

Равчеев Дмитрий Андреевич

**Изучение эволюции регуляторных систем прокариот
методами сравнительно-геномного анализа**

Специальность 03.00.28 – “Биоинформатика”

Диссертация на соискание ученой степени
кандидата биологических наук

Научный руководитель :

кандидат физико-математических наук,
доктор биологических наук, профессор
М.С. Гельфанд

Оглавление

Введение	4
Глава 1. Обзор литературы	8
1.1. Общие принципы регуляции транскрипции бактерий	8
1.1.1. РНК-полимераза: структура и взаимодействие с промотором	8
1.1.2. Факторы транскрипции	10
1.2. Современные методы сравнительной геномики	12
1.2.1. Предсказание функций генов на основе сравнения аминокислотных последовательностей	14
1.2.2. Кластеризация генов на хромосоме	16
1.2.3. Слияние генов	16
1.2.4. Профили встречаемости генов	16
1.2.5. Методы распознавания потенциальных регуляторных сайтов	17
1.2.6. Сравнительная геномика и изучение регуляции	18
1.3. FruR (Cra) – регулятор центрального метаболизма	22
1.4. Регуляция утилизации рибозы	26
1.5. PurR – регулятор биосинтеза пуриновых нуклеотидов	27
1.6. Глобальная регуляция генов дыхания	30
1.6.1. Общие принципы устройства дыхательных цепей бактерий	30
1.6.2. Особенности регуляции дыхания <i>E. coli</i>	32
1.6.3. Fnr: ответ на молекулярный кислород	33
1.6.4. Двухкомпонентная система ArcB-ArcA: ответ на окислительно-восстановительный статус хинонов	34
1.6.5. Регуляция нитрат-нитритного дыхания: двухкомпонентные системы NarX-NarL и NarQ-NarP	37
Глава 2. Материалы и методы	40
2.1. Общие принципы сравнительного подхода к регуляции	40
2.2. Объект исследования и банки данных	41
2.3. Программное обеспечение	42
Глава 3. Исследование эволюции обобщенного FruR (Cra)-регулона	44
3.1. Изучение эволюции регуляторной системы	44
3.2. Построение распознающего правила для поиска потенциальных сайтов связывания FruR	44
3.3. Структура обобщенного FruR-регулона в исследованных геномах	47
3.3.1. Гены белков фосфотрансферазных систем	52
3.3.2. Гены ферментов центрального метаболизма	53

3.3.3. Гены белков дыхательных комплексов	55
3.3.4. Гены ферментов ассимиляции азота	56
3.3.5. Гены транспортных белков	57
3.3.6. Гены регуляторных белков	57
3.4. Эволюция FruR-регулона	58
3.5. Обсуждение и выводы	61
Глава 4. Исследование эволюции обобщенных PurR- и RbsR-регулонов	64
4.1. Изучение эволюции регуляторных систем	64
4.2. Исследование RbsR-зависимой регуляции	65
4.2.1. Построение распознающего правила для поиска потенциальных сайтов связывания RbsR	65
4.2.2. Структура обобщенного RbsR-регулона	67
4.3. Исследование PurR-зависимой регуляции	68
4.3.1. Построение распознающего правила для поиска потенциальных сайтов связывания PurR	68
4.3.2. Структура обобщенного PurR-регулона	70
4.3.2.1. Синтез пуриновых нуклеотидов	70
4.3.2.2. Синтез пиримидиновых нуклеотидов	74
4.3.2.3. Метаболизм азота	74
4.3.2.4. Метаболизм одноуглеродных фрагментов	74
4.3.2.5. Транспортные белки	77
4.3.2.6. Утилизация нуклеотидов и нуклеиновых кислот	80
4.3.2.7. Центральная метаболизм	80
4.3.2.8. Белки с неизвестной функцией	80
4.3.3. Таксон-специфические особенности PurR-регуляции	80
4.4. Исследование регуляции в Pseudomonadales	83
4.4.1. Ген регуляторного белка из Pseudomonadales	83
4.4.2. Построение распознающего правила для поиска потенциальных сайтов RbsR Pseudomonadales	84
4.4.3. Структура обобщенного регулона в Pseudomonadales	84
4.5. Эволюция PurR- и RbsR-регулонов	85
4.6. Обсуждение и выводы	87
Глава 5. Эволюция глобальной регуляции дыхания	90
5.1. Эволюция регуляторных систем	90
5.1.1. Однокомпонентная регуляторная система Fnr	90
5.1.2. Двухкомпонентная система ArcB-ArcA	90
5.1.3. Регуляция нитрат-нитритного дыхания: удвоенная двухкомпонентная система	

NarX-NarL и NarQ-NarP	93
5.2. Построение распознающих правил для поиска сайтов связывания регуляторов дыхания	98
5.2.1. Построение распознающего правила для поиска потенциальных сайтов связывания Fnr	98
5.2.2. Построение распознающего правила для поиска потенциальных сайтов связывания ArcA	101
5.2.3. Построение распознающего правила для поиска потенциальных сайтов связывания NarP	103
5.3. Состав обобщенных Fnr-, ArcA- и NarP-регулонов	104
5.3.1. Белки дыхательных цепей	104
5.3.2. Биосинтез молибдоптеринового кофактора	113
5.3.3. Центральный метаболизм и брожение	113
5.3.4. Метаболизм углеводов	114
5.3.5. Метаболизм жирных кислот	115
5.3.6. Ответ на кислородный стресс	116
5.3.7. Нуклеотидредуктазы	116
5.3.8. Транспортные белки	117
5.3.9. Пептидазы	117
5.3.10. Регуляторы транскрипции	117
5.4. Таксон-специфические особенности глобальной регуляции дыхания	118
5.4.1. Состав обобщенных регулонов в различных таксонах	118
5.4.2. Структура регуляторных каскадов в разных таксонах	119
5.4.3. Регуляция в отдельных таксонах	122
5.5. Обсуждение и выводы	125
Выводы	127
Список работ, опубликованных по теме диссертации	128
Благодарности	130
Список литературы	131
Приложения	162

Введение

Бактерии известны своей способностью приспосабливаться к различным условиям и занимать самые разнообразные экологические ниши. Подобная приспособляемость достигается за счет способности достаточно быстро отвечать на изменение окружающих условий и физиологического состояния клетки, чему микроорганизмы обязаны системе регуляции экспрессии собственных генов. Подобная регуляция осуществляется сразу на многих уровнях: транскрипции, трансляции, ковалентной модификации белков и аллостерической регуляции. Если аллостерическая регуляция позволяет быстро среагировать на резко меняющиеся условия существования, то в основе более эффективного использования имеющихся ресурсов лежит регуляция экспрессии генов на уровне транскрипции. Такая регуляция осуществляется с участием как необходимых компонентов транскрипции, так и дополнительных белков, называемых факторами транскрипции. В настоящее время в молекулярной биологии и смежных областях преобладает мнение, что для жизни микроорганизмов важны не только собственно гены, содержащиеся в геноме, но и профили их регуляции. Именно регуляция позволяет эффективно использовать имеющиеся гены в зависимости от потребностей клетки. В виду столь высокой роли регуляции в жизни микроорганизмов, исследование регуляции в различных бактериях и сравнительный анализ последних позволяют сделать выводы об эволюции как отдельных функциональных систем клетки, так и организмов в целом.

Долгое время изучение регуляции транскрипции осуществлялось только лишь экспериментально. При этом исследования сосредотачивались, как правило, на транскрипции индивидуальных генов, что позволяло собрать массу необходимых сведений, но не давало полной картины регуляторных взаимодействий. Современные методы массового анализа, такие как метод микрочипов или иммунопреципитация хроматина, позволяют исследовать экспрессию сотен и даже тысяч генов, но имеют массу существенных недостатков. Во-первых, для данных методов характерен относительно высокий уровень шума, а во-вторых, благодаря им можно получить лишь косвенные подтверждения регуляции, такие как влияние мутации в гене белка регулятора на уровень экспрессии гена или связывание белка с регуляторной областью.

В последний десяток лет в руках исследователей появился новый мощный инструмент для изучения регуляции, в особенности, у бактерий – методы сравнительного анализа последовательностей геномов. Его использованию способствуют высокие темпы роста количества бактериальных геномов с известной полной последовательностью. В 1995 году впервые была опубликована полная последовательность бактериального генома – это был

геном облигатного паразита, возбудителя менингита, хронического бронхита и других болезней, *Haemophilus influenzae* Rd KW20 [1]. С тех пор определение полной последовательности (секвенирование) геномов стало происходить нарастающими темпами, превратившись в настоящее время в широко развитую индустрию. Так, в базе данных KEGG ([2, 3]) к концу 2008-го года насчитывалось 740 последовательностей полных геномов, причем только за один 2008-й год появилось 167 новых последовательностей. При этом растет и количество геномов, относящихся к одной таксономической группе, часто даже таксонов такого низкого ранга, как вид. Так, к концу 2008-го года известны последовательности геномов 16 штаммов *Escherichia coli*. Понятно, что для анализа такого количества геномов недостаточно одних лишь экспериментальных методов, и необходимо использование биоинформатических подходов.

Здесь следует остановиться на взаимоотношениях биоинформатики и экспериментальной молекулярной биологии. Биоинформатика, как наука, изучающая последовательности нуклеиновых кислот и белков [4], получает от молекулярной биологии собственно последовательность, и, зачастую, ее аннотацию – описание функций определенных участков последовательности. Однако, непрерывный рост полных последовательностей геномов делает невозможной экспериментальную аннотацию всех последовательностей. Поэтому в настоящее время большинство функционально значимых участков последовательностей аннотируются именно методами биоинформатики. Этими методами можно получить информацию о таких функционально значимых участках, как гены, регуляторные сайты, белковые мотивы и других.

В настоящее время аннотация новых геномных последовательностей как правило осуществляется практически исключительно биоинформатическими методами. Большую популярность среди биологов получили такие программы, как BLAST [5] для сравнения последовательностей, CLUSTAL [6] и MUSCLE [7] для множественного выравнивания и выделения функциональных участков, Mfold [8] для предсказания вторичной структуры РНК, TMHMM [9] для идентификации трансмембранных сегментов в белках и другие.

Что касается изучения регуляции методами биоинформатики, то основной ее задачей является выявление последовательностей, ответственных за регуляцию генов: промоторов и терминаторов транскрипции, сайтов связывания регуляторных белков, последовательностей потенциальных белков-регуляторов. В настоящее время изучение регуляции методами биоинформатики распространено крайне широко, и зачастую используется самими экспериментальными биологами, в качестве предварительного исследования или дополнения к эксперименту [10-14].

В настоящее время активно используется и изучение регуляции исключительно методами биоинформатики, без привлечения эксперимента. Наиболее достоверные результаты при этом дают методы, основанные на сравнении нескольких геномных последовательностей. Так, была успешно исследована регуляция биосинтеза аргинина [15] и ароматических аминокислот [16, 17], биосинтеза пуринов, ароматических аминокислот и фиксации азота в археях ([18]), метаболизма углеводов [19, 20] ответа на тепловой шок [21] и устойчивости к ионам тяжелых металлов [22]. В последнее время в практику вошло исследование сразу нескольких функционально близких регуляторных систем. Такой подход хорошо зарекомендовал себя в случаях анализа регуляции метаболизма оксидов азота [23] и жирных кислот [24] и гомеостаза железа и марганца [25]. Также была прослежена эволюция регуляции биосинтеза НАД в протеобактериях [26, 27], группе *Bacillus/Clostridium*, типе *Fusobacteria* и порядке *Thermotogales* [28] и LexA-зависимой регуляции SOS-ответа в различных группах бактерий [29-31]. Исследована регуляция азотфиксации в цианобактериях [32] и *Firmicutes* [33].

Целый ряд исследований посвящен РНК-регуляции. Так, методами биоинформатики изучены РНК-переключатели, регулирующие биосинтез рибофлавина [34-36]), синтез тиамин [37], синтез кобаламина [38], биосинтез метионина и метаболизм S-аденозил метионина [39] и биосинтез, транспорт и катаболизм лизина [40].

В некоторых исследованиях проведен массовый анализ регуляции методами сравнительной геномики. Например, были исследованы сразу 101 регулон для *Rhodopseudomonas palustris* и родственных альфа-протеобактерий [41], 188 регулонов для трех геномов *Bacillus* [42] и 125 регулонов для *Staphylococcus aureus* и других *Bacillales* [43].

В настоящей работе была прослежена эволюция нескольких регуляторных систем в группе гамма-протеобактерий. Таковыми являются регулятор центрального метаболизма FruR (Cra), гомологичные регуляторы биосинтеза пуриновых нуклеотидов и утилизации рибозы, соответственно PurR и RbsR, и глобальные регуляторы дыхания, Fnr, ArcA и NarP.

Глава 1

Обзор литературы

1.1. Общие принципы регуляции транскрипции бактерий

Способность бактерий приспосабливаться к широкому спектру условий обеспечивается, в том числе, экспрессией различных наборов генов, что достигается сложной регуляцией на уровне как транскрипции, так и трансляции. Тем не менее, именно регуляция транскрипции является основным определяющим фактором в изменении экспрессии генов. Следует заметить, что регуляция на уровне транскрипции, в отличие от аллостерической регуляции, представляет собой механизм медленного реагирования на изменения условий и служит скорее для повышения эффективности происходящих в клетке процессов, чем для выживания организма в данный момент времени. Регуляция транскрипции происходит в основном на уровне инициации последней, участниками данного процесса являются как комплекс РНК-полимеразы, осуществляющий транскрипцию, так и дополнительные белки, называемые факторами транскрипции.

1.1.1. РНК-полимераза: структура и взаимодействие с промотором

Главным компонентом регуляции транскрипции бактерий является ДНК-зависимая РНК-полимераза, представляющая собой сложный многосубъединичный комплекс (Рис. 1.1). В структуре РНК-полимеразы принято выделять главный компонент (core) и σ -фактор. Главный компонент является достаточным для элонгации транскрипции, но не может распознавать последовательность промотора в отсутствие σ -фактора. Главный компонент содержит две идентичных α -субъединицы, и по одной копии субъединиц β , β' и ω . Субъединицы β и β' формируют активный центр фермента, связываясь как с матричной цепью ДНК, так и с новосинтезированной РНК [44]. α -субъединица состоит из двух доменов, разделенных линкерной областью из примерно 20 аминокислот. N-концевые домены α -субъединиц обеспечивают их димеризацию и связывание с $\beta\beta'$ -комплексом, тогда как C-концевые необходимы для связывания с ДНК и взаимодействия с различными факторами транскрипции [45]. Роль же ω -субъединицы остается до конца не выясненной [46].

σ -фактор представляет собой субъединицу, обеспечивающую распознавание промоторной последовательности, привлечение на промотор главного компонента РНК-полимеразы и расплетание дуплекса ДНК в области старта транскрипции. σ -фактор является мультидоменным белком, большинство известных к настоящему моменту σ -факторов

содержат четыре домена. Домены 2, 3 и 4 участвуют в распознавании промотора, тогда как роль домена 1 до сих пор не ясна. Более того, в ряде случаев этот домен может отсутствовать [47].

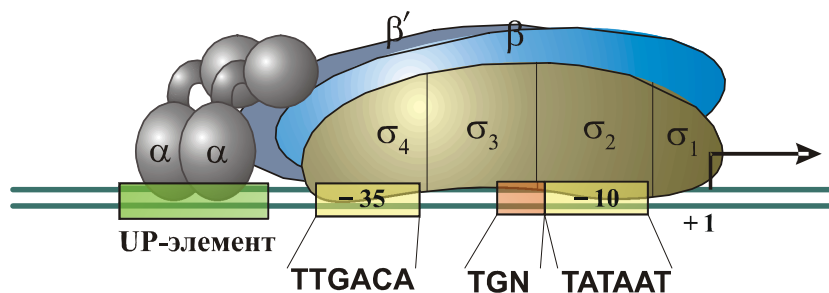


Рис. 1.1. Строение РНК-полимеразы и промотора

Показаны участки промотора, соответствующие UP-элементу, -35 сайту (TTGACA), удлинённому -10 сайту (TGN), -10 сайту (TATAAT), а также старт транскрипции (+1).

В *E. coli* большинство промоторов распознается с помощью основного σ -фактора, называемого σ^{70} . Помимо этого, существуют также альтернативные σ -факторы, обеспечивающие распознавание меньшего количества промоторов, и, как правило, участвующие в ответе на те или иные стимулы. Подавляющее число бактерий имеет несколько альтернативных σ -факторов в дополнение к основному [48].

В случае промоторов для σ^{70} *E. coli* принято выделять четыре основных элемента последовательности. Принципиально важными при этом являются так называемые -35 и -10 последовательности, названные так по положению относительно старта транскрипции. Последовательность -10 сайта имеет консенсус TATAAT и взаимодействует с доменом 2 σ -фактора, тогда как последовательность -35 сайта, с консенсусом TTGACA, взаимодействует с доменом 4. Кроме того, выделяют также два дополнительных элемента промотора. Первый – это так называемый удлинённый -10 сайт, включающий 3-4 п.н. и располагающийся непосредственной перед -10 сайтом. Эта последовательность взаимодействует с доменом 3 σ -субъединицы. Другая последовательность, UP-элемент, располагается перед -35 сайтом и взаимодействует с C-концевым доменом α -субъединицы [49] (Рис. 1.1).

При связывании с ДНК σ -субъединицы и главного компонента РНК-полимеразы, происходит расплетание ДНК в области старта транскрипции, в результате чего образуется так называемый открытый комплекс, на котором и иницируется синтез РНК. По завершении инициации транскрипции, σ -фактор покидает комплекс РНК-полимеразы и элонгация транскрипции осуществляется уже одним главным компонентом [47, 49].

Инициация транскрипции представляет собой сложный пошаговый процесс, что создает условия для ее достаточно точной регуляции.

1.1.2. Факторы транскрипции

Факторами транскрипции называются белки, связывающиеся с промотором и регулирующие транскрипцию, активируя либо репрессируя ее. Большинство факторов транскрипции связываются с ДНК с помощью мотива спираль-поворот спираль (helix-turn-helix, НТН-мотив). Известны и другие ДНК-связывающие структуры, такие как цинковые пальцы (zinc-fingers), цинковые ленты (zinc-ribbons) и антипараллельные β -тяжи, но у бактерий все они встречаются гораздо реже [50, 51].

Всего в геноме *E. coli* было обнаружено около 300 генов, кодирующих известные и предполагаемые факторы транскрипции. Большинство из них характеризуются высокой специфичностью, связываясь лишь с определенными последовательностями ДНК (далее – сайтами). Большинство факторов транскрипции контролируют экспрессию нескольких генов. Так, для *E. coli* известно всего около 60 факторов, регулирующих транскрипцию лишь с одного промотора. При этом в том же организме экспрессия более чем 50% генов регулируется лишь семью факторами: Crp, IHF, Fis, ArcA, Fnr, NarL и Lrp [52].

Количество факторов транскрипции у отдельно взятого организма зависит от различных факторов, таких как место обитания, образ жизни и размер генома. В целом ряде работ исследовалась зависимость числа факторов транскрипции от размеров генома [50, 53-56]. Было показано, что облигатные паразиты и эндосимбионты с малым размером генома, имеют меньшее количество факторов транскрипции (приблизительно 1% от всех генов), чем свободноживущие бактерии или факультативные патогены с геномами большего размера (около 4,5% от всех генов) [56]. В целом, количество факторов транскрипции растет в квадратичной пропорции с увеличением количества генов в геноме, то есть, при удвоении размера генома количество факторов транскрипции возрастает в четыре раза [54]. Теоретически, при неограниченном росте размера генома должен наступить момент, когда все гены в этом геноме будут кодировать факторы транскрипции. Однако, на практике не наблюдается геномов такого размера, чтобы происходило перенасыщение их генами регуляторов транскрипции. Так, для самых крупных из известных геномов, *Pseudomonas aeruginosa* [53] и *Streptomyces coelicolor* [57], отношение числа генов факторов транскрипции к общему числу генов составляет, соответственно, 9,4% и 12,3%. Не исключено, что квадратичное увеличение числа факторов транскрипции является одним из механизмов, ограничивающих размер геномов бактерий.

Факторы транскрипции обеспечивают изменение экспрессии генов в ответ на внешние или внутриклеточные стимулы. В целом можно выделить три механизма, за счет которых осуществляется воздействие этих стимулов на сами факторы. Во-первых, это непосредственное взаимодействие фактора с лигандом, малой молекулой. Примером может служить белок LacI *E. coli*, который, связываясь с молекулой аллолактозы, претерпевает конформационные изменения, в результате чего теряет способность связываться с ДНК, и, соответственно, регулировать экспрессию оперона *lacZYA* [58]. Другим широко распространенным механизмом является ковалентная модификация регуляторного белка. Данный принцип реализуется в двукомпонентных системах, например в системе ответа на фосфатное голодание PhoR-PhoP *Bacillus subtilis*. В случае недостатка фосфатов сенсорная киназа PhoR, мембранный белок, фосфорилирует фактор транскрипции PhoP, который в результате приобретает способность связываться с ДНК [59]. Третий возможный способ – это регуляция гена самого фактора транскрипции. Примером может служить регуляторный белок FixK *Bradyrhizobium japonicum*, экспрессия гена которого активируется белком FixJ в ответ на анаэробные условия [60].

Факторы транскрипции могут активировать либо репрессировать экспрессию с промотора. К настоящему времени известны как белки, действующие исключительно как репрессоры или активаторы, так и белки, способные осуществлять и экспрессию, и активацию, в зависимости от положения их сайта относительно промотора [47].

Принято различать три основных типа активации промоторов. Активаторы типа I связываются с последовательностью перед UP-элементом, взаимодействуя с С-концевым доменом α -субъединицы РНК-полимеразы (Рис. 1.2а). Именно так действует глобальный регулятор Cgr при активации оперона *lacZYA E. coli* [61]. Активаторы класса II связываются с последовательностью, предшествующей -35 элементу промотора, и взаимодействуют с доменом σ -субъединицы (Рис. 1.2.б). По такому принципу устроена регуляция генов ответа на фосфатное голодание белком PhoV *E. coli* [62]. Третий механизм активации осуществляется в случаях, когда белок связывается с последовательностью между -35 и -10 элементами промотора, изменяя таким образом конформацию последнего и создавая условия для связывания σ -фактора (Рис. 1.2в). Подобным образом регулируют экспрессию генов белки MerR-семейства [63].

В случае репрессии генов также возможно несколько механизмов. В самом простом случае фактор транскрипции связывается с последовательностью промотора или в непосредственной близости от нее, препятствуя посадке комплекса РНК-полимеразы (Рис. 1.3а). По такому принципу построено действие репрессора LacI [58]. Другой механизм репрессии заключается в следующем: репрессор связывается с множественными сайтами,

затем образует полимер, что вызывает образование петли ДНК в области промотора и препятствует связыванию РНК-полимеразы (Рис. 1.3б). Таким образом действует репрессор *E. coli* GalR [64]. Третий же механизм включает в себя взаимодействие репрессора с белком, являющимся активатором данного промотора (Рис. 1.3в). Так, белок CytR, садясь на ДНК рядом с Ctr, блокирует активирующее действие последнего [65].

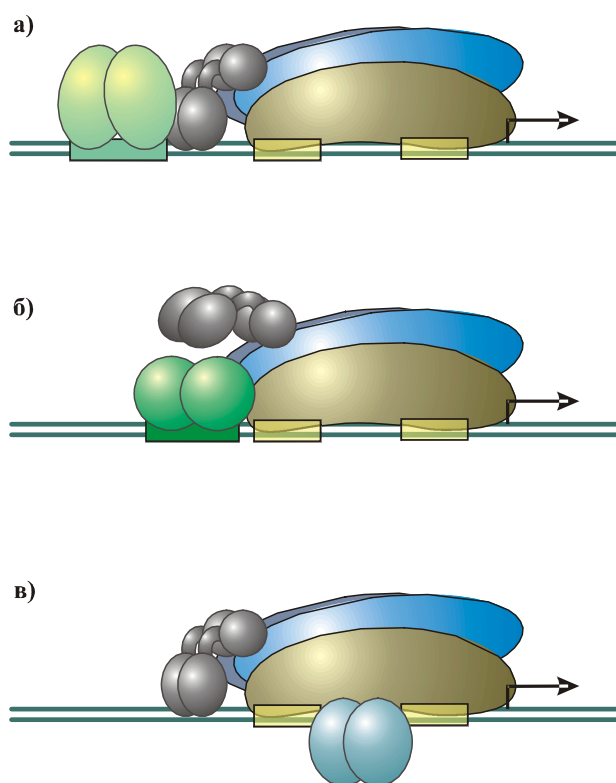


Рис. 1.2. Принципы действия белков-активаторов

а) активатор класса I; б) активатор класса II; в) активатор MerR-семейства.

Таким образом, регуляция экспрессии генов в бактериальной клетке представляет собой достаточно сложный процесс. С другой стороны, специфичность связывания большинства факторов транскрипции дает возможность изучения регуляции не только экспериментально, но и исследуя последовательности геномов.

1.2. Современные методы сравнительной геномики

К настоящему моменту времени экспериментальное изучение регуляции транскрипции бактерий представляет собой тщательно разработанную и широко используемую технологию. Тем не менее, классические методы исследования транскрипции, такие как анализ мутантных штаммов, направленный мутагенез, использование химерных конструкций (fusion construction), замедление ДНК в геле (gel shift assay) и определение защищенных от расщепления ДНК-азами и химическими реагентами участков (footprinting),

весьма трудоемки и позволяют исследовать экспрессию лишь отдельно взятых генов. Методы же массового анализа экспрессии, такие как анализ на микрочипах (microarray analysis) и иммунопреципитация хроматина (ChIP-on-chip), являются крайне дорогостоящими и выдают результат с высоким уровнем шума. Вышеперечисленные трудности вызывают необходимость в использовании методов биоинформатики для изучения репрессии генов, если не вместо экспериментальных технологий, то в дополнение к ним.

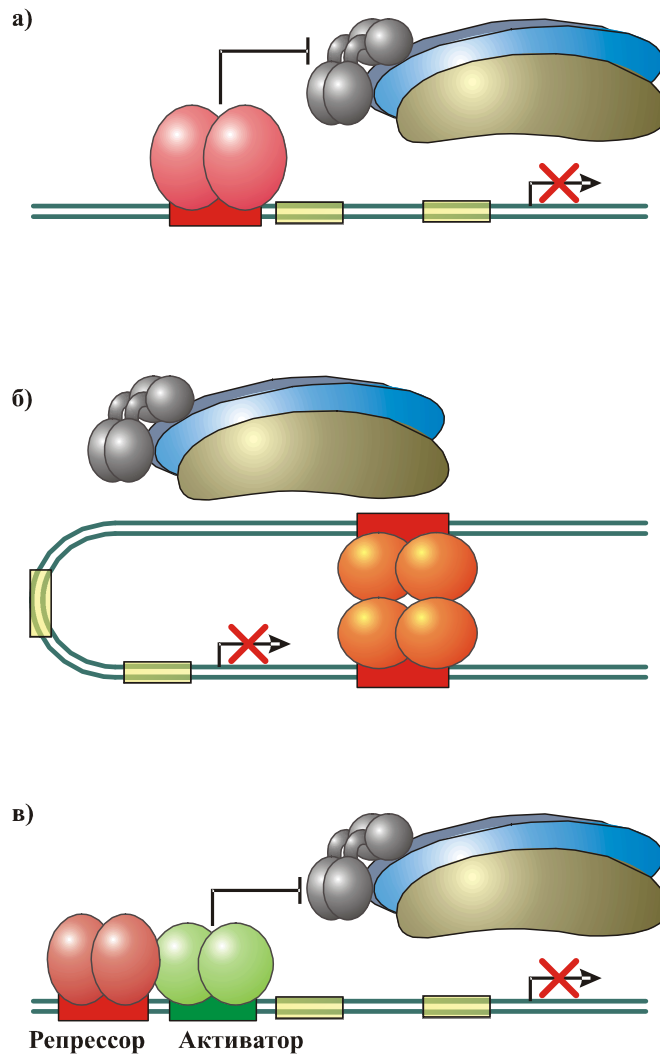


Рис. 1.3. Принципы действия белков-репрессоров

а) репрессия путем препятствия посадке РНК-полимеразы; б) репрессия путем выпетливания ДНК в области промотора; в) репрессия путем деактивации белка-активатора.

Как уже было отмечено ранее, большинство факторов транскрипции связываются с ДНК специфически, то есть имеют предпочтение к определенным последовательностям. Именно предсказание таких последовательностей в геномах бактерий лежит в основе исследования регуляции методами биоинформатики. Сразу же следует отметить, что

изучение регуляции методами сравнительной геномики не ограничивается лишь предсказанием сайтов. Поскольку сайты связывания факторов транскрипции приурочены к определенным генам и оперонам, то исследование регуляции подразумевает также уточнение предсказания границ генов, предсказание оперонной структуры, поиск потенциальных регуляторов и многое другое. Более того, исследование регуляции методами биоинформатики не является самоцелью, а часто служит для решения других задач, таких как предсказание функций регулируемых генов и метаболическая реконструкция. Примерами могут служить идентификация транспортера биотина BioY [66] и предсказание роли рибосомы как депо цинка [67, 68].

В качестве еще одного примера использования изучения регуляции для метаболической реконструкции можно привести предсказание функции белка FadE как ацил-КоА дегидрогеназы при изучении FadR-зависимой регуляции [69], что впоследствии нашло экспериментальное подтверждение [70].

Именно благодаря анализу регуляции методами сравнительной геномики были предсказаны транспортеры кобальта и никеля, соответственно, CbiMNQO и NikMNQO [71].

Ген *uraA* (впоследствии был переименован в *ribU*) был идентифицирован как транспортер рибофлавина в ходе анализа регуляции биосинтеза рибофлавина [37] и в дальнейшем предсказанная функция была подтверждена экспериментально [72].

1.2.1. Предсказание функций генов на основе сравнения аминокислотных последовательностей

В 1995 году была полностью отсеквенирована первая последовательность бактериального генома паразитической бактерии *Haemophilus influenzae* Rd KW20 [1]. В последующие годы процесс секвенирования новых геномов шел по нарастающей, и к настоящему времени насчитывается более 700 бактериальных геномов с полной последовательностью. Однако, само по себе определение полной последовательности является лишь начальным этапом исследований. Полученная последовательность требует также аннотации – то есть определения функционально значимых участков. Как минимум, аннотация бактериального генома требует определения потенциальных белок-кодирующих последовательностей (open reading frame, ORF). В наиболее простом случае производится поиск старт- и стоп-кодонов, отстоящих друг от друга на расстоянии, кратном трем и превышающем заданный минимальный порог. Стандартно за минимальное расстояние принимается 300 п.н., что позволяет учесть большинство белков. По такому принципу организована аннотация ORF в программе Artemis [73]. На следующем этапе определяется

предполагаемая функция полученных рамок считывания. Для этого производится сравнение их последовательностей с таковыми для реальных белков из различных банков данных.

Функция предсказанного белка считается совпадающей с таковой для известного, если доказана их ортологичность. Ортологами называются гены из разных геномов, имеющие общего предшественника, и не подвергшиеся дупликации в процессе эволюции. Гены же, образовавшиеся в результате дупликации предшественника, называются паралогами [74]. При этом функция одного или обоих паралогов может меняться в ходе эволюции, что затрудняет их точную аннотацию.

Среди используемых для аннотации банков последовательностей наиболее известными являются GenBank [75] и EMBL [76]. Также часто используется поиск по белковым банкам данных, таким как UniProt и TrEMBL [76-78]. Для поиска сходных последовательностей по базам данных обычно применяются программы семейства BLAST [5].

Поиск ортологов известных белков в новом геноме дает возможность приписать конкретную функцию примерно 40-65% генам, а анализ гомологов позволяет определить общую функциональную принадлежность для еще 20-30% [79]. Кроме того, в белках, кодируемых в новом геноме, возможно идентифицировать уже известные структурные домены, содержащиеся в таких базах данных, как PROSITE [80], Pfam [81] и SMART [82]. Для поиска по этим и другим банкам данных используется поисковая система InterPro [83]. Для предсказания приблизительной функции найденных генов используют программы поиска отдаленных гомологий, самой популярной из которых является PSI-BLAST [5]. Кроме того, определить приблизительную функцию гена позволяет его принадлежность к какому-либо кластеру ортологичных генов (КОГов) [84, 85].

Дополнительную информацию о функции гена можно получить благодаря идентификации мотивов последовательностей, характерных для определенных функциональных групп. Так, транспортные белки можно выявить по наличию трансмембранных спиралей [9], секретлируемые белки – по наличию сигнальных пептидов [86], а факторы транскрипции – по присутствию в последовательности ДНК-связывающих НТН-мотивов [87].

Несмотря на массу преимуществ, методы, основанные на сходстве аминокислотных последовательностей, не позволяют аннотировать все потенциальные гены в геноме. Поэтому для более качественной аннотации применяются дополнительные методы: анализ кластеризации генов на хромосоме и слияния генов, профили встречаемости генов в полных геномах и анализ регуляции.

1.2.2. Кластеризация генов на хромосоме

В ряде исследований было отмечено, что расположение генов на хромосоме зачастую не является случайным. Так, неоднократно отмечалось, что гены, кодирующие ферменты одного пути, склонны образовывать хромосомные кластеры [88-91]. Поэтому консервативная кластеризация генов на хромосоме может свидетельствовать об их функциональной близости, что позволяет определить функции генов, не имеющих известных гомологов. Примерами предсказаний функций генов с помощью анализа хромосомных кластеров могут служить выявление генов ферментов биосинтеза жирных кислот FabK и FabM в *Streptococcus pneumoniae* [92] и гена шикиматкиназы в археях [93]. На основании кластеризации с геном регулятора RutR генов *rutABCDEFG* был предсказан альтернативный путь деградации пиримидиновых нуклеотидов в нескольких альфа- и гамма-протеобактериях [94]. Можно отметить предсказание специфичности для 36 факторов транскрипции [95]. Возможность автоматизированного анализа хромосомной кластеризации генов реализована в программе STRING [96].

1.2.3. Слияние генов

Как неоднократно отмечалось, функционально близкие, например, кодирующие ферменты одного метаболического пути, гены зачастую сливаются, образуя единую рамку считывания. Такое слияние генов ведет к образованию мультидоменных белков, что дает определенные преимущества, например, ферменты одного метаболического пути оказываются в непосредственной близости, что обеспечивает доступность промежуточных продуктов реакции [97, 98]. Подобные закономерности дают возможность предсказывать функции неизвестных генов на основании слияния их гомологов в других геномах с генами с известной функцией. Примером такого анализа может служить предсказание функции гена *MTH554 Methanobacterium thermoautotrophicum*. Сравнительный анализ показал, что ортолог данного гена в *Caenorhabditis elegans* образует единый белок вместе с гомологом гена *MTH1425*, кодирующим металл-зависимую протеазу и Ser/Тур киназу. Хотя в данном случае невозможно однозначное определение функции гена *MTH554*, можно предположить его принадлежность к системе сигнальных каскадов [99].

1.2.4. Профили встречаемости генов

Данный подход основывается на предположении, что гены, принадлежащие к одному функциональному пути, должны вместе наследоваться в ходе эволюции. Следовательно, присутствие группы генов в одних геномах и отсутствие всех генов из данной группы в других свидетельствует об их функциональной связи [100-102]. В отличие от предыдущих

подходов, данный метод не накладывает ограничений на взаимное расположение генов, а опирается лишь на их одновременное присутствие в геноме. Примером успешного использования данного метода является обнаружение в бактериальных геномах регулятора редуктаз нуклеотидов NrdR [103].

В нескольких случаях профили встречаемости генов были использованы для определения специфичности транспортеров. Так, белок NiaX в бактериях группы *Bacillus/Clostridium* был идентифицирован как транспортер ниацина (витамина PP) на основании встречаемости его гена совместно с генами дополнительного пути биосинтеза НАД *pncA* и *pncB* [27].

Профили встречаемости подразумевают не только совместное присутствие генов в одном геноме, но и случаи, когда в одних геномах присутствует только один набор генов, а в других – только другой, и никогда оба набора генов не обнаруживаются в одном геноме. Такая ситуация сложилась с предсказанием транспортера тиамин (витамин B₁) YuaJ в организмах группы *Bacillus/Clostridium*. Было обнаружено, что данный ген под регуляцией тиаминового РНК-переключателя встречается лишь в геномах *Streptococcus* spp., где отсутствуют гены биосинтеза тиамин [37]. Подобным образом для генов *lysXY* была предсказана функция системы транспорта лизина. Данные гены присутствовали лишь в тех бактериях порядка Lactobacilliales, где отсутствовали известные гены биосинтеза и транспорта лизина [40]. Гену *ribU* в бактериях группы *Bacillus/Clostridium* была приписана функция транспортера рибофлавина (витамин B₂), поскольку этот ген обнаруживался лишь в тех геномах, где отсутствовали гены биосинтеза рибофлавина [35].

1.2.5. Методы распознавания потенциальных регуляторных сайтов

В настоящее время в биоинформатических исследованиях применяются различные методы поиска потенциальных регуляторных сайтов, однако в их основе лежит один общий принцип. Первоначально на основе группы известных сайтов, называемых обучающей выборкой, строится распознающее правило, с помощью которого и производится поиск потенциальных сайтов в регуляторных областях исследуемых генов.

Наиболее простым вариантом распознающего правила является консенсусная последовательность, где каждой позиции соответствует нуклеотид, преобладающий в данной позиции в сайтах обучающей выборки. Однако этот метод представляется крайне неточным, и его применение приводит к значительному количеству ложно-положительных предсказаний. Поэтому в настоящее время чаще применяются распознающие правила, основанные на использовании позиционных матриц. Примером могут служить позиционные матрицы частот (positional frequency matrices), где для каждого нуклеотида учитывается его

частота в каждой позиции [104]. Однако этот метод, хоть и имеет преимущества по сравнению с консенсусом, не учитывает специфики связывания белков с ДНК, поскольку предполагает равную значимость всех позиций сайтов. Более прогрессивным представляется метод позиционных матриц весов (positional weight matrices), где учитывается не только частота нуклеотида, но и консервативность самой позиции [105].

В настоящее время при изучении регуляции исследователи часто сталкиваются с тем, что известных регуляторных сайтов либо недостаточно для построения обучающей выборки, либо информация о них полностью отсутствует. В таком случае производится поиск потенциальных сайтов в регуляторных областях генов, для которых предполагается корегуляция. Наиболее распространенным является метод филогенетического футпринтинга (phylogenetic footprinting), основанный на поиске консервативных участков в выравниваниях регуляторных областей ортологичных генов из разных организмов [106]. Этот метод предсказания потенциальных сайтов реализован в семействе программ VISTA [107].

Однако метод филогенетического футпринтинга не работает в случае слабого сходства регуляторных последовательностей и совершенно не применим для анализа ко-регулируемых, но не ортологичных генов. Поэтому часто используются методы, основанные на поиске мотивов, то есть схожих участков в регуляторных областях ко-регулируемых генов. Различные варианты подобных методов реализованы в таких программах, как, например, AlignACE [108], MEME [109], SignalX [110] и SeSiMCMC [111].

Кроме того, в некоторых алгоритмах используется комплексный подход, совмещающий филогенетический футпринтинг и поиск мотивов. Такие алгоритмы реализованы в программах PhyloGibbs [112] и PhyME [113].

1.2.6. Сравнительная геномика и изучение регуляции

Практика использования различных методов предсказания сайтов показывает, что выбор порога для поиска сайтов представляет значительные трудности – при повышении порога значительное число экспериментально подтвержденных сайтов теряется, тогда как при его понижении сильно возрастает количество ложных предсказаний. В подобной ситуации избавиться от ложных предсказаний можно, используя методы сравнительной геномики.

Метод проверки соответствия (consistency check) заключается в поиске потенциальных сайтов перед ортологичными генами. Если сайты перед ортологичными генами или оперонами, содержащими ортологичные гены, сохраняются в нескольких родственных геномах, то данный ген с большой вероятностью является регулируемым. В случае же, если

сайт перед геном обнаруживается лишь в единичном геноме, то, скорее всего, он представляет собой перепредсказание и ген не является регулируемым.

Впервые данный метод был применен при сравнении регуляции биосинтеза пуриновых нуклеотидов, аргинина и ароматических аминокислот в геномах *Escherichia coli* и *Haemophilus influenzae* [114]. В данной работе были не только выявлены различия в регуляции между двумя геномами, но также предсказан ряд новых членов регулонов, например, была предсказана пуриновая регуляция для ряда транспортеров нуклеотидов. Дальнейшие исследования показали, что метод проверки соответствия применим для исследования регуляции не только бактерий, но и архей. Так, например, были предсказаны гены регуляторов биосинтеза триптофана и ответа на тепловой шок [18]. В дальнейшем метод проверки соответствия получил широкое применение и был использован в ряде исследований регуляции. К таковым можно отнести исследование регуляции биосинтеза аргинина в различных группах бактерий [15], биосинтеза ароматических аминокислот в грам-отрицательных [16] и грам-положительных [17] бактериях, анализ регуляции утилизации углеводов в грам-отрицательных [19] и грам-положительных [20] бактериях, ответа на тепловой шок в различных классах протеобактерий [21], биосинтеза НАД в энтеробактериях [26], устойчивости к ионам тяжелых металлов у бактерий из разных таксонов [22]. Достаточно подробно изучена эволюция LexA-зависимой регуляции SOS-ответа в различных группах бактерий [29-31]. Исследована регуляция азотфиксации фактором транскрипции NtcA в цианобактериях [32] и белками TnrA и GlnR в бактериях типа Firmicutes [33].

Подробно изучалась регуляция биосинтеза НАД в различных группах бактерий. Методами биоинформатики было обнаружено новое семейство регуляторов, получившее название NrtR, причем ДНК-связывающая активность белков NrtR из *Synechocystis* sp. PCC 6803 и *Shewanella oneidensis* MR-1 была подтверждена экспериментально [28]. В группе *Bacillus/Clostridium*, типе Fusobacteria и порядке Thermotogales был предсказан регулятор NiaR, а в бактериях группы *Bacillus/Clostridium* предсказаны новые транспортеры ниацина NiaP, NiaX и NiaY, гены которых входят в регулон NiaR. ДНК-связывающая способность белков *Thermotoga maritima* и *Bacillus subtilis* и транспортная функция белка NiaP из *B. subtilis* были подтверждены экспериментально [27].

Весьма перспективным представляется исследование сразу нескольких функционально близких регуляторных систем, контролирующих экспрессию одной и той же группы генов. Исследование регуляции метаболизма оксидов азота в различных группах бактерий белками HcpR, HcpR2, NsrR, NorR и Dnr выявило, что в случае комплексной регуляция генов в целом сохраняется, хотя регуляция отдельной регуляторной системой может достаточно сильно

меняться от таксона к таксону [23]. Похожие результаты были получены при изучении гомеостаза железа и марганца в альфа-протеобактериях, где была исследована регуляция белками IscR, Fur, Irr, RirA, Mur и MntR. В этом случае также железо-зависимая регуляция сохранялась в достаточно далеких таксонах, но при этом регуляция в разных таксонах осуществлялась разными белками [25]. Еще одним примером комплексной регуляции является контроль генов метаболизма жирных кислот, осуществляемый белками LiuR, LiuQ, PsrA, FadP и FadR [24].

В последнее время появились работы, посвященные массовому анализу регуляции методами сравнительной геномики. Сравнение *Rhodopseudomonas palustris* с геномами других альфа-протеобактерий позволило выделить 101 потенциальный регулон, в том числе регулоны азотфиксации FixK, NnrR, NtrC и RpoN, регулон OhrR, ответственный за ответ на гидропероксидный стресс, регулон SOS-ответа LexA, регулон биосинтеза белков жгутиков FlbD и регулон фотосинтеза PspR [41]. При сравнении геномов *Bacillus subtilis*, *B. halodurans* и *B. stearothermophilus* было выделено 188 потенциальных регулонов, например, были предсказаны сайты связывания регулятора теплового шока CtsR, регулятора центрального метаболизма CsrA, тРНК-зависимый и S-аденозилметиониновый РНК-переключатели, а также PurR-зависимый аттенюатор [42]. Сравнение генома *Staphylococcus aureus* с шестью другими бактериями порядка Bacillales позволило предсказать 125 потенциальных регулонов, в том числе регулоны TnrA, Fnr, Fur, CtsR и LexA [43].

Исследование регуляции методами сравнительной геномики значительно расширяет возможности предсказаний функций генов. Так, функция гена транспортера биотина *bioY* была предсказана на основании его кластеризации с генами синтеза биотина и регуляции его белком BioA [66], впоследствии данная функция была подтверждена экспериментально [115].

Другой пример экспериментального подтверждения предсказаний сравнительной геномики – цинк-зависимая регуляция генов паралогов белков рибосом. Так, в процессе изучения регуляции утилизации цинка была предсказана цинк-зависимая регуляция транскрипции генов, кодирующих не требующие наличия цинка пралоги рибосомных белков [67]. В дальнейшем предсказание получило экспериментальные подтверждения для *Bacillus subtilis* [116, 117] и *Streptomyces coelicolor* [118, 119].

Для потенциального транспортного белка YicE на основании анализа аминокислотной последовательности и предсказанной PurR-зависимой регуляции была предсказана функция транспортера ксантина [114], что в дальнейшем было подтверждено экспериментально [120].

Регуляция генов транспортера олигогалактоуронида *ogtABCD* (впоследствии были переименованы в *togMNAB*) белком KdgR в *Erwinia chrysantemi* первоначально была

предсказана методами сравнительной геномики [121], а затем подтверждена экспериментально [122].

Регулятор метаболизма метионина MetR первоначально был предсказан в геномах бактерий семейства *Streptococcaceae* [39], а впоследствии его предполагаемая функция и потенциальный мотив связывания были подтверждены экспериментально для одного из представителей этого семейства, *Streptococcus mutans* [123].

При исследовании регуляции метаболизма аргинина белком ArgR была предсказана регуляция оперона *uqiXYZ* в *Bacillus subtilis*. Также было выдвинуто предположение, что гены этого оперона кодируют систему транспорта аргинина [15]. Практически одновременно с предсказанием функция данных генов и их ArgR-зависимая регуляция были подтверждены экспериментально [124].

Еще одним примером экспериментально подтвержденной регуляции является регулятор генов нуклеотид редуктаз NrdR. Первоначально этот белок и его роль в регуляции генов *nrdDG* были предсказаны методами биоинформатики [103], а затем были получены экспериментальные подтверждения этого [125, 126].

Интересна также и история исследования РНК-переключателей (riboswitches). РНК-переключатели представляют собой последовательности РНК, способные принимать альтернативные вторичные структуры в зависимости от взаимодействия с малыми молекулами, что позволяет им участвовать в регуляции. Первым известным РНК-переключателем стал РФН-элемент, регулирующий биосинтез рибофлавина (витамина В₂), структура и механизм которого были изначально предсказаны методами биоинформатики [35, 36] а затем уже подтверждены экспериментально [34].

Следующим был изучен РНК-переключатель, регулирующий синтез тиамина (витамина В₁). Как и в предыдущем случае, вначале был предложен механизм его действия и предсказаны регулируемые им гены [37], а затем получены экспериментальные подтверждения [127].

Исследование РНК-переключателя, контролирующего гены синтеза кобаламина (витамина В₁₂) происходило по той же схеме: после предсказанного механизма регуляции [38] было получено экспериментальное подтверждение [128].

РНК-переключатель, регулирующий биосинтез метионина и метаболизм S-аденозил метионина, наоборот, вначале был изучен экспериментально [129-131], а уже затем методами сравнительной геномики были сделаны предсказания о регулируемых им генах в различных бактериях [39].

В случае же *Lys*-элемента, контролирующего гены биосинтеза, транспорта и катаболизма лизина, биоинформатические [40] и экспериментальные [132] исследования были проведены практически одновременно независимыми группами.

Таким образом, сравнительная геномика представляет собой мощный инструмент для исследования регуляции генов прокариот. В настоящей работе внимание было сосредоточено на эволюции регуляторных взаимодействий, и в качестве примеров были рассмотрены следующие системы:

- белок FruR (*Cra*), регулятор центрального метаболизма, эволюционировавший из локального регулятора фруктозного оперона;
- гомологичные белки PurR и RbsR, произошедшие путем дупликации от общего предшественника;
- глобальные регуляторы Fnr, ArcA и NarP, осуществляющие комплексный контроль генов дыхания.

1.3. FruR (*Cra*) – регулятор центрального метаболизма

Белок FruR был первоначально описан как репрессор оперона *fruBKA* бактерии *Salmonella typhimurium*, обеспечивающего транспорт и начальные этапы катаболизма фруктозы, за что и получил свое название [133]. Дальнейшие исследования этого фактора транскрипции показали, что в *E. coli* данный белок контролирует экспрессию большого числа генов, вовлеченных в различные процессы центрального метаболизма [134-136]. На основании новых полученных данных белок был переименован в *Cra* (*c*atabolite *r*epressor-*a*ctivator) [137]. В настоящее время в научной литературе употребляются оба упомянутых названия.

Анализ аминокислотной последовательности белка FruR показал его принадлежность к белковому надсемейству LacI-GntR. Подобно многим другим белкам данного надсемейства, FruR содержит два домена. N-концевой домен является ДНК-связывающим и содержит НТН-последовательность, тогда как С-концевой домен участвует в связывании лигандов и гомологичен лиганд-связывающим доменам периплазматических рецепторов углеводов [137, 138].

В ряде экспериментов было показано, что в отсутствие эффекторов белок FruR способен связываться с собственными сайтами в виде гомотетрамера. Наличие же эффекторов, фруктоза-1-фосфата и/или фруктоза-1,6-дифосфата (Рис. 1.4), препятствует связыванию фактора транскрипции с ДНК, причем эффекторы различаются по степени воздействия ДНК-связывающую способность белка. Так, для предотвращения связывания достаточно миллимолярных концентраций фруктоза-1,6-дифосфата и всего лишь микромолярных

концентраций фруктоза-1-фосфата [137, 139, 140]. В зависимости от положения сайта связывания относительно промотора регулируемого оперона, белок FruR может быть как репрессором, так и активатором [137, 140]. Таким образом, при наличии достаточных концентраций эфффекторов происходит дерепрессия одних генов и деактивация других (Рис. 1.5).

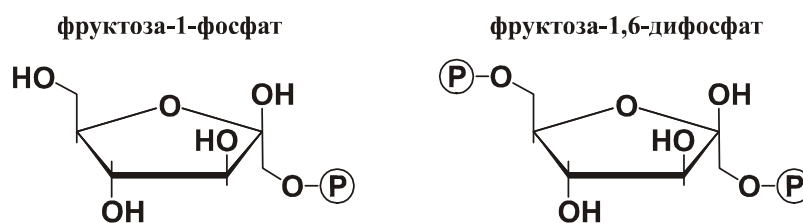


Рис. 1.4. Эфффекторы для белка FruR

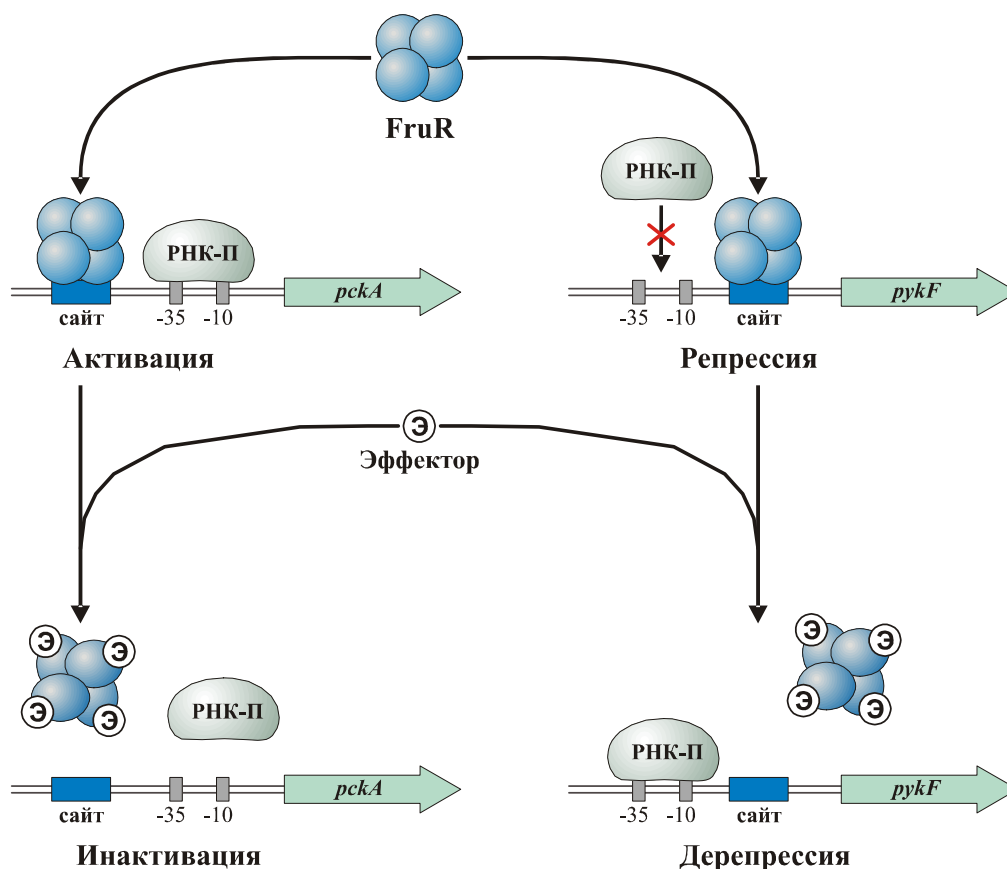


Рис. 1.5. Модель FruR-зависимой регуляции

Условные обозначения: «сайт» - сайт связывания FruR, «-35» и «-10» - последовательность промотора, «РНК-II» - РНК-полимераза.

Наиболее подробно FruR-зависимая регуляция была исследована в *E.coli*, где данный фактор контролирует экспрессию генов фосфотрансферазных систем (*fruBKA*, *ptsHI-crr*, *mtlADR*), гликолиза и глюконеогенеза (*pgk*, *fbaA*, *pfkA*, *glk*, *pykF*, *ppsA*, *pckA*), пути Энтнера-

Дудорова (*edd-eda*), пентозофосфатного пути (*epd*), цикла трикарбоновых кислот (*icdA*), глиоксилатного пути (*aceBAK*), алкоголь дегидрогеназы (*adhE*) и нитрит редуктазы (*nirBDC-cysG*). Подробная информация о генах FruR-регулона представлена в Табл. 1.1 и на Рис. 1.6.

В целом регуляция за счет FruR заключается в следующем. В отсутствие эффекторов белок-регулятор связывается с ДНК, репрессируя гены, необходимые для транспорта и катаболизма сахаров и активируя гены, необходимые для процессов глюконеогенеза. Появление же в клетке эффекторов, свидетельствующих о наличии углеводов в среде, приводит к дерепрессии генов фосфотрансферазных систем, гликолиза, пути Энтнера-Дудорова и пентозофосфатного пути. При этом инактивируется экспрессия генов цикла Кребса, глиоксилатного пути и глюконеогенеза [135, 137, 138]. Таким образом, FruR обеспечивает переключение между анаболизмом и катаболизмом сахаров, в случае недостатка сахаров активируя анаболические пути, а в случае их избытка – катаболические.

Таблица 1.1. Функции генов FruR-регулона *E. coli*

Указаны ссылки на работы, в которых экспериментально подтверждена FruR-зависимая регуляция для данного оперона.

Оперон	Функции генов оперона	Ссылки
Гены, репрессируемые FruR		
<i>fruBKA</i>	Фруктозо-специфичная фосфотрансферазная система; фруктозо-киназа	[140]
<i>epd-pgk-fbaA</i>	Эритроза-4-фосфат дегидрогеназа; фосфоглицерат киназа; фруктоза-фосфат альдолаза	[135]
<i>pfkA</i>	Фруктоза-6-фосфат киназа	[139]
<i>glk</i>	Глюкозо-киназа	[141]
<i>edd-eda</i>	Фосфоглюконат дегидратаза; 2-кето альдолаза	[135]
<i>pykF</i>	Пируват-киназа	[134]
<i>adhE</i>	Алкогольдегидрогеназа	[142, 143]
<i>mtlADR</i>	Маннитол-специфичная фосфотрансферазная система	[135]
<i>ptsHI-crr</i>	Общий компонент фосфотрансферазных систем; глюкозо-специфичная фосфотрансферазная система	[136, 140]
Гены, активируемые FruR		
<i>icdA</i>	Изоцитрат-дегидрогеназа	[140, 144]
<i>aceBAK</i>	Малат-синтаза; изоцитрат-лиаза	[140]
<i>ppsA</i>	Фосфоенолпируват-синтаза	[140]
<i>pckA</i>	Карбоксикиназа фосфоенолпирувата	[135]
<i>nirBDC-cysG</i>	Цитоплазматическая нитрит-редуктаза	[145, 146]

левая часть палиндroma является более консервативной, чем правая, что свидетельствует о разной степени аффинности белка к разным полусайтам. Следует заметить, что такая асимметрия наблюдается как для активаторных, так и репрессорных сайтов [139, 147]. Для большинства регулируемых генов было обнаружено по одному сайту связывания FtuR. Исключение составляет лишь оперон *fruBKA*, имеющий два сайта в промоторной области [138, 140].

1.4. Регуляция утилизации рибозы

Фактор транскрипции RbsR представляет собой белок LacI-GntR надсемейства, имеющий N-концевой ДНК-связывающий и С-концевой лиганд-связывающий домены. В организме *E. coli* данный белок регулирует экспрессию лишь одного оперона *rbsDACBKR* ([148]). Гены *rbsDACB* кодируют высоко-специфичную систему транспорта рибозы [149], тогда как продуктом гена *rbsK* является фермент рибозо-киназа [150]. В отсутствие эффекторов фактор RbsR связывается с промоторной областью *rbs*-оперона, репрессируя его экспрессию. Сайт связывания этого белка в *E. coli* представляет собой палиндром длиной 20 п.н. и имеет последовательность TCAGCGAAACGTTTCGCTGA [148]. В присутствие рибозы, являющейся эффектором, RbsR диссоциирует от ДНК, таким образом снимая репрессию оперона (Рис 1.7).

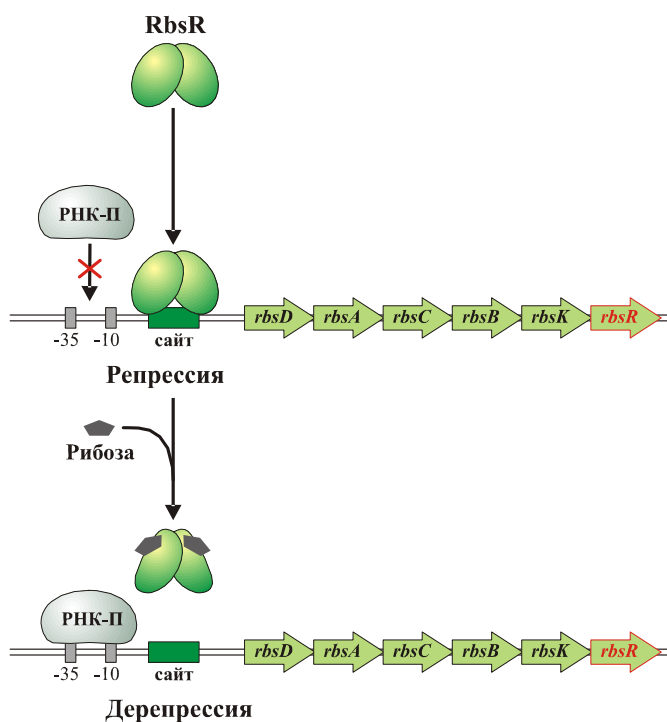


Рис. 1.7. Модель RbsR-зависимой регуляции

Условные обозначения: см. Рис 1.5. Ген репрессора выделен красным шрифтом.

1.5. PurR – регулятор биосинтеза пуриновых нуклеотидов

Белок PurR представляет собой классический фактор транскрипции LacI-GntR надсемейства, имеющий ДНК-связывающий N-концевой и лиганд-связывающий C-концевой домены ([151]). PurR работает только как репрессор, и, в отличие от двух описанных выше белков, связывание с соответствующими сайтами происходит не в отсутствие, а в присутствии лигандов, каковыми являются гуанин и гипоксантин (Рис. 1.8, 1.9). Этот белок связывается с ДНК в форме гомодимера [152, 153].

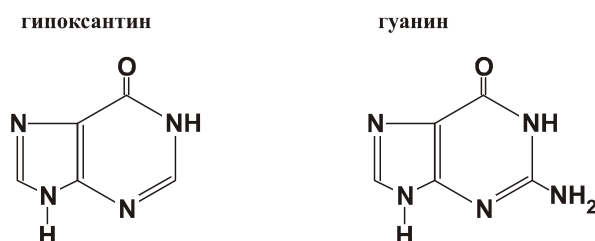


Рис. 1.8. Эффекторы для белка PurR

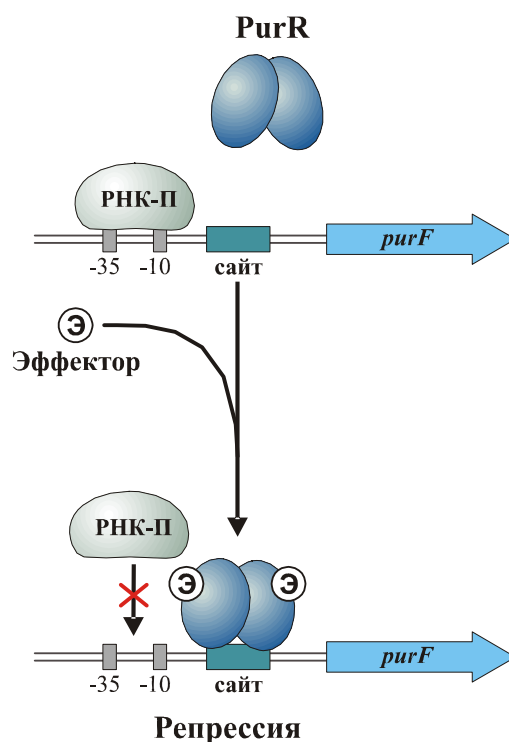


Рис. 1.9. Модель RbsR-зависимой регуляции

Условные обозначения: см. Рис 1.5.

Экспериментальные исследования PurR-зависимой регуляции фактически ограничиваются работами, описывающими регуляцию в *E. coli*, за исключением единственного исследования, проведенного для *Salmonella typhimurium* [154]. В *E. coli* пуриновый репрессор, помимо собственного гена (*purR*) и генов биосинтеза пуриновых

нуклеотидов (*prsA*, *cvpA-purF*, *purHD*, *purMN*, *purT*, *purL*, *purEK*, *purC*, *purB*, *purA*, *guaBA*), контролирует также гены, участвующие в близких биохимических процессах. К таковым относятся гены синтеза пиримидиновых нуклеотидов (*purC*, *purD*, *codBA*), азотного обмена (*speAB*, *glnB*) и фолат-ассоциированных одноуглеродных соединений (*gcvTHP*, *glyA*) (Табл. 1.2, Рис. 1.10).

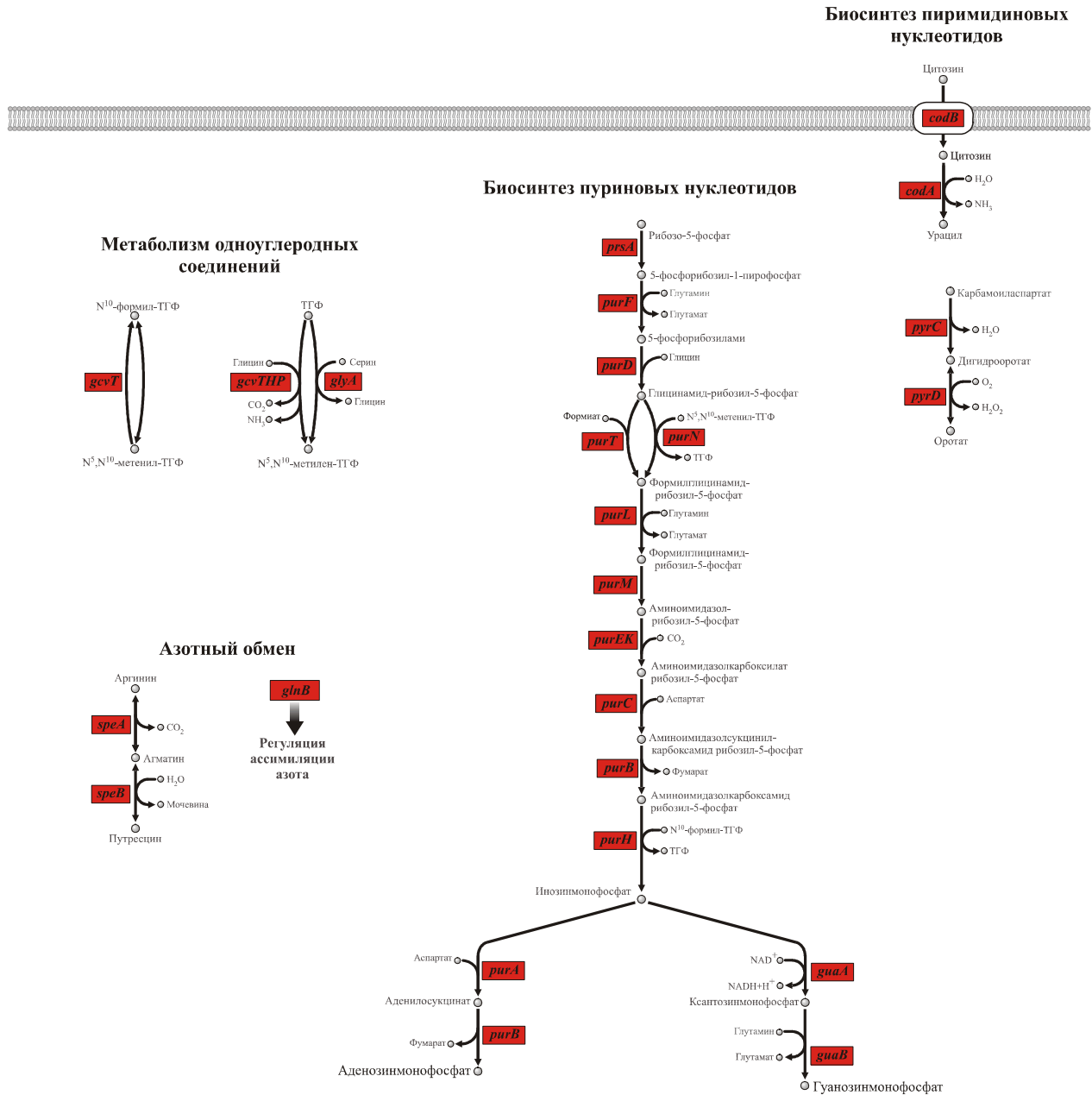


Рисунок 1.10. PurR-регулон *E. coli* на основе экспериментальных данных
Красным помечены репрессируемые PurR гены.

Исследования PurR-зависимой регуляции выявило несколько закономерностей во взаимодействии регулятора с сайтами его связывания. Так, перед всеми генами были обнаружены потенциальные сайты связывания белка, представляющие собой палиндром

длиной 16 п.н. с консенсусом AGCGAAACGTTTCGCT [155]. Перед большинством генов найдено по одному сайту связывания исследуемого белка, исключение составляют гены *purR* и *purA*, перед которыми найдено по два сайта [156, 157]. Все обнаруженные сайты были найдены между стартом транскрипции и стартом трансляции, кроме случая *purB*, где сайт связывания белка располагается внутри кодирующей рамки, в положении +232 п.н. относительно старта транскрипции. В данной ситуации репрессия осуществляется, по всей видимости, за счет ингибирования элонгации транскрипции [158].

Таблица 1.2. Функции генов PurR-регулона *E. coli*

Указаны ссылки на работы, в которых экспериментально подтверждена PurR-зависимая регуляция для данного оперона.

Оперон	Функции генов оперона	Ссылки
<i>purR</i>	Репрессор пуринового регулона	[156, 159]
<i>prsA</i>	Фосфорибозил пиррофосфат (ФРПФ) синтетаза	[11]
<i>svrA-purF</i>	Фактор синтеза колицина V; глутамин-ФРПФ амидотрансфераза	[160, 161]
<i>purHD</i>	Фосфорибозиламиноимдазолкарбоксамид формилтрансфераза, ИМФ циклогидролаза; глицинамид киносинтетаза	[12]
<i>purMN</i>	Аминоимндазолрибозилфосфат-синтетаза; глицинамид-рибозилфосфат формилтрансфераза N	[12, 159]
<i>purT</i>	Глицинамид-рибозилфосфат формилтрансфераза T	[162]
<i>purL</i>	Формилглицинамидин-рибозилфосфат синтетаза	[12, 159]
<i>purEK</i>	N ⁵ -карбоксаминоимдазол-рибонуклеотид мутаза; N ⁵ -карбоксаминоимдазол-рибонуклеотид синтаза	[12]
<i>purC</i>	Фосфорибозиламиноимдазол-сукцинокарбамид синтетаза	[159]
<i>purB</i>	Аденилосукцинат лиаза	[158, 159]
<i>purA</i>	Аденилосукцинат синтаза	[157]
<i>guaBA</i>	ИМФ дегидрогеназа; ГМФ синтетаза	[159]
<i>purC</i>	Карбамоиласпартат дегидратаза	[163]
<i>purD</i>	Дигидрооротат дегидрогеназа	[163]
<i>codBA</i>	Транспортер цитозина; цитозин деаминаза	[164]
<i>speAB</i>	Аргинин декарбоксилаза; агматиназа	[11]
<i>glnB</i>	Регулятор ассимиляции азота	[11]
<i>gcvTHP</i>	Белки системы расщепления глицина	[165]
<i>glyA</i>	Серин гидроксиметилтрансфераза	[166]

Было также замечено, что гены биосинтеза ИМФ из фосфорибозо-пиррофосфата (*prsA*, *svrA-purF*, *purHD*, *purMN*, *purT*, *purL*, *purEK*, *purC*) имеют сайты, более близкие к консенсусу, нежели другие регулируемые гены. Кроме того, было показано, что для генов

синтеза ИМФ PurR снижает интенсивность транскрипции более чем в 10-раз, тогда как для других генов уровень транскрипции в случае репрессии снижается не более, чем в 3 раза [155].

1.6. Глобальная регуляция генов дыхания

1.6.1. Общие принципы устройства дыхательных цепей бактерий

Дыхание представляет собой процесс превращения энергии химических реакций в энергию трансмембранного электрохимического градиента. Впоследствии энергия этого градиента может быть использована клеткой для различных целей, таких как синтез АТФ из АДФ или движения жгутиков.

Одной из отличительных особенностей прокариотических организмов является наличие множественных дыхательных цепей, позволяющих использовать различные органические и неорганические субстраты. Все известные дыхательные цепи бактерий имеют модульное строение, то есть состоят из отдельных заменяемых частей (Рис. 1.11).

Таковыми частями являются:

- дегидрогеназы доноров электронов, осуществляющие реакции окисления;
- хиноны, мембрано-растворимые переносчики электронов;
- оксидоредуктазы акцепторов электронов, осуществляющие реакции восстановления.

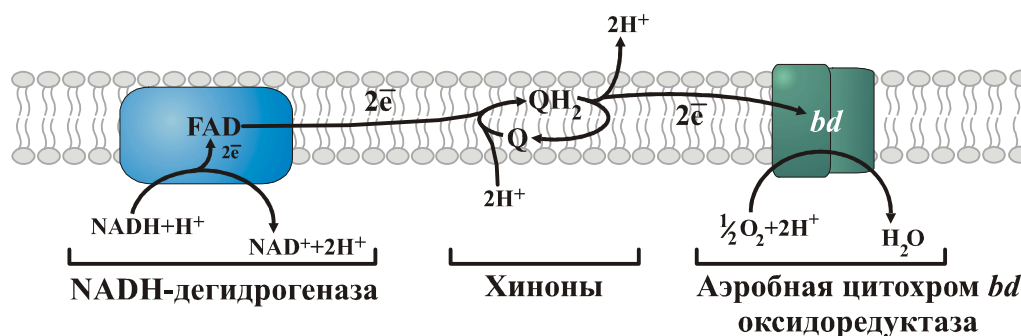


Рисунок 1.11. Пример бактериальной дыхательной цепи

Дегидрогеназы и редуктазы представлены мембранными белковыми комплексами. Как правило, все компоненты одного комплекса кодируются генами, входящими в единый оперон.

За счет особого устройства дыхательных цепей, в клетке реакция окисления-восстановления между донором и акцептором электронов разделяется на две, при этом перенос электронов от донора к акцептору осуществляется посредством хинонов. Фактически, в дыхательной цепи происходит две реакции окисления-восстановления. Первая

реакция, происходящая на дегидрогеназном комплексе, приводит к окислению донора электронов и восстановлению хинона до хинола. Реакция восстановления хинонов требует наличия в среде протонов, которые, в связи расположением активных центров дегидрогеназ вблизи внутренней стороны мембраны, поступают из цитоплазмы. Вторая реакция, осуществляемая оксидоредуктазным комплексом, заключается в окислении хинола до хинона и восстановлении акцептора электронов. Поскольку активные центры дегидрогеназ находятся вблизи внешней стороны мембраны, то освободившиеся протоны выводятся либо в периплазматическое пространство, либо во внешнюю среду (Рис. 1.11). Таким образом, дыхательная цепь обеспечивает перенос протонов из клетки через мембрану, тем самым создавая электрический и химический градиент [167].

Разнообразие дыхательных цепей в клетке осуществляется комбинацией различных дегидрогеназ и оксидоредуктаз. К настоящему моменту для *E. coli* известно более 10 дегидрогеназ и примерно столько же редуктаз. Разнообразие же хинонов не столь велико, известно лишь три типа последних: убихинон, менахинон и диметилменахинон. Использование того или иного типа хинонов зависит от того, какие именно дегидрогеназы и редуктазы присутствуют в мембране [168].

Окислительно-восстановительный потенциал используемых для дыхания реакций может варьировать в широких пределах: от +0.43V (формиат) до +0.03V (сукцинат) для доноров и от +0.82V (кислород) до -0.03V (фумарат) акцепторы. Необычайно широко разнообразие кофакторов, используемых для дыхания. Так, в состав дегидрогеназ и редуктаз могут входить молибдоптериновый кофактор, ионы никеля, ФАД, ФМН, железо-серные кластеры, и геммы *a*, *b*, *c* и *d*. Характерной особенностью дыхательных цепей бактерий является наличие нескольких изоферментов для одного донора, например, формиата или НАД·Н, или акцептора электронов, например, для кислорода или нитрата (Рис. 1.12). Помимо всего прочего, дыхательные ферменты бактерий характеризуются разнообразной топологией. Первая группа белковых комплексов, таких как FdoGHI, NarGHI и FrdABCD имеет как гидрофильную, так и гидрофобную части. Вторая группа представляет собой трансмембранные комплексы с протяженными гидрофобными участками, например, CyoABCD и CyoAB. Третья же группа включает в себя простые дегидрогеназы, состоящие из одного белка, в целом являющегося гидрофильным и заякоренного в мембране лишь несколькими гидрофобными участками. К этой группе относятся такие белки, как GlpD, Ndh, и LctD [169].

1.6.2. Особенности регуляции дыхания *E. coli*

Сложная система дыхательных цепей бактерий требует не менее сложно организованной регуляции, в том числе и на уровне экспрессии генов, кодирующих белки дыхательных комплексов. В ряде исследований было выяснено, что ключевую роль в данной регуляции играет наличие акцепторов электронов. Для бактерии более эффективным является использование акцепторов с наибольшим окислительно-восстановительным потенциалом, каковым является кислород. Поэтому в присутствии молекулярного кислорода *E. coli* активизируется транскрипция оперонов аэробных оксидоредуктаз. В случае же отсутствия кислорода в среде наиболее предпочтительными акцепторами являются нитрат и нитрит, при наличии которых в среде начинают экспрессироваться гены системы нитрат-нитритного дыхания. Таким образом, в *E. coli* регуляция дыхания реализуется по иерархическому принципу, в соответствии с наличием в среде наиболее энергетически выгодного акцептора электронов. Наличие же доноров электронов в среде практически не оказывает влияния на экспрессию генов комплексов дыхательных цепей [170-172].

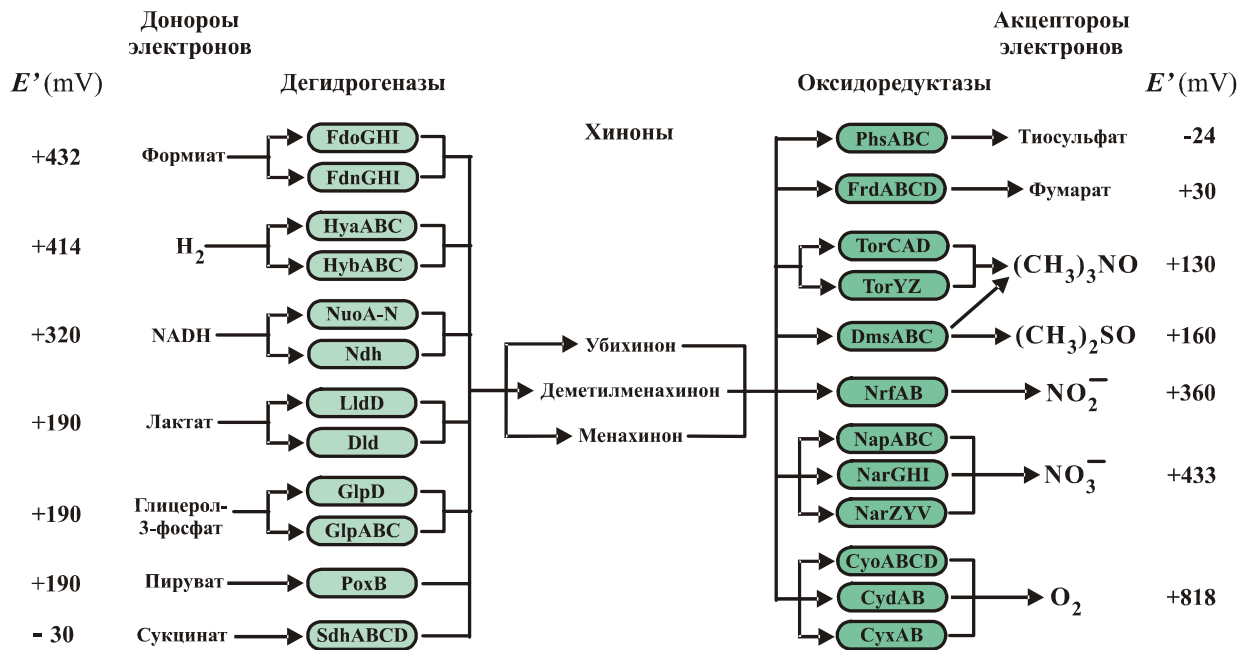


Рисунок 1.12. Многообразие дыхательных цепей *E. coli*

С боков показаны стандартные редокс-потенциалы для соответствующих реакций окисления и восстановления.

Для реализации сложной регуляторной иерархии требуется наличие множества факторов транскрипции, что и реализовано для *E. coli*. Так, ответ на наличие или отсутствие кислорода осуществляется белком Fnr и двухкомпонентной системой ArcB-ArcA, а за реакцию на нитрат и нитрит ответственны двухкомпонентные системы NarX-NarL и NarQ-NarP [168, 169, 173]. Кроме того, существуют также локальные регуляторные системы,

ответственные за реакцию на малоэффективные акцепторы электронов и контролирующие экспрессию всего лишь нескольких генов. К таковым относятся система регуляции дыхания триметил-оксидом азота TorS-TorR [174, 175] и система регуляции фумаратного дыхания DcuS-DcuR [176].

1.6.3. Fnr: ответ на молекулярный кислород

Белок Fnr, первоначально открытый как регулятор нитратного дыхания [177], является одним из глобальных факторов транскрипции, контролирующим переключение между аэробным и анаэробным метаболизмом в клетках *E. coli* [26, 168, 170, 178-180]. Этот белок принадлежит к подсемейству Fnr обширного белкового семейства Fnr-Csr [181]. Он состоит из двух доменов, С-концевого ДНК-связывающего, и N-концевого, отвечающего за димеризацию и чувствительность к концентрациям молекулярного кислорода в среде [182]. Как и большинство белков Fnr-подсемейства, Fnr *E. coli* содержит в С-концевом домене последовательность цистеинов структуры Cys-X₃-Cys-X₂-Cys-X₅-Cys, необходимую для формирования железо-серного кластера [183]. Этот железо-серный кластер формируется в анаэробных условиях, что позволяет белку димеризоваться и связываться с ДНК (Рис. 1.13). В аэробных же условиях, из-за окисления железа происходит сначала перестройка железо-серного кластера, а затем и его разрушение. Оба эти процесса являются обратимыми. В результате белок Fnr претерпевает конформационные изменения и теряет способность к димеризации и связыванию с ДНК [184].

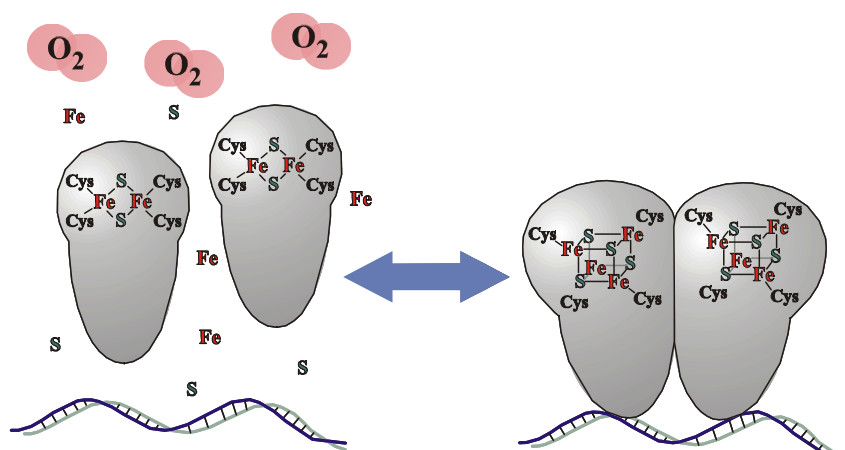


Рисунок 1.13. Механизм активации и инактивации Fnr

В анаэробных условиях Fnr активирует экспрессию генов анаэробного и микроаэробного метаболизма и, одновременно с этим, репрессирует транскрипцию генов аэробного метаболизма [185]. Функции генов Fnr-регулона перечислены в Табл. 1.3. Помимо этого, экспрессия многих генов регулируется Fnr опосредованно. По некоторым оценкам,

количество таких генов превышает 130 [168, 170, 172]. Следует также заметить, что Fnr контролирует экспрессию генов других регуляторов дыхания, таких как *arcA* [186] и *narXL* [187], а также собственного гена [188].

Анализ регуляторных областей контролируемых генов, а также ряд экспериментальных исследований показал, что сайт связывания Fnr представляет инвертированный повтор длиной 14 п.н. с консенсусом TTGAT-4-ATCAA [185, 189].

Таблица 1.3. Функции генов Fnr-регулона *E. coli*

Указаны ссылки на работы, в которых экспериментально подтверждена Fnr-зависимая регуляция для данного оперона.

Оперон	Функции генов оперона	Ссылки
Гены, активируемые Fnr		
<i>dmsABC</i>	Редуктаза диметил-сульфоксида	[13, 190]
<i>hcp-hcr</i>	Система формирования железо-серных кластеров	[191]
<i>narGHJ</i>	Цитоплазматическая нитрат-редуктаза	[13]
<i>narK</i>	Нитрат-нитрит антипортер	[13]
<i>narXL</i>	Двукомпонентная система регуляции нитрат-нитритного дыхания	[187]
<i>nirBDC-cysG</i>	Цитоплазматическая нитрит-редуктаза	[146]
<i>nrdDG</i>	Рибонуклеотид-трифосфат редуктаза	[192]
<i>nrjABCDEFG</i>	Периплазматическая нитрит-редуктаза	[193]
<i>yhjA</i>	Activation	[194]
<i>fdnGHI</i>	Формиат-дегидрогеназа	[195]
<i>napFDAGHBC-csmABCDEFGH</i>	Периплазматическая нитрат-редуктаза	[196]
<i>focA-pflB</i>	Транспортер формиата; пируват лиаза	[197]
<i>hlyE</i>	Гемолизин E	[198]
<i>arcA</i>	Глобальный регулятор дыхания	[186]
Гены, репрессуемые Fnr		
<i>moeAB</i>	Белки биосинтеза молибдоптеринового кофактора	[199]
<i>cydAB</i>	Цитохром <i>bd</i> оксидаза	[200]
<i>fnr</i>	Глобальный регулятор дыхания	[187, 188]
<i>ndh</i>	NADH дегидрогеназа	[201]
<i>yfiD</i>	Стресс-индуцируемая пируват лиаза	[202, 203]
<i>hemA</i>	Глутамил-тРНК редуктаза	[13]
<i>hmp</i>	Дигидроптеридин редуктаза и NO-диоксигеназа	[204]

1.6.4. Двукомпонентная система ArcB-ArcA: ответ на окислительно-восстановительный статус хинонов

Белок ArcA, также являющийся глобальным регулятором дыхания, был открыт при изучении экспрессии *suc*-генов, кодирующих сукцинат дегидрогеназу [205]. Этот белок является частью двукомпонентной системы ArcB-ArcA, где ArcB представляет собой сенсорную киназу. Как было показано, белок ArcB реагирует на редокс-статус растворимых

аналогов убихинона [206]. На основании этого была предложена модель, согласно которой в отсутствие акцепторов электронов происходит накопление восстановленной формы хинонов, хинолов (Рис. 1.14). При этом происходит восстановление цистеиновых остатков ArcV и разрыв дисульфидных мостиков, в результате чего конформация белка изменяется и он становится приобретает киназную активность [206, 207]. При этом происходит фосфорилирование ArcA, который в результате приобретает способность к связыванию ДНК (Рис. 1.15). Стехиометрия данного связывания до сих пор остается спорной, но большинство исследователей склоняются к тому, что ArcA связывается в виде олигомера [208, 209]. Таким образом, ArcA действует как фактор транскрипции в анаэробных условиях, и, по всей видимости, переключает метаболизм с дыхания на брожение, о чем свидетельствует репрессия им транскрипции многих генов аэробного метаболизма (Табл. 1.4).

Структура сайта связывания ArcA до настоящего времени остается спорной. С одной стороны, связывание в виде олигомеров наводит на мысли о симметричных сайтах. Более того, на симметрию сайта указывает и принадлежность данного белка к подсемейству OmpR-PhoB, для белков которого характерны сайты в виде прямых повторов [208]. Тем не менее, компьютерный анализ регуляторных областей генов ArcA-регулона обнаружил последовательность без внутренней симметрии (Рис. 1.16) [210]. Вопрос о структуре сайта связывания ArcA изучен в разделе 5.2.2.

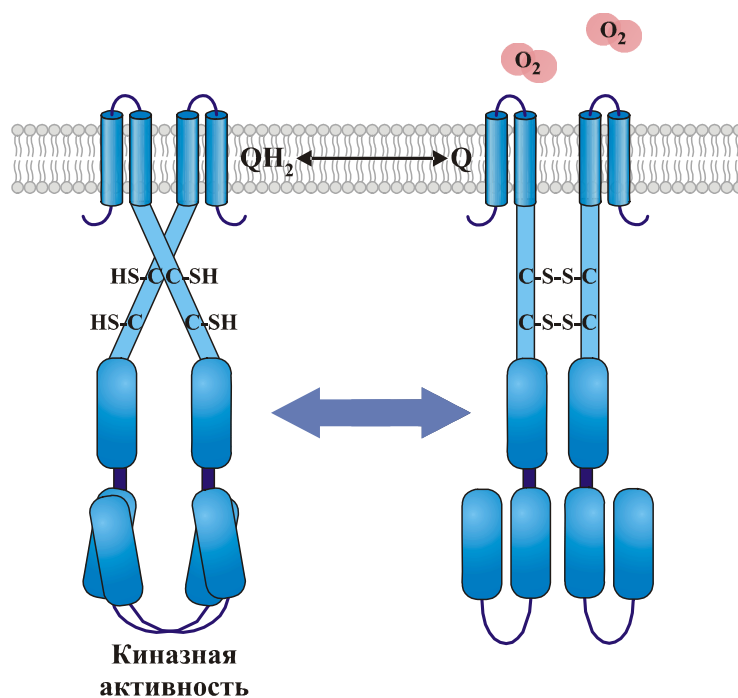


Рисунок 1.14. Механизм активации и инактивации ArcV

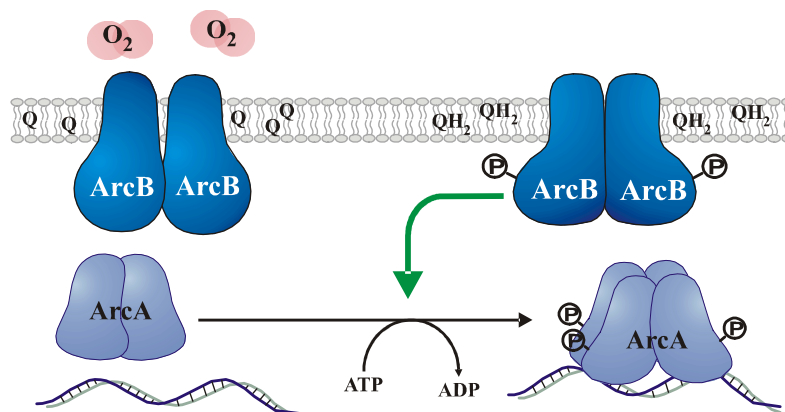


Рисунок 1.15. Механизм взаимодействия ArcB и ArcA

Таблица 1.4. Функции генов ArcA-регулона *E. coli*

Указаны ссылки на работы, в которых экспериментально подтверждена ArcA-зависимая регуляция для данного оперона.

Оперон	Функции генов оперона	Ссылки
Гены, активируемые ArcA		
<i>cydAB</i>	Цитохром <i>bd</i> оксидаза	[211]
<i>moeAB</i>	Белки биосинтеза молибдоптеринового кофактора	[199]
Гены, репрессируемые ArcA		
<i>sdhCDAB-sucABCD</i>	Сукцинат дегидрогеназа; 2-кетоглутарат дегидрогеназа	[211]
<i>gltA</i>	Цитрат синтаза	[211]
<i>fadBA</i>	Комплекс окисления жирных кислот	[212]
<i>fadD</i>	Ацил-КоА синтаза	[212]
<i>icd</i>	Изоцитрат дегидрогеназа	[213]
<i>sodA</i>	Супероксид дисмутаза	[214]
<i>lldPRD</i>	Транспортер лактата; репрессор оперона <i>lldPRD</i> ; лактат дегидрогеназа	[211]
<i>fadL</i>	Транспортер жирных кислот	[212]
<i>fadE</i>	Ацил-КоА дегидрогеназа	[212]
<i>fadIJ</i>	Комплекс окисления жирных кислот	[212]
<i>aldA</i>	Альдегид дегидрогеназа	[215]
<i>glcDEFGBA</i>	Субъединицы гликолат оксидазы; транспортеры гликолата; малат синтаза	[216]
<i>ptsG</i>	Компонент глюкозо-специфичной фосфотрансферазной системы	[217]
<i>rpoS</i>	Альтернативный σ -фактор	[218]
<i>uvrA</i>	Компонент системы эксцизионной репарации оснований	[219]

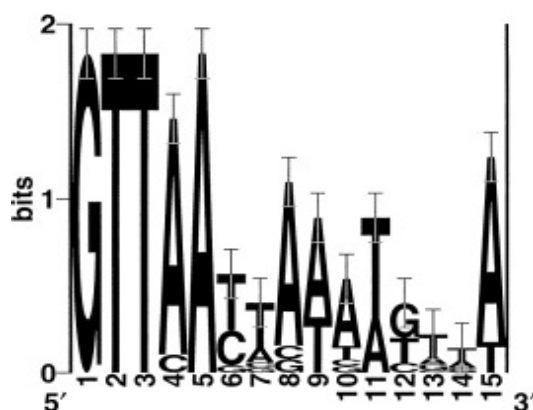


Рисунок 1.16. Диаграмма Лого для сайтов связывания ArgA

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

1.6.5. Регуляция нитрат-нитритного дыхания: двукомпонентные системы NarX-NarL и NarQ-NarP

Системы NarX-NarL и NarQ-NarP представляют собой две попарно гомологичные пары из сенсорной киназы и регулятора [220]. В присутствии эффекторов, которыми являются ионы нитрата и нитрита, сенсорная киназа автофосфорилируется и фосфорилирует регулятор, который в результате приобретает способность связываться с ДНК. Для этой удвоенной двукомпонентной системы характерен сложный характер взаимодействий, включающий в себя три уровня – взаимодействие эффекторов с киназами, взаимодействие киназ с регуляторами и взаимодействие регуляторов с сайтами (Рис. 1.17). Так, киназа NarQ способна взаимодействовать с обеими белками NarL и NarP, тогда как NarX взаимодействует только с NarL. При этом характер взаимодействий зависит от эффектора – в присутствии нитрата NarX фосфорилирует NarL, а в присутствии нитрита – дефосфорилирует, таким образом инактивируя его [221, 222]. Кроме того, белок NarL способен связываться как с сайтами NarP, так и своими собственными, что автоматически делает NarP-регулон частью NarL-регулона [221-223]. В дополнение ко всему, гены белков регуляторных систем связаны между собой сетью регуляторных каскадов [221, 222, 224]. Скорее всего, столь сложная система возникла из-за необходимости отвечать сразу на два эффектора, концентрации которых тесно связаны между собой.

В присутствии эффекторов белки регуляторных систем активируют гены нитрат- и нитрит-редуктаз, а также гены необходимых для формирования дыхательных цепей дегидрогеназ. При этом репрессируется экспрессия генов менее эффективных редуктаз (Табл. 1.4, 1.5). Кроме того, на экспрессию некоторых генов белки NarP и NarL влияют по-разному. Так, транскрипция оперонов *nar* и *nrf* активируется NarP, но репрессируется NarL

[196, 225]. В условиях столь сложной регуляции имеет смысл говорить об общем NarL/NarP-регулоне.

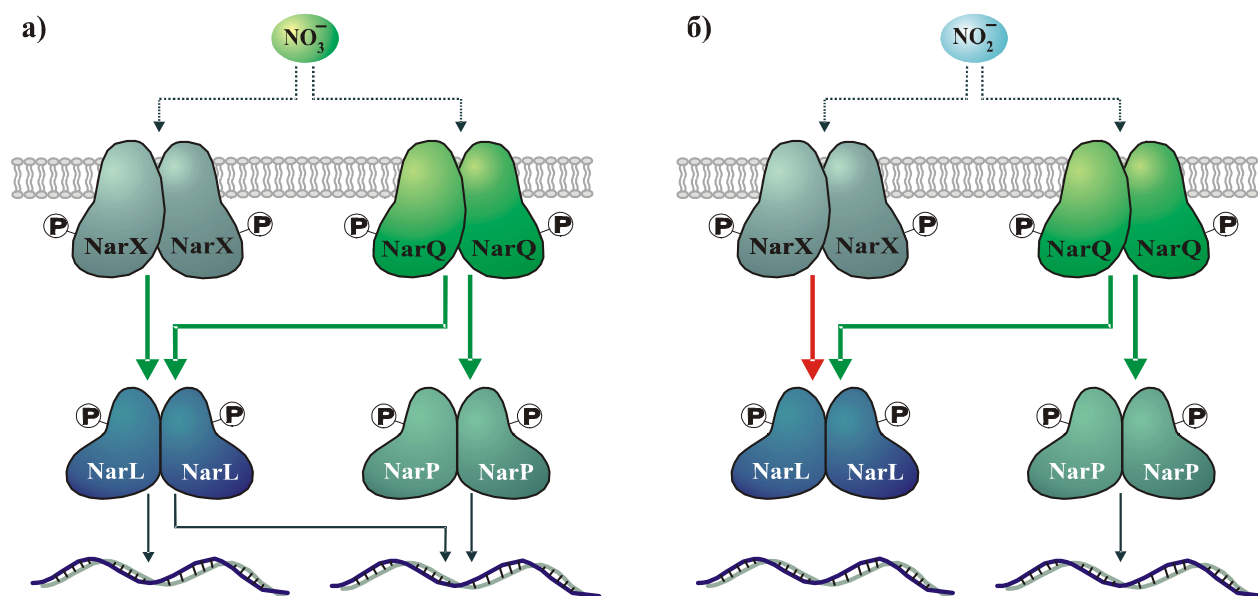


Рисунок 1.17. Взаимодействия в системе регуляции нитрат-нитритного дыхания

Тонкими стрелками показаны взаимодействия между регуляторами и сайтами. Толстыми стрелками – взаимодействия сенсорная киназа-регулятор (зеленым – активация, красным – инактивация). Взаимодействия между эффектором и сенсорной киназой показаны пунктирными стрелками.

Структура сайта связывания белка NarP хорошо известна: он представляет собой инвертированный повтор длиной 16 п.н. с консенсусом TACSSWTNNAWSSGTA [223]. Что же касается структуры NarL-сайта, то несмотря на то, что этот белок был обнаружен почти 30 лет назад [226], до сих пор в литературе нет единого мнения о строении его сайта связывания. Так, по одним сведениям, NarL связывается с одиночными сайтами с консенсусом TACSSWT [227], по другим – его мотив не отличается от такового для NarP [228], по третьим – NarL связывается с повторами одиночных сайтов, находящихся на расстоянии 13-14 п.н. [229]. Некоторые исследователи считают, что данный белок может связываться с любыми комбинациями одиночных сайтов [223].

Таблица 1.5. Функции генов NarL -регулона *E. coli*

Указаны ссылки на работы, в которых экспериментально подтверждена NarL-зависимая регуляция для данного оперона.

Оперон	Функции генов оперона	Ссылки
Гены, активируемые NarL		
<i>nirBDC-cysG</i>	Цитоплазматическая нитрит-редуктаза	[225, 230]
<i>narGHJI</i>	Цитоплазматическая нитрат-редуктаза	[220, 231]
<i>narK</i>	Нитрат-нитрит антипортер	[232, 233]
<i>fdnGHI</i>	Формиат дегидрогеназа	[195]
<i>nuoA-N</i>	NADH дегидрогеназа	[234]
<i>focA-pflB</i>	Транспортер формиата; пируват лиаза	[235]
<i>hcr-hcr</i>	Система формирования железо-серных кластеров	[191]
<i>narXL</i>	Двухкомпонентная система регуляции нитрат-нитритного дыхания	[224]
<i>narP</i>	Регулятор нитрат-нитритного дыхания	[224]
<i>narQ</i>	Сенсорная киназа белков NarL и NarP	[224]
Гены, репрессируемые NarL		
<i>narFDAGHBC-ccmABCDEFGH</i>	Периплазматическая нитрат редуктаза	[196, 236]
<i>nrfABCDEFGF</i>	Периплазматическая нитрит редуктаза	[225, 230]
<i>frdABCD</i>	Фумарат редуктаза	[237]
<i>dmsABC</i>	Редуктаза диметил-сульфоксида	[238]
<i>torCAD</i>	Редуктаза триметил-оксида азота	[239]

Таблица 1.6. Функции генов NarP -регулона *E. coli*

Указаны ссылки на работы, в которых экспериментально подтверждена NarP-зависимая регуляция для данного оперона.

Оперон	Функции генов оперона	Ссылки
Гены, активируемые NarP		
<i>nirBDC-cysG</i>	Цитоплазматическая нитрит-редуктаза	[225, 230]
<i>narFDAGHBC-ccmABCDEFGH</i>	Периплазматическая нитрат редуктаза	[196, 236]
<i>nrfABCDEFGF</i>	Периплазматическая нитрит редуктаза	[225, 230]
<i>fdnGHI</i>	Формиат дегидрогеназа	[195]
<i>nuoA-N</i>	NADH дегидрогеназа	[234]
<i>focA-pflB</i>	Транспортер формиата; пируват лиаза	[235]
<i>hcr-hcr</i>	Система формирования железо-серных кластеров	[191]
<i>narXL</i>	Двухкомпонентная система регуляции нитрат-нитритного дыхания	[224]
Гены, репрессируемые NarP		
<i>frdABCD</i>	Фумарат редуктаза	[237]

Материалы и методы

2.1. Общие принципы сравнительного подхода к регуляции

В настоящей работе для определения принадлежности гена к регулону применялся метод проверки соответствия с различными модификациями. В соответствии с этим методом, ген рассматривался как член регулона, если в его регуляторной области или регуляторной области оперона, содержащего его, обнаружен потенциально сайт связывания, который сохраняется перед ортологичными генами в родственных геномах [18, 114].

При исследовании FruR-зависимой регуляции применялся классический вариант метода проверки соответствия: изначально проводился поиск потенциальных сайтов в базовом геноме, в данном случае в геноме *E. coli*, в результате чего выделялся набор генов, имеющих потенциальные сайты в предполагаемой промоторной области. Далее проводился поиск ортологов в других геномах и проверялось наличие сайтов перед ними. В случае, если сайт находился как минимум в трех геномах, ген считался принадлежащим к обобщенному регулону. Исключение составляли случаи, когда сайты перед геном были обнаружены лишь в геномах *E. coli*, *S. typhi* и *S. typhimurium*. Ввиду того, что эти организмы крайне близки между собой, консервативность сайта зачастую определяется остаточным сходством последовательностей. Поэтому в таких случаях ген не включался в обобщенный регулон.

При исследовании регуляции белками PurR и RbsR был использован модифицированный метод проверки соответствия. В данном случае проводилось попарное сравнение всех геномов внутри одного таксона и ген относился к обобщенному регулону, если перед ним удавалось найти консервативные сайты в большинстве геномов из этого таксона. Если ген был отнесен к обобщенному регулону, то проверялось также наличие сайтов перед его ортологами в других таксонах. Такой подход позволяет выявить не только новые члены регулона, но и зафиксировать случаи таксон-специфичной регуляции.

В случае глобальной регуляции дыхания также проводилось попарное сравнение всех геномов в пределах таксона. Однако, в данной ситуации исследовались сразу три регулятора и, если ген был отнесен как минимум к одному обобщенному регулону минимум в одном таксоне, то наличие сайтов перед ним проверялось для всех трех регуляторов во всех таксонах. В случае регуляции дыхания исследовалось по четыре генома из порядков Pasteurellales и Vibrionales, а из порядка Enterobacteriales – всего два, *Y. pestis* и *Y. enterocolitica*. Поэтому для Pasteurellales и Vibrionales критерием отнесения гена к

обобщенному регулону стало наличие сайта как минимум в трех геномах, а для *Enterobacteriales* – в обоих исследуемых геномах.

Поиск потенциальных сайтов связывания регуляторных белков проводился с использованием метода матриц (профилей) позиционных весов нуклеотидов [18]. Суть метода заключается в следующем: на основе выравнивания регуляторных сайтов, каждый из которых имеет длину L , так называемой обучающей выборки, вычисляется вес $W(b,i)$ каждого нуклеотида b в позиции i . Позиционные веса нуклеотидов вычисляются по формуле:

$$W(b,i) = \log [N(b,i)+0,5] - 0,25 \sum_{b=A,C,G,T} \log [N(b,i)+0,5]$$

где $N(b,i)$ – частота нуклеотида b в позиции i . Используя полученную матрицу, можно поставить в соответствие любой последовательности длины L вес S , равный

$$S = \sum_{i=1..L} W(b_i,i)$$

где b_i – нуклеотид в позиции i . В дальнейшем в качестве потенциальных регуляторных сайтов рассматриваются лишь последовательности, имеющие вес S выше некоего порогового значения. Пороговое значение как правило определяется исходя из весов сайтов, входящих в обучающую выборку. В большинстве случаев в качестве порогового значения используется самый низкий вес для сайтов из обучающей выборки.

В случаях, когда оперонная структура участка последовательности была неизвестна, гены считались принадлежащими к одному оперону, если выполнялся ряд условий: гены имеют одинаковое направление транскрипции, расстояние между ними не превышает 200 п.н. и структура локуса сохраняется в родственных геномах. Данный подход, основанный на сравнении структур локусов в родственных геномах, достаточно хорошо зарекомендовал себя ранее в биоинформатических исследованиях [43].

2.2. Объект исследования и банки данных

В качестве объекта исследований была выбрана группа гамма-протеобактерий. Эта группа в настоящее время представляется наиболее изученной среди всех микроорганизмов. Более того, для данной группы максимально количество организмов с известной полной геномной последовательностью. Так, в соответствии с базой данных KEGG [2], по состоянию 1 апреля 2009 года были известны полные последовательности геномов для 106 видов микроорганизмов из этой группы. Если учитывать все геномы различных штаммов одного вида, то количество полных геномов возрастет до 209. Следует также отметить и большое хозяйственное значение гамма-протеобактерий – данная группа содержит массу симбионтов, паразитов и патогенов как животных, так и растений. Многие штаммы применяются в биотехнологической промышленности.

В целом было исследовано 23 геномных последовательности гамма-протеобактерий, из которых 20 являются полными.

Полные последовательности геномов *Escherichia coli* K12 [240], *Salmonella typhi* Ty2 [241], *Salmonella typhimurium* LT2 [242], *Yersinia pestis* KIM [243], *Yersinia pseudotuberculosis* IP32953 [244], *Yersinia enterocolitica* 8081 [245], *Pectobacterium carotovorum* SCRI1043 ([246]), *Photobacterium luminescens* TTO1 [247], *Pasteurella multocida* PM70 [248], *Haemophilus influenzae* KW20 Rd [1], *Haemophilus ducreyi* 35000HP, *Vibrio cholerae* O1 N16961 [249], *Vibrio fischeri* ES114 [250], *Vibrio parahaemolyticus* O3:K6 [251], *Vibrio vulnificus* CMCP6 [252], *Photobacterium profundum* SS9 [253], *Pseudomonas aeruginosa* PAO1 [53], *Pseudomonas putida* KT2440 [254], *Pseudomonas fluorescens* PfO-1 и *Pseudomonas syringae* DC3000 [255] были взяты из базы данных GenBank [75]. Предварительные последовательности геномов были взяты со следующих интернет-ресурсов: последовательность *Pectobacterium chrysanthemi* 3937 – с сайта The Institute for Genomic Research (<http://www.tigr.org/>), *Serratia marcescens* Db11 – с сайта The Sanger Institute (<http://www.sanger.ac.uk/>), *Actinobacillus actinomycetemcomitans* HK1651 – с сайта University of Oklahoma's Advanced Center for Genome Technology (<http://www.genome.ou.edu/>). Для удобства обозначений каждому из исследованных геномов было присвоено трехбуквенное обозначение (Табл. 2.1).

2.3. Программное обеспечение

Для поиска ортологов и предсказания потенциальных регуляторных сайтов в бактериальных геномах использовался пакет программ Genome Explorer [110]. За критерий ортологичности белков было принято наибольшее сходство при двухстороннем поиске в двух геномах [85]. Для построения матриц позиционных весов была использована программа SignalX [110]. Поиск гомологичных последовательностей по базам данных проводился с помощью программы BLAST [5], при этом в качестве порогового значения e-value было принято e^{-20} . Для выравнивания нуклеотидных и аминокислотных последовательностей, а также для построения филогенетических деревьев по методу объединения соседних пар (neighbour-joining method), использовалась программа ClustalX [6]. Визуализация филогенетических деревьев осуществлялась с помощью программы GeneMaster (А.А. Миронов, неопубл.) Для построения диаграмм Лого, отображающих структуру регуляторного сигнала, была применена программа WebLogo [256, 257]. Существование потенциальных ДНК-связывающих НТН -доменов анализировалось с помощью программы НТН-finder [258].

Таблица 2.1. Условные обозначения исследованных геномов

Геном	Обозначение
Enterobacteriales	
<i>Escherichia coli</i> K12	ECO
<i>Salmonella typhi</i> Ty2	STY
<i>Salmonella typhimurium</i> LT2	STM
<i>Yersinia pestis</i> KIM	YPK
<i>Yersinia pseudotuberculosis</i> IP32953	YPS
<i>Yersinia enterocolitica</i> 8081	YEN
<i>Pectobacterium carotovorum</i> SCRI1043	PCA
<i>Pectobacterium chrysanthemi</i> 3937	PCH
<i>Serratia marcescens</i> Db11	SMA
<i>Photobacterium luminescens</i> TTO1	PLU
Pasteurellales	
<i>Pasteurella multocida</i> PM70	PMU
<i>Actinobacillus actinomycetemcomitans</i> HK1651	AAC
<i>Haemophilus influenzae</i> KW20 Rd	HIN
<i>Haemophilus ducreyi</i> 35000HP	HDU
Vibrionales	
<i>Vibrio vulnificus</i> CMCP6	VVU
<i>Vibrio parahaemolyticus</i> O3:K6	VPA
<i>Vibrio cholerae</i> O1 N16961	VCH
<i>Vibrio fischeri</i> ES114	VFI
<i>Photobacterium profundum</i> SS9	PPR
Pseudomonadales	
<i>Pseudomonas aeruginosa</i> PAO1	PAE
<i>Pseudomonas putida</i> KT2440	PPU
<i>Pseudomonas fluorescens</i> PfO-1	PFO
<i>Pseudomonas syringae</i> DC3000	PST

Исследование эволюции обобщенного FruR (Cra)-регулона

3.1. Изучение эволюции регуляторной системы

С целью выбора геномов гамма-протеобактерий, в которых целесообразно исследование FruR-зависимой регуляции был проведен поиск ортологов (см. “Материалы и методы”) белка FruR из *E. coli* в геномах различных гамма-протеобактерий. В общей сложности, ортологи белка-регулятора были обнаружены в геномах 19 представителей исследуемой группы микроорганизмов, относящихся к четырем порядкам: Enterobacteriales (*E. coli*, *S. typhi*, *S. typhimurium*, *Y. pestis*, *Y. pseudotuberculosis*, *Y. enterocolitica*, *P. carotovorum*, *P. chrysanthemi*, *P. luminescens*, *S. marcescens*), Pasteurellales (*P. multocida*, *A. actinomycetemcomitans*), Vibrionales (*V. vulnificus*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*, *P. profundum*) и Pseudomonadales (*P. aeruginosa*, *P. putida*, *P. fluorescens*, *P. syringae*). Во всех найденных ортологах, за исключением белка из *A. actinomycetemcomitans*, были найдены предполагаемые НТН-мотивы (Приложение 1). Таким образом, исследование FruR-зависимой регуляции целесообразно проводить лишь в 18 геномах.

Как следует из анализа филогенетического дерева для ортологов FruR (Рис. 3.1), эволюционные расстояния между белками из разных организмов невелики. Кроме того, достаточно четко выражены четыре ветви, каждая из которых включает в себя лишь белки из организмов, относящихся к одному порядку.

3.2. Построение распознающего правила для поиска потенциальных сайтов связывания FruR

Поскольку все найденные ортологи FruR достаточно близки эволюционно (Рис. 3.1), для поиска потенциальных сайтов связывания этого белка возможно использование единой матрицы позиционных весов.

Для построения такой матрицы были рассмотрены 5'-некодирующие области генов, регуляция которых была показана экспериментально (Табл. 3.1). В результате с помощью программы SignalX была построена матрица для распознавания потенциальных сайтов связывания белка FruR (Рис. 3.2). Сайт связывания белка FruR представляет собой нестрогий палиндром длиной 16 нуклеотидов (Рис. 3.3), консенсус сайта при этом согласуется с ранее предсказанным [139].

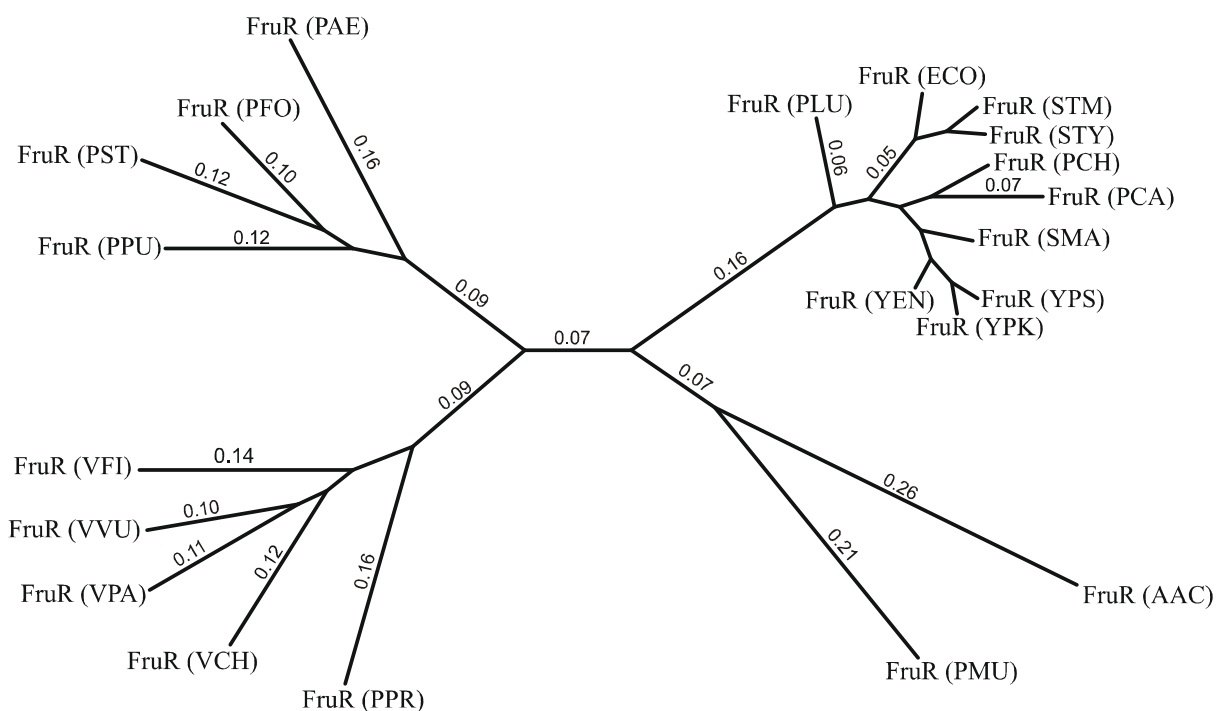


Рис. 3.1. Филогенетическое дерево белков FruR

Для построения использовался метод объединения соседних пар (neighbour-joining). Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,10). В скобках указаны условные обозначения геномов (см. Табл. 2.1).

Таблица 3.1. Сайты связывания белка FruR *E. coli*, вошедшие в обучающую выборку
Указано положение сайта относительно старта трансляции гена. Приводятся ссылки на работы, в которых регуляция гена была показана экспериментально. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Ген	Сайт			
	Положение	Вес	Последовательность	Ссылки
<i>epd</i>	-136	5,21	GCTGAAgCGtTTCAGt	[135]
<i>fruB</i>	-34	5,17	GCTGAATCGtTTCaAt	[140]
<i>icdA</i>	-247	4,93	GCTGAATCGcTTaAcC	[140, 144]
<i>pfkA</i>	-122	4,86	cCTGAATCaATTCAGC	[139]
<i>aceB</i>	-251	4,66	cgTtAAgCGATTCAGC	[140]
<i>glk</i>	-133	4,65	GCTGAAaCGATaaAGt	[141]
<i>edd</i>	-112	4,64	aCTGAAaCGtTTttGC	[135]
<i>pykF</i>	-222	4,61	ctTGAATgGtTTCAGC	[134]
<i>ppsA</i>	-101	4,57	GCTtgAaCGATTCaC	[140]
<i>adhE</i>	-263	4,56	GCTGAAagGtgTCAGC	[142, 143]
<i>mtlA</i>	-124	4,52	agTtAAcCGATTCAGt	[135]
<i>ptsH</i>	-220	4,50	GCTGAATCGATTTtat	[136, 140]
<i>pckA</i>	-287	3,99	GgTGAATCGATaCttt	[135]
<i>nirB</i>	-48	3,92	GCTGAATCGtTaagGt	[145, 146]

Поз.	A	C	G	T	Конс.
1	0.14	-0.01	0.26	-0.39	G
2	-0.22	0.23	0.03	-0.05	C
3	0.10	-0.28	-0.28	0.45	T
4	-0.05	-0.36	0.34	0.07	G
5	0.45	-0.29	-0.08	-0.08	A
6	0.50	-0.03	-0.24	-0.24	A
7	0.11	-0.23	-0.06	0.18	W
8	-0.31	0.41	0.00	-0.10	C
9	-0.10	0.00	0.41	-0.31	G
10	0.18	-0.06	-0.23	0.11	W
11	-0.24	-0.24	-0.03	0.50	T
12	-0.08	-0.08	-0.29	0.45	T
13	0.07	0.34	-0.36	-0.05	C
14	0.45	-0.28	-0.28	0.10	A
15	-0.05	0.03	0.23	-0.22	G
16	-0.39	0.26	-0.01	0.14	C

Рисунок 3.2. Позиционная матрица весов для сайтов связывания FruR
 Указан вес данного нуклеотида в соответствующей позиции. “Поз.” – позиция сайта; “Конс.” – нуклеотид, преобладающий в данной позиции.



Рисунок 3.3. Диаграмма Лого для сайтов связывания FruR

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

В то же время, значимых потенциальных сайтов не обнаружено перед оперонами *acnA*, *acnB* и *cydAB*. Ранее для всех этих оперонов было установлено, что мутация в гене *fruR* влияет на их экспрессию, но связывание белка регулятора с промоторной областью либо не исследовалось [259], либо было показано отсутствие такового [200]. Таким образом, отсутствие потенциальных сайтов связывания подтверждает, что FruR не регулирует напрямую экспрессию трех перечисленных выше оперонов, а влияет на их транскрипцию опосредованно.

Среди сайтов обучающей выборки наименьший вес, 3,92, имеет сайт, обнаруженный перед опероном *nirBDC-cysG*. Однако, в других геномах возможно существование сайтов и с меньшим весом, в связи с потенциальными особенностями взаимодействий белка FruR с ДНК вполне. Поэтому при поиске потенциальных FruR-сайтов с помощью имеющейся матрицы использовалось пороговое значение веса сайта, равное 3,75.

Хотя все экспериментально определенные сайты связывания FruR находятся на расстоянии не более -300 п.н. от старта трансляции гена, в настоящей работе поиск сайтов производился в области от -400 до +100 п.н. Использование столь широкой области поиска обусловлено следующими причинами:

- поскольку фактор транскрипции FruR может являться активатором, сайты связывания данного белка могут находиться на достаточном удалении от старта транскрипции, так как в случае активации сайты связывания белка регулятора должны находиться в области перед промотором;
- в случае репрессии транскрипции гена сайты связывания белка-регулятора могут располагаться в кодирующей области регулируемого гена; кроме того, расположение сайта внутри кодирующей области может представлять собой артефакт, вызванный неправильным определением старта трансляции при аннотации генома.

Сайты, перекрывающиеся с кодирующими областями других генов, не рассматривались.

При этих условиях потенциальные сайты связывания FruR обнаруживаются в каждом геноме приблизительно перед 300 генами. Ясно, что отдельные предсказания таких сайтов недостоверны. Однако, неоднократно было показано, что после применения процедуры проверки соответствия могут остаться лишь единичные ложные показания.

3.3. Структура обобщенного FruR-регулона в исследованных геномах

В соответствии с процедурой поиска членов обобщенного регулона, описанной в главе 2, к обобщенному регулону было отнесено 57 генов, при этом состав регулона довольно сильно отличается в разных таксонах. Обобщенный FruR-регулон включает в себя все гены, входящие в регулон в *E. coli*, и, кроме того, по результатам настоящего исследования к регулону было отнесено еще 28 генов (Табл. 3.2, Рис. 3.4). Последовательности всех сайтов связывания белка FruR приведены в Приложении 2. За счет обнаружения новых членов регулона к трем функциональным группам генов, описанным в Главе 1, добавились еще три новых. Каждая из шести функциональных групп подробно описывается далее.

Таблица 3.2. Гены, включенные в состав обобщенного FruR-регулона.

Условные обозначения: “+” – сайт обнаружен непосредственно перед геном; “*” – сайт обнаружен перед первым геном оперона, в который входит данный ген; “-” – сайта не обнаружено ни перед геном, ни перед опероном; “0” – в исследуемом геноме не найдено ортолога гена. # - регуляция гена была впервые предсказана в настоящей работе. Обозначения геномов: Табл. 2.1.

Геномы																						
Гены	Enterobacteriales										Pasteurellales		Vibrionales					Pseudomonadales				
	ECO	STY	STM	PCA	PCH	YPK	YEN	YPS	SMA	PHL	PMU	AAC	VVU	VPA	VCH	VFI	PBP	PAE	PPU	PFO	PST	
Гены ферментов центрального метаболизма (продолжение)																						
<i>pfkA</i>	+	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	0	0	0	0	
<i>pykF</i>	+	+	+	+	0	+	+	+	+	+	0	-	-	-	-	-	-	0	0	0	0	
<i>pdhR</i> [#]	+	+	+	+	+	+	+	+	-	-	0	-	-	-	-	-	-	-	-	-	-	
<i>aceE</i> [#]	*	*	*	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	
<i>aceF</i> [#]	*	*	*	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	
<i>lpdA</i> [#]	*	*	*	*	*	*	*	*	*	*	*	-	-	-	-	-	-	-	-	-	-	
<i>icdA</i>	+	+	+	+	+	+	+	+	-	-	0	0	0	0	0	0	0	0	0	0	0	
<i>adhE</i>	+	+	+	+	+	+	+	+	+	+	-	-	-	-	-	-	-	0	0	0	0	
<i>aceB</i>	+	+	+	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0	0	
<i>aceA</i>	*	*	*	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0	0	
<i>aceK</i>	*	*	*	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	0	0	0	
Гены блсков дыхательных комплексов																						
<i>nuoA</i> [#]	+	+	+	+	0	+	+	+	+	+	+	+	+	+	+	+	+	0	0	0	0	
<i>nuoB</i> [#]	*	*	*	*	0	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoC</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoD</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoE</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoF</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoF</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoG</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoH</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoI</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoJ</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoK</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0
<i>nuoL</i> [#]	*	*	*	*	-	*	*	*	*	*	*	*	*	*	*	*	*	*	0	0	0	0

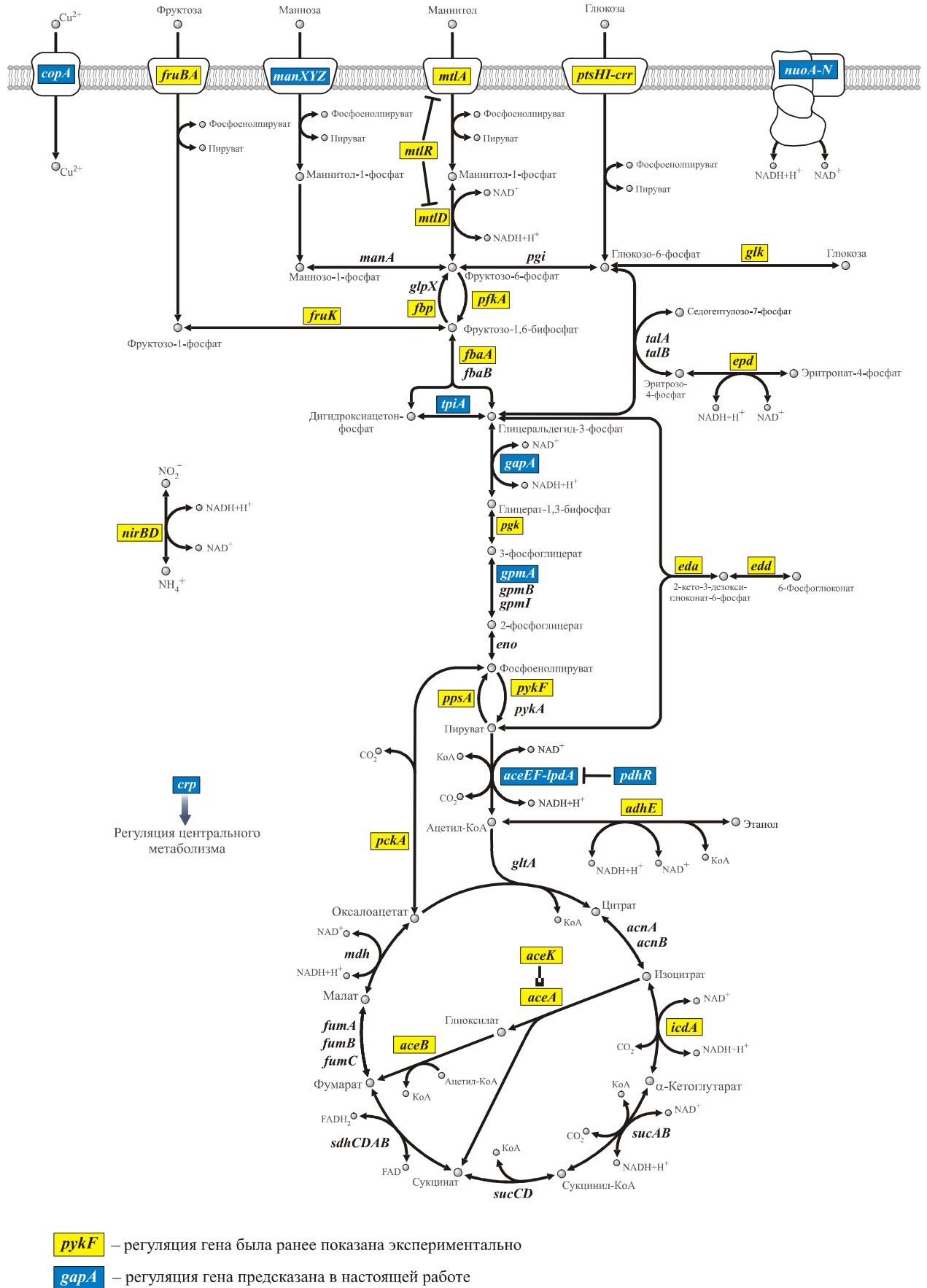


Рисунок 3.4. Новые члены обобщенного FruR-регулона гамма-протеобактерий
 Расшифровка цветowych обозначений дана в нижней части рисунка.

3.3.1. Гены белков фосфотрансферазных систем

Ранее регуляция белком FruR была показана экспериментально для оперонов *fruBKA* [140], *mtlADR* [135] и *ptsHI-crr* [136, 140]. Регуляция оперона *fruBKA* крайне консервативна, потенциальные сайты перед этим опероном были найдены во всех исследованных геномах, за исключением представителей порядка Pasteurellales. FruR-зависимая регуляция оперона *ptsHI-crr* является характерной особенностью Enterobacteriales: почти во всех геномах из данного таксона были обнаружены сайты связывания фактора транскрипции перед этим опероном, тогда как в других таксонах либо не найдено значимых сайтов, либо отсутствуют ортологи генов данного оперона. Регуляция же оперона *mtlADR* сохраняется лишь в крайне узкой таксономической группе, включающей в себя лишь *E. coli* и *Salmonella* spp.

Кроме того, в настоящей работе была предсказана регуляция оперона *manXYZ*. Данный оперон кодирует белки маннозо-специфичной фосфотрансферазной системы [260-263]. Потенциальные сайты связывания FruR обнаружены перед данным опероном в геномах всех Enterobacteriales, за исключением *P. carotovorum*, *S. marcescens* и *P. luminescens*. В геномах Pasteurellaceae найдены ортологи соответствующих генов, но значимых сайтов перед ними не обнаружено, тогда как в геномах Vibrionales и Pasteurellales ортологи для генов *manXYZ* отсутствуют. FruR-зависимая регуляция оперона *manXYZ* представляется вполне обоснованной по ряду причин. Во-первых, данный фактор транскрипции контролирует экспрессию оперонов для других фосфотрансферазных систем. Во-вторых, маннозо-6-фосфат, продукт осуществляемой ферментативным комплексом ManXYZ реакции, может изомеризоваться во фруктозо-6-фосфат, который затем включается в центральный метаболизм [264].

Также для генов фосфотрансферазных систем были обнаружены как оперонные перестройки, так и изменения доменной структуры. Так, в большинстве исследованных геномов гены фруктозо-специфичной фосфотрансферазной системы, *fruB* и *fruA*, и ген фосфофруктокиназы *fruK* образуют один оперон. Однако, в геноме *P. luminescens* такая структура не сохраняется: гены фосфотрансферазной системы образуют оперон *fruBA*, тогда как ген фосфофруктокиназы *fruK* располагается в другом участке хромосомы. Тем не менее, потенциальные сайты связывания FruR в данном геноме обнаружены перед обоими вышеуказанными оперонами.

В геномах бактерий из группы Pseudomonadales были обнаружены доменные перестройки для генов фруктозо-специфичной фосфотрансферазной системы. Такие перестройки были замечены для оперона *fruBKA*. В случае Pseudomonadales первый ген из этого оперона кодирует химерный белок, N- и C-концевые участки которого ортологичны соответственно белкам FruV и PtsI *E. coli* (Рис. 3.5; Приложение 3). Известно, что у *E. coli*

оперон *fruBKA* кодирует белки FruB и FruA, формирующие фруктозо-специфичный компонент фосфотрансферазной системы – так называемый энзиматический комплекс II (PTS enzyme II; EII). Однако для нормального функционирования любой фосфотрансферазной системы требуются такие общие для всех таких систем компоненты, как энзиматический комплекс I (PTS enzyme I; EI) и гистидин-богатый белок (Histidyl phosphorylatable protein; HPr). Данные белки необходимы для переноса фосфатного остатка с фосфоенолпирувата на комплекс EII, который затем переносит соответствующие углеводы в цитоплазму, фосфорилируя их (Рис. 3.6). В геноме *E. coli* EI и HPr кодируются, соответственно, генами *ptsI* и *ptsH* [263, 265]. Кроме того, функцию HPr может, по всей видимости, выполнять С-концевой домен белка FruB, имеющий значительное сходство с PtsH [266]. Таким образом, у *Pseudomonadales* *fru*-оперон кодирует не только гены специфичного EII-комплекса, но также и белки EI, и, возможно, HPr, таким образом включая в себя все гены, необходимые для работы фруктозо-специфичной фосфотрансферазной системы. Возможно, это связано с тем, что в геномах бактерий данной группы гены для других фосфотрансферазных систем отсутствуют.

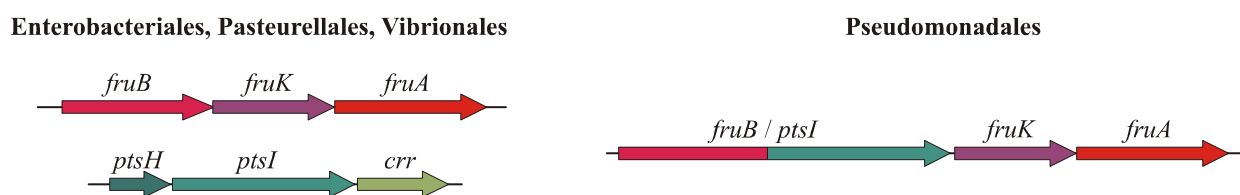


Рисунок 3.5. Оперонная структура *fru*- и *pts*-генов в различных таксономических группах гамма-протеобактерий

3.3.2. Гены ферментов центрального метаболизма

К данной группе были отнесены гены, кодирующие ферменты гликолиза и глюконеогенеза, пентозо-фосфатного пути, пути Энтнера-Дудорова, цикла трикарбоновых кислот, глиоксилатного пути и брожения.

Ранее непосредственная регуляция белком FruR была экспериментально показана для оперонов *glk*, *fbp*, *epd-pgk-fbaA*, *edd-eda*, *pckA*, *ppsA*, *pfkA*, *pykF*, *icdA*, *adhE* и *aceBAK* (см. ссылки в Главе 1).

В ходе исследования выяснилось, что регуляция всех этих оперонов характерна лишь для геномов представителей Enterobacteriales (Табл. 3.2). Так, сайты перед оперонами *ppsA*, *pfkA*, *pykF*, *adhE* и *glk* были найдены практически во всех геномах из данного таксона. В случае гена *glk* наблюдается довольно необычная ситуация, когда сайт перед геном не сохраняется в

наиболее близко родственных к *E. coli* организмах, бактериях рода *Salmonella*, но сохраняется во всех других представителях Enterobacteriales, за исключением *P. carotovorum* и *P. luminescens*.

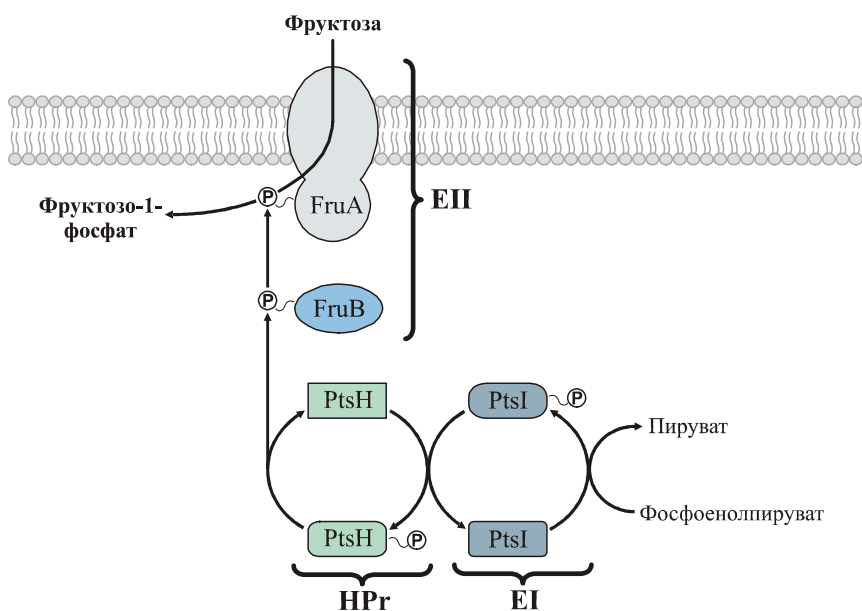


Рисунок 3.6. Строение фруктозо-специфичной фосфотрансферазной системы *E. coli*

Регуляция других оперонов, кодирующих ферменты центрального метаболизма, является менее консервативной. Например, сайты связывания FruR перед опероном *edd-eda* были обнаружены в геномах *E. coli* и *Salmonella* spp. В случае *P. carotovorum* гены *edd* и *eda* располагаются отдельно и сайт обнаруживается только перед геном *edd*. Опероны же *pckA* и *aceBAK* регулируются лишь в *E. coli* и близких к ней бактериях из рода *Salmonella*, то есть их регуляция является характерной особенностью узкой таксономической группы (Табл. 3.2).

В соответствии с процедурой выявления членов обобщенного регулона (Глава 2) к FruR-регулону были отнесены также опероны *tpiA*, *gapA*, *gpmA* и *pdhR-aceEF-lpdA*.

Ген *tpiA* кодирует фермент гликолиза триозофосфат изомеразу ([267]). Потенциальные сайты связывания FruR найдены перед опероном *tpiA* во всех Enterobacteriales, кроме *P. carotovorum*.

Сайты перед генами *gapA* и *gpmA*, кодирующими, соответственно, глицеральдегид-3-фосфат дегидрогеназу [268, 269] и фосфоглицератмутазу [270], обнаружены во всех исследованных Enterobacteriales.

Также в настоящей работе была предсказана регуляция оперона *pdhR-aceEF-lpdA*, кодирующего субъединицы фермента пируватдегидрогеназы и фактор транскрипции PdhR, осуществляющий репрессию этого оперона [271, 272]. Структура данного локуса сохраняется в геномах всех исследованных Enterobacteriales, и потенциальные FruR-сайты

перед геном *pdhR* были найдены во всех геномах из данного таксона, за исключением *S. marcescens* и *P. luminescens*.

Для генов данной функциональной был также выявлен единичный случай неортологичной замены. Так, в *E. coli* функции пируваткиназы выполняют два гомологичных белка, *PykF* и *PykA* [273, 274]. Потенциальные сайты связывания FruR были обнаружены перед ортологами гена *pykF* во всех геномах Enterobacteriales, за исключением *P. chrysanthemi* (Табл. 3.2). Подробный анализ филогенетического дерева белков *PykF* и *PykA* (Рис. 3.7) показывает, что гены, кодирующие эти белки, образовались в результате дупликации, произошедшей до расхождения исследуемых таксонов гамма-протеобактерий. При этом в геномах Pasteurellales произошла делеция, которая привела к утрате гена *pykF*. Что же касается Enterobacteriales, то дуплицированные гены пируваткиназы присутствуют во всех исследованных геномах, кроме *P. chrysanthemi*, где был обнаружен ортолог лишь для *pykA*. Однако, только в этом геноме был обнаружен потенциальный FruR-сайт перед *pykA*. Таким образом, можно предположить, что для бактерий из группы Enterobacteriales важна FruR-зависимая регуляция одного из генов пируваткиназы, и в случае делеции *pykF* в геноме *P. chrysanthemi* регулируемым становится единственный оставшийся ген, *pykA*.

3.3.3. Гены белков дыхательных комплексов

В настоящей работе к обобщенному FruR-регулону был отнесен оперон *nuoABCDEFGHIJKLMN*. Данный оперон кодирует белки комплекса протон-экспортирующей НАД·Н-дегидрогеназы [234, 275-277].

Потенциальные сайты связывания FruR были обнаружены перед *nuo*-опероном во всех исследованных геномах Enterobacteriales, за исключением *P. chrysanthemi* (Табл. 3.2). В последнем случае также отсутствовали ортологи для первых двух генов из оперона. Однако, в настоящее время доступен лишь фрагментарный геном *P. chrysanthemi* и обнаруженные гены *nuo*-оперона располагаются на краю участка с известной последовательностью. Поэтому судить о составе *nuo*-оперона и регуляции его транскрипции в *P. chrysanthemi* будет возможно лишь после выхода полной версии данного генома.

Регуляция экспрессии *nuo*-оперона белком FruR представляется вполне целесообразной, так как многие реакции центрального метаболизма являются обратимыми и сопровождаются восстановлением НАД⁺ до НАД·Н. Поэтому направление этих реакций, и, как следствие, возможность протекания центрального метаболизма в целом, зависит от соотношения окисленной и восстановленной форм НАД. В случае же, если это соотношение сильно смещено в сторону НАД·Н, возможна остановка процессов центрального метаболизма [278-281]. НАД·Н-дегидрогеназа, белковый комплекс, кодируемый генами *nuo*-оперона, окисляет

всего, регуляция данного оперона фактором FruR имеет то же значение, что и описанная выше регуляция *niio*-оперона – таким образом поддерживается баланс между концентрацией окисленной и восстановленной форм НАД. Данное предположение выдвинуто на основе биохимических характеристик периплазматической нитрит-редуктазы. Ферментативный комплекс, образуемый белками NirB и NirD, восстанавливает нитрит до аммония, используя НАД·Н в качестве донора электронов. При этом для восстановления одной молекулы нитрита требуется три молекулы НАД·Н, что достаточно сильно смещает равновесие в сторону НАД⁺, тем самым создавая условия для протекания реакций центрального метаболизма [285, 286].

3.3.5. Гены транспортных белков

В настоящей работе потенциальные сайты связывания FruR были обнаружены перед геном *copA* во всех исследованных геномах Enterobacteriales. Продукт данного гена не был исследован экспериментально, но анализ потенциальной аминокислотной последовательности показал, что белок CopA, по всей видимости, является АТФ-зависимым транспортером ионов меди [287]. Значение возможной FruR-зависимой регуляции данного оперона неясно.

3.3.6. Гены регуляторных белков

Помимо генов транспортных белков, ферментов центрального метаболизма и генов дыхания, FruR-зависимая регуляция была также предсказана для двух генов, кодирующих факторы транскрипции, *fruR* и *crp*.

Авторегуляция *fruR* была предсказана для представителей Vibrionales и Pseudomonadales. Во всех исследованных организмах из этих двух групп ген *fruR* образует дивергон с опероном *fruBKA*, в случае Vibrionales, или *fruB/ptsI-fruKA*, в случае Pseudomonadales. Сайты связывания при этом располагаются в межгенном промежутке. Скорее всего, в данных геномах белок FruR регулирует транскрипцию как оперона фосфотрансферазной системы, так и собственного гена. В случае Pseudomonadales в межгенной области обнаруживаются единичные FruR-сайты, за исключением *P. syringae*, где таких сайтов два (Приложение 4). Для всех же представителей Vibrionales было обнаружено по три сайта в промежутке между *fruR* и *fruB*. Два ближайших к *fruR* сайта расположены в области с высокой степенью консервативности, причем во всех геномах между ними сохраняется расстояние, равное 6 п.н. Возможно, что подобные тандемные сайты необходимы для усиления регуляторного эффекта в ответ на внешний стимул. Третье же

сайты, находящиеся ближе к *fruB*, располагаются в участках, которые не выравниваются между собой (Приложение 5).

Возможно, что геномах Vibrionales и Pseudomonadales FruR регулирует транскрипцию как оперона фосфотрансферазной системы, так и собственного гена. Однако, не следует исключать и гипотезу, что обнаружение сайтов перед геном *fruR* в данных геномах является следствием расположения данного гена в дивергенте с опероном фосфотрансферазной системы, и сайты регулируют экспрессию лишь *fruBKA* и *fruB/ptsI-fruKA* оперонов. Выбор между этими возможностями требует привлечения экспериментальных методов.

Потенциальные сайты связывания FruR перед геном *crp* были обнаружены во всех исследованных геномах Enterobacteriales. Продукт данного гена представляет собой глобальный регулятор транскрипции. Белок Crp переходит в активную форму в присутствии цАМФ (циклического аденозил-монофосфата), повышение концентрации которого свидетельствует о недостатке глюкозы. В описанных условиях Crp контролирует экспрессию более чем ста оперонов, тем самым контролируя процессы катаболизма различных углеводов [61, 181, 263, 288-290]. Таким образом, белок FruR не только напрямую регулирует экспрессию генов центрального клеточного метаболизма, но и модулирует катаболизм различных сахаров посредством FruR-Crp регуляторного каскада.

3.4. Эволюция FruR-регулона

В ходе исследования FruR-регулона в различных гамма-протеобактериях была замечена следующая особенность: состав регулона крайне сильно отличается в различных порядках микроорганизмов (Рис. 3.8).

Так, в Pseudomonadales в состав регулона входит лишь один оперон, *fruB/ptsI-fruKA*. Аналогичная ситуация наблюдается и в Vibrionales, где единственным членом регулона является *fruBKA* оперон. В группе Pasteurellales регулон полностью отсутствует (см. ниже).

У бактерий порядка Enterobacteriales, по сравнению с другими таксонами, наблюдается необычайное расширение регулона. Так, опероны *fruBKA*, *ptsHI-crr*, *manXYZ*, *glk*, *epd-pgk-fbaA*, *ppsA*, *pfkA*, *pykF*, *icdA*, *adhE*, *tpiA*, *gapA*, *gpmA*, *pdhR-aceEF-lpdA*, *nuoABCDEFGHIJKLMN*, *nirBDC-cysG*, *copA* и *crp* входят в состав FruR-регулона в большинстве геномов из этой группы. Другие же опероны, а именно *mtlADR*, *fbp*, *edd-eda*, *prkA* и *aceBAK*, регулируются лишь в *E. coli* и близких к ней *Salmonella* spp.

Таким образом, можно условно разделить все члены обобщенного FruR-регулона на три группы, в соответствии с таксономическим распространением их регуляции:

- гены, регулируемые во всех исследованных геномах, за исключением Pasteurellales;
- гены, регулируемые только в геномах Enterobacteriales;
- гены, регулируемые только в геномах *E. coli* и *Salmonella* spp.

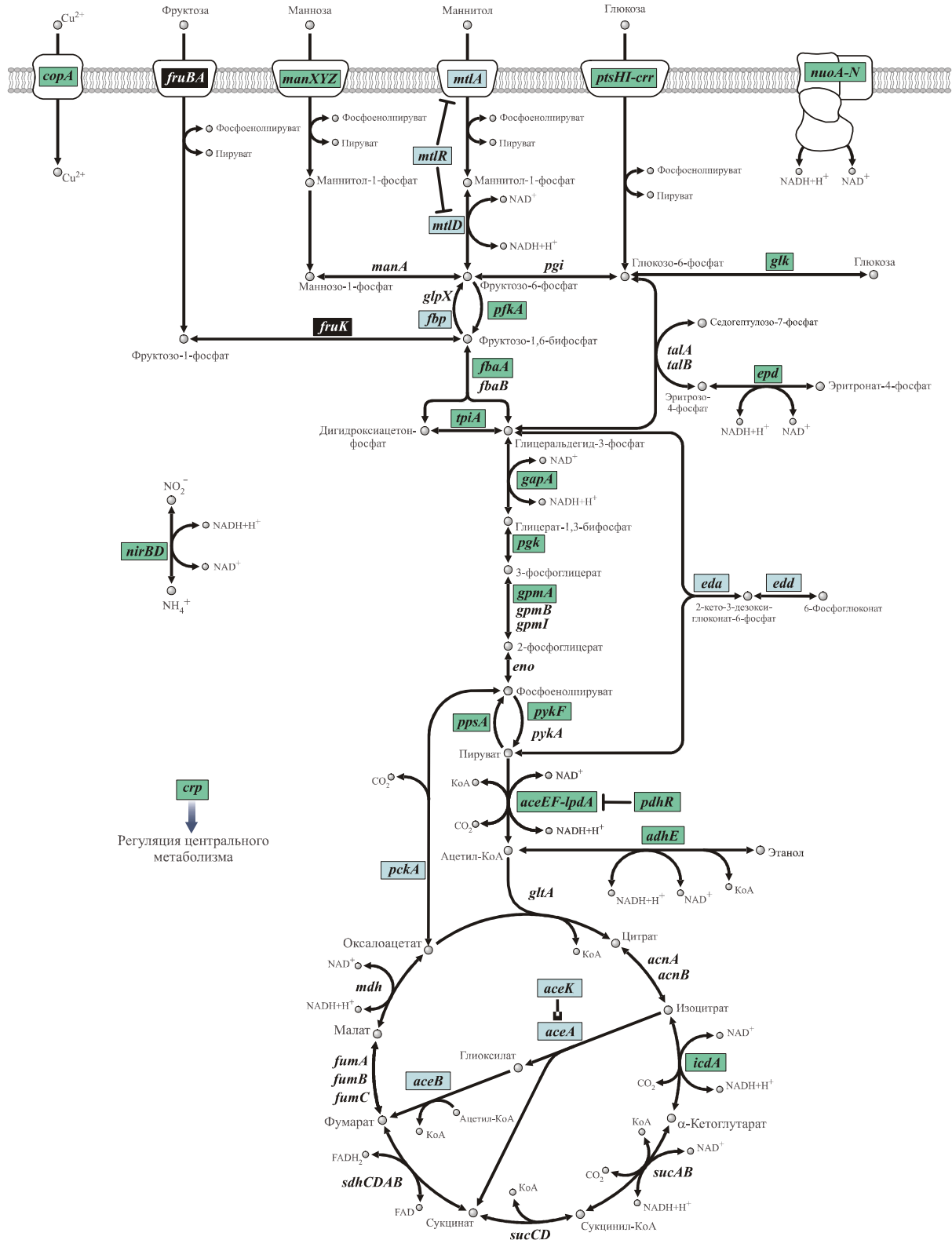
На основании подобного распределения членов регулона по таксономическим группам, был предсказан наиболее вероятный сценарий эволюции FruR-регулона (Рис. 3.9). По всей видимости, первоначально FruR-регуляция возникла у общего предка еще до отделения группы Pseudomonadales, причем первоначально FruR существовал как регулятор единственного *fru*-оперона. Такая ситуация сохранилась в представителях Pseudomonadales и Vibrionales.

В дальнейшем в группах Pasteurellales и Enterobacteriales произошли противоположные процессы. У Pasteurellales регулон разрушился в процессе эволюции, причем в разных геномах можно наблюдать разные стадии этого процесса. У части организмов, таких как *H. influenzae* и *H. dukreyi* ген *fruR* оказался полностью утрачен. У *A. actinomycetemcomitans* произошла делеция лишь части гена, затронувшая участок, соответствующий ДНК-связывающему домену, что сделало невозможным выполнение белком FruR функции регулятора транскрипции. В геноме *P. multocida*, хотя и присутствует полноразмерный ген *fruR*, не было найдено потенциальных сайтов связывания белка ни перед одним геном, входящим в регулон в других исследованных микроорганизмах.

У Enterobacteriales же, наоборот, произошло значительное расширение регулона путем появления новых сайтов связывания FruR перед генами центрального метаболизма, дыхания, ассимиляции азота, транспорта и регуляторных белков. В процессе дальнейшего эволюционного расхождения разных групп Enterobacteriales происходили изменения состава регулона путем исчезновения сайтов перед теми или иными генами.

Дальнейшее расширение FruR-регулона произошло в узкой группе, входящей в состав Enterobacteriales, и включающей в себя *E. coli* и *Salmonella* spp.

К сожалению, не удастся определить, каким же путем происходило появление новых сайтов в ходе эволюции. Интуитивно можно предположить, что сайт в регуляторной области может появиться двумя путями: за счет мононуклеотидных замен, или же за счет вставки последовательности, содержащей сайт. С целью ответа на этот вопрос были построены множественные выравнивания потенциальных регуляторных областей для ортологичных генов-членов регулона из Enterobacteriales. Однако, как выяснилось, регуляторные области ортологичных генов крайне плохо выравниваются между собой, что делает невозможным выдвинуть какие-либо предположения о пути появления сайта.



- fruK*** – регуляция сохраняется у всех представителей Enterobacteriales, Vibrionales и Pseudomonadales
- icdA*** – регуляция сохраняется у большинства представителей
- pckA*** – регуляция сохраняется только у *E. coli* и *Salmonella* spp.

Рисунок 3.8. Состав FruR-регулона в различных группах гамма-протеобактерий
Расшифровка цветowych обозначений дана в нижней части рисунка.

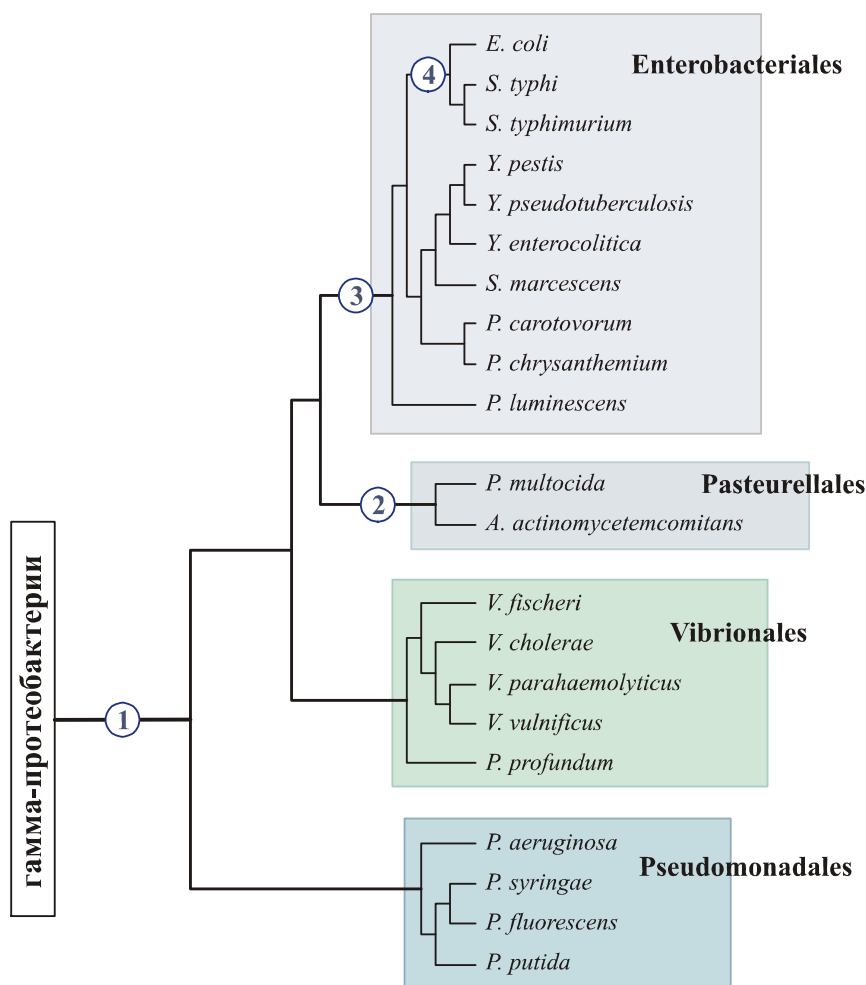


Рисунок 3.9. Предполагаемый сценарий эволюции FruR-регулона гамма-протеобактерий

Цифрами отмечены предполагаемые эволюционные события: 1 – возникновение FruR-регуляции оперона *fru*; 2 – разрушение регулона в Pasteurellales; 3 – расширение регулона в Enterobacteriales; 4 – дальнейшее расширение регулона в группе, включающей *E. coli* и *Salmonella* spp.

3.5. Обсуждение и выводы

С момента открытия фактора транскрипции FruR, осуществляющего цАМФ-независимую регуляцию глюконеогенеза и фруктозо-специфичной фосфотрансферазной системы [133], FruR-регулон был достаточно подробно исследован экспериментальными методами, первоначально в *Salmonella typhimurium* [133, 291], затем в *E. coli* [134-136, 292]. Однако, несмотря на достаточно широкую таксономическую распространенность указанного белка, никаких исследований FruR-зависимой регуляции в других бактериях, по всей видимости, не проводилось.

Ранее было замечено, что в геномах некоторых представителей гамма-протеобактерий, например, в *V. cholerae* и *P. aeruginosa* наблюдается значительное сокращение FruR-регулона по сравнению с *E. coli* (М.С. Гельфанд, личное сообщение). В настоящей работе было

проведено подробное исследование обобщенного FruR-регулона в геномах восемнадцати гамма-протеобактерий, принадлежащих к четырем порядкам.

В ходе исследования были предсказаны новые члены регулона, регуляция которых сохраняется у большинства представителей порядка Enterobacteriales. К таковым относятся гены фосфотрансферазных систем (*manXYZ*), центрального метаболизма (*tpiA*, *gapA*, *gpmA*, *pdhR-aceEF-lpdA*), дыхательных систем (*nuoABCDEFGHIJKLMN*), транспорта (*copA*) и регуляции транскрипции (*crp*). Регуляция других генов фосфотрансферазных систем и центрального метаболизма белком FruR уже была показана ранее, поэтому новые гены лишь дополнили уже сложившуюся картину функциональных составляющих регулона. Регуляция *nuo*-оперона, скорее всего, вызвана необходимостью поддержания баланса между окисленной и восстановленной формами НАД, от которого зависит протекание многих реакций центрального метаболизма.

Регуляция гена *crp* связывает между собой две регуляторные системы. Первая система, Crp, осуществляет контроль процессов катаболизма в клетке в ответ на содержание в среде глюкозы посредством реагирования на уровень цАМФ. Вторая же, FruR, отвечает на присутствие фосфорилированных производных фруктозы. Подобная связь двух глобальных систем регуляции посредством регуляторного Crp-FruR каскада обеспечивает более точный контроль за процессами катаболизма углеводов и центрального метаболизма.

К сожалению, остается неясным значение FruR-зависимой регуляции гена *copA*, по всей видимости, кодирующего транспортер ионов меди.

Другим значимым результатом данной работы является предсказание сценария эволюции FruR-регулона гамма-протеобактерий. На основании данных о структуре обобщенного регулона в различных порядках было выдвинуто предположение, что первоначально, до отделения группы Pseudomonadales, FruR представлял собой узко специфичный регулятор одного оперона, кодирующего фосфотрансферазную систему для фруктозы. В группе же Enterobacteriales произошло достаточно быстрое присоединение к регулону новых оперонов, в результате чего FruR превратился в глобальный регулятор центрального метаболизма. Процесс расширения регулона, судя по всему, происходит весьма динамично, поскольку появление дополнительных членов регулонов наблюдается в достаточно узкой группе организмов, включающей крайне близко родственных *E. coli* и *Salmonella* spp.

В заключение следует упомянуть интересные изменения в строении *fru*-оперона, характерные для бактерий порядка Pseudomonadales. Если в других исследованных геномах в состав этого оперона входят лишь гены, кодирующие фруктозо-специфичные компоненты

фосфотрансферазной системы, то у Pseudomonadales указанный оперон кодирует и остальные компоненты, являющиеся общими для всех фосфотрансферазных систем. Данный факт вполне может рассматриваться как дополнительное подтверждение гипотезы о значительной роли фруктозы в метаболизме древних бактерий, выдвинутой ранее [137].

Исследование эволюции обобщенных PurR- и RbsR-регулонов

4.1. Изучение эволюции регуляторных систем

Как уже упоминалось ранее, PurR и RbsR представляют собой белки с высокой степенью сходства аминокислотных последовательностей. Например, последовательности указанных белков из *E. coli* тождественны на 47,0%. Сходство же последовательностей белков, вычисленное с помощью матрицы аминокислотных замен BLOSUM62 [293], составляет 67,9%. При этом в *E. coli* данные белки являются ближайшими гомологами, то есть сходство между ними значительно выше, чем сходство каждого из них с любым другим белком из семейства LacI. Приведенные факты позволяют предположить, что данные белки произошли в результате относительно недавней дупликации одиночного предкового гена. По этой причине было решено провести одновременное исследование эволюции двух регуляторных систем, PurR и RbsR.

На первом этапе работы был произведен поиск ортологов белков PurR и RbsR из *E. coli* в других геномах гамма-протеобактерий. Ортологи PurR были найдены в 15 геномах бактерий из порядков Enterobacteriales (*E. coli*, *S. typhi*, *S. typhimurium*, *Y. pestis*, *Y. pseudotuberculosis*, *P. carotovorum*, *P. luminescens*), Pasteurellales (*P. multocida*, *H. influenzae*, *H. ducreyi*) и Vibrionales (*V. vulnificus*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*, *P. profundum*). Во всех перечисленных геномах, за исключением *Y. pestis*, *Y. pseudotuberculosis* и *H. ducreyi*, были также обнаружены ортологи белка RbsR. В геномах из группы Pseudomonadales (*P. aeruginosa*, *P. putida*, *P. fluorescens*, *P. syringae*) был обнаружен белок, первоначально аннотированный как RbsR [53], но при этом практически одинаково схожий с PurR и RbsR из *E. coli*. Во всех найденных ортологах PurR и RbsR были найдены предполагаемые ДНК-связывающие НТН-мотивы (Приложение 6).

Для всех найденных белков было построено филогенетическое дерево (Рис. 4.1). Анализ данного дерева позволяет выделить три группы белков:

- ортологи PurR *E. coli*;
- ортологи RbsR *E. coli*;
- белки из бактерий порядка Pseudomonadales.

Внутри групп, соответствующих PurR и RbsR, можно выделить по три ветви, соответствующие белкам из групп Enterobacteriales, Pasteurellales и Vibrionales.

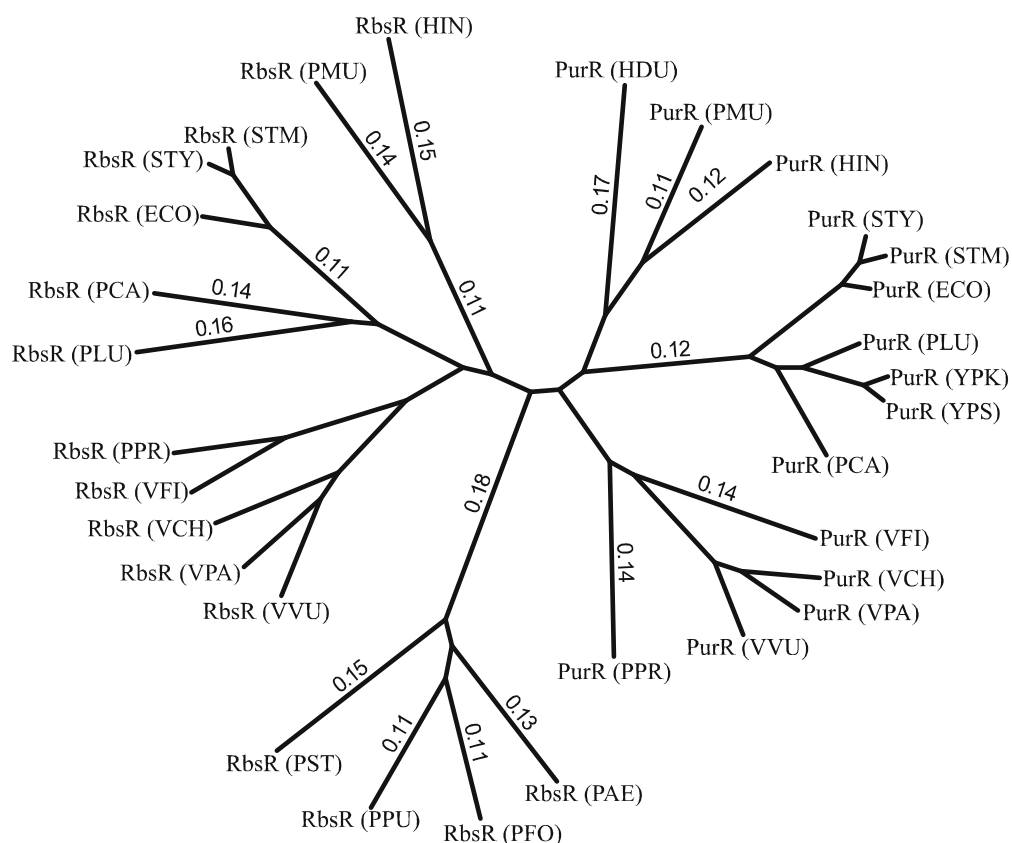


Рис. 4.1. Филогенетическое дерево белков PurR и RbsR

Для построения использовался метод объединения соседних пар (neighbour-joining). Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,10). В скобках указаны условные обозначения геномов (см. Табл. 2.1).

Можно предположить, что белок из Pseudomonadales сохраняет свойства предкового белка, а PurR и RbsR из Enterobacteriales, Pasteurellales и Vibrionales являются потомками копий, образовавшимися в результате дупликации исходного гена. Подобное предположение высказывалось и ранее на основе филогении этих белков и совместного их расположения на хромосоме с другими генами [294].

4.2. Исследование RbsR-зависимой регуляции

4.2.1. Построение распознающего правила для поиска потенциальных сайтов связывания RbsR

Как следует из анализа филогенетического дерева регуляторов (Рис. 4.1), все найденные ортологи RbsR достаточно близки. Поэтому для поиска потенциальных сайтов связывания этого белка возможно использование единой матрицы позиционных весов.

Ранее было показано, что в *E. coli* белок RbsR репрессирует экспрессию оперона *rbsDACBKR*, связываясь с сайтом, расположенном в промоторной области [148]. Кроме того,

сайты перед *rbs*-опероном были предсказаны в геномах *S. typhi*, *P. multocida*, *H. influenzae* и *V. cholerae* [19]. На основе вышеперечисленных сайтов (Табл. 4.1) была построена матрица для распознавания потенциальных сайтов связывания белка RbsR (Рис. 4.2). Сайт связывания белка RbsR представляет собой палиндром длиной 20 нуклеотидов (Рис. 4.3).

Таблица 4.1. Сайты связывания белка RbsR, вошедшие в обучающую выборку

Указано положение сайта относительно старта трансляции гена, также приводятся ссылки на работы, в которых регуляция гена была показана экспериментально. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Ген	Сайт			
		Положение	Вес	Последовательность	Ссылки
<i>E. coli</i>	<i>rbsD</i>	-32	6,93	TCAGCGAAACGTTTCGcTGA	[148]
<i>S. typhi</i>	<i>rbsD</i>	-110	6,39	ctAGCGAAACGTTTCGAcGg	[19]
		-31	6,58	cCAGCGAAACGTTTCGcTag	[19]
<i>P. multocida</i>	<i>rbsD</i>	-45	6,93	TCAtCGAAACGTTTCGATGA	[19]
<i>H. influenzae</i>	<i>rbsD</i>	-38	6,76	TtAtCGAAACGTTTCGATaA	[19]
<i>V. cholerae</i>	<i>rbsD</i>	-36	6,86	TCAtCGAAACGTTTCGATGt	[19]

Поз.	A	C	G	T	Конс.
1	-0.20	0.12	-0.20	0.28	T
2	-0.21	0.23	-0.21	0.18	C
3	0.41	-0.14	-0.14	-0.14	A
4	-0.21	-0.21	0.23	0.18	K
5	-0.14	0.41	-0.14	-0.14	C
6	-0.14	-0.14	0.41	-0.14	G
7	0.41	-0.14	-0.14	-0.14	A
8	0.41	-0.14	-0.14	-0.14	A
9	0.41	-0.14	-0.14	-0.14	A
10	-0.14	0.41	-0.14	-0.14	C
11	-0.14	-0.14	0.41	-0.14	G
12	-0.14	-0.14	-0.14	0.41	T
13	-0.14	-0.14	-0.14	0.41	T
14	-0.14	-0.14	-0.14	0.41	T
15	-0.14	0.41	-0.14	-0.14	C
16	-0.14	-0.14	0.41	-0.14	G
17	0.23	0.18	-0.21	-0.21	M
18	-0.18	0.04	-0.18	0.33	T
19	0.08	-0.03	0.20	-0.25	G
20	0.13	-0.26	0.06	0.06	A

Рисунок 4.2. Позиционная матрица весов для сайтов связывания RbsR

Указан вес каждого нуклеотида в соответствующей позиции. “Поз.” – позиция сайта; “Конс.” – нуклеотид, преобладающий в данной позиции.



Рисунок 4.3. Диаграмма Лого для сайтов связывания RbsR

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

Наименьший вес сайта из обучающей выборки равен 6,39. Однако, учитывая возможность изменения структуры сайта в других геномах, для поиска сайта было принято пороговое значение 5,50. Поиск сайта проводился в области от –300 до +100 п.н. относительно старта трансляции гена. Сайты, перекрывающиеся с кодирующими областями других генов, не рассматривались. При данных условиях потенциальные сайты связывания RbsR обнаруживаются в каждом геноме перед примерно 10 генами.

4.2.2. Структура обобщенного RbsR-регулона

В результате поиска потенциальные сайты связывания RbsR были обнаружены перед опероном *rbsDACBKR* во всех геномах, где присутствует ген регуляторного белка (Приложение 7). Обнаружить консервативные потенциальные сайты перед какими-либо другими оперонами не удалось. Таким образом, в группах Enterobacteriales, Pasteurellales и Vibrionales RbsR является локальным регулятором одного оперона. Поскольку ген самого белка RbsR расположен в том же опероне, данный фактор транскрипции также осуществляет авторепрессию.

В перечисленных группах ген *rbsR* был обнаружен во всех исследованных геномах, за исключением *H. ducreyi* и *Yersinia* spp., что наводит на мысль об относительно недавней утрате регулятора. Поэтому с точки зрения эволюции представляло интерес исследовать судьбу генов *rbs*-оперона в этих геномах.

В геноме *H. ducreyi* не было найдено ортологов для ни одного из генов рибозного оперона, тогда как в *Y. pestis* и *Y. pseudotuberculosis* наблюдалась иная ситуация. В геномах *Yersinia* spp. были обнаружены ортологи для генов *rbsD* и *K*, лежащие в одном

потенциальном опероне, причем перед опероном был найден потенциальный сайт с весьма высоким весом (Рис. 4.4, Приложение 7). Подобные данные указывают на то, что утрата генов произошла относительно недавно. При этом были утеряны ген регулятора и все гены транспортера рибозы, кроме *rbsD*, но сохранился ген *rbsK*, кодирующий киназу.

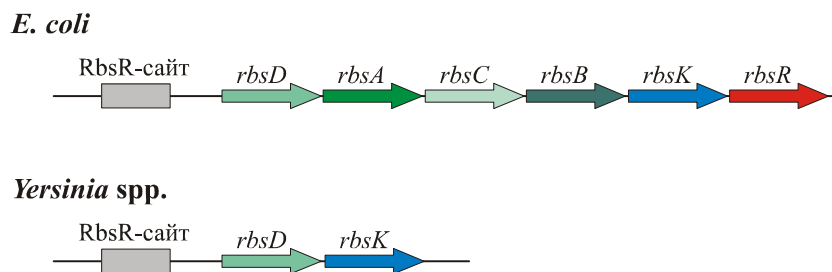


Рисунок 4.4. Сравнение структуры *rbs*-оперона в *E. coli* и *Yersinia spp.*

4.3. Исследование PurR-зависимой регуляции

4.3.1. Построение распознающего правила для поиска потенциальных сайтов связывания PurR

Ввиду эволюционной близости всех найденных ортологов PurR, было сочтено возможным исследование PurR-регуляции во всех геномах с использованием единой матрицы. В геноме *E. coli* ранее были экспериментально установлены сайты связывания пуринового репрессора перед 19 оперонами (Табл. 4.2). Данные сайты были использованы в качестве обучающей выборки для построения соответствующей матрицы (Рис. 4.5). Сайт связывания белка PurR представляет собой нестрогий палиндром длиной 16 нуклеотидов (Рис. 4.6), консенсус сайта при этом согласуется с ранее предсказанным [155].

Поз.	A	C	G	T	Конс.
1	0.31	-0.16	0.01	-0.16	A
2	0.01	0.12	0.01	-0.14	C
3	-0.18	-0.18	0.55	-0.18	G
4	0.16	0.31	-0.34	-0.13	M
5	0.49	-0.23	-0.02	-0.23	A
6	0.55	-0.18	-0.18	-0.18	A
7	0.27	-0.08	-0.18	-0.01	A
8	-0.18	0.55	-0.18	-0.18	C
9	-0.18	-0.18	0.55	-0.18	G
10	0.01	-0.37	0.09	0.27	T
11	-0.23	-0.23	-0.02	0.49	T
12	-0.23	-0.23	-0.02	0.49	T
13	-0.03	-0.14	0.07	0.10	K
14	-0.28	0.43	-0.07	-0.07	C
15	-0.13	0.02	0.17	-0.06	G
16	-0.05	0.11	-0.36	0.30	T

Рисунок 4.5. Позиционная матрица весов для сайтов связывания PurR

Указан вес каждого нуклеотида в соответствующей позиции. “Поз.” – позиция сайта; “Конс.” – нуклеотид, преобладающий в данной позиции.



Рисунок 4.6. Диаграмма Лого для сайтов связывания PurR

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

Таблица 4.2. Сайты связывания белка PurR *E. coli*, вошедшие в обучающую выборку

Указано положение сайта относительно старта трансляции гена, также приводятся ссылки на работы, в которых регуляция гена была показана экспериментально. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Оперон	Сайт			
	Положение	Вес	Последовательность	Ссылки
<i>purR</i>	-61	4,76	AgGCAAACGTTTaCcT	[156, 159]
	27	4,07	gaGCAAACGTTTcCac	
<i>prsA</i>	-357	4,72	AaGaAAACGTTTtCGc	[11]
<i>purC</i>	-168	4,67	ACGCAAACGTgTGCGT	[12, 159]
<i>purA</i>	-122	4,17	AgGaAAACGaTTGgcT	[157]
	-23	3,77	AaGCAAACGgTgattT	
<i>purT</i>	-54	5,08	ACGCAAACGTTTtCGT	[162]
<i>purL</i>	-91	4,94	ACGCAAACGgTTtCGT	[12, 159]
<i>purB</i>	185	4,45	ACGCAAtCGgTTaCcT	[158, 159]
<i>purEK</i>	-86	4,61	ACGCAAcCGTTTtCcT	[12]
<i>cvpA-purF</i>	-71	4,93	ACGCAAACGTTTtCtT	[160, 161]
<i>purMN</i>	-80	4,55	tCGCAAACGTTTGctT	[12, 159]
<i>purHD</i>	-122	4,76	gCGCAAACGTTTtCGT	[12]
<i>guaBA</i>	-68	4,21	AtGCAAtCGgTTaCGc	[159]
<i>pyrC</i>	-66	4,54	AgGaAAACGTTTcCGc	[163]
<i>pyrD</i>	-103	4,51	cgGaAAACGTTTGCGT	[163]
<i>codBA</i>	-237	3,65	gCGCAAgCGTTTtCca	[164]
<i>speAB</i>	-133	4,28	AaGaAAcCGgTTGCGc	[11]
<i>glnB</i>	-82	4,06	AtGCAAACGaTTtCaa	[11]
<i>gcvTHP</i>	-105	4,31	AaGagAACGaTTGCGT	[165]
<i>glyA</i>	-133	4,39	AgGtAAtCGTTTGCGT	[166]

Среди сайтов обучающей выборки наименьший вес, 3,65, имеет сайт, обнаруженный перед опероном *carAB*. Однако, в других геномах в связи с видоспецифичными особенностями ДНК-белковых взаимодействий возможно существование сайтов и с меньшим весом. Поэтому при поиске потенциальных сайтов с помощью имеющейся матрицы использовалось пороговое значение, равное 3,50.

Большинство экспериментальных сайтов располагается в пределах от –300 до +50 п.н. относительно старта трансляции. Сайт перед геном *prsA* находится в позиции –357 п.н. Отдельно следует рассматривать сайт, расположенный в кодирующей области гена *purB* в позиции +185 п.н. Случай регуляции данного гена представляется исключительным, поскольку ген *purB* является вторым в опероне *hflD-purB* и PurR-зависимая регуляция транскрипции осуществляется за счет препятствованию элонгации транскрипции ([158, 295]). Поскольку это единственный подобный случай среди генов PurR-регулона, было решено проводить поиск потенциальных сайтов связывания PurR в области –400..+100 п.н., осуществляя при этом дополнительный поиск сайтов для ортологов *purB* в области –100...+300 п.н. При данных условиях потенциальные сайты связывания белка PurR обнаруживаются приблизительно перед 200 генами на геном.

4.3.2. Структура обобщенного PurR-регулона

В соответствии с процедурой, изложенной в главе “Материалы и методы”, к обобщенному PurR-регулону было отнесено в общей сложности 62 гена, объединенных как минимум в 49 оперонов. Помимо экспериментально известных членов регулона и *E. coli* и генов с ранее предсказанной регуляцией [114], к обобщенному регулону были добавлены 26 новых оперонов, содержащих 30 генов (Табл. 4.3, Рис. 4.7) Последовательности всех сайтов связывания PurR приведены в Приложении 8. На основании функциональной принадлежности все гены, входящие в обобщенный регулон, были разбиты на десять групп.

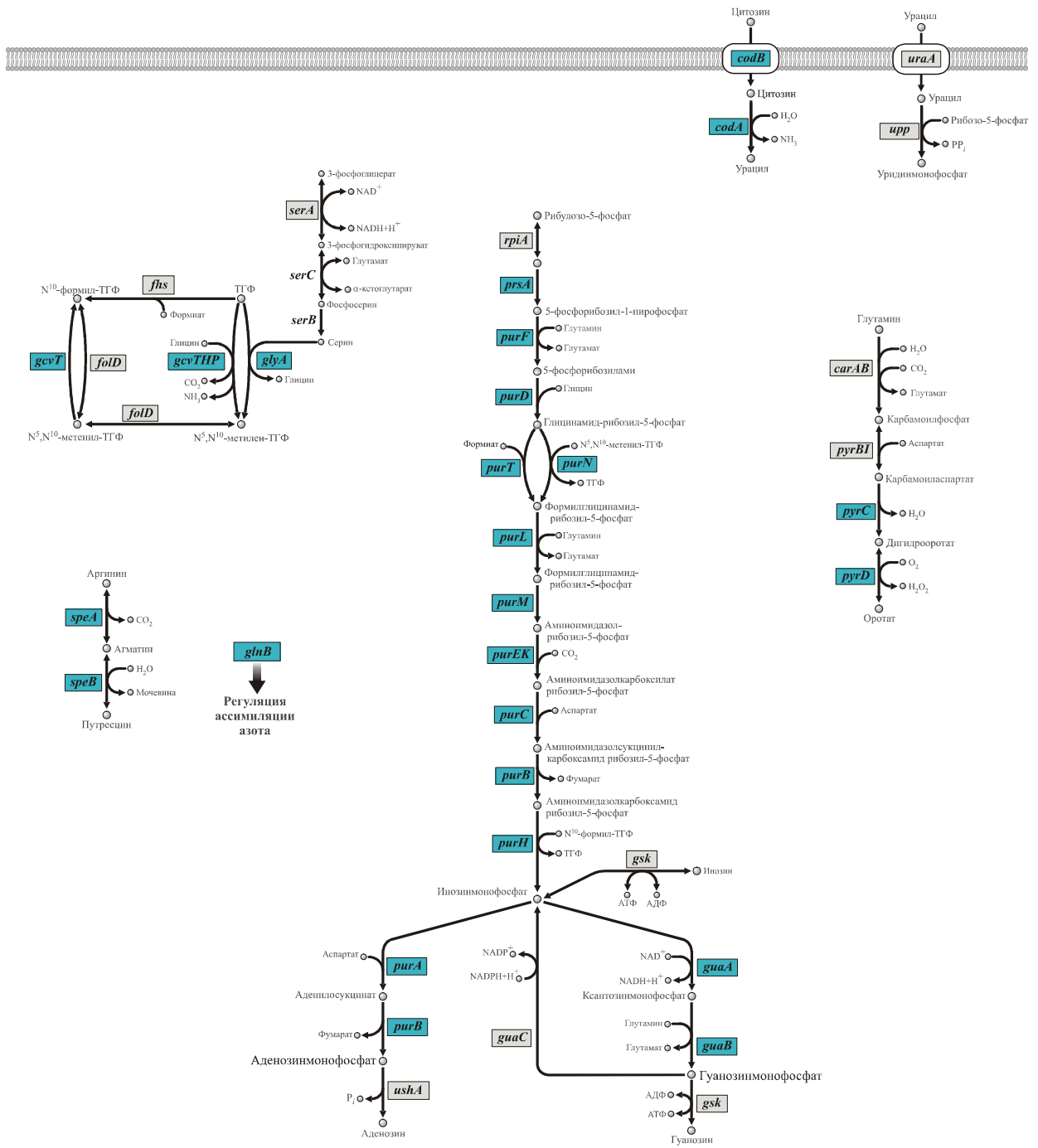
4.3.2.1. Синтез пуриновых нуклеотидов. Ранее PurR-зависимая регуляция была экспериментально обнаружена для оперонов *purR*, *guaBA*, *purB*, *purL* [159], *purHD*, *purC*, *purEK* [12], *prsA* [11], *purA* [157], *purT* [162], *cvpA-purF* [160].

Таблица 4.3. Состав обобщенного PurR-регулона.

Условные обозначения: “+” – перед опероном обнаружен сайт непосредственно перед геном; “-” – сайта перед опероном не обнаружено; “0” – в исследуемом геноме не найдено ортологов генов данного оперона; ¹⁾ регуляция оперона была ранее показана экспериментально; ²⁾ регуляция гена была ранее предсказана [114]. Обозначения геномов: см. Табл. 2.1.

Гены	Геномы														
	Enterobacteriales							Pasteurellales							
	ECO	STY	STM	YPK	YPS	PCA	PLU	PMU	HIN	HDU	VCH	VFI	VPA	VVU	PPR
Синтез пуриновых нуклеотидов															
<i>purR</i> ¹⁾	+	+	+	+	+	+	-	-	-	-	-	-	-	-	-
<i>prsA</i> ¹⁾	+	+	+	-	+	+	+	+	+	+	+	+	+	+	+
<i>purC</i> ¹⁾	+	+	+	-	-	-	-	+	+	+	+	+	+	+	+
<i>purA</i> ¹⁾	+	+	+	-	-	-	-	-	-	-	-	-	-	-	-
<i>purT</i> ¹⁾	+	+	+	+	+	+	0	0	0	+	+	+	+	+	+
<i>purL</i> ¹⁾	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>purB</i> ¹⁾	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>purEK</i> ¹⁾	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>svpA-purF</i> ¹⁾	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>purMN</i> ¹⁾	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>purHD</i> ¹⁾	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>guaBA</i> ¹⁾	+	+	+	+	+	+	+	-	-	-	+	+	+	-	+
<i>gsk</i>	-	-	-	-	-	0	-	0	0	0	+	+	+	+	+
<i>guaC</i>	-	-	-	-	-	-	-	0	0	0	+	+	+	+	+
<i>ushA</i>	-	-	-	-	-	-	-	-	0	0	+	+	-	+	-
Синтез пиримидиновых нуклеотидов															
<i>codBA</i> ¹⁾	+	+	+	0	0	0	0	+	0	0	0	-	-	-	-
<i>urp-uraA</i>	+	+	+	+	+	+	+	-	-	-	+	+	+	+	+
<i>purC</i> ¹⁾	+	+	+	-	-	-	-	0	0	0	-	+	+	+	+
<i>purD</i> ¹⁾	+	+	+	+	+	+	-	-	-	-	+	+	+	-	+
<i>carAB</i>	+	+	+	+	+	+	-	-	0	0	-	-	-	-	-
<i>purLBI</i>	+	-	-	+	+	+	-	0	0	0	+	+	+	-	+
Метаболизм азота															
<i>speAB</i> ¹⁾	+	+	+	-	-	-	+	+	0	0	-	+	+	-	0
<i>glnB</i> ¹⁾	+	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Метаболизм одноуглеродных фрагментов															
<i>gcvTHP</i> ¹⁾	+	+	+	+	+	+	+	+	0	0	-	-	-	-	-
<i>folD</i>	+	+	+	+	+	+	+	+	+	+	+	+	+	+	+
<i>gfyA</i> ¹⁾	+	+	+	+	+	+	+	+	+	0	+	+	+	+	+

Гены	Геномы															
	Enterobacteriales								Pasteurellales				Vibrionales			
	ECO	STY	STM	YPK	YPS	PCA	PLU	PMU	HIN	HDU	VCH	VFI	VPA	VVU	PPR	
Метаболизм одноуглеродных фрагментов (продолжение)																
<i>rpiA</i>	-	-	-	-	-	-	-	+	+	-	-	+	-	-	-	
<i>serA</i>	+	-	-	+	+	+	-	+	+	0	+	+	+	+	-	
<i>fhs</i>	0	0	0	0	0	0	0	0	0	+	+	+	+	-	+	
Трансмембранные белки																
<i>tsx²⁾</i>	+	+	+	0	0	+	0	0	0	0	0	0	0	0	0	
<i>glfS</i>	+	+	+	+	+	0	+	0	-	-	-	+	-	-	-	
<i>yicE²⁾</i>	+	+	+	+	+	+	+	0	0	0	+	+	+	+	+	
<i>yicG²⁾</i>	+	+	+	-	-	-	0	-	0	0	+	+	+	+	-	
<i>yicD²⁾</i>	+	+	+	+	+	+	+	+	+	-	0	+	+	0	+	
Утилизация нуклеотидов и нуклеиновых кислот																
<i>xseA</i>	+	+	+	+	+	+	+	-	+	-	+	+	-	-	+	
<i>cytR</i>	-	-	-	-	-	-	-	0	0	0	+	+	+	+	-	
Центральный метаболизм																
<i>pckA</i>	-	-	-	-	-	-	-	+	-	-	+	+	+	+	-	
<i>glpX</i>	-	-	-	-	-	-	-	+	0	0	+	+	+	-	+	
<i>ppsA</i>	-	-	-	-	-	-	-	0	0	0	+	+	+	+	+	
Белки с неизвестной функцией																
<i>ydiA</i>	-	-	-	-	-	-	-	0	0	0	+	+	+	+	+	
<i>VC2168</i>	0	0	0	0	0	0	0	0	0	0	+	+	+	+	-	
<i>yiiU</i>	-	-	-	0	0	-	-	+	-	-	0	+	+	-	-	
<i>yhhQ</i>	+	+	+	-	-	+	-	-	+	+	+	+	+	+	+	
<i>ydiJ</i>	+	+	+	-	-	+	+	-	0	-	-	-	-	-	-	
<i>HD1120</i>	0	0	0	0	0	0	0	+	+	+	+	0	0	-	0	



- purB* — регуляция была экспериментально показана ранее
- guaC* — регуляция была предсказана в данной работе

Рисунок 4.7. Новые члены обобщенного PurR-регулона гамма-протеобактерий
 Расшифровка цветowych обозначений дана в нижней части рисунка.

Авторегуляция гена *purR* ограничена представителями Enterobacteriales, потенциальные сайты были обнаружены у всех организмов из данного порядка, кроме *P. luminescens*, при этом у *E. coli* и *Salmonella* spp. были найдены парные сайты, что, скорее всего, необходимо для усиления авторепрессии.

Наибольшей консервативностью отличалась регуляция генов пути биосинтеза ИМФ из ФРПП (Рис. 4.7). Все гены этого пути, за незначительными исключениями, имели потенциальные сайты во всех исследованных геномах (Табл. 4.3). Регуляция же оперонов *purB*, *purA* и *guaBA*, ответственных за синтез АМФ и ГМФ из ИМФ, а также гена *prsA*, ответственного за синтез ФРПП из рибозо-5-фосфата, оказалась менее консервативной. Также необязательным оказалось наличие в геноме гена *purT*, поскольку его функция дублируется геном *purN* [296].

Кроме того, в данной работе была предсказана регуляция еще трех генов, продукты которых участвуют в модификации нуклеотидов, *guaC*, *gsk* и *ushA*. Продуктом гена *guaC* является ГМФ редуктаза, осуществляющий необратимое дезаминирование ГМФ, приводящее к формированию ИМФ [297]. Ген *gsk* кодирует инозин/гуанозин киназу, фермент, катализирующий фосфорилирование ИМФ и ГМФ до соответствующих нуклеозид-дифосфатов [298]. Регуляция данных двух генов была предсказана для всех изученных представителей Vibrionales. Ген *ushA* кодирует 5'-нуклеозидазу, фермент, отщепляющий фосфатную группу от нуклеозид-монофосфатов [299]. Сайты перед этим геном были обнаружены в геномах *V. cholerae*, *V. fischeri* и *V. vulnificus*.

4.3.2.2. Синтез пиримидиновых нуклеотидов. Регуляция пуриновым репрессором в *E. coli* была показана ранее для оперонов *purC*, *purD* [163] и *codBA* [164]. Регуляция оперона *codBA* сохраняется лишь в геномах *Salmonella* spp. и *P. multocida*. Несколько более консервативной является регуляция двух других оперонов. Так сайты перед *purC* были обнаружены в *Salmonella* spp., *V. vulnificus*, *V. parahaemolyticus*, *V. fischeri* и *P. profundum*, а сайты перед *purD* – в *Salmonella* spp., *Yersinia* spp., *P. carotovorum*, *V. parahaemolyticus*, *V. fischeri* и *P. profundum*.

В данной работе была также предсказана регуляция для еще трех оперонов, задействованных в биосинтезе пиримидиновых нуклеотидов, *urp-uraA*, *carAB* и *purLBI*.

Продуктами генов *urp* и *uraA* являются, соответственно, фермент урацил фосфорибозилтрансфераза и транспортер урацила [300, 301]. Данные белки осуществляют транспорт урацила из внешней среды с последующим его преобразованием в уридинмонофосфат (Рис. 4.7). Потенциальные сайты перед данным опероном были обнаружены во всех исследованных геномах Enterobacteriales и Vibrionales.

Оперон *carAB* кодирует субъединицы карбамоилсинтазы, тогда как *purLBI* содержит гены субъединиц аспартат карбамоилтрансферазы. Совместно с белками PurC и PurD комплексы CarAB и PurLBI осуществляют серию биохимических реакций, в результате которых из глутамата синтезируется оротат, предшественник пиримидиновых нуклеотидов [302]. Регуляция *carAB* предсказана для всех Enterobacteriales, за исключением *P. luminescens*. Что же касается оперона *purLBI*, то он регулируется в *E. coli*, *Yersinia* spp., *P. carotovorum*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri* и *P. profundum*.

Регуляция биосинтеза пиримидиновых нуклеотидов пуриин-зависимой системой, скорее всего, объясняется тем, что оба типа нуклеотидов входят в состав нуклеиновых кислот, и, соответственно, возникает необходимость в скореллированном биосинтезе пуринов и пиримидинов.

4.3.2.3. Метаболизм азота. К данной группе были отнесены два оперона, *speAB* и *glnB*. Для обоих оперонов регуляция белком PurR была показана экспериментально [11]. Сайты перед опероном *speAB* были найдены в *E. coli*, *Salmonella* spp., *P. luminescens*, *P. multocida* и *V. parahaemolyticus*. Для гена же *glnB* не было найдено ни одного нового сайта, и, возможно, регуляция этого гена является видоспецифичной особенностью *E. coli*.

4.3.2.4. Метаболизм одноуглеродных фрагментов. Ранее для *E. coli* была показана регуляция оперонов *gcvTHP* [165] и *glyA* [166]. Регуляция *gcvTHP* сохраняется во всех Enterobacteriales. Ортологи *glyA* были обнаружены во всех исследованных геномах, за исключением *H. ducreyi*, перед всеми ортологами были найдены потенциальные сайты PurR.

В настоящей работе была также предсказана регуляция оперонов *folD*, *rpiA*, *serA* и *fhs*.

Ген *folD* кодирует метинил-ТГФ циклогидролазу, фермент, участвующий во взаимопревращениях фолат-ассоциированных одноуглеродных соединений [303]. Потенциальные сайты перед этим геном были обнаружены во всех исследованных геномах за исключением *H. ducreyi*.

Продуктом гена *serA* является фермент 3-фосфоглицерат дегидрогеназа, участвующий в биосинтезе серина [304]. Пуриновую регуляцию данного гена можно объяснить тем, что серин является источником одноуглеродных фрагментов, ассоциированных с фолатами [305]. В геномах *E. coli*, *Yersinia* spp., *P. carotovorum*, *V. vulnificus*, *V. parahaemolyticus* и *V. cholerae* потенциальные сайты были обнаружены непосредственно перед геном *serA*. В геномах же *P. multocida*, *H. influenzae* и *V. fischeri* сайты были обнаружены перед потенциальным опероном *rpiA-serA*. Судя по тому, что регуляция гена *rpiA* наблюдается

лишь в тех случаях, где он лежит в одном опероне с *serA*, можно предположить, что именно регуляция *serA* является критичной.

Ген *fhs*, кодирующий формил-ТГФ синтазу, впервые был описан в бактерии *Streptococcus mutans* JH1005 [306]. Ортологи данного гена были найдены в геномах всех исследованных *Vibrionales*, а также *H. ducreyi*. Во всех случаях, за исключением *V. vulnificus*, перед генами *fhs* были найдены потенциальные сайты связывания PurR.

Метаболизм фолат-ассоциированных одноуглеродных фрагментов играет критическую роль в биосинтезе пуриновых нуклеотидов. Так, реакции, осуществляемые белками PurN и PurH требуют присутствия N⁵ N¹⁰-метенил-ТГФ и N¹⁰-формил-ТГФ, соответственно [155]. В *E. coli* биосинтез вышеупомянутых соединений из ТГФ осуществляется в несколько стадий (Рис.4.8а). На первой стадии происходит синтез N⁵ N¹⁰-метилен-ТГФ. Данная реакция катализируется либо белком GlyA, с использованием серина в качестве донора метильной группы, либо комплексом GcvTHP, при этом в качестве донора метильной группы выступает глицин. В дальнейшем, в ходе реакции, катализируемой FolD, N⁵ N¹⁰-метилен-ТГФ преобразуется в необходимый для биосинтеза пуринов N⁵ N¹⁰-метенил-ТГФ. Образование второго необходимого соединения, N¹⁰-формил-ТГФ, осуществляется из N⁵ N¹⁰-метенил-ТГФ двумя возможными вариантами: либо необратимо с помощью GcvT, либо обратимо с помощью FolD [305]. Исходя из данных сравнительной геномики, можно с уверенностью утверждать, что подобный метаболический путь присутствует у всех исследованных *Enterobacteriales*, причем все гены, необходимые для его осуществления, регулируются PurR. Однако, результаты данной работы указывают на то, что описанный метаболический путь претерпевает существенные изменения в других геномах.

Так, в *P. multocida* и *H. influenzae* не обнаружено ортологов *gcv*-генов. По всей видимости, в этих организмах биосинтез необходимых фолат-ассоциированных одноуглеродных соединений протекает за счет GlyA- и FolD-зависимых реакций. Гены данного пути также находятся под регуляцией PurR (Рис. 4.8б).

В *H. ducreyi* синтез N¹⁰-формил-ТГФ возможен напрямую из ТГФ за счет реакции, катализируемой белком Fhs. Поскольку в геноме этого организма отсутствуют гены *gcvTHP* и *glyA*, вполне резонно предположить, что синтез N⁵ N¹⁰-метенил-ТГФ осуществляется из N¹⁰-формил-ТГФ за счет обратимой FolD-зависимых реакции (Рис. 4.8в).

В геномах *Vibrionales* представлены ортологи для *gcvTHP*, *glyA*, *folD* и *fhs*, поэтому в данном случае возможны все вышеописанные пути биосинтеза необходимых соединений (Рис 4.8г).

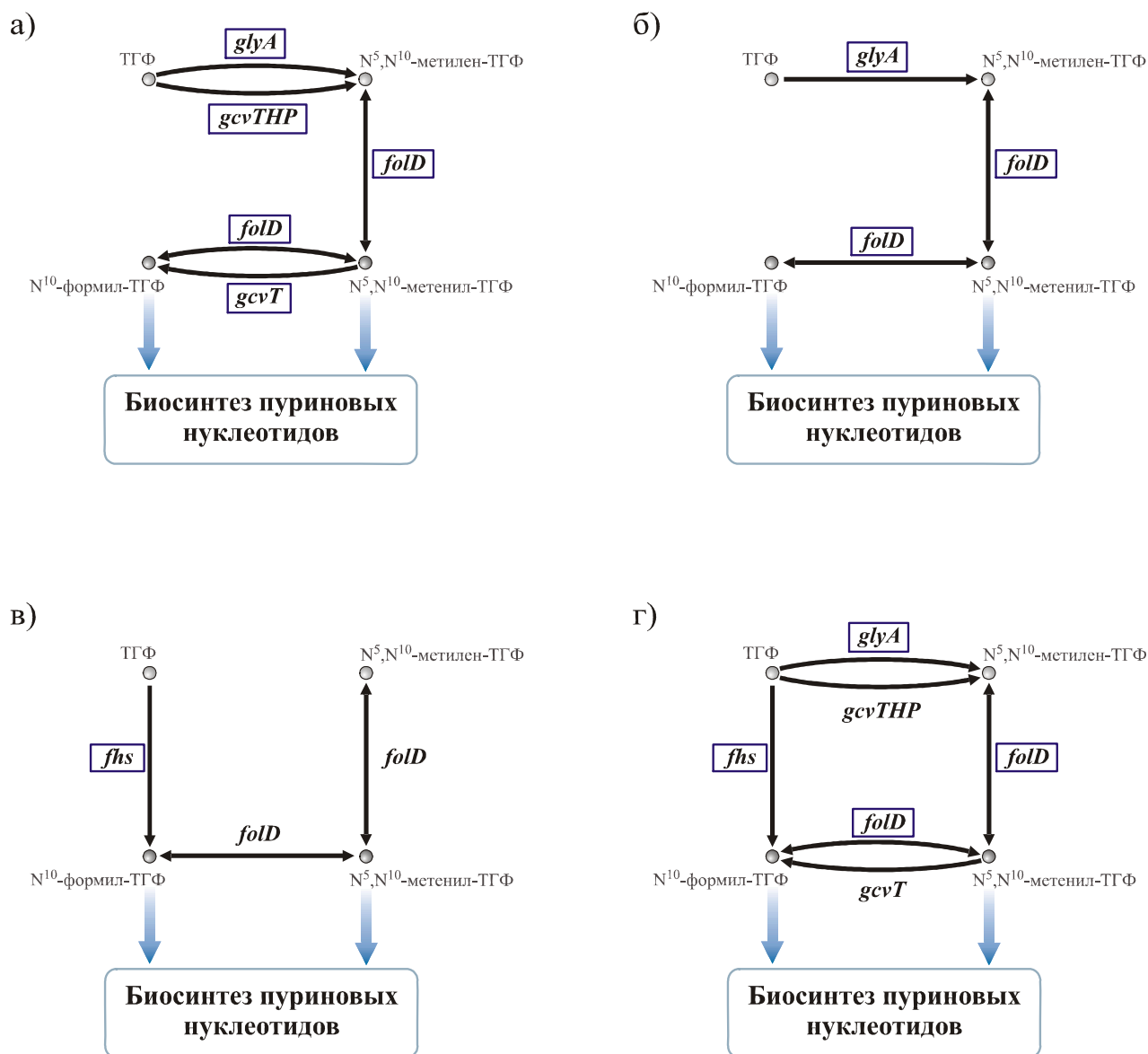


Рисунок 4.8. Метаболизм одноуглеродных соединений в различных таксонах
 Названия генов, регулируемых PurR, обведены. (а) Enterobacteriales, (б) *P. multocida* и *H. influenzae*, (в) *H. ducreyi*, (г) Vibrionales.

4.3.2.5. Транспортные белки. Ранее регуляция пуриновым репрессором была показана экспериментальными методами для гена *codB*, кодирующего транспортер цитозина ([307]). Кроме того, методами сравнительной геномики была предсказана регуляция генов *tsx*, *yicE*, *yieG* и *yjcD* [114]. В настоящей работе была также предсказана регуляция генов транспортных белков *uraA* и *gltS*.

Гены *codB* и *uraA* образуют опероны с генами биосинтеза пиримидиновых нуклеотидов, соответственно, *codBA* и *urp-uraA*. Регуляция данных оперонов подробно описана в разделе 4.3.2.2.

Ген *tsx* кодирует транспортер нуклеозидов и дезоксинуклеозидов [308]. Впервые его регуляция белком PurR была предсказана для *E. coli* [114]. В настоящей работе

потенциальные сайты перед *tsx* были найдены также в *S. typhi*, *S. typhimurium* и *P. carotovorum*, тогда как в остальных геномах ортологов данного гена не обнаружено.

Ген *yicE* кодирует транспортер ксантина [120]. Сайты перед этим геном были найдены во всех изученных геномах Enterobacteriales, Vibrionales, тогда как в Pasteurellales ортологи данного гена отсутствуют.

Продукт гена *yieG* был охарактеризован по аминокислотной последовательности как транспортер аденина [309]. Данный ген, по всей видимости, регулируется в *E. coli*, *Salmonella* spp. и *Vibrio* spp. В остальных геномах либо не обнаружено потенциальных сайтов, либо отсутствуют ортологи данного гена.

Также к регулону был ранее отнесен ген *yjcD*. По аминокислотной последовательности было предсказано, что его продукт, скорее всего, является транспортным белком, но специфичность так и не была установлена [310]. Потенциальные сайты связывания PurR перед *yjcD* найдены во всех Enterobacteriales, а также в *P. multocida*, *H. influenzae*, *V. fischeri*, *V. parahaemolyticus* и *P. profundum*.

В настоящей работе к пуриновому регулону был отнесен ген *gltS*, кодирующий транспортер глутамата [311]. Сайты перед этим геном присутствуют во всех Enterobacteriales, за исключением *P. carotovorum*, где не было найдено ортолога *gltS*. Потенциальный сайт был также найден в геноме *V. parahaemolyticus*.

Ранее было отмечено сходство между аминокислотными последовательностями белков UraA, YieG и YjcD [114]. В ходе настоящей работы было также обнаружено сходство с ними последовательностей белков GltS, CodB и YicE. Для всех упомянутых белков и их паралоги из гамма-протеобактерий было построено филогенетическое дерево (Рис. 4.9).

В соответствии с классификацией по базе данных Swiss-Prot ([77, 78, 312, 313]), все представленные белки распределены по следующим семействам:

- семейство GltS: все обнаруженные ортологи GltS, а также их паралоги из геномов Vibrionales;
- семейство NCS1: белки CodB;
- семейство NCS2: белки UraA и YicE;

семейство YicO/YieG/YjcD: белки YieG и YjcD.

Все семейства представлены на дереве в виде достаточно хорошо разрешенных веток. К сожалению, недостаток экспериментальных сведений не позволяет использовать данное дерево для более точного предсказания функций белков YieG и YjcD. Возможно, для этого необходим более точный анализ с привлечением данных о родственных белках из организмов других групп бактерий.

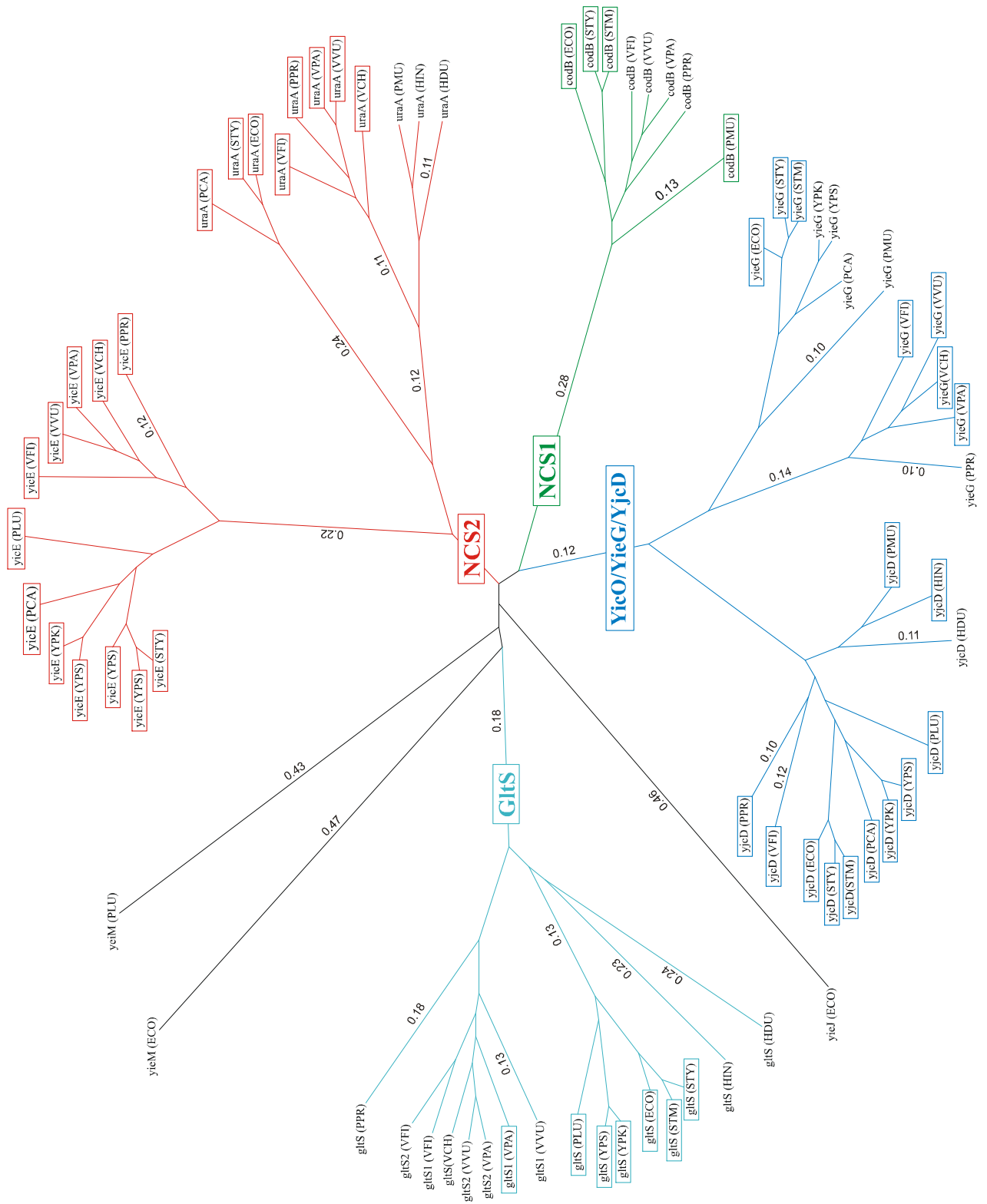


Рис. 4.9. Филогенетическое дерево транспортных белков, гены которых входят в PurR-регулон

Для построения использовался метод объединения соседних пар (neighbour-joining). Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,10). В скобках указаны условные обозначения геномов (см. Табл. 2.1). Регулируемые PurR гены обведены в рамку. На ветвях указаны названия семейств, к которым относятся белки.

4.3.2.6. Утилизация нуклеотидов и нуклеиновых кислот. В настоящей работе была предсказана PurR-зависимая регуляция двух генов, имеющих отношение к утилизации нуклеотидов, *xseA* и *cytR*.

Ген *xseA* кодирует большую субъединицу экзонуклеазы VII, расщепляющей одноцепочечную ДНК [314]. Потенциальные сайты перед геном *xseA* были обнаружены во всех исследованных Enterobacteriales, а также в *H. influenzae*, *V. cholerae* и *V. fischeri*.

Ген *cytR* кодирует регуляторный белок, контролирующей экспрессию генов, ответственных за взаимопревращения и утилизацию нуклеотидов и нуклеозидов [315]. Сайты перед *cytR* были обнаружены в геномах *Vibrio* spp.

4.3.2.7. Центральный метаболизм. В настоящей работе была также предсказана регуляция трех генов, *pckA*, *ppsA* и *glpX*, продукты которых являются ферментами гликолиза и глюконеогенеза [279, 316-318]. Потенциальные PurR-сайты перед геном *pckA* были найдены в *P. multocida* и *Vibrio* spp. Ген *glpX*, судя по всему, регулируется в геномах *P. multocida*, *H. influenzae*, *V. cholerae*, *V. fischeri*, *V. parahaemolyticus* и *P. profundum*. Регуляция же гена *ppsA* предсказана для *V. cholerae*, *V. parahaemolyticus*, *V. vulnificus* и *P. profundum*.

К сожалению, остается неясным, какое значение может иметь пурин-зависимая регуляция генов гликолиза и глюконеогенеза. Не стоит исключать и версию, что найденные сайты принадлежат не PurR, а какому-либо другому родственному белку, контролирующему данные процессы.

4.3.2.8. Белки с неизвестной функцией. Помимо описанных выше генов, в данной работе к пуриновому регулону был также отнесен ряд генов, функция которых в настоящее время неизвестна. Таковыми являются гены *ydiA*, *yiiU*, *yhhQ*, *ydiJ*, *HD1120* и *VC2168* (Табл. 4.3). К сожалению, объяснение принадлежности этих генов к пуриновому регулону станет возможным лишь после того, как их функции будут установлены экспериментально или предсказаны методами биоинформатики.

4.3.3. Таксон-специфические особенности PurR-регуляции

В ходе исследования эволюции пуриновой регуляции гамма-протеобактерий было замечено, что состав регулона достаточно сильно различается в разных видах бактерий. Все входящие в регулон гены можно условно разделить на три группы (Рис. 4.10):

- 1) ядро регулона – гены, регулирующие во всех геномах, или в их подавляющем большинстве;

- 2) гены, регуляция которых слабо консервативна, но при этом встречается в различных таксонах;
- 3) гены с таксон-специфической регуляцией.

К первой группе можно отнести большую часть генов биосинтеза ИМФ из рибозо-5-фосфата. Так, опероны *purT*, *purL*, *purMN*, *purHD* регулируются во всех изученных геномах, тогда как опероны *prsA*, *purC* и *purB* – в подавляющем большинстве геномов.

К этой же группе можно отнести и гены метаболизма одноуглеродных фрагментов, ассоциированных с фолатами. Несмотря на то, что пути биосинтеза фолатов достаточно сильно различаются в разных таксономических группах, в целом путь биосинтеза N⁵ N¹⁰-метенил-ТГФ и N¹⁰-формил-ТГФ из полностью регулируется PurR во всех исследованных геномах за исключением *H. ducreyi* (см. 4.3.2.4). Таким образом, к консервативным членам PurR-регулона можно отнести такие опероны, как *gcvTHP*, *folD*, *glyA* и *fhs*.

Что касается второй группы, то в нее входят гены, не участвующие напрямую в биосинтезе нуклеотидов. К данной группе были отнесены гены азотного метаболизма (*speAB*), метаболизма одноуглеродных фрагментов (*serA*), трансмембранных белков (*gltS*, *yicE* и *yjcD*), утилизации нуклеотидов (*xseA*), а так же ген с неизвестной функцией *yhhQ*.

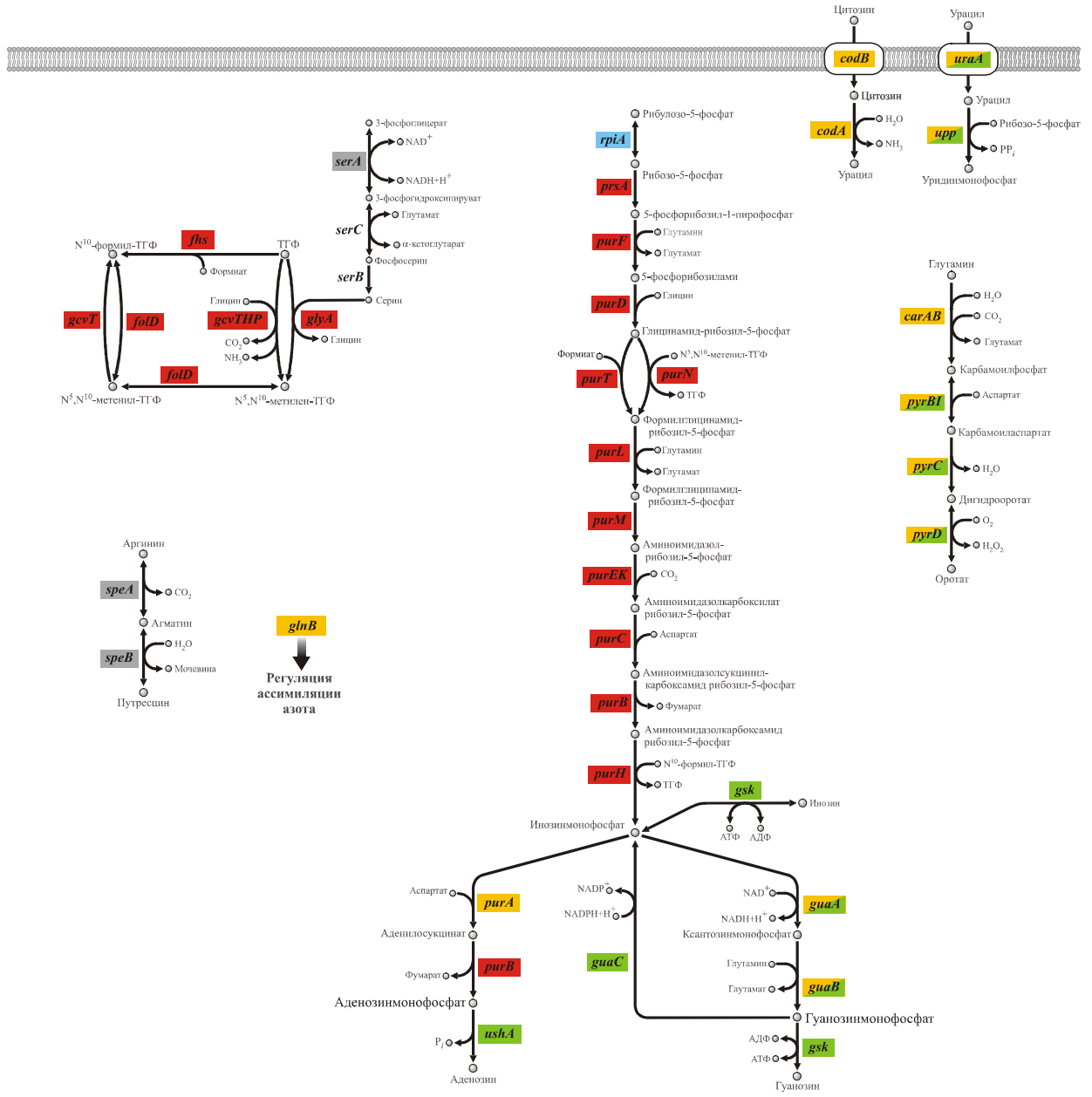
В третью группа генов, регуляция которых специфична для одного или двух таксонов, входят члены различных функциональных групп. Так, для Enterobacteriales характерна регуляция собственно гена белка регулятора PurR, генов, участвующих в модификации нуклеотидов (*purA*, *guaBA*), генов биосинтеза пиримидиновых нуклеотидов (*codBA*, *upp-uraA*, *pyrC*, *pyrD*, *carAB*, *pyrLBI*), гены трансмембранных белков (*tsx*, *yieG*) и ген с неизвестной функцией *ydiJ*.

Что касается группы Pasteurellales, то для нее характерна регуляция гораздо меньшего числа генов, таких как *rpiA*, ответственного за синтез рибозо-5-фосфата, генов центрального метаболизма *pckA* и *glpX*, а так же генов *yiiU* *ydiJ* и *HD1120*, функция которых неизвестна.

Картина, наблюдающаяся в Vibrionales, похожа на таковую для бактерий из группы Enterobacteriales с некоторыми отличиями. Так, для Vibrionales была установлена регуляция генов модификации пуриновых нуклеотидов (*guaBA*, *guaC*, *gsk*, *ushA*), биосинтеза пиримидиновых нуклеотидов (*upp-uraA*, *pyrC*, *pyrD*, *pyrLBI*), трансмембранных белков (*yieG*), центрального метаболизма (*pckA*, *glpX*, *ppsA*) и ряда генов с неизвестными функциями (*ydiA*, *VC2168*, *yiiU*).

Отдельно следовало бы отметить ген *glnB*, регуляция которого является видоспецифичной: из всех исследованных геномов сайт связывания был обнаружен только в *E. coli*. В принципе, подобные случаи видоспецифичной регуляции не могут быть выявлены с помощью применяемой методики и этот ген был отнесен к регулону на основании более

ранних экспериментальных данных [11]. Не исключено, что к обобщенному регулону могут принадлежать и другие гены с видоспецифичной регуляцией, но их выявление используемыми в настоящей работе методами невозможно.



- purB — ядро регулона
- speA — неконсервативная регуляция

Таксон-специфичная регуляция :

- purA — Enterobacteriales
- rpiA — Pasteurellales
- guaC — Vibrionales

Рисунок 4.10. Таксон-специфические особенности PurR-регуляции
 Расшифровка цветowych обозначений дана в нижней части рисунка.

В целом охарактеризовать консервативность PurR-зависимой регуляции можно следующим образом: наиболее консервативной является регуляция генов, участвующих в биосинтезе ИМФ из рибозо-5-фосфата, а так же генов, ответственных за синтез необходимых для этого производных фолатов. Регуляция генов, необходимых для модификаций ИМФ и синтеза пиримидиновых нуклеотидов, так же как и генов центрального метаболизма консервативна на уровне определенных таксонов. Для остальных генов регуляция хоть и встречается в различных таксонах, но консервативной не является.

4.4. Исследование регуляции в Pseudomonadales

4.4.1. Ген регуляторного белка из Pseudomonadales

Как было упомянуто в разделе 4.1, бактерии порядка Pseudomonadales содержат белок, примерно одинаково сходный с белками PurR и RbsR из других порядком гамма-протеобактерий. В ходе анализа локализации данного гена на хромосоме было обнаружено, что он расположен совместно с *rbs*-генами, и, по всей видимости, составляет с ними один оперон. В геномах *P. putida*, *P. fluorescens* и *P. syringae* в оперон входят гены *rbsBACRKD* и ген, по всей видимости, кодирующий нуклеозид-дегидрогеназу. Вывод о таковой функции продукта этого гена был сделан ранее на основе анализа аминокислотной последовательности кодируемого белка [53]. Геном же *P. aeruginosa* содержит укороченный вариант оперона, включающий лишь гены *rbsBACRK* (Рис. 4.11). При сравнении структуры *rbs*-оперона в Pseudomonadales и бактериях из других таксонов было замечено, что в эволюции сохраняется состав оперона, хотя и меняется порядок генов в нем. На основании кластеризации белка-регулятора из Pseudomonadales с генами *rbs*-оперона было выдвинуто предположение, что данный белок выполняет функцию RbsR.

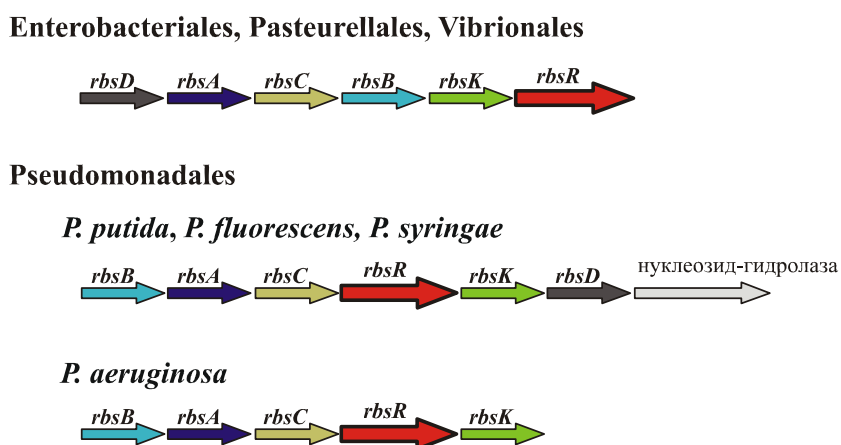


Рисунок 4.11. Сравнение структуры *rbs*-оперона в Pseudomonadales и других порядках

4.4.2. Построение распознающего правила для поиска потенциальных сайтов RbsR Pseudomonadales

В отличие от всех ранее описанных регуляторных белков, для белка RbsR из Pseudomonadales отсутствуют экспериментальные данные о сайтах связывания. По аналогии с другими порядками гамма-протеобактерий, было выдвинуто предположение, что в Pseudomonadales белок RbsR также должен регулировать экспрессию *rbs*-оперона. Исходя из предположения, были исследованы промотерные области последнего и с помощью программы программы SignalX были обнаружены высоко консервативные участки длиной 14 п.н. (Табл. 4). На основании полученных сайтов была построена матрица позиционных весов (Рис. 4.12). Как и в случаях PurR и RbsR других гаммапротеобактерий, сайт имеет структуру палиндрома (Рис. 4.13).

Таблица 4.4. Сайты связывания белка RbsR Pseudomonadales, вошедшие в обучающую выборку

Указано положение сайта относительно старта трансляции гена. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Ген	Сайт		
		Положение	Вес	Последовательность
<i>P. aeruginosa</i>	<i>rbsB</i>	-74	6,45	CGCAAACGTTTGCG
<i>P. putida</i>	<i>rbsB</i>	-51	6,21	CGCAAACGTTTGCT
<i>P. fluorescens</i>	<i>rbsB</i>	-108	6,45	CGCAAACGTTTGCG
<i>P. syringae</i>	<i>rbsB</i>	-82	6,45	CGCAAACGTTTGCG

Наименьший вес сайта из обучающей выборки был равен 6,39. Однако, учитывая возможность изменений в последовательностях сайтов перед другими генами, для поиска было принято пороговое значение 5,50. Поиск сайтов проводился в области от -400 до +100 п.н. относительно старта трансляции гена. Сайты, перекрывающиеся с кодирующими областями других генов, не рассматривались. При данных условиях потенциальные сайты связывания обнаруживаются перед примерно 10 генами в каждом геноме.

4.4.3. Структура обобщенного регулона в Pseudomonadales

В результате поиска консервативные потенциальные сайты связывания RbsR Pseudomonadales были обнаружены во всех исследованных геномах из данного порядка только перед собственно *rbs*-опероном. Ни перед одним из других генов не удалось выявить значимых консервативных сайтов. Таким образом, можно утверждать, что в Pseudomonadales RbsR представляет собой регулятор одного оперона.

Поз.	A	C	G	T	Конс.
1	-0.16	0.47	-0.16	-0.16	C
2	-0.16	-0.16	0.47	-0.16	G
3	-0.16	0.47	-0.16	-0.16	C
4	0.47	-0.16	-0.16	-0.16	A
5	0.47	-0.16	-0.16	-0.16	A
6	0.47	-0.16	-0.16	-0.16	A
7	-0.16	0.47	-0.16	-0.16	C
8	-0.16	-0.16	0.47	-0.16	G
9	-0.16	-0.16	-0.16	0.47	T
10	-0.16	-0.16	-0.16	0.47	T
11	-0.16	-0.16	-0.16	0.47	T
12	-0.16	-0.16	0.47	-0.16	G
13	-0.16	0.47	-0.16	-0.16	C
14	-0.22	-0.22	0.34	0.10	G

Рисунок 4.12. Позиционная матрица весов для сайтов связывания RbsR *Pseudomonadales*

Указан вес каждого нуклеотида в соответствующей позиции. “Поз.” – позиция сайта; “Конс.” – нуклеотид, преобладающий в данной позиции.



Рисунок 4.13. Диаграмма Лого для сайтов связывания RbsR *Pseudomonadales*

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

4.5. Эволюция PurR- и RbsR-регулонов

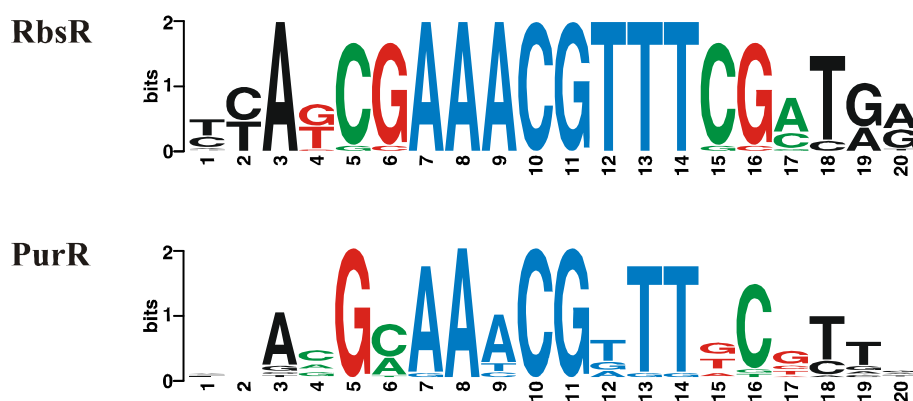
Как было сказано в разделе 4.1, белок RbsR из *Pseudomonadales*, по всей видимости, произошел от предкового белка, имевшегося у общего предка гамма-протеобактерий. В дальнейшем ген этого белка претерпел дупликацию, в результате которой образовались две копии, впоследствии ставшие PurR и RbsR.

Для того, чтобы более детально проследить эволюцию этих регуляторных систем, было проведено сравнение сайтов связывания трех регуляторных белков (Рис. 4.14). Сайты всех исследуемых белков имеют палиндромную структуру, но отличаются по длине: сайт связывания белка RbsR имеет длину 20 п.н., PurR – 16 п.н., а RbsR *Pseudomonadales* – 14 п.н. Таким образом, фланкирующие области сайтов неконсервативны и поэтому не подлежат сравнению. Центральная часть сайтов длиной 8 п.н. весьма консервативна и имеет консенсус AAACGTTT. Наиболее значительные различия наблюдаются как раз в тройках нуклеотидов,

ближайших к центральной консервативной части. Именно в этой области и наблюдаются основные различия между сайтами PurR и RbsR. В то же время, последовательность сайта RbsR Pseudomonadales более похожа на таковую для PurR, чем для RbsR. По всей видимости, сайт связывания PurR сохранил исходную структуру, имевшую место в предковом геноме.

На основании данных о филогении регуляторных белков, структуре сайтов и составе обобщенных регулонов была выдвинута следующая модель эволюции регуляторных систем PurR и RbsR (Рис. 4.15). Судя по всему, первоначально белок RbsR существовал как локальный регулятор, контролирующей экспрессию одного *rbs*-оперона. Однако затем, после отделения предковых форм Pseudomonadales, произошла дупликация гена белка-регулятора. Далее в каждой ветви происходили следующие события. В одной ветви белок сохранила свою функцию локального регулятора, рибозного репрессора, но при этом изменилась структура сайта связывания. Так образовался регулятор RbsR Enterobacteriales, Pasteurellales и Vibrionales. В другой же ветви структура сайта в целом сохранилась, однако, изменилась функция белка. К сожалению, в отличие от FruR, для PurR не наблюдается постепенного расширения регулона. Не исключено, что подобное расширение происходило довольно быстро в эволюционном масштабе. Остается надеяться, что появление новых геномных последовательностей позволит более детально проследить эволюцию двух гомологичных регуляторных систем, PurR и RbsR.

Enterobacteriales, Pasteurellales, Vibrionales



Pseudomonadales



Рисунок 4.14. Сравнение сайтов связывания RbsR, PurR и RbsR Pseudomonadales

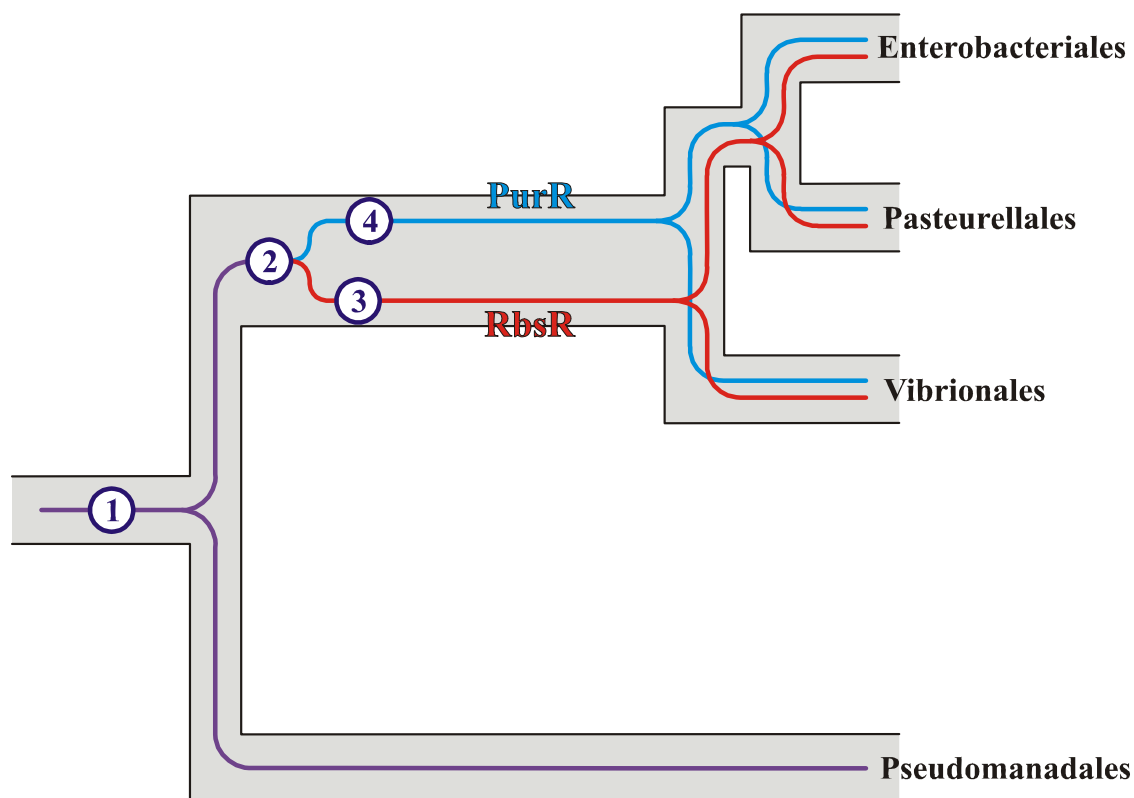


Рисунок 4.15. Предполагаемый сценарий эволюции регулонов RbsR, PurR и RbsR Pseudomonadales

Цифрами отмечены предполагаемые эволюционные события: 1 – RbsR – локальный регулятор *rbs*-оперона; 2 – удвоение гена белка-регулятора; 3 – изменение структуры сайта связывания RbsR с сохранением функции; 4 – изменение специфичности связывания с лигандом PurR при сохранении сайта, возникновение PurR-регулона.

4.6. Обсуждение и выводы

Пуриновый репрессор *E. coli*, регуляторный белок PurR, был впервые описан более двадцати пяти лет назад [319] и к настоящему моменту представляет собой один из наиболее изученных факторов транскрипции. К примеру, в настоящий момент времени насчитывается семь публикаций, посвященных рентгеноструктурному анализу данного белка [151, 152, 320-325]. Тем не менее, все экспериментальные исследования PurR-зависимой регуляции проводились только для *E. coli* и близкородственного вида *S. typhimurium* [326, 327].

Что же касается белка RbsR, то его изучение практически прекратилось после открытия его функции и исследования структуры [148, 328]. Также, как и для PurR, изучение RbsR-зависимой регуляции методами биоинформатики ранее проводилось лишь для небольшого числа представителей гамма-протеобактерий [19]. Белок RbsR Pseudomonadales ранее не изучался экспериментально. Однако на основе исследований методами биоинформатики было виднито предположение о том, что этот белок является предком белков PurR и RbsR

из Enterobacteriales, Pasteurellales и Vibrionales [294], а также предсказан потенциальный сайт связывания этого белка перед *rbs*-опероном в геноме *P. aeruginosa* [19].

В настоящей работе были изучены сразу три регуляторных системы, RbsR, PurR и RbsR Pseudomonadales. Белки RbsR и RbsR Pseudomonadales представляют собой локальные регуляторы, контролирующие экспрессию одного лишь *rbs*-оперона, и ни одного нового потенциального члена этих регулонов так и не было обнаружено. Однако, несмотря на то, что данные белки выполняют одну и ту же функцию, сайты их связывания различаются по структуре. Кроме того, различается также структура регулируемых ими оперонов.

Что же касается белка PurR, то в настоящей работе был выявлен целый ряд новых членов соответствующего регулона, а также обнаружены некоторые таксон-специфические особенности регуляции. Так, было показано, что в регулон входят гены модификации пуриновых нуклеотидов (*gsk*, *guaC*, *ushA*), биосинтеза пиримидиновых нуклеотидов (*upp-uraA*, *carAB*, *pyrLBI*), метаболизма одноуглеродных соединений (*fold*, *rpiA*, *serA*, *fhs*), трансмембранных белков (*gltS*), утилизации нуклеотидов и нуклеиновых кислот (*xseA*, *cytR*), центрального метаболизма (*pckA*, *glpX*, *ppsA*) и белков с неизвестной функцией (*ydiA*, *VC2168*, *yiiU*, *yhhQ*, *ydiJ*, *HD1120*).

Регуляция как ранее известных членов регулона, так и предсказанных в данной работе, может достаточно сильно различаться по степени консервативности в эволюции. Как правило, эти различия коррелируют с функциями генов. Наиболее консервативной является регуляция генов, ответственных за синтез ИМФ из 5-рибозо-фосфата – эти гены регулируются практически во всех изученных геномах. Такая же консервативная регуляция характерна и для генов метаболизма фолат-ассоциированных одноуглеродных фрагментов, а именно – для генов, отвечающих за синтез N⁵ N¹⁰-метинил-ТГФ и N¹⁰-формил-ТГФ, веществ, необходимых для синтеза пуриновых нуклеотидов. Синтез упомянутых соединений может осуществляться различными путями, в том числе возможно и осуществление одной и той же реакции несколькими разными ферментами. Несмотря на то, что метаболические пути достаточно сильно различаются в разных порядках гамма-протеобактерий, было отмечено, что всегда регулируются все гены, позволяющие осуществить путь синтеза N⁵ N¹⁰-метинил-ТГФ и N¹⁰-формил-ТГФ из предшественника, ТГФ. Единственным исключением, где гены данного пути регулируются не полностью, является *H. ducreyi*. Для генов других функциональных групп, таких как гены модификации пуриновых нуклеотидов, синтеза пиримидиновых нуклеотидов, метаболизма азота, центрального метаболизма и генов некоторых транспортных белков, регуляция консервативна лишь в пределах одного или двух порядков.

Еще одним из значимых результатов данной работы является модель эволюции регуляторных систем PurR и RbsR. По всей видимости, изначально существовал единичный белок, выполнявший функцию репрессора рибозного оперона. Вероятно, наблюдаемая в Pseudomonadales картина отражает ситуацию в предковом геноме. Однако, после отделения Pseudomonadales от общего ствола гамма-протеобактерий, произошла дупликация гена белка-регулятора. Одна из копий сохранила свою функцию, претерпев при этом изменения в структуре сайта связывания, и стала RbsR, каковой мы можем наблюдать, например, в *E. coli*. Функция же другой копии гена изменилась, хотя сохранилась структура сайта связывания регуляторного белка. При этом произошло образование нового регулона.

Глава 5

Эволюция глобальной регуляции дыхания

Как было сказано ранее, предварительные исследования Fnr- и ArcA-регулонов ([329, 330]) показали, что глобальная регуляция дыхания является слабо консервативной. По всей видимости, это связано с наличием множественной регуляции у генов, задействованных в дыхании и смежных процессах. В связи с этим было принято решение о комплексном изучении глобальной регуляции дыхания, то есть, об одновременном исследовании Fnr-, ArcA- и Nag-зависимой регуляции.

5.1. Эволюция регуляторных систем

5.1.1. Однокомпонентная регуляторная система Fnr

Ортологи белка Fnr из *E. coli* были найдены в 12 организмах, относящихся к трем порядкам: Enterobacteriales (*S. typhi*, *Y. pestis*, *Y. enterocolitica*, *P. carotovorum*), Pasteurellales (*P. multocida*, *A. actinomycetemcomitans*, *H. influenzae*, *H. ducreyi*), и Vibrionales (*V. vulnificus*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*). Все обнаруженные ортологи имеют потенциальные НТН-мотивы. Анализ консервативности цистеинового кластера показал, что критически важные цистеиновые остатки Cys16, Cys20, Cys23 и Cys29 [331, 332] сохраняются во всех найденных ортологах, за исключением белка из *Y. pestis*, где Cys16 заменен на тирозин (Приложение 9).

Для всех найденных ортологов было построено филогенетическое дерево (Рис. 5.1). Анализ дерева показывает, что белки из разных организмов группируются в соответствии с таксономической принадлежностью последних. Однако, белки из Pasteurellales формируют ветвь, достаточно далеко отстоящую от ветвей для белков из других таксонов. Таким образом, можно ожидать в Pasteurellales достаточно сильных изменений Fnr-зависимой регуляции по сравнению с другими таксонами. В то же время, эволюционные расстояния между всеми ортологами невелики, что позволяет использовать для поиска потенциальных сайтов единое распознающее правило.

5.1.2. Двухкомпонентная система ArcB-ArcA

Ортологи фактора транскрипции ArcA *E. coli* были обнаружены в 12 организмах, относящихся к трем порядкам: Enterobacteriales (*S. typhi*, *Y. pestis*, *Y. enterocolitica*, *P. carotovorum*), Pasteurellales (*P. multocida*, *A. actinomycetemcomitans*, *H. influenzae*, *H. ducreyi*), и Vibrionales (*V. vulnificus*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*). Предполагаемые НТН-

мотивы содержатся во всех найденных (Приложение 10). Незначительные эволюционные расстояния между орологами ArcA (Рис. 5.2) указывают на возможность использования единого распознающего правила для поиска потенциальных сайтов во всех исследованных геномах.

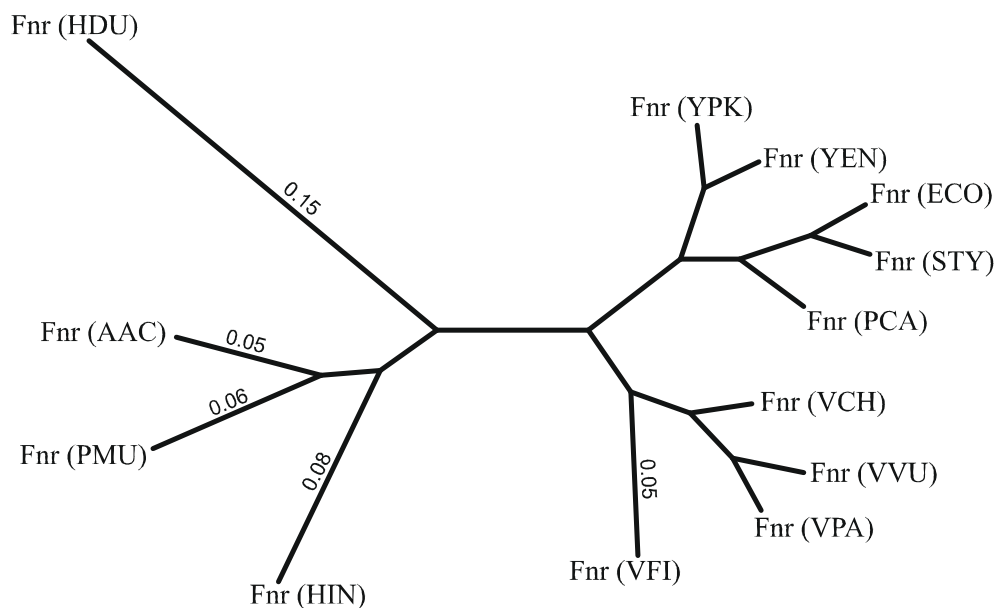


Рис. 5.1. Филогенетическое дерево белков Fnr

Для построения использовался метод объединения соседних пар (neighbour-joining). Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,05). В скобках указаны условные обозначения геномов (см. Табл. 2.1).

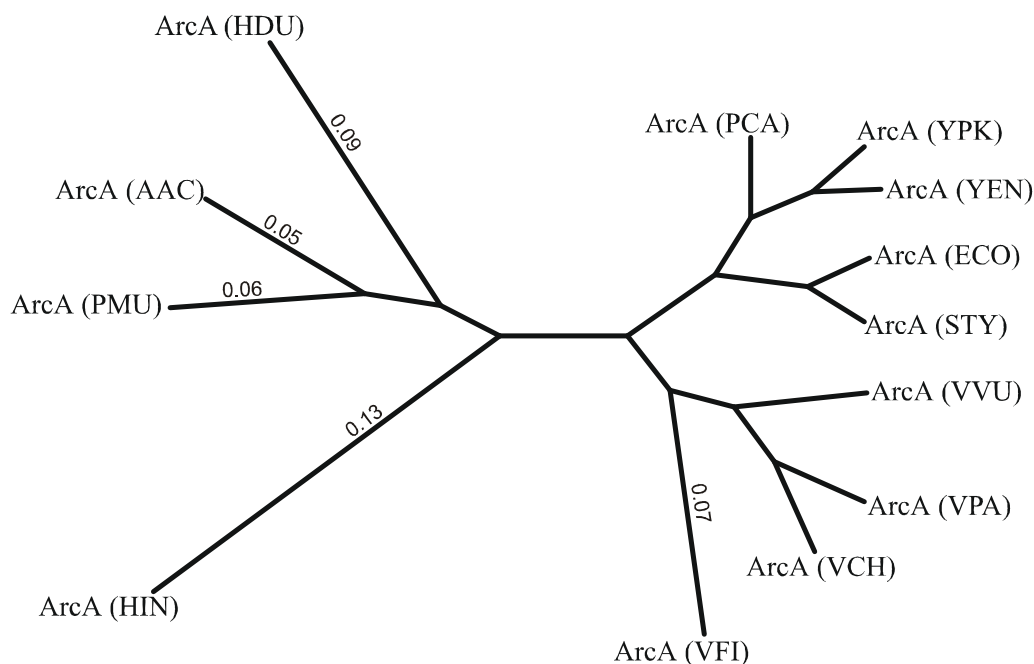


Рис. 5.2. Филогенетическое дерево белков ArcA

Для построения использовался метод объединения соседних пар (neighbour-joining). Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,05). В скобках указаны условные обозначения геномов (см. Табл. 2.1).

Ортологи же сенсорного белка ArcB были обнаружены во всех перечисленных выше геномах, кроме *H. Influenzae* и *H. ducreyi*. Несмотря на то, что в других представителях Pasteurellales, *P. multocida* и *A. actinomycetemcomitans*, ортологи ArcB присутствуют, на филогенетическом дереве данные белки отстоят достаточно далеко от ортологов из других порядков (Рис. 5.3).

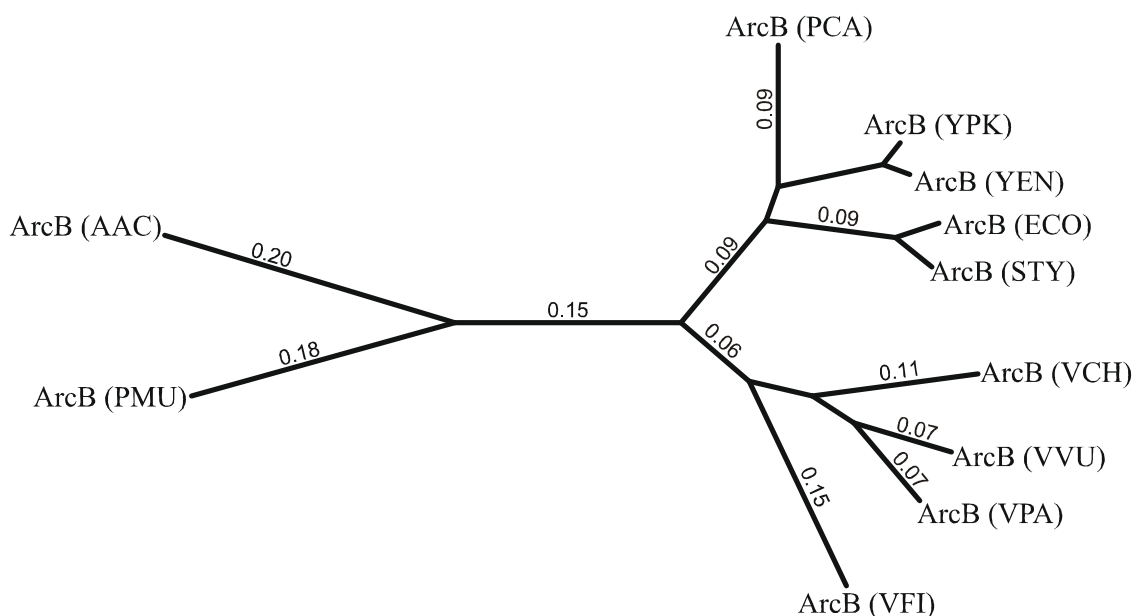


Рис. 5.3. Филогенетическое дерево белков ArcB

Для построения использовался метод объединения соседних пар (neighbour-joining). Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,05). В скобках указаны условные обозначения геномов (см. Табл. 2.1).

Для ортологов ArcB была обнаружена еще одна особенность. Известно, что ключевую роль в димеризации и последующей активации ArcB *E. coli* играют цистеиновые остатки Cys180 и Cys241, расположенные в трансмембранном PAS-домене [206]. Оба цистеиновых остатка консервативны лишь в белках Enterobacteriales и *V. fischeri*, тогда как в других Vibrionales консервативен лишь Cys180. У представителей же Pasteurellales полностью отсутствует PAS-домен. В случае представителей Vibrionales ArcB, скорее всего, функционален, поскольку было показано, что основную роль в димеризации играет Cys180, а замена Cys241 является менее критичной [206]. Что же касается Pasteurellales, то не исключено, что у этих бактерий активация ArcA осуществляется другим сенсорным белком. Подобная ситуация была описана для другого представителя гамма-протеобактерий, *Shewanella oneidensis* MR-1 [333].

Тем не менее, наличие ортологов регуляторного белка ArcA во всех вышеупомянутых организмах позволяет провести исследование соответствующей регуляции.

5.1.3. Регуляция нитрат-нитритного дыхания: удвоенная двукомпонентная система NarX-NarL и NarQ-NarP

Поиск ортологов регуляторных белков показал, что оба регулятора, NarL и NarP, присутствуют лишь в *E. coli*, *S. typhi* и *P. carotovorum*, тогда как в геномах *Yersinia* spp., Pasteurellales и Vibrionales был обнаружен только одиночный регулятор NarP (Приложение 12). Анализ филогенетического дерева для регуляторных белков (Рис. 5.4) показывает, что белки из разных организмов образуют два отдельных кластера, что указывает на удвоение системы до расхождения исследуемых групп организмов.

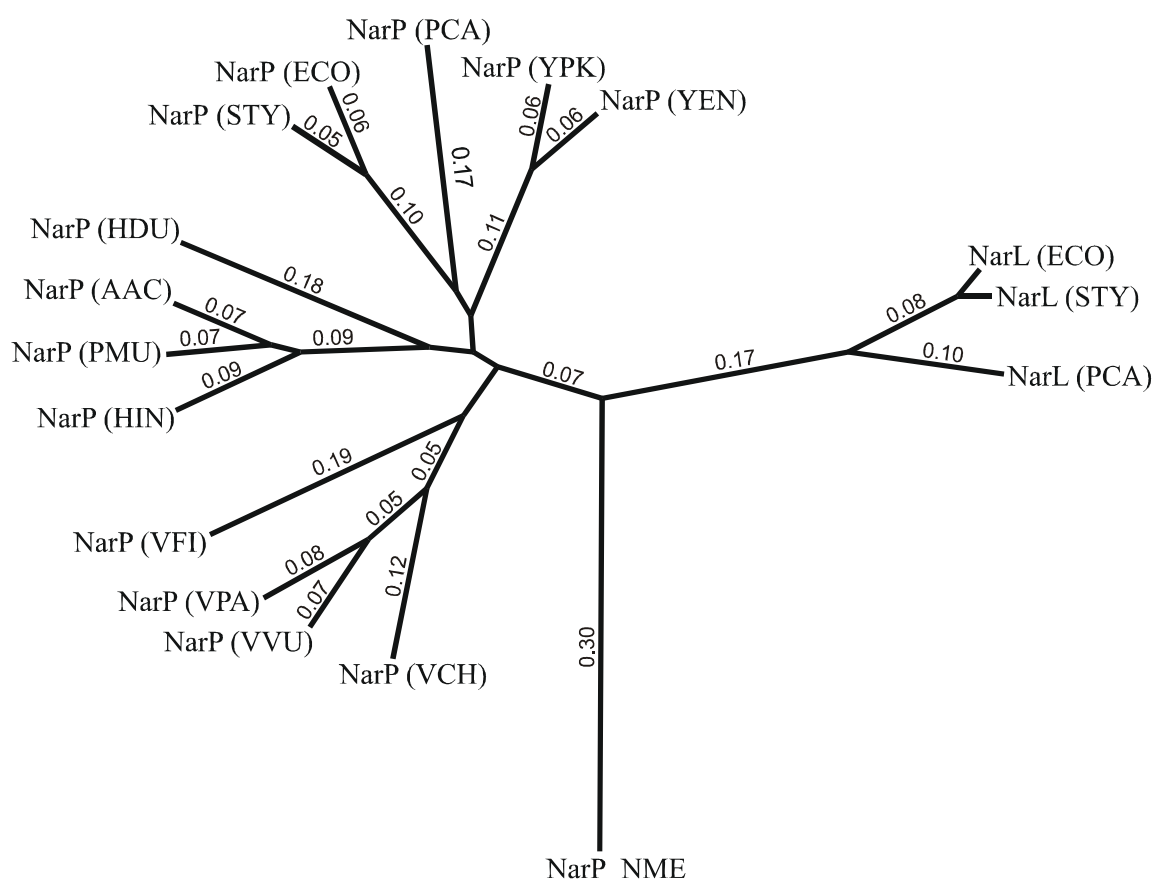


Рис. 5.4. Филогенетическое дерево белков NarL и NarP

Для построения использовался метод объединения соседних пар (neighbour-joining). В качестве внешней группы взят белок *Neisseria meningitidis* MC58, гомологичный NarL и NarP. Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,05). В скобках указаны условные обозначения геномов (см. Табл. 2.1).

Для того, чтобы проверить последнее предположение, были построены множественное выравнивание (Приложение 13) и филогенетическое дерево (Рис. 5.5) для сенсорных белков NarX и NarQ. Как и в случае регуляторных белков, сенсорные белки также образуют на филогенетическом дереве два кластера, причем в геномах, где присутствуют гены для обоих регуляторов, также обнаружены и оба сенсорных белка. В геномах же с одиночной

кроме белков NarX и NarQ. Точная функция центрального неизвестна. Важной особенностью модуля NarX *E. coli*. является кластер цистеинов Cys265, Cys302, Cys308, Cys313 и Cys316. Эти цистеиновые остатки сохраняются в последовательностях NarX из геномов других протеобактерий, и замена любого из них приводит к резкому снижению активности белка [334]. Анализ множественного выравнивания центральных модулей NarX и NarQ из исследуемых организмов (Рис. 5.7) показал, что в последовательностях белков из *Y. pestis* и *Y. enterocolitica* отсутствует характерные для NarX цистеиновые остатки. Следовательно, хотя сенсорные белки из бактерий рода *Yersinia* spp. близки по последовательности к NarX, они, по всей вероятности, выполняют функцию NarQ, то есть реакция в ответ на нитрат и нитрит одинакова. Таким образом, в геномах исследованных организмов присутствуют три типа систем регуляции нитрат-нитритного дыхания: удвоенная система, содержащая сенсоры NarX и NarQ и регуляторы NarL и NarP (Рис. 5.8а), одиночная система с «нестандартной парой» NarX-NarP, функционирующая, однако, как NarQ-NarP (Рис. 5.8б), и одиночная система NarQ-NarP (Рис. 5.8в). Каждая из этих систем характерна для определенной группы родственных организмов.

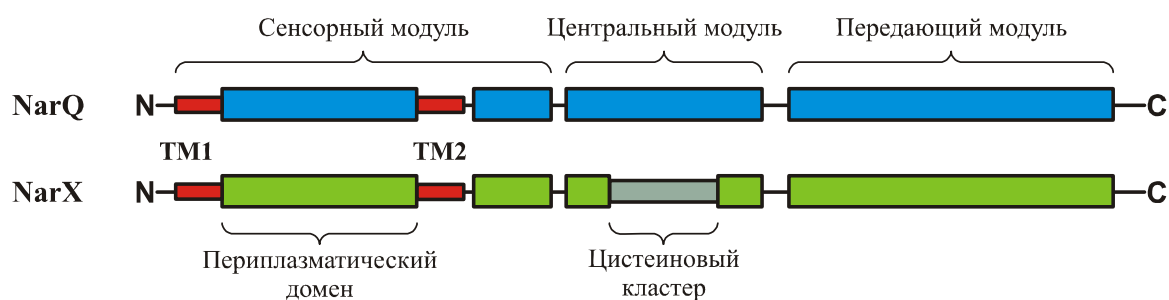


Рис. 5.6. Сравнение первичной структуры белков NarX и NarQ
 TM1 и TM2 – трансмембранные сегменты.

В процессе исследования эволюции системы регуляции нитрат-нитритного дыхания была выявлена корреляция между типом регуляторной системы и структурой системы восстановления нитрата и нитрита. Так, *E. coli* содержит достаточно сложную систему восстановления нитрата и нитрита (Рис. 5.9а). Восстановление нитрата осуществляется двумя ферментативными комплексами, один из которых, NarGHI, располагается в цитоплазме, тогда как другой, NarABCD, – в периплазме. При этом экспрессия оперонов *nar* и *nar* активируется в присутствии нитрата. Восстановление нитрата с помощью Nar-комплекса энергетически менее эффективно, так как при этом в периплазму поступают всего два протона на один перенесенный электрон, тогда как при восстановлении за счет комплекса NarGHI – три. Однако, в последнем случае в цитоплазме накапливается нитрит,

токсичный для клетки. Поэтому в цитоплазме присутствует нитритредуктазный комплекс NirBD [168]. Этот комплекс не сопряжен с мембраной и не участвует в процессах дыхания. Его функцией является восстановление нитрита до аммония, что, во-первых, служит для защиты клетки, а во-вторых, способствует ассимиляции клеткой азота. Помимо этого, нитрит выводится в периплазму с помощью мембранных транспортеров: нитрат-нитритного антипортера NarK и симпортера нитрита и протонов NirC [282]. В периплазме восстановление нитрита осуществляется уже сопряженным с мембраной Nrf-комплексом. Таким образом, восстановление нитрата может осуществляться двумя способами: в периплазме и в цитоплазме. Первый способ менее эффективен с энергетической точки зрения, но более безопасен и используется при низких концентрациях нитрата. При высоких концентрациях данный путь репрессируется с помощью регулятора NarL, и активируется второй путь [336]. Понятно, что подобная система восстановления нитрата и нитрита требует весьма сложной регуляции, что и вызвало удвоение регуляторной системы.

Рис. 5.7. Фрагмент выравнивания аминокислотных последовательностей центрального

NarQ (ECO)	T-----SQI-----DVHCFRHI-LQIVRDNE--AAEYLELN-VGE-NWRISE--G-QPNPE--
NarQ (STY)	T-----SQI-----DVHCFRHI-LQIVREHD--AAWYLELT-VGD-NWRISE--G-TQSPD--
NarQ (PCA)	V-----STI-----DRHCFQOI-LQIVHRYE--TVVCLEMR-VGE-NWLLCE--G-QPDEQ--
NarQ (VPA)	A-----SRI-----NQDNFQAI-LKHIASLEGIKAVKLEIEQLGEPNWILTE--GEECCHD--
NarQ (VVU)	A-----SRI-----SQENFQAI-LRHIVSIEGIVSAKLEIEEIGERNLVLTE--GPKCVGR--
NarQ (VCH)	A-----SRI-----TASNFQTI-LRHWVALEGICALRLEIEEEAGKPLILQE--G-KPSGA--
NarQ (VFI)	A-----SQL-----DEQAFKNI-LDTFTNIEGII SARLIVEEESGGDWEITS--GEPDESP--
NarQ (AAC)	T-----NKI-----DGKILQLV-LQNVMI SE--HLRYLELDVLDAPHWNICL--GVKYDQLE-
NarQ (PMU)	T-----TNV-----DESILQHV-LKNIFISE--HLRYMALVVEGAHWNIRF--GQKQANQD-
NarQ (HIN)	T-----NTI-----NDKILNQV-LNYIFISD--HLNFVKVEVMGAEHWDITL--GKQDANNE-
NarQ (HDU)	SA-----KPL-----NQSPLYV-LQIILDNE--HLRAIEIQVYGADYWNVTI--DNAPAQT--
NarX (ECO)	SRAPL C ERLSPVLNGLQNLTLRLDIELRVYDTDD--EENHQEFT- C -QPD-MT C DDK G QL C PRG-
NarX (STY)	SQAPL C ERLSPVLNGLQNLTLQHDIELRVYDLED--EDNHQEFT- C -QSD-IS C DDK G HL C PRS-
NarX (PCA)	TGAPL C SRLMPILNELPSLTPLRNIQLRLYEDNN--QEQFHQFSD C SQSQPEH C PDNS C QS C GMQ-
NarX (YPK)	THQPL S ERLAPVIEQLQALTTLENVQICLYENHL--YRDHVNHN-A-DAEYLM P K N R P AS L T I SGP
NarX (YEN)	TTQPL S ERLVPVIEQLQTLTPLENVQICLYENHL--YRSHLADH-L-DGEYLP P Q N R P T Q L T I S GP
	260 270 280 290 300 310

модуля белков NarX и NarQ.

Под выравниванием указаны позиции аминокислот для NarX *E. coli*. Консервативные цистеиновые остатки показаны на зеленом фоне, соответствующие им позиции в белках *Yersinia* spp. – на сером фоне. В скобках указаны условные обозначения геномов (см. Табл. 2.1).

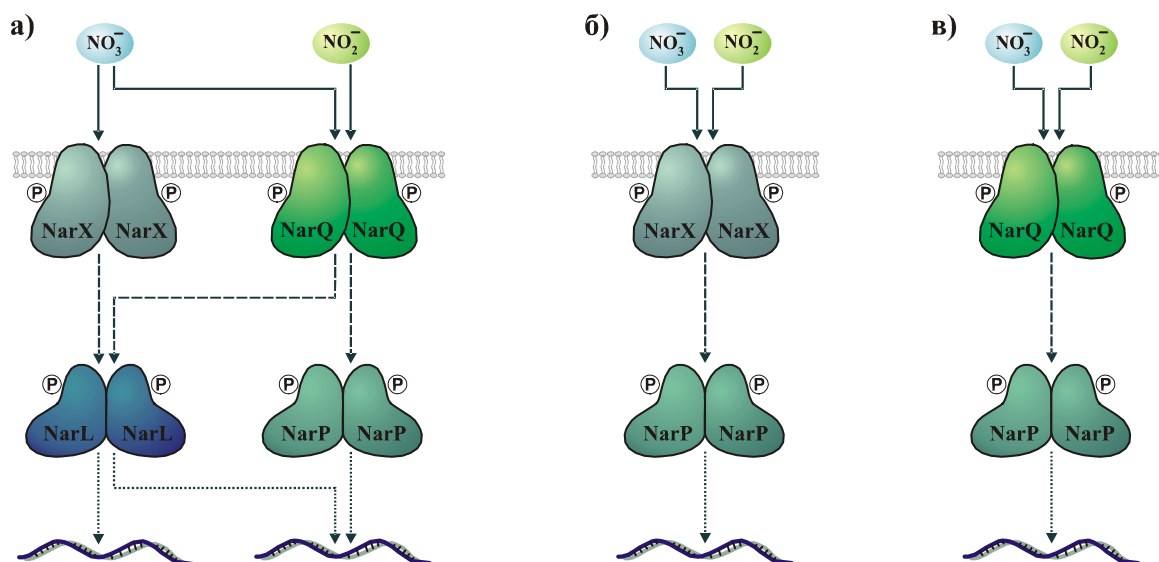


Рис. 5.8. Типы систем регуляции нитрат-нитритного дыхания.

Показано взаимодействие эффекторов с сенсорами (сплошные стрелки), сенсоров с регуляторами (пунктирные стрелки) и регуляторов с сайтами связывания (точечные стрелки). (а) *E. coli*, *S. typhi*, *P. carotovorum*; (б) *Yersinia* spp.; (в) Pasteurellales и Vibrionales.

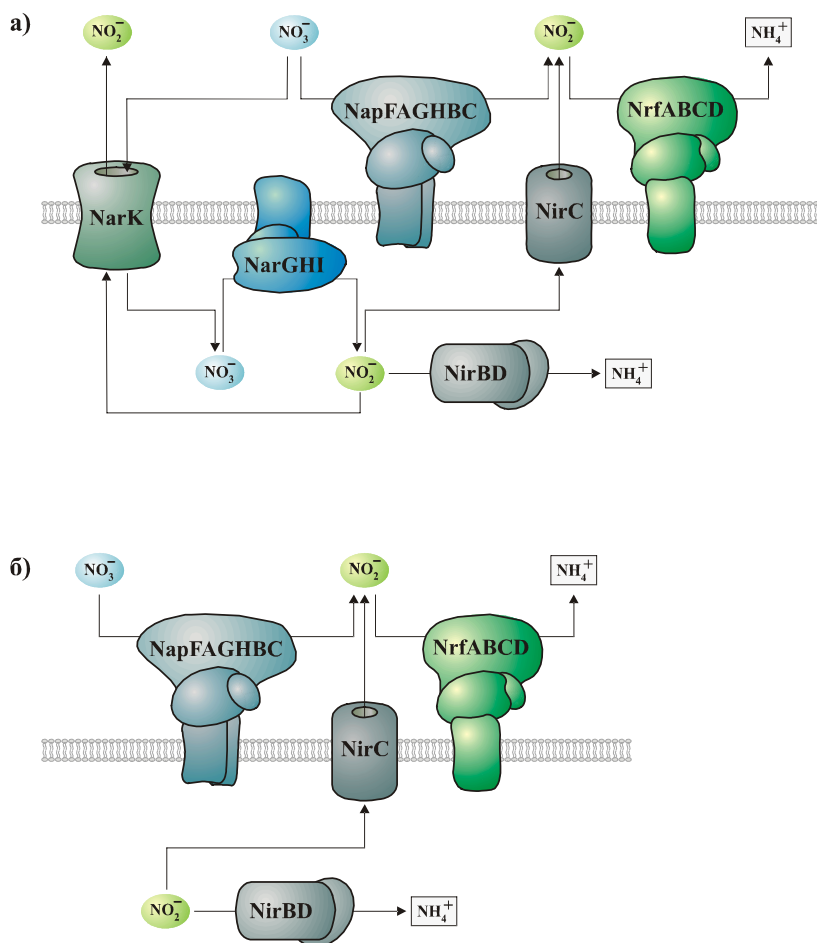


Рис. 5.9. Системы восстановления нитрата и нитрита.

(а) *E. coli*, *S. typhi*, *P. carotovorum*; (б) Pasteurellales и Vibrionales (за исключением *V. cholerae*).

В геномах Pasteurellales и большинства Vibrionales были найдены гены только для периплазматической нитрат-редуктазы, ортологов же генов Nar-комплекса не было обнаружено. Таким образом, можно предположить, что эти организмы выбрали более безопасный путь восстановления нитрата (Рис 5.9б). В этом случае достаточно более простой регуляции, что и выражается в присутствии лишь одной регуляторной системы. Поскольку не требуется дифференцированного ответа на нитрат и нитрит, в качестве сенсорного белка в данной системе действует NarQ. В геномах бактерий рода *Yersinia* были обнаружены ортологи для гена *narX*. Однако, ряд изменений в аминокислотной последовательности, в частности, отсутствие консервативных цистеинов, указывает, что этот белок выполняет, по всей видимости, функции NarQ. Однако, в геномах *Y. pestis* и *Y. enterocolitica* отсутствуют гены для периплазматической нитрит-редуктазы, а в геноме *V. cholerae* вообще не найдено генов ни для одной из нитрит-редуктаз. Поэтому организация системы восстановления нитрата и нитрита в этих трех организмах требует отдельного изучения, что выходит за рамки данного исследования.

5.2. Построение распознающих правил для поиска сайтов связывания регуляторов дыхания

Как было упомянуто ранее, все исследованные в данной части работы геномы содержат факторы транскрипции Fnr и ArcA. В случае же регуляции нитрат-нитритного дыхания наблюдаются существенные различия между геномами. Так, *E. coli*, *S. typhi* и *P. carotovorum* содержат гомологичные регуляторы NarL и NarP, тогда как у представителей *Yersinia* spp., Pasteurellales и Vibrionales был обнаружен только единичный белок NarP. Поскольку данные о структуре сайта связывания NarL противоречивы [223, 227-229], а наличие удвоенной регуляторной системы крайне затрудняет исследование, было решено сосредоточить усилия на исследовании геномов с одиночной системой регуляции нитрат-нитритного дыхания. Таким образом, были изучены 10 геномов, относящихся к следующим трем таксонам: *Yersinia* spp. (*Y. pestis*, *Y. pseudotuberculosis*), Pasteurellales (*P. multocida*, *A. actinomycetemcomitans*, *H. influenzae*, *H. ducreyi*) и Vibrionales (*V. vulnificus*, *V. parahaemolyticus*, *V. cholerae*, *V. fischeri*). Все перечисленные геномы содержат Fnr, ArcA и NarP, но не содержат NarL.

5.2.1. Построение распознающего правила для поиска потенциальных сайтов связывания Fnr

Поскольку генетические расстояния между белками Fnr из разных организмов невелики, возможно проводить поиск сайтов в исследованных геномах, используя единую

матрицу. Такая матрица (Рис. 5.10) была составлена на основе экспериментально определенных сайтов связывания Fng *E. coli* (Табл. 5.1). Fng-связывающий мотив представляет собой инвертированный повтор длиной 14 п.н. (Рис. 5.11), согласующийся с предсказанным ранее консенсусом [183, 185]. Наименьший вес среди сайтов обучающей выборки, 3,87, имеет сайт перед геном *narF*. Для того, чтобы учесть возможность видоспецифичных изменений ДНК-белковых взаимодействий, пороговое значение для весов потенциальных сайтов было принято равным 3,75. Исходя из положения экспериментальных сайтов связывания Fng, поиск сайтов производился в области -500..+100 п.н. относительно старта трансляции гена. При данных параметрах поиска сайты связывания Fng белка были обнаружены перед приблизительно 600 генами на геном.

Поз.	A	C	G	T	Конс.
1	0.04	-0.27	-0.27	0.51	T
2	-0.20	-0.20	-0.20	0.59	T
3	-0.20	-0.20	0.59	-0.20	G
4	0.39	-0.16	-0.16	-0.06	A
5	-0.35	-0.14	0.08	0.41	T
6	0.04	0.00	-0.06	0.02	N
7	0.01	-0.11	-0.11	0.20	T
8	0.27	0.11	-0.42	0.05	A
9	-0.06	0.02	-0.03	0.06	N
10	0.35	-0.01	0.04	-0.39	A
11	0.05	-0.37	-0.06	0.38	T
12	-0.04	0.54	-0.25	-0.25	C
13	0.48	-0.29	0.09	-0.29	A
14	0.48	-0.29	-0.29	0.09	A

Рисунок 5.10. Позиционная матрица весов для сайтов связывания Fng

Указан вес данного нуклеотида в соответствующей позиции. “Поз.” – позиция сайта; “Конс.” – нуклеотид, преобладающий в данной позиции.

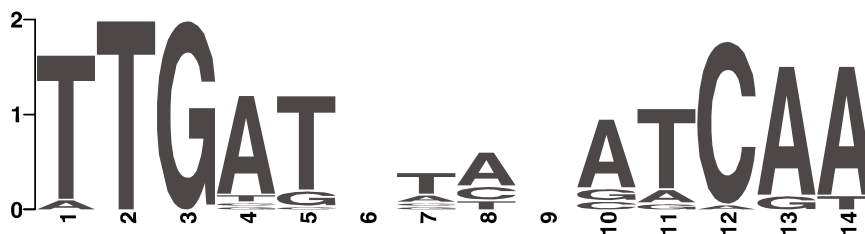


Рисунок 5.11. Диаграмма Лого для сайтов связывания Fng

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

Таблица 5.1. Сайты связывания белка Fnr *E. coli*, вошедшие в обучающую выборку
 Указано положение сайта относительно старта трансляции гена, также приводятся ссылки на работы, в которых регуляция гена была показана экспериментально. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Ген	Сайт			Ссылки
	Положение	Вес	Последовательность	
<i>cydA</i>	-349	5,08	TTGATaTttATCAA	[200]
	-297	4,55	TTGtTcTcgATCAA	
	-252	4,15	TTGATcaccgTCgA	
<i>moeA</i>	19	4,19	TTGATgTcgCTCgA	[199]
<i>hcr</i>	-109	4,25	TTGcgCTAAATCAA	[191]
<i>dmsA</i>	-185	4,41	TTGATaccgAaCAA	[13]
<i>focA</i>	-73	4,79	aTGATcTAtATCAA	
<i>ndh</i>	-151	5,05	TTGATTaAcATCAA	[201]
<i>hlyE</i>	-385	4,36	TTGATaaAcAaCAt	[198]
	-139	4,69	TTGATaTttATCAt	
<i>narX</i>	-278	5,07	TTGATgTAAATCAA	[187]
	-240	4,69	TTGATaTttATCAt	
<i>narK</i>	-112	4,65	aTGATaaAtATCAA	[13]
	-74	5,24	TTGATTTAcATCAA	
<i>narG</i>	-106	4,72	TTGATcgttATCAA	[13]
<i>fnr</i>	-32	4,57	TTGAcAAAtATCAA	[188]
<i>fdnG</i>	-134	4,46	TTGAggTAggTCAA	[195]
	-79	4,23	TTGATTTcgcgCAA	
<i>hmp</i>	-41	4,93	TTGAgATAcATCAA	[204]
<i>yfiD</i>	-173	4,75	TTGATgTAAAaCAA	[202]
	-120	5,16	TTGATTTAAATCAA	
<i>nirB</i>	-73	5,24	TTGATTTAcATCAA	[146]
<i>yhjA</i>	-120	4,81	TTGATTcctATCAA	[194]
<i>nrfA</i>	-142	4,33	TTGATTaAAgaCAA	[193]
<i>nrdD</i>	-248	4,89	TTGAgCTAcATCAA	[192]
<i>arcA</i>	-290	4,99	TTGATaTAtgTCAA	[186]
<i>napF</i>	-45	3,87	TTGgTcgctcTCAA	[196]

5.2.2. Построение распознающего правила для поиска потенциальных сайтов связывания ArcA

В случае ArcA-зависимой регуляции существующих данных о структуре сайтов связывания было явно недостаточно для построения чувствительного распознающего правила, а предсказанный консенсус для данных сайтов отличался слабой специфичностью, что приводило к массовым перепредсказаниям [210]. Поэтому регуляторные области генов, экспрессия которых регулируется ArcA в *E. coli* (Табл. 5.2) были проанализированы с помощью программы SeSiMCMC [111]. Всего перед 14 генами было найдено 18 потенциальных сайтов связывания регулятора, на основании которых была построена матрица позиционных весов (Рис 5.12). Потенциальный мотив связывания ArcA представляет собой несовершенный повтор длиной 15 п.н. (Рис. 5.13). Построенный профиль имеет гораздо большую специфичность, чем составленный ранее [210], поскольку при условии нахождения всех сайтов из обучающей выборки дает меньшее число ложных предсказаний.

Поз.	A	C	G	T	Конс.
1	-0.26	0.06	-0.26	0.45	T
2	0.38	-0.29	-0.29	0.19	A
3	0.30	-0.35	-0.03	0.08	A
4	-0.07	0.42	-0.29	-0.07	C
5	0.49	-0.02	-0.23	-0.23	A
6	0.07	-0.09	-0.19	0.21	T
7	0.10	-0.09	-0.20	0.19	T
8	0.18	-0.39	0.05	0.15	W
9	0.26	-0.04	-0.36	0.15	A
10	0.15	-0.39	0.15	0.09	R
11	-0.23	-0.23	-0.02	0.49	T
12	0.06	-0.26	-0.26	0.45	T
13	0.30	0.03	-0.36	0.03	A
14	0.55	-0.18	-0.18	-0.18	A
15	-0.26	0.45	-0.26	0.06	C

Рисунок 5.12. Позиционная матрица весов для сайтов связывания ArcA

Указан вес данного нуклеотида в соответствующей позиции. “Поз.” – позиция сайта; “Конс.” – нуклеотид, преобладающий в данной позиции.

Наименьший вес среди обучающей выборки был 4,12 (сайт перед геном *fadL*), поэтому для поиска потенциальных сайтов было выбрано пороговое значение, равное 4,00. Поскольку сайты связывания ArcA, в отличие от сайтов упоминавшихся ранее регуляторов, не являются палиндромными, поиск производился по последовательностям для обеих цепей ДНК, и прямой, и обратной. Поиск сайтов проводился в области –500 ...+100 п.н. относительно старта трансляции гена, при этом сайты обнаруживались перед порядка 500 генами на геном.



Рисунок 5.13. Диаграмма Лого для сайтов связывания ArcA

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

Таблица 5.2. Сайты связывания белка ArcA *E. coli*, вошедшие в обучающую выборку

Указано положение сайта относительно старта транскрипции гена и цепь ДНК, на которой он расположен: «→» - прямая, «←» - обратная. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными. В столбце «Ссылки» указаны статьи, подтверждающие, что данный ген в *E. coli* регулируется ArcA.

Ген	Сайт				
	Напр-ие	Положение	Вес	Последовательность	Ссылки
<i>fadE</i>	←	-86	4,41	TA gCAaTat GT tAC	[212]
<i>gltA</i>	→	-393	4,41	c AACATTTAt TT Aat	[211]
<i>sdhC</i>	→	-69	4,78	TA ACAgaa AG TTAAC	[211]
	←	-330	4,41	c AACATTTAt TT Aat	
<i>cydA</i>	←	-388	4,22	T ttCAaTa AG gTAAC	[211]
<i>icd</i>	→	-111	4,51	T tACAaaTca TT AAC	[213]
<i>aldA</i>	→	-49	4,62	TA ACAaTgta TT cAC	[215]
<i>fadD</i>	→	-72	4,67	TA atATTat GT TTAAC	[212]
	←	-74	5,07	TA ACATaata TT AAC	
<i>fadI</i>	←	-65	4,16	TA ACcaTTtt TT tAC	[212]
<i>fadL</i>	→	-238	4,12	TAA aAcaaAa TT cAC	[212]
<i>glcD</i>	→	-171	4,74	T tACATcaAt TT AAC	[216]
	→	-160	5,15	TA ACAT Tg AGTTAAC	
<i>lldP</i>	→	-134	5,25	TA ACATTTAGTTAAC	[211]
<i>fadB</i>	→	-77	4,45	T tgCATaTtt TT AAC	[212]
	→	-66	4,22	TA ACAcAAa T acAC	
<i>sodA</i>	→	-79	5,03	T atCATTTAa TT AAC	[214]
<i>moeA</i>	→	-3	4,08	T ttCAT gg Aa TT tAC	[199]

5.2.3. Построение распознающего правила для поиска потенциальных сайтов связывания NarP

В отличие от других регуляторов дыхания, Fnr и ArcA, в случае NarP использование матрицы, построенной по сайтам связывания данного белка в *E. coli*, было сочтено неправомерным в силу следующих причин. В *E. coli* регуляция генов нитрат-нитритного дыхания осуществляется сразу двумя гомологичными белками, NarL и NarP, причем наблюдается различная специфичность белков-регуляторов к сайтам связывания. Так, NarP связывается с палиндромными сайтами длины 16 п.н., тогда как NarL связывается как с сайтами связывания NarP, так и со множеством других сайтов, точная структура которых неизвестна [223]. Поэтому в геноме *E. coli* сайты связывания NarP являются, по сути, общими сайтами для обоих белков. Поскольку в геномах, исследуемых в данной работе, присутствует лишь единичный регулятор NarP, и ситуация отличается от наблюдаемой в *E. coli*, была построена новая матрица для сайтов связывания NarP для тех геномах, где присутствует единичная система.

Для построения матрицы для поиска сайтов связывания NarP были исследованы регуляторные области оперонов, содержащих *nar*-, *cst*- и *nrf*-гены в исследуемых геномах. Перед большинством данных оперонов были найдены палиндромные последовательности длиной 16 п.н. (Табл. 5.3.), похожие на сайты связывания NarP из *E. coli*. На основании этих последовательностей была построена матрица (Рис. 5.14, 5.15) для распознавания NarP-сайтов в геномах с единичной регуляторной системой.

Поз.	A	C	G	T	Конс.
1	-0.25	0.06	-0.25	0.44	T
2	0.38	-0.08	-0.31	0.00	A
3	0.17	0.33	-0.17	-0.33	C
4	-0.16	0.36	-0.32	0.13	C
5	0.07	0.10	-0.40	0.23	T
6	0.07	0.30	-0.37	0.00	C
7	-0.07	0.02	-0.23	0.29	T
8	0.22	-0.18	-0.34	0.30	W
9	0.30	-0.34	-0.18	0.22	W
10	0.29	-0.23	0.02	-0.07	A
11	0.00	-0.37	0.30	0.07	G
12	0.23	-0.40	0.10	0.07	A
13	0.13	-0.32	0.36	-0.16	G
14	-0.33	-0.17	0.33	0.17	G
15	0.00	-0.31	-0.08	0.38	T
16	0.44	-0.25	0.06	-0.25	A

Рисунок 5.14. Позиционная матрица весов для сайтов связывания NarP

Указан вес данного нуклеотида в соответствующей позиции. “Поз.” – позиция сайта; “Конс.” – нуклеотид, преобладающий в данной позиции.



Рисунок 5.15. Диаграмма Лого для сайтов связывания NarP

По горизонтальной оси указан номер позиции нуклеотида, по вертикальной – информационное содержание позиции в битах. Высота столбца пропорциональна информационному содержанию данной позиции, относительная высота каждой буквы соответствует частоте нуклеотида в данной позиции.

Поскольку наименьший вес сайта из обучающей выборки составлял 3,62 (гена *narF* из *H. ducreyi*), пороговое значение для поиска сайтов было принято равным 3,50. Исключение было сделано для *A. actinomycetemcomitans*: поскольку сайты перед *nar*-, *ccm*- и *nrf*-генами имели низкие веса, для этого генома было принято пороговое значение 3,25. Поиск потенциальных сайтов производился в области –500..+100 п.н. относительно старта трансляции гена. При данных условиях в среднем предсказывалось по 300 сайтов на геном.

5.3. Состав обобщенных Fnr-, ArcA- и NarP-регулонов

В соответствии с процедурой, описаной в Главе 2, в десяти исследуемых геномах были выделены обобщенные регулоны для факторов транскрипции Fnr, ArcA и NarP. Поскольку, как уже отмечалось ранее, регуляция каждой отдельной регуляторной системой слабо консервативна, то имеет смысл рассматривать регуляцию тремя системами в целом. В общей сложности, к трем обобщенным регулонам было отнесено 153 гена, образующих как минимум 68 оперонов (Табл. 5.4). Предсказанные сайты связывания для Fnr, ArcA и NarP приведены в Приложениях 14–16. В соответствии с функцией, все гены были разбиты на 10 групп, каждая из которых описывается ниже.

5.3.1. Белки дыхательных цепей

Ранее регуляция глобальными регуляторами дыхания была показана для таких оперонов, кодирующих белки-компоненты дыхательных цепей, как *cydAB* [337], *nar* [338, 339]), *ccm* [338], *nrf* [225, 230], *nir* [146, 284], *dms* [14, 229], *torCAD* [239], *frd* [237, 239, 340], *fdn* [195, 341], *ndh* [342, 343], *glpABC* [344] и *glpD* [344]. В дополнение к этому, в настоящей работе к обобщенным регулонам был отнесен и ряд других генов, кодирующих необходимые для дыхания белки.

Таблица 5.3. Сайты связывания белка NarP, вошедшие в обучающую выборку

Указано положение сайта относительно старта трансляции гена. Приводятся ссылки на работы, в которых регуляция гена была показана экспериментально. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Ген	Сайт		
		Положение	Вес	Последовательность
<i>Y. pestis</i>	<i>napF</i>	-184	4,48	TAaCTCTaAAGAGtaA
	<i>ccmA</i>	-113	4,67	TACCTCTatAagGGTA
<i>Y. enterocolitica</i>	<i>napF</i>	-182	4,48	TAaCTCTaAAGAGtaA
	<i>ccmA</i>	-113	4,67	TACCTCTatAagGGTA
		-79	3,83	TACtaCgTAAGgatTA
<i>P. multocida</i>	<i>ccmA</i>	-99	4,80	cACCTCTaAAGAGGTA
	<i>nrfA</i>	-289	4,79	TAaCTCTaAAatAGGTA
		-206	4,72	TACtTaTTtAGAGGTA
<i>A. actinomycetemcomitans</i>	<i>nrfA</i>	-403	4,79	TACCTCTTtAtAGtTA
		-343	4,33	TACtaaTaAAatAGGTA
<i>H. influenzae</i>	<i>ccmA</i>	-85	4,79	TACtTCTaAAGAGtTA
<i>H. ducreyi</i>	<i>napF</i>	-81	3,62	TACCTaaatgGAGtgA
	<i>ccmA</i>	-117	3,83	TACtataaaAAaAGGTA
		-86	4,95	TACCTCTaAAGAaGTA
<i>V. vulnificus</i>	<i>napF</i>	-281	4,73	TACCcCtTAAGgGGTA
	<i>ccmA</i>	-147	4,48	TcaCTCTatAGAGGTA
	<i>nrfB</i>	-304	4,92	TACCcCtTaAAGgGGTA
		-202	3,87	TACtaCTaAAGAGtag
	<i>nrfA</i>	-295	3,87	ctaCTCTTtAGtaGTA
		-193	4,92	TACCcCTTtAGgGGTA
<i>V. parahaemolyticus</i>	<i>napF</i>	-124	4,99	TACCTCtTAAGAGGTA
	<i>napG</i>	-82	4,90	TACCTCTTAtGAGGTA
	<i>nrfB</i>	-300	4,89	TACCcCtTaAAGtGGTA
		-196	4,22	TACCTCTTAgGAataA
	<i>nrfA</i>	-294	4,22	TtatTCcTAAGAGGTA
		-190	4,89	TACCaCTTtAGgGGTA
<i>V. cholerae</i>	<i>napF</i>	-153	4,17	cACCTCTTtAGgGcTA
		-79	4,99	TACCTCtTAAGAGGTA
	<i>ccmA</i>	-143	4,23	TAaCTacaAAGgGtTA
<i>V. fischeri</i>	<i>napF</i>	-86	4,25	TAaCcaTTAtGgGGTA
	<i>ccmA</i>	-191	4,52	TACCTacaAAGtGGTA
		-141	3,53	TACaaaTTAcaAGGTA

Таблица 5.3. (Продолжение)

Геном	Ген	Сайт		
		Положение	Вес	Последовательность
<i>V. fischeri</i>	<i>nrfB</i>	-383	4,64	TACt aaTTAAGAGGTA
		-309	4,34	cACCatTaAAGAGGTA
		-293	4,94	TACCTCTTAAGtGtTA
		-205	3,64	TAtCaCaaAAGgaGTA
	<i>nrfA</i>	-280	3,64	TACtcCTTttGtGaTA
		-192	4,94	TAaCaCTTAAGAGGTA
		-176	4,34	TACCTCTTtAatGGTg
		-102	4,64	TACCTCTTAAAttaGTA

Одним из наиболее значимых результатов является предсказание регуляции оперона *atpIBEFHAGCD*. Данный оперон кодирует все субъединицы АТФ-синтетазы, центрального компонента дыхания [345]. Несмотря на всю важность данного оперона, его регуляция плохо изучена, а имеющиеся данные противоречивы. Так, утверждается, что ни Fnr, ни ArcA не влияют на экспрессию оперона *atp* [346]. С другой стороны, было зафиксировано повышение уровня экспрессии различных генов *atp*-оперона как в *fnr*- [347], так и в *arcA*-мутантных штаммах [348] *E. coli*, по сравнению с диким типом. Недавно нами были предсказаны потенциальные сайты связывания Fnr и ArcA перед данным опероном в геноме *E. coli* K-12; эти сайты сохраняются также и в геномах многих других представителей порядка Enterobacteriales [349].

В настоящей работе сайты связывания ArcA перед *atp*-опероном были обнаружены в *Yersinia* spp., тогда как консервативные Fnr-сайты были найдены в Vibrionales. Таким образом, можно утверждать, что экспрессия генов АТФ-синтетазы контролируется глобальными регуляторами дыхания как минимум в некоторых гамма-протеобактериях.

Потенциальные регуляторные сайты были найдены также перед опероном *nqrABCDEF*, кодирующий НАД·Н-дегидрогеназу, экспортирующую ионы Na⁺ [350]. Предполагаемые сайты связывания Fnr были найдены во всех исследованных геномах, тогда как сайты связывания ArcA обнаружались у всех Pasteurellales и Vibrionales. Регуляция же данного оперона белком NarP, судя по всему, является таксон-специфичной особенностью Pasteurellales. Впервые Na⁺-транспортирующая НАД·Н-дегидрогеназа была обнаружена у морской бактерии *Vibrio alginolyticus* [351]. Ортологи генов *nqr*-оперона, как уже упоминалось ранее, были обнаружены во всех исследованных геномах. Для *E. coli* известна другая, протон-экспортирующая, НАД·Н-дегидрогеназа, кодируемая опероном *nuo*, транскрипция которого активируется белками NarL и FNR [234]. Оперон *nuo* найден в

Оперон	Геном										
	<i>Yersinia spp.</i>		Pasteurellales				Vibrionales				
	УРК	YEN	PMU	AAC	HN	HDU	VVU	VPA	VCH	VFI	
Центральный метаболизм и брожение (продолжение)											
<i>pflB</i>	--N	--N	-AN	-AN	-A-	f-N	-AN	--N	fA-	-AN	-AN
<i>yfiD</i>	Fa-	F--	FA-	FA-	FA-	-An	F-N	FAN	FAN	FAN	FAN
<i>mdh</i>	-A-	fA-	f-N	-aN	f-N	-aN	FA-	FA-	FA-	---	---
<i>pckA</i>	---	---	--n	-A-	fAn	fA-	-A-	-An	-A-	Fan	Fan
<i>ppsA</i>	F--	F--	0	0	0	0	-A-	fA-	---	-A-	-A-
<i>aspA</i>	f--	-a-	-A-	f--	-A-	-An	--N	--N	0	fAn	fAn
<i>sfcA</i>	--N	fAn	0	---	0	0	F--	F--	F--	F-n	F-n
<i>sdhCDAB</i>	-A-	-A-	0	0	0	0	-A-	fA-	fA-	-A-	-A-
<i>gltA</i>	-A-	-A-	-an	0	0	0	-A-	fA-	fA-	-A-	-A-
<i>fumC</i>	f--	---	--N	--N	--N	f--	0	-a-	fA-	0	0
<i>sucABCD</i>	--- 7a	--- 7a	-A- 7a	---	--N 7c	fAN 7a	---	7a	---	-a- 7a	-a- 7a
<i>ldhA</i>	---	f--	0	---	-a-	0	-a-	--N	f-N	-aN	-aN
<i>talB</i>	f--	---	-aN	0	f-N	f-N	-a-	f--	---	---	---
Метаболизм углеводов											
<i>mtlADR</i>	fA-	-A-	---	0	0	0	F--	Fa-	F--	F--	0
<i>nagBACD</i>	---	8a	---	8b	---	--n 8b	F-- 8d	F-- 8d	F-- 8d	F-n 8d	F-n 8d
	---	8c	fa-	8c	---	---	f-- 8e	f-- 8e	f-- 8e	---	---
<i>ptsHI-crr</i>	--n	---	-A-	-A-	-A-	fA-	---	--n	-a-	--n	--n
<i>deoCABD</i>	f-- 9a	-a- 9a	---	9b	f-- 9b	f-- 9b	-N 9a	--N 9a	---	-N 9a	-N 9a
<i>malQ-glgBXCAP</i>	---	10a	F-n 10c	---	F-n 10c	0	10d	f-- 10d	---	10d	---
	Fa-	10b	---	10b	---	---	-an 10e	---	---	-a- 10e	-a- 10e
	---	10c	F-- 10c	---	F-- 10c	---	---	---	fa-	10f	f-- 10f
	---	10d	---	10d	---	---	---	---	---	10g	-an 10g
											10h
											10h
Метаболизм жирных кислот											
<i>fadLJ</i>	-A-	-A-	0	0	0	0	FA-	FA-	F--	F--	FA-
<i>fadBA</i>	FAn	FA-	0	0	0	0	-A-	fA-	-A-	-A-	fA-
<i>fadD</i>	fA-	-An	---	-a-	-an	f--	-A-	-A-	-A-	-A-	---
<i>acpP-fabF</i>	-A- 11a	-A- 11a	fAn 11b	---	11b	---	11b	F-- 11a	F-- 11a	F-- 11a	---
											11a
Ответ на кислородный стресс											
<i>sodA</i>	-A-	-A-	FA-	FA-	FAn	FA-	0	0	0	0	0

Оперон	Геном										
	Yersinia spp.				Pasteurellales				Vibrionales		
	YPK	YEN	PMU	AAC	HIN	HDU	VVU	VPA	VCH	VFI	
Нуклеотидпродуктазы											
<i>nrdDG</i>	F--	F--	f--	---	f--	0	F--	Fa-	F--	F--	F--
<i>nrdAB</i>	--n	---	F-n	F--	F--	---	---	---	---	f--	f--
Транспортные белки											
<i>focA</i>	FAN	FAN	FA-	FA-	FA-	FA-	FA-	---	---	fa-	fa-
<i>dcuB</i>	-A-	fA-	F-N	F-N	0	F-N	-a-	---	---	f--	fan
<i>dcuA</i>	-A-	fA-	--N	f-N	f-N	-a-	---	--n	---	---	--n
<i>dcuC</i>	0	f--	-a-	-a-	0	0	Fa-	F--	F--	Fa-	Fa-
<i>glfP</i>	-An	fA-	0	0	0	0	F-n	-A-	-A-	FAn	FAn
<i>glpFK</i>	-An	-A-	f--	-an	-an	fa-	-A-	-An	---	-An	-An
<i>gntXY</i>	---	---	FA-	F--	FA-	FA-	---	---	---	---	--n
<i>fadL</i>	FA-	FA-	---	---	f--	-a-	FA-	-A-	F--	F--	FA-
<i>feoAB</i>	---	f--	0	0	0	0	FAN	FAN	-AN	FA-	FA-
Пептидазы											
<i>pepT</i>	F--	F--	F--	F--	F--	0	F--	Fa-	F--	F--	Fan
Факторы транскрипции											
<i>fir</i>	F--	F--	FAN	FAN	FAN	FA-	FA-	Fa-	F--	F--	F--
<i>arcA</i>	FA-	FA-	f--	---	f-n	-a-	FA-	FA-	FA-	FA-	FA-
<i>narQP</i>	-a- ^{12a}	--n ^{12a}	f-- ^{12a} f-- ^{12b}	--n ^{12a} fa- ^{12b}	--n ^{12a} fa- ^{12b}	---	FAN ^{12c}	FAN ^{12c}	FAN ^{12c}	FAN ^{12c}	FAN ^{12c}
<i>dusB-fis</i>	-a-	---	FA-	-A-	FA-	-A-	---	---	---	---	f--
<i>cpxRA</i>	---	---	FA-	FA-	FA-	FAn	---	---	---	---	-a-
<i>oxyR</i>	---	---	-A-	-A-	-A-	fA-	---	f--	---	---	-a-
<i>fur</i>	---	---	F--	F--	F--	F--	-A-	-A-	-A-	-An	-An

Таблица 5.5. Оперонные перестройки для генов *Fnr*-, *ArcA*- и *NarP*-регулонов.

Обозначения геномов: см. Табл. 2.1. Верхние индексы соответствуют таковым в Табл. 5.4.

Гены	Геном									
	YPK	YEN	PMU	AAC	HIN	HDU	VVU	VPA	VCH	VFI
<i>atpIBEFHAGDC</i>	^{1a} <i>atpIBEFHAGDC</i>			^{1b} <i>atpBEFHAGDC</i>				^{1a} <i>atpBEFHAGDC</i>		
<i>napFDAGHBC</i>	^{2a} <i>napFDABC</i>			^{2b} <i>napFDAGHBC</i>			^{2a} <i>napFDABC</i> ^{2c} <i>napGH</i>		^{2a} <i>napFDABC</i>	
<i>nrjABCDEXFG</i>	0	0	^{3a} <i>nrjABCDEXFG</i>		^{3b} <i>nrjABCD</i> ^{3c} <i>nrjEXFG</i>		^{3d} <i>nrjA</i> ^{3e} <i>nrjBCDEXF</i> ^{3f} <i>nrjG</i>	0	^{3d} <i>nrjA</i> ^{3e} <i>nrjBCDEXF</i> ^{3f} <i>nrjG</i>	
<i>torCAD</i>	0	0	^{4a} <i>torCAD</i>	0	0			^{4b} <i>torCA</i> ^{4c} <i>torD</i>		
<i>moaABCDE</i>	^{5a} <i>moaABCDE</i>			^{5a} <i>moaABCDE</i>				^{5b} <i>moaABCDE</i>		
<i>pdhR-aceEF-lpdA</i>	^{6a} <i>pdhR-aceEF-lpdA</i>			^{6b} <i>aceEF-lpdA</i>				^{6a} <i>pdhR-aceEF-lpdA</i>		
<i>sucABCD</i>	^{7a} <i>sucABCD</i>		^{7a} <i>sucABCD</i>		^{7b} <i>sucAB</i> ^{7c} <i>sucCD</i>	^{7a} <i>sucABCD</i>		^{7a} <i>sucABCD</i>		
<i>nagBACD</i>	^{8a} <i>nagBACD</i>		^{8b} <i>nagBA</i> ^{8c} <i>nagC</i>		^{8b} <i>nagBA</i>			^{8d} <i>nagB</i> ^{8e} <i>nagAC</i>		
<i>deoCABD</i>	^{9a} <i>deoCABD</i>		^{9b} <i>deoD</i>		^{9c} <i>deoCD</i>			^{9a} <i>deoCABD</i>		
<i>malQ-glgBXCAP</i>	^{10a} <i>malQ</i> ^{10b} <i>glgBXCAP</i>		^{10c} <i>malQ-glgBXCAP</i>		0			^{10d} <i>malQ</i> ^{10e} <i>glgB</i> ^{10f} <i>glgX</i> ^{10g} <i>glgC</i> ^{10h} <i>glgA</i>		
<i>acpP-fabF</i>	^{11a} <i>acpP-fabF</i>			^{11b} <i>acpP</i>				^{11a} <i>acpP-fabF</i>		
<i>narQP</i>	^{12a} <i>narP</i>			^{12a} <i>narP</i> ^{12b} <i>narQ</i>				^{12c} <i>narQP</i>		

генах *Yersinia* spp., однако потенциальных сайтов перед ним не обнаружено. Поскольку гены оперонов *nio* и *nqr* негомологичны, в данном случае имеет место негомологичная замена с частичным изменением функции.

Для *E. coli* известно два оперона, кодирующих комплексы редуктазы триметил-оксида азота, *torCAD* и *torYZ* [352, 353]. Эти опероны, по всей видимости, являются следствием относительно недавней дупликации, о чем свидетельствует достаточно высокий уровень попарного сходства между белками TorC–TorY и TorA–TorZ. Экспрессия оперона *torCAD* репрессируется белком NarL и активируется локальным регулятором транскрипции TorR [174, 239], тогда как транскрипция оперона *torYZ* осуществляется постоянно на низком уровне [352]. Однако, в исследованных геномах наблюдается иная ситуация. Так, в *Yersinia* spp. не было найдено ортологов для *tor*-генов. В трех из четырех исследованных геномах оперон *torCAD* отсутствует, а потенциальные сайты для всех регуляторов обнаруживаются перед *torYZ*. В геноме же *P. multocida* наблюдается обратная ситуация – оперон *torYZ* отсутствует, а сайты для всех трех регуляторов найдены перед *torCAD*. В геномах Vibrionales потенциальные сайты связывания присутствуют перед *torYZ*, тогда как оперон *torCAD* распадается на два, *torCA* и *torD*, и перед *torD* обнаруживаются потенциальные NarP-сайты. Таким образом, для данных дублицированных генов характерна регуляция как минимум одной копии.

Сходная ситуация наблюдается и для формиат редуктаз. В *E. coli* данные ферментативные комплексы кодируются двумя гомологичными оперонами, *fdnGHI* и *fdoGHI*, причем транскрипция оперона *fdn* регулируется факторами Fnr, NarL, и NarP [195, 341], тогда как *fdo* экспрессируется конститутивно [354]. Поскольку определение ортологов для генов из этих оперонов вызвало определенные трудности, для гомологов Fdn и Fdo были построены филогенетические деревья (Рис. 5.16). Как следует из анализа данных деревьев, исследованные *Yersinia* spp. содержат ортологи *fdo*-генов. Продукты генов оперона в Pasteurellales являются одинаково близкими и к Fdn, и к Fdo, поэтому данному оперону Pasteurellales было присвоено название *fdx*. В геномах исследованных Vibrionales не было найдено ортологов для данных генов. Скорее всего, у Pasteurellales сохранилось предковое состояние, и оперон претерпел дупликацию у общего предка Enterobacteriales. В геномах *Yersinia* spp. перед опероном *fdo* были обнаружены потенциальные сайты для всех исследованных регуляторов. Регуляция *fdx*-оперона Pasteurellales, по всей видимости, осуществляется белками Fnr и ArcA. Таким образом, можно предположить, что как минимум одна копия оперона формиат редуктазы должна регулироваться соответствующими факторами транскрипции.

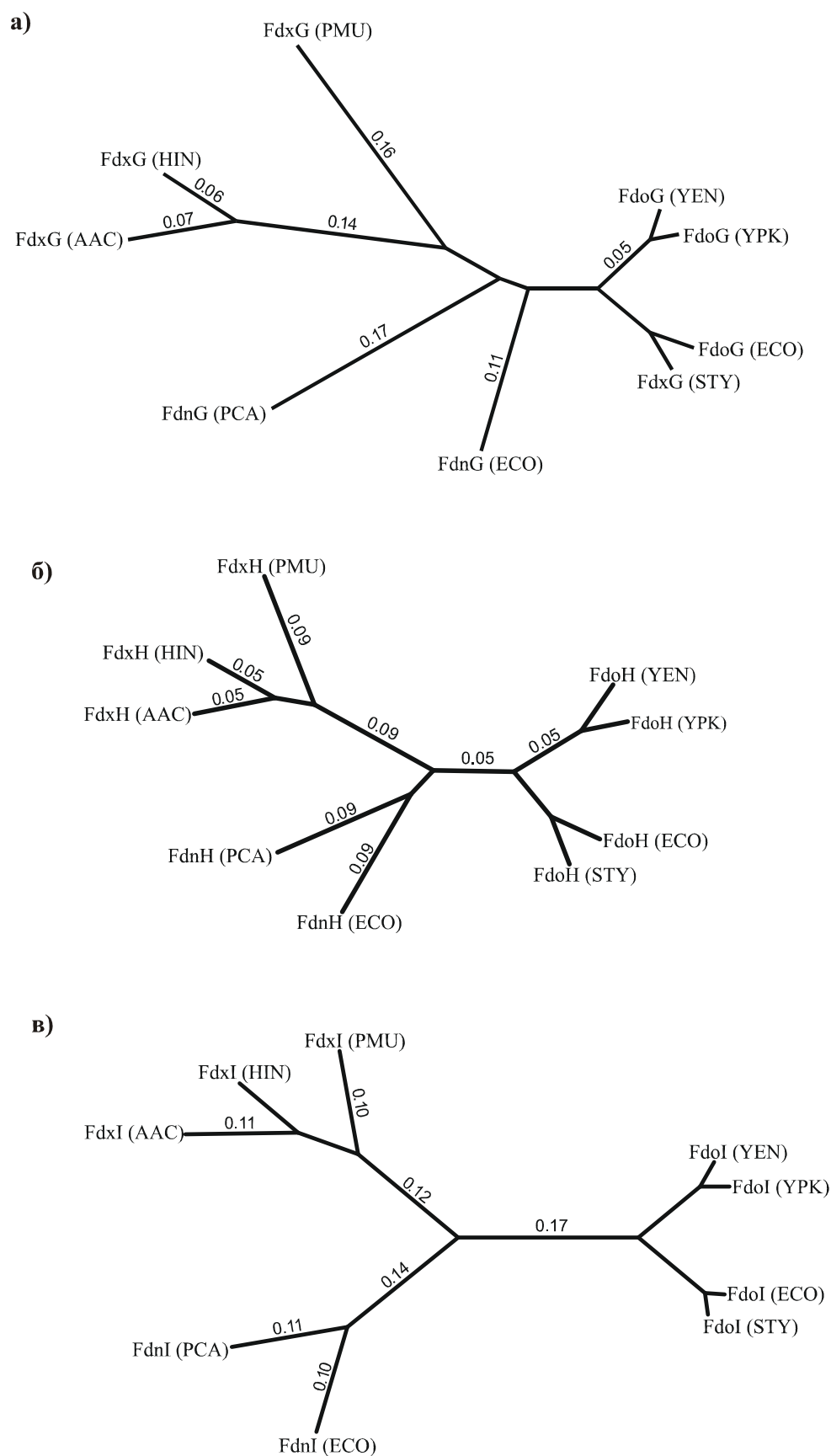


Рис. 5.16. Филогенетические деревья белков Fdn и Fdo

(а) FdnG/FdoG; (б) FdnH/FdoH; (в) FdnI/FdoI. Для построения использовался метод объединения соседних пар (neighbour-joining). Числа на ветвях обозначают ожидаемую долю аминокислотных замен (показаны только доли более 0,05). В скобках указаны условные обозначения геномов (см. Табл. 2.1).

Потенциальные сайты связывания ArcA были обнаружены перед геном *fdhD* в геномах *Yersinia* spp. и Pasteurellales. Кроме того, перед данным опероном в *Yersinia* spp. были обнаружены также сайты для NarP. Ген *fdhD* кодирует белок, необходимых для организации белковых комплексов Fdn и Fdo, но при этом не входит в состав данного комплекса [355, 356]. Примечательно, что регуляция данного гена схожа с таковой для оперонов формиаат дегидрогеназ.

В геномах *Yersinia* spp. и Vibrionales потенциальные сайты связывания ArcA были обнаружены перед опероном *dadAX*, первый ген которого кодирует малую субъединицу мембрано-ассоциированной дегидрогеназы D-аминокислот [357]. Таким образом, мы можем утверждать, что использование аминокислот в качестве доноров электронов регулируется с помощью ArcA в некоторых из исследованных геномов.

5.3.2. Биосинтез молибдоптеринового кофактора

Молибдоптериновый кофактор является необходимым компонентом ряда дегидрогеназ и оксидоредуктаз, таких как комплексы Fdn, Fdo, Nap, Dms, и Tor [168], что объясняет регуляцию генов ферментов его синтеза различными глобальными регуляторами дыхания. Так, ранее было показано, что в *E. coli* экспрессия оперона *moaABCDE* контролируется фактором Fnr [358], тогда как транскрипция оперона *moeAB* регулируется белками Fnr и ArcA [199]. В настоящей работе потенциальные сайты связывания как минимум для одного регулятора были обнаружены перед опероном *moa*, причем регуляция носит ярко выраженный таксон-специфичный характер. Так, для *Yersinia* spp. характерна Fnr-зависимая регуляция, в геномах Pasteurellales были найдены потенциальные сайты для Fnr и NarP, а в Vibrionales были обнаружены предполагаемые NarP-сайты. В случае же оперона *moe* консервативные сайты связывания Fnr и ArcA были найдены только у представителей Pasteurellales.

5.3.3. Центральный метаболизм и брожение

Глобальные регуляторы дыхания, в дополнение к генам белков дыхательных цепей, также контролируют экспрессию целого ряда оперонов, необходимых для осуществления реакций центрального метаболизма. Так, ранее была экспериментально установлена регуляция оперонов *pdhR-aceEF-lpdA* [271], *pflB* [235], *yfiD* [203], *gltA* [359], *sucABCD*, *sdhCDAB* [360], *aspA* [361], *fumC* [362], *mdh* [363], *ldhA* [364] и *adhE* [365, 366]). В настоящей работе регуляция с помощью факторов Fnr, ArcA и NarP была предсказана также для ряда генов, вовлеченных в процессы гликолиза, глюконеогенеза, пентоз-фосфатного пути и

метаболизма пирувата и лактата. Подробнее предсказанная регуляция этих генов описана ниже.

Ген *sfcA* кодирует малат дегидрогеназу, фермент, участвующий в процессе глюконеогенеза [367]. Предполагаемые сайты связывания NarP были обнаружены перед этим геном в *Yersinia* spp., тогда как сайты для Fnr были найдены у представителей Vibrionales.

Другой фермент глюконеогенеза, фосфоенолпируват синтаза, кодируется геном *ppsA* [368]. Потенциальные Fnr-сайты перед этим геном были найдены в *Yersinia* spp., для Vibrionales была же предсказана ArcA-зависимая регуляция.

Регуляция глобальными регуляторами дыхания была предсказана для еще одного гена глюконеогенеза, *pckA*, кодирующего фосфоенолпируват карбоксикиназу [316]. Потенциальные сайты связывания ArcA перед этим геном были обнаружены у всех представителей Vibrionales и Pasteurellales, за исключением *P. multocida*.

В настоящей работе была предсказана регуляция двух генов, продукты которых участвуют в процессах гликолиза: *eno*, кодирующего фермент енолазу, и *pgk*, продуктом которого является фосфоглицерат киназа [279]. В случае гена *eno* наблюдается таксон-специфичная регуляция – консервативные Fnr- и NarP-сайты были обнаружены в геномах Pasteurellales. В случае *pgk* потенциальные сайты для NarP были найдены у Pasteurellales, а сайты связывания Fnr – обнаружены в геномах Vibrionales.

Продуктом гена *talB* является трансальдолаза, фермент пентозфосфатного пути [369]. В геномах Pasteurellales перед данным геном были обнаружены потенциальные NarP-сайты.

Сайты связывания для факторов Fnr и ArcA были обнаружены перед геном *aldB* во всех геномах представителей Vibrionales. Этот ген кодирует фермент альдегид дегидрогеназу, вовлеченный в метаболизм лактата и пирувата [370].

5.3.4. Метаболизм углеводов

В настоящей работе консервативные потенциальные сайты связывания факторов Fnr, ArcA и NarP были обнаружены перед рядом генов метаболизма углеводов. К сожалению, не удалось обнаружить экспериментальных данных о влиянии глобальных регуляторов дыхания на экспрессию таких генов. Тем не менее, подобная регуляция представляется вполне возможной, поскольку метаболизм сахаров тесно связан с процессами центрального метаболизма [368]. Ниже приводится описание генов метаболизма углеводов, регуляция которых была предсказана в данной работе.

Гены *ptsH* и *ptsI* кодируют, соответственно, белки HPr и EI, являющиеся общими компонентами всех фосфотрансферазных систем [263, 265]. В геноме гены составляют единый оперон с геном *crr*, кодирующим ЕПА компонент глюкозо-специфичной

фосфотрансферазной системы [136]. Данная оперонная структура сохраняется во всех исследованных геномах. Потенциальные сайты связывания ArgA перед этим опероном были найдены у всех представителей Pasteurellales.

Оперон кодирует *mtIADR* компонент ЕПАВС маннитол-специфичной фосфотрансферазной системы, фермент маннитол-1-фосфат дегидрогеназу и регуляторный белок MtlR, представляющий собой локальный регулятор, репрессор данного оперона [263, 371]. ArgA-зависимая регуляция оперона *mtl* была предсказана для *Yersinia* spp., а Fng-зависимая – для представителей Vibrionales.

Потенциальные сайты связывания NarP были обнаружены перед опероном *deoCABD* в геномах Vibrionales. Гены этого оперона ответственны за деградацию пиримидиновых дезоксирибонуклеозидов до ацетальдегида и глицерол-3-фосфата [302].

Продуктом гена *nagB* является глюкозамин-6-фосфат деаминаза [372]. Предполагаемые сайты для Fng перед этим геном найдены во всех геномах Vibrionales.

Кластер *glgBXCAP* содержит гены, ответственные за осуществление различных реакций биосинтеза и катаболизма гликогена [373]. Структура кластера существенно отличается в геномах представителей различных таксонов (Табл. 5.5). Так, в *Yersinia* spp. все *glg*-гены находятся в составе одного оперона, перед которым были обнаружены потенциальные сайты Fng. В геномах Pasteurellales гены *glg* колокализированы с геном *malQ*, и, скорее всего, формируют единый оперон *malQ-glgBXCAP*. Перед геном *malQ* Pasteurellales были предсказаны сайты связывания Fng и NarP. Подобная консервативность сайтов при изменении оперонной структуры позволяет предположить, что именно регуляция *glg*-генов является необходимой для бактериальной клетки.

5.3.5. Метаболизм жирных кислот

К настоящему моменту времени экспериментальные данные о влиянии глобальных регуляторов дыхания на транскрипцию генов метаболизма жирных кислот отсутствуют. В настоящей работе была предсказана регуляция ряда генов, продукты которых участвуют в упомянутых метаболических процессах.

Два гомологичных оперона, *fabBA* и *fadIJ*, кодируют белковые комплексы, осуществляющие бета-окисление жирных кислот. Известно, что комплекс FadBA используется для деградации жирных кислот как в присутствии, так и в отсутствие молекулярного кислорода, тогда как комплекс FadIJ используется почти исключительно при анаэробных условиях [374, 375]. Ранее было экспериментально показано, что мутация в гене *arcA* ведет к увеличению уровня экспрессии оперона *fadBA*, но оставалось неясным, является ли данный эффект следствием прямой регуляции [205]. В настоящей работе в геномах

Yersinia spp. потенциальные ArcA-сайты были найдены перед опероном *fadIJ*, а сайты для Fnr и ArcA – перед опероном *fadBA*. У представителей Vibrionales наблюдалась противоположная ситуация: Fnr- и ArcA-сайты были найдены перед опероном *fadIJ*, тогда как перед опероном *fadBA* присутствовали сайты только для ArcA.

Потенциальные сайты связывания ArcA были найдены перед геном *fadD* в *Yersinia* spp. и Vibrionales. Данный ген кодирует ацил-КоА синтазу, белок, участвующий в процессах бета-окисления жирных кислот [376].

Еще одним опероном, для которого была предсказана регуляция, является *acpP-fabF*, кодирующий белки-компоненты синтазы жирных кислот [377]. Консервативные сайты связывания для ArcA были найдены перед этим опероном в *Yersinia* spp., тогда как Fnr-сайты были предсказаны в геномах Vibrionales.

Тем не менее, остается открытым вопрос о значении регуляции метаболизма жирных кислот с помощью глобальных регуляторов дыхания. В данном случае возможны два варианта для объяснения данного явления. Во-первых, метаболизм жирных кислот тесно связан с центральным метаболизмом, например, через ацетил-КоА [377, 378]. Во-вторых, использование тех или иных ферментов окисления жирных кислот зависит от того, в аэробных или анаэробных условиях происходят процессы окисления [375].

5.3.6. Ответ на кислородный стресс

К данной группе был отнесен лишь один ген *sodA*, кодирующий белок супероксиддисмутазу. Ранее была показана регуляция данного гена факторами транскрипции Fnr и ArcA [379]. Подобная регуляция была обнаружена у представителей Pasteurellales, где перед данным геном были найдены сайты для обоих белков Fnr и ArcA. В геномах *Yersinia* spp. были обнаружены сайты только для ArcA, тогда как у Vibrionales ортологов данного гена не было найдено.

5.3.7. Нуклеотидредуктазы

Ранее для оперона *nrdDG* в *E. coli* была экспериментально показана активация в анаэробных условиях при участии фактора Fnr. Также было показано отсутствие влияния Fnr на экспрессию другого оперона, кодирующего белки нуклеотидредуктазного комплекса, *nrdAB* [192]. В настоящей работе потенциальные сайты связывания Fnr перед опероном *nrdDG* были найдены во всех геномах *Yersinia* spp. и Vibrionales. В геномах Pasteurellales, напротив, Fnr-сайты были найдены перед опероном *nrdAB* во всех геномах, за исключением *H. ducreyi*, где отсутствуют гены *nrdDG*. Подобная ситуация свидетельствует о том, что в

случае нуклеотид редуктаз необходимо, чтобы как минимум одна копия оперона экспрессировалась конститутивно.

5.3.8. Транспортные белки

Ранее регуляция глобальными регуляторами была показана для целого ряда оперонов, кодирующих транспортные белки, таких как *focA* [235], *dcuA*, *dcuB* [361], *dcuC* [380], *glpFK* [381] и *feoAB* [382]. Помимо перечисленных выше оперонов, в настоящей работе была предсказана регуляция для ряда других генов транспортных белков.

Потенциальные сайты связывания ArcA были найдены перед геном *glpP* в геномах *Yersinia* spp. и Vibrionales. Продукт данного гена представляет собой транспортер глутамата и аспартата [361]. Данный случай регуляции дополняет уже имеющиеся знания о регуляции генов транспортеров дикарбоновых кислот глобальными регуляторами дыхания. Так, известно, что экспрессия генов *dcuA*, *dcuB* и *dcuC*, кодирующих транспортеры дикарбоновых кислот, в *E. coli* регулируется фактором транскрипции Fnr [361, 380].

Оперон *gntXY* кодирует белки системы транспорта глюконата [383]. Потенциальные сайты связывания ArcA и Fnr были обнаружены перед этим опероном во всех геномах Pasteurellales.

Регуляция была также предсказана для гена транспортера жирных кислот, *fadL* [384]. В геномах *Yersinia* spp. перед *fadL* были найдены потенциальные сайты для Fnr и ArcA, а в Vibrionales – сайты связывания ArcA.

Поскольку регуляция глобальными регуляторами дыхания была предсказана для многих генов метаболизма углеводов и жирных кислот (см. 5.3.6, 5.3.7), регуляция транспортеров для этих веществ представляется вполне целесообразной.

5.3.9. Пептидазы

К этой группе был отнесен единственный ген *pepT*, кодирующий аминопептидазу (пептидаза T). Экспериментально показано, что в геноме *Salmonella typhimurium* экспрессия этого гена положительно регулируется фактором транскрипции Fnr [385, 386]. Такая регуляция сохраняется во всех исследованных геномах, за исключением *H. ducreyi*, где ортологов данного гена найдено не было.

5.3.10. Регуляторы транскрипции

Ранее регуляция была экспериментально показано для таких генов, кодирующих регуляторные белки, как *fnr* [187], *arcA* [387], *narP* и *narQ* [224]. В дополнение к этому, в

настоящей работе регуляция была предсказана для нескольких других генов, кодирующих белки регуляторных систем.

Fis – это белок, регулирующий уровень сверхспирализации ДНК в клетке и участвующий в регуляции экспрессии большого количества генов [224, 388]. Экспериментально установлено, что в *E. coli* Fis совместно с глобальными регуляторами дыхания Fnr и/или ArcA контролирует экспрессию таких оперонов, как *nrf* ([193]), *nir* ([146]), *ndh* [201], *adhE* [366], *yfiD* [202], *sdhCDAB-sucABCD* [202], *acnB* [259] и *narK* [232]. В *E. coli* ген *fis* транскрибируется совместно с геном *dusB* [389] и структура генного кластера *dusB-fis* сохраняется во всех исследованных геномах. В Pasteurellales перед опероном *dusB-fis* были обнаружены консервативные потенциальные сайты ArcA, что позволяет считать, что в данных геномах экспрессия *fis* находится под контролем регуляторов дыхания.

Оперон *spxRA* кодирует белки двухкомпонентной системы, осуществляющей ответ на внешний стресс [390]. Некоторые данные указывают на существование кросс-взаимодействий между системами SpxA-SpxR и ArcB-ArcA в , а именно *E. coli*, на взаимодействия между сенсорным белком SpxA и фактором транскрипции ArcA [391]. В настоящей работе потенциальные сайты для Fnr и ArcA были обнаружены перед опероном *spxRA* в геномах Pasteurellales.

Белок OxyR – это фактор транскрипции, регулирующий ответ на окислительный стресс [392]. Во всех геномах Pasteurellales были обнаружены потенциальные сайты связывания ArcA перед геном *oxyR*. Поскольку ArcA активен как регулятор транскрипции только в условиях отсутствия кислорода, можно предположить, что в данных условиях он репрессирует транскрипцию гена *oxyR*, поскольку в таких условиях сенсора кислорода не требуется. Следует заметить, однако, что в данном случае логика такой регуляции не совсем ясна.

В геномах Pasteurellales консервативные Fnr-сайты были найдены перед геном *fur*. В свою очередь, в геномах Vibrionales перед этим геном были обнаружены потенциальные сайты связывания ArcA. Продуктом гена *fur* является фактор транскрипции, контролирующей экспрессию генов, ответственных за импорт в клетку ионов железа [393]. Подобная регуляция представляется вполне естественной, поскольку железо входит в состав многих оксидоредуктаз и дегидрогеназ [168], а также активной формы белка Fnr [394].

5.4. Таксон-специфические особенности глобальной регуляции дыхания

5.4.1. Состав обобщенных регулонов в различных таксонах

Приведенный выше анализ регуляции показал существенные различия между различными порядками гамма-протеобактерий. Чтобы оценить относительную роль каждого

регулятора в регуляции в пределах отдельного таксона, было посчитано количество генов и оперонов, являющимися членами каждого регулона в каждом порядке. Как и ожидалось, состав обобщенных регулонов сильно варьировал между таксонами.

Так, в *Yersinia* spp. обобщенный регулон NarP оказался намного меньше, чем в других таксонах (Рис. 5.17а) что неудивительно, если принять во внимание, что в *Yersinia* spp. организмах система нитрат-нитритного дыхания сильно редуцирована, и в качестве акцептора электронов может использоваться только нитрат, но не нитрит (см. 5.1.3). Соответственно, поскольку для *Yersinia* spp. характерно уменьшение роли нитрат-нитритного дыхания, то и система регуляции, отвечающая на присутствие в среде нитрата и нитрита, также будет играть много меньшую роль, чем в других таксонах.

Обратная ситуация наблюдается для Pasteurellales, где обобщенный регулон NarP увеличен по сравнению с другими таксонами. Другой отличительной чертой этого порядка является высокая степень перекрывания между обобщенными регулонами для разных регуляторов. Например, консервативные сайты для обоих факторов Fnr и NarP были найдены перед 13 оперонами, в общей сложности содержащими 53 гена (Рис. 5.17б).

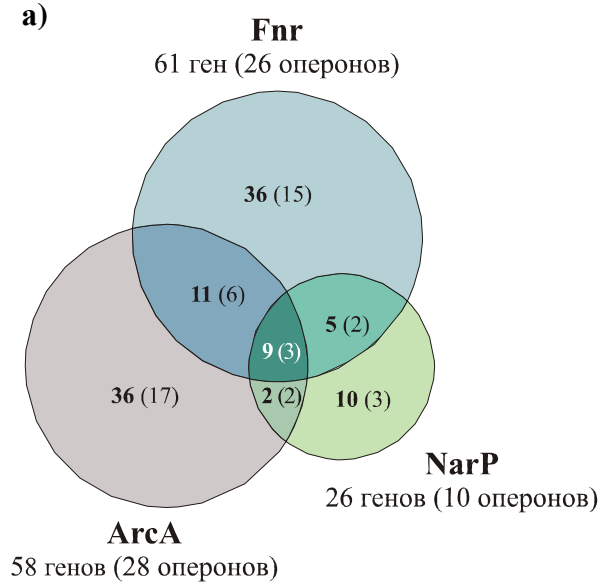
Характерной чертой Vibrionales является увеличение обобщенного ArcA по сравнению с другими группами. У Vibrionales консервативные сайты связывания ArcA были обнаружены перед 47 оперонами, включающими 99 генов (Рис. 5.17в).

5.4.2. Структура регуляторных каскадов в разных таксонах

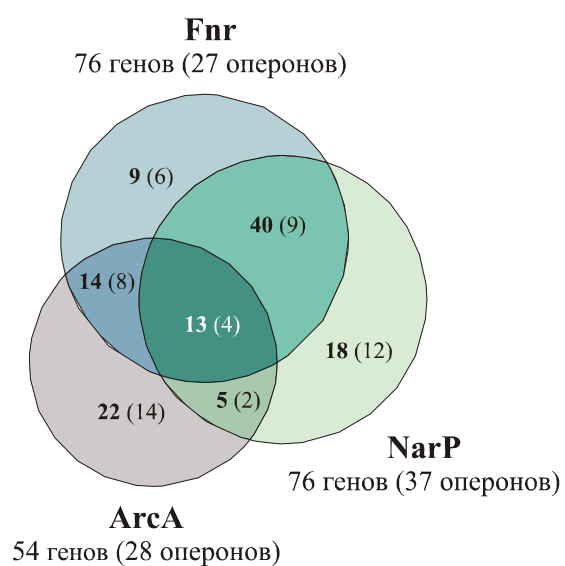
Одним из наиболее важных и при этом неожиданных результатов настоящей работы стало предсказание ряда новых регуляторных каскадов. Предсказания каскадов были сделаны на основе наличия консервативных предсказанных сайтов перед генами *fnr*, *arcA*, и *narP*.

Для *E. coli* регуляторные каскады были известны из экспериментальных работ (Рис. 5.18а). В данном случае Fnr представляет собой главный регулятор дыхания, контролирующей экспрессию генов других регуляторов: Fnr репрессирует транскрипцию оперона *narXL* [187] и регулирует транскрипцию гена *arcA* [187, 387]. Кроме того, сам ген *fnr* является объектом негативной авторегуляции [187]. Для гена *arcA* была предсказана позитивная авторегуляция [330]. Регуляторы нитрат-нитритного дыхания формируют сложную сеть регуляторных взаимодействий, где NarL активирует экспрессию оперонов *narXL*, *narP* и *narQ*, а NarP, в свою очередь, активирует транскрипцию *narXL* [224].

а)



б)



в)

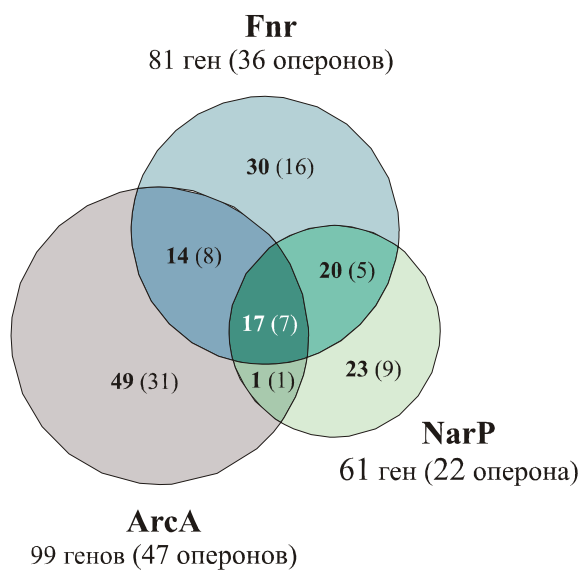


Рис. 5.17. Состав обобщенных регулонов в трех таксонах
(а) *Yersinia* spp.; (б) Pasteurellales; (в) Vibrionales. Указано количество генов и оперонов (в скобках), являющихся потенциальными членами обобщенных регулонов.

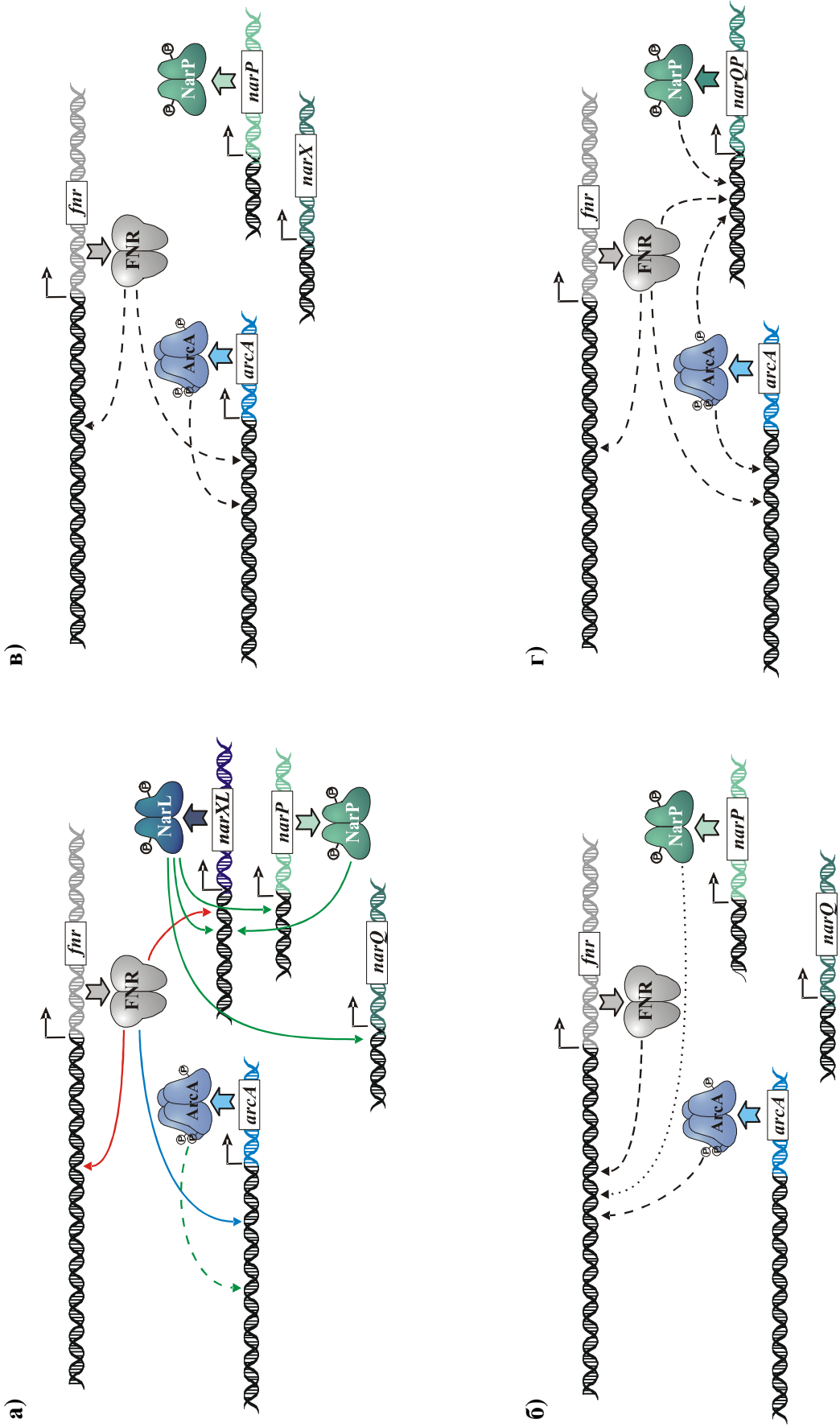


Рис. 5.18. Регуляторные каскады в различных таксонах

(а) *Escherichia coli*; (б) *Yersinia* spp.; (в) Pasteurellales; (г) Vibrionales. Экспериментально выявленная регуляция показана непрерывными стрелками: зеленым – активация, красным – репрессия, синим – амбивалентная активация. Предсказанная регуляция обозначена пунктирными стрелками: зеленым – активация, черным – тип регуляции не определен. Точечной стрелкой обозначено отсутствие данной регуляции у *H. ducreyi*.

Однако, в геномах других гамма-протеобактерий наблюдается иная ситуация. В *Yersinia* spp. Fnr, по всей видимости, регулирует экспрессию *arcA*, и оба гена *fnr* и *arcA* авторегулируются (Рис. 5.18б). Перед генами *narP* и *narX* не было обнаружено потенциальных сайтов для какого-либо из исследованных регуляторов. Таким образом, в *Yersinia* spp. Fnr, как и в *E. coli*, является главным регулятором дыхания. При этом система регуляции нитрат-нитритного дыхания не принимает участия в регуляторных каскадах.

Более значительные изменения в системе регуляторных каскадов были обнаружены в геномах Pasteurellales. В геномах большинства представителей этого порядка потенциальные сайты связывания Fnr, ArcA и NarP были обнаружены перед геном *fnr* (Рис. 5.18в), тогда как перед генами *arcA* и *narP* не было найдено консервативных сайтов. Таким образом, в Pasteurellales роль Fnr в регуляторных каскадах уменьшается по сравнению с *E. coli*. Небольшие отличия от других Pasteurellales наблюдаются в геноме *H. ducreyi*, где не было обнаружено потенциальных сайтов связывания NarP перед геном *fnr* (Рис. 5.18в).

В случае Vibrionales также была предсказана авторегуляция для генов *fnr*, *arcA* и *narP*. Белок ArcA, по всей видимости, регулирует экспрессию гена *narP*, тогда как Fnr контролирует экспрессию генов *arcA* и *narP* (Рис. 5.18г). Следовательно, в данном таксоне роль белка ArcA в регуляторных каскадах возрастает по сравнению с *E. coli*.

Таким образом, структура регуляторных каскадов, характерная для *E. coli*, не сохраняется в других гамма-протеобактериях, и относительная роль каждого регулятора в каскадах может меняться от таксона к таксону. Возможно, что увеличение числа полных геномов позволит более детально проследить эволюцию регуляторных каскадов.

5.4.3. Регуляция в отдельных таксонах

Помимо различий в структуре регуляторных каскадов, в настоящей работе были также обнаружены таксон-специфические различия в функциональном составе обобщенных регулонов (Рис. 5.19).

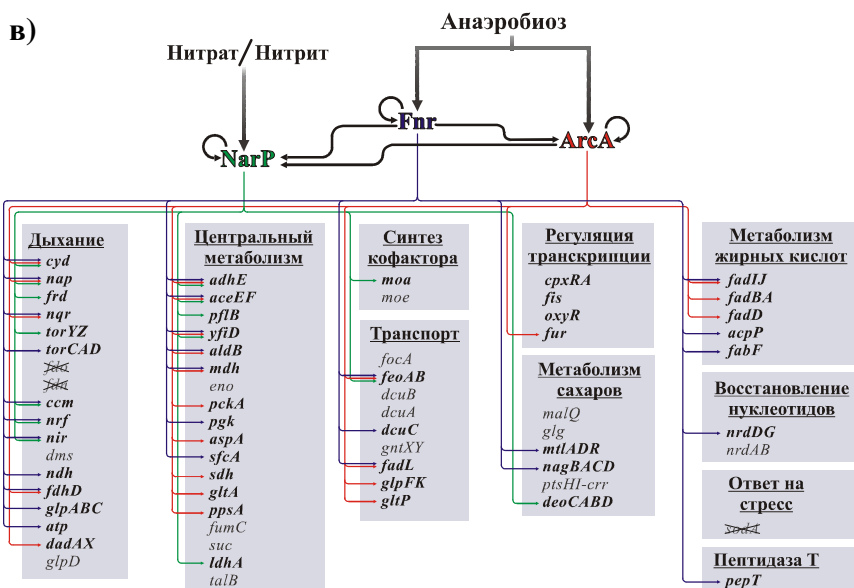
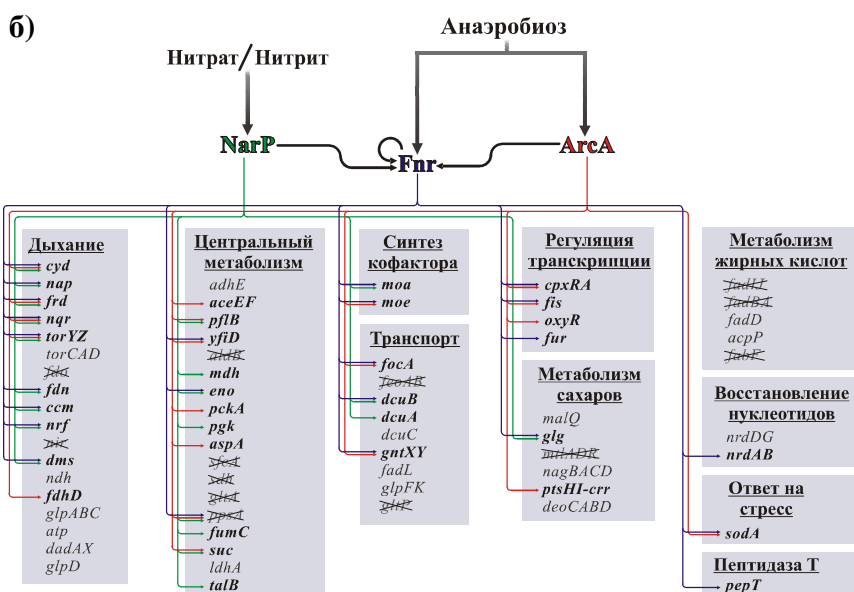
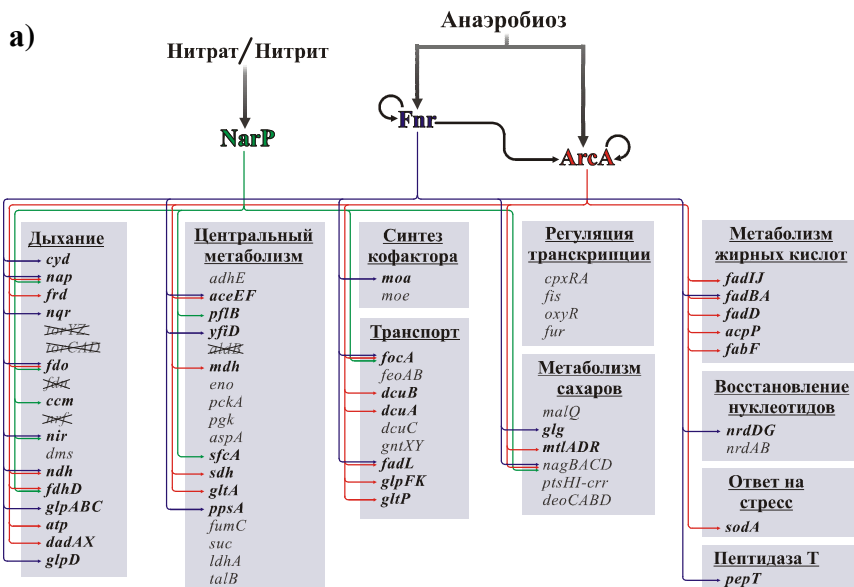


Рис. 5.19. Регуляторные взаимодействия в различных таксонах

(а) *Yersinia* spp.; (б) Pasteurellales; (в) Vibrionales. Регуляторные каскады показаны толстыми черными стрелками. Предполагаемая регуляция транскрипции показана тонкими стрелками: синими для Fnr, красными для ArcA и зелеными для NarP.

Два генома *Yersinia* spp. были выбраны в качестве объектов данной работы ввиду наличия у них одиночной системы регуляции нитрат-нитритного дыхания. Эти бактерии демонстрируют ряд существенных отличий как от *E. coli*, так и от других изученных таксонов (Рис. 5.19а). Наиболее существенной чертой этих организмов является наличие нестардантной пары NarX-NarP, что отличает ее от других исследованных бактерий, имеющих пару NarQ-NarP. Другой отличительной особенностью организмов из данной группы является значительное упрощение системы регуляторных каскадов. Хотя характерный для *E. coli* каскад Fnr-ArcA сохраняется в *Yersinia* spp., регуляторы нитрат-нитритного дыхания не принимают участия в формировании каскадов. Данная особенность коррелирует и со структурой обобщенного NarP-регулона: потенциальные сайты для этого регулятора были найдены только перед генами, необходимыми для нитратного дыхания с использованием формиата в качестве донора электронов (*nar*, *ccm*, *fdo*, *fdhD*) и опероном *nir*, необходимым для защиты клетки от токсичного действия нитрита, продукта нитратного дыхания. В то же время обобщенные Fnr- и ArcA-регулоны достаточно похожи на таковые в других таксонах.

В случае Pasteurellales наблюдаются существенные изменения в структуре регуляторных каскадов. Так, в отличие от других таксонов, в данной группе организмов Fnr не является главным регулятором дыхания (см. 5.4.2). Другой ярко выраженной особенностью является отсутствие гена сенсорного белка ArcB. Подобная ситуация наблюдается и в другой гамма-протеобактерии, *Shewanella oneidensis* MR-1 [333]. Поскольку в Pasteurellales не наблюдается редукции ArcA-регулона, то, следует предположить, что роль сенсора выполняет какой-либо другой белок. Не исключено, что эту роль выполняет белок SrxA, являющийся близким гомологом ArcB. Регуляция оперона *srxRA* факторами Fnr и ArcA в данном таксоне также указывает на такую возможность.

Специфической особенностью Pasteurellales является также наличие дополнительных регуляторных каскадов более низкого уровня. Ряд генов регуляторных систем, таких как *srxRA*, *fis*, *oxyR* и *fur*, имеют консервативные сайты. Все эти опероны, за исключением *fur*, не имеют потенциальных сайтов в геномах из других таксонах. Возможно, что Pasteurellales претерпели в ходе эволюции существенную перестройку регуляторных каскадов с участием регуляторов дыхания, что потребовало возникновения дополнительных регуляторных взаимодействий для более тонкой настройки регуляции дыхания.

Увеличение роли фактора NarP в регуляторных каскадах соответствует расширению обобщенного NarP-регулона (Рис. 5.19б). Подобное расширение может происходить различными путями. Например, перед оперонами *cyd* и *nqr* в *Pasteurellales* NarP-сайты возникают в дополнение к существующим сайтам для Fnr и ArcA. В случае оперонов *mdh*, *eno* и *suc* наблюдается смена регулятора: перед этими оперонами сайты связывания NarP обнаруживаются вместо Fnr- или ArcA-сайтов, наблюдаемых в других таксонах. Кроме того, NarP-сайты перед генами *eno* и *talB* появляются в *Pasteurellales*, тогда как в других геномах перед этими генами не обнаружено регуляторных сайтов для ни одного из трех изученных регуляторов.

В геномах *Vibrionales* Fnr сохраняет роль главного регулятора, но при этом возрастает роль ArcA в регуляторных каскадах. Подобные перестройки каскадов соответствуют расширению обобщенного ArcA-регулона, в основном за счет появления сайтов связывания ArcA перед генами центрального метаболизма (Рис. 5.19в).

5.5. Обсуждение и выводы

К настоящему моменту физиология и регуляция дыхания гамма-протеобактерий представляют собой довольно хорошо изученную область. Тем не менее, экспериментальные исследования проводились только для *E. coli*, за редкими исключениями, посвященными регуляции отдельных генов таких бактерий, как *H. influenzae* [395], *V. cholerae* [396] и *S. oneidensis* [333]. В исследованиях такого рода преобладают два типа подходов. Первый подход основан на изучении регуляции отдельного оперона различными факторами транскрипции. Второй подход основан на изучении регуляции множества оперонов одним регулятором в отдельном геноме. Исследование регуляции дыхания методами сравнительной геномики проводилось только лишь для отдельных регуляторов и для малого количества геномов [329, 330], и лишь показало слабую консервативность отдельных регулонов в эволюции.

В настоящей работе был проведен анализ глобальной регуляции дыхания для большого количества геномов, причем были проанализированы сразу три функционально сходных, но структурно различающихся регуляторных систем. Помимо одновременного исследования нескольких регуляторных систем, данная работа также отличается использованием усовершенствованной методики, основанной на сравнении всех геномов из одного таксона. Одним из важнейших преимуществ данной методики является возможность ее применения для исследования геномов, для которых отсутствуют экспериментальные данные. Кроме того, используемая методика позволяет предсказывать таксон-специфические регуляторные

взаимодействия. Таким образом, появление новых полных геномов не только количественно, но и качественно расширяет возможности сравнительной геномики.

В настоящей работе был предсказан целый ряд новых членов регулонов, кодирующих белки дыхательных цепей (*aptIBEFHAGDC*, *torYZ*, *fdoGHI*, *nqrABCDEFG*, *dadAX*), ферменты центрального метаболизма (*pgk*, *eno*, *aldB*, *pckA*, *ppsA*, *sfcA*, *talB*), метаболизма углеводов (*mtlADR*, *nagBACD*, *ptsHI-crr*, *deoCABD*, *glgBXCAP*), синтеза и катаболизма жирных кислот (*fadIJ*, *fadAB*, *fadD*, *acpP-fabF*), нуклеотидредуктаз (*nrdAB*), транспортных белков (*glpP*, *gntXY*, *fadL*) и факторов транскрипции (*dusB-fis*, *cpxRA*, *oxyR*, *fur*).

В данной работе, помимо ставшего уже стандартным для сравнительной геномики предсказания новых членов обобщенных регулонов, были также предсказаны новые регуляторные каскады, структура которых оказалась таксон-специфичной. При этом были замечены корреляции между структурой регуляторного каскада и составом обобщенного регулона: увеличение относительной роли белка-регулятора в каскаде сопровождается ростом соответствующего регулона, и наоборот. Упрощение структуры каскадов, в свою очередь, компенсируется увеличением степени перекрытия регулонов, что обеспечивает взаимодействие нескольких регуляторов. Таким образом, стабильность метаболической и регуляторной систем может достигаться различными способами. При этом регулируемые гены можно подразделить на ядро регулона, которое стабильно регулируется во всех геномах, и периферию, регуляция которой варьирует от таксона к таксону. Подобные наблюдения были сделаны и для других систем с множественной регуляцией, таких как метаболизм оксидов азота [23] или поддержание гомеостаза железа [25]. Таким образом, можно предположить, что существование ядра регулона, таксон-специфической периферии и перестройки регуляторных каскадов являются общими свойствами всех систем с комплексной регуляцией.

К сожалению, в настоящей работе не удалось обнаружить корреляций между особенностями регуляции дыхания и образом жизни исследованных микроорганизмов.

Выводы

1. Построено распознающее правило для поиска потенциальных сайтов связывания белка фруктозного репрессора FruR и предсказано 26 новых членов обобщенного FruR-регулона.
2. Построена модель эволюции обобщенного FruR-регулона, показана эволюция от локального регулятора оперона *fruVKA* до глобального регулятора метаболизма сахаров.
3. Построены распознающие правила для поиска потенциальных сайтов белков PurR, RbsR и RbsR Pseudomonadales и предсказано 25 новых генов, входящих в обобщенный регулон PurR.
4. Построена модель эволюции, в соответствии с которой PurR и RbsR являются результатом дупликации предкового белка, репрессора рибозного оперона, с последующим изменением функции PurR и структуры мотива связывания RbsR.
5. Построены распознающие правила для поиска потенциальных сайтов связывания факторов транскрипции Fnr, ArgA и NarP, разработана и применена методика анализа комплексной регуляции в родственных геномах. Предсказано 67 новых членов обобщенных регулонов.
6. Предсказаны таксон-специфичные регуляторные каскады, показаны корреляции между перестройкой структуры регуляторных каскадов и размером обобщенного регулона: регуляторы, занимающие относительно более высокое положение в каскаде, имеют большие регулоны.
7. Для глобальной регуляции дыхания показано наличие консервативных ядер регулонов и таксон-специфичной периферической регуляции.

Список работ, опубликованных по теме диссертации

Публикации в научных журналах

1. Равчеев Д.А., Гельфанд М.С., Миронов А.А., Рахманинова А.Б. **Пуриновый регулон гамма-протеобактерий. Детальное описание.** *Генетика*. 2002. **38 (9)**, 1203-1214.
2. Favorov A.V., Gelfand M.S., Gerasimova A.V., Ravcheev D.A., Mironov A.A., Makeev V.J. **A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length.** *Bioinformatics*. 2005. **21**, 2240-2245.
3. Равчеев Д.А., Рахманинова А.Б., Миронов А.А., Гельфанд М.С. **Регуляция нитрат-нитритного дыхания гамма-протеобактерий. исследование методами сравнительной геномики.** *Молекулярная биология*. 2005. **39 (5)**, 832-846.
4. Ravcheev D.A., Gerasimova A.V., Mironov A.A., Gelfand M.S. **Comparative genomic analysis of regulation of anaerobic respiration in ten genomes from three families of gamma-proteobacteria (Enterobacteriaceae, Pasteurellaceae, Vibrionaceae).** *BMC Genomics*. 2007. **8 (1)**. 54.
5. Цыганова М.О., Гельфанд М.С., Равчеев Д.А. **Регуляция дыхания у энтеробактерий: сопоставление данных по экспрессии на биочипах и сравнительно-геномного анализа.** *Молекулярная биология*, 2007, **41 (3)**, 556-571.

Публикации в сборниках трудов конференций

1. Ravcheev D.A., Gelfand M.S., Mironov A.A., Rakhmaninova A.B. **Purine regulon of gamma-proteobacteria.** *Proceedings of The Third International Conference on Bioinformatics of Genome Regulation and Structure*. 2002. **v. 2**. 38-39.
2. Равчеев Д.А., Рахманинова А.Б. **NarP-зависимая регуляция нитрат-нитритного дыхания у гамма-протеобактерий. Исследование эволюции дыхательной системы методами сравнительной геномики.** *Сборник тезисов XI Международной научной конференции студентов, аспирантов и молодых ученых «Ломоносов-2004»*. 2004. **т. 1**. с. 27.
3. Gerasimova A.V., Ravcheyev D.A., Gelfand M.S., Rakhmaninova A.B. **Comparative genomic analysis of respiration switch in gamma-proteobacteria.** *Proceedings of The Fourth International Conference on Bioinformatics of Genome Regulation and Structure*. 2004. **v.2**. 195-198.
4. Равчеев Д.А. **Исследование регуляции нитрат-нитритного дыхания гамма-протеобактерий методами сравнительной геномики.** *Сборник тезисов XII Международной конференции студентов, аспирантов и молодых ученых «Ломоносов»*. 2005. **т. 2**. с. 34-35.
5. Ravcheev D.A., Gerasimova A.V. **How gamma-proteobacteria switch the mode of respiration: A comparative genomic analysis.** *Proceedings of the International Moscow Conference on Computational Molecular Biology*. 2005. **p.** 321-322.

6. Tsiganova M., Ravcheev D.A. **Regulation of respiration in *Escherichia coli*: the comparison of microarray and genomics data.** *Proceedings of the International Moscow Conference on Computational Molecular Biology*. 2005. p. 401-402.
7. Zakirzianova V., Ravcheev D.A. **A comparative genomic analysis of an evolving of regulatory system: how FruR became Cra.** *Proceedings of the International Moscow Conference on Computational Molecular Biology*. 2005. p. 423-424.
8. Равчеев Д.А. **Исследование регуляции нитрат-нитритного дыхания гамма-протеобактерий методами сравнительной геномики.** *Материалы международной школы «Биоинформатика, геномика, протеомика»*. 2006, с. 48-51.
9. Закирзянова В.В., Равчеев Д.А. **Исследование обобщенного FruR-регулона гамма-протеобактерий методами сравнительной геномики.** *Материалы XIII международной конференции студентов, аспирантов и молодых ученых “Ломоносов”*. 2006. т. 4, с. 45-46.
10. Цыганова М.О., Равчеев Д.А. **Регуляция дыхания у энтеробактерий: сопоставление данных по анализу экспрессии на биочипах и сравнительно-геномного анализа.** *Материалы XIII международной конференции студентов, аспирантов и молодых ученых “Ломоносов”*. 2006. т. 4, с. 39-40.
11. Ravcheev D.A. **Regulation of respiration in gamma-proteobacteria: evolution of multiple regulatory systems.** *Fourth Bertinoro Computational Biology Meeting (BCB 2006)*.
12. Tsoy O.V., Zakirzianova W., Ravcheev D.A. **Stories about the evolution of regulators: how FruR became CRA and RbsR became PurR.** *Proceedings of the 3rd Moscow Conference on Computational Molecular Biology*, 2007, p. 300.
13. Цой О.В., Закирзянова В., Равчеев Д.А. **Как PurR стал Cra, а RbsR стал PurR: истории из жизни бактериальных факторов транскрипции.** *Сборник трудов 30-й конференции молодых ученых и специалистов ИППИ РАН “Информационные технологии и системы ИТиС'07”*, 2007, с. 303-306.
14. Фаворов А., Гельфанд М., Герасимова А., Равчеев Д., Кулаковский И., Миронов А., Макеев В. **Алгоритм SeSiMCMC для поиска участков специфического связывания белков – регуляторов транскрипции.** *Сборник трудов 30-й конференции молодых ученых и специалистов ИППИ РАН “Информационные технологии и системы ИТиС'07”*, 2007, с. 334-337.
15. Равчеев Д.А. **Комплексная регуляция дыхания бактерий: исследование методами сравнительной геномики.** *Сборник трудов конференции молодых ученых и специалистов ИППИ РАН “Информационные технологии и системы-2008”*, 2008, с. 273-276.
16. Цыганова М., Равчеев Д. **Регуляция дыхания бактерий: транскриптомика и сравнительная геномика.** *Сборник трудов конференции молодых ученых и специалистов ИППИ РАН “Информационные технологии и системы-2008”*, 2008, с. 332-336.

Благодарности

Автор выражает искреннюю благодарность Михаилу Сергеевичу Гельфанду за чуткое научное руководство, Андрею Александровичу Миронову и Александру Владимировичу Фаворову за любезно предоставленное программное обеспечение, Анне Викторовне Герасимовой, Ольге Цой, Марине Цыгановой и Виоланте Закирзяновой за помощь в выполнении работы и Александре Борисовне Рахманиновой, Дмитрию Александровичу Родионову, Алексею Геннадиевичу Витрещаку и Ольге Николаевне Лайковой за ценные советы и полезное обсуждение.

Список литературы

1. Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.F., Dougherty B.A., Merrick J.M., McKenney K., Sutton G.G., FitzHugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.I., Glodek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D., Utterback T., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S., Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C. **Whole-genome random sequencing and assembly of *Haemophilus influenzae* Rd.** *Science*. 1995, **269**, 5223; 496-512.
2. Kanehisa M., Goto S. **KEGG: kyoto encyclopedia of genes and genomes.** *Nucleic Acids Res.* 2000, **28**, 1; 27-30.
3. Kanehisa M., Goto S., Hattori M., Aoki-Kinoshita K.F., Itoh M., Kawashima S., Katayama T., Araki M., Hirakawa M. **From genomics to chemical genomics: new developments in KEGG.** *Nucleic Acids Res.* 2006, **34**; D354-357.
4. Гельфанд М.С. **Апология биоинформатики.** *Биофизика*. 2005, **50**, 4; 752-766.
5. Altschul S.F., Madden T.L., Schaffer A.A., Zhang J., Zhang Z., Miller W., Lipman D.J. **Gapped BLAST and PSI-BLAST: a new generation of protein database search programs.** *Nucleic Acids Res.* 1997, **25**, 17; 3389-3402 .
6. Thompson J.D., Gibson T.J., Plewniak F., Jeanmougin F., Higgins D.G. **The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools.** *Nucleic Acids Res.* 1997, **25**, 24; 4876-4882.
7. Edgar R.C. **MUSCLE: a multiple sequence alignment method with reduced time and space complexity.** *BMC Bioinformatics*. 2004, **5**; 113.
8. Zuker M. **Mfold web server for nucleic acid folding and hybridization prediction.** *Nucleic Acids Res.* 2003, **31**, 13; 3406-3415.
9. Krogh A., Larsson B., von Heijne G., Sonnhammer E.L. **Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes.** *J. Mol. Biol.* 2001, **305**, 3; 567-580.
10. Darwin A.J., Li J., Stewart V. **Analysis of nitrate regulatory protein NarL-binding sites in the *fdnG* and *narG* operon control regions of *Escherichia coli* K-12.** *Mol. Microbiol.* 1996, **20**, 3; 621-632.
11. He B., Choi K.Y., Zalkin H. **Regulation of *Escherichia coli* *glnB*, *prsA*, and *speA* by the purine repressor.** *J. Bacteriol.* 1993, **175**, 11; 3598-3606.
12. He B., Shiau A., Choi K.Y., Zalkin H., Smith J.M. **Genes of the *Escherichia coli* *pur* regulon are negatively controlled by a repressor-operator interaction.** *J. Bacteriol.* 1990, **172**, 8; 4555-4562.

13. Melville S.B., Gunsalus R.P. **Isolation of an oxygen-sensitive FNR protein of *Escherichia coli*: interaction at activator and repressor sites of FNR-controlled genes.** *Proc. Natl. Acad. Sci. U. S. A.* 1996, **93**, 3; 1226-1231.
14. Tseng C.P., Albrecht J., Gunsalus R.P. **Effect of microaerophilic cell growth conditions on expression of the aerobic (*cyoABCDE* and *cydAB*) and anaerobic (*narGHJI*, *frdABCD*, and *dmsABC*) respiratory pathway genes in *Escherichia coli*.** *J. Bacteriol.* 1996, **178**, 4; 1094-1098.
15. Makarova K.S., Mironov A.A., Gelfand M.S. **Conservation of the binding site for the arginine repressor in all bacterial lineages.** *Genome Biol.* 2001, **2**, 4; RESEARCH0013.
16. Panina E.M., Vitreschak A.G., Mironov A.A., Gelfand M.S. **Regulation of aromatic amino acid biosynthesis in gamma-proteobacteria.** *J. Mol. Microbiol. Biotechnol.* 2001, **3**, 4; 529-543.
17. Panina E.M., Vitreschak A.G., Mironov A.A., Gelfand M.S. **Regulation of biosynthesis and transport of aromatic amino acids in low-GC Gram-positive bacteria.** *FEMS Microbiol. Lett.* 2003, **222**, 2; 211-220.
18. Gelfand M.S., Koonin E.V., Mironov A.A. **Prediction of transcription regulatory sites in Archaea by a comparative genomic approach.** *Nucleic Acids Res.* 2000, **28**, 3; 695-705.
19. Laikova O.N., Mironov A.A., Gelfand M.S. **Computational analysis of the transcriptional regulation of pentose utilization systems in the gamma subdivision of Proteobacteria.** *FEMS Microbiol. Lett.* 2001, **205**, 2; 315-322.
20. Rodionov D.A., Mironov A.A., Gelfand M.S. **Transcriptional regulation of pentose utilisation systems in the *Bacillus/Clostridium* group of bacteria.** *FEMS Microbiol. Lett.* 2001, **205**, 2; 305-314.
21. Permina E.A., Gelfand M.S. **Heat shock (σ_{32} and HrcA/CIRCE) regulons in beta-, gamma- and epsilon-proteobacteria .** *J. Mol. Microbiol. Biotechnol.* 2003, **6**, 3-4; 174-181.
22. Permina E.A., Kazakov A.E., Kalinina O.V., Gelfand M.S. **Comparative genomics of regulation of heavy metal resistance in Eubacteria.** *BMC Microbiol.* 2006, **6**; 49.
23. Rodionov D.A., Dubchak I.L., Arkin A.P., Alm E.J., Gelfand M.S. **Dissimilatory metabolism of nitrogen oxides in bacteria: comparative reconstruction of transcriptional networks.** *PLoS Comput. Biol.* 2005, **1**, 5; e55.
24. Kazakov A.E., Rodionov D.A., Alm E., Arkin A.P., Dubchak I., Gelfand M.S. **Comparative genomics of regulation of fatty acid and branched-chain amino acid utilization in proteobacteria.** *J. Bacteriol.* 2009, **191**, 1; 52-64.
25. Rodionov D.A., Gelfand M.S., Todd J.D., Curson A.R., Johnston A.W. **Computational reconstruction of iron- and manganese-responsive transcriptional networks in alpha-proteobacteria.** *PLoS Comput. Biol.* 2006, **2**, 12; e163.

26. Gerasimova A.V., Gelfand M.S. **Evolution of the NadR regulon in Enterobacteriaceae.** *J. Bioinform. Comput. Biol.* 2005, **3**, 4; 1007-1019.
27. Rodionov D.A., Li X., Rodionova I.A., Yang C., Sorci L., Dervyn E., Martynowski D., Zhang H., Gelfand M.S., Osterman A.L. **Transcriptional regulation of NAD metabolism in bacteria: genomic reconstruction of NiaR (YrxA) regulon.** *Nucleic Acids Res.* 2008, **36**, 6; 2032-2046.
28. Rodionov D.A., De Ingeniis J., Mancini C., Cimadamore F., Zhang H., Osterman A.L., Raffaelli N. **Transcriptional regulation of NAD metabolism in bacteria: NrtR family of Nudix-related regulators.** *Nucleic Acids Res.* 2008, **36**, 6; 2047-2059.
29. Сычева Л.В., Пермина Е.А., Гельфанд М.С. **Таксон-специфичная регуляция SOS-ответа у гамма-протеобактерий.** *Мол. Биол.* 2007, **41**, 5; 908-917.
30. Mazo'n G., Erill I., Campoy S., Corte's P., Forano E., Barbe' J. **Reconstruction of the evolutionary history of the LexA-binding sequence.** *Microbiology.* 2004, **150**, 11; 3783-3795.
31. Erill I., Jara M., Salvador N., Escribano M., Campoy S., Barbe' J. **Differences in LexA regulon structure among Proteobacteria through in vivo assisted comparative genomics.** *Nucleic Acids Res.* 2004, **32**, 22; 6617-6626.
32. Su Z., Olman V., Mao F., Xu Y. **Comparative genomics analysis of NtcA regulons in cyanobacteria: regulation of nitrogen assimilation and its coupling to photosynthesis.** *Nucleic Acids Res.* 2005, **33**, 16; 5156-5171.
33. Дорошук Н.А., Гельфанд М.С., Родионов Д.А. **Регуляция метаболизма азота у грамположительных бактерий.** *Мол. Биол.* 2006, **40**, 5; 919-926.
34. Winkler W.C., Cohen-Chalamish S., Breaker R.R. **An mRNA structure that controls gene expression by binding FMN.** *Proc. Natl. Acad. Sci. U. S. A.* 2002, **99**, 25; 15908-15913.
35. Gelfand M.S., Mironov A.A., Jomantas J., Kozlov Y.I., Perumov D.A. **A conserved RNA structure element involved in the regulation of bacterial riboflavin synthesis genes.** *Trends Genet.* 1999, **15**, 11; 439-442.
36. Vitreschak A.G., Rodionov D.A., Mironov A.A., Gelfand M.S. **Regulation of riboflavin biosynthesis and transport genes in bacteria by transcriptional and translational attenuation.** *Nucleic Acids Res.* 2002, **30**, 14; 3141-3151.
37. Rodionov D.A., Vitreschak A.G., Mironov A.A., Gelfand M.S. **Comparative genomics of thiamin biosynthesis in procaryotes. New genes and regulatory mechanisms.** *J. Biol. Chem.* 2002, **277**, 50; 48949-48959.
38. Vitreschak A.G., Rodionov D.A., Mironov A.A., Gelfand M.S. **Regulation of the vitamin B12 metabolism and transport in bacteria by a conserved RNA structural element.** *RNA.* 2003, **9**, 9; 1084-1097.

39. Rodionov D.A., Vitreschak A.G., Mironov A.A., Gelfand M.S. **Comparative genomics of the methionine metabolism in Gram-positive bacteria: a variety of regulatory system.** *Nucleic Acids Res.* 2004, **32**, 11; 3340-3353.
40. Rodionov D.A., Vitreschak A.G., Mironov A.A., Gelfand M.S. **Regulation of lysine biosynthesis and transport genes in bacteria: yet another RNA riboswitch?** *Nucleic Acids Res.* 2003, **31**, 23; 6748-6757.
41. Conlan S., Lawrence C., McCue L.A. **Rhodopseudomonas palustris regulons detected by cross-species analysis of alphaproteobacterial genomes.** *Appl. Environ. Microbiol.* 2005, **71**, 11; 7442-7452.
42. Terai G., Takagi T., Nakai K. **Prediction of co-regulated genes in Bacillus subtilis on the basis of upstream elements conserved across three closely related species.** *Genome Biol.* 2001, **2**, 11; RESEARCH0048.
43. Alkema W.B., Lenhard B., Wasserman W.W. **Regulog analysis: detection of conserved regulatory networks across bacteria: application to Staphylococcus aureus.** *Genome Res.* 2004, **14**, 7; 1362-1373.
44. Korzheva N., Mustaev A., Kozlov M., Malhotra A., Nikiforov V., Goldfarb A., Darst S.A. **A structural model of transcription elongation.** *Science.* 2000, **289**, 5479; 619-625.
45. Gourse R.L., Ross W., Gaal T. **UPs and downs in bacterial transcription initiation: the role of the alpha subunit of RNA polymerase in promoter recognition.** *Mol. Microbiol.* 2000, **37**, 4; 687-695.
46. Hampsey M. **RNA polymerase comes into focus.** *Trends Genet.* 2000, **16**, 1; 20.
47. Browning D.F., Busby S.J. **The regulation of bacterial transcription initiation.** *Nat. Rev. Microbiol.* 2004, **2**, 1; 57-65.
48. Ishihama A. **Functional modulation of Escherichia coli RNA polymerase.** *Annu. Rev. Microbiol.* 2000, **54**; 499-518.
49. Ross W., Ernst A., Gourse R.L. **Fine structure of E. coli RNA polymerase-promoter interactions: alpha subunit binding to the UP element minor groove.** *Genes Dev.* 2001, **15**, 5; 491-506.
50. Perez-Rueda E., Collado-Vides J. **The repertoire of DNA-binding transcriptional regulators in Escherichia coli K-12.** *Nucleic Acids Res.* 2000, **28**, 8; 1838-1847.
51. Babu M., Teichmann S.A. **Evolution of transcription factors and the gene regulatory network in Escherichia coli.** *Nucleic Acids Res.* 2003, **31**, 4; 1234-1244.
52. Martinez-Antonio A., Collado-Vides J. **Identifying global regulators in transcriptional regulatory networks in bacteria.** *Curr. Opin. Microbiol.* 2003, **6**, 5; 482-489.
53. Stover C.K., Pham X.Q., Erwin A.L., Mizoguchi S.D., Warrenner P., Hickey M.J., Brinkman

- F.S., Hufnagle W.O., Kowalik D.J., Lagrou M., Garber R.L., Goltry L., Tolentino E., Westbrock-Wadman S., Yuan Y., Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R., Smith K., Spencer D., Wong G.K., Wu Z., Paulsen I.T., Reizer J., Saier M.H., Hancock R.E., Lory S., Olson M.V. **Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.** *Nature*. 2000, **406**, 6799; 959-964.
54. van Nimwegen E. **Scaling laws in the functional content of genomes.** *Trends Genet.* 2003, **19**, 9; 479-484.
55. Moreno-Campuzano S., Janga S.C., Pe'rez-Rueda E. **Identification and analysis of DNA-binding transcription factors in *Bacillus subtilis* and other Firmicutes--a genomic approach.** *BMC Genomics*. 2006, **7**; 147.
56. Dobrindt U., Hacker J. **Whole genome plasticity in pathogenic bacteria.** *Curr. Opin. Microbiol.* 2001, **4**, 5; 550-557.
57. Bentley S.D., Chater K.F., Cerden~o-Ta'rraga A.M., Challis G.L., Thomson N.R., James K.D., Harris D.E., Quail M.A., Kieser H., Harper D., Bateman A., Brown S., Chandra G., Chen CW., Collins M., Cronin A., Fraser A., Goble A., Hidalgo J., Hornsby T., Howarth S., Huang CH., Kieser T., Larke L., Murphy L., Oliver K., O'Neil S., Rabinowitsch E., Rajandream M.A., Rutherford K., Rutter S., Seeger K., Saunders D., Sharp S., Squares R., Squares S., Taylor K., Warren T., Wietzorrek A., Woodward J., Barrell B.G., Parkhill J., Hopwood D.A. **Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2).** *Nature*. 2002, **417**, 6885; 141-147.
58. Lewis M. **The *lac* repressor.** *C. R. Biol.* 2005, **328**, 6; 521-548.
59. Hulett F.M., Lee J., Shi L., Sun G., Chesnut R., Sharkova E., Duggan M.F., Kapp N. **Sequential action of two-component genetic switches regulates the PHO regulon in *Bacillus subtilis*.** *J. Bacteriol.* 1994, **176**, 5; 1348-1358.
60. Anthamatten D., Scherb B., Hennecke H. **Characterization of a *fixLJ*-regulated *Bradyrhizobium japonicum* gene sharing similarity with the *Escherichia coli* *fnr* and *Rhizobium meliloti* *fixK* genes.** *J. Bacteriol.* 1992, **174**, 7; 2111-2120.
61. Kolb A., Busby S., Buc H., Garges S., Adhya S. **Transcriptional regulation by cAMP and its receptor protein.** *Annual Review of Biochemistry*. 1993, **62**; 749-795.
62. Kumar A., Grimes B., Fujita N., Makino K., Malloch R.A., Hayward R.S., Ishihama A. **Role of the sigma 70 subunit of *Escherichia coli* RNA polymerase in transcription activation.** *J. Mol. Biol.* 1994, **235**, 2; 405-413.
63. Chen P.R., He C. **Selective recognition of metal ions by metalloregulatory proteins.** *Curr. Opin. Chem. Biol.* 2008, **12**, 2; 214-221.
64. Weickert M.J., Adhya S. **The galactose regulon of *Escherichia coli*.** *Mol. Microbiol.* 1993, **10**, 2; 245-251.
65. Shin M., Kang S., Hyun S.J., Fujita N., Ishihama A., Valentin-Hansen P., Choy H.E. **Repression of *deoP2* in *Escherichia coli* by CytR: conversion of a transcription activator**

- into a repressor. *EMBO J.* 2001, **20**, 19; 5392-5399.
66. Rodionov D.A., Mironov A.A., Gelfand M.S. **Conservation of the biotin regulon and the BirA regulatory signal in Eubacteria and Archaea.** *Genome Res.* 2002, **12**, 10; 1507-1516.
 67. Panina E.M., Mironov A.A., Gelfand M.S. **Comparative genomics of bacterial zinc regulons: enhanced ion transport, pathogenesis, and rearrangement of ribosomal proteins.** *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**, 17; 9912-9917.
 68. Makarova K.S., Ponomarev V.A., Koonin E.V. **Two C or not two C: recurrent disruption of Zn-ribbons, gene duplication, lineage-specific gene loss, and horizontal gene transfer in evolution of bacterial ribosomal proteins.** *Genome Biol.* 2001, **2**, 9; RESEARCH 0033.
 69. Садовская Н.С., Лайкова О.Н., Миронов А.А., Гельфанд М.С. **Изучение регуляции метаболизма длинноцепочечных жирных кислот с использованием компьютерного анализа полных бактериальных геномов.** *Мол. Биол.* 2001, **35**, 6; 1010-1014.
 70. Campbell J.W., Cronan J.E.Jr. **The enigmatic *Escherichia coli* *fadE* gene is *yafH*.** *J. Bacteriol.* 2002, **184**, 13; 3759-3764.
 71. Rodionov D.A., Hebbeln P., Gelfand M.S., Eitinger T. **Comparative and functional genomic analysis of prokaryotic nickel and cobalt uptake transporters: evidence for a novel group of ATP-binding cassette transporters.** *J. Bacteriol.* 2006, **188**, 1; 317-327.
 72. Burgess C.M., Slotboom D.J., Geertsma E.R., Durkens R.H., Poolman B., van Sinderen D. **The riboflavin transporter RibU in *Lactococcus lactis*: molecular characterization of gene expression and the transport mechanism.** *J. Bacteriol.* 2006, **188**, 8; 2752-2760.
 73. Rutherford K., Parkhill J., Crook J., Horsnell T., Rice P., Rajandream M.A., Barrell B. **Artemis: sequence visualization and annotation.** *Bioinformatics.* 2000, **16**, 10; 944-945.
 74. Fitch W.M. **Distinguishing homologous from analogous proteins.** *Syst. Zool.* 1970, **19**, 2; 99-113.
 75. Benson D.A., Boguski M.S., Lipman D.J., Ostell J., Ouellette B.F., Rapp B.A., Wheeler D.L. **GenBank.** *Nucleic Acids Res.* 1999, **27**, 1; 12-17.
 76. Emmert D.B., Stoehr P.J., Stoesser G., Cameron G.N. **The European Bioinformatics Institute (EBI) databases.** *Nucleic Acids Res.* 1994, **22**, 17; 3445-3449.
 77. Boeckmann B., Blatter M.-C., Famiglietti L., Hinz U., Lane L., Roechert B., Bairoch A. **Protein variety and functional diversity: Swiss-Prot annotation in its biological context.** *C. R. Biol.* 2005, **328**, 10-11; 882-899.
 78. Gasteiger E., Jung E., Bairoch A. **SWISS-PROT: connecting biomolecular knowledge via a protein database.** *Curr. Issues Mol. Biol.* 2001, **3**, 3; 47-55.
 79. Koonin E.V., Galperin M.Y. **Prokaryotic genomes: the emerging paradigm of genome-**

- based microbiology.** 1997, **7**, 6; 757-763.
80. Hulo N., Bairoch A., Bulliard V., Cerutti L., Cuče B.A., Castro E., Lachaize C., Langendijk-Genevaux P.S., Sigrist C.J. **The 20 years of PROSITE.** *Nucleic Acids Res.* 2008, **36**, 18003654; D245-D249.
 81. Sonnhammer E.L., Eddy S.R., Birney E., Bateman A., Durbin R. **Pfam: multiple sequence alignments and HMM-profiles of protein domains.** *Nucleic Acids Res.* 1998, **26**, 1; 320-322.
 82. Schultz J., Milpetz F., Bork P., Ponting C.P. **SMART, a simple modular architecture research tool: identification of signaling domains.** *Proc. Natl. Acad. Sci. U. S. A.* 1998, **95**, 11; 5857-5864.
 83. Quevillon E., Silventoinen V., Pillai S., Harte N., Mulder N., Apweiler R., Lopez R. **InterProScan: protein domains identifier.** *Nucleic Acids Res.* 2005, **33**; W116-W120.
 84. Tatusov R.J., Koonin E.V., Lipman D.J. **A genomic perspective on protein families.** *Science.* 1997, **278**, 5338; 631-637 .
 85. Tatusov R.L., Natale D.A., Garkavtsev I.V., Tatusova T.A., Shankavaram U.T., Rao B.S., Kiryutin B., Galperin M.Y., Fedorova N.D., Koonin E.V. **The COG database: new developments in phylogenetic classification of proteins from complete genomes.** *Nucleic Acids Res.* 2001, **29**, 1; 22-28.
 86. Bendtsen J.D., Nielsen H., von Heijne G., Brunak S. **Improved prediction of signal peptides: SignalP 3.0.** *J. Mol. Biol.* 2004, **340**, 4; 783-795.
 87. Kummerfeld S.K., Teichmann S.A. **DBD: a transcription factor prediction database.** *Nucleic Acids Res.* 2006, **34**; D74-D81.
 88. Overbeek R., Fonstein M., D'Souza M., Pusch G.D., Maltsev N. **The use of gene clusters to infer functional coupling.** *Proc. Natl. Acad. Sci. U. S. A.* 1999, **96**, 6; 2896-2901.
 89. Kolesov G., Mewes H.W., Frishman D. **SNAPping up functionally related genes based on context information: a colinearity-free approach.** *J. Mol. Biol.* 2001, **311**, 4; 639-656.
 90. Rogozin I.B., Makarova K.S., Murvai J., Czabarka E., Wolf Y.I., Tatusov R.L., Szekely L.A., Koonin E.V. **Connected gene neighborhoods in prokaryotic genomes .** *Nucleic Acids Res.* 2002, **30**, 10; 2212-2223.
 91. Galperin M.Y., Koonin E.V. **Who's your neighbor? New computational approaches for functional genomics.** *Nat. Biotechnol.* 2000, **18**, 6; 609-613.
 92. Marrakchi H., Choi K.H., Rock C.O. **A new mechanism for anaerobic unsaturated fatty acid formation in *Streptococcus pneumoniae*.** *J. Biol. Chem.* 2002, **277**, 47; 44809-44816.
 93. Daugherty M., Vonstein V., Overbeek R., Osterman A. **Archaeal shikimate kinase, a new member of the GHMP-kinase family.** *J. Bacteriol.* 2001, **183**, 1; 292-300.

94. Osterman A. **A hidden metabolic pathway exposed.** *Proc. Natl. Acad. Sci. U. S. A.* 2006, **11**, 103; 5637-5638.
95. Doerks T., Andrade M.A., Lathe W.3., von Mering C., Bork P. **Global analysis of bacterial transcription factors to predict cellular target processes.** *Trends Genet.* 2004, **20**, 3; 126-131.
96. Snel B., Lehmann G., Bork P., Huynen M.A. **STRING: a web-server to retrieve and display the repeatedly occurring neighbourhood of a gene.** *Nucleic Acids Res.* 2000, **28**, 18; 3442-3444.
97. Enright A.J., Iliopoulos I., Kyrpides N.C., Ouzounis C.A. **Protein interaction maps for complete genomes based on gene fusion events.** *Nature.* 1999, **402**, 6757; 86-90 .
98. Marcotte E.M., Pellegrini M., Ng H.L., Rice D.W., Yeates T.O., Eisenberg D. **Detecting protein function and protein-protein interactions from genome sequences.** *Science.* 1999, **285**, 5428; 751-753.
99. Yanai I., Derti A., DeLisi C. **Genes linked by fusion events are generally of the same functional category: a systematic analysis of 30 microbial genomes.** *Proc. Natl. Acad. Sci. U. S. A.* 2001, **98**, 14; 7940-7945.
100. Huynen M., Snel B., Lathe W.3., Bork P. **Predicting protein function by genomic context: quantitative evaluation and qualitative inferences.** *Genome Res.* 2000, **10**, 8; 1204-1210.
101. Mellor J.C., Yanai I., Clodfelter K.H., Mintseris J., DeLisi C. **Predictome: a database of putative functional links between proteins.** *Nucleic Acids Res.* 2002, **30**, 1; 306-309.
102. Pellegrini M., Marcotte E.M., Thompson M.J., Eisenberg D., Yeates T.O. **Assigning protein functions by comparative genome analysis: protein phylogenetic profiles.** *Proc. Natl. Acad. Sci. U. S. A.* 1999, **96**, 8; 4285-4288.
103. Rodionov D.A., Gelfand M.S. **Identification of a bacterial regulatory system for ribonucleotide reductases by phylogenetic profiling.** *Trends Genet.* 2005, **21**, 7; 385-389.
104. Hertz G.Z., Hartzell G.W., Stormo G.D. **Identification of consensus patterns in unaligned DNA sequences known to be functionally related.** *Comput. Appl. Biosci.* 1990, **6**, 2; 81-92.
105. Berg O.G., von Hippel P.H. **Selection of DNA binding sites by regulatory proteins. Statistical-mechanical theory and application to operators and promoters.** *J. Mol. Biol.* 1987, **193**, 4; 723-750.
106. Shelton D.A., Stegman L., Hardison R., Miller W., Bock J.H., Slightom J.L., Goodman M., Gumucio D.L. **Phylogenetic footprinting of hypersensitive site 3 of the beta-globin locus control region.** *Blood.* 1997, **89**, 9; 3457-3469.
107. Dubchak I., Ryaboy D.V. **VISTA family of computational tools for comparative analysis of DNA sequences and whole genomes.** *Methods Mol. Biol.* 2006, **338**; 69-89 .

108. Chen X., Jiang T. **An improved Gibbs sampling method for motif discovery via sequence weighting.** *Comput. Syst. Bioinformatics Conf.* 2006; 239-247.
109. Bailey T.L., Elkan C. **Fitting a mixture model by expectation maximization to discover motifs in biopolymers.** *proceedings of the Second International Conference on Intelligent Systems for Molecular Biology.* 1994; 28-36.
110. Миронов А.А., Винокурова Н.П., Гельфанд М.С. **Программное обеспечение анализа бактериальных геномов.** *Мол. Биол.* 2000, **34**, 2; 253-262.
111. Favorov A.V., Gelfand M.S., Gerasimova A.V., Ravcheev D.A., Mironov A.A., Makeev V.J. **A Gibbs sampler for identification of symmetrically structured, spaced DNA motifs with improved estimation of the signal length.** *Bioinformatics.* 2005, **21**, 10; 2240-2245.
112. Siddharthan R., Siggia E.D., van Nimwegen E. **PhyloGibbs: a Gibbs sampling motif finder that incorporates phylogeny.** *PLoS Comput. Biol.* 2005, **1**, 7; e67.
113. Sinha S., Blanchette M., Tompa M. **PhyME: a probabilistic algorithm for finding motifs in sets of orthologous sequences.** *BMC Bioinformatics.* 2004, **5**; 170.
114. Mironov A.A., Koonin E.V., Roytberg M.A., Gelfand M.S. **Computer analysis of transcription regulatory patterns in completely sequenced bacterial genomes.** *Nucleic Acids Res.* 1999, **27**, 14; 2981-2989.
115. Hebbeln P., Rodionov D.A., Alfandega A., Eitinger T. **Biotin uptake in prokaryotes by solute transporters with an optional ATP-binding cassette-containing module.** *Proc. Natl. Acad. Sci. U. S. A.* 2007, **104**, 8; 2909-2914.
116. Nanamiya H., Akanuma G., Natori Y., Murayama R., Kosono S., Kudo T., Kobayashi K., Ogasawara N., Park S.M., Ochi K., Kawamura F. **Zinc is a key factor in controlling alternation of two types of L31 protein in the *Bacillus subtilis* ribosome.** *Mol. Microbiol.* 2004, **52**, 1; 273-283.
117. Akanuma G., Nanamiya H., Natori Y., Nomura N., Kawamura F. **Liberation of zinc-containing L31 (RpmE) from ribosomes by its paralogous gene product, YtiA, in *Bacillus subtilis*.** *J. Bacteriol.* 2006, **188**, 7; 2715-2720.
118. Owen G.A., Pascoe B., Kallifidas D., Paget M.S. **Zinc-responsive regulation of alternative ribosomal protein genes in *Streptomyces coelicolor* involves zur and sigmaR.** *J. Bacteriol.* 2007, **189**, 11; 4078-4086.
119. Shin J.H., Oh S.Y., Kim S.J., Roe J.H. **The zinc-responsive regulator Zur controls a zinc uptake system and some ribosomal proteins in *Streptomyces coelicolor* A3(2).** *J. Bacteriol.* 2007, **189**, 11; 4070-4077.
120. Karatza P., Frillingos S. **Cloning and functional characterization of two bacterial members of the NAT/NCS2 family in *Escherichia coli*.** *Mol. Membr. Biol.* 2005, **22**, 3; 251-261.

121. Rodionov D.A., Mironov A.A., Rakhmaninova A.B., Gelfand M.S. **Transcriptional regulation of transport and utilization systems for hexuronides, hexuronates and hexonates in gamma purple bacteria.** *Mol. Microbiol.* 2000 , **38**, 4; 673-683.
122. Rodionov D.A., Gelfand M.S., Hugouvieux-Cotte-Pattat N. **Comparative genomics of the KdgR regulon in *Erwinia chrysanthemi* 3937 and other gamma-proteobacteria.** *Microbiology.* 2004, **150**, 11; 3571-3590.
123. Sperandio B., Gautier C., McGovern S., Ehrlich D.S., Renault P., Martin-Verstraete I., Gue'don E. **Control of methionine synthesis and uptake by MetR and homocysteine in *Streptococcus mutans*.** *J. Bacteriol.* 2007, **189**, 19; 7032-7044.
124. Sekowska A R.S.D.J.H.A.D.A. **Extracting biological information from DNA arrays: an unexpected link between arginine and methionine metabolism in *Bacillus subtilis*.** *Genome Biol.* 2001, **6**; RESEARCH0019.
125. Torrents E., Grinberg I., Gorovitz-Harris B., Lundstro'm H., Borovok I., Aharonowitz Y., Sjo'berg B.M., Cohen G. **NrdR controls differential expression of the *Escherichia coli* ribonucleotide reductase genes.** *J. Bacteriol.* 2007, **189**, 14; 502-521.
126. Grinberg I., Shteinberg T., Gorovitz B., Aharonowitz Y., Cohen G., Borovok I. **The *Streptomyces* NrdR transcriptional regulator is a Zn ribbon/ATP cone protein that binds to the promoter regions of class Ia and class II ribonucleotide reductase operons.** *J. Bacteriol.* 2006, **188**, 21; 7635-7644.
127. Winkler W., Nahvi A., Breaker R.R. **Thiamine derivatives bind messenger RNAs directly to regulate bacterial gene expression.** *Nature.* 2002, **419**, 6910; 952-956.
128. Nahvi A., Barrick J.E., Breaker R.R. **Coenzyme B12 riboswitches are widespread genetic control elements in prokaryotes.** *Nucleic Acids Res.* 2004, **23**, 1; 143-150.
129. Epshtein V., Mironov A.S., Nudler E. **The riboswitch-mediated control of sulfur metabolism in bacteria.** *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**, 9; 5052-5056.
130. McDaniel B.A., Grundy F.J., Artsimovitch I., Henkin T.M. **Transcription termination control of the S box system: direct measurement of S-adenosylmethionine by the leader RNA.** *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**, 6; 3083-3088.
131. Winkler W.C., Nahvi A., Sudarsan N., Barrick J.E., Breaker R.R. **An mRNA structure that controls gene expression by binding S-adenosylmethionine.** *Nat. Struct. Biol.* 2003, **10**, 9; 701-707.
132. Grundy F.J., Lehman S.C., Henkin T.M. **The L box regulon: lysine sensing by leader RNAs of bacterial lysine biosynthesis genes.** *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**, 21; 12057-12062.
133. Chin A.M., Feucht B.U., Saier M.H. **Evidence for regulation of gluconeogenesis by the fructose phosphotransferase system in *Salmonella typhimurium*.** *J. Bacteriol.* 1987, **169**, 2; 897-899.

134. Bledig S.A., Ramseier T.M., Saier M.H. **FruR mediates catabolite activation of pyruvate kinase (*pykF*) gene expression in *Escherichia coli*.** *J. Bacteriol.* 1996, **178**, 1; 280-283.
135. Ramseier T.M., Bledig S., Michotey V., Feghali R., Saier M.H. **The global regulatory protein FruR modulates the direction of carbon flow in *Escherichia coli*.** *Mol. Microbiol.* 1995, **16**, 6; 1157-1169.
136. Ryu S., Ramseier T.M., Michotey V., Saier M.H., Garges S. **Effect of the FruR regulator on transcription of the *pts* operon in *Escherichia coli*.** *J. Biol. Chem.* 1995, **270**, 6; 2489-2496.
137. Saier M.H., Ramseier T.M. **The catabolite repressor/activator (Cra) protein of enteric bacteria.** *J. Bacteriol.* 1996, **178**, 12; 3411-3417.
138. Saier M.H. **Cyclic AMP-independent catabolite repression in bacteria.** *FEMS Microbiol. Lett.* 1996, **138**, 2-2; 97-103.
139. Negre D., Bonod-Bidaud C., Geourjon C., Deleage G., Cozzzone A.J., Cortay J.C. **Definition of a consensus DNA-binding site for the *Escherichia coli* pleiotropic regulatory protein, FruR.** *Mol. Microbiol.* 1996, **21**, 2; 257-266.
140. Ramseier T.M., Negre D., Cortay J.C., Scarabel M., Cozzzone A.J., Saier M.H. ***In vitro* binding of the pleiotropic transcriptional regulatory protein, FruR, to the *fru*, *pps*, *ace*, *pts* and *icd* operons of *Escherichia coli* and *Salmonella typhimurium*.** *J. Mol. Biol.* 1993, **234**, 1; 28-44.
141. Meyer D., Schneider-Fresenius C., Horlacher R., Peist R., Boos W. **Molecular characterization of glucokinase from *Escherichia coli* K-12.** *J. Bacteriol.* 1997, **179**, 4; 1298-1306.
142. Mikulskis A., Aristarkhov A., Lin E.C. **Regulation of expression of the ethanol dehydrogenase gene (*adhE*) in *Escherichia coli* by catabolite repressor activator protein Cra.** *J. Bacteriol.* 1997, **179**, 22; 7129-7134.
143. Kaga N., Umitsuki G., Clark D.P., Nagai K., Wachi M. **Extensive overproduction of the AdhE protein by *rng* mutations depends on mutations in the *cra* gene or in the Cra-box of the *adhE* promoter.** *Biochem. Biophys. Res. Commun.* 2002, **295**, 1; 92-97.
144. Prost J.F., Negre D., Oudot C., Murakami K., Ishihama A., Cozzzone A.J., Cortay J.C. **Cra-dependent transcriptional activation of the *icd* gene of *Escherichia coli*.** *J. Bacteriol.* 1999, **181**, 3; 893-898.
145. Tyson K., Busby S., Cole J. **Catabolite regulation of two *Escherichia coli* operons encoding nitrite reductases: role of the Cra protein.** *Arch. Microbiol.* 1997, **168**, 3; 240-244.
146. Browning D.F., Cole J.A., Busby S.J. **Transcription activation by remodelling of a nucleoprotein assembly: the role of NarL at the FNR-dependent *Escherichia coli* *nir* promoter.** *Mol. Microbiol.* 2004, **53**, 1; 203-215.

147. Cortay J.C., Negre D., Scarabel M., Ramseier T.M., Vartak N.B., Reizer J., Saier M.H., Cozzone A.J. **In vitro asymmetric binding of the pleiotropic regulatory protein, FruR, to the ace operator controlling glyoxylate shunt enzyme synthesis.** *J. Biol. Chem.* 1994, **296**, 21; 14885-14891.
148. Mauzy C.A., Hermodson M.A. **Structural and functional analyses of the repressor, RbsR, of the ribose operon of *Escherichia coli*.** *Protein Sci.* 1992, **1**, 7; 831-842.
149. Bell A.W., Buckel S.D., Groarke J.M., Hope J.N., Kingsley D.H., Hermodson M.A. **The nucleotide sequences of the *rbsD*, *rbsA*, and *rbsC* genes of *Escherichia coli* K12.** *J. Biol. Chem.* 1986, **261**, 17; 7652-7658.
150. Hope J.N., Bell A.W., Hermodson M.A., Groarke J.M. **Ribokinase from *Escherichia coli* K12. Nucleotide sequence and overexpression of the *rbsK* gene and purification of ribokinase.** *J. Biol. Chem.* 1986, **261**, 17; 7663-7668.
151. Schumacher M.A., Choi K.Y., Zalkin H., Brennan R.G. **Crystal structure of LacI member, PurR, bound to DNA: minor groove binding by alpha helices.** *Science.* 1994, **266**, 5186; 763-770.
152. Schumacher M.A., Glasfeld A., Zalkin H., Brennan R.G. **The X-ray structure of the PurR-guanine-*purF* operator complex reveals the contributions of complementary electrostatic surfaces and a water-mediated hydrogen bond to corepressor specificity and binding affinity.** *J. Biol. Chem.* 1997, **272**, 36; 22648-22653.
153. Choi K.Y., Zalkin H. **Structural characterization and corepressor binding of the *Escherichia coli* purine repressor.** *J. Bacteriol.* 1992, **174**, 19; 6207-6214.
154. Zhang H., Wang A. **Functional analysis of three amino acid residues of *purR* repressor, Trp147, Gln-218 and Gln-292 in *Salmonella typhimurium*.** *Sci. China. C. Life. Sci.* 2001, **44**, 2; 184-191.
155. Zalkin, H. and Neidhardt, P. **Biosynthesis of purine nucleotides.** In Neidhardt, F.C. (eds.) *Escherichia coli and Salmonella. Cellular and Molecular Biology*, (1996), ASM Press, Washington D.C., pp. 1325-1333.
156. Rolfes R.J., Zalkin H. **Autoregulation of *Escherichia coli purR* requires two control sites downstream of the promoter.** *J. Bacteriol.* 1990, **172**, 10; 5758-5766.
157. He B., Zalkin H. **Regulation of *Escherichia coli purA* by purine repressor, one component of a dual control mechanism.** *J. Bacteriol.* 1994, **176**, 4; 1009-1113.
158. He B., Smith J.M., Zalkin H. ***Escherichia coli purB* gene: cloning, nucleotide sequence, and regulation by *purR*.** *J. Bacteriol.* 1992, **174**, 1; 130-136.
159. Meng L.M., Kilstrup M., Nygaard P. **Autoregulation of PurR repressor synthesis and involvement of *purR* in the regulation of *purB*, *purC*, *purL*, *purMN* and *guaBA* expression in *Escherichia coli*.** *Eur. J. Biochem.* 1990, **187**, 2; 373-379.

160. Rolfes R.J., Zalkin H. ***Escherichia coli* gene *purR* encoding a repressor protein for purine nucleotide synthesis. Cloning, nucleotide sequence, and interaction with the *purF* operator.** *J. Biol. Chem.* 1988, **263**, 36; 19653-19661.
161. Rolfes R.J., Zalkin H. **Regulation of *Escherichia coli purF*. Mutations that define the promoter, operator, and purine repressor gene.** *J. Biol. Chem.* 1988a, **263**, 36; 19649-19652.
162. Nygaard P., Smith J.M. **Evidence for a novel glycinamide ribonucleotide transformylase in *Escherichia coli*.** *J. Bacteriol.* 1993, **175**, 11; 3591-3597.
163. Wilson H.R., Turnbough C.L.Jr. **Role of the purine repressor in the regulation of pyrimidine gene expression in *Escherichia coli* K-12.** *J. Bacteriol.* 1990, **172**, 6; 3208-3213.
164. Devroede N., Thia-Toong T.L., Gigot D., Maes D., Charlier D. **Purine and pyrimidine-specific repression of the *Escherichia coli carAB* operon are functionally and structurally coupled.** *J. Mol. Biol.* 2004, **336**, 1; 25-42.
165. Wilson R.L., Stauffer L.T., Stauffer G.V. **Roles of the GcvA and PurR proteins in negative regulation of the *Escherichia coli* glycine cleavage enzyme system.** *J. Bacteriol.* 1993, **175**, 16; 5129-5134.
166. Steiert J.G., Kubu C., Stauffer G.V. **The PurR binding site in the *glyA* promoter region of *Escherichia coli*.** *FEMS Microbiol. Lett.* 1992, **78**, 2-3; 299-304.
167. Berry S. **The chemical basis of membrane bioenergetics.** *J. Mol. Evol.* 2002, **54**, 5; 595-613.
168. Gennis, R.B. and Stewart, V. **Respiration.** In F.C. Neidhart (eds.) *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology, (1996), ASM Press, Washington, pp. 217-286.
169. Uden G., Bongaerts J. **Alternative respiratory pathways of *Escherichia coli*: energetics and transcriptional regulation in response to electron acceptors.** *Biochim. Biophys. Acta.* 1997, **1320**, 3; 217-234.
170. Lynch, A.S. and Lin, C.C. **Responses to molecular oxygen.** In F.C. Neidhart (eds.) *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology, (1996), ASM Press, Washington, pp. 1526-1537.
171. Uden G., Becker S., Bongaerts J., Holighaus G., Schirawski J., Six S. **O₂-sensing and O₂-dependent gene regulation in facultatively anaerobic bacteria.** *Arch. Microbiol.* 1995, **164**, 2; 81-90.
172. Spiro S., Guest J.R. **Adaptive responses to oxygen limitation in *Escherichia coli*.** *Trends Biochem. Sci.* 1991, **16**, 8; 310-314.
173. Iuchi S., Lin E.C. **Adaptation of *Escherichia coli* to redox environments by gene expression.** *Mol. Microbiol.* 1993, **9**, 1; 9-15.

174. Simon G., Mejean V., Jourlin C., Chippaux M., Pascal M.C. **The *torR* gene of *Escherichia coli* encodes a response regulator protein involved in the expression of the trimethylamine N-oxide reductase genes.** *J. Bacteriol.* 1994, **176**, 18; 5601-5606.
175. Simon G., Jourlin C., Ansaldi M., Pascal M.C., Chippaux M., Mejean V. **Binding of the TorR regulator to *cis*-acting direct repeats activates *tor* operon expression.** *Mol. Microbiol.* 1995, **17**, 5; 971-980.
176. Golby P., Davies S., Kelly D.J., Guest J.R., Andrews S.C. **Identification and characterization of a two-component sensor-kinase and response-regulator system (DcuS-DcuR) controlling gene expression in response to C4-dicarboxylates in *Escherichia coli*.** *J. Bacteriol.* 1999, **181**, 4; 1238-1248.
177. Lambden P.R., Guest J.R. **Mutants of *Escherichia coli* K12 unable to use fumarate as an anaerobic electron acceptor.** *J. Gen. Microbiol.* 1976, **97**, 2; 145-160.
178. Unden G., Schirawski J. **The oxygen-responsive transcriptional regulator FNR of *Escherichia coli*: the search for signals and reactions.** *Mol. Microbiol.* 1997, **25**, 2; 205-210.
179. Sawers G. **The aerobic/anaerobic interface.** *Curr. Opin. Microbiol.* 1999, **2**, 2; 181-187.
180. Shalel-Levanon S., San K.Y., Bennett G.N. **Effect of ArcA and FNR on the expression of genes related to the oxygen regulation and the glycolysis pathway in *Escherichia coli* under microaerobic growth conditions.** *Biotechnol. Bioeng.* 2005, **92**, 2; 147-159.
181. Korner H., Sofia H.J., Zumft W.G. **Phylogeny of the bacterial superfamily of Crp-Fnr transcription regulators: exploiting the metabolic spectrum by controlling alternative gene programs.** *FEMS Microbiol. Rev.* 2003, **27**, 5; 559-592.
182. Popescu C.V., Bates D.M., Beinert H., Munck E., Kiley P.J. **Mossbauer spectroscopy as a tool for the study of activation/inactivation of the transcription regulator FNR in whole cells of *Escherichia coli*.** *Proc. Natl. Acad. Sci. U. S. A.* 1998, **95**, 23; 13431-13435.
183. Kiley P.J., Beinert H. **Oxygen sensing by the global regulator, FNR: the role of the iron-sulfur cluster.** *FEMS Microbiol. Rev.* 1998, **22**, 5; 341-352.
184. Bauer C.E., Elsen S., Bird T.H. **Mechanisms for redox control of gene expression.** *Annu. Rev. Microbiol.* 1999, **53**; 495-523.
185. Lazazzera B.A., Beinert H., Khoroshilova N., Kennedy M.C., Kiley P.J. **DNA binding and dimerization of the Fe-S-containing FNR protein from *Escherichia coli* are regulated by oxygen.** *J. Biol. Chem.* 1996, **271**, 5; 2762-2768.
186. Cotter P.A., Gunsalus R.P. **Contribution of the *fnr* and *arcA* gene products in coordinate regulation of cytochrome o and d oxidase (*cyoABCDE* and *cydAB*) genes in *Escherichia coli*.** *FEMS Microbiol. Lett.* 1992, **91**, 1; 31-36.
187. Takahashi K., Hattori T., Nakanishi T., Nohno T., Fujita N., Ishihama A., Taniguchi S.

- Repression of in vitro transcription of the *Escherichia coli fnr* and *narX* genes by FNR protein.** *FEBS Lett.* 1994, **340**, 1-2; 59-64.
188. Mettert E.L., Kiley P.J. **Contributions of [4Fe-4S]-FNR and integration host factor to *fnr* transcriptional regulation.** *J. Bacteriol.* 2007, **189**, 8; 3036-3043.
189. Green J., Irvine A.S., Meng W., Guest J.R. **FNR-DNA interactions at natural and semi-synthetic promoters.** *Mol. Microbiol.* 1996, **19**, 1; 125-137.
190. Lamberg K.E., Kiley P.J. **FNR-dependent activation of the class II *dmsA* and *narG* promoters of *Escherichia coli* requires FNR-activating regions 1 and 3.** *Mol. Microbiol.* 2000, **38**, 4; 817-827.
191. Filenko N.A., Browning D.F., Cole J.A. **Transcriptional regulation of a hybrid cluster (prismane) protein.** *Biochem. Soc. Trans.* 2005, **33**, 1; 195-197.
192. Boston T., Atlung T. **FNR-mediated oxygen-responsive regulation of the *nrdDG* operon of *Escherichia coli*.** *J. Bacteriol.* 2003, **185**, 17; 5310-5313.
193. Browning D.F., Grainger D.C., Beatty C.M., Wolfe A.J., Cole J.A., Busby S.J. **Integration of three signals at the *Escherichia coli nrf* promoter: a role for Fis protein in catabolite repression.** *Mol. Microbiol.* 2005, **57**, 2; 496-510.
194. Partridge J.D., Poole R.K., Green J. **The *Escherichia coli yhjA* gene, encoding a predicted cytochrome c peroxidase, is regulated by FNR and OxyR.** *Microbiology.* 2007, **153**, 5; 1499-1507.
195. Li J., Stewart V. **Localization of upstream sequence elements required for nitrate and anaerobic induction of *fdn* (formate dehydrogenase-N) operon expression in *Escherichia coli* K-12.** *J. Bacteriol.* 1992, **174**, 15; 4935-4942 .
196. Darwin A.J., Ziegelhoffer E.C., Kiley P.J., Stewart V. **Fnr, NarP, and NarL regulation of *Escherichia coli* K-12 *napF* (periplasmic nitrate reductase) operon transcription in vitro.** *J. Bacteriol.* 1998, **180**, 16; 4192-4108.
197. Sawers G. **Specific transcriptional requirements for positive regulation of the anaerobically inducible *pfl* operon by ArcA and FNR.** *Mol. Microbiol.* 1993, **10**, 4; 737-747.
198. Kim S.J., Han Y.H., Kim I.H., Kim H.K. **Involvement of ArcA and Fnr in expression of *Escherichia coli* thiol peroxidase gene.** *IUBMB Life.* 1999, **48**, 2; 215-218.
199. Hasona A., Self W.T., Shanmugam K.T. **Transcriptional regulation of the *moe* (molybdate metabolism) operon of *Escherichia coli*.** *Arch. Microbiol.* 2001, **175**, 3; 178-188.
200. Ramseier T.M., Chien S.Y., Saier M.H. **Cooperative interaction between Cra and Fnr in the regulation of the *cydAB* operon of *Escherichia coli*.** *Curr. Microbiol.* 1996, **33**, 4; 270-274.

201. Jackson L., Blake T., Green J. **Regulation of *ndh* expression in *Escherichia coli* by Fis.** *Microbiology*. 2004, **150**, 2; 407-413.
202. Green J., Baldwin M.L., Richardson J. **Downregulation of *Escherichia coli yfiD* expression by FNR occupying a site at -93.5 involves the AR1-containing face of FNR.** *Mol. Microbiol.* 1998, **29**, 4; 1113-1123.
203. Wyborn N.R., Messenger S.L., Henderson R.A., Sawers G., Roberts R.E., Attwood M.M., Green J. **Expression of the *Escherichia coli yfiD* gene responds to intracellular pH and reduces the accumulation of acidic metabolic end products.** *Microbiology*. 2002, **148**, 4; 1015-1026.
204. Seballe B., Poole R.K. **Aerobic and anaerobic regulation of the *ubiCA* operon, encoding enzymes for the first two committed steps of ubiquinone biosynthesis in *Escherichia coli*.** *FEBS Lett.* 1997, **414**, 2; 373-376.
205. Iuchi S., Lin E.C. ***arcA* (*dye*), a global regulatory gene in *Escherichia coli* mediating repression of enzymes in aerobic pathways.** *Proc. Natl. Acad. Sci. U. S. A.* 1988, **85**, 6; 1888-1892.
206. Malpica R., Franco B., Rodriguez C., Kwon O., Georgellis D. **Identification of a quinone-sensitive redox switch in the ArcB sensor kinase.** *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**, 36; 11318-11323.
207. Georgellis D., Kwon O., Lin E.C. **Quinones as the redox signal for the *arc* two-component system of bacteria.** *Science*. 2001, **292**, 5525; 2314-2316.
208. Toro-Roman A., Mack T.R., Stock A.M. **Structural analysis and solution studies of the activated regulatory domain of the response regulator ArcA: a symmetric dimer mediated by the alpha4-beta5-alpha5 face.** *J. Mol. Biol.* 2005, **349**, 1; 11-26.
209. Jeon Y., Lee Y.S., Han J.S., Kim J.B., Hwang D.S. **Multimerization of phosphorylated and non-phosphorylated ArcA is necessary for the response regulator function of the Arc two-component signal transduction system.** *J. Biol. Chem.* 2001, **276**, 44; 40873-40879.
210. McGuire A.M., De Wulf P., Church G.M., Lin E.C. **A weight matrix for binding recognition by the redox-response regulator ArcA-P of *Escherichia coli*.** *Mol. Microbiol.* 1999, **32**, 1; 219-221.
211. Lynch A.S., Lin E.C. **Transcriptional control mediated by the ArcA two-component response regulator protein of *Escherichia coli*: characterization of DNA binding at target promoters.** *J. Bacteriol.* 1996, **178**, 21; 6238-6249.
212. Cho B.K., Knight E.M., Palsson B.O. **Transcriptional regulation of the *fad* regulon genes of *Escherichia coli* by ArcA.** *Microbiology*. 2006, **152**, 8; 2207-2219.
213. Chao G., Shen J., Tseng C.P., Park S.J., Gunsalus R.P. **Aerobic regulation of isocitrate dehydrogenase gene (*icd*) expression in *Escherichia coli* by the *arcA* and *fnr* gene products.** *J. Bacteriol.* 1997, **179**, 13; 4299-4304.

214. Tardat B., Touati D. **Iron and oxygen regulation of *Escherichia coli* MnSOD expression: competition between the global regulators Fur and ArcA for binding to DNA.** *Mol. Microbiol.* 1993, **9**, 1; 53-63.
215. Pellicer M.T., Lynch A.S., De Wulf P., Boyd D., Aguilar J., Lin E.C. **A mutational study of the ArcA-P binding sequences in the *aldA* promoter of *Escherichia coli*.** *Mol. Gen. Genet.* 1999b, **261**, 1; 170-176.
216. Pellicer M.T., Fernandez C., Badi'a J., Aguilar J., Lin E.C., Baldom L. **Cross-induction of *glc* and *ace* operons of *Escherichia coli* attributable to pathway intersection. Characterization of the *glc* promoter.** *J. Biol. Chem.* 1999a, **274**, 3; 1745-1752.
217. Jeong J.Y., Kim Y.J., Cho N., Shin D., Nam T.W., Ryu S., Seok Y.J. **Expression of *ptsG* encoding the major glucose transporter is regulated by ArcA in *Escherichia coli*.** *J. Biol. Chem.* 2004, **279**, 37; 38513-385138.
218. Mika F., Hengge R. **A two-component phosphotransfer network involving ArcB, ArcA, and RssB coordinates synthesis and proteolysis of sigmaS (*RpoS*) in *E. coli*.** *Genes Dev.* 2005, **19**, 22; 2770-2781.
219. Ogasawara H., Teramoto J., Yamamoto S., Hirao K., Yamamoto K., Ishihama A., Utsumi R. **Negative regulation of DNA repair gene (*uvrA*) expression by ArcA/ArcB two-component system in *Escherichia coli*.** *FEMS Microbiol. Lett.* 2005, **251**, 2; 243-249.
220. Lee A.I., Delgado A., Gunsalus R.P. **Signal-dependent phosphorylation of the membrane-bound NarX two-component sensor-transmitter protein of *Escherichia coli*: nitrate elicits a superior anion ligand response compared to nitrite.** *J. Bacteriol.* 1999, **181**, 17; 5309-5316.
221. Rabin R.S., Stewart V. **Dual response regulators (NarL and NarP) interact with dual sensors (NarX and NarQ) to control nitrate- and nitrite-regulated gene expression in *Escherichia coli* K-12.** *J. Bacteriol.* 1993, **175**, 11; 3259-3268.
222. Chiang R.C., Cavicchioli R., Gunsalus R.P. **'Locked-on' and 'locked-off' signal transduction mutations in the periplasmic domain of the *Escherichia coli* NarQ and NarX sensors affect nitrate- and nitrite-dependent regulation by NarL and NarP.** *Mol. Microbiol.* 1997, **24**, 5; 1049-1060.
223. Darwin A.J., Tyson K.L., Busby S.J., Stewart V. **Differential regulation by the homologous response regulators NarL and NarP of *Escherichia coli* K-12 depends on DNA binding site arrangement.** *Mol. Microbiol.* 1997, **25**, 3; 583-595.
224. Darwin A.J., Stewart V. **Expression of the *narX*, *narL*, *narP*, and *narQ* genes of *Escherichia coli* K-12: regulation of the regulators.** *J. Bacteriol.* 1995, **177**, 13; 3865-3869.
225. Wang H., Gunsalus R.P. **The *nrfA* and *nirB* nitrite reductase operons in *Escherichia coli* are expressed differently in response to nitrate than to nitrite.** *J. Bacteriol.* 2000, **182**, 20; 5813-5822.

226. Stewart V. **Requirement of Fnr and NarL functions for nitrate reductase expression in *Escherichia coli* K-12.** *J. Bacteriol.* 1982, **151**, 3; 1320-1325.
227. Baikalov I., Schroder I., Kaczor-Grzeskowiak M., Grzeskowiak K., Gunsalus R.P., Dickerson R.E. **Structure of the *Escherichia coli* response regulator NarL.** *Biochemistry.* 1996, **35**, 34; 11053-11061.
228. Maris A.E., Sawaya M.R., Kaczor-Grzeskowiak M., Jarvis M.R., Bearson S.M., Kopka M.L., Schroder I., Gunsalus R.P., Dickerson R.E. **Dimerization allows DNA target site recognition by the NarL response regulator.** *Nat. Struct. Biol.* 2002, **9**, 10; 771-778.
229. Bearson S.M., Albrecht J.A., Gunsalus R.P. **Oxygen and nitrate-dependent regulation of *dmsABC* operon expression in *Escherichia coli*: sites for Fnr and NarL protein interactions.** *BMC Microbiol.* 2002, **2**, 1; 13.
230. Tyson K.L., Cole J.A., Busby S.J. **Nitrite and nitrate regulation at the promoters of two *Escherichia coli* operons encoding nitrite reductase: identification of common target heptamers for both NarP- and NarL-dependent regulation.** *Mol. Microbiol.* 1994, **13**, 6; 1045-1055.
231. Williams S.B., Stewart V. **Nitrate- and nitrite-sensing protein NarX of *Escherichia coli* K-12: mutational analysis of the amino-terminal tail and first transmembrane segment.** *J. Bacteriol.* 1997, **179**, 3; 721-729.
232. Kolesnikow T., Schroder I., Gunsalus R.P. **Regulation of *narK* gene expression in *Escherichia coli* in response to anaerobiosis, nitrate, iron, and molybdenum.** *J. Bacteriol.* 1992, **174**, 22; 7104-71011.
233. Stewart V., Parales J. **Identification and expression of genes *narL* and *narX* of the *nar* (nitrate reductase) locus in *Escherichia coli* K-12.** *J. Bacteriol.* 1988, **170**, 4; 1589-1597.
234. Bongaerts J., Zoske S., Weidner U., Uden G. **Transcriptional regulation of the proton translocating NADH dehydrogenase genes (*nuoA-N*) of *Escherichia coli* by electron acceptors, electron donors and gene regulator.** *Mol. Microbiol.* 1995, **16**, 3; 521-534.
235. Kaiser M., Sawers G. **Nitrate repression of the *Escherichia coli* *pfl* operon is mediated by the dual sensors NarQ and NarX and the dual regulators NarL and NarP.** *J. Bacteriol.* 1995, **177**, 13; 3647-3655.
236. Grove J., Tanapongpipat S., Thomas G., Griffiths L., Croke H., Cole J. ***Escherichia coli* K-12 genes essential for the synthesis of c-type cytochromes and a third nitrate reductase located in the periplasm.** *Mol. Microbiol.* 1996, **19**, 3; 467-481.
237. Kalman L.V., Gunsalus R.P. **Identification of a second gene involved in global regulation of fumarate reductase and other nitrate-controlled genes for anaerobic respiration in *Escherichia coli*.** *J. Bacteriol.* 1989, **171**, 7; 3810.
238. Cotter P.A., Gunsalus R.P. **Oxygen, nitrate, and molybdenum regulation of *dmsABC* gene expression in *Escherichia coli*.** *J. Bacteriol.* 1989, **171**, 7; 3817-3823.

239. Iuchi S., Lin E.C. **The *narL* gene product activates the nitrate reductase operon and represses the fumarate reductase and trimethylamine N-oxide reductase operons in *Escherichia coli*.** *Proc. Natl. Acad. Sci. U. S. A.* 1987, **84**, 11; 3901-3905 .
240. Blattner F.R., Plunkett G., Bloch C.A., Perna N.T., Burland V., Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F., Gregor J., Davis N.W., Kirkpatrick H.A. , Goeden M.A., Rose D.J., Mau B., Shao Y. **The complete genome sequence of *Escherichia coli* K-12.** *Science.* 1997, **277**, 5331; 1453-1474.
241. Deng W., Liou S.R., Plunkett G., Mayhew G.F., Rose D.J., Burland V., Kodoyianni V., Schwartz D.C., Blattner F.R. **Comparative genomics of *Salmonella enterica* serovar Typhi strains Ty2 and CT18.** *J. Bacteriol.* 2003, **185**, 7; 2330-2337.
242. McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P., Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D., Leonard S., Nguyen C., Scott K. , Holmes A., Grewal N., Mulvaney E., Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M., Waterston R., Wilson R.K. **Complete genome sequence of *Salmonella enterica* serovar Typhimurium LT2.** *Nature.* 2001, **413**, 6858; 852-856.
243. Deng W., Burland V., Plunkett G., Boutin A., Mayhew G.F., Liss P., Perna N.T., Rose D.J., Mau B., Zhou S., Schwartz D.C., Fetherston J.D., Lindler L.E., Brubaker R.R., Plano G.V., Straley S.C., McDonough K.A., Nilles M.L., Matson J.S., Blattner F.R., Perry R.D. **Genome sequence of *Yersinia pestis* KIM.** *J. Bacteriol.* 2002, **184**, 16; 4601-4611.
244. Chain P.S., Carniel E., Larimer F.W., Lamerdin J., Stoutland P.O., Regala W.M., Georgescu A.M., Vergez L.M., Land M.L., Motin V.L., Brubaker R.R., Fowler J., Hinnebusch J., Marceau M., Medigue C., Simonet M., Chenal-Francisque V., Souza B., Dacheux D., Elliott J.M., Derbise A., Hauser L.J., Garcia E. **Insights into the evolution of *Yersinia pestis* through whole-genome comparison with *Yersinia pseudotuberculosis*.** *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**, 38; 13826-13831.
245. Thomson N.R., Howard S., Wren B.W., Holden M.T., Crossman L., Challis G.L., Churcher C., Mungall K., Brooks K., Chillingworth T., Feltwell T., Abdellah Z., Hauser H., Jagels K., Maddison M., Moule S., Sanders M., Whitehead S., Quail M.A., Dougan G., Parkhill J., Prentice M.B. **The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081.** *PLoS Genet.* 2006, **2** , 12; e206.
246. Bell K.S., Sebahia M., Pritchard L., Holden M.T., Hyman L.J., Holeva M.C., Thomson N.R., Bentley S.D., Churcher L.J., Mungall K., Atkin R., Bason N., Brooks K. , Chillingworth T., Clark K., Doggett J., Fraser A., Hance Z., Hauser H., Jagels K., Moule S., Norbertczak H., Ormond D., Price C., Quail M.A., Sanders M., Walker D., Whitehead S., Salmond G.P., Birch P.R., Parkhill J., Toth I.K. **Genome sequence of the enterobacterial phytopathogen *Erwinia carotovora* subsp. *atroseptica* and characterization of virulence factors.** *Proc. Natl. Acad. Sci. U. S. A.* 2004, **101**, 30; 11105-11110.
247. Duchaud E., Rusniok C., Frangeul L., Buchrieser C., Givaudan A., Taourit S., Bocs S., Boursaux-Eude C., Chandler M., Charles J.F., Dassa E., Deroose R., Derzelle S., Freyssinet G., Gaudriault S., Medigue C., Lanois A., Powell K., Siguier P., Vincent R., Wingate V., Zouine M., Glaser P., Boemare N., Danchin A., Kunst F. **The genome sequence of the entomopathogenic bacterium *Photorhabdus luminescens*.** *Nat. Biotechnol.* 2003, **21**, 11;

1307-1313.

248. May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V. **Complete genomic sequence of *Pasteurella multocida*, Pm70.** *Proc. Natl. Acad. Sci. U. S. A.* 2001, **98**, 6; 3460-3465.
249. Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L., Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A., Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D., Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P., McDonald L., Utterback T., Fleishmann R.D., Nierman W.C., White O., Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C., Fraser C.M. **DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.** *Nature.* 2000, **406**, 6795; 477-483.
250. Ruby E.G., Urbanowski M., Campbell J., Dunn A., Faini M., Gunsalus R., Lostroh P., Lupp C., McCann J., Millikan D., Schaefer A., Stabb E., Stevens A., Visick K., Whistler C., Greenberg E.P. **Complete genome sequence of *Vibrio fischeri*: A symbiotic bacterium with pathogenic congeners.** *Proc. Natl. Acad. Sci. U. S. A.* 2005, **102**, 8; 3004-3009.
251. Makino K., Oshima K., Kurokawa K., Yokoyama K., Uda T., Tagomori K., Iijima Y., Najima M., Nakano M., Yamashita A., Kubota Y., Kimura S., Yasunaga T., Honda T., Shinagawa H., Hattori M., Iida T. **Genome sequence of *Vibrio parahaemolyticus*: a pathogenic mechanism distinct from that of *V. cholerae*.** *Lancet.* 2003, **361**, 9359; 743-749.
252. Kim Y.R., Lee S.E., Kim C.M., Kim S.Y., Shin E.K., Shin D.H., Chung S.S., Choy H.E., Progulsk-Fox A., Hillman J.D., Handfield M., Rhee J.H. **Characterization and pathogenic significance of *Vibrio vulnificus* antigens preferentially expressed in septicemic patients.** *Infect. Immun.* 2003, **71**, 10; 5461-5471.
253. Vezzi A., Campanaro S., D'Angelo M., Simonato F., Vitulo N., Lauro F.M., Cestaro A., Malacrida G., Simionati B., Cannata N., Romualdi C., Bartlett D.H., Valle G. **Life at depth: *Photobacterium profundum* genome sequence and expression analysis.** *Science.* 2005, **307**, 5714; 1459-1461.
254. Nelson K.E., Weinel C., Paulsen I.T., Dodson R.J., Hilbert H., Martins dos Santos V.A., Fouts D.E., Gill S.R., Pop M., Holmes M., Brinkac L., Beanan M., DeBoy R.T., Daugherty S., Kolonay J., Madupu R., Nelson W., White O., Peterson J., Khouri H., Hance I., Chris Lee P., Holtzapple E., Scanlan D., Tran K., Moazzez A., Utterback T., Rizzo M., Lee K., Kosack D., Moestl D., Wedler H., Lauber J., Stjepandic D., Hoheisel J., Straetz M., Heim S., Kiewitz C., Eisen J.A., Timmis K.N., Dusterhoft A., Tumbler B., Fraser C.M. **Complete genome sequence and comparative analysis of the metabolically versatile *Pseudomonas putida* KT2440.** *Environ. Microbiol.* 2002, **4**, 12; 799-808.
255. Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T., Gwinn M.L., Dodson R.J., Deboy R.T., Durkin A.S., Kolonay J.F., Madupu R., Daugherty S., Brinkac L., Beanan M.J., Haft D.H., Nelson W.C., Davidsen T., Zafarm N., Zhou L., Liu J., Yuan Q., Khouri H., Fedorova N., Tran B., Russell D., Berry K., Utterback T., Van Aken S.E., Feldblyum T.V., D'Ascenzo M., Deng W.L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K., Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X., Bender C.L., White O., Fraser C.M., Collmer A. **The complete genome sequence of the *Arabidopsis* and tomato pathogen *Pseudomonas syringae* pv. tomato DC3000.** *Proc. Natl. Acad. Sci. U. S. A.* 2003, **100**, 18;

- 10181-10186.
256. Crooks G.E., Hon G., Chandonia J.M., Brenner S.E. **WebLogo: a sequence logo generator.** *Genome Res.* 2004, **14**, 6; 1188-1190.
 257. Schneider T.D., Stephens R.M. **Sequence logos: a new way to display consensus sequences.** *Nucleic Acids Res.* 1990, **18**, 20; 6097-6100.
 258. Dodd I.B., Egan J.B. **Improved detection of helix-turn-helix DNA-binding motifs in protein sequences.** *Nucleic Acids Res.* 1990, **18**, 17; 5019-5026.
 259. Cunningham L., Gruer M.J., Guest J.R. **Transcriptional regulation of the aconitase genes (*acnA* and *acnB*) of *Escherichia coli*.** *Microbiology.* 1997, **143**, 12; 3795-3805.
 260. Stolz B., Huber M., Markovic-Housley Z., Erni B. **The mannose transporter of *Escherichia coli*. Structure and function of the IIABMan subunit.** *J. Bacteriol.* 1993, **268**, 36; 27094-27099.
 261. Huber F., Erni B. **Membrane topology of the mannose transporter of *Escherichia coli* K12.** *Eur. J. Biochem.* 1996, **239**, 3; 810-817.
 262. Plumbridge J. **Control of the expression of the *manXYZ* operon in *Escherichia coli*: Mlc is a negative regulator of the mannose PTS.** *Mol. Microbiol.* 1998, **27**, 2; 369-380.
 263. Postma P.W., Lengeler J.W., Jacobson G.R. **Phosphoenolpyruvate: carbohydrate phosphotransferase systems of bacteria.** *Microbiol. Rev.* 1993, **57**, 3; 543-594.
 264. Miles J.S., Guest J.R., PMID: **Nucleotide sequence and transcriptional start point of the phosphomannose isomerase gene (*manA*) of *Escherichia coli*.** *Gene.* 1984, **32**, 1-2; 41-48.
 265. Postma, P.W., Lengeler, J.W., and Jacobson, G.R. **Phosphoenolpyruvate: carbohydrate phosphotransferase systems.** In Neidhardt F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and molecular biology, (1996), ASM Press, Washington, pp. 1149-1174.
 266. Geerse R.H., Izzo F., Postma P.W. **The PEP: fructose phosphotransferase system in *Salmonella typhimurium*: FPr combines enzyme IIIIFru and pseudo-HPr activities.** *Mol. Gen. Genet.* 1989, **216**, 2-3; 517-525.
 267. Pichersky E., Gottlieb L.D., Hess J.F. **Nucleotide sequence of the triose phosphate isomerase gene of *Escherichia coli*.** *Mol. Gen. Genet.* 1984, **195**, 1-2; 214-320.
 268. Seta F.D., Boschi-Muller S., Vignais M.L., Branlant G. **Characterization of *Escherichia coli* strains with *gapA* and *gapB* genes deleted.** *J. Bacteriol.* 1997, **179**, 16; 5218-5221.
 269. Yang Y., Zhao G., Man T.K., Winkler M.E. **Involvement of the *gapA*- and *epd* (*gapB*)-encoded dehydrogenases in pyridoxal 5'-phosphate coenzyme biosynthesis in *Escherichia coli* K-12.** *J. Bacteriol.* 1998, **180**, 16; 4294-4299 .
 270. Fraser H.I., Kvaratskhelia M., White M.F. **The two analogous phosphoglycerate mutases of**

- Escherichia coli*. *FEBS Lett.* 1999, **455**, 3; 344-348.
271. Quail M.A., Haydon D.J., Guest J.R. **The *pdhR-aceEF-lpd* operon of *Escherichia coli* expresses the pyruvate dehydrogenase complex.** *Mol. Microbiol.* 1994, **12**, 1; 95-104.
272. Quail M.A., Guest J.R. **Purification, characterization and mode of action of PdhR, the transcriptional repressor of the *pdhR-aceEF-lpd* operon of *Escherichia coli*.** *Mol. Microbiol.* 1995, **15**, 3; 519-529.
273. Garrido-Pertierra A., Cooper R.A. **Evidence for two distinct pyruvate kinase genes in *Escherichia coli* K-12.** *FEBS Lett.* 1983, **162**, 2; 420-422.
274. Muirhead H. **Isoenzymes of pyruvate kinase.** *Biochem. Soc. Trans.* 1990, **18**, 2; 193-196.
275. Weidner U., Geier S., Ptock A., Friedrich T., Leif H., Weiss H. **The gene locus of the proton-translocating NADH: ubiquinone oxidoreductase in *Escherichia coli*. Organization of the 14 genes and relationship between the derived proteins and subunits of mitochondrial complex I.** *J. Mol. Biol.* 1993, **233**, 1; 109-122.
276. Steuber J. **The C-terminally truncated NuoL subunit (ND5 homologue) of the Na⁺-dependent complex I from *Escherichia coli* transports Na⁺.** *J. Biol. Chem.* 2003, **278**, 29; 26817-26822.
277. Stolpe S., Friedrich T. **The *Escherichia coli* NADH:ubiquinone oxidoreductase (complex I) is a primary proton pump but may be capable of secondary sodium antiport.** *J. Biol. Chem.* 2004, **279**, 18; 18377-18383.
278. Cronan, J.E. and Laporte, D. **Tricarboxylic acid cycle and glyoxylate bypass.** In Neidhart F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology, (1996), ASM Press, Washington, pp. 206-216.
279. Frankel, D.G. **Glycolysis.** In Neidhardt F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and molecular biology, (1996), ASM Press, Washington, pp. 189-198.
280. Holms W.H. **Control of flux through the citric acid cycle and the glyoxylate bypass in *Escherichia coli*.** *Biochem. Soc. Symp.* 1987, **54**; 17-31.
281. Nimmo H.G., Borthwick A.C., Mansi E.M., Holms W.H., MacKintosh C., Nimmo G.A. **Regulation of the enzymes at the branchpoint between the citric acid cycle and the glyoxylate bypass in *Escherichia coli*.** *Biochem. Soc. Symp.* 1987, **54**; 93-101.
282. Clegg S., Yu F., Griffiths L., Cole J.A. **The roles of the polytopic membrane proteins NarK, NarU and NirC in *Escherichia coli* K-12: two nitrate and three nitrite transporters.** *Mol. Microbiol.* 2002, **44**, 1; 143-155.
283. Page L., Griffiths L., Cole J.A. **Different physiological roles of two independent pathways for nitrite reduction to ammonia by enteric bacteria.** *Arch. Microbiol.* 1990, **154**, 4; 349-354.

284. Tyson K.L., Bell A.I., Cole J.A., Busby S.J. **Definition of nitrite and nitrate response elements at the anaerobically inducible *Escherichia coli nirB* promoter: interactions between FNR and NarL.** *Mol. Microbiol.* 1993, **7**, 1; 151-157.
285. Peakman T., Crouzet J., Mayaux J.F., Busby S., Mohan S., Harborne N., Wootton J., Nicolson R., Cole J. **Nucleotide sequence, organisation and structural analysis of the products of genes in the *nirB-cysG* region of the *Escherichia coli* K-12 chromosome.** *Eur. J. Biochem.* 1990, **191**, 2; 315-323.
286. Jackson R.H., Cornish-Bowden A., Cole J.A. **Prosthetic groups of the NADH-dependent nitrite reductase from *Escherichia coli* K12.** *Biochem. J.* 1991, **193**, 3; 861-867.
287. Fagan M.J., Saier M.H. **P-type ATPases of eukaryotes and bacteria: sequence analyses and construction of phylogenetic trees.** *J. Mol. Evol.* 1994, **38**, 1; 57-99.
288. Zheng D., Constantinidou C., Hobman J.L., Minchin S.D. **Identification of the CRP regulon using in vitro and in vivo transcriptional profiling.** *Nucleic Acids Res.* 2004, **32**, 19; 5874-5893.
289. Busby S., Ebright R.H. **Transcription activation by catabolite activator protein (CAP).** *J. Mol. Biol.* 1999, **293**, 2; 199-213.
290. Tan K., Moreno-Hagelsieb G., Collado-Vides J., Stormo G.D. **A comparative genomics approach to prediction of new members of regulons.** *Genome Res.* 2001, **11**, 4; 566-584.
291. Chin A.M., Feldheim D.A., Saier M.H. **Altered transcriptional patterns affecting several metabolic pathways in strains of *Salmonella typhimurium* which overexpress the fructose regulon.** *J. Bacteriol.* 1989, **171**, 5; 2424-2434.
292. Perrenoud A., Sauer U. **Impact of global transcriptional regulation by ArcA, ArcB, Cra, Crp, Cya, Fnr, and Mlc on glucose catabolism in *Escherichia coli*.** *J. Bacteriol.* 2005, **187**, 9; 3171-3179.
293. Henikoff S., Henikoff J.G. **Performance evaluation of amino acid substitution matrices.** *Proteins.* 1993, **17**, 1; 49-61.
294. Price M.N., Dehal P.S., Arkin A.P. **Horizontal gene transfer and the evolution of transcriptional regulation in *Escherichia coli*.** *Genome Biol.* 2008, **9**, 1; R4.
295. Green S.M., Malik T., Giles I.G., Drabble W.T. **The *purB* gene of *Escherichia coli* K-12 is located in an operon.** *Microbiology.* 1996, **142**, 11; 3219-3230.
296. Marolewski A., Smith J.M., Benkovic S.J. **Cloning and characterization of a new purine biosynthetic enzyme: a non-folate glycinamide ribonucleotide transformylase from *E. coli*.** *Biochemistry.* 1994, **33**, 9; 2531-2537.
297. Andrews S.C., Guest J.R. **Nucleotide sequence of the gene encoding the GMP reductase of *Escherichia coli* K12.** *Biochem. J.* 1988, **255**, 1; 35-43.

298. Kawasaki H., Shimaoka M., Usuda Y., Utagawa T. **End-product regulation and kinetic mechanism of guanosine-inosine kinase from *Escherichia coli***. 2000, **64**, 5; 972-979.
299. Burns D.M., Beacham I.R. **Nucleotide sequence and transcriptional analysis of the *E. coli* *ushA* gene, encoding periplasmic UDP-sugar hydrolase (5'-nucleotidase): regulation of the *ushA* gene, and the signal sequence of its encoded protein product**. *Nucleic Acids Res.* 1986, **14**, 0; 4325-4342.
300. Andersen P.S., Frees D., Fast R., Mygind B. **Uracil uptake in *Escherichia coli* K-12: isolation of *uraA* mutants and cloning of the gene**. *J. Bacteriol.* 1995, **177**, 8; 2008-2013.
301. Andersen P.S., Smith J.M., Mygind B. **Characterization of the *upp* gene encoding uracil phosphoribosyltransferase of *Escherichia coli* K12**. *Eur. J. Biochem.* 1992, **204**, 1; 51-56.
302. Neughard, J. and Kelln, R.A. **Biosynthesis and conversions of pyrimidines**. In Neidhardt F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and molecular biology, (1990), ASM Press, Washington, pp. 580-599.
303. D'Ari L., Rabinowitz J.C. **Purification, characterization, cloning, and amino acid sequence of the bifunctional enzyme 5,10-methylenetetrahydrofolate dehydrogenase/5,10-methenyltetrahydrofolate cyclohydrolase from *Escherichia coli***. *J. Biol. Chem.* 1991, **266**, 35; 23953-23958.
304. Tobey K.L., Grant G.A. **The nucleotide sequence of the *serA* gene of *Escherichia coli* and the amino acid sequence of the encoded protein, D-3-phosphoglycerate dehydrogenase**. *J. Biol. Chem.* 1986, **261**, 26; 12179-12183.
305. Matthews, R.G. **One-carbon metabolism**. In Neidhardt F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology, (1996), ASM Press, Washington, pp. 600-611.
306. Crowley P.J., Gutierrez J.A., Hillman J.D., Bleiweis A.S. **Genetic and physiologic analysis of a formyl-tetrahydrofolate synthetase mutant of *Streptococcus mutans***. *J. Bacteriol.* 1997, **179**, 5; 1563-1572.
307. Danielsen S., Kilstrup M., Barilla K., Jochimsen B., Neuhard J. **Characterization of the *Escherichia coli* *codBA* operon encoding cytosine permease and cytosine deaminase**. *Mol. Microbiol.* 1992, **6**, 10; 1335-1344.
308. Maier C., Bremer E., Schmid A., Benz R. **Pore-forming activity of the Tsx protein from the outer membrane of *Escherichia coli*. Demonstration of a nucleoside-specific binding site**. *J. Biol. Chem.* 1998, **263**, 5; 2493-2499.
309. Burton K. **Adenine transport in *Escherichia coli***. *Proc. Biol. Sci.* 1994, **22**, 255; 153-157.
310. Rapp M., Drew D., Daley D.O., Nilsson J., Carvalho T., Melen K., De Gier J.W., Von Heijne G. **Experimentally based topology models for *E. coli* inner membrane proteins**. *Protein Sci.* 2004, **13**, 4; 937-945.
311. Booth I.R., Kleppang K.E., Kempell K.E. **A genetic locus for the GltII-glutamate**

- transport system in *Escherichia coli*.** *J. Gen. Microbiol.* 1989, **135**, 10; 2767-2774.
312. Bairoch A., Boeckmann B., Ferro S., Gasteiger E. **Swiss-Prot: Juggling between evolution and stability.** *Brief. Bioinform.* 2004, **5**, 1; 39-55.
313. Farriol-Mathis N., Garavelli J.S., Boeckmann B., Duvaud S., Gasteiger E., Gateau A., Veuthey A.-L., Bairoch A. **Annotation of post-translational modifications in the Swiss-Prot knowledge base.** *Proteomics.* 2004, **4**, 6; 1537-1550.
314. Chase J.W., Rabin B.A., Murphy J.B., Stone K.L., Williams K.R. ***Escherichia coli* exonuclease VII. Cloning and sequencing of the gene encoding the large subunit (*xseA*).** *J. Biol. Chem.* 1986, **261**, 32; 14929-14935.
315. Gerlach P., Valentin-Hansen P., Bremer E. **Transcriptional regulation of the *cytR* repressor gene of *Escherichia coli*: autoregulation and positive control by the cAMP/CAP complex.** *Mol. Microbiol.* 1990, **4**, 3; 479-488.
316. Delbaere L.T., Sudom A.M., Prasad L., Leduc Y., Goldie H. **Structure/function studies of phosphoryl transfer by phosphoenolpyruvate carboxykinase.** *Biochim. Biophys. Acta.* 2004, **1697**, 1-2; 271-278.
317. Niersbach M., Kreuzaler F., Geerse R.H., Postma P.W., Hirsch H.J. **Cloning and nucleotide sequence of the *Escherichia coli* K-12 *ppsA* gene, encoding PEP synthase.** *Mol. Gen. Genet.* 1992, **231**, 2; 332-336.
318. Donahue J.L., Bownas J.L., Niehaus W.G., Larson T.J. **Purification and characterization of *glpX*-encoded fructose 1, 6-bisphosphatase, a new enzyme of the glycerol 3-phosphate regulon of *Escherichia coli*.** *J. Bacteriol.* 2000, **182**, 19; 5624-5627.
319. Levine R.A., Taylor M.W. **Regulation of *purE* transcription in a *purE::lac* fusion strain of *Escherichia coli*.** *J. Bacteriol.* 1982, **149**, 3; 1041-1049.
320. Schumacher M.A., Choi K.Y., Lu F., Zalkin H., Brennan R.G. **Mechanism of corepressor-mediated specific DNA binding by the purine repressor.** *Cell.* 1995, **83**, 1; 147-155.
321. Nagadoi A., Morikawa S., Nakamura H., Enari M., Kobayashi K., Yamamoto H., Sampei G., Mizobuchi K., Schumacher M.A., Brennan R.G. **Structural comparison of the free and DNA-bound forms of the purine repressor DNA-binding domain.** *Structure.* 1995, **3**, 11; 1217-1224.
322. Lu F., Schumacher M.A., Arvidson D.N., Haldimann A., Wanner B.L., Zalkin H., Brennan R.G. **Structure-based redesign of corepressor specificity of the *Escherichia coli* purine repressor by substitution of residue 190.** *Biochemistry.* 1998, **37**, 4; 971-982.
323. Arvidson D.N., Lu F., Faber C., Zalkin H., Brennan R.G. **The structure of PurR mutant L54M shows an alternative route to DNA kinking.** *Nat. Struct. Biol.* 1998, **5**, 6; 436-441.
324. Glasfeld A., Koehler A.N., Schumacher M.A., Brennan R.G. **The role of lysine 55 in determining the specificity of the purine repressor for its operators through minor**

- groove interactions. *J. Mol. Biol.* 1999, **291**, 2; 347-361.**
325. Huffman J.L., Lu F., Zalkin H., Brennan R.G. **Role of residue 147 in the gene regulatory function of the *Escherichia coli* purine repressor.** *Biochemistry.* 2002, **41**, 2; 511-520.
326. Yang Z., Lu Z., Wang A. **Adaptive mutations in *Salmonella typhimurium* phenotypic of *purR* super-repression.** *Mutat. res.* 2006, **595**, 1-2; 107-116.
327. Yang Z., Lu Z., Wang A. **Study of adaptive mutations in *Salmonella typhimurium* by using a super-repressing mutant of a trans regulatory gene *purR*.** *Mutat. res.* 2001, **484**, 1-2; 95-102.
328. Lopilato J.E., Garwin J.L., Emr S.D., Silhavy T.J., Beckwith J.R. **D-ribose metabolism in *Escherichia coli* K-12: genetics, regulation, and transport.** *J. Bacteriol.* 1984, **158**, 2; 665-673.
329. Герасимова А.В., Родионов Д.А., Миронов А.А., Гельфанд М.С. **Компьютерный анализ регуляторных сигналов в бактериальных геномах. Участки связывания Fnr.** *Мол. Биол.* 2001, **35**, 6; 1001-1010.
330. Герасимова А.В., Гельфанд М.С., Макеев В.Ю., Миронов А.А., Фаворов А.В. **Регулятор ArcA в гамма-протеобактериях. Определение сайтов связывания и описание регулона.** *Биофизика.* 2003, **48**, 1; 21-25.
331. Kiley P.J., Beinert H. **The role of Fe-S proteins in sensing and regulation in bacteria.** *Curr. Opin. Microbiol.* 2003, **6**, 2; 181-185.
332. Bates D.M., Popescu C.V., Khoroshilova N., Vogt K., Beinert H., Muck E K.P.J. **Substitution of leucine 28 with histidine in the *Escherichia coli* transcription factor FNR results in increased stability of the [4Fe-4S]²⁺ cluster to oxygen.** *J. Biol. Chem.* 2000, **275**, 9; 6234-6240.
333. Gralnick J.A., Brown C.T., Newman D.K. **Anaerobic regulation by an atypical Arc system in *Shewanella oneidensis*.** *Mol. Microbiol.* 2005, **56**, 5; 1347-1357.
334. Stewart V. **Biochemical Society Special Lecture. Nitrate- and nitrite-responsive sensors NarX and NarQ of proteobacteria.** *Biochem. Soc. Trans.* 2003, **31**, 1; 1-10.
335. Stewart V., Chen L.L., Wu H.C. **Response to culture aeration mediated by the nitrate and nitrite sensor NarQ of *Escherichia coli* K-12.** *Mol. Microbiol.* 2003, **50**, 4; 1391-1399.
336. Stewart V., Lu Y., Darwin A.J. **Periplasmic nitrate reductase (NapABC enzyme) supports anaerobic respiration by *Escherichia coli* K-12.** *J. Bacteriol.* 2002, **184**, 5; 1314-1323.
337. Govantes F., Orjalo A.V., Gunsalus R.P. **Interplay between three global regulatory proteins mediates oxygen regulation of the *Escherichia coli* cytochrome *d* oxidase (*cydAB*) operon.** *Mol. Microbiol.* 2000, **38**, 5; 1061-1073.
338. Stewart V., Bledsoe P.J., Williams S.B. **Dual overlapping promoters control *napF***

- (periplasmic nitrate reductase) operon expression in *Escherichia coli* K-12. *J. Bacteriol.* 2003, **185**, 19; 5862-5870.
339. Stewart V., Bledsoe P.J. **Fnr-, NarP- and NarL-dependent regulation of transcription initiation from the *Haemophilus influenzae* Rd *napF* (Periplasmic Nitrate Reductase).** *J. Bacteriol.* 2005, **187**, 20; 6928-6935.
340. Melville S.B., Gunsalus R.P. **Mutations in *fnr* that alter anaerobic regulation of electron transport-associated genes in *Escherichia coli*.** *J. Biol. Chem.* 1990, **256**, 31; 18733-18736.
341. Wang H., Gunsalus R.P. **Coordinate regulation of the *Escherichia coli* formate dehydrogenase *fdnGHI* and *fdhF* genes in response to nitrate, nitrite, and formate: roles for NarL and NarP.** *J. Bacteriol.* 2003, **185**, 17; 5076-5085.
342. Liu X., De Wulf P. **Probing the ArcA-P modulon of *Escherichia coli* by whole genome transcriptional analysis and sequence recognition profiling.** *J. Biol. Chem.* 2004, **279**, 13; 12588-12597.
343. Green J., Guest J.R. **Regulation of transcription at the *ndh* promoter of *Escherichia coli* by FNR and novel factors.** *Mol. Microbiol.* 1994, **12**, 3; 433-444.
344. Iuchi S., Cole S.T., Lin E.C. **Multiple regulatory elements for the *glpA* operon encoding anaerobic glycerol-3-phosphate dehydrogenase and the *glpD* operon encoding aerobic glycerol-3-phosphate dehydrogenase in *Escherichia coli* : further characterization of respiratory control.** *J. Bacteriol.* 1990, **172**, 1; 179-184.
345. Nielsen J., Jorgensen B.B., van Meyenburg K.V., Hansen F.G. **The promoters of the *atp* operon of *Escherichia coli* K12.** *Mol. Gen. Genet.* 1984, **193**, 1; 64-71.
346. Kasimoglu E., Park S.J., Malek J., Tseng C.P., Gunsalus R.P. **Transcriptional regulation of the proton-translocating ATPase (*atpIBEFHAGDC*) operon of *Escherichia coli*: control by cell growth rate.** *J. Bacteriol.* 1996, **178**, 19; 5563-5567.
347. Salmon K., Hung S.P., Mekjian K., Baldi P., Hatfield G.W. **Global gene expression profiling in *Escherichia coli* K12. The effects of oxygen availability and FNR.** *J. Biol. Chem.* 2003, **278**, 32; 29837-29855.
348. Salmon K.A., Hung S.P., Steffan N.R., Krupp R., Baldi P., Hatfield G.W., Gunsalus R.P. **Global gene expression profiling in *Escherichia coli* K12: The effects of oxygen availability and ArcA.** *J. Biol. Chem.* 2005, **280**, 15; 15084-15096.
349. Цыганова М.О., Гельфанд М.С., Равчев Д.А. **Регуляция дыхания у энтеробактерий: сопоставление данных по экспрессии на биочипах и сравнительно-геномного анализа.** *Мол. Биол.* 2007, **41**, 3; 556-571.
350. Beattie P., Tan K., Bourne R.M., Leach D., Rich P.R., Ward FB. **Cloning and sequencing of four structural genes for the Na⁺-translocating NADH-ubiquinone oxidoreductase of *Vibrio alginolyticus*.** *FEBS Lett.* 1994, **356**, 2-3; 333-338.

351. Tokuda H., Nakamura T., Unemoto T. **Potassium ion is required for the generation of pH-dependent membrane potential and delta pH by the marine bacterium *Vibrio alginolyticus*.** *Biochemistry*. 1981, **20**, 14; 4198-4203.
352. Gon S., Patte J.C., Mejean V., Iobbi-Nivol C. **The *torYZ* (*yecK-bisZ*) operon encodes a third respiratory trimethylamine N-oxide reductase in *Escherichia coli*.** *J. Bacteriol.* 2000, **182**, 20; 5779-5786 .
353. Mejean V., Iobbi-Nivol C., Lepelletier M., Giordano G., Chippaux M., Pascal M.C. **TMAO anaerobic respiration in *Escherichia coli*: involvement of the *tor* operon.** *Mol. Microbiol.* 1994, **11**, 6; 1169-1179.
354. Abaibou H., Pommier J., Benoit S., Giordano G., Mandrand-Berthelot M.A. **Expression and characterization of the *Escherichia coli* *fdo* locus and a possible physiological role for aerobic formate dehydrogenase.** *J. Bacteriol.* 1995, **177**, 24; 7141-7149.
355. Stewart V., Lin J.T., Berg. B.L. **Genetic evidence that genes *fdhD* and *fdhE* do not control synthesis of formate dehydrogenase-N in *Escherichia coli* K-12.** *J. Bacteriol.* 1991, **173**, 14; 4417-4423.
356. Schlindwein C., Giordano G., Santini C.L., Mandrand M.A. **Identification and expression of the *Escherichia coli* *fdhD* and *fdhE* genes, which are involved in the formation of respiratory formate dehydrogenase.** *J. Bacteriol.* 1990, **172**, 10; 6112-6121.
357. Lobočka M., Hennig J., Wild J., Kłopotowski T. **Organization and expression of the *Escherichia coli* K-12 *dad* operon encoding the smaller subunit of D-amino acid dehydrogenase and the catabolic alanine racemase.** *J. Bacteriol.* 1994, **176**, 5; 1500-1510.
358. Anderson L.A., McNairn E., Lubke T., Pau R.N., Boxer D.H. **ModE-dependent molybdate regulation of the molybdenum cofactor operon *moa* in *Escherichia coli*.** *J. Bacteriol.* 2000, **182**, 24; 7035-7043.
359. Park S.J., McCabe J., Turna J., Gunsalus R.P. **Regulation of the citrate synthase (*gltA*) gene of *Escherichia coli* in response to anaerobiosis and carbon supply: role of the *arcA* gene product.** *J. Bacteriol.* 1994, **176**, 16; 5086-5092.
360. Park S.J., Chao G., Gunsalus R. **Aerobic regulation of the *sucABCD* genes of *Escherichia coli*, which encode alpha-ketoglutarate dehydrogenase and succinyl coenzyme A synthetase: roles of ArcA, Fnr, and the upstream *sdhCDAB* promoter.** *J. Bacteriol.* 1997, **178**, 13; 4138-4142.
361. Golby P., Kelly D.J., Guest J.R., Andrews S.C. **Transcriptional regulation and organization of the *dcuA* and *dcuB* genes, encoding homologous anaerobic C4-dicarboxylate transporters in *Escherichia coli*.** *J. Bacteriol.* 1998, **180**, 24; 6586-6596.
362. Park S.J., Gunsalus R.P. **Oxygen, iron, carbon, and superoxide control of the fumarase *fumA* and *fumC* genes of *Escherichia coli*: role of the *arcA*, *fnr*, and *soxR* gene products.** *J. Bacteriol.* 1995, **177**, 21; 6255-6262.
363. Park S.J., Cotter P.A., Gunsalus R.P. **Regulation of malate dehydrogenase (*mdh*) gene**

- expression in *Escherichia coli* in response to oxygen, carbon, and heme availability.** *J. Bacteriol.* 1995, **177**, 22; 6652-6656.
364. Jiang G.R., Nikolova S., Clark D.P. **Regulation of the *ldhA* gene, encoding the fermentative lactate dehydrogenase of *Escherichia coli*.** *Microbiology.* 2001, **147**, 9; 2437-2446.
365. Chen Y.M., Lin E.C. **Regulation of the *adhE* gene, which encodes ethanol dehydrogenase in *Escherichia coli*.** *J. Bacteriol.* 1991, **173**, 24; 8009-8013.
366. Membrillo-Hernandez J., Lin E.C. **Regulation of expression of the *adhE* gene, encoding ethanol oxidoreductase in *Escherichia coli*: transcription from a downstream promoter and regulation by *fnr* and *RpoS*.** *J. Bacteriol.* 1999, **181**, 24; 7571-7579.
367. Rest M.E., Frank C., Molenaar D. **Functions of the membrane-associated and cytoplasmic malate dehydrogenases in the citric acid cycle of *Escherichia coli*.** *J. Bacteriol.* 2000, **182**, 24; 6892-6899.
368. Saier, M.H., Ramseier, T.M., and Reizer, J. **Regulation of carbon utilization.** In Neidhart F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology, (1996), ASM Press, Washington, pp. 1325-1333.
369. Sprenger G.A., Schorken U., Sprenger G., Sahm H. **Transaldolase B of *Escherichia coli* K-12: cloning of its gene, *talB*, and characterization of the enzyme from recombinant strains.** *J. Bacteriol.* 1995, **177**, 20; 5930-5936.
370. Xu J., Johnson R.C. ***aldB*, an *RpoS*-dependent gene in *Escherichia coli* encoding an aldehyde dehydrogenase that is repressed by *Fis* and activated by *Crp*.** *J. Bacteriol.* 1995, **177**, 11; 3166-3175.
371. Figge R.M., Ramseier T.M., Saier M.H. **The mannitol repressor (*MtlR*) of *Escherichia coli*.** *J. Bacteriol.* 1994, **176**, 3; 840-847.
372. Plumbridge J.A., Cochet O., Souza J.M., Altamirano M.M., Calcagno M.L., Badet B. **Coordinated regulation of amino sugar-synthesizing and -degrading enzymes in *Escherichia coli* K-12.** *J. Bacteriol.* 1993, **175**, 16; 4951-4956.
373. Yang H., Liu M.Y., Romeo T. **Coordinate genetic regulation of glycogen catabolism and biosynthesis in *Escherichia coli* via the *CsrA* gene product.** *J. Bacteriol.* 1996, **178**, 4; 1012-1017.
374. Clark, D.P. and Cronan, J.E. **Two-carbon compounds and fatty acids as carbon sources.** In Neidhart F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and Molecular Biology, (1996), ASM Press, Washington, pp. 343-357.
375. Campbell J.W., Morgan-Kiss R.M., Cronan J.E. **A new *Escherichia coli* metabolic competency: growth on fatty acids by a novel anaerobic beta-oxidation pathway.** *Mol. Microbiol.* 2003, **47**, 3; 793-805.

376. Weimar J.D., DiRusso C.C., Delio R., Black P.N. **Functional role of fatty acyl-coenzyme A synthetase in the transmembrane movement and activation of exogenous long-chain fatty acids. Amino acid residues within the ATP/AMP signature motif of *Escherichia coli* FadD are required for enzyme activity and fatty acid transport.** *J. Biol. Chem.* 2002, **277**, 33; 29369-29376.
377. Cronan, J.E. and Rock, C.O. **Biosynthesis of membrane lipids.** In Neidhardt F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and molecular biology, (1996), ASM Press, Washington, pp. 612-636.
378. Lin, E.C.C. **Amino acids as carbon sources.** In Neidhardt F.C. (eds.) *Escherichia coli* and *Salmonella*. Cellular and molecular biology, (1996), ASM Press, Washington, pp. 358-379.
379. Hassan H.M., Sun H.C. **Regulatory roles of Fnr, Fur, and Arc in expression of manganese-containing superoxide dismutase in *Escherichia coli*.** *Proc. Natl. Acad. Sci. U. S. A.* 1992, **89**, 8; 3217-3221.
380. Zientz E., Janausch I.G., Six S., Uden G. **Functioning of DcuC as the C4-dicarboxylate carrier during glucose fermentation by *Escherichia coli*.** *J. Bacteriol.* 1999, **181**, 12; 3716-3720.
381. Ailion M., Bobik T.A., Roth J.R. **Two global regulatory systems (Crp and Arc) control the cobalamin/propanediol regulon of *Salmonella typhimurium*.** *J. Bacteriol.* 1993, **175**, 22; 7200-7208.
382. Kammler M., Schon C., Hantke K. **Characterization of the ferrous iron uptake system of *Escherichia coli*.** *J. Bacteriol.* 1993, **175**, 19; 6212-6219.
383. Porco A., Alonso G., Isturiz T. **The gluconate high affinity transport of GntI in *Escherichia coli* involves a multicomponent complex system.** *J. Basic Microbiol.* 1998, **38**, 5-6; 395-404.
384. Black P.N. **Primary sequence of the *Escherichia coli* fadL gene encoding an outer membrane protein required for long-chain fatty acid transport.** *J. Bacteriol.* 1991, **173**, 2; 435-442.
385. Lombardo M.J., Lee A.A., Knox T.M., Miller C.G. **Regulation of the *Salmonella typhimurium* pepT gene by cyclic AMP receptor protein (CRP) and FNR acting at a hybrid CRP-FNR site.** *J. Bacteriol.* 1997, **179**, 6; 1909-1917.
386. Strauch K.L., Lenk J.B., Gamble B.L., Miller C.G. **Oxygen regulation in *Salmonella typhimurium*.** *J. Bacteriol.* 1985, **161**, 2; 673-680.
387. Compan I., Touati D. **Anaerobic activation of arcA transcription in *Escherichia coli*: roles of Fnr and ArcA.** *Mol. Microbiol.* 1994, **11**, 5; 955-964.
388. Schneider R., Lurz R., Luder G., Tolksdorf C., Travers A., Muskhelishvili G. **An architectural role of the *Escherichia coli* chromatin protein FIS in organising DNA.** *Nucleic Acids Res.* 2001, **29**, 24; 5107-5114 .

389. Ninnemann O., Koch C., Kahmann R. **The *E.coli fis* promoter is subject to stringent control and autoregulation.** *EMBO J.* 1992, **11**, 3; 1075-1083.
390. Raivio T.L. **Envelope stress responses and Gram-negative bacterial pathogenesis.** *Mol. Microbiol.* 2005, **56**, 5; 1119-1128.
391. Iuchi S., Furlong D., Lin E.C. **Differentiation of *arcA*, *arcB*, and *cpxA* mutant phenotypes of *Escherichia coli* by sex pilus formation and enzyme regulation.** *J. Bacteriol.* 1989, **171**, 5; 2889-2893.
392. Christman M.F., Storz G., Ames B.N. **OxyR, a positive regulator of hydrogen peroxide-inducible genes in *Escherichia coli* and *Salmonella typhimurium*, is homologous to a family of bacterial regulatory proteins.** *Proc. Natl. Acad. Sci. U. S. A.* 1989, **86**, 10; 3484-3488 .
393. Bagg A., Neilands J.B. **Ferric uptake regulation protein acts as a repressor, employing iron (II) as a cofactor to bind the operator of an iron transport operon in *Escherichia coli*.** *Biochemistry.* 1987, **26**, 17; 5471-5477.
394. Sutton V.R., Mettert E.L., Beinert H., Kiley P.J. **Kinetic analysis of the oxidative conversion of the [4Fe-4S]²⁺ cluster of FNR to a [2Fe-2S]²⁺ cluster.** *J. Bacteriol.* 2004, **186**, 23; 8018-8025.
395. Georgellis D., Kwon O., Lin E.C., Wong S.M., Akerley B.J. **Redox signal transduction by the ArcB sensor kinase of *Haemophilus influenzae* lacking the PAS domain.** *J. Bacteriol.* 2001, **183**, 24; 7206-7212.
396. Sengupta N., Paul K., Chowdhury R. **The global regulator ArcA modulates expression of virulence factors in *Vibrio cholerae*.** *Infect. Immun.* 2003, **71**, 10; 5583-5589.

Приложение 1.

Множественное выравнивание для белков FruR. Участки, соответствующие последовательности НТН, показаны белым шрифтом на темном фоне. Условные обозначения: “*” – абсолютно консервативная позиция; “:” – позиция с высокой степенью консервативности; “.” – позиция с низкой степенью консервативности. Обозначения геномов: см. Табл. 2.

```

FruR (PMU)  MKLKLVELAKLAGVSRTTASYVVNGKAKAYRVSDKTIKVEKLIKQYDFKPNAMAAGLRAGKSRTIGLIVPDFENA
FruR (AAC)  -----MAASLRVGKSKTIGLVIPDFENI
FruR (YPK)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (YPS)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (YEN)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (SMA)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (PCA)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (PCH)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (STY)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (STM)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (ECO)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (PLU)  --MKLDEIARLAGVSRTTASYVINGKAKQYRVSDKTVDKVMAVVREHNYHPNAVAAGLRAGRTRSIGLVIPDLENT
FruR (VFI)  --MTLDEIAKLAGVSKTTASYVINGKAQKYRISEKTLQKVMVAVVEHNYRDPDHAASALRAGSSRSFGLIIPDLENT
FruR (VCH)  --MTLDEIAKLAGVSKTTASYVINGKAQKYRISEKTLQKVMVAVVEHNYRDPDHAASALRAGNSRSFGLIIPDLENT
FruR (VPA)  --MTLDEIAKLAGVSKTTASYVINGKAQKYRISEKTLQKVMVAVVEHNYRDPDHAASALRAGNSRSFGLIIPDLENT
FruR (VVU)  --MTLDEIAKLAGVSKTTASYVINGKAKKYRISEKTLQKVMVAVVEHNYRDPDHAASALRAGNSRSFGLIIPDLENT
FruR (PPR)  --MTLDEIAKLAGVSRTTASYVINGKASKYRISEKTLQKVMVAVVEHNYRDPDHAASALRAGNSRSFGLIIPDLENS
FruR (PST)  --MKLSDIARLAGVSRTTASYVINGKAEQQRISPATVERVKAVVEQYDFRPNPQAAGLSRHSKTLGFIIPDLENP
FruR (PFO)  --MKLSDIARLAGVSRTTASYVINGKAEQQRISTATVERVRAVVDLHGFTPNPQAAGLSRHSRTRTLGFIIPDLENP
FruR (PPU)  --MKLSDIARLAGVSRTTASYVINGKAEQQRISNSTVERVRAVVEAHGFTPNPQAAGLSRHSRTRTLGFIIPDLENP
FruR (PAE)  --MKLSDIARLAGVSRTTASYVINGKAAQRRISAATVERVLAVVEEHGYQPDQQAAGLRRGQTRTLGFIIPDLENP
                                                    *:.** :.:*:.: **:*

```

```

FruR (PMU)  SFARIATHLEAGFRQKGYQLVIACSNNDNPETEMQCAKHLIQRQIDALIVSTTLPADTDFYQRHAN--IPIICFDRI
FruR (AAC)  SYAKIANNLENRLREKGYQLLIACSNNDKVENEQDCVKHLLQRQVDALLVSSALPNTDFYQ-NAS--VPIIGFDRR
FruR (YPK)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCVEHLLQRQVDALIVSTALPPEHPFYQRWAKDPFPIIALDRA
FruR (YPS)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCVEHLLQRQVDALIVSTALPPEHPFYQRWAKDPFPIIALDRA
FruR (YEN)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCIEHLLQRQVDALIVSTALPPEHPFYQRWANDPLPIIALDRA
FruR (SMA)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCIEHLLQRQVDALIVSTALPPEHPFYQRWANDPLPIIALDRA
FruR (PCA)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCIEHLLQRQVDALIVSTALPPEHPFYQRWINGGLPIIALDRA
FruR (PCH)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCIEHLLQRQVDALIVSTALPPEHPFYQRWANDPLPIIALDRA
FruR (STY)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCIEHLLQRQVDALIVSTSLPPEHPFYQRWANDPFIIVALDRA
FruR (STM)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCIEHLLQRQVDALIVSTSLPPEHPFYQRWANDPFIIVALDRA
FruR (ECO)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCIEHLLQRQVDALIVSTSLPPEHPFYQRWANDPFIIVALDRA
FruR (PLU)  SYTRIANYLERQARQRYQLLIACSEDQPDNEMRCVEHLLQRQVDALIVSTALPPEHPFYQRWASRSLPIIALDRA
FruR (VFI)  SYAKLAKLLEFNSRKAGYQILIGSDDDPETEMNVAHALISRRIDALFVASCLPNGSDYYLGIQEKGTPIIAIDRS
FruR (VCH)  SYARLAKLLEQNSRQAGYQILIGSDDDPQIEMAAAEALVSRRIDALFVASGIPSAEYYLKLQSGTPVIAIDRA
FruR (VPA)  SYARLAKLLEQNSRKAGYQILIGSDDDPETERKVAEALISRRIDALFVASGMPANEYYLKLQNSGTPVIALDRP
FruR (VVU)  SYARLAKLLEQNSRKAGYQILIGSDDDPETERKVDALVSRRIDALFVATSMADANEFYLAMQKSGTPVIALDRP
FruR (PPR)  SYAKLAKLLEARNARKAGYQLIISCSDDDPDTEMKVAETLLSRRIDALLVASVLPADNDFYPRIQTGGVVPVIALDRA
FruR (PST)  SYARIAKLEQGARSLGYQLLIASDDDPVSELQLLQFRARRCDALLVASCLPPSDASYRDLQKGMPIIAIDRE
FruR (PFO)  SYARIAKLEQGARARGYQLLIASDDAPDSERQLLQFRARRCDALIVASCLPAGDSDSYRELQDKGLPIIAIDRV
FruR (PPU)  SYARIAKLEQGARARGYQLLIASDDQPDSEQLLQFRARRCDALFVASCLPPEDDSYRELQDKGLPIIAIDRR
FruR (PAE)  SYARLAKLEQGARARGYQLLIACSDDEPETERQVNLFRARRCDALIVASCLPAGDDTHARLVAEGLPVVIAIDRQ
*:.:*. : * *:.:*. * * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

```

FruR (PMU) LDLNK-IRVLTDDDEGDSYRLATHLGR--HCEGKRMFLFGALPKLGLSRKRELGFRQALSD-KVNQVDY-LYASQFH
 FruR (AAC) ISSPKVVNALTDDDEGDAYRLTKELLLKIPH-Q-KRILFFGALPDLPTSREREEGFRSALRY-RKASAKY-LYGDMFR
 FruR (YPK) LDREHFISVVGADQEDAQMLAQELRSFPA---ESVLYLGALPELSVSFLREMGFREAWKD-DPRKVEY-LYANSYE
 FruR (YPS) LDREHFISVVGADQEDAQMLAQELRSFPA---ESVLYLGALPELSVSFLREMGFREAWKD-DPRKVEY-LYANSYE
 FruR (YEN) LDREHFISVVGADQEDAQMLAQELRSFPA---ESVLYLGALPELSVSFLREMGFREAWKD-DPRKVDY-LYANSYE
 FruR (SMA) LDREHFISVVGADQEDAQMLAQELRSFPA---ESVLYLGALPELSVSFLREMGFRQAWQE-DPRHVDY-LYANSYE
 FruR (PCA) LDREHFTSVVGADLEDAEMLAQELRKMPA---KSVLYLGALPELSVSFLREMGFRQAWAG-DPNY-LYSNSYE
 FruR (PCH) LDREHFTSVVGADMEDAEMLAQELRKVPA---ESVMYLGALPELSVSFLREMGFRRAWAE-DPRQVNY-IYANSYE
 FruR (STY) LDREHFTSVVGADQDDAEMLAELRKFFPA---ETVLYLGALPELSVSFLREMGFRRTAWKD-DPNF-LYANSYE
 FruR (STM) LDREHFTSVVGADQDDAEMLAELRKFFPA---ETVLYLGALPELSVSFLREMGFRRTAWKD-DPNF-LYANSYE
 FruR (ECO) LDREHFTSVVGADQDDAEMLAELRKFFPA---ETVLYLGALPELSVSFLREMGFRRTAWKD-DPHF-LYANSYE
 FruR (PLU) LENEHFISVVGDDFEDAKMLAQEFQFPA---ESVLYLGALPELSVSFLREMGFRREAWKD-DPRKVMY-LYSNSYE
 FruR (VFI) LDDEHFGCVI SEDFDAAYELTKSVNVENV---KSIGLIGALPELNI SRERQLGF EAKAKE-K-GLASIMGYGEHFN
 FruR (VCH) LDDEYFSCVISEDFGAAFELTRSVLTQDV---HSVGLVGALPELNVSREREQGFAMAVKQ-R-GLPTTLGYGEHFN
 FruR (VPA) MDDEHFCVISEDFDAAFELTESVLSPEI---KTIGLIGALQDLQVSKERELGFRSACQK-D-DTAMTVGYGLHFS
 FruR (VVU) LDDEFFSCVISEDFGAAFELTSSLMSSEI---QTVLIGALPELNI SRERHQGYLSAVKMMG-KQAQS-GYGEHFI
 FruR (PPR) LDDEIFASVISEDLEGAFELTQTLNKGV---NSIGLVGAVPELVGSKEREQGF LTAIRRHQPDINMH IAYGDHFS
 FruR (PST) MDPNFCSVVSDDHDASLQLTRSL LQTP---RHIALIGARPELSISRARAAGFEHALDGF A-GTTIV-EHGEAFS
 FruR (PFO) MDAEHFCSVVSDDREASLHLTESL LDPQP---KQIVLLGARPELSISQERAAGFKQALAGFK-GEVLI-EHAESFS
 FruR (PPU) LDPAHFCSVVSDDRDASRLAASLSSAP---RSIALIGARPELSVQARAGGFDEALQGYT-GEVRR-YQGEAFS
 FruR (PAE) LDPQRFASVVSDDQDAARRLTETLLQPAP---RHIALLGARADLIISHDRAAGFRQALKDFR-GEVIV-EHAEVFS
 :. :. * : * : . : .** . * * * * :

FruR (PMU) KLASAKAFHHWLEK-HEMPEAIFATSLTLLHG VFQILIKRY-HRIPRSLI IATFGHHMDL DLENKVICCVQNH EK
 FruR (AAC) KNEAAQCF AQWLEN-NPLPTAIFTTSFTLLEGLVLLVLLQKL-GQVPKDLVVATFGHVEMLELLENRVLCSVQDYDK
 FruR (YPK) REAAGVLF AEWLKT-HPMPQALFTTSFQLLQGVM DVTLKQS-GR LPCNLAIATFGDNELLD FLECPVLA VAQRHRD
 FruR (YPS) REAAGVLF AEWLKT-HPMPQALFTTSFQLLQGVM DVTLKQS-GR LPCNLAIATFGDNELLD FLECPVLA VAQRHRD
 FruR (YEN) REAAGVLF AQWLKT-HPMPQALFTTSFQLLQGVM DVTLKQS-GR LPCNLAIATFGDNELLD FLECPVLA VAQRHRD
 FruR (SMA) REAAGALFAEWLKT-HPMPQALFTTSF SLLQGVM DVTLKQR-GR LPTDLAIATFGDHELLD FLECPVLA VAQRHRD
 FruR (PCA) RVASAAAFMDYLQN-NPMPEALFTTSF PLLQGVM DVNLKIN-GR LPNNLAIATFGDNELLD YLECPVLSVAQRHRE
 FruR (PCH) RSAAA AVFEEYLKT-HPMPQALFTTSF PLLQGVM DVTLRQS-GR LPNNLAIATFGDNELLD FLECPVLSVAQRHRE
 FruR (STY) REAAAQLFEKWLET-HPMPQALFTTSF ALLQGVM DVTLRRD-GKLP SDLAIATFGDHELLD FLQCPVLA VAQRHRD
 FruR (STM) REAAAQLFEKWLET-HPMPQALFTTSF ALLQGVM DVTLRRD-GKLP SDLAIATFGDHELLD FLQCPVLA VAQRHRD
 FruR (ECO) REAAAQLFEKWLET-HPMPQALFTTSF ALLQGVM DVTLRRD-GKLP SDLAIATFGDHELLD FLQCPVLA VAQRHRD
 FruR (PLU) RDAAA EVFTKWLET-HPMPQALFTTSF ALLQGVM DVTLKHS-GR LPNQLVIATFGDHELLD FLECPVLSVAQRHRE
 FruR (VFI) SESGKAIFEQWIKN-DTIPDAIITTSY ILLEGILDV LIEKP-ELMTR-IKLGTFGDNRLLD FLPIKVN SLPQQFEL
 FruR (VCH) REEGRKVF AKWVAN-DQLPDAVATS YTLLEGILDV LLEQP-ELMQK-VRLATFGDNRLLD FLPIRVN SLPQQFEL
 FruR (VPA) REEGKVL DKKVQA-DSL PDAIVTTSY TLLEGVLDV MLEKP-ELMSQ-VRLATFGDNRLLD FLPVKIN SLPQQFEL
 FruR (VVU) REEGYRIMHEWLLA-DTIPDAIVTTSY TLLEGVLDV LLEQP-ELVNK-VKLATFGDNRLLD FLPFKVN SLPQQFEL
 FruR (PPR) QEQQQKQVQQWIDS-EHFPDAI LATS YTLFEGILDCLLAN-ELMSK-IHLATFGDNRLLD FLPIKIQSLPQQFEL
 FruR (PST) RQCGQRLMSEMLQRQGHFPDALITTSYVLLQGVFDV LQSQSLNTTS--LHLGTFGDTQLLD FLPLPVNAMAQQHQ L
 FruR (PFO) RECCKQLMEELLQRLGHL PDALITTSYVLLQGVFDV LQSDPLKSRP--LRLGTFGDTQLLD FLPLPVNAMAQQHQ L
 FruR (PPU) RECCKQLMQQLIDDLGGLPDALVTTSYVLLQGVFDV LQARPVDSRQ--LQLGTFGDNQLLD FLPLPVNAMAQQHQ GQ
 FruR (PAE) RECGRRLMEQLLERLGYLPDALVT TAYVLLLEGVFDV FQARA-DGWPEGLRVATFGDTQLLD FVPLKVN SIYQQHAL
 . . . : : * * : : * : * : : : : . * .

FruR (PMU) IAQTLLHLVLSRM--KSKNVKSDEVI IREIIE-----SRI----
 FruR (AAC) VSQSLTLTTFDAMSKKHKN-SVNTPLTRS I IK-----YHC---
 FruR (YPK) VAERVLELVLASLDEPHKP-KPGLTRIRRNLFRRGTL SLSRR---
 FruR (YPS) VAERVLELVLASLDEPHKP-KPGLTRIRRNLFRRGTL SLSRR---
 FruR (YEN) VAERVLELVLASLDEPHKP-KPGLTRIRRNLFRRG--SLSR----
 FruR (SMA) VAERVLELVLASLDEPRKP-KPGLTRIRRNLFRRG--SLSR----
 FruR (PCA) VAERVLELVLASLDEPKRP-KPGLNRIRRNLYRRG--SLSRM---
 FruR (PCH) VAERVLELVLASLDEPRKP-KPGLYRIRRNLYRRG--FLSRL---
 FruR (STY) VAERVLEIVLASLDEPRKP-KPGLTRIRRNLYRRG--ILSRS---
 FruR (STM) VAERVLEIVLASLDEPRKP-KPGLTRIRRNLYRRG--ILSRS---
 FruR (ECO) VAERVLEIVLASLDEPRKP-KPGLTRIKRNLYRRG--VLSRS---
 FruR (PLU) IAERVLELVLASLDEKQKP-GPGITRIHRDL CRRG--SLSRKI--
 FruR (VFI) IADSALALGLNA--SAKRY-QAGIELIPRKIVI-----
 FruR (VCH) IADSALALALNA--SAKRY-QTGIELIPRQLKV-----RT---
 FruR (VPA) IADSAMELALNA--SAKRY-KPGVELIPRKIVK-----RQ---
 FruR (VVU) IADSALELALNA--SAKRY-KPGVELIPRKIVR-----RF---
 FruR (PPR) IADNALELALNA--AAGRY-RTGVEVVP RKIKR-----R---
 FruR (PST) IADKALQALAAI-EKDDY-QPGVHAIARAFKQ-----RIHGV
 FruR (PFO) IADKALELALAAV-EQSDY-QPGVQAIARTFKQ-----RIHRD
 FruR (PPU) IAATALELALAAI-EEKRY-EPGVHAVGRTFKQ-----RISVA
 FruR (PAE) IAANALELALRAV-EENDY-RPGLHAVPRQLKL-----R-D--
 :: : : : . : * :

Приложение 2.

Последовательности потенциальных сайтов связывания белка FruR. Указано положение сайта относительно старта транскрипции гена. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>E. coli</i>	<i>fruBKA</i>	-101	4,95	ctTGAAaCGtTTCAGC
		-34	5,17	GCTGAATCGtTTCAat
	<i>mtlADR</i>	-124	4,52	agTtAAcCGATTCAGt
	<i>ptsHI-crr</i>	-246	3,93	aaTtAATCGtTaCAGg
		-220	4,50	GCTGAATCGATTTtat
	<i>manXYZ</i>	-83	3,86	tCTGAATCGATTCgat
	<i>glk</i>	-133	4,65	GCTGAAaCGATaaAGt
	<i>fbp</i>	-89	4,14	GCTGgcgCGATTCacC
	<i>epd-pgk-fbaA</i>	-136	5,21	GCTGAAgCGtTTCAGt
	<i>edd-eda</i>	-112	4,64	aCTGAAaCGtTTtGC
	<i>tpiA</i>	-160	4,68	GtTGAAcCGATTaAGC
	<i>gapA</i>	-88	3,91	atTGAcacCGATTCcGC
	<i>gpmA</i>	54	3,90	agaaAAcCGtTTCacC
	<i>pckA</i>	-287	3,99	GgTGAATCGATaCttt
	<i>ppsA</i>	-101	4,57	GCTtgAaCGATTCacC
	<i>pfkA</i>	-122	4,86	cCTGAATCaATTCAGC
	<i>pykF</i>	-222	4,61	ctTGAATgGtTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-252	4,51	agTGAATCGgTTCAat
	<i>icdA</i>	-247	4,93	GCTGAATCGcTTaAcC
	<i>adhE</i>	-263	4,56	GCTGAAagGtgTCAGC
	<i>aceBAK</i>	-251	4,66	cgTtAAgCGATTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-282	3,91	aCgGAcacCGATTCaAc
	<i>nirBDC-cysG</i>	-48	3,92	GCTGAATCGtTaagGt
<i>copA</i>	-55	4,09	GCTGgAagGtTTaAcC	
<i>crp</i>	-221	4,14	GCTGAAgCGAgacAcC	
<i>S. typhi</i>	<i>fruBKA</i>	-224	3,81	GCTGAAgCctTTtAtg
		-102	4,95	ctTGAAaCGtTTCAGC
		-36	5,17	GCTGAATCGtTTCAat
	<i>mtlADR</i>	-123	4,33	agTtgAaCGATTCAGt
	<i>ptsHI-crr</i>	-244	3,93	aaTtAATCGtTaCAGg
		-218	4,50	GCTGAATCGATTTtat
	<i>manXYZ</i>	-82	3,94	tCTGAATCGATTCget
	<i>fbp</i>	-88	4,14	GCTGgcgCGATTCacC
	<i>epd-pgk-fbaA</i>	-136	5,21	GCTGAAgCGtTTCAGt
	<i>edd-eda</i>	-114	4,56	cCTGAAaCGATTTtGC
	<i>tpiA</i>	-164	4,68	GtTGAAcCGATTaAGC
<i>gapA</i>	-86	3,91	atTGAcacCGATTCcGC	
<i>gpmA</i>	52	4,29	agaGAAcCGtTTCacC	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>S. typhi</i>	<i>pckA</i>	-286	4,05	GgTGAATCGATattGt
	<i>ppsA</i>	-104	4,57	GCTtgAaCGATTCAcC
	<i>pfkA</i>	-121	4,86	cCTGAATCaATTCAGC
	<i>pykF</i>	-220	4,61	ctTGAATgGtTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-254	4,30	GtTGAATCGgTTCAGa
	<i>icdA</i>	-245	4,66	GCTGAATCGcTTaAcg
	<i>adhE</i>	-266	4,56	GCTGAAagGtgTCAGC
	<i>aceBAK</i>	-251	4,67	cCTaAAgCGtTTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-282	3,91	aCgGAcaCGATTCAaC
	<i>nirBDC-cysG</i>	-46	3,92	GCTGAATCGtTaagGt
	<i>copA</i>	-245	3,92	GtTtAAatGcTTCtGC
<i>crp</i>	-224	4,14	GCTGAAgCGAgaCAcC	
<i>S. typhimurium</i>	<i>fruBKA</i>	-221	3,81	GCTGAAgCctTTtAtg
		-100	4,95	ctTGAAaCGtTTCAGC
		-34	5,17	GCTGAATCGtTTCaAt
	<i>mtlADR</i>	-125	4,33	agTtgAaCGATTCAgT
	<i>ptsHI-crr</i>	-246	3,93	aaTtAATCGtTaCAGg
		-220	4,50	GCTGAATCGATTttat
	<i>manXYZ</i>	-84	3,94	tCTGAATCGATTcgt
	<i>fbp</i>	-88	4,14	GCTGgcgCGATTCAcC
	<i>epd-pgk-fbaA</i>	-136	5,21	GCTGAAgCGtTTCAGt
	<i>edd-eda</i>	-112	4,56	cCTGAAaCGATTtGC
	<i>tpiA</i>	-162	4,68	GtTGAAcCGATTaAGC
	<i>gapA</i>	-86	3,91	atTGAcacCGATTcGC
	<i>gpmA</i>	54	4,29	agaGAAcCGtTTCAcC
	<i>pckA</i>	-288	4,05	GgTGAATCGATattGt
	<i>ppsA</i>	-104	4,57	GCTtgAaCGATTCAcC
	<i>pfkA</i>	-123	4,86	cCTGAATCaATTCAGC
	<i>pykF</i>	-220	4,61	ctTGAATgGtTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-254	4,30	GtTGAATCGgTTCAGa
	<i>icdA</i>	-247	4,66	GCTGAATCGcTTaAcg
	<i>adhE</i>	-266	4,56	GCTGAAagGtgTCAGC
	<i>aceBAK</i>	-251	4,67	cCTaAAgCGtTTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-282	3,91	aCgGAcaCGATTCAaC
	<i>nirBDC-cysG</i>	-48	3,92	GCTGAATCGtTaagGt
	<i>copA</i>	-243	3,92	GtTtAAatGcTTCtGC
	<i>crp</i>	-85	4,09	GCTGgAagGtTTaAcC
		-226	4,14	GCTGAAgCGAgaCAcC
	<i>P. carotovorum</i>	<i>fruBKA</i>	-103	4,57
-36			5,29	GCTGAAaCGATTCAaC
<i>epd-pgk-fbaA</i>		-118	5,21	GCTGAAgCGtTTCAGt
<i>edd</i>		-389	4,29	GCTGAATCGAgTaAag
<i>gapA</i>		-94	3,91	atTGAcacCGATTcGC
<i>gpmA</i>		-149	4,36	GCTGAAaCGgTaaAGC
<i>ppsA</i>	-118	4,44	GgTtgATCGATTCAcC	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>P. carotovorum</i>	<i>pfkA</i>	-135	4,86	cCTGAATCaATTCAGC
	<i>pykF</i>	-240	4,76	atTGAATgGtTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-253	4,49	atTtAAaCGgTTCAGC
	<i>icdA</i>	-175	4,98	GCTGAATCGtTTaAct
	<i>adhE</i>	-339	4,63	GCTGAATgGtgTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-257	4,60	aCTGAcTCGATtTAGC
	<i>crp</i>	-297	4,06	GCTGAAgCGAgaCAaC
<i>P. chrysanthemum</i>	<i>fruBKA</i>	-105	4,57	ttTGAAaCGtTTCAGC
		-37	5,02	GCTGAAaCGATTCAag
	<i>manXYZ</i>	-195	4,27	cCTGAATCGATTcAta
		47	3,99	GCTGAAatGATTcTtg
	<i>glk</i>	-253	3,76	aCTGAAgCtgTTtAGC
	<i>epd-pgk-fbaA</i>	-154	5,21	GCTGAAgCGtTTCAGt
	<i>tpiA</i>	-178	4,68	GtTGAAcCGATTaAGC
	<i>gapA</i>	-88	3,91	atTGAcCaCGATTcGC
	<i>gpmA</i>	-146	4,77	GCTGAAaCGATaaAGC
	<i>ppsA</i>	-109	4,44	GgTtgATCGATTCAcC
	<i>pfkA</i>	-117	4,86	cCTGAATCaATTCAGC
	<i>pykA</i>	-144	4,10	GgaagAaCGATTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-254	4,16	GgTtgAaCGgTTCAGC
	<i>icdA</i>	-174	5,18	GCTGAATCGtTTaAGt
	<i>adhE</i>	-314	4,63	GCTGAATgGtgTCAGC
	<i>crp</i>	-307	4,06	GCTGAAgCGAgaCAaC
<i>Y. pestis</i>	<i>fruBKA</i>	-101	5,10	atTGAAaCGtTTCAGC
		-35	5,17	GCTGAAaCGATTCAat
	<i>ptsHI-crr</i>	-331	5,05	GCTGAATCGATTtAcC
	<i>manXYZ</i>	-158	4,80	cCTGAATCGATTcAtt
		59	3,77	GCTGAAatGATTttag
	<i>glk</i>	-126	4,36	GCTGAAaCGtTTttat
	<i>epd-pgk-fbaA</i>	-160	5,21	GCTGAAgCGtTTCAGt
	<i>tpiA</i>	-165	4,68	GtTGAAcCGATTaAGC
	<i>gapA</i>	-79	3,91	atTGAcCaCGATTcGC
	<i>gpmA</i>	-194	4,56	GCaGAATCGATTcGC
	<i>ppsA</i>	-194	4,44	GgTtgATCGATTCAcC
	<i>pfkA</i>	-143	4,58	ctTGAATCaATTCAGC
	<i>pykF</i>	-257	4,88	GtTGAATgGtTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-201	4,29	aCTatAaCGcTTCAGC
	<i>icdA</i>	-125	4,64	GCTGAATCGgTTaAct
	<i>adhE</i>	-308	4,63	GCTGAATgGtgTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-258	4,60	aCTGAcTCGATtTAGC
	<i>nirBDC-cysG</i>	-167	4,35	GCTGAtaCctTTCAGC
	<i>copA</i>	-182	4,09	GCTGgAagGtTTaAcC
<i>crp</i>	-254	4,59	GCTGAAgCGATaCAaC	
	-155	3,85	agTGAAggGtTTctat	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>Y. enterocolitica</i>	<i>fruBKA</i>	-103	4,95	ctTGAAaCGtTTCAGC
		-37	5,17	GCTGAAaCGATTCAat
	<i>ptsHI-crr</i>	-286	4,80	GCTGAATCGATTtAtC
	<i>manXYZ</i>	-221	4,80	cCTGAATCGATTCAat
	<i>glk</i>	-131	4,36	GCTGAAaCGtTTttat
	<i>epd-pgk-fbaA</i>	-159	5,21	GCTGAAgCGtTTCAGt
	<i>tpiA</i>	-162	4,68	GtTGAAcCGATTaAGC
	<i>gapA</i>	-88	3,91	atTGAcacCGATTcGC
	<i>gpmA</i>	-212	3,84	atTaAATCGATTcGaC
	<i>ppsA</i>	-163	4,44	GgTtgATCGATTcAcC
	<i>pfkA</i>	-145	4,58	ctTGAAATCaATTcAGC
	<i>pykF</i>	-258	4,88	GtTGAAATgGtTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-285	4,29	aCTatAaCGcTTCAGC
		-49	4,16	GCaGAgacCGATTcAct
	<i>icdA</i>	-176	4,64	GCTGAATCGgTTaAct
	<i>adhE</i>	-310	4,63	GCTGAATgGtgTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-257	4,60	aCTGAacTCGATTtAGC
	<i>nirBDC-cysG</i>	-168	5,04	aCTGAAaCcATTcAGC
	<i>copA</i>	-141	4,09	GCTGtAagGtTTaAcC
	<i>crp</i>	-255	4,59	GCTGAAgCGATaCAaC
-156		3,85	agTGAAggGtTTCtat	
<i>Y. pseudotuberculosis</i>	<i>fruBKA</i>	-103	5,10	atTGAAaCGtTTCAGC
		-37	5,17	GCTGAAaCGATTCAat
	<i>ptsHI-crr</i>	-333	5,05	GCTGAATCGATTtAcC
	<i>manXYZ</i>	-156	4,80	cCTGAATCGATTCAat
	<i>glk</i>	-127	4,36	GCTGAAaCGtTTttat
	<i>epd-pgk-fbaA</i>	-160	5,21	GCTGAAgCGtTTCAGt
	<i>tpiA</i>	-163	4,68	GtTGAAcCGATTaAGC
	<i>gapA</i>	-86	3,91	atTGAcacCGATTcGC
	<i>gpmA</i>	-276	4,56	GCaGAATCGATTcGcGC
	<i>ppsA</i>	-230	4,44	GgTtgATCGATTcAcC
	<i>pfkA</i>	-145	4,58	ctTGAAATCaATTcAGC
	<i>pykF</i>	-258	4,88	GtTGAAATgGtTTCAGC
	<i>pdhR-aceEF-lpdA</i>	-273	4,29	aCTatAaCGcTTCAGC
		-37	4,16	GCaGAgacCGATTcAct
	<i>icdA</i>	-174	4,64	GCTGAATCGgTTaAct
	<i>adhE</i>	-308	4,63	GCTGAATgGtgTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-256	4,60	aCTGAacTCGATTtAGC
	<i>nirBDC-cysG</i>	-169	4,35	GCTGAtaCctTTCAGC
	<i>copA</i>	-182	4,09	GCTGgAagGtTTaAcC
	<i>crp</i>	-256	4,59	GCTGAAgCGATaCAaC
-157		3,85	agTGAAggGtTTCtat	
<i>S. marcescens</i>	<i>fruBKA</i>	-99	4,95	ctTGAAaCGtTTCAGC
		-33	5,17	GCTGAAaCGATTCAat
	<i>ptsHI-crr</i>	-254	4,93	GCTGAATCGATTtAct
	<i>glk</i>	-125	4,04	GCTGAAaCGtTTtttg

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>S. marcescens</i>	<i>epd-pgk-fbaA</i>	-208	5,21	GCTGAAgCGtTTCAGt
	<i>tpiA</i>	-164	4,68	GtTGAAcCGATTaAGC
	<i>gapA</i>	-88	3,91	atTGAcCaCGATTCcGC
	<i>gpmA</i>	-150	4,04	GCctgAaCGATTCAGC
	<i>ppsA</i>	-132	4,44	GgTtgATCGATTCaC
	<i>pfkA</i>	-141	4,58	ctTGAATCaATTCAGC
	<i>pykF</i>	-237	4,88	GtTGAATgGtTTCAGC
	<i>adhE</i>	-306	4,56	GCTGAAagGtgTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-270	4,53	aCTGAcCaCGATTtAGC
	<i>nirBDC-cysG</i>	-167	4,75	aCTGAcTCGcTTCAGC
<i>crp</i>	-301	4,06	GCTGAAgCGAgCaAaC	
<i>P. luminescens</i>	<i>fruBA</i>	-105	5,10	GtTGAAaCGtTTCAGt
		-39	4,70	aCTGAAaCGATTctat
	<i>fruK</i>	-87	5,05	aCTGAAaCGATTCAat
	<i>epd-pgk-fbaA</i>	-140	5,00	GCTGAATCGtTTCatt
	<i>tpiA</i>	-177	4,56	atTGAAcCGATTaAGC
	<i>gapA</i>	-89	3,91	atTGAcCaCGATTCcGC
	<i>gpmA</i>	54	4,17	agaGAAcCGtTTCAct
	<i>ppsA</i>	-113	4,44	GgTtgATCGATTCaC
	<i>pfkA</i>	-140	4,85	GtTGAATCaATTCAGC
	<i>pykF</i>	-344	4,88	GtTGAATgGtTTCAGC
	<i>adhE</i>	-163	4,42	GCTGAAagGAaTCAGC
	<i>nuoABCDEFGHIJKLMN</i>	-253	4,46	aCTGAcCaCGtTtAGC
<i>V. vulnificus</i>	<i>fruBKA</i>	-386	5,52	GCTGAATCGATTCAGt
		-364	5,16	GCTGAAagGATTCAGC
		-214	5,23	GCTGAATCGgTTCAGC
	<i>fruR</i>	-259	5,23	GCTGAAcCGATTCAGC
		-109	5,16	GCTGAATCctTTCAGC
	-87	5,52	aCTGAATCGATTCAGC	
<i>V. parahaemolyticus</i>	<i>fruBKA</i>	-293	5,11	GCTGAAcCGATTCAGt
		-271	5,16	GCTGAAagGATTCAGC
		-113	5,52	GCTGAATCGATTCAGt
	<i>fruR</i>	-267	5,52	aCTGAATCGATTCAGC
		-109	5,16	GCTGAATCctTTCAGC
-87		5,11	aCTGAATCGgTTCAGC	
<i>V. cholerae</i>	<i>fruBKA</i>	-265	5,52	GCTGAATCGATTCAGt
		-243	5,16	GCTGAAagGATTCAGC
		-116	5,23	GCTGAAcCGATTCAGC
	<i>fruR</i>	-238	5,23	GCTGAATCGgTTCAGC
		-111	5,16	GCTGAATCctTTCAGC
-89		5,52	aCTGAATCGATTCAGC	
<i>V. fischery</i>	<i>fruBKA</i>	-211	5,52	GCTGAATCGATTCAGt
		-189	5,16	GCTGAAagGATTCAGC
		-80	5,52	GCTGAATCGATTCAGt

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. fischery</i>	<i>fruR</i>	-221	5,52	aCTGAATCGATTCAGC
		-112	5,16	GCTGAATCctTTCAGC
		-90	5,52	aCTGAATCGATTCAGC
<i>P. profundum</i>	<i>fruBKA</i>	-241	5,45	GCTGAAaCGATTCAGt
		-219	5,09	GCTGAAagGtTTCAGC
		-152	5,57	GCTGAAaCGATTCAGC
	<i>fruR</i>	-204	5,57	GCTGAATCGtTTCAGC
		-137	5,09	GCTGAAaCctTTCAGC
		-115	5,45	aCTGAATCGtTTCAGC
<i>P. aeruginosa</i>	<i>fruB/ptsI-fruKA</i>	-250	5,32	agTGAATCGATTCAGC
	<i>fruR</i>	-93	5,32	GCTGAATCGATTCact
<i>P. putida</i>	<i>fruB/ptsI-fruKA</i>	-130	4,78	GaTtAAaCGtTTCAGC
	<i>fruR</i>	-170	4,78	GCTGAAaCGtTTaAtC
<i>P. fluorescens</i>	<i>fruB/ptsI-fruKA</i>	-163	4,78	GaTtAAaCGtTTCAGC
	<i>fruR</i>	-158	4,78	GCTGAAaCGtTTaAtC
<i>P. syringae</i>	<i>fruB/ptsI-fruKA</i>	-198	3,90	aCaGAATCGAgTaAcg
		-167	4,85	GaTtAAaCGATTCAGC
	<i>fruR</i>	-203	4,85	GCTGAATCGtTTaAtC
		-172	3,90	cgTtAcTCGATTCtGt

Приложение 3.

Множественное выравнивание для белков FruB и PtsI из *E. coli* и химерного белка из Pasteurellales. Выделены последовательности, соответствующие белкам FruB и PtsI *E. coli*. Обозначения геномов: см. Табл. 2.

FruB

ECO MFQLSVQDIHPGKAGDKKEEAIRQVAAALVQAGNVAEGYVNGMLAREQQSTFLGNGIAIPHGTTDTRDQVLKTVQVQVQFP
PFO MLELTVQISMGSQSAVDKPAALQLLASHLVADGLVADGYLSGLQAREAQQSTFLGQGIAPHGTPETRDQVFTTGVRMLQFP
PST MLELTLGQISMAQSAVDKTAALKLLADHLVADGLVAEGYLTGLMNRQQGSTFLGQGIAPHGTPETRDVFTTGVRMLQFP
PPU MLELANEQIAMGQKAADKAEALRLLADRLVADGLVAEGYLQGLQAREAQQSTFLGQGIAPHGTPQTRDLVYATGVRLLQFP
PAE MLELDTRQIRMQRADKAEALRLLGAALVADGLAAPGYAELKAREAQGSTYLGQGIAPHGTPDRELVFSTGVRLLQFP
: * : * . : * ** * : : . * * * . * ** * : * ** * ** : ** * ** : ** : ** : ** : ** : ** : **

FruB

ECO EGVTVGQVAYVAIGIAASSDEHLGRLQLTHVLSDSVAEQKLSATTAEELRALLMGEKQSEQLKLDNEMLTLDIVASDI
PFO DGVDWGDGQIVYLAIGIAAKSDEHLRLLQLLTRALGETDLGQALRRASSPEALLKLLQG--APQELALDAQMI GLGVSADDF
PST EGVTVGQVAYVAIGIAAKSDEHLRLLQLLTRALGETDLGQALRRASSPEALLKLLQG--APQELALDAQMI SLGVSADDF
PPU EGVTVGQVAYVAIGIAAKSDEHLRLLQLLTRALGETDLAEALRRASSAEALLKLLQG--APQALALDAQLVGLNVPADF
PAE EGVTVGQVAYVAIGIAAKSDEHLRLLQLLTRALGETDLGQALRRASSAEALLKLLQG--APQELALDAQLVGLGQNAEDL
: ** * ** * . * : ** : ** * ** * ** * ** * : ** : * * : * * * . : * * * : : * . * * :

FruB

ECO LTLQALNAARLKEAGAVDATFVTKAINEQPLNLGQGIWLSDSAEGNLRSAIAVSRAANAFDVGGETAAMLVSVAMNDDQPIA
PFO EELVWRGARLLRQADCVSNGFAGVLQQVEALPLGDGLWHLHSEQTVKRPGLAFVTPDKPMRYLGQPLSGLFCLASLGEAHQA
PST EELVWRGARLLRQADCVSNGFAAVLQQVEALSGLDGLWHLHSEQTVKRPGLAFVTPDKPIRYLGQPLTGLFCLASLGEAHQA
PPU DELAWRGARLLQRADCVDSGFAAVLQQAEPLPLGEGWLHSEQRVVRPGLAFITPQQPLRYRQPLNGLFCLASLGEAHQA
PAE DELAWLGARLLKAGCVENGFAAVLQQTEPLPLGDGLWHLHSEQLVLRPGLAFVTPAQPLQHQGLVLTGLFCLASLGEAHQA
* . * * : * . * . * . : : * ** : * . * . : : * . : : : * . : : * . * . *

FruB

ECO VLKRLADLLDNKADRLKADAATLLALLTSDDAPTDDVLSAEFVVRNEHGLHARPGTMLVNTIKQFNSDITVTLNLDGTGKPK
PFO LLERLCALLIEGRGHGELGRATSSRKVLEVLGGELP-ADWPSARIALANAHGLHARPAKILAQLAKSFDGEIRVRIVDGDQSA
PST LLERLCALLIEGRGHGELGHATNSRVVLEALGGEVPAEWPTQQIQLANAHGLHARPAKILAQLAKEFEGEVRVRLVETGESA
PPU LLERLCALLIEGRGHGELGHATNSRVVLEALGGEVPAEWPTQQIQLANAHGLHARPAKILAQLAKEFEGEVRVRLVETGESA
PAE LLERLCALLIEGRGHGELGHATNSRVVLEALGGEVPAEWPTQQIQLANAHGLHARPAKILAQLAKEFEGEVRVRLVETGESA
: * : * . : * : * . : : * : : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

FruB

ECO ANGRSLMKVVALGVKKGHRLRFTAQGDAAEQALKAIGDAIAAGLGEGA
PFO VSVKLSKLLSLGARRGQVLEIIAEPSSIAADALPALLAAIEEGLGEEVEPLPAVSQ--QIADIAEVLASAPASGSLQA
PST VSAKLSKLLSLGARRGQVLEIIAEPSSIAADALPALLAAIEEGLGEEVEPLPAASEPEAAPTVAVAKPINAPA-AGSVLSA
PPU VSVKLSKLLSLGARRGQVLEIIAEPSSIAADALPALLAAIEEGLGEEVEPLPAVSQ-PIADDVPEVLQAPA-AGSRIQG
PAE VSAKLSKLLALGARRGQVLEIIAEPSSIAADALPALLAAIEEGLGEEVEALAEALP--DAVGEAEDARPAPLRAGERLQA
.. : * * : : * : : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : * : *

PtsI

MISC

PtsI

ECO TLASPGIAFGKALLLKEDEIVDRKKISADQVDQEVERFLSRAKASAQLETIKTKAGETFGEKEAIFEGHIMLLEDEELE
PFO IPAAPGIAIGPAHIQVQQAIDYPLR---GESAAIERERLQALADVRQDIQGLIERSKA---KAIREIFITHQEMLDDPELT
PST VPASPGIAIGPAHVQVLQTFDYPQK---GESVAAERERLQTAIGEVRRDIENLIQRSKS---KAIREIFITHQEMLDDPELI
PPU VQAAPGIAIGPAHVQVQVREFDYPLR---GESCALERQKREALATVNGELQALVLRSD---KAIGEIVFVTHQEMLADPALT
PAE IAASPGIASGPAHVQVQVREFDYPLR---GESPAHERERLLRAKRAVDEEIVGLVERSTV---KAIREIFVTHREMLDDPELA
: * : * * * * * : : : : * : : : : : : : : : : * * * : * * *

PtsI

ECO QEIIALIKDKHMTADAAAHEVIEGQASALEELDDEYLKERAADVRDIGRLLRNILGLKIIDLSAIQDEVILVAADLTPSET
PFO DEVDTRLK-QGESAAAWMAVIEAAAKQQESLHDALLAERAADLRDIGRRLVLAQLCGVQTP--AEPEQPYILVMDEVGPSDV
PST EQRLN-DNESAAAATVIEAAAVQEQKLDALLAERAADLRDIGRRLVLAQLCGVETV--AAPDEPYILVMDEVGPSDV
PPU DDVEQRLA-QGESAAAAMAVIEAAARQQEALHDALLAERAADLRDIGRRLVLAQLCGVQAQ--VEPEQPYILVMDEVGPSDV
PAE EQVQLRLN-RGESAAAWSRVVEDSAAQQEALHDALLAERAADLRDIGRRLVLAQLCGVEAP--REPEQPYILVMDEVGPSDV
: : : : * * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * * : * * * * *

PurR (STY) -FTD TVIDNAFAGGYMAGRYLVERGHRDIGVIPGPLERNTGAGRLAGFMKAMEEALINVPDNWIVQGDFEPESGYHA
PurR (STM) -FTD TVIDNAFAGGYMAGRYLVERGHRDIGVIPGPLERNTGAGRLAGFMKAMEEALINVPDNWIVQGDFEPESGYHA
PurR (ECO) -FTDAVIDNAFAGGYMAGRYLIERGHRDIGVIPGPLERNTGAGRLAGFMKAMEEAMIKVPESWIVQGDFEPESGYRA
PurR (YPS) -FTDSIIDNAFAGGYLAGRYLIERGHRDIGAIPGQLARNTGGGRHQGFLKALEEANIPVREEWIVQGDFEPESGYKA
PurR (YPK) -FTDSIIDNAFAGGYLAGRYLIERGHRDIGAIPGQLARNTGGGRHQGFLKALEEANIPVREEWIVQGDFEPESGYKA
PurR (PLU) -FTDTIIDNAFHGGYLAGRYLIERGHRDIGAIPGQLARNTGGGRHQGFLKALKEANIPKKEEWIVQGDFEPESGYKA
PurR (PCA) -FTDTIIDNAFAGGYMAGRYLIERGHRDIGAIPGQIQRNTGSGRYLGF LKALKEADITVRNEWVVQDFEPESGYKA
PurR (PMU) -HTDIIQDHSFEGGYLATKYLIDHGKDIGIAGELTKTTAKTRYEGFIKAMNEAGLKINPDWVMEGFFEPEDGYEC
PurR (HIN) -NTDVIDDHSFDGGYLATKHLIECGHKKIGICGELNKTARTRYEGFEKAMEEAKLTINPSWVLEGADEFEPEDGYEC
PurR (HDU) -KNDRIIDNSFNAGYLA TRHLIDNGHKDIGVISGYLTKTICSERLEGF IKAMKEANLP IREEWILEGDFEPESGFEQ
PurR (VPA) -RADKIIDNSEEGGYLATKYLIDNGHTDIACLSGHFVKACQERIQGFRAMAEAKLPINEDWILEGNFECDTAVLV
PurR (VCH) -HADKIIDNSEEGGYLATKYLTDRGHTIACLSGHFVKAACQERIQGFRAMAEAKLTVNEDWILEGNFECDTAVLA
PurR (VVU) -QADKIIDNSEEGGYLATKYLIDNGHTKIACLSGHFEKAACQERIQGFRAMKEANITLNNEWILEGNFECDTAVMA
PurR (VFI) -RTDKIIDNSEEGGYLATKHLIENGHTHIACITGQADKVTCKERVGRFERAHIDANLTFNPEWILEGDFECASASRA
PurR (PPR) -LTDKIQDNAEQGGYVATKHFIDNGHEKICLSGHSEKSTCRERLKGFNKAMAEAGITVNNWIIDGDFECESAVEA
RbsR (PPR) -PSDKIQDNSHHGGYLA TRHLIAQGHQTQIGCLTGPLDKLPAQQRLSGFVQAMEEAGLTINKSWIASGNFECGEGKA
RbsR (VFI) -PSDKIQDNSHHGGYLATKHLIEQGHTIIGCLTGPLNKLQAQQRLSGFVQAMEESGLEINKNWIASGNFECGEGEAA
RbsR (VVU) -SSDKIQDNSLRGGYLAANHLIQMGHKEIGCITGPLIKHQAMRYEGYKRALNEHGLFPNPAWIVEADFECEGGYQA
RbsR (VPA) -TSDKIQDNSLRGGYLAAYKYLIDCGHTIIGCITGPLIKHQAMRYEGYKRAMNEAGLEFNANWIIESDFECGGYQA
RbsR (VCH) -ASDKIQDNSLQGGYMAAKHLIECGHKEIGCITGPLIRHQAMRYEGYKRALAEAGIAPNDWIVESDFECGGYQA
RbsR (STY) GSDSLIQDNSLLGGDLATQHLIDKGFTRIACTIPLDKTPARLRLEGYRAAMKRAGLAIPQGYEITGDFEFGGGDA
RbsR (STM) GSDSLIQDNSLLGGDLATQYLLIDKGFTRIACTIPLDKTPARLRLEGYRAAMKRAGLAIPQGYEITGDFEFGGGDA
RbsR (ECO) -DSDLIQDNSLLGGDLATQYLLIDKGFTRIACTIPLDKTPARLRLEGYRAAMKRAGLNIPDGYEVTGDFEFNGGDA
RbsR (PCA) -AIDVIKDNSLLGGEIATNYLISRGYKKIACIAGPKDKTTAYNRLEGYRQAMQLAGLSVPTDYEIFGDFEFETGYRA
RbsR (PLU) -VGDIVIQDNSLLGEMATNHLISCGFKKIACIAGPDKTPANHRLKGYRKAMANAGLTPDGYEIIYSDFEFAGGFKS
RbsR (PMU) -KADKIYENSEYGGYLATKQLIEKGHQAIIITGNLKKSLAQNRDLGYYKALKEANIPREDCIIESHFDFEGGKRG
RbsR (HIN) -NADKILENSALGGYLATKALIDAGHRKIGIITGNLKSVAQNRLQYKALSEAKIALNPHWIVESHDFEGGVLG
RbsR (PFO) --ADLVRIDHEQGAYLATRHLLELGHRIATIGGPASTSVAQMRQAGFARALKEAGVEVAPSRMLESDFSTGGYNA
RbsR (PPU) --ADLVRIDHEQGAYLATRHLLELGHRIATIGGPASTSVAQMRQAGFARALKEAGVEVAPSRMLESDFSTGGYNA
RbsR (PST) --ADMVRIDHEQGAYLATRHLLELGHRIATIGGPASTSVAQMRQAGFARALKEAGVEVAPSRMLESDFSTGGYNA
RbsR (PAE) --ADRVQIDHEMGAYLATRHLLELGHRIATIGGPASTSVAQMRQAGFARALKEAGVEVAPSRMLESDFSTGGYNA
* : . * . : * . : * * : *

PurR (STY) MQQILSQS--HRPTAVFCGGDIMAMGALCAADEMGLRVPQDVSVIGYDNVRNARFFTPALTTIHQPKDSLGETAFNM
PurR (STM) MQQILSQS--HRPTAVFCGGDIMAMGALCAADEMGLRVPQDVSVIGYDNVRNARYFTPALTTIHQPKDSLGETAFNM
PurR (ECO) MQQILSQP--HRPTAVFCGGDIMAMGALCAADEMGLRVPQDVSLIGYDNVRNARYFTPALTTIHQPKDSLGETAFNM
PurR (YPS) MHQILTQK--HRPTAVFCGGDIMAMGAI CAADELGLRVPQDISVIGYDNVRNARYFSPALTTIHQPKERLGETAFAM
PurR (YPK) MHQILTQK--HRPTAVFCGGDIMAMGAI CAADELGLRVPQDISVIGYDNVRNARYFSPALTTIHQPKERLGETAFAM
PurR (PLU) MHQILNQK--QRPTAVFCGGDVMAMGAI CAADELGLRVPQDISVIGYDNIRNARYFSPALTTIHQPKERLQGMAFSM
PurR (PCA) MHQILAQK--QRPTAVFCGGDIMAMGAI CAADELGLRVPQDISVIGYDNVRHARFFTPALTTIHQPKERLQGSASFMS
PurR (PMU) MNKILAQD--TLPTAVFCNDVMALGAI SAIGEGLRVPEDISVIGYDNIHARSFYSPPLTTIHQSKSRLGARAVTL
PurR (HIN) MNRLLTQE--KLPTALFCCNDVMALGAI SALTEKGLRVPEDMSI IGYDDIHASRFYAPPLTTIHQSKLRLGRQAINI
PurR (HDU) MNNLLRLD--KLPTALFCSDTIALGVLSALSEKGIAPVQDISI IGYDNTHSSRFFSPPLTTIHQSKSRLGVKALNL
PurR (VPA) ADKIIAME--KRPTAVFCFNDTALGLMSRLQQNGIKVPDDVSVIGYDNIELAEYFSPPLTTIHQPKRRVGKNAFEI
PurR (VCH) ADKIIAMD--KRPTAVFCFNDTALGLMSRLQQKGRIPEDMSVIGYDNIELAEYFSPPLTTIVHQPKRRVGKNAFEI
PurR (VVU) ADKIASWQ--DRPTAVFCFNDTALGLMSRLQQGLRIPEDISVIGYDNIELAEYFSPPLTTIHQPKRRVGKNAFEI
PurR (VFI) VDKILAI EENKRPTALFCFNDIMALAAI SKIQSGLRVPEDISVIGYDNIELSAFYSPPLTTIHQPKRRVGKTAVEI
PurR (PPR) ANQYIAMK--DRPTALFCFNDIMAMALISTFEQAGVRVPDDISVIGYDNI DLAPYFSPPLTTIHQPKRRLGKTAIEI
RbsR (PPR) FEELYARG--SLPTALFVCNDMMAMGVINTANKKGISVNPDISVIGYDDIKLAKYITPSLTTIHQPKHRLGQQAVDT
RbsR (VFI) FNEIHAKG--KLPTALFVCNDMMAMGMINCASKKGISVNPDISVIGYDDIKMAKYITPSLTTIHQPKHRLGQQAVDS
RbsR (VVU) FKKIVAKG--PLPSALFVCNDMMAMGVINAANEVGV EIPHDLIS IGYDDIHIKFMTPSLTTIHQPKYRLGQA AVET
RbsR (VPA) FKKMAERG--TLPSIFVSNDDMMAMGVINAANELGIKVPDDLSI IGYDDIHIKFMSPSLTTIHQPKYRLGQA AVET
RbsR (VCH) FEKLYQRG--KLPSALFVSNDDMMAMGV IQAASQRGLRVPDDLSLIGYDDVHIKFMTPALTTIHQPKYRLGKA AVDT
RbsR (STY) MQALLAHQ--QRPQAVFTGNDAMAVGAYQALYQAGLRIPQDMAVIGYDDIELARYMTPPLTTIHQPKDELGEL AIDV
RbsR (STM) MQALLAHQ--QRPQAVFTGNDAMAVGAYQALYQAGLRIPQDMAVIGYDDIELARYMTPPLTTIHQPKDELGEL AIDV
RbsR (ECO) MRQLLSHP--LRPQAVFTGNDAMAVGVYQALYQAE LQVPQDI AVIGYDDIELASFMTPLTTIHQPKDELGEL AIDV
RbsR (PCA) MQQLLALQ--DKPEAVFTSNDAMAVGVYHALYQAGLSIPQDMAVIGYDDIELARYMSPPLTTIVHQPDELGEL AIDV
RbsR (PLU) MQELLQHP--EPPEAVFAGNDAMAVGAYQALYQAGLSV PDDISI IGYDDIDLAPYMIPLTTIHQPKDELGKLA VDT
RbsR (PMU) MEKVLALQ--PRPTAVFACSDSIAFGAYQTVWHQGLH IHPQDISI IGYDNI TLAEYMSPLTTINQPKAELGKLA VET
RbsR (HIN) IQSLLTQS--SRPTAVFCCSDTIAVGAYQAIQQQGLR IHPQDISI IGYDNI TLAEYMSPLTTINQPKAELGKLA VET
RbsR (PFO) ASILL--ER-NPPSAIFAGNDMIGIVLRAAAERNR VPDEL SVIGYDDIELARYMTPPLTTIVGQSL LGEMAAEV
RbsR (PPU) AAQVL--EG-KRPSAIFAGNDMIGIVLRAAAERNR VPDEL SVIGYDDIELARYMTPPLTTIVGQSL LGEMAAEV
RbsR (PST) AGRLL--AE-HAPTALFAGNDVIAIGVLRAAAERS IRV PDL SVIGYDDIQISRYVYPALTTIVGQSL IRLGEMAAEV
RbsR (PAE) ALGLLGEG--ERPTAIVAGNDMIGIVLRAAAELNIR VPDEL SVIGYDDIELSRVFPALTTIVGQSL IRLGEMAAEV
* : . * . : * . : * * : *

```

PurR (STY)  LLDRIV--NKRE---ESQSIEVHPRLVERRSVADGPFDRDYRR-
PurR (STM)  LLDRIV--NKRE---ESQSIEVHPRLVERRSVADGPFDRDYRR-
PurR (ECO)  LLDRIV--NKRE---EPQSIEVHPRLIERRSVADGPFDRDYRR-
PurR (YPS)  LLDRIV--SKRE---DPQTIEVHPKLVERRSVADGPFDRDYRR-
PurR (YPK)  LLDRIV--SKRE---DPQTIEVHPKLVERRSVADGPFDRDYRR-
PurR (PLU)  LLDRII--NKRE---DAQTIEVHPRLVERRSVVDGPFVDYRR-
PurR (PCA)  LLDRIT--SKRE---DAHVIEVHPTLIERRSVADGPFDRDYRR-
PurR (PMU)  LFERIN--EKSE---ERAVIEMHPELVIRKSVKSRL-----
PurR (HIN)  LLERIT--HKDEGVQQYSRIDITPELIIKSVKSIL-----
PurR (HDU)  LLERIKLGHKEY---QPHTIEFHPELVLRHSVRNLNKSEE---
PurR (VPA)  LLERIK--DKEH---EKRVFEMQPEIVIRNTVK-KLN-----
PurR (VCH)  LLERIK--DKEH---ERRIFEMHPELIVERDVK-DLTKS----
PurR (VVU)  LLERIK--DKEH---DKRVFEMHPELVVRDVK-KIN-----
PurR (VFI)  LLERIK--DKDH---ERRVFEMQPEVVTSSVSNRLK-----
PurR (PPR)  LMERVK--DKNH---ERRVFEMNPELVIRKSVK-DLN-----
RbsR (PPR)  LLDKIQ--SQRE---TNLI IQLEPTLVERDSVKKM-----
RbsR (VFI)  LLEKIA--TKSE---TNRVFQLEPTLMVRDSVKTLSN-----
RbsR (VVU)  LLKKIN--KETA---EVQVVQLEPTLIERN SVVEWHDNEI---
RbsR (VPA)  LVRRLD--DKSN---EAQVVQLEPTLVVRNSVTNFS-----
RbsR (VCH)  LLYRLE--NPDT---TAQVVQLEPTLVVRNSVRKLG-----
RbsR (STY)  LIHRMA--QPAL---QQQRLQLTPVLTVRGSV-----
RbsR (STM)  LIHRMA--QPAL---QQQRLQLTPVLTVRGSV-----
RbsR (ECO)  LIHRIT--QPTL---QQQRLQLTPILMERGSA-----
RbsR (PCA)  LLYRLE--HPNT---EPNVLVLTPELMVRQSVQ-----
RbsR (PLU)  LLYRMD--NPES---DPKLLVLTPTLIERSVSRSS-----
RbsR (PMU)  LLQRIK--NPTA---KHRTLLLEPQLVWRESVRQLYQEKRG--
RbsR (HIN)  LLQRIK--NPNE---NYRTLVLLEPTCVLRESI---YSLK----
RbsR (PFO)  LLRRIATP-DLA----TDQRIVTPSIVLRESTAPLSGVFTEYR
RbsR (PPU)  LLTRIATPRQGA----AEQRIVAPRIVLRESTGPRPDLFNDYR
RbsR (PST)  LLSRIASPHATP----VEKRLVKPCVVVRESTAAPNTSSNE--
RbsR (PAE)  LLQRIAER-EEP----MQRRVVAPALVLRRESTAPPLVRV----
* . : : . * * :

```


Приложение 7.

Последовательности потенциальных сайтов связывания белка RbsR. Указано положение сайта относительно старта транскрипции гена. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными. ¹⁾ Сайты, обнаруженные в геномах, где отсутствует ген регулятора.

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>E. coli</i>	<i>rbsDACBKR</i>	-32	6,91	TCAGCGAAACGTTTCGcTGA
<i>S. typhi</i>	<i>rbsDACBKR</i>	-110	6,39	ctAGCGAAACGTTTCGAcGg
		-31	6,56	cCAGCGAAACGTTTCGcTag
<i>S. typhimurium</i>	<i>rbsDACBKR</i>	-112	5,53	ctAGCGAAACGTTTCGAcGg
		-33	6,30	cCAGCGAAACGTTTCGCTAg
<i>P. carotovorum</i>	<i>rbsDACBKR</i>	-110	6,36	TtAGCGAAACGTTTCGATGg
		-33	6,36	TtAGCGAAACGTTTCGCTGg
<i>P. luminescens</i>	<i>rbsDACBKR</i>	-111	6,36	TtATCGAAACGTTTCGATAg
		-33	5,96	atAGCGAAACGTTTCGCTAA
<i>P. multocida</i>	<i>rbsDACBKR</i>	-45	6,91	TCAtCGAAACGTTTCGATGA
<i>H. influenzae</i>	<i>rbsDACBKR</i>	-38	6,74	TtAtCGAAACGTTTCGATaA
<i>V. cholerae</i>	<i>rbsDACBKR</i>	-36	6,84	TCAtCGAAACGTTTCGATGt
<i>V. fischeri</i>	<i>rbsDACBKR</i>	-72	5,84	TtAcgcAAACGTTTCGATGA
<i>V. parahaemolyticus</i>	<i>rbsDACBKR</i>	-225	6,43	cCATCGAAACGTTTCGATGA
		-95	6,16	gCATCGAAACGTTTCGATGA
<i>V. vulnificus</i>	<i>rbsDACBKR</i>	-73	6,69	TCATCGAAACGTTTCGATGA
<i>P. profundum</i>	<i>rbsDACBKR</i>	-119	5,78	cCATCGAAACGTTTgcgTAA
<i>Y. pestis</i> ¹⁾	<i>rbsDK</i>	-32	6,36	TtAGCGAAACGTTTCGCTcT
<i>Y. pseudotuberculosis</i> ¹⁾	<i>rbsDK</i>	-32	6,36	TtAGCGAAACGTTTCGCTcT

Приложение 8.

Последовательности потенциальных сайтов связывания белка PurR. Указано положение сайта относительно старта трансляции гена. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>E. coli</i>	<i>purR</i>	-61	4,76	AgGCAAACGTTTaCcT
		27	4,07	gaGCAAACGTTTcCac
	<i>prsA</i>	-357	4,72	AaGaAAACGTTTtCGc
	<i>purC</i>	-168	4,67	ACGCAAACGTgTGCGT
	<i>purA</i>	-122	4,17	AgGaAAACGaTTGgcT
		-23	3,77	AaGCAAACGgTgattT
	<i>purT</i>	-54	5,08	ACGCAAACGTTTtCGT
	<i>purL</i>	-91	4,94	ACGCAAACGgTTtCGT
	<i>purB</i>	185	4,45	ACGCAAtCGgTTaCcT
	<i>purEK</i>	-86	4,61	ACGCAAcCGTTTtCcT
	<i>cvpA-purF</i>	-71	4,93	ACGCAAACGTTTtCtT
	<i>purMN</i>	-80	4,55	tCGCAAACGTTTGctT
	<i>purHD</i>	-122	4,76	gCGCAAACGTTTtCGT
	<i>guaBA</i>	-68	4,21	AtGCAAtCGgTTaCGc
	<i>yhhP</i>	-76	3,98	ACGCAAtCGTTaaCcT
	<i>codBA</i>	-83	4,68	ACGaAAACGaTTGctT
	<i>upp-uraA</i>	-233	4,63	AaGCAAACGTTTGCGa
	<i>pyrC</i>	-66	4,54	AgGaAAACGTTTcCGc
	<i>pyrD</i>	-103	4,51	cgGaAAACGTTTGCGT
	<i>carAB</i>	-237	3,65	gCGCAAgCGTTTtCca
	<i>pyrLBI</i>	-91	3,79	AataAAAcCGTTTGCGc
	<i>speAB</i>	-133	4,28	AaGaAAAcCGgTTGCGc
	<i>glnB</i>	-82	4,06	AtGCAAACGaTTtCaa
	<i>gcvTHP</i>	-105	4,31	AaGagAACGaTTGCGT
	<i>folD</i>	22	3,62	ggt aAAACGaTTGCGc
	<i>glyA</i>	-133	4,39	AgGtAAAtCGTTTGCGT
	<i>serA</i>	-94	3,72	ACGCAAACGTTcataT
	<i>tsx</i>	-153	4,57	ACGCAAtCGaTTaCGT
	<i>gltS</i>	-226	4,48	AaGCAAACGTTTGcCa
	<i>yieG</i>	-114	4,21	ACGCAAtCGTTgcCGT
	<i>yjcD</i>	-273	4,22	ACGatAACGTTTGCGc
		-131	3,87	AgctAAACGTTTGctT
<i>rnt</i>	28	3,79	gtGCgAcCGTTTtCGT	
<i>xseA</i>	-109	3,68	gCGtAAcCGaTTGCaT	
<i>yicE</i>	-69	4,46	tgGCAAACGTTTGctT	
<i>yhhQ</i>	-160	3,87	AgGttAACGaTTGCGT	
<i>ydiJ</i>	-223	4,16	ACGCAAACGaTTtacT	
<i>S. typhi</i>	<i>purR</i>	-57	4,76	AgGCAAACGTTTaCcT
		29	4,07	ggGCAAACGTTTcCac
	<i>prsA</i>	-472	4,72	AaGaAAACGTTTtCGc
	<i>purC</i>	-132	4,67	ACGCAAACGTgTGCGT
	<i>purA</i>	-121	4,17	AgGaAAACGaTTGgcT
-23		3,77	AaGCAAACGgTgattT	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>S. typhi</i>	<i>purT</i>	-50	5,08	ACGCAAACGTTTtCGT
	<i>purL</i>	-91	4,94	ACGCAAACGgTTtCGT
	<i>purB</i>	183	4,71	ACGCAAACGgTTaCcT
	<i>purEK</i>	-88	4,61	ACGCAAcCGTTTtCcT
	<i>cvpA-purF</i>	-73	4,93	ACGCAAACGTTTtCtT
	<i>purMN</i>	-77	4,55	tCGCAAACGTTTGcT
	<i>purHD</i>	-124	4,76	gCGCAAACGTTTtCGT
	<i>guaBA</i>	-63	4,29	AtGCAAcCGaTTaCGT
	<i>yhhP</i>	-78	3,98	ACGCAAtCGTTaaCcT
	<i>codBA</i>	-75	4,85	ACGaAAACGTTTGcT
	<i>upp-uraA</i>	-351	4,63	AaGCAAACGTTTGCGa
	<i>pyrC</i>	-66	4,63	ACGaAAACGTTTcCGc
	<i>pyrD</i>	-100	4,21	AgGaAAACGaTTGcTa
	<i>carAB</i>	-237	3,65	gCGCAAgCGTTTtCta
	<i>speAB</i>	-133	3,67	AaGaAAcCGgTcGCGc
	<i>gcvTHP</i>	-107	4,48	AaGagAACGTTTGCGT
	<i>folD</i>	20	3,62	ggt aAAACGaTTGCGc
	<i>glyA</i>	-133	4,39	AgGtAAtCGTTTGCGT
	<i>tsx</i>	-155	4,57	ACGCAAtCGaTTaCGT
	<i>gltS</i>	-273	4,15	gCGCAAcCGgTTGCGc
		-175	4,46	AaGCAAACGaTTGCGa
	<i>yicE</i>	-186	4,48	AaGCAAACGTTTGcTa
		-67	4,29	tCGCAAtCGTTTGcT
	<i>yieG</i>	81	4,37	ACGCAAACGTTTcCta
	<i>yjcD</i>	-272	4,22	ACGatAACGTTTGCGc
		-129	3,87	AgctAAACGTTTGcT
<i>rnt</i>	30	3,79	gtGCgAcCGTTTtCGT	
<i>xseA</i>	-107	4,09	ACGtAAtCGgTTGCaT	
<i>yhhQ</i>	-107	3,87	AgGttAACGaTTGCGT	
<i>ydiJ</i>	-101	4,06	ACGCAAACGaTTaaGc	
<i>S. typhimurium</i>	<i>purR</i>	-57	4,76	AgGCAAACGTTTaaCcT
		29	4,07	ggGCAAACGTTTcCac
	<i>prsA</i>	-474	4,72	AaGaAAACGTTTtCGc
	<i>purC</i>	-130	4,67	ACGCAAACGTgTGCGT
	<i>purA</i>	-121	4,17	AgGaAAACGaTTGgcT
		-23	3,77	AaGCAAACGgTgattT
	<i>purT</i>	-38	5,08	ACGCAAACGTTTtCGT
	<i>purL</i>	-89	4,94	ACGCAAACGgTTtCGT
	<i>purB</i>	185	4,71	ACGCAAACGgTTaCcT
	<i>purEK</i>	-86	4,61	ACGCAAcCGTTTtCcT
	<i>cvpA-purF</i>	-71	4,93	ACGCAAACGTTTtCtT
	<i>purMN</i>	-15	4,55	tCGCAAACGTTTGcT
	<i>purHD</i>	-122	4,76	gCGCAAACGTTTtCGT
	<i>guaBA</i>	-79	4,42	AgGCAAcCGaTTaCGT
	<i>yhhP</i>	-76	3,98	ACGCAAtCGTTaaCcT
	<i>codBA</i>	-77	4,86	AaGCAAACGTTTGcT
	<i>upp-uraA</i>	-350	4,63	AaGCAAACGTTTGCGa
	<i>pyrC</i>	-64	4,63	ACGaAAACGTTTcCGc

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>S. typhimurium</i>	<i>pyrD</i>	-102	4,21	AgGaAAACGaTTGCt a
	<i>carAB</i>	-237	3,65	gCGCAAAGCGTTTTtCt a
	<i>speAB</i>	-133	3,67	AaGaAAcCGgTcGCGc
	<i>gcvTHP</i>	-107	4,48	AaGagAACGTTTGCCT
	<i>folD</i>	-40	4,44	t aGCAAACGTTTTtCtT
	<i>glyA</i>	-125	4,62	AtGCAAtCGTTTGCCT
	<i>gltS</i>	-272	4,15	gCGCAAACGgTTGCGc
		-191	4,75	AaGCAAtCGTTTGCCT
	<i>yicE</i>	-69	4,29	tCGCAAtCGTTTGCtT
	<i>yieG</i>	-117	4,21	ACGCAAtCGTTgcCGT
	<i>yjcD</i>	-272	4,22	ACGatAACGTTTGCc
		-129	3,87	AgctAAACGTTTGCtT
	<i>mt</i>	30	3,79	gtGCgAcCGTTTTtCGT
	<i>xseA</i>	-109	4,09	ACGtAAAtCGgTTGCaT
<i>yhhQ</i>	-109	3,87	AgGttAACGaTTGCCT	
<i>ydiJ</i>	-101	4,06	ACGCAAACGaTTaaGc	
<i>Y. pestis</i>	<i>purR</i>	-61	3,61	AgGCAAACGaTTaaca
	<i>purT</i>	-86	4,82	ACGCAAtCGTTTTtCGT
	<i>purL</i>	-131	4,94	ACGCAAACGgTTtCGT
	<i>purB</i>	194	4,32	ACGCAAACGcTTaCcT
	<i>purEK</i>	-115	4,61	ACGCAAACGTTTTtCcT
	<i>cvpA-purF</i>	-75	4,93	ACGCAAACGTTTTtCtT
	<i>purMN</i>	-241	4,55	tCGCAAACGTTTGCcT
	<i>purHD</i>	-110	4,65	ACGCAAtCGTTTTtCGc
	<i>guaBA</i>	-90	4,31	AgGCAAtCGaTTaCGc
	<i>yhhP</i>	-30	3,66	ACGCAAACGaTTaaGg
	<i>upp-uraA</i>	-138	4,63	AgGCAAACGTTTGCga
	<i>pyrD</i>	-253	4,45	tCGCAAACGTTTTaCtT
	<i>carAB</i>	-212	3,96	tCGCAAACGTTTTaCtc
	<i>pyrBI</i>	-323	3,86	AgGtAAAtCGTTTGCt a
	<i>gcvTHP</i>	-97	3,60	AaGgtAtCGTTTGCCT
	<i>folD</i>	22	3,62	ggtaAAACGaTTGCGc
	<i>glyA</i>	-324	4,12	AgcCAAtCGTTTGCCT
	<i>serA</i>	-9	4,08	AgGCAAtCGTTTGT aT
	<i>gltS</i>	-183	4,58	AaGCAAtCGTTTGCc
	<i>yicE</i>	-19	4,46	gCGCAAACGaTTGCtT
	<i>yjcD</i>	-205	3,61	AgatAAAtCGTTTGCcT
		-72	4,29	ACGgAAACGaTTaCGT
	<i>xseA</i>	-107	3,84	gCGtAAAtCGaTTGCcT
<i>Y. pseudotuberculosis</i>	<i>purR</i>	-63	3,61	AgGCAAACGaTTaaca
	<i>purT</i>	-84	4,82	ACGCAAtCGTTTTtCGT
	<i>purL</i>	-129	4,94	ACGCAAACGgTTtCGT
	<i>purB</i>	183	4,32	ACGCAAACGcTTaCcT
	<i>purEK</i>	-115	4,61	ACGCAAACGTTTTtCcT
	<i>cvpA-purF</i>	-73	4,93	ACGCAAACGTTTTtCtT
	<i>purMN</i>	-281	4,55	tCGCAAACGTTTGCcT
	<i>purHD</i>	-179	4,65	ACGCAAtCGTTTTtCGc
	<i>guaBA</i>	8	4,31	AgGCAAtCGaTTaCGc

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>Y. pseudotuberculosis</i>	<i>yhhP</i>	-171	3,66	ACGCAAACGaTTaaGg
	<i>upp-uraA</i>	-169	4,63	AgGCAAACGTTTGCGa
	<i>pyrD</i>	-255	4,45	tCGCAAACGTTTaaCtT
	<i>carAB</i>	-241	3,96	tCGCAAacCGTTTaaCt c
	<i>pyrBI</i>	-323	3,86	AgGtAAaCGTTTGCTa
	<i>gcvTHP</i>	-97	3,60	AaGgtAtCGTTTGCGT
	<i>folD</i>	22	3,62	ggt aAAACGaTTGCGc
	<i>glyA</i>	-322	4,12	AgcCAAaCGTTTGCGT
	<i>serA</i>	-111	4,08	AgGCAAaCGTTTGtaT
	<i>gltS</i>	-185	4,58	AaGCAAaCGTTTGCGc
	<i>yicE</i>	-68	4,46	gCGCAAACGaTTGCtT
	<i>yjcD</i>	-205	3,61	AgatAAaCGTTTGCCt
	<i>xseA</i>	-72	4,29	ACGgAAACGaTTaCGT
	-109	3,84	gCGtAAaCGaTTGCcT	
<i>P. carotovorum</i>	<i>purR</i>	-68	3,61	AgGCAAACGaTTaaca
	<i>prsA</i>	-353	4,89	AaGaAAACGTTTtCGT
	<i>purT</i>	-60	5,08	ACGCAAACGTTTtCGT
	<i>purL</i>	-112	4,94	ACGCAAACGgTTtCGT
	<i>purB</i>	183	4,40	ACGCAAACGcTTtCcT
	<i>purEK</i>	-131	4,61	ACGCAAacCGTTTtCcT
	<i>cvpA-purF</i>	-89	4,93	ACGCAAACGTTTtCtT
	<i>purMN</i>	-128	4,38	tCGCAAACGaTTGCcT
	<i>purHD</i>	-158	4,91	ACGCAAACGTTTtCGc
	<i>guaBA</i>	-78	4,31	AgGCAAaCGaTTaCGc
	<i>yhhP</i>	-81	4,33	gCGCAAACGTTgtCGT
	<i>upp-uraA</i>	-241	4,37	AgGCAAaCGTTTGCGa
	<i>pyrD</i>	-125	3,93	cCGCAAACGTTTtacT
	<i>carAB</i>	-238	4,04	taGCAAacCGTTTaaCtT
	<i>pyrBI</i>	-288	4,14	AaGCAAaCGaTTGCGa
	<i>gcvTHP</i>	-63	3,87	AaGatAtCGTTTGCGc
	<i>folD</i>	20	3,62	ggt aAAACGaTTGCGc
	<i>glyA</i>	-151	3,76	AgctAAaCGTTTGCGT
	<i>serA</i>	-97	4,08	AaGCAAaCGTTTGtaT
	<i>tsx</i>	-130	4,67	AgGCAAaCGTTTtCGT
	<i>yicE</i>	-73	4,46	taGCAAACGTTTGcT
	<i>yjcD</i>	-85	4,85	ACGCAAACGTTTaaCtT
	<i>xseA</i>	-110	3,84	gCGtAAaCGaTTGCcT
<i>yhhQ</i>	-151	4,22	ACGacAAACGTTTGCGc	
<i>ydiJ</i>	-238	3,60	gCGCAcACGaTTGCGa	
<i>P. luminescens</i>	<i>prsA</i>	-294	4,71	AaGaAAACGTTTcCGT
	<i>purL</i>	-90	4,94	ACGCAAACGgTTtCGT
	<i>purB</i>	197	4,32	ACGCAAACGcTTaCcT
	<i>purEK</i>	-85	4,67	ACGCAAaCGTTTtCcT
	<i>cvpA-purF</i>	-71	5,08	ACGCAAACGTTTtCGT
	<i>purMN</i>	-81	4,55	tCGCAAACGTTTGcT
	<i>purHD</i>	-282	5,08	ACGCAAACGTTTtCGT
	<i>guaBA</i>	-79	4,42	AgGCAAaCGaTTaCGT
	<i>upp-uraA</i>	-161	4,63	AaGCAAACGTTTGCGa

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>P. luminescens</i>	<i>speAB</i>	-98	3,89	ttGCAAACGTTTtctT
	<i>gcvTHP</i>	-107	4,04	AaGatAtCGTTTGCGT
	<i>folD</i>	22	4,25	ggGaAAAACGaTTGCGc
	<i>glyA</i>	-125	4,62	AtGCAAtCGTTTGCGT
	<i>gltS</i>	-191	4,75	AaGCAAtCGTTTGCGT
	<i>yicE</i>	-67	4,78	ACGCAAACGaTTGcT
	<i>yjcD</i>	-132	3,76	AgctAAAtCGTTTGCGT
	<i>xseA</i>	-109	4,19	ACGtAAAtCGgTTGcT
	<i>ydiJ</i>	-128	3,97	AgGtAAACGcTTGcT
<i>P. multocida</i>	<i>prsA</i>	-310	3,91	AtGCAAACGaTTGtGa
	<i>purC</i>	-84	4,73	AtGCAAACGTTTGcT
	<i>purL</i>	-84	5,08	ACGCAAACGTTTtCGT
	<i>purB</i>	183	4,59	AgGCAAACGaTTaCtT
	<i>purEK</i>	-77	4,60	AgGCAAtCGTTTGcT
	<i>cvpA-purF</i>	-68	4,93	ACGCAAACGTTTtCtT
	<i>purMN</i>	-73	4,55	tCGCAAACGTTTGcT
	<i>purHD-glyA</i>	-74	4,78	gCGCAAACGTTTGCGT
	<i>codBA</i>	-77	4,86	AaGCAAACGTTTGcT
	<i>speAB</i>	-98	3,89	ttGCAAACGTTTtctT
	<i>folD</i>	-40	4,44	taGCAAACGTTTtCtT
	<i>rpiA-serA</i>	-103	4,34	gCGCAAACGaTTaCGc
	<i>yjcD</i>	-113	4,06	AgcCAAACGaTTGcT
	<i>deoD</i>	-127	4,65	AaGtAAACGTTTGCGT
	<i>pckA</i>	59	3,50	AaGaAAAtCGTTTataa
	<i>glpX</i>	-65	4,31	ACGCAAtCGTTTGcT
	<i>yiiU</i>	-166	4,44	taGCAAACGaTTGCGT
	<i>HD1120</i>	-101	4,01	ttGCAAACGTTTGcT
<i>H. influenzae</i>	<i>prsA</i>	-283	3,60	AtaaAAAACGaTTGCGa
	<i>purC</i>	-31	4,46	taGCAAACGTTTGcT
	<i>purL</i>	-1	4,73	AtGCAAACGTTTGcT
	<i>purEK</i>	-77	4,46	taGCAAACGTTTGcT
	<i>cvpA-purF</i>	-84	4,93	ACGCAAACGTTTtCtT
	<i>purMN</i>	-62	4,55	tCGCAAACGTTTGcT
	<i>purHD</i>	-65	5,01	AaGCAAACGTTTGCGT
	<i>folD</i>	-44	3,91	taGCAAACGcTTtCtT
	<i>glyA</i>	-333	3,77	ACGagAAAgtTTtCtT
	<i>rpiA</i>	-36	4,59	ACGCAAtCGTTTaCtT
	<i>yjcD</i>	-85	4,85	ACGCAAACGTTTaCtT
	<i>glpX</i>	-75	4,29	AtGCAAACGaTTaCtc
	<i>yiiU</i>	-102	3,82	gaGtAAAtCGTTTGCaT
	<i>HD1120</i>	-94	4,19	ttGCAAACGgTTGcT
<i>H. ducreyi</i>	<i>prsA</i>	-271	3,60	ggGgAAAACGaTTGCGa
	<i>purC</i>	-91	4,20	taGCAAtCGTTTGcT
	<i>purT</i>	-40	4,95	ACGCAAACGTTTGcT
	<i>purL</i>	-1	4,47	AtGCAAtCGTTTGcT
	<i>purB</i>	185	4,59	AgGCAAACGaTTaCtT
	<i>purE</i>	-68	3,55	AaaCAAtCGTTTGtT

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>H. ducreyi</i>	<i>cvpA-purF</i>	-56	4,93	ACGCAAACGTTTTtCtT
	<i>purHD</i>	-180	5,01	AaGCAAACGTTTGCGT
	<i>purM</i>	-73	4,23	cCGCAAacCGTTTGcTtT
	<i>fhs</i>	-42	4,08	AaGCAAtCGgTTGCca
	<i>deoC</i>	-89	4,86	AaGCAAACGTTTGcTtT
	<i>yhhQ</i>	-135	4,45	ACGCAAACGaTTaCGa
	<i>HD1120</i>	-85	4,32	ACGCAAACcTTTGcTtT
<i>V. cholerae</i>	<i>prsA</i>	-294	4,45	ACGaAAACGaTTGCGa
	<i>purC</i>	-9	4,86	AaGCAAACGTTTGcTtT
	<i>purT</i>	-79	4,84	AaGCAAACGTTTtCtT
	<i>purL</i>	-171	4,81	ACGCAAACGgTTGcTtT
	<i>purB</i>	52	4,11	Agt aAAACGaTTGCGT
	<i>purEK</i>	-88	4,86	AaGCAAACGTTTGcTtT
	<i>cvpA-purF</i>	-65	4,91	AaGaAAACGTTTGCGT
	<i>purMN</i>	-80	4,93	ACGCAAACGTTTtCcT
	<i>purHD</i>	-186	3,64	tgcCAAACGaTTGCGc
	<i>guaBA</i>	-113	3,74	AgtCAAACGaTTGCcT
	<i>gsk</i>	-94	4,41	gtGCAAACGTTTGcTtT
	<i>guaC</i>	-316	3,52	ggGCAAtCGaTTGttc
	<i>ushA</i>	-261	4,23	AgtCAAACGTTTGcCtT
	<i>upp-uraA</i>	-102	4,91	AgGaAAACGTTTGCGT
	<i>pyrBI</i>	-155	3,56	AtGtAAAtCGcTTGCGc
	<i>folD</i>	-11	4,86	AaGCAAACGTTTGcCtT
	<i>glyA</i>	-15	5,00	ACGaAAACGTTTGCGT
	<i>serA</i>	-98	4,76	AgGaAAACGTTTGcTtT
	<i>fhs</i>	3	4,61	gCGCAAACGTTTtCcT
	<i>yicE</i>	-274	4,86	AaGCAAACGTTTGcTtT
		-217	4,53	gCGCAAACGTTTtCcT
	<i>yieG</i>	-38	4,53	ACGCAAACGTTTtCcT
	<i>xseA</i>	-104	3,86	AgGCAAtCGgTTGacT
	<i>cytR</i>	-151	3,56	AgtgAAAtCGTTTtCGT
	<i>pckA</i>	-193	3,98	gCGCAAAGTTTGCGc
	<i>glpX</i>	-74	4,76	AgGCAAACGTTTtCcT
	<i>ppsA</i>	-74	4,76	AgGCAAACGTTTtCcT
	<i>ydiA</i>	-139	4,20	tggCAAtCGTTTGcTtT
	<i>VC1168</i>	-71	4,44	taGCAAACGTTTtCcT
	<i>yhhQ</i>	-160	3,91	ACGCcAtCGTTTGcTc
<i>HD1120</i>	-289	3,62	cattAAACGTTTGCGT	
<i>V. fischeri</i>	<i>prsA</i>	-243	4,19	ACGtAAACGaTTGCGa
	<i>purC</i>	-91	4,20	taGCAAtCGTTTGcTtT
	<i>purT</i>	-73	4,61	taGCAAACGTTTGCGT
	<i>purL</i>	-183	4,81	ACGCAAACGgTTGcTtT
	<i>purB</i>	146	4,36	ACGCAAACGaTTGCGg
	<i>purEK</i>	-87	4,93	ACGCAAACGTTTtCcT
	<i>cvpA-purF</i>	-62	4,91	AaGaAAACGTTTGCGT
	<i>purMN</i>	-81	4,75	ACGCAAACGTTTtCcT
	<i>purHD</i>	-207	4,28	AtGCAAtCGaTTGCGc
	<i>guaBA</i>	-121	3,63	AgcCAAAGCGTTTGCGc

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. fischeri</i>	<i>gsk</i>	-106	4,16	gaGCAAACGTTTGCTa
	<i>guaC</i>	-342	4,36	taGCAAACGTTTaCtT
	<i>ushA</i>	-221	4,60	ACGtAAACGgTTGCGT
	<i>upp-uraA</i>	-142	4,47	AgGgAAACGTTTGCGT
	<i>pyrC</i>	-62	3,99	AaGCAAACGaTTaacT
	<i>pyrD</i>	-246	4,06	AgcCAAACGTTTGCTc
	<i>pyrBI</i>	-227	4,20	AtGCAAACGcTTGCTT
	<i>folD</i>	-87	4,78	ACGCAAACGTTTGCTc
	<i>glyA</i>	-65	4,74	ACGtAAACGTTTGCGT
	<i>rpiA-serA</i>	-133	4,13	cCGaAAcCGTTTGCTT
	<i>fhs</i>	44	4,61	AtGaAAACGaTTGCGT
	<i>yicE</i>	-187	4,69	AaGCAAACGaTTGCTT
	<i>yieG</i>	-38	4,53	ACGCAAACGTTTaCtT
	<i>yjcD</i>	-104	3,63	ACGgtAACGaTTGCTT
	<i>rnt</i>	33	3,94	AaaaAAcCGTTTtCGT
	<i>xseA</i>	-103	3,68	gCGCAAACGcTTGgcT
	<i>cytR</i>	-47	4,18	gaGCAAtCGTTTaCtT
	<i>pckA</i>	-193	3,66	gCGCAAtCGaTTttaT
	<i>glpX</i>	-75	4,66	AaGCAAACGTTTcCcT
<i>yiiU</i>	-308	4,32	AgGgAAACGTTTGCTT	
<i>V. parahaemolyticus</i>	<i>prsA</i>	-290	3,87	gCGtAAACGaTTGCGa
	<i>purC</i>	-133	4,86	AgGCAAACGTTTGCTT
	<i>purT</i>	-82	4,86	AaGCAAACGTTTGCTT
	<i>purL</i>	-160	4,81	ACGCAAACGgTTGCTT
	<i>purB</i>	-88	4,61	gCGCAAACGTTTGCGc
	<i>purEK</i>	-89	4,95	ACGCAAACGTTTGCTT
	<i>cvpA-purF</i>	-62	4,91	AaGaAAACGTTTGCGT
	<i>purMN</i>	-83	4,57	ACGCAAACGTTTGCTa
	<i>purHD</i>	-197	3,64	tgcCAAACGaTTGCGc
	<i>guaBA</i>	-138	3,63	AgtCAAAGCGaTTGCGT
	<i>gsk</i>	-95	4,08	caGCAAACGTTTGCTa
	<i>guaC</i>	-299	4,30	ACGCAAAGGTTTGCGc
	<i>upp-uraA</i>	-113	4,61	taGCAAACGTTTGCGT
	<i>pyrC</i>	-76	3,88	gCGCAAAGCGTTTGCTc
	<i>pyrD</i>	-193	3,80	ctGaAAAtCGaTTGCTT
	<i>pyrBI</i>	-252	3,56	AtGtAAAtCGcTTGCGc
	<i>speAB</i>	95	3,60	ACGCAGACGTTgtgGT
	<i>folD</i>	-86	5,01	AgGCAAACGTTTGCGT
	<i>glyA</i>	-75	5,00	ACGaAAACGTTTGCGT
	<i>serA</i>	-107	3,73	AaGgAAACGTTTgtcc
	<i>fhs</i>	-84	4,61	gCGCAAACGTTTGCGc
	<i>gltS</i>	-296	4,09	ACGCAAtCGgaTGCGT
	<i>yicE</i>	-165	4,29	taGCAAACGaTTGCTT
	<i>yieG</i>	-45	4,33	ACGCAAACGTTTagaT
	<i>yjcD</i>	-104	3,63	ACGgtAACGaTTGCTT
	<i>cytR</i>	-334	3,76	taGCAAACGTTTaacT
	<i>pck</i>	-190	3,81	gCGCAAAGGaTTGCGc

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. parahaemolyticus</i>	<i>glpX</i>	-74	4,16	AgGCAAACGTTTaaCT
	<i>pps</i>	-76,00	4,33	ACGtAAtCGTTTGcCT
	<i>ydiA</i>	-138	4,74	AgGCAAACGaTTaCGT
	<i>VC1186</i>	-194,00	4,46	taGCAAACGTTTGcCT
		-62,00	3,63	AaGaAcAcTtTTtCGT
	<i>yiiU</i>	-287	3,87	AgttAAACGTTTGcCT
	<i>yhhQ</i>	-189	3,70	ggatAAACGTTTGCGT
<i>V. vulnificus</i>	<i>prsA</i>	-302	4,01	ACGgAAACGaTTGCGa
	<i>purC</i>	-135	4,86	AgGCAAACGTTTGcCT
	<i>purT</i>	-82	4,86	AaGCAAACGTTTGcCT
	<i>purL</i>	-158	4,81	ACGCAAACGgTTGcCT
	<i>purB</i>	-90	4,61	gCGCAAACGTTTGCGc
	<i>purEK</i>	-88	4,86	AaGCAAACGTTTGcCT
	<i>cvpA-purF</i>	-64	4,91	AaGaAAACGTTTGCGT
	<i>purMN</i>	-80	4,46	gCGCAAACGTTTGcCT
	<i>purHD</i>	-200	3,64	tgcCAAACGaTTGCGc
	<i>guaC</i>	-288	3,82	gCGCAAAGCGTTTGCGa
	<i>gsk</i>	-92	4,08	taGCAAACGTTTGcTa
	<i>ushA</i>	-243	4,23	AgtCAAACGTTTGcCT
	<i>upp</i>	-108	4,52	gaGCAAACGTTTGCGc
	<i>uraA</i>	-66	3,72	taGCAAtCGTTTGcAa
	<i>pyrC</i>	-49	3,83	ACGgAAAGCGTTTGcCT
	<i>folD</i>	-85	5,01	AaGCAAACGTTTGCGT
	<i>glyA</i>	-72	5,00	ACGaAAACGTTTGCGT
	<i>serA</i>	-104	3,73	AaGgAAACGTTTGtcc
	<i>yicE</i>	-185	4,29	taGCAAACGaTTGcCT
	<i>yieG</i>	-39	4,53	ACGCAAcCGTTTaCtT
	<i>cytR</i>	-230	3,98	taGCAAACGTTTaCca
	<i>pck</i>	-189	4,15	gCGCAAAGGTTTGCGT
	<i>pps</i>	-33	4,01	gCGtAAtCGTTTGcCT
	<i>ydiA</i>	-138	4,57	AgGCAAACGaTTaCGc
	<i>VC1186</i>	-151	3,82	taGCAAtCGTTTGcTa
	<i>yiiU</i>	-230	3,98	taGCAAACGTTTaCca
	<i>yhhQ</i>	-186	3,67	gttCAAtCGTTTGCGT
<i>P. profundum</i>	<i>prsA</i>	-287	4,45	ACGaAAACGaTTGCGa
	<i>purC</i>	-142	4,86	AgGCAAACGTTTGcCT
	<i>purT</i>	-72	5,01	AgGCAAACGTTTGCGT
	<i>purL</i>	-199	4,64	gCGCAAACGgTTGCGT
	<i>purB</i>	-192	4,27	taGCAAACGaTTGCGc
	<i>purE</i>	61	4,57	ACGCAAtCGaTTaCGT
	<i>cvpA-purF</i>	-62	4,65	AgGtAAACGTTTGCGT
	<i>purMN</i>	-80	4,75	ACGCAAACGTTTcCCT
	<i>purHD</i>	-169	4,69	gaGCAAACGTTTGCGT
	<i>guaBA</i>	-83	4,06	AgcCAAACGaTTGcCT
	<i>gsk</i>	-125	3,98	taGCAAAGGTTTGCGT
	<i>guaC</i>	-64	4,58	ACGaAAACGaTTaCcT
	<i>upp</i>	-125	4,47	AgGgAAACGTTTGCGT
		<i>uraA</i>	-65	4,07

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>P. profundum</i>	<i>pyrC</i>	-113	4,21	AgcCAAACGaTTGCGT
	<i>pyrD</i>	-185	4,30	ACGgAAatCGTTTGCGT
	<i>folD</i>	-87	4,84	AaGCAAACGTTTGCGc
	<i>glyA</i>	-95	5,00	ACGaAAACGTTTGCGT
	<i>fhs</i>	-97	4,61	gCGCAAACGaTTGCGT
	<i>yicE</i>	-68	4,46	gCGCAAACGaTTGcT
	<i>yjcD</i>	-94	4,17	taGCAAACGaTTaCGc
		-82	4,68	ACGCAAACGTTTaCtc
	<i>xseA</i>	-191	4,18	AaGCAAtCGTTTGgcT
	<i>glpX</i>	-22	4,15	ACGCAAACGTTTcacT
	<i>pps</i>	-66	4,42	AtGCAAtCGTTTcCGT
	<i>ydiA</i>	-134	4,14	ACGgAAACGaTTGCaT
	<i>yhhQ</i>	-155	4,69	AgGCAAACGaTTGcT

Приложение 9.

Множественное выравнивание для белков Fnr. Участки, соответствующие последовательности НТН, показаны белым шрифтом на темном фоне. Критически важные цистеиновые остатки показаны на голубом фоне. Условные обозначения: “*” – абсолютно консервативная позиция; “:” – позиция с высокой степенью консервативности; “.” – позиция с низкой степенью консервативности. Обозначения геномов: см. Табл. 2.

```

Fnr (ECO) -----MIPEKRIIRRIQSGGCAIHCQDCCSISQLCIPFTLNEHELDQLDNI IERKKPIQKGQTLF
Fnr (STY) ----MLKLTNINYGLSRPMIPEKRIIRRIQSGGCAIHCQDCCSISQLCIPFTLNEHELDQLDNI IERKKPIQKGQTLF
Fnr (PCA) -----MIPEKRIIRRIQSGGCAIHCQDCCSISQLCIPFTLNEHELDQLDNI IERKKPIQKGQALF
Fnr (YPK) -----MIPEKRVIRRIQSGGYAIHCQDCCSISQLCIPFTLNEHELDQLDNI IERKKPIQKGQALF
Fnr (YEN) -----MIPEKRIIRRIQSGGCAIHCQDCCSISQLCIPFTLNEHELDQLDNI IERKKPIQKGQALF
Fnr (VPA) -----MISEKPATKRIQSGGCAIHCQDCCSISQLCIPFTLNESELDQLDQI IERKKPIQKGQELF
Fnr (VVU) -----MISEKPATKRIQSGGCAIHCQDCCSISQLCIPFTLNEAEELDQLDQI IERKKPIQKGQELF
Fnr (VCH) -----MISEKPAAKRIQSGGCAIHCQDCCSISQLCIPFTLNESELDQLDQI IERKKPIQKGQELF
Fnr (VFI) -----MMSDNSANKRIQSGGCAIHCQDCCSISQLCIPFTLNDSELDQLDEI IERKKPY-KGQELF
Fnr (PMU) MILFEIDQFYI IFVVLIKVL FMDTKI GRKVQSGGCAIHCQDCCSISQLCIPFTLNEQELDQLDNI IERKKPIQKSQVLF
Fnr (AAC) -----MKILATETKLGRRVQSGGCAIHCQNCISISQLCIPFTLNQHELDQLDNI IERKKPIQKSQVLF
Fnr (HIN) -----MKNFVAGAQSGRKIQSGGCAIHCQDCCSISQLCIPFTLNEQELDQLDNI IERKKPIQKSQVLF
Fnr (HDU) -----MKIISDAKNVGR TACTIHCQNCISISQLCIPFTLNEHELTQLDNI IERKKPVQKSQIIF
      .:. :****:*****:*****: * * *:***** *.* :*

```

```

Fnr (ECO) KAGDELKSLYAIRSGTIKSYTITEQGDEQITGFHLAGDLVGFDAIGSGHHSFAQALETSMVCEIPFETLDDLSGKMP
Fnr (STY) KAGDELKSLYAIRSGTIKSYTITEQGDEQITGFHLAGDLVGFDAIGSGHHSFAQALETSMVCKIPFETLDDLSGKMP
Fnr (PCA) KAGDELRSLYAIRSGTIKSYTITEQGDEQITGFHLAGDLVGFDAIGTAQHPSFAQALETSMVCEIPFETLDDLSGKMP
Fnr (YPK) KAGDELKSLYAIRSGTIKSYTITEEGDEQITGFHLAGDLVGFDAISNLQHP SFAQALETSMVCEIPFDLDDLSGKMP
Fnr (YEN) KAGDELKSLYAIRSGTIKSYTITEEGDEQITGFHLAGDLVGFDAISNLQHP SFAQALETSMVCEIPFDLDDLSGKMP
Fnr (VPA) KAGDELKSLYAIRSGTIKSYTITEQGDEQITAFHLAGDLVGFDAITGDCHPSFAQALETSMVCEIPYEILDDLSGKMP
Fnr (VVU) KAGDDLKSLYAIRSGTIKSYTITEQGDEQITAFHLAGDLVGFDAITGDSHPSFAQALETSMVCEIPSEILDDLSGKMP
Fnr (VCH) KAGDELRSLYAIRSGTIKSYTITEQGDEQITAFHLAGDLVGFDAITGDQHP SFAQALETSMVCEIPYEILDDLSGKMP
Fnr (VFI) KAGDELKCLYAIRSGTIKSYTITEQGDEQITAFHLAGDLVGFDAITEAQHPSFAQALETSMVCEIPYEILDDLSGKMP
Fnr (PMU) KAGDPLNSLYAIRSGTIKTYTISETGEEQITSFQLPGDLVGFDAIMNQHP SFAQALETSMVCEIPFDLDDLSGKMP
Fnr (AAC) KAGDELTSLYAIRSGTIKSYTISETGEEQITSFHLPGLVGFDAIMNQHP SFAQALETAMVCEIPFDLDDLSGKMP
Fnr (HIN) KAGBSLNSIYAIRSGTIKSYTISESGEEQITSFHLPGLVGFDAITQMHP SFAQALETAMVCEIPFDLDDL LAGKMP
Fnr (HDU) QSGDELRSIYAIRSGTIKSYTISESGEEQITAFHLPGNLVGFDAITNMRHLGFAQALETSMICEIPFDLDDL LAGKMP
      :.* * .:*****:****:* :****.*:*:*:***** * .*****:*.***: * :****:****:

```

```

Fnr (ECO) NLRQQMRLMSGEIKGDQDMILLSSKNAEERLAAFIYNLSRRFAQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (STY) NLRQQMRLMSGEIKGDQDMILLSSKNAEERLAAFIYNLSRRFAQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (PCA) NLRQQMRLMSGEIRGDQDMILLSSKNAEERLAAFVYNLSRRFAQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (YPK) NLRQQIMRLMSGEIKGDQDMILLSSKNAEERLAAFIYNLSHRFAQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (YEN) NLRQQMRLMSGEIKGDQDMILLSSKNAEERLAAFIYNLSRRFAQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (VPA) KLRQQIMRLMSNEIKGDQEMILLSSKNAEERLAAFLYNLSTRFSQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (VVU) KLRQQIMRLMSNEIKGDQEMILLSSKNAEERLAAFLYNLSTRFSQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (VCH) KLRQQIMRLMSNEIKGDQEMILLSSKNAEERLAAFLYNLSTRFSQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (VFI) KLRQQIMRLMSNEIKGDQEMILLSSKNAEERLAAFLYNLSTRFHQRGFS PREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (PMU) KLRQQIMRLMSNEIKNDQEMILLSSKNAEERLAAFIYNLSRRYSARGFSAREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (AAC) KLRQQIMRLMSNEIKSDQEMILLSSKNAEERLAAFIYNLSQRY SARGFSAREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (HIN) KLRQQILRLMSSEIKSDQEMILLSSKNAEERLAAFIHNLSKRY SARGFSAREFRLT MTRGDIGNYLGLTVETISRLL
Fnr (HDU) KIRHQIMRLMSNEIKSDQEMILLSSKNAEERLAAFIHNLSQRYAERGFSAREFRLT MTRGDIGNYLGLT IETISRLL
      :.*:*:****.**:.*:***** * * :*****:****:.* * : * * .*****:*****:*****

```

```

Fnr (ECO) GRFQKSGMLAVKGYIT IENNDAL AQLAGHTRNVA-----
Fnr (STY) GRFQKSGMLAVKGYIT IENSDAL AALAGHTRNVA-----
Fnr (PCA) GRFQKSGTLAVKGYIT IENIDAL SELAGSSRK-----
Fnr (YPK) GRFQKSNILSVKGYIT IENNEAL AQLAGNPKPRL-----
Fnr (YEN) GRFQKSDILSVKGYIT IENTEAL AQLAGNPKPSCN-----
Fnr (VPA) GRFQKSEILSVKGYIT ILDHDA LMELAGVTKE-----
Fnr (VVU) GRFQKSEILSVKGYIT IILDHDA LMELAGVSKE-----
Fnr (VCH) GRFQKSEILSVKGYIT IILNHA ELMELAGVSKE-----
Fnr (VFI) GRFQKTEMLTVKGYIT IINDHDA LAELAGSAKEIK-----
Fnr (PMU) GRLQKMGILSVQGYIT IINDMVA LIELSGTTKTNIKMG I--
Fnr (AAC) GRFQKLGVL SVQGYIT IINNMA ELIELSGTNKNKIQLI---
Fnr (HIN) GRFQKLGVISVQGYIT IINDLNGLIELTGTNKTKITLVK--
Fnr (HDU) GRFQKSGMITVQGYIT IINHMDK LIDMAGVTRSQISVTQSA
      **:* * :*:****.* * . * :***:

```

Приложение 10.

Приложение 11.

Множественное выравнивание для белков ArcB. Участки, соответствующие последовательности НТН, показаны белым шрифтом на темном фоне. Критически важные цистеиновые остатки показаны на голубом фоне. Условные обозначения: “*” – абсолютно консервативная позиция; “:” – позиция с высокой степенью консервативности; “.” – позиция с низкой степенью консервативности. Обозначения геномов: см. Табл. 2.

```

ArcB (ECO)    MKQIRLLAQYYVDLMMKLGLVRF SMLLALALVVLAI VVQMAVTMVLHGQVESIDVIRSIFFGLLITPWAVYFLSVVV
ArcB (STM)    MKQIRMLAQYYVDLMMKLGLVRF SMLLALALVVLAI VVQMAVTMVLHGQVESIDVIRSIFFGLLITPWAVYFLSVVV
ArcB (STY)    MKQIRMLAQYYVDLMMKLGLVRF SMLLALALVVLAI VVQMAVTMVLHGQVESIDVIRSIFFGLLITPWAVYFLSVVV
ArcB (YPK)    MKQIRVLAQYYVDLMVKLGLVRF SLLLASALVLLAMVVQMAVTFVLRGQSVETL DLVRSIFFGLLITPWAVYFLSVVV
ArcB (YEN)    MKQIRVLAQYYVDLMVKLGLVRF SLLLASVLLAMVVQMAVTFVLRGQSVETL DLVRSIFFGLLITPWAVYFLSVVV
ArcB (PCA)    MKQIRLLAQYYVDLMVKLGLVRF SLLLASVLLAMVVQMAVTFVLRGQSVETL DLVRSIFFGLLITPWAVYFLSVVV
ArcB (VPA)    MKPMKNLAQYYVDLLVVKLGIVRF SLLLALALVALAVVVQV GITLALKGRVDDIDIVRSVFFGLLITPWAVYFLSVVV
ArcB (VVU)    MKPMKNLAQYYVDLLVVKLGIVRF SLLLALALVALAVVVQV GITLVLKGNVDDIDIVRSVFFGLLITPWAVYFLSVVV
ArcB (VCH)    MKPMKNLAQYYVDLLVVKLGLRFS ILLALALVALAVVVQV GITLLINGFVDDIDII RSVFFGLLITPWAVYFLSVVV
ArcB (VFI)    MKQLKILAQFYVDLMVRLGIVRF SLLLAIALVALAVIVQV GITLVL SGEVNDLDIIRSVFFGLLITPWAVFFLSVVV
ArcB (AAC)    MKNLKDFVRRYVDWVIRLGRIRFS LLGVMVLAVLALCTQ ILLSLLVMGKILWADVARIVFGLISAPFVIYFFTVLV
ArcB (PMU)    MKNIKYFAQKYVDWVIRLGRIRFS ILGVIVLAILALCTQ ALLSFL IIGEVYVWDMIRSVVFG LISAPFVIYFFTVLV
      * : : : * * : : * * : * * * . . * . * : : * : * : * : * * : * * * : * : : * * *
ArcB (ECO)    EQLEESRQRLSRLVQKLEEMRERDLSLVNQLKDNIAQLNQE IAVREKAEAEELQETF GQKLEIKEREETQIQLEQQS
ArcB (STM)    EQLEESRQRLSRLVQKLEEMRERDLSLVNQLKDNIAQLNQE IADREKAEAEELQETF EQLKVEIKEREEAQIQLEQQS
ArcB (STY)    EQLEESRQRLSRLVQKLEEMRERDLSLVNQLKDNIAQLNQE IADREKAEAEELQETF EQLKVEIKEREEAQIQLEQQS
ArcB (YPK)    EQLEESRQRLSRLVDKLEVMRHRDLELNKQLTENITQLNQE IVEREKAEKAHLQVVDK LKEEMGHREQAQIELGQQS
ArcB (YEN)    EQLEESRQRLSRLVDKLEVMRYRDLELNKQLTENIAQLNQE IVEREKAEKAHLQVVDK LKEEMGHREQAQVELGQQS
ArcB (PCA)    EQLEESRQRLSKLVAKLEEMRHRDLELNVLQENIAQLNQE IADRIKAEEARVLVMSR LKEEMSRREQAQIELEQQS
ArcB (VPA)    DQLEESRQR LTKLVSKLKDMRSR DQELNFQLQONIEKLNLEIEERIKAE EAREEAMQDLENEVFQRERTQLELAERT
ArcB (VVU)    DQLEESRQR LTKLVSKLKDMRSR DQELNVQLQONIEQLNQEIEERIRAE EAREEAMQDLENEVFQRERTQLELAERT
ArcB (VCH)    DQLEESRQR LTKLVSKLKDMRSR DQELNQLQONIVKLNQEIEERIKAE EAREEAMADLENEVYQREKTQVELAERT
ArcB (VFI)    EHLED SRQR LSKLVSKLEEMRTRDLDLNK KLGKGNVKKLNQEIEERIKAE EAREEAMADLENEVYQREKAQLELAEQT
ArcB (AAC)    EKLEHSRQALS-----CSVEELCQERVS AEKKLSEALDN-----
ArcB (PMU)    EKLEHSRGLA-----KMVENLRKEV SERVLA EKKLSIALDN-----
      : : * * * * : : : * * * : * * * * * * * * * * * * * * * * * * * * * * * *
ArcB (ECO)    SF LRSFLDASPD LVFYRNE DKEFSG CNRAME LLTGKSEKQLVHLKPADVYSPEAAAKVIETD EKVFRHNVS LTYEQW
ArcB (STM)    SF LRSFLDASPD LVFYRNE DKEFSG CNRAME LLTGKSEKQLVHLKPEDVYSPEAAEKVIETD EKVFRHNVS LTYEQW
ArcB (STY)    SF LRSFLDASPD LVFYRNE DKEFSG CNRAME LLTGKSEKQLVHLKPEDVYSPEAAEKVIETD EKVFRHNVS LTYEQW
ArcB (YPK)    ALLRSFLDASPD LVYYRNE DNEFSG CNRAME LLTGKSEKQLVGLTPKDVYAPDIAEKMETD EKVFRHNVS LTYEQW
ArcB (YEN)    ALLRSFLDASPD LVYYRNE DNEFSG CNRAME LLTGKSEKQLVGLTPKDVYAPDIAEKMETD EKVFRHNVS LTYEQW
ArcB (PCA)    ALLRSFLDASPD LVYYRNE EKEFSG CNRAME LLVGK SQQLIGLTPQDVYAPDIAEKMETD EKVFRHNVS LTYEQW
ArcB (VPA)    ALLRSFIDASPD LIYYRNA DGVSFG CNRAME EELTGKKEHQVGLSPWDVYSKEIAQP IVDTDQKVS DNQAITYEQW
ArcB (VVU)    ALLRSFIDASPD LIYYRNA EGVFSG CNRAME EELTGKKEHQVGLSPWDVYSKEVAQQ IVDTDQV FANNRALT YEQW
ArcB (VCH)    ALLRSFIDASPD LIYYRNA KGEFSG CNRAME EELTGKRESELVGLTPWDVYRKEIAQS IVETDQV FNNNTAITYEQW
ArcB (VFI)    ALLRSYLDTS PDIIYYRNDLGLFSG CNKAME QLTGKLEKELIGLSPKDVYDAEVAAKVIDTDRQVLQENQALTYEQW
ArcB (AAC)    -----
ArcB (PMU)    -----
ArcB (ECO)    LDYDPGRKACFEIRKVPYYDRVGKRHGLMGFGRDITERKRYQDALERASRDKTTFFISTISHELRTP LINGIVGLSRIL
ArcB (STM)    LDYDPGRKACFEIRKVPYYDRVGKRHGLMGFGRDITERKRYQDALERASRDKTTFFISTISHELRTP LINGIVGLSRIL
ArcB (STY)    LDYDPGRKACFEIRKVPYYDRVGKRHGLMGFGRDITERKRYQDALERASRDKTTFFISTISHELRTP LINGIVGLSRIL
ArcB (YPK)    LVYDPGRKACFE LRKVPFYDRVGKRHGLMGFGRDITERKRYQDALENASRDKTTFFISTISHELRTP LINGIVGLSRIL
ArcB (YEN)    LVYDPGRKACFE LRKVPFYDRVGKRHGLMGFGRDITERKRYQDALENASRDKTTFFISTISHELRTP LINGIVGLSRIL
ArcB (PCA)    LVYDPGRKACFE LRKVPFYDRMGKRHGLMGFGRDITERKRYQDALENASRDKTTFFISTISHELRTP LINGIVGLSRIL
ArcB (VPA)    LEYDPGRKHYFELR KVPFYSKDG RHLGLVGFGRDITERKRHEESLEKASRDKTTFFISTISHELRTP LINGIVGLSRIL
ArcB (VVU)    LEYDPGRKHYFELR KVPFYSKDG RHLGLVGFGRDITERKRHEESLEKASRDKTTFFISTISHELRTP LINGIVGLSRML
ArcB (VCH)    LEYDPGRKSFFELR KVPFYNKDGR HGLGLVGFGRDITERKRHEESLEKASRDKTTFFISTISHELRTP LINGIVGLSRML
ArcB (VFI)    LDYDPGRKACFE LNKIPFYSKDGRRLGLVGFGRNITERKEYEDALQKASRDKTTFFISTISHELRTP LINGIVGLSRML
ArcB (AAC)    -----LEKINRDKTTLMTTISHELRTP LINGIIGLSRIL
ArcB (PMU)    -----IEKTSRDKTRLMATISHELRTP LINGIIGLSRIL
      : : . . * * * * : : : * * * * * * * * * * * * * * * * * * * * * * * *

```

ArcB (ECO) LDTELTAEQEKEYLKT IHVSAVTLGNIFNDI IDMDKMERRKVQLDNQPVDFTSFLADLENLSALQAQQKGLRFNLEPT
 ArcB (STM) LDTDLTAEQEKEYLKT IHVSAVTLGNIFNDI IDMDKMERRKVQLDNQPIDFTSFMADLENLSGLQAQQKGLRFNLEPT
 ArcB (STY) LDTDLTAEQEKEYLKT IHVSAVTLGNIFNDI IDMDKMERRKVQLDNQPVDFTSFMADLENLSGLQAQQKGLRFNLEPT
 ArcB (YPK) LDTELDSEQLKYLKT IHVSAITLGNIFNDI IEMDKLERRKVQLDNQPIDFTGFMADLENLSGLLVQPKGLKFIPEPQ
 ArcB (YEN) LDTELDSEQLKYLKT IHVSAITLGNIFNDI IEMDKLERRKVQLDNQPIDFTGFMADLENLSGLLVQPKGLKFIPEPQ
 ArcB (PCA) LDTQLDLEQQKYLKT IHVSAITLGNIFNDI IEMDKQERRKVQLDNQPIDFTGFLVDLENLGGLLAEQKGLKIMDQH
 ArcB (VPA) LDSQLTEEQQRKHMQT INVSAITLGNIFNDI IDMDKFDRRKLELLPAPLNFEFVAEIESI SALMAEQKGLRFDLERL
 ArcB (VVU) LDTRLTDEQRKYMQT INVSAITLGNIFNDI IDMDKFDRRKLELLPAPLNFEFVAEIESI SALMAEQKGLRFDLERL
 ArcB (VCH) LDTPLSGEQRKHLQT INVSAITLGNIFNDI IDMDKFDRRKLELFPALNFEFVAEIEVLAALMAEQKGLRFDLERL
 ArcB (VFI) LDSRLDDEQHNYMRT INVSAITLGNIFNDI IDMDKFDRRKLELFPKPLDFISFVTEMENISGLMAEQKGLRFDLERL
 ArcB (AAC) LEENPSEQQNYLKT INSSTLSLAHFSDI IDLEKIDAKRIELNRKATDLYALLNDIANFALLMTEEKHLQFQLVCP
 ArcB (PMU) LDEKLTETQQNYLKT IHVSAISLGHIFSDI IDLEKIDARRIELNAKAVELNSFLNDI SNVARLLAQQKGLAFQLVTD
 * : * : : * : * : : * : * : : * : * : : * : * : : * : * : : * : * : : * : * : : * : * : :

ArcB (ECO) LPLPHQVITDGTRLRQILWNLISNAVKFT-----QQGQVTVRVRYDEGD---MLHFEVEDSGIGIPQDELDKIFAM
 ArcB (STM) LPLPHKVITDGTRLRQILWNLISNAVKFT-----QQGQVTVRARYDEGD---MLHFEVEDSGIGIPQDEQDKIFAM
 ArcB (STY) LPLPHKVITDGTRLRQILWNLISNAVKFT-----QQGQVTVRARYDEGD---MLHFEVEDSGIGIPQDEQDKIFAM
 ArcB (YPK) LPLPEKVIADGTRLRQILWNLIGNAVKFT-----QQGKIVVRVRREGND---RLIFEVEDSGMGIPEDQDKIFAM
 ArcB (YEN) LPLPEKVIADGTRLRQILWNLIGNAVKFT-----QQGKIVVRVRRESND---RLIFEVEDSGMGIPEDQDKIFAM
 ArcB (PCA) QPLPQKVIADGTRLRQILWNLISNAVKFTPKGEGEKGIVVRVWHEKGD---RLRFEVEDSGMGIPADELEKIFAM
 ArcB (VPA) SDLPKAEVDATRLRQVIWNLVSNAMKFT-----KEGGVMTVVSADVEEDFATIIIEVEDTGIGIPEAELEKIFAM
 ArcB (VVU) SELPKAEVDATRLRQVIWNLISNAMKFT-----KEGGVMTVVSSEVEEDFATIIIEVEDTGIGIPEAELEKIFAM
 ArcB (VCH) TDLPPLVEVDSTRLRQVLWNLISNAMKFT-----KEGGVMTVSAECDQAHAETIEVEDTGIGIPEAEQEKIFAM
 ArcB (VFI) TDLPPLHISVDSTRLRQVLWNLVGNATKFT-----KEGGVMSVSADIHQDSAEITFEIEDTGIGIPEEELEKIFAM
 ArcB (AAC) PTLPNWMLDGVRLNQVLWNLINNAVKFT-----QQGSVTLSEVQTAEE---EFALRVTDTGIGIVEQDLQKIFEL
 ArcB (PMU) KNLPTWFMVDYARLSQILWNLIGNAVKFT-----QEGRITLNVKRVSEY---QISFSVSDTGIGIPANELPRVFEM
 * * . * . * * * : : *

ArcB (ECO) YYQVKDSHGGKPATGTGIGLAVSRRLAKNMGGDITVTVSEQKGGSTFTLTIHAPSAVEEVDADFEDDDMPLPALNVLL
 ArcB (STM) YYQVKDSNGGKPATGTGIGLAVSRRLAKNMGGDITVSSLPKGGSTFTLTVHAPAVAEVEEDADFEDDDMPLPALNVLL
 ArcB (STY) YYQVKDSHGGKPATGTGIGLAVSRRLAKNMGGDITVSSLPKGGSTFTLTVHAPAVAEVEEDADFEDDDMPLPALNVLL
 ArcB (YPK) YYQVKDRNGGRPATGTGIGLAVSKRLAQSMGGDITVKSTQGVGSCFTLTIKAPAVQEAASN-APSGDDMPLPALNVLL
 ArcB (YEN) YYQVKDSNGGRPATGTGIGLAVSKRLAQSMGGDITVKSQVQAGSCTFTTIKAPAVQAASS-APSGDDTGLPALNVLL
 ArcB (PCA) YYQVKDQHGGKPATGSGIGLAVSKRLAQSMGGDITVSSQKGGSCFTLTVIAPSADEAGSGLEDLDDLPALNVLL
 ArcB (VPA) YYQVKSQKDNLHAGVTGIGLAVSKQLINMMDDGDI TVSEEGFGSTFTVSIHVPIVELEEPVIDAKRDN---VHLNIFM
 ArcB (VVU) YYQVKSQKDNLHAGVTGIGLAVSRQLINMMDDGDI TVSSEEGFGSTFTI SIHVPIVIESAPSTVDKPKTQ--LHLSIFM
 ArcB (VCH) YYQVKSQKDNLHAGVTGIGLAVSRQLIKLMGGDITVSNSEEGFGSTFTVTVIRVPLLAEP I E IEPQEPQS--TELNIFM
 ArcB (VFI) YYQVKQKDNLHAGVTGIGLAVSKELINLMGGDITVTVSEVDFGSTFTVTVINPLVDHVEEEEAMPETKLSKSLRIFM
 ArcB (AAC) YYQAGS--DANKSLGSGIGLSVSKTIAQLMGGDLTVSSEVKGSTFLFTFKARQAIPVE---EDEHMLPLKNILL
 ArcB (PMU) YYQVQS--SGHKPLGSGIGLAISQTI AQLMGGDL YATSEMKGSCFTLMLPVKEVEKPV---VQNQMPKLKNILL
 * * * . . . * : *

ArcB (ECO) VEDIELNVIVARSVLEKLGNSVDVAMTGAAALEMFKPGYDLVLLDIQLPDMTGLDISRELTKRYPR---EDLPPLV
 ArcB (STM) VEDIELNVIVARSVLEKLGNSVDVAMTGAAALEMFPGEYDLVLLDIQLPDMTGLDIARELTRRHTR---EDLPPLV
 ArcB (STY) VEDIELNVIVARSVLEKLGNSVDVAMTGAAALEMFPGEYDLVLLDIQLPDMTGLDIARELTRRHTR---EDLPPLV
 ArcB (YPK) VEDIELNVIVARSVLEKLGNSVDVAMNGHDALAMFNPEDFDLVLLDIQLPDMSGLDIARQIRAEYK---QSLPPLV
 ArcB (YEN) VEDIELNVIVARSVLEKLGNSVEVAMNGHDALAMFNPEDFDLVLLDIQLPDMSGLDIARQIRADYQ---QSLPPLV
 ArcB (PCA) VEDIELNVVARSVLEKLGNSVDVAMTQGEALDMFDPDFDLVLLDIQLPDMTGLDVARQLRSRYG---RSLPPLV
 ArcB (VPA) VEDIELNVTVAKSLESLSGHVSVTAMTGEEALVNFVDPDYDLALLDIQLPDMTGFVDAKYREHYD---HLPLPLV
 ArcB (VVU) VEDIAALNVTVARSLESLSGHKVTAMTGDEAIRSFDPSRYDLVLLDIQLPDMTGFVDAAYRKHYS---QLPPLV
 ArcB (VCH) VEDIELNITVARSLESLSMGHKVTAMTGEEAIQGFNPTEYDLVFLDIQLPDMTGFDI AHYYRTHYS---SLPPLV
 ArcB (VFI) LEDIELNVTVACSLLESLSGHTVDVAMCGEDA I KQFDP T KYDLALLDIQLPDMTGFVDAKILRDKFD---NLPPLI
 ArcB (AAC) VEDIENVVVAKSMLEKLGQIDIAMTGAEAIRKFEQNYDLVFLDIQLPDMSGFDIAAHFRQNYENGVYDFLPLLI
 ArcB (PMU) VEDIENVVVAKSVLEKLGYYVDVAMTQQAIEKFEQYDYDLVLLDIQLPDMSGFAVAKYLRGVEDGRYDFLPLLI
 : * * * : * : * * * : * : * * * : * : * * * : * : * * * : * : * * * : * : * * * : * : * * * : * : * * * :

ArcB (ECO) ALTANVLKDKQEYLNAGMDDVLSKPLSVPALTAMIKKFWDTQDD--EESTVTTEEN-----SKSEALLDIPMLE
 ArcB (STM) ALTANVLKDKKEYLDAGMDDVLSKPLSVPALTAMIKKFWDATDK--EESTVTPEES-----DKAQALLDIPMLE
 ArcB (STY) ALTANVLKDKKEYLDAGMDDVLSKPLSVPALTAMIKKFWDATDK--EESTVTPEES-----DKAQALLDIPMLE
 ArcB (YPK) ALTANVLKDKKEYLDAGMDDVLSKPLSVPALTAMIKQFWDK---PSSAVKQKQEHKVM-----QTHESLLDIPMLE
 ArcB (YEN) ALTANVLKDKKEYLDAGMDDVLSKPLSVPALTAMIKQFWDHK---PSSAAKQEHKVM-----QTHESLLDIPMLE
 ArcB (PCA) ALTANVLKDKREYLDAGMDDVLSKPLSVPALTAVIKQHWDTHTVWTEESVVEEGDEMA-----KVKEDLLDIPMLE
 ArcB (VPA) ALTANVMKDRREYLENGMDDAISKPLAVHAIQSVIDKFFVD--RVDKPAVIVKPEVKEEAAPSPVIETHLLDIPMLE
 ArcB (VVU) ALTANVLKDKREYLDNGMDEAISKPLAVQAIQDVIEKLIK--QESTTIEMEEAEVVEITTS--HVENHLLDIPMLE
 ArcB (VCH) ALTANVLKDKKEYRQKGMDAAI SKPLSVAAVIAKMTQH--HAGESVAKVKSNEKEKELPD--DLYQQLLDLEMLQ
 ArcB (VFI) ALTANLLKEKDEYFKNGMDEALS KPLSVKALTEVLEQFSMSEKKT DIPSVPQEQEVSEPTENDQLIECLLDIPMLE
 ArcB (AAC) ALTANVVQKQKEYLAQGMDDVHVKPLSLEELRHLCHLDYFGEELTQFNLP--SNKPQ-----AESVELDTKMLT
 ArcB (PMU) ALTANVMHDKKEYLAQGMDDVLRKPLSIEELTACLHVEYFGHDVSLCLETPTTHQIT-----QDATHFDVKMLS
 * * * * : : : * * * * : * : * * * : * : * * * : * : * * * : * : * * * : * : * * * : * : * * * :

ArcB (ECO) QYLELVGPKLITDGLAVFEKMMPGYVSVLESNLTAQ-----DKKGIVEEGHKIKGAAGSVGLRHLQQLGQQIQSPDL
 ArcB (STM) QYIELVGPKLITDGLAVFEKMMPGYLSVLESNLTAQ-----DKKGVVEEGHKIKGAAGSVGLRHLQQLGQQIQSPDL
 ArcB (STY) QYIELVGPKLITDGLAVFEKMMMLGYLSVLESNLTAQ-----DKKGVVEEGHKIKGAAGSVGLRHLQQLGQQIQSPDL
 ArcB (YPK) QYIDLVGPKLIHQSLFMFEQMMPGYLAVLDSNMTAR-----DQKGITEEAHKIKGAAGSVGLRHLQQLGQQIQTPTL
 ArcB (YEN) QYIDLVGPKLIHQSLFMFEQMMPSYLAVLDSNMTAR-----DQKGITEEAHKIKGAAGSVGLRHLQQLGQQIQTPTL
 ArcB (PCA) QYLDLVGPKLIHQSLFMFEQMMPGYLAILDSNMTAR-----DQKGITEEGHKIKGAAGSVGLRHLQQLGQQIQTSLL
 ArcB (VPA) SYVDIVGVKPVYDSIKMFEDMMPDYIEVLDSNMTAK-----DQDGIIVSEAHKIKGAAGSIGLKRISVAQKAQSPDM
 ArcB (VVU) SYVEIVGPPVLDLSIRMFEEMMPEYLEILDSNMIAK-----DQPGIVSEAHKIKGAAGSIGLKRISVAQKAQSPDM
 ArcB (VCH) SYVEIVGSPVIDSVHLFEQSMPAYLAVLDSNMVAK-----DQEGIVSEAHKIKGAAGSVGLKRISVAQKAQSPDM
 ArcB (VFI) SYIEIVGPKPIYANIEMFEKSMKEYIQILDNTMTAR-----DQDGIIVFETHKMKGAAGSIGLKHISVAQKAQSPDM
 ArcB (AAC) ELVEMLGADVFVNNLILFEQTMQDYVAELQQAYQAYLNDPHTQPEVLSIVHKIKGALASVGLKRLQWIAAQAQNDT
 ArcB (PMU) ELVEMLGTDVFLKNLALFKETMPEYMQALMTAYSVYQQDESQKADVLSCAHKIKGAAASVGLKRISVAQKAQDPEQ
 . : : : * . : . : : * . * * : * . : : : * * * * . * : * * : * .

ArcB (ECO) PAWEDNVGEWIEEMKEEWRHDVEVLKAWVAKATKK--
 ArcB (STM) PAWEDNVVEWIEEMKQEWQHDVAVLKAWVANAEEK--
 ArcB (STY) PAWEDNVVEWIEEMKQEWQHDVAVLKAWVANAEEK--
 ArcB (YPK) PAWWDNVQDWVDELKLEWRNDVQVLREWAAEVEKK--
 ArcB (YEN) PAWWDNVQDWVDELKSEWRNDVQVLREWAAEAEKK--
 ArcB (PCA) PAWWDNVQEWIDELKHDWRHDVQVLRDWVAKAEKES--
 ArcB (VPA) PAWWENISDWVEEIKNEYQNDIQVLKSWLDQQRK----
 ArcB (VVU) PAWWENIADWVEEIKNEYQNDIQILKSWLEQR-----
 ArcB (VCH) PAWWENISDWVEEIKNEYQSDIALKRWLTQQANAKQ
 ArcB (VFI) PAWWENIDWVEEIKNEYEHDIQILKSWLEQQDK---
 ArcB (AAC) ADWQGNIAHWVNLAKEWQTDVAKLREWLAGY-----
 ArcB (PMU) GEWTENIHDWVQQLAQYWLLDVEALERYLNKSV----
 * * : . * : : : : * : * . :

Приложение 12.

Множественное выравнивание для белков NarL и NarP. Участки, соответствующие последовательности НТН, показаны белым шрифтом на темном фоне. Условные обозначения: “*” – абсолютно консервативная позиция; “:” – позиция с высокой степенью консервативности; “.” – позиция с низкой степенью консервативности. Обозначения геномов: см. “Условные обозначения и сокращения”.

```

NarP (YPK)  MLFELQGELEFIMTKSHTILIVDDHPLMRRGIKQLLEGLDSRFVVAEANNNGSDAITEAAKQPQDVIILLDLNMKGMSGL
NarP (YEN)  -----MTKSHTIMIVDDHPLMRRGIKQLLELDSNFDVVAEANCSDAIEAAKQPQDVIILLDLNMKGMSGL
NarP (ECO)  -----MPEATPFQVMIVDDHPLMRRGVRQLELDPGSEVVAEAGDGASAIIDLANRLDIDVILLDLNMKGMSGL
NarP (STY)  -----MPEVTPFQVIVDDHPLMRRGIRQLLELDPAFYVVAEAGDGASAIIDLANRIEPLDILLDLNMKGMSGL
NarP (PCA)  -----MSEKPSYRVLIVDDHPLMRRGIRQLLATDAIFDVIGEASNGVEALSLANRDSPLIILLDLNMKGMSGL
NarP (PMU)  -----MSEKMKVLIIDHPLMRRGIKQLELAPNFVEVADAGNGSEGISLALQTPDLIILLDLNMKGMSGL
NarP (AAC)  -----MTDKLVLIIDHPLMRRGIKQLELEEGFEVVGAGNGTEGINLALQTPDLIILLDLNMKGMSGL
NarP (HIN)  -----MQDKLVLLIDHPLMRRGIKQLELDDNFVADVSSGTEGISVALQTPDVIILLDLNMKGMSGL
NarP (HDU)  -----MTESYKILLIDHPLMRRGIRQILELEDNFTIVGEASNGNDGIALALDTPDLIILLDLNMKGMSGL
NarP (VPA)  -----MCKVMLVDDHPLMRRGIHQLESFEPEFEVVAEASNGADAVAKAHELELDLVLLDLNMKGMSGL
NarP (VVU)  -----MLVDDHPLMRRGINQLLSFEEQFEVVAEASNGTEAVAKAHENELDLIILLDLNMKGMSGL
NarP (VCH)  -----MLVDDHPLMRRGINQLLSFEEQFEVVAEASNGAEALAHLDLQPDILLDLNMKGMSGL
NarP (VFI)  -----MWNVIVDDHPLMRRGINQLLSFEEFTLTGEASNGTDAVALIQDPELDLVLLDLNMKGMSGL
NarL (ECO)  -----MSNQEPATILLIDHPLMRLTGKQLISMADPITVVEASNGEQGIELAESLDPDLIILLDLNMPGMNGL
NarL (STY)  -----MNNQEPATILLIDHPLMRLTGKQLVSMADPISVVEASNGEQGIDLAEPLDILLDLNMPGMNGL
NarL (PCA)  -----MINEDAATLLIDHPLMRLNGKQLISMDELQVAGEASHGEQGVELAEQLDPLIILLDLNMPGMNGL
          :::*:*:* * : * : : : . . . * . : : . * : : * * * * * * * * :

NarP (YPK)  DTLKALRHNGSDARI I I LTVSDARSVDYAMIDAGADGYLLKDCPEILLENIRQAARGENVFSDEVIQYLSR----
NarP (YEN)  DTLKALRNEGIDARI I V LTVSDARSVDYAMIDAGADGYLLKDCSEPEILLENIRLASKGENVFSDAVTQYLSR----
NarP (ECO)  DTLNALRRDGVTAQI I I LTVSDASDV FALIDAGADGYLLKDS DPEVLL EAIRGAKGSKVFSERVNQYLREREMF-
NarP (STY)  DTLNALRRDGVTAQI I I LTVSDASDIYALIDAGADGYLLKDS DPEVLL EAIRKGANGGKVFSDRVNEYLRERERF-
NarP (PCA)  DTLHALRRDGI CARV I V LTVSDAPSDIYALMDAGADGYLLKDS APEHLLDSIRS----SDAFSEQVRDALRHRIAI-
NarP (PMU)  DTLKALRAEGIDARI I I LTVSDLKSDIFTLIDAGADGYLLK DTEPDALLS QIQR I AQGEVILSDS I KDALIER----
NarP (AAC)  DTLKALRAEGVDARI I I LTVSDAKNDIYTLIDAGADGYLLK DTEPD TLLS QLKRIARGEVILSDS I KNLLLER----
NarP (HIN)  DTLKGLRAEGVDARI I I LTVSDAKNDIYTLIDAGADGYLLK DTEPD TLL E QIKRI AQGEVILSDS I KNLLLER----
NarP (HDU)  DTLRALRKQGVDSRI I V LTVASDEQSDIFASMDAGVDGYLLKDS DTS E LRENIRKAVLG ETVLSDS I RAHLLNR----
NarP (VPA)  DTLKALRADGCEAR I V I LTVSDSPADIEAIVRSGADGYLLK DTEPDEL VELLE KQAH QGDKAYSQEVAKYLSER----
NarP (VVU)  DTLKALRADGCEAS I V I LTVSDSPADIEAIVKAGADGYLLK DTEPDEL VELLE KAH LGDKAYSSV VAKYLLQR----
NarP (VCH)  DTLTA-RTDECDAYV I LTVSDSADIEALVKAGADGYLLK DTEPDQLVALLKQAMQGEKAFSESVERYLLNR----
NarP (VFI)  DTLHALRNEGVTCP I V I LTVSDNKQDIKTLIKAGADGYLLK DSEPDEL IALLKEAMKGGKAYSEQVKTCL EED----
NarL (ECO)  ETLDKLREKSLSGRI V V FSVSNHEEDVVTALKR GADGYLLK DMEPEDLLKALHQAAAGEMVLSEALTPVLAASLRAN
NarL (STY)  ETLDKLREKALSGRI V V FSVSNHEEDVVTALKR GADGYLLK DMEPEDLLKALQQAAGEMVLSEALTPVLAASLRAN
NarL (PCA)  ETLNRLREKSLSGRI V V FSVSNHEEDVVTALKR GADGYLLK DMEPEDLLAALHQAAAGEMVLSETLTPILAASLRAN
          : ** * . : : : : * : * : : * . * * * * * * . . * . : . * . : *

NarP (YPK)  -HEQVNPFS ELTERELDVLQEVARGMSNKQVAFELHI SEETVKVHIRNLLRKLNVRSRVAATIMYLENKKY----
NarP (YEN)  -DEQVNPFS ELTERELDVLQEVARGMSNKQVAFELHI SEETVKVHIRNLLRKLNVRSRVAATIMYLENKQY----
NarP (ECO)  -GAEEDPFSV LTERELDVLHEL AQGLSNKQIASV LNI SEQTVKVHIRNLLRKLNVRSRVAATILFLQQRGAQ----
NarP (STY)  -GAQEDPFS I LTERELDVLHEL AQGLSNKQIASV LNI SEQTVKVHIRNLLRKLNVRSRVAATILFLQTRGMQ----
NarP (PCA)  -KETPSPTV LTERELDVLQEVASGLSNKQIAAVLY I SEETVKVHIRNMLRKLNVRSRVAATVLYLETRGQ----
NarP (PMU)  -QMNADPIYS L TDREMDVLR LIATGLSNKQIAGQLF I SEETVKVHIRNLLRKLNVHSRVAATVLYFEHKGR----
NarP (AAC)  -QSAQEP IYS L TDREMGVLR LIATGLSNKQIAGQLF I SEETVKVHIRNLLRKLNVHSRVAATVLYFEHKN----
NarP (HIN)  -THEDNPLDSL TDREMGVLR QIATGLSNKQIAAQLF I SEETVKVHIRNLLRKLNVHSRVAATVLF FFEQNR----
NarP (HDU)  -RPENDPLSV L TDREMDVLR QIATGLSNKQIAGQLF I SEETVKVHIRNLLRKLNVHSRVAATILYLEQQR LKQVK
NarP (VPA)  -SDHEDVFD S L TDRETQILAR GFRNKQIADRLF I SESTVKVHMK SLLKKLQVPSRTAATVLYLERFGDIK--
NarP (VVU)  -EENTDVFDEL TERERQILQEVARGFRNKQIADRLF I SESTVKVHMK SLLKKLKVP SRTAATVLF LERYGELK--
NarP (VCH)  -EHTSNPF EAL TEREMQIMSEVAKGYRNQIADRLF I SESTVKVHMK SLLKKL DVP SRTAATILYLERY-----
NarP (VFI)  -AQGDDKLSQL TARELEILQH VAKGMRNKQVADKLF I SEATVKVHMK SLLKKL NVKSRVAATLLYLDS-----
NarL (ECO)  RATTDRDVNQLTPRERDILKLI AQGLPNKMIARRLDI TESTVKVHV K HMLKMKMLKSRVEAAVVMHQERIF----
NarL (STY)  RATTDRDVTQLTPRERDILKLI AQGLPNKMIARRLDI TESTVKVHV K HMLKMKMLKSRVEAAVVMHQERIF----
NarL (PCA)  RHSSDRDIQLLTPRERDILKLLA QGLSNKVIARKLTI TESTVKVHV K HLLKMKMLKSRVEAAVVMVLEQKVI----
          . ** * * : : * * * : : * * * * * * * : : * : : : * . * : :

```


Приложение 13.

Множественное выравнивание для белков NarX и NarQ. Условные обозначения: “*” – абсолютно консервативная позиция; “.” – позиция с высокой степенью консервативности; “:” – позиция с низкой степенью консервативности. Обозначения геномов: см. “Условные обозначения и сокращения”.

```

NarQ (ECO) -----MIVKRPVSASLARAFFYIVLLSILSTGIALLTASSLRDAEAINIAGSLRMQSYRLGYDLQSGSPQ
NarQ (STY) -----MTVKRPVSASLAKAFFYIVLLSILSTGSALLTLTSSLRDAEAINIAGSLRMQSYRLGYDLQSRSPQ
NarQ (PCA) -----MMVKRPVSTSLARAFFYIALLSFLTGTGIALLTASSLRDAEAINVAGSMRMQSYRMGYDLQGDRAA
NarQ (VPA) -----MFKNVKKSVTRTIASAMLLILLLSVATTGFAIFTLASSLNDAEAVNVAGSMRMQSYRLAHDIQIRSV
NarQ (VVU) -----MGLILLLSIATTSFAIINLASSLNDAEAVNVAGSMRMQSYRLAHDIQINAQD
NarQ (VCH) MVINQRETALFTVPRHSLTRTMAKAMLLILFLSALTGTVAIVTLASSLNDAEAVNVAGSMRMQSYRLAYDIQTQSHD
NarQ (VFI) -----MKHNKTSILTRTIAQVMLLIVMISITITSLALITLSSSLKDAEAVNIAGSLRMQSYRLAYDIETNSPE
NarQ (AAC) -MTRFLPRLLVILKIKHSVSTRIANYLIVIIIFVGIIGALSALMAGNKSAYAEAINVSGSLRMQSYRILYEIEHEPQM
NarQ (PMU) -----MKIKHSVSTKIANYLFIIMIFAGVITSLSLMIMASNKSDAESINVSGSLRMQSYRLLHELEHHP
NarQ (HIN) -----MYTKGSVSTRIAKYLFILIVAGVSSLSLAIMSNKYDAEAINISGSLRMQSYRLLYEMQEPES
NarQ (HDU) -----MYTKYSIARRIGFYFLIMIGFASLISGILSGIMWSNKS DAGLINVSGSLRMQSYRFLYEMEHHPNS
NarX (ECO) -----MLKRCLSPPLTLVNQVALIVLLSTAIGLAGMAVSGWLQVGVQSAHAINKAGSLRMQSYRLLAAVPLSEKD
NarX (STY) -----MFKRCFSPLTLVNQLALIVMLSTAIGVAGMAVSGWLQVGVQSAHAINKAGSLRMQSYRLLAAVPLDAKD
NarX (YPK) MVKGISHAMKRYLTP--LVNQIALLMILLGMLGLAGMVISSWMAQSIQNAHA INQAGSLRMQSYRLLSMIPLDSNE
NarX (YEN) -----MKRYLIP--LVNQIALLMILLGMLGLAGMTISSWMAQSIQNAHA INKAGSLRMQSYRLLSMVPLDKGD
NarX (PCA) -----MLKRFLPLSLVNQVALMLLLGLLGIAGMSVSSWMSQSIQNAHA INKAGSLRMQSYRLLSMVPLSAEN
      . . . : * : * : * : * : * : * : * :
      . . . : * : * : * : * : * : * : * :

NarQ (ECO) LNAHRQLFQQALHSP--VLTNLNVWVPEAVKTRYAHLNANWLEMN--RLSKGDLWPYQANINNYVNQIDLFVLAL
NarQ (STY) INAHRQLFQHALNSP--VLQNLNAWYVPQAVKTRYARLHANWLEMNS--RLQDGDIVWYQSNINNYVDQIDLFVLAL
NarQ (PCA) LEAHRHLYAQTLDSP--VFHLLDRWYVPRDVRDRYSALLQSWKTMDE--RLSAGDRVWYQDNIVRYVTQIDLFVLAL
NarQ (VPA) YSSHIDAFEHSIYSS--SMKALQHWSPEDITHDYRILMRWHELKS--VLRGEDPSQYQLLVAGFVQQIDDFVFKL
NarQ (VVU) YSHIYLFERSIYSP--SMKALQNWTVPDDITHDYRILMRWHELKE--VLQSKDRERYLVLVAGFVQSIDAFVFKL
NarQ (VCH) YKAHIFLFENSLYSP--SMLALLDWTVPSDIQDQDYQLIERWHELK--VLNSDQKAQYLDQVAPFVSLVDGFVLKL
NarQ (VFI) LIEHLDKFSQSIESP--SFKSLDQWFPDDIEDYEQWLWLSLQP--SLLSDNKQIYLNKVSLFVDEIDHFVRL
NarQ (AAC) VESSLRQYRVSLHSP--SLVEINYQSFASQDVKDSYHNLIKRWEVMER--LAREQNFTEYQKNIANVEQVDQFVFTL
NarQ (PMU) VETGLRQYRISLHSP--ALLEID--QWFPVDEVKQSYHQLIKRWERMER--YAKQREINAYKNEIADYVAQVDQFVFTL
NarQ (HIN) VETNLRRYHISLHSS--ALLEVQNFQTPNVLKHSYQNILQRWTNMEK--YARQDVKNYSKQLTDYVADVDYFVVEL
NarQ (HDU) IPSRLAEYRQSLNSPEIKESLEYKFLPTTEVTERYDNLKQSWQAMEG--FIINGDRDAYVANLEGYARQVNNFVWSL
NarX (ECO) KPLIKEMEQTAFSAE-----LTRAERDGLAQQLQGLQDYWRNELIPALMRAQNRRETVSADVSQFVAGLDQLVSGF
NarX (STY) QKLLDEMEQTAFSPE-----LTRAERDGLQKQKALQDYWHNELSPRLQHAQNAHAVAEDI TRFVAGLDRLVTSF
NarX (YPK) LSYLVALEQDKTSHD-----LQALLREGLMGQYQQIERYWQDTLKPQLQAKHPDEVAASVADFVHQLDALVLAI
NarX (YEN) LPYLAALEQDKTSD-----LQHALQREGLTRQYQQIERYWQNTLKPQLLQAKQPDVAANVADFVHQLDALVLAI
NarX (PCA) EIYLQEELEDEISSD-----LQAVRREGLSEQFTALRVFWDNLQPHLRQATHSADASADVARFVKQLDDLVSAI
      : * : . : : : * :
      : * : . : : : * :

NarQ (ECO) QHYAERKMLLVVAISLAGGIGIFTLVFFTLRRIRHQVAPLNQLVTASQRIEHGQFDSPPDLTNPNELGLLAKTFN
NarQ (STY) QHYAERKMLLVVAISLAGGIGIFTLVFFTLRRIRQVVRPLNQLVTASQRIEHGQFAPLPLDTSPLNELGLLAKTFS
NarQ (PCA) QHYSERKMMIVVVTIIGSITIIYILIFFTLRRIRQVVVPLNKLMAACASIEKGRFTHSRLDVNPNELGLLAQTF
NarQ (VPA) QNFSEQKLINLAWIGGLGGLGILCASFVVFHIRLEVVRPLRALVFASERIKNRSFD--INLAVSSDNEMGILTRTFN
NarQ (VVU) QHFSEQKLIKLAWVGGGLGGLGILLVSI FVVHFRVVTPLKAMVTASEQIQSRSFD--VSLNVTSDNEMGILTRTFN
NarQ (VCH) QRFSEKLI ALA AVGSFGLGGIFAVSLFVVIYIREQVVRPLNAMVDASEKIKNRNFD--VMLEETSTSTEMGILSRFTN
NarQ (VFI) QEFSEKQLQLLSLIGALGSLIVFSSIFIIFTQRKIVSPLHHLVAASHAMTKGNYS--VQVDLNSDNELGQLGNF
NarQ (AAC) QHSAEEKWIWALCVIIFSMLLIVIMVSYVNVYTRKAVVRPLEQLTRASTQVQMGQFKHIPLDVRKEDELGHLARTFT
NarQ (PMU) QRFAEKKWIIAVLIMSLSMLSIIAMVAVVWYARKEVVMPLKQLTQASIQVQMRQF SHVKVDTQKEDELGNLARAFS
NarQ (HIN) QRFSEQKWILGVSVLGFAMLLILLMVSYVIWYTNVPLHLMTKASMQVQMRQFNHIIPLDTRKQNELGTLARVFT
NarQ (HDU) QSF AELKCLKIAIGVIIVSMLLIIGLAYIGVWYTRKKIIAPLNQLVVASRQIKNEDFDHVRVLTPEPNELGFLSSTFT
NarX (ECO) DRTEMRIETVVLVHRVMAVFMALLLVFTI IWLRLRLLQPWRLQLLAMASAVSHRDFTRANISG--RNEMAMLGTALN
NarX (STY) DHTTELRIERVVLVHRVMAIFMALLLVFTI IWLRLRLLQPWRLQLLSMARAVSQRDFTRANISG--RNEMAALGSALN
NarX (YPK) DHKTEQRLLQVTLIPLVFLSLLTSLLLAMIYYLRRRLLRPWLKLI SMANAIGRGDFSQRYTLDYQRDEMGMLGTALN
NarX (YEN) DHKTEQRLLLVMTMIQLVFI VLTTLGLMLATIYYLRRRLLRPWLQLI SMANAIGRGDFSKRFSLPYQRDEMGMLGTALN
NarX (PCA) DHKTEQRLLMVTLVQRIFIGIMFILLIATFFYLRRRLLTPWRRLLVSMQAIGQGDFTRVAITG--HDEM SKLGQALN
      : : * : . : : * : . : : * : * : * :

```


NarQ (ECO) SEIAVSCVTAPDGNHTVYIRDNGIGIGEPKEPEGHYGLNIMRERAERLGGTLTFSQPSGGGTLVSI SFRS-AEGE
 NarQ (STY) SEIAVSCVTAPDGDHTVYIRDNGIGIGEPHEPAGHYGLNIMRERAERLGGTLNFSQPSGGGTLVSI SFRS-SNKE
 NarQ (PCA) TTIIVNCCTQPDGSHCV@DDGCGI INQEEPAGHYGLNIMQERAERLGGRLAIGLHPEGGTQVSVRFSTPTAR
 NarQ (VPA) DHIHIQCLDC-DGKVTVTVSDDGVGFQDEKINHYGMT IMQERAARLHADLQIEAS INKGCTVKLEFQHSKEVN
 NarQ (VVU) ANIRIECCEE-GEWVTVSVSDDGVGFQDCNQKLNHYGMS IMQERAARLNGSLSVTAS PQNGCAVVLYQRIKETR
 NarQ (VCH) TKIKVNCQEQ-QGRVQVTIEDDGVGFDLQDHKINHYGMS IMQERAARLRGELKVESAPGKGCKVVLTYQQSEKN
 NarQ (VFI) DNI I I TCFQE-GEQGCINIEDDGVGFDPTEEKINHYGLRIMQERANCVHGKLSITSEIDNGCTVNVMFPLNQ---
 NarQ (AAC) TLIEVIAHTNEDGEYELVVRDNGIGIPSLEEDPGHYGLNIMHERSIQLNAKLTIANRTDGGTEVKVTLAHTL---
 NarQ (PMU) NRIEVIAHINEDGEYEFIVRDDGIGIPSLEEEPEGHYGLNIMHERATQLNAHLRIANRPSGGTEIKLTLHQNRNLT
 NarQ (HIN) TVIEI SARINAEGEYEILVEDDGVGIPNLEEEPEGHYGLNIMAERCQLNAQLHIHRREQGGTQVKITLPHTLY--
 NarQ (HDU) TQIEVIAETNQDGEQCLIVRDNGNGIVNGIEPDDHYGLT IMKERAAQLKGEFTI INRAEGGVQVMVTLPLNTL---
 NarX (ECO) SEVVVTVAQN-DNQVKLTVQDNGCGVPENAI RSNHYGMI IMRDRAQSLRGDCRVRRESGGTEVVVTFIPEKTFT
 NarX (STY) DDVVVTVTQC-GKQVKLVQDNGCGVPENAERSNHYGMI IMRDRAQSLRGDCQVRRRETGGTEVTVTFIPETNFT
 NarX (YPK) SEATVNVHLE-QGDVVMSITDNGCGIGQASEQPNHYGLMIMQDRANSLHGHC SINQKEGGGTEVVRVQFTPEN---
 NarX (YEN) TEAAVKVTAE-QGDVVMSILDNGCGIGQASKRPNHYGLMIMQDRATSLHGHC@QREEGGTEVVRVFTPDN---
 NarX (PCA) TQVDITLQLR-QGYIELSVTDNGIGIPDDASRANHYGLI IMRDRARGLHGECIVRRRPSGGTEVNVVSFLSEYRHR

: . : * : * * . . * * : * * : * . : . . * : :

NarQ (ECO) ESQLM-----
 NarQ (STY) ESQLT-----
 NarQ (PCA) ERDNTNNLYPE
 NarQ (VPA) FDSV-----
 NarQ (VVU) I-----
 NarQ (VCH) KNEL-----
 NarQ (VFI) -----
 NarQ (AAC) -----
 NarQ (PMU) -----
 NarQ (HIN) -----
 NarQ (HDU) -----
 NarX (ECO) DVQG-DTHE--
 NarX (STY) ETQG-DTHE--
 NarX (YPK) -----NDLD--
 NarX (YEN) -----DDL D--
 NarX (PCA) LPLTGETHD--

Приложение 14.

Последовательности потенциальных сайтов связывания белка Fnr. Указано положение сайта относительно старта трансляции гена. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>Y. pestis</i>	<i>adhE</i>	-400	4,17	TTttTgTtAATCAA
	<i>arcA</i>	-299	4,59	TTGATaTAtgTCAA
	<i>ccmABCDEFGH</i>	-303	3,84	TgGATTTttAaCAA
	<i>cydAB</i>	-313	4,84	TTGtTcTAAATCAA
		-262	4,48	TTGtcgTAGATCAA
	<i>dmsABC</i>	-200	3,97	TTGATTccAgaCAA
	<i>fadBA</i>	-144	4,27	aTGATaaAgATCAA
	<i>fadD</i>	-51	4,52	TTGtTTaAAAaCAA
	<i>fadL</i>	-166	4,07	TTttTTgAgATCAA
	<i>fnr</i>	-33	4,24	TTGAcgcAtATCAA
	<i>focA</i>	-81	3,94	aTGATccAtATCAA
	<i>fumC</i>	-482	4,11	TgGATcTtgATCAA
	<i>glgBXCAP</i>	-26	3,80	aTGCtATtAATCAA
	<i>glpABC</i>	-146	4,23	TTGATcacAATCAat
		-92	4,20	TTGtTTcttATCAA
		-88	3,77	TTctTaTcAATCAA
	<i>glpD</i>	-215	4,10	aTGATgTAAATCgA
		20	3,84	TTGATTacAATCgt
	<i>moaACDE</i>	-309	3,99	gTGATTTtAAaCAA
	<i>mtlADR</i>	-218	4,19	gTGATcTtAATCAA
	<i>napFDABC</i>	-212	3,93	TTGATcTtgcgCAA
		-120	3,85	TTGATTTcAATaAt
	<i>ndh</i>	-149	4,90	TTGATaTAtATCAA
	<i>nirBDC-cysG</i>	-70	5,15	TTGATTTAcATCAA
	<i>nqrABCDEF</i>	17	3,92	aTGATTaAAATaAA
	<i>nrdDG</i>	-191	4,45	TTGtTcTAGgTCAA
	<i>pdhR-aceEF-lpdA</i>	-64	4,04	aTGATTTcggTCAA
<i>pepT</i>	-24	3,99	TTGATcatgATCAc	
<i>ppsA</i>	-4	4,16	TcGATgTccAaCAA	
<i>talB</i>	-419	3,76	TTGAcTgAcgTCag	
	-125	4,22	TTttTTTTtAATCAA	
<i>yfiD</i>	-112	4,19	TTGATaTAAAaCAat	
	-59	4,75	TTGATTTgAATCAA	
<i>Y. enterocolitica</i>	<i>arcA</i>	-471	4,80	TTGATTTtAAaCAA
		-297	4,59	TTGATaTAtgTCAA
	<i>atpIBEFHAGDC</i>	-499	4,06	TTGAagatgATCAA
	<i>cydAB</i>	-377	3,90	TTtATTaAcAgCAA
		-313	4,84	TTGtTcTAAATCAA
		-262	4,31	TTGtTcTggATCAA
<i>dcuA</i>	-25	3,86	TTcATTTggATCAA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>Y. enterocolitica</i>	<i>dcuB</i>	-286	4,34	TTtATTTAAAaCAA
	<i>dcuC</i>	-172	3,82	TTGATagAtgTCAt
	<i>fadBA</i>	-144	4,35	aTGATaaAAATCAA
	<i>fadL</i>	-162	3,96	TTttTTgcAATCAA
	<i>feoAB</i>	-318	4,17	TTGATTcAAATCAt
	<i>fnr</i>	-33	4,24	TTGAcgcAtATCAA
	<i>focA</i>	-78	4,04	aTGATccAcATCAA
	<i>frdABCD</i>	-271	3,86	TTGAggTAgtTCAA
	<i>glgBXCAP</i>	-113	4,03	TTGAccTtgATCAt
	<i>glpABC</i>	-146	4,23	TTGATcacAATCAt
		-92	4,12	TTGtTccttATCAA
	<i>glpD</i>	-453	3,79	TTGAgcacAATCAg
		-66	3,84	TTGATTacAATCgt
	<i>glpP</i>	-256	3,79	TTttTTaAtAaCAA
	<i>mdh</i>	-255	3,79	TTGATgTAAAggAA
	<i>moaA</i>	-351	3,99	gTGATTTtAAaCAA
	<i>napFDABC</i>	-210	4,48	TTGATcTtgcTCAA
	<i>ndh</i>	-150	4,90	TTGATaTAtATCAA
	<i>nirBDC-cysG</i>	-72	5,15	TTGATTTAcATCAA
	<i>nqrABCDEF</i>	1	3,77	aTGATaaAAATaAA
	<i>nrdDG</i>	-196	4,45	TTGtTcTAggTCAA
	<i>pdhR-aceEF-lpdA</i>	-397	3,84	TTtATAatggTCAA
		-148	4,11	aTGATTTtgTCAA
	<i>pepT</i>	-400	3,94	TTGATgaAAAaCAc
-53		3,95	TTGATggtgATCAc	
<i>ppsA</i>	-2	4,16	TcGATgTccAaCAA	
<i>sfcA</i>	-466	3,98	TTtATTTtcATCgA	
<i>yfiD</i>	-180	4,19	TTGATaTAAAaCAt	
	-127	4,89	TTGATTTAagTCAA	
<i>P. multocida</i>	<i>acpP</i>	-95	3,86	TTGAccaccgaCAA
	<i>adhE</i>	-115	4,19	TTGATcaAAATCAc
	<i>arcA</i>	-152	3,81	aTGATcTAtATCAt
	<i>atpB</i>	-28	3,97	TTcATaaAtATCAA
	<i>cctABCDEFGH</i>	-127	4,84	TTGtTcTAAATCAA
	<i>cpxRA</i>	-394	3,92	TTGATgTcccgCAA
		-172	4,15	TTGAccTAgATaAA
		-32	3,99	TTctTTTtAATCAA
	<i>cydAB</i>	-305	4,21	TTGAccTgtATCAA
	<i>dcuB</i>	-131	4,43	TTGATTTAAAcCAA
	<i>dmsABC</i>	-240	4,30	TTGATTacAgaCAA
		-51	4,47	TTGATTTAtATCAt
		-1	3,95	TTGATcTAtcTCAt
	<i>eno</i>	-108	4,34	TaGATgTAAATCAA
<i>fdhD</i>	-204	3,80	TTGATagtcgTaAA	

Геном	Оперон	Сайт			
		Положение	Вес	Последовательность	
<i>P. multocida</i>	<i>fdxGHI</i>	-52	3,80	TTtAcgactATCAA	
	<i>fnr</i>	9	4,53	TTGAgaTAgATCAA	
	<i>focA</i>	-212	3,94	TTGtTgaAAATtAA	
	<i>frdABCD</i>	-133	3,82	TTGAggTAgATCAc	
	<i>fur</i>	-118	3,90	TTGATgaAgATCgt	
	<i>fur</i>	-31	3,83	TTGcTTaAtAgCAA	
	<i>glpFK</i>	-271	4,04	cTGATgTggATCAA	
	<i>gntXY</i>	-24	3,83	TTtATgTcAAgCAA	
	<i>malQ-glgBXCAP</i>		-469	4,10	TTGATggAcATtAA
			-97	3,83	TTGAcgaAgATCAc
	<i>nagC</i>	-246	3,85	TTGcgTTAcgTCAA	
	<i>napFDAGHBC</i>	-120	4,32	TTGATcTAGcaCAA	
	<i>narQ</i>	-89	4,32	TTGtgcTAgATCAA	
	<i>nqrABCDEFGF</i>	-332	3,87	TTGAccTtctTCAA	
	<i>nrdAB</i>		11	4,02	TTGATggtcAcCAA
			48	4,01	TTGATcTggATaAA
	<i>nrdDG</i>	-217	4,23	TTGATcTAGATCAc	
	<i>nrjABCDEFXG</i>	-142	3,93	TTGATcaAgcgCAA	
	<i>pepT</i>		-475	4,15	TTGtTTTAccaCAA
			-74	3,87	TTGtgaTAtgTCAA
	<i>sodA</i>	-84	4,42	aTGATcTtAATCAA	
	<i>torCAD</i>		-246	3,96	TTGATccttcTCAA
			-200	3,83	TTtcTTatAATCAA
<i>yfiD</i>		-138	5,12	TTGATcTAAATCAA	
		-96	4,45	TTGtTcTAGgTCAA	
<i>A. actinomycetemcomitans</i>	<i>ccmABCDEFGH</i>	-126	4,84	TTGtTcTAAATCAA	
	<i>cpxRA</i>		-368	3,93	TTGATgTgccaCAA
			-230	3,98	TTtAcTTAcAaCAA
	<i>cydAB</i>		-282	4,37	TTGATcTtgAgCAA
			-237	4,24	TTGATTTAtATCtA
	<i>dcuA</i>	-420	3,95	TaGATcaccATCAA	
	<i>dcuB</i>	-85	4,06	TTGATTTAtcgCAA	
	<i>dmsABC</i>		-261	3,72	TTaATTgtgAaCAA
			-118	4,42	TTGATTTtgATCat
	<i>eno</i>		-405	3,93	TTGATTgcgATCcA
			-118	3,86	TgGATccAAATCAA
	<i>fdhD</i>	-348	4,01	TTGATagttATaAA	
	<i>fdxGHI</i>	-44	4,01	TTtATAactATCAA	
	<i>fnr</i>	-59	4,60	TTGAgcTAgATCAA	
	<i>focA</i>	-183	4,26	TTGAcAaAgAaCAA	
	<i>frdABCD</i>	-262	3,87	TTGAgtTAgATCAc	
	<i>fur</i>	-117	3,67	TTGATgaAgATCgc	
	<i>glpD</i>	34	3,66	caGATTTAtATCAA	
	<i>gntXY</i>	-136	4,07	TTtATTTAAAgCAA	
	<i>moaACDE</i>	-203	3,65	aTaATcTAGATCAA	
	<i>ndh</i>	-95	3,74	TTttTAcAAATCAA	
	<i>nqrABCDEFGF</i>	-316	3,68	TTGATTTtgTCAA	
	<i>nrdAB</i>	14	4,02	TTGATggtcAcCAA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>A. actinomycetemcomitans</i>	<i>nrfABCDEFXG</i>	-258	4,72	TTGATcaAAAaCAA
	<i>sodA</i>	-74	3,98	aTGATTccAAATCAA
	<i>torYZ</i>	-110	4,70	TTGAcTgAAATCAA
	<i>yfiD</i>	-117	4,34	TTGAcgTgcATCAA
<i>H. influenzae</i>	<i>aspA</i>	-66	4,14	gTGATcTtCATCAA
	<i>ccmABCDEFGH</i>	-167	4,54	TTGATTTAgATCAat
		-113	4,92	TTGtTTTAAATCAA
	<i>cpxRA</i>	-381	4,11	TTGATgTAccgCAA
	<i>cydAB</i>	-293	4,81	TTGATcTAAgTCAA
	<i>dmsABC</i>	-87	4,23	TTGATTTggcTCAA
	<i>dusB-fis</i>	-145	4,10	TTttTgTcAAATCAA
	<i>eno</i>	-267	3,75	TTctTcTAAAaCAA
		-118	4,04	TaGATaaAgATCAA
		-14	4,01	TTGAggaAAATaAA
	<i>fnr</i>	-49	4,13	TTGcgTTAgATCAA
	<i>focA</i>	-320	4,09	TTGtgaatAAATCAA
	<i>frdABCD</i>	-142	3,82	TTGAggTAgATCAc
	<i>glgB</i>	-76	3,80	TTtATTTttAgCAA
	<i>glpABC</i>	-447	4,09	TTGtgcTAAgTCAA
	<i>gntXY</i>	-267	3,79	TTGATaaAcAagAA
	<i>malQ-glgBXCAP</i>	-86	3,83	TTGAcgaAgATCAc
	<i>mdh</i>	-162	3,91	TTcAcTTtCATCAA
	<i>moaACDE</i>	-343	4,62	aTGATTTAAATCAA
		-221	3,84	TTGATTTttAagAA
		-195	4,45	cTGATTTtCATCAA
	<i>napFDAGHBC</i>	-125	4,69	TTGATcTATaCAA
	<i>nqrABCDEF</i>	-267	3,91	TTGATaTATcgCAA
	<i>nrdAB</i>	-136	3,89	TTGtcaagAAATCAA
	<i>nrfABCDEFXG</i>	-207	4,34	TTtATTTAAAaCAA
		-157	4,48	TTGATcaAgcTCAA
	<i>pckA</i>	65	3,80	TTGtTTacAAATCcA
	<i>pepT</i>	-72	4,62	TTGtTaTATATCAA
	<i>pgk</i>	-92	4,07	TTGATcTAAATCgt
		-149	4,92	TTGATTTAAAaCAA
	<i>sodA</i>	-95	4,54	aTGATcTAAATCAA
-230		3,87	TTcATTatAAaCAA	
<i>yfiD</i>	-143	4,31	TTaATTTAgATCAA	
<i>H. ducreyi</i>	<i>ccmABCDEFGH</i>	-135	4,41	TTGATTcctATCAA
	<i>cydAB</i>	-320	5,03	TTGATgaAAATCAA
	<i>dcuB</i>	-310	3,76	TTGATcaAAAcCgA
	<i>dusB-fis</i>	-495	3,96	TTaATTTtAgTCAA
	<i>focA</i>	-100	4,71	TTGATTTAccTCAA
	<i>frdABCD</i>	-125	3,82	TTGAggTAgATCAc
	<i>gntXY</i>	-172	4,04	TTcATcaAtATCAA
	<i>hlyX</i>	-228	3,83	TTGAgcTttATaAA
		-196	4,15	TTttTTTcAAATCAA
		-44	4,50	TTGcTaTAAATCAA
<i>moaACDE</i>	-210	3,92	TTtATTTtAAATCAat	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>H. ducreyi</i>	<i>napFDAGHBC</i>	-119	4,48	TTGATTcttATCAA
	<i>nqrABCDEF</i>	-265	4,28	TTGATTTAttTCAA
	<i>nrfABCDEF</i>	-332	3,91	TTtATTTttcTCAA
		-188	4,28	TTGATTTAttTCAA
		-184	3,85	TTtATTTcAATaAA
	<i>pckA</i>	-262	3,94	TTtcTcTAcATCAA
	<i>pflB</i>	11	4,04	TcGATgaAgcTCAA
	<i>sodA</i>	-92	4,24	TTGATTTAtATtAA
<i>torYZ</i>	-228	3,92	aTGATTaAAATaAA	
<i>V. vulnificus</i>	<i>acpP-fabF</i>	-144	4,23	aTGATTTAggTCAA
	<i>adhE</i>	-445	4,16	TTGATaaAAAaCAA
		-176	3,78	cTGATccAAgTCAA
	<i>aldB</i>	-468	3,76	TcGgTcaAAATCAA
		-111	3,94	TTaATTaAcAaCAA
	<i>arcA</i>	-327	5,05	TTGATaTAAATCAA
	<i>atpIBEFHAGDC</i>	-232	4,81	TTGATTTAggTCAA
		-7	3,98	TTGtTaatgATCgA
	<i>cydAB</i>	-280	5,12	TTGATTTAgATCAA
		-230	4,75	TTGATTcAAATCAA
	<i>dcuC</i>	-232	4,23	TTGATgcAAATCgA
		-212	4,20	TTGATaTttATCat
		-107	3,99	TTGcTTTAgATCat
	<i>fadIJ</i>	-400	4,35	TTGATcTAAATCAA
		-237	4,00	TTGtTTTAtAcCAA
	<i>feoAB</i>	-91	4,42	TTGATTTcAgaCAA
	<i>fnr</i>	-29	4,69	TTGAcgTAtATCAA
	<i>glpABC</i>	-491	3,85	TTtATTTAAATCAA
		-360	3,91	TTGtTTTAtAaaaAA
	<i>mdh</i>	-100	4,89	TTGATTgtAATCAA
	<i>mtlADR</i>	-138	3,84	TTGATTgAgAaCtA
	<i>nagB</i>	-93	4,53	TTGAccTAAAaCAA
	<i>napFDABC</i>	-155	4,84	TTGATcTAAAaCAA
	<i>napG</i>	-128	4,68	TTGATcTAAcTCAA
	<i>narQP</i>	-158	4,84	TTGtTTTAgATCAA
	<i>ndh</i>	-225	4,78	TTGATaaAtATCAA
	<i>nirBDC-cysG</i>	-405	3,78	aTGAccTgAATCAA
		-209	3,90	TTGcTTTAAAaCgA
		-76	4,57	TTGATTTAcATCat
	<i>nqrABCDEF</i>	-55	3,73	aTGATTacAATaAA
	<i>nrfA</i>	-141	4,57	TTGATcTAAAgCAA
	<i>nrfBCDEF</i>	-354	4,57	TTGcTTTAgATCAA
	<i>pdhR-aceEF-lpdA</i>	-112	4,23	aTGATTTAggTCAA
	<i>pepT</i>	41	3,77	TTGgTcTcggTCAA
	<i>pgk</i>	-393	4,26	aTGAcgTAAATCAA
	<i>sfcA</i>	-88	5,00	TTGATgTAtATCAA
	<i>torD</i>	-154	4,47	TTGATTcAAAaCAA
	<i>yfiD</i>	-164	4,76	TTGATTTAAcTCAA
		-111	4,81	TTGAccTAAATCAA

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. parahaemolyticus</i>	<i>acpP</i>	-143	4,31	aTGATTTAAgTCAA
	<i>adhE</i>	-160	3,78	cTGATccAAgTCAA
	<i>aldB</i>	-107	3,94	TTaATTaAcAaCAA
	<i>arcA</i>	-322	5,15	TTGATgTAAATCAA
	<i>atpIBEFHAGDC</i>	-180	4,81	TTGATTTAggTCAA
	<i>cydAB</i>	-250	4,59	TTGATcTggATCAA
		-200	5,00	TTGATgTAtATCAA
	<i>dcuC</i>	-402	4,04	TTGAgcgtcAaCAA
		-116	4,26	TTGtTTTAgATCAt
	<i>dmsABC</i>	-443	4,00	aTGATTgAtAaCAA
		-273	4,35	TTGATagtggTCAA
		-53	3,84	TTGtTaTtAATtAA
	<i>fadBA</i>	-148	3,88	TTGAaaTtAAaCAA
	<i>fadL</i>	-389	4,20	TTGATcTAttTCAA
	<i>feoAB</i>	-103	4,29	TTGATTTcAcaCAA
	<i>fnr</i>	-29	5,00	TTGATgTAtATCAA
	<i>focA</i>	-166	4,65	TTGtTTaAtATCAA
	<i>focA</i>	-127	4,02	TTGAgTTttATCgA
	<i>gltA</i>	26	3,78	aTgTgAggTCAA
	<i>mdh</i>	-100	4,89	TTGATTgtAATCAA
	<i>mlADR</i>	-185	3,85	TTGATcgttATCAc
		-136	3,87	TTGATcacgAaaAA
		42	3,80	TcGtTTTctgTCAA
	<i>nagAC</i>	-377	3,95	TTGtTTgtgATCAt
	<i>nagB</i>	-316	4,49	TTGAgTgAgATCAA
		-92	4,53	TTGAccTAAAaCAA
	<i>napFDABC</i>	-109	4,44	TTGAcTTgAATCAA
	<i>napG</i>	-152	4,39	TTGATccAAAaCAA
	<i>narQP</i>	-109	4,39	TTGtTTTggATCAA
	<i>ndh</i>	-262	4,78	TTGATaaAtATCAA
	<i>nirBDC-cysG</i>	-78	4,57	TTGATTTAcATCAt
	<i>nqrABCDEFG</i>	1	3,73	aTGATTacAATaAA
	<i>nrdDG</i>	-135	5,00	TTGATcaAAATCAA
	<i>nrfA</i>	-138	4,57	TTGATcTAAAgCAA
	<i>nrfBCDEF</i>	-350	4,57	TTGcTTTAgATCAA
	<i>oxyR</i>	-33	4,60	TTGAcTatcATCAA
	<i>pdhR-aceEF-lpdA</i>	-408	4,05	TTGATTTgtAgCAA
		-316	3,84	TTGtTaaAAATtAA
		-110	4,23	aTGATTTAggTCAA
	<i>pepT</i>	-496	4,04	TTGtcTatcAaCAA
	<i>pgk</i>	-406	4,26	aTGAcgTAAATCAA
	<i>sdhCDAB</i>	-498	3,78	TTtAcgaAcgTCAA
		-435	3,78	TTGAccTtcAaCAt
	<i>sfcA</i>	-86	4,34	TTGATgTgcgTCAA
	<i>talB</i>	-397	4,02	aTgTtTcTccATCAA
		-7	3,86	TcGATTatgAgCAA
	<i>torD</i>	-162	4,47	TTGATTcAAAaCAA
<i>yfiD</i>	-163	4,76	TTGATTTAAcTCAA	
	-110	4,89	TTGAcTTAAATCAA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. cholerae</i>	<i>acpP</i>	-48	4,47	aTGATTTAtATCAA
	<i>adhE</i>	-160	3,78	cTGATcTgAgTCAA
	<i>aldB</i>	-400	4,24	TTGtTgTcAcTCAA
	<i>arcA</i>	-263	5,15	TTGATgTAAATCAA
	<i>atpIBEFHAGDC</i>	-179	5,20	TTGATTTAAATCAA
		46	3,76	TTGtTaTtgATCca
	<i>cydAB</i>	-302	3,97	TTGATTgtttTCAA
		-236	4,54	TTGATTTAgATCAat
		-186	5,12	TTGATTTAgATCAA
		-6	3,91	TTGtTcTAtgTCgA
	<i>dcuB</i>	-241	4,76	TTGAccTAcATCAA
	<i>dcuC</i>	-101	3,99	TTGcTTTAgATCAat
	<i>fadIJ</i>	-207	4,35	TTGATcTAAATCAA
	<i>fadL</i>	-378	4,11	TTGATcTtgATgAA
	<i>fnr</i>	-29	4,79	TTGAcgTAcATCAA
	<i>fsfcA</i>	-311	3,80	TTGcTaTAtAgCAA
	<i>fumC</i>	-118	4,44	TTGATggAgcTCAA
	<i>glgX</i>	-62	4,22	TTGAccTAAgCAA
	<i>glpABC</i>	-372	3,89	TTttTgTtAAaCAA
	<i>glpD</i>	-111	4,34	aTGATcaAgATCAA
	<i>gltA</i>	-77	3,88	TTGAcTTtgATtAA
		-71	3,95	TTGATTaAgATCgg
	<i>ldhA</i>	-316	3,87	TTGATagAAcgCAA
	<i>mdh</i>	27	4,31	TTGATTgtAATCAg
	<i>mlADR</i>	-350	3,97	gTGATaaAtATCAA
	<i>nagB</i>	-140	4,53	TTGAccTAAAaCAA
	<i>napFDABC</i>	-107	4,39	TTGATccAAAaCAA
	<i>narQP</i>	-64	4,39	TTGtTTTggATCAA
	<i>ndh</i>	-198	4,78	TTGATaaAtATCAA
	<i>nqrABCDEFGF</i>	49	3,73	aTGATTacAATaAA
	<i>pdhR-aceEF-lpdA</i>	-119	4,23	aTGATTTAggTCAA
	<i>pepT</i>	-120	3,74	cTGATTTgtAaCAA
		-7	3,74	TTGtTTatgAaaAA
	<i>pflB</i>	-293	3,89	TTGtccTtcATCgA
	<i>pgk</i>	-376	4,26	aTGAcgTAAATCAA
	<i>sdhCDAB</i>	-332	3,95	ccGATcTtAATCAA
-326		3,88	TTaATcaAAgTCAA	
<i>torD</i>	-131	4,15	TTGATgcAAAgCAA	
<i>yfiD</i>	-164	4,29	TTGAcacAAATCAA	
	-111	4,81	TTGAccTAAATCAA	
<i>V. fischeri</i>	<i>adhE</i>	-363	4,09	cTGATcTgAATCAA
		47	3,98	TTcATcTAcAaCAA
	<i>aspA</i>	-96	3,80	acGATaTtgATCAA
	<i>atpIBEFHAGDC</i>	-40	5,20	TTGATTTAAATCAA
	<i>ccmABCDEFGH</i>	-72	3,84	TTGtTTaAtATCtA
	<i>dmsABC</i>	-307	3,81	TTGATTTAAATCtt
-271		4,13	TTGtTcTcgTCAA	
-137		4,72	TTGATcaAAAaCAA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. fischeri</i>	<i>fadBA</i>	-141	3,86	TTGtgaTcAAaCAA
	<i>fadIJ</i>	-417	4,20	TTGATcTAttTCAA
	<i>frdABCD</i>	-225	3,76	TTGAcggAgATCAc
		-164	3,94	TTtATccAtATCAA
	<i>glgX</i>	63	3,85	TTGcTgTtcATCAc
	<i>glpABC</i>	-81	4,13	TTGcgTTagATCAA
	<i>glpD</i>	-393	4,10	TTGATTatcATgAA
		-145	4,42	TTGATTaAgATCAg
	<i>nagB</i>	-96	3,98	TTGAgcaAAATCAc
	<i>napFDABC</i>	-114	4,84	TTGATcTAAAaCAA
	<i>narQP</i>	-87	4,84	TTGtTTTAgATCAA
	<i>nirBDC-cysG</i>	-71	4,04	TTGATccAcATCAc
	<i>nqrABCDEFG</i>	-231	4,72	TTGATTTtgAaCAA
	<i>nrfA</i>	-140	5,15	TTGATgTAAATCAA
	<i>nrfBCDEF</i>	-343	5,15	TTGATTTAcATCAA
	<i>pckA</i>	33	4,21	TcGATcTAAcTCAA
	<i>pdhR-aceEF-lpdA</i>	-145	4,17	aTGATTcAAATCAA
	<i>pgk</i>	-403	4,16	aTGAcAATAATCAA
	<i>sfcA</i>	-113	4,00	TTGATTgAtAaaa
	<i>torD</i>	-140	4,47	TTGATTcAAAaCAA
<i>yfiD</i>	-152	4,53	TTGATTTAggaCAA	
	-99	4,59	TTGAcATAtATCAA	

Приложение 15.

Последовательности потенциальных сайтов связывания белка ArcA. Указано положение сайта относительно старта трансляции гена а также цепь ДНК, на которой он расположен: «→» - прямая, «←» - обратная. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>Y. pestis</i>	<i>glpFK</i>	→	-87	4,22	TAACgаТаАаTTtAC
	<i>dusB-fis</i>	→	-331	4,24	TAACATTTcagTAtC
	<i>fadBA</i>	→	-79	4,57	TtgCATTTcaTTAAC
	<i>gltP</i>	→	-377	4,06	TAACAaaataaaaAAC
	<i>dcuA</i>	→	-48	4,14	acACAaTTttTTAAC
		→	-17	4,04	aACAaggAGTaAAC
		←	-8	4,34	GTaAACactATGaTA
	<i>frdABCD</i>	→	-180	4,65	TAtCATcaAGTTAAC
	<i>mdh</i>	→	-73	4,22	TcACAaTaAaTTAAg
		←	-94	4,22	GTTAAAtTtAtTGTgt
	<i>pdhR-aceEF-lpdA</i>	←	-32	4,85	GTgAAacttTGTTA
	<i>dcuB</i>	→	-343	4,30	TAAaAaTgAaTTAAAt
	<i>nqrABCDEFG</i>	←	-443	4,13	GcTAAAtTctAcGTTA
	<i>ydiF</i>	←	-420	4,11	GTTAAaTAtATcTTt
		←	-68	4,01	GTgAAaaAAtTGaTt
	<i>narP</i>	←	-474	4,25	GaTAAAtTAttTtTTA
	<i>napFDABC</i>	→	-340	4,25	TAAaAaTAaTTAtC
	<i>fadL</i>	→	-28	4,04	TAACATagttTaAAAt
		←	-191	4,24	GTgAgtTtttTGTTA
	<i>fadIJ</i>	→	-91	4,04	aAACgTTTgGTTAAC
	<i>acpP-fabF</i>	→	-62	4,27	cAACATTTtaTAcAC
	<i>ndh</i>	→	-31	4,94	TAACAaTTgGTTcAC
	<i>adhE</i>	←	-395	4,21	GTTAAAtcAAtaGTaA
	<i>dadAX</i>	←	-118	4,18	GTaAAAtAAAtTGagA
		←	-55	4,46	GTTAAAtTttcTGTcA
	<i>fadD</i>	→	-22	4,19	TAAaATaTtGTTAAAt
		←	-24	5,14	GTTAAaatAtTGTTA
	<i>focA</i>	→	-410	4,09	TAAaAaTaAaTAcAC
		←	-286	4,34	GTaAAAtaAAAtaGTTA
	<i>moaACDE</i>	←	-104	4,15	GTaAAaTAtAaGTTA
	<i>cydAB</i>	←	-474	4,83	GTTAAaaaAgtTGTTA
		←	-300	4,40	GTTtAtTcttTGTcA
	<i>sdhCDAB</i>	→	-424	4,40	TgACAaagAaTaAAC
→		-250	4,83	TAACAacTttTTAAC	
<i>gltA</i>	←	-179	4,20	GTTAAaaaAAATGTgg	
<i>arcA</i>	←	-12	4,22	GTTAcCacAtTGTgA	
<i>glgBXCAP</i>	→	-84	5,50	TAtCATTTAaTTAAC	
<i>fdoH</i>	→	-325	4,02	TcACATTTtaaaaAAC	
<i>fdhD</i>	→	-88	4,28	tTTAAaTAAAtTGTTc	

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>Y. enterocolitica</i>	<i>glpFK</i>	→	-86	4,22	ТААСGaTaAaTTtAC
	<i>glpABC</i>	←	-187	4,41	GTTAACattttTGaaA
	<i>fadBA</i>	→	-79	4,51	TtgCAaTTcaTTAAC
		→	-68	4,10	ТААСAaaataTccAC
	<i>gltP</i>	→	-378	4,06	ТААСAaaatacaAAC
		→	-248	4,05	ТААСAaaTAtTTAct
	<i>dcuA</i>	←	-8	4,34	GTaAACactATGaTA
	<i>aspA</i>	→	-51	4,22	cAtaATTTAGTTAAC
		→	-40	4,08	ТААСAaaTAcAAAC
	<i>frdABCD</i>	→	-146	4,02	TacCAaTcAGTTAgC
	<i>mdh</i>	→	-226	4,10	TtActaTTAaTTAAt
	<i>deoCABD</i>	←	-116	4,85	GTgAAactttTGTTA
	<i>pdhR-aceEF-lpdA</i>	→	-243	4,09	TtAtATTctaTTAAC
		→	-6	4,12	ТААtATTatGaTAAC
		←	-8	4,12	GgTAAatattATGaTA
	<i>dcuB</i>	→	-24	4,33	ТААСATTaAtTaAAa
		←	-187	4,24	GTgAgtTtttTGTTA
	<i>fadL</i>	→	-89	4,04	aAAAtTTTgGTTAAC
		←	-102	4,62	GTTtACcAAATGTaA
	<i>focA</i>	→	-145	4,30	aAtCAaaaAGTTAAC
	<i>acpP-fabF</i>	→	-62	4,27	cAACATTTtaTAcAC
	<i>ndh</i>	→	-31	4,94	ТААСAaTTgGTTcAC
		←	-366	4,04	GTTAAagActgGTTA
		←	-211	4,22	GcTAACTAttTGaaA
	<i>dadAX</i>	←	-119	4,56	GTaAAAtAAAtTGTgA
		←	-56	4,46	GTTAAAtTttcTGTcA
	<i>fadD</i>	→	-78	4,19	TAAaATaTtGTTAAa
		←	-80	5,14	GTTAAaatAtTGTTA
	<i>cydAB</i>	→	-133	4,48	ТААСctTaTgGTTAAC
		→	-163	4,14	tTgAAAtAAAtTGTcA
		→	-135	4,24	tTTAACTtAtgGTTA
		←	-124	4,42	GTTAACTAAcTGccA
	<i>sfcA</i>	→	-164	4,10	aAAaAaTTttTTAAC
		→	-338	4,48	tTTAAAtAAAtTcTTA
		→	-270	4,12	aTTgAaTAAcTGTTA
	<i>sdhCDAB</i>	←	-473	4,34	GTTAcaTAgtTGTTA
		←	-462	4,05	GTTAACgAttTGTcg
		←	-299	4,40	GTTtAtTcctTGTcA
	<i>gltA</i>	→	-365	4,40	TgACaagAaTaAAC
		→	-202	4,05	cgACAaaTcGTTAAC
		→	-191	4,34	ТААСAacTAtgTAAC
<i>fdoGHI</i>	→	-180	4,22	TcACAaTgtGgTAAC	
<i>fdhD</i>	←	-45	4,22	GTTAcCacAtTGTgA	
<i>sodA</i>	→	-86	5,50	TAtCATTTAaTTAAC	
<i>mlADR</i>	→	-339	4,04	ТААСctTgTgaTTAAC	

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>Y. enterocolitica</i>	<i>arcA</i>	→	-379	3,92	TAACAТаattTaAcC
	<i>napFDABC</i>	→	-291	3,73	tТаAACCaAAtcGTTA
	<i>atpIBEFHAGDC</i>	→	-156	3,65	aTTgAaTAAATGTTc
<i>P. multocida</i>	<i>sodA</i>	←	-63	4,53	GTTAAaatctTGTTA
	<i>yfiD</i>	→	-28	4,05	TAACAaaTcGccAAC
		→	36	4,34	TgACAaTTtaTTAAa
		←	-202	5,01	GTTAAaTAAATGTTt
		←	-191	5,14	GTTtAtTtAtTGTTA
	<i>focA</i>	→	-202	4,09	TtAaAaTattTTAAC
		→	-191	4,22	TAACAaTaAaTTgAt
		→	-80	4,16	TAAaATTTttTTAAa
		←	-251	4,53	GTTAAaatctTGTTA
		→	-229	4,41	tTTAAagtAATGTTA
		←	-218	4,15	GTTAAAtTtgtTGaaA
	<i>pflB</i>	←	-55	4,30	GTTActTttATGTaA
		←	9	4,49	GTcAAATAAATGaaA
	<i>dcuC</i>	←	41	4,16	GTTAttacAtTGTCa
	<i>gltA</i>	→	-167	4,91	TAACAgagcaTTAAC
		→	-112	5,07	TAACATTTtaTTtAC
	<i>sucABCD</i>	←	-220	5,07	GTaAAaAAAATGTTA
		←	-165	4,91	GTTAAAtgctcTGTTA
	<i>fnr</i>	→	-237	4,27	TAACAgtAaAaTaAtC
	<i>fumC</i>	→	-19	4,13	aAACAAagAGgTAAC
	<i>aceEF-lpdA</i>	←	-140	4,58	GTTAAaaaAttTGtgA
		→	-67	4,30	aTTAAAtTcAtTGTTt
	<i>ptsHI-crr</i>	←	-227	5,32	GTTAAAtTgAAATGTTA
	<i>cydAB</i>	←	-334	4,40	GTTgctaAAAATGTTA
	<i>dusB-fis</i>	←	-179	4,02	GTTAAgTAAAtTtTTg
	<i>aspA</i>	→	-231	5,29	TAACAaagtGTTAAC
		→	-233	4,57	tTTAACaAAgTGTTA
	<i>nqrABCDEFGH</i>	→	-258	4,22	TAttATTTAGTTAAa
		→	-227	4,53	TAtCAaaaattTTAAC
		←	-414	4,04	GaTAACaAAAATGTat
	<i>oxyR</i>	←	-71	4,14	GTcAAAAtttATGaTA
	<i>atpBEFHAGDC</i>	←	-210	4,48	GTTAACaAcAttTTA
	<i>gntXY</i>	←	-255	4,24	GaTAttgAAAATGTTA
	<i>talB</i>	←	-353	4,01	GTTAAaTtgATGaaA
	<i>cpxRA</i>	→	-27	4,02	tTTAAAtcAAAATGTgA
	<i>acpP</i>	→	-142	4,09	TAACtaTattTTcAC
	<i>fdxGHI</i>	←	-299	3,99	GTTgctTtttTGTTA
	<i>fdhD</i>	→	42	3,99	TAACAaaaAagcAAC
	<i>torCAD</i>	→	-342	3,95	tTTAttTtttTGTTA
	<i>A. actinomycetemcomitans</i>	<i>sodA</i>	→	-139	4,24
←			-95	5,26	GTTAAAtaAAATGTaA
<i>dusB-fis</i>		←	-66	4,18	GTTAAaagAATGaaA

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>A. actinomycetemcomitans</i>	<i>nqrABCDEF</i>	→	-241	4,22	TAttATTTAGTTAAa
		→	-211	4,46	TAgCAaaaAtTTAAC
	<i>atpBEFHAGDC</i>	→	-14	4,53	TtACAaaggGTTAAC
	<i>oxyR</i>	←	-80	4,22	GTgAAAtATAgGTaA
	<i>talB</i>	→	-205	4,34	atACAaaTAtTTAAC
		→	-194	5,38	TAACATTaAtTTAAC
	<i>dcuC</i>	→	-362	4,54	aTTAAAtTAttTGaTA
	<i>aceEF-lpdA</i>	←	-149	4,83	GTTAAaaAttTGTaA
	<i>fnr</i>	→	-318	4,20	TcACATTgAtTTAgC
	<i>torYZ</i>	←	-306	4,10	GTTAAAtTcAcTtTTt
	<i>fdhD</i>	→	-119	4,91	TAtCAaaaAaTTAAC
		←	-246	5,19	GTTAAaTAtcTGTTA
		←	-235	4,16	GTTAtacAttTGTTA
	<i>fdxGHI</i>	→	-158	4,16	TAACAaaTgtaTAAC
		→	-147	5,19	TAACAgaTAtTTAAC
		←	-274	4,91	GTTAAAtTtttTGaTA
	<i>sucABCD</i>	→	-126	4,25	TAtCAgTTAGTaAAa
		←	-60	4,71	GTTAAaTtttTGaTA
	<i>pflB</i>	←	-37	4,15	GTagAtaAttTGTTA
	<i>focA</i>	→	-124	4,31	gAAaTaTAaTTAAC
		←	-28	4,39	GTTAAaaAAATaTaA
	<i>frdABCD</i>	→	-32	3,86	TAAaATTgctTTAAt
	<i>cpxRA</i>	←	-354	3,97	GgTAACTAttTGgaA
<i>aspA</i>	←	-33	3,78	GTTAgaTcAAATtTaA	
<i>ptsHI-crr</i>	←	-215	4,42	GcTAACTcTtTGaTA	
<i>H. influenzae</i>	<i>fdhD</i>	→	-124	4,77	TAACgaTTttTTAAC
		→	-49	4,65	TAAaAaTaAtTTAAC
		→	-27	4,16	TtACATTTAacTAAt
	<i>fdxGHI</i>	←	-484	4,47	cTTAAAtgAAtTGTaA
		←	-258	4,16	aTTAgtTAAATGTaA
		←	-236	4,65	GTTAAaTtAtTtTTA
	<i>yfiD</i>	←	-161	4,77	GTTAAaaAAAcGTTA
		→	-124	4,48	TAAaAaTTAaTTAAa
		←	-196	5,38	GTTAAaTtAAATGTTA
	←	-185	4,94	GTTAAaTtAtTGaTA	
	<i>ldhA</i>	←	-236	4,79	GTTAAAtacAtTtTTA
	<i>nqrABCDEF</i>	→	-423	4,61	TAtCAaTTAtTgAAC
		←	-287	4,45	GTTAtaaAAATGTaA
	<i>pflB</i>	→	-21	4,10	TAtttaTTAaTTAAC
	<i>focA</i>	→	-468	4,24	TAACcaaattTTAAC
		←	-470	4,04	aTTAACcAAATtTTA
	<i>fur</i>	←	-470	4,39	GTgAttTAAAtTGaTA
<i>napFDAGHBC</i>	←	-299	4,39	GTgAAaaAAAtgGTTA	
	←	-253	4,16	GcgAAATAttTGaTA	
<i>fadD</i>	→	-52	4,45	tTgAAATAAATGaTA	

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>H. influenzae</i>	<i>gntXY</i>	←	-352	4,19	GaTAACCaAtATcTTA
	<i>pgk</i>	←	-155	4,25	aTTAttTAttTGTTA
	<i>aspA</i>	→	-26	4,22	TgACGaTaAaTTAAC
	<i>torYZ</i>	←	-496	4,10	aTTAACTttcTGTaA
	<i>glpFK</i>	→	-435	4,15	TAAAtATTgtGTacAC
	<i>ndh</i>	→	-54	4,23	TAtCATTacaTTAAt
		→	-49	4,33	TtACATTaAtgTTAAC
	<i>pckA</i>	→	-73	4,16	TAACATTTttTTAta
		←	-104	4,16	cTTAAaaAAATGTTt
	<i>frdABCD</i>	←	-91	4,36	GTTAttaAttTGTaA
	<i>cpxRA</i>	→	-192	4,02	TAAaATTTgaTcAAC
		→	27	4,40	TgAtATTgAaTTAAC
		←	-460	4,39	GTTAAaacAcTtTTA
	<i>nrfEXFG</i>	←	-164	4,71	aTTAAagAAATGTTA
	<i>sucCD</i>	→	-432	4,83	TAtCAaTTAGTTAAa
	<i>aceEF-lpdA</i>	←	-88	4,20	GTTAAaTgttTGTgA
	<i>moeAB</i>	←	-455	4,04	GTTAgCgtAATGTcA
	<i>ptsHI-crr</i>	→	-171	4,06	TttCATtagtTTAAC
<i>oxyR</i>	←	-303	3,98	GcTAAAtTcgtTGTaA	
<i>cydAB</i>	←	-212	3,93	tTTtACTAttTtTTA	
<i>sodA</i>	→	-104	3,91	aAACATTTAaTgAtC	
<i>H. ducreyi</i>	<i>dcuA</i>	←	-498	4,24	GTTAAataccAttTTA
	<i>frdABCD</i>	→	-246	4,36	TAgCAgTTAaTTtAC
	<i>napFDAGHBC</i>	→	-145	4,01	TAAAtAaaAaTTAAa
		←	-464	4,36	aTTAAAtTcAAcTTA
		←	-289	4,22	aTTAtCTAAATGTaA
		←	-202	4,07	GgTtAggAAATGTTA
	<i>ptsHI-crr</i>	→	-51	4,05	tTTAAAtTAtcTGTTt
	<i>mdh</i>	→	-76	4,08	TcACATTaAtTTAAa
	<i>arcA</i>	→	-122	4,10	aAAaAgTgAaTTAAC
		→	-63	4,02	cAACCaTaAtTcAAC
	<i>sodA</i>	→	-496	4,01	TAAaAaTgAtTAcAC
	<i>nrfABCD</i>	←	-257	4,34	GgTAAaTAAaAGTTA
	<i>nqrABCDEFGF</i>	←	-325	4,16	tTTAAaTcAAcGTTA
	<i>dusB-fis</i>	→	-419	4,95	TAAaAaTTAtTTAAC
	<i>oxyR</i>	←	-449	4,20	GTaAACTcAccGTTA
		←	-181	4,19	GTgtAtgAAATGTcA
	<i>ccmABCDEFGH</i>	←	-274	4,62	GcTAAAtTtttTGTTA
	<i>pgk</i>	→	21	4,20	TgACcTTgAtTTAAC
	<i>focA</i>	←	-185	4,10	GTaAtCatAtTGTTA
	<i>yfiD</i>	→	-33	4,57	TAAaAaTTAtTaAAC
	<i>fadL</i>	←	11	4,74	GTTtAaTtAcTGTTA
<i>fdhD</i>	→	-393	4,07	TAtCATTTttTTcAg	
<i>glpFK</i>	←	-347	4,18	tTTAAAtTtAtTtTTA	

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>H. ducreyi</i>	<i>glpABC</i>	←	-154	4,22	aTTAACTAttTGaaA
	<i>pckA</i>	←	-10	4,05	GgTgAtTttATGTTA
	<i>sucABCD</i>	←	-115	4,24	GTTAttattcTGTTA
	<i>cpxRA</i>	→	25	4,17	TgACcTagAaTTAAC
		←	-144	4,34	tTTAAaTAAATGTTc
	<i>aceEF-lpdA</i>	→	-37	4,03	TAtCATTactaTAAC
	<i>moeAB</i>	←	-119	4,15	GTTAAaatttTGaaA
	<i>cydAB</i>	→	-216	5,32	TAACATTcAaTTAAC
		→	-205	4,65	TAACATTTtaTTcAg
	<i>gntXY</i>	←	-438	3,72	tTTAAaaAAtgGaTA
	<i>aspA</i>	←	-339	3,86	GTaAAaaAAATGcaA
	<i>ndh</i>	→	-276	3,96	TAcCATacctTTAAC
	<i>moaACDE</i>	←	-324	3,93	GaTAACaAtcTtTTA
	<i>torYZ</i>	→	-115	3,93	TAAaAgaTtGTTAtC
<i>fnr</i>	←	-486	3,92	GaTAAtcAAAtTtTTA	
<i>V. vulnificus</i>	<i>fadD</i>	→	-133	4,53	TAACAaaattTaAAC
		←	-135	4,71	GTTAACaAAAtTaA
	<i>feoAB</i>	←	-322	4,41	GTTcACTAtAaGTTA
		←	-191	4,46	GTaAACaAAAcGTTA
	<i>sdhCDAB</i>	→	-117	4,20	TAACAagTttTTtAC
		←	-264	4,34	GTTAAaaAtcTGTTc
		←	-179	4,04	GTTAACTcctTaTaA
		←	-73	5,88	GTTAAtTAAATGTTA
	<i>gltA</i>	→	-337	5,88	TAACATTTAaTTAAC
		→	-231	4,04	TtAtAaggAGTTAAC
		→	-146	4,34	gAACAgatTtTTAAC
		←	-293	4,20	GTaAAaaActTGTTA
	<i>fur</i>	←	-171	4,59	GTTAtttaaAAtTGTaA
	<i>dadAX</i>	→	-54	4,08	TAcCAagTtGTaAAC
		←	-56	4,04	GTTAcCaAgtTGTaA
	<i>arcA</i>	←	-353	4,57	GaTtAaTAAtTGTTA
	<i>mdh</i>	←	-105	4,88	GTTAAtTgAtTGTaA
	<i>pckA</i>	←	-50	4,30	aTTAAtaAAAtTcTTA
	<i>fadBA</i>	←	-158	4,07	GTTtAaaAAATGTat
	<i>frdABCD</i>	→	-248	4,04	cAACATccAaTTAAC
	<i>pdhR-aceEF-lpdA</i>	←	-197	4,89	GTTtACTAttTGTaA
	<i>glpFK</i>	←	-289	5,50	GTTAAtTAAATGTaA
	<i>nqrABCDEFGH</i>	→	-422	4,17	TAACAcTgttTcAAC
	<i>ccmABCDEFGH</i>	←	-129	4,38	tTTAAaTAAATGTgA
		←	-108	4,40	GTTAACaAAAtTcTcA
	<i>fadIJ</i>	→	-353	4,32	TgtCATcTAGTTAAC
<i>pflB</i>	→	-134	4,02	TAtCgTTTtaTgAAC	
<i>cydAB</i>	→	-138	5,15	TAACAaTTAacTAAC	
	←	-304	4,20	GTTAACaAAcTGTct	
<i>ldhA</i>	←	-180	4,53	GTTAACccAAATGTTt	

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>V. vulnificus</i>	<i>fnr</i>	→	-389	4,22	TAACATTTAGaaAAg
		←	-276	4,04	tTTAACTcAATGaTt
	<i>nrfBCDEF</i>	→	-220	4,04	aAtCATTgAGTTAAa
		←	-107	4,22	cTTttCTAAATGTTA
	<i>nrfA</i>	→	-261	4,58	TAACtTTTaaTTtAC
		→	-241	4,59	TAtCAaTgAaTTAgC
		←	-46	4,34	tTTtAaTAttTGTTA
	<i>adhE</i>	→	-109	4,16	atACAaaTAaTaAAC
	<i>ppsA</i>	←	-487	4,40	tTTAAAtTcAATGTcA
	<i>glpABC</i>	→	-123	5,20	TtACATTaAaTTAAC
	<i>fadL</i>	→	-267	4,59	TtACAaaaAGTaAAC
		→	-130	4,07	TttCcgTgAGTTAAC
		←	-405	4,10	tTTAACcAcATGTTA
	<i>dcuB</i>	←	54	4,34	GTTActTcAtTGTcA
	<i>dcuC</i>	←	-340	4,02	GTaAAgatttTGTTA
	<i>talB</i>	→	-451	4,14	TAACAaTgAtTTgAg
→		-246	4,14	aTTAAaTAAcTGTcA	
<i>narQP</i>	→	-117	5,50	TtACATTTAaTTAAC	
<i>aldB</i>	←	-266	4,37	GTgAACTcAATGagA	
<i>fnr</i>	→	-21	3,97	TAtCAaaTAGTgAAa	
<i>napFDABC</i>	→	-61	4,14	TgACAgTTtaTTAAT	
<i>V. parahaemolyticus</i>	<i>fadBA</i>	→	-49	4,01	TAACAaTagGTatAC
		←	-162	4,07	GTTtAaaAAATGTat
	<i>mdh</i>	←	-105	4,88	GTTAAAtTgAtTGTaA
	<i>mtlADR</i>	→	-195	4,14	TAtCAaTTAaTTgAt
	<i>arcA</i>	←	-267	4,10	GTTAACaAtATaaaA
	<i>yfiD</i>	→	-168	4,07	TtttATTgAtTTAAC
		←	-221	4,05	aTTAAaTAttTtTTA
		←	-106	4,16	cTTAAaTcAATaTTA
	<i>dadAX</i>	←	-58	4,65	GTTAAAtaAAAAGTaA
	<i>gltA</i>	→	-336	5,32	TAACATTTgaTTAAC
		→	-290	4,57	aAACAcTTtGTTAAC
		→	-144	4,34	gAACAgatTTtTTAAC
		←	-292	4,66	GTaAACacttTGTTA
	<i>sdhCDAB</i>	→	-118	4,66	TAACAaagtGTTtAC
		←	-266	4,34	GTTAAaaAtcTGTTc
		←	-120	4,57	GTTAACaAAgTGTTt
		←	-74	5,32	GTTAAAtcAAATGTTA
	<i>feoAB</i>	→	6	4,13	TAtCAgaaAtTaAAC
←		-216	4,16	GTTAAgTtttTGTTt	
<i>fadD</i>	→	-120	4,53	TAACAaaattTaAAC	
	←	-122	4,71	GTTAACaAAATtTaA	
<i>cydAB</i>	→	-108	5,15	TAACAaTTAacTAAC	
	←	-326	4,48	GTTAACTcgATGTTt	
<i>aldB</i>	→	-113	5,50	TtACATTTAaTTAAC	

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>V. parahaemolyticus</i>	<i>adhE</i>	←	-31	4,34	tTTtAaTAttTGTTA
	<i>fadIJ</i>	→	-53	4,17	TAACAaaTtGaTtAC
	<i>ccmABCDEFGH</i>	→	-115	4,75	TgACAgTaAGTTAAC
	<i>nqrABCDEF</i>	→	-376	4,18	aAAaAaTgtGTTAAC
	<i>cpxRA</i>	→	-39	4,59	TAAAtATTaAGTaAAC
	<i>fumC</i>	←	-451	4,18	GcTAAAtTtAtTcTTA
	<i>atpIBEFHAGDC</i>	→	-96	4,35	TAataATTgcaTTAAC
	<i>ppsA</i>	←	-161	4,16	GTaAACatAATaTTA
	<i>fadL</i>	→	-125	4,42	aAACATacAaTTAAC
		→	-114	4,71	TAACATTacGTTtAC
	<i>nrdDG</i>	→	-327	4,35	TAACAaaTAGTgAtC
	<i>narQP</i>	→	-196	4,14	aTTAAataAAcTGTcA
	<i>napFDABC</i>	→	-66	4,14	TgACAgTTtaTTAAt
	<i>pckA</i>	←	-49	3,92	aTTAAaacAtTcTTA
	<i>fnr</i>	→	-21	3,97	TAtCAaaTAGTgAAa
	<i>glpFK</i>	←	-116	3,98	GTTAACaAAtTcaTg
	<i>glpP</i>	→	-308	3,87	TcACAcTgcaTaAAC
→		-150	3,97	TacaATTTAGTTAAa	
<i>V. cholerae</i>	<i>mdh</i>	→	-125	4,01	TAAaATTgtcTTcAC
		←	22	4,88	GTTAAtTgAtTGTaA
	<i>dadAX</i>	←	-51	4,53	GTTAAtgAAAaGTaA
	<i>ptsHI-crr</i>	→	-31	4,53	TtACAaaaAaTaAAC
	<i>fumC</i>	→	-424	4,16	TAtCATTTAaaTAAa
	<i>aldB</i>	→	-184	4,04	gAtCcTTTtGTTAAC
	<i>cydAB</i>	→	-94	5,15	TAACAaTTAacTAAC
	<i>pflB</i>	→	-132	4,01	TtttAaTTttTTAAC
	<i>fadD</i>	→	-136	4,53	TAACAaaattTaAAC
		←	-138	4,71	GTTAACaAAATtTaA
	<i>adhE</i>	→	-28	4,16	TAACATTTttTTAtt
		←	-217	4,10	GTTtAaTttATGTTt
	<i>feoAB</i>	←	-8	4,10	GTTAcCTctATGaaA
	<i>sdhCDAB</i>	←	-263	4,34	GTTAAaaAtcTGTTc
		←	-72	5,05	GTTgAaTAAATGTTA
	<i>gltA</i>	→	-332	5,05	TAACATTTAtTcAAC
		→	-141	4,34	gAACAgAtttTTAAC
	<i>fur</i>	←	-287	4,94	GTTAAaTtAtTGTaA
	<i>nqrABCDEF</i>	→	-401	4,11	gAgCAaaTtaTTAAC
	<i>yfiD</i>	←	-222	4,31	cTTAACTAttTtTTA
	<i>arcA</i>	←	-96	4,18	tTTActTtAtTGTTA
	<i>pckA</i>	→	-55	4,02	TAACAaTTacaTAAa
	<i>fadBA</i>	←	-125	4,07	GTTtAaaAAATGTat
<i>gltP</i>	←	-156	4,36	GcTAAaacttTGTTA	
<i>narQP</i>	→	-383	4,04	TAACAaaacGTgcAC	
	→	-56	4,47	gAtCAaTaAaTTAAC	

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>V. cholerae</i>	<i>napFDABC</i>	←	-116	4,47	GTTAAAtTtAtTGaTc
	<i>glgX</i>	←	-85	4,09	GaTAAaTAttTGTgA
	<i>torD</i>	→	-109	3,91	TAACAgaTtaacAAC
	<i>pdhR-aceEF-lpdA</i>	←	-433	3,97	aaTAACTAAATGTgA
<i>V. fischeri</i>	<i>fadBA</i>	→	-118	4,13	aACAaAaTtGTTAAa
		→	-45	4,21	TAAAtAaTgAaTAcAC
		←	-120	5,07	GTaAACaAAtTGTTA
	<i>nqrABCDEF</i>	→	-312	4,30	TAACATaaAaTaAAt
	<i>gltA</i>	→	-334	4,38	TAACATTTAtgcAAC
	<i>sdhCDAB</i>	←	-68	4,38	GTTgcaTAAATGTTA
	<i>sucABCD</i>	←	-277	4,14	GTTggtTAAAtTGaTA
	<i>cydAB</i>	→	-151	5,25	TgACATTTAaTTAAC
	<i>gltP</i>	→	-159	4,09	aACAaTaAaTAcAC
	<i>ldhA</i>	→	-380	4,46	TAcCAaaTtGTaAAC
	<i>dcuB</i>	→	-34	4,16	aAtCAaTTAGTTAAa
	<i>ppsA</i>	→	-111	4,01	TAAataaaAGTTAAC
		←	-197	4,74	GTTAAAtTccttTtTTA
	<i>pflB</i>	→	-441	4,21	TAACAaTggtTTAAt
	<i>moeAB</i>	←	-472	4,04	GTTAACgtAATaTcA
		←	-73	4,26	GTctAtTAgaTGTTA
		←	-39	4,14	GTTcAaTcAtTaTTA
	<i>dcuC</i>	→	-181	4,32	TAACAaTTAtTcAtC
	<i>fadIJ</i>	←	-262	4,36	GTTAAAtTgttTtTTA
	<i>napFDABC</i>	←	-186	4,14	GTTtgtTAAAtTGTgA
	<i>narQP</i>	→	-16	4,14	TcACAaTTAacaAAC
	<i>yfiD</i>	→	-128	4,01	cTgAAAtTAAAtTGaaA
	<i>arcA</i>	→	-114	4,65	TtAcCtTTTtaTTAAC
	<i>pdhR-aceEF-lpdA</i>	←	-176	4,89	GTTAACTAAATGTat
	<i>frdABCD</i>	→	-499	4,11	TAAaAaagAaTgAAC
		→	-351	4,14	TAcCtaTTttTTAAC
		←	-353	4,31	GTTAcCTAttTtTTA
	<i>cpxRA</i>	←	-194	4,08	GTTgAagAAATcTTA
	<i>aspA</i>	→	-292	4,04	TAAaAgagtGTaAAC
	<i>glpFK</i>	→	-463	4,51	cAtCATTTAGTTcAC
	<i>glpD</i>	→	-271	4,04	TtACATaTgaTTAtC
	<i>pepT</i>	→	-191	4,07	TAtCAgTaAtTTAAt
	<i>glgB</i>	←	-194	4,69	GTaAAaTtAtTGTTA
		←	-69	4,34	GTTAACTtAATcTcA
	<i>glgC</i>	←	-153	4,20	GTaAAAtTgtcTGTTA
	<i>focA</i>	→	-413	4,92	TAACAaaTAaaTAAC
	<i>feoAB</i>	←	-102	3,93	GTcAAAtTtAcTGaaA
	<i>nrfBCDEXF</i>	→	-449	3,72	TtttAaTTAaTTAgC
		←	-393	3,72	GTaAAAtTAtATacTA
	<i>nrfA</i>	→	-91	3,72	TAgTATaTAaTTtAC
		←	-35	3,72	GcTAAAtTAAAtTaaaA

Геном	Оперон	Сайт			
		Направление	Положение	Вес	Последовательность
<i>V. fischeri</i>	<i>pckA</i>	←	-320	3,94	GTTgcaaAAcTGTTA
	<i>fadL</i>	→	-342	5,18	TttCATTTAGTTAAC
		→	-143	4,10	TAACgaTTttTTAgC
		←	-209	4,04	GTTAAagctcTaTTA
		→	-145	4,01	tTTAACgAttTtTTA

Приложение 16.

Последовательности потенциальных сайтов связывания белка NarP. Указано положение сайта относительно старта трансляции. Позиции сайта, совпадающие с консенсусом, показаны прописными буквами, несовпадающие – строчными.

Геном	Оперон	Сайт			
		Положение	Вес	Последовательность	
<i>Y. pestis</i>	<i>glpFK</i>	-469	3,74	c ACCTtTatAaAGGTg	
	<i>fadBA</i>	-482	3,80	T AaCctTaAAGgGa T A	
	<i>gltP</i>	-346	3,67	T AaaaCa T AAttGG T A	
	<i>dmsABC</i>	-178	3,90	ct CC T Cga A AGAGG T A	
	<i>napFDABC</i>	-184	4,48	T Aa C T C Ta A AGAG t a A	
	<i>ptsHI-crr</i>		-392	3,71	T Ag t at T T A atgGG T A
			-319	3,73	T AC C Ta T T t GcaG T A
	<i>ccmABCDEFGH</i>	-113	4,67	T AC C T C TatAagGG T A	
	<i>sfcA</i>	-23	3,63	T Aa C gt T T A AGAG t a A	
	<i>pflB</i>		-366	3,52	T AC t T t TattGAGG c A
			47	3,54	T taaca T T A AAGG T A
	<i>nrdAB</i>	-26	3,51	a AaCaCa T tAaAGG T A	
	<i>cydAB</i>	-372	4,02	T Aa C at T T A atG t T A	
	<i>nirBDC-cysG</i>	-99	4,75	T AC C a C c T tAGAGG T A	
	<i>fdhD</i>	-10	3,64	T AC C aCa T tgtgat T A	
<i>focA</i>	-24	3,39	T Aa C c C TatAagaGag		
<i>fdoGHI</i>	-171	3,64	T Aatcaca A tG t GG T A		
<i>Y. enterocolitica</i>	<i>narP</i>	-338	4,48	T ta C T T t T AGAG t T A	
	<i>napFDABC</i>	-182	4,48	T Aa C T C Ta A AGAG t a A	
	<i>ccmABCDEFGH</i>		-113	4,67	T AC C T C TatAagGG T A
			-79	3,83	T AC t a C g T AAGgat T A
	<i>torYZ</i>	-180	4,35	T Aa C c t TatAGAG t T A	
	<i>pflB</i>	-23	3,70	T tCaca T T A AAGG T A	
	<i>focA</i>	-21	3,85	T Aa C c C T T AAGAGag A	
	<i>fadD</i>	66	3,76	T t C t T C T TtgG t GGa A	
	<i>sfcA</i>	-22	3,55	T Aa C gt T T A AGAG t g A	
	<i>nirBDC-cysG</i>	-101	4,75	T AC C a C c T tAGAGG T A	
	<i>fdoGHI</i>	-183	3,64	T Aatcaca A tG t GG T A	
	<i>fdhD</i>	-43	3,64	T AC C aCa T tgtgat T A	
<i>atpIBEFHAGDC</i>	10	3,72	T c C T t TatAG t GGg A		
<i>P. multocida</i>	<i>ccmABCDEFGH</i>	-99	4,80	c AC C T C Ta A AGAGG T A	
	<i>nrfABCDEFGF</i>		-289	4,79	T Aa C T C Ta A atAGG T A
			-206	4,72	T AC t Ta T T t AGAGG T A
	<i>pflB</i>	-17	3,71	T Aat T aa T AgaAGG T A	
	<i>frdABCD</i>	-141	4,36	T Aa C T t T t tGAGG T A	
	<i>gltA</i>	-290	3,51	T Aa t aa T aAgGAG t g A	
	<i>sucABCD</i>	-43	3,51	T ca C T C c T tAttat T A	
	<i>fdxGHI</i>	-234	3,68	T t C C T ta T tAatGG T A	
	<i>fdhD</i>	-24	3,68	T AC C at T a T aAGG a A	
	<i>malQ</i>	-38	3,62	T Aa C Ta T T A atcat T A	
	<i>mdh</i>	-21	4,25	T Aa C cCa T AAGgaG T A	
	<i>fnr</i>	-23	3,72	c AC t T C T T AAttG t a A	
<i>nrdAB</i>	-149	3,60	T AC t Ta T T t AtAaa T A		

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>P. multocida</i>	<i>fumC</i>	-64	3,65	TAaCcCgcAAGtGtTA
	<i>nqrABCDEFBCDEF</i>	-376	4,43	TAaCcCTTtAtAaGTA
		-351	4,13	TACtcaTTAAGAGtaA
	<i>dcuB</i>	-105	4,06	TACCTaTTtAttGcTA
	<i>adhE</i>	-77	3,76	TACtaCTTttcAGGTA
	<i>pckA</i>	-79	3,60	TAaCagTTtttAGGTA
		-18	3,70	TAaCTCTatgagGtaA
	<i>talB</i>	-238	3,60	aAaCaCTTAgGAGGaA
	<i>dmsABC</i>	19	3,92	TACCTCTTtttgGtTg
	<i>torCAD</i>	-268	4,14	TAaCTCTTAAaAGaTA
	<i>pgk</i>	-68	3,57	TAgtTagaAAGgGGTA
	<i>eno</i>	-22	3,60	TtCaTtTaAAGAGGaA
<i>dcuA</i>	-105	3,65	TtCtTaTaAAaAatTA	
	-22	4,56	TAaCatTTtAGAGGTA	
<i>napFDAGHBC</i>	-251	3,24	TgatgaTaAAGAGGTA	
<i>A. actinomycetemcomitans</i>	<i>nrfABCDEFXG</i>	-403	4,79	TACCTCTTtAtAGtTA
		-365	3,55	TtCCcCTTtAatGaTA
		-343	4,33	TACtaaTaAatAGGTA
	<i>ccmABCDEFGH</i>	-99	3,57	TACCgtgaAAGAGtTA
	<i>frdABCD</i>	-311	3,26	cAatatTTAAGAataA
		-270	4,20	TAaCTtTTtGAGtTA
		-182	3,25	TAaCagTctAaAGtTA
	<i>narP</i>	7	3,29	TAaaTtaaAAGtatTA
	<i>eno</i>	-20	3,70	TAaCTtaTAtGAGGaA
	<i>nqrABCDEFBCDEF</i>	-351	4,36	TACCTCTTtgGtatTA
	<i>dmsABC</i>	-495	3,70	gACCTCTTtAGcGtTA
		-44	3,68	TcaCcaTaAAtgGtTA
	<i>atpBEFHAGDC</i>	-477	3,69	TACCaCTTtttgAGTg
	<i>dcuA</i>	-19	3,45	TtatTCtaAgGAGGcA
	<i>fumC</i>	-20	4,51	TAatcaTTAAGAGGTA
	<i>mdh</i>	-18	3,29	TACtaCaTgAGgaGaA
	<i>aceEF-lpdA</i>	-257	3,77	TACCaaTTtgtgatTA
	<i>fnr</i>	-78	3,41	TAatgaaTtAGtGGTA
	<i>moaACDE</i>	-296	4,28	TAatTaTTAtGAGGTA
	<i>torYZ</i>	-63	3,84	TAaCTCgTtgtAGtTA
		-43	3,53	TtCCTtTTAgttatTA
	<i>fdhD</i>	-182	3,44	TAagTaaTAAtAGtTA
	<i>fdxGHI</i>	-212	3,44	TAaCTaTTAttActTA
	<i>dcuB</i>	-392	3,85	TtCtTCTaAttgGGTA
		-56	4,64	TACtaCTTAAGAaGTA
	<i>adhE</i>	-50	4,13	TAatTCTaAgtAGtTA
	<i>cydAB</i>	-107	3,44	TcatTaTaAtaAGGTA
	<i>pflB</i>	-18	3,62	TAatctaaAAGAGGaA
	<i>napFDAGHBC</i>	-153	3,47	TACtcaTTtgtAGtgA
	<i>ptsHI-crr</i>	-23	4,34	TAaCTtTTtAGAGGaA
<i>H. influenzae</i>	<i>napFDAGHBC</i>	-87	3,92	TACCTaTTttGgGtTg
	<i>fadD</i>	52	3,69	TACCTCaaAAtAtGaA
	<i>pgk</i>	-71	3,50	TAgCaCgaAgtAGGTA

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>H. influenzae</i>	<i>torYZ</i>	-47	4,13	TAaCTaTaAgtAGtTA
	<i>glpFK</i>	-254	3,50	gAatTCCTAAaAGGaA
		-20	3,68	cACtTCtaAgGAGtgA
	<i>narP</i>	-18	3,59	aACCaCTaAAGgaGaA
	<i>dcuA</i>	-76	3,79	TACCTaTTtgttGcTA
	<i>ndH</i>	-135	3,59	TAaCTCaaAAaAataA
	<i>pckA</i>	-125	3,97	TcaCTaaaAAGAGGTA
	<i>frdABCD</i>	-150	4,02	TAgCTtTTttGAGGTA
		-47	3,55	TACCctTTtAGgGGat
	<i>arcA</i>	-283	3,66	TACTtTtAAAaAGGTt
		-82	3,72	cACtTtTTtAttGtTA
		-27	3,55	TACgcCaaAatAaGTA
	<i>eno</i>	-22	3,54	TAaCTtaattGAGGaA
	<i>dmsABC</i>	-437	3,98	TcCCcaTTAAatGGTA
		-240	3,55	TAgTtaaTAatAGtTA
		-104	4,28	TACCTaTTAAatgatTA
		-46	3,75	TAaCTCgTtttAGtTA
	<i>nrfABCD</i>	-188	4,23	TACCTCTtttgaGTA
	<i>cydAB</i>	-133	3,60	TAaCctTTAAaAataA
	<i>sodA</i>	-179	4,79	TAaCTCTTtAGaAGTA
	<i>ccmABCDEFGH</i>	-85	4,79	TACTTCtaAAGAGtTA
	<i>talB</i>	-71	3,79	TACCctTTtAatGcTA
	<i>mdh</i>	-384	3,52	TtCCatTTtgGtatTA
-20		4,22	TAaCaCaTAAGgaGTA	
<i>malQ</i>	-29	3,57	TcCCTaacAAGAGtTA	
<i>fumC</i>	-62	3,62	TAaCTCggAAGtGtTA	
<i>fnr</i>	-71	3,58	aAaCTaaTtAGtGGTA	
<i>sucAB</i>	-18	3,94	TcataaTaAAGAGGTA	
<i>moaACDE</i>	-323	3,59	TACCTCaTAatAatcA	
	-305	3,59	TAaaaaTTtAtgGtTA	
	-269	3,50	TAtCaCTaAtGAGcTA	
<i>H. ducreyi</i>	<i>napFDAGHBC</i>	-81	3,62	TACCTaaatgGAGtgA
	<i>dcuB</i>	-97	4,34	TACCTCTcttGAGGTA
	<i>mdh</i>	-373	3,52	TtCCatTTtgGtatTA
		-26	3,91	TAaCcaTTtAtgatTA
	<i>nqrABCDEFBCDEF</i>	-308	4,27	TACTaCTTttGAGtTA
	<i>aspA</i>	-66	3,73	TtaCTCgaAatAGtTA
	<i>ccmABCDEFGH</i>	-117	3,83	TACTataaAAaAGGTA
		-86	4,95	TACCTCTaAAGaAGTA
	<i>pgk</i>	-72	3,50	TcgCcCaaAatAGGTA
		-33	3,60	TACaTtTTAAGActTA
	<i>ndh</i>	-97	3,94	TAaCaCTaAAaAGtgA
		-47	3,61	TACTtTtTttGAttTA
	<i>pflB</i>	-204	3,62	TcaCaCTTtAttatTA
		-21	3,57	TAaCatTTAAGAGGac
<i>yfiD</i>	48	3,98	TAaCTCTtctgGtTA	
<i>sucABCD</i>	-127	3,66	gACCTCTTAtttGtTA	
<i>moaACDE</i>	-239	3,57	TAaaaCTatAaAatTA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>H. ducreyi</i>	<i>torYZ</i>	-201	3,57	TAatTtTatAGtttTA
	<i>aceEF-lpdA</i>	-276	3,60	TAaCaCTTggtgaGTA
	<i>cydAB</i>	-152	4,05	TtCtTaTTtAGGtTA
		-111	3,51	TACCTCTattatataA
<i>nrfABCD</i>	-182	3,32	TAttTCaatAaAaGTA	
<i>V. vulnificus</i>	<i>yfiD</i>	-158	3,51	TAaCTCaatAaAataA
	<i>aspA</i>	-224	4,89	TACCcCTTtAGAGtTA
	<i>frdABCD</i>	-85	3,54	TACTaCTTttGgGGag
	<i>deoCABD</i>	-369	3,54	TAtCTaTaAtGtaGTA
		<i>ccmABCDEFGH</i>	-147	4,48
	<i>pflB</i>	-23	4,11	TtCtaCTaAAaAGGTA
	<i>cydAB</i>	-389	4,95	TACCTCTaAatAGGTA
	<i>napGH</i>	-102	3,64	TtaCcCTTttGgGGTg
		-101	3,73	TACCctTTtgGgGtgA
	<i>nrfBCDEFG</i>	-304	4,92	TACCcCTaAAGgGGTA
		-202	3,87	TACTaCTaAAGAGtag
	<i>nrfA</i>	-295	3,87	ctaCTCTTtAGtaGTA
		-193	4,92	TACCcCTTtAGgGGTA
	<i>moaABCDE</i>	-255	3,89	TACTcacaAgGtGGTA
	<i>adhE</i>	-207	4,24	cACCaCTTtgGgGGTA
	<i>nirBDC-cysG</i>	-105	4,00	TACCcCTTtAGGgtag
		-17	4,47	TAaCcCTTAgGAaGTA
	<i>narQP</i>	-188	4,73	TACCcCTTAgGgGGTA
	<i>glgB</i>	-191	3,89	TACCagTaAAatGGTA
	<i>glpP</i>	-420	3,61	TACCTtaaAAaAaGTg
-83		3,58	TACTtTtTtAttttTA	
<i>feoAB</i>	-242	3,32	TACTctTaAgttGtTg	
<i>napFDABC</i>	-281	4,73	TACCcCcTAAGgGGTA	
<i>V. parahaemolyticus</i>	<i>pckA</i>	-223	3,60	TAatTtTaAgtAatTA
	<i>dadAX</i>	-6	3,55	TACCTCaatgGAaGtc
	<i>feoAB</i>	-50	3,52	TACTctTaAatAataA
	<i>pflB</i>	-23	4,11	TtCtaCTaAAaAGGTA
	<i>torD</i>	-67	3,50	TtaCaCTTAgGgGaTA
	<i>cydAB</i>	-359	4,04	TAataCTaAttAGGTA
	<i>nrfBCDEFG</i>	-300	4,89	TACCcCTaAAGtGGTA
		-196	4,22	TACCTCTTAgGAataA
	<i>nrfA</i>	-294	4,22	TtatTCcTAAGAGGTA
		-190	4,89	TACCaCTTtAGgGGTA
	<i>moaABCDE</i>	-217	3,77	TACTcacTAAGcGGTA
	<i>adhE</i>	-191	3,98	cACTaCTTtgGtGGTA
	<i>ccmABCDEFGH</i>	-145	3,53	TAgCcactTAtgGtTA
	<i>glpFK</i>	-326	4,47	TACCTCTTAgGgatTA
	<i>deoCABD</i>	-372	3,54	TAtCTaTaAtGtaGTA
	<i>pdhR-aceEF-lpdA</i>	-52	3,59	cAaCTtaaAAttGGTA
	<i>frdABCD</i>	-87	3,54	TACTaCTTttGgGGag
	<i>aspA</i>	-226	4,26	TACCcacaAAGgGtTA
	<i>dcuA</i>	-22	3,65	aAaaaCTaAAGAGGTA
	<i>ldhA</i>	-18	3,52	TACTtTtTggGAGGgA

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. parahaemolyticus</i>	<i>nirBDC-cysG</i>	-107	4,00	TACCcCTTtAGgGtag
		-17	4,47	TAaCcCTTAGGAaGTA
	<i>narQP</i>	-139	4,99	TACCTCTTAgGAGGTA
	<i>napFDABC</i>	-124	4,99	TACCTCcTAAGAGGTA
	<i>napGH</i>	-82	4,90	TACCTCTTAatGAGGTA
	<i>yfiD</i>	-27	3,35	cACCaCgTAAGtGGTc
<i>V. cholerae</i>	<i>dadAX</i>	-147	3,63	TACCTgcatAaAaGTA
	<i>cydAB</i>	-101	3,52	TACaaCTTAaAatTA
	<i>torYZ</i>	-45	4,56	TACCcCaaAAGgGGTA
	<i>adhE</i>	-191	4,03	TACCaCTTgttAGGTA
	<i>ccmABCDEFGH</i>	-143	4,23	TAaCTacaAAGgGtTA
		-322	3,74	TAatcCTatAGActTA
	<i>focA</i>	-6	3,76	TACCTCTatgaAatTg
		-257	3,50	TtaCaaTTttGtaGTA
	<i>yfiD</i>	-220	3,65	TAaCTaTTtttatTA
		-61	3,59	cAaCTtaaAAttGGTA
	<i>pdhR-aceEF-lpdA</i>	-134	4,36	TACCaCTTAatgGGTg
	<i>frdABCD</i>	-173	3,54	TAaaTCaattGtGGTA
	<i>ldhA</i>	-53	3,89	TtCCcacTAatgGGTA
	<i>narQP</i>	-94	4,99	TACCTCTTAgGAGGTA
		-20	4,17	TAgCcCTaAAGAGGTg
<i>napFDABC</i>	-153	4,17	cACCTCTTtAGgGcTA	
	-79	4,99	TACCTCcTAAGAGGTA	
<i>V. fischeri</i>	<i>gntXY</i>	-50	3,72	TgCCaaTaAAGAGGTg
	<i>deoCABD</i>	-241	3,90	TgCCTCaaAAGAaGTA
	<i>nqrABCDEFBCDEF</i>	-251	3,69	TgCtaCTTAtGgGGTA
	<i>nirBDC-cysG</i>	-114	4,25	TACCaCcTtAGAGcTA
		-100	3,61	TACCcCTTtcGAGtag
	<i>fur</i>	-209	3,71	TACaatTTttGgGGTA
	<i>adhE</i>	-408	3,56	ctaCcCgTAAGgGGTA
		-394	3,66	TAagTCgaAAGtGGTA
	<i>cydAB</i>	-170	3,59	TtCtataTtAGAGtTA
		-118	4,07	TAaCaCTTtgtgGtTA
	<i>glpP</i>	-203	3,51	ctCtcCTaAAttGtTA
		-57	3,62	TAataaaTAAGAGcTA
	<i>ldhA</i>	-241	3,55	aACCTCTTAAGcatTA
	<i>nrfBCDEFG</i>	-383	4,64	TACtaaTTAAGAGGTA
		-309	4,34	cACCatTaAAGAGGTA
		-293	4,94	TACCTCTTAAGtGtTA
		-205	3,64	TAtCaCaaAAGgaGTA
	<i>nrfA</i>	-280	3,64	TACTcCTTttGtGaTA
-192		4,94	TAaCaCTTAAGAGGTA	
-176		4,34	TACCTCTTtAatGGTg	
-102		4,64	TACCTCTTAAttaGTA	
<i>sfcA</i>	-431	3,51	TACCaCTTtActttTA	
	-221	3,68	TAaaTtTTtAtgGtTA	
	62	3,52	ctaCTCaatAaAGGTA	
<i>pflB</i>	-270	4,22	TAataCTTAtGgGGTA	

Геном	Оперон	Сайт		
		Положение	Вес	Последовательность
<i>V. fischeri</i>	<i>pflB</i>	-25	4,11	TtCtaCTaAAaAGGTA
	<i>ccmABCDEFGH</i>	-191	4,52	TACCTacaAAGtGGTA
		-141	3,53	TACaaaTTAcaAGGTA
	<i>ptsHI-crr</i>	-179	3,54	TAaCaaTTttGAactA
	<i>napFDABC</i>	-86	4,25	TAaCcaTTAtGgGGTA
	<i>narQP</i>	-117	4,25	TACCcCaTAAtgGtTA
	<i>dcuB</i>	-262	3,55	TAaCTCaTtttgatTA
	<i>yfiD</i>	-287	3,80	TACCTacaAttgGtTA
	<i>pdhR-aceEF-lpdA</i>	-399	4,39	TACCcCcaAatAGtTA
	<i>frdABCD</i>	-195	3,84	TgCCcCaTtAGtGGTA
	<i>aspA</i>	-141	4,64	TACCaCTaAAGAGGTg
	<i>dcuA</i>	-67	3,60	TtaCTaTatAtAGcTA
		-34	3,71	TAatTaTTtAatatTA
	<i>nagB</i>	-461	3,81	TAatatTaAtGAGtTA
	<i>pckA</i>	-126	4,50	TACCaCTTtAtgGtTA
	<i>torYZ</i>	-65	4,46	TAaCcCcaAAGtGGTA
	<i>glpFK</i>	-90	3,50	TAaaTaTTAAGAatgA
	<i>glpABC</i>	-32	4,14	TACaTCaaAAGtGGTA
	<i>pepT</i>	-163	3,56	TtCaTCTTAAaAGcTA
	<i>torCA</i>	-492	4,51	TtaCTCTTtAGgGGTA
		-491	3,66	TACTctTTAgGgGtaA
		-34	3,77	TAaCgaTTtAGAatTA
	<i>glgC</i>	-232	3,54	TACaTaTTttaAaGTA