

Predikce genů kódujících proteiny

- **Prokaryotické geny**
 - Nepřerušované úseky DNA mezi **startovním kodonem** (ATG, GTG, TTG, CTG) a **stop kodonem** (TAA, TGA, TAG).
- **Eukaryotické geny**
 - Přerušovány **introny**. Průměrná délka exonu je 50 kodonů, některé jsou mnohem kratší.
 - Některé introny extrémně dlouhé, geny zabírají mbp v genomové DNA.

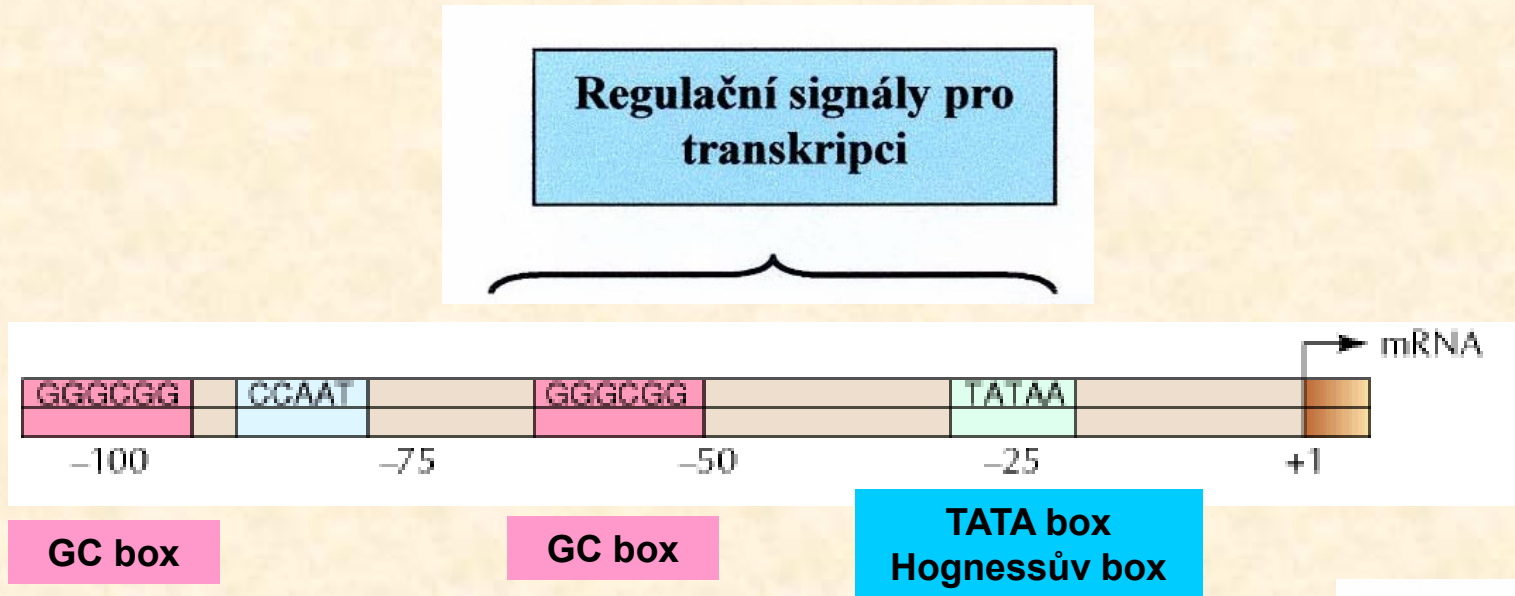
**Predikce eukaryotických genů je
mnohem složitější než predikce
genů prokaryotických a
představuje **STÁLE**
NEVYŘEŠENÝ problém!**

Eukaryotické geny

Jednobuněčná eukaryota

- **Genomy jednobuněčných eukaryot se výrazně liší** (frekvence intronů, jak velká část genomu je tvořena geny kódujícími proteiny).
- *Saccharomyces cerevisiae* – 67% genomu je protein-kódující, jen 4% obsahují introny.
- Hlenky – průměrný gen obsahuje 3,7 intronu.
- **Pro některá jednobuněčná eukaryota (kvasinky) je možné použít stejné postupy jako pro prokaryota.**

Translační a transkripční signální sekvence



Promotor RNA-polymerasy II



(gcc)gccRccAUGG

Kozak sequence
Sekvence Kozakové

Eukaryota

Eukaryotické geny

Mnohobuněčná eukaryota

- **Rozpoznání exonů/intronů**

Identifikace míst sestřihu: **GT** na 5 konci, **AG** na 3 konci.

- **Chyby při rozpoznávání exonů/intronů**

Velké množství chyb. Dlouhé introny – určeny jako intergenové úseky. Krátké intergenové useky – určeny jako introny.