

# A ABSTRACÇÃO É NOSSA AMIGA

---

Prever estruturas, comparar e procurar  
motivos no RNA

[Robert Giegerich, Faculty of Technology Bielefeld University, Vienna 2007]



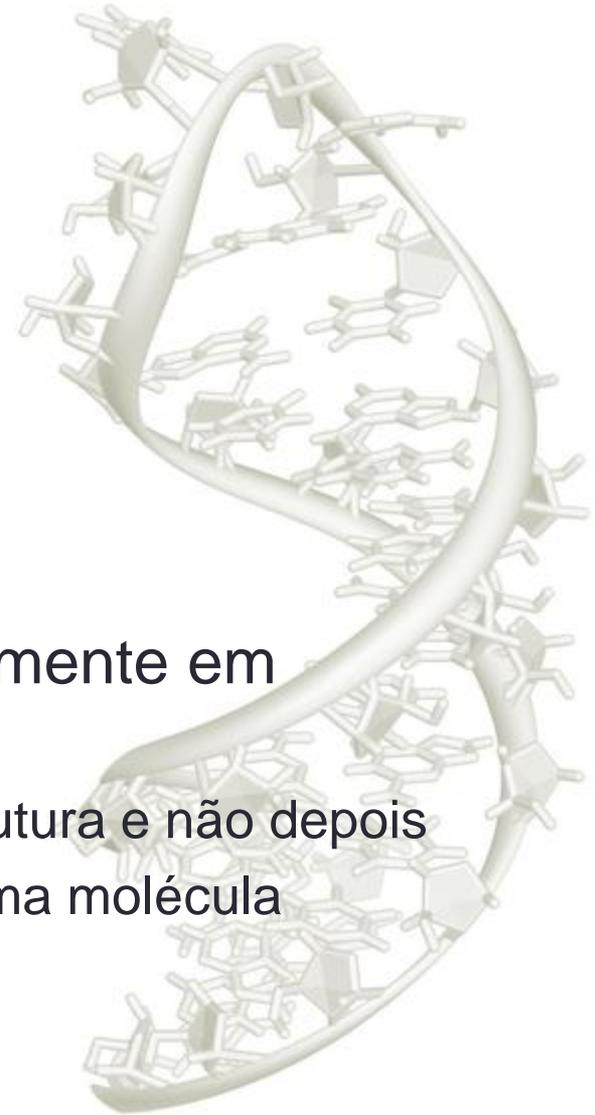
# Conteúdos

- Motivação
- RNA
  - classes, motivos e ncRNA
- Alinhamento Estrutural
- Abstracção da forma:
  - aplicações, níveis, conceitos matemáticos e conceitos gerais
- Aplicações práticas da abstracção da forma
- Conclusões



# Motivação

- Atribuição de um papel activo ao RNA
  - Novas funções atribuídas ao ncRNA
- Abstracção da forma integra-se perfeitamente em programação dinâmica
  - Pode ser aplicada durante a previsão da estrutura e não depois
  - Consegue prever a tendência estrutural de uma molécula

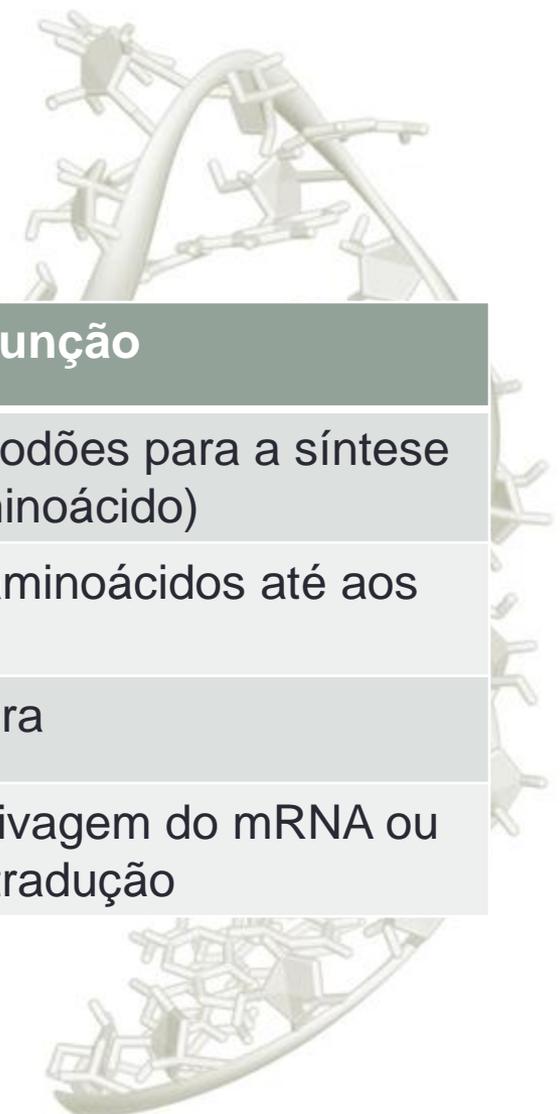


# RNA – O que é?

- Responsável pela síntese de proteínas da células
- Cadeia simples de nucleotídeos:
  - Adenina (A)
  - Guanina (G)
  - Citosina (C)
  - Uracilo (U)
- Estruturalmente complexo



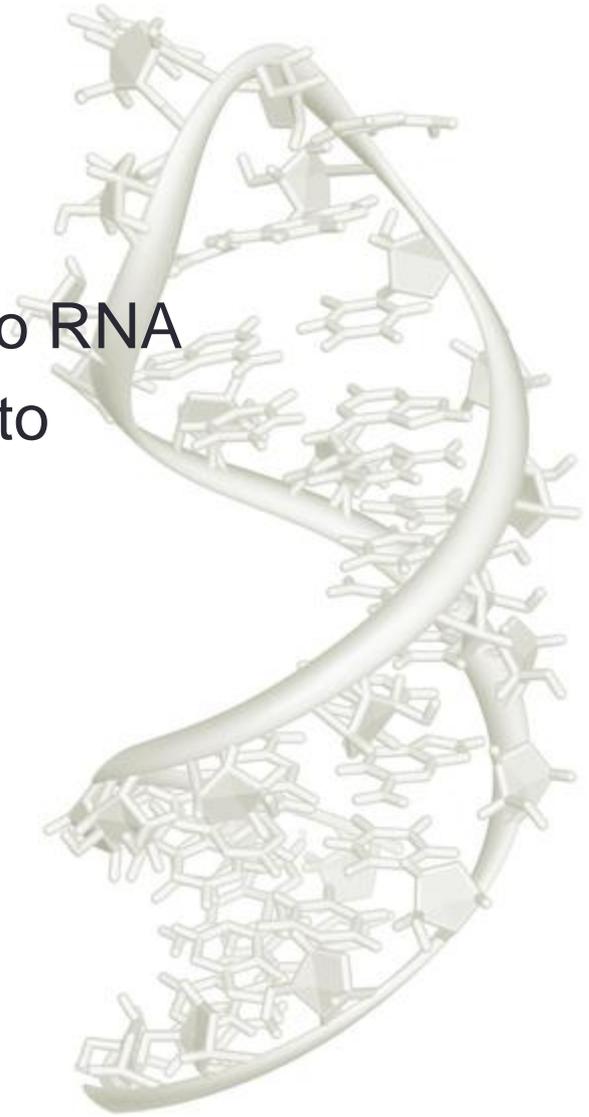
# RNA – Classes



Tipos de RNA	Função
RNA Mensageiro (mRNA)	Informação em codões para a síntese de proteínas (aminoácido)
RNA Transferência (tRNA)	Transporte dos aminoácidos até aos ribossomas
RNA não codificante (ncRNA)	Função reguladora
Micro RNA (miRNA)	Direccionam a clivagem do mRNA ou reprimem a sua tradução

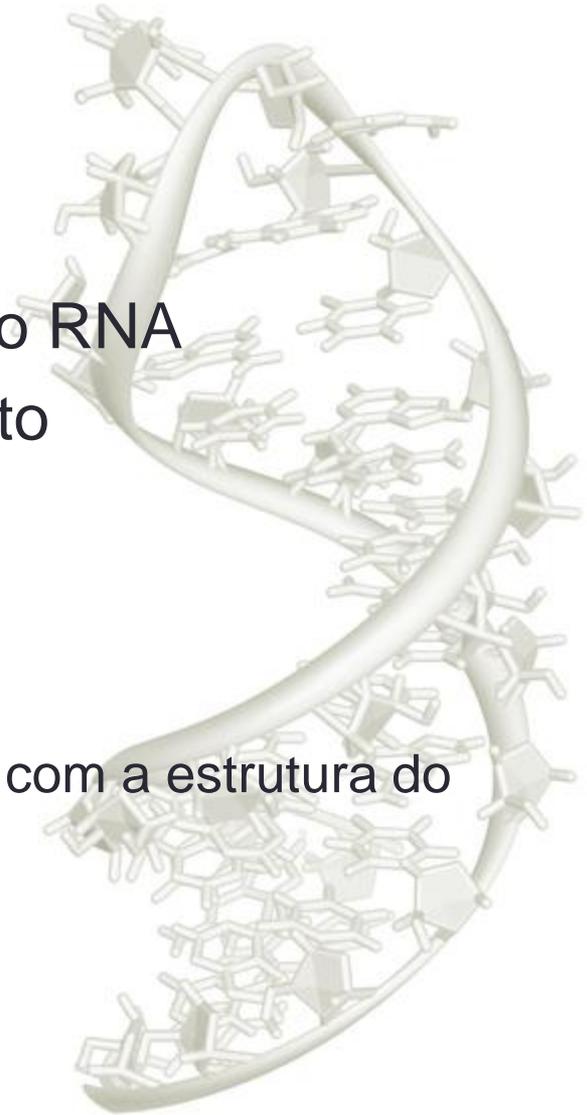
# Motivos RNA

- Conferem funcionalidades específicas ao RNA
- Podem ser determinados por alinhamento



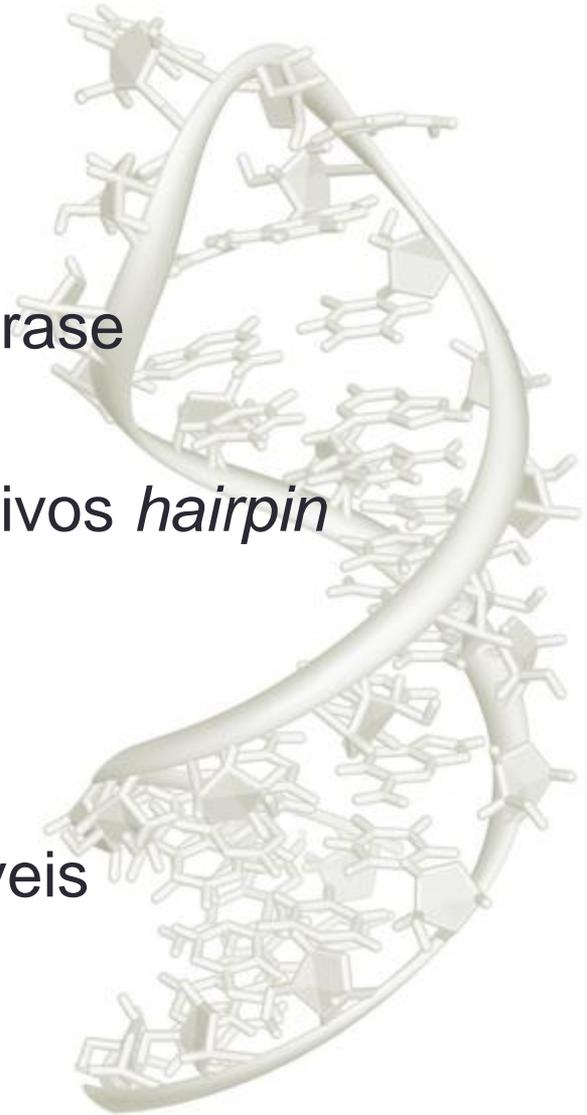
# Motivos RNA

- Conferem funcionalidades específicas ao RNA
- Podem ser determinados por alinhamento
- **PROBLEMA:**
  - Grande parte dos motivos estão relacionados com a estrutura do RNA
  - **MAS** a informação encontra-se na sequência



# Características do ncRNA

- Papel importante na síntese pela polimerase
- São, geralmente, terminadores com motivos *hairpin*
- Estrutura secundária bem definida
- Motivos de sequência e estrutura imutáveis

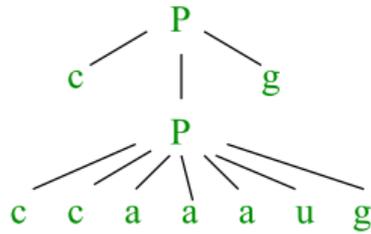
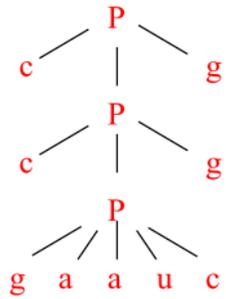


# Novas técnicas

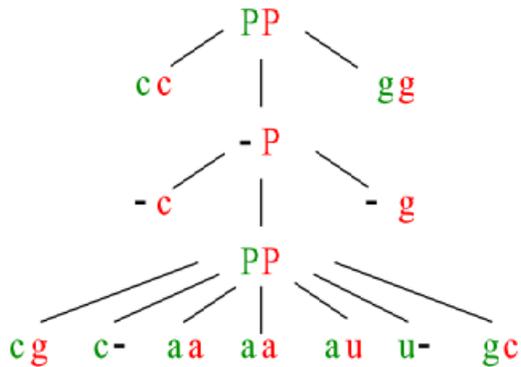
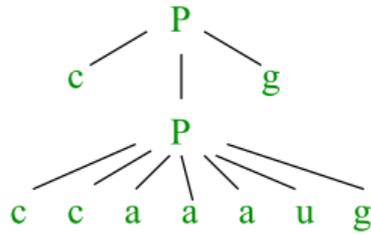
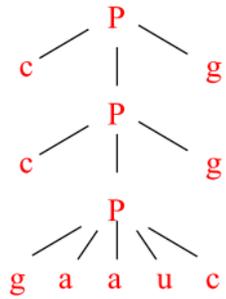
- Motivos 3D
- Avaliação sistemática
  - utilizadas ferramentas de avaliação independentes (modelo Rfam considerado *standard*)
- Previsão de genes
  - RNAz e EVOFOLD
- *Clustering* por similaridade estrutural
- Localização de motivos utilizando *Matchers* termodinâmicos.
  - Locomotif



# Alinhamento Estrutural



# Alinhamento Estrutural



c-ccaaaug-g  
 (-.....)-  
 (((-...-)))  
 ccg-aau-cgg



# Abstracção da Forma – Aplicações

- Prever alternativas estruturais
- Classificar por forma dominante
- Prever estrutura, por comparação
- Acelerar pesquisas

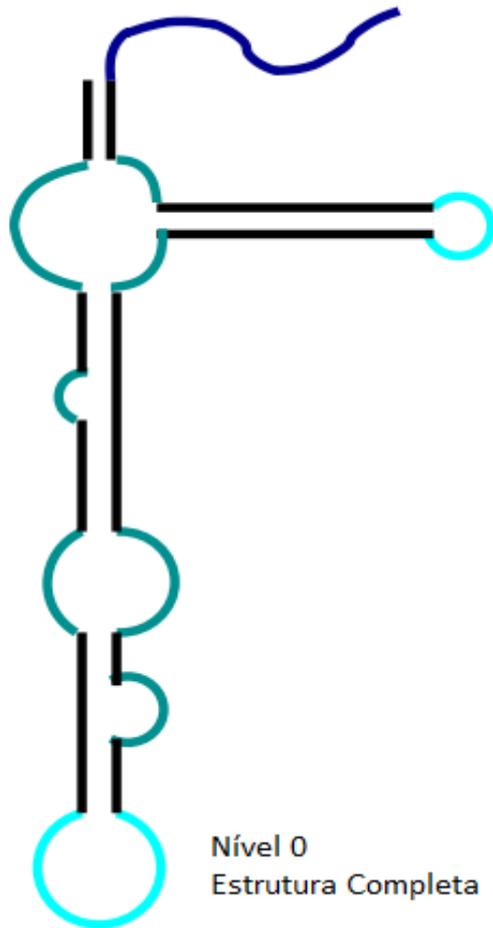


# Abstracção da Forma

- Mantém nidificação e adjacência das hastes
- Despreza os tamanhos
- Mantém ou ignora a presença e o tipo de *loops* internos e/ou terminais



# Níveis de Abstracção



Nível 0  
Estrutura Completa

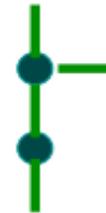
Nível 1  
*Loops*



Nível 3  
Cortes de hélices



Nível 4  
*loops* internos e múltiplos



Nível 5  
arranjo das hastes



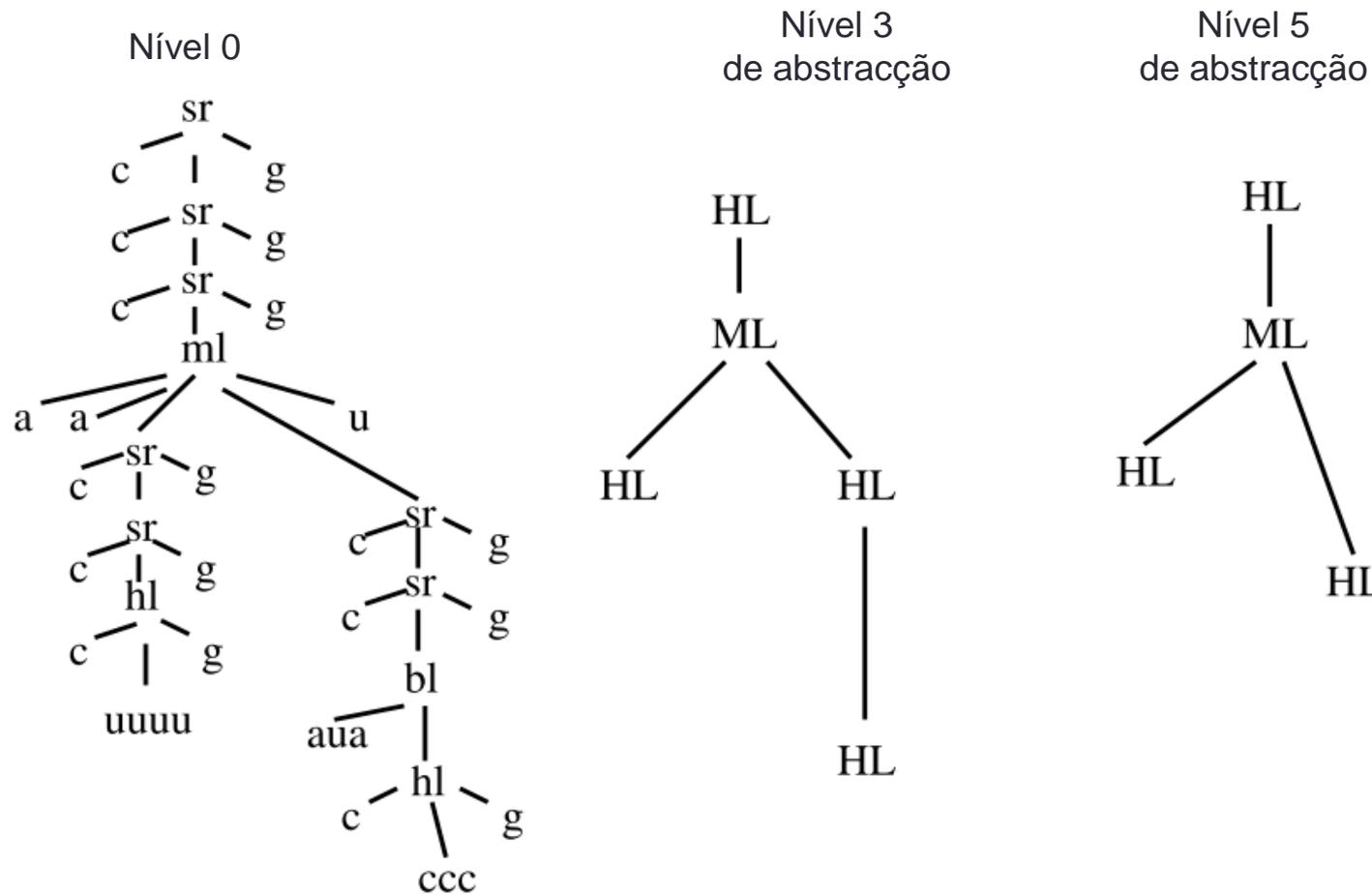
Nível 6  
ignorar *hairpins*



Nível 7

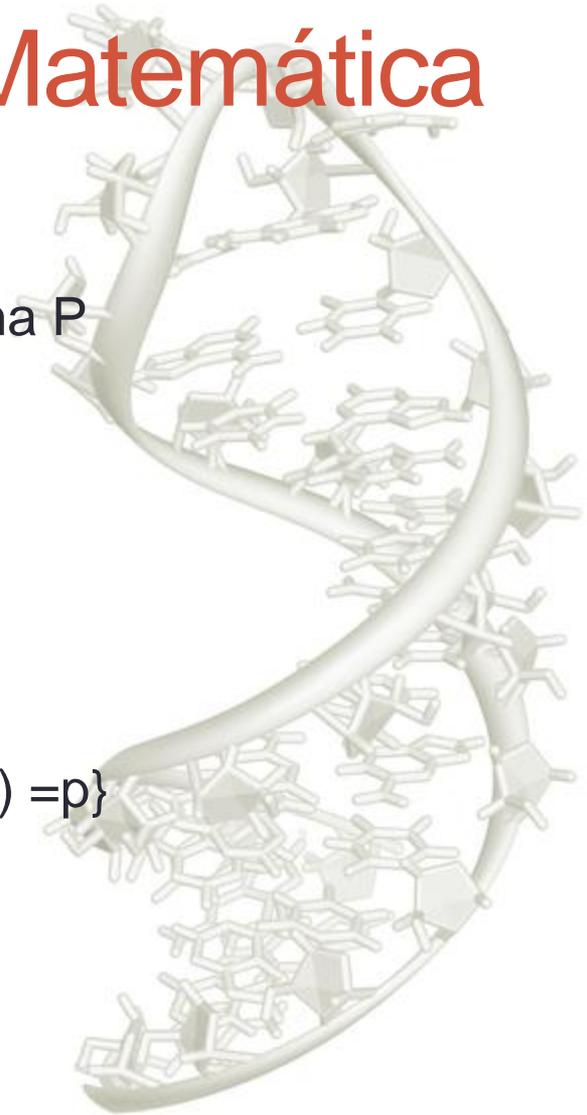


# Níveis de Abstracção – Exemplo



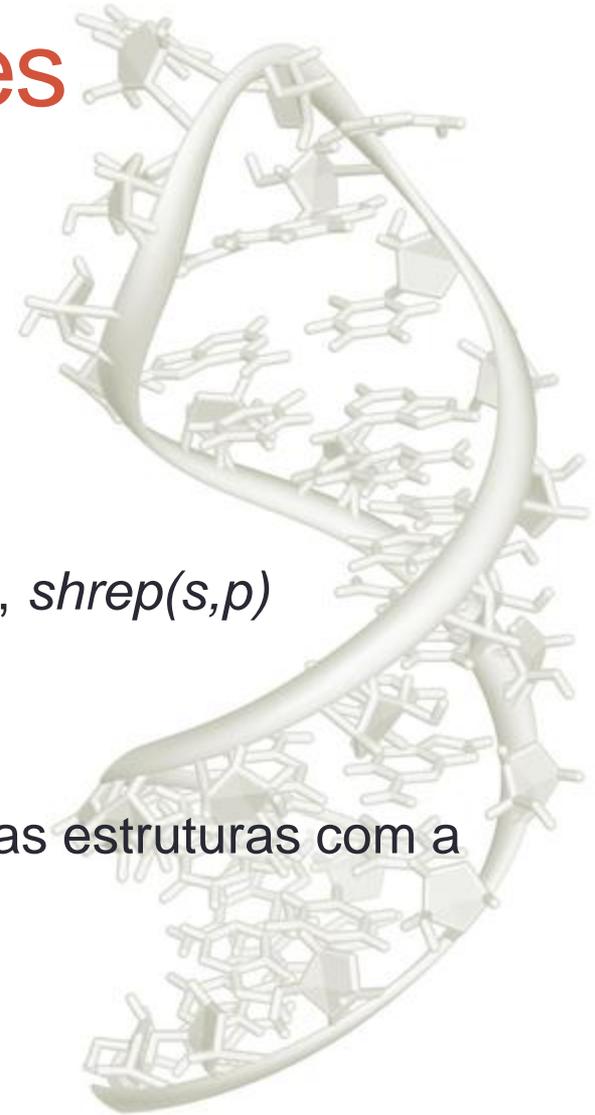
# Abastracção da Forma – Matemática

- Generalidade:
  - Domínio do tipo árvore com estrutura  $F$  e forma  $P$
  - Homomorfismo na árvore  $\pi : F \rightarrow P$
- E para cada sequência de RNA  $s$ :
  - *folding space* de  $s$ :  $F(s)$
  - *shape space* de  $s$ :  $P(s) = \pi(F(s))$
  - *Shape class* de  $p \in P(s)$ :  $f(x,p) = \{x \mid x \in F(s), \pi(x) = p\}$



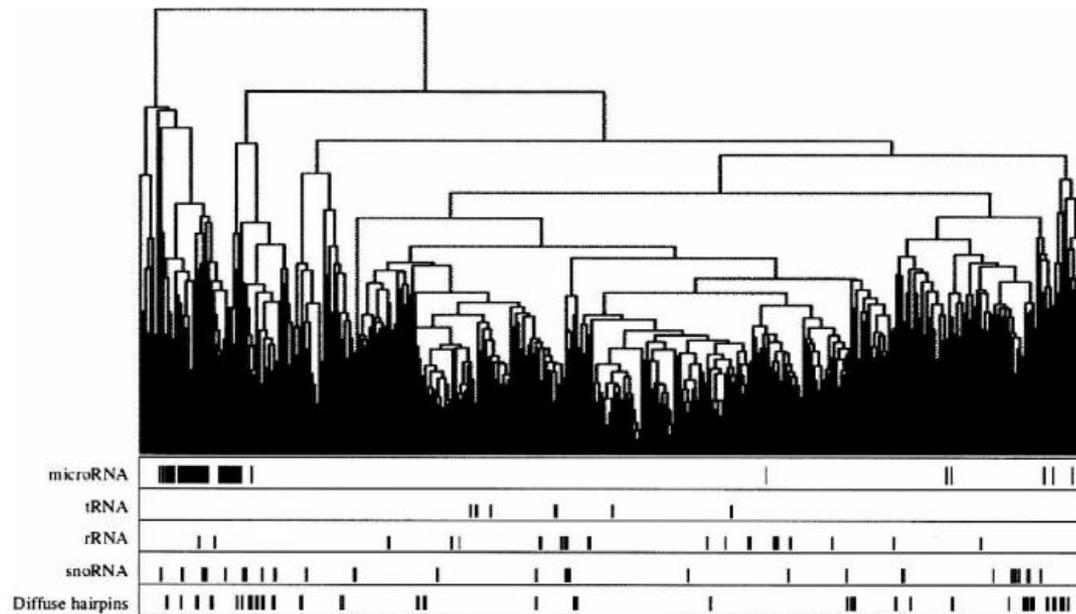
# Três Conceitos Importantes

- Função de abstracção da forma
- Estrutura representativa da forma
  - *shrep* = classe com o mínimo de energia livre,  $shrep(s,p)$
- Probabilidade da forma:
  - *Prob(p)* = probabilidade acumulada de todas as estruturas com a forma  $p$



# Aplicação Prática I

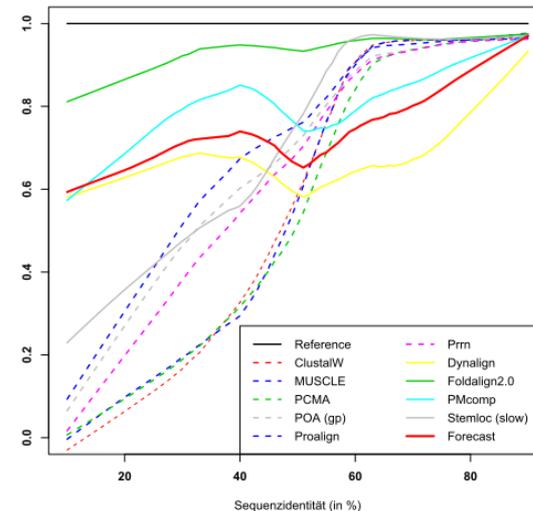
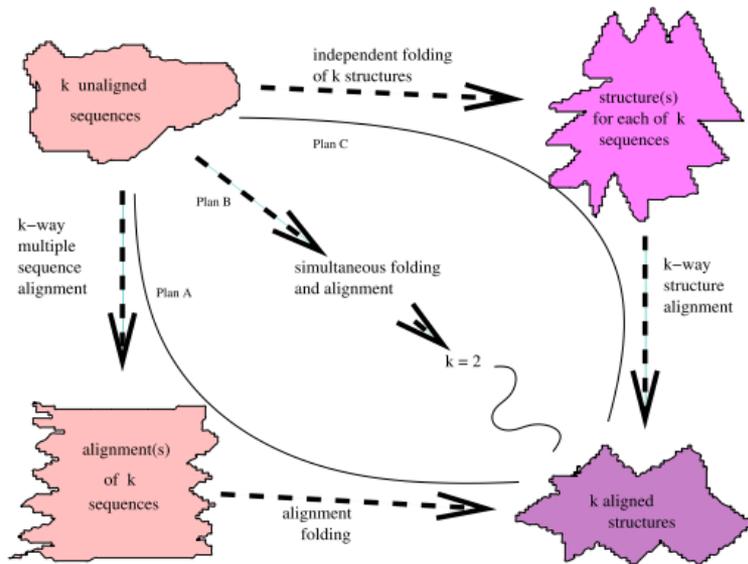
- Previsão de estruturas alternativas para a classificação do miRNA



- Classificação utilizando um top 3 ranking de *shreps*, um *cluster* de miRNA e um cluster de *hairpins* difusos

# Aplicação Prática I

- Previsão de estruturas alternativas para a classificação do miRNA



- Utilizando um top 100 de formas para cada sequência, para um *matching* mais correcto

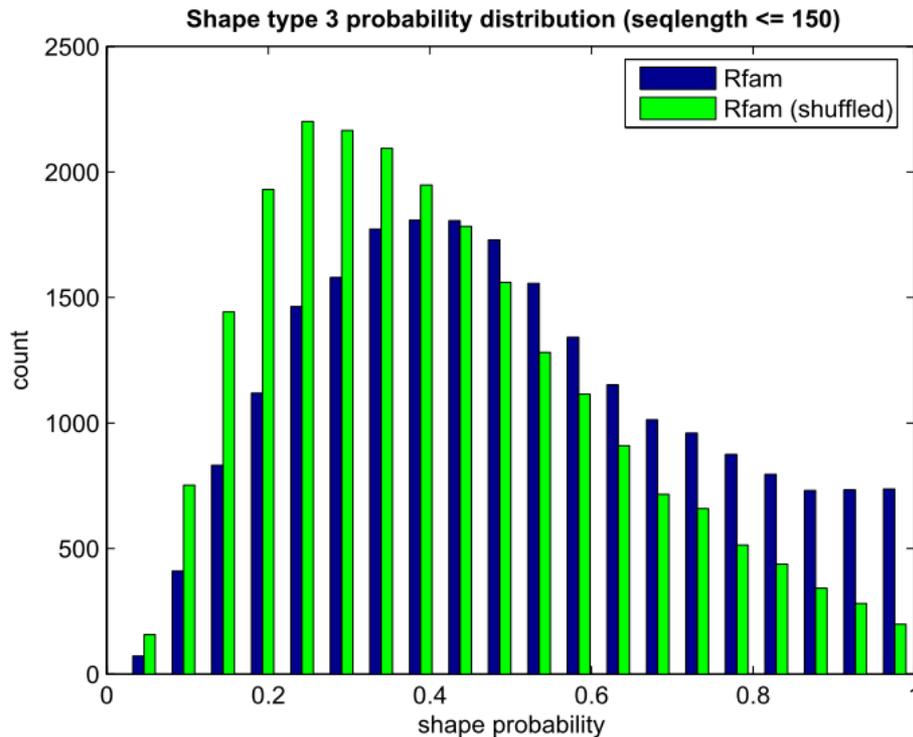
# Aplicação Prática II

- A forma  $p$  é dominante se  $Prob(p) > 0.5$
- Domínio da forma:
  - Correlacionado com o MFE
  - Independente da composição da sequência
  - Independente do tamanho da sequência



# Aplicação Prática II

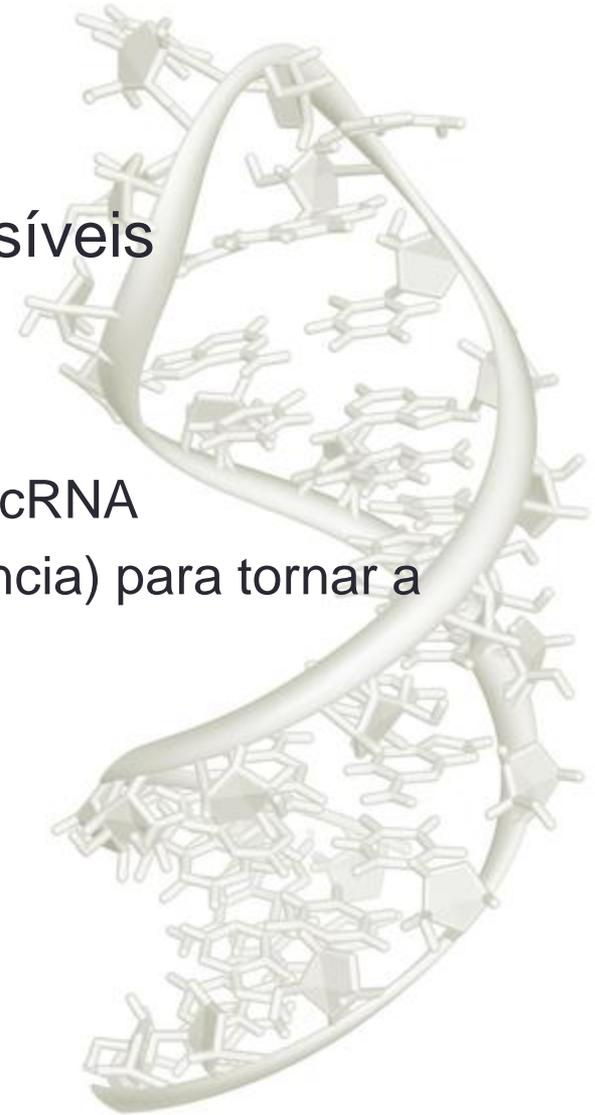
- No nível 3 de abstracção, os valores de probabilidade mais elevados





# Aplicação Prática IV

- Tornar as pesquisas usando Rfam acessíveis
- RAVENNA
  - procura estruturas homologas (motivos) em ncRNA
  - Utiliza abstracção do CM (modelo de covariância) para tornar a pesquisa mais rápida



# Aplicação Prática IV

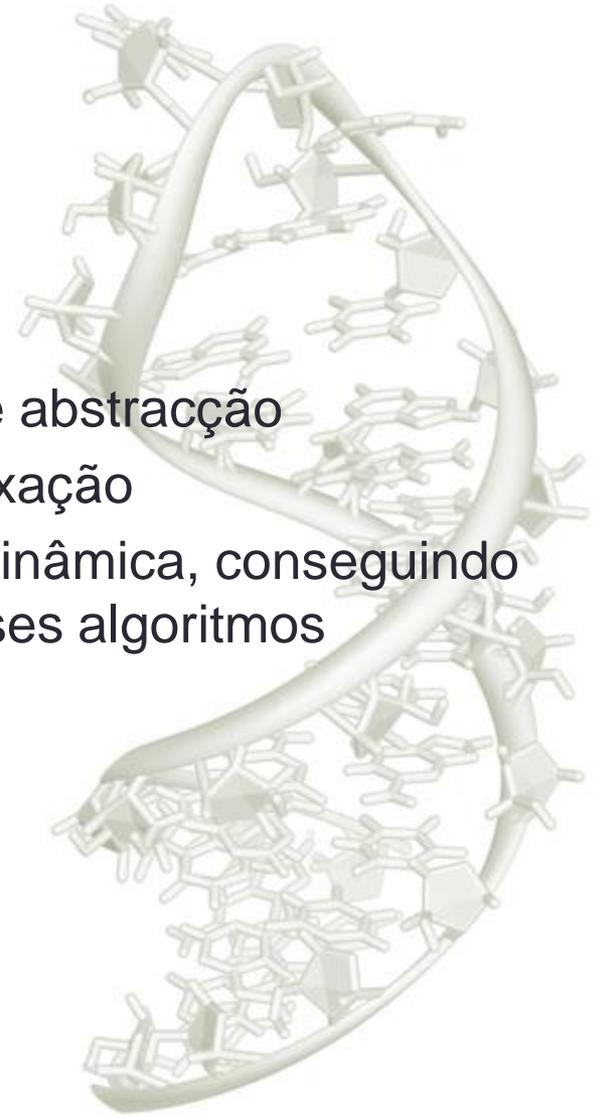
- Tornar as pesquisas usando Rfam acessíveis
- RAVENNA
  - procura estruturas homologas (motivos) em ncRNA
  - Utiliza abstracção do CM (modelo de covariância) para tornar a pesquisa mais rápida
- **Nova aplicação: *RNA sifter***
  - *Rfam shape index*: computado para cada Rfam
  - *Query shape spectrum*: computado com complexidade  $O(kn^3)$
  - Comparação de índices:  $O(kl)$ , onde  $l$  é o tamanho da sequência





# Conclusão

- Com a abstracção da forma:
  - Possibilidade de escrever a própria função de abstracção
  - Aplicar, *a posteriori*, para classificação e indexação
  - Integrar com os algoritmos de programação dinâmica, conseguindo uma forma estrutural *sensata* para aplicar esses algoritmos
- Não necessita de heurísticas



# Questões?

