

Лекция 19-ая «Конформационные особенности сайтов связывания транскрипционных факторов»

Спецкурс "Информационная Биология"

Лектор Д.Ю.Ощепков

[Slide 1]

Тема лекции предполагает что читателя, прежде всего, необходимо ознакомить с некоторыми общеизвестными положениями из молекулярной биологии, после чего можно будет перейти к более глубокому изложению материала. Поэтому почти каждый раздел изложения будет включать материал призванный напомнить читателю некоторую необходимую информацию.

[Slide 2]

Прежде всего, необходимо остановиться на сведениях о структуре ДНК. Существует 4 уровня организации ДНК в хромосоме. Хромосомы - это нуклеопротеиновые тела, в которых хранится и передается потомству наследственная информация, и, поскольку они являются главным "хранилищем" ДНК, то с их непосредственным участием протекают процессы, в котором участвуют транскрипционные факторы (ТФ). Первичная последовательность ДНК - это цепь из 4 четырех чередующихся в определенном порядке нуклеотидов - А, Т, G, С. Две спиральные (комплементарные) полинуклеотидные цепи закручены вправо вокруг одной оси и формируют регулярную двойную спираль - второй уровень организации. Далее спираль "намотана" на нуклеосомы - диски, состоящие из 4 белковых глобул - гистонов. Эта 3-ий уровень организации хроматина. В свою очередь, компактизованные таким образом ДНК-белковые образования закручиваются в еще более сложные структуры, формируя хромосомы.

[Slide 3]

Комплементарные взаимодействия двух цепей ДНК происходят за счет водородных связей, двух в случае взаимодействия тимин-аденин, и трех в случае гуанин - цитозин. Надо отметить, что это находит отражение в свойствах ДНК, так, например, АТ- тракты отличаются меньшей по сравнению с GC- богатыми участками температурой плавления, что в свою очередь используется факторами и ферментами, разделяющими цепи ДНК.

[Slide 4]

Наиболее распространена В-форма ДНК, представляющая собой правозакрученную двойную спираль, состоящую из двух цепей, положение которых фиксируется за счет формирования водородных связей между парами комплементарных оснований. Каждому шагу спирали соответствует одна комплементарная пара оснований. Для В-формы ДНК на один виток спирали приходится в среднем 10.5 п.о.; поперечный размер спирали составляет 24 А, длина витка - 34 А, ширина большой и малой бороздок - 12 А и 6 А соответственно.

[Slide 5] Вдобавок к этой классической структуре известно несколько других форм ДНК, и на данный момент ясно, что молекула ДНК может принимать различные структуры в зависимости от среды и первичной последовательности. Различные формы ДНК определяются как А-, В-, С- и т.д. формы ДНК. На данный момент остаются только буквы F, Q, U, V и Y для описания вновь найденных структур ДНК.

А-форма, так же как и В-форма представляет собой правозакрученную двойную спираль ДНК. А-форма характеризуется 11-12 п.о. на виток и поперечным размером 23 А. Гибрид ДНК-РНК может существовать только в форме А-ДНК.

Z-форма ДНК, в отличие от А- и В-форм является левозакрученной двойной спиралью. Эту

форму могут образовывать полимеры вида $(RY)_n$, в которых пурины ($R=A, G$) чередуются с пиримидинами ($Y=C, T$). Двойная спираль Z-формы имеет выраженную зигзагообразную конформацию. Это означает, что угол спирального вращения длинной оси симметрии пары оснований ДНК (угол твист) различен для двух соседних пар оснований: при этом чередуются два значения угла спирального вращения. Биологическая роль Z-ДНК ещё не вполне ясна. Показано существование ДНК в Z-форме в полиденных хромосомах плодовой мушки. Вероятно также, что Z-форма ДНК может выполнять регуляторную функцию.

[Slide 6]

Коды ДНК

Помимо классического кода передачи информации от РНК к белку, геномная ДНК включает в себя множество других, нетриплетных кодов для реализации других генетических процессов. Подобно триплетному коду, каждый из них является вырожденным, что позволяет им сосуществовать в одной и той же последовательности ДНК, так что одно и то же основание ДНК часто (если не всегда) принадлежит нескольким различным не перекрывающимся типам кодов. Таким образом, геномная ДНК представляет собой пример естественной многозначности текста, где важность каждого отдельного символа усиливается его одновременным использованием во многих перекрывающихся генетических сообщениях.

Так в качестве примера можно перечислить ряд важнейших методов кодирования геномной информации:

1. Код укладки РНК. Около половины генома эукариот транскрибируется в виде молекул РНК, имеющих компактную вторичную и третичную структуру, которая формируется на основе комплементарных взаимодействий между основаниями, а также комплементарных взаимодействий с различными белками. Часто для правильного действия этих молекул требуется укладка в уникальную вторичную структуру, на уровне ДНК обусловленную кодом укладки.
2. Хроматиновый код. Эукариотическая ДНК практически полностью включена в хроматин и почти вся упакована в нуклеосомы. Каждая элементарная структурная единица хроматина состоит из гистонового коря и закрученного вокруг него фрагмента ДНК размером 146 п.н. Любая последовательность ДНК может формировать нуклеосомные комплексы с гистоновыми октамерами. После накопления данных о том, что нуклеосомы способны позиционироваться относительно последовательности ДНК стало ясно, что расположение значительной части. А, скорее всего, и всех нуклеосом является контекст - специфичным, однако сигналы позиционирования нуклеосом являются достаточно слабыми и не содержат специфических последовательностей в фиксированных позициях. Можно предполагать, что существует другой способ кодирования этих сигналов.

3. Конформационный код.

Определяемая контекстом локальная конформация молекул ДНК является одним из факторов специфичности ДНК - белкового узнавания. Нарастающий объем экспериментальных данных дают основания считать, что функционирование сайтов связывания транскрипционных факторов в значительной степени определяется конформационными и физико - химическими свойствами ДНК.

[Slide 7]

Впервые зависимость конформации ДНК от контекста была обнаружена Dickerson and Drew (1981), путем рентгеноструктурного анализа додекамеров ДНК. Растущее число данных структурного анализа показало неоднородность конформационных и физико-химических свойств и их зависимость от нуклеотидной последовательности. Затем различными авторами были показаны ряд физико - химических особенностей ДНК, зависящих от контекста, и эти значения были приведены в соответствие к динуклеотидному коду.

[Slide 8]

Так, если мы определим систему координат для каждой из пар оснований (см. рис), то положение следующей пары оснований можно полностью описать системой из трех углов (Roll, Tilt, Twist) и трех расстояний (Slide, Shift, Rise) - по одному для каждой из осей. Кроме того, необходимо описать взаимное положение нуклеотидов в паре, что также можно сделать с помощью угла (propeller twist). Надо отметить, что это не все конформационные свойства, которые можно применить для описания структуры ДНК. Так, некоторые базы по структуре ДНК насчитывают десятки конформационных свойств, таких, например, как глубина и ширина бороздок спирали и много других свойств.

[Slide 9, 10]

Как видно из табл., свойства имеют величины, различные для разных динуклеотидов, что и определяет существование конформационного кода.

Кроме того, можно показать насколько сильно зависит форма ДНК от этих величин [Slide 11].

[Slide 12]

Чем же определяется различие конформационных параметров между динуклеотидами? Как было показано, зависимость конформационных параметров ДНК объясняется стерическими взаимодействиями экзоциклических групп NH_2 и O в бороздках спирали. В случае отсутствия этих экзоциклических групп угол закручивания спирали приближался бы к значению 42° . Однако вследствие конкуренции за место больших экзоциклических групп в бороздках спираль раскручивается и приобретает специфические конформационные свойства в зависимости от того, какие группы "сталкиваются" в бороздках.

[Slide 13]

Помимо конформационных свойств важную роль в ДНК - белковых взаимодействиях играют физико - химические свойства ДНК. Ранее уже упоминалось такое свойство как температура плавления. Помимо него важными оказываются такие свойства как изгибная жесткость ДНК, причем эта жесткость может определяться по разным направлениям - как в сторону большой бороздки, так и в сторону малой, персистентная длина (длина, на которой спираль можно приближенно считать цилиндром без явных изгибов и деформаций), распределение зарядов - электростатический заряд ДНК, дипольный момент и т.д.

Перейдем теперь к рассмотрению второго участника взаимодействий - белку, а именно к понятию **ТРАНСКРИПЦИОННЫЙ ФАКТОР** [Slide 14]

Согласно современным представлениям у эукариот наиболее значимым уровнем регуляции экспрессии генов является транскрипция, которая осуществляется с участием огромного числа белков, называемых транскрипционными факторами (ТФ). В основе функционирования ТФ лежит способность этих белков к специфическим взаимодействиям с ДНК промоторов, энхансеров и других регуляторных элементов генов, а также к белковым взаимодействиям с другими ТФ и белками эукариотической клетки. У прокариот основной механизм контроля экспрессии генов состоит в подавлении обычно активной транскрипционной системы. В отличие от них, эукариотам для выполнения генетических программ, обеспечивающих их функционирование, необходимо осуществлять гораздо более сложную регуляцию экспрессии генов. В значительной степени эта регуляция основана на координированной активации групп генов, продукты которых необходимы клеткам и тканям эукариотических организмов в различных ситуациях, зависящих от функционального состояния организмов, стадий их развития, условий окружающей среды и т.д.

Транскрипционные факторы, связывающиеся с ДНК, могут влиять на транскрипцию генов через несколько механизмов:

- В большинстве изученных к настоящему моменту случаев ТФ стимулируют формирование комплекса предынициации на ТАТА-боксе / инициаторном элементе за счет взаимодействия их транс - активирующих доменов с компонентами базального транскрипционного комплекса, либо непосредственно, либо через коактиваторы / медиаторы.
- Некоторые ТФ вызывают изменения структуры хроматина, делая его более доступным для РНК - полимераз.
- Некоторые ТФ являются вспомогательными, создавая оптимальную конформацию ДНК для действия других ТФ
- Известны ТФ, которые подавляют транскрипцию за счет непосредственного действия своих ингибирующих доменов, либо нарушая совместное функционирование комплекса ТФ внутри регуляторной области гена (промотора, энхансера)
- Наконец, существуют ТФ, которые сами не связываются с ДНК, а объединяются в более сложные комплексы посредством белок - белковых взаимодействий

Таким образом, ТРАНСКРИПЦИОННЫЙ ФАКТОР - это белок, который после его перемещения в ядро клетки регулирует транскрипцию, специфически взаимодействуя с ДНК, либо стехиометрически взаимодействуя с другим белком, который может образовывать специфичный к последовательности ДНК комплекс "белок - ДНК" [Slide 15].

Рассмотрение основных классов ДНК - связывающих доменов транскрипционных факторов невозможно без знания основных понятий о строении глобулярных белков.

[Slide 16]

Подобно ДНК, в белках разделяют первичную, вторичную, третичную и четвертичную структуры. Первичная последовательность белков состоит из 20 аминокислот.

[Slide 17]

Каждой из таких аминокислот можно приписать ряд свойств, таких как полярность, заряд, гидрофобность, гидрофильность и т.д. В естественных условиях клеточного окружения белок сворачивается в глобулу, структура которой определяется последовательностью его аминокислот, и далее выполняет специфические функции в составе клетки. Так, одними из основных блоков вторичной структуры белков являются альфа - спирали и бета - складчатые листы.

[Slide 18]

Как альфа - спирали, так и бета - складчатые листы формируются за счет водородных связей между группами NH и CO основной цепи аминокислот первичной последовательности. Эти элементы вторичной структуры играют важную роль в понимании структуры ДНК - связывающих доменов ТФ.

Структура ТФ подвергалась подробному изучению в течение последних нескольких десятков лет, поскольку является ключевой в понимании кода узнавания своих сайтов связывания, что в итоге является необходимым для понимания основ регуляции транскрипции, и, в итоге, расшифровки генома.

[Slide 19]

Так, Э. Вингендером была предложена схема классификации ТФ, которая на данный момент является наиболее употребительной в кругах специалистов, занимающихся изучением ТФ (см. табл.)

[Slide 20]

Как видно, одним из важнейших пунктов классификации является определение типа ДНК-связывающего домена (ДСД). Согласно этой классификации принято делить ДСД на 4 суперкласса. Рассмотрим основные суперклассы ДСД

[Slide 21].

Контакты первого из суперклассов ДСД - "основного" - формируются за счет электростатических взаимодействий. [Slide 22] Общей особенностью белков этого типа является то, что они контактируют с ДНК посредством основной области. В свободном состоянии (в растворе) она имеет неупорядоченную структуру, а при связывании с ДНК переходит в конформацию Альфа-спирали. Наиболее представленными в этом суперсемействе являются белки трех классов: мотив типа лейциновой застежки, мотив типа основной спираль - петля - спираль, и мотив смешанного типа спираль - петля - спираль + лейциновая застежка. Помимо них к этому классу относятся ряд других, менее представленных классов. Консенсусный мотив, общий для всех основных доменов этого суперсемейства, отсутствует (Вингендер, 1997). Даже основные области факторов типа лейциновой застежки и типа основной спираль - петля - спираль сильно отличаются, несмотря на то, что способ их взаимодействия с ДНК очень похож: димеризационный домен, имеющий структуру Альфа-спирали, обеспечивает связь между двумя субъединицами, входящими в состав ТФ, сближая их основные области, которые при взаимодействии с ДНК переходят из неупорядоченной в Альфа-спиральную конформацию (по-видимому, вследствие нейтрализации положительных зарядов боковых групп основных районов при взаимодействии с сахарофосфатным остовом ДНК несущим отрицательный заряд). ДСД типа "лейциновая застежка" получил свое название благодаря любопытному способу, которым он связывается с ДНК. Этот тип ДНК - связывающих доменов был изначально определен по своему димеризационному мотиву, Альфа-спиральному сегменту, который обычно называют "лейциновой застежкой". Этот сегмент длиной примерно 30 а.к. формирует С-концевой участок, который содержит положительно заряженный "основной" район и N концевой домен. Димеризация происходит через формирование "скрученной спирали" (coiled coil) из двух Альфа-спиралей, как бы застегиваясь подобно замку - zipperу. При этом ДНК зажимается между концами этой застежки, образуя специфические контакты с обоими альфа - спиралями в большой бороздке.

[Slide 23] Действие следующего суперкласса аналогично принципу связывания предыдущего суперкласса - домены, в качестве структурного элемента содержащие ион Zn^{+} . Интересен принцип, по которому сами ионы цинка связываются с цепью белка.

[Slide 24] Цепь образует специфическую петлю - "палец", в котором в соответствующих местах расположены два остатка цистеина и два - гистидина, формирующие связи с ионом цинка. Поэтому часто такие структуры еще называют "цинковыми пальцами".

Известно три типа таких доменов. Первый из них состоит из примерно 30 аминокислотных остатков, ион цинка при этом в качестве лиганда присоединяется к двум гистидиновым и двум цистеиновым остаткам. Последовательности около двухсот доменов этого класса удовлетворяют следующему консенсусу: $o x C x_{2-4} C x_3 o x_5 o x_2 H x_{3-4} H x_{2-6}$

где o - гидрофобный остаток, а x - любой остаток. Инвариантные Hys и Cys могут быть использованы как определяющие для такой структуры.

Второй вид доменов этого типа - последовательность порядка 70 а.к. где два иона цинка связаны с четырьмя цистеинами. Третий вид имеет также 2 иона цинка, связанные с шестью цистеинами. Большинство белков имеющие в своем составе такие модули включают в себя последовательно три и более домена типа цинковых пальцев.

[Slide 25] Следующий суперкласс - ДСД типа спираль - поворот - спираль является одним из самых распространенных и, возможно, самым древним из доменов. Большинство прокариотических транскрипционных факторов, а также эукариотические гомеодомены попадают в эту группу. Этот домен обычно состоит из 20 а.к. с двумя альфа - спиральями, которые пересекаются под углом примерно 120^0 . В разных случаях могут быть дополнительные остатки в составе спиралей как на N конце первой, так и на C конце второй спиралей. Вторая из этих спиралей называется узнающей, поскольку при связывании с ДНК она ложится в большую бороздку, а также потому что эксперименты по замене остатков на внешней стороне спирали показали их важность для специфического узнавания.

Считается, что вторая альфа - спираль выполняет функции ориентации и стабилизации комплекса относительно спирали ДНК. Как видно из следующего слайда, [Slide 26] на данный момент известно множество модификаций ДСД этого суперкласса в составе большого количества ТФ, благодаря чему можно сказать, что этот ДСД является одним из самых часто встречающихся. [Slide 25] Для классов ТФ, входящих в следующий суперкласс трудно выделить одно общее свойство, однако можно указать ряд характерных признаков. В действительности, не все факторы этого класса содержат Бета-скэффолд, и не все факторы в любых ситуациях осуществляют стабильные контакты с ДНК по малой бороздке. Однако это наиболее типичные характеристики факторов, входящих в этот суперкласс. При попарном сравнении последовательностей любых классов внутри этого суперкласса выявлено значительное сходство. Сходные признаки не характерны для членов других суперклассов. Эти обстоятельства и послужили для выделения этих ТФ в один суперкласс. Большинство ДСД этого класса содержат достаточно выраженный скэффолд из Бета-нитей, на котором располагаются элементы вторичной структуры, непосредственно контактирующие с ДНК, такие как ?-спирали, петли и сами Бета-нити. Факторы этого типа образуют с ДНК обширные контакты. Одни из них могут охватывать, по крайней мере, половину витка ДНК, как, например, факторы Rel-класса, мультимеры р53, ТАТА - Вох связывающий белок (ТВР), и HMG-домены. В других случаях ДНК закручена вокруг белкового ядра, как показано для MADS-бокс содержащих факторов и вирусного регулятора E2. Общим свойством многих ТФ этого класса является то, что они связываются с АТ богатыми участками, окруженными по краям последовательностями, богатыми CG парами, и при этом взаимодействия с центральной частью сайта осуществляются путем контактов по малой бороздке ДНК. Подобные взаимодействия сильно искажают конформацию ДНК (возможным исключением являются белки класса Rel).

На данный момент огромное число работ направлено на расшифровку кода сайтов связывания, с целью подробного изучения процессов регуляции транскрипции. Строятся различные программы распознавания сайтов связывания, основанные на анализе контекста. В рамках нашей лаборатории ведется работа по созданию методов распознавания, основанных, в том числе на использовании конформационных и физико - химических свойств. [Slide 27] Основой для этого подхода стало предположение о том, что для различных вариантов геномных последовательностей, взаимодействующих с конкретным регуляторным белком необходимо сохранение определенных конформационных и физико-химических характеристик.

Сохранение свойств может быть обеспечено несколькими механизмами, которые необходимо учитывать при анализе последовательностей функциональных сайтов и регуляторных районов ДНК:

1. Консервативность нуклеотидного контекста, проявляющегося в том, что различные варианты сайта содержат фиксированный набор нуклеотидов в определенных позициях. Функциональные сайты такого типа могут анализироваться с помощью методов, основанных на контекстном анализе, таких, как консенсус, весовые матрицы и т.д.
2. Консервативность некоторых конформационных и физико-химических свойств в

определенных участках сайта, обусловленные таким характером нуклеотидных замен, которые существенно не меняют этих свойств. Для выявления таких участков функциональных сайтов могут применяться методы, основанные на рассмотрении контекстно - зависимых конформационных и физико-химических свойств ДНК. Таким образом, консервативность нуклеотидного контекста, по сути, может являться лишь следствием ограничений, накладываемых на значение конкретных свойств в определенных участках последовательности.

При такого рода анализе возможно два подхода: Во-первых, можно рассматривать консервативность среднего значения свойства для протяженного участка последовательности длиной 10-30 нуклеотидов, находящейся во взаимодействии либо в непосредственной близости от связанного с ней белка. Этот подход оказался очень плодотворным и позволил построить ряд методов распознавания и определения активности. Подобный подход реализован в системе B-DNA-Video [Ponomarenko J.V et al., 1999]. Однако, в случае существования отдельных важных для взаимодействия конформационных особенностей, но локализованных на протяжении лишь коротких участков исследуемой последовательности этот подход не даст возможности это обнаружить. Во-вторых, можно рассматривать последовательность сайта связывания более детально, рассматривая консервативность свойств на более коротких участках функциональной последовательности. В этом случае анализ более сложного паттерна конформационных особенностей, локализованных по последовательности сайта, может пролить свет на характер молекулярных механизмов взаимодействия анализируемых функциональных последовательностей с взаимодействующим с ними белком. Вообще говоря, такое исследование имеет целью найти те допустимые конформации участка ДНК, для которого связывание проходит наиболее эффективно, т.е. в итоге определить те участки, которые и являются потенциальными сайтами связывания соответствующего регуляторного элемента. При анализе выборок выравненных нуклеотидных последовательностей ДНК сайтов использовались 38 конформационных и физико - химических свойств ДНК из БД property (<http://www.mgs.bionet.nsc.ru/mgs/gnw/bdna/>).

[Slide 28] Наш подход заключается в следующем. Рассматривается выборка из N выравненных (сфазированных) последовательностей ДНК длиной L (без гэпов). Каждому динуклеотиду сопоставляется величина определенного физико-химического или конформационного свойства f . В результате формируется матрица размером $N \times (L-1)$. Элемент этой матрицы $f(l, m)$ соответствует величине определенного свойства f динуклеотида в m -ой позиции последовательности l .

Среднее значение свойства i в позиции l :

$$\overline{F_{il}} = \frac{1}{N} \sum_{k=1}^N F_{ikl}, \quad (1)$$

В качестве меры консервативности i -го свойства для каждой позиции l используется величина дисперсии:

$$\sigma^2_{F_{il}} = \frac{1}{N-1} \sum_{k=1}^N (F_{ikl} - \overline{F_{il}})^2, \quad (2)$$

Предполагается, что в случае важности конкретного свойства в определенном участке последовательности для функционирования этого сайта связывания величина данного свойства сохраняется для всех последовательностей выборки, что и обеспечивает малую величину дисперсии по сравнению с выборкой случайных последовательностей. Таким образом, малая величина дисперсии конкретного свойства говорит о его консервативности в данной позиции.

[Slide 29]

Благодаря этому подходу были проанализированы ряд сайтов связывания, в том числе сайт связывания гетеродимера E2F/DP. В результате анализа выборки сайтов связывания E2F/DP и проверки всех 38 свойств на консервативность был обнаружен ряд особенностей конформации ДНК, из которых можно выделить 2. Для свойств major/minor groove width/depth, наблюдается консервативность для всех позиций сайта, (8 п.н при значимости 99% по критерию хи-квадрат).

[Slide 30]

При этом консервативные значения свойств major groove width для участка TTTCGCGCG и minor groove width для участка TTTCGCGCG имеют значения, существенно отклоняющиеся в большую сторону от средних значений этих свойств. Наблюдения о размерах большой и малой бороздок согласуются с данными рентгеноструктурного анализа. При связывании гетеродимера E2F/DP в большую бороздку ДНК помещается две узнающих Альфа-спирали, по одной от каждого из составляющих гетеродимера. При этом Альфа -спираль ДНК - связывающего домена E2F помещается в большую бороздку в области TTTCGCGCG, одновременно Альфа -спираль ДНК - связывающего домена DP помещается в большую бороздку в области TTTCGCGCG, таким образом, увеличенная ширина большой бороздки в области TTTCGCGCG, видимо, является необходимым условием для распознавания гетеродимером E2F/DP сайта связывания. Следует отметить, что такого рода зависимость является безусловной особенностью характерной для сайтов связывания гетеродимера E2F/DP, поскольку узнающие Альфа-спирали обоих белков в пространстве расположены одна за другой и между контактирующими с ними основаниями нет перерыва.