

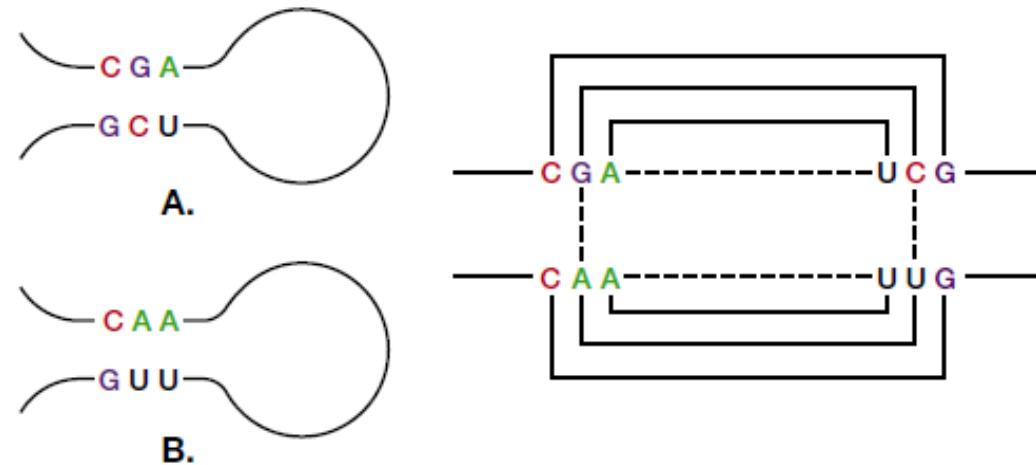
КОМПЮТЪРНИ МОДЕЛИ НА СТРУКТУРАТА НА РНК

ПРОФ. БОРОВСКА

Предсказване на вторичната структура на РНК

- ▶ Целта на подравняването на протеините и ядрените аминокиселини е да се определят множества от символи на секвенцията, които са подредени по същия начин в други секвенции
- ▶ Ядрените аминокиселини, които специфицират молекулите на РНК, трябва да бъдат сравнявани по различен начин
- ▶ При вариациите на секвенциите РНК се наблюдават устойчиви комбинации от двойки бази, които формират **участъци от двойни нишки(вторичната структура)** в молекулата
- ▶ Подреждането на две секвенции, които специфицират еднакви молекули РНК показват ковариации при взаимодействащите базови двойки
- ▶ Освен ковариабилните позиции, генетичните секвенции на РНК могат да съдържат редове от сходни нуклетоди, които показват **общия произход на гените**

Комплементарните секвенции на молекули РНК поддържат вторичната структура на РНК



Проста стволова верижна структура, формирана в резултат на нагъването на нишката РНК обратно върху себе си

- Молекулата А зависи от наличието на две комплементарни последователности CGA и UCG , които са свързани (basepaired) в структурата.
- При молекулата В, има две промени в секвенцията, $G \rightarrow A$ и $C \rightarrow U$, които поддържат същата вторична структура като в молекулата А.
 - подравняването на молекули РНК изисква локализацията на такива участъци с ковариация, които биха могли да осъществят сдвояване на нуклеотидни бази в съответната вторична структура на РНК.

Програми и guest sites за анализ на РНК

- ▶ RNA secondary structures, Group I introns - <http://www.rna.cccb.utexas.edu/>
- ▶ RNA structure database - <http://www.rnabase.org/>
- ▶ RNA world at IMB Jena - <http://www.imb-jena.de/RNA.html>
- ▶ Small RNA database - <http://mbcr.bcm.tmc.edu/smallRNA/smallrna.html>
- ▶ Vienna RNA package for RNA secondary structure prediction and comparison - <http://www.tbi.univie.ac.at/ivo/RNA/>
- ▶ A list of RNA Web sites and databases is available at <http://bioinfo.math.rpi.edu/zukerm/>
- ▶ and at <http://pundit.colorado.edu:8080/>

Предсказване на структурата на РНК (RNA STRUCTURE PREDICTION BASICS)

Изчислителен метод 1

- ▶ От множеството възможни комплементарни секвенции, които потенциално могат да бъдат сформирани, се избират структурите, при които **молекулите са енергийно стабилни**.
- ▶ В резултат на анализа могат да се определят структурите с най-стабилните (устойчиви) състояния, както и да се прогнозират участъците, които са най-стабилни по отношение на енергийното състояние.
- ▶ Разработват се и методи за предсказване на третичната структура на РНК

Предсказване на структурата на РНК

Изчислителен метод 2

- ▶ При този метод се вземат предвид **устойчиви модели на сдвояване на нуклеотидните бази, които са се съхранили през еволюцията** на даден клас молекули РНК
 - ▶ При еволюцията се наблюдава, също така, промяна в позициите на базите, които се сдвояват, по такъв начин, че да се запази структурната цялост (интегритет) на молекулите на РНК.
- ▶ Така например, ако две позиции G и C формират двойка бази при определен тип молекула, то секвенциите, при които са налице позиции C и G, или - A и U или U и A, в съответните позиции, се приемат за разумни съчетания за сдвояване.

Предсказване на структурата на РНК

Изчислителен метод 2

- ▶ Тези *модели на ковариация (съизменения) при молекулите РНК* се третират като индикация за вторична структура и се използват за прогнозиране на структурата на молекулата РНК.
- ▶ **Метод 2- откриване на ковариабилните позиции на фона на другите изменения в секвенцията.**

Характеристики на вторичната структура на РНК

- ▶ Вторичната структура на РНК може лесно да бъде разглеждана като междинен етап в образуването на тримерна структура
- ▶ Вторичната структура на РНК се състои главно от двойно-нишкови РНК региони, оформени чрез прегъване на едно-нишкова молекула обратно на себе си
- ▶ За създаването на такива двойно-нишкови участъци, участък от нишката на РНК секвенцията трябва да бъде комплементарен на усуканата част, така че се получава сдвояване на комплементарните вериги (двойката комплементарни нуклеотида G/C и A/U (аналогично на G/C и A/T в ДНК))

Енергийна стабилност на молекулата

- ▶ Двойката нуклеотидни бази G/C допринасят най-много за енергийната стабилност на молекулата
- ▶ Двойката бази A/U обуславят по-малка енергийна стабилност от G/C
- ▶ Нестабилната двойка бази G/U обуславят най-малката енергийна стабилност на молекулата
- ▶ Предсказването на структурата на РНК обхваща участъци със сдвоени и несдвоени нуклеотидни бази в рамките на участъци с контури и възли

Типове едно и двойно-нишкови участъци във вторичната структура на РНК

- ▶ Еднонишковите молекули на РНК се прегъват и формират двунишкови спирали от комплементарни секвенции.
- ▶ Дадена нуклеотидна база може да не бъде сдвоена, (пример А), *или може да бъде сдвоена с друга база* (пример В)
- ▶ Двойно-нишковите участъци се формират с най-голяма вероятност на местата, където единият участък нуклеотиди в секвенцията може да се сдвои с отсрещния участък от комплементарни нуклеотиди.
- ▶ **Енергията на сдвоените бази предопределя повишаване на енергийната стабилност на структурата**

A. Single-stranded RNA

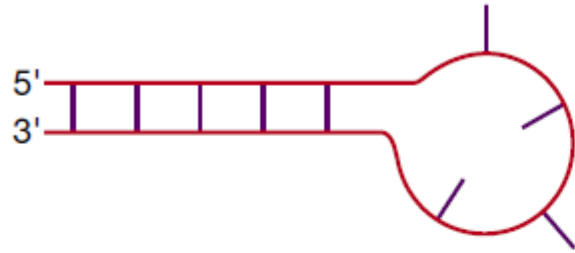


B. Double-stranded RNA helix of stacked base pairs

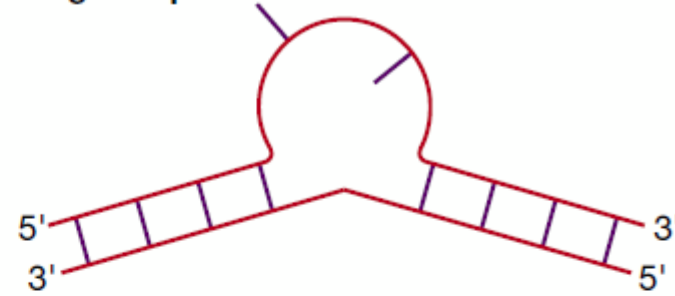


Вторична структура на РНК

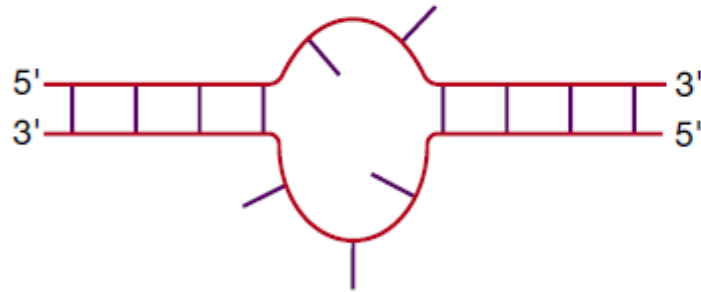
C. Stem and loop or hairpin loop.



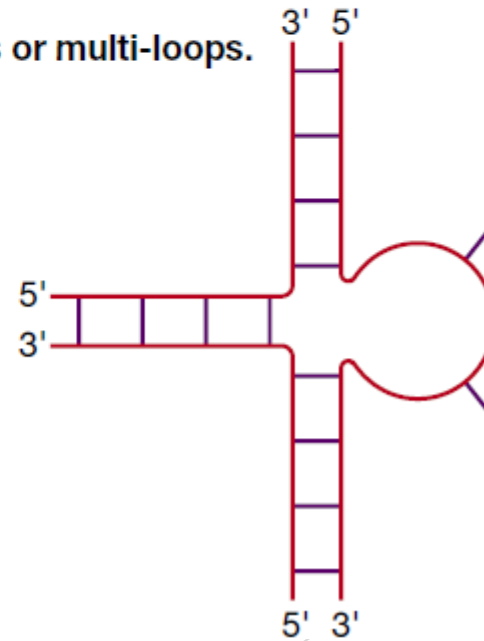
D. Bulge loop



E. Interior loop

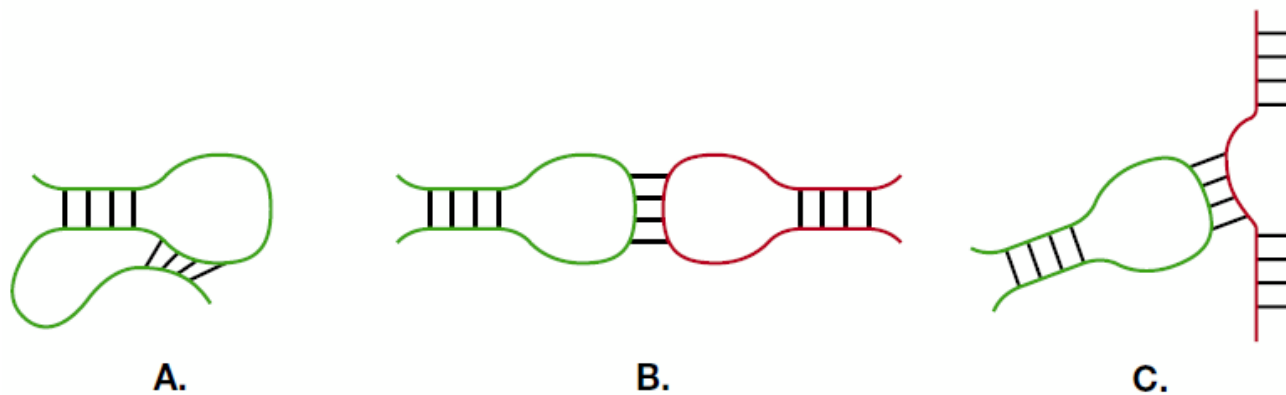


F. Junctions or multi-loops.

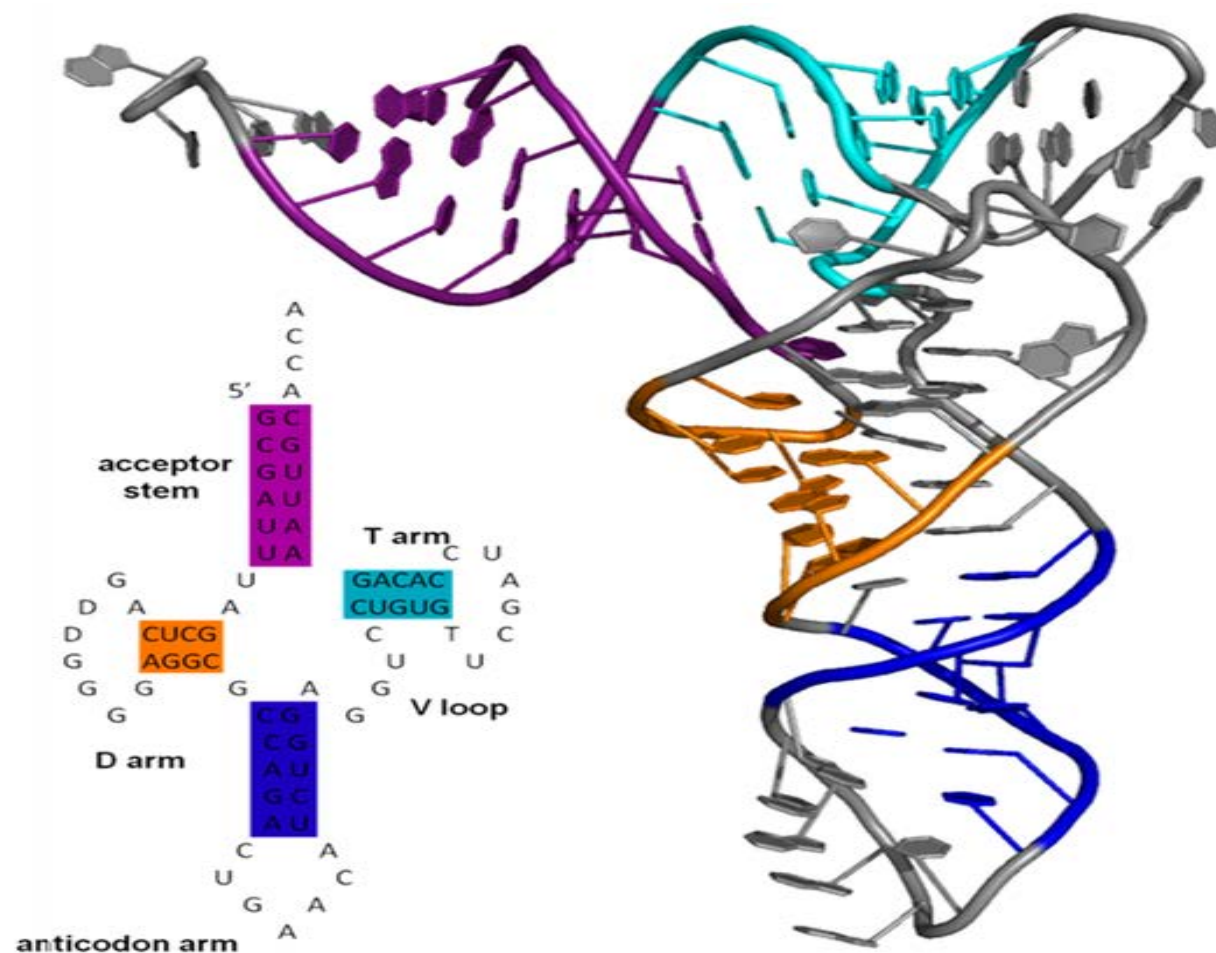


Третични взаимодействия

- ▶ Освен вторичните структурни взаимодействия при РНК, съществуват също така, и третични взаимодействия.
- ▶ Третичната структура на РНК не може да бъде предсказана с програмите за предсказване на вторичната структура.
- ▶ Третичната структура на РНК може да бъде предсказана на основата на прецизен ковариантен анализ.



Втори́чна и трети́чна структура на tRNA

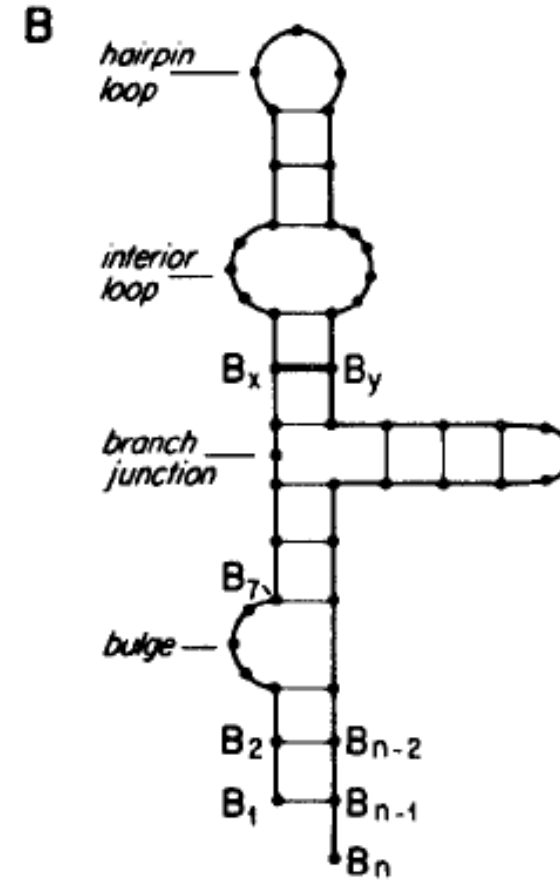
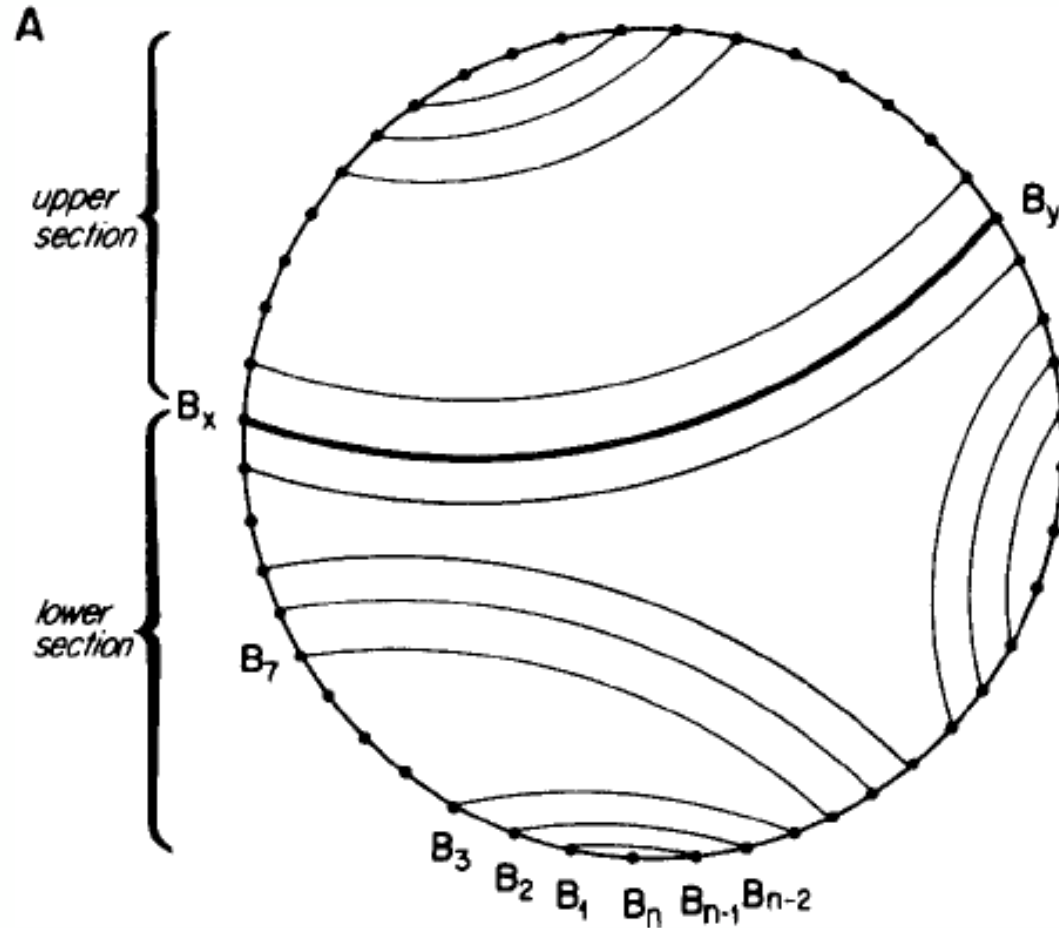


Ограничения при предсказването на вторичната структура на РНК

► Допускания за опростяване:

1. Най-вероятната структура е енергийно най-стабилната структура
2. Енергията, свързана с произволна позиция в структурата се определя само от локалната секвенция и структура (енергията, свързана с определена базова двойка в двунишков участък се приема, че се влияе от предходната базова двойка, и не се влияе от следващите двойки или кои да са други базови двойки в структурата)
3. Приемаме, че структурата се формира посредством прегъване на нишката и сдвояване сама със себе си по начин по който не се получават възли (knots).

Представяне на базовите двойки във вторичната структура на РНК чрез кръгова диаграма



Представяне на базовите двойки във вторичната структура на РНК чрез кръгова диаграма

- ▶ Предсказаната структура с минимална свободна енергия се *представя графично посредством кръгова диаграма, при която дъгите свързват* съответните бази в секвенцията
- ▶ Някои от дъгите не трябва да се пресичат, като това е графично представяне на структура, която не съдържа възли
- ▶ Най-добрият начин за представяне на структура без възли е да се начертае секвенцията във формата на кръг. След това сдвоените бази се свързват посредством дъги, които не бива да се пресичат

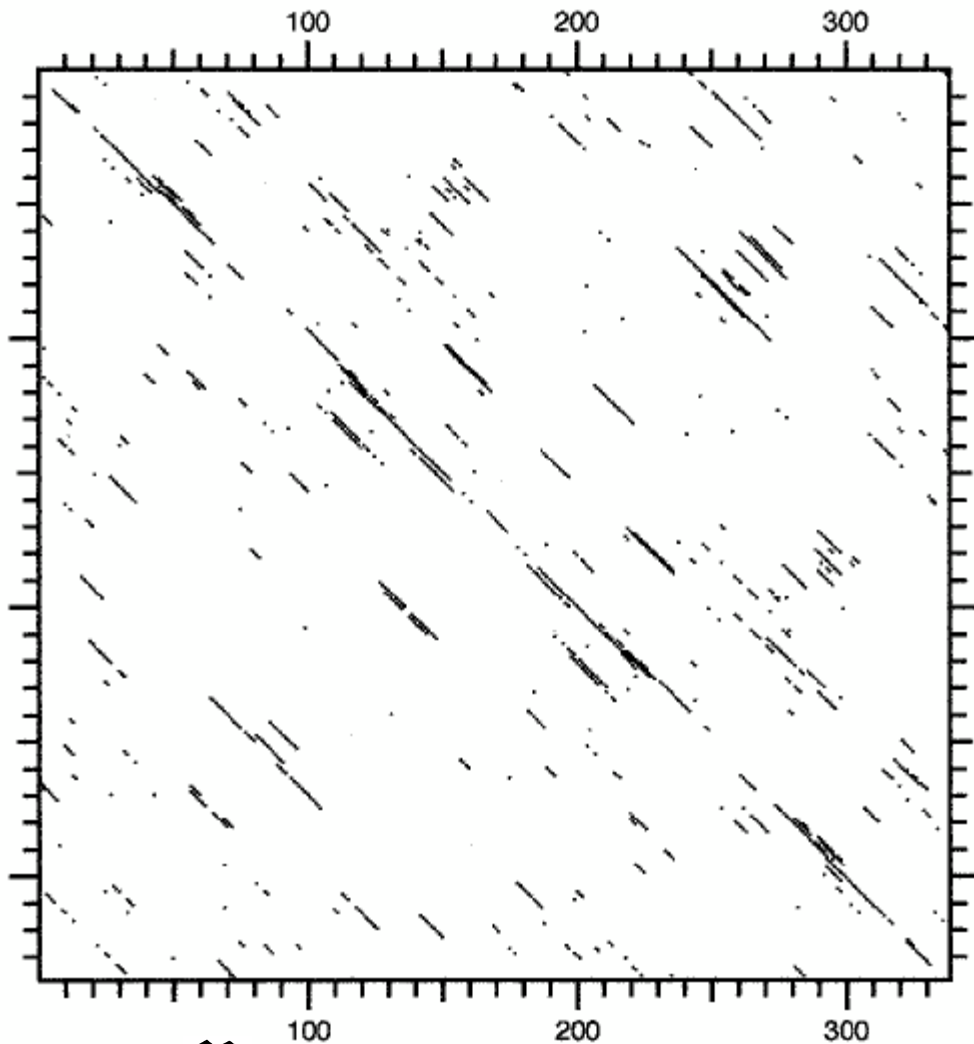
Методи за предсказване на вторичната структура на РНК

1. Метод на точковата матрица за сравняване на секвенции за самотопълващи се участъци
2. Метод на минималната свободна енергия
3. Субоптимално предсказване на структурата с помощта на MFOLD и използване на енергийни диаграми
4. Предсказване на най-вероятната вторична структура на РНК
5. Ковариация на секвенцията (съизменения)

Метод на точковата матрица за сравняване на секвенции за самодопълващи се участъци

- ▶ Всички видове структурен анализ на вторичната структура на РНК започват с откриването на самодопълващи се участъци, след което, съвместимите участъци могат да бъдат използвани за предсказване на структура с минимална енергия
- ▶ Подобрена версия на метода с точковата матрица може да се използва за определяне на частите на молекулата с най-висока енергия
- ▶ Само допълващите се участъци в РНК могат да бъдат открити посредством анализ с метода на точковата матрица, като секвенцията се сравнява сама със себе (вместо да са сравняват две различни секвенции)

Анализ на вторичната структура на РНК по метода на точковата матрица посредством функцията DNA MATRIX на софтуера DNA Strider



- Появата на диагонал от центъра на матрицата до горния ѝ ляв край, и огледален образ на този диагонал в долния десен край.
- Появата на този диагонал показва наличието на голяма самодопълваща се секвенция, и следователно, цялата молекула потенциално може да се нагъне сама на себе си до образуването на структура „фиба“.

Метод на минималната енергия

- ▶ Първо, всяка база се сравнява с всяка от останалите бази посредством анализ, много близък до анализа с точкова матрица
- ▶ Наличието на ред от съвпадения в матрицата на РНК показва последователност от самодопълващи се бази, които потенциално могат да формират самонагънат (сдвоен) участък
- ▶ Енергията на всяка предсказана структура се оценява на базата на правилото на най-близкия съсед (**the nearest-neighbor rule**) – сумират се отрицателните енергии на съседни двойки в нагънатите участъци и се прибавят оценените положителни енергии на дестабилизиращите участъци като примки (loops) в края на “фиби” (hairpins), издутини (bulges) в рамките “фиби”, вътрешни издутини, и други несдвоени участъци

Метод за анализ на линейна структура на РНК на основата на динамичното програмиране за откриване на най-благоприятната конфигурация в енергийно отношение

- ▶ Сравняват се първите позиции i_1 и j_2 за потенциално сдвоена база, и ако е възможно сдвояване, се изчислява стойността на енергията, като се записва в матрицата на енергията (energy matrix) W в позиция 1,2.
- ▶ След това, се сравняват базите в позиции i_2 и j_3 , и т.н., до изчерпване на всички комбинации на базите по протежение на диагонала.
- ▶ После се извършват сравнения по следващия диагонал горе в дясно.
- ▶ След изчерпването на всички двойки бази, се изчислява енергията, която е оптималната до този етап на сравнението.
- ▶ В най-простия случай, ако i_1 се сдвоява с j_1 , и i се сдвоява с j , и ако тази структура е най-благоприятната до този момент, енергията на двойката бази i/j се добавя към тази на двойката бази i_1/j_1

Метод на минималната свободна енергия

MINIMUM FREE-ENERGY METHOD

A. Base comparisons

5'	A	C	G	U	3'
A					
C					
G					
U					
-					
-					
G		C/G		U/G	
C			G/C		
G		C/G		U/G	
U	A/U	C/U	G/U		
3'					

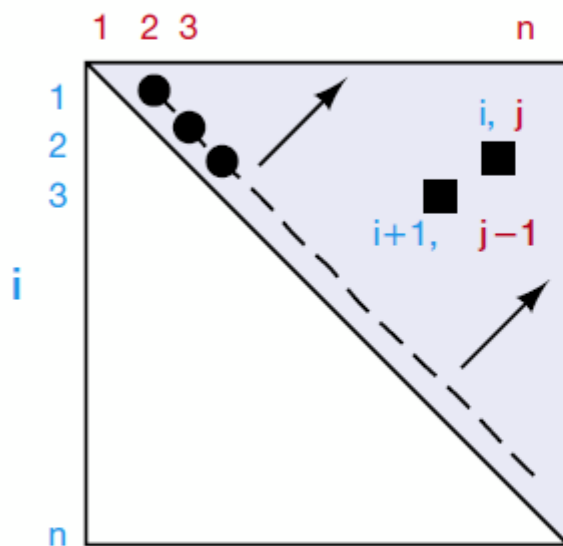
Сравнение на базите

B. Free energy calculations

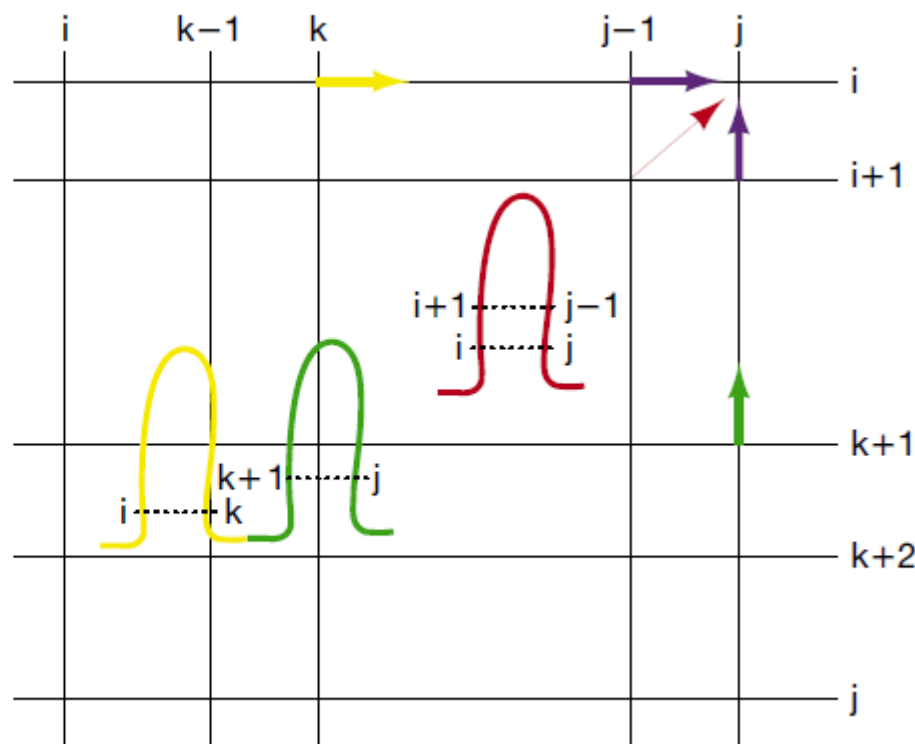
5'	A	C	G	U	3'
A					
C					
G					
U					
-					
-					
G				-6.4	
C			-5.2		
G		-1.8			
U					
3'					

Изчисляване на свободната енергия

MINIMUM FREE-ENERGY METHOD



A.



B.

- Освен оценъчната матрица на енергията W , се използват и допълнителни оценъчни матрици за да се следи помощната информация като напр., най-добрата енергия до i, j където i и j формират двойка, както и влиянието на изпъкнали примки, вътрешни примки, и други дестабилизиращи енергии.
- Важна е и втората матрица $V(i, j)$, която проследява всички подструктури в интервала i, j , при които i формира двойка (сдвояване) с j .

Софтуерен пакет Vienna RNA

- ▶ <http://www.tbi.univie.ac.at/RNA/>
- ▶ Vienna RNA Package съдържа библиотека от код на C и няколко самостоятелни програми за предсказване for и сравнение на вторичната структура на РНК
- ▶ В рамките на този пакет предсказването на вторичната структура на РНК на основата на минимизиране на енергията е най-често използваната функция в пакета
- ▶ *За предсказването на структурата се използват 3 вида алгоритми на динамичното програмиране:*
- ▶ (1) **алгоритъм с минималната свободна енергия**, който дава една оптимална структура,
- ▶ (2) **алгоритъм с функция на разделяне** (the partition function algorithm), който изчислява вероятностите за сдвоени бази в термодинамичен аспект, и
- ▶ (3) **алгоритъм за субоптимални нагъвания** (the suboptimal folding algorithm, който генерира всички възможни субоптимални структури в зададен диапазон на оптимална енергия.
- ▶ За сравняването на вторични структури, пакетът съдържа няколко мерки за дистанция (различия) като се използва подреждане на стринг или редактиране на дърво
- ▶ Алгоритъм за дизайн на секвенция с предефинирана (зададена) структура (инверсно нагъване - inverse folding).

Vienna RNA software library

- ▶ **RNAfold** - предсказва структури с минимална свободна енергия и вероятностите за сдвояване на базите
- ▶ **RNAeval** - изчислява енергията на вторични структури на РНК
- ▶ **RNAheat** -- изчислява специфичната топлина (кривата на топене) на зададена секвенция на РНК
- ▶ **RNAinverse** - инверсно сгъване - по зададена предефинирана структура генерира дизайн на секвенция на РНК
- ▶ **RNAdistance** - сравнение на вторични структури
- ▶ **RNApdist** - сравнява вероятностите за сдвояване на бази
- ▶ **RNAsubopt** - генериране на всички субоптимални нагъвания
- ▶ **RNAplot** - изчертаване на структурата на РНК с PostScript, SVG, or GML
- ▶ **RNAcofold** - предсказване на хибридната структура на две секвенции
- ▶ **RNA duplex** - предсказване на вероятни сайтове на хибридизация между две секвенции
- ▶ **RNAalifold** - предсказване на консесусната структура на няколко подредени секвенции
- ▶ **RNAfold** -- предсказване на локално стабилна структура на дълги секвенции
- ▶ **RNAplfold** - изчислява средните вероятности за сдвояване на локалните бази в дълги секвенции
- ▶ **RNApaln** - бързо структурно подреждане на секвенции РНК посредством подреждане на стрингове
- ▶ Няколко малки, но полезни **Perl Utilities**

RNAssoft

- ▶ **RNAssoft** – софтуер за предсказване и дизайн на вторичната структура на РНК/ДНК <http://www.rnasoft.ca/>
- ▶ **PairFold** предсказва вторичната структура с минимална свободна енергия формирана от две входни ДНК или РНК молекули. **PairFold** може да се използва, напр., за предсказване на взаимодействията между пробна и целева молекула РНК, или между две нишки при биомолекулярни наноструктури.
- ▶ **CombFold** предсказва коя нишка от зададено комбинаторно множество ДНК или РНК нишки, се нагъва във вторична структура с най-ниската минимална свободна енергия. С помощта на CombFold може, напр., да се тества ефективно, че в рамките на голяма библиотека не съществува нишка, която да формира нежелана вторична структура
- ▶ **RNA Designer** генерира дизайн на секвенция РНК, която се нагъва като зададена (предефинирана) вторична структура на РНК. Софтуерният инструмент е предназначен за дизайнери на молекули РНК със специфични структурни или функционални свойства.
- ▶ **HotKnots** предсказва вторични структури на РНК с псевдовъзли. Софтуерният инструмент е предназначен за обработка на къси секвенции РНК, които имат тенденцията да образуват псевдовъзли.

MFOLD и енергийни диаграми

- ▶ Първоначално, програмата FOLD на M. Zuker може да предсказва само една структура с минимална свободна енергия.
- ▶ Промените дори в един нуклеотид на секвенцията, обаче, могат да доведат до драстични промени в предсказаната структура
- ▶ По-късна версия, наречена MFOLD, осигурява по-добро предсказване на взаимодействията на несдвоените бази и генерира няколко структури с енергия, близка до минималната свободна енергия
 - ▶ MFOLD използва нова оценъчна матрица, при която секвенцията се представя с две тандемни копия както по вертикалната, така и по хоризонталната ос

Модел на вторичната структура на РНК на пре-терминален протеин на човешки аденовирус

A.

```

      10      20      30      40
CKCG |C  -----AK      A      AAAA      CU AU
      GUU CAG      GUUGC GC GCGC      AGUG CC G
      CAG GUC      CGGCGCG CGCCG      UCGC GG G
-ARA ^C  SUGNCCUA      -      -GUC      AG CU
      560      550      330      50

                        60      70      80      90
C      UC      GC      U      GA      UCUAGAC
UGGCCGG AGGC GCGCAG CGUU CGC      C
ACCGGUU UUCG CGCGUC GUAG GCG      G
-      UU      AU      -      -      -----
      320      310      300

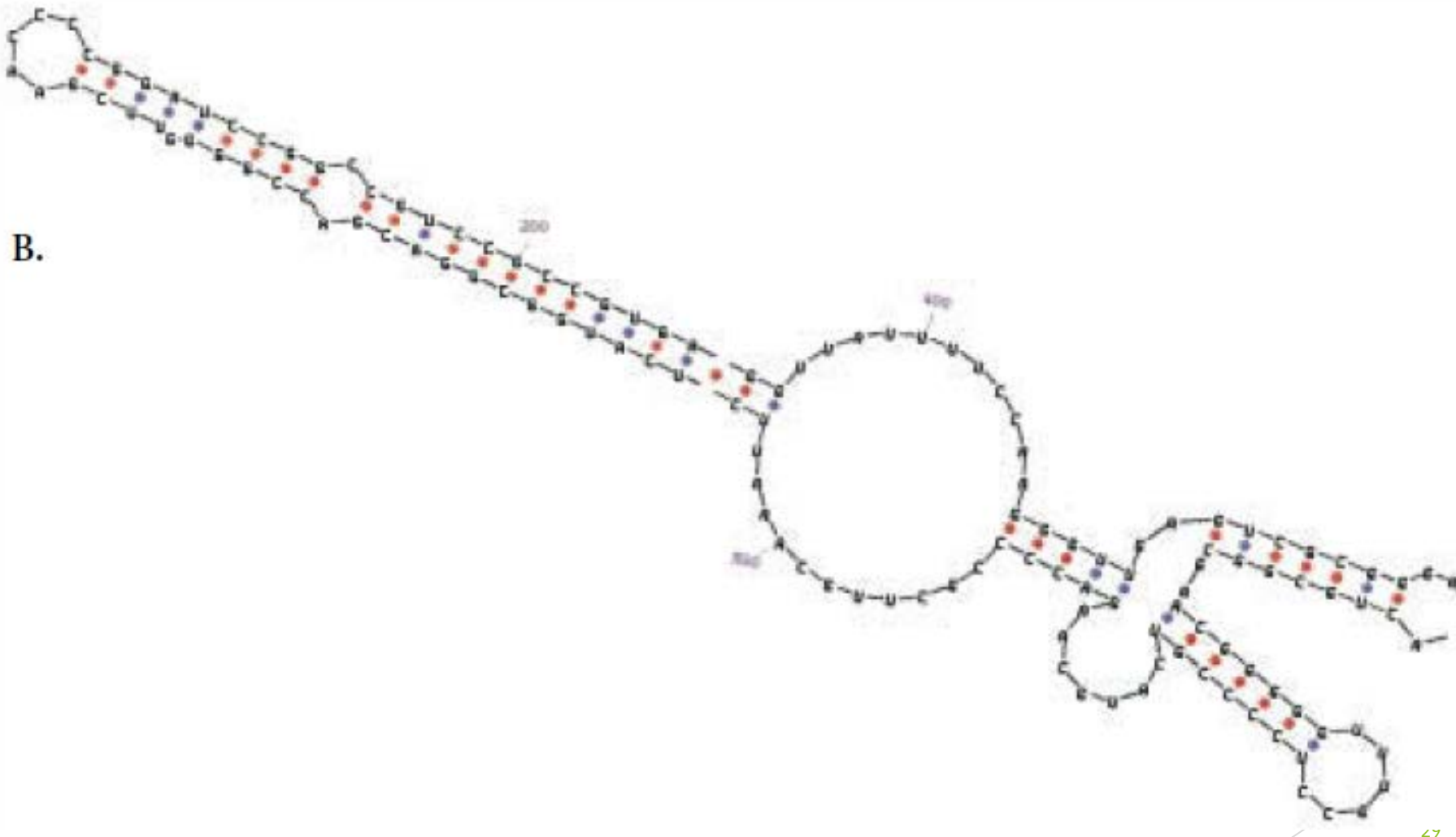
      100      110      120      130      140      150
..... --AA G      UGU C      A      G      -----
      GUGC      AAGGA AGCC      AAG GGGC CUCUCCGU GUCU      GGUGG UAA UCGCA G
      CGCG      UUCCU UCGG      UUC CUCG GAGGGGGCA CAGA      CCGCC AUU GCGCU C
..... GACC      -      -UU      -      C      A      CUGCA      -      -      -A
      290      280      270      260      250      220      210

      160      170      180
..... U      A      G      AA
      G AUCAUGGCGGACG CCGGG UUCG
      C UAGUGCCGCCUGC GGCCU AGGC C
..... -      C      -      CC
      200      190
```

Текстово представяне на една от
предсказаните вторични структури
от MFOLD



По-подробно представяне на част от предсказаните структури



Софтуерен пакет Amber

- ▶ <http://ambermd.org/>
- ▶ ***Assisted Model Building with Energy Refinement***
- ▶ "Amber" включва две основни части:
- ▶ (1) множество от молекулярни механични силови полета за симулация на биомолекули (които са публично достъпни, и се използват в широка гама компютърни симулации);
- ▶ (2) пакет от програми за молекулярни симулации, който включва сорс кодове и демонстрации
- ▶ Текущата версия на кода е Amber v 11
- ▶ Дистрибуцията на Amber съдържа 2 части: AmberTools и Amber11.
- ▶ Пакетът включва около 50 програми, които работят добре заедно.

Използване на ковариационен метод на секвенциите за предсказване на структурата

- ▶ ко-вариация – съизменение – едновременно промени във фиксирани позиции на секвенциите
- ▶ Този метод изследва секвенции на една и съща молекула РНК от различни биологични видове по отношение на позициите, които се изменят едновременно по начин, който им дава възможност да създават двойки бази във всички молекули
- ▶ При двойно-нишковите участъци на молекулите РНК, промените в секвенциите, които се осъществяват при еволюцията, не трябва да променят сдвоените бази
- ▶ При промените в примките на секвенцията и еднонишковите участъци не съществува такова ограничение
- ▶ Методът за анализ преглежда позициите в секвенцията, при които ковариациите (съизмененията) запазват свойствата на сдвоените бази непроменени

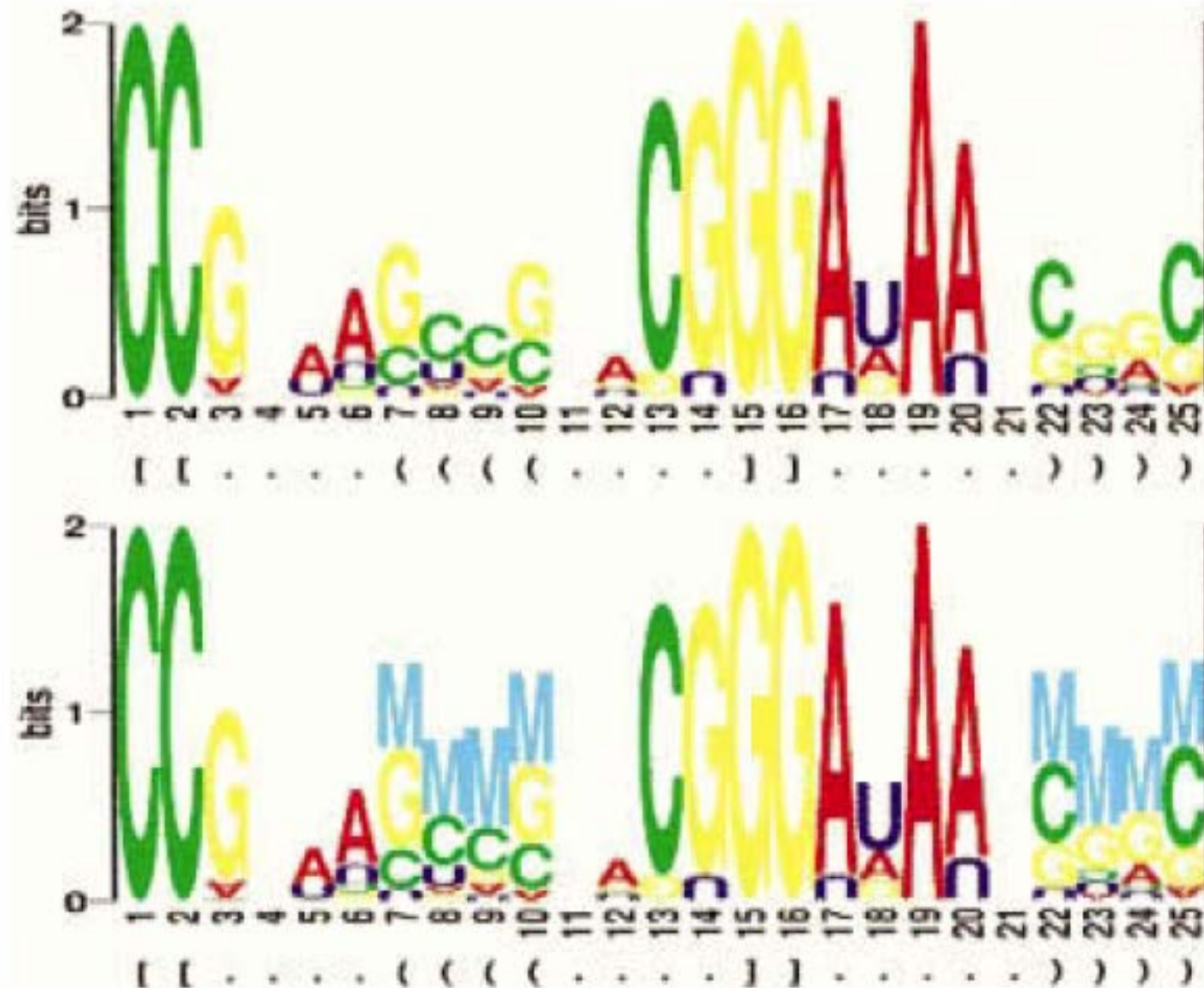
Ковариационен анализ на секвенции РНК

- ▶ Когато в определена позиция в едната база от сдвоената двойка настъпи промяна, то настъпва промяна в съответната позиция на другата база в двойката
- ▶ Напр., ако две бази G и C са сдвоени, то секвенциите, които имат C и G, или A и T или T и A в съответните позиции, също могат да се смятат за възможни съвпадения
- ▶ Ковариацията на секвенциите се използва за предсказване на термодинамичната структура

RNA structure logo

- ▶ Горният панел представя лого на секвенция, което показва размера на всяка база пропорционално на приноса на тази база по отношение на информацията в съответната колона на множественото подреждане на секвенциите
- ▶ Долният панел включва същата информация плюс информация за двойките колони
- ▶ Количеството информация се индицира от символа M , а съвпадащите колони са представени чрез множество вложени скоби.
- ▶ Всички секвенции имат C в колона 1 както и насрещно G в колона 16.
- ▶ Сходните колони 2 и 15 могат да формират втора двойка сдвоени бази върху първата двойка.

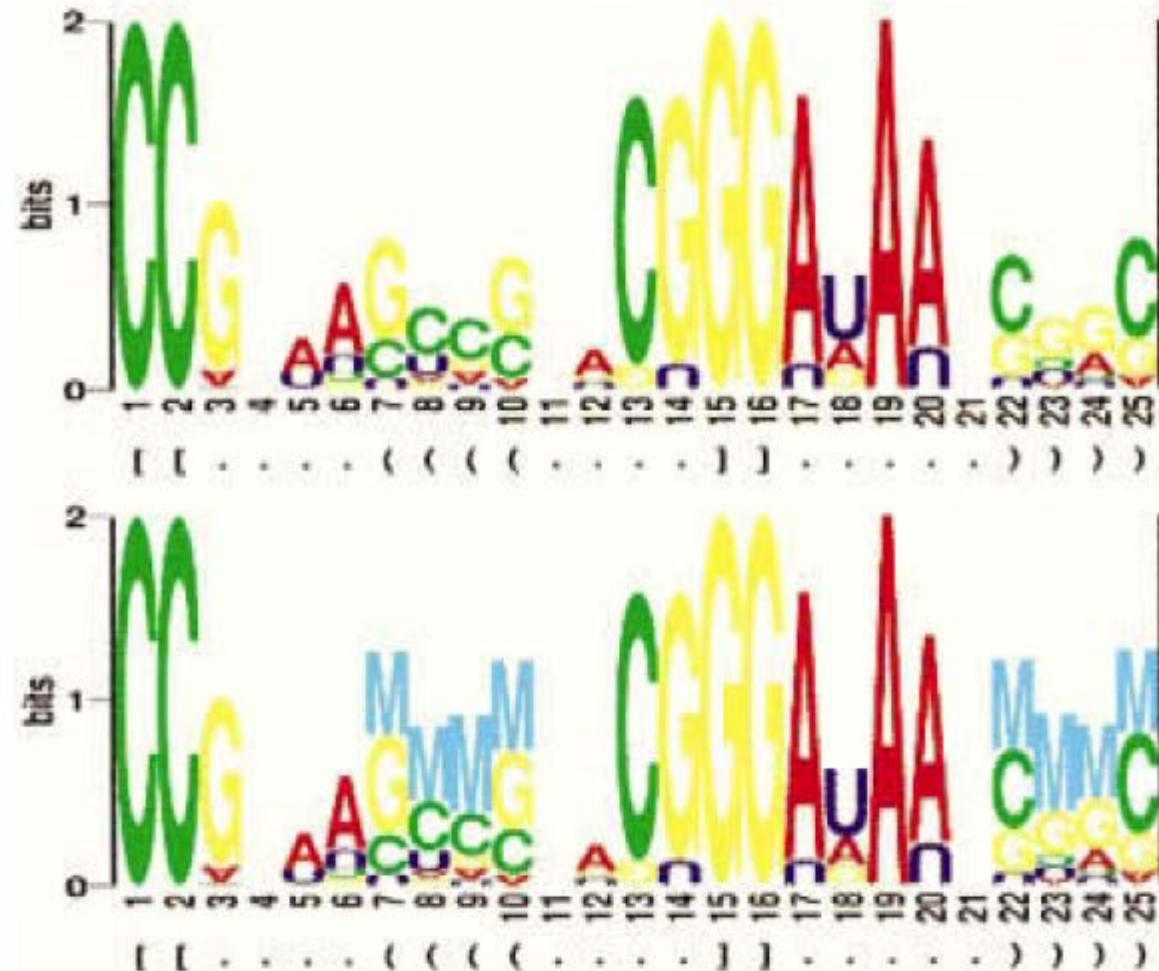
Лого на структурата на РНК



Структурно лого на РНК

RNA structure logo

- Всички секвенции имат С в колона 1, както и насрещно G в колона 16.
- Сходните колони 2 и 15 могат да формират втора двойка сдвоени бази върху първата двойка.



Структурно лого на РНК RNA structure logo

- Колони 7–10 и 25–22 също могат да формират сдвоени бази G/C.
- Секвенции с G в колона 7 често имат C в колона 25, а тези с C в колона 7 могат да имат G в колона 25.

