

PCC146 - Reconhecimento de Padrões

Sávio Gonçalves Carvalho

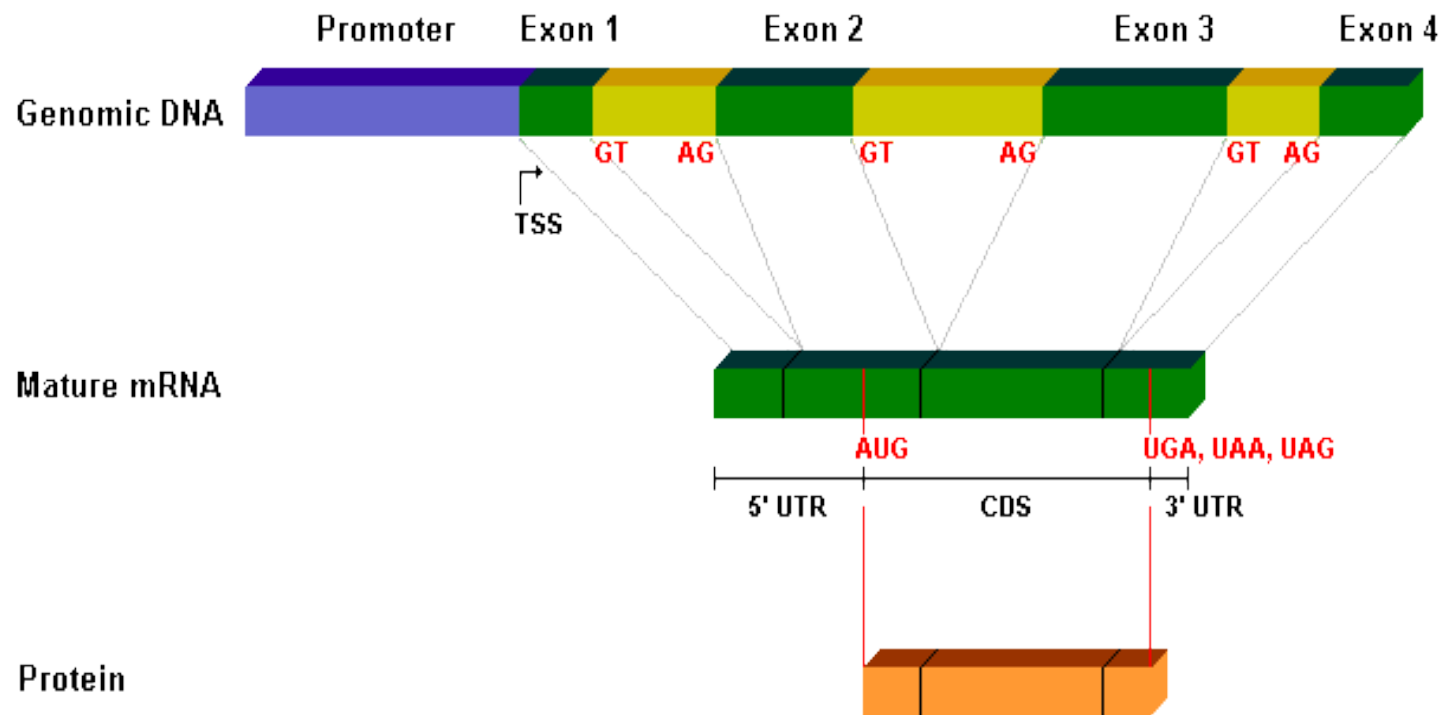
1º Seminário sobre estado da arte:

*“Large-scale structural analysis of the core
promoter
in mammalian and plant genomes”*

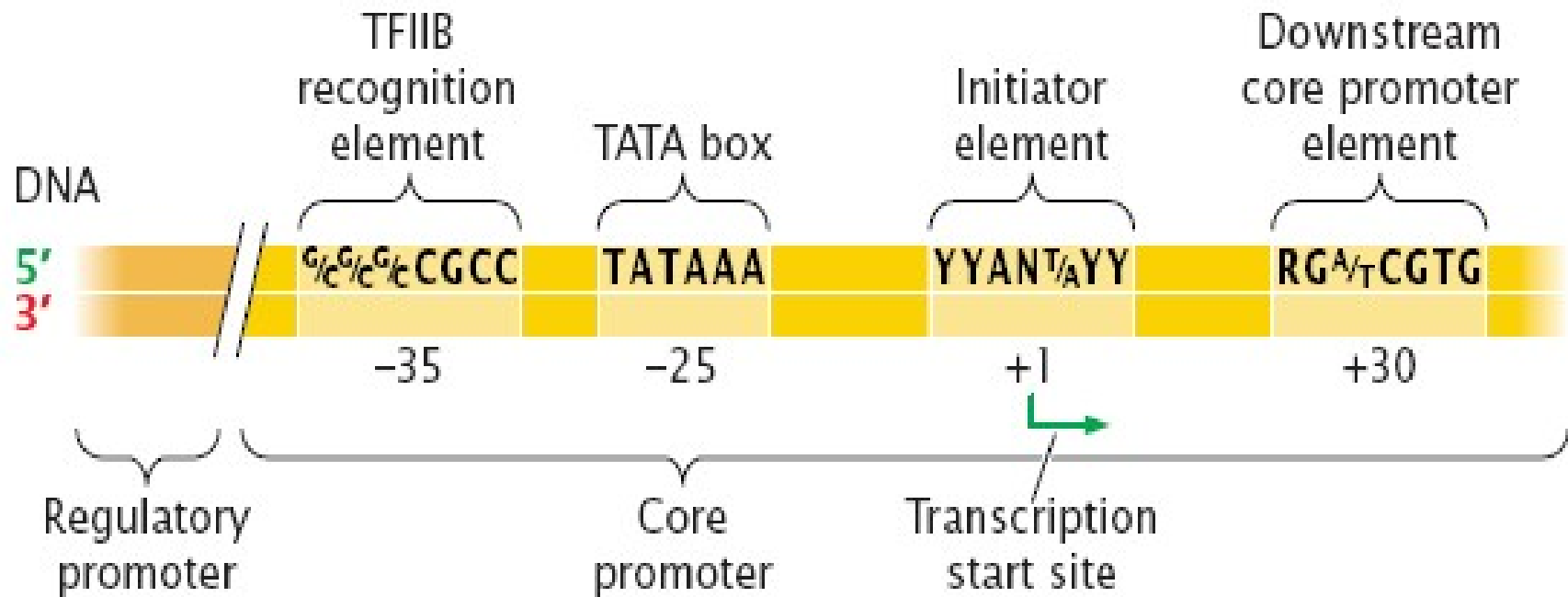
Kobe Florquin, Yvan Saeys, Sven Degroeve, Pierre Rouze and Yves Van
de Peer

Nucleic Acids Research, 2005, Vol. 33, No. 13 4255–4264

Introdução

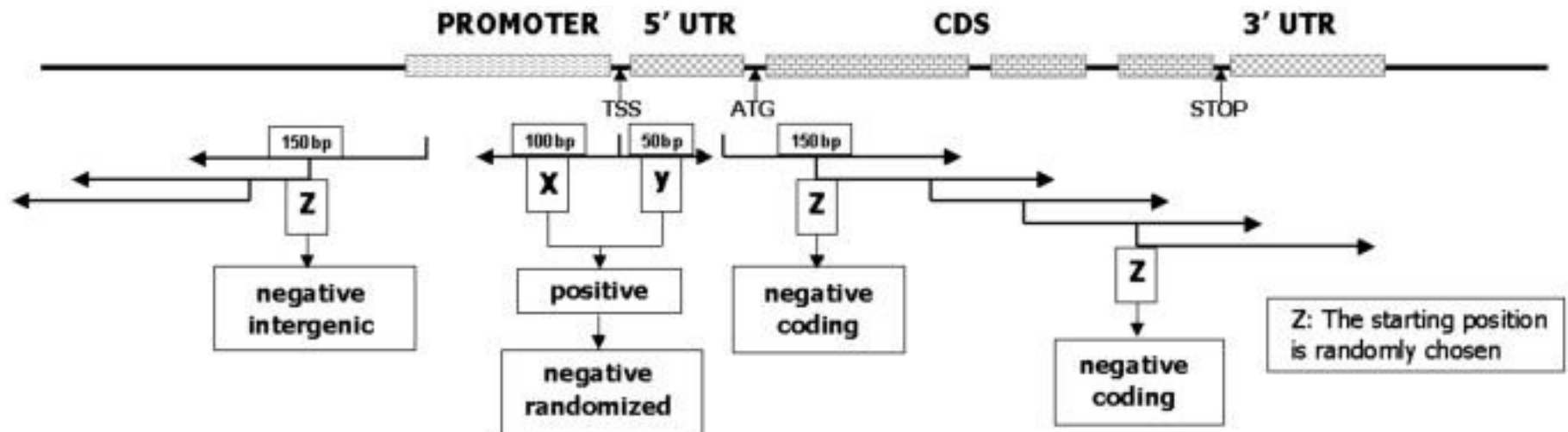


Introdução



Core-promoter datasets

- Tamanho utilizado: 150 pb upstream e 50 downstream
- Base positiva:
 - DBTSS (8793 para humano e 6875 para rato)
 - ARAPROM (7088 para Arabdopsis)
 - RICEPROM (2195 para arroz)
- Base negativa:
 - Regiões codificantes, intergênicas ou embaralhadas

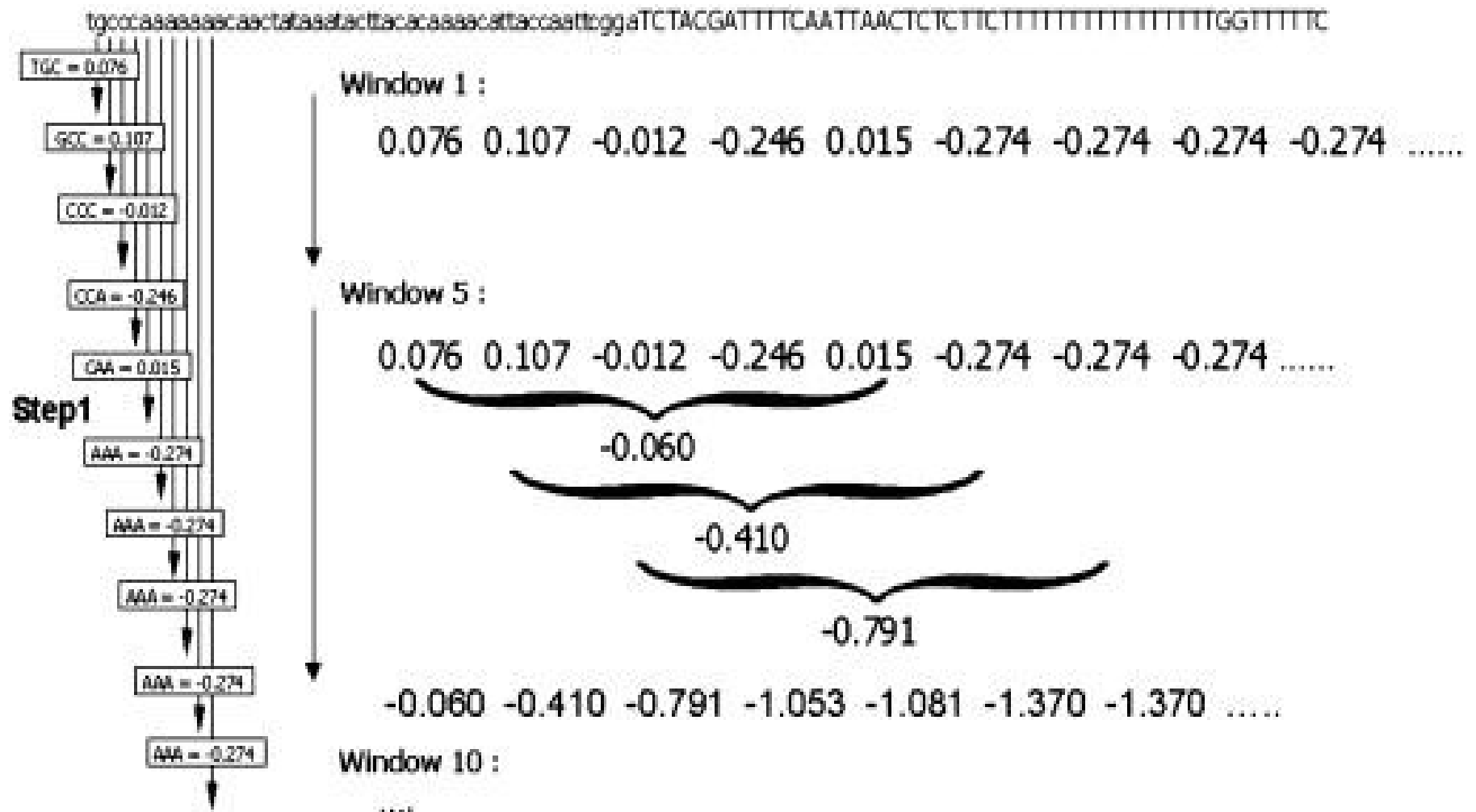


Structural models

- Perfis estruturais que podem ser calculados através da interação entre nucleotídeos.
- Utilizam 2 ou 3 nucleotídeos
- Ex: Stacking energy, Propeller twist, Nucleosome position preference, Bendability, Duplex disrupt energy, Duplex free energy, DNA denaturation, etc.

Structural models

- Trinucleotide bendability model



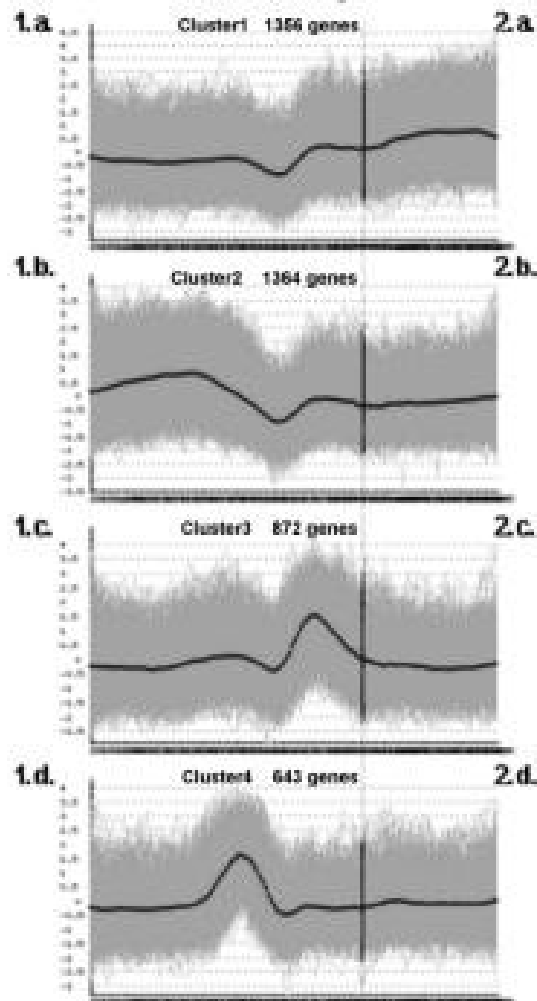
Clustering and classification

- Método de clusterização: ABQC (adaptive quality-based clustering)
 - Motivações para utiliza-lo
 - Não necessita de estipular número de clusters
 - Não força a entrada de toda sequência em algum dos clusters
 - As sequências convertidas são utilizadas
- Resultados:
 - Aproximadamente a mesma quantidade de clusters gerados para cada perfil estrutural
 - Sequências de um mesmo cluster apresentam baixa (ou nenhuma) similaridade entre si

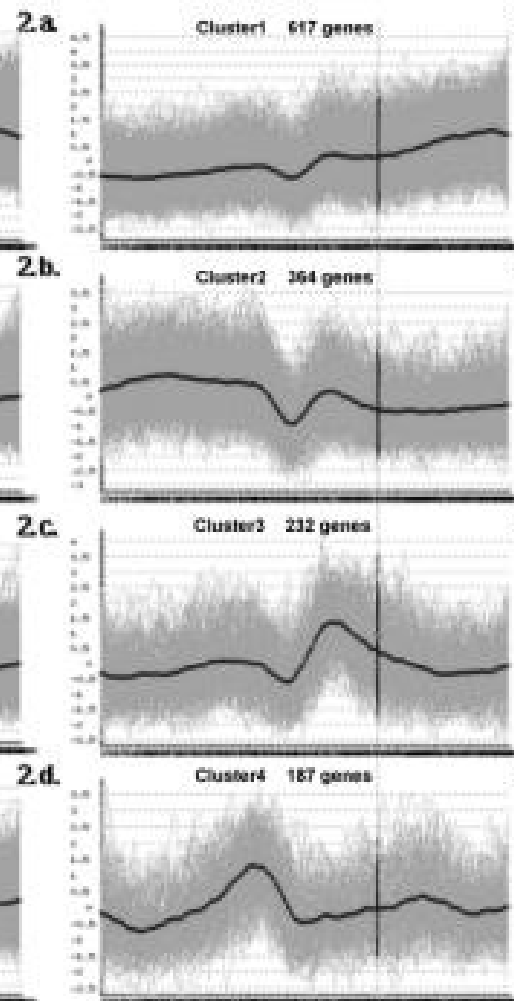
Structural models

- Duplex disrupt energy
- Window size 10

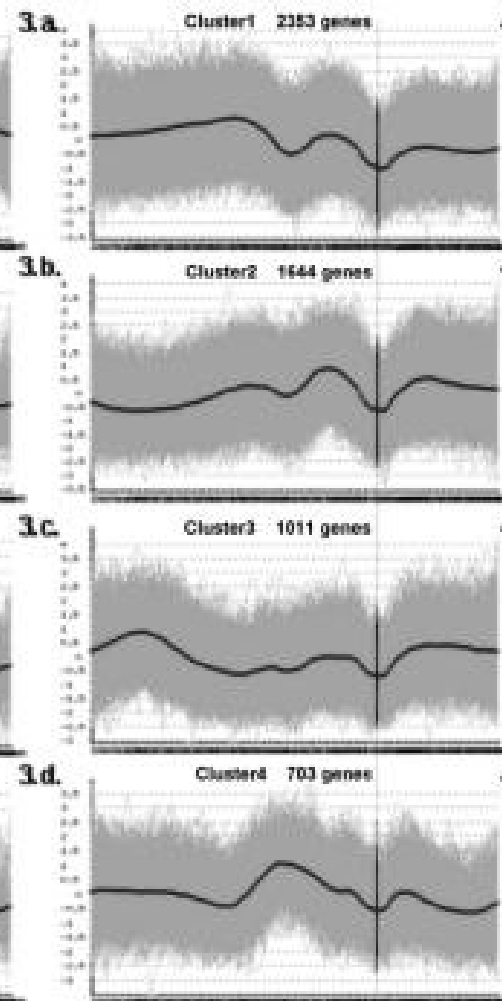
Arabidopsis



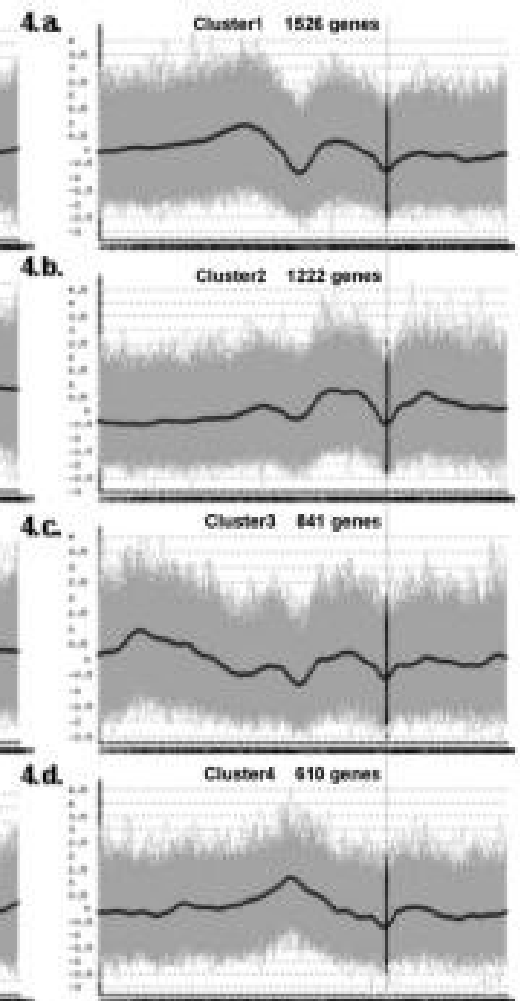
Rice



Human



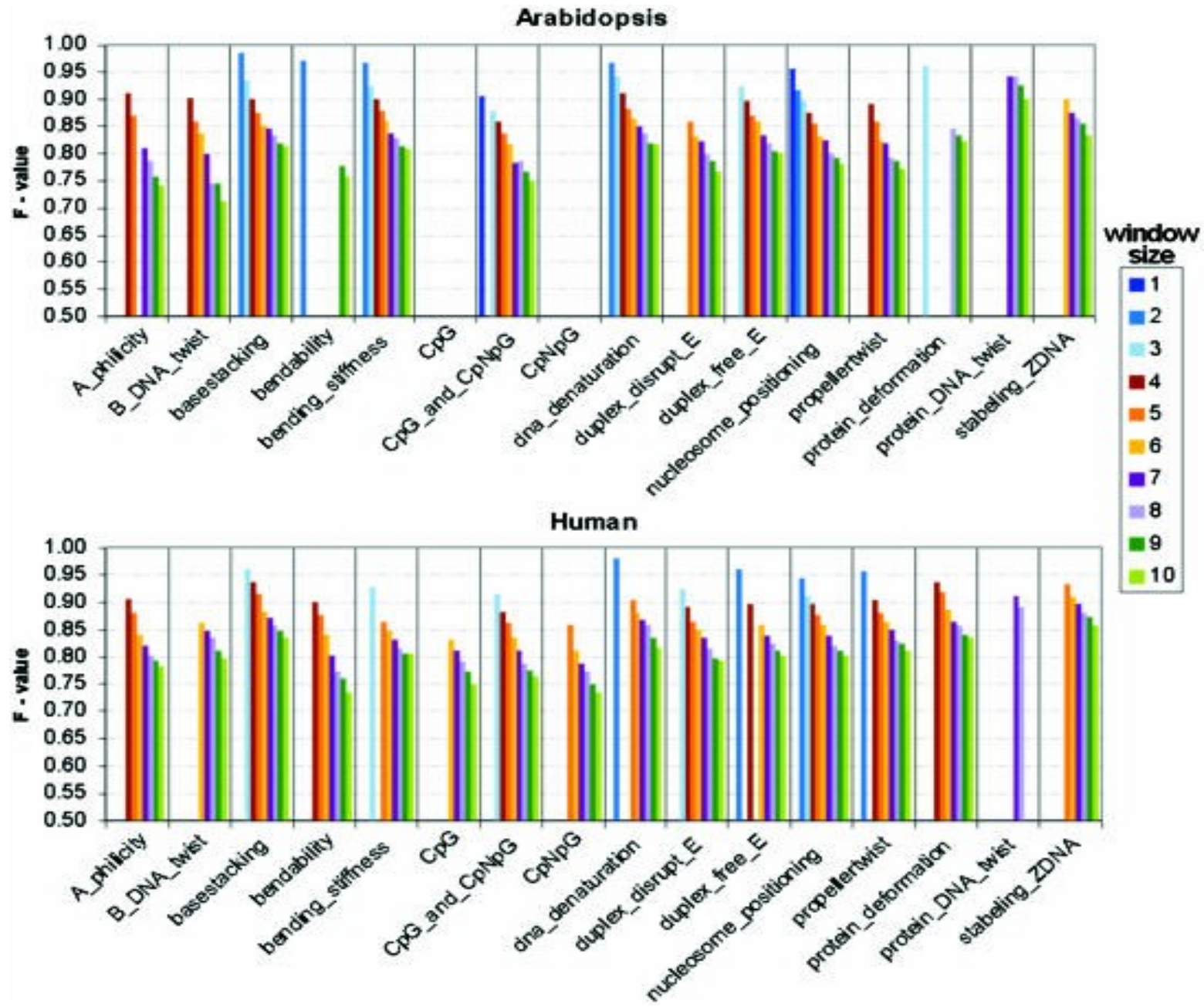
Mouse



Clustering and classification

- Método de Classificação: Linear SVM
 - Classificar nova sequência como promotor ou não-promotor
 - Base positiva: Sequências do cluster de melhor qualidade para cada perfil estrutural (25%)
 - Janela de tamanho 5 → 6 são capazes de diferenciar core-promoters de não-core-promotores
 - A-philicity
 - DNA bending stiffness
 - DNA denaturation
 - duplex disrupt energy
 - nucleosome positioning preference
 - propeller twist
 - Performance para os modelos foi de cerca de 82%

Clustering and classification



Conclusão

- A utilização de modelos estruturais é capaz de melhorar as ferramentas de classificação de promotores, assim como a consideração da existência das diferentes classes