

Acessando a regulação transcricional da formação de biofilme em *Escherichia coli* utilizando bibliotecas genômica e de promotores

Palavras-chave: 1. Biofilme. 2. *Escherichia coli*. 3. Promotores 4. Regulação da expressão gênica 5. SortSeq

Resumo

Biofilmes são estruturas complexas formadas por comunidades bacterianas da mesma ou de espécies diferentes, embebidas em uma matriz extracelular composta por substâncias extracelulares poliméricas (EPS), como polissacarídeos, proteínas (como estruturas curli) e ácidos nucleicos. As redes reguladoras da formação do biofilme são compostas pela integração de estímulos ambientais e intracelulares. Uma de suas camadas é a regulação transcricional, que é capaz de alterar drasticamente a expressão gênica bacteriana, convertendo-a de um estilo de vida livre para um comportamento sésil. Para entender melhor como os genes relacionados ao biofilme são regulados por diferentes fatores de transcrição, decidimos aplicar a técnica de *SortSeq* para investigar a fundo a arquitetura de promotores centrais relacionados à formação de biofilme. Iniciamos este trabalho com 21 promotores, associados diretamente ao desenvolvimento do biofilme e motilidade bacteriana. Construímos e testamos uma biblioteca contendo esses 21 promotores mutados aleatoriamente a uma taxa de 10%. Os testes preliminares mostraram que esta biblioteca possuía muitas mutações e truncamentos das sequências promotoras, e poucas variantes, inviabilizando a aquisição de resultados com ela. Desta forma, decidimos focar em dois principais promotores de genes relacionados à síntese de *curli*: *csgB_{Ap}* e *csgD_p*. Seus promotores estão restritos a uma mesma região intergênica e possuem uma relação estrita, uma vez que CsgD é um regulador mestre da transcrição do operon *csgBA*. Dividimos essa região intergênica em quatro partes e construímos bibliotecas com mutações focalizadas e aleatórias. Essas quatro bibliotecas foram inseridas individualmente nos promotores originais, os quais já haviam sido clonados em um plasmídeo de baixo número de cópias e modulam a expressão dos genes repórteres *sfgfp* e *mCherry* (plasmídeo *csgBD*-repórter). Para acessar como as mutações do promotor podem afetar a expressão de cada gene, transformamos as quatro bibliotecas na linhagem de *E. coli* W3110 RpoS⁺ e as cultivamos até a fase estacionária. Reproduzindo a técnica *SortSeq*, as culturas foram analisadas

por citometria de fluxo e classificadas em quatro tubos de acordo com sua fluorescência para sfGFP, mCherry, sfGFP e mCherry, e negativo (sem expressão para ambos os repórteres). Triplicatas deste experimento tiveram seus plasmídeos extraídos e sequenciados usando a plataforma Illumina. Este novo design de biblioteca e abordagem SortSeq nos permitiu entender melhor como cada região promotora pode afetar a expressão de ambos os genes individual e conjuntamente, ajudando-nos a confirmar ou negar a atuação de sítios de ligação a fatores de transcrição já descritos. Também encontramos novos possíveis sítios de ligação para fatores de transcrição ainda não descritos, os quais foram utilizados para a predição de novos reguladores para esses promotores. Para expandir a investigação dos promotores *csgBA* e *csgD* construímos uma biblioteca genômica da cepa *E. coli* W3110 RpoS⁺, usando um plasmídeo contendo a transposase Tn5 e sequências de DNA como código de barras. A transposase Tn5 presente no plasmídeo insere aleatoriamente um código de barras único em uma região única do DNA genômico. Em seguida, mapeamos o código de barras para sua região genômica correspondente, permitindo a identificação do sítio mutado apenas pelo sequenciamento do código de barras. Transformamos nosso plasmídeo *csgBD*-reporter (sem qualquer inserção de biblioteca) na biblioteca e usamos a abordagem SortSeq para dividir as bactérias nos mesmos quatro fenótipos e sequenciar o código de barras inserido por meio da técnica NGS. Esta segunda abordagem levou à identificação de novos atores na regulação gênica desses genes tão importantes associados à formação do biofilme. Combinando esses dois tipos de bibliotecas, descobrimos possíveis novas regulações dos promotores de *csgBA* e *csgD* que serão testados e confirmados futuramente.

Accessing transcriptional regulation of *Escherichia coli* biofilm formation using promoter and genomic libraries

Key-words: 1. Biofilm. 2. *Escherichia coli*. 3. Promoters 4. Gene expression regulation 5. SortSeq

Abstract

Biofilms are complex structures formed by bacterial communities of the same or different species, embedded in an extracellular matrix composed of polymeric extracellular substances (EPS), such as polysaccharides, proteins (as *curli* structures), and nucleic acids. The regulatory networks of biofilm formation are composed of the integration of environmental and intracellular stimuli. One of its layers is transcriptional regulation, which is capable of drastically altering bacterial gene expression, converting it from a free-living style to a sessile behavior. To better understand how biofilm-related genes are regulated by different transcription factors, we aimed to apply the SortSeq technique to deeply investigate the architecture of central promoters of biofilm formation. We started this work with 21 promoters of interest, associated directly with biofilm formation and motility. A library containing these 21 promoters mutated randomly at a 10% rate was constructed and tested. The preliminary tests showed that this library had too many mutations and truncations per promoter sequence, and few promoter variants, making unfeasible the acquisition of results with it. This way, we decided to focus on two main promoters of genes for biofilm formation, related to *curli* synthesis: *csgBA_p*, and *csgD_p*. Their promoters are restricted to the same intergenic region between their genes and have a strict relation, once *csgD* is a master regulator of transcription of the *csgBA* operon. We divided this intergenic region into four parts and built libraries with region-focused random mutations. These four libraries were individually inserted into the original promoters, which were assembled in a low-copy number plasmid and modulate the expression of *sfGFP* and *mCherry* reporter genes (*csgBD*-reporter plasmid). To access how the promoter mutations could affect the expression of each gene, we transformed the four libraries into the *E. coli* W3110 RpoS⁺ and cultivated it to the stationary phase. Reproducing the SortSeq technique, the cultures were analyzed by flow cytometry and sorted into four tubes according to their fluorescence for sfGFP, mCherry, sfGFP, and mCherry, and Negative (no expression for both reporters). Triplicates of this experiment were mini-prepped and deep-sequenced using

Illumina platform. This new library design and SortSeq approach allowed us to better understand how each promoter region can affect the expression of both genes individually and at the same time, helping us to confirm or deny binding sites already predicted. We also found new possible binding sites for transcription factors not described yet, which were used for the prediction of new regulators to these promoters. This library gave us some insights into the logic involved in *csgB* bistability regulation. To expand the investigation of *csgBA* and *csgD* promoters we constructed a genomic library of the *E. coli* W3110 RpoS⁺ strain, using a barcoded-Tn5 plasmid. The Tn5 transposase present in the plasmid inserts randomly a unique barcode in a unique region of the genomic DNA. Then, we map the barcode to its corresponding genomic region, allowing the mutated site identification only by sequencing the barcode. We transformed our *csgBD*-reporter plasmid (without any library insertion) into it and used the SortSeq approach to sort the bacteria in the same four phenotypes and sequence the inserted barcode through the NGS technique. This second approach led to the identification of new players in the gene regulation of these important genes associated with biofilm formation. Combining these two types of libraries, we had new insights about the *csgBA* and *csgD* promoters' regulation that will be tested and confirmed in the future.