УНИВЕРЗИТЕТ У БЕОГРАДУ

Јелена Н. Гузина

БИОИНФОРМАТИЧКА АНАЛИЗА МЕХАНИЗАМА ТРАНСКРИПЦИОНЕ ИНИЦИЈАЦИЈЕ КОД БАКТЕРИЈСКИХ ECF σ ФАКТОРА

докторска дисертација

Београд, 2017.

UNIVERSITY OF BELGRADE

Jelena N. Guzina

BIOINFORMATICS ANALYSIS OF TRANSCRIPTION INITIATION MECHANISMS IN THE GROUP OF BACTERAL ECF σ FACTORS

Doctoral Dissertation

Belgrade, 2017.

ПОДАЦИ О МЕНТОРИМА:

др Марко Ђорђевић, ментор
ванредни професор Биолошког факултета, Универзитета у Београду

др Магдалена Ђорђевић, коментор
научни саветник Института за физику, Универзитета у Београду

ПОДАЦИ О ЧЛАНОВИМА КОМИСИЈЕ:

др Мирослав Живић, ванредни професор Биолошког факултета Универзитета у Београду

др Гордана Павловић-Лажетић
редовни професор Математичког факултета, Универзитета у Београду

др Слађана Спасић, виши научни сарадник Института за мултидисциплинарна истраживања Универзитета у Београду

ДАТУМ И ВРЕМЕ ОДБРАНЕ: _____

Захвалница

Ова докторска дисертација урађена је на Катедри за општу физиологију и биофизику, Института за физиологију и биохемију на Биолошком факултету, у оквиру пројеката ***"Биоинформатичке предикције промотера и теоријско моделовање генских кола код бактерија"*** *(ОИ173052) и* ***"Bioinformatics and modeling of bacterial immune systems - understanding control of CRISPR/Cas"*** *(SCOPES project IZ73Z0_152297), под непосредним руководством проф. др Марка Ђорђевића.*

Ментору проф. др Марку Ђорђевићу и коментору др Магдалени Ђорђевић, дугујем највећу захвалност на указаном поверењу, стрпљивом и брижљиво осмишљеном руковођењу, на сталној доступности и спремности да разреше моју било коју недоумицу, несебично пружајући ми на располагање своје огромно знање и искуство. Захваљујем им на слободи која ми је у раду увек указивана, а највише на пренетом ентузијазму и љубави за бављење науком.

Члановима комисије, проф. др Гордани Павловић-Лажетић, проф. др Мирославу Живићу и др Слађани Спасић захваљујем на пажљивом читању дисертације и корисним смерницама, које су значајно допринеле побољшању овог рукописа.

Драгој колегиници Анђели Родић захвална сам на увек пријатној радној атмосфери и добром дружењу на заједничким путовањима.

Пријатељима захваљујем на разумевању, подршци и стрпљењу, што су увек уз мене да поделе радост сваког успеха.

Родитељима, Николи и Милени, сестри Наташи, Драгану, Јакову и Лидији, мом Стевану захваљујем на безрезервној подршци, разумевању и љубави. Њима је ова дисертација и посвећена.

БИОИНФОРМАТИЧКА АНАЛИЗА МЕХАНИЗАМА ТРАНСКРИПЦИОНЕ ИНИЦИЈАЦИЈЕ КОД БАКТЕРИЈСКИХ ECF σ ФАКТОРА

Алтернативни σ фактори су неопходни за преживљавање бактеријске ћелије у стресним условима, као и за пролазак кроз стадијуме морфолошке диференцијације. Међу њима, ECF σ фактори представљају најбројнију, најразноврснију, али и најмање изучену групу за коју се сматра да препознаје ригидне, добро конзервиране промоторе. Насупрот овоме, механизам "mix-and-match" у групи примарних σ фактора (са еквивалентном промоторском структуром) подразумева флексибилно препознавање промотора, чији се елементи међусобно надопуњују. Будући да је тренутна парадигма о функционисању ECF σ фактора резултат малог броја изучених чланова, главни циљ тезе је опсежна биоинформатичка анализа доступних ECF σ фактора.

Комбиновањем биоинформатичких метода са биофизички-заснованом анализом, екстензивно су проучавани (квалитативни и квантитативни) аспекти флексибилности у препознавању ECF промотора. Анализа фагних ECF σ фактора довела је до нове процедуре за предвиђање фагних промотора, као и до првих (квалитативних) примера механизма "mix-and-match" у групи ECF. Анализа бактеријских ECF представника довела је до препознавања "неканонских" интеракција са промоторским секвенцама, у којима учествују конзервирани елементи у спејсеру, као и протеински мотиви изван домена за интеракцију са ДНК (σ_2 and σ_4). Коначно, установљено је значајно функционално надопуњавање промоторских елемената у групи ECF ("mix-and-match") које је, супротно очекивањима, јаче него у промоторима примарних σ фактора (и упућује на другачији кинетички профил транскрипционе иницијације). Будући да су ECF најразличитији од примарних σ фактора у оквиру σ^{70} групе, добијени резултати указују да је "mix-and-match" вероватно заједнички механизам препознавања промотора у целокупној σ^{70} фамилији.

Кључне речи: ECF σ фактори, бактеријски/бактериофагни промотори, транскрипциона иницијација, фамилија σ^{70}, механизам "mix-and-match"

ОБЛАСТ: БИОФИЗИКА; **УЖА ОБЛАСТ:** БИОИНФОРМАТИКА

BIOINFORMATICS ANALYSIS OF TRANSCRIPTION INITIATION MECHANISMS IN THE GROUP OF BACTERAL ECF σ FACTORS

Alternative σ factors are indispensible in bacterial cell for coping with stress or development. ECFs are the most numerous and diverse, but the least studied group of alternative σ factors, which are assumed to interact with rigid, well conserved promoters. On the contrary, "mix-and-match" mechanism of housekeeping σ factors (with equivalent promoter organization), implies flexibility in promoter recognition, where elements complement each other strengths. Since the current ECF paradigm is supported by limited information on few (canonical) representatives, our goal was to validate it through a comprehensive computational analysis of all available ECF σs.

By combining bioinformatics methods with biophysics-based analysis, we extensively studied (qualitative and quantitative) aspects of flexibility in ECF promoter recognition. Analysis of phage outliers gave a novel procedure for detecting phage-specific promoters, and also the first (qualitative) example of mix-and-matching in ECF group. We also found examples of "non-canonical" interactions between ECFs and their promoters, exhibited by the conserved promoter spacer elements and the σ-motifs outside of the main σ_2 and σ_4 DNA-binding domains. Finally, we provided quantitative evidence of substantial promoter element complementation (mix-and-matching) in ECFs, which is unexpectedly much stronger (and points to a different kinetic profile of the initiation process) than in housekeeping promoters. As ECFs and housekeeping σ factors are the most divergent σ^{70} groups, this implies that "mix-and-match" may function as a common promoter recognition mechanism in the entire σ^{70} family.

Keywords: ECF σ factors, bacterial/bacteriophage promoters, transcription initiation, σ^{70} family, "mix-and-match" mechanism

САДРЖАЈ

1. Увод

Појам генске експресије односи се на све процесе укључене у синтезу функционалних протеинских (и РНК) молекула, чиме се декодира наследна информација садржана у ДНК, т.ј. остварује њен хоризонтални проток кроз ћелију/организам. Регулисани проток наследне информације, у чијој је основи диференцијална генска експресија, ћелији омогућава да у сваком тренутку део репертоара својих функција интегрише у ефикасан физиолошки одговор на променљиве срединске услове, који могу сигнализирати наступање одређене врсте стреса, прелазак у следећу развојну етапу итд. [1]. Управо зато за ћелију је од изузетног значаја да прецизно контролише када ће и који скуп гена активирати, при чему сваки од корака генске експресије може да функционише као контролна тачка за регулацију.

Код бактерија главна тачка за регулацију генске експресије је њен почетни корак – транскрипција [2], током које ензим полимераза РНК (РНКП) катализује формирање фосфодиестарских веза међу рибонуклеотидима у растућем ланцу РНК, на основу редоследа нуклеотида у матричном ланцу ДНК. Транскрипција се дели на три фазе: иницијацију, током које се врши избор гена за активацију и почетак синтезе молекула РНК; елонгацију, током које се целокупан ген са ДНК преписује у молекул РНК; и терминацију, током које се синтетисани ланац РНК ослобађа из комплекса од РНКП и ДНК. И док фазу елонгације и терминације РНКП може самостално да обавља, за иницијацију транскрипције неопходан је σ фактор, протеин малих димензија са којим РНКП ступа у холоензимски комплекс, а који обезбеђује избор специфичних, регулаторних секвенци ДНК на које ће се ензим регрутовати [3-4].

Секвенце на ДНК, за које се σ фактор специфично везује, а које се називају промотори, локализоване су непосредно узводно у односу на гене [5-6], што омогућава правилно дефинисање транскрипционих јединица, т.ј. експримирање целокупне информације кодиране извесним геном. Различити σ фактори имају различит специфицитет за промоторске секвенце, при чему су функционално повезани бактеријски гени (нпр. везани за одговор на одређени облик стреса) често транскрибовани истим σ фактором. Ово

омогућава ћелији да ефикасно и економично усклади транскрипциони профил са динамиком промена срединских/метаболичких услова, будући да од врсте σ фактора који је у датом тренутку асоциран са РНКП зависи који скуп гена у ћелији може бити активиран [7-8].

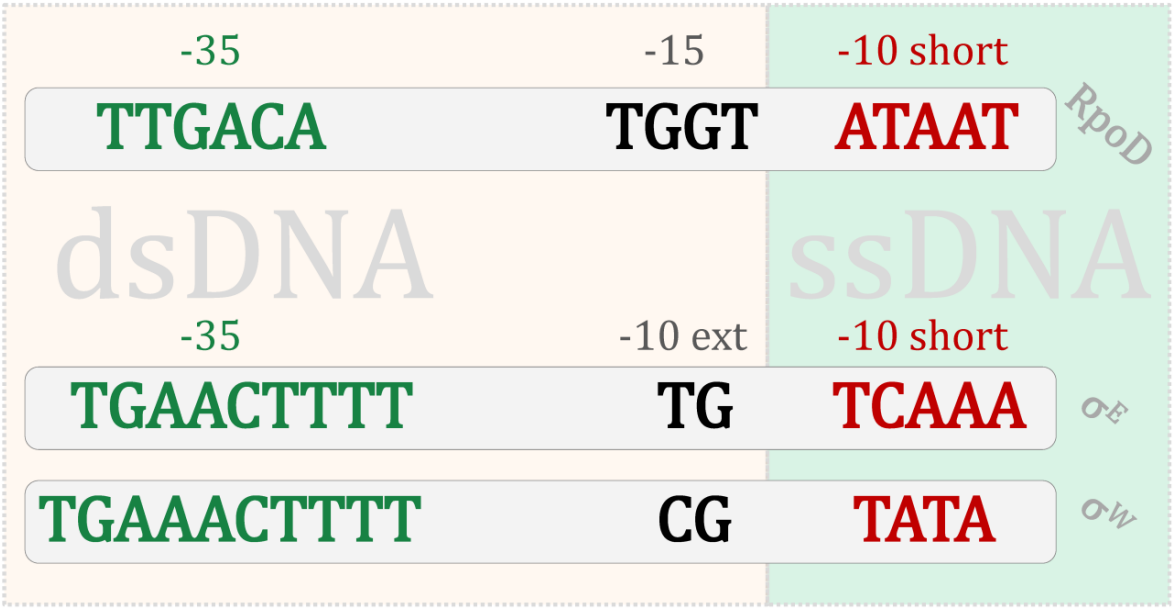
Сем регрутовања РНКП на промоторске секвенце, који представљају главну одредницу регулације генске експресије код бактерија [9], σ фактори остварују интеракције и са активаторским протеинима, учествују у раздвајању ланаца ДНК, стимулишу прелазак у фазу елонгације, а у извесним случајевима могу да модулишу и динамику елонгације [10-15]. Стога је изучавање механизма интеракције σ фактора са промоторским секвенцама током иницијације транскрипције незаобилазан корак за разумевање регулације генске експресије код бактерија.

1.1. σ^{70} фамилија – физиолошка улога и главне карактеристике

Већина познатих σ фактора припада σ^{70} фамилији која, сходно, заузима највећи удео у транскрипцији бактеријских гена [11]. Чланови σ^{70} фамилије препознају промоторе заједничке опште организације, чији су основни мотиви хексамери, -35 и -10 елемент, где називи одговарају типичној удаљености ових елемената у односу на место почетка транскрипције [16]. Сем елемената -10 и -35, σ^{70} промотори могу садржати и додатне одреднице специфицитета, попут -15, тј. продуженог -10 елемента [17-19]. Протеинске секвенце σ^{70} фактора имају модуларну структуру, т.ј. сегменти који интерагују са промоторским елементима и РНКП организовани су у хеликалне домене, којих највише може бити четири. Од тога, домени σ_2 и σ_4, који су присутни у целокупној фамилији [20], учествују у препознавању -10 и -35 елемента.

Последица препознавања промотора исте опште организације је да фактори σ^{70} фамилије иницирају транскрипцију кроз исте биофизичке кораке, који су описани истим кинетичким параметрима [21]. Први корак, који се назива формирање затвореног комплекса, јесте регрутовање холоензима РНКП на промотор реверзибилним везивањем σ фактора за дволанчану ДНК (dsDNA) (Слика 1). Енергија ове интеракције одређује афинитет везивања – K_B, биофизички параметар формирања затвореног комплекса. Наредни корак транскрипционе иницијације је формирање отвореног комплекса, т.ј.

раздвајање два ланца ДНК (тзв. топљење промотора), током којег σ фактор интерагује са једноланчаном ДНК (ssDNA). При топљењу промотора у нивоу -10 елемента иницира се транскрипциони мехур термалним флуктуацијама које локално раздвајају молекул ДНК, а које су помогнуте интеракцијом насталих једноланчаних сегмената ДНК са σ фактором [22]. Овај корак транскрипционе иницијације биофизички се описује стопом преласка затвореног у отворени комплекс – k_f, која зависи од енергије интеракције σ фактора са промоторским једноланчаним сегментом.



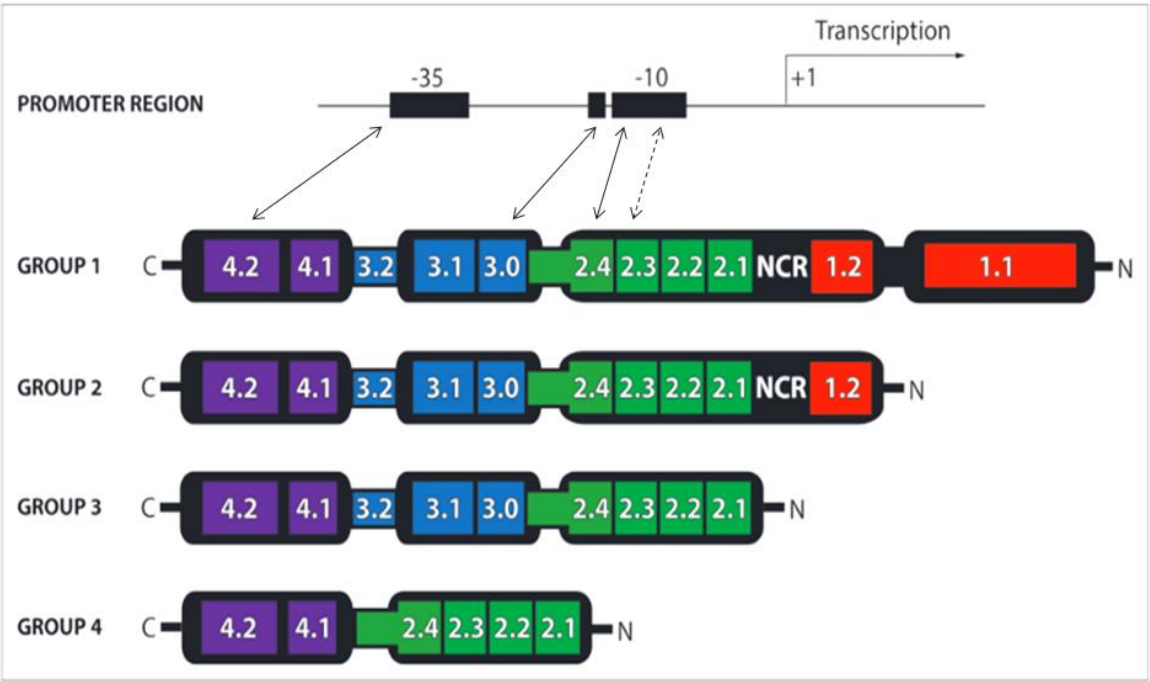
Слика 1. Структура промотора у фамилији σ^{70}: На слици је приказана организација промотора које препознају σ фактори из група RpoD (RpoD из *E. coli* - горњи део слике) и ECF (σ^E из *E. coli* и σ^W из *B. subtilis* - доњи део слике). Промоторски елементи који са σ фактором интерагују као dsDNA приказани су у зеленој/црној боји и осенчени розе правоугаоником. Промоторски елементи који са σ фактором интерагују као ssDNA приказани су у црвеној боји и осенчени зеленим правоугаоником.

Према улози у бактеријској физиологији, σ^{70} фактори деле се на примарне и алтернативне, при чему су први неопходни за функционисање ћелије под оптималним условима, док други покрећу специјализоване одговоре на различите стресне сигнале, могу да учествују у вируленцији код патогених врста, као и прогресији кроз различите етапе морфолошке диференцијације ћелије [23-25]. Од особености животног циклуса бактерије зависи расподела транскрипционог простора између примарног и алтернативних σ фактора, као и број алтернативних σ фактора који ће бактеријски геном кодирати [20]. Тако геном бактерије *Streptomyces coelicolor*, код које прилагођавање на врло варијабилне услове животне средине захтева сложене физиолошке одговоре, поред примарног, кодира и преко

шездесет алтернативних σ фактора. Са друге стране, геном *Streptococcus pyogenes*, бактерије окружене униформним миљеом епитела хуманог респираторног тракта, кодира свега два алтернативна σ фактора [9, 11, 23].

Припадници σ^{70} фамилије јасно су и структурно издиференцирани, што је, уз пратеће функционалне разлике, послужило као критетријум за даљу класификацију на четири субфамилије (Групе I до IV) [7]. Група I, која се још назива и RpoD, обухвата примарне σ^{70} факторе, који су најбројнији и структурно најкомплекснији припадници фамилије [11, 23]. RpoD σ фактори регулишу експресију највећег броја бактеријских, тзв. housekeeping гена, од чије активности зависи функционисање ћелије у експоненцијалној фази раста, т.ј. контрола базалне транскрипције. Протеинске секвенце RpoD σ фактора, које су организоване у четири домена, препознају промоторе са високо варијабилним елементима, т.ј. широким спектром јачина. Наиме, транскрипционе активности RpoD промотора мерене у условима in-vitro разликују се и до три реда величине, док у условима in-vivo ова разлика расте на чак пет редова величине [3, 20].

Варијабилност RpoD промотора не огледа се само у дивергенцији елемената у односу на њима одговарајуће консензусе, већ и у могућности за функционалну комбинаторику различитих елемената при иницијацији транскрипције [26]. Најпознатији пример ове флексибилности у RpoD групи је могућност постизања промоторске активности у одсуству -35 елемента, које се компензује интеракцијама са јаким продуженим -10 елементом [7, 19]. Комбиновано, наведени нивои варијабилности обезбеђују широку плејотропију σ факторима RpoD групе, која уз учешће додатних регулатора омогућава прецизну контролу експресије RpoD регулона састављених од неколико хиљада гена [7], тј. ефикасну координацију транскрипционе активности у експоненцијалној фази раста бактерија.



Слика 2. Доменска организација протеинских секвенци у σ^{70} фамилији. На слици је приказана доменска организација за 4 групе σ^{70} од најкомплексније ка најједноставнијој (горе ка доле), при чему је на самом горњем делу слике назначена интеракција са одговарајућим елементима промотора за приказане домене; слика преузета и адаптирана из [9].

Преостале σ^{70} субфамилије састоје се од алтернативних σ фактора, чија структурна комплексност опада од Групе II ка Групи IV (Слика 2). σ фактори Групе II по доменској организацији, али и примарној структури делова секвенце који интерагују са ДНК, врло су слични RpoD σ факторима, што се огледа у готово идентичном специфицитету за промоторе код ове две субфамилије [7, 27]. За разлику од RpoD σ фактора који контролишу транскрипцију под оптималним условима, σ фактори Групе II активни су током стационарне фазе раста, при чему често покрећу алтернативне метаболичке програме као и генерализовани одговор на стрес, експресијом регулона величине ~ 500 гена [23, 28-29].

Представници Група III и IV су по структурним карактеристикама и начину функционисања знатно удаљенији од примарних RpoD σ фактора те се ови σ фактори могу сматрати алтернативним у најужем смислу речи [23, 30]. За њих је карактеристично да, активирани специфичним сигналима, покрећу уско усмерене физиолошке реакције, од којих су најучесталији механизми за превазилажење различитих стресних услова [7]. Под таквим околностима, за ћелију је императив да што брже оствари специфичан одговор, због чега регулони ових σ фактора ретко садрже преко стотину, а некада и свега неколико

гена. Упркос ограниченом броју гена чију експресију регулишу, алтернативни σ фактори могу да реализују и врло комплексне физиолошке процесе, интегрисањем активности већег броја индивидуалних σ фактора на хијерархијски вишем функционалном нивоу. Тако нпр. прецизно оркестрираним каскадама активације σ фактори Групе III (σ^E, σ^F, σ^G, σ^H, σ^K) регулишу процес формирања спора код бактерија из рода *Bacillus* [20]. Сем спорулације, σ фактори Групе III контролишу и биосинтезу флагела, одговор на топлотни стрес, а у Грам-позитивним бактеријама активира их и осмотски стрес, неповољан енергетски статус, присуство етанола, итд. [9].

Протеинске секвенце σ фактора Групе III састоје се од домена σ_2, σ_3 и σ_4, а специфицитет промотора које препознају разликује се у односу на групу RpoD [7]. Иако промотори Групе III такође садрже продужени -10 елемент који се препознаје доменом σ_3 на нивоу протеинксе секвенце, структура им је ригиднија у односу на RpoD промоторе, а елементи знатно више конзервирани, чиме је и одређен горњи лимит на величину контролисаног регулона.

За σ факторе Групе IV сматра се да препознају промоторе још једноставније организације будући да поседују само домене σ_2 и σ_4 за интеракцију са промоторским елементима -10 и -35 [30-31]. Ова, по доменској архитектури најједноставнија субфамилија, бројношћу знатно превазилази остале алтернативне σ факторе. Бројност чланова субфамилије одражава се у високој хетерогености њихових протеинских секвенци, на основу чега је извршена даља класификација на више од 40 подгрупа, при чему су извесне подгрупе широко распрострањене у геномима различитих бактерија, док је другима присуство ограничено на специфичне филогенетске групе [32].

У различитим подгрупама субфамилије IV промоторски специфицитет одређен је у највећој мери -10 елементом, што је условљено присуством варијабилне аминокиселинске петље у домену σ_2, која остварује специфичне интеракције са различитим позицијама у оквиру -10 елемента [33]. Насупрот овоме, кристалографска студија, у којој је изучавана интеракција домена σ_4 σ^E фактора *E. coli* са -35 промоторским елементом, указује да се овај елемент препознаје на основну локалне геометрије молекула ДНК [34]. При том, уочено је да је мотив "AAC" – главна одредница овог препознавања – распрострањен у промоторима већине анализираних представника Групе IV.

Још једна готово универзална карактеристика изучених представника Групе IV је да регулони које контролишу садрже ген датог σ фактора [35] па позитивна повратна спрега доводи до брзе и интензивне експресије контролисаног регулона, што је у складу са улогом коју σ фактори Групе IV имају у бактеријској физиологији [30, 36-38]. Наиме, активношћу ових σ фактора, бактерије одговарају на различите облике стреса које изазива: присуство тешких метала и других токсичних молекула, неправилно савијених протеина у мембрани, неповољан оксидативни статус, недостатак хранљивих материја, итд. [9, 39].

Сигнали који активирају σ факторе Групе IV најчешће потичу из ванћелијског простора, а спроводе се преко мембранских и компонената ћелијског зида, на основу чега им је додељен алтернативни назив – ECF (*E*xtra*C*ytoplasmic *F*unction) σ фактори [31]. Код патогених врста, компоненте ћелијског зида служе и као контактна површина за интеракцију са домаћином па тако ECF σ фактори, путем очувања интегритета ћелијског омотача, узимају учешће и у вируленцији [36]. Пример наведеног је ECF σ фактор *Pseudomonas aeruginosa*, AlgU, који на имунски одговор домаћина и примену антибиотика – узрочнике периплазматског стреса – покреће синтезу мукоидног омотача и тиме доводи до преласка бактерије у патогени фенотип, који изазива тешке плућне инфекције код имунокомпромитованих пацијената [38, 40]. Сличну улогу у патогенези ECF σ фактори остварују и код представника рода *Mycobacterium*, регулацијом синтезе молекула за адхеренцију на ћелије домаћина и раста у миљеу хуманих макрофага [41]. Стога, проучавање механизама функционисања ECF σ фактора, сем значаја за разумевање контроле бактеријске транскрипције у условима стреса, има и значајан потенцијал за практичне примене у медицини и фармакологији.

Попут представника Групе III, и активност више ECF σ фактора може бити интегрисана ради остваривања функција хијерархијски већег степена сложености. Истраживања на бактерији *Pseudomonas aeruginosa*, чији геном кодира двадесет пет алтернативних σ фактора (двадесет један из групе ECF), показала су да директно преклапање два различита регулона може да обухвати и до 20% припадајућих гена, за које је утврђено да имају улогу у сложеним физиолошким процесима попут хемотаксе, адхеренције, секреције протеина, итд. Такође, уочена је и појава индиректног преклапања активности различитих σ фактора, где обухваћени гени учествују у процесима централног метаболизма, чиме се есенцијални ћелијски процеси усклађују са динамиком промена у условима животне средине [35].

Функционално преклапање припадника групе ECF јавља се и код бактерија рода *Bacillus*, где омогућава да сигнали мембранског стреса активирају исти централни скуп гена, којим се остварају опште особине одбрамбеног одговора. Истовремено, специфичне одлике одговора у зависности су од конкретног сигнала који покреће одбрамбени механизам, тј. од активности непреклапајућих гена из регулона ECF σ фактора. Сумарно, уочени баланс редундантности и специфичности интеракција са промоторима обезбеђује високу пластичност одговора на мембрански стрес у роду *Bacillus*, који настањује врло хетерогене еколошке нише [37, 42].

Да би се утврдио опсег функционалног преклапања различитих σ фактора приликом остваривања комплексних физиолошких процеса, неопходно је познавање њиховог промоторског специфицитета [7]. Сем наведеног, значај утврђивања промоторског специфицитета је и у томе што представља најкраћи пут ка функционалној карактеризацији нових σ фактора, чији је велики број предвиђен захваљујући интензивној акумулацији доступних секвенци бактеријских генома. Ови нови σ фактори готово у целости припадају групи ECF која, сразмерно величини, има најмањи број изучених представника у фамилији σ^{70} што захтева озбиљну, систематску анализу интеракције са промоторским секвенцама.

Општије, анализа интеракције различитих σ фактора са промоторским секвенцама може указати и на који начин се структурне разлике у фамилији σ^{70} одражавају на функционалном нивоу, што уједно значи и преусмеравање истраживачког фокуса са добро проучених, примарних σ фактора на испитивање механизама транскрипционе иницијације код алтернативних σ фактора [23]. Преглед досадашњих сазнања о механизмима иницирања транскрипције изложен је у наредном поглављу како за примарне, тако и за алтернативне σ^{70} факторе.

1.2. Механизам транскрипционе иницијације у σ^{70} фамилији: примарни vs. алтернативни σ фактори

Централно питање у изучавању механизма транскрипционе иницијације је на који начин су специфичне секвенце промоторских елемената повезане са јачином транскрипционог одговора промотора, тј. како се кодирањем варијабилних елемената прецизно регулише кинетика иницијационог процеса. Одговор на ово питање пружа могућност да се за секвенцу било ког промоторског елемента прецизно одреди квантитативни удео у постигнутој промоторској активности. Природно, оваква истраживања увек почињу од интеракције између промоторских елемената и σ фактора, као главних одредница специфичности у интеракцији између холоензима РНКП и ДНК.

И поред врло израженог структурног и функционалног диверзитета σ фактора из фамилије σ^{70}, механизам транскрипционе иницијације детаљно је изучен само у групи примарних RpoD σ фактора [3, 7, 18-19]. Протеинске секвенце ових σ фактора најсложеније су структурне организације у фамилији, при чему сваки од четири присутна домена остварује специфичне интеракције са елементима ДНК у промоторском региону. Енергије ових интеракција могу бити међусобно снажно условљене, чиме се успоставља прецизна контрола над кинетиком иницијације, током које ензим РНКП и промоторска ДНК пролазе кроз сложене конформационе промене [3, 43]. И спољашњи регулатори (транскрипциони фактори) често чине део инфраструкутре за прецизно подешавање транскрипционог одговора RpoD промотора [44], који, као што је већ наглашено, опсегом може да варира и до 5 редова величине у условима in-vivo.

Основни кораци у којима учествују сви σ^{70} фактори током иницијације транскрипције су регрутовање ензима РНКП на ДНК и топљење низводног промоторског региона ради ослобађања матричног ланца за преписивање генетичке информације у молекул РНК. У регрутовању ензима РНКП на промотор посредује интеракција σ фактора са више промоторских елемената у дволанчаној форми, при чему се први контакт успоставља између домена $\sigma_{4.2}$ RpoD σ фактора (видети Слику 2) и -35 елемента [45]. Након препознавања -35 елемента долази до конформационе промене у молекулу ДНК, која омогућава региону узводно у односу на -35 елемент да оствари разгранату мрежу неспецифичних интеракција са површинским делом холоензимског комплекса и тиме

додатно стабилизује затворени комплекс у настанку. Интеракција -35:$\sigma_{4.2}$ има важну улогу у позиционирању домена $\sigma_{2.3}$ у односу на -10 елемент, који успостављањем специфичне интеракције појачавају афинитет везивања РНКП за промоторску ДНК у затвореном комплексу. Наиме, растојање између домена $\sigma_{4.2}$ и $\sigma_{2.3}$ и дужина спејсерске секвенце која раздваја елементе -35 и -10 су подударни, при чему чак и мање варијације дужине спејсера могу смањити стабилност затвореног комплекса, јер изазивају ротацију домена $\sigma_{4.2}$ и саме спејсерске секвенце [3]. При том, опсег ротације спејсера утиче и на стабилност низводних корака у инцијацији, јер од ње зависи утицај домена $\sigma_{1.1}$ на брзинску константу формирања и структуру/стабилност отвореног комлекса [46]. Постојање ове регулаторне везе сматра се главном "капијом" за спречавање непромоторске ДНК да оствари стабилан отворени комплекс и уметне у активно место ензима један од ланаца ДНК као матрицу за синтезу РНК молекула.

Иако је на примеру дужине спејсера очигледно да особине елемената укључених у почетне кораке транскрипционе иницијације могу испољити снажан регулаторни утицај на касније фазе процеса, линија разграничења између формирања затвореног и отвореног комплекса може се повући у нивоу -10 елемента. Наиме, -10 елемент, који је у условима in-vitro довољан за покретање транскрипционог одговора са промотора, најузводнијом базом на позицији -12 (најчешће је то T) учествује у формирању затвореног комплекса [47]. Непосредно узводно у односу на позицију -12 у неким промоторима налази се и продужени -10 (тј. такозвани -15) елемент, који преко интеракције са доменом σ_3 такође повећава афинитет везивања РНКП за дволанчану промоторску ДНК [18].

Након успостављања интеракција у затвореном комплексу дешава се нова конформациона промена промоторске ДНК, непосредно низводно у односу на позицију -12. На позицији -11 (где се у највећем броју RpoD промотора јавља база A) почиње савијање и топљење молекула ДНК, што је предуслов за смештање матричног ланца у активни канал ензима, тј. настанак отвореног комплекса. База A на позицији -11 (A_{-11}) избацује се из поретка наслојавања у дуплексу тако што упада у хидрофобни џеп домена $\sigma_{2.3}$, у ком остварује снажну интеракцију са присутним аминокиселинским остацима [47-48]. Увођењем тачке дестабилизације у дуплекс значајно се олакшава савијање низводне ДНК, док топљење почиње независно термалним флуктуацијама у низводном сегменту -10 елемента, а додатно се стабилизује интеракцијом σ фактора са базама комплементарног ланца у

једноланчаној форми. На стабилизацију насталог отвореног комплекса значајно утиче и база T_{-7}, која се такође изврће помоћу хидрофобног џепа домена $\sigma_{2.3.}$ Дакле, мрежа интеракција, која се између σ фактора и молекула ДНК у нивоу -10 елемента успоставља од узводног ка низводном сегменту, важна је одредница преласка из затвореног у отворени комплекс. У складу са тим, секвенца конкретног -10 елемента значајно утиче на кинетичке параметре који описују настанак оба комплекса.

σ фактор посредством домена $\sigma_{1.2}$ може остварити специфичне интеракције и са сегментом низводно од -10 елемента, који се назива дискриманторски регион [49]. Јачина ове интеракције утиче на време трајања отвореног комплекса, што је у директној вези са ефикасношћу иницијационог процеса. Наиме, дуже време трајања отвореног комплекса даје већи обрт циклуса абортивних синтеза, а самим тим и већу вероватноћу преласка у фазу елонгације. Стабилизација отвореног комплекса у нивоу дискриминаторског региона дешава се по истој матрици као у низводном сегменту -10 елемента, тј. кључни догађај представља извртање најбоље конзервиране базе у елементу од стране хидрофобног џепа домена $\sigma_{1.2}$ [3].

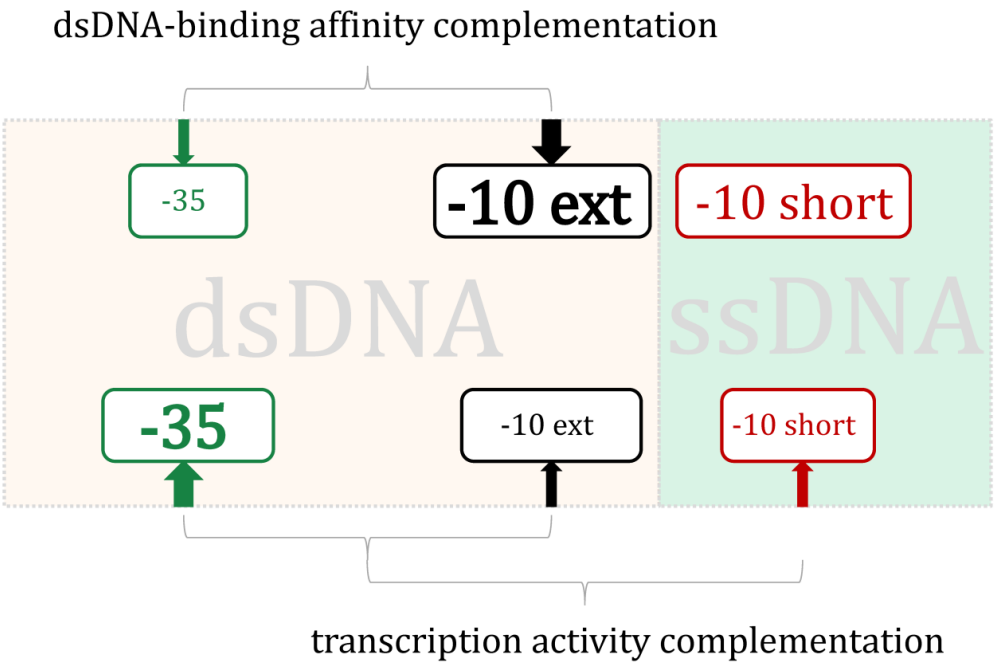
Специфичне интеракције са промоторском ДНК у мањем опсегу могу се успоставити и посредством одређених делова ензима РНКП. Најпознатији пример представља интеракција домена α-CTD са бипартитним UP елементом, који се у промоторима јавља у виду низова богатих нуклеотидима А и Т узводно од -35 елемента [50-51]. На извесним промоторима показано је да енергије интеракције ових најузводнијих сегмената (које укључују и неспецифичне интеракције са ензимом РНКП узводно од -35 елемента) значајно утичу на стабилност отвореног комплекса, што је још један пример регулаторног утицаја узводних, "раних" интеракција, у нисходним корацима иницијационог процеса. Додатно, специфични контакти између ензима РНКП и промотора могу припадати и једноланчаним интеракцијама, за шта је пример препознавање елемента CRE (енг. *C*ore *R*ecognition *E*lement), позиционираног готово симетрично око места почетка транскрипције, од стране ензимске субјединице β [52]. Ова интеракција такође утиче на стабилизацију отвореног комплекса, при чему постоје назнаке да испољава регулаторну улогу и током фазе елонгације [43, 53].

Конформационе промене кроз које промоторска ДНК пролази приликом иницирања транскрипције праћене су и опсежном просторном прерасподелом покретних модула

ензима РНКП, при чему је описано чак 5 региона "окидача" преко којих се иницира реструктуирање ензима [3]. Кинетика ових прерасподела у снажној је спрези са енергијама интеракције различитих промоторских елемената са σ фактором, при чему постаје очигледно да прецизно нивелисање активности експримираних гена мора почивати на комплексној регулаторној мрежи, где усходни кораци процеса могу утицати на динамику нисходних, чак и када на крупном плану нису повезани директним просторно/временским односима.

У врло рудиментарној форми описани феномен први пут је уочен квалитативном анализом RpoD промоторских секвенци [19]. Наиме, примећено је да између промоторских елемената који учествују у регрутовању РНКП на ДНК (тј. интерагују са σ фактором у форми dsDNA) постоји комплементација функционалне активности, чиме се постиже неопходна ефикасност првог корака у транскрипционој иницијацији. Прецизније, уочено је да промотори са јаким продуженим -10 елементом често имају слаб (или потпуно одсутан) -35 елемент, или дужину спејсера која није оптимална. Описано опажање формулисано је као механизам "mix-and-match", за који компензација одсуства -35 елемента продуженим -10 елементом представља екстреман (квалитативни) пример.

Систематска квантитативна анализа промоторских секвенци RpoD σ фактора из *E. coli*, која је заснована на биофизичком моделу транскрипционе иницијације, показала је да оригинални концепт механизма "mix-and-match" може да се прошири на кинетичке параметре процеса транскрипционе иницијације (Слика 3) [18]. Резултати анализе указали су да се најјача комплементација енергија интеракције са σ фактором дешава између једноланчаних и дволанчаних промоторских елемената, што је уједно подржано и резултатима биохемијских мерења [54-55]. Ово указује да је укупна транскрипциона активност, а не афинитет везивања РНКП за дволанчану промоторску ДНК, као што је најпре формулисано, главно обележје кинетике иницијационог процеса на RpoD промоторским секвенцама.



Слика 3. Механизам "mix-and-match" проширен на кинетичке параметре. Промоторски елементи који се функционално надопуњују деле се на дволанчане (dsDNA; осенчени розе на слици) и једноланчане (ssDNA; осенчени зелено на слици); У горњем делу слике приказана је комплементација снага дволанчаних елемената која води ка постизању довољног нивоа афинитета РНКП за дволанчане промоторске елементе, док је у доњем делу слике приказана комплементација између дволанчаних и једноланчаних елемената за постизање довољног нивоа укупне транскрипционе активности; снаге промоторских елемената означене су величином одговарајућег фонта и придружене стрелице.

Флексибилна промоторска структура, која укључује могућност надопуњавања енергија интеракције различитих промоторских елемената са σ фактором, тренутно се сматра искључивим својством групе примарних RpoD σ фактора, док је за алтернативне σ факторе на снази парадигма о ригидном промоторском специфицитету са добро конзервираним елементима [7, 47]. Нпр, у групи ECF σ фактора се сматра да је за промоторску активност обавезно присуство и -35 и -10 елемента, који при том исказују врло ограничену варијабилност. Такође, уобичајено је становиште да промотори ове групе немају продужени -10 елемент, будући да секвенце ECF σ фактора не поседују препознатљив домен σ_3 [23, 30]. При том, парадигма ригидног промоторског специфицитета тумачи се као засебан механизам препознавања промотора, који је у потпуној супротности са механизмом "mix-and-match", чврсто установљеним у групи примарних RpoD σ фактора.

Највероватнији разлог за *a priori* одбацивање механизма "mix-and-match" у групама σ^{70} фактора, који препознају промоторе са добро конзервираним елементима, је интуитивно

довођење у везу комплементације снага различитих промоторских елемената са степеном плејотропије σ фактора, будући да механизам "mix-and-match" експлоатише управо варијабилност промоторских елемената, који се функционално надопуњују. Стога се чак и за σ факторе Групе III, који имају продужене -10 елементе у промоторима, сматра да не испољавају механизам "mix-and-match", јер их одликује висока специфичност за промоторске секвенце које препознају [7]. За ECF σ факторе, који су високо специфични и у највећој мери дивергентни у односу на примарне RpoD σ факторе, управо на овој премиси (одсуство механизма "mix-and-match") заснивају се сва објашњења о карактеристичним својствима физиолошких одговора које дата група регулише.

Међутим, нејасно је са којом сигурношћу се може прихватити тренутна парадигма о препознавању промотора у ECF групи, будући да је заснована на врло ограниченом скупу σ фактора за које је експериментално анализиран промоторски специфицитет. Расположиви подаци за преостале чланове ECF групе резултат су биоинформатичких предвиђања, која за полазне премисе узимају обележја промоторског специфицитета експериментално проучених ECF σ фактора (нпр. обавезно присуство -35 елемента, парадигма о ауторегулацији итд.) [30]. Оваквим приступом може се доћи само до сличних биоинформатичких предвиђања, што даље може конвергирати ка успостављању нетачних парадигми. Додатно, чак и за добро проучене ECF σ факторе (σ^E из *E coli* и σ^W из *B. subtilis*), за које је расположив већи број експериментално потврђених промотора, до сада није рађена квантитативна анализа да би се утврдило у којој мери њихови промоторски елементи (не) испољавају "mix-and-match".

Иако међу члановима фамилије σ^{70} постоје значајне разлике у структури и функцији, не треба занемарити чињеницу да приликом иницијације транскрипције све четири групе σ^{70} фактора пролазе кроз исте кораке, чија се кинетика биофизички описује истим параметрима [21]. Имајући ово у виду, разумна је претпоставка да за све групе постоји обједињујући биофизички механизам интеракције са промоторским секвенцама (нпр. "mix-and-match"), при чему би различите структуре σ фактора могле бити узрок различитих кинетичких профила иницијационог процеса (нпр. различитих кинетичких параметара чија минимална вредност мора да се оствари). Додатно, карактеристика свих група σ^{70} фактора је да, у зависности од услова под којима регулишу генску експресију, подешавају транскрипциони профил тако да се омогући оптимално физиолошко

функционисање бактеријске ћелије. Ово значи да је за ћелију значајно одржавање постојећих транскирпционих веза у сваком сигмулону (регулон σ фактора), а нарочито у сигмулонима алтернативних σ фактора, који регулишу ћелијски одговор на стрес мобилисањем активности малог броја гена. У овом контексту, механизам "mix-and-match" постаје ефикасна платформа за компензовање дејства мутација у промоторским елементима, које смањују енергију интеракције елемента са σ фактором. Наиме, концепт надопуњавања снага промоторских елемената значи да се мутације у једном промоторском елементу, које воде смањењу кинетичког параметра који је услов промоторске активности, могу компензовати увођењем мутација у било који од преосталих промоторских елемената, уколико исте доводе до поновног постизања вредности тог кинетичког параметра. Са механистичког аспекта, механизам "mix-and-match" би се у различитим групама фамилије σ^{70} могао спровести кроз комплементације снага различитих промоторских елемената, у зависности од конкретних структура протеинских секвенци σ фактора, а са циљем произвођења физиолошког одговора карактеристичног за сваку групу σ^{70} фактора.

Будући да је група ECF најразличитија у односу на примарне RpoD σ факторе [30], њени чланови су најподеснији модел за изучавање присуства механизма "mix-and-match" у целокупној фамилији σ^{70}, при чему би оваква анализа дала допринос и у проширивању увида у механизме транскрипционе иницијације у овој најбројнијој, а веома слабо изученој, групи алтернативних σ фактора. Међутим, да би механизам "mix-and-match" могао квантитативно да се тестира, али и да се прошири знање о било ком аспекту функционисања ECF σ фактора, неопходно је прецизно одредити све елементе који улазе у састав промотора, као и интеракције између ових елемената и различитих домена σ фактора. Ово намеће потребу за систематском биоинформатичком анализом промоторског специфицитета у субфамилији, као почетним извором значајно различитих података о механизму препознавања промотора у односу на тренутна сазнања, ограничена на информације о експериментално изученим представницима. С друге стране, експериментално предвиђене промоторске секвенце могу директно да се користе за тестирање механизма "mix-and-match", при чему би корелације снага различитих промоторских елемената и њихово поређење са референтним вредностима из групе RpoD

могле да укажу на кинетичке параметре, који описују функционалност промотора у овим различитим групама фамилије σ^{70}.

1.3. Бактериофагни σ фактори као модел систем за изучавање специфицитета у субфамилији ECF

Погодан модел систем за анализу промоторског специфицитета и проналажење значајно различитих података за σ факторе из субфамилије ECF могу бити геноми бактериофага, вируса који инфицирају бактеријске ћелије. Још од почетка развоја молекуларне биологије као научне дисциплине, многе важне парадигме о структури генома и регулацији генске експресије успостављене су кроз експериментална истраживања на бактериофазима као модел системима. Експеримент који су Hershey и Chase спровели на бактериофагу T2 довео је до открића да је молекул ДНК, а не протеини, носилац генетичке информације [56]. Експерименти са бактериофагом λ као модел системом допринели су разумевању операторског концепта у регулацији генске експресије на нивоу иницијације транскрипције [57-59]. Једноставност животног циклуса бактериофага, одређена малом количином генетичке информације, чији проток регулише врло ограничен број чинилаца, омогућава лаку манипулацију у лабораторији и основни је разлог широке применљивости ове групе вируса у молекуларно-биолошким истраживањима.

Данас, бактериофази су погодни модел системи и за биоинформатичке анализе, посебно узевши у обзир експоненцијални пораст броја секвенцираних фагних генома, похрањених у базама података. Овај растући број секвенци бактериофагног порекла доступних за анализу део је опште тенденције, узроковане продором све напреднијих, а приступачних технологија за секвенцирање. Будући да су бактериофази по бројности и динамици најдоминантнији биолошки ентитети у микробијалним заједницама [60], нови сојеви се изолују са великом учесталошћу, а незаобилазан корак у карактеризацији добијених изолата представља управо секвенцирање фагних генома. Стицање нових биолошких увида из масивне количине података, која се овим путем производи, зависи од ефикасне квантитативне/биоинформатичке анализе [61]. Међутим, и поред растућег броја изолованих бактериофага и секвенцираних фагних генома, расветљавање биолошког диверзитета присутног у популацијама ове групе вируса још увек је у зачетку. Стога,

биоинформатичка анализа сваке нове геномске секвенце носи потенцијал за уочавање другачијих парадигми за регулацију животног циклуса бактериофага, које се најчешће испољавају управо на нивоу иницијације транскрипције.

Великој разноврсности бактериофагних популација доприносе и честе генетичке размене, што за последицу има мозаичну структуру фагних генома [60]. У пракси, ово значи да велики број предвиђених гена (~80%) у новосеквенцираном фагном геному не може функционално да се анотира. Чак и када се претрагом база предвиде извесни погоци, најбољи међу њима (~ мање од 10) односе се на готово идентичне секвенце сродних бактериофага, док погоци из других филогентских група (нпр. бактерија) показују веома низак ниво сличности са упоређиваном фагном секвенцом. Услед овога, за изучавање фагних секвенци најчешће је неопходан независан, *de novo* приступ, будући да је компаративном анализом могуће добити врло ограничене информације о управо секвенцираном бактериофагу. Управо у овом контексту, бактериофази су значајни за анализу σ фактора субфамилије ECF, где могућност добијања података о промоторском специфицитету, који се неће ослањати на почетне премисе пореклом од изучених бактеријских представника, има првостепени значај.

Додатно, бактериофази поседују врло кратке геноме, реда величине неколико десетина/стотина килобаза (кбп), што их додатно чини погодним модел системом за биоинформатичке анализе [62-63]. Наиме, краћа секвенца значи мању количину података које биоинформатички алгоритам обрађује, што сем убрзања претраге често доприноси и већој тачности. Такође, дужина фагних генома, која је приближно два реда величине мања од просечне дужине бактеријског генома, подразумева мање сложену контролу протока генетичке информације. Наиме, транскрипциона иницијација се код бактериофага у највећој мери регулише помоћу бактеријске РНКП и два σ фактора, од којих један кодира геном бактериофага, а други је део бактеријског холоензимског комплекса, или помоћу две РНКП – једне бактеријске, друге фагне [64-65]. Додатно учешће транскрипционих фактора, који могу да ступе у сложене међуодносе са основним регулаторним компонентама, може да се јави код бактериофага са комплекснијим животним циклусом, као што су лизогени бактериофази [66]. Међутим, стандардна парадигма регулације експресије бактериофагних гена, као што је наглашено, подразумева учешће малог броја регулатора [65, 67], што знатно олакшава њено изучавање.

Проучавање процеса транскрипционе иницијације код бактериофага аналогно је изучавању инфективне стратегије, будући да се контрола животног циклуса ових вируса одвија на нивоу транскрипционе иницијације [67]. Регулацијом транскрипције успостављају се каракетеристични временски обрасци активности функционално повезаних група гена, што корелише са сменама ране, средње и касне фазе у животном циклусу бактериофага [64]. У раној фази синтетишу се фактори који касније у инфекцији блокирају бактеријску, а стимулишу фагну транскрипциону активност. У средњој фази дешава се умножавање вирусног генома у великом броју копија, што је праћено интензивном синтезом ензима неопходних за репликацију ДНК и метаболизам нуклеотида. Напослетку, у касној фази синтетишу се структурни елементи вирусног омотача и помоћни фактори, који учествују у формирању нових вирусних честица и њиховом ослобађању из бактеријске ћелије.

Регулаторна парадигма којом се постиже прецизна синхронизација описаних процеса почива на сукцесивним променама специфицитета РНКП за промоторе. Групе гена са истим временским обрасцем експресије повезане су промоторским секвенцама истог специфицитета [64, 68], при чему је сценарио најчешће такав да се рани фагни гени експримирају са промотора које препознаје примарни (RpoD) σ фактор бактерије домаћина. Остале групе фагних гена (средња и касна) експримирају се са промотора које препознају σ фактори (или РНКП) кодирани фагним геномом, чија активност обезбеђује висок ниво вирусне транскрипције у одмаклим фазама инфекције. σ фактори које кодирају фагни геноми генерално су слабо проучени и, као што је претходно истакнуто, могу бити врло дивергентни у односу на бактеријске σ факторе из истих група. Стога, разумно је претпоставити да проучавање специфицитета фагних σ фактора из субфамилије ECF може бити извор нових, независних података о механизму препознавања промотора за ову важну, али недовољно проучену, групу алтернативних σ фактора.

Анализа специфицитета фагних σ фактора се, у највећој мери, своди на предвиђање непознатих промоторских секвенци у вирусном геному, што одговара *ab initio* претрази регулаторних елемената на ДНК која се заснива на алгоритмима за локално поравнање већег броја секвенци (MLSA – *M*ultiple *L*ocal *S*equence *A*lignment). Ова тема у биоинформатици се још увек сматра отвореним проблемом, при чему предвиђање фагних промотора не потпада под уобичајену примену MLSA алгоритама, јер се ови елементи

типично јављују као неколико мотива дужине приближно 10 базних парова (бп) у геномима реда величине 50 – 100 кбп [69]. Промотори овакве организације тешко се препознају стандардним MLSA алгоритмима, који су оптимизовани за проналажење мотива присутних у већини претраживаних секвенци. Стога, изучавање специфицитета фагних σ фактора уједно подразумева и отимизацију приступа за *ab initio* предвиђања регулаторних елемената, који су присутни у врло малом броју понављања у односу на укупан број претраживаних секвенци.

Додатни интерес за изучавање механизама транскрипционе иницијације/регулације генске експресије код бактериофага је терапеутски потенцијал ових организама за борбу против растућег броја патогених бактерија, резистентних на антибиотике [70-72]. За алтернативне третмане најподеснији су протеини које бактериофаг у природним условима експримира ради блокаде метаболичких активности ћелије домаћина, међу којима се нарочито издвајају инхибитори бактеријског холоензима РНКП. Ови молекули су важни за прелазак из ране у средњу фазу животног циклуса бактериофага па анализа генске експресије током инфекције представља први корак у разумевању механизма њиховог функционисања. Стандардни приступ за анализу фагне генске експресије обухата врло сложену експерименталну методологију (утврђивање временског обрасца генске експресије помоћу микрочипа и биохемијску карактеризацију промоторских елемната), која се комбинује са биоинформатичким предвиђањима [64, 73]. Овакав приступ неекономичан је са становишта утрошеног времена и ресурса па је процена ефикасности директне биоинформатичке анализе у изучавању инфективне стретегије бактериофага још једно важно питање, на које се може дати одговор у контексту анализе специфицитета σ фактора из групе ECF.

1.4. Биоинформатичке методе за изучавање специфицитета σ фактора

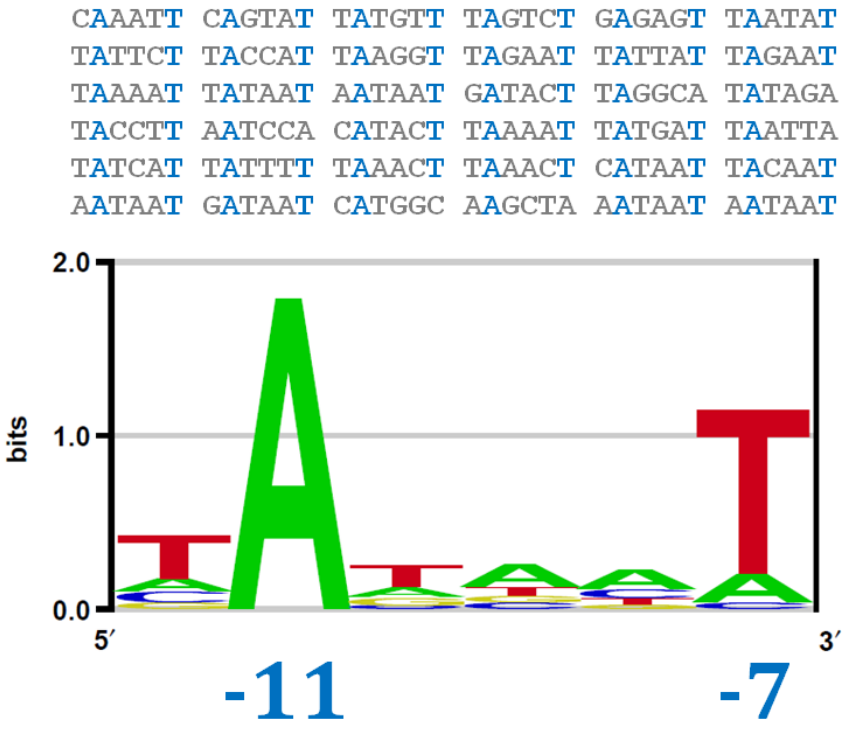
1.4.1. Надгледана претрага промоторских елемената

Још од успостављања првих парадигми о регулацији генске експресије, изучавање интеракције између транскрипционих фактора и одговарајућих везивних секвенци на ДНК представља веома активну област истраживања у биофизици и молекуларној биологији. Подробно познавање специфицитета транскрипционих фактора може бити основ за процену функционалне активности конкретних регулаторних елемената, као и ефекта мутација на степен дате активности. Међутим, познавање специфицитета транскрипционог фактора најчешће се користи као основ за надгледану биоинформатичку претрагу чији је циљ предвиђање нових везивних места у геному, тј. потпуније описивање регулона транскрипционог фактора на основу прикупљених сазнања о новим циљним генима. Додатно, анализом предвиђања добијених за различите транскрипционе факторе, чији регулаторни елементи коегзистирају на истим/приближним геномским локацијама, може се стећи увид у хијерархијски принцип организације регулаторних модула вишег реда [74], преко којих се прецизно регулишу комплексни физиолошки процеси у складу са тренутним метаболичким захтевима ћелије.

Да би информација о специфицитету транскрипционог фактора, најчешће доступна у виду ограниченог скупа експериментално утврђених везивних места, могла да се искористи у алгоритму као основ за узорковање нових места, неопходно је наћи одговарајућу математичку репрезентацију специфицитета [75] којом ће бити описана својства интеракције транскрипционог фактора са везивним местима на ДНК. Узевши у обзир изразиту варијабилност секвенци ДНК за које транскрипциони фактори могу да се вежу високим афинитетом, налажење шаблона/репрезентације, који ће обезбедити оптималну осетљивост и специфичност при предвиђању нових регулаторних елемената у геному, представља не тако једноставан биоинформатички задатак [76].

У биолошком контексту, варијабилност везивних секвенци транскрипционог фактора представља основ за испољавање плејотропије, чији се опсег мери бројем циљних гена у регулону и распоном на њима остварене транскрипционе активности. У складу са

наведеним, врло висок степен варијабилности треба очекивати међу везивним секвенцама σ^{70} фактора (нарочито примарних), који представљају глобалне регулаторе генске експресије у бактеријској ћелији [7]. Претходно се лако илуструје на примеру -10 елемента, који препознаје RpoD σ фактор из *E. coli*, у чијем саставу само две, од укупно шест позиција, испољавају висок степен конзервације једног од четири нуклеотида (Слика 4). Као што се уочава на Слици 4, конзервиране су управо позиције (A_{-11} и T_{-7}) на којима се у току формирања отвореног комплекса нарушава поредак базног наслојавања, као последица интеракције са хидрофобним џепом домена $\sigma_{2.3}$ [47, 77], што указује на постојање везе између природе интеракције протеин-ДНК и уочене учесталости/конзервације нуклеотида на различитим позицијама везивног елемента.



Слика 4. Поравнање (лого) -10 елемента за бактеријски RpoD σ фактор. У горњем делу слике приказане су секвенце конкретних -10 елемената (преузете из базе RegulonDB [78]), са позицијама -7 и -11 означеним плавом бојом за свако уочено присуство консензусних нуклеотида (Т и А); На доњем делу слике приказан је лого направљен од датих секвенци, ради бољег уочавања учесталости консензусних нуклеотида на позицијама -7 и -11.

Метод који за репрезентацију специфицитета транскрипционог фактора експлоатише ово својство, и који се и данас најшире примењује за надгледану претрагу регулаторних (нпр. промоторских) елемената, су матрице тежине [79-80]. У матрицама тежине дескриптори специфицитета изводе се из опажених учесталости нуклеотида у почетном скупу података

[81], пружајући за сваку позицију у елементу процену релативног квантитативног доприноса различитих нуклеотида енергији интеракције протеин-ДНК. Наиме, у раду [82] показано је на основу теорије статистичке механике, да енергијски доприноси за различите позиције везивног елемента одговарају негативним вредностима логаритама опажених учесталости нуклеотида на датим позицијама, уз основну претпоставку адитивности, тј. одсуства међузависности појединачних енергијских доприноса (тзв. апроксимација независних нуклеотида). Детаљнији увид у парамтеризацију матрица тежине и начин узорковања анализираних секвенци дат је у наставку.

1.4.1.1. Матрице тежине

Матрица тежине, као израз специфицитета транскрипционог фактора, чија везивна места на ДНК садрже L позиција, представљена је елементима – $W(\alpha, i)$, при чему је α = A, C, G или T, док је $i = 1,...,L$. Као што је наглашено, елементи матрице тежине (којих укупно има $4\times L$) дају процену релативног енергијског доприноса нуклеотида α на позицији i укупној енергији интеракције између транскрипционог фактора и секвенци ДНК. Стога, скор матрице за било коју анализирану секвенцу једноставно се добија сабирањем елемената у матрици, који одговарају присуству конкретних нуклеотида на позицијама у анализираној секвенци, и представља процену енергије њене интеракције са транскрициоиним фактором [79]:

$$\left(S_j|W\right) = \sum_{\alpha,i} W(\alpha, i)S(\alpha, i)). \quad (1.1)$$

У изразу (1.1) са $W(\alpha, i)$ означени су сви елементи матрице тежине, док $S(\alpha, i)$ представља матричну репрезентацију анализиране секвенце S_j (где је j = 1,...,N), у којој се присуство нуклеотида α на позицији i у секвенци кодира са 1, док се одсуство кодира помоћу 0 – тј. тражени скор представља скаларни производ матричне репрезентације секвенце и матрице тежине.

Кључни корак у претрагама заснованим на матрицама тежине је одређивање конкретних вредности елемената матрице, односно њена параметризација на основу скупа познатих везивних места S. Ова парметризација се заснива на разлици између својстава секвенци из скупа S, и тзв. "позадине" (под којом се најчешће подразумева преостала некодирајућа ДНК у геному). Уколико пођемо од претпоставке да постоји пробабилистички модел Θ,

којим су генерисана везивна места из скупа *S*, примена апроксимације независних нуклеотида [76] омогућава параметризацију овог модела помоћу матрице чији елементи – $\Theta_{\alpha,i}$ представљају вероватноћу налажења нуклеотида α на позицији i у секвенци, која је генерисана моделом Θ. Елементи вероватноће $\Theta_{\alpha,i}$ процењују се на основу критеријума максималне веродостојности, што значи да је потребно наћи такве вредности $\Theta_{\alpha,i}$ које ће максимизовати очекивање да су везивне секвенце из скупа *S* заиста генерисане моделом мотива Θ, уместо моделом позадине Θ_0 (чији елементи носе вероватноће појављивања четири нуклеотида у некодирајућој ДНК), тј. тежи се максимизацији односа:

$$\frac{\prod_{j=1}^{N} P\left(S_j|\,\Theta\right)}{\prod_{j=1}^{N} P\left(S_j|\,\Theta_0\right)} \tag{1.2}$$

што доводи до процене:

$$\Theta_{\alpha,i} = \nu_{(\alpha,i)} = \frac{1}{N}\sum_{j=1}^{N} S_j\,(\alpha, i). \tag{1.3}$$

Као што је наглашено, веза између параметара $\Theta_{\alpha,i}$, тј. опажених учесталости нуклеотида на различитим позицијама у елементу и доприноса укупној енергији везивања има упориште у статистичкој механици [82] и укратко је сумирана у наставку. Полазећи од скупа везивних места (S_j), за које је познато да имају висок афинитет за дати транскрипциони фактор, и процене нуклеотидног састава геномске секвенце из које везивна места потичу, може се наћи матрица $T(\alpha, i)$, чији елементи представљају удео у укупној енергији везивања нуклеотида α на позицији i. Налажењем скора матрице *T* за анализирану секвенцу, може се одредити вероватноћа да ће дата секвенца бити везана од стране транскрипционог фактора ($P_b(S)$, при чему b означава "bound", односно везано стање):

$$P_b(S) = \frac{e^{-\sum_{\alpha,i} T(\alpha,i)S(\alpha,i))}}{Z} \tag{1.4}$$

где $S(\alpha,\ i)$ представља матричну репрезентацију секвенце *S*, док је са *Z* означена партициона функција, односно сума експонената енергија везивања за сваку могућу секвенцу у геному, која има дужину анализиране секвенце *S*. Будући да је унапред познато да везивна места која се користе за параметризацију матрице $T(\alpha, i)$ имају висок афинитет за транскрипциони фактор, тј. високу вероватноћу да за њега буду везана, примена критеријума максималне веродостојности за параметризацију непознатих елемената матрице овде добија биофизичку интерпретацију, јер се датим поступком сада

максимизује вероватноћа везаног стања $P_b(S)$ за сваку секвенцу из скупа S_j. Узимањем у обзир претпоставке да геномски сегменти, чија дужина одговара везивном месту S, прате учесталост нуклеотидног састава генома, одређивање елемената матрице $T(\alpha, i)$ који максимизују очекивање за $P_b(S)$, врши се по формули:

$$T(\alpha, i) = -ln\frac{\nu_{(\alpha,i)}}{p_\alpha} \tag{1.5}$$

где p_α представља учесталост нуклеотида α у геномској секвенци. Претходно значи да, полазећи само од скупа познатих везивних места, можемо наћи елементе матрице тежине који ће дати процену енергијског доприноса за сваки нуклеотид, на свакој позицији у мотиву. На основу овако параметризованих елемената, узорковањем било које секвенце, чија дужина одговара броју везивних позиција кодираних матрицом, можемо добити процену енергије интеракције, односно вероватноће да је дата секвенца везивно место транскрипционог фактора, чији је специфицитет моделован матрицом тежине. Овде је неопходно напоменути да узорковање матрицом не даје и процену прага, на основу ког се могу раздвојити специфична од неспецифичних везивних места за протеин, већ се ова вредност најчешће одређује емпиријски [83]. Такође, да би се при параметризацији елемената матрице избегла појава логаритамске дивергенције услед ограничене величине скупа података из којих се одређује матрица, неопходно је увођење тзв. "псеудобројева" $\nu_{(\alpha,i)}$, који представљају корекцију за случајеве када присуство нуклеотида α на позицији i није уочено у везивним местима из скупа S:

$$\nu_{(\alpha,i)} = \frac{\sum_{j=1}^{N} S_j(\alpha,i) + b_\alpha}{N + \sum_\alpha b_\alpha} \tag{1.6}$$

при чему је b_a псеудоброј за базу a, а вредност му се најчешће подешава тако да буде пропорционална или једнака позадинској учесталости базе p_a. Порастом N утицај псеудобројева постаје мањи, али исто тако и вероватноћа да се база a не уочи на позицији i у поравнању секвенци скупа S.

Из претходног је јасно да тачност матрице тежине увелико зависи од својстава почетног скупа секвенци, на основу којих се њени елементи параметризују, при чему се додатним ограничењем може сматрати и то што, сем процене енергије везивања секвенце ДНК за транскрипциони фактор, скор матрице тежине не пружа никакав увид у механизам дате интеракције [79]. Такође, за извесне транскрипционе факторе апроксимација независних нуклеотида, тј. независних енергијских доприноса на различитим позицијама у елементу,

не мора у потпуности да буде реалистична [76]. Апроксимација независних нуклеотида може да се заобиђе генерализацијом концепта матрице тежине, која дозвољава да елементи матрице кодирају и друга својства специфицитета, сем појединачних енергијских доприноса различитих позиција у елементу, попут условљености појаве нуклеотида α на позицији i присуством различитих нуклеотида на суседним позицијама [79]. Међутим, за овакав приступ, неопходно је обезбедити довољно обиман почетни скуп везивних места, из кога може адекватно да се изведе знатно већи број параметара, који матрице вишег реда захтевају (у супротном, матрица са већим бројем параметара ће заправо водити ка мањој тачности предвиђања).

За прокариотске σ факторе, међутим, матрице тежине првог реда генерално су добра апроксимација правог специфицитета [84-85], што се огледа и у врло широкој примени за предвиђање промоторских елемената, које препознају примарни (RpoD) σ фактори. Овај приступ биће примењен за предвиђање RpoD бактеријских промотора при анализи бактериофагне геномске секвенце, у склопу изучавања вирусне транскрипционе стратегије, али и за изучавање промоторског специфицитета у групи ECF, у случајевима када је информација о везивним местима унапред доступна. На крају, неопходно је напоменути да предвиђање секвенце која носи велику вероватноћу да буде везивно место за σ фактор, не значи нужно и проналажење промоторске активности на секвенци, због чега претрага промоторских елемената заснована на матрицама тежине често доводи до великог броја лажно позитивних предвиђања. Концепт којим се омогућава превазилажење овог проблема изложен је у поглављу 1.5.

1.4.2. Ненадгледана (*ab initio*) претрага промоторских елемената

У претходном поглављу изложен је приступ за биоинформатичко предвиђање нових регулаторних елемената у геномској секвенци, на основу познатог специфицитета транскрипционог фактора (нпр. претрага промоторских елемената за примарни RpoD σ фактор). Информација о специфицитету, међутим, често није унапред доступна услед чега је неопходно развити методе за проналажење мотива, конзервираних у датом скупу анализираних секвенци, за које се претпоставља да садрже заједнички мотив. Пример претходног су секвенце локализоване узводно од гена, чију активност регулише исти

транскрипциони фактор. Услед вероватноће да већина ових секвенци садржи места за везивање датог транскрипционог фактора, управо ова везивна места представљају заједнички мотив у секвенцама, чије откривање би било циљ ненадгледане претраге.

Слично процесу параметризације матрица тежине, главни принцип ненадгледане претраге је да се максимизује извесна мера конзервације мотива, који тежимо да предвидимо. При том, као критеријум квалитета поравнања најчешће се истиче тзв. информациони садржај ("information content"), који представља логаритам количника вероватноћа да је добијено поравнање генерисано моделом мотива, у односу на модел позадине [86-87].

Да би се локализовало поравнање, које у претраживаним секвенцама (типичне дужине ~100 бп) максимизује информациони садржај, неопходно је истражити енормно велики број комбинација сегмената, који по дужини одговарају траженом мотиву. Због тога се претрага вишемодалне функције вероватноће (са великим бројем локалних максимума) врши апроксимативном процедуром, која се заснива на репетитивном узорковању секвенци [88]. Ово је основни концепт у раду пробабилистичких алгоритама за локално поравнавање већег броја секвенци (MLSA – *M*ultiple *L*ocal *S*equence *A*lignment), где разликујемо два основна приступа – детерминистички и хеуристички. Пример детерминистичког узорковања у MLSA алгоритмима је метод *EM* (*E*xpectation *M*aximization), који се заснива на критеријуму максималне веродостојности, док Гибсова претрага, заснована на методу Монте-Карло, представља хеуристички облик репетитивног узорковања [87, 89].

У *EM* претрази алгоритам циклично пролази кроз два корака, при чему се у првом кораку параметри модела мотива Θ иницијализују на случајну вредност, ради локализовања поравнања, у које улазе сегменти претраживаних секвенци са најбољим скором матрице тежине, описане моделом Θ. У другом кораку локализовано поравнање користи се за поновну процену параметара модела Θ, након чега се описани кораци понављају све до конвергенције ка најближем локалном максимуму функције вероватноће. Конвергенција ка локалном максимуму је загарантована за алгоритам *EM*, будући да исти представља детерминистичку процедуру [90].

Чињеница да алгоритам *EM* конвергира ка најближем (у односу на почетне услове претраге) локалном максимуму мере конзервације мотива води ка томе да глобални максимум не мора да буде достигнут у *EM* претрази. Наиме, биолошке секвенце

представљају податке велике сложености па њима типично одговарају мере конзервације мотива са великим бројем локалних максимума. Стога, с порастом броја анализираних секвенци расте и димензионалност проблема, што доводи до веће осетљивости алгоритма *EM* на иницијалне услове претраге, а такође и до значајног пораста временског интервала потребног за конвергенцију алгоритма. Да би се овај проблем делимично превазишао, *EM* алгоритам се у пракси (репетитивно) иницира са већим броја различитих почетних услова, међутим, глобални максимим ни у овом случају не мора бити досегнут.

С друге стране, у хеуристичкој Гибсовој претрази процењује се заједничка расподела параметара позиције и модела, док се локализација оптималног поравнања обавља касније у претрази (тј. након корака узорковања) [89]. Код Гибсове претраге користе се насумични Монте-Карло бројеви, тако да се локације мотива у сваком кораку алгоритма генеришу са вероватноћом пропорционалном коришћеној мери конзервације мотива. Будући да представља хеуристичку процедуру, Гибсова претрага је мање осетљива на велики број локалних максимума, који се типично јављају у поравнању секвенци. Такође, у односу на алгоритам *EM*, Гибсова претрага знатно брже доспева до траженог решења, што представља још једну погодност са аспекта практичности примене. Узевши у обзир претходно, Гибсова претрага биће метод избора за ненадгледану претрагу промотора у анализи специфицитета ECF σ фактора, услед чега у наставку поглавља следи детаљнији увид у принцип рада овог алгоритма.

1.4.2.1. Гибсова претрага

У основном облику, алгоритам заснован на Гибсовој претрази анализира скуп секвенци (S_1,…,S_N) у циљу проналажења заједничког (тј. највероватнијег) мотива унапред дефинисане дужине L [91]. Предвиђању мотива претходи пролазак алгоритма кроз итерације, састављене од два корака, где је сваком од њих придружена по једна растућа структура података. У првом кораку развија се пробабилистички модел мотива, док се у другом кораку на основу њега проналази најбоље поравнање, избором највероватнијих почетних позиција мотива у анализираним секвенцама. Модел мотива (Θ) описује се параметрима $q_{\alpha,i}$ који означавају учесталост нуклеотида α на i-тој позицији мотива, при чему је α = A, C, G или T, док је i = 1,...,L. У састав прве структуре података улази и модел

позадине (Θ_0), који као параметре има $p_{\alpha,}$ тј. учесталости нуклеотида (α = A, C, G или T) на позицијама анализираних секвенци, ван поравнатих мотива. Друга структура података представљена је скупом параметара A_k, који се састоји од почетних позиција заједничког мотива (a_k ; $k = \{1, \dots, N\}$) у анализираном скупу секвенци (S_1,…,S_N).

Иницијализација рада алгоритма подразумева издвајање једне од секвенци (коју означавамо са z) из скупа (S_1,…,S_N) (редослед издвојене секвенце је небитан) и, паралелно, додељивање позиције мотива a_k преосталим секвенцама насумичним избором (т.j. насумичним Монте-Карло бројевима). На основу овако дефинисаног поравнања (A_k) рачунају се учесталости $q_{\alpha,i}$ по формули:

$$q_{\alpha,i} = \frac{c_{\alpha,i} + b_\alpha}{N-1+B} \tag{1.7}$$

где N - 1 представља укупан број нуклеотида у поравнању на i-тој позицији мотива (секвенца z која је издвојена из поравнања је искључена из процене $q_{\alpha,i}$), $c_{\alpha,i}$ означава уочени број нуклеотида α на i-тој позицији у поравнању, b_α је корекција (псеудоброј, објашњено у поглављу 1.4.1.1) за учесталост нуклеотида α, док B означава збир псеудобројева за сва четири нуклеотида. Позадинске вероватноће p_α се процењују на основу позиција које у анализираним секвенцама не улазе у поравнање мотива.

У другом кораку се на основу описаних вероватноћа $q_{\alpha,i}$ и p_α, претражује издвојена секвенца z, где се за сваки сегмент t дужине L (дужина мотива), рачуна количник вероватноћа да је дати сегмент генерисан моделом мотива, односно моделом позадине:

$$W_t \propto \prod_{i=1}^{L} \left(\frac{\theta_i}{\theta_0}\right)^{b_{t+i-1}} \tag{1.8}$$

при чему b има вредност 1 уколико је нуклеотид α присутан на одговарајућој позицији у сегменту, а 0 у супротном. Нормализацијом количника добија се вероватноћа да се мотив описан моделом Θ налази на датој (t) позицији у секвенци z:

$$W_t / \sum_{t=1}^{Lz-L+1} W_t \tag{1.9}$$

Из скупа овако добијених вероватноћа позиција мотива a_k у секвенци z одабира се стохастичким узорковањем, при чему већа нормализована вредност W_t носи већу вероватноћу узорковања сегмента t као позиције мотива a_k у секвенци z.

Након завршене једне итерације, ново поравнање A_k служи за поновно израчунавање параметара модела мотива и позадине ($q_{\alpha,i}$ и p_α), за којим следи узорковање a_k у следећој секвенци z (рачунањем вредности W_t, али сада на основу нове процене о учесталостима $q_{\alpha,i}$

и p_α). Главна идеја је да, што је прецизнији опис мотива у претходном кораку, то ће тачније бити одређена његова позиција у следећем кораку, и обратно, јер се постиже боље раздвајање модела мотива од модела позадине. Будући да је узорковање које се обавља стохастичко, до првог избора тачне позиције a_k алгоритам може да дође кроз различит број итерација. Међутим, избором првог тачног a_k, у наредним итерацијама долази до избора наредних тачних позиција, које заузврат још више појачавају дискриминатна својства модела мотива, што најчешће доводи до оптималног поравнања A_k, односно налажења глобалног максимума.

За разлику од алгоритма *EM*, који гарантује конвергенцију ка најближем (најчешће локалном) максимуму, у односу на услове под којима је претрага покренута, у Гибсовој претрази то није случај услед стохастичке природе узорковања. Ову мању вероватноћу остајања у локалном максимуму Гибсова претрага, међутим, може да "плати" пријављивањем субоптималних решења што, уз одсуство процене статистичког значаја добијених поравнања, изискује понављање претраге ради процене робусности пријављених мотива.

Као што је на почетку истакнуто, оба приступа се заснивају на претпоставци да је заједнички (варијабилни) мотив присутан у већини претраживаних секвенци. Међутим, постоје случајеви где наведена претпоставка не може да се примени, при чему је један од најилустративнијих примера препознавање промотора за σ факторе унапред непознатог специфицитета, који су кодирани бактериофагним геномима. У односу на број претраживаних секвенци (~100 интергенских региона), ови мотиви се типично јављају у врло ограниченом броју копија (~10), због чега у претрази помоћу алгоритама MLSA врло лако могу бити маскирани сегментима сличним са мотивима које препознају дати σ фактори, а који се насумично јављају у геномским секвенцама. Имајући то у виду, изучавање промоторског специфицитета за бактериофагне ECF σ факторе уједно значи и процену способности MLSA алгоритама (конкретно Гибсове претраге) да препозна фагне промоторе, као и поређење ових алгоритама са алтернативно конципираним процедурама. Такође, Гибсова претрага ће бити екстензивно коришћена и у изучавању специфицитета бактеријских ECF σ фактора, чији највећи део чине експериментално неизучени представници.

1.4.3. Глобално поравнавање већег броја секвенци

Сем локалних поравнања већег броја секвенци, у којима је циљ предвиђање кратких функционалних мотива од интереса (нпр. промоторских елемената), важан извор информација о функционисању σ фактора могу дати и глобална вишеструка поравнања, у којима је могуће паралелно анализирати функционалне целине различитих размера (од домена до кратких конзервираних мотива). Наиме, глобално поравнавање већег броја протеинских секвенци ECF σ фактора представља комплементаран приступ анализи промоторског специфицитета, јер омогућава изучавање својстава ДНК-интерагујућих домена, који дефинишу специфицитет канонских елемената промотора. С друге стране, метод омогућава и потенцијално предвиђање функционалних/конзервираних мотива за успостављање "неканонских" интеракција са промоторском секвенцом, чија појава (као обележја промоторског специфицитета) није искључена за најхетерогеније и најслабије изучене представнике фамилије σ^{70}.

Најчешће коришћен програм за глобално поравнавање групе (протеинских) секвенци је ClustalW [92], у којем се вишеструко глобално поравнање заснива на хијерархијском методу. Прецизније, алгоритам у почетку пореди анализиране секвенце у паровима, након чега добијена поравнања пролазе кроз корак кластеровања ради хијерархијског груписања секвенци на основу степена сличности забележеног у (парним) поравнањима. Овако груписане секвенце означавају се појмом "водеће дрво", јер је информација која је у њему садржана основ за развијање крајњег поравнања. Окосница за иницирање финалног поравнања су најсличније секвенце, којима се поступно додају све удаљенији чланови из анализираног скупа. Премда је, због хијерархијског развоја поравнања, метод осетљив на пропагацију потенцијалних грешака у поравнању (направљних у било ком кораку процеса), правилан избор секвенци за анализу већином води ка поузданим поравнањима, услед чега ће бити екстензивно коришћен у проучавању доступних представника ECF групе.

1.5. Кинетика транскрипционе иницијације

У претходним поглављима дат је кратак преглед биоинформатичких метода које се најчешће користе за изучавање специфицитета σ фактора, при чему се за неке методе временски распон између најновијих имплементација и првобитне примене основног алгоритма мери деценијама. И поред интензивног, континуираног рада на унапређењу препознавања промотора, тачност добијених биоинформатичких предвиђања је и даље веома ниска, при чему се као главни проблем јавља велики број лажно позитивних предвиђања [93-94]. Илустративан пример је претрага RpoD промоторских секвенци у интергенским регионима *E. coli*, заснована на матрицама тежине, која даје број предвиђања за два реда величине изнад укупног броја гена овог организма – тј. горњег лимита очекиваног броја промотора у геномској секвенци [95].

Већина истраживачких напора у области биоинформатичког изучавања високоафинитетних интеракција регулаторних протеина са ДНК била је усмерена на осмишљање напреднијих алгоритама за прецизније моделовање специфицитета и претрагу нових везивних места на ДНК. Међутим, у домену предвиђања промотора, овај приступ није довео до задовољавајућег побољшања тачности претрага. Пример претходног били су ранији покушаји моделовања специфицитета RpoD σ фактора *E. coli* приступом заснованим на примени неуронске мреже, што представља алгоритамски комплекснију алтернативу матрицама тежине [80]. Међутим, већа прецизност претраге добијена је само у скупу секвенци помоћу којих је модел параметризован, док се ефикасност претраге на независном скупу секвенци није разликовала у односу на приступ заснован на матрицама тежине.

Чињеница да до данас проблем великог броја лажно позитивних предвиђања није решен, сугерише да повећање тачности претраге промотора не може да се сведе само на прецизније моделовање специфицитета σ фактора, већ да захтева додатне информације од биолошког значаја. Наиме, транскрипциона иницијација је вишестепени и интензивно регулисан процес [43], где на опсег активности промотора снажно утиче кинетика корака нисходно од формирања затвореног комплекса. Уколико за пример узмемо претрагу засновану на матрицама тежине, то значи да побољшање предиктивне моћи скорова матрица за различите промоторске елементе подразумева успостављање директне

квантитативне везе између специфицитета и кинетичких параметара иницијационог процеса, којима је одређена функционална активност промотора [96]. Управо је ово постигнуто биофизичким моделом транскрипционе иницијације, чије су основне карактеристике изложене у наставку поглавља.

1.5.1. Биофизички модел транскрипционе иницијације

Представљени биофизички механизам транскрипционе иницијације на једноставан начин повезује енергије интеракције различитих промоторских елемената са σ фактором и кинетичке параметре, који описују сукцесивне кораке иницијационог процеса. Процена енергија интеракције различитих промоторских елемената са σ фактором изводи се из одговарајућих елемената матрица тежине, што значи да модел пружа могућност директног изучавања кинетике транскрипционе инцијације за било који скуп промоторских секвенци, под чији оквир спада и тестирање механизма "mix-and-match".

1.5.1.1. Кинетичка схема и основни параметри

Биофизички модел транскрипционе инцијације, који је развијен од стране [21], заснован је на општој схеми процеса [97]:

$$[RNAP]+[P]\underset{k_{off}}{\overset{k_{on}}{\rightleftarrows}}[RNAP-P]_c \xrightarrow{k_f} [RNAP-P]_o \xrightarrow{k_e} [RNAP]_e+[P] \qquad (1.10)$$

у којој су РНКП, ДНК промотора, затворени и отворени комплекс означени са [*RNAP*], [*P*], $[RNAP\text{-}P]_c$ и $[RNAP\text{-}P]_o$. Брзинске константе настајања и раскидања затвореног комплекса означене су са k_{on} и k_{off}; стопа преласка из затвореног у отворени комплекс са k_f, док је брзинска константа корака у ком РНКП напушта промотор означена са k_e. Први корак на схеми представља реверзибилно везивање РНКП за промотор, након чега следи корак раздвајања ланаца ДНК и формирање отвореног комплекса и, на крају, иреверзибилно напуштање промотора тј. прелазак РНКП у фазу елонгације.

Динамику формирања/раскидања затвореног комплекса карактерише временска скала резолуције ~1s, што је за два реда величине изнад резолуције временске скале динамике формирања отвореног комплекса (~100s) [97]. Узевши у обзир претходно, као и чињеницу

да је РНКП само део укупног времена везана за промотор (тзв. "несатурисана апроксимација") [96], израз за транскрипциону активност може да се поједностави на [21]:

$$\varphi \approx [RNAP] K_B k_f \tag{1.11}$$

где [*RNAP*] означава концентрацију слободне РНКП у ћелији. У једначини (1.11) уочавамо да је транскрипциона активност директно пропорционална производу K_B и k_f (афинитета везивања за dsDNA и стопе преласка из затвореног у отворени комплекс), што одговара уобичајеној мери за снагу промотора [97].

1.5.1.2. Веза кинетичких параметара са енергијама интеракције

Као што је изложено у поглављу 1.2, кинетички параметри транскрипционе иницијације директно зависе од енергија интеракције σ фактора са промоторским елементима на ДНК. Прецизније, афинитет везивања РНКП за дволанчану ДНК зависи од енергије интеракције σ фактора са -35 елементом, узводним делом -10 елемента у форми дволанчане ДНК и спејсерском секвенцом између ова два елемента [21], [12]:

$$\log\left(K_B(S)\right) \sim c - \frac{\Delta G_{ds}\left(S_{(-35)}\right) + \Delta G(\gamma) + \Delta G_{ds}\left(S_{(-10)}^{(ds)}\right)}{k_B T} \tag{1.12}$$

где $S_{(-35)}$, $S_{(-10)}^{(ds)}$ и γ означавају, редом, секвенце -35 елемента, дела -10 елемента који интерагује са РНКП у форми дволанчане ДНК, и дужину спејсерске секвенце, при чему је c константа која не зависи од секвенце промотора. $\Delta G(S_{(-35)})$, $\Delta G(S_{(-10)}^{(ds)})$ и $\Delta G(\gamma)$ представљају, редом, енергије интеракције σ фактора са -35 елементом, dsDNA-сегемтном -10 елемента и спејсерском секвенцом, која зависи од њене дужине.

Да би се и брзинска константа формирања отвореног комплекса – k_f повезала са енергијама интеракције између σ фактора и релевантних промоторских елемената, користи се механистички модел формирања отвореног комплекса [21]:

$$\log\left(k_f\right) = c + \frac{\Delta G_m\left(S_{(-10)}^{(ss)}\right) - \Delta G_{ss}\left(S_{(-10)}^{(ss)}\right)}{k_B T} \tag{1.13}$$

где $S_{(-10)}^{(ss)}$ означава сегмент -10 елемента који интерагује са σ фактором у форми једноланчане ДНК, тј. сегмент у ком се иницира формирање транскрипционог мехура,

$\Delta G_m(S_{(-10)}^{(ss)})$ је енергија топљења сегмента $S_{(-10)}^{(ss)}$ у одсуству РНКП, док $\Delta G_{ss}(S_{(-10)}^{(ss)})$ представља енергију интеракције датог сегмента са σ фактором у отвореном кмплексу.

На основу претходних израза, транскрипциона активност промоторске секвенце S изражава се преко енергија интеракције σ фактора са промоторским елементима на следећи начин:

$$\log\left(\varphi\left(S\right)\right)=c+\frac{-\Delta G_{ds}\left(S_{(-35)}\right)-\Delta G\left(\gamma\right)-\Delta G_{ds}\left(S_{(-10)}^{(ds)}\right)+\Delta G_m\left(S_{(-10)}^{(ss)}\right)-\Delta G_{ss}\left(S_{(-10)}^{(ss)}\right)}{k_BT} \qquad (1.14)$$

где су сви чланови једначине претходно дефинисани.

1.5.1.3. Параметризација модела помоћу матрица тежине

Параметризација једначина које повезују кинетичке параметре транскрипционе иницијације са енергијама интеракције између σ фактора и промоторских елемената, ослања се на статистичко-механичке прорачуне који показују да енеријски допринос интеракције протеин-ДНК на одређеној позицији промоторског елемента одговара параметру матрице тежине којим је дата позиција кодирана [82]. У складу са претходним, $K_B(S)$ и $\varphi(S)$ могу се изразити на следећи начин:

$$\log\left(K_B\left(S\right)\right)\sim\sum_{i=-35}^{-30}\sum_{\alpha=1}^{4}w_{i\alpha}^{(-35)}S_{i\alpha}^{(-35)}+\sum_{j=1}^{5}w_j^{(\gamma)}\delta_{j\gamma}+\sum_{i=-15}^{-12}\sum_{\alpha=1}^{4}w_{i\alpha}^{(-10)}S_{i\alpha}^{(-10)} \qquad (1.15)$$

$$\log\left(\varphi\left(S\right)\right)\sim\sum_{i=-35}^{-30}\sum_{\alpha=1}^{4}w_{i\alpha}^{(-35)}S_{i\alpha}^{(-35)}+\sum_{j=1}^{5}w_j^{(\gamma)}\delta_{j\gamma}+\sum_{i=-15}^{-12}\sum_{\alpha=1}^{4}w_{i\alpha}^{(-10)}S_{i\alpha}^{(-10)}+\sum_{i=-11}^{-7}\sum_{\alpha=1}^{4}w_{i\alpha}^{(-10)}S_{i\alpha}^{(-10)} \qquad (1.16)$$

У горњим једначинама $w_{i,\alpha}$ представљају матрице тежине, при чему суперскрипти ((-35), (-10) и (γ)) означавају конкретне промоторске елементе на које се матрице тежине односе. Индекс i означава различите позиције у оквиру -35 елемента и -10 елемента, док индекс j означава пет могућих растојања између ова два елемента (дужина спејсера).

Сумарно, једначина (1.15) показује да скорови матрица тежина промоторских елемената који интерагују са σ фактором у форми dsDNA адитивно доприносе логаритму афинитета везивања за дволанчану ДНК – $\log(K_B(S))$. Слично, једначина (1.16) показује да, ако у збир укључимо скорове матрица тежина оних елемената промотора који са σ фактором интерагују у форми једноланчане ДНК, добијамо логаритам укупне транскрипционе

активности – $\log(\varphi(S))$. Ова директана веза између скорова матрица тежина и кинетичких параметара иницијационог процеса омогућава тестирање механизма "mix-and-match" за било који скуп промоторских секвенци, са прецизним поравнањем промоторских елемената. Наиме, корелисањем скорова матрица тежине којим су дефинисани различити промоторски елементи, уочава се да ли је међу њима присутна комплементација "снага" (енергија интеракције са σ фактором) и, уколико јесте, ка ком кинетичком параметру је усмерена, при чему је детаљнији увид у начин изучавања кинетичких особина процеса транскрипционе иницијације за конкретну (ECF) групу σ фактора дат у поглављу 4.3.

2. ХИПОТЕЗА И ЦИЉЕВИ

Биофизички модел транскрипционе иницијације, као и биохемијске и молекуларно-биолошке карактеристике представника фамилије (детаљно описани у Уводу), послужили су као основ за формулисање хипотезе докторске дисертације, која претпоставља да механизам "mix-and-match" (доказан у групи RpoD [18-19]) може да се прошири и на остале представнике σ^{70} фамилије, потенцијално представљајући универзалан начин њиховог функционисања. Наиме, биофизички модел указује да је за транскрипциону активност (φ) на σ^{70} промоторима неопходно постизање минималних вредности кинетичких параметара (K_B, k_f), који описују кораке формирања затвореног и отвореног комплекса, при чему се комплементација енергија интеракције различитих промоторских елемената са σ фактором јавља као ефикасна платформа за постизање датог циља. Хипотеза је, такође, у складу са заједничком општом организацијом промотора у σ^{70} фамилији (Слика 1 и Слика 2, Увод), која показује да еквивалентни промоторски елементи у истој форми интерагују са σ фактором (као дволанчани или једноланчани) током иницијације транскрипције. Постојање обједињујућег механизма интеракције са промоторским елементима, међутим, не искључује могућност остваривања специфичних физиолошких одговора у различитим σ^{70} групама, будући да разлике у структурној организацији протеинских секвенци σ^{70} фактора функционално могу да се одразе на кинетички профил иницијационог процеса (нпр. да у σ^{70} групама исти кинетички параметри у различитом опсегу утичу на транскрипциони одговор промотора).

Насупрот нашој хипотези, тренутна парадигма о механизму транскрипционе иницијације у σ^{70} фамилији подразумева да разлике у структури протеинских секвенци чланова фамилије нужно воде ка фундаментално различитом кинетичком механизму иницијације транскрипције [7]. Прецизније, група примарних (RpoD) σ фактора препозната је као једина у којој механизам "mix-and-match" може да се испољи, јер се сматра да алтернативни σ фактори за то не поседују неопходну доменску организацију протеинских секвенци, која је једноставнија него код RpoD σ фактора и, додатно, праћена значајно мањом варијабилношћу промоторских елемената. Стога је за разлику од "mix-and-match" механизма, за алтернативне σ факторе формулисана парадигма ригидног специфицитета,

која подразумева одсуство флексибилности у интеракцији са промоторским елементима током иницијације транскрипције.

Овде је важно приметити да присуство (тј. одсуство) механизма "mix-and-match" може да се утврди само на основу квантитативне анализе, услед неопходности процене енергија интеракције различитих промоторских елемената са σ фактором. Сем за примарне RpoD σ факторе [18] оваква анализа, међутим, никада није спроведена у σ^{70} фамилији, што доводи у питање валидност парадигме о ригидном промоторском специфицитету код алтернативних σ фактора. Будући да је на успостављање ове парадигме у значајној мери утицао јак контраст између физиолошких процеса (и њихове регулације) који су везани за примарне, односно алтернативне σ факторе, најподеснији модел за процену валидности ове парадигме су σ фактори групе ECF, као структурно и функционално најдивергентнији у односу на групу RpoD. Додатно, оваква анализа дала би важан допринос расветљавању механизама транскрипционе инцијације у групи ECF, којом је обухваћен највећи број физиолошки врло значајних, али истовремено слабо проучених, алтернативних σ фактора, посебно пошто је ово и најбројнија група алтернативних σ фактора.

Да би се проверила парадигма о ригидном промоторском специфицитету у групи ECF неопходно је:

i) систематско биоинформатичко проучавање протеинских и ДНК мотива, који су укључени у интеракције ECF σ фактора са промотором током иницијације транскрипције;

ii) утврђивање присуства негативних корелација између енергија интеракције релевантних промоторских елемената са σ фактором (за које је неопходан већи број промоторских секвенци).

Систематско биоинформатичко проучавање промоторског специфицитета значајно је јер представља први корак ка могућности (ширег) квантитативног испитивања "mix-and-match"-а у групи ECF. Наиме, да би комплементација енергија интеракције са σ фактором између различитих промоторских елемената могла да се тестира, прво је неопходно предвидети све елементе којима је описан промоторски специфицитет, тј. који потенцијално могу функционално да се надопуњују. Додатно, већ на нивоу изучавања промоторског специфицитета могуће је уочити (драстичне) примере механизма "mix-and-match", као што је компензација одсуства -35 елемента помоћу јаког продуженог -10

елемента, која се јавља у RpoD промоторима. Такође, у групи ECF, као структурно најудаљенијој у односу на групу RpoD, оправдано је очекивати присуство квалитативно другачијих (тзв. "неканонских") интеракција између σ фактора и промоторске ДНК, а за које је неопходна независна, *de novo* анализа специфицитета.

Да би анализа специфицитета заиста била независна и пружила квалитативно другачије увиде у односу на тренутна сазнања о интеракцији ECF σ фактора са промоторском ДНК, на којима је заснована парадигма о ригидности, почетни корак је изучавање промоторског специфицитета бактериофагних ECF σ фактора. Фагни σ фактори представљају адекватан систем за уочавање другачијих регулаторних парадигми функционисања у ECF групи, пошто су удаљени од добро проучених бактеријских ECF представника. Како за бактериофагне σ факторе промоторски специфицитет најчешће није унапред познат, важан циљ дисертације је и развој биоинформатичких метода, које омогућавају предвиђање фагних промотора директно из геномске секвенце вируса, а које ће бити ефикасна алтернатива стандардном приступу за *ab initio* предвиђања регулаторних елемената (тј. алгоритмима MLSA). Уско повезано са овим циљем је и (што је могуће потпуније) разумевање транскрипционе стратегије бактериофага директно из геномске секвенце, а која је у непосредној вези са распоредом промотора у секвенци бактериофагног генома.

Кроз компаративну анализу добијених предвиђања за фагне ECF σ факторе са специфицитетом бактеријских представника биће започета систематска анализа интеракција, које ECF σ фактори из различитих бактеријских подгрупа остварују са промоторским секвенцама. Прецизније, биоинформатичким методама за поређење и анализу протеинских и секвенци ДНК тежи се предвиђању мотива, који су укључени у иницијацију транскрипције, при чему ће случајеви који указују на комплементацију канонских (нпр. са -35 елементом, као у групи RpoD) и присуство "неканонских" интеракција бити засебно анализирани, као могући показатељи флексибилности у функционисању ECF σ фактора.

Сем проучавања промоторског специфицитета у групи ECF, важан део дисертације обухвата директно тестирање механизма "mix-and-match", које је могуће само за добро проучене (канонске) ECF представнике, за које је доступан већи број (експериментално предвиђених) промоторских секвенци. Наиме, тестирање механизма "mix-and-match"

заснива се на биофизичком моделу транскрипционе иницијације, који омогућава процену енергија интеракције промоторских елемената са σ фактором, као и доприноса ових енергија интеракције кинетичким параметрима иницијације транскрипције. С тим у вези, неопходан је довољно обиман скуп промоторских секвенци како за парамтеризацију матрица, тако и за директно корелисање снага различитих промоторских елемената. Овакву могућност пружа ECF представник σ^E из *E. coli*, за који је доступан релативно велики број промоторских секвенци (~60) [98]. Сем σ^E, још један канонски ECF σ представник (σ^W) поседује већи број експериментално утврђених промотора (~30), што га такође чини погодним кандидатом за анализу [99]. Резултати корелационе анализе за ECF σ факторе биће упоређени са резултатима доступним за примарне RpoD σ факторе, који представљају добро проучену референцу за процену значаја добијених резултата. Наиме, поређење резултата за групе ECF и RpoD пружа увид у то како различита структурна и функционална ограничења у најдивергентнијим σ^{70} групама утичу на испољавање механизма "mix-and-match", првенствено кроз уочавање значаја различитих кинетичких параметара у иницијацији транскрипције за одређене групе σ^{70} фактора.

Важан аспект корелационе анализе у ECF групи (прецизније, поређења добијених резултата са групом RpoD) је успостављање везе између опсега испољавања "mix-and-match" ефекта (који се квантификује одговарајућим корелационим коефицијентима) и величине сигмулона испитиваног σ фактора. Овим се проверава интуитивна претпоставка, по којој је степен плејотропије σ фактора директно повезан са опсегом испољавања "mix-and-match" ефекта, а на којој се заснива и претпоставка о одсуству ових ефеката у групама алтернативних σ фактора. Иако феномен "mix-and-match" инхерентно јесте повезан са варијабилношћу елемената, корелациона анализа нам даје процену опсега његовог испољавања у популацији промоторских секвенци, тј. увид у којој мери је механизам експлоатисан (физиолошки релевантан) за функционисање конкретног σ фактора у целини. Прецизније, постизање јаких корелација (опсега комплементације) није одређено само присуством веће апсолутне варијабилности промоторских елемената, већ, још битније, и учесталошћу са којом уочавамо коегзистенцију слабије и боље конзервираних елемената на нивоу појединачних промоторских секвенци. С тим у вези, мања варијабилност промоторских елемената алтернативних (ECF) σ фактора не значи нужно и

мање изражену/одсутну комплементацију њихових снага, у односу на промоторске секвенце примарних RpoD σ фактора.

Сумарно, постављени циљеви докторске дисертације представљају основ за стицање значајно комплетнијег увида у механизам функционисања σ фактора ECF групе, који у бактеријским ћелијама одржавају хомеостазу у условима дејства широког спектра стресора, порекла углавном из спољашње средине, као и за могућу ревизију неких од тренутних парадигми о иницирању транскрипције код ове групе. Уједно, постављени циљеви су и полазна тачка за даља истраживања (биоинформатичка и експериментална), усмерена ка потврђивању "mix-and-match"-а као обједињујућег биофизичког механизма иницирања транскрипције у целокупној σ^{70} фамилији.

3. Методе

У наредним поглављима изложен је методолошки приступ за изучавање квалитативних и квантитативних аспеката механизма транскрипционе иницијације у групи алтернативних ECF σ фактора. Анализа почиње проучавањем бактериофагних ECF представника као најдивергентнијих у групи, у склопу биоинформатичког изучавања вирусне транскрипционе стратегије током инфекције. С обзиром на доступност анотираних геномских секвенци за проучаване бактериофаге на GenBank-у, биоинформатичка анализа транскрипционе стратегије своди се на предвиђања промотора, која интегрисана са информацијом о функционалној анотацији гена дају слику временског обрасца вирусне генске експресије. Промотори за бактеријски RpoD σ фактор (у геному бактериофага 7-11) једноставно се предвиђају стандардном надгледаном претрагом помоћу матрица тежине [96], док је за предвиђање промотора за фагне σ факторе (и РНКП) развијен нов методолошки приступ, који се заснива на поравнању интергенских региона у паровима [100].

На предвиђања промоторског специфицитета за фагне ECF σ факторе надовезује се систематска анализа квалитативних аспеката функционисања у групи ECF, која обухвата све доступне бактеријске представнике (канонске чланове σ^E и σ^W и представнике осталих ~40 ECF подгрупа) [101], за које се сем промоторске структуре детаљно изучава и сама протеинска секвенца. За изучавање (предвиђање) ECF промотора познатог специфицитета користи се надгледана претрага помоћу матрица тежине, док се за изучавање промотора непознатог специфицитета примарно користе алгоритми за локално поравнање већег броја секвенци (MLSA), који представљају стандарне методе за *ab initio* предвиђање мотива. Додатно, за предвиђање ECF промотора чији мотиви нису унапред познати, а који могу бити присутни у малом броју високо конзервираних поновака, примењује се метод поравнавања секвенци у паровима, развијен за биоинформатичку анализу фагних генома. Поравнавање секвенци у паровима (локално и глобално) екстензивно се користи и за изучавање протеинских секвенци ECF σ фактора, конкретно, за предвиђање и компаративну анализу ДНК-интерагујућих домена (σ_2 и σ_4). Протеинске секвенце σ

фактора у бактеријским ECF подгрупама глобално се поравнавају и у већем броју, ради јасног предвиђања функционалних (конзервираних) мотива од интереса.

С друге стране, квантитативна анализа механизма функционисања у групи ECF (тј. тестирање механизма "mix-and-match") врши се само на експериментално предвиђеним промоторима канонских представника, σ^E и σ^W, помоћу корелационе анализе засноване на биофизичком моделу транскрипционе иницијације. На основу (*de novo*) поравнатих промоторских елемената конструишу се матрице тежине, које омогућавају процену енергија интеракције σ фактора са дволанчаним и једноланчаним елементима анализираних промотора. Корелисањем добијених вредности добија се систематска процена о степену функционалног надопуњавања промоторских елемената и кинетичким својствима промоторског одговора, што представља битну информацију о опсегу и начину испољавања "mix-and-match" ефекта у групи ECF σ фактора.

3.1. Скупови анализираних секвенци

3.1.1. Секвенце ДНК

За биоинформатичку анализу транскрипционе стратегије бактериофага, односно препознавање одговарајућих промоторских секвенци, коришћени су интергенски региони, издвојени из геномских секвенци 7-11, phiEco32 и Xp10, за које је анотација доступна на GenBank-у. Предвиђени промотори за фагне ECF σ факторе 7-11 и phiEco32 су даље коришћени у анализи специфицитета у ECF групи, заједно са промоторима које препознају: σ^E из *E. coli*, σ^W из *B. subtilis* и ECF σ фактори бактеријских подгрупа ECF28 и ECF32, при чему се промотори RpoD σ фактора из *E. coli* користе као добро проучена референца.

Скуп промоторских секвенци које препознаје σ^E из *E. coli* састоји се од 60 експериментално предвиђених промотора и садржи поравнате -35 и -10 елементе [98], при чему је доступна и информација о њиховој активности под in-vitro условима.

Промоторске секвенце које препознаје σ^W повучене су из базе података DBTBS, која обједињује информације о промоторима и σ факторима бактерије *Bacillus subtilis* [99].

Овај скуп секвенци састоји од 34 експериментално предвиђена промотора, од којих је један (узводно од гена *ywbLMN*) избачен из даље анализе, због тешкоће у поравнавању његовог -35 елемента (који се разликује за најмање 5 бп у односу на консензус).

Скуп RpoD промотора садржи 322 секвенце са експериментално предвиђеним почецима транскрипције, које су систематски поравнате *de novo* уз помоћ Гибсове претраге [18].

Скупови фагних ECF промотора садрже само поравнате -10 елементе, који су за бактериофаг 7-11 предвиђени биоинформатичком анализом геномске секвенце, док је за бактериофаг phiEco32 шест анализираних промотора добијено из експерименталних резултата [64].

У бактеријској подгрупи ECF28, анализирана су два скупа секвенци ДНК без унапред поравнатих промоторских елемената. Први скуп коришћен је за ненадгледану претрагу ECF28 промотора и састоји се од секвенци дужине 50 бп, локализованих узводно од гена који кодирају изабране представнике (са међусобно довољно различитим промоторским секвенцама; сви идентификациони бројеви – GenInfo Identifier [GI] приложени су у Табели M1 на крају Метода). Проширивањем првог скупа секвенци са 50 бп на пуну дужину интергенских региона добијен је други скуп секвенци за надгледану претрагу ECF28 промотора помоћу матрица тежине (видети испод поглавље 3.3.6).

У подгрупи ECF32 анализирани су промотори које препознају ECF32 чланови са конзервираним протеинским мотивом N-терминално од домена σ_4 (погледати поглавље 4.2.5 у Резултатима). Део анализираних ECF32 промотора преузет је из референце [30]. Други део промотора предвиђен је помоћу MLSA алгоритма, при чему конзервирани мотиви нису нађени у свим секвенцама па овакве секвенце нису коришћене у даљој анализи.

3.1.2. Протеинске секвенце

Скуп анализираних протеинских секвенци садржи бактеријске ECF σ факторе класификоване у 43 различите подгрупе [30], од којих су детаљно анализиране подгрупе: ECF01 која садржи σ^W; ECF02 која садржи σ^E; ECF28 која је насроднија фагном ECF σ фактору 7-11; и подгрупа ECF32.

Анализиране су такође секвенце ECF σ фактора, које кодирају геноми бактериофага 7-11 и phiEco32, а које су повучене из GenBank-а. Анализиране протеинске секвенце (~44,000 σ фактора) класификоване као чланови групе RpoD, такође су повучене из GenBank-а.

3.2. Издвајање интергенских региона из бактериофагних геномских секвенци

Интергенски региони из геномске секвенце бактериофага 7-11 издвојени су према координатама из GenBank анотације и подељени на 3 скупа секвенци, који одговарају секвенцама узводно од i) свих гена у геному, ii) гена са "+" транскрипционом орјентацијом, iii) гена са "-" транскрипционом орјентацијом. У свакој групи, орјентација директног ланца одговара смеру у ком се низводни гени транскрибују па су надаље претраживани само директни ланци издвојених интергенских региона (промоторски мотиви немају палиндромску симетрију). Узевши у обзир типичну дужину промотора, издвојени су само интергенски региони дужи од 50 бп. Будући да у извесним случајевима промоторски елементи могу да се преклопе са 3' крајем узводног гена, 5' крајеви интергенских региона продужени су узводно за 30 бп, при чему су ови дужи интергенски региони коришћени за предвиђање промоторских елемената. На исти начин су издвојени и интергенски региони из геномских секвенци бактериофага Xp10 и phiEco32, на којима је тестирана ефикасност новог метода за предвиђање промоторског специфицитета фагних σ фактора и РНКП.

3.3. Поравнања секвенци ДНК

3.3.1. Поравнања већег броја секвенци

За локална поравнања већег броја секвенци ДНК коришћена је Гибсова претрага (програм Gibbs Motif Sampler) [91] као стандардни алгоритам MLSA за *ab initio* предвиђања кратких дегенерисаних мотива. Програм Gibbs Motif Sampler коришћен је у опцијама Site Sampler и Motif Sampler. Покретањем опције Site Sampler, алгоритам проналази тачно један мотив у свакој претраживаној секвенци, што је оптимално за случај када се сматра да велика

већина анализираних секвенци садржи конзервирани мотив. С друге стране, опција Motif Sampler омогућава препознавање мотива који нису присутни у свим претраживаним секвенцама, односно узима се у обзир могућност да неке од секвенци не садрже конзервиран мотив, док друге секвенце могу садржати више понављања мотива. При овоме, алгоритам у последњем циклусу претраге, у крајњем поравнању додаје/одузима мотиве из претраживаних секвенци у зависности од доприноса информационом садржају поравнања (који се узима за стандардну меру конзервације мотива). Стога, укупан број предвиђених мотива по претраживаној секвенци у завршном поравнању може бити и већи и мањи од 1.

При сваком покретању програма Gibbs Motif Sampler претраживан је само директни ланац ДНК, при чему је број очекиваних мотива по претраживаној секвенци подешен на 1. Дужина мотива подешавана је на више различитих вредности при анализи истог скупа података (да би се проверила постојаност предвиђеног мотива), док су остали параметри били на подразумевано подешеним вредностима у алгоритму. Постојаност предвиђања фагних промотора (у геномима 7-11, phiEco32 и Xp10) тестирана је и покретањем претраге у програму BioProspector, чији се рад такође базира на Гибсовој претрази [102].

3.3.2. Предвиђање фагних промотора помоћу алгоритама MLSA

Први приступ предвиђању промоторског специфицитета фагних σ фактора је коришћење два програма MLSA, заснована на Гибсовој претрази – Gibbs Motif Sampler-a и BioProspector-a. Програм Gibbs Motif Sampler коришћен је у опцији Motif Sampler, при чему је дужина мотива подешена на 9 (за интергенске регионе бактериофага 7-11 и phiEco32) и 18 бп (за интергенске регионе бактериофага Xp10). За бактериофаге 7-11 и phiEco32 дужина мотива је изабрана у складу са дужином -10 елемента (односно збиром дужина продуженог и кратког -10 елемента) за RpoD σ факторе, будући да и RpoD и фагни σ фактори припадају протеинској фамилији σ^{70}. За бактериофаг Xp10 дужина мотива подешена је према дужини промотора РНКП коју кодира бактериофаг T7, који је сличан РНКП бактериофага Xp10. У програму BioProspector су за претрагу дефинисане исте дужине мотива као за Gibbs Motif Sampler; у оба програма је претраживан само директни

ланац ДНК. Остали параметари претраге одговарају подразмевано подешеним вредностима.

3.3.3. Предвиђање фагних промотора поравнавањем секвенци у паровима

Као алтернативна стратегија за предвиђање фагних промотора развијен је приступ који се заснива на поравнању фагних интергенских региона у паровима. Приступ је реализован помоћу BLAST-а (верзија blastn, опција blast2seq) [103]. Овај алтернативни приступ је добро прилагођен особини фагних промотора да се у геномима типично јављају у малом броју поновака, који су добро конзервирани. Програм BLAST коришћен је са унапред задатим параметрима, уз изузетак подешавања минималне дужине поравнања на 7 бп. Будући да су гени бактериофага 7-11 организовани у два дивергентно транскрибована кластера (кластери "+" и "-"), претпоставка је да би први (најузводнији) интергенски регион у барем једном кластеру требало да садржи макар један примерак фагног промотора – у супротном, гени узводно од промотора локализованих унутар генског кластера не би били транскрибовани. Сходно томе, у претрази се поравнава први интергенски регион из кластера "+" (који садржи структурне, највероватније касне гене) са самим собом и преосталим интергенским регионима. Такође, и преостали интергенски региони кластера "+" међусобно се поравнавају, ради препознавања фагних промотора у случају њиховог одсуства из првог интергенског региона структурног кластера. Као последњи корак, на свим интергенским регионима обавља се и надгледана претрага, ради препознавања и оних промоторских поновака који су због слабије конзервације или краће дужине могли бити пропуштени у иницијалном циклусу претраге.

3.3.4. Логои ДНК секвенци

За прављење логоа ДНК секвенци коришћен је програм enoLOGOS [104], као и одговарајућа функција из MATLAB-а (MathWorks). enoLOGOS и MATLAB коришћени су са подразумевано подешеним параметрима, уз изузетак GC-састава који је у програму enoLOGOS засебно подешаван за сваки анализиран скуп секвенци.

Програмом enoLOGOS су прављени логои за фагне (7-11 и phiEco32), ECF σ^E и σ^W, и RpoD -10 елементе, за које су скупови података дефинисани у почетном поглављу Метода. Поравнања промотора σ^E и σ^W прво су проверена програмом Gibbs Motif Sampler (при чему је у оба случаја добијено поравнање које се подудара са поравнањем из изворног скупа секвенци), а тек потом коришћена за логое -10 елемената уз додатак 3 узводна бп. За промоторе 7-11 лого је прављен од -35 елемената, препознатих узводно од две фагне промоторске секвенце. -35 елемент за трећи фагни промотор изостављен је из логоа због изражене дегенерације, која значајно нарушава поравнање. Ради приказа неканонских интеракција између ECF σ фактора и њихових промотора, логои су прављени у MATLAB-у за промоторе σ^E *неактивне* in-vitro, промоторе σ^W и промоторе представника групе ECF32. За σ^W и σ^E промоторе лого је направљен од одговарајућих -10 елемената, заједно са секвенцом спејсера, која се протеже до низводне границе -35 елемента, док је за промоторе ECF32 лого прављен за читаву секвенцу промотора.

3.3.5. Предвиђање конзервације у узводним секвенцама фагних phiEco32 промотора

Да би се испитало присуство конзервираног мотива у свим фагним phiEco32 промоторима, узводне промоторске секвенце анализиране су помоћу програма Gibbs Motif Sampler (опција Site Sampler), у складу са претходно описаном процедуром. Дужина мотива подешена је на 8, 7 и 6 бп (што одговара типичној дужини -35 елемента у σ^{70} фамилији) да би се проверила постојаност предвиђеног мотива.

3.3.6. Предвиђање промотора у подгрупи ECF28

Ненадгледана (*ab initio*) претрага била је први избор за предвиђање промотора у подгрупи ECF28; претходно дефинисан скуп секвенци ДНК анализиран је програмом Gibbs Motif Sampler у обе опције (Site Sampler и Motif Sampler; погледати објашњење изнад), при чему је дужина мотива подешена на 7, 6 и 5 бп (што одговара типичној дужини -10 елемента) зарад провере постојаности пронађених мотива. Алтернативно, исти скуп анализиран је

такође ненадгледаном претрагом, али помоћу метода поравнавања интергенских региона у паровима (видети 3.3.3 изнад).

Други избор за препознавање промотора у подгрупи ECF28 била је надгледана претрага помоћу матрица тежине за -35 елемент σ^W (рачунање елемената матрице описано је у поглављу 3.5). Мотиви са најбољим скором у свакој од анализираних секвенци означени су као -35 елементи, при чему су низводни сегменти издвојени за даљу претрагу програмом Gibbs Motif Sampler, у циљу предвиђања -10 елемената.

3.3.7. Предвиђање промотора у подгрупи ECF32

За чланове подгрупе ECF32 у чијим протеинским секвенцама је предвиђен конзервирани мотив, а за које нису доступне промоторске секвенце у реф. [30], специфицитет је предвиђен помоћу алгоритама MLSA. Коришћен је програм Gibbs Motif Sampler у опцији Motif Sampler, при чему је дужина мотива подешена на типичну дужину -10 елемента, за претраживање секвенци које се састоје од 100 узводних бп, у односу на почетке транслације одговарајућих гена за ECF σ факторе.

3.4. Поравнања протеинских секвенци

За глобална поравнања већег броја протеинских секвенци коришћен је програм ClustalW [92], за поравнања у паровима BLAST, а за препознавање домена CD-Search [105]. Програм ClustalW коришћен је са подразумеваним вредностима параметара у алгоритму. Програм BLAST коришћен је у верзији blastp са опцијом за поравнавање две или више секвенци (осталим параметрима нису мењане унапред задате вредности). Програм CD-Search коришћен је такође са унапред задатим параметрима, са изузетком прага за E-вредност, који је у више корака спуштан (до крајње вредности 10) при предвиђању домена σ_4 у протеинској секвенци фагног σ фактора phiEco32.

При утврђивању сличности између фагних ECF (7-11 и phiEco32), канонских бактеријских ECF (σ^W и σ^E), и σ фактора RpoD из *E. coli*, протеинске секвенце анализиране су програмом CD-Search, након чега су предвиђени домени (σ_2 и σ_4) поравнати помоћу BLAST-а, уз изузетак фагног σ фактора phiEco32 за који је цео домен ECF коришћен у анализи.

За утврђивање сличности између фагног σ фактора 7-11 и припадника група ECF и RpoD (скупови секвенци дефинисани у поглављу 3.1.2), протеинске секвенце поравнате су у паровима помоћу програма BLAST, по претходно описаној процедури. При упоређивању фагног σ фактора 7-11 са σ факторима из групе ECF, анализа је засебно спроведена за сваку од 43 различите подгрупе.

Фагни σ фактор 7-11 додатно је упоређиван са одабраним представницима из подгрупа ECF01 и ECF28 (дати у Табели М1) глобалним поравнавањем већег броја секвенци у програму ClustalW. При поређењу са подгрупом ECF01, скуп анализираних представника сужен је на σ факторе чије су протеинске секвенце најсличније ECF σ фактору 7-11 (што је процењено на основу E-вредности поравнања из BLAST-а). С друге стране, при поређењу са подгрупом ECF28, скуп секвенци је сужен тако да садржи по једног представника из сваког бактеријског рода, који се јавља у подгрупи ECF28. Глобалним поравнањем анализирана је и целокупна подгрупа ECF28, након чега је направљен лого (помоћу програма enoLOGOS) од сегмената, који се пружају од C-терминуса домена σ_2 до појаве прве празнине у поравнању (што је у непосредној близини N-терминуса домена σ_4). Истим поступком (глобално поравнање програмом ClustalW) анализиране су протеинске секвенце одабраних представника подгрупа ECF02 и ECF32, при чему је за обе подгрупе направљен лого помоћу програма enoLOGOS за секвенце у којима је при поравнању препознат конзервирани мотив. При прављењу логоа секвенцама предвиђених мотива прикључени су и околни низводни сегменти, ради бољег уочавања степена конзервације предвиђеног мотива.

3.5. Конструисање матрица тежине

Сем за надгледану претрагу промотора, матрице тежине користе се и за корелациону анализу канонских ECF промотора. С тим у вези, матрице су дефинисане за промоторске секвенце које препознају фактори ECF σ^E и σ^W (скупови секвенци описани у поглављу 3.1.1), при чему су за σ^E коришћена поравнања -35 елемента, продуженог и кратког -10 елемента, као и поравнање мотива предвиђеног у спејсеру (видети поглавље 4.2.5 у Резултатима), док је за σ^W коришћено само поравнање -35 елемента. Матрице су, такође, дефинисане за промоторе бактеријског RpoD σ фактора.

Матрице за бактеријски RpoD σ фактор, које се користе за претрагу промотора у геномској секвенци бактериофага 7-11, конструисане су ради веће тачности претраге на основу новог, прецизнијег поравнања RpoD промоторских елемената *E. coli* [18]. Ово ново поравнање омогућава: i) прецизно дефинисање матрица тежине за -10 и -35 промоторске елементе, ii) узимање у обзир различитих скорова који одговарају променљивом растојању између наведених елемената, iii) формирање засебне матрице тежине за продужени -10 елемент (тј. -15 елемент, ако се у матрицу укључи и најузводнија база -10 елемента [18-19]. Праг претраге за предвиђања се емпиријски одређује, што је за предивиђања RpoD промотора у геному 7-11 детаљније продискутовано у поглављу 4.1.2 у Резултатима.

Елементи матрице тежине – $w_{\alpha,i}$, који дефинишу квантитативни допринос базе α на позицији i у мотиву, рачунају се на основу израза [80, 106]:

$$w_{\alpha,i} = \log\left(\frac{nv_{a,i} + p_a}{p_a(n+1)}\right) \tag{3.1}$$

где је са n означен број мотива у поравнању по ком се матрица формира, а са $v_{a,i}$ учесталост појављивања базе α на позицији i у поравнању. Са p_a је означена позадинска учесталост базе, док члан p_a у бројиоцу (тзв. "псеудоброј") представља корекцију за мале вредности n (видети Увод 1.4.1.1).

Тежине које одговарају различитим дужинама спејсера рачунају се у складу са изразом [18]:

$$w_i = \log(v_i) \tag{3.2}$$

при чему w_i означава тежину спејсера дужине i, док v_i означава учесталост са којом се у поравнању појављује спејсер дате дужине.

Финално, од сваке колоне у матрици одузима се вредност за базу која се на датој позицији појављује у консензусној секвенци мотива, чиме се постиже да скор матрице за консензус има вредност 0. Остали скорови имају негативне вредности, при чему вредност ближа 0 означава бољи скор.

3.6. Корелациона анализа снага промоторских елемената

Пирсонови корелациони коефицијенти (на другим местима у тексту само корелациони коефицијенти) и њима одговарајуће P-вредности израчунате су у MATLAB-у по процедури која се заснива на насумичној пермутацији тачака из анализираног скупа података. Корелациони коефицијенти рачунају се за сваку насумичну пермутацију, при чему се статистички значај за разлику између корелационих коефицијената у изворном и пермутованом скупу података процењује помоћу t-теста.

Табела М1. GI бројеви σ фактора коришћених у студији

Подгрупа ECF28 секвенце коришћене за вишеструко поравнање са фагним 7-11 σ фактором; претрага ECF28 промотора	**Подгрупа ECF01** секвенце коришћене за вишеструко поравнање са фагним 7-11 σ фактором	**Подгрупа ECF02** секвенце коришћене за вишеструко поравнање	**Подгрупа ECF32** секвенце коришћене за прављење протеинског логоа; секвенце у *italic-у* одговарају σ факторима са доступним промоторским секвенцама (од којих је конструисан промоторски лого)
119944707	71281354	120609877	*50121015*
78486112	109900043	121595577	*4581629*
24374616	114562024	115351057	*9885630*
71279207	52784027	194563837	*28868612*
88795792	16077241	107022203	*51103045*
109897977	56418685	78065711	*76574797*
54308153	67940793	53725364	42601254
28901545	51893181	53720042	8515859
	51891307	83718976	21311396
	148657909	33602727	
	134299216	33593421	
	121594756	16130498	
	94971553	16272571	
	116626287	85712966	
	60681313	114320495	
	108763948	121996850	
	146300502	110834500	
	90023535	88703559	
		83644624	
		67153338	

4. РЕЗУЛТАТИ

4.1. Биоинформатичка анализа транскрипционе стратегије бактериофага 7-11

Бактериофаг 7-11, чији је домаћин ентеробактерија *Salmonella enterica* [107], погодан је модел систем за процену ефикасности биоинформатичких метода у предвиђању вирусне транскрипционе стератегије током инфекције, услед недостатка експерименталних информација о механизмима којим се регулише генска експресија. Геном бактериофага 7-11, који се састоји од ~90 кбп дволанчане ДНК, кодира 30 гена на директном (транскрипциона орјентација "+") и 121 ген на реверзном ланцу (транскрипциона орјентација "-"). Гени исте транскрипционе орјентације организовани су у неиспрекидане кластере, који су раздвојени дугачким узводним интергенским регионом. Кластер на директном ланцу састоји се од структурних гена, који се код већине бактериофага експримирају касније у инфекцији и учествују у изградњи вирусних партикула, док кластер на реверзном ланцу чине функционални гени, са улогом у одржавању и експресији генома. Међу њима, гени који кодирају σ фактор и анти-σ фактор код већине бактериофага представљају кључне елементе у регулацији транскрипционе активности, будући да доводе до потпуне блокаде експресије бактеријских гена, уз преусмеравање транскрипционе машинерије на фагне промоторе, локализоване узводно од дела вирусних гена [100]. Стога, предвиђање промоторског специфицитета за фагне σ факторе представља окосницу за реконструкцију обрасца генске експресије током инфекције, где битан додатни извор информација дају предвиђања за промоторе σ фактора бактерије домаћина, као и функционална анотација гена предвиђених у фагном геному.

Услед ниске ефикасности стандардних биоинформатичких метода (алгоритми MLSA) за предвиђање промоторског специфицитета фагних σ фактора [69], овом проблему неопходно је приступити са алтернативно конципираном стратегијом за претрагу, због чега могућност потврде добијених биоинформатичких предвиђања представља њен врло значајан аспекат. За бактериофаг 7-11 могућност провере добијених предвиђања последица је високог нивоа сличности σ фактора, који кодира, са σ фактором колифага

phiEco32, чији је промоторски специфицитет одређен детаљном експерименталном анализом [64]. Додатно, бактериофаг phiEco32, заједно са још једним експериментално изученим бактериофагом Xp10, може послужити за процену ефикасности алтернативне стратегије претраге, при чему је на оба примера већ показано да стандардна биоинформатичка анализа не доводи до валидних предвиђања промоторског специфицитета фагних σ фактора/РНКП [64, 68].

Сем за директну биоинформатичку анализу вирусне транскрипционе стратегије, бактериофаг 7-11 погодан је модел систем и за изучавање промоторског специфицитета у групи ECF, будући да је ECF σ фактор 7-11 сличнији бактеријским припадницима групе у односу на експериментално изучени ECF σ фактор phiEco32, што га чини подеснијим за компаративну анализу. ECF σ фактор бактериофага 7-11 можемо сматрати кариком која повезује најдивергентније чланове групе (ECF σ фактор phiEco32), за које се са највећом вероватноћом очекује уочавање квалитативно другачијих регулаторних парадигми, и бактеријске представнике, чије је изучавање до сада било засновано на ограниченим информацијама о специфицитету малог броја канонских (експериментално анализираних) представника. Стога, анализа промоторског специфицитета за σ факторе кодиране геномима 7-11 и phiEco32 може бити извор значајних информација о механизму иницирања транскрипције у важној групи алтернативних σ фактора, али и окосница ефикасније анализе инфективног циклуса за растући број бактериофага са секвенцираним геномима.

4.1.1. Предвиђање фагних ECF промотора у геномској секвенци бактериофага 7-11

Први избор за предвиђање промотора фагног ECF σ фактора у геномској секвенци 7-11 била је стандардна претрага заснована на алгоритмима MLSA. Услед недостатка процене статистичког значаја [102, 108] уобичајени критеријум веродостојности MLSA резултата јесте постојаност мотива, које за исти скуп података пријављују различити програми конципирани на истом типу алгоритма, или различити циклуси претраге у случају коришћења само једног програма. За предвиђање ECF промотора у геномској секвенци 7-11 примењен је први приступ – два програма заснована на Гибсовој претрази, Gibbs Motif

Sampler и BioProspector, независно су претраживали скупове интергенских региона дивергентне транскрипционе орјентације. Ни у једном скупу секвенци претрага није предвидела постојане мотиве, што је очекивана последица ниске учесталости са којом се фагни промотори јављају у геномској секвенци вируса, због чега у MLSA претрагама лако бивају маскирани случајним мотвима, који се са већом учесталошћу јављају у геному.

Као алтернатива алгоритмима MLSA, за предвиђање промоторског специфицитета фагних σ фактора осмишљен је приступ где се интергенски региони поравнавају у паровима, а који је боље прилагођен за предвиђање мотива код којих мали број поновака прати и мања дегенерација секвенце, што је готово универзална особина промотора за σ факторе кодиране фагним геномима [69]. Конкретно, приступ подразумева поравнавање интегренских региона структурног (најчешће касно експримираног) кластера, услед очекивања да његов најузводнији интергенски регион садржи барем један примерак фагног промотора који обезбеђује експресију структурних гена у каснијим фазама инфекције, након блокаде бактеријског холоензима. Овде треба приметити да је најузводнији интергенски регион структурног кластера управо онај који раздваја дивергентно транскрибоване гене (са орјентацијама "+" и "-") и који најчешће садржи промоторе за експресију дивергентно транскрибованих кластера (архитектура карактеристична за добро проучени бактериофаг λ [109-110]).

Описаним приступом предвиђена су четири потенцијална ECF промотора у интергенском региону узводно од структурног кластера (Табела 1) дужине 12 бп, која почињу динуклеотидом "TG", за којим следи централни конзервирани сегмент "TGATGT" и продужетак у виду елемента "TATA". Предвиђена четири поновка дужине 12 бп статистички су високо значајна, будући да им је E-вредност процењена на $\sim 10^{-8}$ за секвенцу дужине ~4400 бп, што одговара укупној дужини интергенских региона у структурном генском кластеру. Централни конзервирани мотив предвиђених промотора искоришћен је за додатни циклус претраге на свим интергенским регионима, с циљем предвиђања промоторских секвенци које су због краће дужине могле бити пропуштене у иницијалној претрази. Овим приступом предвиђено је шест додатних мотива, статистички такође значајних ($P < 0.05$), који су лоцирани узводно од гена датих у Табели 1. Узевши у обзир значајну разлику у дужини, промотори предвиђени у почетном и накнадном циклусу претраге класификовани су, редом, као дуги и кратки поновци.

Табела 1. Предвиђени промоторски елементи за фагни ECF σ фактор 7-11. У првој колони приказан је број низводног гена, потом следи секвенца предвиђеног мотива, док је у последњој колони дата геномска координата која одговара почетку мотива - 5' краја "TG" сегмента за дуге мотиве и централног елемента "TGATGT"' за кратке мотиве

Ген	Предвиђени мотив	Геномска координата
Дуги поновци - "касни" промотори		
1	**tgtgatgttata**	995 бп
1	**tgtgatgttata**	1091 бп
1	**tgagatgttata**	1056 бп
1	**ggggatgttata**	844 бп
Кратки поновци – "средњи" промотори		
1	**tgatgt**gtaa	1741 бп
1	**tgatgt**agtc	91 бп
25	**tgatgt**ttgg	26138 бп
88	**tgatgt**atct	33069 бп
116	**tgatgt**agac	17568 бп
122	**tgatgt**aact	12935 бп

Две групе предвиђених мотива (дуги и кратки) највероватније су две засебне класе промотора, на шта указује и јасна разлика у геномској локализацији. Наиме, дуги поновци предвиђени су искључиво узводно од структурног кластера, највероватније експримираног у каснијим фазама инфекције, услед чега су ови мотиви означени као "касни" промотори. С друге стране, кратки мотиви сем структурних транскрибују и низводне гене функционалног кластера, који се најчешће експримирају пре касних, а након гена транскрибованих бактеријским холоензимом, услед чега су ови мотиви означени као "средњи" промотори. На крају, важно је нагласити да су сви предвиђени мотиви локализовани низводно од гена који кодира анти-σ фактор, што је у складу са очекивањем да транскрипција регулисана фагним σ фактором може да отпочне тек након блокаде активности бактеријског холоензима, а коју узрокује управо анти-σ фактор.

4.1.2. Предвиђање бактеријских RpoD промотора у геномској секвенци бактериофага 7-11

Након предвиђања промоторског специфицитета за фагни ECF σ фактор у геномској секвенци 7-11, уследило је предвиђање промотора које препознаје бактеријски холоензим уз помоћ побољшаних матрица тежине за RpoD σ фактор. Као критеријум за дефинисање прага претраге, чијим се преласком секвенце сврставају у категорију предвиђених промотора, послужило је очекивање да бактеријски RpoD промотори треба да буду локализовани на реверзном ланцу ДНК. Овај критеријум је последица могућности бактеријског σ фактора да регулише експресију само раних функционалних гена, будући да анти-σ фактор, као један од производа раних гена, касније у инфекцији блокира активност бактеријског холоензима.

Табела 2. Предвиђени промоторски елементи за бактеријски RpoD σ фактор. У првој колони приказан је број низводног гена, што је праћено -35 елементом, дужином спејсера и -10 елементом предвиђених промотора; потом следи геномска координата која одговара почетку мотива (5' крају -35 елемента), док је у последњој колони дат скор матрице тежине за свако предвиђање. Позиције на којима се предвиђени промотрски елементи подударају са одговарајућом конснезусном секвенцом обојене су црвено.

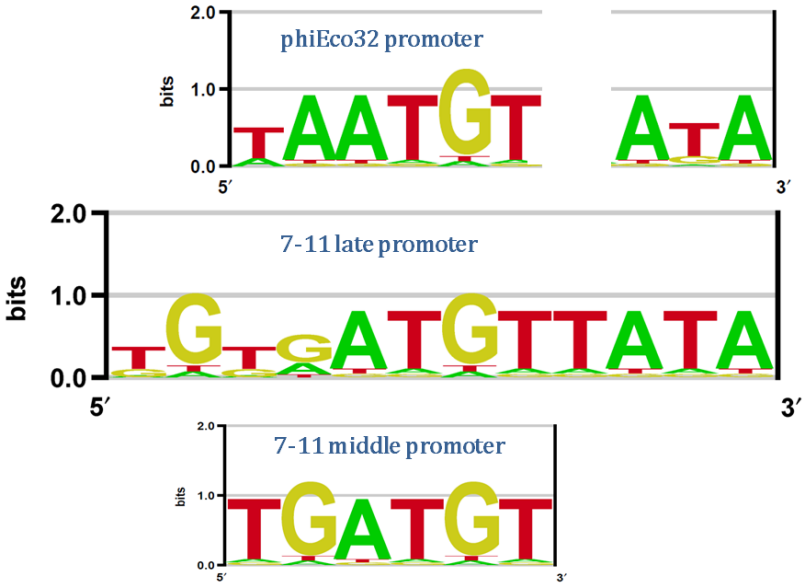
Ген	-35 елемент **ttgaca**	Дужина спејсера	-10 елемент **tataat**	Геномска координата	Скор мотива
151	ttgaca	18 бп	tatagt	89788 бп	-2.78
151	ttgaca	18 бп	taatct	89695 бп	-2.94
151	ttgcaa	18 бп	taatat	89601 бп	-3.24
151	ttgccg	18 бп	tagagt	887 бп	-3.27
94	atgaaa	19 бп	tacaat	30132 бп	-3.66
96	ttgcct	17 бп	tatatt	28300 бп	-3.96
151	ttagta	18 бп	taaaat	89566 бп	-3.98

Описаном претрагом предвиђено је осам бактеријских RpoD промотора на реверзном ланцу ДНК (Табела 2), који су локализовани непосредно узводно или унутар функционалног генског кластера, укључујући и позицију узводно од гена за анти-σ фактор. Очекиван просторни профил добијених предвиђања значајна је потврда њихове функционалности. Додатно, у Табели 2 уочава се висока сличност два најбоља предвиђања са консензусним секвенцама одговарајућих RpoD промоторских елемената, што указује да активност ова два промотора има највећи удео у транскрипцији раних гена, тј. да се функционални кластер експримира као дугачак оперон. Преостали промотори могу додатно да појачају транскрипциону активност фагних гена, која је типично знатно јача у поређењу са стопом транскрипције бактеријских гена, при чему промотори

локализовани унутар самог кластера највероватније прецизно нивелишу експресију специфичних гена (нпр. гена за анти-σ фактор).

4.1.3. Поређење генома бактериофага 7 - 11 и phiEco32

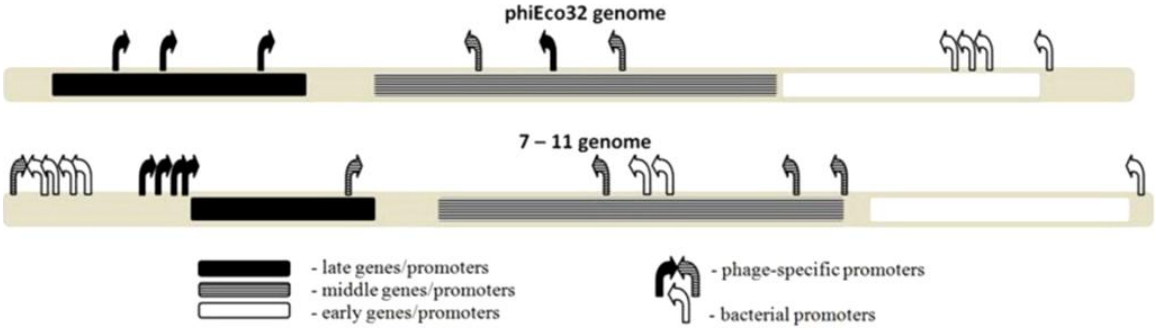
Претходно је истакнуто да бактериофази 7-11 и phiEco32 кодирају сопствене σ и анти-σ факторе, као и да постоји висок степен сличности између њихових ECF σ фактора. Сличност између бактериофага 7-11 и phiEco32 присутна је и на нивоу организације геномске секвенце, о чему сведоче еквивалентне позиције гена у припадајућим кластерима. Стога се потврђивању предвиђања за фагне промоторе у геномској секвенци 7-11 може приступити путем компаративне анализе са бактериофагом phiEco32, чија је транскрипциона регулација експериментално детаљно окарактерисана.



Слика 5. Поређење логоа фагних промотора: На слици су, од врха до дна, редом приказани логои следећих промоторских секвенци: i) експериментално предвиђених фагних промотора у геномској секвенци бактериофага phiEco32 [6], ii) предвиђених дугих мотива ("касни" фагни промотори) у геномској секвенци бактериофага 7-11 (Табела 1), iii) предвиђених кратких мотива ("средњи" фагни промотори) у геномској секвенци бактериофага 7-11 (Табела 1). Логои промоторских секвенци су поравнати, при чему је на једној позицији у логоу промотора phiEco32 уведена празнина ради лакшег уочавања сличности у специфицитету бактериофагних ECF σ фактора.

На Слици 5 приказано је поређење логоа за предвиђене промоторе ECF σ фактора 7-11 и експериментално одређене промоторе ECF σ фактора phiEco32. Премда само поравнање није тривијално (уочити уметнуту базну позицију у логоу phiEco32 промотора), сличност

промоторских специфицитета је очигледна, што представља јаку додатну потврду промоторске активности за предвиђене поновке у геному 7-11.



Слика 6. Поређење распореда промотора у геномској секвенци и груписање гена по временском обрасцу експресије током инфекција код бактериофага phiEco32 и 7-11: Геномске секвенце бактериофага phiEco32 и 7-11, са назначеним распоредом предвиђених промотора и временским обрасцем генске експресије, приказане су у горњем и доњем реду на слици; Боје којима су кодиране различите класе промотора и групе заједнички експримираних гена назначене су у легенди слике.

Сем по специфицитету, ECF промотори бактериофага 7-11 и phiEco32 слични су и по геномској локализацији, што се јасно уочава на Слици 6. У геномима оба бактериофага структурни гени транскрибују се са касних ECF промотора; узводни сегмент функционалног генског кластера транскрибује се са бактеријских RpoD промотора, док низводни сегмент истог кластера транскрибују средњи ECF промотори. Еквивалентна геномска локализација промотора додатно потврђује предвиђени специфицитет за ECF σ фактор 7-11, али и поделу промоторских секвенци на средњу и касну класу. Овде је битно нагласити да је предвиђање специфицитета фагних ECF промотора у геномској секвенци 7-11 вршено без претходног коришћења било каквих информација о транскрипционој регулацији бактериофага phiEco32; тј. ова информација је коришћена искључиво за потврду већ добијених предвиђања.

4.1.4. Предвиђање фагних промотора у геномским секвенцама бактериофага phiEco32 и Xp10

У претходном поглављу представљени су резултати компаративне анализе, који потврђују предвиђени промоторски специфицитет фагног ECF σ фактора 7-11. Будући да је препознавање ECF промотора у геномској секвенци 7-11 резултат новог приступа за анализу специфицитета фагних σ фактора/РНКП, погодност овог приступа тестирана је на геномима експериментално анализираних бакериофага, Xp10 и phiEco32. Као и за

бактериофаг 7-11, паралелно су тестирани и стандардни MLSA алгоритми, што омогућава потпунију процену значаја новог приступа за предвиђање промоторског специфицитета у фагним геномима.

Интергенски региони бактериофага phiEco32, који припадају кластерима "+" и "-", првобитно су засебно, а потом и заједно, анализирани уз помоћ програма BioProspector и Gibbs Motif Sampler, при чему је добијено да за сва три анализирана скупа секвенци поновљена Гибсова претрага пријављује мотиве врло различите по специфицитету. Постојана предвиђања добијају су тек по формирању скупа за претрагу од секвенци интергенских региона кластера "+" и интергенских региона кластера "-", за које мерења експресије генским чипом указују припадност средњој временској класи фагних гена. Будући да мотив пријављен у овом скупу секвенци одговара консензусу експериментално одређених ECF промотора – "TAATGTATA", јасно је да алгоритми MLSA дају валидна промоторска предвиђања за phiEco32 искључиво уз ослањање на резултате експерименталних анализа. С друге стране, приступ заснован на поравнавању интергенских региона у паровима самостално доводи до предвиђања мотива дужине 9 бп – "tAATGTAtA", узводно од гена: 6, 13, 26, 40, 58 и 68 (Табела 3), а који се поклапа се експериментално предвиђеним ECF промоторима у геномској секвенци phiEco32 [64].

Табела 3. Предвиђени промоторски елементи за фагни ECF σ фактор phiEco32. У првој колони приказан је број низводног гена, потом следи секвенца предвиђеног мотива, док је у последњој колони дата геномска координата која одговара почетку (5' крају) предвиђеног мотива. У консензусној секвенци, која одговара предвиђеним промоторима, апсолутно конзервиране позиције приказане су великим словима, док су позиције на којима се јавља (једно) одступање приказане малим словима. Позиције које се у предвиђеним промоторима подударају са консензусном секвенцом обојене су црвено.

Ген	Фагни промотор phiEco32	Геномска координата
	tAATGTAtA	
6	taatgtaga	1677 бп
13	taatgtata	9101 бп
26	taatgtata	26014 бп
40	taatgtata	40915 бп
58	taatgtata	34610 бп
68	aaatgtata	30685 бп

У геному бактериофага Xp10 алгоритми MLSA такође не дају постојана предвиђања за фагне промоторе, које препознаје РНКП овог бактериофага. Слични резултати доступни су и у студији [68], у којој су фагни промотори успешно предвиђени тек након

укључивања експерименталне информације о временском обрасцу генске експресије током инфекције. Насупрот претходном, поравнавањем интергенских региона у паровима јасно се издвајају два статистички високо значајна, идентична поновка дужине 43 бп, која су локализована у дугом интергенском региону, узводно од структурног генског кластера. Ова предвиђања су у складу са парадигмом да бактериофагни геноми типично поседују мали број високо конзервираних фагних промотора, који транскрибују низводне гене као део врло дугих оперона. Експериментални подаци такође потврђују функционалност добијених предвиђања, будући да два Xp10 промотора дужине ~20 бп фигуришу као саставни део поновака предвиђених поравнавањем интергенских региона у паровима.

Сумарно, добијени резултати потврђују значајно бољу прилагођеност новог приступа за предвиђање промоторског специфицитета фагних σ фактора и РНКП, у поређењу са стандардним приступом. Конкретно, поравнавање у паровима интергенских региона бактериофага phiEco32 и Xp10 довело је до тачног препознавања експериментално предвиђених фагних промотора, што је за стандардне методе (алгоритми MLSA) могуће само уз претходно коришћење експерименталне информације о временском обрасцу експресије фагних гена током инфекције.

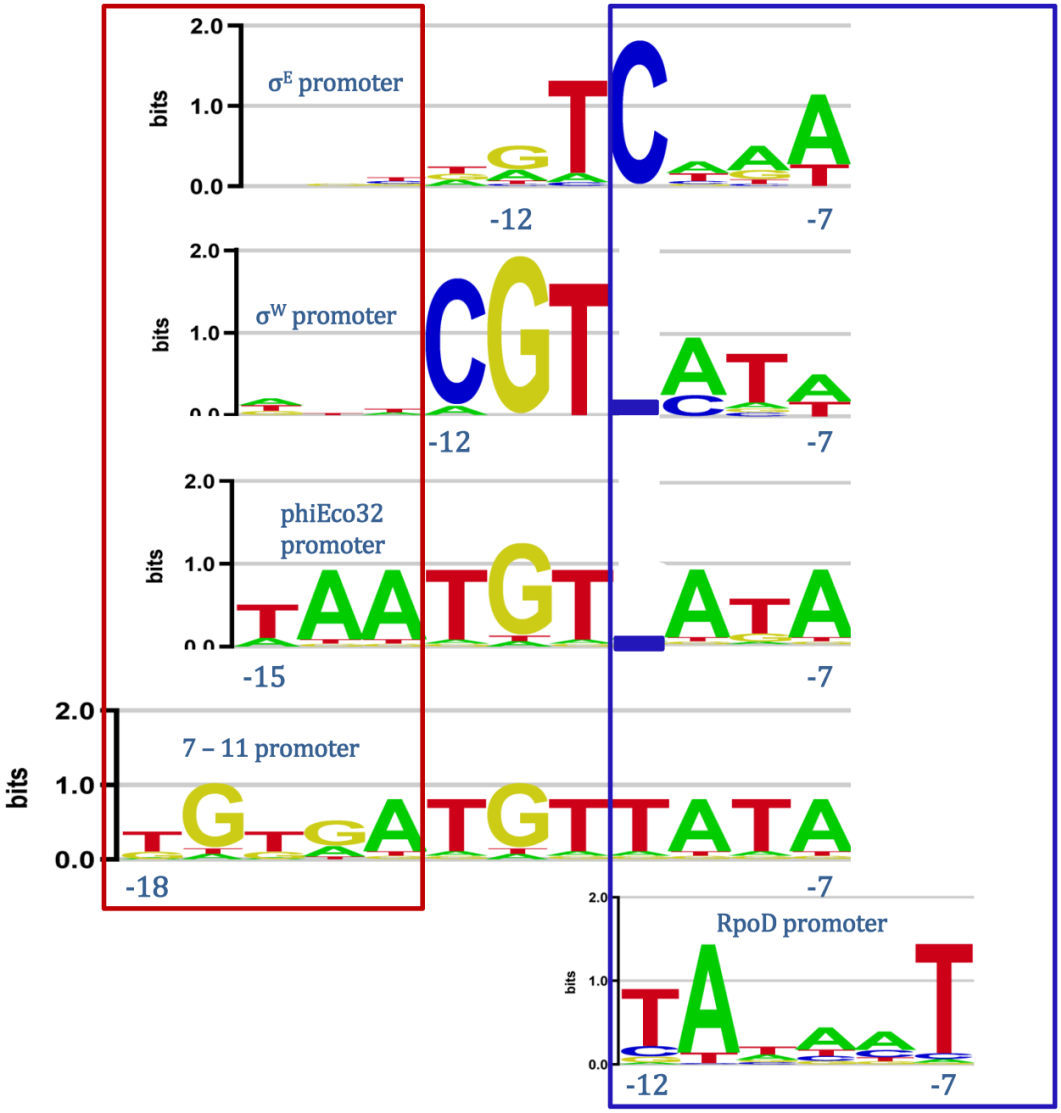
4.2. Промоторски специфицитет у групи ECF σ фактора

Тренутни закључци о механизму препознавања промоторских секвенци у групи ECF последица су проучавања свега неколицине бактеријских представника, због чега бактериофагни ECF σ фактори 7-11 и phiEco32 представљају добру полазну тачку за независну анализу промоторског специфицитета у групи. Конкретно, поређење промотора које препознају фагни ECF σ фактори, као најразличитији чланови групе, са промоторима добро проучених ECF представника (σ^E из *E. coli*, σ^W из *B. subtilis*) омогућило би уочавање потенцијалних разлика у функционисању, које би указале на одступање од парадигме о ригидним интеракцијама ECF σ фактора са промоторском секвенцом. Последица оваквог резултата била би опсежна систематска анализа специфицитета у групи, са тежиштем на препознавању протеин-ДНК интеракција путем којих ECF σ фактори иницирају

транскрипцију, односно ревизију парадигме о промоторском специфицитету. Важан допринос ревизији тренутне парадигме може да пружи и поређење ECF промотора са специфицитетом RpoD σ фактора, који остварују флексибилне интеракције са промоторима и представљају најудаљеније чланове фамилије σ^{70} у односу на σ факторе групе ECF. Наиме, овакво поређење нам омогућава да испитамо да ли промотори различитих σ^{70} група имају еквивалетну организацију промоторских елемената и одговарајућих интеркација са σ фактором. Ово би указало на сличност у механизму препознавања ових промотора, односно на чињеницу да, насупрот тренутним претпоставкама, одлучујући фактор механизма препознавања σ^{70} промотора нису протеинске секвенце интерагујућих σ^{70} фактора, већ заједнички биофизички механизам иницирања транскрипције.

4.2.1. Компаративна анализа промоторског специфицитета репрезентативних фактора σ^{70} фамилије

На Слици 7 приказано је поравнање -10 елемената које препознају бактеријски (σ^E и σ^W) и бактериофагни (phiEco32 и 7-11) ECF σ фактори, као и канонски представник групе RpoD. На поравнању се јасно уочава да је сличност између бактеријских и фагних ECF -10 елемената ограничена на низводне сегменте (упоредити промоторски лого за σ^W и фагне ECF σ факторе). У узводним деловима бактеријски и фагни -10 елементи се значајно разликују, што се најбоље види на узводним сегментима фагних -10 елемената, који су знатно дужи него одговарајући сегменти σ^W и σ^E промотора. Конкретно, на Слици 7 примећујемо одсуство конзервације у секвенцама узводно од -10 елемената σ^E и σ^W, која је са друге стране врло изражена у узводним сегментима фагних -10 елемената (уочити секвенце уоквирене црвеном бојом на слици).



Слика 7. Поравнање логоа промоторских -10 елемената које препознају различити σ^{70} фактори: На поравнању су приказани логои -10 елемената које препознају, редом: бактеријски ECF σ фактори (σ^E и σ^W), фагни ECF σ фактори (бактериофага phiEco32 и 7-11) и σ фактор Групе I (RpoD из *E. coli*). Логои са -10 елементима које препознају σ^E и σ^W продужени су за узводна 3 бп да би се уочило одсуство конзервације у овим продужењима. Елемент TATA, који је саставни део промотора Групе I (RpoD) и аналогни промоторски сегменти у ECF σ промоторима означени су плавим правоугаоником. Секвенце које представљају узводна продужења -10 елемента у фагним промоторима и њима одговарајуће позиције у промоторима σ^E и σ^W означене су црвеним правоугаоником. Координате узводних и низводних граница промоторских елемената дате су у односу на удаљеност од места почетка транскрипције.

Продужење фагних -10 елемената највероватније има функцију еквивалентну установљеној у групи RpoD, где дати продужетак (тзв. -15 елемент [18]) може да компензује одсуство -35 елемента, што је класично обележје механизма "mix-and-match" у RpoD промоторима [19]. На сличност у функционисању RpoD и фагних ECF промотора упућују и низводни сегменти њихових -10 елемената, богати нуклеотидима T и A, што

представља још једно од класичних обележја RpoD промоторског специфицитета [10]. Узевши у обзир претходно, неопходно је испитати организацију фагних ECF промотора (нарочито у погледу -35 елемента), са чим је уско повезано и налажење мотива у протеинској секвенци, који препознаје неуобичајено дуге продужетке фагних -10 елемената. Додатно, за препознати протеински мотив потребно је утврдити у којој мери је заступљен у преосталим члановима ECF групе, што омогућава почетну процену опсега одступања ECF σ фактора од парадигме о ригидним интеракцијама са промоторским секвенцама. С тим у вези, даља анализа промоторског специфицитета у групи ECF усмерена је на потрагу за класичним примерима механизма "mix-and-match" (компензација одсуства -35 елемента продуженим -10 елементом) у промоторима ECF σ фактора. На ово се директно надовезује детаљна анализа протеин-ДНК интеракција, преко којих се иницира транскрипција у групи ECF, која сем предвиђања протеинских мотива за препознавање продужених -10 елемената, носи потенцијал за предвиђање до тада непрепознатих елемената промоторског специфицитета у најобимнијој и најдивергентнијој групи σ^{70} фактора.

4.2.2. Компаративна анализа протеинских секвенци одабраних σ^{70} фактора

Након поређења промоторских -10 елемената за одабране σ^{70} факторе уследила је компаративна анализа њихових протеинских секвенци да би се испитало да ли су сличности уочене на протеинском и ДНК нивоу међусобно конзистентне. Поређењу протеинских секвенци фагних σ фактора (7-11 и phiEco32) са добро проученим бактеријским представницима σ^{70} фамилије (σ^{W}, σ^{E} и RpoD) претходило је предвиђање ДНК-везујућих домена (σ_2 и σ_4) у анализираним секвенцама, који су као носиоци промоторског специфицитета били главни предмет компаративне анализе. Важно је нагласити да у протеинској секвенци ECF σ фактора phiEco32 домен σ_4 није предвиђен чак ни уз значајно спуштање прага претраге, услед чега је целокупан ECF домен phiEco32 коришћен у анализи.

Уопштено, резултати поређења за домене σ_2 и σ_4, сумирани у Табелама 4 и 5, међусобно су конзистентни. Анализа јасно указује да су фагни σ фактори врло различити од RpoD σ фактора *E. coli*, али и од σ^E који припада ECF σ групи. Статистички значајана поравнања добијају се једино између фагних ECF σ фактора и σ^W, што је у складу са већом сличношћу њихових -10 елемената. При том, за ECF σ фактор 7-11 утврђена је већа сличност са σ^W, у односу σ фактор phiEco32, што је очекиван резултат узевши у обзир одсуство домена σ_4 у протеинској секвенци ECF σ фактора, који кодира phiEco32. Сходно томе, σ фактор phiEco32 (као најдивергентнији представник групе ECF) могао би дати пример другачије регулаторне парадигме, при чему се његов механизам функционисања може поредити са осталим члановима групе ECF преко σ фактора 7-11 (као сличнијег бактеријским ECF σ факторима).

Табела 4. Поређења домена σ_2 за одабране σ^{70} факторе.

	E-вредност за поравнања домена σ_2			
домен σ_2	*E. coli* RpoD	ECF 7-11	ECF phiEco32	*E. coli* σ^E
ECF 7-11	0.046			
ECF phiEco32	без поравнања	$2e^{-7}$		
E. coli σ^E	$4e^{-6}$	0.86	0.61	
B. subtilis σ^W	$6e^{-4}$	$1e^{-6}$	$3e^{-5}$	$2e^{-15}$

Табела 5. Поређења домена σ_4 за одабране σ^{70} факторе. (X) - За фагни σ phiEco32 фактор цео ECF домен (уместо само σ_4 домена) је коришћен у анализи.

	E-вредност за поравнања домена σ_4			
домен σ_4	*E. coli* σ^E	ECF phiEco32	ECF 7-11	*E. coli* RpoD
B. subtilis σ^W	$2e^{-15}$	X	$2e^{-4}$	без поравнања
E. coli σ^E		X	без поравнања	0.42
ECF phiEco32			X	X
ECF 7-11				без поравнања

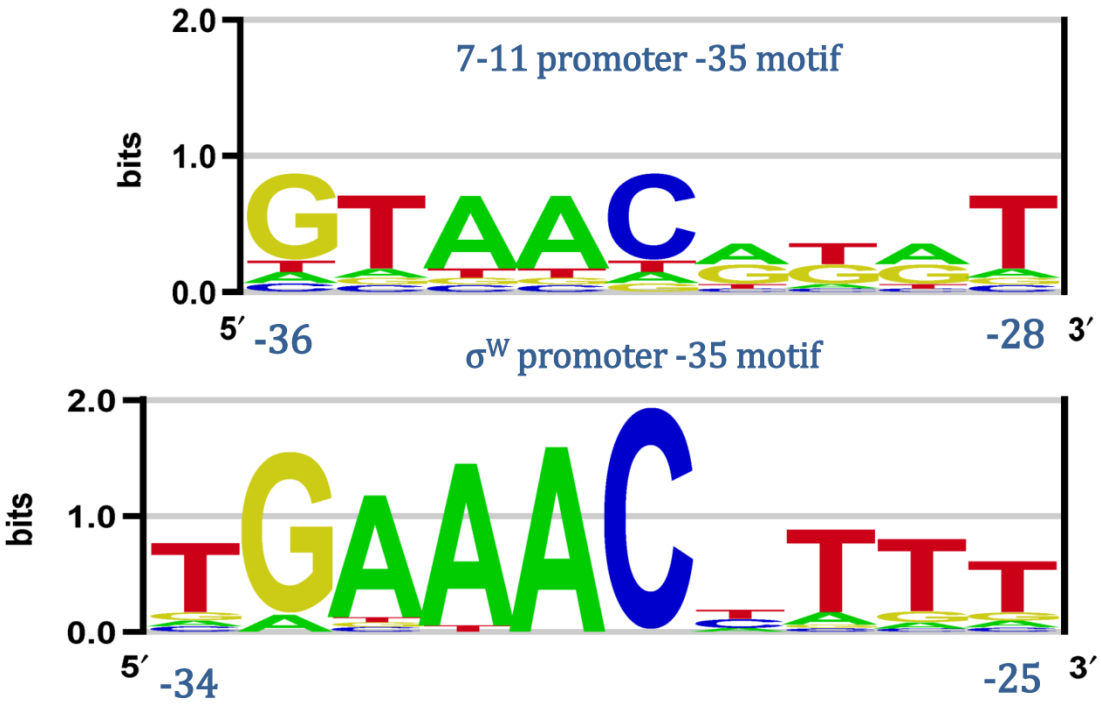
У складу са претходним, уследила је шира компаративна анализа протеинских секвенци у групи ECF. У овој анализи фагни σ фактор 7-11 поравнат је у паровима са познатим бактеријским ECF представницима ради препознавања ECF подгрупе, која је најсличнија фагним σ факторима. Добијена поравнања су већином статистички значајна (нека и високо значајна), при чему подгрупа ECF28 показује највећу сличност са фагним σ факторима. Од добро проучених ECF подгрупа, највећа сличност са σ фактором 7-11 уочена је за подгрупу ECF01, која садржи σ^W из *B. subtilis*. Додатно, секвенце оба фагна ECF σ фактора

поравнате су и са члановима групе RpoD (~44,000 протеинских секвенци анотираних у GenBank-у), при чему су за велику већину анализираних секвенци добијена поравнања без статистичког значаја, што је у складу са резултатима анализе домена σ_2 и σ_4.

Сумарно, резултати компаративне анализе протеинских секвенци у σ^{70} фамилији сведоче о врло израженим разликама између њених чланова, које се најјасније уочавају поређењем фагних ECF и бактеријских RpoD σ фактора. Насупрот овоме, σ^{70} промоторе одликује иста општа организација, при чему је нарочито битно уочити сличност у организацији -10 елемената управо фагних ECF и бактеријских RpoD промотора. Ово нас упућује на закључак да је заједничка структура σ^{70} промотора вероватно условљена заједничким механизмом интеракције са σ факторима (нпр. "mix-and-match"), чија структура очигледно не испољава доминатан утицај при обликовању кључних механистичких аспеката иницирања транскрипције.

4.2.3. Предвиђање -35 елемената у фагним ECF промоторима

Присуство дугих узводних продужетака -10 елемента у фагним ECF промоторима, које у RpoD групи могу да компензују одсуство -35 елемента [19], указало је на значај потраге за могућим -35 елементима у сегменатима узводно од предвиђених фагних -10 елемената. Анализом узводних промоторских секвенци бактериофага 7-11, чији -10 елемент показује сличност са експериментално предвиђеним -10 елементом σ^W, успешно је предвиђен -35 елемент, који је такође налик -35 елементу σ^W. Екстензивна сличност, која се протеже од узводне ивице -35 елемента, преко карактеристичног мотива "AAC" па све до низводног сегмента елемента богатог нуклеотидима T, јасно се уочава на Слици 8.



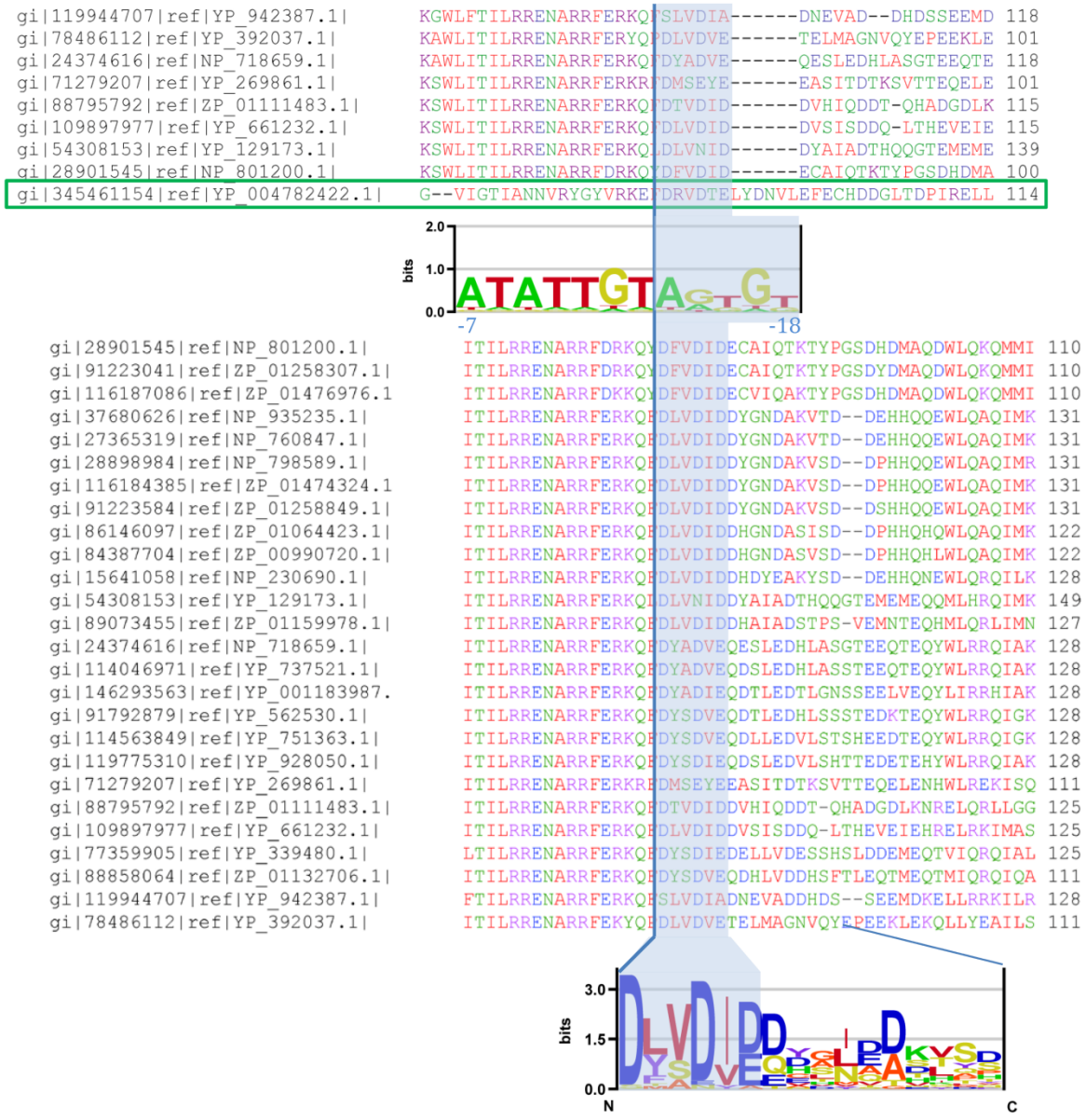
Слика 8. Поравнање логоа промоторских -35 елемената које препознају фагни σ фактор 7-11 и σ^W: (изнад) Лого је направљен на основу поравнања секвенци узводно од -10 елемента које препознаје фагни σ фактор 7-11; (испод) Лого је направљен од поравнатих -35 елемената које препознаје σ^W. Координате узводних и низводних граница -35 елемената дате су у односу на удаљеност од места почетка транскрипције.

С друге стране, у анализи узводних промоторских секвенци бактриофага phiEco32 није препознат -35 елемент, што је конзистентно са одсуством домена σ_4 (укљученог у интеракције са -35 елементом) у протеинској секвенци phiEco32 σ фактора. Међутим, код мањег дела анализираних секвенци (две од укупно шест), које одговарају касним фагним промоторима, предвиђен је конзервирани мотив ("AAGACCT") у три поновка. Предвиђени мотив је могуће место везивања транскрипционог фактора (вероватно кодираног фагним геномом), који би могао имати улогу у успостављању различитог временског обрасца експресије за средње и касне фагне гене. Значај овог резултата за ревизију регулаторне парадигме у групи ECF детаљније је изложен у Дискусији.

4.2.4. Протеински мотиви за препознавање продужетка -10 елемента

Присуство продужетка -10 елемента у фагним ECF промоторима изискује препознавање мотива на нивоу протеинске секвенце, који интерагује са овим продужетком, а чија се појава за фагни σ фактор 7-11 очекује C-терминално у односу на границу домена σ_2. Конкретно, за σ фактор 7-11 граница домена σ_2 (који интерагује са -10 елементом) предвиђена је на основу секвенци канонских ECF представника (σ^E и σ^W), који не препознају -10 елементе са узводним продужецима, услед чега се мотив који интерагује са овим продужетком у фагним σ факторима очекује управо C-терминално од границе домена σ_2. На функционалност предвиђене интеракције додатно би указало и присуство предвиђеног мотива у бар неким члановима бактеријских ECF подгрупа, али и паралелно одсуство у ECF σ факторима блиским σ^W (подгрупа ECF01), који је сличан фагним σ факторима, али не препознаје узводне продужетке -10 елемента. У складу са изнетим претпоставкама, C-терминални сегмент у односу на домен σ_2 фагног σ фактора 7-11 упоређен је са:

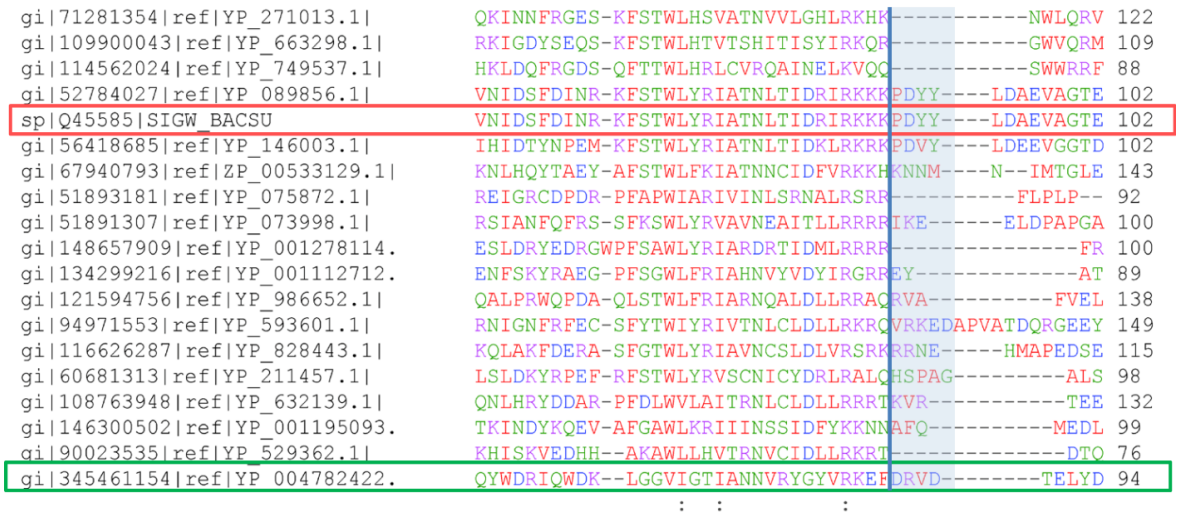
(i) одговарајућим сегментима бактеријских ECF σ фактора, нарочито са члановима (најсличније) подгрупе ECF28;
(ii) са представницима подгрупе ECF01, чији је члан канонски ECF σ фактор σ^W.



Слика 9. Поравнање протеинских секвенци фагног σ фактора 7-11 и чланова подгрупе ECF28: (изнад) Поравнање више секвенци које укључује фагни σ фактор 7-11 (уоквирен зеленим правоугаоником) и одабране (види секцију 3.4 Метода) представнике подгрупе ECF28. Део поравнања приказан на слици одговара С-терминалном крају домена σ_2 (вертикална линија на поравнању означава С-терминалну границу домена) и секвенце које се налазе С-терминално у односу на означену границу домена σ_2. Осенчена површина на поравнању одговара делу протеинске секвенце σ фактора који интерагује са продужетком -10 елемента у промоторима које препознаје фагни σ фактор 7-11. Испод поравнања приказан је лого фагног промотора 7-11, са осенченим продужетком -10 елемента за који су приказане и координате у односу на место почетка транскрипције. Идентификациони бројеви (GI) поравнатих представника подгрупе ECF28 дати су у Табели М1 у Методама. (испод) Поравнање више секвенци које укључује све σ факторе из подгрупе ECF28 са осенченим конзервираним мотивом који интерагује са продужетком -10 елемента. Испод поравнања приказан је лого конзервираног мотива заједно са околним секвенцама, које се протежу од С-терминалне границе мотива до појаве прве празнине у поравнању.

У горњем делу Слике 9, који приказује поравнање σ фактора 7-11 и изабраних представника подгрупе ECF28, уочава се добро поравнат сегмент C-терминално од границе домена σ_2 7-11 (означена вертикалном линијом на поравнању) дужине 6 амино киселина (АК), што приближно одговара сегменту посредством ког RpoD σ фактор интерагује са продуженим -10 елементом [111]. Да би се испитало присуство и степен конзервације овог сегмента у целокупној подгрупи ECF28, урађено је глобално поравнање свих њених представника и конструисан лого за сегменте C-терминално од домена σ_2, што је приказано на средишњем и доњем делу Слике 9. На логоу се јасно издваја разлика у конзервацији сегмента који одговара предвиђеном мотиву и секвенце C-терминално од њега, која се протеже до N терминуса домена σ_4, што представља додатну потврду функционалности предвиђеног мотива, тј. улоге у препознавању продуженог -10 елемента. С друге стране, на Слици 10 приказано је глобално поравнање фагног σ фактора 7-11 са представницима подгрупе ECF01 (која садржи канонски фактор σ^W), у којима се јасно уочава одсуство конзервираног мотива предвиђеног у подгрупи ECF28, што је илустровано лошим поравнањем секвенци са доста празнина у поравнању. Оба резултата су у складу са почетним претпоставкама, тј. потврђују да је конзервирани сегмент C-терминално од домена σ_2 код одређених представника групе ECF задужен за интеракцију са продужетком -10 елемента.

Ради утврђивања степена заступљености конзервираног мотива у целокупној групи ECF, у глобално поравнање подгрупе ECF28 и фагног σ фактора 7-11 укључиване су секвенце представника осталих ECF подгрупа, при чему је анализа указала на одсуство предвиђеног конзервираног мотива у остатку ECF субфамилије. Дакле, појава предвиђеног конзервираног мотива у секвенцама ECF σ фактора је уско специфична, што је у складу са претпостављном функцијом препознавања продужетка -10 елемента, која не представља широко распрострањен феномен међу промоторима групе ECF (будући да одсуствује у канонским промоторима σ^E и σ^W). У складу са овим, подгрупа ECF28 може се међу бактеријским ECF представницима сматрати могућим примером одступања од тренутне регулаторне парадигме, услед чега је у наставку анализиран промоторски специфицитет њених представника.

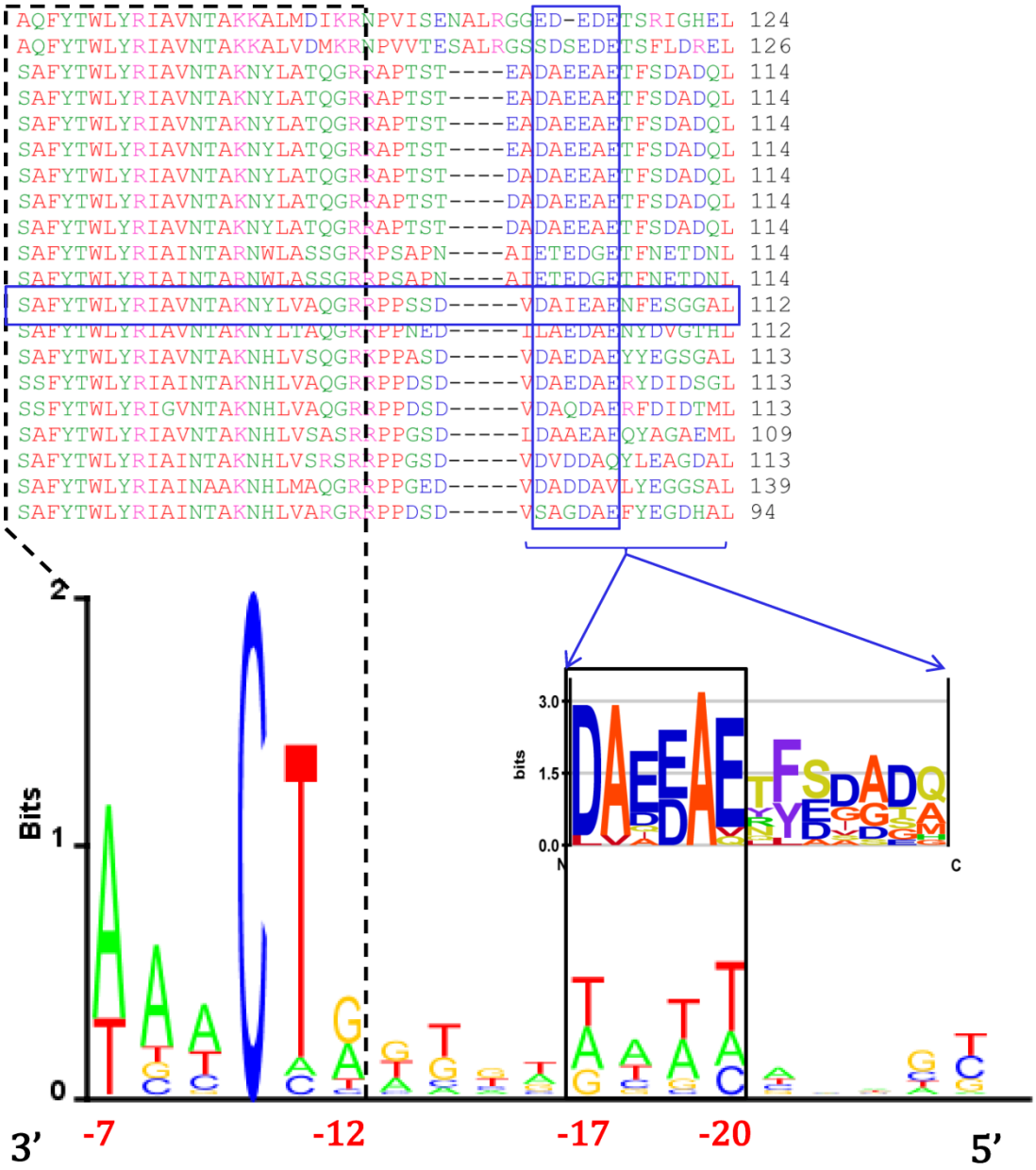


Слика 10. Поравнање протеинских секвенци фагног σ фактора 7-11 и чланова подгрупе ECF01: Приказано је поравнање више секвенци које укључује одабране представнике подгрупе ECF01, укључујући σ^W (секвенца означена црвеним правоугаоником), који препознају промоторе без продужетка -10 елемента, као и фагни σ 7-11 (секвенца означена зеленим правоугаоником). Део поравнања приказан на слици одговара С-терминалном крају домена σ_2 (вертикална линија на поравнању означава С-терминалну границу домена) и секвенце које се налазе С-терминално у односу на означену границу домена σ_2. Осенчена површина на поравнању одговара делу протеинске секвенце σ фактора који интерагује са продужетком -10 елемента у промоторима које препознаје фагни σ фактор 7-11. Идентификациони бројеви (GI) поравнатих представника подгрупе ECF01 дати су у Табели М1 у Методама.

Будући да се подгрупа ECF28 састоји од у потпуности неизучених чланова, почетна тачка за анализу била је парадигма о ауторегулацији ECF σ фактора, тј. за претрагу су одабрани интергенски региони узводно од гена који кодирају анализиране σ факторе, у којима су потом промоторски елементи претраживани и надгледаном и ненадгледаном претрагом. Ни један од примењених приступа није дао валидна предвиђања за ECF28 промоторе, за шта могу постојати два објашњења: једно је да чланови анализиране групе не подлежу ауторегулацији па тражени промоторски елементи нису ни присутни у претраживаним секвенцама, док према другом промотори чланова анализиране групе не садрже препознатљиви ECF -35 елемент, чији је специфицитет окосница надгледане претраге. Обе могућности указују на одступање од тренутне парадигме о механизму препознавања промотора у групи ECF, чији је значај детаљније изложен у Дискусији.

4.2.5. Анализа промоторског специфицитета бактеријских ECF подгрупа

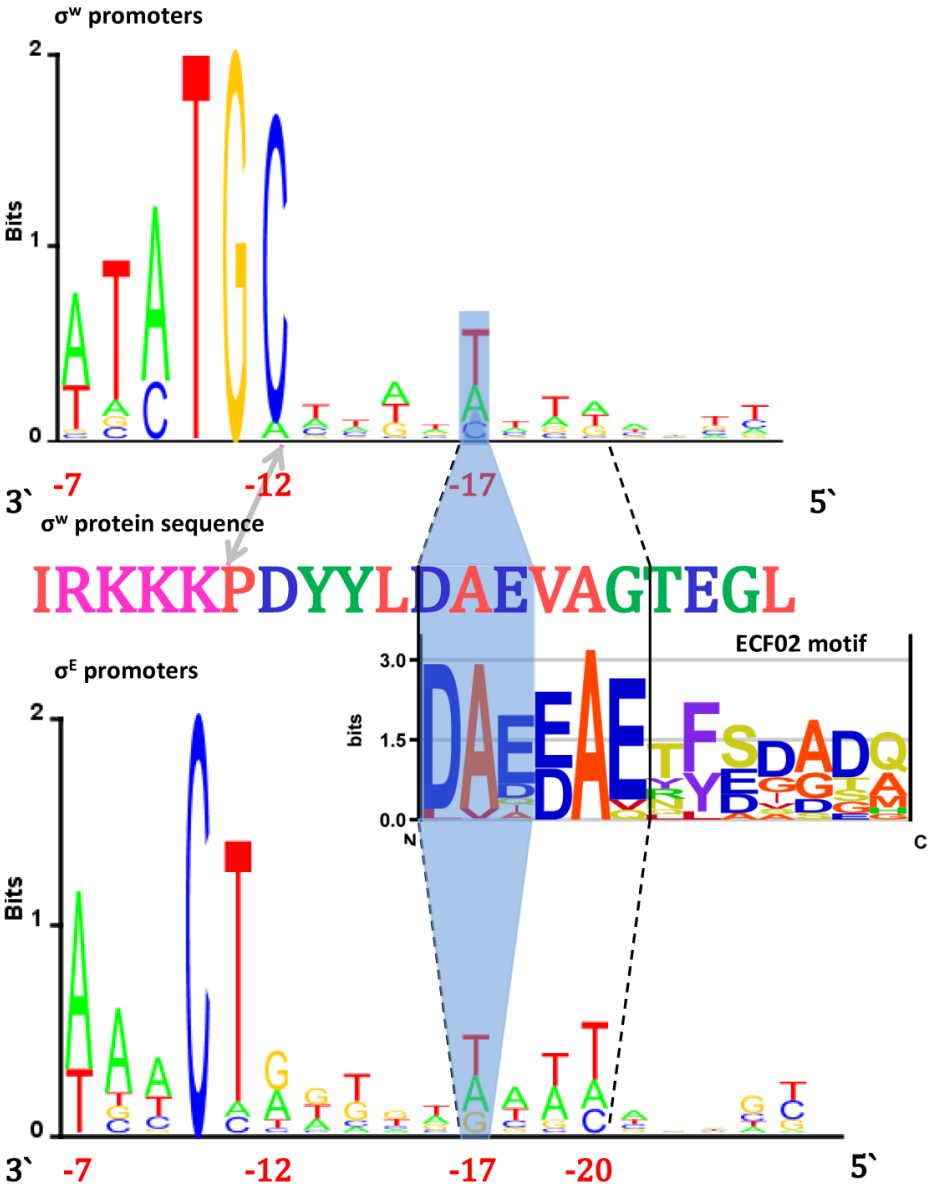
Узевши у обзир уочена одступања од парадигме о ригидним интеракцијама са промотором и код бактеријских ECF представника, у наставку су анализирани канонски чланови групе ECF, σ^E и σ^W, за које је доступан значајан број експериментално потврђених промоторских секвенци, и за које би интеракције ван канонских -35 и -10 елемената могле пружити додатне примере флексибилности у ECF субфамилији. Поравнање промотора σ^E довело је до предвиђања конзервираног мотива у секвенци спејсера (између -10 и -35 елемента), који је нарочито изражен у групи промотора који су неактивни у условима in-vitro (Слика 11, доњи лого). Предвиђени мотив налази се у близини узводне границе -10 елемента (позиције од -17 до -20 на логоу), услед чега се не може сматрати његовим директним продужетком. Насупрот овоме, предвиђени мотив одсуствује у промоторским секвенцама са јаком транскрипционом активношћу, при чему је значај овог резултата детаљније изложен у Дискусији.



Слика 11. Неканонске интеракције протеин-ДНК за ECF фактор σ^E: (изнад) Приказано је поравнање више секвенци које укључује фактор σ^E (секвенца означена плавим правоугаоником) и друге изабране представнике подгрупе ECF02. Део поравнања приказан на слици одговара C-терминалном крају домена σ_2 (уоквирен испрекиданом линијом заједно са промоторским -10 елементом) и секвенце које се налазе C-терминално у односу на приказану границу домена σ_2. Предвиђени конзервирани протеински мотив означен је плавим правоугаоником. Испод поравнања приказан је лого конзервираног протеинског мотива са пратећим (неконзервираним) секвенцама, које су локализоване C-терминално у односу на мотив. (испод) Лого промоторског -10 елемента заједно са секвенцом спејсера, које препознаје σ^E, приказан је у орјентацији 3' - 5', са координатама назначеним у односу на место почетка транскрипције. Предвиђена интеракција између конзервираног протеинског мотива и конзервираног мотива у спејсеру промотора уоквирена је пуном црном линијом. Идентификациони бројеви (GI) поравнатих представника подгрупе ECF02 дати су у Табели М1 у Методама.

Да би се у протеинској секвенци σ фактора предвидео мотив који интерагује са конзервираним спејсерским елементом, урађено је вишеструко поравнање секвенци, и то секвенце σ^E са одабраним представницима подгрупе ECF02, у потрази за конзервацијом С-терминално у односу на домен σ_2. Описаном претрагом предвиђен је конзервирани мотив дужине 6 АК у подгрупи ECF02, локализован у непосредној близини С-терминуса домена σ_2, али не као директни наставак, (Слика 11), што га чини погодним за интеракцију са конзервираним спејсерским елементом у промоторима σ^E.

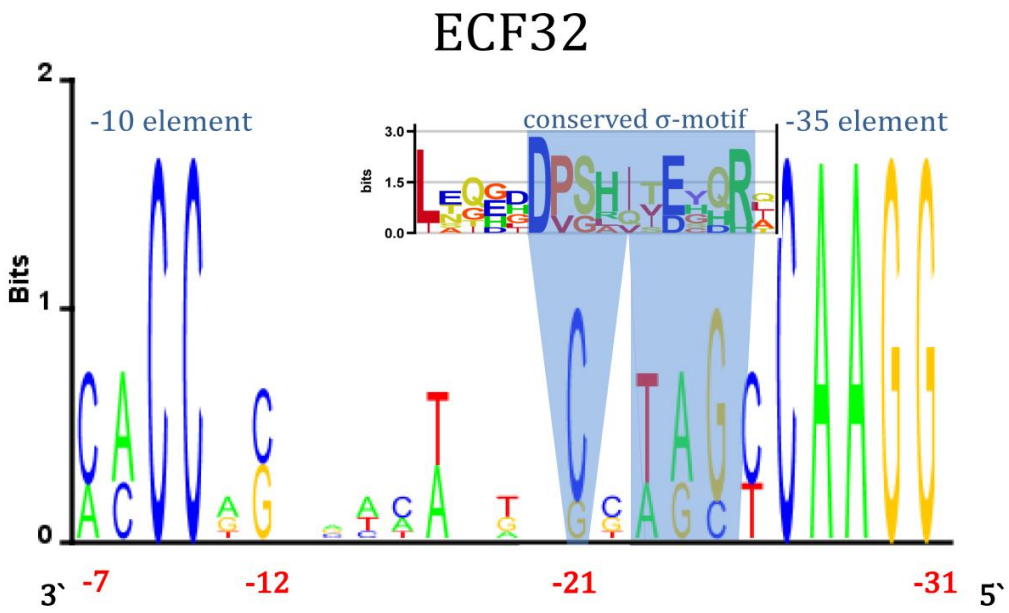
Поређењем секвенци σ^E и σ^W, може се уочити да је узводна половина мотива предвиђеног код σ^E (DAE) присутна и у секвенци σ^W, такође С-терминално од домена σ_2 (Слика 12). У складу са овим, у промоторским секвенцама σ^W, уочава се конзервација нуклеотида Т на позицији -17, што одговара најнизводнијој бази конзервираног спејсерског мотива код σ^E – дакле, делимичну конзервацију протеинског мотива σ^E у секвенци σ^W прати делимична конзервација предвиђене протеин-ДНК интеракције. При том, одсуство потпуне конзервације ове интеракције је у складу са одсуством конзервираног протеинског мотива, које је претходно уочено при поравнању чланова подгрупе ECF01.



Слика 12. Поређење неканонских интеракција протеин-ДНК за ECF факторе σ^E и σ^W: (изнад) Лого промоторских секвенци за σ^W (приказан у орјентацији 3' -5') који приказује -10 елемент и узводну секвенцу спејсера. (средина) Део протеинске секвенце фактора σ^W који укључује С-терминални крај домена σ_2 и пратећу секвенцу која се налази С-терминално од границе домена σ_2. Граница домена σ_2 и промоторског -10 елемента назначени су сивом стрелицом. Испод протеинске секвенце фактора σ^W приказан је лого конзервираног протеинског мотива, који је предвиђен у подгрупи ECF02. (испод) Лого промоторских секвенци за σ^E, (приказан у орјентацији 3' -5') који приказује -10 елемент и узводну секвенцу спејсера. Границе неканонске интеракције протеин-ДНК предвиђене за σ^E означене су испрекиданом линијом у свим приказаним секвенцама, при чему је осенчен део предвиђене интеркације који је конзервиран код σ^W. Координате промотрских елемената назначене су у односу на место почетка транскрипције.

Будући да резултати добијени на канонским ECF представницима σ^E и σ^W указују на значајну флексибилност промоторског специфицитета у групи ECF, која обухвата интеракције конзервираних мотива одвојене од канонских промоторских елемената,

приступило се предвиђању могућих интеракција ове врсте и у преосталим бактеријским ECF подгрупама. Треба, међутим, имати на уму да је овај покушај знатно отежан ограниченим скупом доступних ECF промотора, што онемогућава спровођење детаљне анализе, по угледу на канонске ECF представнике. Стога, претрага је ограничена на уочавање конзервације између протеинских домена σ_2 и σ_4, која би била праћена аналогно позицинираном конзервацијом у промоторској секвенци спејсера (ван -35 и -10 елемента). Описаном претрагом уочена је подгрупа ECF32 као још један пример ECF σ фактора, чије тзв. "неканонске" интеракције промотора нису ограничене на продужетке домена σ_2/-10 елемента. Нарочито је интересантна чињеница да је у групи ECF32 конзервирани ДНК елемент локализован на супротном крају секвенце спејсера, тј. у близини низводне границе -35 елемента, што на нивоу протеинске секвенце прати присуство конзервираног мотива N-терминално у односу на домен σ_4 (Слика 13). При том, експериментално је показано да мутирање базе у конзервираном елементу промотора ECF32 (конзервирано G на позицији -25) значајно смањује транскрипциону активност промотора [112], што додатно сугерише функционалност предвиђеног конзервираног мотива у спејсеру. Свеукупно, "неканонске" интеракције предвиђене за σ факторе групе ECF указују на већу квалитативну флексибилност у функционисању у поређењу са групом RpoD, а тиме и на постојање ширег репертоара елемената преко којих механизам "mix-and-match" може да се испољи и чије је присуство у групи ECF тестирано у наставку анализе.



Слика 13. Интеракције између σ фактора и спејсерске секвенце у промотору у подгрупи ECF32. Горњи лого на слици одговара представницима групе у којима је пронађен конзервирани протеински мотив (засенчен светло плаво) N-терминално од границе домена σ_4 (крај логоа се поклапа са датом границом). Доњи лого на слици одговара промоторима за дате представнике (које је било могуће предвидети), при чему је у спејсерској секвенци низводно од -35 елемента засенчен конзервирани елемент, коме је предвиђена интеракција са конзервираним протеинским мотивом; координате на слици означене су у односу на почетак транскрипције, при чему су GI бројеви за анализиране ECF32 σ факторе дати у Табели М1 у Методама.

4.3. Испитивање механизма "mix-and-match" у групи ECF σ фактора

Ради тестирања механизма "mix-and-match" у групи алтернативних ECF σ фактора спроведена је корелациона анализа на промоторима канонских представника σ^E и σ^W (еквивалентна анализи у групи RpoD), која се заснива на биофизичком моделу транскрипционе иницијације изложеном у Уводу. Окосница анализе је корелисање скорова матрица тежине за испитиване промоторске елементе (дволанчане и једноланчане) [18], који дају процену енергија интеракције са σ фактором у условима "несатурисане апроксимације" (видети Увод 1.5.1.1) [21, 113]. Према моделу изложеном у Уводу, сабирањем енергија интеракције дволанчаних промоторских елемената са σ фактором добијамо процену логаритма афинитета везивања РНКП за промотор [21, 96], док укључивањем у збир и енергетског доприноса једноланчаних елемената добијамо процену логаритма укупне снаге промотора, односно процену његове транскрипционе активности [21, 97].

У дволанчане промоторске елементе спадају -35 и продужени -10 елемент, док од једноланчаних у анализи учествује кратки -10 елемент. Дужина спејсера (тзв. "spacer weights") такође фигурише у анализи као један од дволанчаних елемената и тиме улази у процену афинитета везивања РНКП за промотор и укупне снаге промотора. Корелисањем снага једноланчаних и дволанчаних елемената, као и релевантних кинетичких параметара директно се тестира механизам "mix-and-match", јер се комплементација слабијег јачим елементом испољава кроз негативне корелације између одговарајућих скорова матрица тежине. Додатно, функционална комплементација између елемената највећим уделом бива усмерена ка кинетичком параметру који је пресудан за функционалност промотора, због чега се корелационом анализом паралелно изучава и кинетички профил промоторског одговора. На крају, неопходно је нагласити да се при поређењу промоторског елемента са кинетичким параметром на који утиче (нпр. -35 елемент са афинитетом везивања РНКП), снага датог елемента искључује из процене вредности параметра да би се избегла аутокорелација.

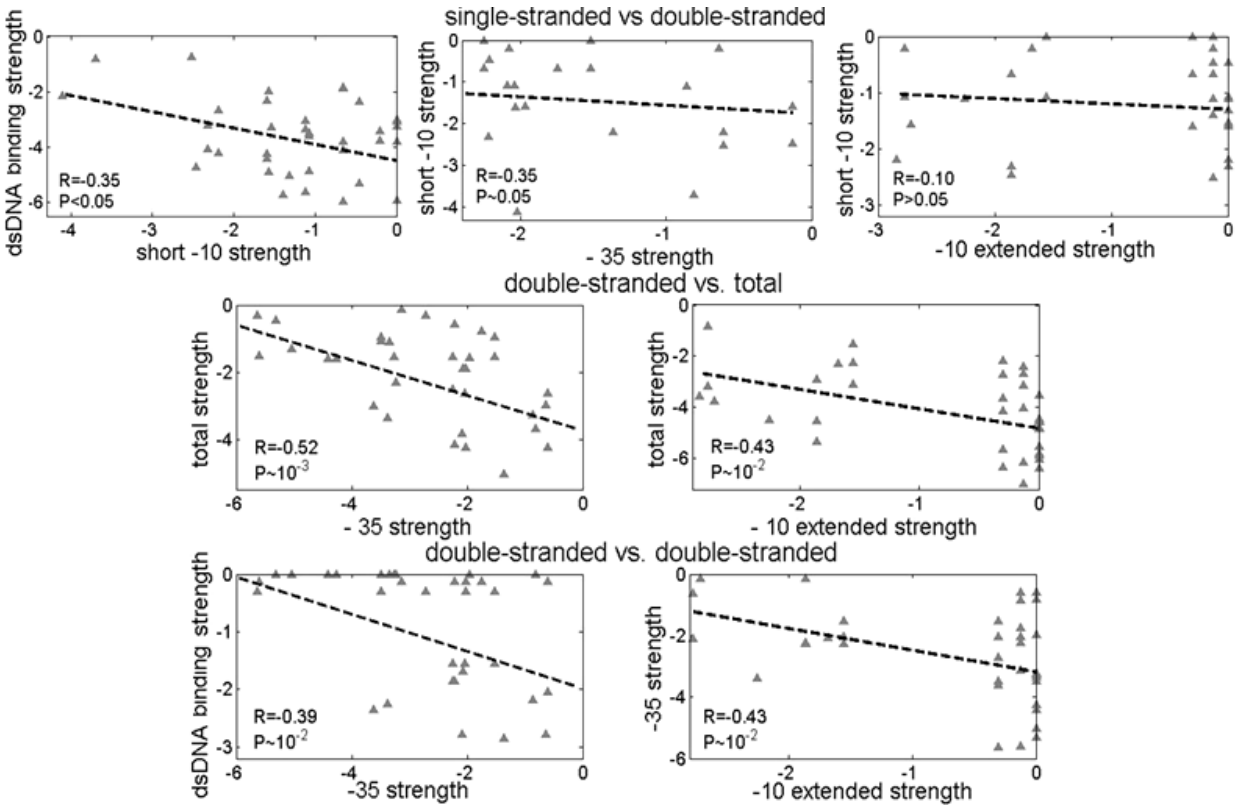
Ради једноставније интерпретације, резултати корелационе анализе графички су организовани на следећи начин: корелације између снага једноланчаних и дволанчаних промоторских елемената, које указују на комплементацију ка укупној транскрипционој активности, приказане су у првом реду панела; корелације између снага дволанчаних елемената и укупне јачине промотора приказане су у другом реду панела; док су корелације између снага дволанчаних елемената, које указују на комплементацију ка афинитету везивања РНКП, приказане у трећем реду панела. Већа негативна корелација указује на већи ниво комплементације, при чему се значајним називају искључиво корелације статистички значајне са P-вредношћу мањом од нивоа значајности 0,05.

4.3.1. Корелација снага σ^E промоторских елемената

Први скуп тестираних секвенци, за које су резултати приказани на Слици 14, састоји се од промотора σ^E *активних* у условима in-vitro. Јака транскриционa активност ових промотора одређена је искључиво особинама базалних (канонских) промоторских елемената, који су испитивани у анализи. Присуство статистички значајних корелација уочава се на Слици

14 за готово све тестиране комбинације елемената/параметара, које по опсегу превазилазе корелације добијене на промоторским RpoD секвенцама *E. coli* [18]. Конкретно, у групи RpoD најјачи ниво комплементације између промоторских елемената не прелази вредност -0.2 и усмерен је ка постизању довољног нивоа укупне транскрипционе активности. С друге стране, у групи ECF на комплементацију ка укупној транскрипционој активности упућују знатно јаче негативне корелације са максималном вредношћу од чак -0.52, што је илустровано првим и другим редом панела на Слици 14.

Ради правилног тумачења добијених корелација неопходно је још једном нагласити да и укупна снага промотора и афинитет везивања РНКП за промотор укључују снаге више појединачних промоторских елемената, које се у знатној мери преклапају. Директна последица претходног је јача корелација између снаге -35 елемента и укупне снаге промотора (други ред панела), у поређењу са корелацијом између снаге -35 и кратког -10 елемента (први ред панела). Наиме, укупна снага промотора сем кратког -10 елемента, укључује и снаге продуженог -10 елемента и спејсера, што нам говори да висок ниво комплементације ка укупној транскрипционој активности представља последицу јаких корелација између дволанчаних промоторских елемената (трећи ред панела).

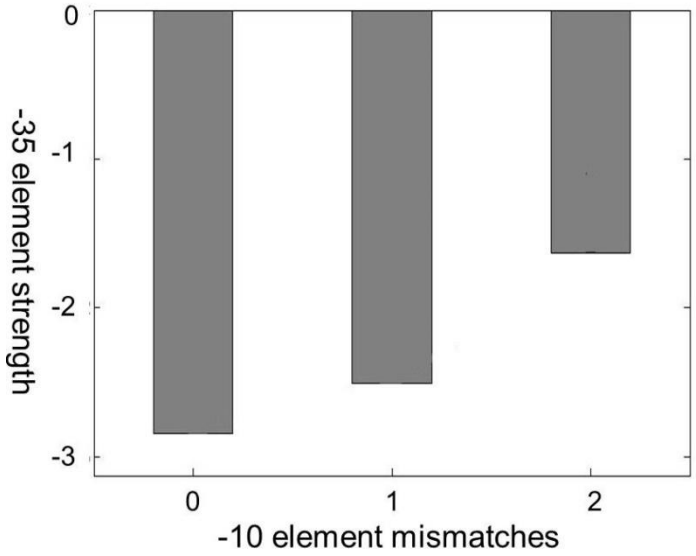


Слика 14. "Mix-and-match" у промоторским секвенцама σ^E *активним* in-vitro: На панелима слике приказане су корелације између различитих комбинација промоторских елемената за секвенце *активне* in-vitro. За сваки панел, на осама су назначени елементи чије су снаге укључене у корелације. Корелациони коефицијенти и P-вредности за приказане корелације назначене су на сваком панелу. Снаге појединачних промоторских елемената процењене су на основу скорова матрица тежине. Укупна снага промотора, као и афинитет везивања за dsDNA, добијају се сабирањем снага свих промоторских елемената, односно снага елемената који са σ фактором интерагују као dsDNA, при чему се снага елемента, са којим се дати кинетички параметар корелише, искључује из збира за његов прорачун. Први ред одговара корелацијама између елемената који са σ фактором интерагују као dsDNA и ssDNA; други ред одговара корелацијама између елемената dsDNA и укупне снаге промотора; трећи ред одговара корелацијама између елемената dsDNA.

Треба приметити и да се на трећем реду панела уочава субпопулација промотора са јаким продуженим -10 елементима и високим афинитетима везивања РНКП за промотор (приметити да високи скорови одговарају вредностима ~0), који не корелишу са снагама -35 и продуженог -10 елемента – дата промоторска субпопулација доприноси визуелној дисперзији извесног броја тачака на два плота. Овај резултат наводи на закључак да присуство врло јаког продуженог -10 елемента чини снагу -35 елемента много мање битном [19], што је у складу са добро познатим резултатом у групи RpoD, где јак продужени -10 елемент може да компензује одсуство -35 елемента, тј. да обезбеди довољан афинитет везивања РНКП за промотор, независно од снаге -35 елемента. Међутим, и поред присуства описане промоторске субпопулације, јаке негативне

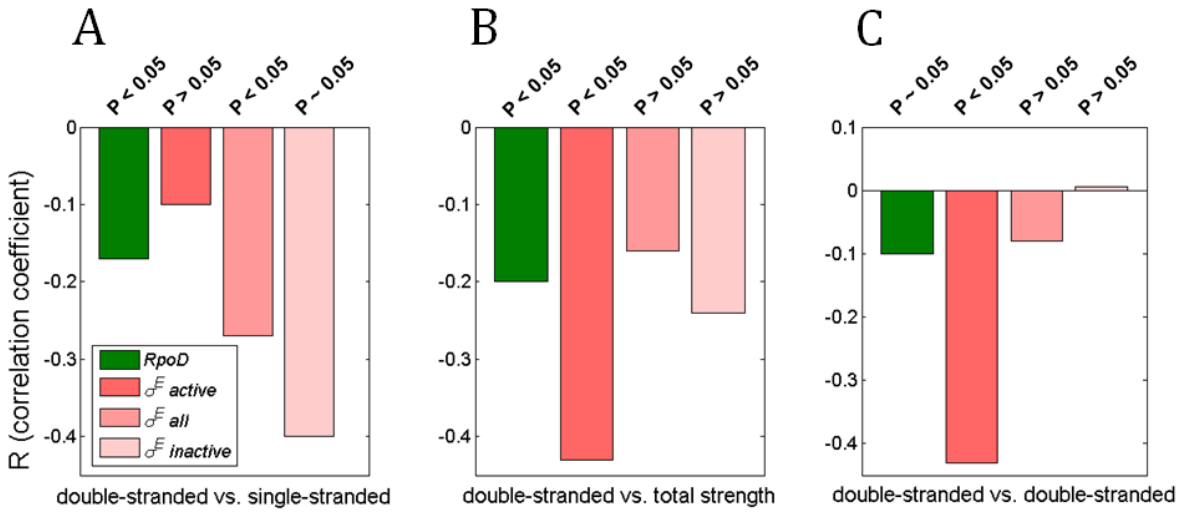
корелације јављају се између дволанчаних елемената у σ^E промоторима, што претходно није забележено у групи RpoD, и самим тим представља главну разлику у испољавању механизма "mix-and-match" између примарних RpoD и алтернативних ECF σ фактора.

Тенденција ка комплементацији, која се јавља у скупу σ^E промотора, присутна је и међу промоторима σ^W *B. subtilis*-а, иако је анализа у овом случају отежана мањим бројем доступних промоторских секвенци, које су при том знатно боље конзервиране [99]. Конкретно, -10 елемент у промоторима σ^W јавља се са највише два одступања у односу на консензусну секвенцу. Слично овоме, продужени -10 елемент готово је у потпуности конзервиран, о чему сведочи присуство једног неподударања у свега неколико промоторских секвенци. Стога су матрице тежине конструисане само за -35 елемент, који једини испољава задовољавајући ниво варијабилности, док су секвенце -10 елемената подељене у три групе: са два, једним и без неподударања у односу на консензус. За сваку од три промоторске групе процењена је (просечна) снага -35 елемента, при чему се показало да су најслабији -35 елементи удружени са консензусним -10 елементима, нешто јачи -35 елементи одговарају -10 елементима са једним неподударањем, док се најјачи -35 елементи јављају у комбинацији са најслабијим -10 елементима (Слика 15). Дакле, -10 елемент са већим бројем неподударања (тј. слабији) асоциран је са јачим -35 елементом, што је јасна тенденција ка испољавању механизма "mix-and-match".



Слика 15: "Mix-and-match" у промоторима σ^W. На слици је приказана корелација између просечне снаге дволначаних промоторских елемената (скорови матрице тежине на *y*-оси) и једноланчаних промоторских елемената (број неподударања у односу на консензус у одговарајућим -10 елементима на *x*-axis). Приметити да јачим -35 елементима одговара скор матрице тежине ближи 0.

Ради увида како на механизам "mix-and-match" утиче повећање хетерогености промоторских секвенци (тј. јачина промотора), првобитно анализираним σ^E промоторима придружене су секвенце *неактивне* у условима in-vitro, што је дало два нова скупа промотора за тестирање корелација, од којих су у првом присутни сви σ^E промотори (*активни* и *неактивни* in-vitro), док други чине само in-vitro *неактивне* секвенце. На Слици 16 упоређене су корелације за два нова скупа σ^E промотора са корелацијама добијеним на in-vitro *активном* секвенцама, при чему је поређење дато за комплементацију између дволанчаних и једноланчаних промоторских елемената (панел A), дволанчаних елемената и укупне снаге промотора (панел B) и дволанчаних елемената међусобно (панел C). Одабране корелације за три σ^E промоторска скупа упоређене су такође са одговарајућим корелацијама за RpoD промоторе (представљене левим баром на свим панелима).



Слика 16. Промена вредности корелација од in-vitro *активних* ка in-vitro *неактивним* промоторским секвенцама (поређење са промоторима RpoD). На панелима су приказане комплементације између: **A)** елемената dsDNA (продужетак -10 елемента) и ssDNA (кратки -10 елемент) **B)** елемената dsDNA (продужетак -10 елемента) и укупне снаге промотора**, C)** елемената dsDNA (продужетак -10 елемента) и dsDNA (-35 елемент) за промоторске секвенце σ^E *активне* in-vitro, *све* и *неактивне* in-vitro. Релевантне корелације (назначене на *y*-оси) упоређене су са одговарајућим вредностима за промоторске секвенце RpoD (у којима је корелисана снага -15 елемента). Информације о одговарајућим скуповима промоторских секвенци дате су у легенди слике.

Као што је на почетку истакнуто, у RpoD промоторским секвенцама најјаче корелације су зарад постизања довољног нивоа укупне транскрипционе активности [18], при чему сада на Слици 16A (упоредити први и други бар) уочавамо да је одговарајућа корелација у in-vitro *активним* σ^E секвенцама слабија (-0.1 за σ^E у поређењу са -0.17 за RpoD). Преласком ка in-vitro *неактивним* секвенцама (упоредити трећи и четврти бар) уочава се јачање дате

корелације (од -0.1 до -0.4), што је највероватније последица постојања притиска у скупу in-vitro *неактивних* секвенци да се повећа инхерентно ниска транскрипциона активност, кроз функционално надопуњавање промоторских елемената.

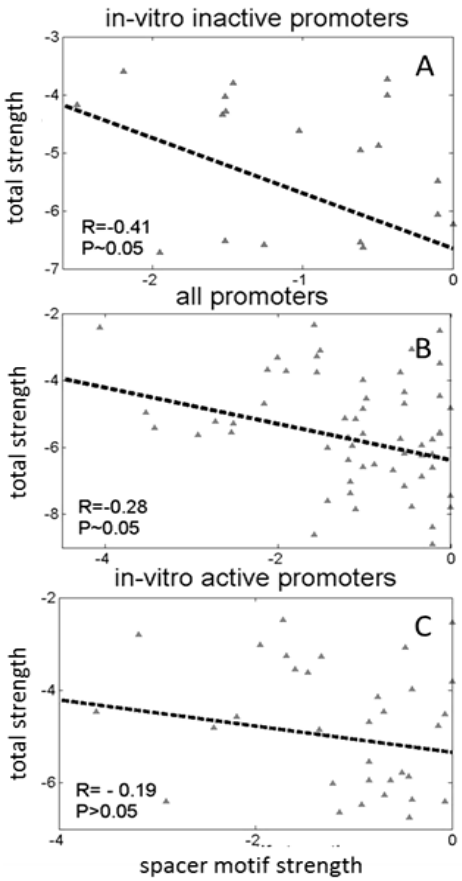
На Слици 16C уочавамо супротну тенденцију за комплементацију ка постизању довољно високог афинитета везивања РНКП за промотор. Овде уочавамо знатно јаче негативне корелације између дволанчаних елемената за σ^E in-vitro *активне* промоторе, у односу на еквивалентне корелације у RpoD промоторима, чиме је јасно наглашена основна разлика у транскрипционој кинетици ове две групе промотора. Конкретно, за RpoD секвенце параметар који у највећој мери одређује одговор промотора је укупна транскрипциона активност, док у σ^E секвенцама ова улога припада афинитету везивања РНКП за промотор. Додатно, корелације између дволанчаних елемената опадају од in-vitro *активних* ка *неактивним* секвенцама (упоредити барове 2-4 на Слици 16C), што даље потврђује да је промоторска активност σ^E уско повезана са афинитетом везивања РНКП за промотор. Прецизније, корелације између дволанчаних елемената су врло јаке (-0.43) у in-vitro *активним* секвенцама, што прелази у потпуно одсуство корелација у in-vitro *неактивним* секвенцама. Уочени образац промене корелација између дволанчаних елемената води ка сличном тренду промене корелација између дволанчаних елемената и укупне снаге промотора, што је илустровано на слици Слици 16B (централни панел).

Значајно опадање корелација, које су уочава на Сликама 16B и 16C, при преласку са in-vitro *активних* ка *неактивним* σ^E промоторима може бити узрокавано утицајем спољашњих регулатора на транскрипциону активност in-vitro слабих промотора, чији се удео у "mix-and-match"-у, међутим, не може утврдити праћењем корелација између канонских промоторских елемената. Један од могућих спољашњих носилаца транскрипционе активности (прецизније, афинитета везивања РНКП за промотор) у σ^E промоторима је конзервирани спејсерски елемент [114], што је кроз корелисање са снагама канонских промоторских елемената размотрено у наредном поглављу.

4.3.2. Корелације снага спејсерског и канонских σ^E елемената

Информације о улози спејсерског мотива у функционисању σ^E промотора нам такође даје корелациона анализа, којом се утврђује степен функционалног надопуњавања између овог могућег спољашњег регулатора и канонских промоторских елемената. Корелационом анализом обухваћена су сва три скупа σ^E промоторских секвенци, са промоторима *неактивним* in-vitro, свим промоторима и промоторима *активним* in-vitro. При том, сем корелација између спејсерског елемента и осталих елемената/параметара, наново су процењене и корелације између канонских промоторских елемената, по укључивању снаге спејсерског елемента у релевантне кинетичке параметре.

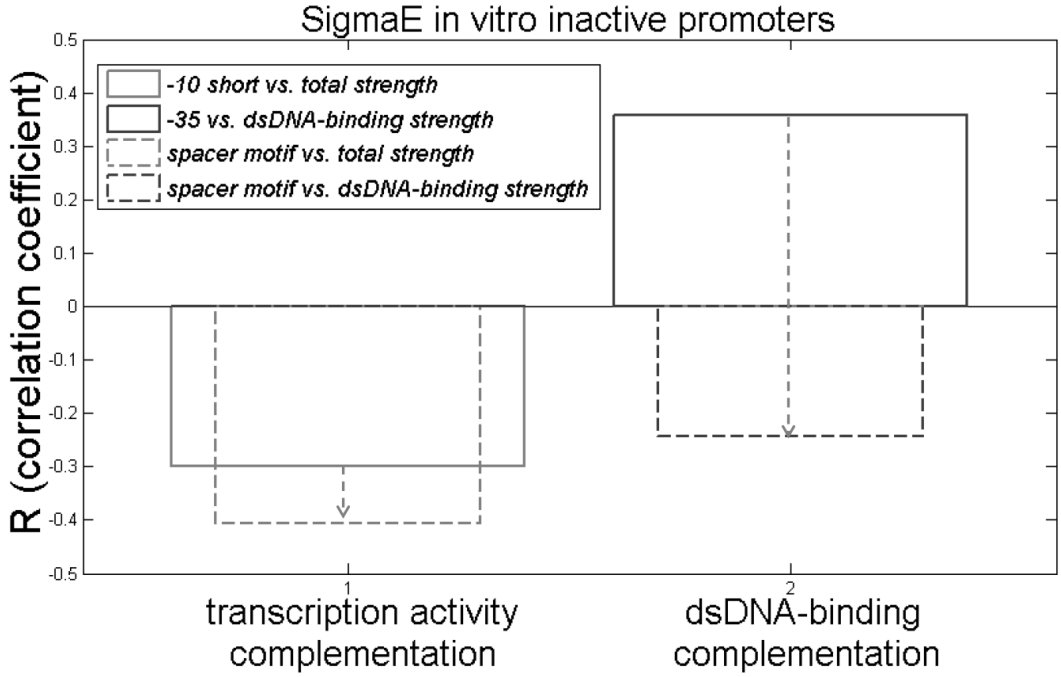
У скупу in-vitro *неактивних* секвенци спејсерски елемент остварује изражене негативне корелације са свим промоторским елементима (вредности корелационих коефицијената се крећу између -0.24 и -0.41; резултати нису приказани). Једини изузетак је позитивна корелација која постоји између спејсерског и продуженог -10 елемента, што указује на могућност да ова два елемента заједнички надопуњују снаге осталих промоторских елемената. Најјача негативна корелација (-0.41) добијена је са укупном снагом промотора (Слика 17A), док је најслабија корелација добијена са дволанчаним промоторским елементима (-0.24). Свеукупно, уочавање јаких негативних корелација потврђује почетну претпоставку да конзервирани елемент спејсера у in-vitro слабим промоторима учествује у "mix-and-match"-у са канонским елементима зарад појачавања инхерентно ниске транскрипционе активности промотора.



Слика 17. Корелације између конзервираног елемента у спејсеру и укупне снаге промотора: Горњи, средњи и доњи панел на слици приказују, редом, корелације у промоторским секвенцама σ^E *неактивним* in-vitro, свим промоторима и промоторским секвенцама *активним* in-vitro. На сва три панела приказана је корелација између снаге конзервираног мотива у спејсеру и укупне снаге промотора. Корелациони коефицијент и P-вредност приказане су на сваком панелу. Укупна снага промотора одговара збиру снага свих промоторских елемената (изузев снаге мотива у спејсеру са којим се корелише).

Даље је испитивано како спејсерски елемент утиче на промену претходно добијених корелација између канонских елемената у in-vitro *неактивним* σ^E промоторима, при чему на Слици 18 уочавамо врло снажно повећање комплементације за афинитет везивања РНКП за промотор (промена од -0.6) и знатно слабије повећање комплементације за укупну транскрипциону активност (промена од -0.1). Овај резултат јасно потврђује предложену улогу за спејсерски елемент, која подразумева надопуњавање снага канонских елемената у правцу постизања довољног афинитета везивања РНКП за слабе σ^E промоторе. Овде треба нагласити да је, аналогно са резултатима анализе канонских елемената, највећа апсолутна корелација која постоји између спејсерског елемента и укупне снаге промотора заправо последица јаких негативних корелација спејсерског са готово свим (већином дволанчаним) канонским елементима. Наново процењене

корелације између канонских елемената и укупне снаге промотора (по укључивању снаге спејсерског елемента) такође се одликују порастом – нпр. корелација између дволанчаних и кратког -10 елемента расте са -0.3 на -0.42, што још једном потврђује улогу спејсерског мотива у комплементацији снага осталих промоторских елемената. Као једини изузетак јавља се продужени -10 елемент, чије корелације са осталим промоторским елементима слабе по укључивању спејсерског мотива, а што је у складу са претходно изнетом претпоставком да дата два елемента кроз "mix-and-match" заједнички надопуњују снаге осталих.



Слика 18. Поређење корелација у σ^E промоторима *неактивним* in-vitro са и без укључивања снаге конзервираног елемента у спејсеру: На панелима су приказане комплементације (ка укупној транскрипционој активности – лево; ка афинитету везивања за dsDNA – десно) између канонских промоторских елемената и њихове вредности након укључивања снаге конзервираног мотива у спејсеру у корелације. Опсег промене корелационог коефицијента R приказан је испрекиданом стрелицом на оба панела, док су апсолутне вредности одговарајућих корелационих коефицијената приказане на *y*-оси. Информације о промоторским елементима/кинетичким параметрима, који су укључени у приказане корелације дате су у легенди слике.

Даље, у скупу секвенци који обједињује све σ^E промоторе такође су уочене негативне корелације, али у мањем обиму у односу на скуп in-vitro *неактивних* промотора. На пример, корелација спејсерског елемента са укупном снагом промотора опада са -0.41 на -0.28 (Слика 17B), премда и даље има статистички значај. Наново процењене корелације канонских елемента расту по укључивању спејсерског елемента у анализу, што прати тренд уочен на in-vitro *неактивним* секвенцама.

Финално, у скупу in-vitro *активних* секвенци добијају се слабије корелације у поређењу са осталим скуповима σ^E промотора, које додатно губе и статистички значај. На пример, на Слици 17C уочава се да корелација спејсерског мотива са укупном снагом промотора износи свега -0.19, при чему није статистички значајна. Пад корелација, који се уочава од in-vitro *неактивних* ка *активним* секвенцама, сведочи о мањем функционалном значају спејсерског мотива у јаким промоторима, што је у складу са слабијим присуством датог елемента у in-vitro *активним* у односу на *неактивне* секвенце [114]. Сумарно, можемо закључити да се већа квалитативна флексибилност у функционисању чланова групе ECF очитава и на квантитативном нивоу, за шта расподела јаких корелација између канонских σ^E промоторских елемената и спољашњег регулатора (спејсерски елемент) представља јасан пример.

5. Дискусија

У поглављу које следи резултати анализе специфицитета и механизма "mix-and-match" у групи ECF σ фактора биће основ за ревизију тренутне парадигме о ригидном промоторском специфицитету. Везано за предвиђање специфицитета у ECF σ групи, на почетку ће бити показано како предвиђања промотора за фагне ECF σ факторе, у комбинацији са применом стандардних биоинформатичких метода за анализу геномских секвенци, могу довести до приближног разумевања инфективне стратегије вируса. У контексту процене парадигме о ригидности, као главни резултат у предвиђањима за фагне промоторе биће препознато присуство класичног (екстремног) примера механизма "mix-and-match" (компензација одсуства -35 елемента јаким продуженим -10 елементом), јер представља први доказ флексибилности у интеракцијама ECF σ фактора са промоторским секвенцама. На ово се надовезују резултати шире анализе специфицитета бактеријских ECF представника, код којих су предвиђене тзв. неканонске интеракције са промоторским секвенцама, што је препознато као израженија квалитативна флексибилност у функционисању у односу на RpoD σ факторе. Потом следе резултати корелационе анализе, који потврђују испољавање механизма "mix-and-match" код (канонског) ECF представника – σ^E, и то у већем опсегу него у групи RpoD, што је у Дискусији објашњено помоћу једноставног квалитативног модела. Додатно, уочене разлике у важности кинетичких параметара за постизање транскрипционе активности на промоторима ECF и RpoD дискутоване су као последица различитих структурно-функционалних ограничења у групама σ^{70} фамилије. Будући да добијени резултати представљају снажну назнаку универзалности механизма "mix-and-match" у фамилији σ^{70}, Дискусија је закључена усмерењима ка даљим (експерименталним) истраживањима, којима би ова хипотеза могла да се потврди.

5.1. Транскрипциона стратегија бактериофага 7-11

Предвиђање специфицитета σ фактора, кодираних фагним геномима, најзахтевнији је корак у биоинформатичком изучавању регулације фагне транскрипције, који уз стандардну надгледану претрагу бактеријских промотора даје целовиту слику просторне организације промотора у геному. Заједно са резултатима функционалне анотације гена, добија се оквир на основу ког можемо реконструисати временски образац вирусне генске експресије током инфекције, што је у наставку показано на примеру бактериофага 7-11.

У почетној фази инфекције, бактеријски холоензим РНКП са јаких RpoD промотора иницира транскрипцију бактериофагних "раних" гена, који чине узводни део функционалног генског кластера. Како се предвиђени RpoD промотори већински налазе узводно од "раних" гена, функционални кластер се највероватније транскрибује као дугачак оперон, што је типична одлика стратегије генске ескпресије великог броја бактериофага [73]. Прелазак са "ране" на "средњу" фазу инфекције дешава се након експресије анти-σ фактора – протеинског производа једног од узводних гена функционалног кластера – чија је улога да иницира промену промоторског специфицитета ензима РНКП. Последица активности анти-σ фактора је блокада бактеријског холоензима, услед чега наступа гашење транскрипције свих бактеријских и "раних" вирусних гена.

Транскрипција преосталих вирусних гена ("средњи" и "касни") наставља се са фагних промотора помоћу хетерогеног холоензимског комплекса, у коме место бактеријског RpoD σ фактора заузима ECF σ фактор кодиран бактериофаним геномом. Како су фагни промотори локализовани узводно од гена за ECF σ фактор, почетне количине овог протеина, које су неопходне за покретање фагне транскрипције, највероватније се експримирају са RpoD промотора у склопу дугачког оперона, чији узводни део чине "рани" гени. Након блокаде транскрипције са RpoD промотора, експресија фагног σ фактора наставља се са ECF промотора, што му обезбеђује високу активност и у почетним, и у каснијим фазама инфекције. Гени са описаним транскрипционим обрасцем класификују се као "средњи", заузимају низводни сегмент функционалног кластера и у највећој мери транскрибују се са средњих фагних промотора (предвиђени кратки ECF мотиви). Ова генска класа најчешће кодира протеине који блокирају транскрипционе и транслационе процесе бактеријске ћелије, због чега предвиђање транскрипционе

стратегије бактериофага, чији су домаћини патогени сојеви, може имати медицински значај [72].

Да би инфективни циклус бактериофага био завршен, неопходно је експримирати и "касне" гене, који кодирају делове вирусног омотача и протеине са улогом у склапању нових фагних партикула. Локација ових гена ограничена је на структурни кластер, узводно од ког се налазе једини ECF (фагни) промотори са додатним елементом "TATA" (предвиђени дуги поновци). У складу са овим, разумно је претпоставити да управо елемент "TATA" обезбеђује довољну јачину дугих (касних) ECF промотора, која омогућава транскрипциону активност у последњој фази инфекције, када су ограничене количине и ензимских и градивних ресурса на којима почива генска експресија. Преферентним коришћењем касних, у односу на средње ECF промоторе, успоставља се оптималан однос структурних компонената вирусног омотача и насинтетисаних копија генома, што доводи до стварања великог броја нових вирусних партикула спремних да уђу у наредне циклусе инфекције.

Као што је претходно истакнуто, најзахтевнији корак у реконструкцији вирусне транскрипционе стратегије је предвиђање промоторског специфицитета за σ факторе (или РНКП) кодиране фагним геномима где, за разлику од остатка анализе геномске секвенце бактериофага, примена стандардних метода (алгоритми MLSA) не даје поуздана предвиђања. Стога, значајан резултат је успешна потврда новог методолошког приступа, који је развијен на предвиђању фагних промотора у геному бактериофага 7-11, а затим потврђен и на геномима phiEco32 и Xp10, у којима се предвиђени мотиви подударају са експериментално установљеним фагним промоторима [64, 68, 100]. У овом контексту, важно је нагласити да је на геномима phiEco32 и Xp10 још једном потврђена слаба прилагођеност алгоритама MLSA за препознавање промотора фагних σ фактора, сем када се претрага конципира на експерименталној информацији о подели фагних гена на временски различите експресионе класе. На основу претходног, може се закључити да поравнавање интергенских региона у паровима у многим случајевима представља преферентни методолошки приступ за предвиђање специфицитета фагних σ фактора (и РНКП), који типично интерагују са добро конзервираним промоторима, али статистички слабо издиференцираним у односу на позадину (тј. јављају се у много мањем броју понављања, у односу на укупан број претраживаних секвенци).

Сем новог приступа за предвиђање фагних промотора, потврђени су и промоторски мотиви у геному 7-11 помоћу компаративне анализе са бактериофагом phiEco32, при чему је уочена конзистентност и у геномској локализацији, и у специфицитету промотора за високо сличне ECF σ факторе два бактериофага. У контексту изучавања промоторског специфицитета у групи ECF, компаративна анализа, међутим, дала је много значајнији резултат, јер се поређењем фагних промотора 7-11 и phiEco32 први пут уочава флексибилност промоторске структуре у групи ECF. Наиме, у промоторима оба бактериофага присутни су врло изражени продужеци -10 елемента, који у групи RpoD типично компензују слаб (или одсутан) -35 елемент. У ECF промоторима phiEco32 продужетке -10 елемента заиста прати одсуство -35 елемента, док је у промоторима 7-11 -35 елемент удружен само са дугим поновцима (тзв. "касним" промоторима), што су врло јасни примери одступања од тренутне парадигме о ригидном промоторском специфицитету.

Флексибилност у структури фагних ECF промотора потврда је полазног становишта о неопходности независног приступа за изучавање промоторског специфицитета у групи ECF, где је највећа вероватноћа за уочавање квалитативно другачијих регулаторних парадигми била приписана управо представницима који се највише разликују у односу на остале чланове σ^{70} фамилије (односно бактериофагним ECF σ факторима). Такође, значај добијених резултата за (даље) изучавање механизма функционисања ECF σ фактора током иницијације транскрипције афирмише употребу бактериофага као модел система за бионформатичка истраживања базичних процеса у молекуларној биологији, чиме се уз смањен утрошак времена (и ресурса) може значајно повећати вероватноћа за успостављање нових парадигми.

5.2. Промоторски специфицитет ECF σ фактора

Значај добијених предвиђања за фагне ECF промоторе дискутован је у претходном поглављу са више аспеката. Од тога, нарочито битним показале су се импликације за даље изучавање промоторског специфицитета у групи ECF, услед врло јасног одступања фагних промотора од тренутне парадигме о ригидним интеракцијама између ECF σ фактора и њихових промоторских секвенци [7, 23, 115]. Наиме, уочавање првог примера флексибилности у функционисању ECF σ фактора иницирало је систематску анализу интеракција протеин-ДНК у групи, која је обухватила бактеријске представнике различитих ECF подгрупа, као и проширену анализу фагних ECF σ фактора, за које дискусија следи у наставку поглавља.

На основу поравнања промотора фагних и канонских ECF представника (σ^W и σ^E), у које је укључен и RpoD промотор из *E. coli* као добро проучен представник, уочено је да се код промотора фагних ECF σ фактора јављају најдужи продужеци -10 елемента. Наиме, -10 елементи канонских ECF σ фактора, σ^W и σ^E, интерагују са σ фактором у форми dsDNA посредством сегмента дужине 2 бп, док у RpoD промоторима одговарајући мотив (-15 елемент) има дужину 4 бп [18]. У односу на њих, продужеци -10 елемента у фагним ECF промоторима су готово дуплирани, будући да износе 5 и 7 бп.

У промоторима σ фактора phiEco32 врло изражен продужетак -10 елемента је праћен недостатком -35 елемента, што је у складу са одсуством домена σ_4 у протеинској секвенци. Будући да су узводно од -10 елемента у делу phiEco32 промотора предвиђени конзервирани мотиви, недостајуће σ_4:-35 интеракције могле би бити надомештене регрутовањем транскрипционог фактора. У том случају, комплементација одсутног домена σ_4 посредством интеракција са транскрипционим фактором представљала би још један (могући) пример флексибилности у функционисању ECF σ фактора, будући да је регулација активности ECF промотора помоћу транскриционог фактора опажена у врло малом броју случајева [7, 116]. Додатно, оваква регулација активности дела phiEco32 промотора имала би јасан значај и у успостављању временског обрасца експресије фагних гена (раздвајање средње и касне генске класе), будући да су само промотори касних гена удружени са узводним конзервираним мотивима [114].

Продужетке -10 елемента, који у фагним ECF промоторима достижу највећу дужину, RpoD σ фактори препознају у значајној мери посредством домена σ_3 [7], који одсуствује из протеинских секвенци ECF σ фактора. Препознавање функционалног аналога домена σ_3 у фагним ECF σ факторима омогућила је екстензивна компаративна анализа њихових протеинских секвенци са представницима различитих бактеријских подгрупа, којом је уочен конзервирани мотив у фагним (7-11) и њима најсроднијим бактеријским ECF σ факторима (подгрупа ECF28). Препознати конзервирани мотив представља директан продужетак домена σ_2, будући да се наставља на предвиђену границу домена у C-терминалној орјентацији, што је идеална позиција за интеракцију са продужетком -10 елемента. Овде је интересантно приметити да σ фактори RpoD групе продужетке -10 елемента препознају комбинованим интеракцијама C-терминалног дела домена σ_2 и N-терминалног дела домена σ_3, док се у групи ECF аналогни мотив чак веће дужине у потпуности акомодира посредством C-терминалног продужетка домена σ_2.

Додатна назнака улоге препознатог продужетка домена σ_2 у интеракцији са продужетком -10 елемента је одсуство овог конзервираног мотива у ECF подгрупама, чији представници не препознају продужене -10 елементе (σ^E и σ^W). Сходно томе, овај мотив представља специфично обележје фагних и σ фактора подгрупе ECF28, чијом анализом би могло значајно да се унапреди разумевање механизма функционисања у субфамилији, будући да је ECF28 у потпуности неизучена подгрупа бактеријских ECF σ фактора. Међутим, промотори које ови σ фактори препознају нису предвиђени ни *ab initio* претрагом регулаторних мотива, као ни надгледаном претрагом на основу матрица тежине за карактеристични ECF -35 елемент. Сходно томе, узрок неуспешног предвиђања промоторског специфицитета за чланове подгрупе ECF28 може бити изостанак ауторегулације или карактеристичног ECF -35 елемента из промоторских секвенци, где обе могућности представљају одступање од тренутне парадигме о механизму функционисања у субфамилији ECF. Овде треба приметити да ауторегулација, која у систем уводи позитивну повратну спрегу и тако повећава ефикасност транскрипционог одговора са ECF промотора, постаје сувишна у случају појаве других механизама за појачавање транскрипционе активности. Будући да је у подгрупи ECF28 предвиђен продужетак домена σ_2, изостанак ауторегулације био би логична последица интеракције

овог протеинског мотива са продужетком -10 елемента, услед постојања додатног енергетског доприноса укупној снази промотора.

Поред препознавања C-терминалног продужетка домена σ_2 као платформе за интеракцију са дугим продужецима -10 елемента, глобална компаративна анализа протеинских секвенци у групи ECF омогућила је уочавање додатних носилаца промоторског специфицитета, који интерагују са секвенцама између елемената -10 и -35 (тзв. конзервирани спејсерски мотиви). За разлику од групе RpoD, где су интеракције између промотора и σ фактора у спејсерском региону ограничене искључиво на продужетке домена σ_2 и -10 елемента, у групи ECF се новооткривене интеракције протежу дуж целог региона спејсера. Конкретно, у секвенцама чланова подгрупе ECF02 (којој припада и σ^E) уочен је конзервирани протеински мотив у близини, али не и директном наставку, C-терминалне границе домена σ_2, који по специфицитету наликује конзервираном мотиву предвиђеном у секвенцама σ фактора ECF28 и 7-11. Да уочени протеински мотив узима учешће у иницијацији транскрипције сугерише паралелно присуство конзервираног ДНК мотива у промоторима σ^E, који се јавља узводно од -10 елемента, али поново не као директан продужетак. Слични примери флексибилности (тј. "неканонске" интеракције) примећени су и у подгрупи ECF32 и делимично код фактора σ^W, при чему су за ECF32 σ факторе неке од датих интеракција и експериментално потврђене. У подгрупи ECF32 ови "неканонски" носиоци промоторског специфицитета локализовани су у близини N-терминалног дела домена σ_4 и низводно од -35 елемента, дакле имају супротан поларитет у односу на класичне продужетке домена σ_2/-10 елемента, који се јављају у RpoD групи [6, 18]. На крају, важно је нагласити да је предвиђени спејсерски мотив у σ^E промоторима јаче изражен у секвенцама са ниском транскрипционом активношћу (in-vitro *неактивни* промотори), на основу чега закључујемо да је појава датог мотива повезана са повећањем иначе ниске снаге промотора. Улогу у појачавању транскрипционе активности спејсерски мотив могао би да оствари комплементацијом снага канонских промоторских елемената, што је још један потенцијални показатељ експлоатисања механизма "mix-and-match" од стране ECF σ фактора.

Свеукупно, резултати анализе промоторског специфицитета у групи ECF надовезују се на предвиђања фагних ECF промотора, чиме су оправдане полазне назнаке које доводе у сумњу важење парадигме о ригидном препознавању промотора током иницијације

транскрипције. При том, присуство "неканонских" интеракција код бактеријских ECF σ фактора не представља само још једно у низу одступања од тренутне парадигме, већ и јасан показатељ веће квалитативне флексибилности у функционисању у односу на примарне RpoD σ факторе, тј. присуства разноврснијег репертоара промоторских елемената, преко којих се испољавају "mix-and-match" ефекти.

5.3. Механизам "mix-and-match" у групи ECF σ фактора

Тумачењем резултата анализе специфицитета за групу ECF потврђено је да механизам функционисања ECF σ фактора значајно одступа од парадигме ригидних интеракција са промоторским секвенцама, што је првобитно уочено на фагним промоторима phiEco32 у виду квалитативног испољавања механизма "mix-and-match". Овај драстичан пример функционалног надопуњавања промоторских елемената оповргава тренутно становиште по ком групе алтернативних σ^{70} фактора, услед упрошћене структуре протеинских секвенци и веће конзервације промоторских елемената, не испољавају "mix-and-match" механизам [7]. Међутим, резултати систематске анализе специфицитета у групи ECF указују да је репертоар промоторских елемената, чије снаге могу међусобно да се надопуњују, код алтернативних σ фактора чак и разноврснији, што пружа потенцијал за већу квантитативну флексибилност у функционисању у поређењу са примарним σ факторима. Тренутна парадигма, међутим, и ову могућност *a priori* не узима у обзир, јер опсег испољавања "mix-and-match" ефекта третира као директну последицу величине регулона, која је за алтернативне σ факторе значајно мања у поређењу са примарним σ факторима групе RpoD.

Насупрот претходном, резултати корелационе анализе на канонским ECF представницима [117] јасно указују да мање димензије регулона нису нужно асоциране са слабијим испољавањем "mix-and-match" ефекта, што је могућа последица додатних физиолошких ограничења која утичу на транскрипциони одговор промотора. У зависности од удела којим различити фактори учествују у обликовању функционалне активности промотора, остварени "mix-and-match" ефекат може значајно да се разликује у односу на максималну могућност за функционалну комплементацију промоторских елемената, која је за дати σ

фактор одређена опсегом промоторске варијабилности (тј. величином регулона). У складу са претходним, корелациони коефицијенти представљају показатељ физиолошког значаја механизма "mix-and-match" за функционисање тестираних σ фактора, што је у наставку објашњено увођењем једноставног квалитативног модела, који чини окосницу за тумачење резултата корелационе анализе за групу ECF (σ^E).

Корелациона анализа за σ^E промоторе даје (већином) статистички значајне корелационе коефицијенте, које по опсегу превазилазе вредности добијене у групи RpoD, где је механизам "mix-and-match" додатно валидиран на основу биохемијских мерења. Конкретно, најјача забележена корелација у групи ECF износи -0.52, док у групи RpoD корелациони коефицијенти не прелазе праг од -0.2. Као што је претходно истакнуто, тумачење овог наизглед неинтуитивног резултата, где боље конзеривирани промоторски елементи испољавају већи ниво функционалне комплементације, омогућава једноставан квалитативни модел (схематизован на Слици 19) који транскрипциону активност третира као објекат различитих физиолошких ограничења, услед чега скуп промоторских секвенци које дефинишу регулон σ фактора може да се подели на више категорија. За одређени удео промотора примарно ограничење може бити постизање врло јаке укупне транскрипционе активности, која искључује могућност појаве слабих промоторских елемената; жељени ниво транскрипционе активности за други скуп промотора може бити нижи, али уско дефинисаног опсега; док за део промоторских секвенци ограничавајући фактор може бити и афинитет везивања РНКП за дволанчану ДНК, што свеукупно доводи до врло прецизног нивелисања генске експресије у регулону. Од тога, механизам "mix-and-match" испољиће доминатан утицај на транскрипциону активност само оних промоторских секвенци, које су довољно близу линије раздвајања специфичне од неспецифичне интеракције ДНК секвенце са σ фактором, јер се у оваквим промоторима помоћу функционалне комплементације елемената вредности релевантних кинетичких параметара (а тиме и промоторска активност) одржавају изнад минималног нивоа.

У складу са претходним, простор са промоторским секвенцама подељен је на Слици 19 у три региона:

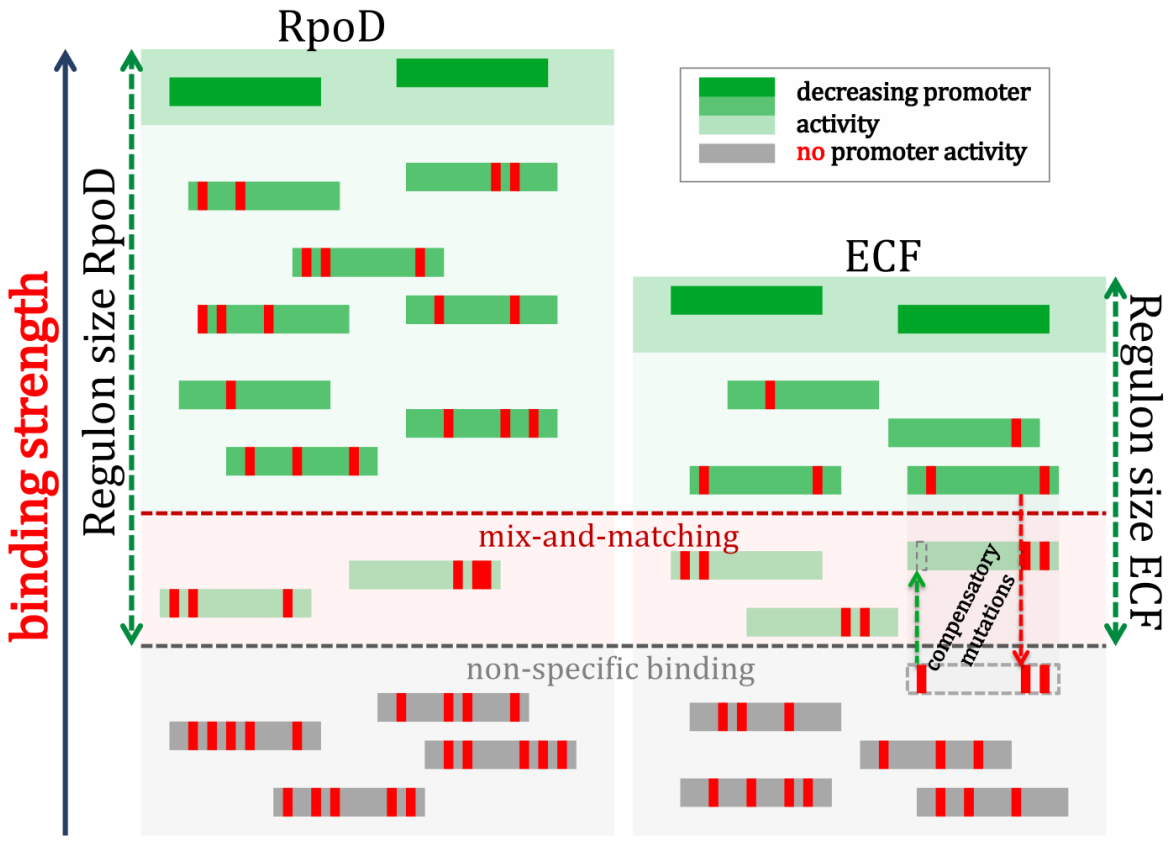
i) регион високе промоторске активности, где су промоторске секвенце далеко од границе неспецифичне интеракције са σ фактором и где акумулација мутација доводи до диференцирања на жељене нивое транскрипционе активности (приметити

на слици да увођење мутација у односу на консензусне промоторске елементе корелише са назначеним падом промоторске активности);

ii) регион "mix-and-match"-а, где су промоторске секвенце у близини границе неспецифичне интеракције са σ фактором, услед чега појава мутација у елементима брзо води ка изостанку транскрипционе активности. Очување успостављених транскрипционих веза у овом региону остварује се појавом компензаторних мутација (назначено на слици), које појачавањем енергије интеракције другог промоторског елемената са σ фактором одржавају ниво битног кинетичког параметра на вредности која гарантује транскрипциону активност.

iii) регион неспецифичне интеракције секвенце ДНК са σ фактором, где довољно велики број акумулираних мутација доводи до изостанка промоторске активности.

На слици се лако уочава да модел не предвиђа нужно јаче негативне корелације између промоторских елемената у регулонима већег распона, тј. мањег специфицитета (на слици илустровани помоћу σ фактора RpoD). Чак напротив, у сигмулонима мањег распона (на слици илустровани помоћу σ фактора ECF) већи удео промоторских секвенци наћи ће се у региону "mix-and-match"-а, услед постојања ужег "прозора" за акумулирање мутација, у ком промотори остају далеко од границе неспецифичне интеракције са σ фактором. Дакле, модел предвиђа јаче негативне корелације за σ факторе већег специфицитета, чиме објашњава однос добијених корелација у групама ECF и RpoD.



Слика 19. Веза између испољавања ефекта mix-and-match и специфицитета σ фактора: Приказани су регулони за σ фактор ниског (RpoD - лево) и високог специфицитета (ECF - десно); промоторске секвенце представљене су зеленим правоугаоницима, при чему најјача нијанса зелене боје одговара консензусним промоторским секвенцама. Мутације у промоторским елементима у односу на консензусну секвенцу назначене су вертикалним правоугаоницима. Зона "mix-and-match", у којој секвенце одржавају промоторску активност увођењем компензаторних мутација осенчена је црвеном бојом. Зона неспецифичне интеракције са σ фактором, у којој секвенце немају промоторску активност, означена је сивом бојом.

Сем правилне процене значаја механизма "mix-and-match" за функционисање алтернативних ECF σ фактора, поређење корелација добијених у овој групи са одговарајућим вредностима у групи RpoD, пружа увид у то како различита структурно-функционална ограничења за σ^{70} факторе утичу на кинетички профил промоторског одговора. Поређењем корелација између дволанчаних елемената у промоторима σ^{E}, *активним* in-vitro, и промоторима RpoD (-0.43 vs. -0.1) уочава се да је афинитет везивања РНКП за дволанчане промоторске елементе врло битна кинетичка одредница јаких ECF промотора, што се подудара са карактеристикама физиолошког одговора за који су у бактеријској ћелији задужени ECF σ фактори. Наиме, ови σ фактори покрећу брзе и уско усмерене реакције на сигнале везане за стресне услове, услед чега ефикасно регрутовање

ензима РНКП на промоторе постаје главна одредница транскрипционе активности. Насупрот овоме, у групи RpoD укупна транскрипциона активност одговара највећем нивоу функционалне комплементације промоторских елемената [18], што је такође у складу са физиологијом примарних σ фактора, где брза мобилизација транскрипционе машинерије, заснована на високом афинитету везивања за дволанчану ДНК, не представља кључни фактор у произношењу разноврсних физиолошких одговора са врло широког спектра RpoD промотора.

Механистичка подлога за уочене разлике у кинетичким профилима промоторског одговора у фамилији σ^{70} садржана је у функционалној комплементацији различитих комбинација промоторских елемената. У групи RpoD најјаче корелације уочавају се између -15 елемента и осталих промоторских елемената, док је комплементација канонских интеракција (-35 и -10 елемената) слабо изражена и без статистичког значаја [18]. При том, истакнута улога -15 елемента у RpoD промоторима може бити директна последица интеракција које остварује са доменом σ_3 у протеинским RpoD секвенцама [23]. С друге стране, у групи ECF корелације приближних вредности јављају се између свих промоторских елемената, што може бити у вези са једноставнијом структурном организацијом протеинских секвенци ових σ^{70} фактора, који поседују само два домена за интеракцију са молекулом ДНК (σ_2 и σ_4) [30].

Претходно је истакнуто да је афинитет везивања за dsDNA пресудан кинетички параметар у транскрипционој активности на јаким σ^E промоторима, о чему сведоче изражене негативне корелације између дволанчаних елемената. Уколико, међутим, кренемо од in-vitro *активних* ка in-vitro *неактивним* σ^E промоторима, ове корелације значајно опадају, за шта је највероватнији узрок зависност активности слабих промотора од спољашњих фактора, на које се преноси део комплементација са канонских промоторских елемената. Интересантно је, међутим, приметити да су у in-vitro *неактивним* секвенцама негативне корелације између једноланчаних и дволанчаних промоторских елемената и даље високе, што се може тумачити као последица притиска ка повећању инхерентно слабе транскрипционе активности у овој групи промоторских секвенци. Сличан тренд комплементације између једноланчаних и дволанчаних промоторских елемената уочен је и код промоторских секвенци које препознаје фактор σ^W.

Идеју да спољашњи фактори могу имати удела у функционалној комплементацији са канонским промоторским елементима, ради постизања одговарајуће вредности релевантног кинетичког параметра на слабим промоторима, подржавају резултати корелационе анализе са конзервираним спејсерским σ^E елементом. Најјаче негативне корелације уочене су управо у скупу in-vitro *неактивних* промотора, при чему се комплементација ка афинитету везивања за dsDNA одликује највећим релативним порастом, у поређењу са одговарајућим корелацијама добијеним између канонских елемената. У скупу in-vitro *неактивних* промотора јасан пораст корелација уочава се и за остале комбинације промоторских елемената, након укључивања спејсерског елемента у анализу, што додатно потврђује његов значај у иницијацији транскрипције са слабих σ^E промотора. У истом контексту треба тумачити и готово потпуно одсуство спејсерског σ^E елемента (и корелација асоцираним с њим) у промоторским секвенцама *активним* in-vitro, за које је уочен најјачи степен комплементације између канонских елемената, што регрутовање спољашњих фактора ради појачања транскрипционе активности чини излишним.

Будући да транскрипциона иницијација подразумева координисану активност већег броја компонената холоензимског комплекса, спољашњим доприносом за појачање промоторске активности могу се сматрати и интеракције централног дела ензима РНКП са промотором, попут интеракција субјединице α-CTD са узводним UP елементима [118], или субјединица β и β’ са промоторским низводним сегментима у дуплексној форми [119-120]. Овде не треба занемарити ни значај интрамолекуларних реаранжмана у нивоу субјединица β и β’, који на сличан могу стабилизовати формирање отвореног промоторског комплекса у комбинацији са различитим σ факторима [121], што додатно сведочи о универзалном карактеру механизма "mix-and-match" у фамилији σ^{70}. На крају, треба такође рећи да допринос транскрипционој активности може пружити и функционална комплементација промоторских елемената са везивним местима транскрипционих фактора, премда овакви примери регулације транскрипционе активности у групи ECF нису довољно изучени [122-123].

На почетку је истакнуто да добијени корелационе коефицијенти дају процену физиолошког значаја механизма "mix-and-match" за целисходно функционисање испитиваног σ фактора. Комплементаран приступ за изучавање овог механизма била би

биохемијска анализа помоћу транскрипционих in-vitro есеја, која полазећи од специфичне промоторске секвенце уводи у њене елементе појединачне базне измене, а потом прати да ли биофизички компензаторне мутације наново успостављају нарушену транскрипциону активност. Оваква анализа даје увид у механистичка ограничења за испољавање "mix-and-match" ефекта и већ је спроведена на промоторима примарних RpoD σ фактора. Њена примена на промоторе ECF групе, као и на промоторе преосталих алтернативних σ фактора, могла би да пружи додатни експериментални доказ у корист нове парадигме о флексибилним интеракцијама са промоторским секвенцама, као и да успостави "mix-and-match" као обједињујући механизам за препознавање промотора у целокупној σ^{70} фамилији.

6. Закључак

У докторској дисертацији изложени су резултати опсежне рачунске анализе механизма функционисања физиолошки важне, али слабо проучене, групе ECF σ^{70} фактора. Примењени приступ, који комбинује биоинформатичке методе са биофизичким моделом транскрипционе инцијације, указао је на нову парадигму о препознавању промотора у ECF σ субфамилији, која подразумева применљивост "mix-and-match" механизма. Будући да је овај механизам карактеристичан за примарне RpoD σ факторе, који су у σ^{70} фамилији најразличитији у односу на субфамилију ECF, добијени резултати су јака назнака универзалног присуства механизма "mix-and-match" у σ^{70} фамилији. Премда су анализом готово у потпуности обухваћени бактеријски σ фактори, важан део спроведен је на бактериофагним σ факторима, чиме је показано да ови организми и у будућности носе велики потенцијал за успостављање нових парадигми везаних за експресију генома.

Конкретно, биоинформатичка анализа геномске секвенце бактериофага 7-11 показала се као ефикасан начин за предвиђање вирусне инфективне стратегије, при чему је главни резултат препознавање промотора бактериофагног σ фактора. При овоме је осмишљен нов методолошки приступ, који је значајно боље прилагођен за *ab initio* предвиђање специфицитета фагних σ фактора, у односу на стандардне MLSA алгоритме. Будући да се заснива на поравнавању интергенских региона у паровима, процедура се једноставно имплементира помоћу широко заступљених биоинформатичких алгоритама (нпр. BLAST, LALIGN).

Откривени специфицитет ECF σ фактора 7-11 омогућио је да се по први пут уочи флексибилност у промоторској структури ECF σ фактора, односно присуство дугачког продужетка -10 елемента у бактериофагним промоторима, који компензује недостајући -35 елемент. Следственом детаљном анализом доступних протеинских секвенци бактеријских ECF σ фактора и њима одговарајућих промотора, предвиђене су протеин-ДНК интеракције изван канонских промоторских елемената, које нису ограничене на продужетке домена σ_2/-10 елемента, што представља већи ниво флексибилности у препознавању промотора, у односу на добро проучену групу примарних RpoD σ фактора.

Корелациона анализа снага промоторских ECF елемената (σ^E из *E. coli*), заснована на биофизичком моделу транскрипционе иницијације, и кантитативно је потврдила флексибилност при препознавању промотора у ECF σ^{70} субфамилији, тј. присуство механизма "mix-and-match", при чему су добијене јаче негативне корелације у односу на групу RpoD за коју је дати механизам добро установљен. Додатно, утврђено је да су у различитим σ^{70} групама различити кинетички параметри битни за транскрипциону активност промотора – тј. афинитет везивања за dsDNA у групи ECF, а укупна транскрипциона активност у групи RpoD – што се механистички постиже различитим комбинацијама промоторских елемената, који су укључени у "mix-and-match". Такође, код промотора са слабом базалном транскрипционом активношћу, спољашњи фактори, попут новооткривеног спејсерског елемента у промоторима σ^E, могу да надопуњавају снаге канонских промоторских елемената у смеру постизања довољног нивоа релевантног кинетичког параметра (афинитет везивања за dsDNA у промоторима σ^E *неактивним* in-vitro).

Додатно, резултати корелационе анализе мотивисали су успостављање модела, којим се повезује опсег испољавања "mix-and-match" ефекта у популацији промоторских секвенци и специфицитета σ фактора који дате промоторе препознаје (тј. величине контролисаног регулона). Модел се заснива на претпоставци да само за промоторске секвенце, које су у непосредној близини границе, која раздваја специфичну од неспецифичне интеракције промотора са σ фактором, "mix-and-match" постаје битан фактор за очување транскрипционе активности. Насупрот интиуитивним очекивањима, а у складу са резултатима корелационе анализе, модел предвиђа да је мања величина регулона асоцирана са већим опсегом испољавања ефекта "mix-and-match".

Сумарно, резултати анализе припадника ECF субфамилије, потврђују полазну хипотезу о универзалности механизма "mix-and-match" током иницијације транскрипције у фамилији σ^{70}, при чему се разлике у физиолошком одговору у оквиру фамилије постижу кроз различит допринос кинетичких параметара транскрипционом одговору промотора. Добијени резултати представљају полазну тачку за даља изучавања промоторског специфицитета и механизама функционисања физиолошки веома битних, али недовољно истражених, алтернативних σ фактора.

7. РЕФЕРЕНЦЕ

1. Balleza, E., et al., *Regulation by transcription factors in bacteria: beyond description.* FEMS Microbiol Rev, 2009. **33**(1): p. 133-51.
2. Browning, D.F. and S.J. Busby, *The regulation of bacterial transcription initiation.* Nat Rev Microbiol, 2004. **2**(1): p. 57-65.
3. Ruff, E.F., M.T. Record, Jr., and I. Artsimovitch, *Initial events in bacterial transcription initiation.* Biomolecules, 2015. **5**(2): p. 1035-62.
4. Burgess, R.R., et al., *Factor stimulating transcription by RNA polymerase.* Nature, 1969. **221**(5175): p. 43-6.
5. Reznikoff, W.S., et al., *The regulation of transcription initiation in bacteria.* Annu Rev Genet, 1985. **19**: p. 355-87.
6. Shultzaberger, R.K., et al., *Anatomy of Escherichia coli s 70 promoters.* Nucleic Acids Res, 2007. **35**(3): p. 771-788.
7. Feklístov, A., et al., *Bacterial sigma factors: a historical, structural, and genomic perspective.* Annual review of microbiology, 2014. **68**: p. 357-376.
8. Osterberg, S., T. del Peso-Santos, and V. Shingler, *Regulation of alternative sigma factor use.* Annu Rev Microbiol, 2011. **65**: p. 37-55.
9. Souza, B.M., et al., *sigma(ECF) factors of gram-positive bacteria: a focus on Bacillus subtilis and the CMNR group.* Virulence, 2014. **5**(5): p. 587-600.
10. Murakami, K.S. and S.A. Darst, *Bacterial RNA polymerases: the wholo story.* Curr Opin Struct Biol, 2003. **13**(1): p. 31-39.
11. Paget, M. and J. Helmann, *The sigma70 family of sigma factors.* Genome biology, 2003. **4**(1): p. 203.
12. Borukhov, S. and K. Severinov, *Role of the RNA polymerase sigma subunit in transcription initiation.* Res Microbiol, 2002. **153**(9): p. 557-562.
13. Harden, T.T., et al., *Bacterial RNA polymerase can retain sigma70 throughout transcription.* Proc Natl Acad Sci U S A, 2016. **113**(3): p. 602-7.
14. Mooney, R.A., S.A. Darst, and R. Landick, *Sigma and RNA polymerase: an on-again, off-again relationship?* Mol Cell, 2005. **20**(3): p. 335-45.
15. Gross, C.A., et al., *The functional and regulatory roles of sigma factors in transcription.* Cold Spring Harb Symp Quant Biol, 1998. **63**: p. 141-55.
16. Harley, C.B. and R.P. Reynolds, *Analysis of E. coli promoter sequences.* Nucleic Acids Res, 1987. **15**(5): p. 2343-2361.
17. Barne, K.A., et al., *Region 2.5 of the Escherichia coli RNA polymerase sigma70 subunit is responsible for the recognition of the 'extended-10' motif at promoters.* EMBO J., 1997. **16**(13): p. 4034-40.
18. Djordjevic, M., *Redefining Escherichia coli σ70 promoter elements:− 15 motif as a complement of the− 10 motif.* Journal of bacteriology, 2011. **193**(22): p. 6305-6314.
19. Hook-Barnard, I.G. and D.M. Hinton, *Transcription initiation by mix and match elements: flexibility for polymerase binding to bacterial promoters.* Gene Regulation and Systems Biology, 2007. **1**: p. 275.
20. Gruber, T.M. and C.A. Gross, *Multiple sigma subunits and the partitioning of bacterial transcription space.* Annu Rev Microbiol, 2003. **57**: p. 441-66.

21. Djordjevic, M. and R. Bundschuh, *Formation of the Open Complex by Bacterial RNA Polymerase—A Quantitative Model.* Biophysical Journal, 2008. **94**(11): p. 4233-4248.
22. Liu, X., D.A. Bushnell, and R.D. Kornberg, *Lock and key to transcription: sigma-DNA interaction.* Cell, 2011. **147**(6): p. 1218-9.
23. Paget, M.S., *Bacterial Sigma Factors and Anti-Sigma Factors: Structure, Function and Distribution.* Biomolecules, 2015. **5**(3): p. 1245-65.
24. Manganelli, R., *Sigma Factors: Key Molecules in Mycobacterium tuberculosis Physiology and Virulence.* Microbiol Spectr, 2014. **2**(1): p. MGM2-0007-2013.
25. Durre, P., *Physiology and Sporulation in Clostridium.* Microbiol Spectr, 2014. **2**(4): p. TBS-0010-2012.
26. Busby, S.J.W. *Investigations of the modular structure of bacterial promoters*. 2006.
27. Paget, M.S. and J.D. Helmann, *The sigma70 family of sigma factors.* Genome Biol, 2003. **4**(1): p. 203.
28. Battesti, A., N. Majdalani, and S. Gottesman, *The RpoS-mediated general stress response in Escherichia coli.* Annu Rev Microbiol, 2011. **65**: p. 189-213.
29. Schellhorn, H.E., *Elucidating the function of the RpoS regulon.* Future Microbiol, 2014. **9**(4): p. 497-507.
30. Staroń, A., et al., *The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family.* Mol Microbiol, 2009. **74**(3): p. 557-581.
31. Helmann, J.D., *The extracytoplasmic function (ECF) sigma factors.* Adv Microb Physiol, 2002. **46**: p. 47-110.
32. Ho, T.D. and C.D. Ellermeier, *Extra cytoplasmic function sigma factor activation.* Curr Opin Microbiol, 2012. **15**(2): p. 182-8.
33. Campagne, S., et al., *Structural basis for− 10 promoter element melting by environmentally induced sigma factors.* Nature structural & molecular biology, 2014. **21**(3): p. 269-276.
34. Lane, W.J. and S.A. Darst, *The structural basis for promoter− 35 element recognition by the group IV σ factors.* PLoS biology, 2006. **4**(9): p. e269.
35. Schulz, S., et al., *Elucidation of sigma factor-associated networks in Pseudomonas aeruginosa reveals a modular architecture with limited and function-specific crosstalk.* PLoS Pathog, 2015. **11**(3): p. e1004744.
36. Bashyam, M.D. and S.E. Hasnain, *The extracytoplasmic function sigma factors: role in bacterial pathogenesis.* Infect Genet Evol, 2004. **4**(4): p. 301-8.
37. Helmann, J.D., *Bacillus subtilis extracytoplasmic function (ECF) sigma factors and defense of the cell envelope.* Curr Opin Microbiol, 2016. **30**: p. 122-32.
38. Raivio, T.L. and T.J. Silhavy, *Periplasmic stress and ECF sigma factors.* Annu Rev Microbiol, 2001. **55**: p. 591-624.
39. Potvin, E., F. Sanschagrin, and R.C. Levesque, *Sigma factors in Pseudomonas aeruginosa.* FEMS Microbiol Rev, 2008. **32**(1): p. 38-55.
40. Brooks, B.E. and S.K. Buchanan, *Signaling mechanisms for activation of extracytoplasmic function (ECF) sigma factors.* Biochim Biophys Acta, 2008. **1778**(9): p. 1930-45.
41. Newton-Foot, M. and N.C. Gey van Pittius, *The complex architecture of mycobacterial promoters.* Tuberculosis (Edinb), 2013. **93**(1): p. 60-74.

42. Radeck, J., G. Fritz, and T. Mascher, *The cell envelope stress response of Bacillus subtilis: from static signaling devices to dynamic regulatory network.* Curr Genet, 2016.
43. Feklistov, A., *RNA polymerase: in search of promoters.* Ann N Y Acad Sci, 2013. **1293**: p. 25-32.
44. Zhou, D. and R. Yang, *Global analysis of gene transcription regulation in prokaryotes.* Cellular and Molecular Life Sciences CMLS, 2006. **63**(19-20): p. 2260-2290.
45. Murakami, K.S., et al., *Structural Basis of Transcription Initiation: An RNA Polymerase Holoenzyme-DNA Complex.* Science, 2002. **296**(5571): p. 1285-1290.
46. Hook-Barnard, I.G. and D.M. Hinton, *The promoter spacer influences transcription initiation via sigma70 region 1.1 of Escherichia coli RNA polymerase.* Proc Natl Acad Sci U S A, 2009. **106**(3): p. 737-42.
47. Feklistov, A. and S.A. Darst, *Structural basis for promoter-10 element recognition by the bacterial RNA polymerase sigma subunit.* Cell, 2011. **147**(6): p. 1257-69.
48. Zhang, Y., et al., *Structural basis of transcription initiation.* Science, 2012. **338**(6110): p. 1076-80.
49. Haugen, S.P., et al., *Fine structure of the promoter-sigma region 1.2 interaction.* Proc Natl Acad Sci U S A, 2008. **105**(9): p. 3292-7.
50. Chen, H., H. Tang, and R.H. Ebright, *Functional interaction between RNA polymerase alpha subunit C-terminal domain and sigma70 in UP-element- and activator-dependent transcription.* Mol Cell, 2003. **11**(6): p. 1621-33.
51. Gourse, R.L., W. Ross, and T. Gaal, *UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition.* Mol Microbiol, 2000. **37**(4): p. 687-95.
52. Vvedenskaya, I.O., et al., *Interactions between RNA polymerase and the core recognition element are a determinant of transcription start site selection.* Proc Natl Acad Sci U S A, 2016. **113**(21): p. E2899-905.
53. Petushkov, I., et al., *Mutations in the CRE pocket of bacterial RNA polymerase affect multiple steps of transcription.* Nucleic Acids Res, 2015. **43**(12): p. 5798-809.
54. Thouvenot, B., B. Charpentier, and C. Branlant, *The strong efficiency of the Escherichia coli gapA P1 promoter depends on a complex combination of functional determinants.* Biochem J, 2004. **383**(Pt 2): p. 371-82.
55. Hook-Barnard, I., X.B. Johnson, and D.M. Hinton, *Escherichia coli RNA polymerase recognition of a sigma70-dependent promoter requiring a -35 DNA element and an extended -10 TGn motif.* J Bacteriol, 2006. **188**(24): p. 8352-9.
56. Hershey, A.D. and M. Chase, *Independent functions of viral protein and nucleic acid in growth of bacteriophage.* J Gen Physiol, 1952. **36**(1): p. 39-56.
57. Johnson, A.D., et al., *lambda Repressor and cro--components of an efficient molecular switch.* Nature, 1981. **294**(5838): p. 217-23.
58. Ptashne, M. and N. Hopkins, *The operators controlled by the lambda phage repressor.* Proc Natl Acad Sci U S A, 1968. **60**(4): p. 1282-7.
59. Eisen, H., et al., *Regulation of repressor expression in lambda.* Proc Natl Acad Sci U S A, 1970. **66**(3): p. 855-62.
60. Hatfull, G.F. and R.W. Hendrix, *Bacteriophages and their genomes.* Curr Opin Virol, 2011. **1**(4): p. 298-303.
61. Pallen, M.J., *Microbial bioinformatics 2020.* Microb Biotechnol, 2016. **9**(5): p. 681-6.

62. Zhang, J., et al., *Complete genomic sequence of the Vibrio alginolyticus lytic bacteriophage PVA1.* Arch Virol, 2014. **159**(12): p. 3447-51.
63. Li, Y., et al., *Complete Genomic Sequence of Bacteriophage H188: A Novel Vibrio kanaloae Phage Isolated from Yellow Sea.* Curr Microbiol, 2016. **72**(5): p. 628-33.
64. Pavlova, O., et al., *Temporal regulation of gene expression of the Escherichia coli bacteriophage phiEco32.* Journal of molecular biology, 2012. **416**(3): p. 389-399.
65. Hinton, D.M., *Transcriptional control in the prereplicative phase of T4 development.* Virol J, 2010. **7**: p. 289.
66. Shea, M.A. and G.K. Ackers, *The OR control system of bacteriophage lambda. A physical-chemical model for gene regulation.* J Mol Biol, 1985. **181**(2): p. 211-230.
67. Yang, H., et al., *Transcription regulation mechanisms of bacteriophages: recent advances and future prospects.* Bioengineered, 2014. **5**(5): p. 300-4.
68. Semenova, E., et al., *The tale of two RNA polymerases: transcription profiling and gene expression strategy of bacteriophage Xp10.* Molecular Microbiology, 2005. **55**(3): p. 764.
69. Guzina, J. and M. Djordjevic, *Bioinformatics as a first-line approach for understanding bacteriophage transcription.* Bacteriophage, 2015. **5**(3): p. e1062588.
70. Joerger, R.D., *Alternatives to antibiotics: bacteriocins, antimicrobial peptides and bacteriophages.* Poult Sci, 2003. **82**(4): p. 640-7.
71. Haq, I.U., et al., *Bacteriophages and their implications on future biotechnology: a review.* Virol J, 2012. **9**: p. 9.
72. Cisek, A.A., et al., *Phage Therapy in Bacterial Infections Treatment: One Hundred Years After the Discovery of Bacteriophages.* Curr Microbiol, 2016.
73. Djordjevic, M., et al., *Quantitative analysis of a virulent bacteriophage transcription strategy.* Virology, 2006. **354**(2): p. 240-51.
74. Ishihama, A., *Prokaryotic genome regulation: multifactor promoters, multitarget regulators and hierarchic networks.* FEMS Microbiol Rev, 2010. **34**(5): p. 628-45.
75. van Nimwegen, E., *Finding regulatory elements and regulatory motifs: a general probabilistic framework.* BMC Bioinformatics, 2007. **8 Suppl 6**: p. S4.
76. GuhaThakurta, D., *Computational identification of transcriptional regulatory elements in DNA sequence.* Nucleic Acids Res, 2006. **34**(12): p. 3585-98.
77. Sevostyanova, A., et al., *Specific recognition of the-10 promoter element by the free RNA polymerase sigma subunit.* J Biol Chem, 2007. **282**(30): p. 22033-9.
78. Gama-Castro, S., et al., *RegulonDB version 7.0: transcriptional regulation of Escherichia coli K-12 integrated within genetic sensory response units (Gensor Units).* Nucleic acids research, 2011. **39**(suppl 1): p. D98.
79. Stormo, G.D., *Modeling the specificity of protein-DNA interactions.* Quant Biol, 2013. **1**(2): p. 115-130.
80. Stormo, G.D., *DNA binding sites: representation and discovery.* Bioinformatics, 2000. **16**(1): p. 16-23.
81. Mathelier, A. and W.W. Wasserman, *The next generation of transcription factor binding site prediction.* PLoS Comput Biol, 2013. **9**(9): p. e1003214.
82. Berg, O.G. and P.H. von Hippel, *Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters.* J Mol Biol, 1987. **193**(4): p. 723-50.

83. Mehta, P., D.J. Schwab, and A.M. Sengupta, *Statistical Mechanics of Transcription-Factor Binding Site Discovery Using Hidden Markov Models.* J Stat Phys, 2011. **142**(6): p. 1187-1205.
84. Benos, P.V., M.L. Bulyk, and G.D. Stormo, *Additivity in protein–DNA interactions: how good an approximation is it?* Nucleic Acids Res, 2002. **30**(20): p. 4442-4451.
85. Benos, P.V., A.S. Lapedes, and G.D. Stormo, *Is there a code for protein-DNA recognition? Probab(ilistical)ly.* Bioessays, 2002. **24**(5): p. 466-75.
86. Stormo, G.D. and G.W. Hartzell, 3rd, *Identifying protein-binding sites from unaligned DNA fragments.* Proc Natl Acad Sci U S A, 1989. **86**(4): p. 1183-7.
87. Sheng, Q., et al., *Applications of Gibbs sampling in bioinformatics.* Optimization Methods and Software, 2005.
88. Thijs, G., *Probabilistic methods to search for regulatory elements in sets of coregulated genes.* Leuven: Faculty of Applied Sciences, Katholieke Universiteit Leuven, 2003.
89. Das, M.K. and H.K. Dai, *A survey of DNA motif finding algorithms.* BMC Bioinformatics, 2007. **8 Suppl 7**: p. S21.
90. Bailey, T.L. and C. Elkan, *Unsupervised learning of multiple motifs in biopolymers using expectation maximization.* Machine learning, 1995. **21**(1-2): p. 51-80.
91. Lawrence, C. and S. Altschul, *Detecting subtle sequence signals: A Gibbs sampling strategy for multiple alignment.* Science, 1993. **262**(5131): p. 208.
92. Larkin, M.A., et al., *Clustal W and Clustal X version 2.0.* Bioinformatics, 2007. **23**(21): p. 2947-2948.
93. Djordjevic, M., *Integrating sequence analysis with biophysical modelling for accurate transcription start site prediction.* J Integr Bioinform, 2014. **11**(2): p. 240.
94. Gordon, J.J., et al., *Improved prediction of bacterial transcription start sites.* Bioinformatics, 2006. **22**(2): p. 142-8.
95. Robison, K., A. McGuire, and G. Church, *A comprehensive library of DNA-binding site matrices for 55 proteins applied to the complete Escherichia coli K-12 genome.* Journal of molecular biology, 1998. **284**(2): p. 241-254.
96. Stormo, G.D. and D.S. Fields, *Specificity, free energy and information content in protein-DNA interactions.* Trends Biochem. Sci, 1998. **23**: p. 109–113.
97. Wagner, R., *Transcription regulation in prokaryotes*. 2000, Oxford: Oxford University Press.
98. Rhodius, V.A. and V.K. Mutalik, *Predicting strength and function for promoters of the Escherichia coli alternative sigma factor, σE.* Proceedings of the National Academy of Sciences, 2010. **107**(7): p. 2854-2859.
99. Ishii, T., et al., *DBTBS: a database of Bacillus subtilis promoters and transcription factors.* Nucleic acids research, 2001. **29**(1): p. 278-280.
100. Guzina, J. and M. Djordjevic, *Inferring bacteriophage infection strategies from genome sequence: analysis of bacteriophage 7-11 and related phages.* BMC evolutionary biology, 2015. **15**(Suppl 1): p. S1.
101. Staroń, A., et al., *The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) σ factor protein family.* Molecular microbiology, 2009. **74**(3): p. 557-581.
102. Liu, X., D. Brutlag, and J. Liu. *BioProspector: discovering conserved DNA motifs in upstream regulatory regions of co-expressed genes*. 2001.

103. Tatusova, T.A. and T.L. Madden, *BLAST 2 Sequences, a new tool for comparing protein and nucleotide sequences.* FEMS microbiology letters, 1999. **174**(2): p. 247-250.
104. Workman, C.T., et al., *enoLOGOS: a versatile web tool for energy normalized sequence logos.* Nucleic Acids Res, 2005. **33**(Web Server issue): p. W389-92.
105. Marchler-Bauer, A. and S.H. Bryant, *CD-Search: protein domain annotations on the fly.* Nucleic acids research, 2004. **32**(suppl 2): p. W327-W331.
106. Hertz, G.Z. and G.D. Stormo, *Identifying DNA and protein patterns with statistically significant alignments of multiple sequences.* Bioinformatics, 1999. **15**: p. 563-577.
107. Kropinski, A.M., E.J. Lingohr, and H.W. Ackermann, *The genome sequence of enterobacterial phage 7-11, which possesses an unusually elongated head.* Arch Virol, 2011. **156**(1): p. 149-51.
108. Thompson, W., E.C. Rouchka, and C.E. Lawrence, *Gibbs Recursive Sampler: finding transcription factor binding sites.* Nucleic Acids Res, 2003. **31**(13): p. 3580-3585.
109. Dodd, I.B., K.E. Shearwin, and J.B. Egan, *Revisited gene regulation in bacteriophage lambda.* Curr Opin Genet Dev, 2005. **15**(2): p. 145-52.
110. Oppenheim, A.B., et al., *Switches in bacteriophage lambda development.* Annu Rev Genet, 2005. **39**: p. 409-29.
111. Borukhov, S. and E. Nudler, *RNA polymerase holoenzyme: structure, function and biological implications.* Curr Opin Microbiol, 2003. **6**(2): p. 93-100.
112. Nissan, G., et al., *Analysis of promoters recognized by HrpL, an alternative sigma-factor protein from Pantoea agglomerans pv. gypsophilae.* Mol Plant Microbe Interact, 2005. **18**(7): p. 634-43.
113. Djordjevic, M., A.M. Sengupta, and B.I. Shraiman, *A biophysical approach to transcription factor binding site discovery.* Genome Res, 2003. **13**(11): p. 2381-2390.
114. Guzina, J. and M. Djordjevic, *Promoter recognition by ECF sigma factors: analyzing DNA and protein interaction motifs.* J Bacteriol, 2016.
115. Campagne, S., et al., *Structural basis for -10 promoter element melting by environmentally induced sigma factors.* Nat Struct Mol Biol, 2014. **21**(3): p. 269-76.
116. Abellon-Ruiz, J., et al., *The CarD/CarG regulatory complex is required for the action of several members of the large set of Myxococcus xanthus extracytoplasmic function sigma factors.* Environ Microbiol, 2014. **16**(8): p. 2475-90.
117. Guzina, J. and M. Djordjevic, *Mix-and-matching as a promoter recognition mechanism by ECF sigma factors.* BMC Evol Biol, 2017. **17**(Suppl 1): p. 12.
118. Estrem, S.T., et al., *Bacterial promoter architecture: subsite structure of UP elements and interactions with the carboxy-terminal domain of the RNA polymerase α subunit.* Genes & development, 1999. **13**(16): p. 2134-2147.
119. Mekler, V., et al., *Coupling of Downstream RNA Polymerase–Promoter Interactions with Formation of Catalytically Competent Transcription Initiation Complex.* Journal of molecular biology, 2014. **426**(24): p. 3973-3984.
120. Mekler, V., L. Minakhin, and K. Severinov, *A critical role of downstream RNA polymerase-promoter interactions in the formation of initiation complex.* Journal of Biological Chemistry, 2011. **286**(25): p. 22600-22608.
121. Chakraborty, A., et al., *Opening and closing of the bacterial RNA polymerase clamp.* Science, 2012. **337**(6094): p. 591-595.

122. Abellón-Ruiz, J., et al., *The CarD/CarG regulatory complex is required for the action of several members of the large set of Myxococcus xanthus extracytoplasmic function σ factors.* Environmental microbiology, 2014. **16**(8): p. 2475-2490.
123. Guzina, J. and M. Djordjevic, *Inferring bacteriophage infection strategies from genome sequence: analysis of bacteriophage 7-11 and related phages.* BMC evolutionary biology, 2015. **15**(1): p. 1.

БИОГРАФИЈА КАНДИДАТА

Јелена Гузина рођена је 28.02.1990. у Требињу (БиХ). Школске 2008/2009. уписала је Биолошки факултет Универзитета у Београду, смер Молекуларна биологија и физиологија, који је завршила са просечном оценом 9.76. По завршетку мастер студија на Биолошком факултету са просеком 9.86, уписала је мултидисциплинарне докторске студије Биофизике при Универзитету у Београду, где је положила све испите са просеком 10.0. Током основних и мастер студија била је добитник већег броја Републичких и Градских стипендија за најбоље студенте. Од априла 2014. стипендиста је за докторске студије Министарства просвете, науке и технолошког развоја, када је и бирана у истраживача приправника на Биолошком факултету. На истом факултету запослена је од марта 2015. године, где је у јулу 2015. изабрана у истраживача сарадника и где је сарадник у настави на три предмета (Биоинформатика, Основи системске биофизике, Физика). Као члан организационог одбора учествовала је у припреми међународне конференције из биоинформатике BelBi2016, одржане јуна 2016. у Београду. Сем на научним скуповима (TABIS2013 Београд, RBC2014 Smolenice, Bacteriophages2015 London, PROKAGenomics2015 Goettingen, BelBi2016 Београд, BGRS2016 Novosibirsk), резултате истраживања излагала је и на домаћим семинарима и колоквијумима (Биоинформатички семинар на МАТФ-у, Недеља биофизике на Коларцу). Добитник је и награде за научноистраживачки рад младог истраживача на Универзитету у Београду – Биолошком факултету у школској 2016/2017.

Изјава о ауторству

Име и презиме аутора Јелена Гузина

Број индекса 24/2013

Изјављујем

да је докторска дисертација под насловом

"Биоинформатичка анализа механизама транскрипционе иницијације код бактеријских ECF σ фактора"

- резултат сопственог истраживачког рада;
- да дисертација у целини ни у деловима није била предложена за стицање друге дипломе према студијским програмима других високошколских установа;
- да су резултати коректно наведени и
- да нисам кршио/ла ауторска права и користио/ла интелектуалну својину других лица.

Потпис аутора

У Београду, 19.05.2017.

Јелена Гузина

Изјава о истоветности штампане и електронске верзије докторског рада

Име и презиме аутора Јелена Гузина

Број индекса 24/2013

Студијски програм Биофизика

Наслов рада "Биоинформатичка анализа механизама транскрипционе иницијације код бактеријских ECF σ фактора"

Ментор проф. др Марко Ђорђевић, др Магдалена Ђорђевић

Изјављујем да је штампана верзија мог докторског рада истоветна електронској верзији коју сам предао/ла ради похрањена у **Дигиталном репозиторијуму Универзитета у Београду.**

Дозвољавам да се објаве моји лични подаци везани за добијање академског назива доктора наука, као што су име и презиме, година и место рођења и датум одбране рада.

Ови лични подаци могу се објавити на мрежним страницама дигиталне библиотеке, у електронском каталогу и у публикацијама Универзитета у Београду.

Потпис аутора

У Београду, 19.05.2017.

Јелена Гузина

Изјава о коришћењу

Овлашћујем Универзитетску библиотеку „Светозар Марковић“ да у Дигитални репозиторијум Универзитета у Београду унесе моју докторску дисертацију под насловом:

"Биоинформатичка анализа механизама транскрипционе иницијације код бактеријских ECF σ фактора"

која је моје ауторско дело.

Дисертацију са свим прилозима предао/ла сам у електронском формату погодном за трајно архивирање.

Моју докторску дисертацију похрањену у Дигиталном репозиторијуму Универзитета у Београду и доступну у отвореном приступу могу да користе сви који поштују одредбе садржане у одабраном типу лиценце Креативне заједнице (Creative Commons) за коју сам се одлучио/ла.

1. Ауторство (CC BY)

2. Ауторство – некомерцијално (CC BY-NC)

3. Ауторство – некомерцијално – без прерада (CC BY-NC-ND)

4. Ауторство – некомерцијално – делити под истим условима (CC BY-NC-SA)

5. Ауторство – без прерада (CC BY-ND)

6. Ауторство – делити под истим условима (CC BY-SA)

(Молимо да заокружите само једну од шест понуђених лиценци.
кратак опис лиценци је саставни део ове изјаве).

Потпис аутора

У Београду, 19.05.2017.

1. **Ауторство**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце, чак и у комерцијалне сврхе. Ово је најслободнија од свих лиценци.

2. **Ауторство – некомерцијално**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела.

3. **Ауторство – некомерцијално – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца не дозвољава комерцијалну употребу дела. У односу на све остале лиценце, овом лиценцом се ограничава највећи обим права коришћења дела.

4. **Ауторство – некомерцијално – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца не дозвољава комерцијалну употребу дела и прерада.

5. **Ауторство – без прерада**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, без промена, преобликовања или употребе дела у свом делу, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце. Ова лиценца дозвољава комерцијалну употребу дела.

6. **Ауторство – делити под истим условима**. Дозвољавате умножавање, дистрибуцију и јавно саопштавање дела, и прераде, ако се наведе име аутора на начин одређен од стране аутора или даваоца лиценце и ако се прерада дистрибуира под истом или сличном лиценцом. Ова лиценца дозвољава комерцијалну употребу дела и прерада. Слична је софтверским лиценцама, односно лиценцама отвореног кода.