На правах рукописи

Осипов Александр Александрович

ЭЛЕКТРОСТАТИЧЕСКИЕ СВОЙСТВА ГЕНОМНОЙ ДНК

03.00.28 – Биоинформатика

Автореферат
диссертации на соискание ученой степени
кандидата биологических наук

Москва 2009

Работа выполнена в Учреждении Российской академии наук Институте биофизики клетки РАН.

Научные руководители: доктор биологических наук
Камзолова Светлана Григорьевна

кандидат физико-математических наук
Сорокин Анатолий Александрович

Официальные оппоненты: кандидат физико-математических наук, доктор биологических наук, профессор
Миронов Андрей Александрович
Московский государственный университет имени М.В. Ломоносова

кандидат физико-математических наук
Есипова Наталия Георгиевна
Учреждение Российской академии наук Институт молекулярной биологии им. В.А. Энгельгардта РАН

Ведущая организация: Учреждение Российской академии наук Институт теоретической и экспериментальной биофизики РАН

Защита диссертации состоится 20 марта 2009 года в 14 часов на заседании диссертационного совета Д002.077.02 при Учреждение Российской академии наук Институте проблем передачи информации им. А.А. Харкевича РАН по адресу:
127994, г. Москва, ГСП-4, Большой Каретный переулок, д.19. стр. 1.

С диссертацией можно ознакомиться в библиотеке Учреждения Российской академии наук Института проблем передачи информации им. А.А. Харкевича РАН

Автореферат разослан 19 февраля 2009 года

Ученый секретарь диссертационного совета
доктор биологических наук, профессор Рожкова Г.И.

Общая характеристика работы

Актуальность темы

На данный момент существует дисбаланс между большим и постоянно растущим количеством секвенированных геномов и недостатком их биологического описания. Невозможность эффективного биохимического и генетического изучения такого количества геномов, лишь отчасти компенсируемая современными высокопроизводительными методами исследований, диктует необходимость развития методов анализа и интерпретации текстов первичной последовательности ДНК. Одним из направлений такого анализа является предсказание функций по первичной структуре специфических участков ДНК. Было разработано много инструментов, основанных на текстовом анализе последовательности ДНК, для предсказания некоторых ключевых свойств, таких как распределение и функции открытых рамок считывания, промоторов и других регуляторных элементов.

Однако, несмотря на накопленную информацию о структуре последовательностей, до сих пор представляется затруднительным выделить исключительно на ее основе регуляторные элементы, такие как промоторы, или предсказать их функциональные характеристики. Множество алгоритмов поиска промоторов, основанных на текстовом анализе последовательностей, неудовлетворительно справляются с этой задачей.

Известно, что дополнительная информация для распознавания и модуляции активности промоторов может заключаться в физических свойствах ДНК, таких как общая геометрия двойной спирали, ее деформируемость, температурная стабильность и динамические свойства. В нашей лаборатории был предложен новый подход к этой проблеме на основе анализа электростатических свойств промоторной ДНК (Сорокин А.А., 2001, Джелядин Т.Р., 2001), для чего был разработан упрощенный метод вычисления распределения электростатического потенциала вокруг молекул ДНК величиной до целых геномов (Polozov R.V., 1999). С его помощью были проведены исследования электростатических свойств некоторых геномов, которые показали важность электростатических взаимодействий промоторной ДНК и РНК-полимеразы для регуляции функций промоторов. Электростатические свойства промоторной ДНК характеризуются выраженными паттернами, специфичными для различных групп промоторов, которые могут играть роль сигнальных элементов в дифференциальном распознавании соответствующих промоторов РНК-полимеразой.

Другим важным результатом было открытие нелинейной зависимости профиля потенциала от последовательности ДНК, означающей, что данное свойство обусловлено всей последовательностью целиком, в том числе фланкирующими регионами, нежели ее текстом в непосредственной точке рассмотрения, и для некоторых систем было показано, что биохимические свойства промоторов имеют гораздо лучшую корреляцию с их электростатикой, чем с текстом последовательностей.

Таким образом, электростатические свойства геномной ДНК весьма важны для ее биологических функций, и информация о них имеет большое значение для функциональной, сравнительной и эволюционной геномики, будучи представлена для значительного количества геномов, особенно интегрированной с возможно более полной аннотацией уже известных для них биохимических функций.

Цель и задачи исследования

В соответствии с обозначенной проблемой были установлены следующие цели:

1. создать инструмент, предоставляющий доступ к биологическим и электростатическим свойствам ДНК, и набор инструментов для анализа этих свойств
2. исследовать закономерности формирования электростатических свойств ДНК и общие электростатические свойства природных геномов
3. исследовать электростатические свойства промоторной ДНК Т7-подобных бактериофагов

Для достижения этих целей были сформулированы конкретные задачи:

1. разработать базу данных, содержащую последовательности геномов с биологической аннотацией и систематическим положением, и их электростатические свойства
2. разработать инструменты для визуализации электростатических свойств последовательностей геномов, сопоставления с аннотацией, проведения анализа и представления результатов
3. оценить взаимосвязь нуклеотидного состава последовательности ДНК и ее электростатических свойств и влияние на них окружения последовательности
4. провести исследование общих электростатических свойств природных геномов
5. провести исследование связи биологической функции и электростатических свойств последовательности на примере промоторов Т7-подобных бактериофагов, взаимодействующих с РНК-полимеразой бактерии-хозяина и с нативной фаговой РНК-полимеразой
6. провести исследование роли электростатических свойств в дифференциальном распознавании промоторов РНК-полимеразами фагов Т7 и Т3 на примере описанного в литературе эксперимента с мутантом Т7, приспособившимся к росту на РНК-полимеразе фага Т3

Научная новизна и практическое значение

Впервые создана база данных, содержащая электростатические свойства ДНК природных геномов, включающая сведения о всех полностью секвенированных бактериальных и вирусных геномах, а также ряде расчетных последовательностей.

Впервые исследованы сравнительные электростатические свойства полных геномов и обнаружена близкая к линейной зависимость среднего потенциала природных геномов и сбалансированных случайных последовательностей от содержания в них GC пар, а также рассчитаны ее параметры для разных таксономических групп. Установлено, что величина этой зависимости коррелирует с содержанием GC пар.

Установлен ряд закономерностей формирования электростатического потенциала вокруг молекулы ДНК природных геномов и случайных и регулярных последовательностей. Показана неоднозначная зависимость потенциала и его разброса от содержания GC пар и ее зависимость от сбалансированности и трековости последовательности, а также возможность формирования принципиально различающегося потенциала идентичными по составу фрагментами ДНК и идентичного – разными.

Впервые произведен количественный анализ и выявлена степень влияния фланкирующих участков и единичных замен на формирование потенциала в области рассмотрения. Показано, что окружение способно полностью видоизменить электростатический профиль участков ДНК, равных известным консервативным регуляторным последовательностям, а единичные замены могут как проходить бесследно, так и полностью менять профиль. Продемонстрировано благоприятное влияние естественного окружения на примере промоторов бактериофага Т3.

Высказана гипотеза, что в сдвиг распределения природных геномов в АТ-богатую область могли внести вклад б**о**льшие возможности формирования выраженных электростатических элементов АТ-обогащенными последовательностями по сравнению с GC-обогащенными.

Показано различие в масштабах, на которых проявляются закономерности распределения потенциала для промоторов, взаимодействующих с бактериальными и фаговыми полимеразами, величины которых различается почти на порядок, что отражает физическую картину взаимодействия ДНК с белком.

Показано, что приспособление промоторов бактериофага Т7 к взаимодействию с РНК-полимеразой бактериофага Т3 сопровождается изменением электростатического потенциала в районе 0 – -5 п.о., приводящим к формированию профиля, идентичного промоторам Т3, что свидетельствует о возможной зависимости от него дифференциального распознавание промоторов РНК-полимеразой Т3, при этом указанные отличия потенциала мало влияют на распознавание промоторов РНК-полимеразой Т7, но играют для нее регуляторную роль.

Результаты работы могут быть использованы при создании искусственных геномов с заданными свойствами, в частности, при разработке экспрессионных систем, а также при проведении научных исследований в области биофизики, биологии клетки, биоинформатики и сравнительной, функциональной и эволюционной геномики.

Апробация работы

Материалы диссертации докладывались на следующих конференциях:

III съезд биофизиков России. Воронеж, 2004; 4-я международная конференция «Bioinformatics of genome regulation and structure» BGRS-2004, 2004, Новосибирск; XII симпозиум по межмолекулярному взаимодействию и конформациям молекул. Пущино, 2004; Albany 2005, The 14th Conversation, 2005; International Moscow Conference on Computational Molecular Biology (MCCMB'05), Moscow, Russia, 2005; XIII Симпозиум по межмолекулярному взаимодействию и конформациям молекул. Санкт-Петербург, 2006; The fifth international conference on bioinformatics of genome regulation and structure (BGRS-2006), 2006; International Moscow Conference on Computational Molecular Biology (MCCMB'07), Moscow, Russia, 2007; Albany 2007, The 15th Conversation, 2007; International Workshop on Integrative Bioinformatics, 4th annual meeting, University of Ghent, Belgium, 2007; 11 Международная Пущинская школа-конференция молодых ученых «Биология наука XXI века» (2007 г, Пущино); International Conference on Computational Phylogenetics and Genosystematics, Moscow, Russia, 2007; European conference on synthetic biology (ECSB): Design, programming and optimisation of biological systems, Spain, 2007; XIV Симпозиум по межмолекулярному взаимодействию и конформациям молекул, 2008, Челябинск; Межинститутский научный семинар ИБК РАН и ИТЭБ РАН, 2008, Пущино; 16 Международная конференция "Математика. Компьютер. Образование", Пущино, 2009 г.

По материалам диссертации опубликовано 13 статей в рецензируемых журналах, 1 раздел в монографии, 3 статьи в научных сборниках и периодических научных изданиях и 20 публикаций в материалах научных мероприятий.

Структура и объем диссертации

Диссертация включает в себя обзор литературы, описание методов, 3 главы, посвященные изложению результатов и их обсуждению, заключение, выводы, список литературы и приложение. Работа изложена на 102 страницах и содержит 4 таблицы и 25 рисунков. Список литературы содержит 153 наименования.

Содержание работы

Материалы и методы

Для разработки базы данных электростатических свойств геномной ДНК (DEPPDB) и анализа данных использовались следующие материалы и методы.

Нуклеотидные последовательности и элементы геномов и их аннотации

Последовательности всех полных секвенированных бактериальных и вирусных геномов и их аннотации взяты из базы данных NCBI RefSeq (ftp://ftp.ncbi.nih.gov/refseq/) и частично из BioCyc (http://BioCyc.org). Данные в форме текстовых файлов взяты с ftp сайта и разбирались специально написанным набором программ на языке Perl. Ряд данных был получен из литературных источников и внесен в базу через интерфейс ее управления, также написанный на Perl.

Таксономический раздел

Описания таксонов и идентификаторы, позволяющие сформировать иерархическую древовидную структуру раздела и приписать геномы таксонам, взяты из базы данных NCBI Taxonomy (ftp://ftp.ncbi.nih.gov/pub/taxonomy/) в виде текстовых файлов и разбирались специально написанным набором программ на языке Perl.

Генерация случайных и регулярных последовательностей ДНК

С помощью специально написанной программы было рассчитано по 10 случайных последовательностей с содержанием каждого нуклеотида с шагом в 10% и длиной последовательности от 1000 до 100000 с шагом в порядок, результат статистических расчетов сохранен в базе, а также по одной последовательности длиной 1000000 с сохранением текста последовательностей для дальнейшего изучения, и набор последовательностей с равным содержанием всех 4 нуклеотидов.

С помощью специально написанной программы на языке Perl был рассчитан набор регулярных (периодических) последовательностей следующего вида: полинуклеотиды с периодом в 1 и 2 пары каждого вида, и все перестановки из 4, 8 и 12 пар с равным количеством нуклеотидов A, T, G и C. Из анализа исключались циклические перестановки (дающие при повторении одинаковые последовательности), из поли-12 нуклеотидов брались по 100 вариантов, имеющих максимальные и минимальные значения среднего потенциала.

Расчет электростатических свойств ДНК

Электростатический потенциал (ЭП) вокруг молекул геномной ДНК рассчитывался с помощью оригинального метода (Polozov et al., 1999), использующего расчет по закону Кулона полноатомной модели ДНК с использованием подгоночных параметров зарядов и диэлектрической проницаемости для согласования с расчетами, полученными решением уравнения Пуассона-Больцмана.

Вычислялось значение электростатического потенциала на поверхности соосного двойной спирали молекулы ДНК цилиндра, радиусом 15 ангстрем, что составляет около 5 ангстрем от ее поверхности, то есть примерно соответствует расстоянию, на котором, предположительно, белки неспецифически взаимодействуют с ДНК. Далее значение потенциала усреднялось по угловой переменной для получения одномерного распределения потенциала вдоль молекулы ДНК, т.е. профиля ЭП, который и использовался для заполнения базы и дальнейшего анализа.

Для получения линейных координат пар оснований вдоль молекулы ДНК генома и усредненных по углу значений электростатического потенциала вокруг молекулы ДНК в линейных координатах вдоль молекулы (т.е. профиля ЭП), использовалась программа А. Сорокина (Sorokin, 2001), модифицированная для пакетной обработки целых геномов и вычисления ряда дополнительных параметров распределения электростатического потенциала.

Вычислялись следующие показатели распределения усредненного потенциала вдоль целой последовательности: минимум, максимум, среднее арифметическое, геометрическое и гармоническое, медиана, дисперсия и стандартное отклонение, коэффициент асимметрии и эксцесс распределения.

Программное обеспечение СУБД, публикации данных и инструментов обработки и анализа

Хранение данных

Большая часть данных хранится в реляционной базе под управлением СУБД MySQL v5.0 в таблицах типа MyISAM.

Заголовочные части записей БД NCBI RefSeq, относящиеся к геному, хранятся в текстовых файлах операционной системы в формате ASCII, по одной записи на файл.

Тексты последовательностей хранятся в текстовых файлах в формате ASCII, непрерывной строкой, по одной последовательности на файл.

Линейные координаты (в ангстремах) пар оснований вдоль молекулы ДНК генома хранятся в бинарных файлах форматом 4 байта на основание.

Усредненные по углу значения электростатического потенциала вокруг молекулы ДНК в линейных координатах вдоль молекулы хранятся в нормализованном виде в бинарных файлах форматом 2 байта на 1 ангстрем.

Доступ к данным и инструменты анализа: веб-публикация

Пользовательский доступ к данным и инструментам анализа осуществляется через веб-интерфейс по протоколу http с помощью динамической системы публикаций, основанной на веб-сервере Apache v.2.2, СУБД MySQL v5.0 и программах, написанных на языке Perl. Система включает стандартную поставку ActiveState Perl v. 5.8 с рядом дополнительных модулей, один из которых модифицирован, и набор скриптов, написанных для БД DEPPDB.

Динамически генерируемые страницы в формате html содержат ряд интерактивных элементов, написанных на языке Javascript v.1.2 и тестировались в браузерах MS IE vv. 6,7, Mozilla Firefox vv. 2,3, Opera v. 9 и Google Chrome v.1.0.154.36. Графики строятся «на лету» в формате PNG с помощью модулей Perl GD и GD::Graph.

Кроме того, часть инструментов анализа используют расширение языка Perl PDL (Perl Data Language) v. 2.4.3 с графическим модулем PGPLOT v.2.19.

База данных доступна для академического использования через веб-интерфейс по адресу http://promodel.icb.psn.ru.

Следует отметить, что некоторые намеченные оптимизации программного и аппаратного обеспечения позволят кардинально улучшить возможности обработки данных.

Представление данных в работе

На всех рисунках, представляющих профили ЭП, по вертикальной оси отложена величина ЭП в единицах заряда электрона на ангстрем (e/A), по горизонтальной –

расстояние вдоль оси молекулы ДНК в ангстремах. Вертикальной линией по центру отмечена точка, по которой выравнивались последовательности.

Выравнивание по номеру нуклеотида не соответствует выравниванию в физическом пространстве из-за разницы расстояния между парами оснований. Все графики, в т.ч. и содержания GC пар, строились в реальном физическом пространстве.

В случае, когда на графике присутствуют 3 панели, на верхней дан электростатический потенциал, горизонтальные линии – среднее значение потенциала всего генома(ов); на средней – стандартные отклонения для каждой группы, горизонтальные линии – среднее значение для каждого генома (группы); на нижней – содержание GC пар в процентах для каждой группы. Для отображения GC состава делалось усреднение окном в несколько пар вокруг каждой точки.

Результаты и обсуждение

1. База данных свойств электростатического потенциала геномной ДНК DEPPDB

Общее описание данных

DEPPDB – база данных смешанного реляционно-файлового типа, содержащая информацию о геномной ДНК и ее свойствах, прежде всего электростатических, и ряд инструментов для работы с этой информацией. На данный момент база охватывает все полные секвенированные прокариотические – бактериальные, включая плазмиды, и вирусные геномы.

Основной объект базы – геном, представляющий собой непрерывную последовательность секвенированной единой молекулы ДНК биологического организма, с известными свойствами. Геном имеет набор общих свойств, характеризующих его как целое, и ряд свойств, приписанных его элементам, которые определяются позициями на его последовательности.

Геномы организованы с помощью таксономического раздела базы, основным объектом которого является таксон в общепринятом биологическом смысле. Таксоны организованы в иерархическую древовидную структуру, при этом с каждым таксоном связаны все геномы, относящиеся непосредственно к нему, а также всем дочерним таксонам.

Кроме того, база содержит ряд данных, рассчитанных для набора случайных и регулярных последовательностей ДНК, включая некоторые из самих последовательностей.

Общая характеристика генома

Общая характеристика генома включает идентификаторы (DEPPDB, NCBI RefSeq, GenBank); характеристику записи в базе NCBI RefSeq: дата; код раздела; характеристики геномной последовательности: длину; тип нуклеиновой кислоты, количество нитей, топология молекулы, количество оснований A, T, G и C; процентное содержание GC, рассчитанное для всего генома, длину молекулы в ангстремах. Для геномов, содержащих в своей последовательности неопределенные позиции, приведен их список.

Описание генома и организма включает определение генома, название организма и его систематическое положение, аннотацию генома как целого (из БД NCBI RefSeq), включая описания литературных источников.

Последовательность молекулы ДНК генома (для РНК-содержащих вирусов записана эквивалентная последовательность ДНК).

Линейные координаты пар оснований вдоль молекулы ДНК генома (для РНК-содержащих вирусов рассчитаны для эквивалентной последовательности ДНК).

Усредненные по углу значения электростатического потенциала вокруг молекулы ДНК в линейных координатах вдоль молекулы (для РНК-содержащих вирусов рассчитаны для эквивалентной последовательности ДНК).

Свойства распределения усредненного по углу электростатического потенциала вокруг молекулы ДНК: минимум; максимум; медиана; среднее арифметическое, геометрическое и гармоническое; дисперсия и стандартное отклонение; эксцесс; асимметрия.

Элементы генома

Элементы генома взяты из БД NCBI RefSeq и, как правило, имеют какую-либо выраженную биологическую функцию или структурную особенность.

Для каждого экземпляра элемента указаны его координаты в последовательности: ведущая или обратная цепь; позиции в п.о. относительно начала последовательности начала и конца всех сегментов последовательности, к которым относится данный элемент; общее начало и конец элемента; количество концов сегментов для протяженных (1 – для точечных) элементов.

Кроме того, для каждого экземпляра элемента приведена его полная структурированная аннотация по БД NCBI RefSeq, включающая описание его биологических функций и структурных особенностей, названия генов и белков и их транслируемые последовательности, экспериментальные сведения, комментарии и пр. В отдельную таблицу также вынесены ссылки из аннотации на внешние БД.

Таксономия

Таксономический раздел базы служит для организации геномов по таксономическому принципу. Основным объектом раздела является таксон в общепринятом биологическом смысле. Для каждого таксона имеется идентификатор в БД DEPPDB, NCBI Taxonomy его самого и родительского таксона; основное научное и ряд дополнительных названий с указанием их типа; ранг; код раздела; ссылка на литературное описание и ряд других параметров, взятых из БД NCBI Taxonomy.

Ссылка на родительский таксон и указание ранга позволяет организовать иерархическую древовидную структуру всего таксономического дерева.

Непосредственная принадлежность геномов таксонам низкого ранга определяется прямым соответствием идентификатора таксона идентификатору нуклеотидной последовательности по GenBank, взятыми из БД NCBI Taxonomy. Как правило, таким таксонам соответствуют полные геномы индивидуальных организмов, в свою очередь, геномы плазмид, особенно неспецифичных по отношению к хозяину, могут непосредственно принадлежать таксонам более высокого ранга.

Кроме того, все таксоны высшего ранга содержат ссылки на все геномы дочерних таксонов. Это позволяет рассчитывать и показывать для каждого таксона ряд обобщенных свойств входящих в него геномов и их элементов, что может служить для решения задач сравнительной и эволюционной геномики.

Инструменты анализа данных

Инструмент визуализации и анализа множественных электростатических профилей

Данный инструмент служит для визуализации и анализа множественных графиков электростатических профилей избранных участков геномов. В анализ входит вычисление и построение арифметического среднего для графиков и их групп, стандартного отклонения, среднего взвешенного и среднего взвешенного стандартного отклонения для всех участвующих в анализе геномов по профилям электростатического потенциала, а также арифметического среднего по содержанию GC пар с выбором окна предварительного усреднения для индивидуальных последовательностей.

Инструмент визуализации электростатических профилей (упрощенный вариант)

Данный инструмент строит графики электростатических профилей для выбранных геномов и позиций, позволяющие выбрать различные режимы сглаживания, расчета содержания GC пар и ряд других параметров.

Инструмент визуализации и анализа отношений среднего потенциала генома к содержанию GC пар для множественных геномов

Данный инструмент строит графики отношений среднего потенциала генома к содержанию GC пар для множественных геномов и рассчитывает прямые линейной регрессии для выбранных наборов геномов. Инструмент доступен по прямой ссылке с главной страницы и на страницах описаний таксонов. Интерфейс инструмента позволяет выбирать наборы данных, включать и выключать расчет прямых линейной регрессии и регулировать размеры графика и величину точек, представляющих каждый геном.

Основная статистика

Статистика по геномам и таксонам на текущий момент в базе приведена в таблице 1.

Таблица 1. Сводная статистика по геномам и таксонам базы.

Объекты	**Количество**
Индивидуальные таксоны всего	4393
Индивидуальные геномы всего	4528
Геномы с рассчитаными электростатическими свойствами	4266
Бактерии и плазмиды	1533
Вирусы	2733

2. Зависимость электростатических свойств последовательности ДНК от ее состава

Для выяснения закономерностей формирования электростатического потенциала были исследованы природные геномы и ряд последовательностей, рассчитанных по заранее заданным свойствам, таким как содержание нуклеотидов A, T, G и C и равномерность их распределения. Анализировалась зависимость среднего потенциала от содержания GC пар для разных вариантов его распределения. Результаты анализа представлены в таблице 2 и на рисунке 1.

Случайные последовательности ДНК

Случайные последовательности ДНК демонстрируют близкую к линейной зависимость среднего электростатического потенциала от содержания GC пар с весьма значительным разбросом значений (рис. 1, серые точки, таблица 2). При этом разброс значений среднего электростатического потенциала плавно уменьшается с ростом содержания GC пар от 0 до 100% с 1.59 до 1.0023 e/A, т.е. более чем в полтора раза. Таким образом, выявляется закономерность, заключающаяся в том, что большее содержание AT пар обеспечивает большие возможности изменений значения электростатического потенциала последовательности.

Рассмотрим зависимость формирования среднего потенциала от сбалансированности последовательности, т.е. отношения содержания A к T и G к C. Случайные последовательности, имеющие эти отношения в пределах 0.5 – 2 (рис. 1, синие точки, таблица 2), что близко к наблюдаемому в природных геномах, имеют в среднем в 5-6 раз меньший разброс и демонстрируют значительно более пологий график зависимости потенциала от GC состава, чем выходящие за эти пределы. При этом становится заметной отклонение от линейной зависимости в сторону уменьшения ее в AT-богатой области и увеличения в GC-богатой, что демонстрирует еще один аспект закономерности зависимости возможности изменений значения электростатического потенциала от GC состава последовательности.

Следует отметить также (очевидную) обратную зависимость величины разброса от величины исследуемой случайной последовательности, тем более выраженную, чем более сбалансирована последовательность.

Очевидным следствием случайного характера последовательностей является повышение вероятности образования треков нуклеотидов одного вида для несбалансированных последовательностей по сравнению со сбалансированными, что указывает на еще одну закономерность при рассмотрении разбросов – чем больше треков, тем более выражена зависимость потенциала от GC состава последовательности.

Регулярные последовательности ДНК

Наиболее интересным результатом анализа регулярных последовательностей является демонстрация возможностей формирования сильно различающегося потенциала идентичными по составу последовательностями. Диапазон средних потенциалов последовательностей, содержащих равное количество A, T, G и C (рис. 1, вертикальный ряд точек в положении GC=50%, таблица 2) в пределах соседних 4 нуклеотидов равен диапазону среднего значения для природных геномов от 27 до 65% GC (1.3064 e/A); диапазон для 8 – перекрывает вообще все значения для всех природных геномов (2.0834 e/A), достигая 2.2343 e/A, а размах минимальных и максимальных значений для 12 (3.9776) больше среднего размаха минимумов и максимумов (3.9374) внутри индивидуальных природных геномов. При этом характерные величины амплитуды особенностей электростатического потенциала, исследованных для регуляторных элементов природных геномов, составляют от 0.3 до 2-2.5 e/A.

Таким образом, однородные по составу даже в пределах 4 (тем более – 8) соседних пар нуклеотидов последовательности способны сформировать электростатические элементы, по выраженности равные природным регуляторным структурам или превосходящие их. При этом однородность включает не просто процент GC пар, а распространяется на содержание каждого нуклеотида в равной пропорции.

Рассмотрим другой крайний случай, а именно последовательности, содержащие только один или два нуклеотида. Диапазон средних значений для последовательностей, построенных исключительно из A и T (поли-A и поли-AT) равен 2.7959, что в

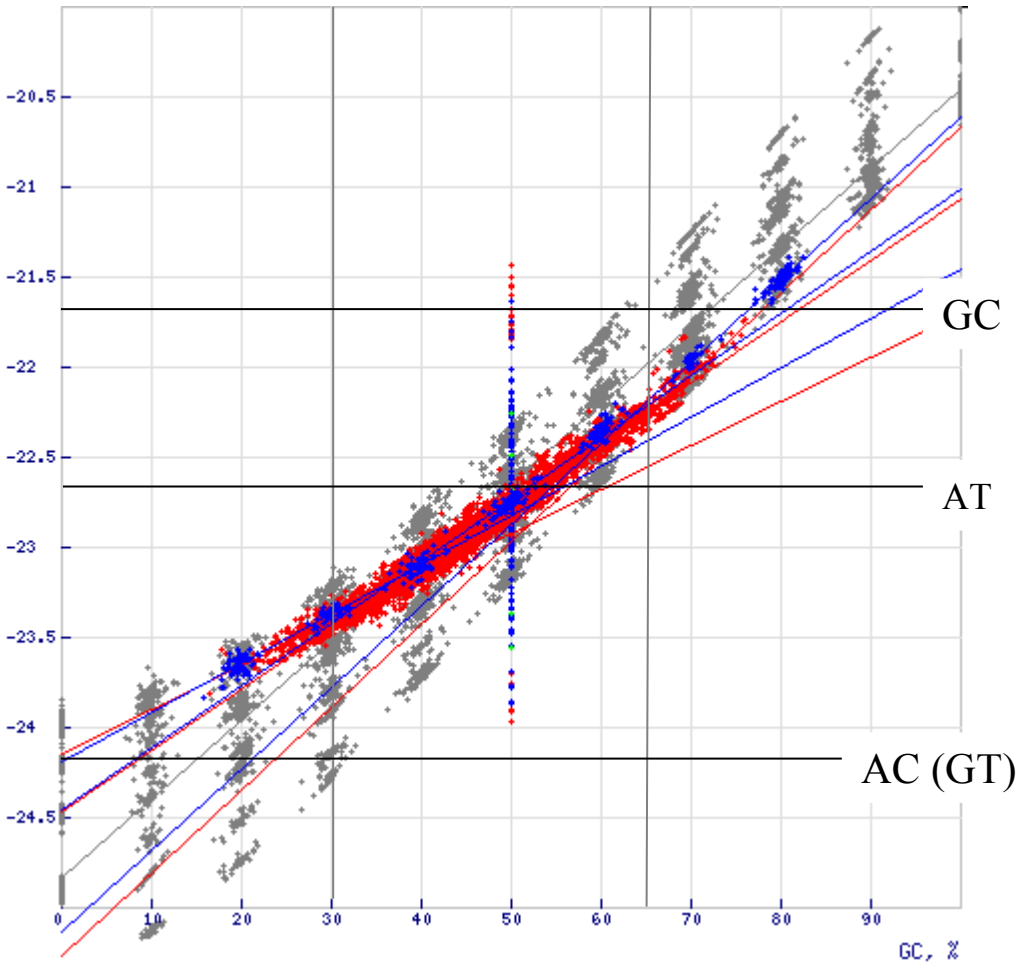

Рисунок 1. Средний электростатический потенциал и содержание GC пар для различных групп последовательностей.
Серый – все случайные последовательности, синий – сбалансированные; красный – природные геномы. Вертикальный ряд точек в положении GC=50% – регулярные последовательности с соотношением A/T/G/C = 1/1/1/1. Каждая точка соответствует одной последовательности.
Вертикальные линии показывают деление на группы по 30 и 65 процентам, горизонтальные – значения потенциала полинуклеотидов указанного состава. Наклонные прямые линии – графики линейной регрессии для соответствующих групп. По правому краю сверху вниз: все случайные, сбалансированные случайные с GC>65%, природные с GC>65%, сбалансированные случайные с 30%<GC<65%, природные с 30%<GC<65%, сбалансированные случайные с GC<30%, природные с GC<30%/
По вертикальной оси – значение среднего потенциала последовательности в e/A, по горизонтальной – содержание GC пар в процентах.

1.4 раза больше, чем для G и C (1.9898), что подтверждает б**о**льшую гибкость в формировании электростатического потенциала AT последовательностями, чем GC. Показательно, что поли-AC и поли-GT (50% GC) имеют значительно больший средний потенциал (-24.1867), чем поли-AT (-22.6485, 0% GC), см. рис.1.

Большая по сравнению с соответствующими случайными последовательностями величина диапазона еще раз указывает на значение «трековости» для формирования потенциала. Рассмотрим этот показатель подробнее на примере полинуклеотидов с длиной повтора 12 п.о. У последовательностей с наименьшим потенциалом нуклеотиды G и C организованы в треки, а A и T – перемежаются, с наибольшим – вся последовательность равномерно перемешана с преобладанием сочетаний AC и GT, что демонстрирует совокупное действие факторов трековости, гибкости формирования

потенциала парами AT и GC, а также большей величины потенциала для поли-AC (поли-GT), чем поли-AT.

Все это демонстрирует недостаточность учета одного лишь содержания GC для анализа формирования электростатических свойств.

ДНК природных геномов

Природные геномы (рис. 1, красные точки, таблица 2) демонстрируют близкую к линейной зависимость среднего значения электростатического потенциала от содержания GC пар с линейными коэффициентами, близкими для разных групп геномов, незначительно выделяющимися у архебактерий.

Разброс значений небольшой, с размахом в 0.5065 e/A (от -0.2249 до 0.2816) и стандартным отклонением 0.0545.

Следует отметить, что распределение геномов по содержанию GC пар несимметрично и имеет сдвиг в область пониженного содержания (среднее 45.0056, стандартное отклонение 10.1139, минимум 17, максимум 76, размах 59, все в процентах GC). При этом для геномов с пониженным содержанием GC пар зависимость среднего значения электростатического потенциала от него выражена менее, чем в среднем, а с повышенным – более, аналогично сбалансированным случайным последовательностям, однако эта закономерность более выражена у природных геномов, незначительно в области высокого содержания GC и сильно – в области низкого. Можно предположить, что в сдвиг распределения природных геномов в AT-богатую область внесли вклад большие возможности формирования выраженных электростатических элементов AT-обогащенными последовательностями по сравнению с GC-обогащенными.

Группа последовательностей	**a**	**b**
Природные геномы		
Бактерии	0.0336	-24.4403
Архебактерии	0.0307	-24.2732
Плазмиды	0.0337	-24.4474
Вирусы	0.0337	-24.4555
Все геномы	0.0336	-24.4480
Сравнение со сбалансированными случайными последовательностями		
Природные с содержанием 30%< GC< 65%	0.0339	-24.4670
Случайные с содержанием 30%< GC< 65%	0.0344	-24.4569
Природные с содержанием GC <30%	0.0245	-24.1549
Случайные с содержанием GC <30%	0.0273	-24.1981
Природные с содержанием 65%< GC	0.0460	-25.2738
Случайные с содержанием 65%< GC	0.0452	-25.1393
Несбалансированные случайные	0.0483	-25.0789
Все случайные	0.0437	-24.8360
Регулярные с n=1, 2	0.0340	-24.5733
Регулярные с n=1, 2, 4, 8, 12	0.0340	-24.5155

Таблица 2. Линейные коэффициенты зависимости среднего значения электростатического потенциала от содержания GC пар для различных групп последовательностей. Уравнение зависимости **y=ax+b**, где **y** – среднее значение потенциала последовательности и **x** - содержание GC пар в процентах. Чем больше **a**, тем больше наклон прямой и сильнее выражена зависимость.

Электростатический потенциал пар A, T, G и C

Электростатический потенциал в центрах пар A, T, G и C анализировался для первых 100000 п.о. генома *E. coli* (GC=51%) и случайной последовательности с соотношением A/T/G/C = 1/1/1/1, а также целого генома бактериофагов T7(GC=48%), phiMFV1 (GC=25%) и mu1/6 (GC=71%). Данные представлены на рис.2. Абсолютные значения максимумов, минимумов и средних выше для пар A и T и ниже для пар G и C в пределах одного организма, однако средние отличаются менее, чем на величину стандартного отклонения (0.2 – 0.5), а у организмов, сильно отличающихся по среднему содержанию GC пар, эти параметры могут быть для пар A и T даже ниже (на > 1), чем для G и C.

Таким образом еще раз демонстрируется крайне слабая зависимость величины потенциала от нуклеотидного состава в точке рассмотрения, недостаточная для формирования электростатических элементов, по выраженности равных природным регуляторным структурам.

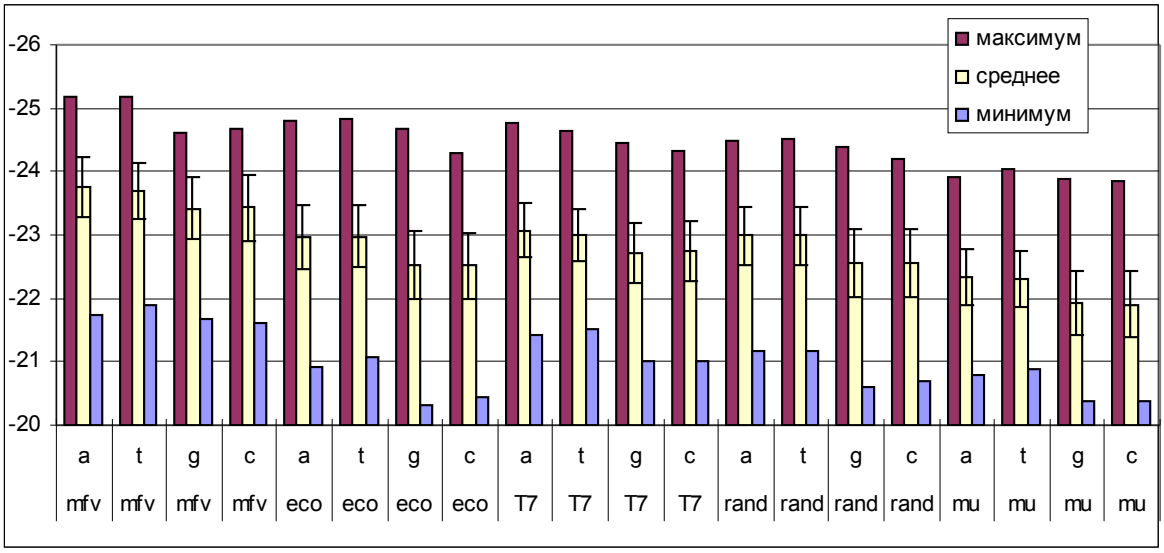

Рисунок 2. Электростатический потенциал в центрах пар A, T, G и C в последовательностях ДНК: максимум, минимум и среднее со стандартным отклонением. Последовательности: геномы бактериофага phiMFV1 (GC=25%), *E. coli* (GC=51%), бактериофага T7 (GC=48%), случайная последовательность с соотношением A/T/G/C = 1/1/1/1, бактериофаг mu1/6 (GC=71%).
По вертикальной оси – значение потенциала в e/A

Зависимость от контекста в природных геномах и случайных последовательностях

Для изучения влияния окружения на формирование электростатического потенциала последовательности, была исследована зависимость от длины фрагмента разброса потенциала в центре разных экземпляров одинаковых фрагментов ДНК. Анализировался электростатический потенциал в центрах фрагментов ДНК длиной от 1 до 40 п.о. для первых 100000 п.о. генома *E. coli* и случайной последовательности с соотношением A/T/G/C = 1/1/1/1, а также целого генома бактериофага T7. Для анализа отбирались фрагменты, представленные не менее чем в трех экземплярах. Данные представлены на рис.3.

Абсолютные значения разницы между максимумами и минимумами для всех экземпляров каждого вида фрагментов и их среднее стандартное отклонение плавно уменьшаются с ростом длины фрагмента. Следует отметить, что в природных последовательностях эти показатели для длинных фрагментов маскируются их включением

в участки больших повторов или дупликаций, что заметно при сравнении величин их размахов.

Показательно, что максимальные разбросы сохраняют величину, достаточную для формирования электростатических элементов, по выраженности равных известным природным регуляторным структурам, вплоть до максимальной исследованной длины в 40 п.о., а до длины 8 п.о. – превосходящих большинство из них.

Анализ консенсусного промотора бактериофага T3 (23 п.о.), помещенного в контекст ряда различных регулярных последовательностей, показал, что средний разброс в его центре составил около 0.2 e/A, а максимальный – около 1.1 e/A, притом что величина между максимумом и минимумом усредненного профиля реальных промоторов T3 в контексте его генома составляет около 1 e/A, средний разброс в центре – 0.15, а максимальный – 0.5, то есть в два раза меньше, несмотря на имеющиеся отличия этих промоторов от консенсуса. Хотя в среднем промотор сохраняет свою характерную картину распределения электростатического потенциала, в ряде случаев она меняется кардинально, гораздо более, чем необходимо для потери узнавания нативной РНК-полимеразой (~ 0.5 e/A от среднего). Показательно, что в данном случае средний и максимальный разброс не уменьшаются равномерно к центру, а имеют минимум в точке старта (6 п.о. от края), и небольшой локальный максимум в области -3 п.о. от точки старта, что говорит о сложном характере влияния окружения на потенциал последовательности.

Таким образом, окружение последовательности длиной более, чем длина консервативных участков известных регуляторных элементов, способно сформировать в ней электростатические элементы, по выраженности равные природным регуляторным структурам или превосходящие их, или помешать формированию таких элементов.

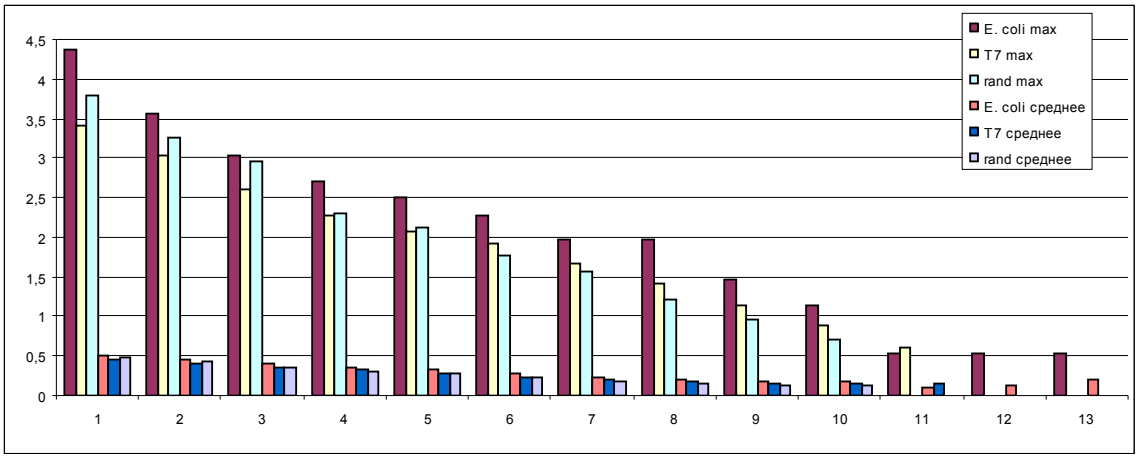

Рисунок 3. Разброс значений электростатического потенциала в центрах одинаковых фрагментов ДНК длиной **n** от 1 до 40 п.о.
Столбцы: 1-10 – n=1-10, 11-13 – n=20, 30, 40.
Последовательности: геномы *E. coli*, бактериофага T7 и случайная последовательность с соотношением A/T/G/C = 1/1/1/1; большие столбцы – максимальный разброс, малые – среднее стандартное отклонение.
По вертикальной оси – значение потенциала в e/A.

3. Особенности электростатических свойств промоторов ряда T7-подобных фагов и рибосомальных промоторов E.coli

Рассмотрим особенности организации генома T7-подобных фагов на примере бактериофага T7, заражающего *E. coli.* Во время инфекции *E. coli* ранняя область T7 генома транскрибируется хозяйской РНК-полимеразой (Eσ70) с трех тандемно распо-

ложенных на левом конце сильных промоторов A1, A2, A3. Одним из основных генных продуктов этой области является фаговая РНК-полимераза, которая осуществляет транскрипцию средних (класс II) и поздних (класс III) генов Т7-ДНК.

Известно, что более 20% из всех ~4000 промоторов *E. coli* также расположены тандемно, в частности, таковы сильные рибосомальные промоторы. Такая организация промоторной зоны, по-видимому, способствует повышению надежности системы распознавания генетических элементов, особенно важных для организма.

Следует отметить, что в отличие от мультисубъединичной РНК-полимеразы *E. coli*, являющейся одним из самых больших бактериальных белков, Т7-специфичный фермент состоит из одной небольшой субъединицы. Соответственно и промоторы, нативные к этим двум ферментам, отличаются прежде всего по своим размерам. Если для РНК-полимеразы *E. coli* контактная промоторная площадка составляет >150 п.о. (~510 ангстрем), то для Т7-специфичного фермента она равна 23 нуклеотидным парам (~80 ангстрем), к тому же находящимся в составе нуклеотидспецифичного консенсусного элемента. Это априори указывает на принципиальное отличие в характере электростатических взаимодействий при узнавании нативных промоторов этими двумя ферментами.

Ранние промоторы Т7-подобных фагов и рибосомальные промоторы *E.coli*

Результаты, полученные при анализе особенностей электростатического потенциала промоторных участков бактериофага Т7, показали наличие выраженных неоднородностей профиля в районе группы ранних промоторов, взаимодействующих с бактериальной РНК-полимеразой организма хозяина на ранних этапах фагового заражения, проявляющихся в виде характерной серии изменений потенциала с большой амплитудой при малой частоте, с наиболее ярко выраженной первой волной, где каждая волна соответствует сильному промотору (рис. 4.а.). Такой характер электростатического профиля промоторной зоны, по-видимому, служит для повышения надежности узнавания промоторов хозяйской полимеразой, что, в свою очередь, повышает шансы успешности фаговой инфекции. Изменения потенциала находятся в одном масштабе с контактной площадкой молекулы бактериальной РНК-полимеразы (~500 ангстрем).

Анализ ранних областей геномов группы Т7-подобных фагов Т3, phiA1122, phiYeO3-12 (рис. 4.а.), K1-5 и SP6 (рис. 4.б.) показал у них наличие точно такой же картины, причем для фагов Т7 и Т3 в базе данных NCBI RefSeq, откуда бралась исходно биологическая аннотация, указано наличие промоторов бактерии-хозяина, для фага phiYeO3-12 оно указывалось как предположительное, а для фагов K1-5, SP6 и phiA1122 такой информации не было вовсе. Тем не менее, для них для всех было обнаружено поразительное сходство картины распределения электростатического потенциала, что позволяет предположить, что данные промоторы там присутствуют и выполняют свою биологическую функцию.

Интересно сравнить профиль этих областей с районами рибосомальных промоторов *E.coli*, для которых характерно наличие тандема из двух сильных промоторов и перед которыми также стоит задача максимизации надежности их узнавания. Хорошо видно, что профили этих областей имеют между собой определенное сходство (рис. 4.в.). Это может отражать общность их биологических функций.

Следует отметить, что нуклеотидные последовательности всех этих районов значительно различаются между собой, что указывает на важность анализа физических свойств в дополнении к традиционному текстовому анализу нуклеотидной последовательности.

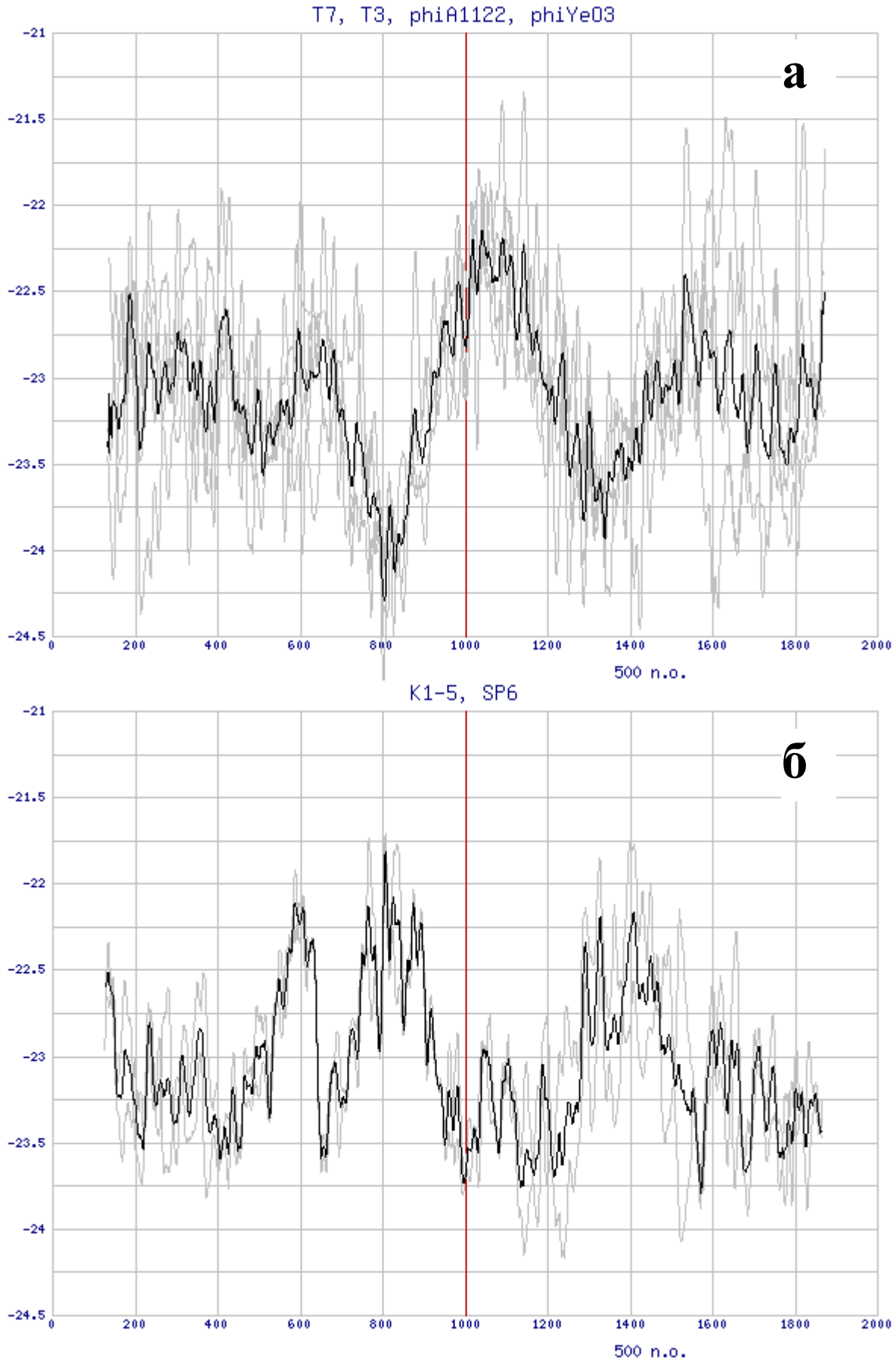

Рисунок 4 а, б. Электростатические профили ранних промоторов фагов T7, T3, phiA1122 и phiYeO3-12 (а) и K1-5 и SP6 (б), взаимодействующих с бактериальной РНК-полимеразой. Серый – индивидуальные профили, черный – усредненные, длина участков – 500 п.о. По вертикальной оси – величина ЭП в е/А, по горизонтальной – расстояние вдоль оси молекулы ДНК в ангстремах.

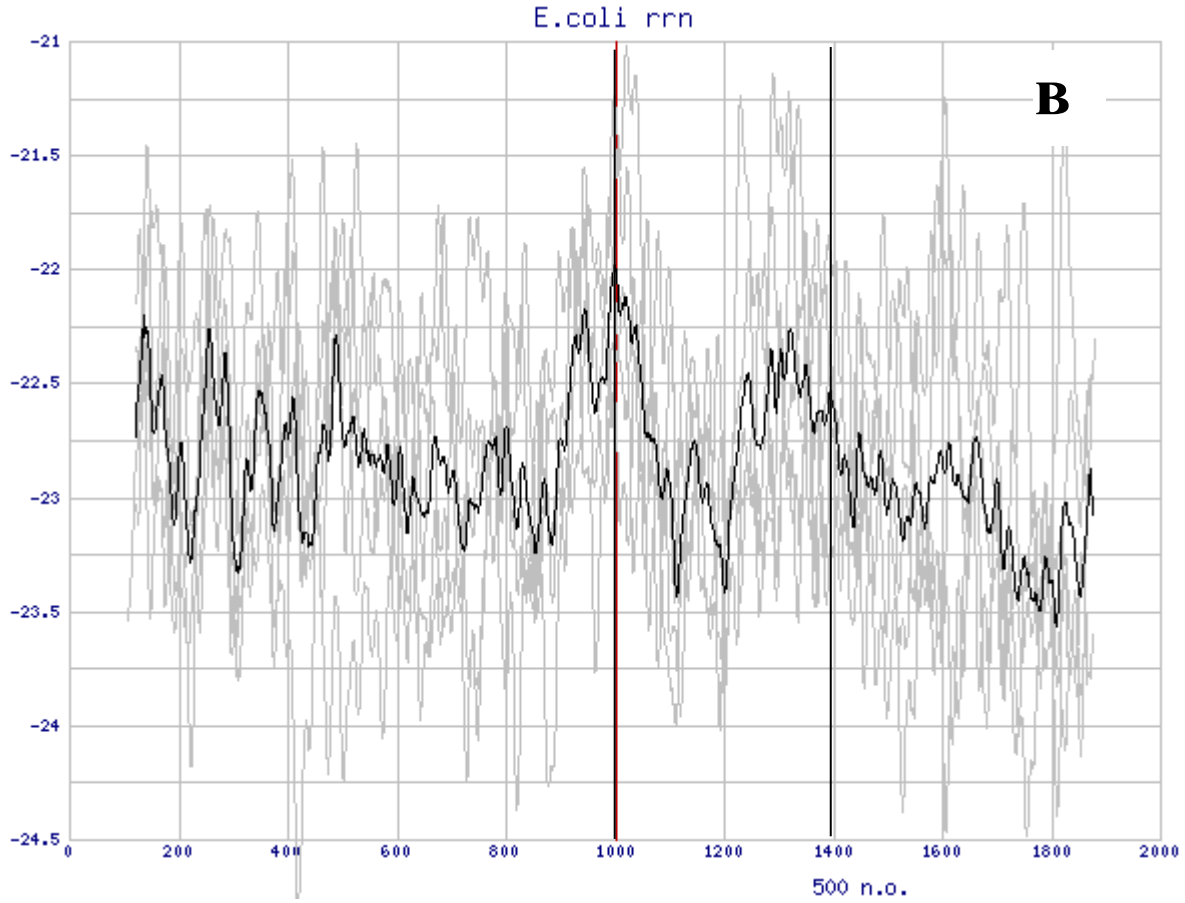

Рисунок 4 в. Электростатические профили тандемных рибосомальных промоторов *E. coli*, вертикальными линиями отмечены положения первых и вторых промоторов тандемных пар.
Серый – индивидуальные профили, черный – усредненные, длина участков – 500п.о.
По вертикальной оси – величина ЭП в e/A, по горизонтальной – расстояние вдоль оси молекулы ДНК в ангстремах.

Промоторы Т7-подобных фагов, взаимодействующие с фаговой РНК-полимеразой

Общая характеристика ЭП промоторов фагов Т7, Т3, phiYeO3-12 и VP4

При анализе распределения электростатического потенциала вокруг промоторов, взаимодействующих с нативными фаговыми РНК-полимеразами, выявляется общая картина сходства, выражающегося в наличие синхронизированных неоднородностей (подъемов и спадов) одного масштаба с молекулой полимеразы и, в частности, в переходе графика потенциала от спада непосредственно перед точкой старта к его подъему сразу за ней, а для фагов Т3, phiYeO3-12, VP4 - двух волн таких переходов (рис. 5). Наряду со сходством, видно различие однородности характеристической картины для разных фагов, от наиболее однородного фага Т3 до наименее – Т7.

На примере фагов Т3 и phiYeO3-12 (рис. 6) хорошо видно различие в устойчивости картины электростатического профиля к нуклеотидным заменам на разных его участках. В районе точки старта профиль устойчив к присутствующим там единичным и двойным заменам, в районе первого upstream пика (-30 ангстрем от точки старта) – гиперчувствителен даже к единичной замене С на А, которая совершенно элиминируют этот пик (рис. 6, линия 2). Это указывают на необходимость дополнения тестового анализа нуклеотидных последовательностей анализом физических свойств ДНК.

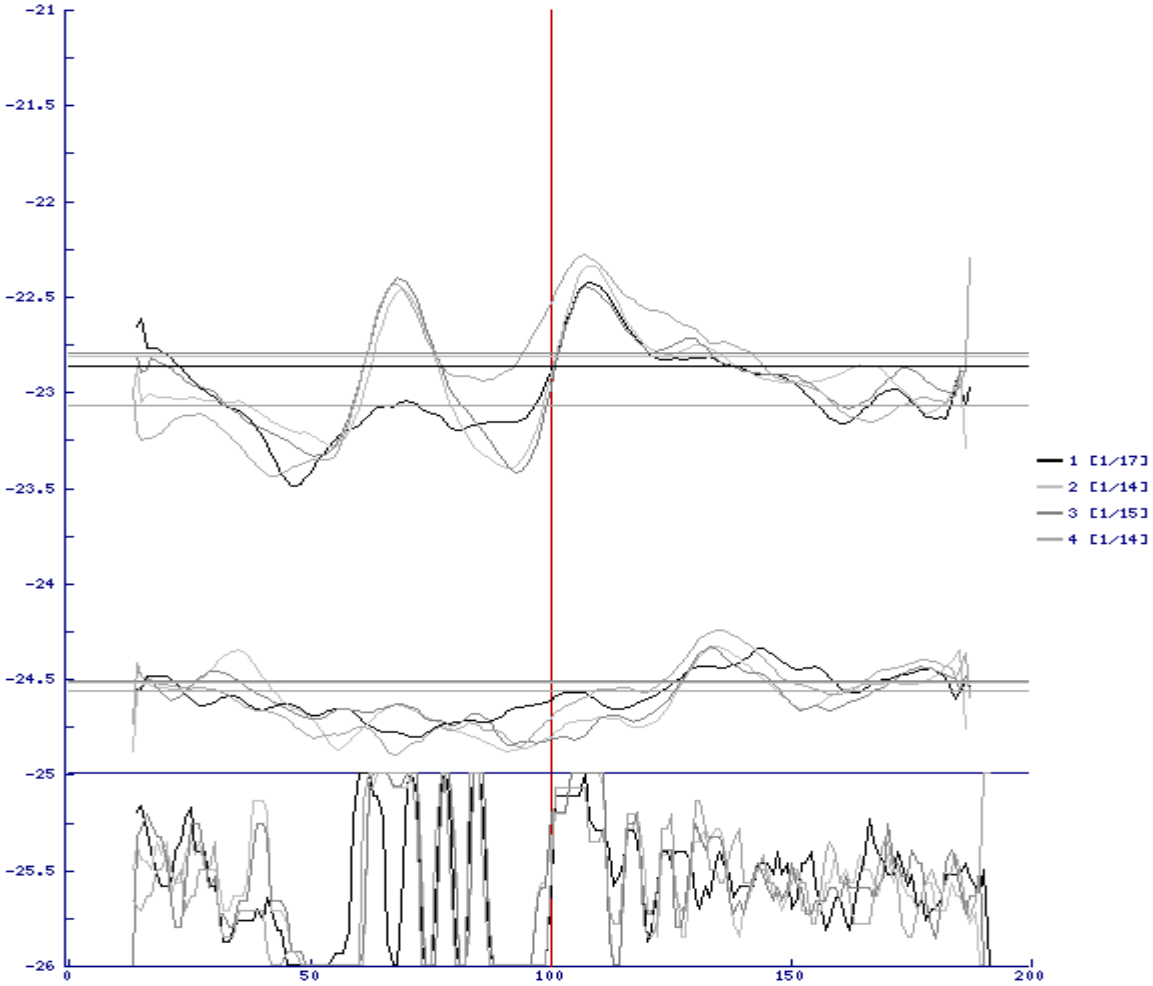

Рисунок 5. Усредненные ЭП нативных промоторов бактериофагов T7 (1), T3 (2), phiYeO3-12 (3) и VP4 (4). Длина участков – 50 п.о., окно GC состава – 1 п.о.
По вертикальной оси – величина ЭП в e/A, по горизонтальной – расстояние вдоль оси молекулы ДНК в ангстремах.

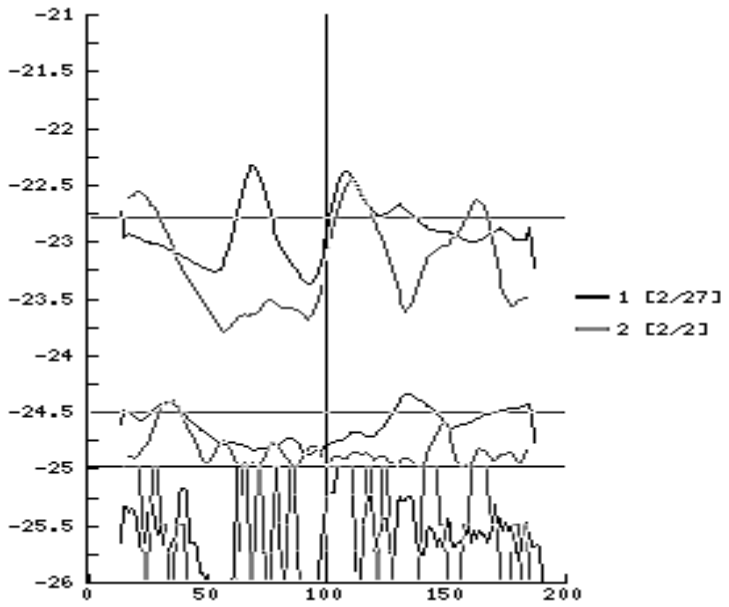

Рисунок 6. Усредненные ЭП промоторов с C (1) и A (2) в -10 позиции. Длина участков – 50 п.о., окно GC состава – 1 п.о.
По вертикальной оси – величина ЭП в e/A, по горизонтальной – расстояние вдоль оси молекулы ДНК в ангстремах.

Промоторы мутантного штамма бактериофага Т7, приспособленного к РНК-полимеразе бактериофага Т3

Для выявления электростатических элементов, потенциально могущих иметь значения для функционирования промоторов, распознающихся фаговыми РНК-полимеразами, был выбран мутантный штамм бактериофага Т7, приспособившийся к росту на РНК-полимеразе родственного бактериофага Т3 (J.J. Bull et al., 2007).

Как известно, РНК-полимеразы этих фагов крайне слабо взаимодействуют с промоторами друг друга, тем не менее, рост фага Т7 с делецией гена РНК-полимеразы возможен в клетках *E. coli*, экспрессирующих РНК-полимеразу фага Т3. В эксперименте с такой системой было показано (J.J. Bull et al., 2007), что при исключении возможности приспособительных мутаций в гене РНК-полимеразы происходит постепенное восстановление жизнеспособности (рис. 7) мутантного фага за счет накопления мутаций в промоторных регионах.

Авторами была произведена делеция области 3343-5878 генома бактериофага Т7, включающая ген РНК-полимеразы и ранний промотор phi1.1A (рис. 9).

В ходе приспособления к росту на РНК-полимеразе бактериофага Т3, у мутантного штамма бактериофага Т7 произошли следующие дополнительные мутации (таблица 3), затрагивающие промоторные области (следует обратить внимание на переход к консенсусному промотору бактериофага Т3):

1. промотор репликации (phiOL) – в позиции 394 А –> С и делеция Т в позиции 403 (обе – к консенсусу Т3);
2. промотор класса II phi1.5 – в позиции 7768 А –> С (к консенсусу Т3);
3. промотор класса II phi1.6 – в позиции 7884 Г –> А (не к консенсусу Т3);
4. промотор класса II phi2.5 – в позиции 9105 Т –> А (к консенсусу Т3);
5. промотор класса III phi6.5 – в позиции 18534 Г –> А (не к консенсусу Т3) и в позиции 18543 Т –> А (к консенсусу Т3);
6. промотор класса III phi9 – в позиции 21863 Т –> А (к консенсусу Т3);
7. промотор класса III phi10 – в позиции 22893 Г –> А (не к консенсусу Т3) и в позиции 22902 Т –> А (к консенсусу Т3);
8. промотор класса III phi13 – в позиции 27265 С –> Т (*от консенсуса Т3*);
9. промотор репликации (phiOR) – в позиции 39218 Г –> А (не к консенсусу Т3) и в позиции 39227 Т –> А (к консенсусу Т3).

Таким образом, в промоторных областях возникло 13 мутаций (рис. 10, 11), только 7 из которых являются переходом к консенсусному промотору бактериофага Т3, а одна из оставшихся 6 – переходом от консенсусного к неконсенсусному. При этом в известной своей консервативности и считающейся важной для различения промоторов полимеразами Т3 и Т7 позиции -11 от точки старта только одна из 6 мутаций была в сторону консенсуса Т3.

Следует отметить, что полученный мутант хорошо рос на нативной полимеразе Т7.

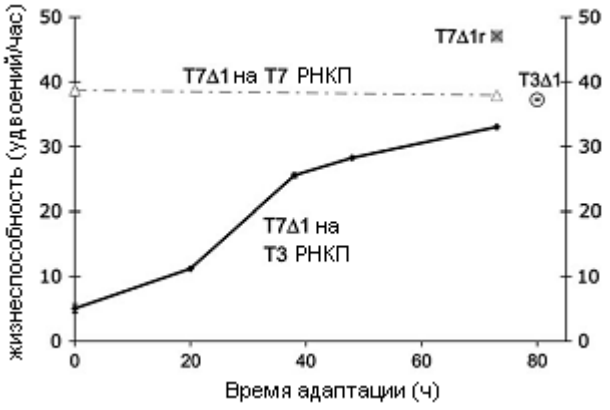

Рисунок 7. Восстановление жизнеспособности мутантного фага во время эксперимента (из J.J. Bull et al., 2007, с изменениями)

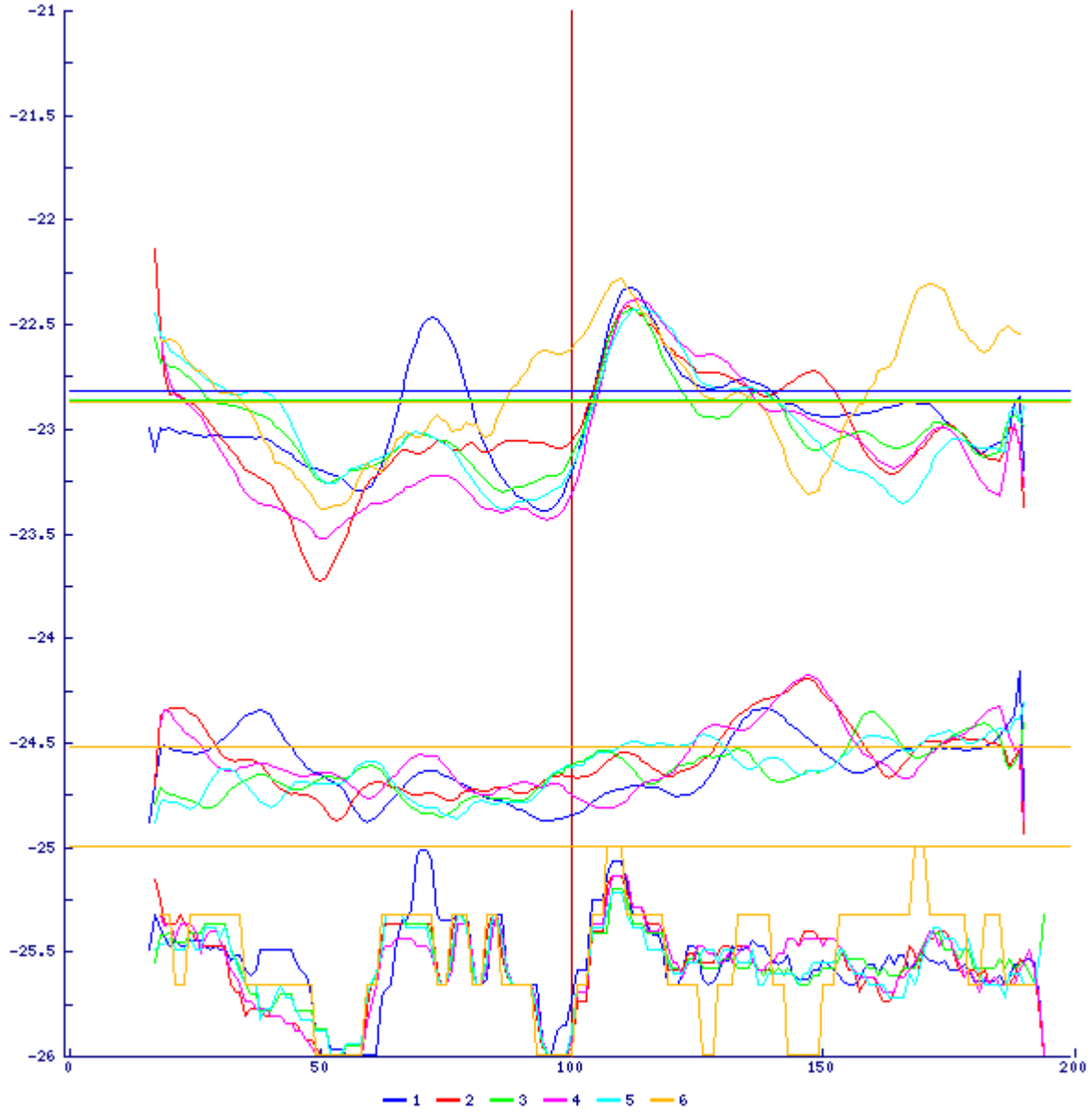

Рисунок 8. Графики электростатического потенциала вокруг промоторов бактериофага T7 и его мутанта и бактериофага T3, усредненные по следующим группам:
1. Промоторы бактериофага T3 (14 шт.)
2. Промоторы бактериофага T7, мутировавшие в ходе эксперимента (9 шт.)
3. Промоторы бактериофага T7, не мутировавшие в ходе эксперимента (7 шт.)
4. Промоторы мутанта T7, мутировавшие в ходе эксперимента (9 шт.)
5. Промоторы мутанта T7, не мутировавшие в ходе эксперимента (без phi1.3) (6 шт.)
6. Промотор phi1.3 мутанта T7

По вертикали:
Верхняя часть графика – электростатический потенциал в e/A, горизонтальные линии – среднее значение потенциала всего генома исследованных фагов
Средняя часть графика – стандартные отклонения для каждой группы, горизонтальные линии – среднее значение для каждого генома исследованных фагов
Нижняя часть графика – содержание GC пар в процентах для каждой группы, усреднение в окне по 3 пары вокруг каждой точки
По горизонтали:
Расстояние вдоль последовательности 50 п.о. вокруг точки старта (+1, вертикальная линия), в ангстремах
Цвет каждого графика соответствует своей группе.

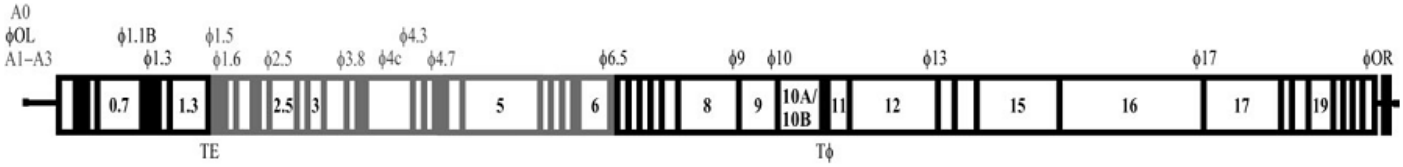

Рисунок 9. Диаграмма генома мутантного штамма бактериофага T7 (из J.J. Bull et al., 2007).

Элемент	Функция	Мутация
		ins #1 340
øOL	промотор репликации	394 A→C
		del 403 T
0.6B	незначимая	1878 A→G Q81Q
1[b]	РНК полимераза	del 3343-5878
ø1.5	промотор класса II	7768 A→C
ø1.6	промотор класса II	7884 G→A
1.7	незначимая	8335 A→C N57T
ø2.5	промотор класса II	9105 T→A
2.5	1нДНК-связывающий белок	9529 A→C H124P
3.5	лизоцим	10736 T→C S11P
5	ДНК полимераза	14811 A→C E153D
ø6.5	промотор класса III	18534 G→A
		18543 T→A
7.3	инициация инфекции	19632 G→A G33D
8	соединитель головки и хвоста	21749 G→A G504S
ø9	промотор класса III	21863 T→A
ø10	промотор класса III	22893 G→A
		22902 T→A
*10B**	малый капсидный белок	24088 G→A E375K*
Tø*	терминатор фаговой РНАП	ins 24200 G*
*12**	белок хвоста	26586 C→T A582V*
ø13	промотор класса III	27265 C→T
*13**	внутренний белок головки	27706 A→C I134L*
*14**	внутренний коровый белок	27782 G→A G19S*
*16**	внутренний коровый белок	33068 A→C K825T*
		34042 T→C F1150L*
*17.5**	холин	36414 T→C V24A*
19	терминаза	37715 C→T P116S
IG*19.5*		38793 A→G
øOR	промотор репликации	39218 G→A
		39227 T→A

Таблица 3. Список всех мутаций исследуемого штамма бактериофага T7. Позиции нуклеотидов указаны для дикого штамма T7 (Genbank V01146). Звездочками помечены мутации не строго компенсаторные для взаимодействия с РНАП T3. Также указана начальная делеция гена 1 (РНАП). Вставка ins #1 после позиции 340 составляет ACTACATAAAGACCAGACCTAAAGAC.
(из J.J. Bull et al., 2007, с изменениями)

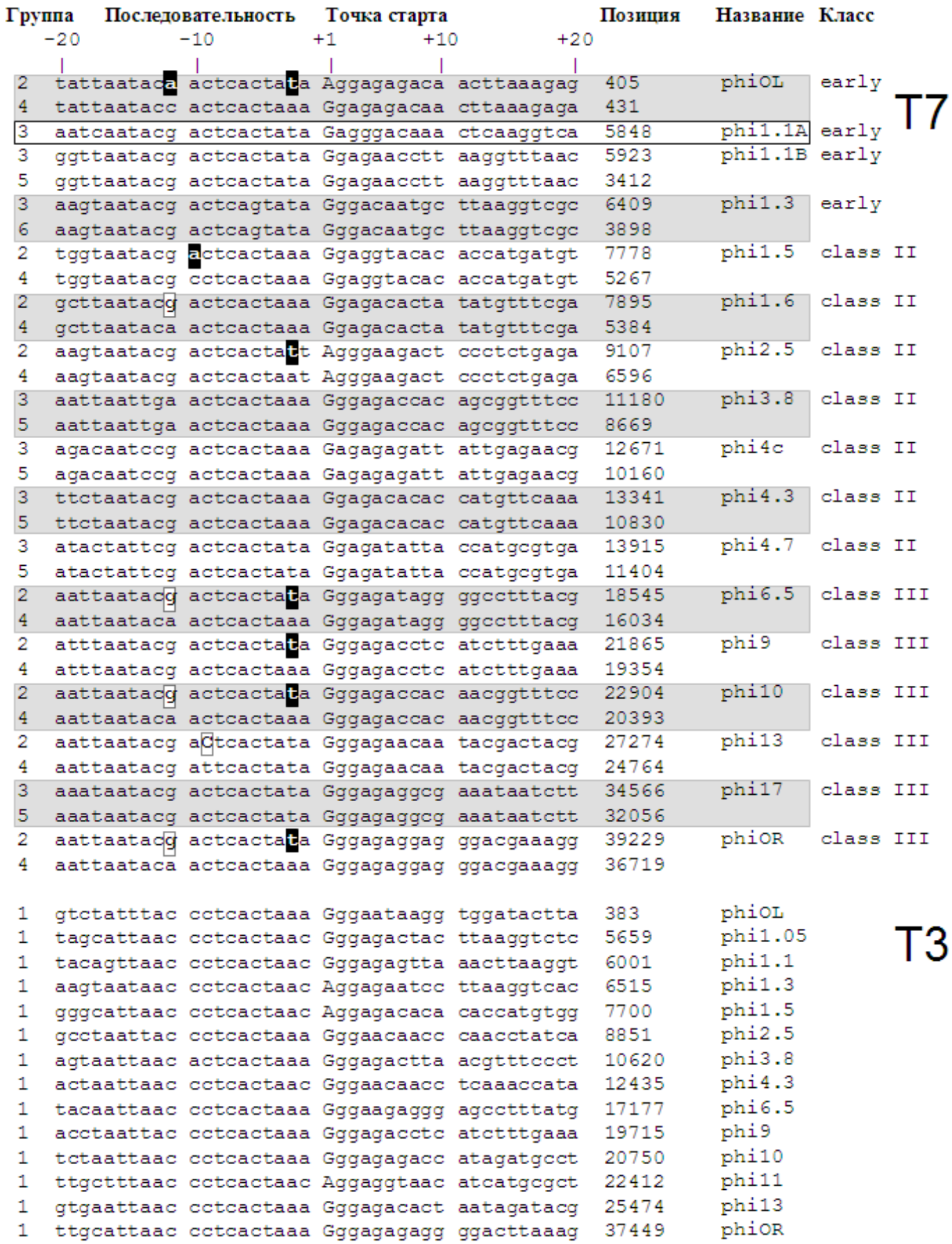

Рисунок 10. Последовательности промоторов бактериофага T7 и его мутанта (попарно), и бактериофага T3.
Прямоугольниками выделены мутации, черными – к консенсусу T3

Рисунок 11. Логотип промоторов дикого типа бактериофагов T7 и T3.
Величина букв пропорциональна частоте встречаемости данного нуклеотида в обозначенной позиции.
Под ним указано количество мутаций в этой позиции, в верхней строке – общее, в нижней – к консенсусу T3 (из J.J. Bull et al., 2007).

Появившиеся в ходе эксперимента мутации в промоторах T7 не могут быть полностью объяснены переходом к консенсусному по тексту последовательности промотору для РНАП T3, поэтому мы произвели анализ свойств электростатического потенциала вокруг промоторов бактериофага T3, бактериофага T7 и его мутанта, полученного в ходе эксперимента. Результаты анализа отображены на рис. 8.

Основные отличия электростатических профилей промоторов бактериофага T7, с одной стороны, и промоторов бактериофага T3, а также мутантного штамма T7, с другой, находятся в области точки старта. На графике (рис. 8) хорошо видно, что в районе -5 – -15 ангстрем от точки старта, что соответствует примерно -2 – -5 парам оснований, электростатический профиль промоторов бактериофага T7 значительно выше, чем у T3 и мутантного штамма (см. также рис. 5, линии 1 и 2).

Мутировавшая группа промоторов (рис. 8, линия 2) демонстрирует наибольшее отклонение потенциала, вне зависимости от позиции мутации. Показательно, что промоторы, не подвергшиеся мутациям, изначально (рис. 8, линия 3) демонстрировали профиль, близкий к профилю T3, однако они также изменили профиль в сторону соответствия T3, хотя и весьма незначительно (рис. 8, линия 5), несмотря на отсутствие мутаций в самих этих промоторах, что показывает влияние на электростатические свойства последовательности ее окружения.

Особняком стоит не мутировавший ранний промотор phi1.3 (рис. 8, линия 6). Картина его электростатического профиля изменяется в сторону от общей, что приводит к ухудшению показателей соответствующей группы. Однако следует отметить, что его роль в жизнедеятельности фага, по-видимому, незначительна, так как он расположен в самом конце ранней области генома T7, как известно находящейся под контролем сильных промоторов бактериальной РНК-полимеразы, и считывающейся

ею в составе одного транскрипта, что допускает возможность выделения его в отдельную группу.

Принимая во внимание интегральный показатель активности промоторов мутировавшего штамма, выраженный не в их силе непосредственно, а в общем влиянии на жизнеспособность, можно тем не менее сделать вывод, что дифференциальное распознавание промоторов РНК-полимеразой бактериофага Т3 возможно зависит от характеристик электростатического потенциала в районе -5 – -15 ангстрем от точки старта, что соответствует примерно -2 – -5 парам оснований, при этом электростатический потенциал должен быть достаточно высок, в то время как указанные отличия потенциала мало влияют на распознавание промоторов РНК-полимеразой бактериофага Т7, но, по-видимому, играет для нее регуляторную роль.

Выводы

1. Разработана база данных электростатических свойств ДНК природных геномов («DEPPDB» - DNA Electrostatic Potential Properties DataBase), содержащая интегрированные данные о последовательности, биологическую аннотацию, таксономическое положение и электростатические свойства всех полностью секвенированных бактериальных и вирусных геномах, а также ряда расчетных случайных и регулярных последовательностей.
2. На основе DEPPDB разработаны инструменты, позволяющие проводить сравнительный, функциональный и эволюционный анализ свойств электростатического потенциала генома и его элементов на уровне как отдельных геномов, так и целых таксономических групп.
3. Обнаружена близкая к линейной зависимость среднего потенциала природных геномов от содержания в них GC пар. Установлено, что величина этой зависимости коррелирует с содержанием GC пар.
4. Изучены закономерности формирования электростатического потенциала вокруг молекулы ДНК. Выявлена сложная зависимость распределения потенциала от состава и организации ее последовательности. Показано влияние окружающих последовательностей на формирование локального потенциала в области рассмотрения.
5. Высказана гипотеза, что в сдвиг распределения природных геномов в АТ-богатую область могли внести вклад большие возможности формирования выраженных электростатических элементов АТ-обогащенными последовательностями по сравнению с GC-обогащенными.
6. Показано различие в масштабах, на которых проявляются закономерности распределения потенциала для промоторов, взаимодействующих с бактериальными и фаговыми полимеразами, что отражает физическую картину взаимодействия ДНК с белком.
7. Показано, что приспособление промоторов бактериофага T7 к взаимодействию с РНК-полимеразой бактериофага T3 сопровождается изменением электростатического потенциала в районе 0 – -5 п.о., приводящим к формированию профиля, идентичного промоторам T3, что свидетельствует о возможной зависимости от него дифференциального распознавание промоторов РНК-полимеразой T3. При этом указанные отличия потенциала мало влияют на распознавание промоторов РНК-полимеразой T7, но играют для нее регуляторную роль.

Я выражаю глубокую благодарность Светлане Григорьевне Камзоловой за чуткое научное руководство, понимание и всестороннюю поддержку, Анатолию Александровичу Сорокину за саму неоценимую возможность работать над данной темой, а также искреннюю признательность Элеоноре Григорьевне Савельевой за постоянную помощь в работе, Тимуру Рустемовичу Джелядину за ценные советы при подготовке диссертации и всем моим коллегам за плодотворные дискуссии и критические замечания, а жене, друзьям и детям– за поддержку.

А еще – спасибо моему деду и моей маме – без них ничего бы не было.

Список работ, опубликованных по теме диссертации

публикации в рецензируемых журналах:

1. Osypov A.A., Beskaravainy P.M. Sorokin A.A., Kamzolova S.G. Electrostatic Potential Map of the Whole Genome DNA of T7 Bacteriophage. Electrostatic Properties and Function of its Promoter Regions. J. Biomol. Struct. Dyn, 2007, 24(6), p. 714-715
2. Sorokin A.A., Osypov A.A., Kamzolova S.G. Comparative Analysis of Electrostatic and Functional Properties of Some Synthetic A-Tracts Containing Promoters. J. Biomol. Struct. Dyn, 2007, 24(6), p.657-658
3. Камзолова С.Г., Осипов А.А., Бескаравайный П.М., Джелядин Т.Р., Сорокин А.А., Регуляция активности промоторной ДНК через электростатические взаимодействия с РНК-полимеразой. Биофизика, 2007, 52(2), с.228-236.
4. Сорокин А.А., Осипов А.А., Бескаравайный П.М., Камзолова С.Г., Анализ распределения нуклеотидной последовательности и электростатического потенциала генома E.coli. Биофизика, 2007, 52(2), с.223-227.
5. Камзолова С.Г., Бескаравайный П.М., Осипов А.А., Джелядин Т.Р., Сорокин А.А., Сравнительный анализ электростатических и функциональных свойств σ70-специфичных промоторов, содержащих олиго dA-последовательности. Вестник биотехнологии им. Овчинникова, 2006, 2(1), с.5-10.
6. Sorokin A.A., Osypov A.A., Dzhelyadin T.R., Beskaravainy P.M., Kamzolova S.G., Electrostatic properties of promoters recognized by E.coli RNA polymerase Eσ70 J. Bioinf. Comp. Biol., 2006, vol.4, no.2, p.455-467.
7. Sorokin A.A., Osypov A.A., Beskaravainy P.M., Kamzolova S.G., Oligonucleotide analysis of E.coli promoters recognized by σ70-RNA polymerase., J.Biomol.Struct.Dyn., 2005, v.22(6), p.821.
8. Камзолова С.Г., Сорокин А.А., Осипов А.А., Бескаравайный П.М., Общие закономерности формирования σ70-специфичных промоторов в геноме E.coli на основе электростатических характеристик промоторной ДНК, Биофизика, 2005, 50(3), с.444-449.
9. Kamzolova S.G., Sorokin A.A., Dzhelyadin T.R., Beskaravainy P.M., Osypov A.A., Electrostatic potentials of E.coli genome DNA., J.Biomol.Struct.Dyn., 2005, v.23(3), p.341-346.
10. Sorokin A.A., Beskaravainy P.M., Osypov A.A., Kamzolova S.G., Electrostatic map of E.coli genome DNA. Specific features of electrostatic potential of promoter and nonpromoter regions., J.Biomol.Struct.Dyn., 2005, v.22(6), p.791-792.
11. Камзолова С.Г., Сорокин А.А., Осипов А.А., Бескаравайный П.М., Сравнительный анализ электростатических и функциональных свойств промоторов Т7 ДНК, взаимодействующих с РНК-полимеразой E coli. Вестник биотехнологии и физико-химической биологии им. Ю.А.Овчинникова, т.4 №1, стр. 5-13, 2008
12. Камзолова С.Г., Сорокин А.А., Осипов А.А., Бескаравайный П.М., Электростатическая карта генома бактериофага Т7. 1. Сравнительный анализ электростатических свойств σ70-специфических промоторов Т7 ДНК, взаимодействующих с РНК-полимеразой E.coli., Биофизика, 2008, (в печати).
13. Камзолова С.Г., Бескаравайный П.М., Осипов А.А., Сорокин А.А., Электростатическая карта генома бактериофага Т7. 2. Сравнительный анализ электростатических свойств промоторов Т7 ДНК, контролируемых Т7 РНК-полимеразой., Биофизика, 2008, (в печати).

раздел в монографии:

1. Kamzolova S.G., Sorokin A.A., Beskaravainy P.M., Osypov A.A., Comparative analysis of electrostatic patterns for promoter and non promoter DNA in E.coli. In: Bioinformatics of Genome Regulation and Structure II. Eds. N.Kolchanov and Hofestaedt. Springer Science Business Media Inc., 2005, p.67-74.

статьи в научных сборниках и периодических научных изданиях:

1. Kamzolova S.G., Osypov A.A., Dzhelyadin T.R., Beskaravainy P.M., Sorokin A.A., Context-Dependent Effects of Upstream A-Tracts on Promoter Electrostatic Properties and Function, Proceedings of the fifth international conference on bioinformatics of genome regulation and structure, BGRS-2006, Vol.1, p.56-60, 2006.
2. Kamzolova S.G., Sorokin A.A., Dzhelyadin T.R., Beskaravainy P.M., Osypov A.A. Electrostatic properties of E.coli genome DNA. Proceedings of the 4th international conference of bioinformatics of genome regulation and structure BGRS-2004, Vol.1, p. 80-83, 2004.
3. Kamzolova S.G., Sorokin A.A., Dzhelyadin T.R., Osypov A.A., Beskaravainy P.M. Analysis of oligonucleotide composition in DNA of E.coli genome and promoter sites. Proceedings of the 4th international conference of bioinformatics of genome regulation and structure, BGRS-2004, Vol.1, p. 77-79, 2004.

публикации в материалах научных мероприятий:

1. Осипов А.А., Бескаравайный П.М., Джелядин Т.Р., Камзолова С.Г., Сорокин А.А., Электростатические свойства промоторов ДНК мутантного штамма бактериофага Т7, приспособившегося к РНК-полимеразе Т3. 16 Международная конференция "Математика. Компьютер. Образование", Пущино, 19-24 января 2009 г. Тезисы, вып.16, ч.1, стр.277
2. Бескаравайный П.М., Осипов А.А., Джелядин Т.Р., Камзолова С.Г., Сорокин А.А., Электростатические свойства промоторов Т7 ДНК, взаимодействующих с РНК-полимеразой E.coli. 16 Международная конференция "Математика. Компьютер. Образование", Пущино, 19-24 января 2009 г. Тезисы, вып.16, ч.1, стр.233
3. Джелядин Т.Р., Бескаравайный П.М., Осипов А.А., Камзолова С.Г., Сорокин А.А., Сравнительный анализ электростатических и функциональных свойств промоторов Т7 ДНК, контролируемых Т7 РНК-полимеразой. 16 Международная конференция "Математика. Компьютер. Образование", Пущино, 19-24 января 2009 г. Тезисы, вып.16, ч.1, стр.241
4. С.Г.Камзолова, А.А.Осипов, П.М.Бескаравайный, А.А.Сорокин, Сравнительный анализ электростатических свойств σ70-специфичных промоторов Т7 ДНК, взаимодействующих с РНК-полимеразой E.coli. В сборнике тезисов XIV Симпозиума по межмолекулярному взаимодействию и конформациям молекул. (15-21 июня 2008 г., Челябинск), стр. 141 (Б-43).
5. А.А.Осипов, С.Г.Камзолова, П.М.Бескаравайный, Исследование промоторных областей группы Т7-подобных фагов. В сборнике тезисов XIV Симпозиума по межмолекулярному взаимодействию и конформациям молекул. (15-21 июня 2008 г., Челябинск), стр. 142 (Б-44).
6. А.А.Осипов, С.Г.Камзолова, П.М.Бескаравайный, Анализ электростаических свойств Т3 ДНК, контролируемых Т3 РНК-полимеразой и сравнение с Т7 фагом. В сборнике тезисов XIV Симпозиума по межмолекулярному взаимодействию и конформациям молекул. (15-21 июня 2008 г., Челябинск), стр. 143 .

7. Osypov, A. A., Beskaravainy, P. M., Sorokin, A. A., Kamzolova, S. G. DEPPDB - DNA electrostatic potential properties database and its use in functional, comparative and evolutionary genomics, International Conference on Computational Phylogenetics and Genosystematics, conference proceedings, p.1-3, 2007
8. Осипов А.А., Бескаравайный П.М., Сорокин А.А, Камзолова С.Г. База данных свойств электростатического потенциала ДНК. 11 Международная Пущинская школа-конференция молодых ученых «Биология наука XXI века», Тезисы, стр. 13, 2007.
9. Бескаравайный П.М., Осипов А.А., Сорокин А.А, Камзолова С.Г Электростатическая карта генома бактериофага Т7. 11 Международная Пущинская школа-конференция молодых ученых «Биология наука XXI века», Тезисы, стр.6, 2007.
10. Осипов А.А., Панюков В.В. Вычислительный подход к анализу свойств профиля электростатического потенциала ДНК. 11 Международная Пущинская школа-конференция молодых ученых «Биология наука XXI века», Тезисы, стр. 59-60, 2007.
11. Osypov, A. A., Beskaravainy, P. M., Sorokin, A. A., Kamzolova, S. G. DEPPDB - DNA Electrostatic Potential Properties Database. International Workshop on Integrative Bioinformatics, 4th annual meeting, abstract №249, 2007
12. Osypov A.A., Beskaravainy P.M., Kamzolova S.G., Sorokin A.A., DNA Electrostatic Potential Database. Proceedings of the third Moscow Conference on Computational Molecular Biology (MCCMB'07), p.241-243, 2007
13. Osypov A.A., Panjukov V.V., Computational approach to the analysis of the properties of electrostatic potential profile of genome DNA. Moscow Conference on Computational Molecular Biology (MCCMB'07), p.243-244, 2007
14. А.А. Сорокин, А.А. Осипов, П.М. Бескаравайный, С.Г. Камзолова, Анализ распределения нуклеотидной последовательности и электростатического потенциала генома E.coli, Тезисы докладов XIII Симпозиума по межмолекулярному взаимодействию и конформациям молекул, с.177, 2006.
15. С.Г. Камзолова, А.А. Осипов, П.М. Бескаравайный, Т.Р. Джелядин, А.А. Сорокин, Регуляция активности промоторной ДНК через электростатические взаимодействия с РНК-полимеразой. Тезисы докладов XIII Симпозиума по межмолекулярному взаимодействию и конформациям молекул, с.66, 2006.
16. Sorokin A.A., Osypov A.A., Beskaravainy P.M., Kamzolova S.G., Promoter recognition by electrostatic properties of DNA helix. Proceedings of the International Moscow conference on Computational Molecular Biology, p.379-380, 2005.
17. Sorokin A.A., Dzhelyadin T.R., Osypov A.A., Beskaravainy P.M., Kamzolova S.G., Electrostatic properties of promoters recognized by RNA polymerase Eσ70. New promoter determinants. Proceedings of the International Moscow conference on Computational Molecular Biology, p.381-382, 2005.
18. А.А. Сорокин, Т.Р. Джелядин, П.М. Бескаравайный, А.А. Осипов, С.Г. Камзолова. Общие закономерности электростатических взаимодействий промоторной ДНК с РНК-полимеразой E.coli., Тезисы докладов XII симпозиума по межмолекулярному взаимодействию и конформациям молекул, стр.115, 2004.
19. А.А. Сорокин, Т.Р. Джелядин, П.М. Бескаравайный, А.А. Осипов, С.Г. Камзолова. Сравнительный анализ электростатических свойств различных промоторов рибосомальных оперонов E. coli. III съезд биофизиков России. Тезисы докладов. т. II, с. 797-798, 2004.
20. А.А. Сорокин, Т.Р. Джелядин, А.А. Осипов, П.М. Бескаравайный, С.Г. Камзолова. Электростатические свойства нуклеотидной последовательности генома E. coli. III съезд биофизиков России, Тезисы докладов. т. II, с. 798, 2004

Осипов Александр Александрович

ЭЛЕКТРОСТАТИЧЕСКИЕ СВОЙСТВА ГЕНОМНОЙ ДНК

Электростатические свойства геномной ДНК влияют на ее взаимодействие с различными белками, в частности, могут принимать участие в регуляции транскрипции РНК-полимеразами. Была создана база данных электростатических свойств ДНК всех полных секвенированных геномов прокариот и вирусов DEPPDB, с биологической аннотацией их элементов, и инструменты анализа этих данных.

Сравнения электростатических свойств полных геномов выявило близкую к линейной зависимость их среднего потенциала от содержания GC пар. Были рассчитаны ее параметры для разных таксономических групп. Установлено, что величина этой зависимости коррелирует с содержанием GC пар.

Изучены закономерности формирования электростатического потенциала ДНК. Выявлена сложная зависимость распределения потенциала от состава и организации ее последовательности. Показано влияние флангов на локальный потенциал.

Высказана гипотеза, что в сдвиг распределения природных геномов в АТ-богатую область могла внести вклад большая гибкость формирования электростатических элементов АТ-обогащенными последовательностями.

Показано различие в масштабах электростатических элементов промоторов бактериальных и фаговых полимераз, что отражает физические свойства этих белков.

Показано, что приспособление промоторов бактериофага Т7 к РНК-полимеразе Т3 сдвигает электростатический потенциала в районе 0 – -5 п.о. в сторону Т3, что свидетельствует о возможной зависимости от него дифференциального распознавание промоторов РНК-полимеразой Т3.

Osypov Alexander Alexandrovich

ELECTROSTATIC PROPERTIES OF GENOME DNA

Electrostatic properties of genome DNA influence its interactions with different proteins, in particular transcription regulation by RNA-polymerases. We developed DNA Electrostatic Potential Properties Database, DEPPDB, of all complete genomes of procaryotes and viruses, with biological annotations of their elements, and analytical tools.

Electrostatic properties of all available natural genomes were compared that revealed close to linear dependence of the genome mean potential on the GC content and the correlation of the dependence strength with the GC content. DNA electrostatic potential formation principles were studied and its complicated dependence on the composition and structure of the sequence was revealed. The flanking regions influence on the local potential is shown.

We hypothesize the possibility of the AT-rich sequences flexibility in the electrostatic potential formation contribution to the distribution shift of the natural genomes to AT-richness.

The scale difference of the electrostatic potential elements of the promoters for bacterial and viral RNA-polymerases was shown that reflects physical properties of the proteins.

Also was shown that T7 bacteriophage with its RNA polymerase deleted and supplied in trans by T3 one, evolved its promoters so that their electrostatic profiles moved to that of T3 in 0 – -5 .b.p. region, suggesting its possible role in their differential recognition.