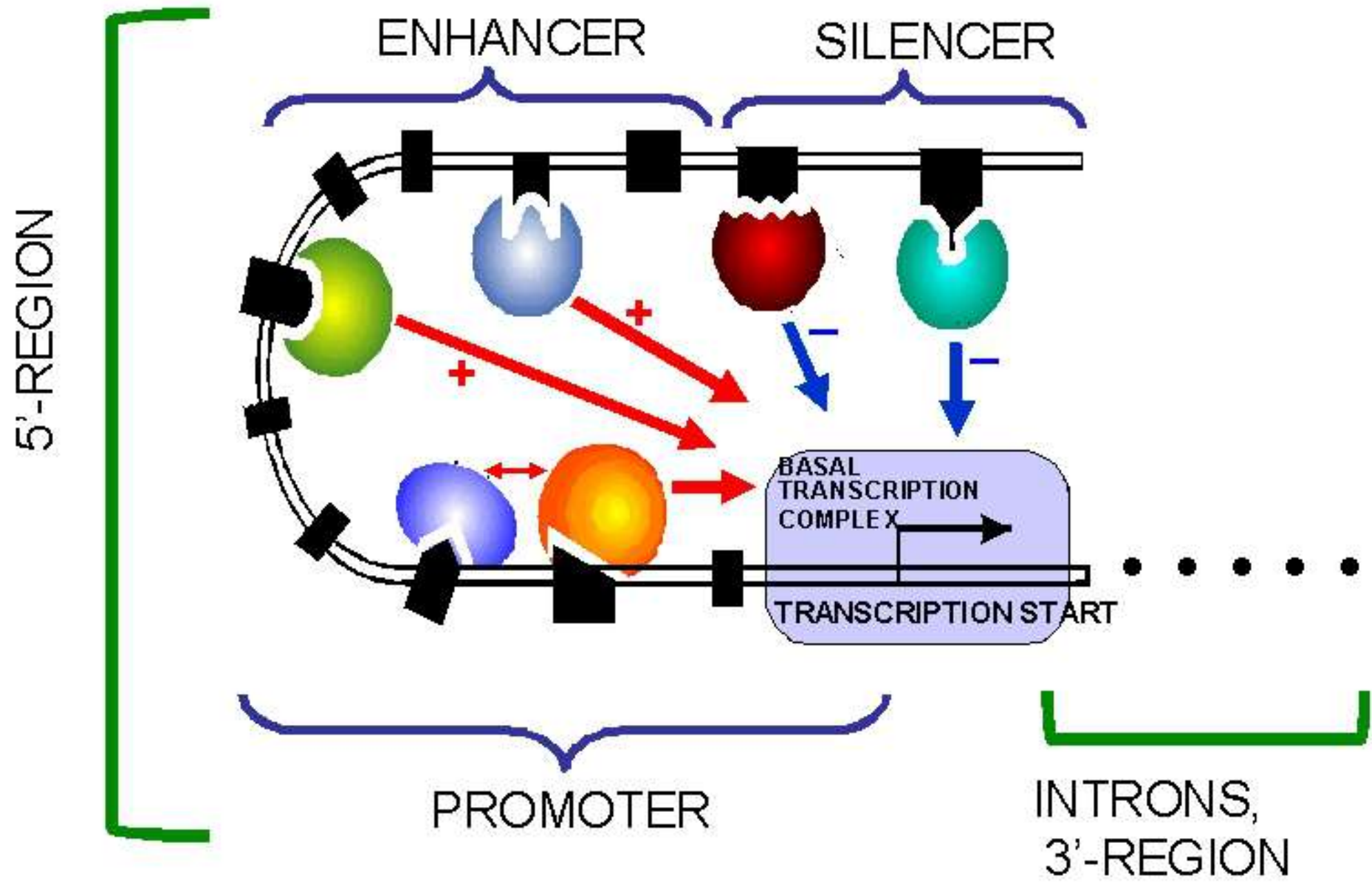
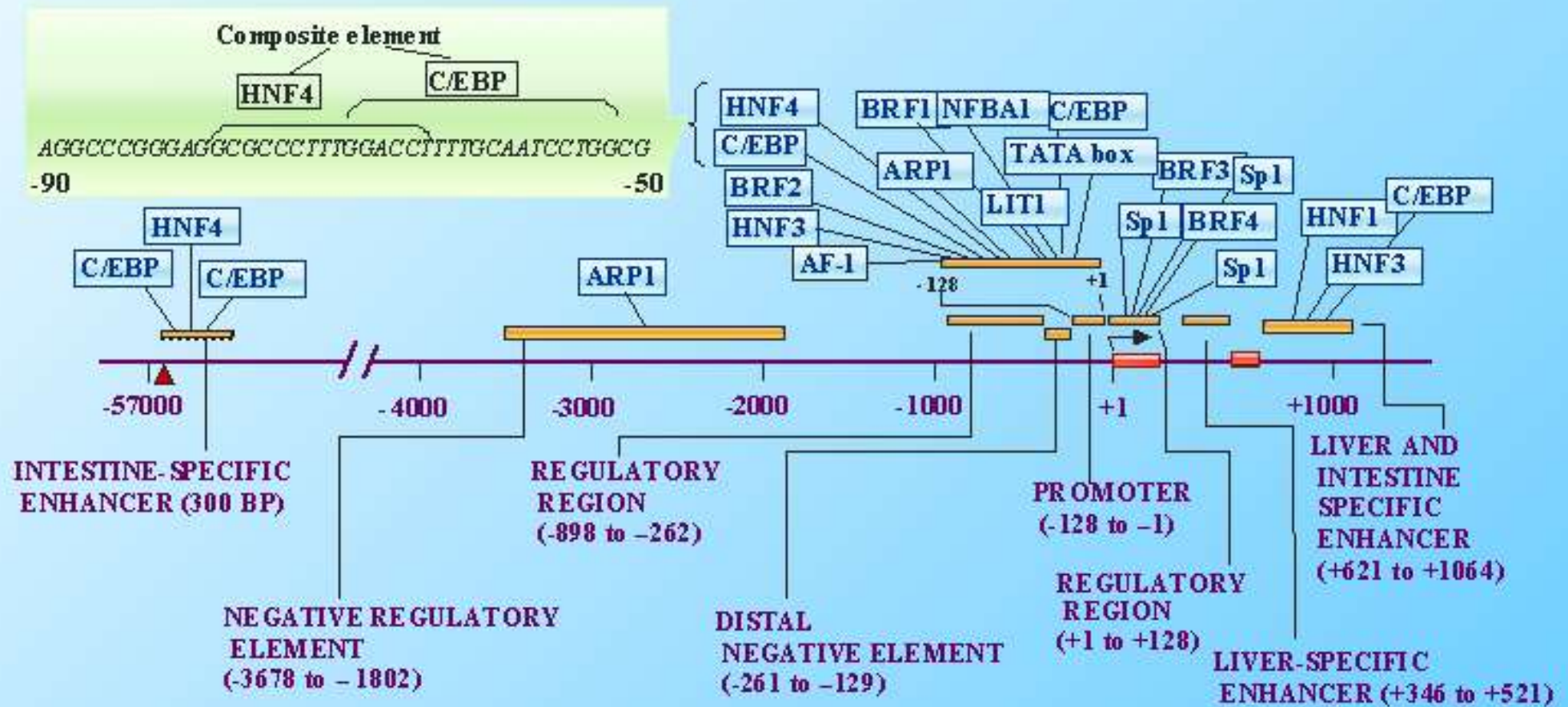


GENERAL MODEL OF EUKARYOTIC GENE TRANSCRIPTION REGULATION



ORGANIZATION OF THE REGULATORY REGIONS CONTROLLING TRANSCRIPTION OF THE GENE FOR HUMAN APOLIPOPROTEIN B.

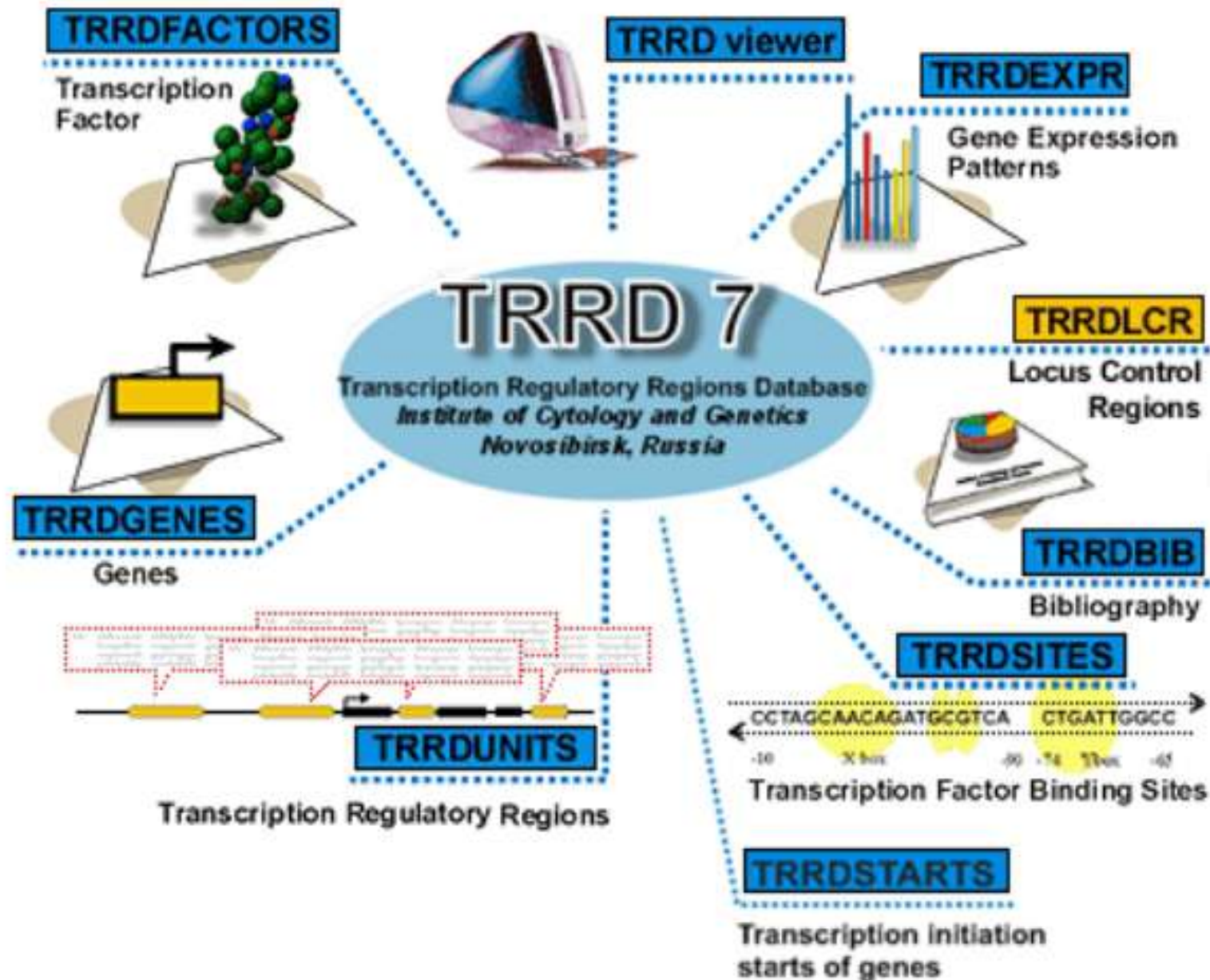


8 regulatory units

23 transcription factor binding sites

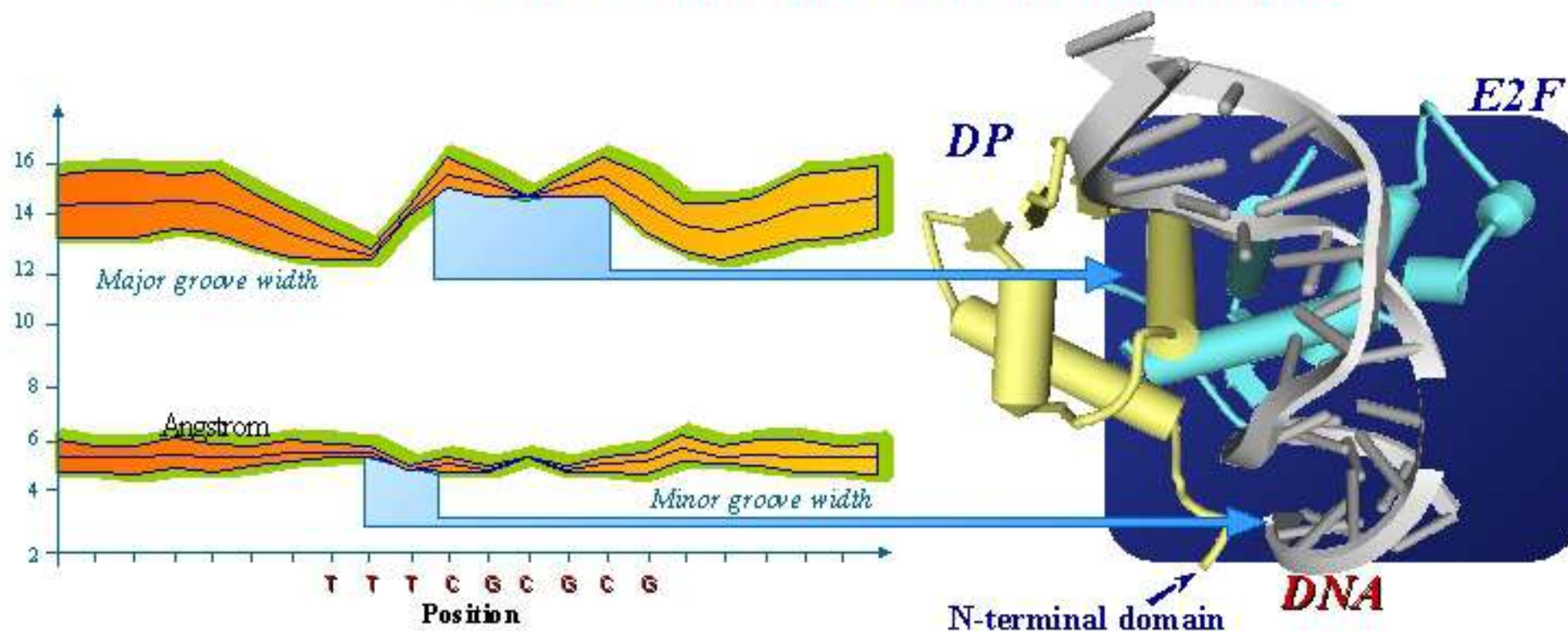
TRRD is an informational resource comprising a family of databases

<http://www.bionet.nsc.ru/trrd/>





SITECON: a method for study of the conserved conformational and physicochemical properties in short regions of DNA sites
ANALYSIS OF E2F/DP BINDING SITES



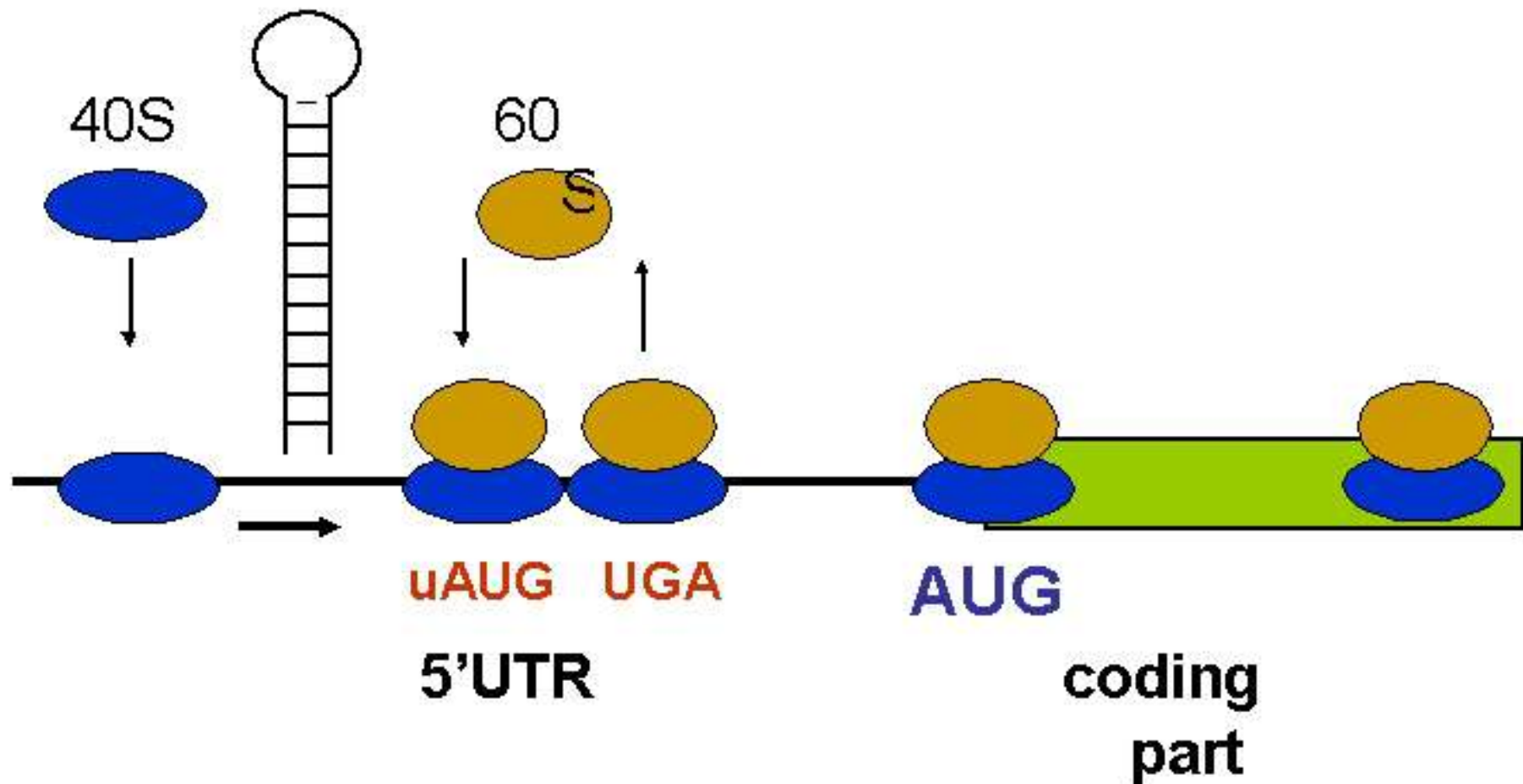
Conformational similarity value

$$P_{\Sigma} = \frac{\sum_{i=0, l=0}^{I, L} \delta_{il} \frac{1}{\sqrt{2\pi\sigma_{F_{il}}}} \exp(-((\bar{F}_{il} - F_{il})/\sigma_{F_{il}})^2)}{\sum_{i=0, l=0}^{I, L} \delta_{il}}$$

$$\sum_{i=0, l=0}^{I, L} \delta_{il}$$

Negative 5'UTR features decreasing translation rate:

- stable hairpins slowing the 40S movement
- upstream AUG triplets (false start codons)



Promoters prediction: general view

I. *Search by content*

Dinucleotide or oligonucleotide frequencies distribution, hidden markov models, discriminant analysis, neural networks

II. *Search by signals*

Potential transcription factor binding sites (TFBS) prediction by position weight matrix (PWM)

Identification of statistically overrepresented 'motifs', each presumably corresponding to particular TFBS.

III. Complex – combination of I & II methods

IV. Additional DNA or structural features:

CpG islands, DNA conformational, physical or chemical properties. nucleosome positioning, etc.

This classification is just an approximation, for example in some cases methods I and II are very similar

Better performance may be achieved by integration with results of other programs for prediction of different gene structural components (coding

Suffix trees approach

Sequences:

aaba

abaa

aabb

Required support $K=2$

Without mismatches

