 10mM de imidazol; 50mM – Solução tampão 50mM Hepes a pH 7,5 com 50mM de imidazol; 100mM – Solução tampão 50mM Hepes a pH 7,5 com 100mM de imidazol; 250mM – Solução tampão 50mM Hepes a pH 7,5 com 250mM de imidazol; M corresponde ao marcador de pesos moleculares da NZYTech.



**Figura 3.4** - Resultados da análise eletroforética em condições nativas da purificação por IMAC dos complexos Coh-Doc. As concentrações de imidazol utilizadas para eluição das respectivas frações estão apresentadas no topo da figura; A BSA é apresentada como referência de corrida.

A observação do gel nativo acima apresentado permite-nos verificar que cada complexo Coh-Doc é eluído maioritariamente com uma concentração de 100mM de imidazol. Porém é ainda perceptível a presença de algumas bandas abaixo da correspondente a cada complexo, pertencente a algumas outras proteínas ainda presente na fração solúvel e aos respectivos módulos coesina. Sendo assim, foi necessário, em purificações futuras, a realização de mais passos de purificação para a obtenção do complexo Coh-Doc a um nível elevado de pureza.

Em suma, todos os testes de expressão realizados permitiram a determinação das condições mais favoráveis de expressão para cada complexo, bem como a elaboração de uma estratégia de purificação para os mesmos.

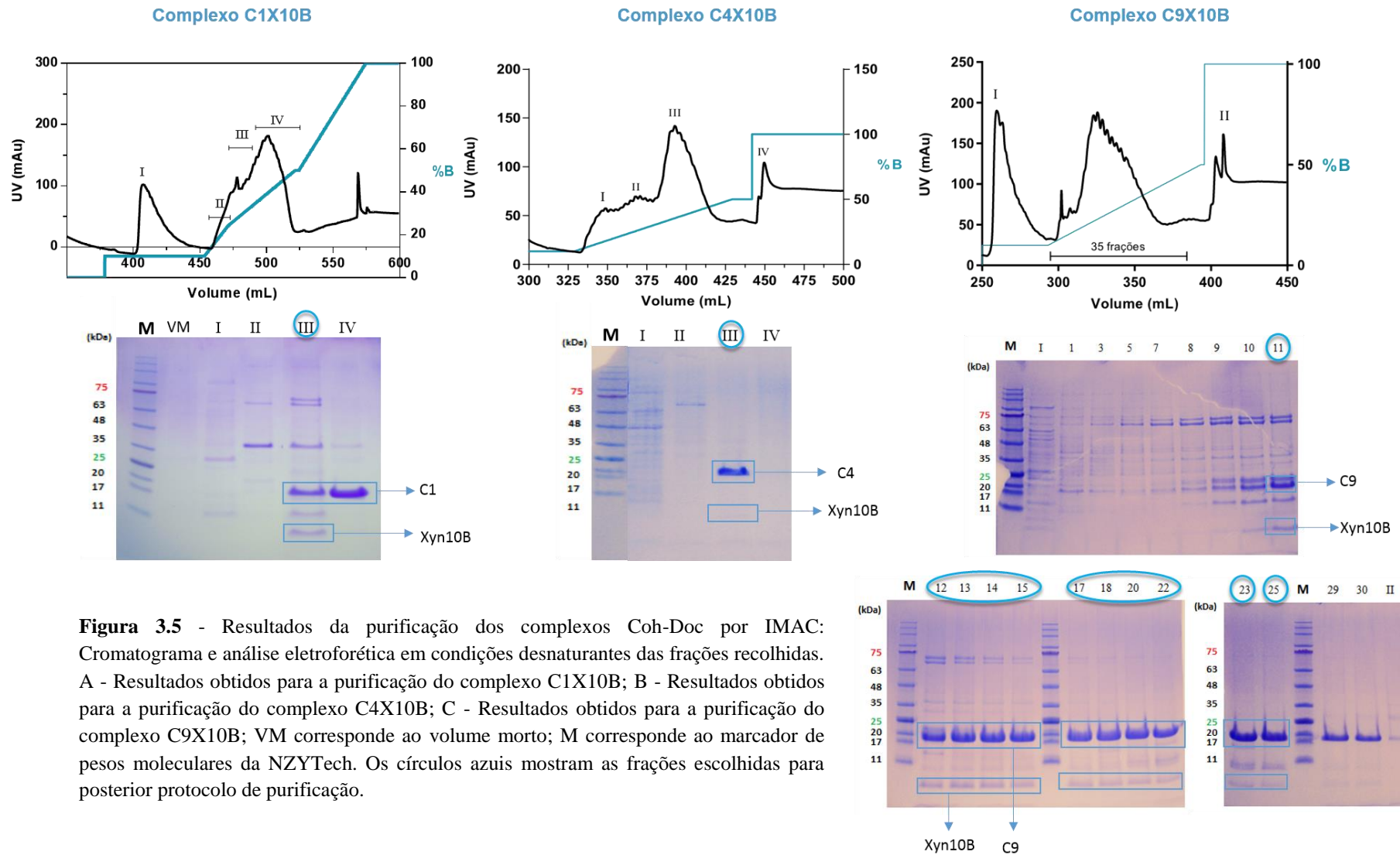
Para realização dos ensaios de cristalização os complexos C1X10B, C3X10B, C4X10B e C9X10B, foram primeiro expressos em grande escala, em frações de 500mL, num volume total de 4 litros, seguindo-se o protocolo apresentado na secção 3.2.3. Após protocolo de expressão, procedeu-se à purificação destes complexos recorrendo a vários protocolos de cromatografia, de forma a, numa primeira etapa, separar o complexo Coh-Doc e a coesina sobre-expressa das restantes proteínas de *E.coli* e, numa segunda etapa, separar esta coesina do complexo em estudo. Para isso, utilizaram-se um total de três protocolos de purificação: cromatografia de afinidade por iões metálicos imobilizados, cromatografia de exclusão molecular e cromatografia de permuta iónica. Todos os complexos foram purificados, exceto o complexo C3X10B. A Tabela 3.4 apresenta os tampões utilizados em cada protocolo de purificação.

**Tabela 3.4** - Soluções tampão utilizadas nos vários protocolos de purificação.

Cromatografia de afinidade por iões metálicos imobilizados	Cromatografia de exclusão molecular	Cromatografia de permuta iónica
50mM HEPES a pH 7.5, 1M NaCl, 10mM Imidazol	50mM HEPES, 150mM NaCl, 2mM CaCl <sub>2</sub>	20mM Tris-HCl pH8, 150mM NaCl e 2mM CaCl <sub>2</sub>
50mM HEPES a pH 7.5, 1M NaCl, 250mM Imidazol		20mM Tris-HCl pH8, 1M NaCl e 2mM CaCl <sub>2</sub>

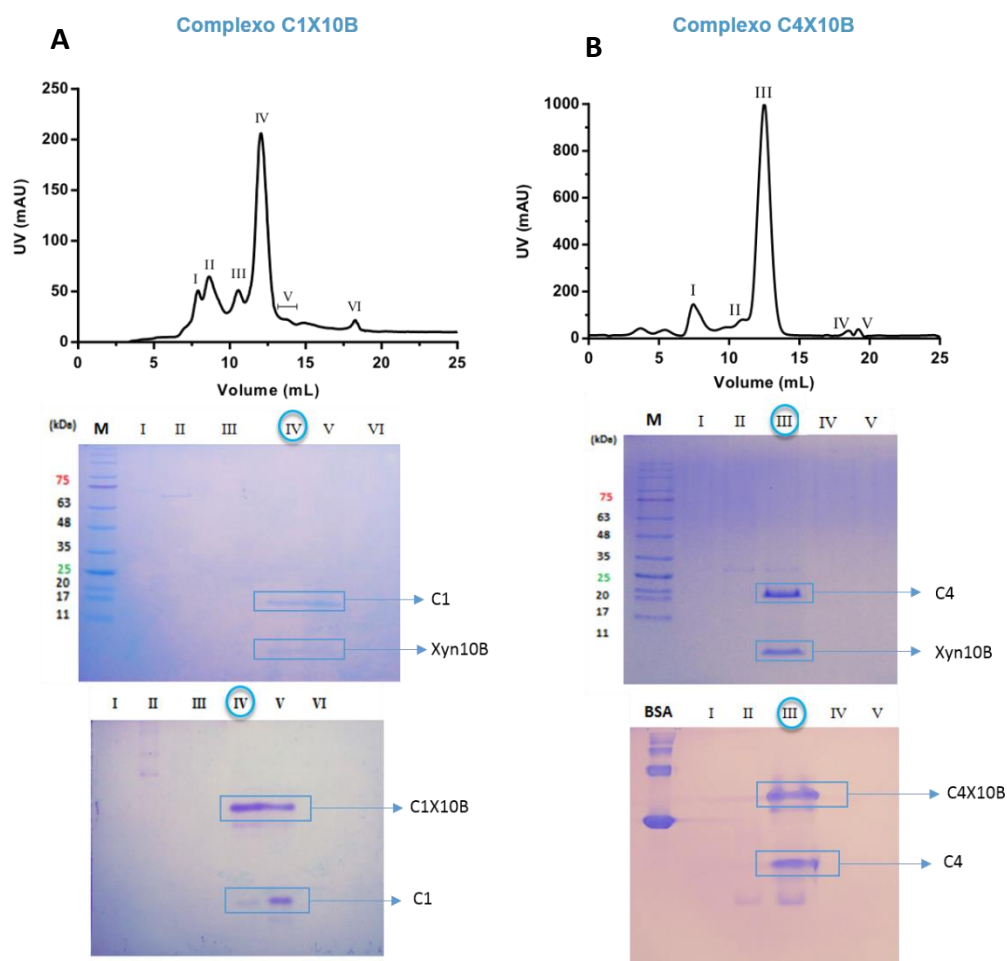
### 3.3.2. Purificação dos complexos por cromatografia de afinidade por iões metálicos imobilizados

A purificação por IMAC dos complexos Coh-Doc foi realizada numa coluna *HisTrap*<sup>TM</sup> (GE-Healthcare) de 5mL, acoplada ao cromatógrafo ÄKTASTART. O protocolo de purificação foi descrito na secção 3.2.3.1, e teve por objetivo separar o conjunto complexo Coh-Doc – coesina em excesso, das restantes proteínas presentes no extrato. Todas as frações recolhidas foram analisadas por eletroforese em condições desnaturantes, permitindo-nos selecionar as correspondentes ao respetivo complexo purificado (Figura 3.5.). Estas mesmas frações foram posteriormente dessalinizadas e concentradas no tampão do segundo protocolo a ser realizado, usando concentradores Vivaspin® 15R, da GE Healthcare.



### 3.3.3. Purificação dos complexos por cromatografia de exclusão molecular

A purificação por cromatografia de exclusão molecular foi realizada numa coluna Superdex 75™ 10/300 GL (GE-Healthcare) de 25 mL, acoplada ao cromatógrafo ÄKTAFPLC. O protocolo de purificação foi descrito na secção 3.2.3.2. A cromatografia de exclusão molecular foi escolhida para segundo passo de purificação pois é possível verificarmos a existência de outras proteínas com pesos moleculares distintos. Teve por objetivo a separação do complexo Coh-Doc da coesina em excesso e das restantes proteínas ainda presentes nas frações escolhidas anteriormente. Todas as frações recolhidas foram analisadas por eletroforese em condições desnaturantes e nativas, permitindo-nos seleccionar as correspondentes ao respetivo complexo purificado (Figura 3.6.). Estas mesmas frações foram posteriormente dessalinizadas e concentradas no tampão do respetivo seguinte protocolo a ser realizado.



**Figura 3.6** -Resultados da purificação dos complexos Coh-Doc por cromatografia de exclusão molecular: Cromatograma e análise eletroforética em condições desnaturantes e não desnaturantes das frações recolhidas. A - Resultados obtidos para a purificação do complexo C1X10B; B - Resultados obtidos para a purificação do complexo C4X10B; M corresponde ao marcador de pesos moleculares da NZYTech. BSA é apresentada como referência de corrida. Os círculos azuis mostram as frações escolhidas para posterior protocolo de purificação

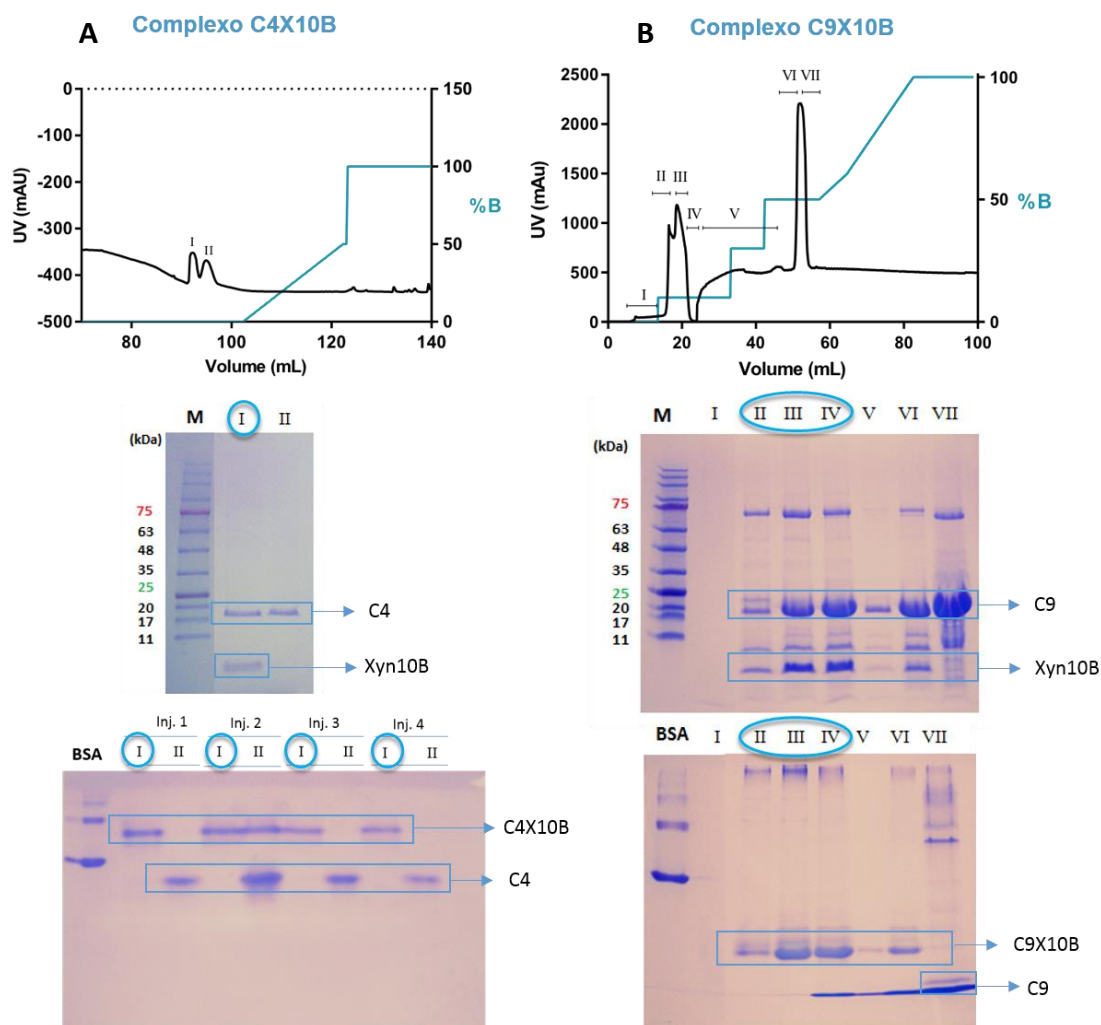
Para o complexo C1X10B foram feitas 7 injeções de 500µL de um volume total de 3,5mL. Por outro lado, para o complexo C4X10B foram realizadas 4 injeções de 500µL, num volume total de 2mL. A Figura 3.6 reflete os resultados para a primeira injeção do protocolo de purificação de cada complexo, possuindo as restantes o mesmo perfil cromatográfico. Os resultados dão conta que não conseguimos, em nenhum dos dois protocolos, separar a coesina em excesso do respetivo complexo.

No final do protocolo, e após dessalinização e concentração, percebemos que a quantidade de complexo C1X10B não permitia a realização de um terceiro protocolo de purificação a fim de separá-lo a coesina em excesso ainda presente. Neste caso, prosseguiu-se diretamente para os ensaios de cristalização.

### **3.3.4. Purificação dos complexos por cromatografia de permuta iónica**

A purificação por cromatografia de permuta iónica foi realizada numa coluna MonoQ™ 5/50 GL (GE-Healthcare), de 1 mL, acoplada ao cromatógrafo ÄKTAFPLC. O protocolo de purificação foi descrito na secção 3.2.3.3. A cromatografia de permuta iónica foi escolhida como passo de purificação para se tentar definitivamente separar o complexo Coh-Doc da coesina em excesso. Todas as frações recolhidas foram analisadas por eletroforese em condições desnaturantes e nativas, permitindo-nos seleccionar as correspondentes ao respetivo complexo purificado (Figura 3.7.).

Para o complexo C4X10B foram feitas 4 injeções de 2mL para um volume total de 8mL, não tendo sido necessário gradiente de NaCl para a sua eluição e separação da coesina em excesso. Por outro lado, para o complexo C9X10B foi realizada apenas uma injeção de 2mL, tendo sido necessários a realização de três patamares (0,1; 0,3 e 0,5M) de NaCl para a eluição do mesmo e separação de parte da coesina em excesso. A observação do resultado da análise eletroforética (Figura 3.7 – B) permitiu-nos verificar a presença de coesina em excesso e de proteínas de *E.coli*, que não foram separadas anteriormente. Decidiu-se contudo prosseguir com os ensaios de cristalização com as frações assinaladas tendo em conta que a proteína apresentava um grau de pureza de aproximadamente 90%. No final do protocolo, as frações correspondentes aos complexos em estudo foram dessalinizadas e concentradas para posteriores ensaios de cristalização.



**Figura 3.7** - Resultados da purificação dos complexos Coh-Doc por cromatografia de permuta iônica: Cromatograma e análise eletroforética em condições desnaturantes e nativas das frações recolhidas. A - Resultados obtidos para a purificação do complexo C4X10B; B - Resultados obtidos para a purificação do complexo C9X10B; M corresponde ao marcador de pesos moleculares da NZYTech. BSA é apresentada como referência de corrida. Os círculos azuis mostram as frações escolhidas para posteriores ensaios de cristalização.

### 3.3.5. Ensaios de cristalização das amostras em estudo, experiência de difração de raios-X e resolução de estruturas tridimensionais

A identificação das condições de cristalização dos complexos Coh-Doc C1X10B, C4X10B e do módulo coesina C9 foi realizada utilizando concentrações apresentadas na Tabela 3.2, da secção 3.2.4. e várias soluções de cristalização. Após a obtenção de cristais procedeu-se à otimização de algumas dessas condições. Os cristais a serem testados, quer em sincrotrão, quer *in house*, foram previamente criopreservados em azoto-líquido na presença de uma solução de *harvesting buffer* e de uma solução crioprotetora. A primeira possui a mesma composição que a solução precipitante, porém com uma concentração mais elevada de agente precipitante tendo por objetivo evitar que o cristal se dissolva.

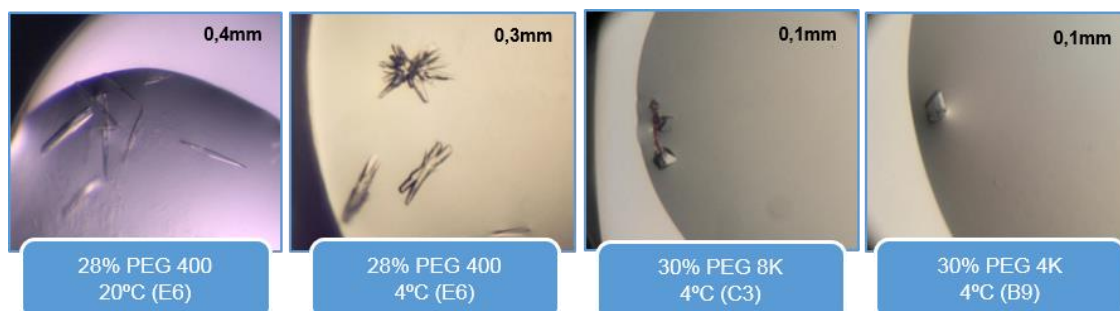
Por outro lado, a solução crio-protetora previne a formação de gelo no cristal aquando da sua congelação em azoto-líquido. Esta solução advém da adição de uma percentagem de glicerol à solução de *harvesting buffer*.

### 3.3.5.1. Complexo coesina-doquerina C1X10B

#### 3.3.5.1.1. Ensaio de cristalização e experiência de difração de raios-X

Os ensaios de cristalização para o complexo C1X10B foram realizados às temperaturas de 4 e 20°C, utilizando o *robot* Oryx8 e recorrendo à técnica da gota assente em todos os ensaios. Testaram-se as seguintes formulações pré-preparadas (*screens*): cristalização 80! (preparado *in-house* e baseado na formulação desenvolvida por Jancarik et al., 1991)<sup>[53]</sup>, PEG ION (1 e 8K); JBS (5,6,7 e 8), e JCSG+ (formulações comerciais). A proteína foi utilizada à concentração de 1,8mg/mL, testando-se a proporção de proteína-precipitante de 1:1.

Nos dias seguintes à preparação das gotas, na maioria das condições, não foi observado qualquer formação de precipitado, o que indica que o equilíbrio entre proteína-precipitante se encontrava maioritariamente na zona de sub-saturação do diagrama de fases da cristalização (Figura 1.18.), tendo ocorrido apenas mais tarde a formação de pequenos cristais nas condições apresentadas da Figura 3.8, referentes ao *screen* de 80!.



**Figura 3.8** - Resultados obtidos para os ensaios de cristalização do complexo C1X10B utilizando o *screen* de 80!. O agente precipitante utilizado para a cristalização bem como os tamanhos dos cristais obtidos compreendidos entre os 0.1 e 0.4 mm estão apresentados na figura.

Destas condições, os cristais com melhores características morfológicas foram testados no ESRF e SOLEIL. A Tabela 3.5 descreve as percentagens de PEG e de glicerol com os quais os cristais escolhidos foram previamente crio-preservados.

**Tabela 3.5** - Percentagens de agente precipitante e de glicerol utilizadas para a realização da crio-preservação dos cristais.

Condição	% (V/V) PEG	% (V/V) Glicerol
E6 – 28% PEG 400; 0.2M CaCl <sub>2</sub> ; 0.1M HEPES pH 7.5	30	5
C3 - 30% PEG 8K; 0.2M Tartarato K/Na; 0.1M Mes pH 6.5	32	15
B9 – 30% PEG 4K; 0.2M CaCl <sub>2</sub> ; 0.1M HEPES pH 7.5	32	20

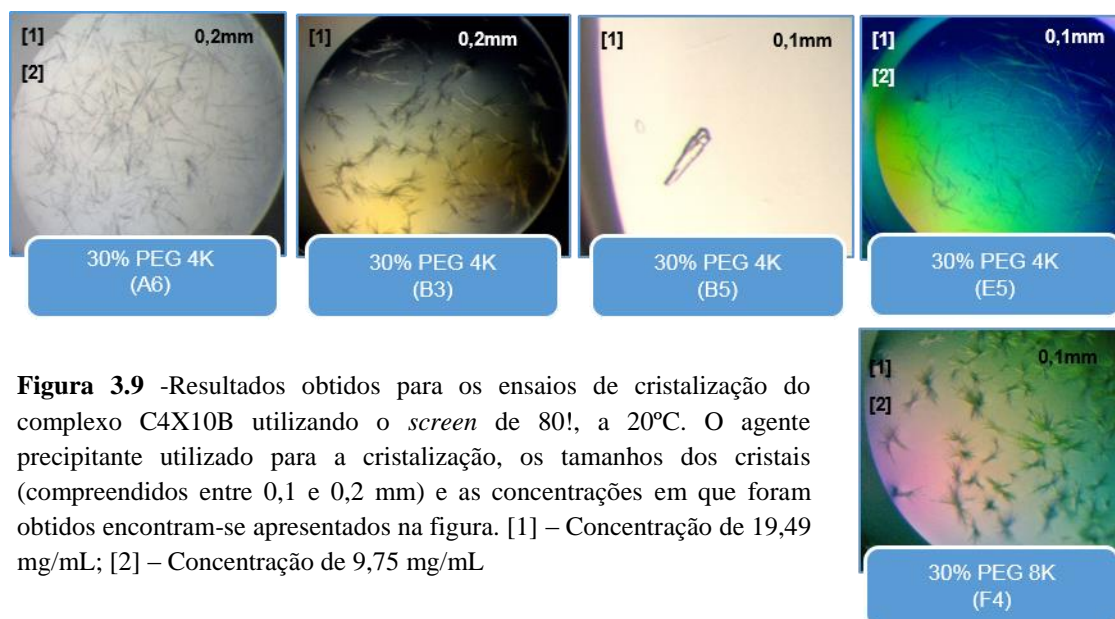
A recolha de dados de difração de raios-X nestes sincrotrões mostrou que a totalidade destes cristais era do composto iónico presente na mistura proteína-precipitante. A concentração obtida para este complexo parece ser o fator limitante para a obtenção de cristais de proteína. Para o melhor prosseguimento do trabalho desta dissertação, e uma vez que a cristalização das outras amostras foi bem-sucedida, foi decidido não insistir na pesquisa das condições de cristalização deste complexo.

### 3.3.5.2. Complexo coesina-doquerina C4X10B

#### 3.3.5.2.1. Ensaio de cristalização e experiência de difração de raios-X

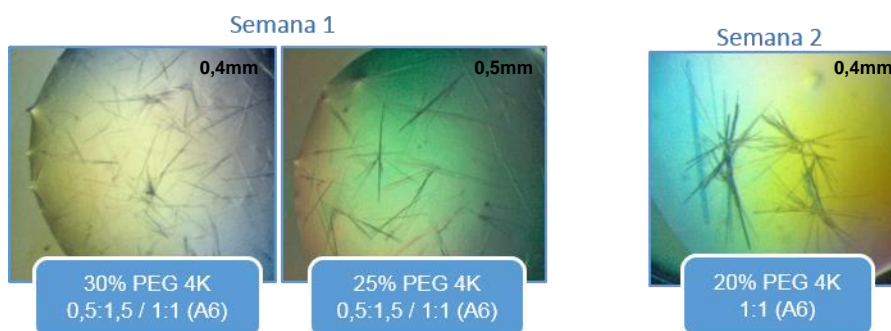
Os ensaios de cristalização para o complexo C4X10B foram efetuados utilizando-se o *screen* de 80!, com a proteína nas concentrações de 19,48 e 9,75 mg/mL. A formação de cristais foi testada à temperatura de 20°C, utilizando o *robot* Oryx8 e, novamente, recorrendo à técnica da gota assente em todos os ensaios. Durante os vários protocolos de otimização variou-se a proporção de proteína-precipitante e as percentagens de agente precipitante.

Nos dias seguintes à preparação das gotas, cerca de 3 dias depois, foi observada a formação de precipitado em algumas condições, em ambas as concentrações, indicando que, nestes casos, o equilíbrio proteína-precipitante se encontra na zona de sobressaturação do diagrama de fases da cristalização (Figura 1.18). Houve formação de cristais em 5 condições (Figura 3.9).



Face a estes resultados foram escolhidas as três condições (A6, B5 e E5) com melhores perspectivas de otimização em que se obtiveram cristais com melhor morfologia. As restantes condições não foram escolhidas pois verificou-se a agregação de cristais de pequenas dimensões sob forma de “ourios”. A observação da fotografia dos cristais obtidos para as condições A6 e E5 revela uma elevada taxa de nucleação, com a presença de vários cristais em forma de agulhas muito finas. A otimização, feita em *robot*, teve por objetivo a diminuição da taxa de nucleação bem como a obtenção de cristais maiores.

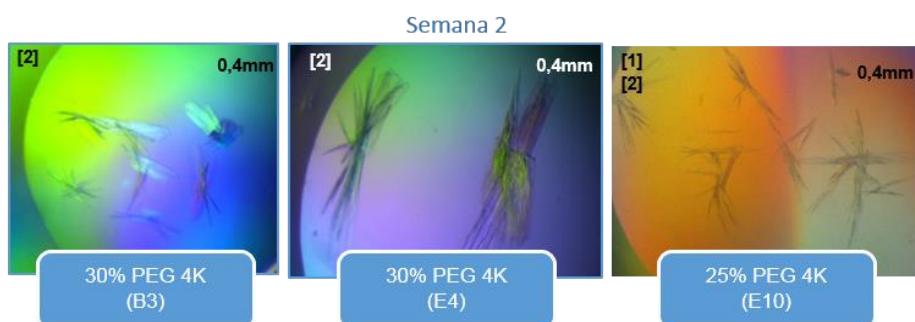
A otimização da condição B5 teve como único objetivo obter cristais com maiores dimensões. Para isso testámos, para as condições A6 e E5, diferentes proporções proteína:precipitante (1:1; 0,5:1,5) e diferentes concentrações do agente precipitante PEG 4K (10, 15, 20, 25 e 30%). Para a condição B5 apenas fizemos a repetição da proporção 1:1, cerca de 8 vezes com o objetivo de ver a reprodutibilidade da condição. É de referir que esta optimização foi realizada com a concentração de 9,75 mg/mL.



**Figura 3.10** - Resultados obtidos para as primeiras duas semanas de observação da otimização realizada a 20°C, para a condição A6. As proporções de proteína:precipitante, as percentagens de agente precipitante e os tamanhos dos cristais (compreendidos entre 0,4 e 0,5 mm) estão esquematizados na figura.

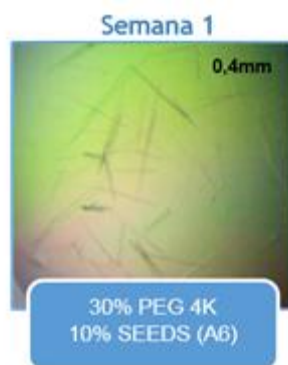
A observação destes cristais obtidos permitiu-nos verificar que conseguimos diminuir a taxa de nucleação e obter cristais de maiores dimensões. Verificamos ainda que a condição com 30% de PEG 4K é reprodutível e que conseguimos obter cristais a concentrações de precipitante menores que a inicial. Até à data, não obtivemos reprodutibilidade das condições E5 e B5 nesta tentativa de otimização.

É de referir que se continuou a observar o *screen* inicial de 80!, a 20°C, onde se verificou o aparecimentos de novos cristais em diversas condições (Figura 3.11).



**Figura 3.11** - Resultados obtidos para os ensaios de cristalização do complexo C4X10B utilizando o *screen* de 80!, a 20°C. As percentagens de agente precipitante e os tamanhos dos cristais, 0,4 mm, estão esquematizados na figura. [1] – Concentração de 19,49 mg/mL; [2] – Concentração de 9,75 mg/mL.

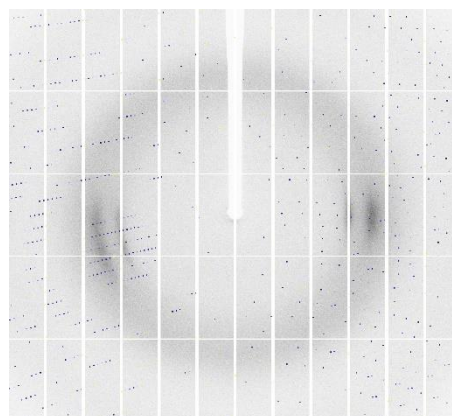
Tendo em conta estes novos resultados, e os obtidos anteriormente, decidiu-se realizar uma nova tentativa de otimização, novamente em robot, para as condições de cristalização A6 e E4 do *screen* do 80!. Decidiu-se variar novamente a percentagem de PEG 4K para a condição A6 (21, 23, 25 e 30%) e para a condição E4 (15, 20, 25 e 30%), utilizando a proporção proteína:precipitante de 1:1. Nesta nova otimização foram também utilizados 1 e 10% de uma solução de micro-cristais pré-formados (*seeds*), para a condição A6 a 30% PEG 4K. Até à data apenas se observou um resultado positivo para o aparecimento de cristais na condição A6 com 10% de *seeds* (Figura 3.12).



**Figura 3.12** - Resultado obtido para a primeira semana de observação da nova otimização da condição A6, a 20°C. A percentagem de agente precipitante e o tamanho dos cristais (0,4 mm) estão esquematizadas. A concentração de proteína utilizada tem o valor de 9,75 mg/mL.

Foram escolhidos um total 6 cristais de entre as condições A6 (0.2M CaCl<sub>2</sub>; 0.1M Acetato pH 4,5) com 20% PEG 4K (1:1) e A6 com 25% PEG 4K (1:1), para serem testados e medidos na linha ID29 do sincrotrão ESRF. Para isso, os mesmos foram previamente crio-preservados na presença de uma maior percentagem de PEG (neste caso entre 25 e 30% de PEG 4K, respetivamente) e de 15% de glicerol. As condições a testar foram escolhidas tendo em conta a morfologia dos cristais obtidos.

Foram recolhidos dois conjuntos de dados de difração de raios-X para este complexo C4X10B, com uma resolução de 2 e 1,83 Å, para cristais da condição A6 com 25% PEG 4K. Escolhemos o conjunto de dados com a melhor resolução (1,83 Å) para prosseguirmos com a análise de dados. A análise do padrão de difração (Figura 3.13) realizou-se aplicando os programas descritos na secção 3.2.5.



**Figura 3.13** - Padrão de difração do cristal do complexo C4X10B.

O grupo espacial do cristal foi determinado como sendo P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>, pelo programa IMOSFLM, que processou os dados de difração de modo a obtermos o ficheiro mtz que contém toda a informação deste conjunto de dados (índices de Miller *hkl*, as intensidades das reflexões ( $I_{hkl}$ ) e os respetivos erros associados ( $\sigma_{hkl}$ ). O AIMLESS permitiu a confirmação do grupo espacial referido anteriormente e a

indicação dos parâmetros da célula unitária  $a=49,97$ ,  $b=62,15$ ,  $c=131,99$  e  $\alpha=\beta=\gamma=90^\circ$ , podendo assim afirmar-se que o cristal pertence ao sistema cristalino ortorrômbico. Para além de processar os dados, o AIMLESS fornece um conjunto de estatísticas sobre a recolha dos dados (Tabela 3.6). Foi também estimada a percentagem de solvente pelo programa MATTHEWS.

**Tabela 3.6** -Estatísticas de recolha de dados para o cristal do complexo Coh-Doc C4X10B.

Parâmetro	Resultados globais	Camada externa de resolução
Limites de resolução (Å)	30,25- 1,81	1,85-1,81
$\langle I/\sigma(I) \rangle$	7,3	3,4
Multiplicidade	4,8	5,0
Completeness (%)	99,3	99,6
$R_{\text{merge}}$	0,136	0,398
Coefficiente de Matthews (Å)	1,97	
Conteúdo em solvente (%)	37,62	

A análise dos dados fornecidos pelo AIMLESS e MATTHEWS permitem, numa primeira visualização, verificar que os dados são confiáveis pelos valores de alta resolução e de  $I/\sigma(I)$ . O parâmetro global de  $\langle I/\sigma(I) \rangle$  indica que a intensidade é, aproximadamente, 7,3 vezes superior ao ruído-de-fundo associado.

A multiplicidade relaciona o número total de reflexões existentes num conjunto de dados em estudo e o número de reflexões únicas. Quanto maior é esse valor, maior é o número de reflexões únicas e mais confiável é o conjunto de dados. Por outro lado, a plenitude dos dados - *completeness* - compara a percentagem de reflexões medidas com o número de reflexões totais que são possíveis para uma dada resolução. O valor de *completeness* para os dados globais e dados da camada externa revelam que foram medidas praticamente todas as reflexões possíveis tendo em conta a nossa resolução.

O  $R_{\text{merge}}$  reflecte a concordância entre as reflexões simetricamente equivalentes, pois estas devem possuir uma intensidade semelhante. Por conseguinte, o valor de  $R_{\text{merge}}$  não deve ser elevado. Os dados são tipicamente truncados quando o valor de  $R_{\text{merge}}$  excede  $\sim 0,6$  a  $0,8$ . Estando os nossos valores, quer global quer de camada externa, abaixo deste intervalo de referência, podemos considerar mais uma vez que estamos perante um conjunto de dados de elevada qualidade.

Por último, o coeficiente de Matthews relaciona o volume da célula unitária com o peso molecular da proteína em estudo, tal como do número de unidades assimétricas presentes. Este parâmetro permite calcular o conteúdo em solvente presente no cristal, que deve estar compreendido entre os valores de 30 e 70%. Para o cristal do C4X10B temos a presença de duas moléculas por unidade assimétrica e um conteúdo em solvente de 40%.

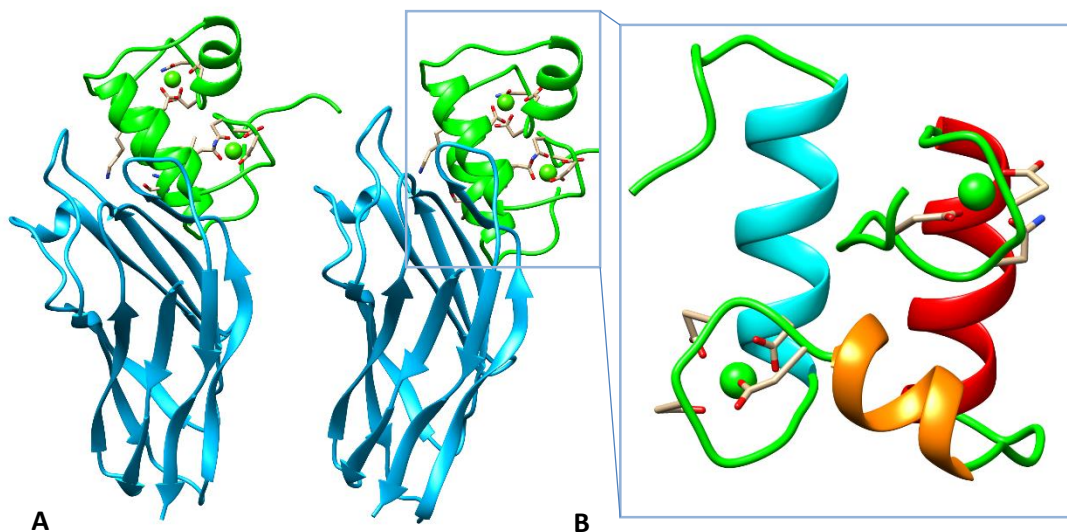
Em suma, a análise destas estatísticas apontaram para a boa qualidade dos dados de difração, seguindo-se assim para a determinação da estrutura tridimensional do complexo.

#### **3.3.5.2.2. Resolução da estrutura tridimensional do complexo C4X10B**

A partir do ficheiro mtz proveniente do AIMLESS e do programa PHASER, foi possível determinar as fases utilizando a coesina do modelo 2CCL do PDB (*Protein Data Bank*), para a resolução por Substituição Molecular. Este modelo foi o escolhido pois além de pertencer ao mesmo organismo (*C.thermocellum*), possui uma semelhança estrutural visto o módulo doquerina ser o mesmo e por isso estar na mesma orientação. O programa orienta e tenta posicionar a coesina do modelo 2CCL na densidade eletrónica calculada através de operações de translação e de rotação. O PHASER encontrou uma solução que englobou duas moléculas por unidade assimétrica, No final foi gerado um novo ficheiro mtz, contendo as fases determinadas e os fatores de estrutura, assim como um ficheiro pdb, que contém a informação sobre a possível posição dos átomos do nosso complexo Coh-Doc. Após melhorarmos as fases da estrutura obtida com ajuda do programa DM, o programa BUCCANEER permitiu fazer o rastreio da cadeia da doquerina no mapa de densidade eletrónica calculado. Em resultado a esta procura obtivemos resultado positivo para dois complexos C4X10B, com duas cadeias (A – coesina C4; B – doquerina Xyn10B) cada um. No final foi gerado um novo ficheiro mtz e um novo ficheiro pdb com as fases melhoradas e as novas posições atómicas, respetivamente. Estes ficheiros foram visualizados em conjunto no programa COOT. A estrutura foi refinada pelo programa REFMAC5. Este melhora a posição dos átomos na densidade eletrónica. Durante os vários ciclos de refinamento a estrutura continuou a ser visualizada e ajustada com auxílio do programa COOT. Comparando os valores de factor R obtidos no primeiro refinamento (0,2136) e o valor do último refinamento (0,1901) verificou-se um decréscimo do factor R. Este valor está nos parâmetros ideais (menor que 0,2), sendo indicador de um modelo tridimensional do complexo C4X10B que melhor se aproxima dos dados experimentais medidos por difração de raios-X.

#### **3.3.5.2.3. Resolução da estrutura tridimensional do complexo C4X10B**

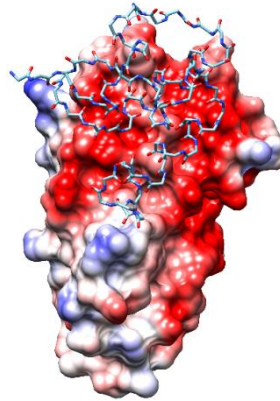
A Figura 3.14-A permite a visualização da estrutura obtida para cada molécula encontrada na unidade assimétrica. A azul está destacado o módulo coesina C4 e a cor verde o módulo doquerina, Xyn10B.



**Figura 3.14** - A - Representação das duas moléculas de C4X10B presentes na unidade assimétrica. B - Representação da estrutura tridimensional do módulo Xyn10B. Este apresenta um enrolamento em três hélices  $\alpha$  (Ser17B-Val27B – vermelho ; Asp35B-Ala41B - laranja; Ala51B-Leu61B - Azul). As imagens criadas foram conseguidas através da utilização do programa Chimera.

A semelhança das duas estruturas do complexo C4X10B encontradas na unidade assimétrica são confirmadas pelo valor baixo de RMSD (0,383 Å). Estas estão relacionadas por uma operação de translação. Cada módulo coesina C4 está organizado numa estrutura em barril  $\beta$  de 12 cadeias  $\beta$  dispostas em *jellyroll* de duas folhas  $\beta$  e de uma hélice  $\alpha$  entre as cadeias  $\beta$  5 e 6. Já o módulo doquerina Xyn10B está organizado em três hélices  $\alpha$  (Ser17B-Val27B; Asp35B-Ala41B; Ala51B-Leu61B), estas dispostas segundo uma conformação *loop-helix-loop*, seguida de *helix-loop-helix*, conetados por nove resíduos (Figura 3.14-B). Estão presentes dois iões  $\text{Ca}^{2+}$  coordenados pelos resíduos pertencentes à estrutura *EF-hand*. Um dos iões cálcio está próximo do N-terminal, sendo coordenado pelos resíduos Asp8B, Asn10B, Asp12B, Thr14B e Asp19B. O segundo ião cálcio é coordenado pelos resíduos (Asp42B, Asp44B, Asn46B, Ser48B e Asp53B) pertencentes ao *loop* presente entre as hélices 2 (laranja) e 3 (azul).

Os contactos entre os módulos C4 e Xyn10B que promovem a formação do complexo Coh-Doc estão localizados na face correspondente às folhas  $\beta$ 4,  $\beta$ 6,  $\beta$ 7 e  $\beta$ 10 do módulo coesina. A presença de resíduos hidrofóbicos também promove a interação Coh-Doc (Leu79A da coesina C4 e Leu62B). Estes dois módulos interagem também por uma série de ligações por pontes de hidrogénio (Tabela 3.9), mas também por pontes salinas, determinados pela ferramenta PDBePisa, da EMBL-EBI.



**Figura 3.15** - Superfície de potencial eletroestático do módulo coesina C4 do complexo C4X10B. O módulo Xyn10B está representado pela sua cadeia principal.

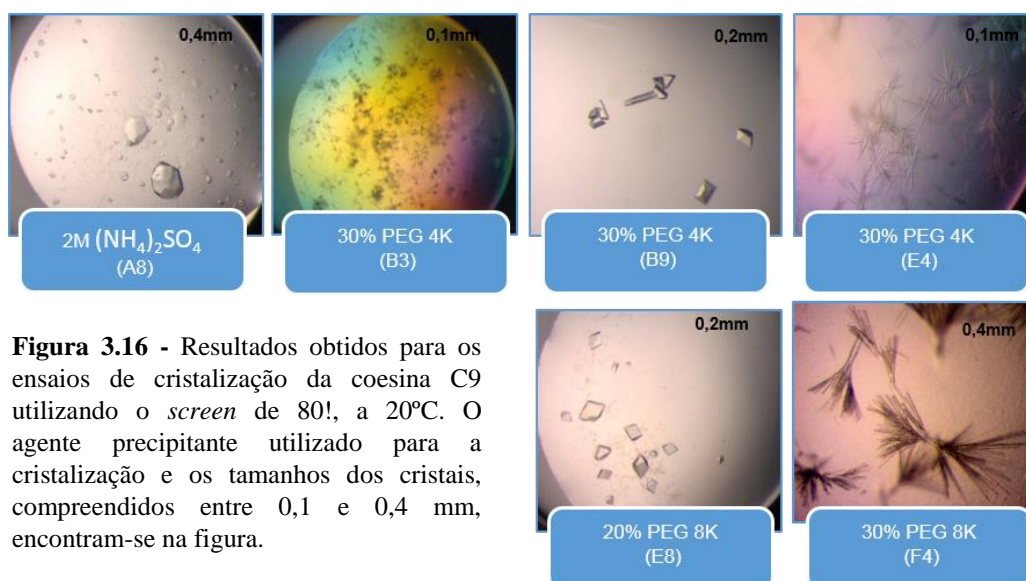
A Figura 3.15 representa o potencial electroestático calculado para o módulo coesina C4, revelando que a face que interage com o módulo Xyn10B possui, predominantemente, um potencial carregado negativamente.

### **3.3.5.3. Módulo coesina C9**

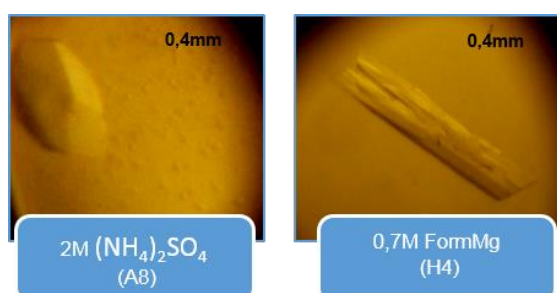
#### **3.3.5.3.1. Ensaio de cristalização e experiência de difração de raios-X**

Os ensaios de cristalização para esta amostra foram efetuados utilizando-se o *screen* de 80 condições com a proteína na concentração de 11,14 mg/mL. A formação de cristais foi testada às temperaturas de 4 e 20°C, utilizando o *robot Oryx8* da Douglas Instruments (técnica da gota assente) e as placas de cristalização de 24 reservatórios (técnica da gota suspensa). Durante os vários protocolos de otimização variou-se a proporção de proteína-precipitante e as percentagens de agente precipitante. Recordamos que as experiências de cristalização visavam a cristalização do complexo C9X10B, mas, após a resolução da estrutura tridimensional, verificou-se que apenas o módulo coesina C9 integrou os cristais obtidos.

Durante a primeira semana desde da preparação das gotas, foi observada a formação de precipitado em algumas condições, em ambas as temperaturas, indicando que, mais uma vez, nestes casos, o equilíbrio proteína-precipitante se encontra na zona de sobressaturação. Foram também visualizadas várias condições com a presença de cristais (Figura 3.16 e 3.17).

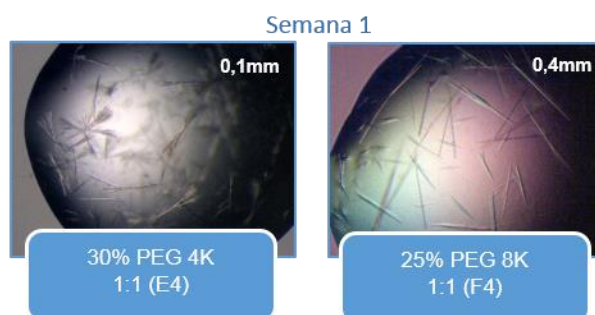


**Figura 3.16** - Resultados obtidos para os ensaios de cristalização da coesina C9 utilizando o *screen* de 80!, a 20°C. O agente precipitante utilizado para a cristalização e os tamanhos dos cristais, compreendidos entre 0,1 e 0,4 mm, encontram-se na figura.



**Figura 3.17** - Resultados obtidos para os ensaios de cristalização utilizando o *screen* de 80!, a 4°C. O agente precipitante utilizado para a cristalização e os tamanhos dos cristais (0,4 mm) encontram-se na figura.

Face a estes resultados foram escolhidas três condições (B3, E4 e F4) para otimização em que se obtiveram cristais com morfologia semelhante aos cristais do complexo C4X10B. Os cristais das condições A8, B9 e E8 foram utilizados em posteriores testes de difração *in house*. A observação das fotografias dos cristais obtidos para as condições B3 e E4 revela uma elevada taxa de nucleação e a presença de cristais em forma de agulhas muito finas e de pequenas dimensões. Na condição F4 estamos perante cristais de maiores dimensões mas difíceis de serem manipulados devido à morfologia do conjunto de cristais presentes. A otimização teve por objetivo a diminuição da taxa de nucleação tal como a tentativa de obter cristais com maiores dimensões. Para isso testámos, para todas estas condições, diferentes proporções proteína:precipitante (1:1; 0,5:1,5) e diferentes concentrações do agente precipitante PEG (10, 15, 20, 25 e 30%). Para cada uma das três condições fizemos a repetição da proporção, 1:1, da condição inicial com o objetivo de ver a reprodutibilidade da condição. É de referir que esta otimização foi realizada com a concentração de 11,14mg/mL e foi realizada em *robot* e manualmente, novamente de modo a verificarmos a reprodutibilidade das condições.



**Figura 3.18** - Resultados obtidos para a otimização das condições E4 e F4 realizadas manualmente, para a temperatura de 20°C. As proporções de proteína:precipitante e as percentagens de agente precipitante estão esquematizadas na figura. Os tamanhos dos cristais estão compreendidos entre os 0,1 e 0,4 mm. A concentração de proteína utilizada tem o valor de 11,14mg/mL

A observação da placa de otimização feita no *robot* levou-nos a concluir que não houve reprodutibilidade nas diferentes condições pois, até à data não houve aparecimento de cristais. Por outro lado, nas placas realizadas manualmente, conseguimos reproduzir a condição E4 inicial (30% PEG 4K, 1:1) e ainda obter novos cristais para a condição F4 (25% PEG 8K, 1:1) (Figura 3.18). Os cristais obtidos para a condição E4 continuaram morfologicamente semelhantes aos obtidos no *screen* inicial de 80! para a temperatura de 20°C. Os novos cristais obtidos para a condição F4 revelam que diminuámos a nucleação conseguindo assim uma otimização positiva.

A próxima etapa passou pelo teste e medição, quer *in house* (D8 Venture, Bruker), quer na linha ID23-1 do sincrotrão ESRF, dos cristais escolhidos para a experiência de difração. Esta etapa teve por objetivo, além da recolha de dados, a comparação da qualidade dos dados usando fontes diferentes de raios-X para os mesmos cristais. Para isso, os mesmos foram previamente criopreservados na presença de diferentes concentrações de agente precipitante e percentagens de glicerol (Tabela 3.7).

**Tabela 3.7** -Condições escolhidas para a experiência de difração e respectivos concentrações de agente precipitante e percentagens de glicerol utilizadas para a realização da criopreservação dos cristais.

Condições	Concentração de agente precipitante	% (V/V) Glicerol
A8 (0,2M Citrato de sódio) com 2M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> (1:1)	2,2M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub>	30
B9 (0,2M CaCl <sub>2</sub> ; 0,1M Hepes pH 7,5) com 30% PEG 4K (1:1)	35% PEG 4K	15
E8 (0,2M Acetato de magnésio; 0,1M Cacodilato pH 6,5) com 20% PEG 8K (1:1)	25% PEG 8K	15
F4 (0,2M (NH <sub>4</sub> ) <sub>2</sub> SO <sub>4</sub> ) com 25% PEG 4K (1:1)	30% PEG 4K	10

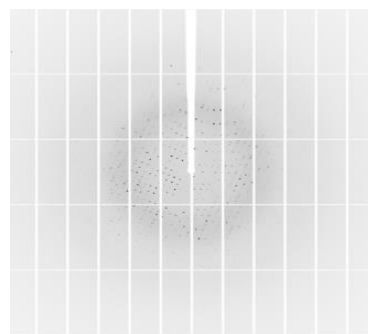
A recolha *in house* permitiu verificarmos a qualidade dos cristais escolhidos e confirmar se seriam constituídos por proteína. Foi recolhida uma imagem para cada cristal das condições A8, B9 e E8. Foi possível assim verificar que os cristais da condição B9 não eram de proteína, mas sim do respetivo composto iónico. Para a experiência de difração na linha ID23-1 no sincrotrão ESRF foram levados 1 cristal da condição A8, 1 cristal da condição F4 e 4 cristais para a condição E8 para serem testados. O

cristal crio-preservedo para a condição F4 não foi encontrado no *loop* aquando da recolha de dados. No ESRF foram recolhidos cinco conjuntos de dados, um para a condição A8 e os outros quatro para a condição E8. A análise dos conjuntos de dados realizou-se aplicando os programas discriminados na secção 3.2.5.

O programa IMOSFLM permitiu uma primeira avaliação dos conjuntos recolhidos, determinando os grupos espaciais como sendo  $P2_12_12_1$ . É de referir que um dos conjuntos de dados (x2)

foi avaliado pelo XDS, pois não foi possível realizar a integração das imagens no iMOSFLM, conseguindo-se confirmar o mesmo grupo espacial. Os programas utilizados após esta primeira avaliação originaram diferentes conclusões. A Tabela 3.8 resume todos os passos realizados para cada conjunto de dados.

A análise da tabela permitiu-nos verificar que, dos cinco conjuntos de dados, apenas três são destacados pela sua possibilidade de progressão de estudo, são eles: x2, x3 e x5. De forma a escolhermos o melhor conjunto para dar continuidade à determinação da estrutura da proteína em questão, decidimos comparar a qualidade dos mesmos. O AIMLESS permitiu a confirmação do grupo espacial para os três conjuntos de dados e a indicação dos parâmetros da célula unitária, podendo assim afirmar-se que os cristais pertencem ao sistema cristalino ortorrômbico. Foi também comparado o grupo espacial destes conjuntos com o dos dados obtidos para o complexo C4X10B ( $a=49,97$ ,  $b=62,15$ ,  $c=131,99$  e  $\alpha=\beta=\gamma=90^\circ$ ), no intuito de percebermos qual poderia ser a melhor opção de estudo. Para o cristal x6 não foi possível indexar os dados pois as constantes da célula mudavam consoante cada imagem, podendo assim tratar-se de um cristal múltiplo. Observando a tabela verificamos que o conjunto de dados x2 é o que tem as constantes da célula unitária mais semelhantes ao complexo C4X10B. Por outro lado, o conjunto de dados x3 e x5 possuem uma elevada semelhança. No entanto é o conjunto de dados x5 que melhor resolução apresenta.



**Figura 3.19**-Padrão de difração do cristal do módulo C9.

**Tabela 3.8** - Tabela resumo dos programas e respectivos resultados obtidos na análise de cada conjunto de dados.

Conjunto de dados	IMOSFLM	XDS	AIMLESS	MATTHEWS	PHASER
x2	-	P2 <sub>1</sub> 2 <sub>1</sub> 2	a= 50,97; b= 63,99 c=126,08; $\alpha=\beta=\gamma$ Resolução máxima=2,04Å	2 moléculas na unidade assimétrica 40% solvente	Pesquisa do complexo modelo (2CCL) e da coesina (2CCL)
x3	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	-	a= 40,67; b= 75,51 c=93,43; $\alpha=\beta=\gamma$ Resolução máxima=2,05Å	1 molécula na unidade assimétrica 55% de solvente	-
x4	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	-	Cristais múltiplos	-	-
x5	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>	-	a=41,14; b= 78,22 c=95,13; $\alpha=\beta=\gamma$ Resolução máxima=1,84Å	1 molécula na unidade assimétrica 62% solvente	Pesquisa do complexo modelo (2CCL) e da coesina (2CCL)
x6	Não foi possível indexar os dados				

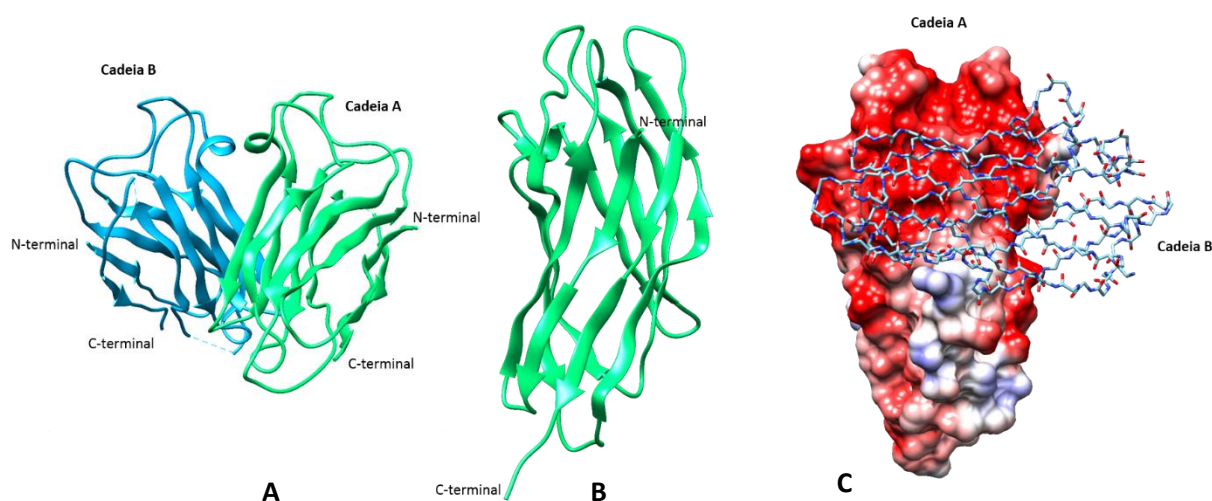
### 3.3.5.3.2. Resolução da estrutura tridimensional do módulo coesina C9

Até percebermos qual o melhor conjunto de dados, se o x2 ou o x5, foi decidido continuar a análise dos dois. A partir do ficheiro mtz proveniente do AIMLESS e do programa PHASER, tentou-se determinar as fases, utilizando novamente o modelo 2CCL do PDB (*Protein Data Bank*), para a resolução por Substituição Molecular. O programa tentou orientar e posicionar o complexo do modelo 2CCL na densidade eletrónica calculada através de operações de translação e de rotação, sem qualquer sucesso. Ao efetuar o mesmo com o módulo coesina C2 do modelo 2CCL obtivemos um resultado positivo para o conjunto de dados x5. Voltou-se a correr o PHASER, desta vez utilizando a doquerina Xyn10B do modelo 2CCL. Esta procura não teve sucesso. O PHASER encontrou assim apenas uma solução que englobava duas cópias de coesina C9 por cada célula unitária. Chegou-se assim à conclusão que poderíamos estar na presença de duas moléculas de C9 na unidade assimétrica e não perante o complexo C9X10B. No final foi gerado um novo ficheiro mtz, contendo as fases determinadas e os fatores de estrutura, assim como um ficheiro pdb, que contém a informação sobre a possível posição dos átomos das duas coesinas C9. Após melhorarmos as fases com ajuda do

programa DM, ainda se efetuou um rastreio da cadeia da doquerina, pelo programa BUCCANEER, no mapa de densidade eletrónico obtido, mas mais uma vez a doquerina não foi encontrada. O ficheiro mtz e o ficheiro pdb com as fases melhoradas, provenientes do programa DM e do PHASER foram visualizados em conjunto no programa COOT. A estrutura foi refinada pelo programa REFMAC5. Durante os vários ciclos de refinamento a estrutura continuou a ser visualizada e ajustada com auxílio do programa COOT. Comparando os valores de fator R obtidos no primeiro refinamento (0,2907) e o valor do último refinamento (0,2529) verificou-se um decréscimo do fator R, Este valor não está nos parâmetros ideais (menor que 0,2), mas o seu decréscimo é indicador de uma correta construção da estrutura tridimensional das duas moléculas de C9 na unidade assimétrica.

### 3.3.5.3.3. Caracterização da estrutura tridimensional do módulo C9

A Figura 3.20 permite a visualização da estrutura obtida para cada molécula encontrada na unidade assimétrica. A ferramenta PDBePisa permitiu verificar que as duas moléculas não formem um dímero fisiológico, confirmando que não se está perante um complexo formado entre dois módulos coesina C9.



**Figura 3.20** - Representação das duas moléculas de C9 presentes na unidade assimétrica; B - Representação da estrutura tridimensional do módulo C9; C - Superfície de potencial electrostático do módulo coesina C9. As imagens foram conseguidas através da utilização do programa Chimera.

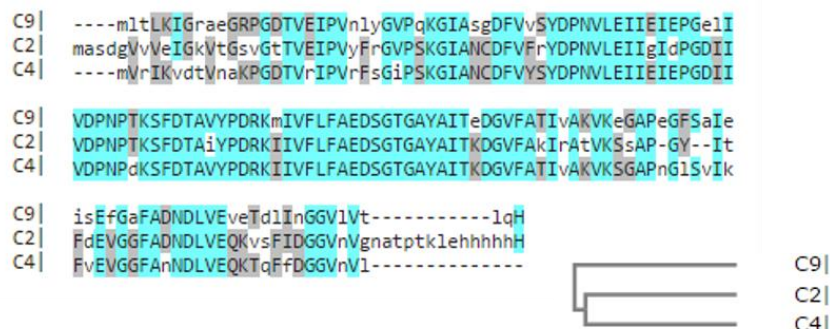
Conforme descrição realizada para o complexo C4X10B, verificamos a semelhança das duas estruturas presentes na unidade assimétrica pelo valor baixo de RMSD (0,414 Å). Cada módulo coesina C9 está organizado numa estrutura em barril  $\beta$  de 12 cadeias  $\beta$ , dispostas em *jellyroll* e de uma hélice  $\alpha$  entre as folhas  $\beta 5$  e  $\beta 6$ , tal como visto anteriormente para o complexo caracterizado anteriormente. Estes dois módulos estão relacionados por um eixo de rotação, estando viradas para a mesma interface (formada pelas cadeias  $\beta 4$ ,  $\beta 6$ ,  $\beta 7$ , e  $\beta 10$ ). O cálculo do potencial electrostático de

superfície (Figura 3.20-C) revela que a interface de interação com o segundo módulo coesina está, predominantemente, carregada negativamente. Acredita-se que o excesso de coesina C9 ainda presente, nas frações utilizadas nos ensaios de cristalização, tenha diminuído a probabilidade de haver formação e cristalização do complexo C9X10B. Por outras palavras, a formação deste dímero não cristalográfico impediu a ligação do doquerina Xyn10B.

### 3.3.6. Comparação das estruturas C2-Xyn10B, C4X10B e C9

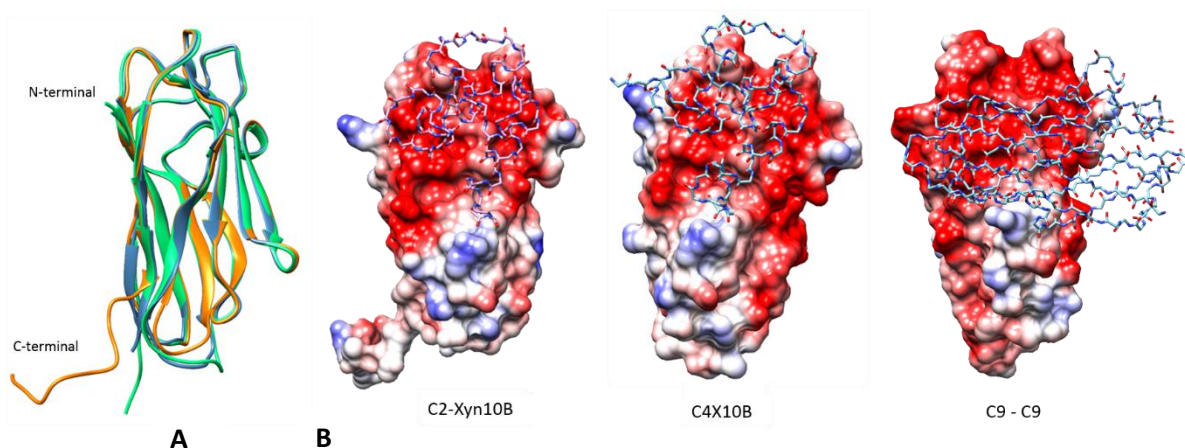
A resolução das estruturas descritas anteriormente, C4X10B e C9, foram realizadas considerando-se o complexo C2-Xyn10B (PDB:2CCL) como modelo. Nesta secção é feita a comparação entre as três estruturas - C2-Xyn10B, C4X10B e C9 – de modo a identificarmos possíveis semelhanças e diferenças que poderão estar implícitas na montagem/organização do celulosoma em estudo.

Numa primeira etapa comparou-se as três estruturas. O alinhamento das respetivas sequências (Figura 3.21) na ferramenta bioinformática MUSCLE (EMBL-EBI) permite-nos confirmar a coerência das percentagens de identidade que a Figura 6 da secção 1.3.1.2, verificando que a maioria do alinhamento reflete o mesmo conteúdo em resíduos.



**Figura 3.21** - Alinhamento múltiplo das sequências dos módulos coesina C2, C4 e C9 pela ferramenta bioinformática MUSCLE, e respetiva árvore filogenética

A sobreposição das estruturas obtidas (Figura 3.22-A) mostra-nos que são estruturalmente semelhantes, tendo-se obtido um valor de RMSD = 0,677Å. Por outro lado, o cálculo do potencial eletroestático (Figura 3.22-B) permitiu verificar que todos eles possuem, predominantemente, um potencial carregado negativamente.



**Figura 3.22** - A- Sobreposição das estruturas C2-Xyn10B (Laranja), C4X10B (Azul) e C9 (verde); B-Superfície de potencial electrostático dos módulos coesina C2, C4 e C9. Nas figuras representantes dos complexos C2-Xyn10B e C4X10B, o módulo Xyn10B está representado pela sua cadeia principal.

Numa segunda etapa, comparou-se, no caso dos dois complexos C2-xyn10B e C4X10B, a participação dos resíduos de cada sequência na formação do complexo Coh-Doc. A Tabela 3.9 apresenta os resíduos do módulo do querina Xyn10B envolvidos nas ligações por pontes de hidrogénio e os respetivos resíduos do módulo coesina, de cada complexo, a que está associada essa ligação. O PDBePisa (EMBL-EBI) foi a ferramenta utilizada para a construção da tabela.

**Tabela 3.9** - Ligações por pontes de hidrogénio entre o módulo Xyn10B e os módulos C2 e C4

C2	Xyn10B	C4
A:GLU 131[ OE2]	B:ASN 10[ ND2]	...
A:ASP 39[ OD1]	B:SER 17[ N ]	A:ASP 35[ OD1]
A:ASP 39[ OD1]	B:SER 17[ OG ]	A:ASP 35[ OD1]
A:GLY 123[ O ]	B:SER 17[ OG ]	A:GLY 122[ O ]
A:ASN 37[ OD1]	B:SER 17[ OG ]	A:ASN 33[ OD1]
A:ASP 39[ N ]	B:SER 17[ OG ]	A:ASP 35[ N ]
A:GLU 131[ OE2]	B:THR 18[ OG1]	A:GLU 130[ OE1]
A:ASN 37[ OD1]	B:THR 18[ OG1]	A:ASN 33[ OD1]
...	B:THR 21[ OG1]	A:ALA 32[ O ]
A:ASN 37[ ND2]	B:THR 21[ OG1]	A:ASN 33[ ND2]
A:LYS 63[ O ]	B:LYS 24[ NZ ]	...
A:ALA 85[ O ]	B:LYS 24[ NZ ]	A:ALA 81[ O ]
A:ASN 127[ O ]	B:ARG 25[ NH1]	A:ASN 126[ O ]
A:GLU 86[ OE2]	B:ARG 25[ NH2]	A:GLU 82[ OE2]
A:ASN 127[ OD1]	B:ARG 25[ NH2]	A:ASN 126[ OD1]
A:GLU 120[ OE1]	B:ARG 57[ NH2]	...
A:TYR 74[ OH ]	B:VAL 58[ N ]	...
A:TYR 74[ OH ]	B:LEU 62[ O ]	A:TYR 70[ OH ]
A:ARG 77[ NH1]	B:ARG 63[ O ]	A:ARG 73[ NH1]
A:ARG 77[ NH2]	B:ARG 63[ O ]	A:ARG 77[ NH2]

A visualização da tabela permite-nos verificar que o complexo C4X10B consegue assegurar 15 dos 18 contactos efetuados pelo complexo C2-Xyn10B. No caso da estrutura obtida para o módulo C9, tentou-se perceber que resíduos poderiam estar envolvidos na formação deste dímero não-cristalográfico. Numa primeira análise verificámos a existência de um contacto hidrofóbico entre os resíduos Leu79A da coesina C9 e Leu79B da segunda coesina. Este contacto parece substituir o contacto entre os resíduos Leu79A da coesina C4 e Leu62B da doquerina Xyn10B. Construiu-se também uma tabela que ilustra as ligações por pontes de hidrogénio que podem ser formadas entre os dois módulos C9 (Tabela 3.10.)

**Tabela 3.10** - Ligações por pontes de hidrogénio entre os módulos C9.

C9 - cadeia A	C9 - cadeia B
A:TYR 70[ OH ]	B:PHE 36[ N ]
A:ASP 35[ OD2]	B:TYR 70[ OH ]
A:GLU 119[ OE1]	B:PRO 71[ N ]
A:GLU 119[ OE1]	B:ARG 73[ N ]
A:PHE 36[ N ]	B:TYR 70[ OH ]
A:LYS 63[ NZ ]	B:ASP 83[ OD1]
A:LYS 63[ NZ ]	B:ASP 83[ O ]
A:TYR 70[ OH ]	B:ASP 35[ OD2]
A:ASP 72[ N ]	B:GLU 119[ OE1]
A:ASP 72[ N ]	B:GLU 119[ OE2]
A:ARG 73[ N ]	B:GLU 119[ OE1]
A:ARG 73[ NH1]	B:ARG 73[ O ]
A:ARG 73[ NH2]	B:SER 118[ OG ]

Comparando assim as ligações por pontes de hidrogénio encontradas nos complexos Coh-Doc e no dímero C9-C9, podemos verificar que: a ponte de hidrogénio encontrada entre os resíduos Asp35A da coesina C4 e Ser17 da doquerina Xyn10B é coincidente com a que é encontrada entre os resíduos Asp35A da coesina C9 e Tyr70B da segunda coesina C9; a encontrada entre os resíduos Tyr70A da coesina C4 e Leu62 da doquerina Xyn10B é coincidente com a que encontramos entre os resíduos Tyr70A da coesina C9 e Asp35B da segunda coesina C9; e ainda que a ponte de hidrogénio entre os resíduos Arg73A da coesina C4 e Arg63 da doquerina Xyn10B é coincidente com as encontradas entre o resíduo Arg73A da coesina C9 e os resíduos Arg73, Ser118 e Glu119 da segunda coesina C9.

Em suma, a ligação da segunda coesina C9, à mesma interface onde o módulo doquerina deveria estar ligado, parece compensar as ligações encontradas no complexo C4X10B, levando assim a crer que a formação do complexo C9-Xyn10B não foi conseguido devido ao excesso de coesina presente na amostra. Podemos, no entanto, inferir que esta interface coesina-coesina não-cristalográfica coincide com a interface que reconhecerá a respetiva doquerina na organização celulosomal.

### 3.4. Conclusão

Nos últimos anos, o estudo dos mecanismos de associação entre os vários módulos coesina e doquerina tem sido realizado por Cristalografia de Raios-X em combinação com estudos de mutagénesse. É neste contexto que o trabalho desenvolvido no âmbito da presente dissertação procurou a caracterização estrutural dos vários módulos coesina da proteína de agregação CipA em complexo com o módulo doquerina Xyn10B. Estudos prévios usando géis nativos tinham revelando diferenças na especificidade e pretendia-se determinar a razão estrutural para essas diferenças.

O primeiro passo consistiu na expressão e purificação dos complexos formados pelos módulos coesina e doquerina escolhidos para o estudo. Prosseguiu-se com ensaios de cristalização de modo a obter cristais de qualidade para a experiência de difração. De entre as várias condições testadas obtiveram-se apenas cristais nas condições 0.2M Acetato de amónio + 0.1M Citrato pH 5.5 + 25% PEG 4K para o complexo C4-Xyn10B e condições 0.2M Acetato de magnésio + 0.1M Cacodilato pH 6.5 + 20% PEG 8K para a coesina C9. A recolha de dados em sincrotrão dos cristais obtidos permitiu a determinação de duas estruturas tridimensionais, uma para o complexo C4-Xyn10B (1,83 Å) e uma para o módulo coesina C9 (1,84 Å). A determinação das fases foi realizada pelo método de Substituição Molecular, tendo sido possível localizar os respetivos módulos.

A análise da estrutura tridimensional obtida para o complexo C4-Xyn10B permitiu confirmarmos a ligação a uma das faces da doquerina (hélice 1) à face da coesina C4 composta pelas folhas  $\beta$ 4,  $\beta$ 6,  $\beta$ 7 e  $\beta$ 10. Já a análise tridimensional da estrutura obtida para o dímero não-fisiológico C9-C9, revelou que a formação deste dímero levou à não-formação do complexo C9-Xyn10B. Foi possível assim identificar a interface da coesina C9 (composta pelas mesmas folhas  $\beta$ 4,  $\beta$ 6,  $\beta$ 7 e  $\beta$ 10), semelhante à face dos módulos coesina C4 e C2 e que poderá estar envolvida na formação do complexo. A comparação do módulo coesina nestas duas estruturas (complexo C4-Xyn10B e dímero C9-C9) com o módulo coesina no modelo do complexo C2-Xyn10B (PDB 2CCL) possibilitou, numa primeira abordagem, verificar as semelhanças entre os módulos coesina C2, C4 e C9 na distribuição das estruturas secundárias presentes na estrutura tridimensional. De seguida, fez-se a comparação das estruturas tridimensionais referentes aos complexos C4X10B e C2-Xyn10B, o que permitiu confirmar que a ligação Coh-Doc é realizada pela mesma hélice da doquerina, ou seja, que esta apresenta a mesma face de ligação.

Em conclusão, o trabalho apresentado nesta primeira parte da dissertação forneceu informações estruturais sobre dois módulos coesina, C4 e C9, da CipA, até então desconhecidas, permitindo a futura caracterização da montagem/organização do celulosoma em estudo. Por outro lado, facultou-nos informações sobre a formação do complexo com o módulo Xyn10B.

Capítulo

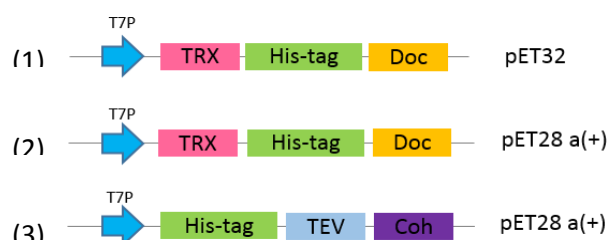
4

**Análise das interações coesina-doquerina no *C. thermocellum* e *R. flavefaciens* usando a tecnologia de *Microarrays***



## 4.1. Introdução

Neste capítulo, os módulos coesina e doquerina em estudo foram anteriormente clonados, pelo Professor Carlos Fontes e pela Doutora Joana Armada Brás (Faculdade de Medicina Veterinária), utilizando dois vetores, pET28a(+) e pET32, permitindo a inserção de uma cauda de 6 histidinas (*His-Tag*) no terminal N. Além da cauda, as doquerinas foram clonadas com um módulo tiorredoxina (TRX) e as coesinas com a sequência muito específica de reconhecimento para a protease do vírus do tabaco (Tobacco Etch Virus – TEV) (Figura 4.1). O módulo TRX está presente nos constructos das doquerinas pois confere-lhes estabilidade e permite a sua deteção por anticorpos monoclonais anti-TRX durante o ensaio nos *microarrays*. Já a sequência TEV, presente no constructo das coesinas, permite a remoção específica da cauda de histidinas da restante proteína. A utilização do vetor pET28a(+) é condicionado pelo uso do antibiótico canamicina (KAN), enquanto a utilização do vetor pET32 é condicionado pelo antibiótico ampicilina (AMP).



**Figura 4.1** - Constructos utilizados para expressão dos módulos coesina-doquerina. Os termos Doc e Coh correspondem aos módulos doquerina e coesina, respetivamente. As coesinas apresentam no seu N-terminal um módulo TEV e uma *His-tag*. As doquerinas possuem um módulo TRX e uma *His-tag* no seu N-terminal.

Durante este trabalho tentou-se caracterizar as diferentes afinidades e especificidades entre coesinas e doquerinas de duas espécies, *Clostridium thermocellum* (*C. thermocellum*) e *Ruminococcus flavefaciens* (*R. flavefaciens*). Para isso escolheram-se 26 coesinas (14 do *C. thermocellum* e 7 do *R. flavefaciens*) e 46 doquerinas (40 do *C. thermocellum* e 6 do *R. flavefaciens*). A lista de todos os módulos coesinas e doquerinas envolvidas neste estudo encontram-se no Anexo 7B.I. Estudos anteriores por eletroforese de poliacrilamida em condições nativas entre coesinas e doquerinas, usados para observar a formação de complexos coesina-doquerina em solução, mostraram diferenças de especificidade entre alguns destes módulos.<sup>[46]</sup> A técnica de *microarrays*, acoplada à eletroforese de poliacrilamida em condições nativas, poderá permitir uma primeira determinação de interações coesina-doquerina mais rápida.

## 4.2. Materiais e Métodos

### 4.2.1. Transformação de células competentes

A transformação de células competentes foi realizada tendo em conta o método descrito na secção 3.2.1. Neste capítulo, os transformantes foram selecionados através da adição do antibiótico KAN ou AMP ao meio LB, nas concentrações de 50 e 100µg/mL, respetivamente, consoante vetor empregue.

#### **4.2.2. Expressão dos módulos coesina e doquerina de tipo I e II**

Os módulos coesina e doquerina foram sobre-expressos em células de *E.coli* BL21 (DE3) em meio de cultura LB. Alguns módulos coesina e doquerina foram expressos num meio LB onde se substituiu a triptona por peptona.

Cada clone proveniente da transformação foi inoculado em meio de cultura LB, condicionado com o respetivo antibiótico, 50µg/mL KAN ou 100µg/mL AMP. O pré-inóculo foi incubado a 37°C, a uma rotação de 210 rpm durante a noite (Orbital Shaker-Incubator ES-20, da Grant.bio). No dia seguinte adicionou-se 2,5mL de cada pré-inóculo a um volume final de meio LB condicionado com KAN ou AMP, respeitando-se a proporção 1:200 e 1:100 para expressão dos módulos coesina e doquerina, respetivamente. Cada cultura foi incubada a 37°C, a uma rotação de 210 rpm, (Shaker IS-971R da Lab Companion), até se alcançar uma D.O.<sub>600nm</sub> de 0,6-0,8. Uma vez atingido este valor, cada cultura foi induzida com IPTG a uma concentração final de 1mM e incubada a 19°C, com uma agitação de 150 rpm durante a noite (Shaker IS-971R da Lab Companion). No terceiro dia, as culturas foram centrifugadas durante 15 minutos a uma temperatura de 4 ou 8°C, consoante rotor usado – velocidades de 8000 rpm (Rotor JA-14, Avanti J-26 XPI da Beckman Coulter) e 7000 rpm (Rotor GSA, RC 5C plus da Sorvall®) – respetivamente. O sobrenadante foi descartado e o *pellet* guardado a -20°C para posterior purificação e de dessalinização.

#### **4.2.3. Purificação dos módulos coesina e doquerina de tipo I e II**

Os módulos coesina e doquerina do tipo I e II foram purificados por uma cromatografia de afinidade por iões metálicos imobilizados (IMAC) e posteriormente dessalinizados.

Após recolha das células por centrifugação estas foram ressuspensas em tampão 50mM HEPES a pH 7,5, 1M NaCl, 5mM CaCl<sub>2</sub> e 10mM de imidazol (1g/10mL) juntamente com 1µg/µL de DNase I (Sigma-Aldrich), 300µg/µL de lisozima (Sigma-Aldrich), 10mM MgCl<sub>2</sub> e uma pastilha de inibidores de proteases (*SIGMAFAST™ Protease Inhibitor Tablets*). Posteriormente as células contendo os módulos coesina foram lisadas por ultrassons em 3 ciclos de 1 minutos (UP50H da Lab Device), recolhendo-se a fração solúvel após centrifugação (Rotor F-34-6-38, Centrífuga 5810R) a 10000 rpm, durante 30 minutos a 4°C. Já as células com os módulos doquerina foram submetidas a 2 ciclos de 1 minuto, recolhendo-se a fração solúvel por centrifugação (Rotor F-45-30-11, Centrífuga 5810R) a 12000 rpm, durante 30 minutos a 4°C.

##### **4.2.3.1. Cromatografia de afinidade por iões metálicos imobilizados**

A cromatografia de afinidade por iões metálicos imobilizados, tal como referido na secção 3.2.3.1. é conhecida pela sua elevada eficácia na separação de proteínas recombinantes às quais foram adicionadas uma cauda de histidinas. Cada módulo coesina e doquerina foi purificado por IMAC em colunas com matriz de Sepharose™ carregada de níquel, His GraviTrap™ de 1 mL, da GE-Healthcare.

Durante este protocolo foram realizados vários patamares de concentrações diferentes de imidazol (10, 60 e 250mM) em que se incluem os passos de equilíbrio da coluna, lavagem e de eluição das proteínas em estudo. Para todos os passos utilizou-se a mesma solução tampão base, 50mM HEPES a pH 7,5, 1M NaCl, 5mM CaCl<sub>2</sub>. As colunas His GraviTrap™ foram primeiramente lavadas com cerca de 10mL de água Mili-Q e posteriormente equilibradas com 10mL de tampão com 10mM de imidazol. Cada extrato foi carregado na sua respectiva coluna. Procedeu-se a uma lavagem inicial com 10mL do tampão com 60mM de imidazol. Eluiu-se os módulos coesina e doquerina com um volume de 5 e 2,5mL de tampão com 300mM de imidazol, respetivamente. Foi realizado o teste de Bradford para confirmar a presença das proteínas no produto eluído.

#### **4.2.4. Ensaios de ligação e de competição entre módulos coesina e doquerina**

A capacidade de coesinas e doquerinas interagirem entre si foi determinada por *microarray* de proteínas e eletroforese em gel de poliacrilamida, em condições nativas.

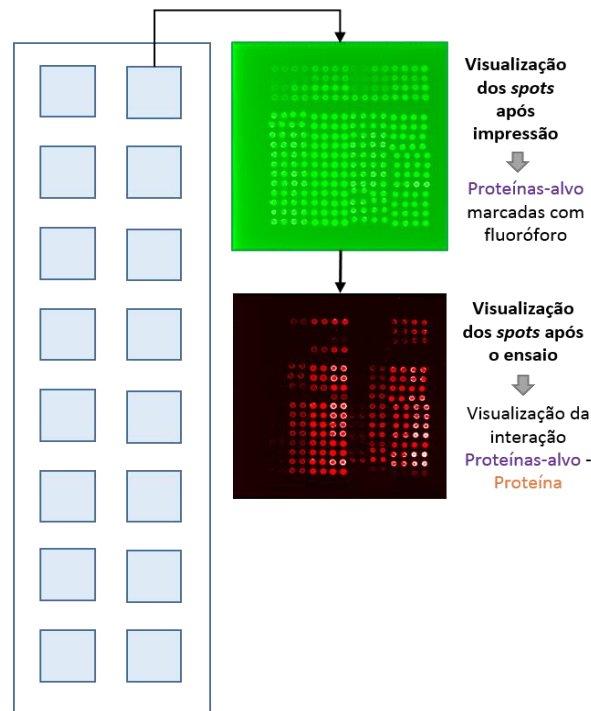
##### **4.2.4.1. Ensaios de ligação e de competição através da técnica de *microarray***

A técnica de *microarrays*, tal como referido na introdução (secção 1.4.1), pode ser dividida, após expressão e purificação das proteínas alvo, em 3 etapas: 1) Desenho do padrão da disposição das proteínas e a impressão destas mesmas proteínas num suporte sólido seguindo o padrão estabelecido; 2) Realização do ensaio de ligação; 3) Quantificação e análise dos resultados.

##### **1) Desenho do padrão da disposição das proteínas e a impressão destas mesmas proteínas num suporte sólido seguindo o padrão estabelecido**

Na primeira etapa foram imobilizados os módulos coesina em estudo, segundo o padrão apresentado em anexo (Anexos 7B.II e 7B.III), a diferentes concentrações (7,5, 30, 75 e 300µg/mL) em slides com uma matriz sólida de nitrocelulose. Nos vários ensaios foram utilizados slides com diferentes números (1, 2 e 16) de blocos de nitrocelulose onde são feitas as experiências de ligação. Nos slides com mais de um bloco é possível fazer mais do que uma experiência, pois cada bloco corresponde a uma experiência. A etapa de impressão, para os slides de 1 e 2 blocos foi realizada manualmente utilizando um *arrayer* manual (MicroCaster™ System da Whatman®, Schleider&Schuell). A impressão dos slides com 16 blocos foi realizada por um *arrayer* robotizado no laboratório da Professora Ten Feizi no Imperial College em Londres. Os módulos coesina imobilizados possuem 5µg/mL de Cy3 ( $\lambda_{excitação}=532nm$ ) na sua composição, um fluoróforo que nos permite verificar a localização e qualidade dos *spots* impressos antes de iniciarmos o ensaio, através da sua digitalização. Para um bom resultado de impressão é necessário que haja equidistância entre os vários *spots* e que estes tenham sempre o mesmo diâmetro (Figura 4.2). Esta visualização é realizada através da digitalização do respetivo slide impresso num equipamento especializado chamado de *scanner* de fluorescência de *microarrays*.

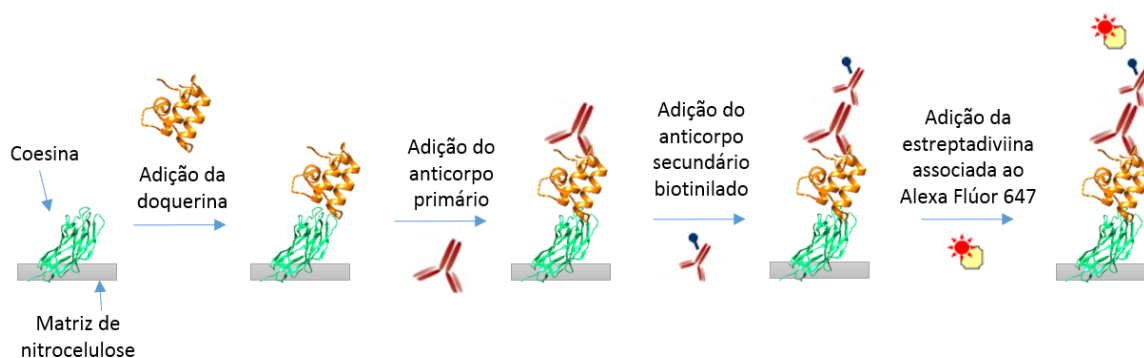
É importante referir que num ensaio de *microarrays* de proteínas a detecção pode ser feita por fluorescência. Entre os marcadores fluorescentes mais utilizados, também denominados de fluoróforos, encontramos as cianinas Cy3 e Cy5. Estes são amplamente utilizados nestes ensaios devido à sua mínima interação com outras biomoléculas e a sua alta compatibilidade com os *lasers* disponíveis nos equipamentos para *microarrays*.<sup>[41][45]</sup>



**Figura 4.2-** Representação esquemática da visualização do sinal de fluorescência emitido por um bloco após impressão e após o ensaio.

## 2) Realização do ensaio de microarrays

O método da biotina-estreptavidina (Figura 4.3) é muito utilizado para a detecção de um ensaio de *microarrays*. Passa pela incubação de um primeiro anticorpo, que ao reconhecer a proteína que interage com a proteína alvo imobilizada, forma um complexo proteína-anticorpo primário. É posteriormente adicionado um segundo anticorpo marcado por ligação à proteína biotina. A presença desta proteína leva a que a interação proteína alvo-proteína-anticorpo primário-anticorpo secundário seja reconhecida por uma outra proteína incubada posteriormente – a estreptavidina. A forte interação biotina-estreptavidina permite a detecção da interação proteína alvo-proteína, sendo esta visualizada por nova digitalização do slide após o ensaio. É ainda importante referir que esta visualização e posterior quantificação são apenas conseguidas pois a estreptavidina está, também ela, marcada com um fluoróforo.



**Figura 4.3** - Representação esquemática dos principais passos de um ensaio de *microarrays* de proteínas.

Cada ensaio deve ser acompanhado de pelo menos uma experiência de controle. Um deles passa pela detecção das coesinas imobilizadas através do anticorpo primário anti-histidinas ( $\alpha$ -His), produzido em ratinho, da Sigma-Aldrich. Neste controle  $\alpha$ -His, o anticorpo ligar-se-á às coesinas imobilizadas, pois estas possuem uma cauda de histidinas. A consequente ligação do anticorpo secundário e interação biotina-estreptavidina permitem também neste caso visualização dos *spots* e confirmação da correta imobilização dos módulos coesina. O outro controle passa pela realização de uma experiência em que não foi adicionado nenhum módulo doquerina. O segundo passo passa pela adição do anti-tiorredoxina ( $\alpha$ -TRX). Ao utilizarmos este controle  $\alpha$ -TRX e, não havendo adição de qualquer doquerina e consequente formação do complexo coesina-doquerina, não ocorrerá ligação de nenhum dos anticorpos, e não haverá qualquer emissão de fluorescência nesses *spots*.

Neste trabalho, os complexos coesina-doquerina e os módulos doquerinas, utilizados nos ensaios de competição e de ligação, respetivamente, foram testados às concentrações de 0,01 e 0,1 $\mu$ g/mL.

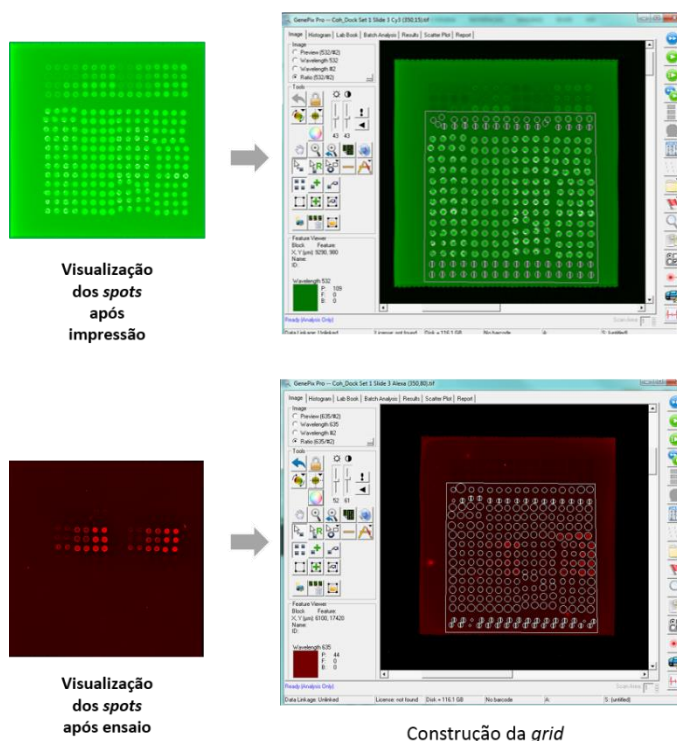
Após impressão e digitalização do(s) slide(s) cada bloco foi humidificado com água Mili-Q e o excesso retirado por movimentos suaves ou por inversão do mesmo. Cada slide foi posteriormente exposto à solução de bloqueio (3% BSA, 1% de caseína) que permite bloquear as ligações inespecíficas e promover as ligações coesina-doquerina. O passo seguinte passou pela adição do módulo doquerina a testar na respetiva concentração em estudo. A detecção é conseguida por adição de 1 $\mu$ g/mL de estreptavidina conjugada com o fluoróforo Alexa Flúor 647 ( $\lambda_{excitação}=635$ nm) após incubação inicial dos anticorpos primário – anticorpo monoclonal anti-tiorredoxina ( $\alpha$ -TRX), produzido em ratinho, da Sigma-Aldrich – e secundário - anticorpo IgG–Biotilado, da Sigma-Aldrich. Os módulos doquerinas ao possuírem um módulo tiorredoxina ligado, são inicialmente reconhecidos por um anticorpo primário. A seguir, o anticorpo secundário biotilado é reconhecido pela estreptavidina. A forte interação biotina-estreptavidina permitiu, após *scan* final do(s) slide(s), visualizar os resultados e realizar a posterior análise e quantificação. É ainda importante referir que entre cada passo de adição, o(s) slide(s) devem ser incubados à temperatura ambiente, protegidos da luz direta. Por outro lado, todas as diluições necessárias à preparação das soluções foram realizadas

em 1% BSA, 0,1% caseína em tampão HBS (5 mM Hepes pH 7.4; 150 mM NaCl, CaCl<sub>2</sub> 5 mM). O protocolo detalhado encontra-se apresentado em anexo (Anexo 7B.IV).

### 3) Quantificação e análise dos resultados

Tal como mencionado anteriormente, no final de cada ensaio é realizado uma segunda digitalização - *scanner* GenePix<sup>®</sup> 4300A (Molecular Devices) - do sinal de fluorescência para a energia do *laser* selecionado (635 nm) no intuito de visualizarmos e quantificar os resultados obtidos. Os programas associados aos equipamentos de *scanner* de fluorescência de *microarrays*, tais como o programa GenePix Pro da Molecular Devices, permitem a análise das imagens obtidas. Para a digitalização do slide é necessário a escolha do comprimento de onda do *laser* pretendido, tendo em conta os fluoróforos utilizados durante o ensaio. Após este primeiro passo de aquisição de imagem, o processamento da quantificação pode ser dividido em três etapas: 1) criação de uma matriz/grelha que engloba todos os *spots* testados; 2) quantificação das intensidades de fluorescência de cada *spot* e 3) apresentação dos resultados.

Numa primeira fase, é necessária a criação de uma matriz/grelha que engloba todos os *spots* de cada bloco. A grelha/matriz é criada usando a primeira digitalização como referência (após impressão do slide), e depois ajustada à segunda digitalização para depois ser feita a quantificação (no final do ensaio) – Figura 4.4.



**Figura 4.4** - Representação esquemática da construção da grelha para cada digitalização do slide, através do programa GenePix Pro da Molecular Devices.

Esta mesma matriz/grelha, guardada como ficheiro .gps, vai permitir analisar todas as intensidades de fluorescência presentes em cada bloco. Por outras palavras, o programa irá determinar a intensidade na região ocupada pelo *spot*, bem como na região fora do *spot* (*background*). A intensidade final de cada *spot* é calculada tendo em conta a região do mesmo, o *background* e a área do *spot*, cuja intensidade de *pixéis* é maior do que a dos *pixéis* que se encontram fora do *spot*, mas dentro da região da matriz/grelha. Uma vez realizada esta análise, o programa fornece-nos um novo ficheiro (.gpr) que contém toda a informação adquirida. A apresentação dos resultados pode ser feita por meio de gráficos de barras, tabelas ou ainda matrizes. Quanto menor o número de experiências realizadas por slide, menor é a complexidade dos resultados, facilitando a apresentação dos mesmos. Nalguns casos, o ficheiro .gpr pode ser transposto para programas como o GraphPad Prism, a partir do qual apresentamos os resultados por meio de gráficos. Noutros casos em que o número de experiências é elevado ou que estes estão no seguimento de outros obtidos anteriormente, podem tornar esta tarefa mais morosa. Para isso, programas associados a bases de dados de *microarrays* foram desenvolvidos no grupo da Prof Ten Feizi (Imperial College, Londres), onde é possível realizar a análise dos resultados de uma forma semi-automatizada. Além de facilitarem a representação dos resultados, permitem que todas as experiências, com o respetivo método utilizado, sejam guardadas para posterior consulta.

#### 4.2.4.2. Ensaios de ligação e competição em gel de poliacrilamida em condições nativas

A utilização de géis de poliacrilamida em condições nativas permite de uma forma geral, a verificação da estabilidade de cada um dos módulos coesina e doquerina, mas também permite a realização de ensaios de ligação e de competição entre diferentes módulos, validando os ensaios em *microarray*. A Figura 4.5 representa o esquema de um ensaio de ligação e de competição.



**Figura 4.5** - Esquema do *template* de ensaio de ligação e de competição. Doc: doquerina; Coh1: coesina 1; Coh2: coesina 2; BSA: albumina sérica bovina.

Em todos os ensaios é incluída uma proteína como controlo positivo de corrida com um padrão de corrida já conhecido, a albumina sérica bovina (BSA). Os módulos coesina e doquerina são misturados em proporções equimolares e depois analisados no gel.

Cada gel correspondente a um ensaio de ligação tem como controlos não só a BSA, mas também os módulos coesina (Coh1 e Coh2) e doquerina (Doc), corridos cada um separadamente em poços diferentes. Os seus perfis de migração foram utilizados para comparar e analisar os resultados obtidos na corrida no gel da mistura composta por uma coesina e uma doquerina (Coh1-Doc e Coh2-Doc).

Para os ensaios de competição e destabilização do complexo foram utilizados 6 controlos positivos corridos separadamente no gel: a BSA, os dois módulos coesina (Coh1 e Coh2) que vão competir pela ligação ao módulo doquerina, o módulo doquerina (Doc) e os dois complexos coesina-doquerina (Coh1-Doc e Coh2-Doc) resultantes que serão testados. No ensaio de destabilização do complexo, cada módulo coesina (Coh1 e Coh2) é incubado com o módulo doquerina. Depois de 1 hora de incubação é adicionada um segundo módulo coesina (Coh1 ou Coh2). A revelação do gel permite verificarmos qual o complexo destabilizado e/ou qual a ligação coesina-doquerina que é mais estável.

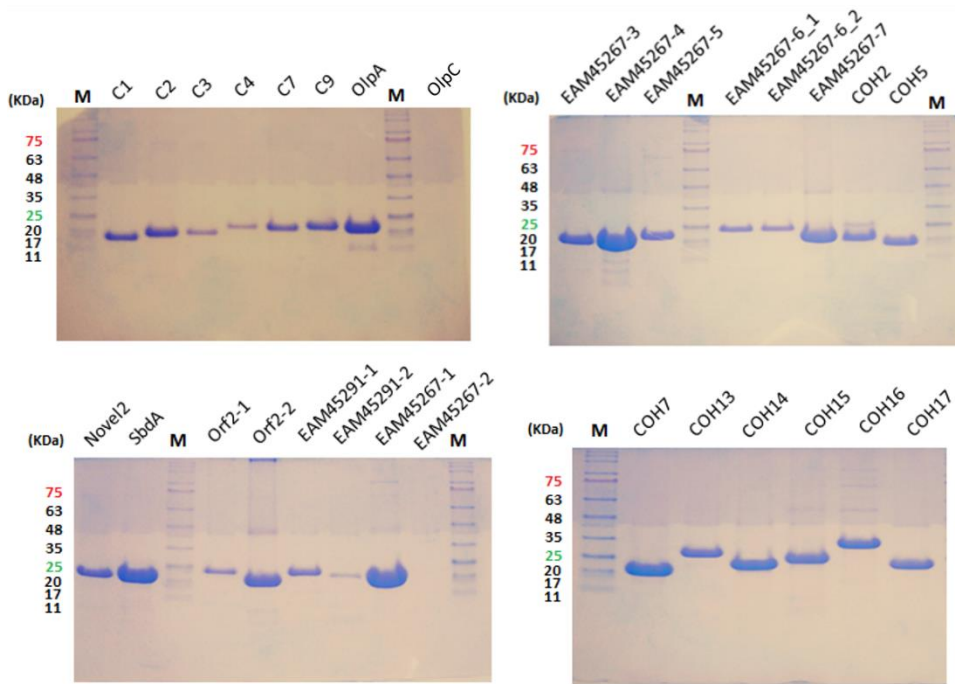
### **4.3. Apresentação e discussão dos resultados**

#### **4.3.1. Expressão e purificação dos módulos coesina e doquerina de tipo I e II**

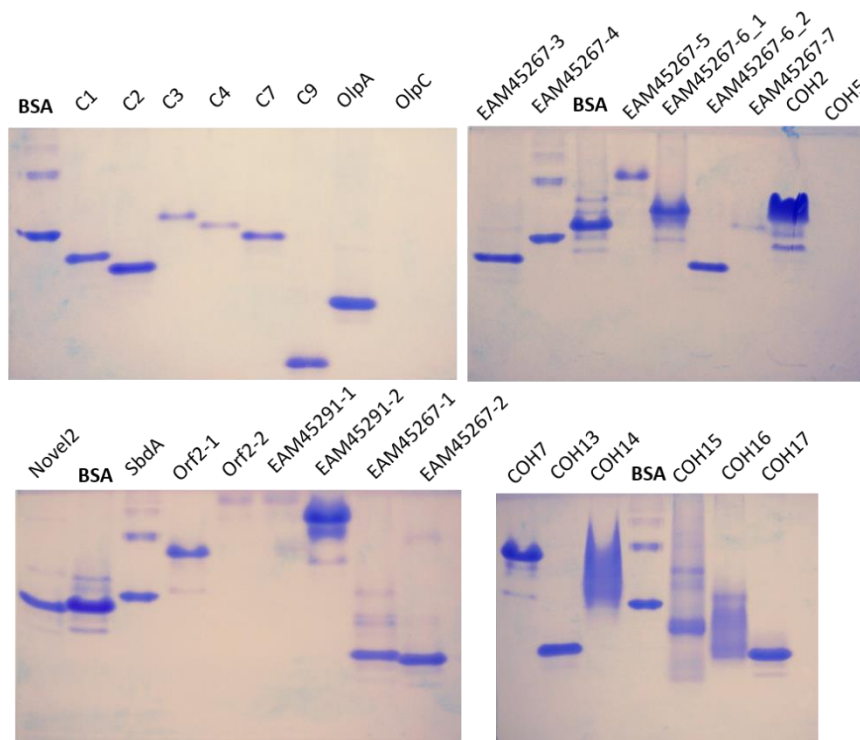
Para estudos de especificidade e de afinidade entre módulos coesina e doquerina do tipo I e II foi necessária a expressão e purificação de 30 módulos coesina e de 42 dos 46 módulos doquerina listados em anexo.

##### **4.3.1.1. Expressão e purificação dos módulos coesina de tipo I e II**

Os 30 módulos coesina foram expressos em meio LB, através da indução por IPTG, seguindo-se o procedimento experimental apresentado na secção 4.2.2. É de referir que os módulos coesinas C1, C3, C4 e OlpC, foram expressas em meio LB com triptona. Todos os restantes foram expressos em meio LB com peptona. Após expressos os 30 módulos coesina, procedeu-se à purificação por IMAC, seguida de um processo de dessalinização. A IMAC foi efetuada em colunas His GraviTrap™ de 1 mL, da GE-Healthcare, tendo as frações sido dessalinizadas. O respetivo protocolo experimental de purificação encontra-se na secção 4.2.3. A análise eletroforética de todos os módulos coesina expressos foram realizados em condições desnaturantes e em condições nativas (Figuras 4.6 e 4.7).



**Figura 4.6** - Resultados da análise eletroforética em condições desnaturantes, após dessalinização, dos 30 módulos coesina expressos e purificados. M corresponde ao marcador de pesos moleculares da NZYTech.



**Figura 4.7** - Resultados da análise eletroforética, em condições nativas, após protocolo de dessalinização dos 30 módulos coesina expressos. A BSA é apresentada como referência de corrida.

Pela observação das Figuras 4.6. e 4.7. verificamos o sucesso da expressão e purificação dos 30 módulos coesina em estudo. Algumas bandas não estão visíveis por possuírem uma baixa concentração. Todas as restantes bandas encontram-se entre os 17 e os 26 kDa, o que coincide com os pesos moleculares previstos (Anexo 7B.I). As concentrações de todos os módulos coesina em estudo foram determinadas através da lei de Lambert-Beer, utilizando o Nanodrop SpectraMax 190 da Molecular Devices, acoplado ao *software* SoftMax® Pro Softwares.

#### 4.3.1.2. Expressão e purificação das doquerinas de tipo I e tipo II

Tentou-se otimizar o protocolo de expressão e purificação utilizado para os módulos coesina, a fim de obtermos a maior quantidade de doquerina no menor volume de crescimento possível. Para isso, escolhemos 9 módulos doquerina e testamos diferentes volumes de expressão (5, 50 e 250mL) sempre respeitando a proporção pré-inóculo:inóculo de 1:100; dois tipos de lise (lise lenta e lise rápida); duas colunas de IMAC (His SpinTrap™ e His GraviTrap™, ambas da GE-Healthcare); diferentes volumes de eluição (200µL, 1, 2,5 e 5mL) e duas colunas de dessalinização distintas (GE-25 e PD-10, ambas da GE-Healthcare). A Figura 4.8. representa o esquema de todas as condições testadas. É de referir que todos os resultados foram analisados eletroforéticamente, em condições desnaturantes, tendo sido escolhidos três possíveis protocolos:

- (1) Expressão num volume de 50mL de meio LB, seguido de purificação em coluna His GraviTrap™ de 1mL, com volume de eluição de 2,5mL e dessalinização em colunas PD-10;
- (2) Expressão num volume de 50mL de meio LB, seguido de purificação em coluna His GraviTrap™ de 600µL, com volume de eluição de 1mL e dessalinização em colunas PD-10;
- (3) Expressão num volume de 250mL de meio LB, seguido de purificação em coluna His GraviTrap™ de 1mL, com volume de eluição de 2,5mL e dessalinização em colunas PD-10;

Após observação da análise eletroforética que permitiu a escolha destes três protocolos, procedeu-se ao cálculo da concentração de um módulo doquerina expresso e purificado segundo estes três protocolos acima mencionados, de forma a escolhermos o que seria empregue para os restantes 41 módulos doquerina. A Tabela 4.1, que apresenta estes resultados.

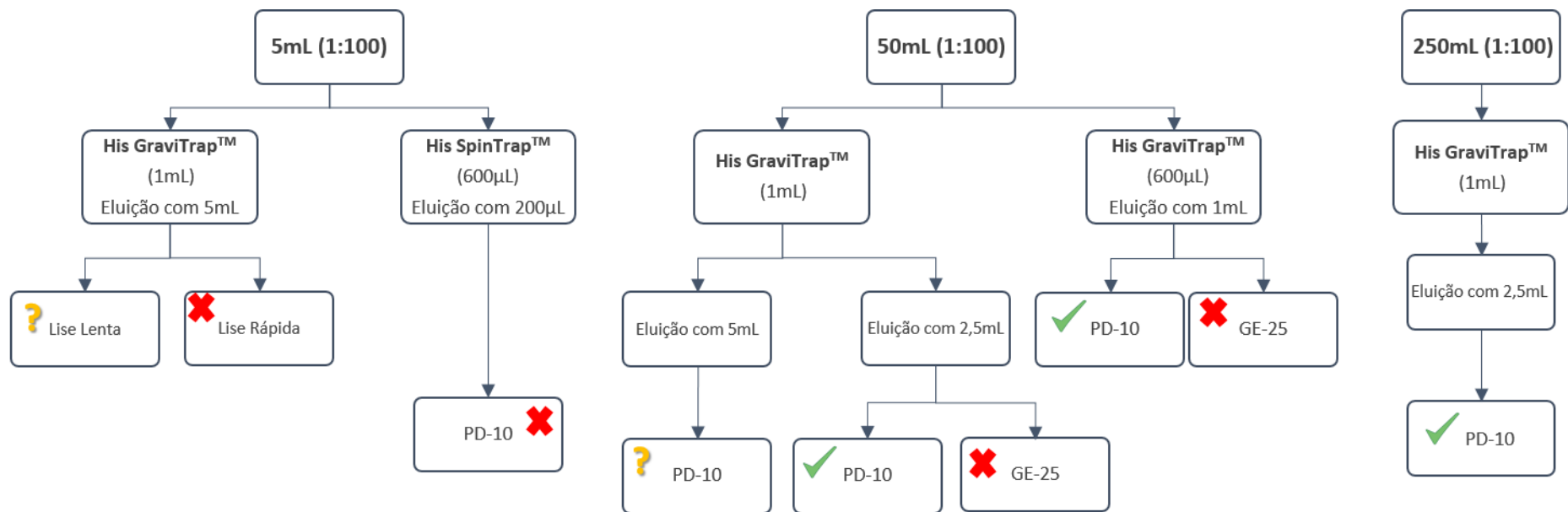
**Tabela 4.1** -Concentrações calculadas após realização dos protocolos (1), (2) e (3).

Doquerina	Concentração (µM)		
	Protocolo (1)	Protocolo (2)	Protocolo (3)
CelA	4,51	7,26	14,54

O módulo CelA proveniente do protocolo (1) foi ainda testado nos *microarrays*, com o objetivo de determinarmos se a quantidade de proteína obtida nesse volume (50mL) era suficiente, e se com o processo de armazenamento ocorreria uma alteração na sua capacidade de interação com as coesinas imobilizadas nos *microarrays*. Como decorria ainda a implementação dos *microarrays* de coesina no

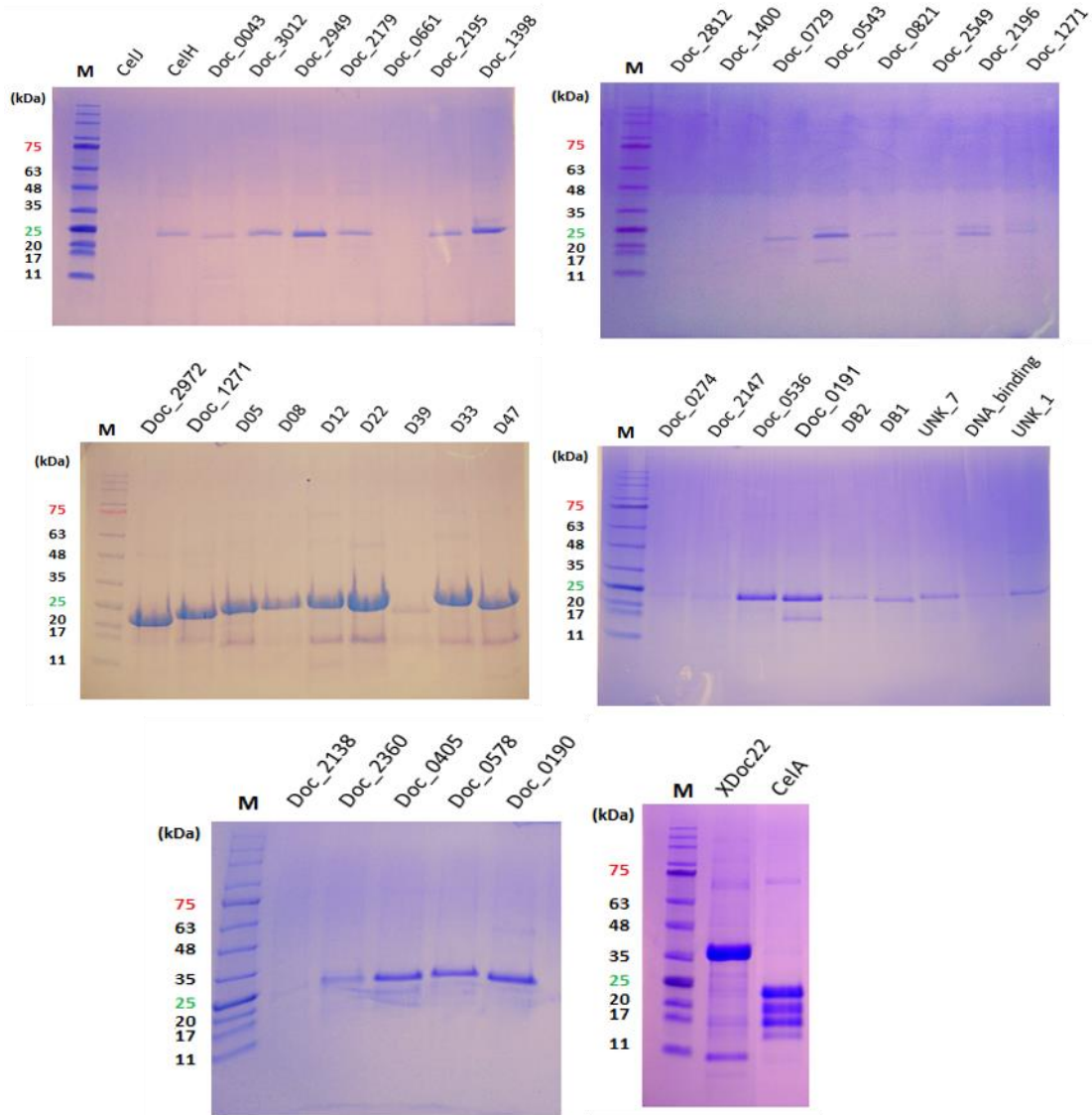
nosso laboratório, decidiu-se inicialmente testar a estabilidade das doquerinas expressas e purificadas aquando do seu armazenamento às temperaturas de -80 e 4°C.

As condições de armazenamento testadas, a -80 e 4°C, foram: em tampão; em tampão com 1mg/mL BSA e em tampão com 5% glicerol. Os resultados deste ensaio permitiram verificarmos que não existem alterações da estabilidade destes módulos nestas diferentes condições. Decidiu-se assim que estas seriam apenas armazenadas em tampão a -80°C, sabendo que o seu armazenamento a 4°C também é possível por pequenos períodos de tempo.



**Figura 4.8** - Esquema da otimização dos protocolos de expressão, purificação e de dessalinização das 46 doquerinas em estudo. A figura descreve para cada volume de expressão utilizado, os vários protocolos testados, isto é, que colunas de purificação utilizadas (His GraviTrap™ e His SpinTrap™) e os respetivos volumes de coluna (600µL e 1mL); os volumes de eluição (200µL, 1, 2,5 e 5mL); e as colunas de dessalinização utilizadas (PD-10 e GE-25). Os resultados dos vários protocolos estão assinalados com um certo (a verde) aquando da obtenção de um resultado favorável; com um ponto de interrogação (a laranja) na presença de resultados inconclusivos; e com uma cruz (a vermelho) aquando da obtenção de um resultado não favorável à otimização do protocolo.

Em suma, e tendo em conta todos os resultados anteriormente referidos, escolhemos expressar e purificar todos os módulos doquerina seguindo o protocolo (3): Expressão num volume de 250mL de meio LB, seguido de purificação em coluna His GraviTrap™ de 1mL, com volume de eluição de 2,5mL e dessalinização em colunas PD-10. A análise eletroforética em condições desnaturantes é apresentada na Figura 4.9.



**Figura 4.9** - Resultados da análise eletroforética, em condições desnaturantes, após protocolo de dessalinização das 42 doquerinas expressas. M corresponde ao marcador de pesos moleculares da NZYTech

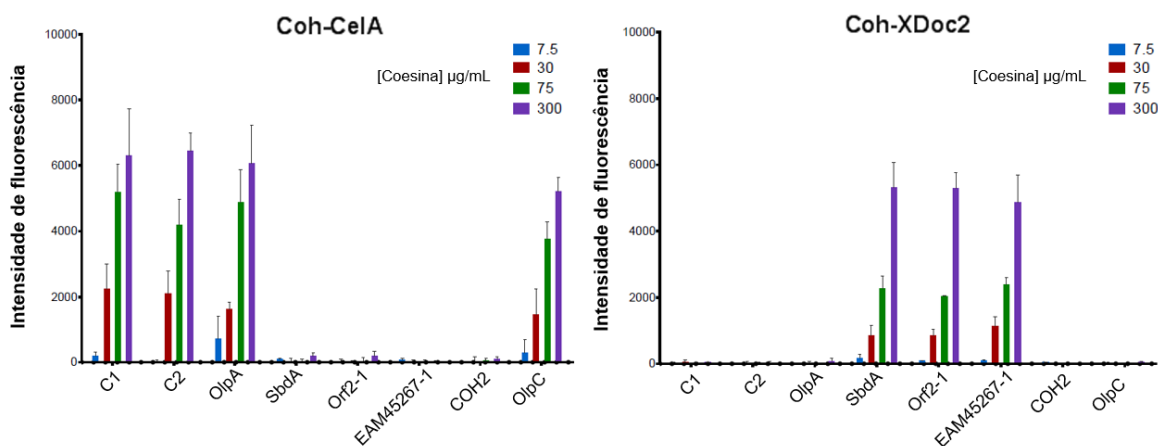
#### 4.3.2. Ensaio de ligação e de competição entre os módulos coesina e doquerina

O estudo das interações entre módulos coesina e doquerina (Coh-Doc) do tipo I e II foi realizado com o intuito de se determinar as interações entre os diferentes módulos Coh-Doc, na tentativa de se decifrar a organização/montagem do celulosoma do organismo *C. thermocellum*. Para isso, realizou-se estudos de afinidade e especificidade utilizando 24

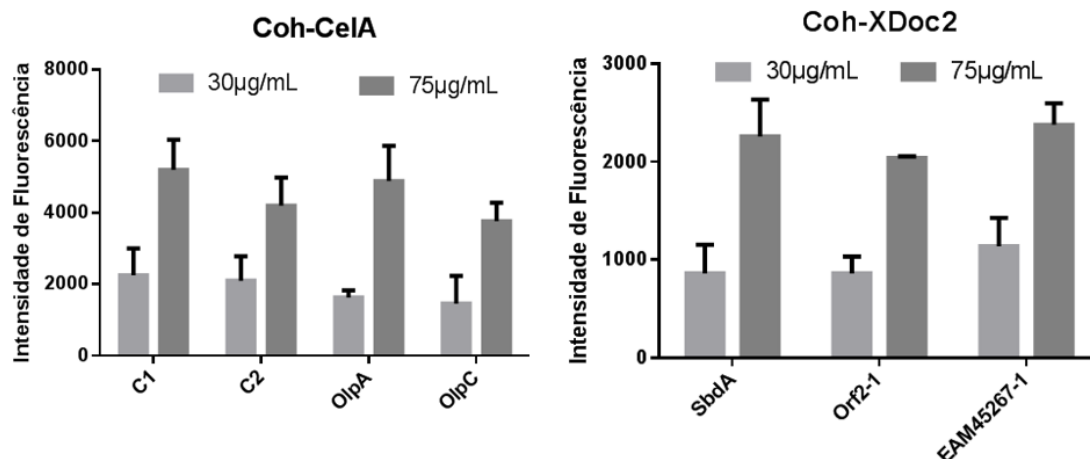
módulos coesina (17 do organismo *C. thermocellum* e 7 do organismo *R. flavefaciens*) e 46 módulos doquerina distintas (39 do organismo *C. thermocellum* e 7 do organismo *R. flavefaciens*). Tal como descrito na secção 4.2.4.1, os módulos coesina foram imobilizados, quer manualmente, quer robóticamente, em quatro concentrações distintas, 7,5, 30, 75 e 300 µg/mL. É ainda de referir que foram realizados triplicados nos slides de 1 bloco e duplicados nos slides de 2 e 16 blocos, para cada uma destas concentrações. Os módulos doquerina foram testados em duas concentrações distintas (0,01 ou 0,1 µM).

#### Ensaio de ligação em *microarrays* e validação

Os primeiros ensaios foram realizados utilizando slides de 1 bloco, recorrendo à imobilização manual, onde testamos a ligação dos módulos CelA e XDoc2, doquerinas do tipo I e II, respetivamente, a oito módulos coesina, quatro do tipo I (C1, C2, OlpA, COH 2 e OlpC) e três do tipo II (SdbA, Orf2-1, EAM45267-1). Neste estudo, a concentração final de doquerina testada foi de 0,1 µM. As Figuras 4.10 e 4.11 descrevem os resultados da análise quantitativa, sob a forma de gráficos de barras.

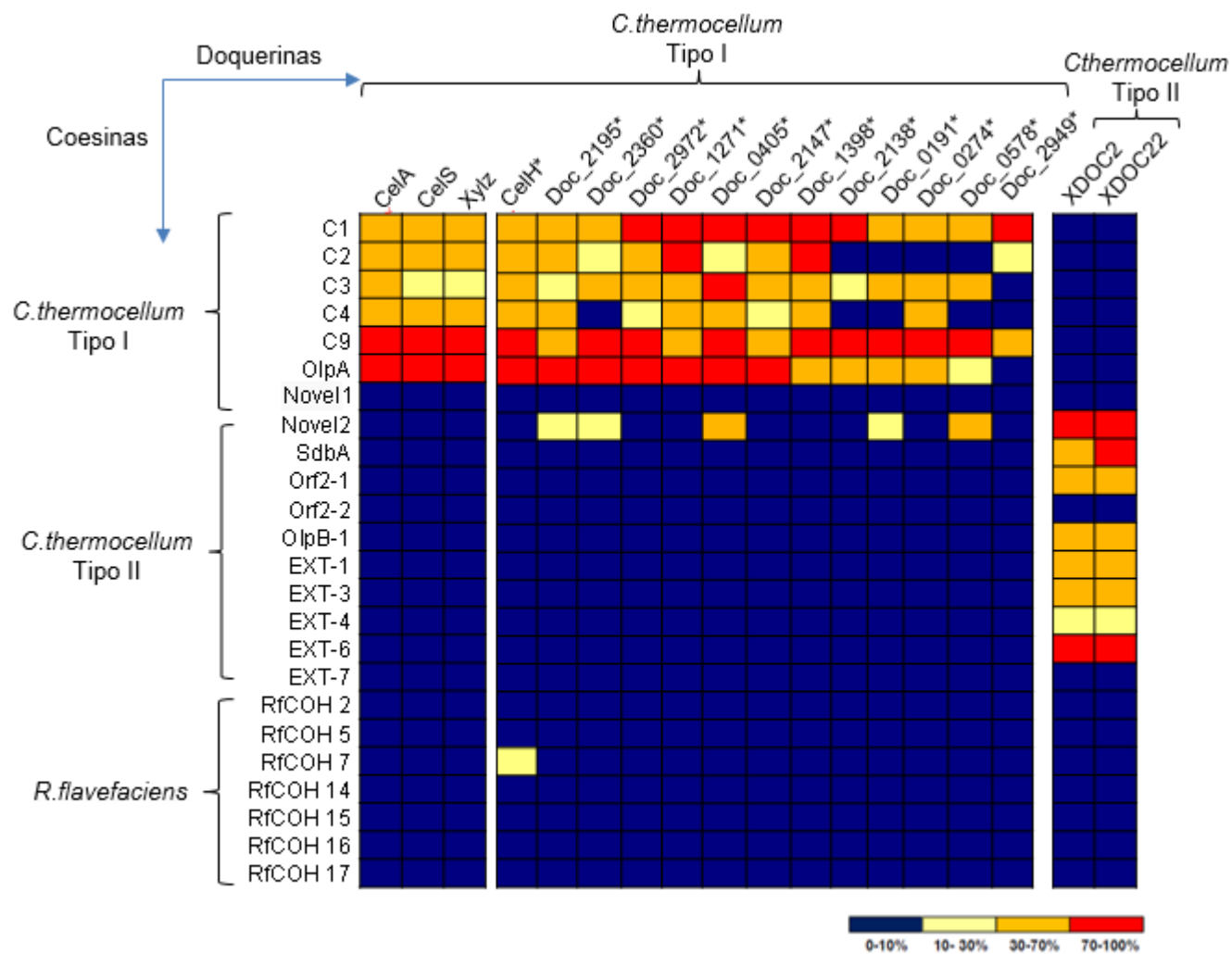


**Figura 4.10** - Gráficos obtidos para a quantificação dos ensaios de ligação pela técnica de *microarray* para as quatro concentrações dos módulos coesina (C1, C2, OlpA, OlpC, SdbA, Orf2-1, EAM45267-1 e COH2) e para a concentração de 0,1 µM de doquerina (CelA e XDoc2).



**Figura 4.11** - Gráficos obtidos para a quantificação dos ensaios de ligação pela técnica de *microarray* para as concentrações de 30 e 75 µg/mL dos módulos coesina (C1, C2, OlpA, OlpC, SbdA, Orf2-1 e EAM45267-1) e para a concentração de 0,1 µM de doquerina (CelA e XDoc2).

Face aos resultados gerais obtidos neste ensaio e visualizando a Figura 4.10., conseguimos verificar a especificidade entre os módulos coesina e doquerina do mesmo tipo. Isto é, a doquerina do tipo I (CelA) apenas interagiu com as coesinas do tipo I. O mesmo aconteceu para a doquerina do tipo II (Xdoc2), que apenas interagiu com as coesinas do tipo II. Verificamos ainda que esta especificidade apenas ocorre entre módulos Coh-Doc do mesmo organismo, isto porque o módulo coesina COH2 (*R. flavefaciens*) não interagiu com nenhum dos módulos doquerina testados. A Figura 4.11. apenas possui os resultados obtidos para as concentrações de 30 e 75 µg/mL para uma mais fácil interpretação dos mesmos. Estes novos gráficos permitem-nos inferir que as interações Coh-Doc, quer do tipo I, quer do tipo II, possuem forças de interação equivalentes com pequenas diferenças. Os restantes ensaios de ligação foram realizados utilizando slides de 16 bloco, nos quais testamos os 46 módulos doquerinas com 24 dos 30 módulos coesina selecionados inicialmente cuja imobilização foi realizada roboticamente. O resumo da análise dos dados encontra-se apresentado na Figura 4.12. sob a forma de uma de matriz, na qual apresentamos os diferentes valores de fluorescência obtidos convertidos em valores de percentagem (0-100%). É ainda de referir que apenas são apresentadas na matriz as doquerinas que obtiveram resultados positivos nos ensaios realizados (18 módulos doquerinas num total de 46).



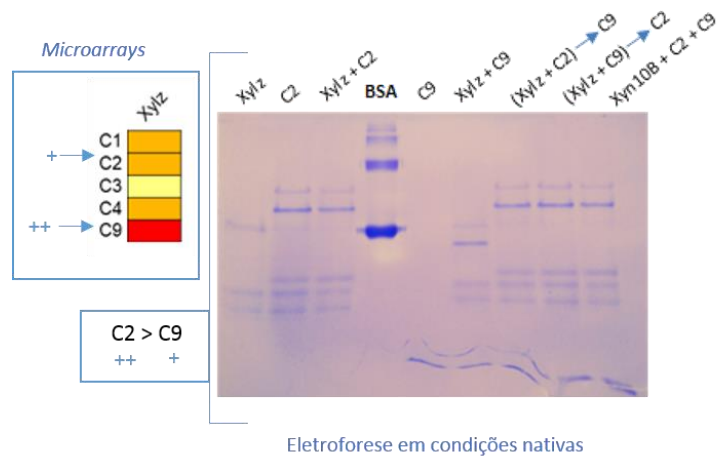
**Figura 4.12**-Resultados da análise quantitativa, sob a forma de matriz, dos ensaios de ligação pela técnica de microarrays para os módulos doxerina e coesina do tipo I e II. \* - Sinais de ligação correspondem às concentrações de coesina de 75µg/mL ou 300 µg/mL.

Esta análise permitiu a verificação da alta especificidade das interações Coh-Doc: que doquerinas de um organismo usualmente só se ligam a coesinas desse mesmo organismo. Contudo mesmo dentro da mesma espécie, doquerinas de um tipo só se ligam a coesinas desse mesmo tipo, isto é, doquerinas do tipo I só se ligam a coesinas do tipo I. No entanto, conseguimos visualizar algumas exceções, sendo o caso das doquerinas Doc\_2195, Doc\_2360, Doc\_0405, Doc\_0191 e Doc\_0578 (tipo I) que interagem com a coesina Novel2 (tipo II); da doquerina CelH (tipo I do *C. thermocellum*) que interage com a coesina COH 7 (*R. flavefaciens*). Verificou-se também perfis de ligação diferentes das doquerinas testadas para as várias coesinas pertencentes a um mesmo tipo. Alguns destes ensaios terão de ser novamente realizados, quer em *microarrays*, quer em géis de poliacrilamida em condições nativas, para posterior validação e esclarecimento dos perfis.

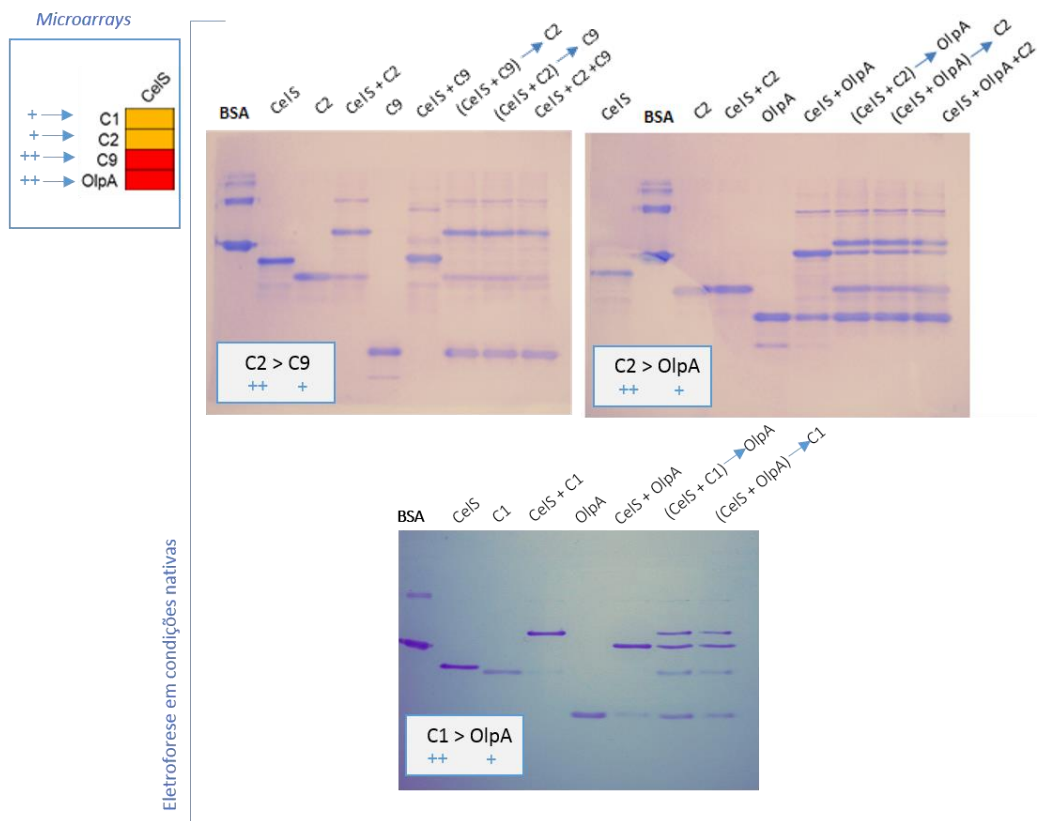
Os géis de poliacrilamida em condições nativas foram utilizados para avaliar e validar alguns dos resultados obtidos e apresentados na matriz, nomeadamente confirmar se os novos constructos (pET28a) dos módulos coesina e doquerina mantêm especificidades e afinidades semelhantes a constructos (pET21a e pET32a) utilizados anteriormente na clonagem destes mesmos módulos. Para isso utilizaram-se módulos coesina e doquerina que tinham sido previamente estudados e cujas especificidades/afinidades eram já conhecidas. Para isso, em cada ensaio foi testada a competição entre dois módulos coesina do tipo I pela ligação a um módulo doquerina do tipo I, de maneira a percebermos qual dos módulos coesina tem maior afinidade por um módulo doquerina específico. Os ensaios de ligação e de competição decorreram consoante o método apresentado na secção 4.2.4.2, testando-se os seguintes pares Coh-Doc:

- (1) Módulo doquerina CelS, com os módulos coesina C1, C2, C3 da proteína estrutural CipA, e OlpA (um módulo coesina presente na parede celular do *C. thermocellum*);
- (2) Módulo doquerina Xylz, com os módulos coesina C2 e C9 da proteína estrutural CipA;
- (3) Módulo doquerina Xyn10B, com os módulos coesina C2, C3, C4 e C9 da proteína estrutural CipA.

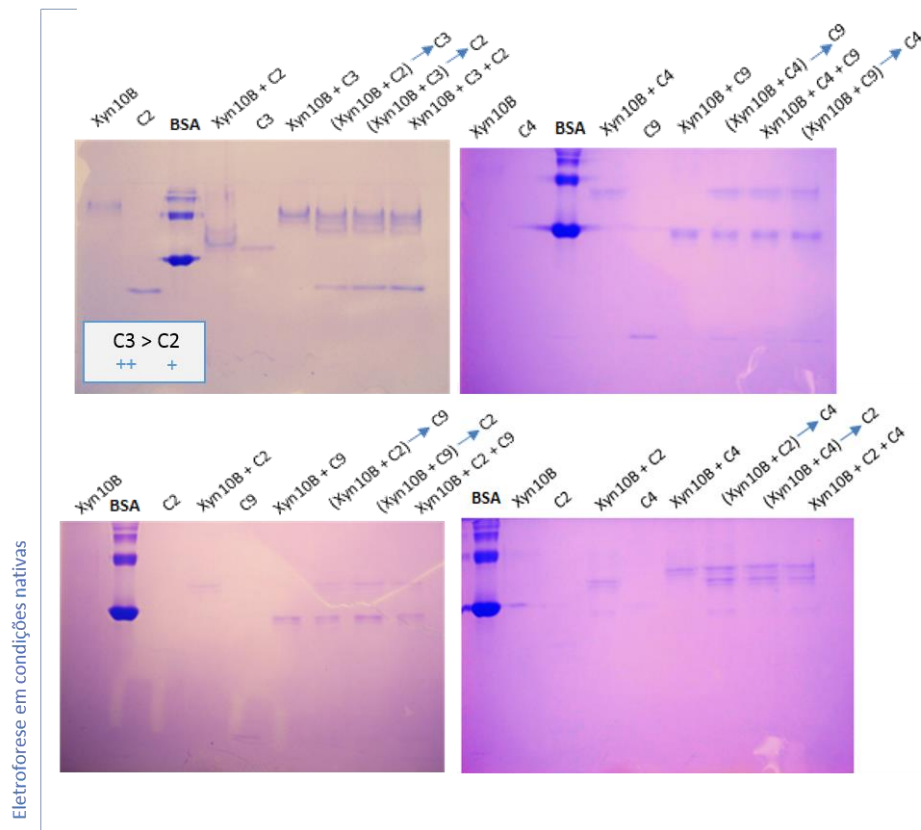
Os resultados destes ensaios encontram-se apresentados nas figuras 4.13 a 4.15.



**Figura 4.13** - Resultados dos ensaios de ligação e de competição realizados por *microarrays* e eletroforese em condições nativas (entre os módulos coesina C2 e C9, e o módulo doquerina Xylz). A BSA é apresentada como referência de corrida.



**Figura 4.14** - Resultados dos ensaios de ligação e de competição realizados por *microarrays* e por eletroforese em condições nativas, para a doquerina CelS (entre os módulos coesinas C2 e C9, C2 e OlpA, e C1 e OlpA). A BSA é apresentada como referência de corrida.



**Figura 4.15** - Resultados dos ensaios de ligação e de competição realizados por eletroforese em condições nativas, para a doquerina Xyn10B (entre os módulos coesina C2 e C3, C2 e C4, C4 e C9, e C2 e C9). A BSA é apresentada como referência de corrida.

A observação dos géis acima apresentados permitiu verificar a existência de diferentes afinidades entre diferentes pares de Coh-Doc da espécie *C. thermocellum*. No caso do estudo (3) – Figura 4.15. – os resultados apenas foram conclusivos para o ensaio onde foram testados simultaneamente os módulos C2 e C3. Nos restantes, foi difícil observar diferenças de afinidade do módulo doquerina Xyn10B para os diferentes módulos coesina C2, C4 e C9. No conjunto das 9 experiências em que se testou combinatoriamente estes 3 módulos coesina, nunca houve a formação prevalente de um complexo. A diferença de resultados pode dever-se a uma diferença nos constructos utilizados, visto não termos obtido qualquer ligação para esta proteína nos *microarrays*, revelando que esta poderá não estar funcional.

Realizou-se uma tabela resumo de modo a concentrar a informação recolhida pela visualização dos resultados dos ensaios realizados por eletroforese em condições nativas.

**Tabela 4.2** - Resumo dos resultados obtidos nos ensaios de ligação e de competição entre pares Coh-Doc realizados por eletroforese em, condições nativas.

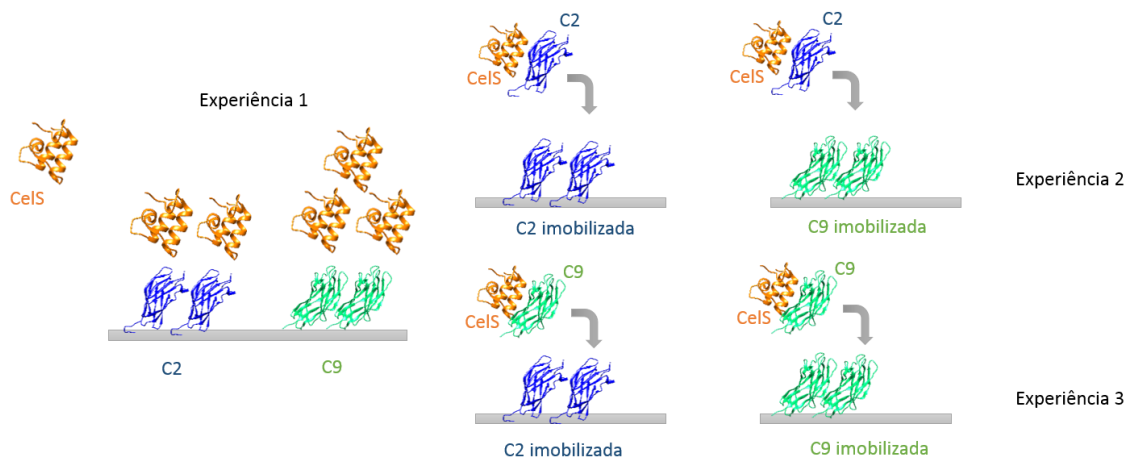
Doquerina	Coesina 1	Coesina 2	Resultado
CelS	C1	OlpA	C1 > OlpA
CelS	C2	C9	C2 > C9
CelS	C2	OlpA	C2 > OlpA
Xylz	C2	C9	C2 > C9
Xyn10B	C2	C3	C3 > C2

Comparando os resultados obtidos utilizando géis nativos e *microarrays*, verificamos que a imobilização parece influenciar a força de ligação dos módulos doquerina, pois encontramos diferenças no comportamento do módulo coesina quando se encontra em solução ou imobilizado. Por exemplo, o módulo CelS parece possuir uma maior força de ligação para o módulo C9, quando este se encontra imobilizado, mas uma maior afinidade para o módulo C2 quando em solução. A mesma observação foi feita para o módulo doquerina XylZ.

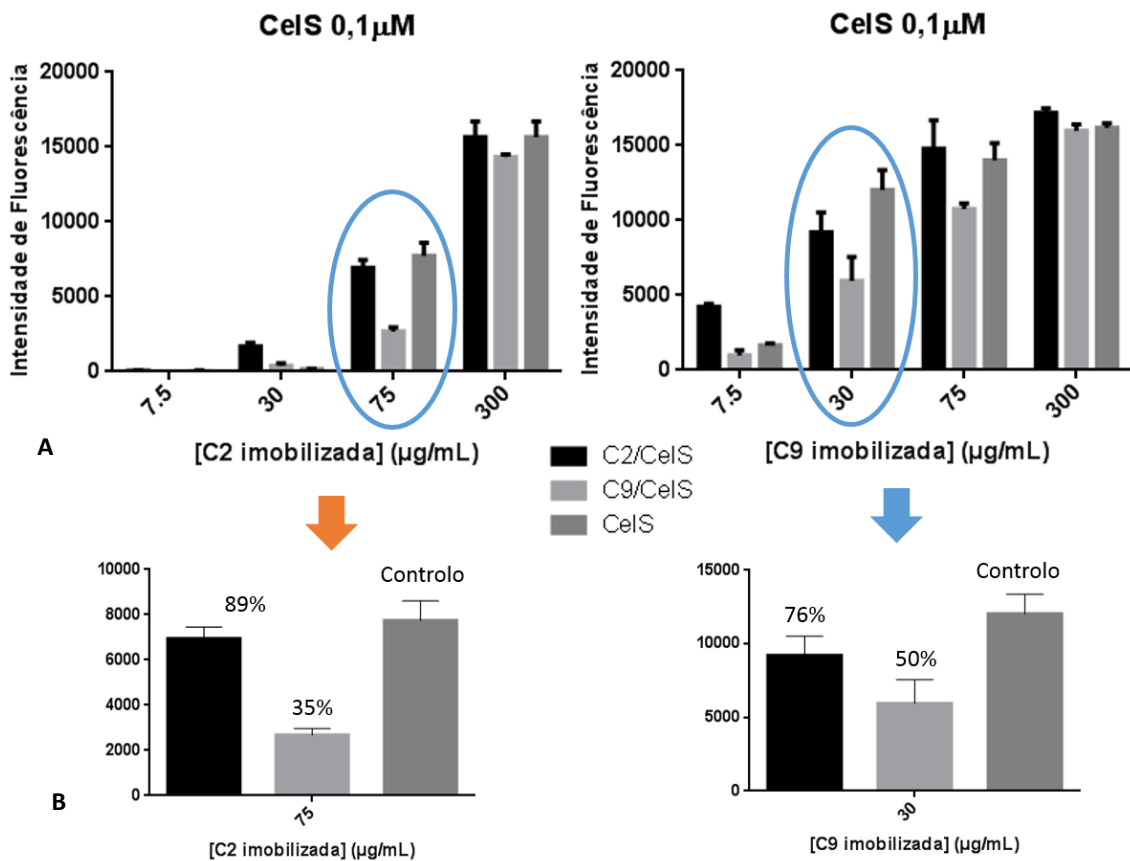
#### Ensaio de competição em *microarrays* e validação

Numa tentativa de tentarmos perceber se esta diferença de afinidade poderá ser influenciada pela imobilização, decidiu-se realizar um ensaio de competição entre os módulos coesina C2 e C9, para a doquerina CelS.

O ensaio de competição entre estes módulos C2 e C9 foi realizado, nos nossos *microarrays*, recorrendo a três slides de 2 blocos. A impressão das coesinas foi realizada manualmente e em triplicados nas concentrações de 7,5, 30, 75 e 300µg/mL. Este ensaio contempla 3 experiências distintas. A primeira experiência serviu de controlo, tendo-se testado a doquerina CelS a uma concentração de 0,1 µM para os dois módulos coesina imobilizados. Nas experiências 2 e 3 testou-se o módulo CelS em complexo com uma das coesinas (C2 ou C9), que fará uma inibição/competição para os módulos coesina imobilizados. Nestes ensaios a coesina foi incubada previamente com a doquerina CelS e numa concentração duas vezes maior (0,2µM), fazendo com a doquerina esteja sempre em concentração limitante. As Figura 4.16 e 4.17 esquematizam as três experiências realizadas, bem como os resultados do ensaio, respetivamente.



**Figura 4.16** - Esquema das três experiências realizadas no ensaio de competição entre os módulos coesina C2 e C9 em complexo com o módulo doquerina CelS.



**Figura 4.17** - Resultados relativos aos ensaios de competição pela técnica de *microarrays* de proteínas para as concentrações de 7,5, 30, 75 e 300 µg/mL de coesina imobilizada: A – Gráficos referentes aos resultados para os módulos C2 e C9 imobilizados; B – Gráficos referentes à menor concentração onde se verificou a capacidade de competição entre os módulos em estudo.

Os resultados acima apresentados levam-nos a concluir que para a concentração mais elevada de coesina imobilizada – 300 µg/mL – não existe sensibilidade suficiente para verificarmos a capacidade de competição dos módulos em estudo. A visualização destes resultados permitem ainda concluir que o fenómeno de competição/inibição existe nas concentrações de 75 µg/mL para a coesina C2 imobilizada e de 30 e 75 µg/mL para a coesina C9 imobilizada, tendo sido obtido uma maior competição/inibição por parte do módulo C9, quando em complexo com o módulo CelS e em solução. Ou seja, o módulo C9 inibe a ligação do módulo doquerina Cels aos módulos coesina imobilizados. Os resultados obtidos nas restantes concentrações não permitem aferir conclusões.

A Figura 4.17-B mostra-nos os resultados da capacidade de competição para a menor concentração de coesina imobilizada. Para o módulo C2 imobilizado, a 75 µg/mL, o módulo C2 em solução inibe a ligação em 11% enquanto que o módulo C9 tem uma inibição maior, de 65%. No caso da imobilização do módulo C9, a 30 µg/mL, verificamos que existe uma inibição de ligação de 24% por parte do módulo C2, em complexo com o módulo doquerina Cels e em solução, e uma inibição da ligação de 50% do módulo C9, também em complexo com a CelS e também em solução. Estes resultados revelam uma maior capacidade de competição/inibição por parte da coesina C9 em relação à coesina C2 em solução. Verificou-se também que a diferença de inibição entre as duas coesinas em solução foi maior para a C2 imobilizada (6×) do que para a C9 (2×). Para a melhor compreensão destes resultados, mais ensaio terão de ser realizados.

#### **4.4. Conclusão**

A utilização da técnica de *microarrays* de proteínas permite obter informações acerca de forças de ligação, afinidades e especificidades das interações proteína-proteína. É neste contexto que a segunda parte deste trabalho procurou estudar estes parâmetros entre os inúmeros módulos coesina e doquerina existentes nos celulosomas dos organismos *C. thermocellum* e *R. flavefaciens* recorrendo aos *microarrays* de proteína desenvolvidos no nosso laboratório.

O primeiro passo consistiu na expressão e purificação de todos os módulos a serem testados. Prosseguiu-se com ensaios de ligação e de competição entre os mesmos, conseguindo determinar diferentes especificidades e percentagens de afinidade. Verificou-se que a interação coesina-doquerina é específica para cada espécie, possuindo diferentes forças de ligação e afinidades para os diferentes pares coesina-doquerina. A validação realizada para alguns pares Coh-Doc revelou que a imobilização dos módulos coesina poderá influenciar a sua função, tendo sido observadas diferenças na afinidades e força de ligação para o módulo quando em solução ou quando imobilizado.

Capítulo

5

## **Conclusão geral e perspectivas futuras**



## 5. Conclusão geral e perspectivas futuras

A celulose, juntamente com outros polissacáridos da parede celular, são os compostos orgânicos mais abundantes na biosfera. Constituem além de uma fonte de energia renovável, um material que pode ser empregue em inúmeras aplicações biotecnológicas. Por esta razão, a sua hidrólise em açúcares mais simples pela ação de diferentes enzimas provenientes de diferentes organismos tem vindo a ser alvo de grande interesse. O conhecimento destes sistemas biológicos permitirá um melhor aproveitamento destas matérias-primas. Neste trabalho pretendeu-se estudar o celossoma, em especial o do microrganismo anaeróbio *C. thermocellum* através da caracterização e avaliação das interações coesina-doquerina, cruciais para a organização/montagem deste complexo multienzimático. Para isso utilizaram-se duas técnicas distintas: a cristalografia de raios-X e os *microarrays* de proteína.

No capítulo 3 foram apresentados os resultados relativos à caracterização das estruturas do complexo C4-Xyn10B e do módulo C9, na qual comparamos a interação dos módulos coesina à mesma interface do módulo doquerina Xyn10B. A comparação destes módulos e da interação C4-Xyn10B com o modelo C2-Xyn10B permitiu verificar que a localização destes módulos na CipA poderá não influenciar diretamente a interação com o módulo Xyn10B, não conseguindo determinar a razão estrutural para as diferenças de afinidade encontradas aquando da realização dos estudos prévios usando os géis nativos. Com isto, e de modo a perceber se estas semelhanças na interação Coh-Doc se mantêm, pretende-se futuramente continuar os estudos estruturais de outros módulos coesina da CipA, quer complexados a este módulo Xyn10B, quer a outros módulos doquerina.

No capítulo 4, apresentamos os resultados relativos ao estudo de diversas interações Coh-Doc, a nível da especificidade, afinidade e força de ligação. Pretendeu-se assim complementar o estudo realizado no capítulo 3 utilizando uma técnica que permite avaliar um elevado número de proteínas num único ensaio. Nestes resultados confirmamos que a maioria das interações Coh-Doc é realizada entre módulos pertencentes ao mesmo tipo e à mesma espécie, salvo algumas exceções que precisam ainda de confirmação. Por outro lado, não se conseguiu ver interação aos módulos coesina nos *microarrays* por parte do módulo doquerina Xyn10B. Esta incongruência precisa de ser verificada de modo a completar o estudo dos complexos C4-Xyn10B e do módulo C9 por Cristalografia de raios-X.

Deste modo pretende-se validar os resultados obtidos nos *microarrays*, confirmando e verificando-os por eletroforese em gel de poliacrilamida em condições nativas e também através da realização dos ensaios de competição/inibição, ainda em processo de otimização. Para algumas doquerinas do *C. thermocellum* e todas do *R. flavefaciens* testadas até agora, as condições do ensaio terão de ser optimizadas, como por exemplo, através do aumento da concentração da doquerina

utilizada. Pretende-se no futuro utilizar esta técnica robusta para o estudo de interações coesina-  
doquerina em outros organismos celulolíticos.

Capítulo



## **Bibliografía**



## 6. Bibliografia

- [1] D. J. Cosgrove, "Growth of the plant cell wall," *Nat. Rev. Mol. Cell Biol.*, vol. 6, pp. 850–861, 2005.
- [2] Rose JC; Bennett AB, "Cooperative disassembly of the cellulose–xyloglucan network of plant cell walls: parallels between cell expansion and fruit ripening," *Trends Plant Sci.*, vol. 4, pp. 176–183, 1999.
- [3] Z. A. Popper, "Evolution and diversity of green plant cell walls.," *Curr. Opin. Plant Biol.*, vol. 11, pp. 286–292, 2008.
- [4] S. Ohmae, M.; Kobayashi, "Enzyme-catalyzed synthesis of natural and unnatural polysaccharides," *J. Polym. Sci. Part A Polym. Chem.*, vol. 44, pp. 5014–5027, 2006.
- [5] A. Klemm, D.; Heublein, B.; Fink, HP.; Bohn, "Cellulose: Fascinating Biopolymer and Sustainable Raw Material," *Angew. Chemie Int. Ed.*, vol. 44, pp. 3358–3393, 2005.
- [6] P. Béguin and M. Lemaire, "The cellulosome: an exocellular, multiprotein complex specialized in cellulose degradation.," *Crit. Rev. Biochem. Mol. Biol.*, vol. 31, no. 3, pp. 201–236, 1996.
- [7] L. J. Gibson, "The hierarchical structure and mechanics of plant materials," *J. R. Soc. Interface*, vol. 9, no. 76, pp. 2749–2766, 2012.
- [8] J. Béguin, P.; Aubert, "The biological degradation of cellulose," *FEMS Microbiol Rev.*, vol. 13, pp. 25–28, 1994.
- [9] R. Heredia, A.; Jiménez, A.; Guillén, "Composition of plant cell walls.," *Z Leb. Unters Forsch.*, vol. 200, pp. 24–31, 1995.
- [10] H. J. Gilbert and G. P. Hazlewood, "Genetic modification of fibre digestion.," *Proc. Nutr. Soc.*, vol. 50, no. 2, pp. 173–186, 1991.
- [11] L. R. et al. Lynd, "Microbial Cellulose Utilization: Fundamentals and Biotechnology," *Microbiol Mol Biol Rev.*, vol. 66, pp. 505–577, 2002.
- [12] R. P. de Vries, J. Visser, P. Ronald, de Vries, R., and P., "Aspergillus Enzymes Involved in Degradation of Plant Cell Wall Polysaccharides," *Microbiol. Mol. Biol. Rev.*, vol. 65, no. 4, pp. 497–522, 2001.
- [13] J. G. Ng. T.K.; Weimer. P.J., Zeikus, "Cellulolytic and physiological properties of *Clostridium thermocellum*," *Arch. Microbiol.*, vol. 114, pp. 1–7, 1977.
- [14] M. et al. Jindou, S.; Brulc, JM.; Levy-Assaraf, "Cellulosome gene cluster analysis for gauging the diversity of the ruminal cellulolytic bacterium *Ruminococcus flavefaciens*," *FEMS Microbiol Lett.*, vol. 285, pp. 188–194, 2008.
- [15] H. W. Bergquist, J.A.; Gibbs, M.D.; Morris, D.D.; Te'o, V.S.J.; Saul, D.J. e Morgan, "Molecular diversity of thermophilic cellulolytic and hemicellulolytic bacteria.," *FEMS Microbiol. Ecol.*, vol. 28, pp. 99–110.

- [16] K. N. Wood, T.M.; Wilson, C.A., McCrae, S.I. e Joblin, “A highly active extracellular cellulase from the anaerobic rumen fungus *Neocallimastix frontalis*,” *FEMS Microbiol. Lett.*, vol. 34, pp. 37–40, 1986.
- [17] G. P. Eberhardt, R.Y.; Gilbert, H.J e Hazlewood, “Primary sequence and enzymatic properties of two modular endoglucanases, Cel5A and Cel45A, from the anaerobic fungus *Piromyces equi*,” *Microbiology*, vol. 146, pp. 1999–2008, 2000.
- [18] C. M. G. a Fontes and H. J. Gilbert, “Cellulosomes: highly efficient nanomachines designed to deconstruct plant cell wall complex carbohydrates,” *Annu. Rev. Biochem.*, vol. 79, pp. 655–681, 2010.
- [19] and C. F. Pinheiro, B. A., J. L. A. Bras, S. Najmudin, AL Carvalho, LMA Ferreira, JAM Prates, “Flexibility and specificity of the cohesin-dockerin interaction: implications for cellulosome assembly and functionality,” *Biocatal. Biotransformation*, vol. 30, pp. 309–315, 2012.
- [20] H. Akinosho, K. Yee, D. Close, and A. Ragauskas, “The emergence of *Clostridium thermocellum* as a high utility candidate for consolidated bioprocessing applications,” *Front. Chem.*, vol. 2, no. August, pp. 1–18, 2014.
- [21] W. J. Demain AL, Newcomb M, “Cellulase, clostridia, and ethanol,” *Microbiol. Mol. Biol. Rev.*, vol. 69, pp. 124–154, 2005.
- [22] I. A. Uversky, V.N.; Kataeva, “Cellulosome,” *Hauppauge, NY Nov. Sci. Publ.*, 2006.
- [23] E. a Bayer and R. Lamed, “Ultrastructure of the Cell-Surface Cellulosome of *Clostridium-Thermocellum* and Its Interaction with Cellulose,” *J. Bacteriol.*, vol. 167, no. 3, pp. 828–836, 1986.
- [24] Y. Shoham, R. Lamed, and E. a. Bayer, “The cellulosome concept as an efficient microbial strategy for the degradation of insoluble polysaccharides,” *Trends Microbiol.*, vol. 7, no. 7, pp. 275–281, 1999.
- [25] E. a. Bayer, E. Setter, and R. Lamed, “Organization and distribution of the cellulosome in *Clostridium thermocellum*,” *J. Bacteriol.*, vol. 163, no. 2, pp. 552–559, 1985.
- [26] E. a Bayer, L. J. Shimon, Y. Shoham, and R. Lamed, “Cellulosomes-structure and ultrastructure,” *J. Struct. Biol.*, vol. 124, no. 2–3, pp. 221–234, 1998.
- [27] L. J. Shimon, E. a Bayer, E. Morag, R. Lamed, S. Yaron, Y. Shoham, and F. Frolow, “A cohesin domain from *Clostridium thermocellum*: the crystal structure provides new insights into cellulosome assembly,” *Structure*, vol. 5, no. 3, pp. 381–390, 1997.
- [28] S. Salamiou, O. Raynaud, M. Lemaire, M. Coughlan, P. Beguin, and J. P. Aubert, “Recognition specificity of the duplicated segments present in *Clostridium thermocellum* endoglucanase CelD and in the cellulosome-integrating protein CipA,” *J. Bacteriol.*, vol. 176, no. 10, pp. 2822–2827, 1994.
- [29] S. Chauvaux, P. Beguin, J. P. Aubert, K. M. Bhat, L. a Gow, T. M. Wood, and a Bairoch, “Calcium-binding affinity and calcium-enhanced activity of *Clostridium thermocellum* endoglucanase D,” *Biochem. J.*, vol. 265, no. 1, pp. 261–265, 1990.

- [30] yer EA; Henrissat B.; Lamed R, “The cellulosome: A natural bacterial strategy to combat biomass recalcitrance,” *Biomass Recalcitrance*, pp. 407–426, 2008.
- [31] A. L. Carvalho, F. M. V Dias, J. a M. Prates, T. Nagy, H. J. Gilbert, G. J. Davies, L. M. a Ferreira, M. J. Romão, and C. M. G. a Fontes, “Cellulosome assembly revealed by the crystal structure of the cohesin-dockerin complex,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 100, no. 24, pp. 13809–13814, 2003.
- [32] A. L. Carvalho, F. M. V Dias, T. Nagy, J. a M. Prates, M. R. Proctor, N. Smith, E. a Bayer, G. J. Davies, L. M. a Ferreira, M. J. Romão, C. M. G. a Fontes, and H. J. Gilbert, “Evidence for a dual binding mode of dockerin modules to cohesins.,” *Proc. Natl. Acad. Sci. U. S. A.*, vol. 104, no. 9, pp. 3089–3094, 2007.
- [33] W. R. A. J. Boraston A. B., McLean B. W. Kormos J. M., Alam M., Gilkes N. R., Haynes C. A., Tomme P., Kilburn D. G., “Carbohydrate-binding modules: diversity of structure and function,” *Gilbert H. J., Davies G. J., Henrissat B., Svensson B., Ed. Recent Adv. Carbohydr. Bioeng. Cambridge R. Soc. Chem.*, pp. 202–211, 1999.
- [34] J. Tormo, R. Lamed, a J. Chirino, E. Morag, E. a Bayer, Y. Shoham, and T. a Steitz, “Crystal structure of a bacterial family-III cellulose-binding domain: a general mechanism for attachment to cellulose.,” *EMBO J.*, vol. 15, no. 21, pp. 5739–5751, 1996.
- [35] [http://www.nobelprize.org/nobel\\_prizes/chemistry/laureates/1964/speedread.html](http://www.nobelprize.org/nobel_prizes/chemistry/laureates/1964/speedread.html). [Accessed: 01-Aug-2015].
- [36] J. Kendrew, “M.f. perutz,” vol. 1962, pp. 2684–2685, 1997.
- [37] M. J. Romão, “Cristalografia de Proteínas: metodologias e aplicações me Bioquímica,” *Boletim de Biotecnologia*, vol. 53. pp. 18–36, 1996.
- [38] G. Rhodes, *Crystallography Made Crystal Clear - A Guide for Users of Macromolecular Models*, Third edit. Academic Press, 2006.
- [39] G. Taylor, “The phase problem,” *Acta Crystallogr. - Sect. D Biol. Crystallogr.*, vol. 59, no. 11, pp. 1881–1890, 2003.
- [40] L. H. Augenlicht, D. Kobrin, a Pavlovec, and M. E. Royston, “Elevated expression of an endogenous retroviral long terminal repeat in a mouse colon tumor.,” *J. Biol. Chem.*, vol. 259, no. 3, pp. 1842–1847, 1984.
- [41] D. a Hall, J. Ptacek, and M. Snyder, “Protein microarray technology.,” *Mech. Ageing Dev.*, vol. 128, no. 1, pp. 161–7, 2007.
- [42] M. Zhu, H. Bilgin, M., Snyder, “Proteomics,” *Annu. Rev. Biochem.*, vol. 72, pp. 783–812, 2003.
- [43] S. Q. Lesaichere, M.L., Lue, R.Y., Chen, G.Y., Zhu, Q., Yao, “Inteinmediated biotinylation of proteins and its application in a protein microarray,” *J. Am. Chem. Soc.*, vol. 124, pp. 8768–8769, 2002.
- [44] W. Kusnezow, A. Jacob, A. Walijew, F. Diehl, and J. D. Hoheisel, “Antibody microarrays: An evaluation of production parameters,” *Proteomics*, vol. 3, no. 3, pp. 254–264, 2003.

- [45] R. Jara-acevedo, S. Matarraz, and S. Paradinas, “Nanotecnología en proteómica;” pp. 6–11, 2011.
- [46] B. a Pinheiro, H. J. Gilbert, K. Sakka, K. Sakka, V. O. Fernandes, J. a M. Prates, V. D. Alves, D. N. Bolam, L. M. a Ferreira, and C. M. G. a Fontes, “Functional insights into the role of novel type I cohesin and dockerin domains from *Clostridium thermocellum*.” *Biochem. J.*, vol. 424, no. 3, pp. 375–384, 2009.
- [47] T. G. G. Batty, L. Kontogiannis, O. Johnson, H. R. Powell, and A. G. W. Leslie, “iMOSFLM: A new graphical interface for diffraction-image processing with MOSFLM,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. 4, pp. 271–281, 2011.
- [48] W. Kabsch, “Xds,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 2, pp. 125–132, 2010.
- [49] A. J. McCoy, R. W. Grosse-Kunstleve, P. D. Adams, M. D. Winn, L. C. Storoni, and R. J. Read, “Phaser crystallographic software,” *J. Appl. Crystallogr.*, vol. 40, no. 4, pp. 658–674, 2007.
- [50] K. Cowtan, “Recent developments in classical density modification,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 66, no. 4, pp. 470–478, 2010.
- [51] K. Cowtan, “Fitting molecular fragments into electron density,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 64, no. 1, pp. 83–89, 2007.
- [52] G. N. Murshudov, P. Skubák, A. a. Lebedev, N. S. Pannu, R. a. Steiner, R. a. Nicholls, M. D. Winn, F. Long, and A. a. Vagin, “REFMAC5 for the refinement of macromolecular crystal structures,” *Acta Crystallogr. Sect. D Biol. Crystallogr.*, vol. 67, no. 4, pp. 355–367, 2011.
- [53] J. J. S. H. Kim, “Sparse matrix sampling: a screening method for crystallization of proteins,” *J. Appl. Crystallogr.*, vol. 24, pp. 409–411, 1991.

Capítulo



**Anexos**



## 76A – Anexos referentes ao estudo estruturais da proteína CipA por Cristalografia de Raios-X

### 7A.I – Composição dos meios de cultura

**Tabela 6A.1-** Composição dos meios de expressão utilizados nos testes de expressão em E. coli. Os meios foram preparados com água destilada. Foram de seguida autoclavados a 120°C durante 20 minutos.

Composição/Reagente				
	LB	TB	2xYP	Marca
Peptona	-	12 g/L	16 g/L	Fluka
Triptona	10 g/L	-	-	Bacto™ Tryptone BD
Extracto de levedura	5 g/L	24 g/L	10 g/L	Fluka
Cloreto de sódio	10 g/L	-	5 g/L	Scahrlau

### 7A.II - Protocolo de purificação por cromatografia de afinidade por iões metálicos imobilizados

- 1) Lavagem da coluna His-Trap™ com cerca de 50 mL água Milli-Q a um fluxo de 5 mL/min;
- 2) Equilibrar a coluna His-Trap™ com solução tampão 50mM Hepes a pH 7.5, 1M NaCl a 10mM imidazol por cerca de 20 mL continuando com um fluxo de 5 mL/min;
- 3) Injectar extracto proteico alterando o fluxo para 2,5mL/min;
- 4) Lavagem da coluna His-Trap™ com a solução tampão 50mM Hepes a pH 7.5, 1M NaCl a 10mM imidazol a 2,5 mL/min até valores de absorvância estabilizarem;
- 5) Realizar primeiro patamar de lavagem da coluna com 10% da solução tampão 50mM Hepes a pH 7.5, 1M NaCl a 250mM imidazol (2,5mL/min);
- 6) Realizar gradiente de lavagem da coluna até à concentração de 250mM imidazol a 2,5mL/min;
- 7) Lavagem da coluna His-Trap™ com solução tampão 50mM Hepes a pH 7.5, 1M NaCl a 250mM imidazol por cerca de 50 mL continuando com um fluxo de 2,5 mL/min;
- 8) Lavagem da coluna His-Trap™ com água Milli-Q por cerca de 50 mL com um fluxo de 5 mL/min.
- 9) Guardar a coluna a 4°C em etanol 20%.

### 7A.III -Protocolo de purificação por cromatografia de exclusão molecular

- 1) Lavagem da coluna Superdex 75™ 10/300 GL com água Milli-Q por cerca de 50 mL com um fluxo de 0,5 mL/min;
- 2) Equilibrar a coluna Superdex 75™ 10/300 GL com solução tampão 50mM Hepes a pH 7,5, 150mM NaCl e 2mM CaCl<sub>2</sub> por cerca de 50 mL continuando com um fluxo de 0,5 mL/min;
- 3) Injectar a mistura de proteínas em fracções de 500µL com a ajuda de uma seringa;

- 4) Lavagem da coluna Superdex 75<sup>TM</sup> 10/300 GL com solução tampão 50mM Hepes a pH 7,5, 150mM NaCl e 2mM CaCl<sub>2</sub> por cerca de 5 mL a um fluxo de 0,8 mL/min e continuando os restantes 20mL a um fluxo de 0,5 mL/min;
- 5) Lavagem da coluna Superdex 75<sup>TM</sup> 10/300 GL com 50 mL de água Milli-Q a um fluxo de 5 mL/min.
- 6) Guardar a coluna a 4°C em etanol 20%.

#### **7A.IV - Protocolo de purificação por cromatografia de permuta iónica**

- 1) Lavagem da coluna MonoQ<sup>TM</sup> 5/50 GL com cerca de 5 mL água Milli-Q a um fluxo de 2mL/min;
- 2) Equilibrar a coluna MonoQ<sup>TM</sup> 5/50 GL com solução tampão 20mM Tris a pH 7,5, 2mM CaCl<sub>2</sub> e 150mM NaCl por cerca de 5 mL continuando com um fluxo de 2 mL/min;
- 3) Equilibrar a coluna MonoQ<sup>TM</sup> 5/50 GL com solução tampão 20mM Tris a pH 7,5, 2mM CaCl<sub>2</sub> e 1M NaCl por cerca de 5 mL continuando com um fluxo de 2 mL/min;
- 4) Voltar a equilibrar a coluna MonoQ<sup>TM</sup> 5/50 GL com solução tampão 20mM Tris a pH 7,5, 2mM CaCl<sub>2</sub> e 150mM NaCl por cerca de 5 mL continuando com um fluxo de 2 mL/min;
- 5) Injectar a mistura de proteínas em fracções de 1,8mL com a ajuda de uma seringa;
- 4) Lavagem da coluna MonoQ<sup>TM</sup> 5/50 GL com solução tampão 20mM Tris a pH 7,5, 2mM CaCl<sub>2</sub> e 150mM NaCl por cerca de 10 mL continuando com um fluxo de 2 mL/min, até os valores de absorvância estabilizarem;
- 5) Realizar primeiro patamar de lavagem da coluna, por cerca de 10mL, com 10% da solução tampão 20mM Tris a pH 7,5, 2mM CaCl<sub>2</sub> e 1M NaCl (2mL/min);
- 5) Realizar gradiente de lavagem da coluna até se atingir a concentração máxima de 1M de NaCl a 2mL/min;
- 6) Lavagem da coluna MonoQ<sup>TM</sup> 5/50 GL com solução 1M NaOH por cerca de 5 mL;
- 7) Lavagem da coluna MonoQ<sup>TM</sup> 5/50 GL com água Milli-Q por cerca de 20 mL com um fluxo de 0,5 mL/min;
- 6) Guardar a coluna a 4°C em etanol 20%.

#### **7A.V - Screen de cristalização PEG/ION 1K**

**PEG/Ion 1K (1)** – 20% PEG 3350; 0.2M Fluoreto de sódio

**PEG/Ion 1K (2)** – 20% PEG 3350; 0.2M Fluoreto de potássio

**PEG/Ion 1K (3)** – 20% PEG 3350; 0.2M Fluoreto de amónio

**PEG/Ion 1K (4)** – 20% PEG 3350; 0.2M Cloreto de lítio

**PEG/Ion 1K (5)** – 20% PEG 3350; 0.2M Cloreto de magnésio

**PEG/Ion 1K (6)** – 20% PEG 3350; 0,2M Cloreto de sódio

**PEG/Ion 1K (7)** – 20% PEG 3350; 0,2M Cloreto de cálcio

**PEG/Ion 1K (8)** – 20% PEG 3350; 0.2M Cloreto de potássio

**PEG/Ion 1K (9)** – 20% PEG 3350; 0.2M Cloreto de amônio  
**PEG/Ion 1K (10)** – 20% PEG 3350; 0.2M Iodeto de sódio  
**PEG/Ion 1K (11)** – 20% PEG 3350; 0.2M Iodeto de potássio  
**PEG/Ion 1K (12)** – 20% PEG 3350; 0.2M Iodeto de amônio  
**PEG/Ion 1K (13)** – 20% PEG 3350; 0.2M Tiocianato de sódio  
**PEG/Ion 1K (14)** – 20% PEG 3350; 0.2M Tiocianato de potássio  
**PEG/Ion 1K (15)** – 20% PEG 3350; 0.2M Nitrato de lítio  
**PEG/Ion 1K (16)** – 20% PEG 3350; 0.2M Nitrato de magnésio  
**PEG/Ion 1K (17)** – 20% PEG 3350; 0.2M Nitrato de sódio  
**PEG/Ion 1K (18)** – 20% PEG 3350; 0.2M Nitrato de potássio  
**PEG/Ion 1K (19)** – 20% PEG 3350; 0.2M Nitrato de amônio  
**PEG/Ion 1K (20)** – 20% PEG 3350; 0.2M Formato de magnésio  
**PEG/Ion 1K (21)** – 20% PEG 3350; 0.2M Formato de sódio  
**PEG/Ion 1K (22)** – 20% PEG 3350; 0.2M Formato de potássio  
**PEG/Ion 1K (23)** – 20% PEG 3350; 0.2M Formato de amônio  
**PEG/Ion 1K (24)** – 20% PEG 3350; 0.2M Acetato de lítio  
**PEG/Ion 1K (25)** – 20% PEG 3350; 0.2M Acetato de magnésio  
**PEG/Ion 1K (26)** – 20% PEG 3350; 0.2M Acetato de zinco  
**PEG/Ion 1K (27)** – 20% PEG 3350; 0.2M Acetato de sódio  
**PEG/Ion 1K (28)** – 20% PEG 3350; 0.2M Acetato de cálcio  
**PEG/Ion 1K (29)** – 20% PEG 3350; 0.2M Acetato de potássio  
**PEG/Ion 1K (30)** – 20% PEG 3350; 0.2M Acetato de amônio  
**PEG/Ion 1K (31)** – 20% PEG 3350; 0.2M Sulfato de lítio  
**PEG/Ion 1K (32)** – 20% PEG 3350; 0.2M Sulfato de magnésio  
**PEG/Ion 1K (33)** – 20% PEG 3350; 0.2M Sulfato de sódio  
**PEG/Ion 1K (34)** – 20% PEG 3350; 0.2M Sulfato potássio  
**PEG/Ion 1K (35)** – 20% PEG 3350; 0.2M Sulfato amônio  
**PEG/Ion 1K (36)** – 20% PEG 3350; 0.2M Tartarato dissódico  
**PEG/Ion 1K (37)** – 20% PEG 3350; 0.2M Tartarato de sódio e potássio  
**PEG/Ion 1K (38)** – 20% PEG 3350; 0.2M Tartarato diamônico  
**PEG/Ion 1K (39)** – 20% PEG 3350; 0.2M Fosfato monossódico  
**PEG/Ion 1K (40)** – 20% PEG 3350; 0.2M Fosfato dissódico  
**PEG/Ion 1K (41)** – 20% PEG 3350; 0.2M Fosfato monopotássico  
**PEG/Ion 1K (42)** – 20% PEG 3350; 0.2M Fosfato dipotássico  
**PEG/Ion 1K (43)** – 20% PEG 3350; 0.2M Fosfato de amônio  
**PEG/Ion 1K (44)** – 20% PEG 3350; 0.2M Fosfato diamônico  
**PEG/Ion 1K (45)** – 20% PEG 3350; 0.2M Citrato de lítio

**PEG/Ion 1K (46)** – 20% PEG 3350; 0.2M Citrato de sódio  
**PEG/Ion 1K (47)** – 20% PEG 3350; 0.2M Citrato de potássio  
**PEG/Ion 1K (48)** – 20% PEG 3350; 0.2M Citrato diamónico

#### **7A.VI - Screen de cristalização PEG/ION 8K**

**PEG/Ion 8K (1)** – 20% PEG 8K; 0.2 M Fluoreto de sódio; pH 7.1  
**PEG/Ion 8K (2)** – 20% PEG 8K; 0.2 M Fluoreto de potássio; pH 7.2  
**PEG/Ion 8K (3)** – 20% PEG 8K; 0.2 M Fluoreto de amónio; pH 6.2  
**PEG/Ion 8K (4)** – 20% PEG 8K; 0.2 M Cloreto de lítio anidro; pH 6.7  
**PEG/Ion 8K (5)** – 20% PEG 8K; 0.2 M Cloreto de magnésio hexahidratado; pH 5.8  
**PEG/Ion 8K (6)** – 20% PEG 8K; 0.2 M Cloreto de sódio; pH 6.9  
**PEG/Ion 8K (7)** – 20% PEG 8K; 0.2 M Cloreto de cálcio dihidratado; pH 5.1  
**PEG/Ion 8K (8)** – 20% PEG 8K; 0.2 M Cloreto de potássio; pH 6.9  
**PEG/Ion 8K (9)** – 20% PEG 8K; 0.2 M Cloreto de amónio; pH 6.3  
**PEG/Ion 8K (10)** – 20% PEG 8K; 0.2 M Iodeto de sódio; pH 6.9  
**PEG/Ion 8K (11)** – 20% PEG 8K; 0.2 M Iodeto de potássio; pH 6.8  
**PEG/Ion 8K (12)** – 20% PEG 8K; 0.2 M Iodeto de amónio; pH 6.2  
**PEG/Ion 8K (13)** – 20% PEG 8K; 0.2 M Tiocianato de sódio; pH 6.9  
**PEG/Ion 8K (14)** – 20% PEG 8K; 0.2 M Tiocianato de potássio; pH 7  
**PEG/Ion 8K (15)** – 20% PEG 8K; 0.2 M Nitrato de lítio; pH 7.1  
**PEG/Ion 8K (16)** – 20% PEG 8K; 0.2 M Nitrato de magnésio hexahidratado; pH 5.8  
**PEG/Ion 8K (17)** – 20% PEG 8K; 0.2 M Nitrato de sódio; pH 6.8  
**PEG/Ion 8K (18)** – 20% PEG 8K; 0.2 M Nitrato de potássio; pH 6.9  
**PEG/Ion 8K (19)** – 20% PEG 8K; 0.2 M Nitrato de amónio; pH 6.3  
**PEG/Ion 8K (20)** – 20% PEG 8K; 0.2M Formato de magnésio; pH 5.9  
**PEG/Ion 8K (21)** – 20% PEG 8K; 0.2 M Formato de sódio; pH 7.2  
**PEG/Ion 8K (22)** – 20% PEG 8K; 0.2 M Formato de potássio; pH 7.3  
**PEG/Ion 8K (23)** – 20% PEG 8K; 0.2 M Formato de amónio; pH 6.6  
**PEG/Ion 8K (24)** – 20% PEG 8K; 0.2M Acetato de lítio dihidratado; pH 7.8  
**PEG/Ion 8K (25)** – 20% PEG 8K; 0.2M Acetato de magnésio tetrahidratado; pH 7.7  
**PEG/Ion 8K (26)** – 20% PEG 8K; 0.2M Acetato de zinco dihidratado; pH 6.3  
**PEG/Ion 8K (27)** – 20% PEG 8K; 0.2M Acetato de sódio trihidratado; pH 7.9  
**PEG/Ion 8K (28)** – 20% PEG 8K; 0.2M Acetato de cálcio hidratado; pH 7.3  
**PEG/Ion 8K (29)** – 20% PEG 8K; 0.2M Acetato de potássio; pH 7.8  
**PEG/Ion 8K (30)** – 20% PEG 8K; 0.2M Acetato de amónio; pH 7.1  
**PEG/Ion 8K (31)** – 20% PEG 8K; 0.2M Sulfato de lítio monohidratado; pH 6.4  
**PEG/Ion 8K (32)** – 20% PEG 8K; 0.2M Sulfato de magnésio heptahidratado; pH 5.9

**PEG/Ion 8K (33)** – 20% PEG 8K; 0.2M Sulfato de sódio decahidratado; pH 6.6  
**PEG/Ion 8K (34)** – 20% PEG 8K; 0.2M Sulfato de potássio; pH 6.7  
**PEG/Ion 8K (35)** – 20% PEG 8K; 0.2M Sulfato de amônio; pH 6  
**PEG/Ion 8K (36)** – 20% PEG 8K; 0.2M Tartarato di-sódio dihidratado; pH 7.2  
**PEG/Ion 8K (37)** – 20% PEG 8K; 0.2M Tartarato de potássio/sódio tetrahidratado; pH 7.2  
**PEG/Ion 8K (38)** – 20% PEG 8K; 0.2M Tartarato di-amônio; pH 6.6  
**PEG/Ion 8K (39)** – 20% PEG 8K; 0.2M Fosfato dihidrogénio de sódio monohidratado; pH 4.5  
**PEG/Ion 8K (40)** – 20% PEG 8K; 0.2M Fosfato hidrogénio de disódio dihidratado; pH 9.1  
**PEG/Ion 8K (41)** – 20% PEG 8K; 0.2M Fosfato de potássio dihidrogénio; pH 4.7  
**PEG/Ion 8K (42)** – 20% PEG 8K; 0.2M Fosfato hidrogénio di-potássio; pH 9.2  
**PEG/Ion 8K (43)** – 20% PEG 8K; 0.2M Fosfato dihidrogénio de amônio; pH 4.6  
**PEG/Ion 8K (44)** – 20% PEG 8K; 0.2M Fosfato hidrogénio di-amônio; pH 7.9  
**PEG/Ion 8K (45)** – 20% PEG 8K; 0.2M Citrato tri-lítio tetrahidratado; pH 8.1  
**PEG/Ion 8K (46)** – 20% PEG 8K; 0.2M Citrato tri-sódio dihidratado; pH 8.2  
**PEG/Ion 8K (47)** – 20% PEG 8K; 0.2M Citrato tri-potássio monohidratado; pH 8.3  
**PEG/Ion 8K (48)** – 20% PEG 8K; 0.2M Citrato hidrogénio di-amônio; pH 5

#### **7A.VII - Screen de cristalização JBS 5**

**JBS5 (1)** – 12% PEG 8K; 5% Glicerol; 0.1M Cloreto de potássio  
**JBS5 (2)** – 12% PEG 8K; 10% Glicerol; 0.1M Cloreto de potássio  
**JBS5 (3)** – 15% PEG 8K; 0.2M Sulfato de amônio  
**JBS5 (4)** – 15% PEG 8K; 0.5M Sulfato de lítio  
**JBS5 (5)** – 15% PEG 8K; 0.1M Mes pH6.5; 0.2M Acetato de sódio  
**JBS5 (6)** – 15% PEG 8K; 0.05M Sulfato de amônio; 0.1M Citrato de sódio  
**JBS5 (7)** – 18% PEG 8K; 0.1M Hepes pH 7.5; 0.2M Acetato de cálcio  
**JBS5 (8)** – 18% PEG 8K; 0.1M Hepes pH 7.5; 2% isopropanol; 0.1M Acetato do sódio  
**JBS5 (9)** – 18% PEG 8K; 0.1M Tris pH 8.5; 0.2M Sulfato de lítio  
**JBS5 (10)** – 20% PEG 8K; 0.1M Hepes pH 7.5  
**JBS5 (11)** – 20% PEG 8K; 0.1M Hepes pH 6.5; 0.2M Acetato de magnésio  
**JBS5 (12)** – 20% PEG 8K; 0.1M Tris pH 9.5  
**JBS5 (13)** – 22% PEG 8K; 0.1M Mes pH 6.5; 0.2M Sulfato de amônio  
**JBS5 (14)** – 25% PEG 8K; 0.2M Cloreto de lítio  
**JBS5 (15)** – 30% PEG 8K; 0.2M Sulfato de amônio  
**JBS5 (16)** – 8% PEG 10K; 0.1M Acetato pH 4.5  
**JBS5 (17)** – 14% PEG 10K; 0.1M Imidazol pH 8  
**JBS5 (18)** – 16% PEG 10K; 0.1M Tris pH 8.5  
**JBS5 (19)** – 18% PEG 10K; 0.1M Tris pH 8.5

- JBS5 (20)** – 20% PEG 10K; 0.1M Hepes pH 8.5; 20% Glicerol; 0.1M Cloreto de sódio  
**JBS5 (21)** – 30% PEG 10K; 0.1M Hepes pH 7.5  
**JBS5 (22)** – 30% PEG 10K; 0.1M Tris pH 8.5  
**JBS5 (23)** – 10% PEG 20K; 0.1M Mes pH 6.5  
**JBS5 (24)** – 17% PEG 20K; 0.1M Tris pH 8.5; 0.1M Cloreto de magnésio

#### **7A.VIII – *Screen* de cristalização JBS 6**

- JBS6 (1)** – 0.5M Sulfato de amônio; 1M Sulfato de lítio; 0.1M Citrato de sódio  
**JBS6 (2)** – 1M Sulfato de amônio  
**JBS6 (3)** – 1M Sulfato de amônio; 0.1M Acetato pH 4.5  
**JBS6 (4)** – 1M Sulfato de amônio; 0.1M Hepes pH 7.5; 2% PEG 400  
**JBS6 (5)** – 1M Sulfato de amônio; 0.1M Tris pH 8.5  
**JBS6 (6)** – 1.2M Sulfato de amônio; 3% Isopropanol; 0.05M Citrato de sódio  
**JBS6 (7)** – 1.5M Sulfato de amônio; 0.1M Tris pH 8.5; 15% Glicerol  
**JBS6 (8)** – 1.6M Sulfato de amônio; 0.1M Cloreto de lítio  
**JBS6 (9)** – 1.6M Sulfato de amônio; 1M Sulfato de lítio  
**JBS6 (10)** – 1.6M Sulfato de amônio; 0.1M Hepes pH 7.5; 0.2M Cloreto de sódio  
**JBS6 (11)** – 1.6M Sulfato de amônio; 0.1M Hepes pH 7.5; 2% PEG 1K  
**JBS6 (12)** – 1.8M Sulfato de amônio; 0.1M Mes pH 8.5  
**JBS6 (13)** – 2M Sulfato de amônio; 2M Cloreto de sódio  
**JBS6 (14)** – 2M Sulfato de amônio; 0.1M Acetato pH 4.5  
**JBS6 (15)** – 2M Sulfato de amônio; 0.1M Mes pH 6.5; 5% PEG 400  
**JBS6 (16)** – 2M Sulfato de amônio  
**JBS6 (17)** – 2.2M Sulfato de amônio  
**JBS6 (18)** – 2.2M Sulfato de amônio; 20% Glicerol  
**JBS6 (19)** – 2.4M Sulfato de amônio; 0.1M Citrato de sódio  
**JBS6 (20)** – 3M Sulfato de amônio; 1% MPD  
**JBS6 (21)** – 3M Sulfato de amônio; 10% Glicerol  
**JBS6 (22)** – 3.5M Sulfato de amônio; 0.1M Hepes pH 7.5  
**JBS6 (23)** – 3.5M Sulfato de amônio; 0.1M Mes pH 6.5; 1% MPD  
**JBS6 (24)** – 3.5M Sulfato de amônio

#### **7A.IX – *Screen* de cristalização JBS 7**

- JBS7 (1)** – 10% MPD; 0.1M Hepes pH 7.5, 0.1M Citrato de sódio  
**JBS7 (2)** – 12% MPD; 0.1M Tris pH 8.5; 0.05M Cloreto de magnésio  
**JBS7 (3)** – 15% MPD; 0.1M Acetato pH 4.5; 0.02M Cloreto de cálcio  
**JBS7 (4)** – 15% MPD; 0.1M Imidazol pH 8; 5% PEG 4K

**JBS7 (5)** – 15% MPD; 0.1M Citrati pH 5.5; 0.2M Acetato de amónio  
**JBS7 (6)** – 15% MPD; 0.1M Mes pH 6.5; 0.2M Cloreto de magnésio  
**JBS7 (7)** – 15% MPD; 0.1M Hepes pH 7.5; 0.2M Citrato de sódio  
**JBS7 (8)** – 20% MPD; 0.1M Hepes pH 7.5; 0.1M Citrato de sódio  
**JBS7 (9)** – 20% MPD; 0.1M Imidazol pH 8  
**JBS7 (10)** – 20% MPD; 4% Glicerol; 0.2M Cloreto de sódio  
**JBS7 (11)** – 30% MPD; 0.1M Acetato pH 4.5; 0.02M Cloreto de cálcio  
**JBS7 (12)** – 30% MPD; 0.1M Citrato pH 5.5; 0.2M Acetato de amónio  
**JBS7 (13)** – 30% MPD; 0.1M Mes pH 6.5; 0.2M Acetato de magnésio  
**JBS7 (14)** – 30% MPD; 0.1M Hepes pH 7.5; 0.5M Sulfato de amónio  
**JBS7 (15)** – 30% MPD; 0.1M Hepes pH 7.5; 0.2M Citrato de sódio  
**JBS7 (16)** – 30% MPD; 0.1M Hepes pH 7.5; 5% PEG 4K  
**JBS7 (17)** – 30% MPD; 0.1M Imidazol pH 8; 10% PEG 4K  
**JBS7 (18)** – 30% MPD; 20% Etanol  
**JBS7 (19)** – 35% MPD  
**JBS7 (20)** – 35% MPD; 0.1M Imidazol pH 8  
**JBS7 (21)** – 40% MPD; 0.1M Tris pH 8.5  
**JBS7 (22)** – 47% MPD; 0.1M Hepes pH 7.5  
**JBS7 (23)** – 50% MPD

#### **7A.X – Screen de cristalização JBS 8**

**JBS8 (1)** – 50% MPD; 15% Etanol; 0.01M Acetato de sódio  
**JBS8 (2)** – 50% MPD; 2% Isopropanol; 0.05M Acetato de sódio; 0.05M Cloreto de sódio  
**JBS8 (3)** – 50% MPD; 0.1M Tris pH 8.5; 0.2M Fosfato de amónio  
**JBS8 (4)** – 55% MPD  
**JBS8 (5)** – 60% MPD; 0.1M Acetato pH 4.5; 0.01M Cloreto de cálcio  
**JBS8 (6)** – 60% MPD; 0.02M Acetato de sódio  
**JBS8 (7)** – 70% MPD; 0.1M Mes pH 6.5  
**JBS8 (8)** – 70% MPD; 0.1M Tris pH 8.5; 0.01M Cloreto de cálcio  
**JBS8 (9)** – 2% Etanol; 0.1M Tris pH 8.5  
**JBS8 (10)** – 5% Etanol; 0.1M Hepes pH 7.5; 5% MPD  
**JBS8 (11)** – 5% Etanol; 0.1M Tris pH 8.5; 5% MPD; 0.2M Cloreto de sódio  
**JBS8 (12)** – 10% Etanol; 0.1M Tris pH 8.5  
**JBS8 (13)** – 12% Etanol; 0.1M Acetato pH 4.5; 4% PEG 400  
**JBS8 (14)** – 14% Etanol; 0.1M Tris pH 8.5  
**JBS8 (15)** – 18% Etanol; 0.1M Tris pH 8.5  
**JBS8 (16)** – 20% Etanol

- JBS8 (17)** – 20% Etanol; 10% Glicerol  
**JBS8 (18)** – 30% Etanol; 10% PEG 6K; 0.1M Acetato de sódio  
**JBS8 (19)** – 45% Etanol  
**JBS8 (20)** – 50% Etanol; 0.01M Acetato de sódio  
**JBS8 (21)** – 60% Etanol; 1.5% PEG 6K; 0.05M Acetato de sódio  
**JBS8 (22)** – 60% Etanol; 0.1M Cloreto de sódio  
**JBS8 (23)** – 2% Isopropanol; 0.1M Tris pH 8.5; 0.01M Sulfato de magnésio

#### **7A.XI – Screen de cristalização JCSG+**

- JCSG (1)** – 50% PEG 400; 0.1 M Acetato de sódio pH=4.5; 0.2 M Sulfato de lítio  
**JCSG (2)** – 20% PEG 3000; 0.1 M Acetato de sódio pH=5.5; 200 mM Sulfato de amónio  
**JCSG (3)** – 30% PEG 3350; 0.2 M Citrato di-amónio hidrogénio  
**JCSG (4)** – 30% MPD; 0.1 M Acetato de sódio pH=4.6; 0.02 M Cloreto de cálcio  
**JCSG (5)** – 20% PEG 3350; 0.2 M Formato de magnésio  
**JCSG (6)** – 20% PEG 1000; 0.1 M Fosfato/citrato pH=4.2; 0.2 M Sulfato de lítio  
**JCSG (7)** – 20% PEG 8000; 0.1 M CHES pH=9.5  
**JCSG (8)** – 20% PEG 3350; 0.2 M Formato de amónio  
**JCSG (9)** – 20% PEG 3350; 0.2 M Cloreto de amónio  
**JCSG (10)** – 20% PEG 3350; 0.2 M Formato de potássio  
**JCSG (11)** – 50% MPD; 0.1 M Tris-HCl pH=8.5; 0.2 M Fosfato dihidrogénio de amónio  
**JCSG (12)** – 20% PEG 3350; 0.2 M Nitrato de potássio  
**JCSG (13)** – 0.1 M Citrato pH=4; 0.8 M Sulfato de amónio  
**JCSG (14)** – 20% PEG 3350; 0.2 M Tiocianato de sódio  
**JCSG (15)** – 20% PEG 6000; 0.1 M Bicina pH=9  
**JCSG (16)** – 10% PEG 6000; 10% PEG 8000; 8% Etilenoglicol; 0.1 M HEPES pH=7.5  
**JCSG (17)** – 40% MPD; 5% PEG 8000; 0.1 M Cacodilato de sódio pH=6.5  
**JCSG (18)** – 40% Etanol; 5% PEG 1000; 0.1 M Fosfato/citrato pH=4.2  
**JCSG (19)** – 8% PEG 4000; 0.1 M Acetato de sódio pH=4.6  
**JCSG (20)** – 10% PEG 8000; 0.1 M Tris-HCl pH=7; 0.2 mM Cloreto de magnésio  
**JCSG (21)** – 20% PEG 6000; 0.1 M citrato pH=5; 100 mM Acetato de sódio  
**JCSG (22)** – 50% PEG 200; 0.1 M Cacodilato de sódio pH=6.5; 0.2 M Cloreto de magnésio  
**JCSG (23)** – 1.6 M Citrato tri-sódio  
**JCSG (24)** – 20% PEG 3350; 0.2M Citrato tri-potássio  
**JCSG (25)** – 20% PEG 8000; 0.1 M Fosfato/citrato pH=4.2; 0.2 M Cloreto de sódio  
**JCSG (26)** – 20% PEG 6000; 0.1 M Citrato de sódio pH=4; 1 M Cloreto de lítio  
**JCSG (27)** – 20% PEG 3350; 0.2 M Nitrato de amónio  
**JCSG (28)** – 10% PEG 6000; 0.1 M NaHEPES pH=7; 0.8 M Fosfato dihidrogénio de sódio

**JCSG (29)** – 0.8 M Fosfato dihidrogénio de sódio; 0.8 M Fosfato dihidrogénio de potássio; 0.1 M NaHEPES pH=7.5

**JCSG (30)** – 40% PEG 300; 0.1 M Fosfato/citrato pH=4.2

**JCSG (31)** – 10% PEG 3000; 0.1 M Acetato de sódio pH=4.5; 0.2 M Acetato de zinco

**JCSG (32)** – 20% Etanol; 0.1 M Tris-HCl pH=8.5

**JCSG (33)** – 25% 1,2-Propanodiol; 10% Glicerol; 0.1 M Fosfato sódio/potássio pH=6.2

**JCSG (34)** – 10% PEG 20000; 2% Dioxano; 0.1 M Bicina pH=9

**JCSG (35)** – 2 M Sulfato de amónio; 0.1 M Acetato de sódio pH= 4.6

**JCSG (36)** – 10% PEG 1000; 10% PEG 8000

**JCSG (37)** – 24% PEG 1500; 20% Glicerol

**JCSG (38)** – 30% PEG 400; 0.1M Na HEPES pH=7.5; 0.2 M Cloreto de magnésio

**JCSG (39)** – 50% PEG 200; 0.1 M Fosfato de sódio/potássio pH=6.2; 0.2 M Cloreto de sódio

**JCSG (40)** – 30% PEG 8000; 0.1 M Acetato de sódio pH=4.5; 0.2 M Sulfato de lítio

**JCSG (41)** – 70% MPD; 0.1 M HEPES pH=7.5

**JCSG (42)** – 20% PEG 8000; 0.1 M Tris-HCl pH=8.5; 0.2 M Cloreto de magnésio

**JCSG (43)** – 40% PEG 400; 0.1 M Tris-HCl pH=8.5; 0.2 M Sulfato de lítio

**JCSG (44)** – 40% MPD; 0.1 M Tris-HCl pH=8

**JCSG (45)** – 25.5% PEG 4000; 15% Glicerol; 0.17 M Sulfato de amónio

**JCSG (46)** – 40% PEG 300; 0.1 M Cacodilato de sódio pH=6.5; 0.2 M Acetato de cálcio

**JCSG (47)** – 14% 2-propanol; 30% Glicerol; 0.07 M Acetato de sódio pH=4.6; 0.14 M Cloreto de cálcio

**JCSG (48)** – 16% PEG 8000; 20% Glicerol; 0.04 M Fosfato dihidrogénio de potássio

#### **7A.XII – Screen de cristalização 80!**

- 80! (1)** - 0,1 M Acetato pH=4,5; 0,2 M Cloreto de cálcio; 30% MPD
- 80! (2)** - 0,1 M MES pH=6,5; 1 M Tartarato de potássio/sódio;
- 80! (3)** - 0,4 M Fosfato de amónio
- 80! (4)** - 0,1 M Tris-HCl pH=8,5; 3 M Sulfato de amónio
- 80! (5)** - 0,1 M HEPES pH=7,5; 0,2 M Citrato de sódio; 30% MPD
- 80! (6)** - 0,1 M Acetato pH=4,5; 0,2 M Cloreto de magnésio; 30% PEG 4K – **A6**
- 80! (7)** - 0,1 M HEPES pH=7,5; 1,2 M Citrato de sódio
- 80! (8)** - 0,2 M Citrato de sódio; 2M Sulfato de amónio – **A8**
- 80! (9)** - 0,1 M Citrato pH=5,5; 0,2 M Acetato de amónio; 30% PEG 400
- 80! (10)** -0,1 M Acetato pH=4,5; 1,5 M Fosfato de amónio
- 80! (11)** -0,1 M HEPES pH=7,5; 1,5 M Fosfato hidrogénio di-potássio; 1,5 M Fosfato hidrogénio di-sódio; 0,2 M Sulfato de amónio
- 80! (12)** -0,1 M Tris-HCl pH=8,5; 0,2 M Citrato de sódio; 20% PEG 400

- 80! (13)** -0,1 M HEPES pH=7,5; 0,2 M Cloreto de cálcio; 25% PEG 4K
- 80! (14)** -0,1 M MES pH=6,5; 0,1 M Cloreto de magnésio; 30% PEG 8K
- 80! (15)** -0,1 M Citrato pH=5,5; 0,2M Sulfato de lítio; 30% PEG 4K – **B3**
- 80! (16)** -0,1 M Acetato pH=4,5; 1 M Sulfato de lítio
- 80! (17)** -0,1 M Tris-HCl pH=7,5; 0,2 M Fosfato de amónio; 30% MPD – **B5**
- 80! (18)** -0,1 M Tris-HCl pH=7,5; 0,2 M Acetato de amónio; 1,5 M Fosfato hidrogénio di-potássio; 1,5 M Fosfato dihidrogénio de sódio
- 80! (19)** -0,1 M Citrato pH=5,5; 0,1 M Sulfato de amónio; 30% PEG 8K
- 80! (20)** -0,1 M MES pH=6,5; 30% MPD
- 80! (21)** -0,1 M HEPES pH=7,5; 0,2 M Cloreto de magnésio; 30% PEG 4K – **B9**
- 80! (22)** -0,1 M Tris-HCl pH=8,5; 0,2 M Acetato de sódio; 30% PEG 4K
- 80! (23)** -0,1 M Tris-HCl pH=7,5; 1 M Tartarato de potássio/sódio;
- 80! (24)** -0,1 M Tris-HCl pH=8,5; 0,2 M Cloreto de cálcio
- 80! (25)** -0,1 M Citrato pH=5,5; 0,5 M Acetato de amónio; 30% MPD
- 80! (26)** -0,1 M MES pH=6,5; 2 M Acetato de sódio
- 80! (27)** -0,1 M MES pH=6,5; 0,2 M Tartarato de potássio/sódio; 30% PEG 8K – **C3**
- 80! (28)** -0,1 M HEPES pH=7,5; 1 M Tartarato de potássio/sódio
- 80! (29)** -0,1 M Acetato pH=4,5; 0,2 M Sulfato de amónio; 30% PEG 400
- 80! (30)** -0,1 M HEPES pH=7,5; 0,1 M Sulfato de amónio; 20% PEG 4K
- 80! (31)** -0,1 M MES pH=6,5; 2 M Sulfato de amónio
- 80! (32)** -0,1 M MES pH=6,5; 0,2 M Cloreto de sódio; 30% Etanol
- 80! (33)** -0,1 M HEPES pH=7,5; 0,2 M Cloreto de magnésio; 30% Etanol
- 80! (34)** -0,1 M Tris-HCl pH=8,5; 0,2 M Acetato de amónio; 30% Etanol
- 80! (35)** -0,1 M Acetato pH=4,5; 0,2 M Cloreto de cálcio; 30% Etanol
- 80! (36)** -0,1 M HEPES pH=7,5; 0,2 M Acetato de sódio; 30% Etanol
- 80! (37)** -0,1 M HEPES pH=7,5; 0,2 M Cloreto de magnésio; 30% Isopropanol
- 80! (38)** -0,1 M Cacodilato pH=6,5; 30% MPD
- 80! (39)** -0,1 M Acetato pH=4,5; 2 M Formato de sódio
- 80! (40)** -0,1 M Cacodilato pH=6,5; 0,2 M Citrato de sódio; 40% Isopropanol
- 80! (41)** -0,1 M HEPES pH=7,5; 20% PEG 400; 10% Isopropanol
- 80! (42)** -0,1 M HEPES pH=7,5; 1 M Sulfato de lítio
- 80! (43)** -0,1 M Tris-HCl pH=8,5; 0,2 M Sulfato de lítio; 30% PEG 4K
- 80! (44)** -0,1 M Cacodilato pH=6,5; 0,2 M Sulfato de amónio; 30% PEG 6K
- 80! (45)** -0,1 M Acetato pH=4,5; 1,5 M Acetato de sódio
- 80! (46)** -0,1 M Citrato de sódio; 1 M Fosfato dihidrogénio de amónio
- 80! (47)** -4M Formato de sódio

- 80! (48)** -0,1 M HEPES pH=7,5; 1,2 M Citrato de sódio
- 80! (49)** -0,4 M Tartarato de potássio/sódio
- 80! (50)** -0,1 M Tris-HCl pH=8,5; 0,2 M Cloreto de magnésio; 30% PEG 4K
- 80! (51)** -0,1 M Cacodilato pH=6,5; 1,4 M Acetato de sódio
- 80! (52)** -0,1 M Citrato pH=5,5; 0,2 M Acetato de amónio; 30% PEG 4K
- 80! (53)** -0,1 M Citrato pH=4,5; 0,2 M Acetato de amónio 30% PEG 4K – **E5**
- 80! (54)** -0,1 M HEPES pH=7,5; 0,2 M Cloreto de cálcio; 28% PEG 400 – **E6**
- 80! (55)** -0,1 M Cacodilato pH=6,5; 0,2 M Sulfato de amónio; 30% PEG 8K
- 80! (56)** -0,1 M Cacodilato pH=6,5; 0,2 M Acetato de magnésio; 30% PEG 8K – **E8**
- 80! (57)** -0,1 M Tris-HCl pH=8,5; 0,2 M Acetato hidrogénio de amónio; 30% Isopropanol
- 80! (58)** -0,1 M Acetato pH=4,5; 0,2 M Sulfato de amónio; 25% PEG 4K
- 80! (59)** -0,1 M Cacodilato pH=6,5; 0,2 M Acetato de magnésio; 30% MPD
- 80! (60)** -0,1 M Acetato pH=4,5; 0,2 M Cloreto de cálcio; 20 % Isopropanol
- 80! (61)** -0,1 M Imidazol pH=7; 1 M Acetato de sódio;
- 80! (62)** -0,1 M Cacodilato pH=6,5; 0,2 M Citrato de sódio; 20 % Isopropanol
- 80! (63)** -0,1 M Cacodilato pH=6,5; 0,2 M Acetato de sódio; 30% PEG 8K
- 80! (64)** -0,2 M Sulfato de amónio; 30% PEG 8K – **F4**
- 80! (65)** -0,2 M Sulfato de amónio; 30% PEG 4K
- 80! (66)** -66. 0,1 M HEPES pH=7,5; 1,6 M Fosfato hidrogénio di-potássio; 1,6 M Fosfato dihidrogénio de sódio
- 80! (67)** -0,1 M Tris-HCl pH=8,5; 8% PEG 8K
- 80! (68)** -0,1 M Acetato pH=4,5; 8% PEG 4K
- 80! (69)** -0,1 M HEPES pH=7,5; 1,8 M Fosfato dihidrogénio de amónio; 2% PEG 400
- 80! (70)** -0,1 M Citrato pH=5,5; 20% PEG 4K; 20 % Isopropanol
- 80! (71)** -0,05 M Fosfato hidrogénio di-potássio; 20% PEG 8K
- 80! (72)** -30% PEG 1K
- 80! (73)** -0,2 M Formato de magnésio
- 80! (74)** -0,1 M Cacodilato pH=6,5; 0,2 M Acetato de zinco; 18% PEG 8K
- 80! (75)** -0,1 M Cacodilato pH=6,5; 0,2 M Acetato de cálcio; 18% PEG 8K
- 80! (76)** -0,1 M Acetato pH=4,5; 2 M Sulfato de amónio
- 80! (77)** -0,1 M Tris-HCl pH=8,5; 2 M Sulfato de amónio
- 80! (78)** -1 M Sulfato de lítio; 2% PEG 8K
- 80! (79)** -0,5 M Sulfato de lítio; 18% PEG 8K
- 80! (80)** -0,1 M Citrato pH=5,5; 0,2 M Acetato de amónio; 20 % PEG 4K; 20 % Isopropanol

## 7B – Anexos referentes ao estudo das interações coesina-doquerina no *C. thermocellum* e *R. flavofaciens* por microarrays

### 7B.I - Informação sobre os módulos coesina e doquerina

**Tabela 7B.1** - Informação sobre os módulos coesina em estudo do *C. thermocellum*. A tabela apresenta o número de resíduos de aminoácidos (aa) presentes na sequência, o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta ExPASy

ID do gene	Proteína	Domínio a que a proteína está associada	Número de aa	PM (Da(gmol <sup>-1</sup> ))	$\epsilon$ (M <sup>-1</sup> .cm <sup>-1</sup> )
Cthe_1307	C1	Tipo I – Proteína de ancoragem CipA	164	17346,7	5960,0
Cthe_1307	C2	Tipo I - Proteína de ancoragem CipA	164	17688,8	8940,0
Cthe_1307	C3	Tipo I - Proteína de ancoragem CipA	167	18089,4	7450,0
Cthe_1307	C4	Tipo I - Proteína de ancoragem CipA	167	18104,3	7450,0
Cthe_1307	C7	Tipo I - Proteína de ancoragem CipA	167	18123,4	7450,0
Cthe_1307	C9	Tipo I - Proteína de ancoragem CipA	168	17998,1	7450,0
s/i	OlpA	Tipo I - Proteína de ancoragem CipA	172	18999,5	5960,0
s/i	OlpC	Tipo I – Proteína OlpC (parede celular)	177	19781,3	10430,0
Cthe_0735	Novel2	Tipo II – Proteína Novel 2 (parede celular)	191	20812,3	18910,0
s/i	SdbA	Tipo II - Proteína SdbA (parede celular)	189	20704,3	15930,0
Cthe_3079	Orf2-1	Tipo II – Proteína Orf2 (parede celular)	187	20764,5	17420,0
Cthe_3079	Orf2-2	Tipo II - Proteína Orf2 (parede celular)	186	20518,2	24410,0
Cthe_3078	EAM45291-1	Tipo II – Proteína OlpB (parede celular)	187	20469,9	14440,0
Cthe_3078	EAM45291-2	Tipo II - Proteína OlpB (parede celular)	188	20392,9	24410,0
Cthe_0736	EAM45267-1	Tipo II – Proteína Ext	187	20623,1	20400,0
Cthe_0736	EAM45267-2	Tipo II - Proteína Ext	184	20076,7	17420,0
Cthe_0736	EAM45267-3	Tipo II - Proteína Ext	186	20243,7	15930,0
Cthe_0736	EAM45267-4	Tipo II - Proteína Ext	188	20427,9	20400,0
Cthe_0736	EAM45267-5	Tipo II - Proteína Ext	184	20419,1	21890,0
Cthe_0736	EAM45267-6	Tipo II - Proteína Ext	185	20672,2	23380,0
Cthe_0736	EAM45267-7	Tipo II - Proteína Ext	183	20244,6	20400,0

**Tabela 7B.2** - Informação sobre os módulos do querina em estudo do *C. thermocellum*. A tabela apresenta o número de resíduos de aminoácidos presentes na sequência, o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta ExPASy.

ID do gene	Domínio a que a proteína está associado	Proteína	Número de aa	PM (Da (g mol <sup>-1</sup> ))	$\epsilon$ (M <sup>-1</sup> cm <sup>-1</sup> )
Cthe_3077	Tipo II – Proteína de ancoragem	XDoc2	284	30612,7	31400,0
Cthe_0624	Beta-glicosidase - Família 9	CELJ	194	21316,0	22920,0
Cthe_0918	Proteína de função desconhecida	UNK_1	190	20780,0	18450,0
Cthe_0912	Beta-glicosidases - Família 10	Xyn10B	193	20891,0	15470,0
Cthe_0435	Módulo com homologia a proteínas de ligação ao DNA	DNA_binding	195	21534,5	25440,0
Cthe_0258	Proteína de função desconhecida	UNK_7	194	21096,2	22460,0
Cthe_2271	Proteína de função desconhecida	DB1	194	21235,1	24410,0
Cthe_2271	Proteína de função desconhecida	DB2	196	21254,9	21430,0
Cthe_1806	Imunoglobulina - Grupo 2	XDoc22	291	31896,9	28420,0
Cthe_1963	Beta-glicosidase – Família 10	XylZ	193	20869,7	16960,0
Cthe_1472	Beta-glicosidase – Família 5	CelH	199	21938,0	16960,0
Cthe_0269	Beta-glicosidase – Família 8	CelA	193	21055,9	19940,0
Cthe_2089	Beta-glicosidase – Família 48	GH48 (CelS)	192	20869,5	18450,0
Cthe_0043	Beta-glicosidase – Família 9	Doc_0043	190	20373,1	15470,0
Cthe_3012	Módulo de ligação a hidratos de carbono – Família 6	Doc_3012	191	20653,4	19940,0
Cthe_2949	Esterase de pectinas	Doc_2949	193	21134,0	19940,0
Cthe_2179	Liase de ácido poligalacturônico	Doc_2179	192	21042,9	18450,0
Cthe_0661	Ricina – Família B das Lectinas	Doc_0661	190	20782,5	16960,0
Cthe_2195	Módulo de ligação a hidratos de carbono – Família 6	Doc_2195	191	20860,8	15470,0
Cthe_1398	Proteína de função desconhecida	Doc_1398	195	21385,3	25440,0
Cthe_2812	Beta-glicosidase – Família 9	Doc_2812	191	21001,8	19940,0
Cthe_1400	Beta-glicosidase – Família 53	Doc_1400	195	21323,2	22920,0
Cthe_2972	Beta-glicosidase – Família 11	Doc_2972	193	20828,6	16960,0
Cthe_0729	Proteína de função desconhecida	Doc_0729	192	21057,8	22920,0
Cthe_0543	Beta-glicosidase – Família 9	Doc_0543	198	21926,9	19940,0
Cthe_0821	Proteína com homologia ao domínio do fator de coagulação tipo 5/8	Doc_0821	194	21158,1	22460,0
Cthe_2549	Proteína de função desconhecida	Doc_2549	194	21287,1	22920,0
Cthe_2196	Beta-glicosidase – Família 43	Doc_2196	192	20934,7	15470,0
Cthe_0640	Proteína de função desconhecida	Doc_0640	193	21055,7	18450,0
Cthe_1271	Módulo de ligação a hidratos de carbono – Família 6	Doc_1271	191	20613,2	15470,0
Cthe_2138	Beta-glicosidase – Família 43	Doc_2138	194	21063,9	23950,0
Cthe_0191	- Inibidor protease – Família serpina I4	Doc_0191	192	21083,9	18450,0
Cthe_0536	Beta-glicosidase – Família 5	Doc_0536	189	20363,1	16960,0
Cthe_2147	Beta-glicosidase – Família 5	Doc_2147	193	20870,6	16960,0
Cthe_0274	Beta-glicosidase – Família 9	Doc_0274	198	21662,6	19940,0
Cthe_2360	Beta-glicosidase – Família 9	Doc_2360	191	20977,9	16960,0
Cthe_0405	Beta-glicosidase – Família 5	Doc_0405	188	20403,2	19480,0
Cthe_0578	Beta-glicosidase – Família 9	Doc_0578	198	21619,6	21430,0
Cthe_0190	Inibidor protease – Família serpina I4	Doc_0190	190	20768,6	18450,0

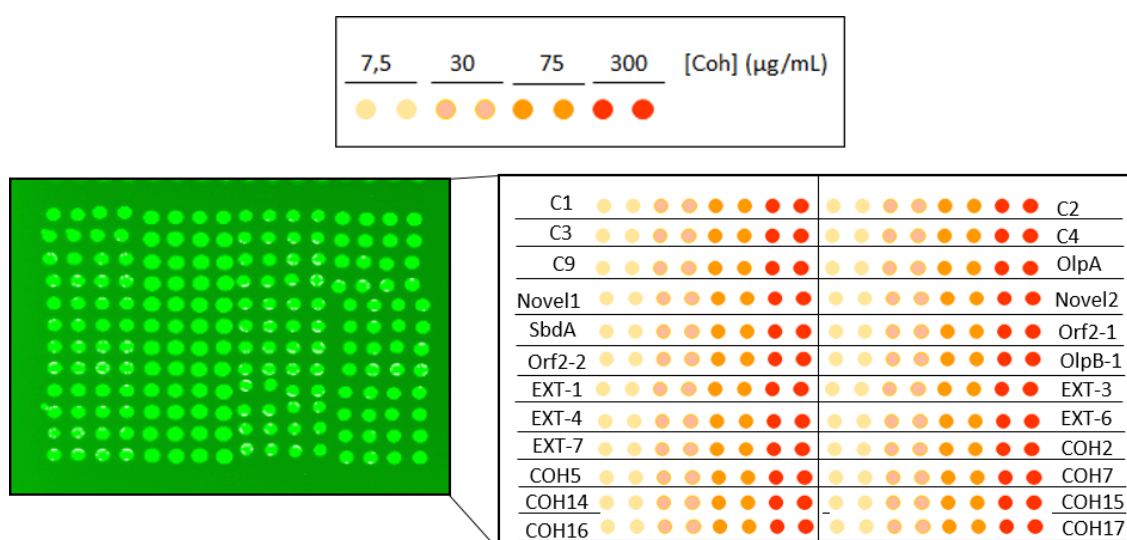
**Tabela 7B.3** - Informação sobre os módulos coesina em estudo do *R. flavefaciens*. A tabela apresenta o número de resíduos de aminoácidos presentes na sequência o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta ExPASy.

Proteína	Domínio a que a proteína está associado	Número de aa	PM (Da(gmol <sup>-1</sup> ))	$\epsilon$ (M <sup>-1</sup> .cm <sup>-1</sup> )
COH 2	ScaA (WP_009986657.1)	165	17639,0	6020,0
COH 5	ScaB (WP_009986658.1)	165	17643,0	5960,0
COH 7	ScaB (WP_009986658.1)	163	16956,0	18510,0
COH13	ScaE (WP_009986662.1)	221	23556,0	32890,0
COH14	ScaA (WP_09984551.1)	188	20119,0	26990,0
COH15	s/i	217	23501,0	21560,0
COH16	ScaC (WP_009986682.1)	236	26468,0	31920,0
COH17	WP_009988208.1	187	19907,0	21490,0

**Tabela 7B.4** - Informação sobre os módulos doquerina em estudo do *R. flavefaciens*. A tabela apresenta o número de resíduos de aminoácidos presentes na sequência, o peso molecular (PM) e o coeficiente de extinção molecular da mesma ( $\epsilon$ ). Estes parâmetros foram determinados pela ferramenta ExPASy.

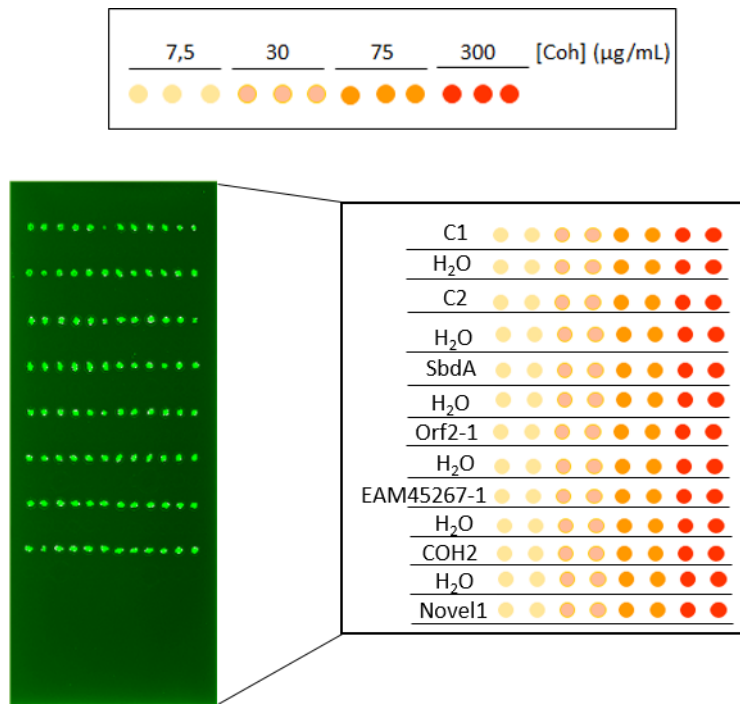
Proteína	Número de aa	PM (Da (gmol <sup>-1</sup> ))	$\epsilon$ (M <sup>-1</sup> .cm <sup>-1</sup> )
D05	220	23781	20190
D08	213	22939,8	22585
D12	214	23078,1	20190
D22	202	21980	15595
D33	220	23575,7	28210
D239	210	22710,8	21095
D47	212	23023,2	20065

## 7B.II – Padrão de impressão dos módulos coesina realizado roboticamente

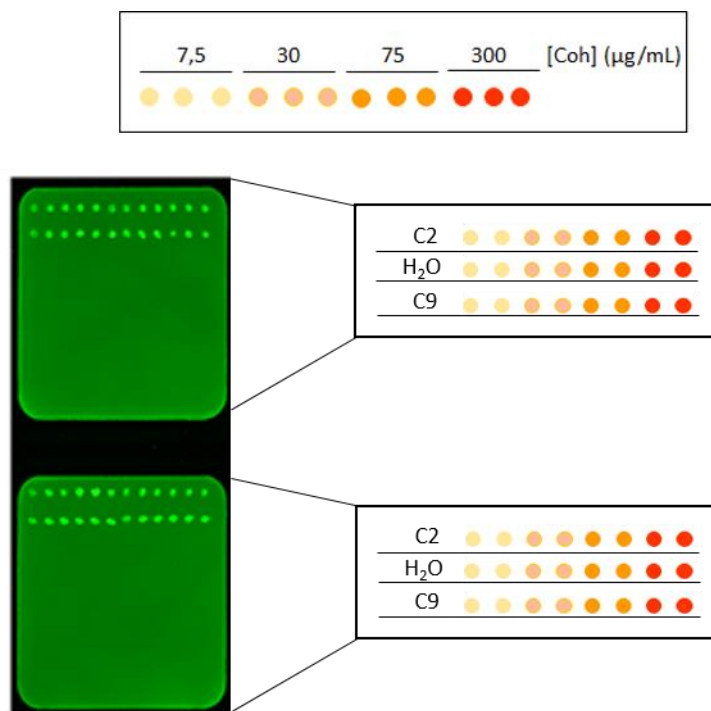


**Figura 7B.1** – Representação esquemática da impressão dos slides de 16 blocos dos módulos coesina. Cada módulo foi impresso em quatro concentrações diferentes (7,5, 30, 75 e 300 µg/mL).

**7B.III – Padrão de impressão dos módulos coesina realizado manualmente**



**Figura 7B.2** - Representação esquemática da impressão dos slides de 1 *bloco* dos módulos coesina. Cada módulo foi impresso em quatro concentrações diferentes (7,5, 30, 75 e 300 µg/mL).

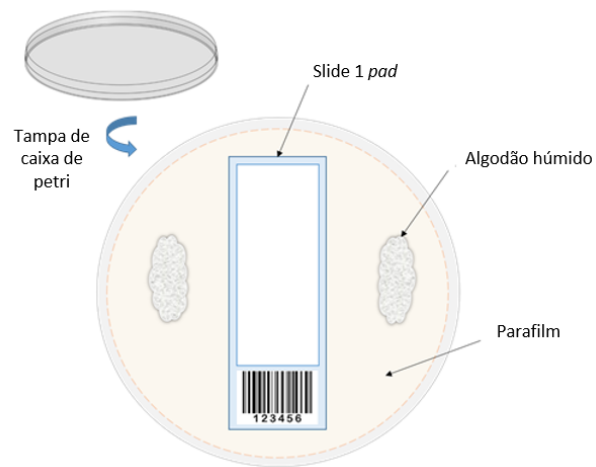


**Figura 7B.3** - Representação esquemática da impressão dos slides de 2 *bloco*s dos módulos coesina C2 e C9. Cada módulo foi impresso em quatro concentrações diferentes (7,5, 30, 75 e 300 µg/mL).

#### 7B.IV – Protocolo de Microarrays

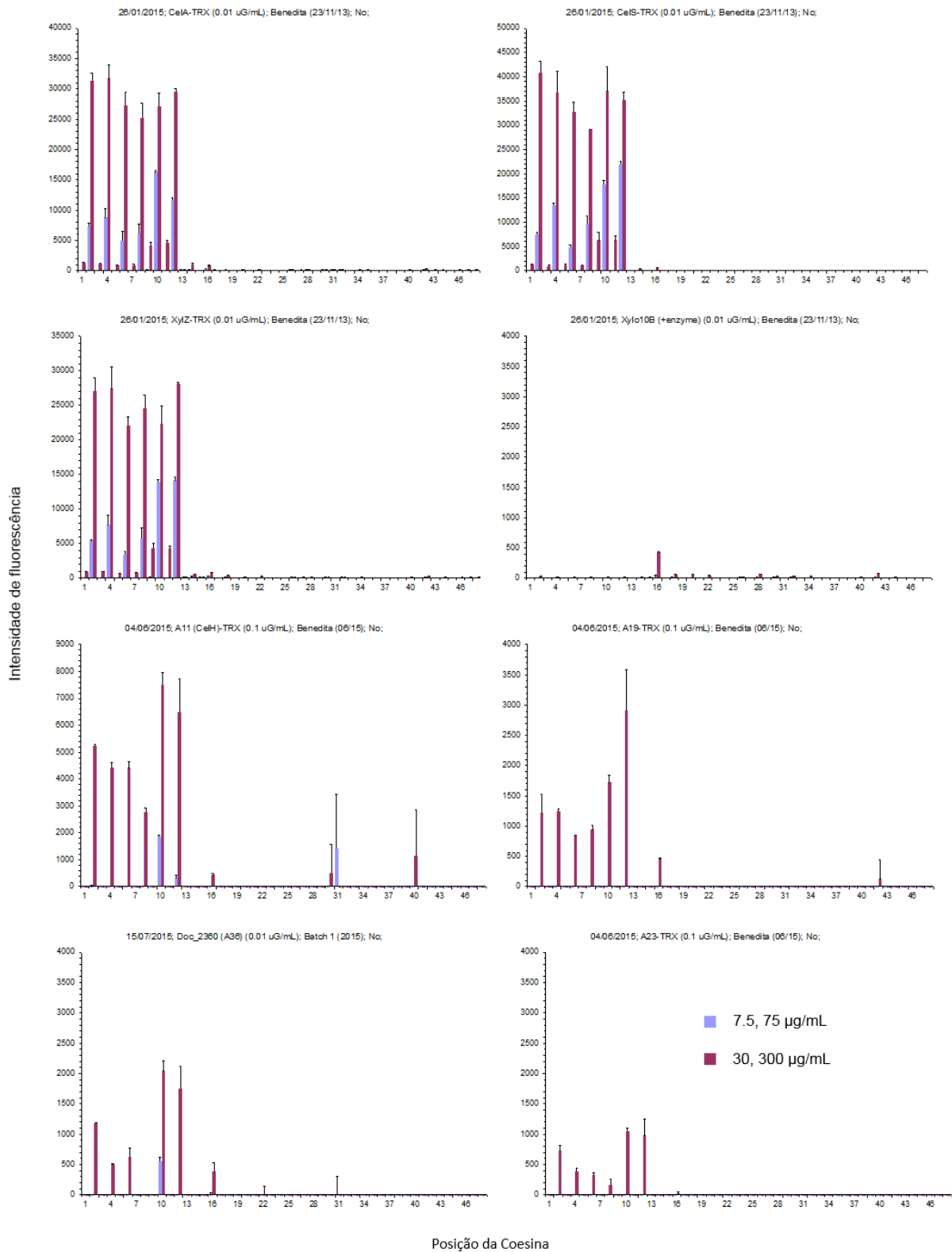
- 1) Digitalizar o slide para a fluorescência do fluoróforo Cy3 ( $\lambda_{\text{excitação}}=532\text{nm}$ );
- 2) No caso de utilizarmos slides de 1 ou 2 blocos, colocar o mesmo numa tampa de caixa de *petri*, de acordo com a Figura 7B.4. No caso de utilizarmos slides de 16 blocos, colocar o mesmo no suporte Fast QQ, com o respetivo *gasket*;
- 3) Lavar cada bloco com água Mili-Q. Remover o excesso com a micropipeta, tendo cuidado para não tocar a nitrocelulose;
- 4) Adicionar 1mL (slides de 1 ou 2 *blocos*) ou 150  $\mu\text{L}$  (slides de 16 *blocos*) de solução de bloqueio (3% BSA, 1% de Caseína). Cobrir com papel de alumínio e incubar durante 1 hora à temperatura ambiente;
- 5) Remover a solução de bloqueio e lavar o slide com água Mili-Q, segundo o passo 3);
- 6) Adicionar 1mL (slides de 1 ou 2 *blocos*) ou 90 $\mu\text{L}$  (slides de 16 *blocos*) do módulo doquerina (0,1 ou 0,01 $\mu\text{M}$ ) no *bloco* correspondente, preparado na solução de diluição (0,1% BSA, 0,1% Caseína em tampão HBS (5 mM Hepes pH 7.4; 150 mM NaCl, CaCl<sub>2</sub> 5 mM). Cobrir e incubar durante 1 hora à temperatura ambiente;
- 7) Remover o excesso de cada módulo doquerina e lavar cada bloco com 1mL (slides de 1 ou 2 *blocos*) ou 150 $\mu\text{L}$  (slides de 16 *blocos*) de tampão HBS. Repetir esta lavagem 4 vezes;
- 8) Adicionar 1mL (slides de 1 ou 2 *blocos*) ou 90 $\mu\text{L}$  (slides de 16 *blocos*) da anticorpo primário (3 $\mu\text{g}/\text{mL}$ ) em cada *bloco*, preparado na solução de diluição. Cobrir e incubar durante 1 hora à temperatura ambiente;
- 9) Remover o excesso de anticorpo primário (e lavar cada bloco com 1mL (slides de 1 ou 2 *blocos*) ou 150 $\mu\text{L}$  (slides de 16 *blocos*) de tampão HBS. Repetir esta lavagem 4 vezes;
- 10) Adicionar 1mL (slides de 1 ou 2 *blocos*) ou 90 $\mu\text{L}$  (slides de 16 *blocos*) da anticorpo secundário (1 $\mu\text{g}/\text{mL}$ ) em cada *bloco*, preparado na solução de diluição. Cobrir e incubar durante 1 hora à temperatura ambiente;
- 11) Remover o excesso de anticorpo secundário e lavar cada bloco com 1mL (slides de 1 ou 2 *blocos*) ou 150 $\mu\text{L}$  (slides de 16 *blocos*) de tampão HBS. Repetir esta lavagem 4 vezes;
- 12) Adicionar 1mL (slides de 1 ou 2 *blocos*) ou 90 $\mu\text{L}$  (slides de 16 *blocos*) de estreptavidina (1 $\mu\text{g}/\text{mL}$ ) em cada *bloco*, preparado na solução de diluição. Cobrir e incubar durante 45 minutos à temperatura ambiente;
- 13) Remover o excesso de estreptavidina e lavar cada bloco com 1mL (slides de 1 ou 2 *blocos*) ou 150 $\mu\text{L}$  (slides de 16 *blocos*) de tampão HBS. Repetir esta lavagem 3 vezes. Lavar com água Mili-Q (2vezes);

14) Retirar o slides na caixa de petri ou do suporte, conforme slide utilizado, e secar o mesmo com secador. Após estar seco, digitalizar segundo a fluorescência do fluoróforo Alexa 647 ( $\lambda_{excitação}=635nm$ ).

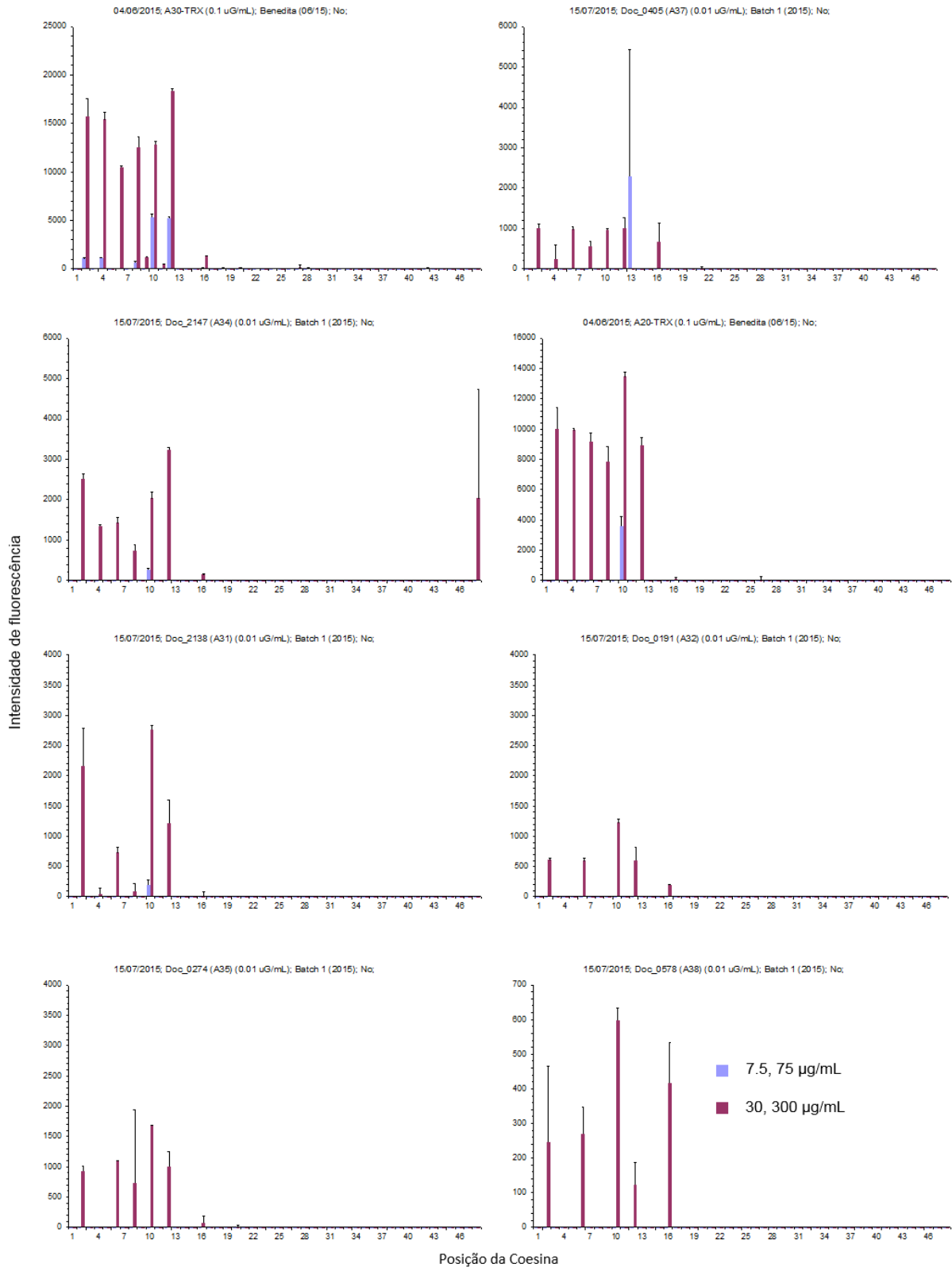


**Figura 7B.4** - Representação esquemática da colocação dos slides de 1 ou 2 blocos na tampa de caixa de petri. O uso de algodão húmido proporciona um ambiente húmido, evitando assim a secagem da matriz de nitrocelulose. O parafilm promove a estabilização dos slides durante todo o ensaio.

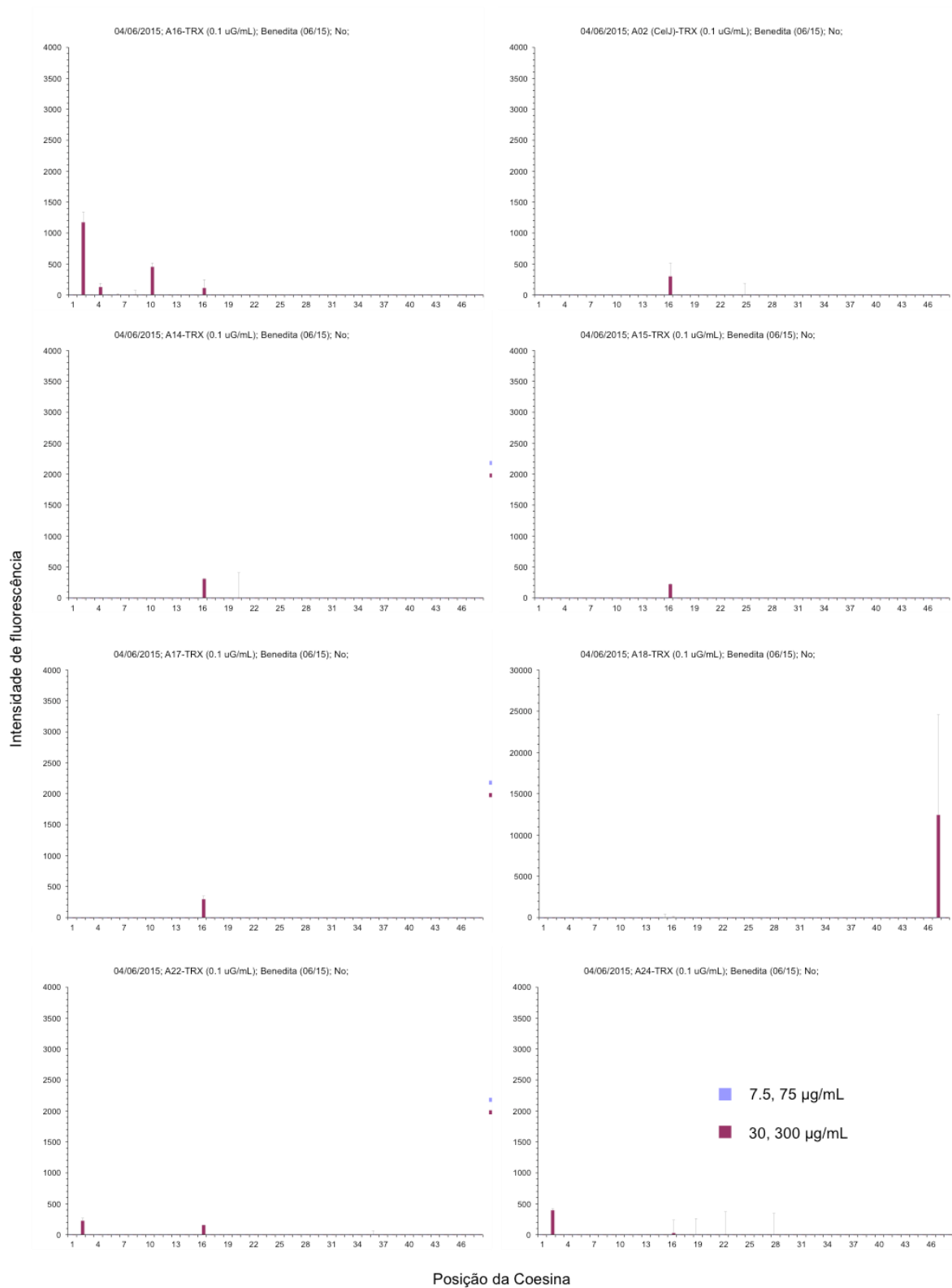
## 7B.V – Gráficos – Intensidades de fluorescência em função da posição do módulo coesina



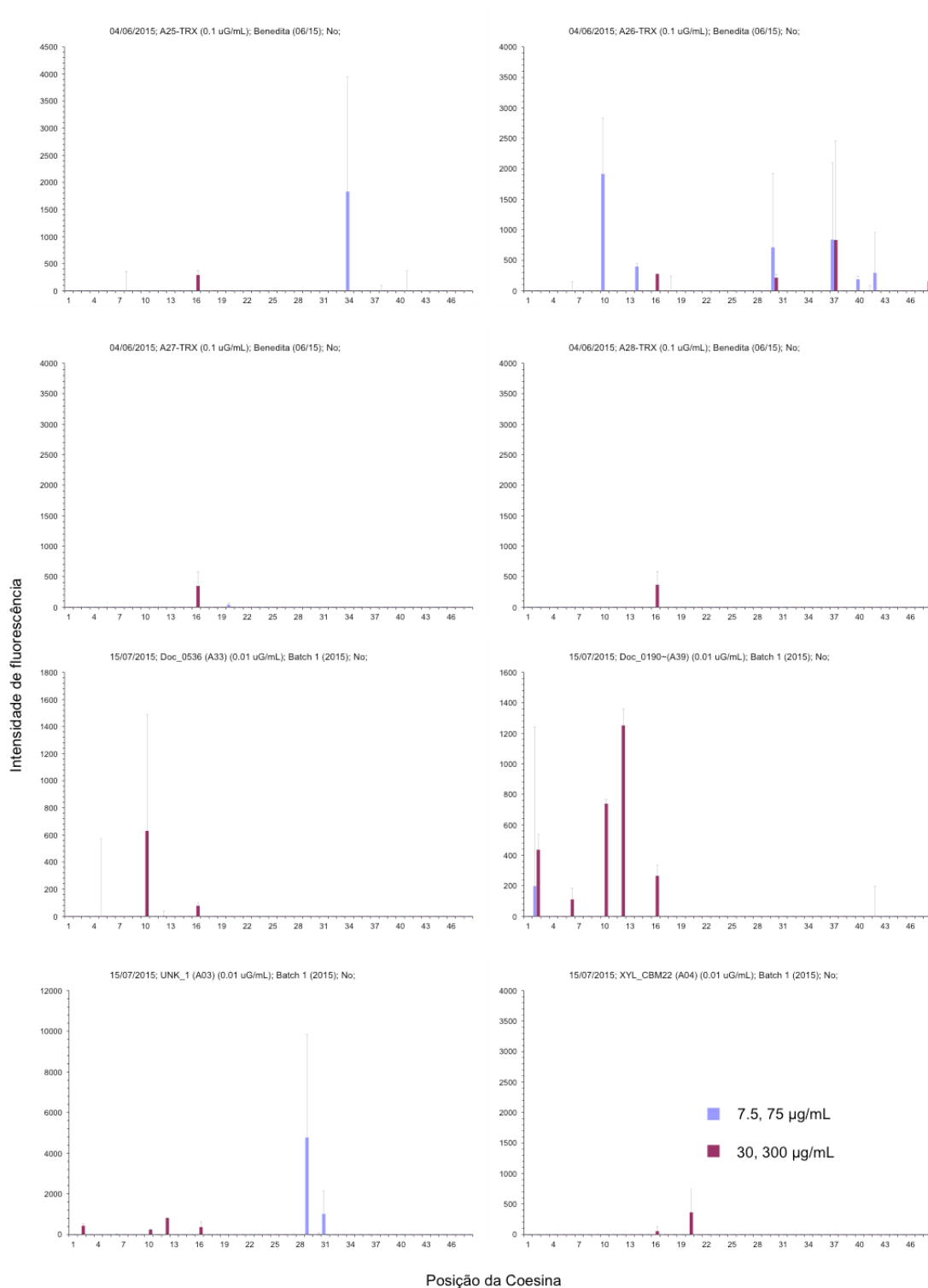
**Figura 7B.5** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente.



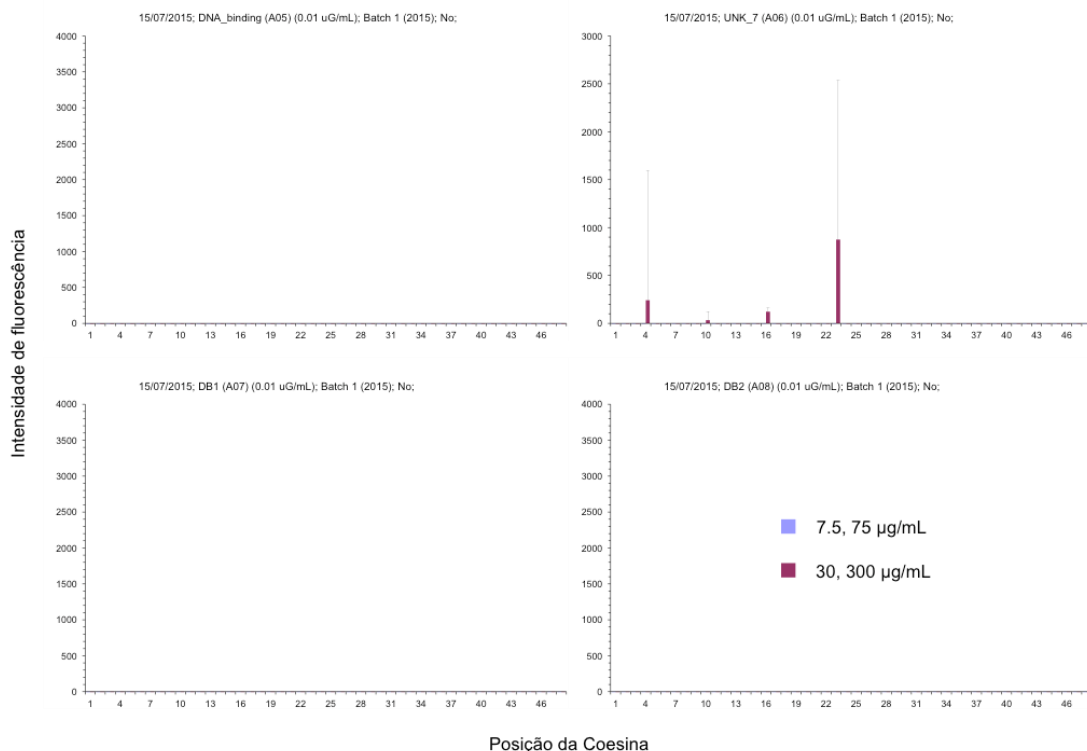
**Figura 7B.6** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente.



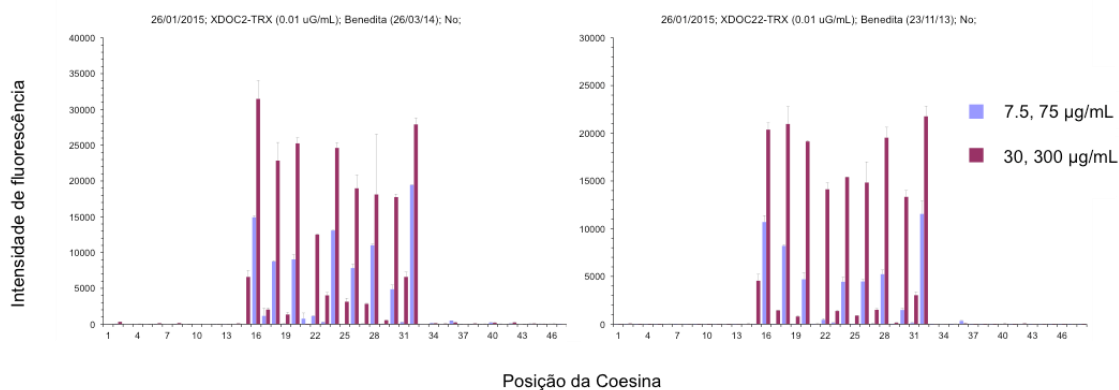
**Figura 7B.7** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarray*s correspondentes a uma concentração diferente.



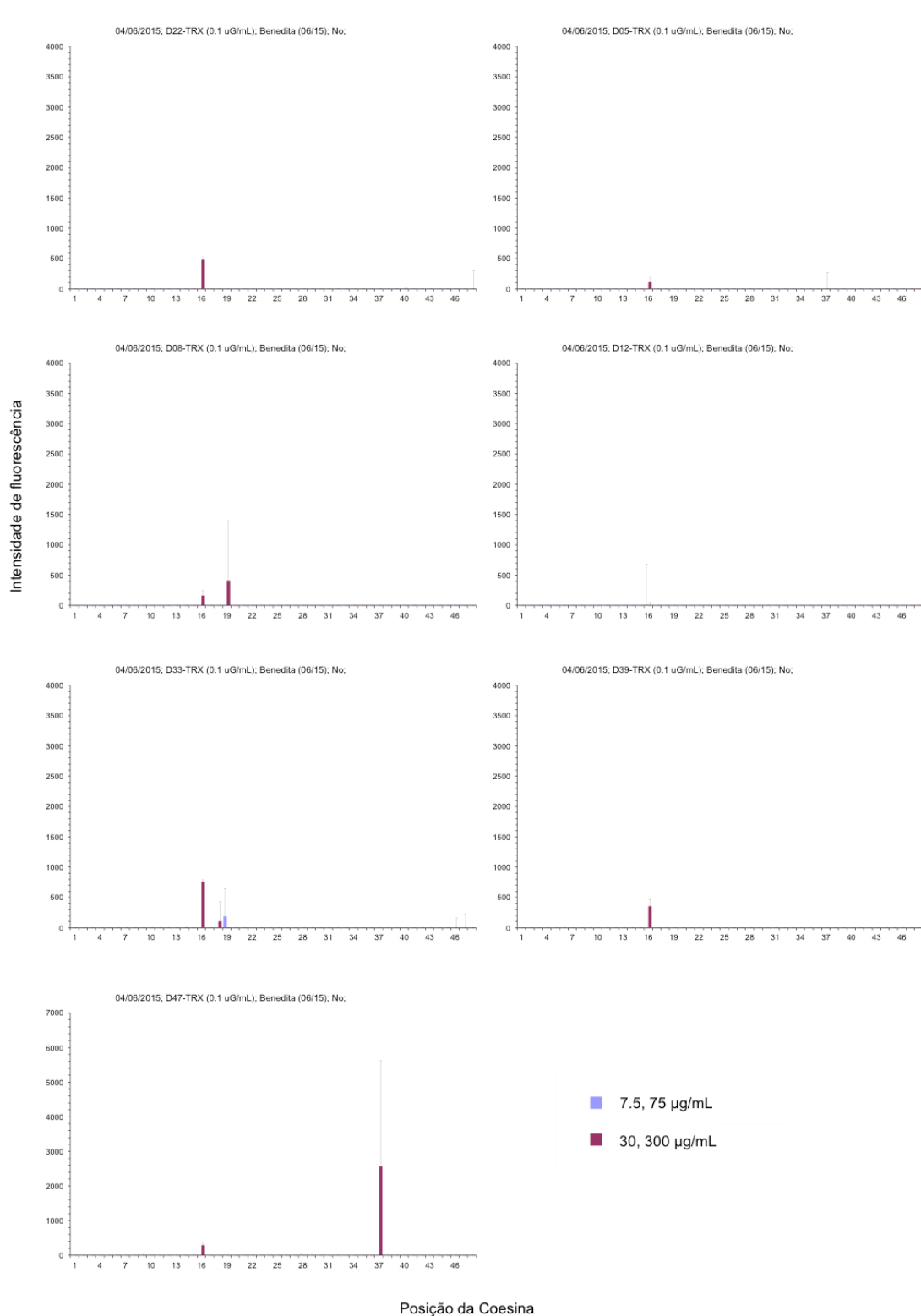
**Figura 7B.8** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente.



**Figura 7B.9** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo I do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente.



**Figura 7B.10** - Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do tipo II do *C. thermocellum* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente.



**Figura 7B.11-** Gráficos das intensidades de fluorescência obtidas após o ensaio com várias doquerinas do *R. Flavefaciens* no *microarray* de coesinas. Cada gráfico corresponde a uma doquerina e cada barra a uma posição no *microarray*. Uma coesina tem 4 posições no *microarrays* correspondentes a uma concentração diferente.