Modeling and predicting the structure of transmembrane proteins

Jérôme Waldispühl [123], Jean-Marc Steyaert [2]

Jerome.Waldispuhl@polytechnique.edu

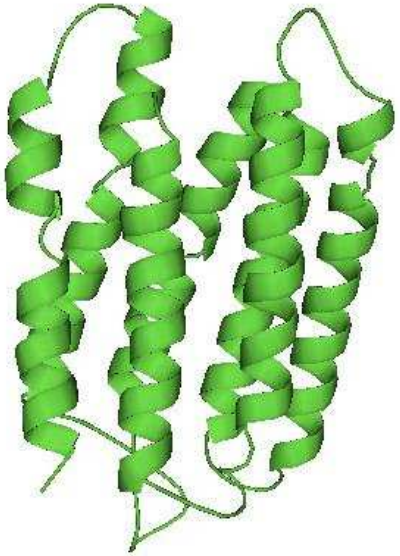
[1] Department of Biology, Boston College, USA

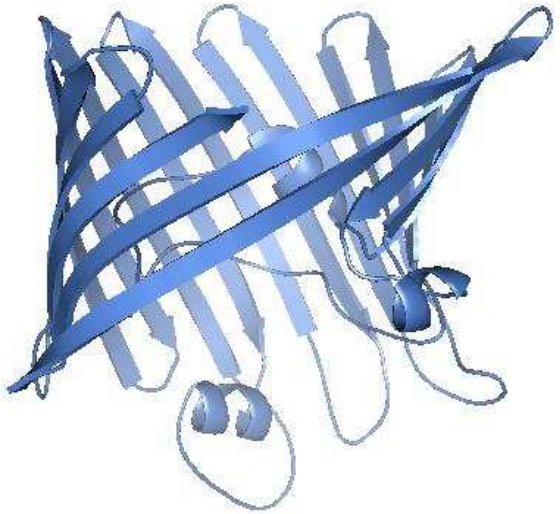
[2] LIX, École polytechnique, France

[3] LIAFA, Université de Paris 7, France

Objectives

Modeling and predicting the structure of transmembrane proteins





Outlines

- Summary of language theory (multi-tape S-attribute grammars),
- some notions of biology,
- an approximate physical model,
- grammatical modeling,
- performance evaluation,
- conclusion.

S-attribute grammars

Definition *Context-free grammars*

$$G = \{V_T, V_N, P, S\}$$

- V_T *is the set of terminals,*
- V_N *is the set of non-terminals,*
- P *is the set of productions rules* $(A \rightarrow \alpha)$*,*
- S *is the axiom,*

S-attribute grammars

Definition *S-attribute grammars*

$$G = \{V_T, V_N, P, S, \mathcal{A}, \lambda_{\mathcal{A}}, F_P\}$$

- V_T *is the set of terminals,*
- V_N *is the set of non-terminals,*
- P *is the set of productions rules,*
- S *is the axiom,*
- $\mathcal{A}$ *is the set of attributes,*
- $\lambda_{\mathcal{A}}$ *is the set of evaluation functions for the terminals,*
- F_P *is the set of the fonctions used to compute the non-terminals attributes.*

Example (1) : arithmetical expressions

$$\left.\begin{array}{l} V_T = \{0, \cdots, 9, +, \times\}, \\ V_N = S, \\ P = \left\{ \begin{array}{l} S \to S + S \\ S \to S \times S \\ S \to 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \end{array} \right. \end{array}\right) \textit{Context-free grammar}$$

Example (1) : arithmetical expressions

$$
\left.
\begin{array}{l}
V_T = \{0, \cdots, 9, +, \times\}, \\
V_N = S, \\
P = \left\{ \begin{array}{l} S \to S + S \\ S \to S \times S \\ S \to 0 \mid 1 \mid 2 \mid 3 \mid 4 \mid 5 \mid 6 \mid 7 \mid 8 \mid 9 \end{array} \right.
\end{array}
\right\}
$$

Context-free grammar

$$
\left.
\begin{array}{l}
\mathcal{A} = \mathbb{N} \\
\lambda_{\mathcal{A}} = \left\{ \begin{array}{l} S_{\mathcal{A}}(0) = 0 \\ \vdots \\ S_{\mathcal{A}}(9) = 9 \\ S_{\mathcal{A}}(+, \times) = 0 \end{array} \right. \\
F_P = \left\{ \begin{array}{l} f_{S \to S+S}(xyz) = x + z \\ f_{S \to S \times S}(xyz) = x \times z \\ f_{S \to a \in V_T}(x) = x \end{array} \right.
\end{array}
\right\}
$$

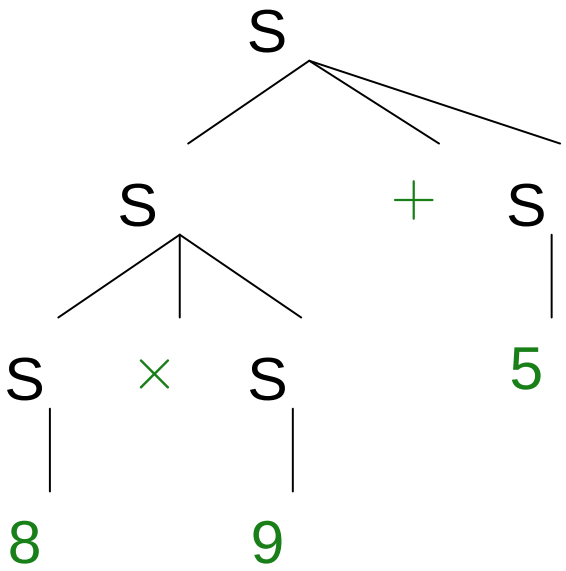
Attribute system

Exemple (1) : Arithmetical expressions

$$8 \times 9 + 5$$

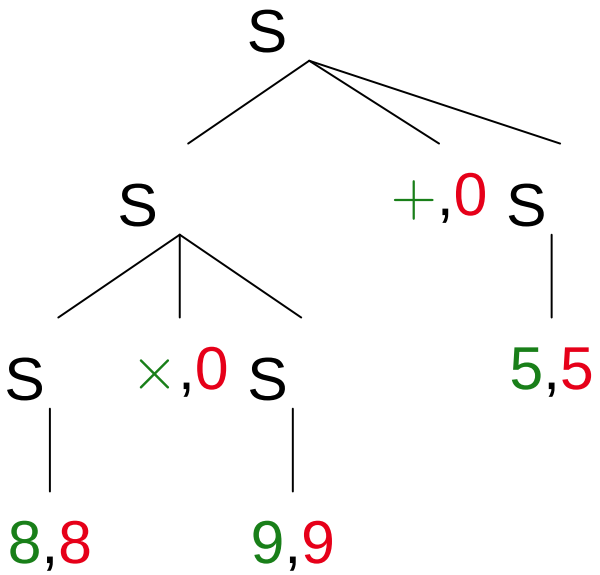
Exemple (1) : Arithmetical expressions

$8 \times 9 + 5$



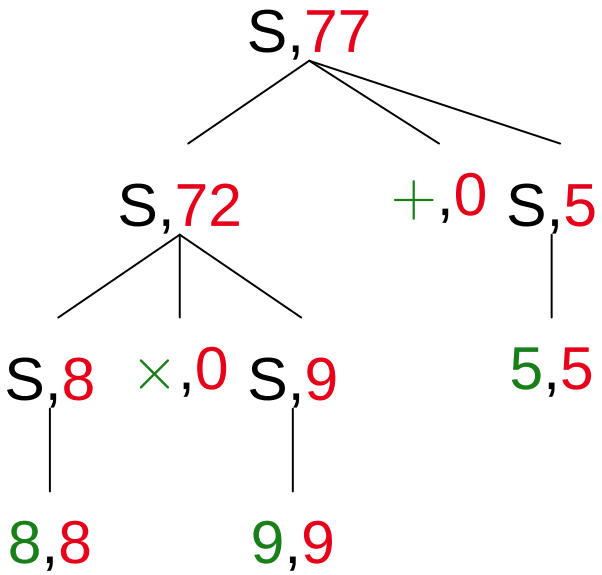
Exemple (1) : Arithmetical expressions

$8 \times 9 + 5$



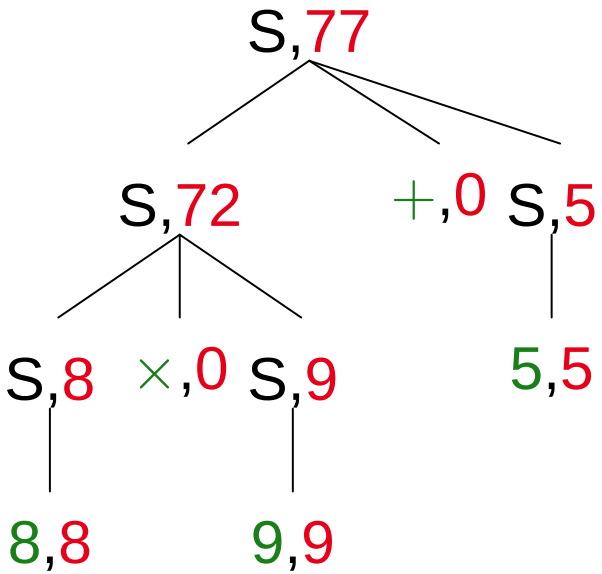
Exemple (1) : Arithmetical expressions

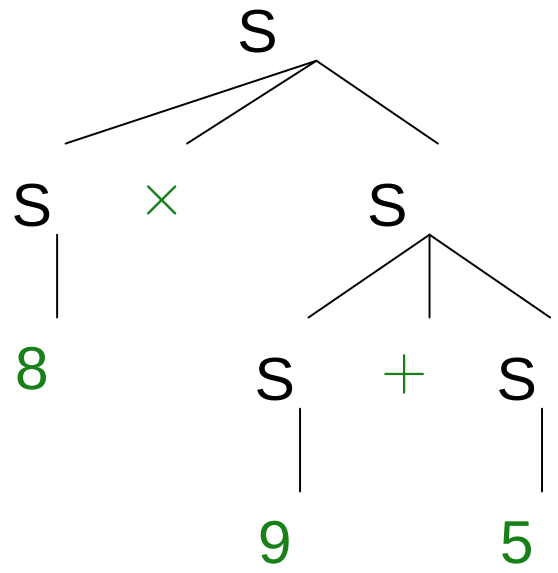
$8 \times 9 + 5$



Exemple (1) : Arithmetical expressions

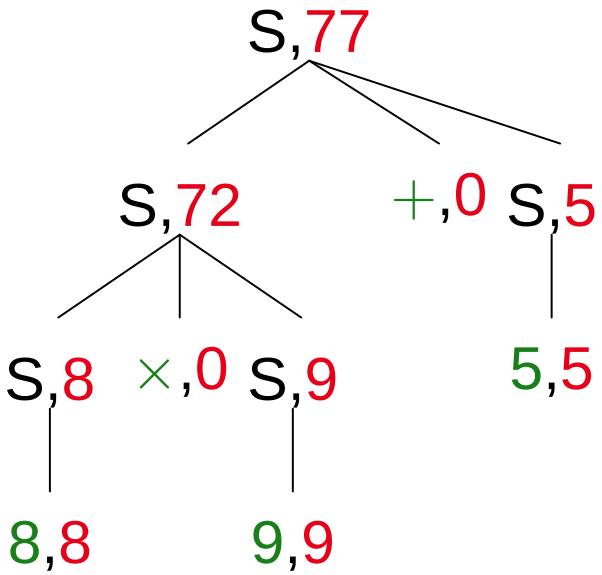
$8 \times 9 + 5$

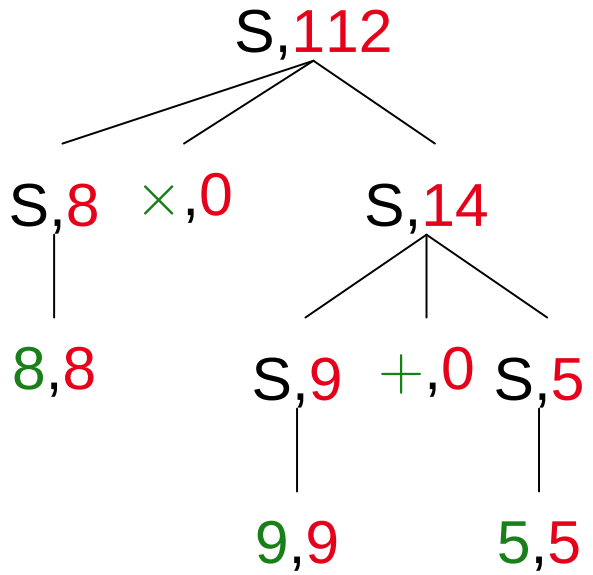




Exemple (1) : Arithmetical expressions

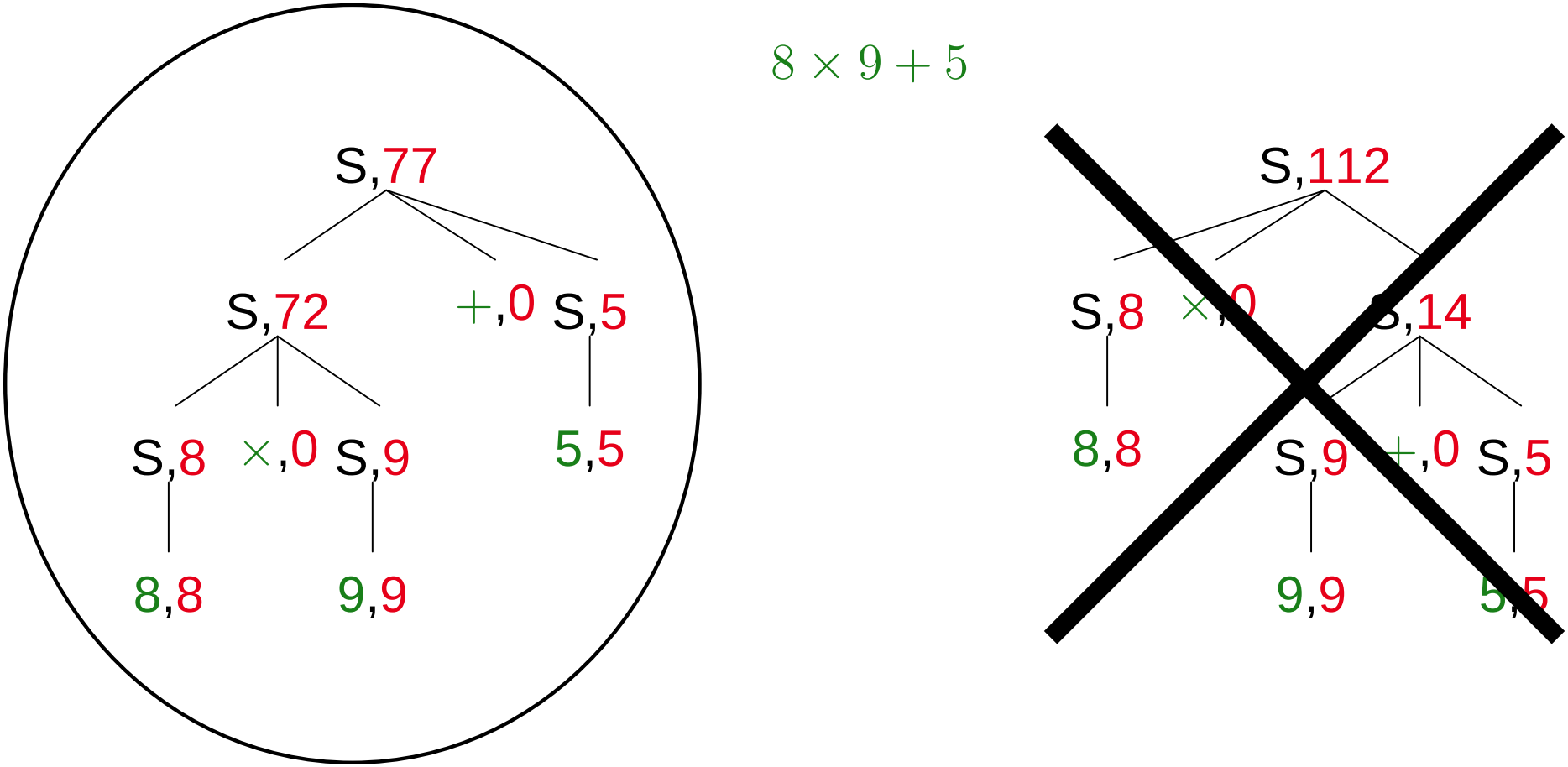
$8 \times 9 + 5$





Exemple (1) : Arithmetical expressions

$8 \times 9 + 5$



...Only the derivation tree with the optimal attribute is conserved.

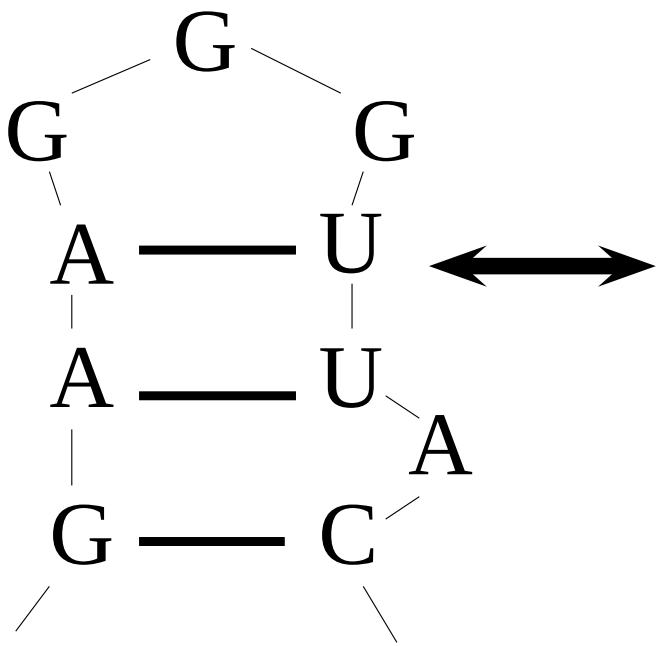
Optimization constraint

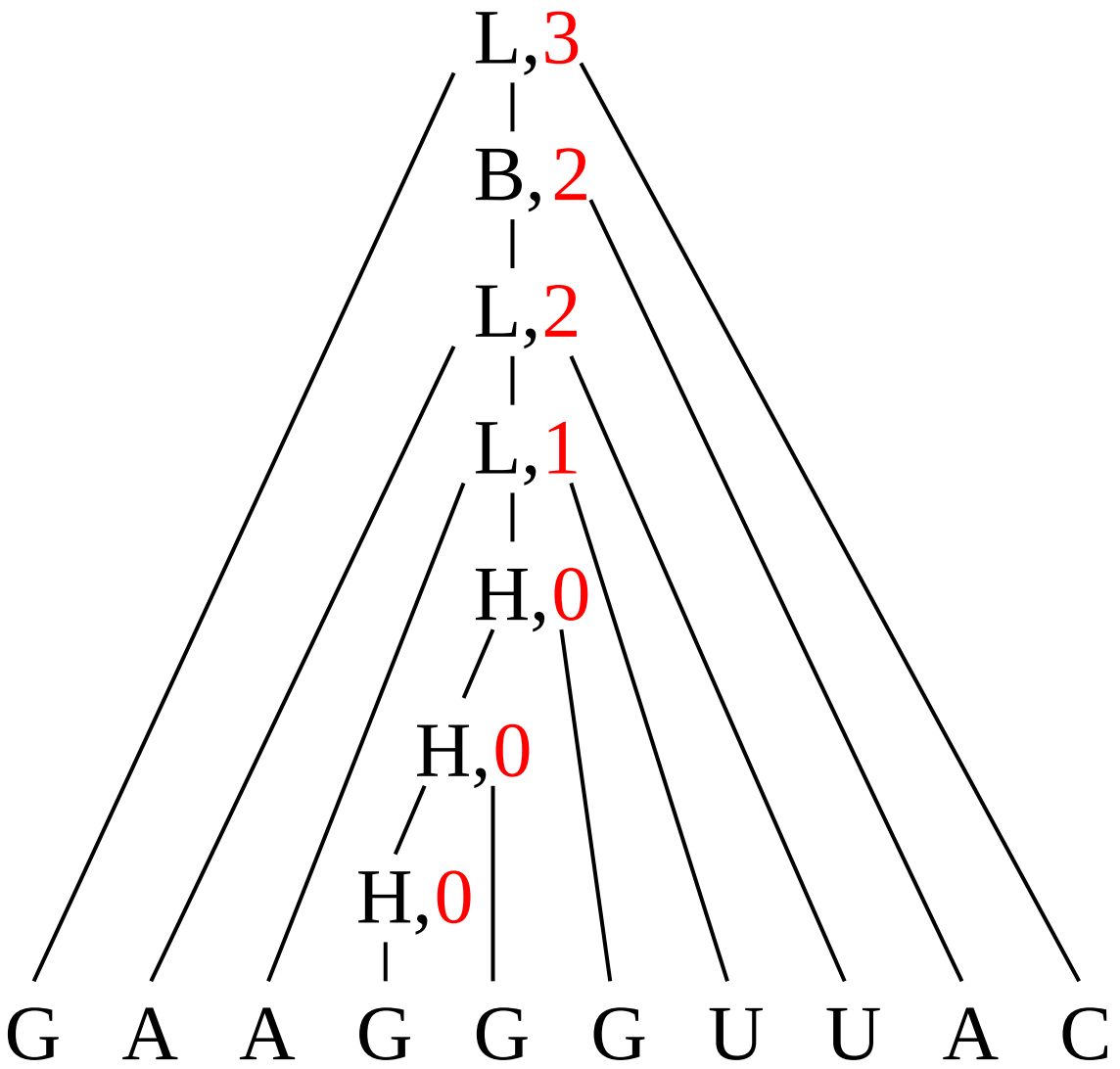
Definition *Optimization constraint*

$$\mathcal{C}(x, \lambda_x, y, \lambda_y)$$

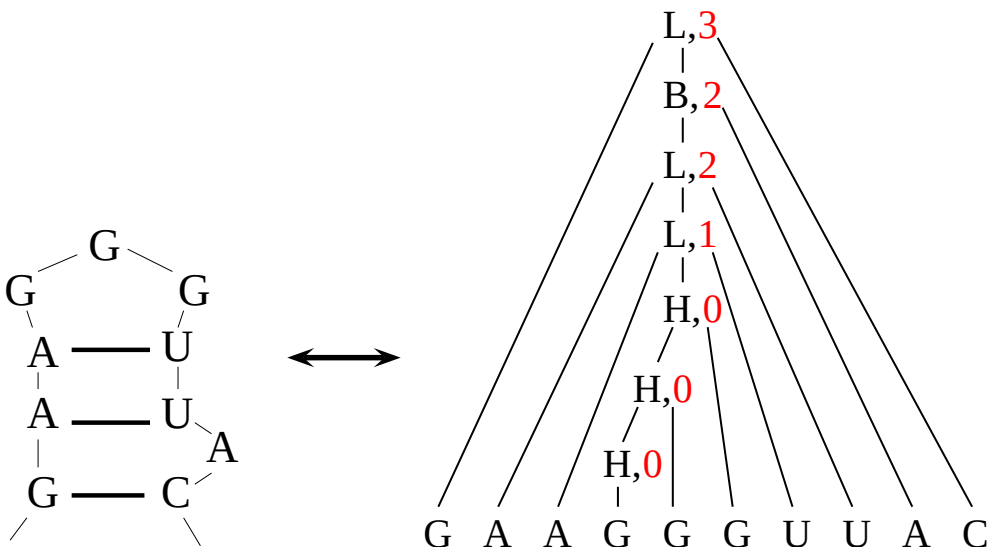
- x *and* y $\in V_T \cup V_N$,
- λ_x *and* $\lambda_x \in \mathcal{A}$,
- *return the pair* (z, λ_z) *such that* λ_z *is optimal.*

Example (2) : RNA secondary structure





Example (2) : RNA secondary structure



- secondary structure = derivation tree
- folding energy = attributes
- secondary structure with the minimum free energy (Zuker) = derivation tree with the optimal attribute

S-attribute grammars

How to find the optimal derivation tree ?

principle : dynamic programming

algorithm : Cocke-Kazamy-Younger, Earley, GCP...

implementation : *mtsag2c* (F. Lefebvre, 1997)

Multi-tape S-attribute grammar

Definition *Multi-tape alphabet*

$$\Sigma = \bigotimes_{i=1\cdots m} (\Sigma^{(i)} \cup \{\varepsilon\})$$

Definition *Multi-tape Context-free grammar*

$$G = \{V_T, V_N, P, S\} \text{ where } V_T \text{ is an } m\text{-tape alphabet.}$$

Definition *Multi-tape S-attribute grammar*

$$G = \{V_T, V_N, P, S, \mathcal{A}, \lambda_{\mathcal{A}}, F_P\} \text{ where } V_T \text{ is an } m\text{-tape alphabet.}$$

Example : RNA sequence alignment

$$
\begin{cases}
S \rightarrow SS \mid mat \mid del \mid ins \mid mut \\
mat \rightarrow \begin{bmatrix} a \\ a \end{bmatrix} \mid \begin{bmatrix} u \\ u \end{bmatrix} \mid \begin{bmatrix} g \\ g \end{bmatrix} \mid \begin{bmatrix} c \\ c \end{bmatrix} \\
del \rightarrow \begin{bmatrix} - \\ a \end{bmatrix} \mid \begin{bmatrix} - \\ u \end{bmatrix} \mid \begin{bmatrix} - \\ g \end{bmatrix} \mid \begin{bmatrix} - \\ c \end{bmatrix} \\
ins \rightarrow \begin{bmatrix} a \\ - \end{bmatrix} \mid \begin{bmatrix} u \\ - \end{bmatrix} \mid \begin{bmatrix} g \\ - \end{bmatrix} \mid \begin{bmatrix} c \\ - \end{bmatrix} \\
mut \rightarrow \begin{bmatrix} a \\ u \end{bmatrix} \mid \begin{bmatrix} a \\ g \end{bmatrix} \mid \begin{bmatrix} a \\ c \end{bmatrix} \mid \begin{bmatrix} u \\ a \end{bmatrix} \\
\qquad\quad \mid \begin{bmatrix} u \\ g \end{bmatrix} \mid \begin{bmatrix} u \\ c \end{bmatrix} \mid \begin{bmatrix} g \\ a \end{bmatrix} \mid \begin{bmatrix} g \\ u \end{bmatrix} \\
\qquad\quad \mid \begin{bmatrix} g \\ c \end{bmatrix} \mid \begin{bmatrix} c \\ a \end{bmatrix} \mid \begin{bmatrix} c \\ u \end{bmatrix} \mid \begin{bmatrix} c \\ g \end{bmatrix}
\end{cases}
$$

Example : RNA sequence alignment

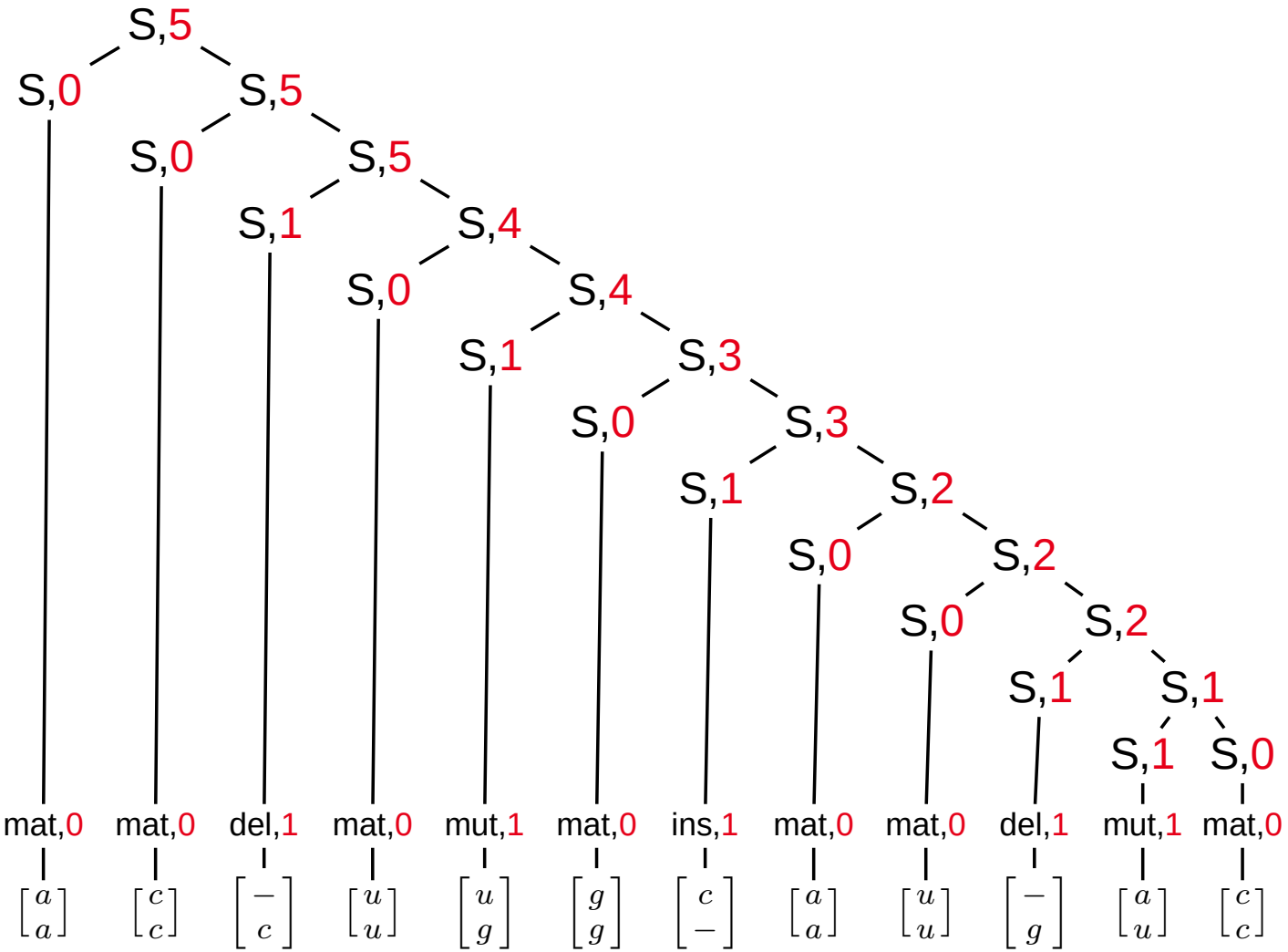
$$
\begin{cases}
S & \to & SS \mid mat \mid del \mid ins \mid mut \\
mat & \to & \begin{bmatrix} a \\ a \end{bmatrix} \mid \begin{bmatrix} u \\ u \end{bmatrix} \mid \begin{bmatrix} g \\ g \end{bmatrix} \mid \begin{bmatrix} c \\ c \end{bmatrix} \\
del & \to & \begin{bmatrix} - \\ a \end{bmatrix} \mid \begin{bmatrix} - \\ u \end{bmatrix} \mid \begin{bmatrix} - \\ g \end{bmatrix} \mid \begin{bmatrix} - \\ c \end{bmatrix} \\
ins & \to & \begin{bmatrix} a \\ - \end{bmatrix} \mid \begin{bmatrix} u \\ - \end{bmatrix} \mid \begin{bmatrix} g \\ - \end{bmatrix} \mid \begin{bmatrix} c \\ - \end{bmatrix} \\
mut & \to & \begin{bmatrix} a \\ u \end{bmatrix} \mid \begin{bmatrix} a \\ g \end{bmatrix} \mid \begin{bmatrix} a \\ c \end{bmatrix} \mid \begin{bmatrix} u \\ a \end{bmatrix} \\
 & & \mid \begin{bmatrix} u \\ g \end{bmatrix} \mid \begin{bmatrix} u \\ c \end{bmatrix} \mid \begin{bmatrix} g \\ a \end{bmatrix} \mid \begin{bmatrix} g \\ u \end{bmatrix} \\
 & & \mid \begin{bmatrix} g \\ c \end{bmatrix} \mid \begin{bmatrix} c \\ a \end{bmatrix} \mid \begin{bmatrix} c \\ u \end{bmatrix} \mid \begin{bmatrix} c \\ g \end{bmatrix}
\end{cases}
$$

$$\mathcal{A} = \mathbb{Z}$$

$$\lambda(\bullet) = 0$$

$$
F_P = \begin{cases}
f_{S \to SS}(xy) & = & x + y \\
f_{S \to del|ins|mut}(x) & = & x \\
f_{mat \to \bullet}(x) & = & 0 \\
f_{del \to \bullet}(x) & = & 1 \\
f_{ins \to \bullet}(x) & = & 1 \\
f_{mut \to \bullet}(x) & = & 1
\end{cases}
$$

Example : RNA sequence alignment

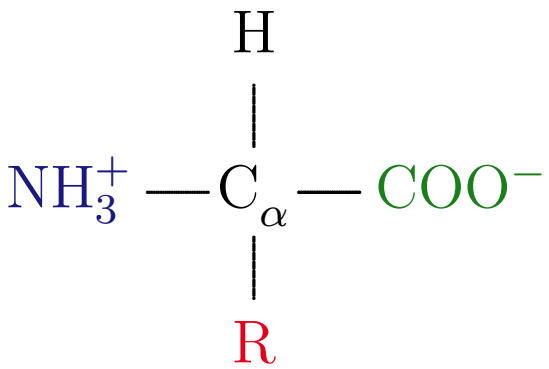


Notions of Biology

- Definition of a protein
- Structure of proteins
- Transmembrane channels

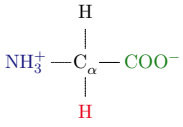
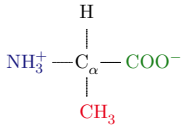
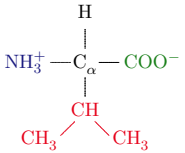
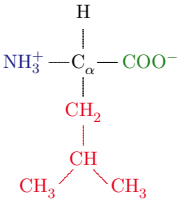
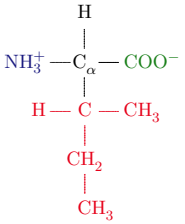
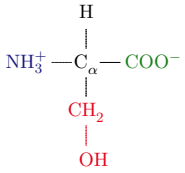
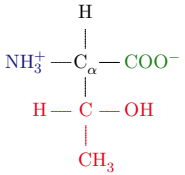
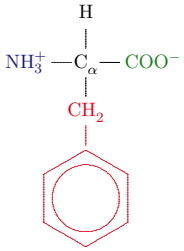
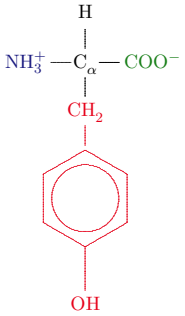
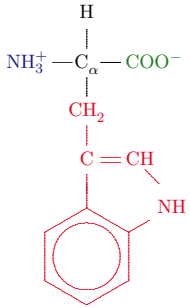
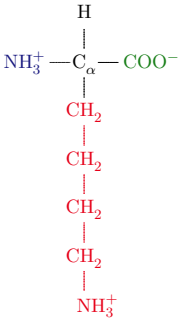
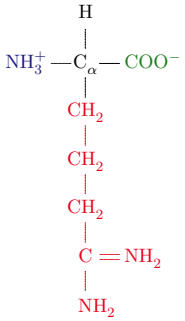
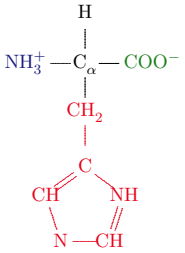
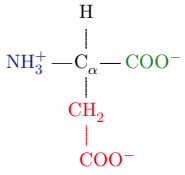
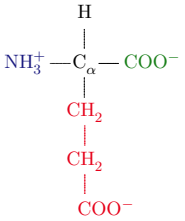
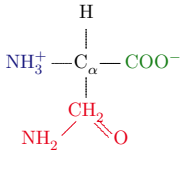
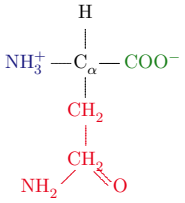
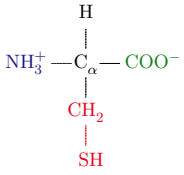
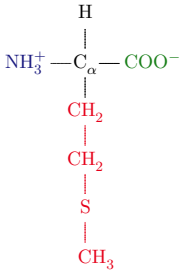
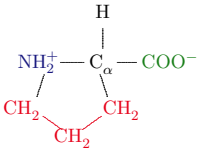
Definition of a protein

Amino acid chemical formula :



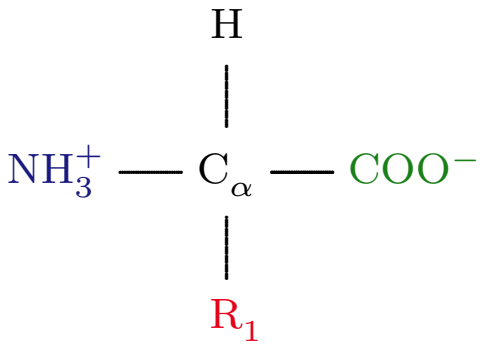
Definition of a protein

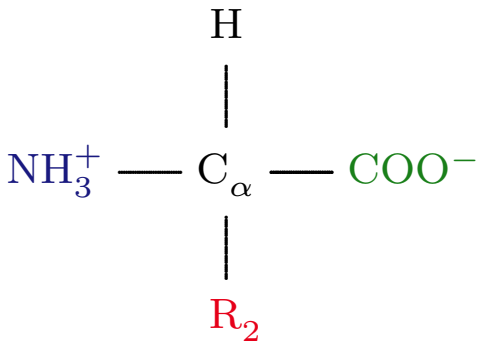
The 20 amino acids :

 *Glycine (Gly/G)*	 *Alanine (Ala/A)*	 *Valine (Val/V)*	 *Leucine (Leu/L)*	 *Isoleucine (Ile/L)*	 *Serine (Ser/S)*	 *Threonine (Thr/T)*	 *Phenylalanine (Phe/F)*	 *Tyrosine (Tyr/Y)*	 *Tryptophane (Trp/W)*
 *Lysine (Lys/K)*	 *Arginine (Arg/R)*	 *Histidine (His/H)*	 *Aspartate (Asp/D)*	 *Glutamate (Glu/E)*	 *Asparagine (Asn/N)*	 *Glutamamine (Gln/Q)*	 *Cystéine (Cys/C)*	 *Méthionine (Met/M)*	 *Proline (Pro/P)*

Definition of a protein

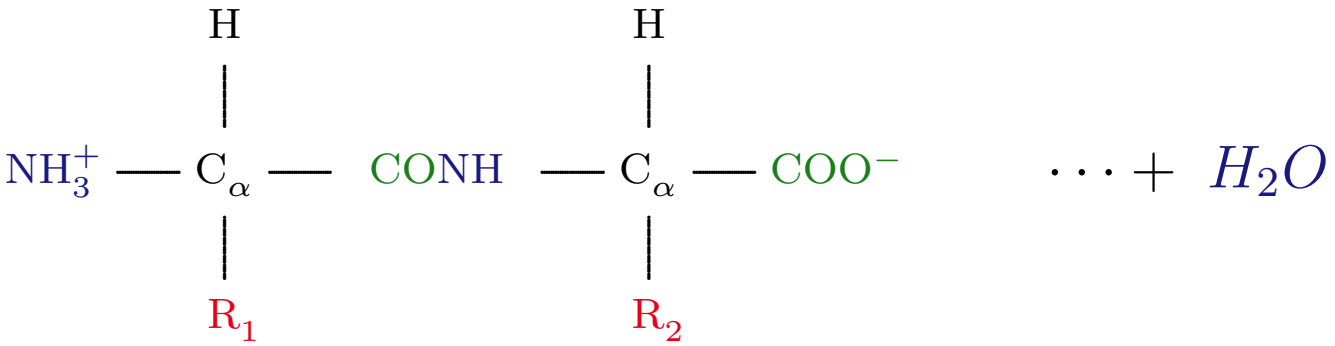
The peptid bond :





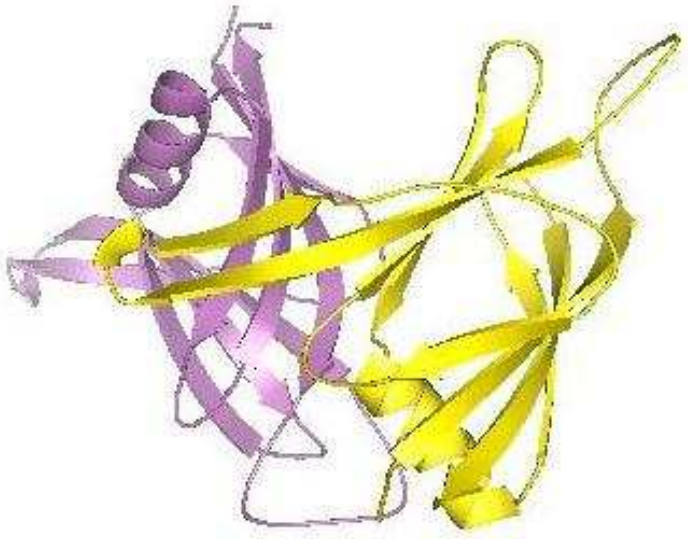
Definition of a protein

The peptid bond :



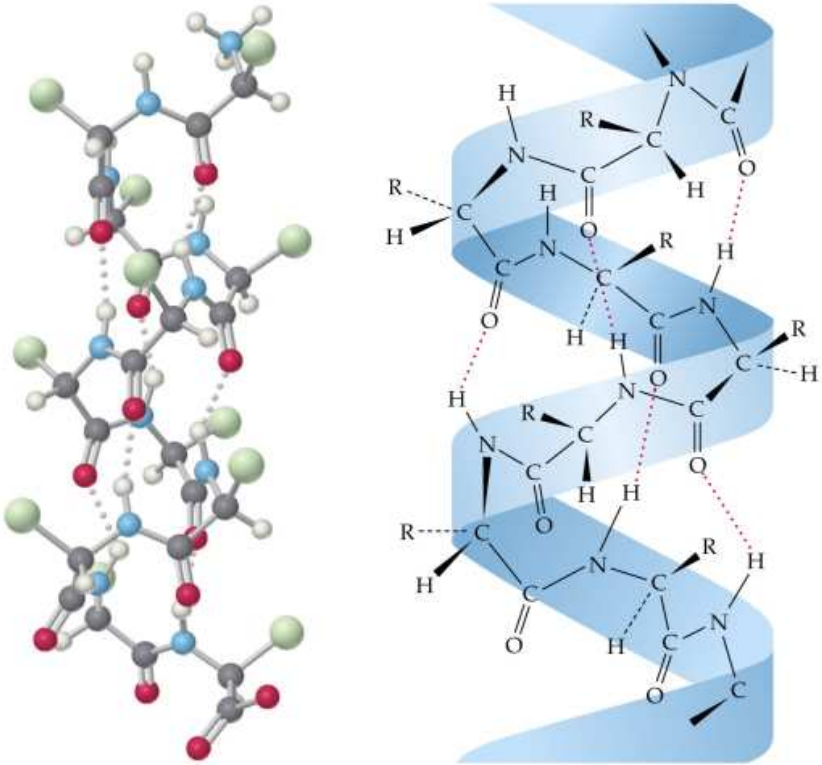
Structure of proteins

Kynase C



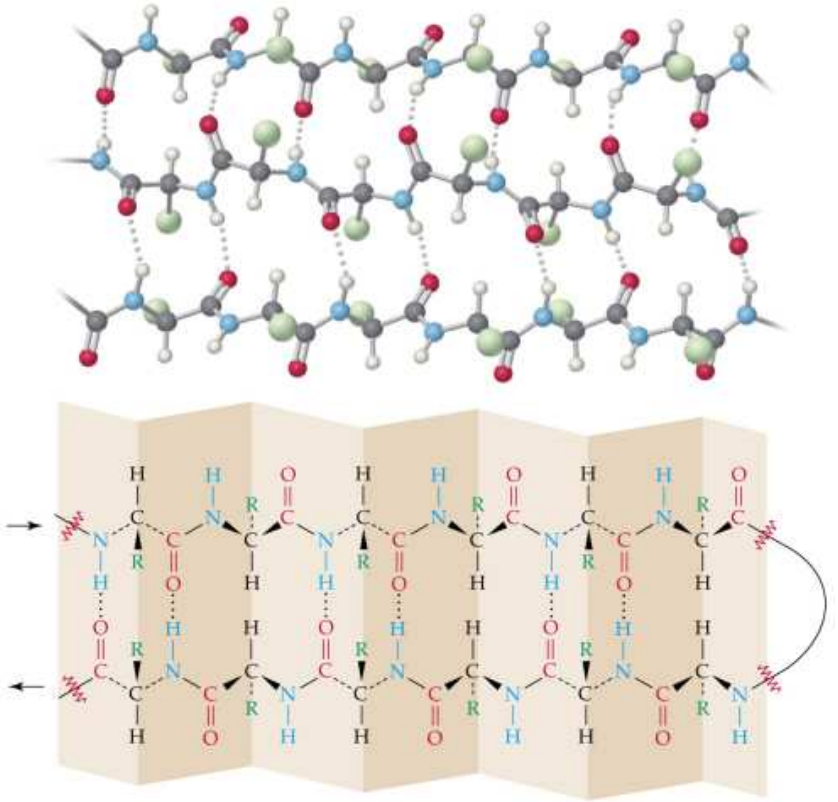
α-helix

- 3.6 amino acids per turn,
- hydrogen bond between residus n and $n+4$.



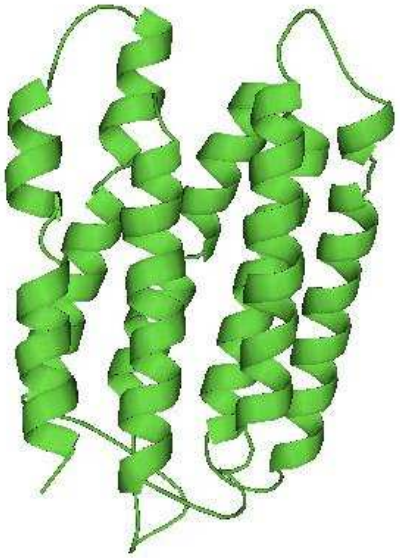
β-sheet

- composed of β-strands
- 2 amino acids per turn,
- hydrogen bond between residues of paired β-strands.

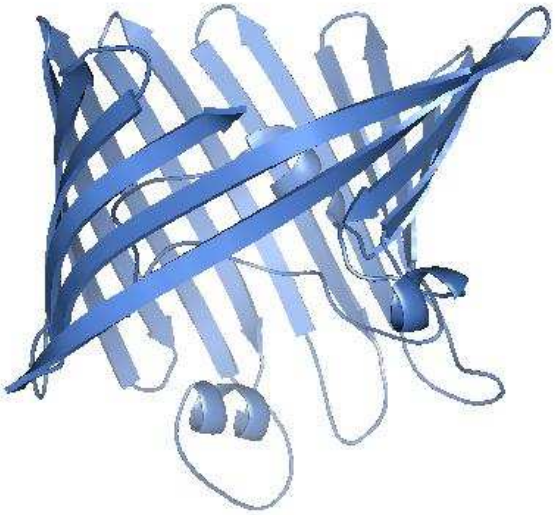


Transmembrane channels

Bacteriorhodopsin

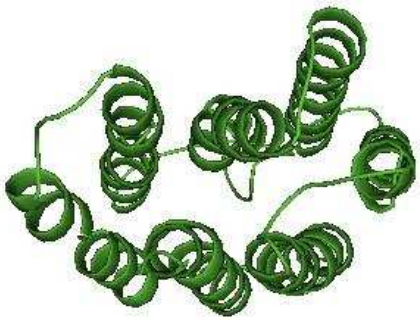


Porin



Transmembrane channels

Bacteriorhodopsin



Porin

Transmembrane channel

Why ?

Transmembrane channel

Why ?

- Simple topologies (only parallel or anti-parallel pairings),
- strong contraints from the environment,
- Some parameters are (much) more important than the others (hydrophobicity)

Transmembrane channel

Why ?

- Simple topologies (only parallel or anti-parallel pairings),
- strong contraints from the environment,
- Some parameters are (much) more important than the others (hydrophobicity)

Interest ?

Transmembrane channel

Why ?

- Simple topologies (only parallel or anti-parallel pairings),
- strong contraints from the environment,
- Some parameters are (much) more important than the others (hydrophobicity)

Interest ?

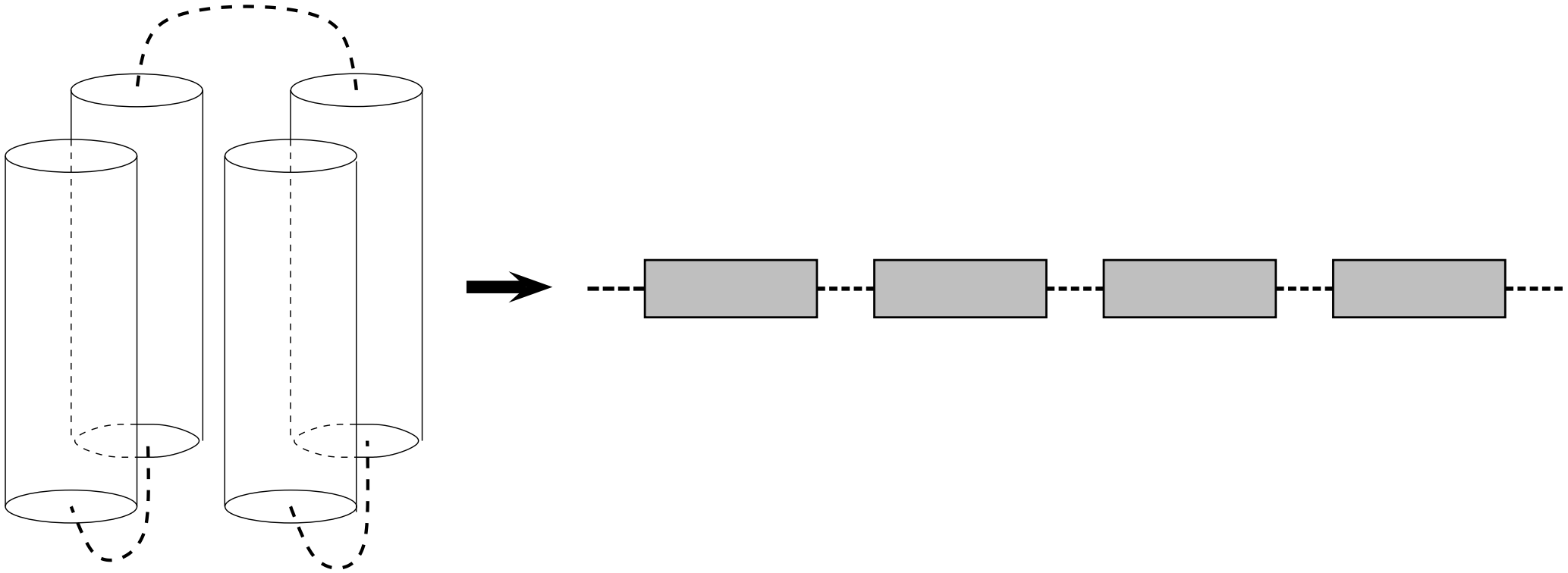
- nearly 40% of the proteome,
- functional importance (allows communication between inner and outer milieu of cell),
- difficult to be observe experimentaly.

Approximate physical model for transmembrane channels

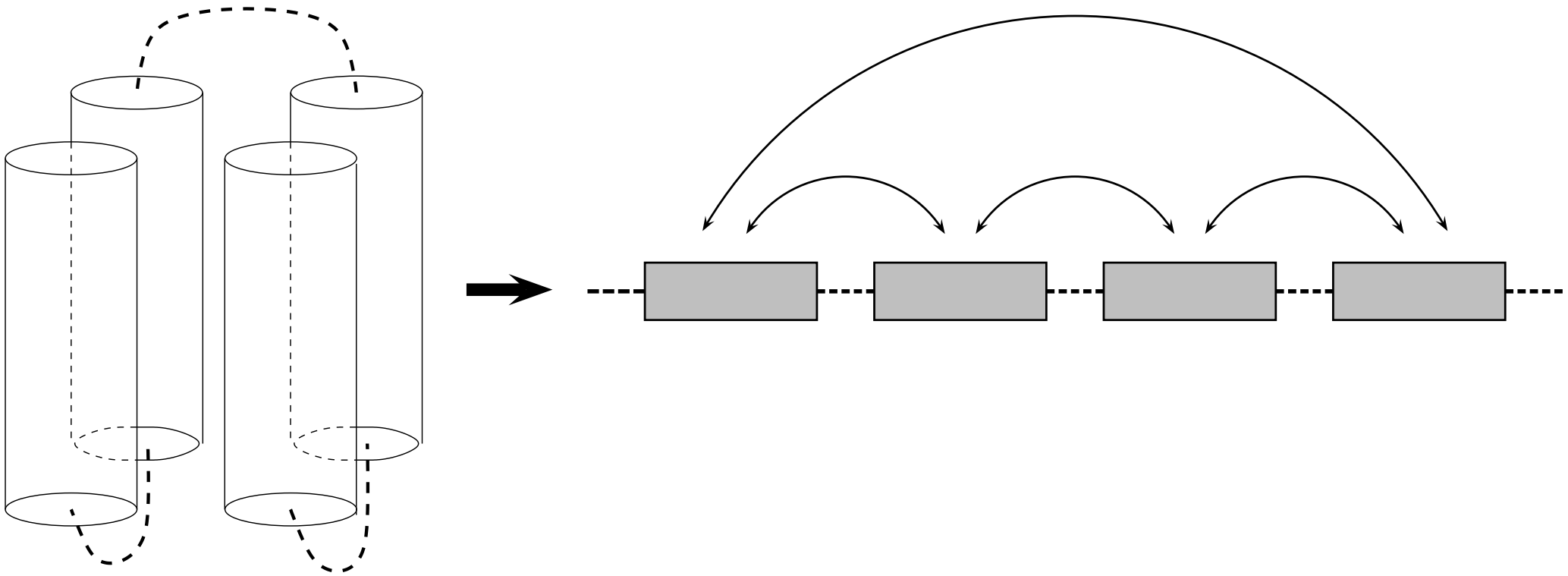
Approximate physical model for α-transmembrane channels

- Modeling the overall structure of α-channel,
- modeling anti-parallel pairing of α-helices,
- modeling the local structure of α-helices,
- pseudo folding energy of α-channels.

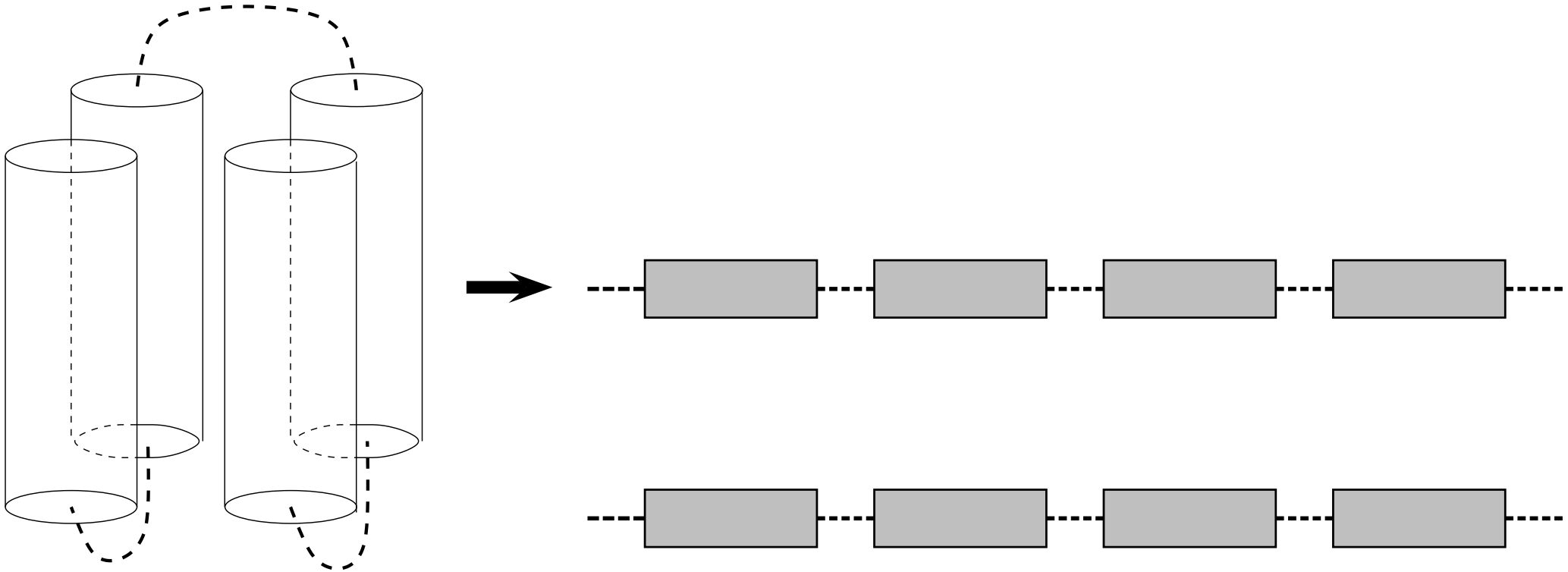
Modeling the overall structure of α-channel



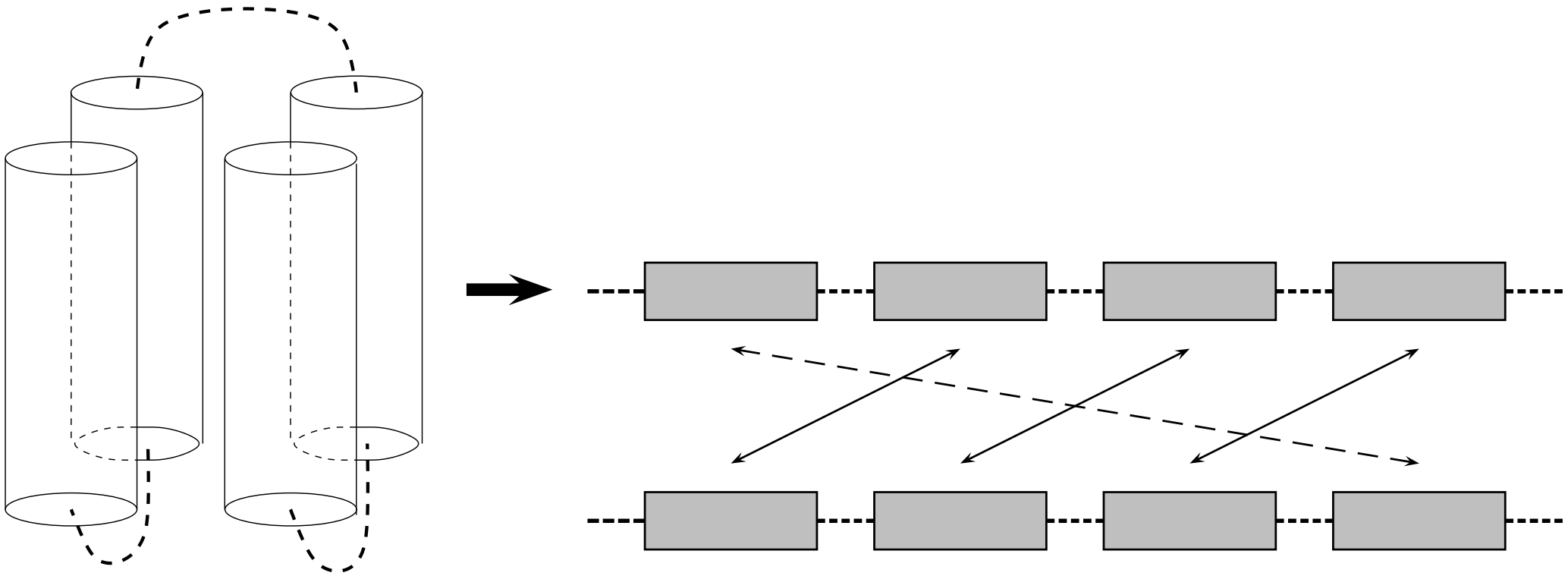
Modeling the overall structure of α-channel



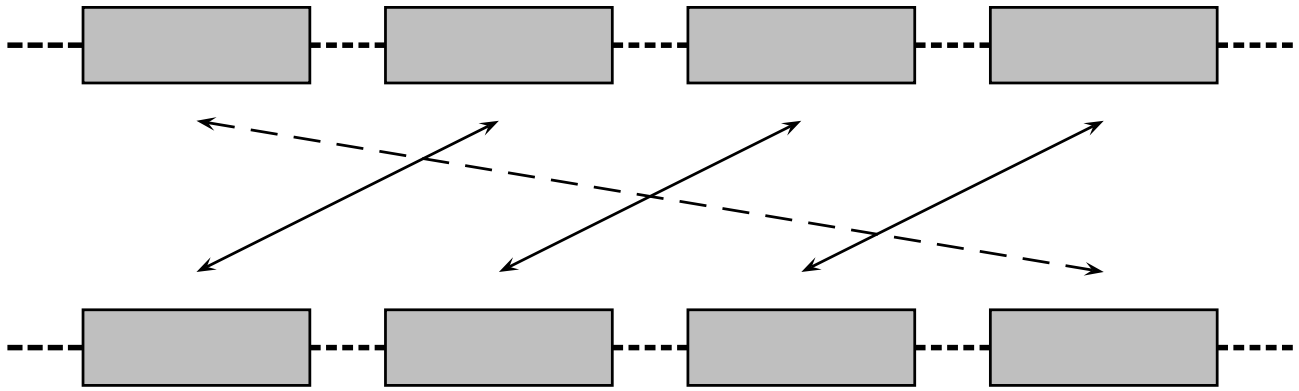
Modeling the overall structure of α-channel



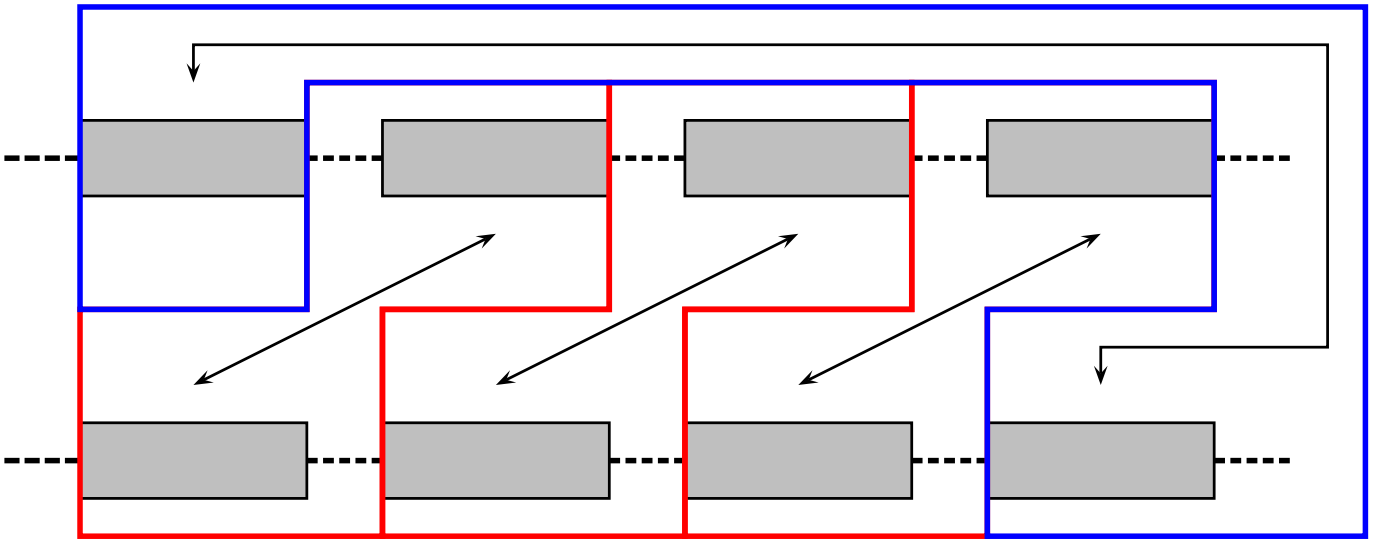
Modeling the overall structure of α-channel



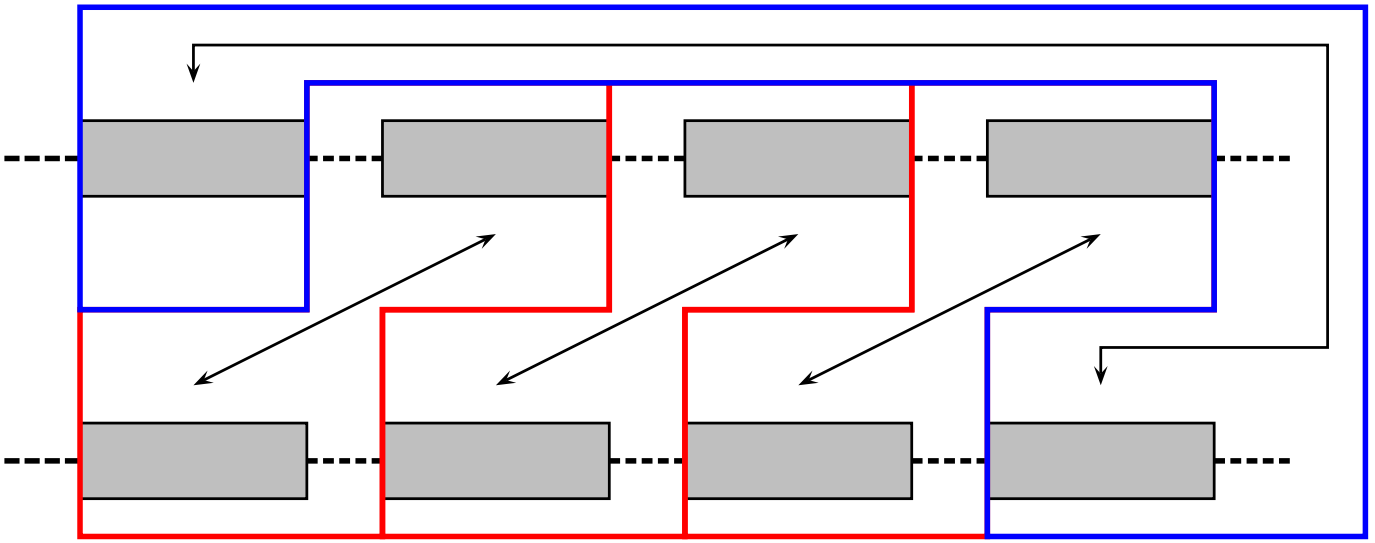
Modeling the overall structure of α-channel



Modeling the overall structure of α-channel

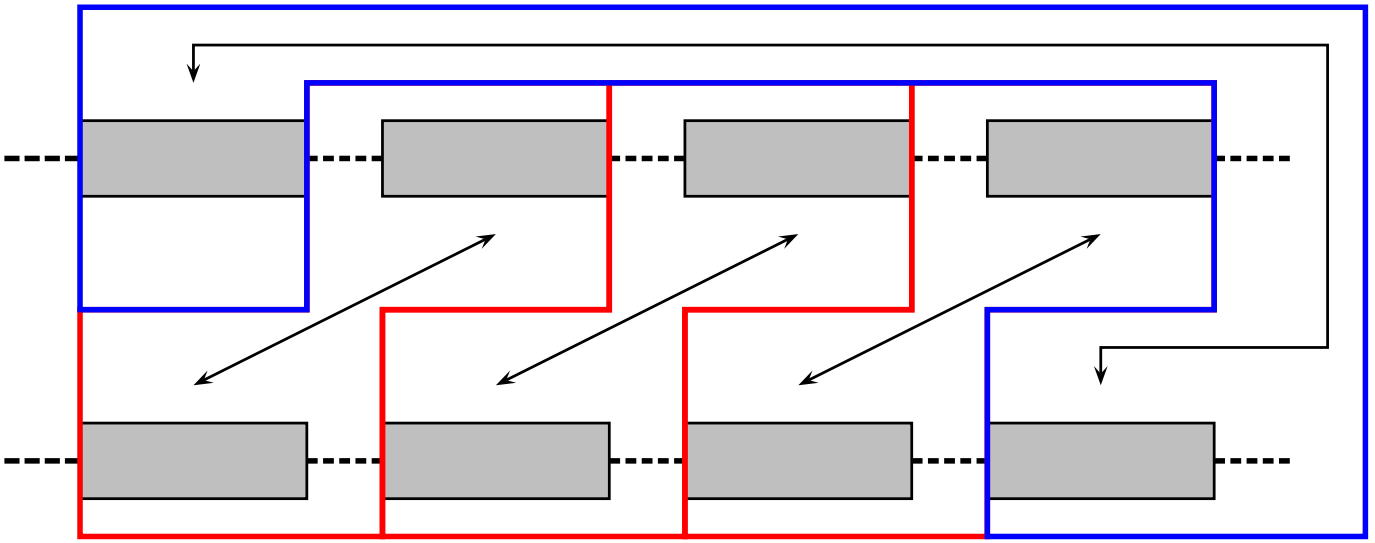


Modeling the overall structure of α-channel



Description of α-channels with $only$ simple anti-parallel pairings.

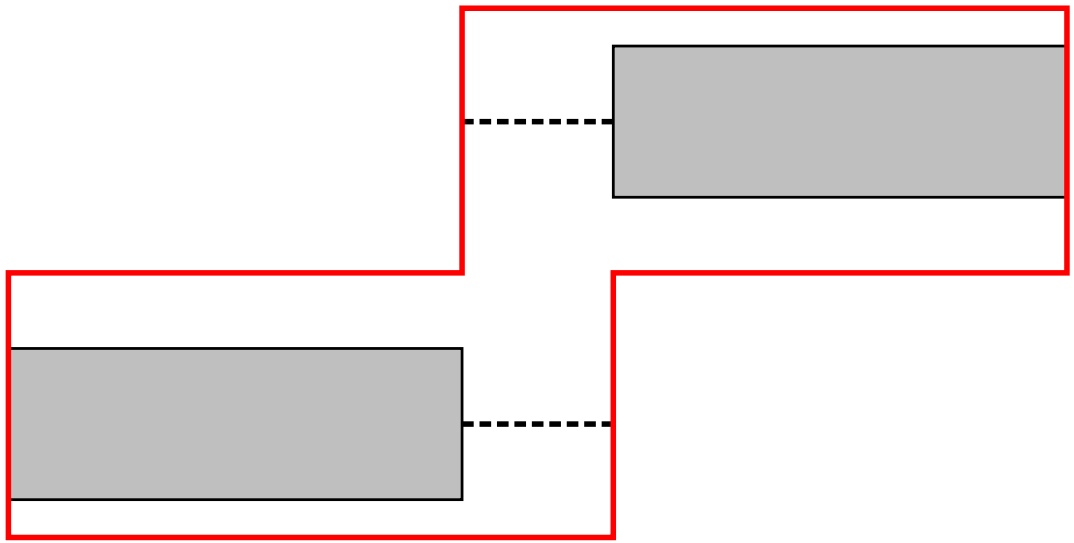
Modeling the overall structure of α-channel



An α-channel is a concatenation of simple anti-parallel pairings.

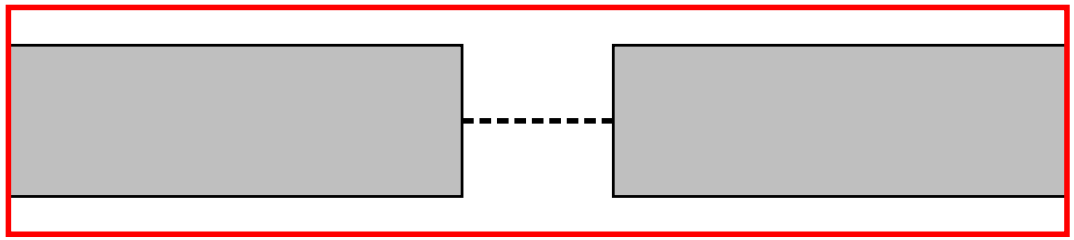
Modeling anti-parallel pairing of α-helices

Let's go back to a linear description :

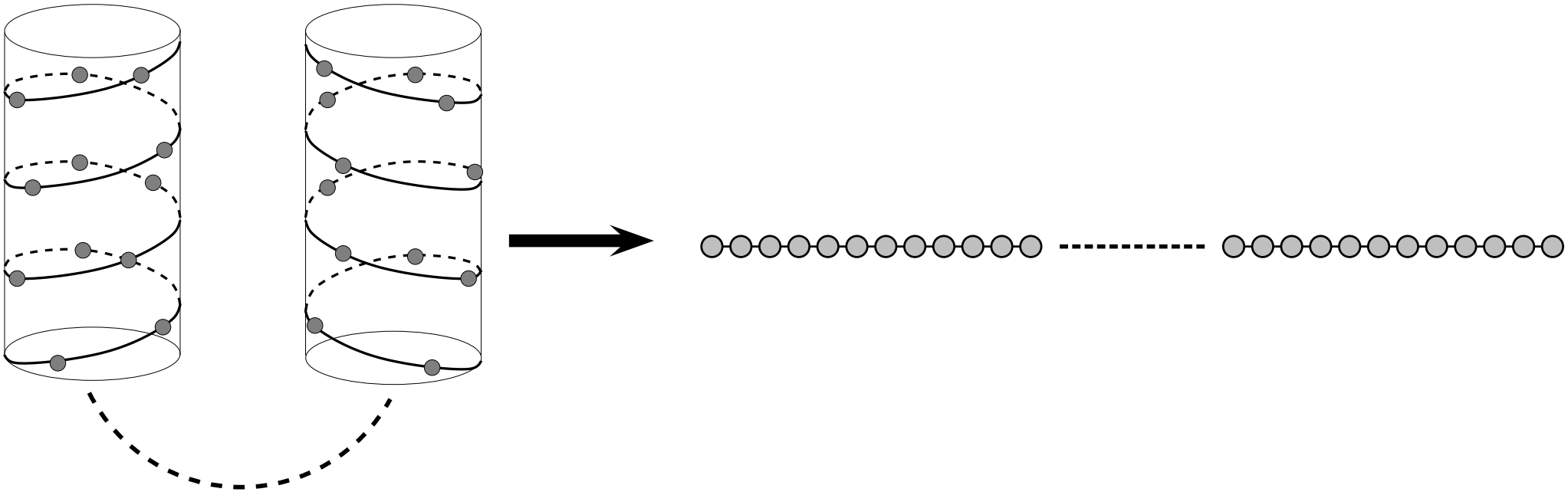


Modeling anti-parallel pairing of α-helices

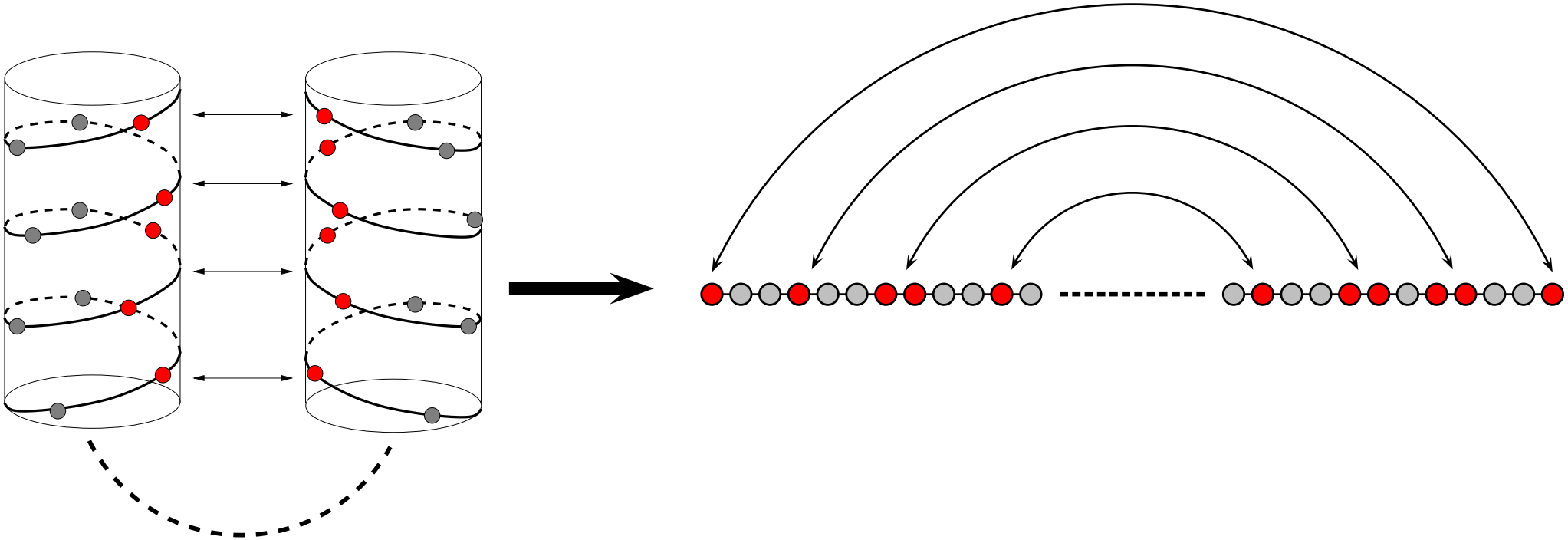
Let's go back to a linear description :



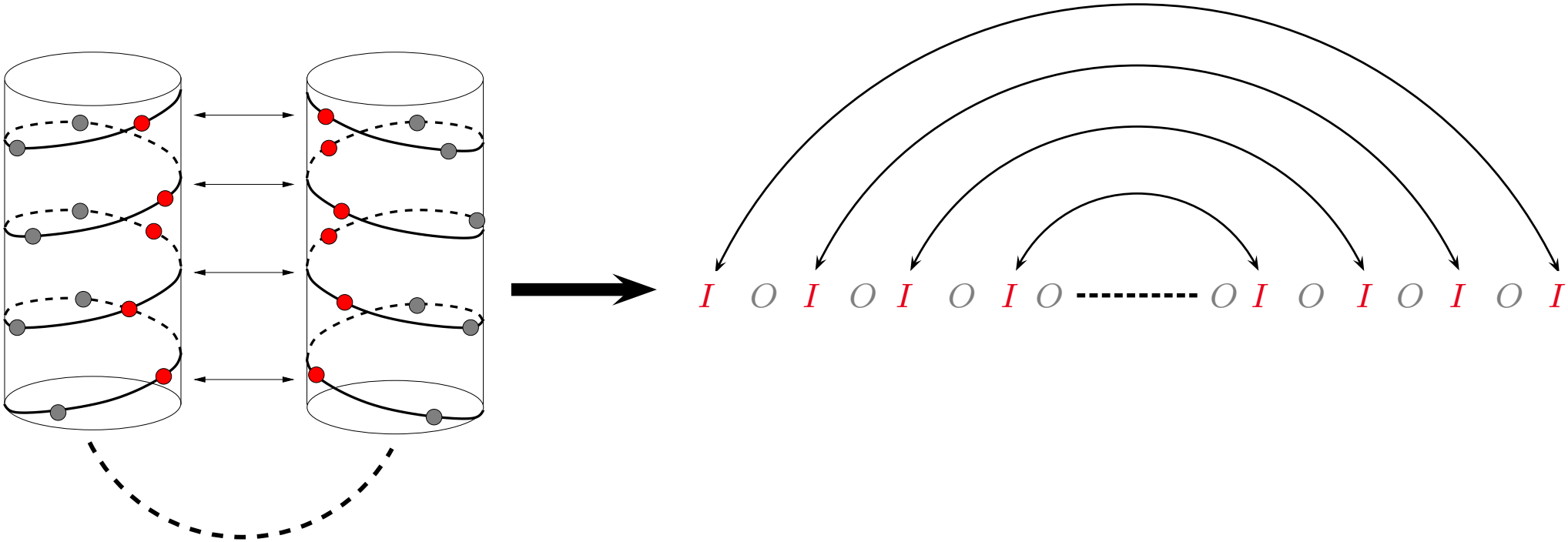
modeling anti-parallel pairing of α-helices



modeling anti-parallel pairing of α-helices

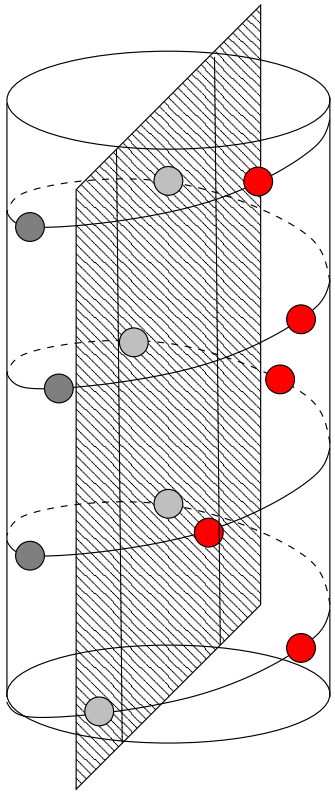


modeling anti-parallel pairing of α-helices



Modeling the local structure of α-helices

O

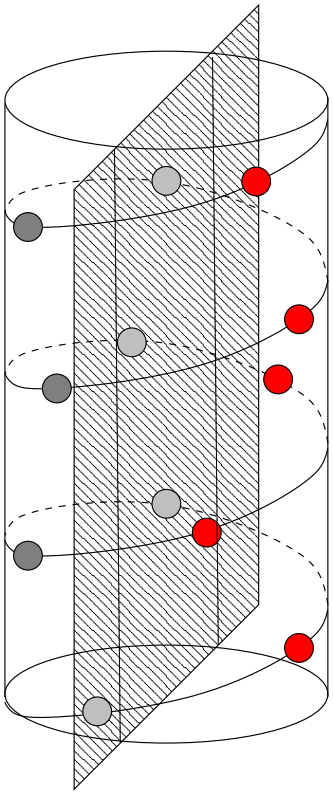


I

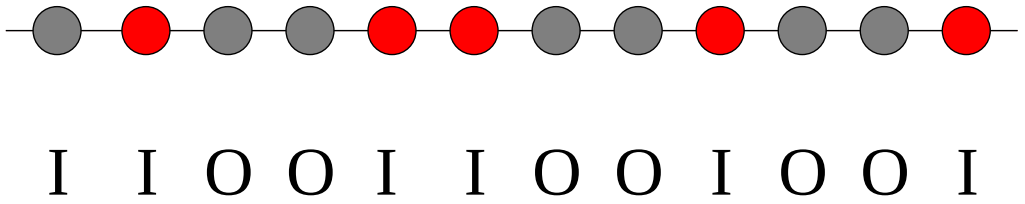
- A $helix$ is a stacking of helix turns,
- A $helix\ turn$ is a sequence of residues corresponding to a complete turn around the helix axis,
- A $helical\ face$ is a subsequence of consecutive amino acids of a helix turn,
  - face I is involved in pairing,
  - face O is opposed to pairing.

Modeling the local structure of α-helices

O



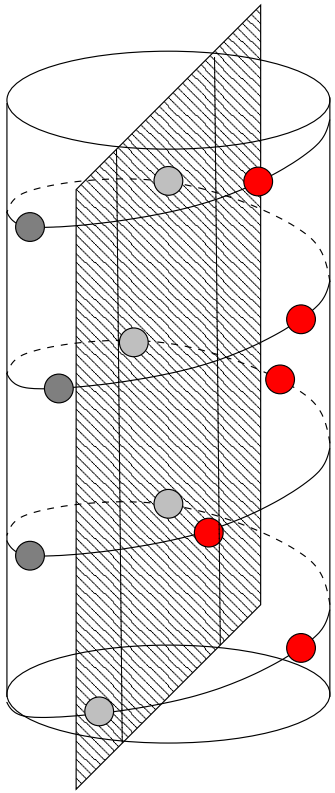
I



I I O O I I O O I O O I

Modeling the local structure of α-helices

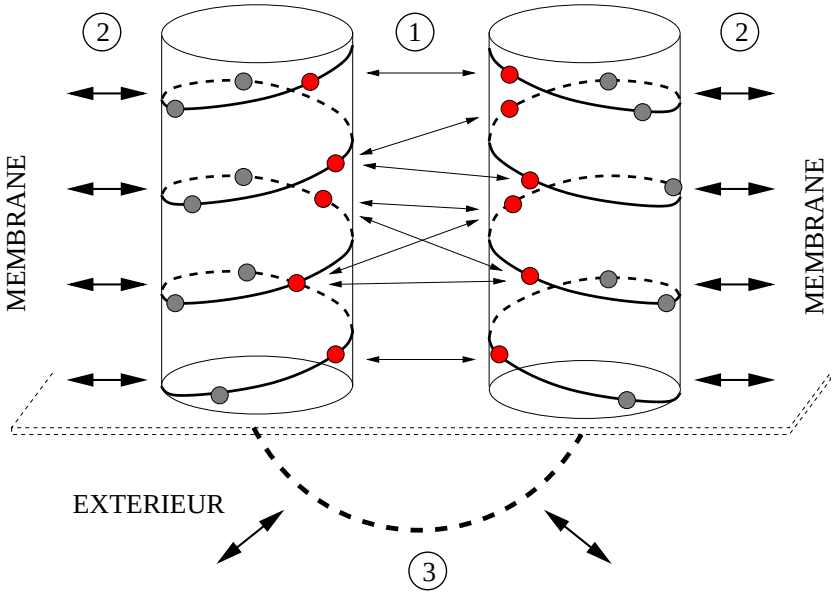
O



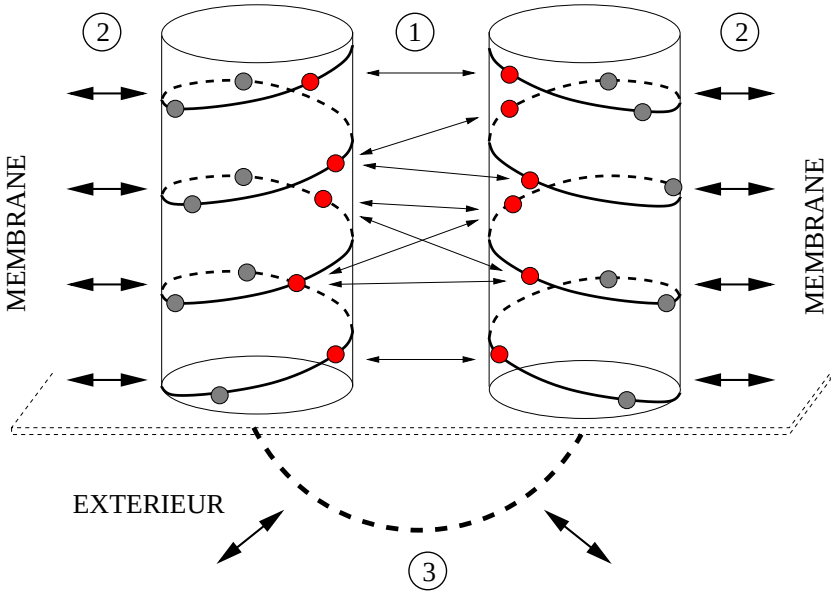
I

- A helix is an alternate sequence of residues belonging to a I or O face,
- A helix turn is composed by 3 or 4 consecutive amino acids,
- A helical face has 1 or 2 amino acids,
- On average, 3.6 residus per turn.

Pseudo folding energy



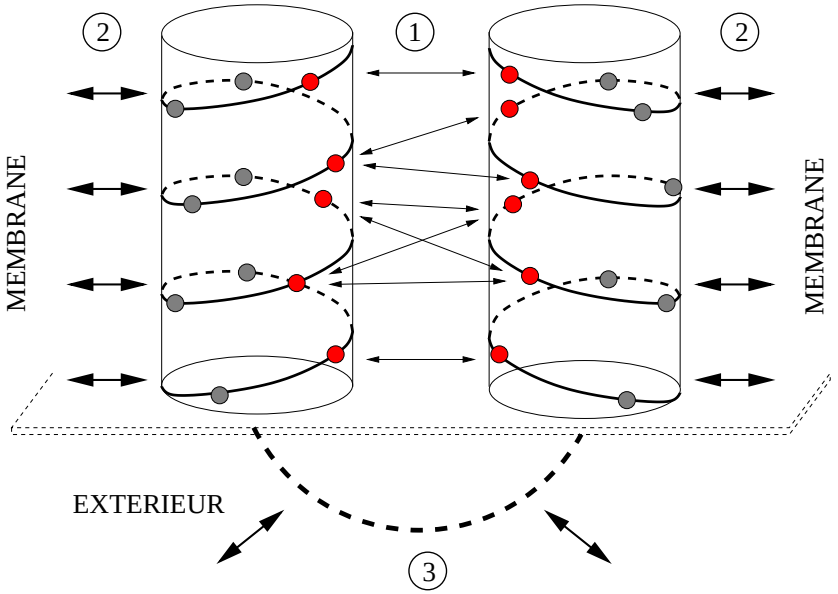
Pseudo folding energy



1. $E_{contact}$ residu interaction energy,

$$E_{contact} = \sum_{i=0}^{n} f(I_i^k, I_i^{k+1}), \text{ où } f(I_i^k, I_i^{k+1}) = \sum_{\omega_j \in I_i^k} \cdot \sum_{\omega_{j'} \in I_i^{k+1}} \frac{\lambda_j \cdot \lambda_{j'}}{\sqrt{\#I_i^k \cdot \#I_i^{k+1}}}$$

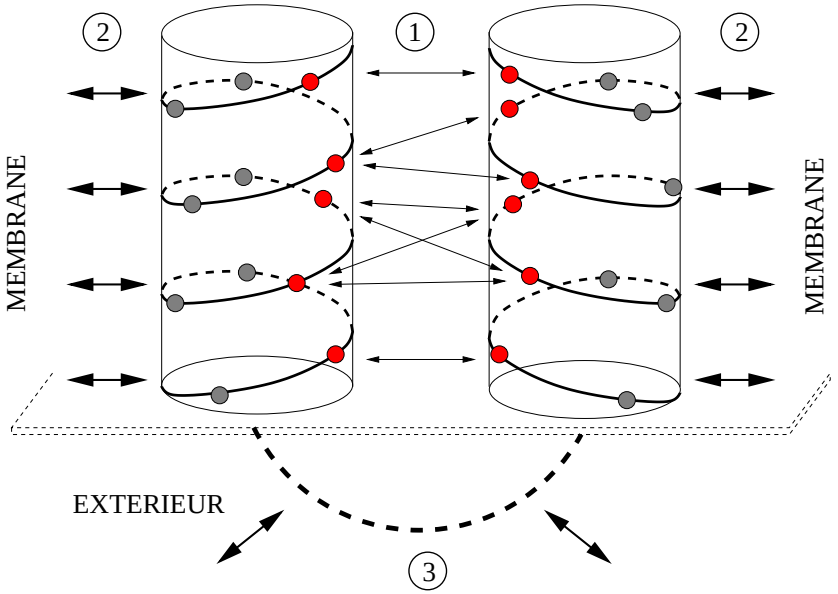
Pseudo folding energy



1. $E_{contact}$ residu interaction energy,
2. E_{memb} membrane interaction energy,

$$E_{memb} = \sum_{\omega_i \in O_i^k} \mathcal{K}_{memb} \cdot \lambda_i$$

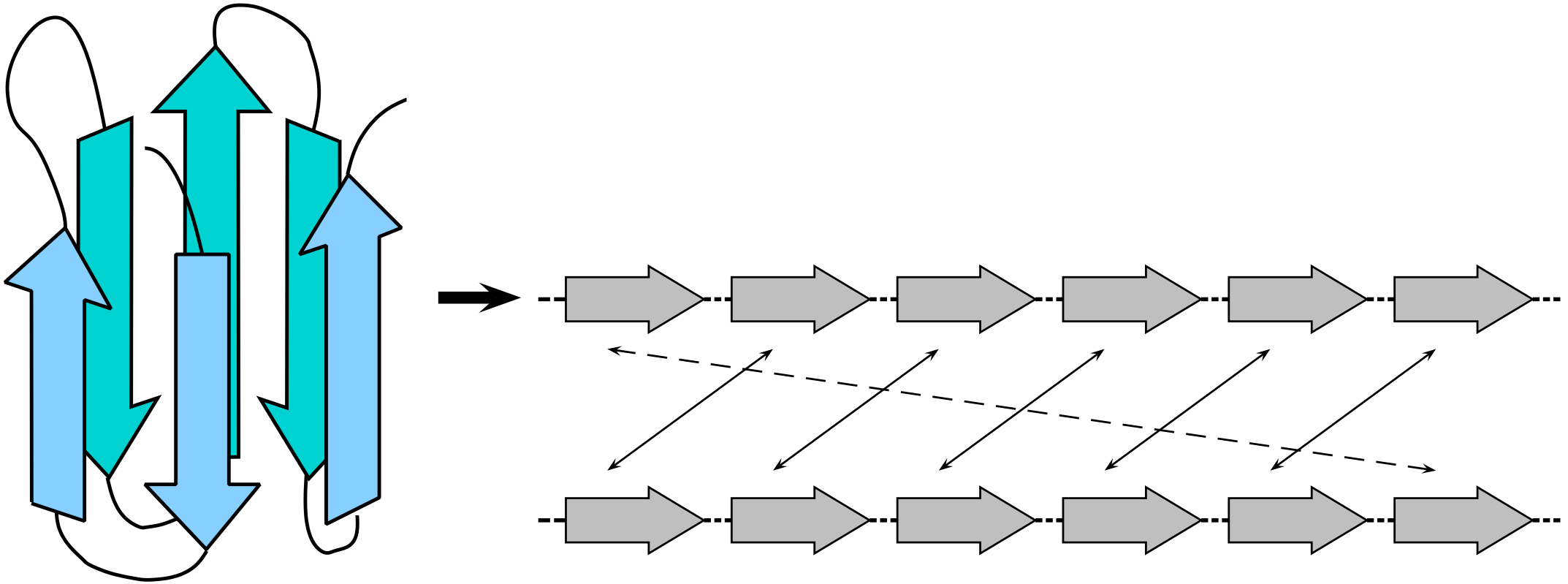
Pseudo folding energy



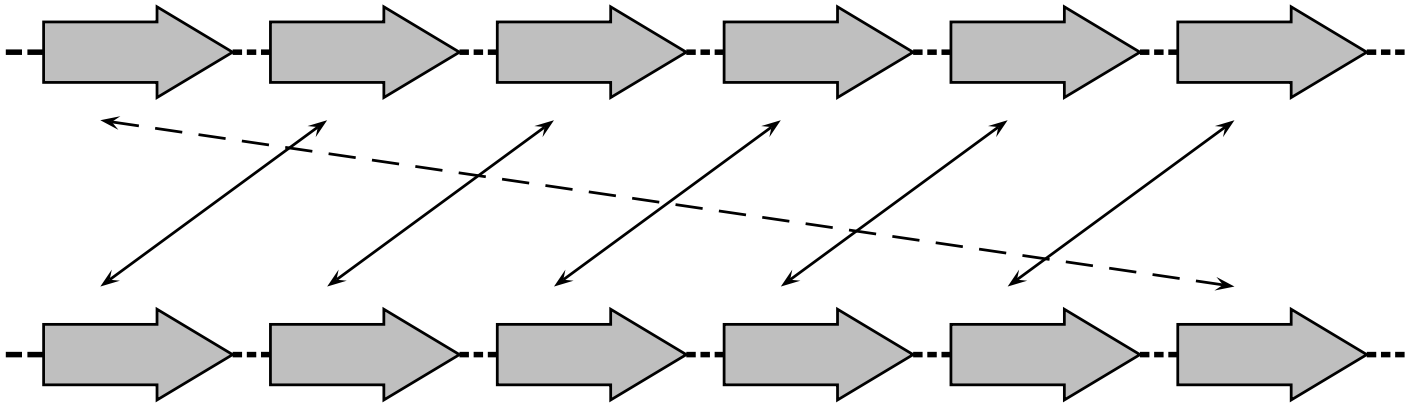
1. $E_{contact}$ residu interaction energy,
2. E_{memb} membrane interaction energy,
3. E_{turn} turn energy.

$$E_{turn} = \mathcal{T}(n-m) + \sum_{i=m}^{n} \mathcal{K}_{cyt/per} \cdot \lambda_i$$

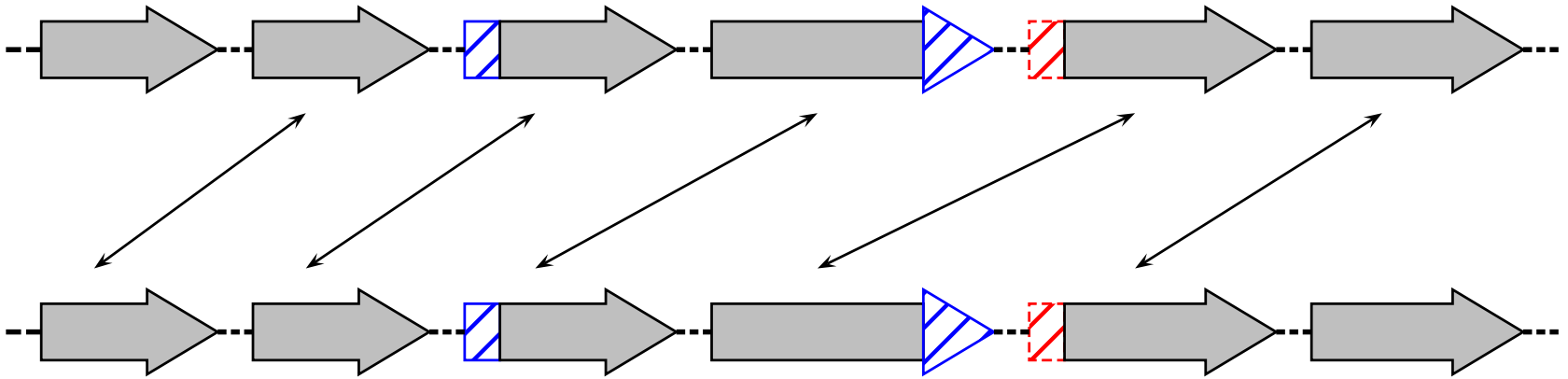
Modeling the overall structure of α-channel



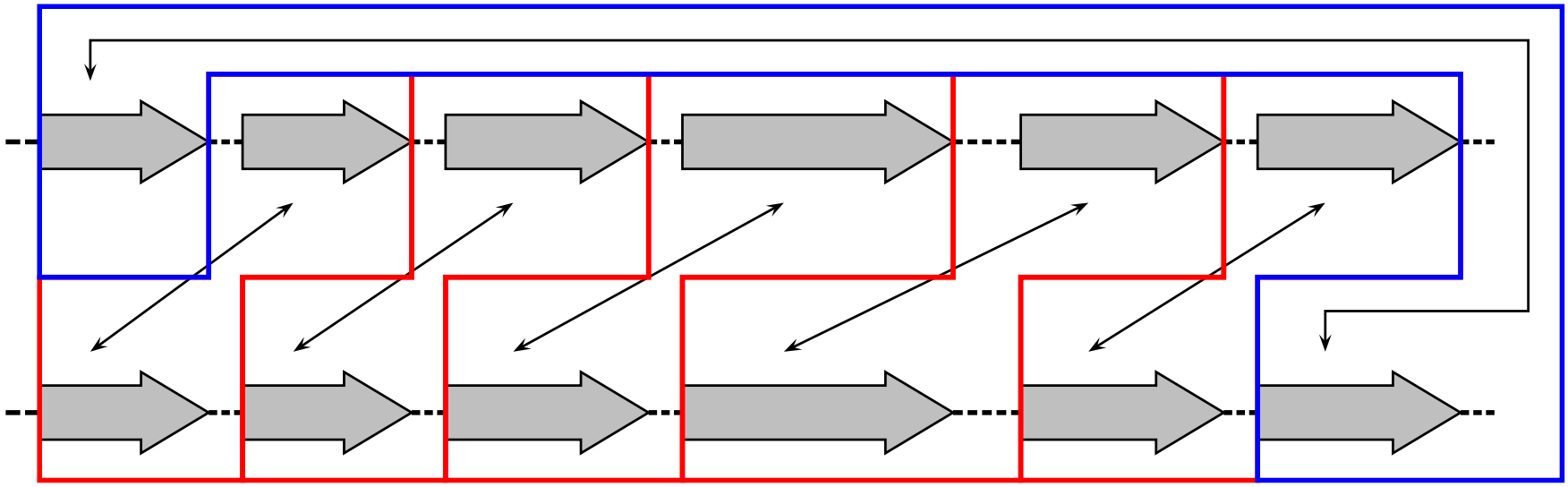
Modeling the overall structure of α-channel



Modeling the overall structure of α-channel



Modeling the overall structure of α-channel



A β-channel is a concatenation of anti-parallel pairings of β-strands.

Grammatical modeling of Transmembrane channels

- Local structure of the secondary structures : rational grammar
- Secondary structure pairing : context-free grammar
- Overall structure of a TM channel : multi-tape context-free grammar
- pseudo folding energy : attributes

Grammatical modeling of TM α-channels

- Regular grammar for α-helix,
- Context-free grammar for α-helix pairings,
- Multi-tape context-free grammar for α-channel,
- Muti-tape S-attribute grammar for α-channel,

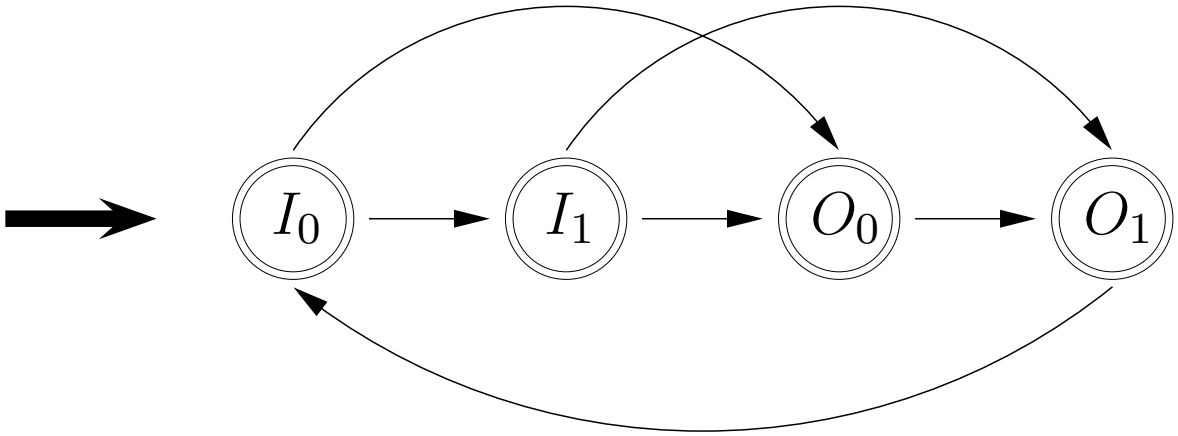
Regular grammar for α-helix

- A helix is an alternate sequence of residues belonging to a I or O face,
- A helix turn is composed by 3 or 4 consecutive amino acids,
- A helical face has 1 or 2 amino acids,
- On average, 3.6 residus per turn.

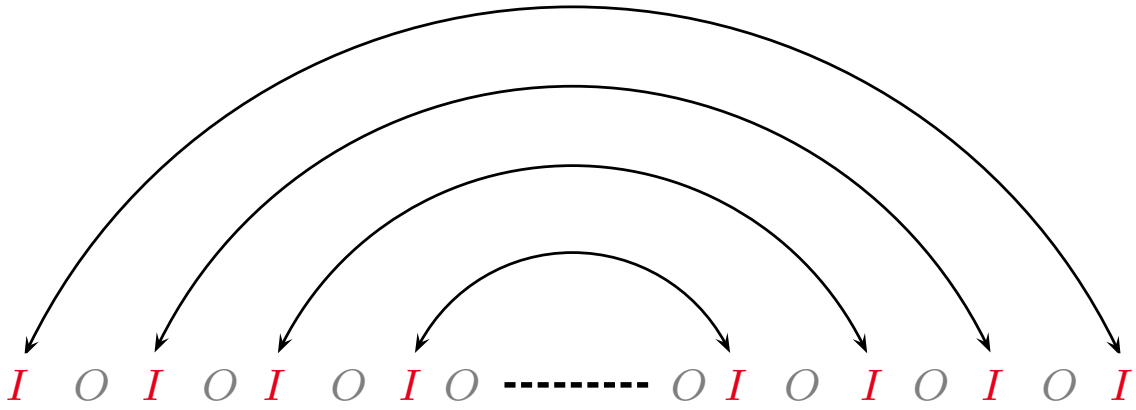
$$P_{helice} = \begin{cases} S_{helice} \rightarrow I_0 \mid I_1 \mid O_0 \mid O_1 \\ I_0 \rightarrow \bullet\ I_1 \mid \bullet\ O_0 \\ I_1 \rightarrow \bullet\ O_0 \mid \bullet\ O_1 \\ O_0 \rightarrow \bullet\ O_1 \\ O_1 \rightarrow \bullet\ I_0 \end{cases}$$

Regular grammar for α-helix

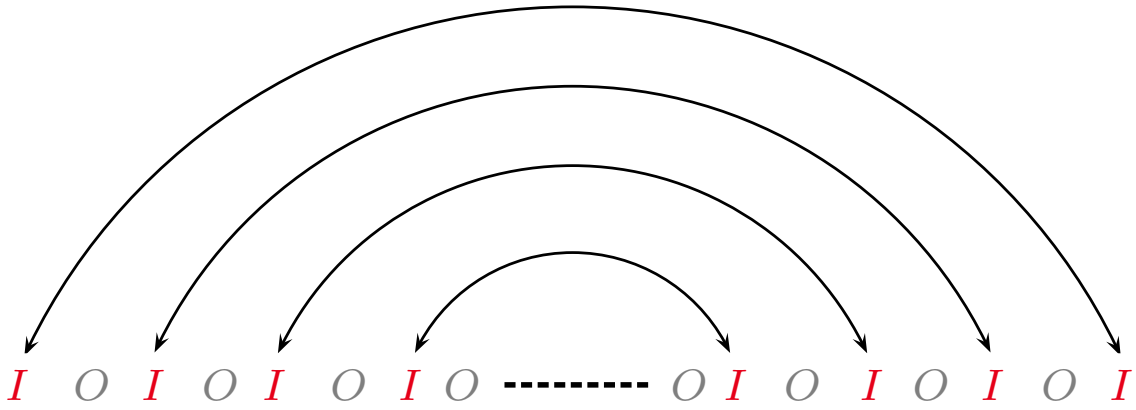
- A helix is an alternate sequence of residues belonging to a I or O face,
- A helix turn is composed by 3 or 4 consecutive amino acids,
- A helical face has 1 or 2 amino acids,
- On average, 3.6 residus per turn.



CFG for α-helix anti-parallel pairings

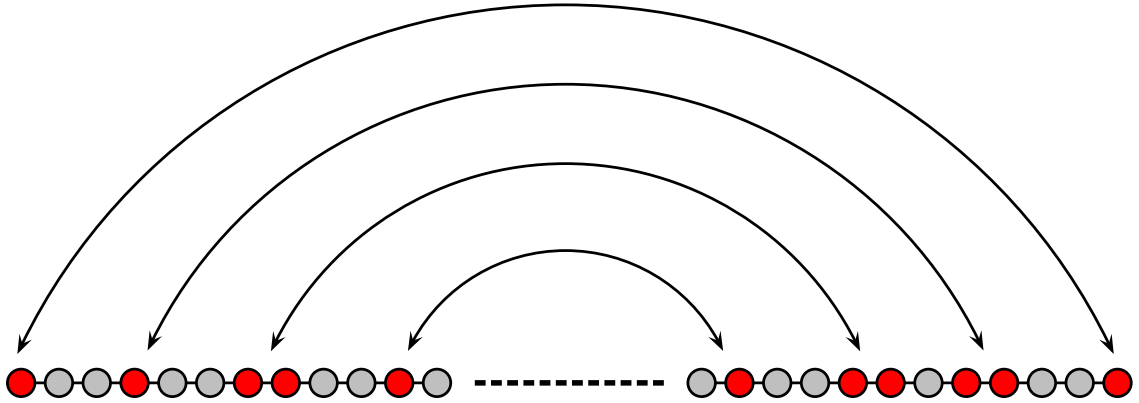


CFG for α-helix anti-parallel pairings



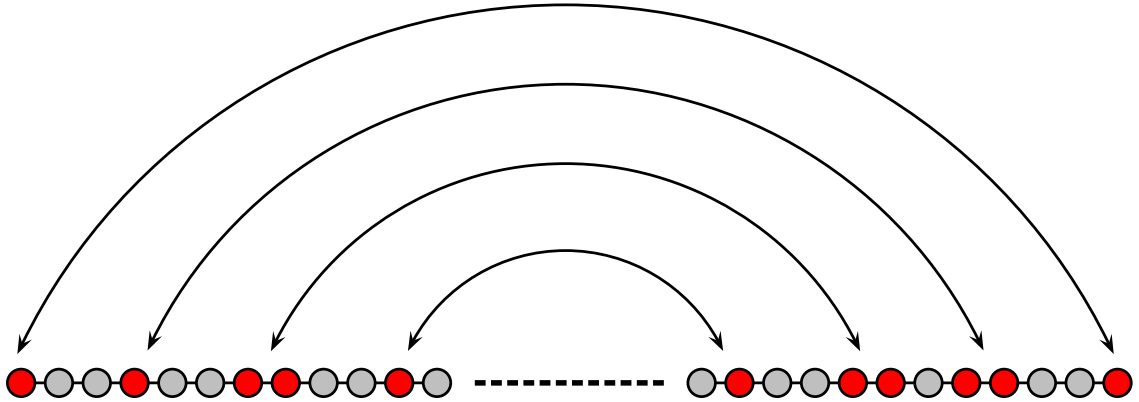
$$P_{ap} = \begin{cases} S^{\alpha}_{ap} & \rightarrow & F^{\alpha}_{O} \mid F^{\alpha}_{I} \\ F^{\alpha}_{I} & \rightarrow & I\ F^{\alpha}_{O}\ I \mid C^{\alpha}_{cyt} \mid C^{\alpha}_{per} \\ F^{\alpha}_{O} & \rightarrow & O\ F^{\alpha}_{I}\ O \mid C^{\alpha}_{cyt} \mid C^{\alpha}_{per} \\ C^{\alpha}_{cyt} & \rightarrow & i\ C^{\alpha}_{cyt} \mid i \\ C^{\alpha}_{per} & \rightarrow & o\ C^{\alpha}_{per} \mid o \end{cases}$$

CFG for α-helix anti-parallel pairings



$$P_{ap} = \begin{cases} S^{\alpha}_{ap} & \rightarrow & F^{\alpha}_{O} \mid F^{\alpha}_{I} \\ F^{\alpha}_{I} & \rightarrow & I\ F^{\alpha}_{O}\ I \mid C^{\alpha}_{cyt} \mid C^{\alpha}_{per} \\ F^{\alpha}_{O} & \rightarrow & O\ F^{\alpha}_{I}\ O \mid C^{\alpha}_{cyt} \mid C^{\alpha}_{per} \\ C^{\alpha}_{cyt} & \rightarrow & i\ C^{\alpha}_{cyt} \mid i \\ C^{\alpha}_{per} & \rightarrow & o\ C^{\alpha}_{per} \mid o \end{cases}$$

CFG for α-helix anti-parallel pairings

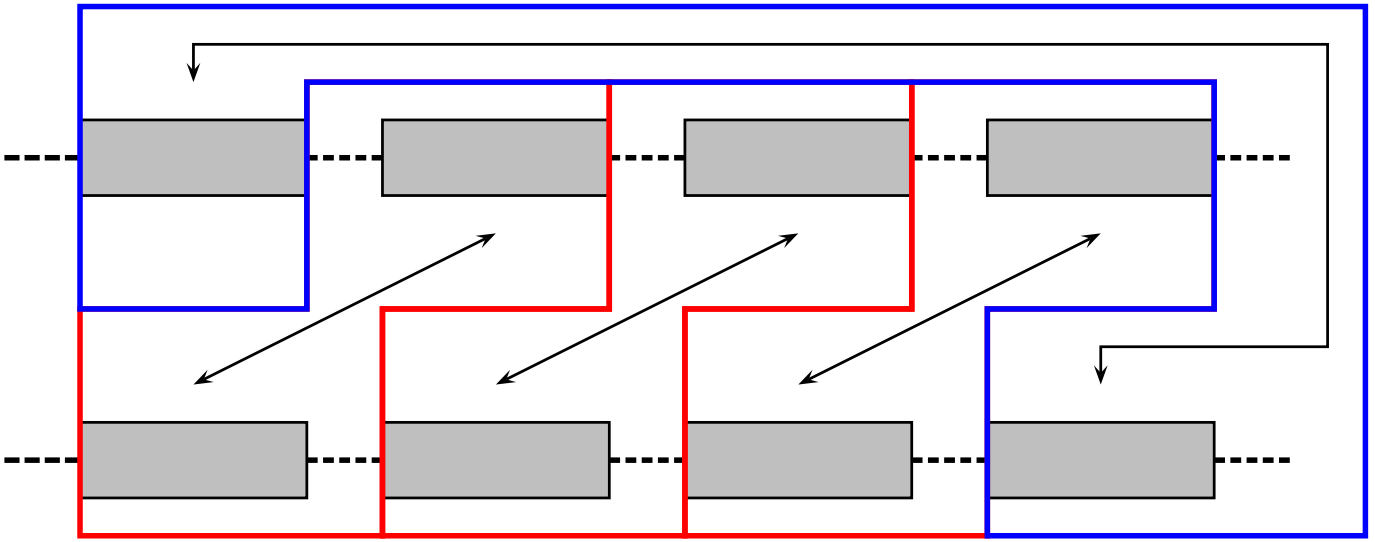


$$P_{ap} = \begin{cases} S^{\alpha}_{ap} & \rightarrow & F^{\alpha}_{O} \mid F^{\alpha}_{I} \\ F^{\alpha}_{I} & \rightarrow & I\ F^{\alpha}_{O}\ I \mid C^{\alpha}_{cyt} \mid C^{\alpha}_{per} \\ F^{\alpha}_{O} & \rightarrow & O\ F^{\alpha}_{I}\ O \mid C^{\alpha}_{cyt} \mid C^{\alpha}_{per} \\ C^{\alpha}_{cyt} & \rightarrow & i\ C^{\alpha}_{cyt} \mid i \\ C^{\alpha}_{per} & \rightarrow & o\ C^{\alpha}_{per} \mid o \end{cases} \quad \bigcirc \quad P_{helice} = \begin{cases} S_{helice} \rightarrow I_0 \mid I_1 \mid O_0 \mid O_1 \\ I_0 \rightarrow \bullet\ I_1 \mid \bullet\ O_0 \\ I_1 \rightarrow \bullet\ O_0 \mid \bullet\ O_1 \\ O_0 \rightarrow \bullet\ O_1 \\ O_1 \rightarrow \bullet\ I_0 \end{cases}$$

CFG for α-helix anti-parallel pairings

$$
P^{\alpha}_{ap} = \begin{cases}
S_{ap} \rightarrow F^{1,1}_{I} \mid F^{1,1}_{O} & 1 \\
F^{1,1}_{I} \rightarrow \bullet\bullet F^{1,1}_{O} \bullet\bullet \mid \bullet F^{2,1}_{O} \bullet\bullet \mid \bullet\bullet F^{1,2}_{O} \bullet \mid \bullet F^{1,1}_{O} \bullet \mid C^{\alpha}_{cyt} & 2 \\
F^{2,1}_{I} \rightarrow \bullet\bullet F^{1,1}_{O} \bullet\bullet \mid \bullet\bullet F^{1,2}_{O} \bullet \mid C^{\alpha}_{cyt} & 3 \\
F^{1,2}_{I} \rightarrow \bullet\bullet F^{1,1}_{O} \bullet\bullet \mid \bullet F^{2,1}_{O} \bullet\bullet \mid C^{\alpha}_{cyt} & 4 \\
F^{2,2}_{I} \rightarrow \bullet\bullet F^{2,2}_{O} \bullet\bullet \mid C^{\alpha}_{cyt} & 5 \\
F^{1,1}_{O} \rightarrow \bullet\bullet F^{1,1}_{I} \bullet\bullet \mid \bullet F^{2,1}_{I} \bullet\bullet \mid \bullet\bullet F^{1,2}_{I} \bullet \mid \bullet F^{2,2}_{I} \bullet \mid C^{\alpha}_{cyt} & 6 \\
F^{2,1}_{O} \rightarrow \bullet\bullet F^{1,1}_{I} \bullet\bullet \mid \bullet\bullet F^{1,2}_{I} \bullet \mid C^{\alpha}_{cyt} & 7 \\
F^{1,2}_{O} \rightarrow \bullet\bullet F^{1,1}_{I} \bullet\bullet \mid \bullet F^{2,1}_{I} \bullet\bullet \mid C^{\alpha}_{cyt} & 8 \\
F^{2,2}_{O} \rightarrow \bullet\bullet F^{1,1}_{I} \bullet\bullet \mid C^{\alpha}_{cyt} & 9 \\
C^{\alpha}_{cyt} \rightarrow \bullet C^{\alpha}_{cyt} \mid \bullet & 10
\end{cases}
$$

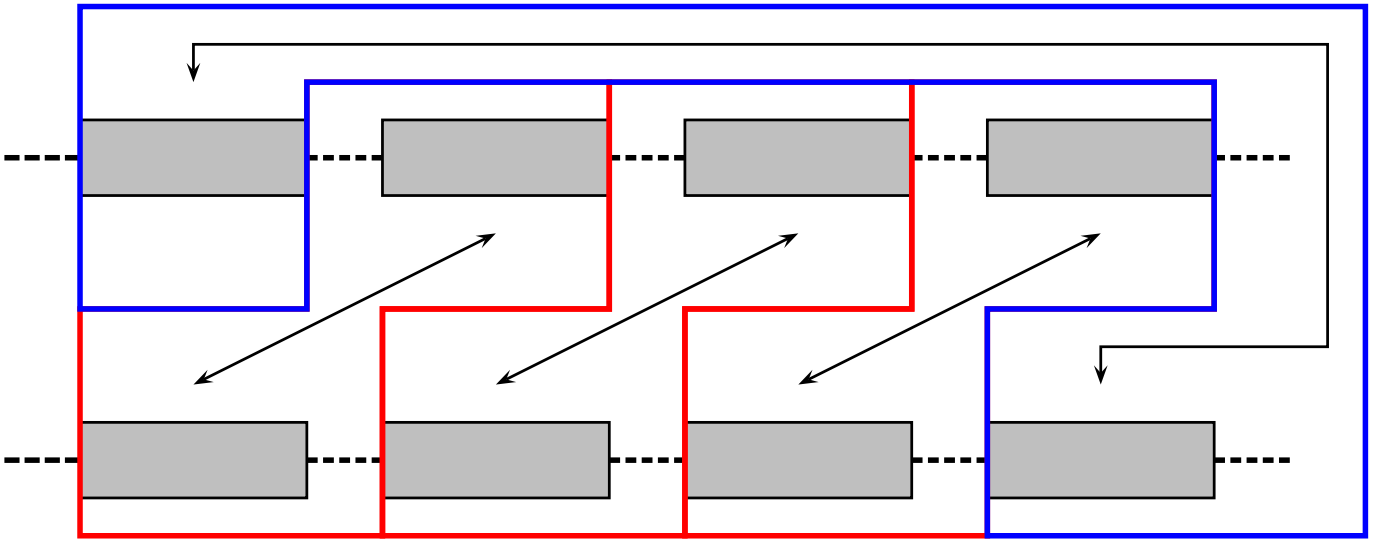
MTCFG for α-channels



A TM-channel is represented by a 2-tape word :

```
))))))iii))))))oo))))))ii))))))ooo))))))
((((((iii((((((oo((((((ii((((((ooo((((((
```

MTCFG for α-channels



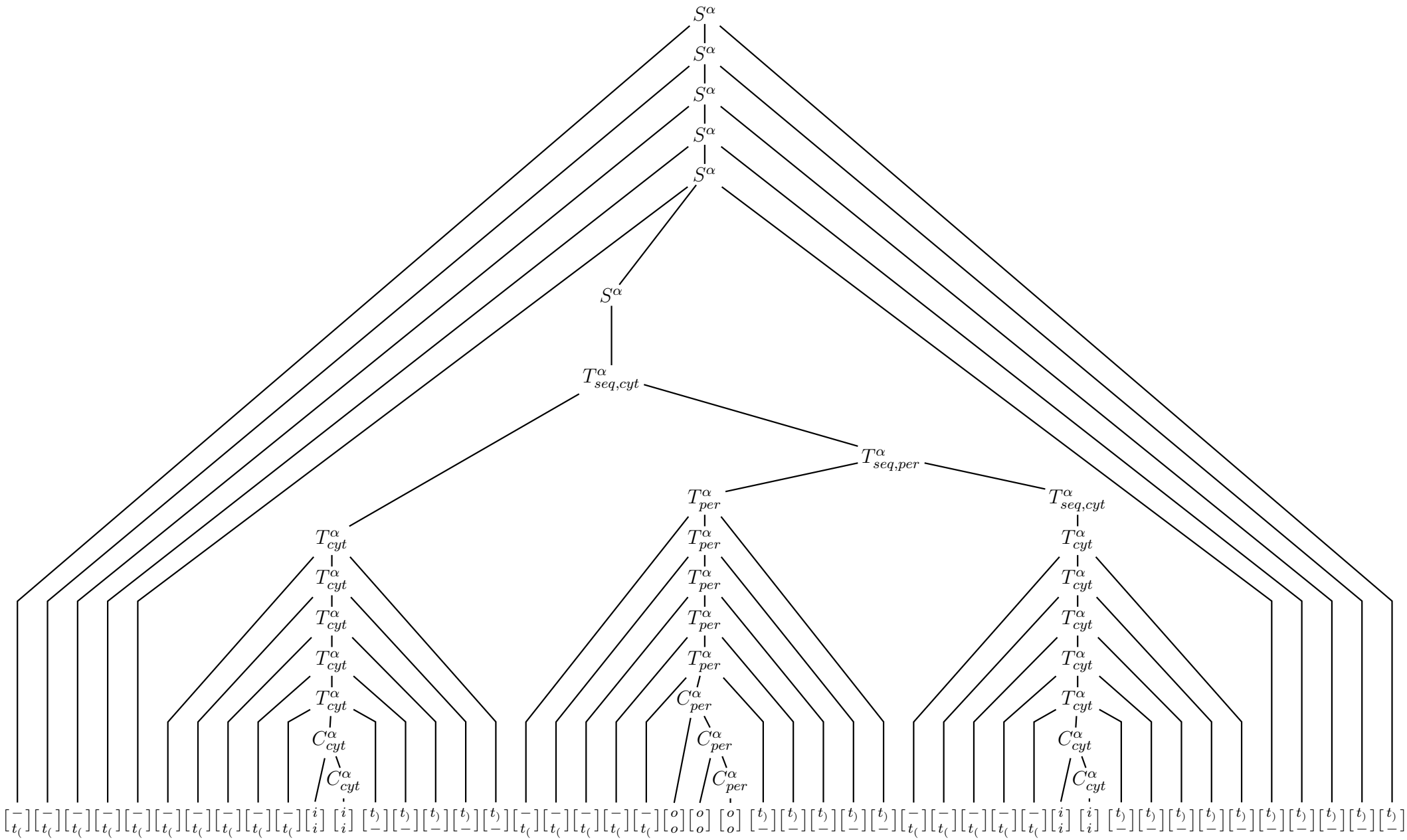
A TM-channel is represented by a 2-tape word :

```
))))))------iii))))))------oo))))))------ii))))))------
------((((((iii------((((((oo------((((((ii------((((((
```

MTCFG for α-channels

$$
P_{canal} = \left\{
\begin{array}{lcll}
S_\alpha & \rightarrow & \begin{bmatrix} t_) \\ - \end{bmatrix} S_\alpha \begin{bmatrix} - \\ t_(\end{bmatrix} \mid T^\alpha_{seq,cyt} \mid T^\alpha_{seq,per} & 1 \\
T^\alpha_{seq,cyt} & \rightarrow & T^\alpha_{cyt}\ T^\alpha_{seq,per} \mid T^\alpha_{cyt} & 2 \\
T^\alpha_{seq,per} & \rightarrow & T^\alpha_{per}\ T^\alpha_{seq,cyt} \mid T^\alpha_{per} & 3 \\
T^\alpha_{cyt} & \rightarrow & \begin{bmatrix} - \\ t_(\end{bmatrix} T^\alpha_{cyt} \begin{bmatrix} t_) \\ - \end{bmatrix} \mid C^\alpha_{cyt} & 4 \\
T^\alpha_{per} & \rightarrow & \begin{bmatrix} - \\ t_(\end{bmatrix} T^\alpha_{per} \begin{bmatrix} t_) \\ - \end{bmatrix} \mid C^\alpha_{per} & 5 \\
C^\alpha_{cyt} & \rightarrow & \begin{bmatrix} i \\ i \end{bmatrix}\ C^\alpha_{cyt} \mid \begin{bmatrix} i \\ i \end{bmatrix} & 6 \\
C^\alpha_{per} & \rightarrow & \begin{bmatrix} o \\ o \end{bmatrix}\ C^\alpha_{per} \mid \begin{bmatrix} o \\ o \end{bmatrix} & 7
\end{array}
\right.
$$

MTCFG for α-channels



MTCFG for α-channels

How to integrate the pairing rules ?

```
))))iii))))oo))))ii))))oo))))
((((iii((((oo((((ii((((oo((((
```

MTCFG for α-channels

How to integrate the pairing rules ?

MPPMMPMMPPMMiiiPPMPPMMPMMPPooMMPPMMPMMPMMiiPPMPPMMPMMPPooMMPMMPMMPPMM
PMMPPMMPPMMPiiiMPPMMPMMPPMMooPMMPPMPPMMPPiiMPPMMPPMMPPMooMPPMMPPMPPMM

MTCFG for α-channels

$$
P^{\alpha}_{ap} = \left\{
\begin{array}{lcll}
S_{ap} & \rightarrow & F_I^{1,1} \mid F_O^{1,1} & 1 \\
F_I^{1,1} & \rightarrow & \bullet\ \bullet\ F_O^{1,1}\ \bullet\ \bullet \mid \bullet\ F_O^{2,1}\ \bullet\ \bullet \mid \bullet\ \bullet\ F_O^{1,2}\ \bullet \mid \bullet\ F_O^{1,1}\ \bullet \mid C^{\alpha} & 2 \\
F_I^{2,1} & \rightarrow & \bullet\ \bullet\ F_O^{1,1}\ \bullet\ \bullet \mid \bullet\ \bullet\ F_O^{1,2}\ \bullet \mid C^{\alpha} & 3 \\
F_I^{1,2} & \rightarrow & \bullet\ \bullet\ F_O^{1,1}\ \bullet\ \bullet \mid \bullet\ F_O^{2,1}\ \bullet\ \bullet \mid C^{\alpha} & 4 \\
F_I^{2,2} & \rightarrow & \bullet\ \bullet\ F_O^{2,2}\ \bullet\ \bullet \mid C^{\alpha} & 5 \\
F_O^{1,1} & \rightarrow & \bullet\ \bullet\ F_I^{1,1}\ \bullet\ \bullet \mid \bullet\ F_I^{2,1}\ \bullet\ \bullet \mid \bullet\ \bullet\ F_I^{1,2}\ \bullet \mid \bullet\ F_I^{2,2}\ \bullet \mid C^{\alpha} & 6 \\
F_O^{2,1} & \rightarrow & \bullet\ \bullet\ F_I^{1,1}\ \bullet\ \bullet \mid \bullet\ \bullet\ F_I^{1,2}\ \bullet \mid C^{\alpha} & 7 \\
F_O^{1,2} & \rightarrow & \bullet\ \bullet\ F_I^{1,1}\ \bullet\ \bullet \mid \bullet\ F_I^{2,1}\ \bullet\ \bullet \mid C^{\alpha} & 8 \\
F_O^{2,2} & \rightarrow & \bullet\ \bullet\ F_I^{1,1}\ \bullet\ \bullet \mid C^{\alpha} & 9 \\
C^{\alpha} & \rightarrow & \bullet\ C^{\alpha} \mid \bullet & 10
\end{array}
\right.
$$

MTCFG for α-channels

$$
P^{\alpha} = \left\{
\begin{array}{lcll}
S_{\alpha} & \rightarrow & \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\ S_{\alpha}\ \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix} \mid Canal & 1 \\
Canal & \rightarrow & F_I^{1,1}\ Canal \mid F_O^{1,1}\ Canal \mid F_I^{1,1} \mid F_O^{1,1} & 2 \\
F_I^{1,1} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{2,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{1,2}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{2,2}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 3 \\
F_I^{2,1} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{1,2}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 4 \\
F_I^{1,2} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{2,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 5 \\
F_I^{2,2} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_O^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 6 \\
F_O^{1,1} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{2,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{1,2}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{2,2}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 7 \\
F_O^{2,1} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{1,2}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 8 \\
F_O^{1,2} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{2,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 9 \\
F_O^{2,2} & \rightarrow & \begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\begin{bmatrix}\varepsilon\\ \bullet\end{bmatrix}\ F_I^{1,1}\ \begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix}\begin{bmatrix}\bullet\\ \varepsilon\end{bmatrix} \mid C^{\alpha} & 10 \\
C^{\alpha} & \rightarrow & \begin{bmatrix}\bullet\\ \bullet{=}1\end{bmatrix}\ C^{\alpha} \mid \begin{bmatrix}\bullet\\ \bullet{=}1\end{bmatrix} & 11
\end{array}
\right.
$$

Multi-tape S-attribute grammar for α-channels

To each production rule, associate a functions which allows a recursive computation of the energy.

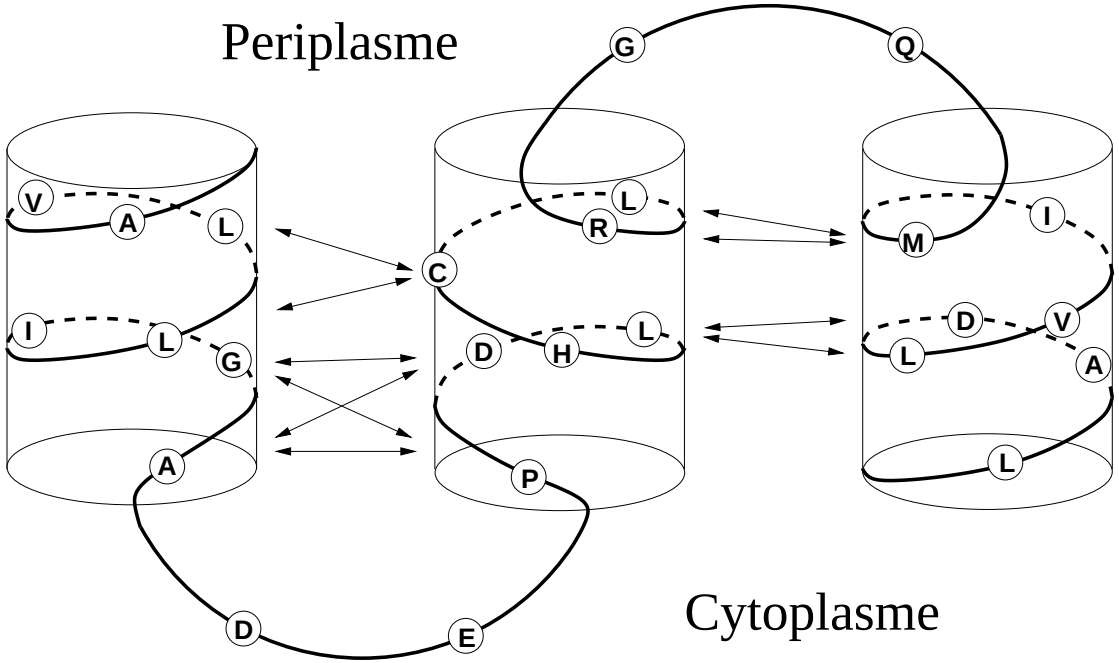
MTSAG for α-channels

$$
F^{\alpha}_{ap} = \left\{ \begin{array}{lcl}
f^{energy}_{F^{1,1}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} + \frac{(u.\text{hp}+v.\text{hp})\cdot(y.\text{hp}+z.\text{hp})}{2} \\
f^{energy}_{F^{1,1}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{2,1}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uxyz) & = & x.\text{energy} + \frac{(u.\text{hp})\cdot(y\ \text{hp}+z.\text{hp})}{\sqrt{2}} \\
f^{energy}_{F^{1,1}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,2}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxy) & = & x.\text{energy} + \frac{(u.\text{hp}+v.\text{hp})\cdot(y.\text{hp})}{\sqrt{2}} \\
f^{energy}_{F^{1,1}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{2,2}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uxy) & = & x.\text{energy} + u\ \text{hp} \cdot y.\text{hp} \\
f^{energy}_{F^{1,1}_I \to C^{\alpha}}(x) & = & x.\text{energy} \\
f^{energy}_{F^{2,1}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} + \frac{(u.\text{hp}+v.\text{hp})\cdot(y.\text{hp}+z.\ \text{hp})}{2} \\
f^{energy}_{F^{2,1}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,2}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxy) & = & x.\text{energy} + \frac{(u.\text{hp}+v.\text{hp})\cdot(y.\text{hp})}{\sqrt{2}} \\
f^{energy}_{F^{2,1}_I \to C^{\alpha}}(x) & = & x.\text{energy} \\
f^{energy}_{F^{1,2}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} + \frac{(u.\text{hp}+v.\text{hp})\cdot(y.\text{hp}+z.\ \text{hp})}{2} \\
f^{energy}_{F^{1,2}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{2,1}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uxyz) & = & x.\text{energy} + \frac{(u.\text{hp})\cdot(y\ \text{hp}+z.\text{hp})}{\sqrt{2}} \\
f^{energy}_{F^{1,2}_I \to C^{\alpha}}(x) & = & x.\text{energy} \\
f^{energy}_{F^{2,2}_I \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_O \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} + \frac{(u.\text{hp}+v.\text{hp})\cdot(y.\text{hp}+z.\ \text{hp})}{2} \\
f^{energy}_{F^{2,2}_I \to C^{\alpha}}(x) & = & x.\text{energy}
\end{array} \right.
$$

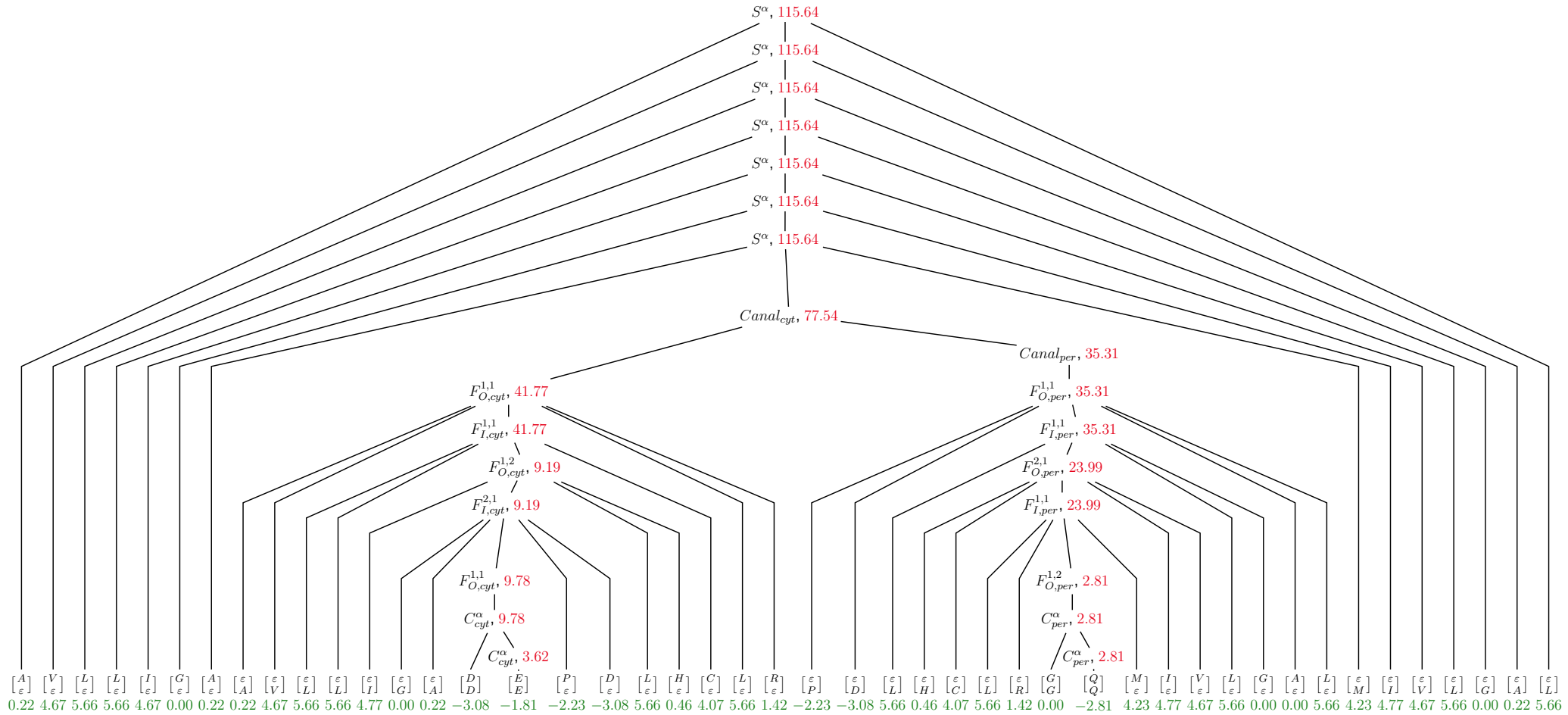
$$
\begin{array}{lcl}
f^{energy}_{F^{1,1}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} \\
f^{energy}_{F^{1,1}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{2,1}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uxyz) & = & x.\text{energy} \\
f^{energy}_{F^{1,1}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,2}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxy) & = & x.\text{energy} \\
f^{energy}_{F^{1,1}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{2,2}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uxy) & = & x.\text{energy} \\
f^{energy}_{F^{1,1}_O \to C^{\alpha}}(x) & = & x.\text{energy} \\
f^{energy}_{F^{2,1}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} \\
f^{energy}_{F^{2,1}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,2}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxy) & = & x.\text{energy} \\
f^{energy}_{F^{2,1}_O \to C^{\alpha}}(x) & = & x.\text{energy} \\
f^{energy}_{F^{1,2}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} \\
f^{energy}_{F^{1,2}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{2,1}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uxyz) & = & x.\text{energy} \\
f^{energy}_{F^{1,2}_O \to C^{\alpha}}(x) & = & x.\text{energy} \\
f^{energy}_{F^{2,2}_O \to \left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right]\left[\begin{smallmatrix}\varepsilon\\\bullet\end{smallmatrix}\right] F^{1,1}_I \left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]\left[\begin{smallmatrix}\bullet\\\varepsilon\end{smallmatrix}\right]}(uvxyz) & = & x.\text{energy} \\
f^{energy}_{F^{2,2}_O \to C^{\alpha}}(x) & = & x.\text{energy} \\
f^{energy}_{C^{\alpha} \to \left[\begin{smallmatrix}\bullet\\\bullet=1\end{smallmatrix}\right] C^{\alpha}}(xy) & = & x.\text{hp} \cdot \mathcal{K}_{milieu} + y.\text{energy} \\
f^{energy}_{C^{\alpha} \to \left[\begin{smallmatrix}\bullet\\\bullet=1\end{smallmatrix}\right]}(x) & = & x.\text{hp} \cdot \mathcal{K}_{milieu}
\end{array}
$$

An example !

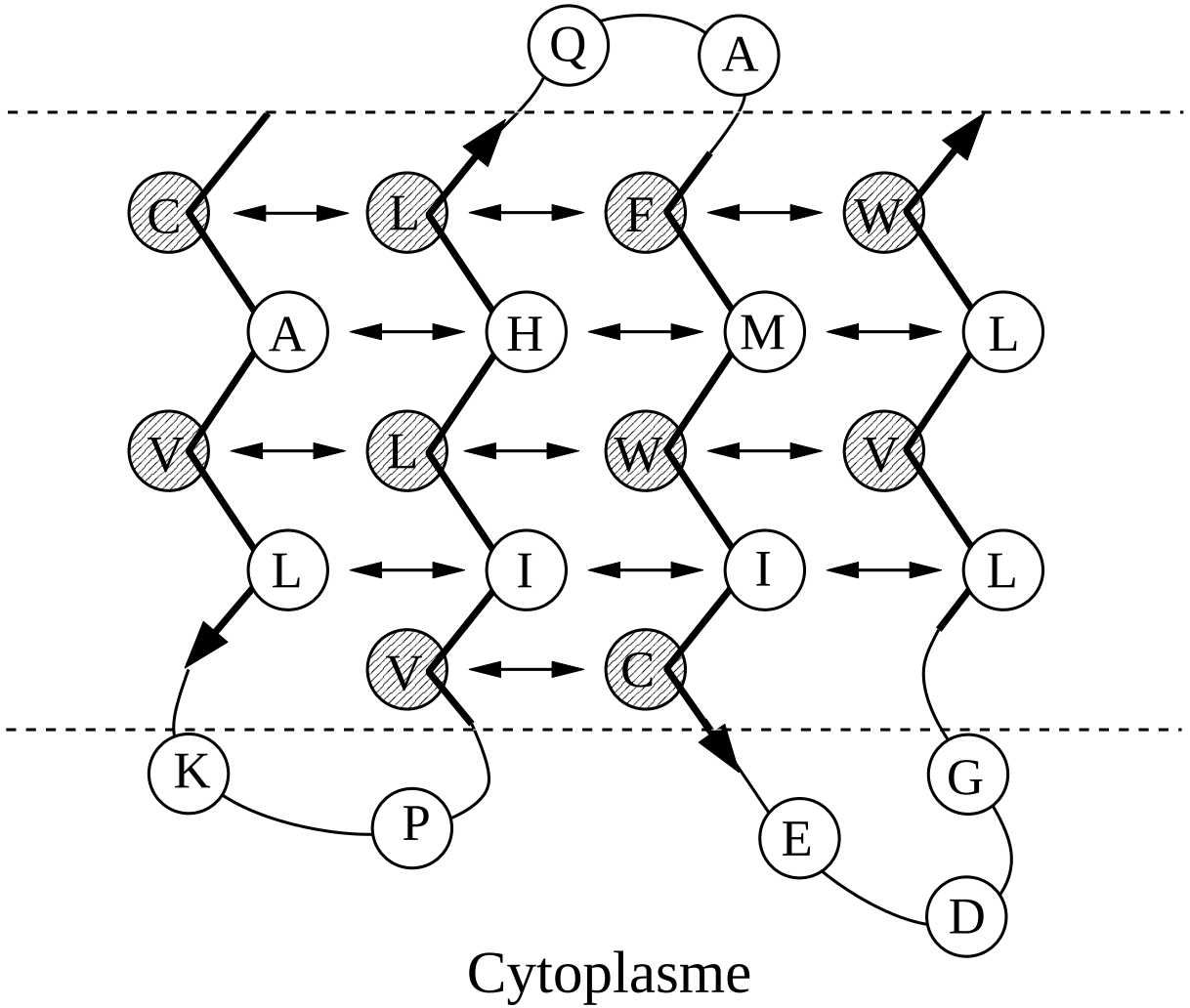
MTSAG for α-channels



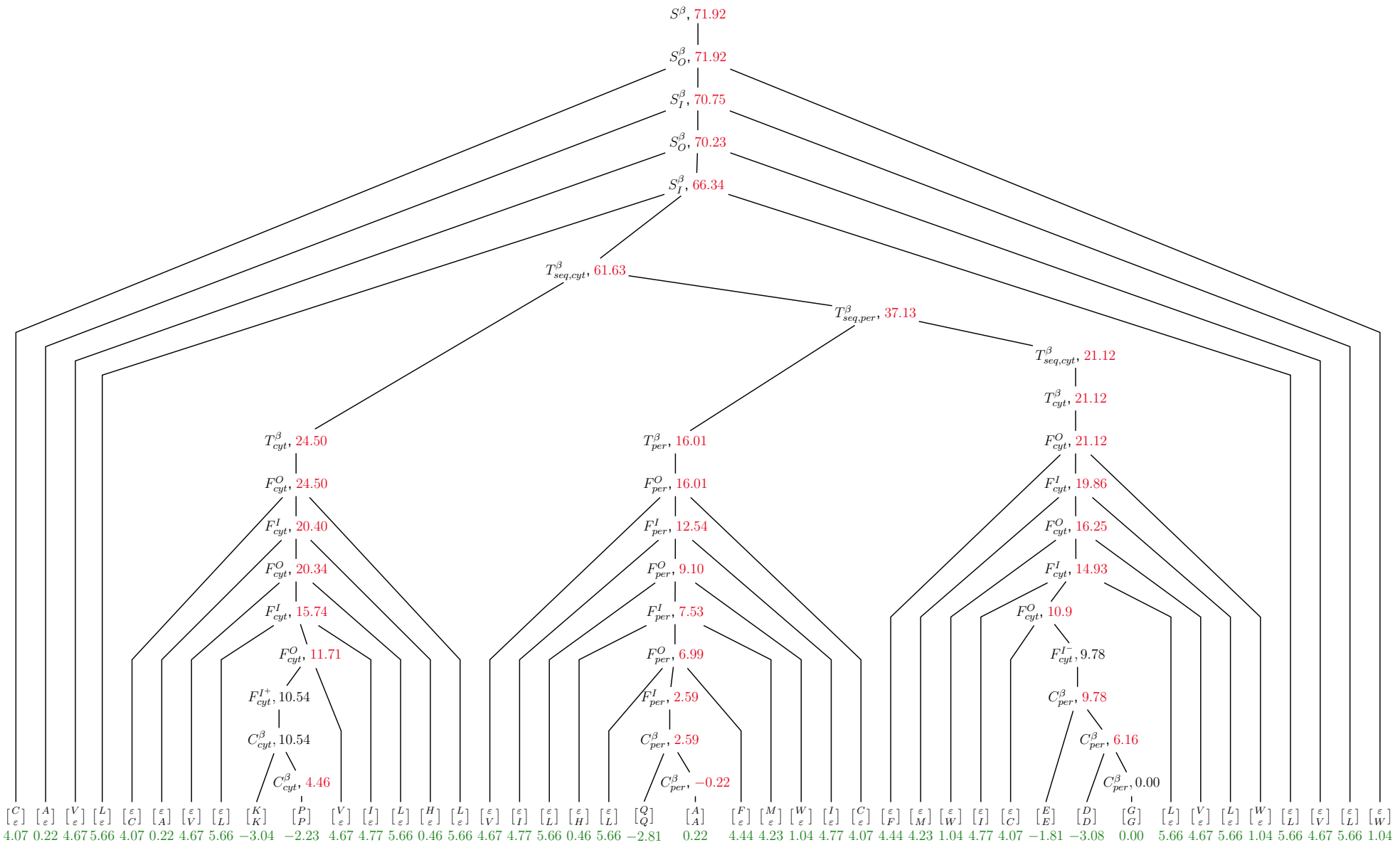
MTSAG for α-channels



MTSAG for β-channels



MTSAG for β-channels



MTSAG for TM channels

What has not been said :

- TM-channel closure,
- TM α-helix selection,
- turn selection (between secondary structures),
- constraints on the overlapping of the motifs.

Performance evaluation

- How to realize a structure prediction ?
- How to evaluate a prediction ?
- Results.

How to realize a structure prediction ?

- syntax analysis (GCP algorithm),
- implementation using *mtsag2c* (F. Lefebvre),
- software *tmmtsag*... and now ASTRiD (web interface).

How to realize a structure prediction ?

Example of an α-channel : Bacteriorhodopsin

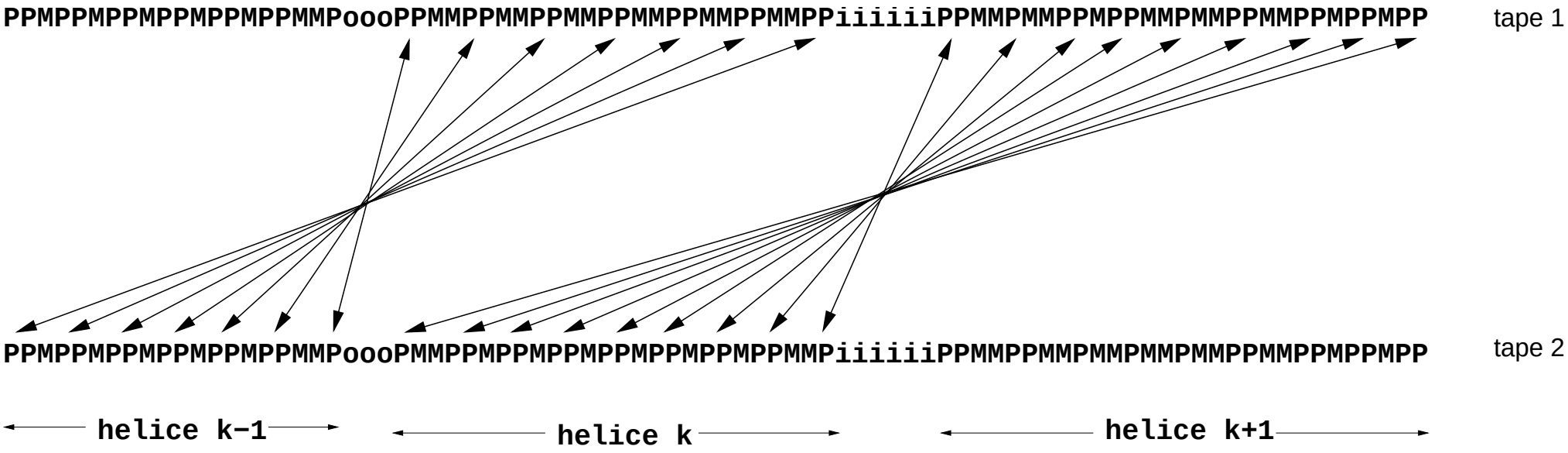
```
QAQITGRPEWIWLALGTALMGLGTLYFLVKGMGVSDPDAKKFYAITTLVPAIAFTMYLSMLLGYGLTMVPFGGEQNPIYWARYADWLFTTPLLLLDLALL
.......TTHHHHHHHHHHHTTHHHHHHHHSS..S.HHHHHHHHHHHHTHHHHHHHHHHHHTT.....SSS.SSS....STTHHHHTTTHHHHTTTTSTT
............MMMMMMMMMMMMMMMMMMMMiiiiiiiPPMPPMPPMMPPMMPPMMPPMMPPMMPPooooooooPPMPPMPPMPPMMPPMPPMPPMP
............PMMPMMPMMPMMPMMPMMPMMPiiiiiiiPMMPMMPPMMPPMMPPMMPPMMPPMMPPMMPPooooooooPPMMPPMMPPMMPPMMPMMPMMP

VDADQGTILALVGADGIMIGTGLVGALTKVYSYRFVWWAISTAAMLYILYVLFFGFTSKAESMRPEVASTFKVLRNVTVVLWSAYPVVWLIGSEGAGIVP
TT..HHHHHHHHHHHHHHHHHHHHHS..SSS.HHHHHHHHHHHHHHHHHHTTTTTTT..TT.SHHHHTTHHHHHHHHHHHHHHHHHHTTTTSSSSSS.
PiiiiiiPPMPPMPPMPPMPPMPPMMPoooPPMMPPMMPPMMPPMMPPMMPPMMPPiiiiiiPPMMPMMPPMPPMMPMMPPMMPPMPPMPPooooooPPM
PiiiiiiPPMPPMPPMPPMPPMPPMMPoooPMMPPMPPMPPMPPMPPMPPMPPMMPiiiiiiPPMMPPMMPMMPMMPMMPPMMPPMPPMPPooooooMMM

LNIETLLFMVLDVSAKVGFGLILLRSRAIFGEAEAPEPSAGDGAAATS
SHHHHHHHHHHHHHHTHHHHTTTT.........................
PPMPPMPPMPPMPPMMPMMPPMPP.........................
MMMMMMMMMMMMMMMMMMMMMMM.........................

pseudo folding energy : 1583.92
```

How to realize a structure prediction ?



How to realize a structure prediction ?

Example of a β-channel : Porin

```
MAPKDNTWYTGAKLGWSQYHDTGLINNNGPTHENKLGAGAFGGYQVNPYVGFEMGYDWLGRMPYKGSVENGAYKAQGVQLTAKLGYPITDDLDIYTRLGG
....TT.EEEEEEEEEES.S.....SS.......EEEEEEEEEEE.BTTEEEEEEEEEEEE.....SS....EEEEEEEEEEEEEEEESSSSEEEEEEEEE
...............EEEEEEEEEEooooooooooCBCBCBCBCiiiiiiBCBCBCBCBCBCooooooooooooooBCBCBCBCBiiiiiiiiBCBCBCB
...............CBCBCBCBCBooooooooooBCBCBCBCBiiiiiiBCBCBCBCBCBCooooooooooooooBCBCBCBCBiiiiiiiiCBCBCBC

MVWRADTYSNVYGKNHDTGVSPVFAGGVEYAITPEIATRLEYQWTNNIGDAHTIGTRPDNGMLSLGVSYRFG
EEEEEEE..SSS..EEEEEEEEEEEEEEEEESSSSEEEEEEEEEE......SS........EEEEEEEEEE.
CBCBCooooooooooooooCBCBCBCBCiiiCBCBCBCBCBCooooooooooooooooooooBCBCBCBCBC.
BCBCBooooooooooooooCBCBCBCBCiiiCBCBCBCBCBCooooooooooooooooooooEEEEEEEEEE.

pseudo folding energy : 402.15
```

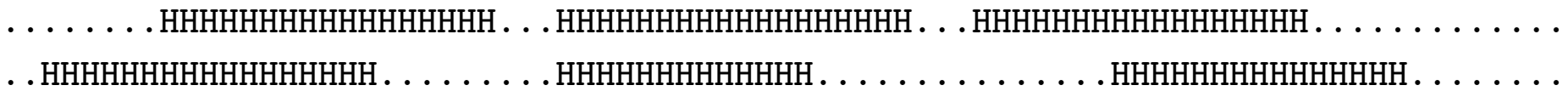
How to evaluate a prediction ?

```
observed : ......HHHHHHHHHHHHHHHHHH...HHHHHHHHHHHHHHHHHH...HHHHHHHHHHHHHHHH........
predition : ..HHHHHHHHHHHHHHHH.........HHHHHHHHHHHH.............HHHHHHHHHHHHHHHH...
```

Definition *A secondary structure is said to be predicted, if it intersects one and only one observed secondary structure.*

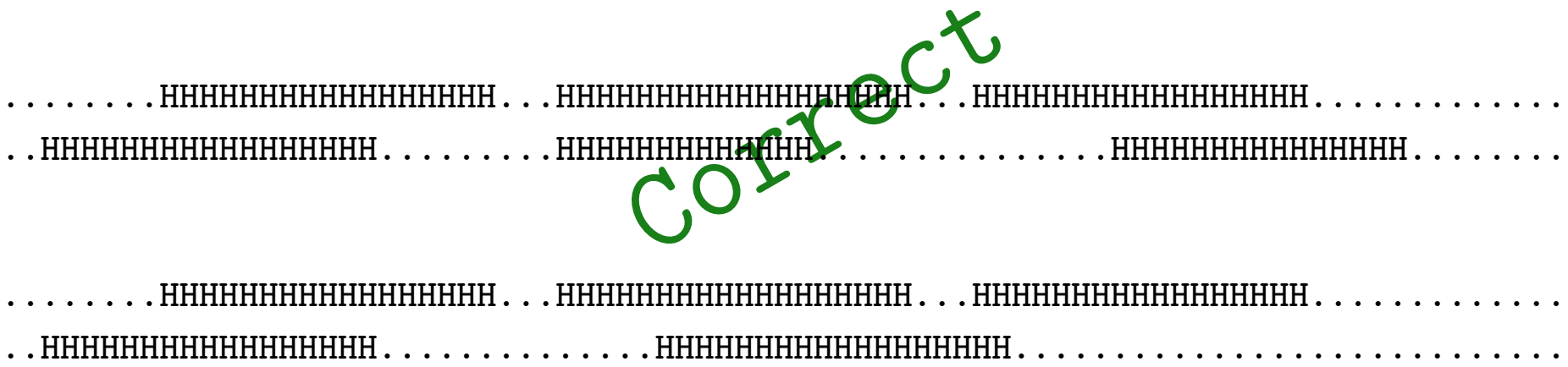
Definition *A structure is correctly predicted if all its secondary structures are predicted, almost predicted if the non-predicted secondary structures do not intersect any observed secondary structures, and non-predicted otherwise.*

How to evaluate a prediction ?

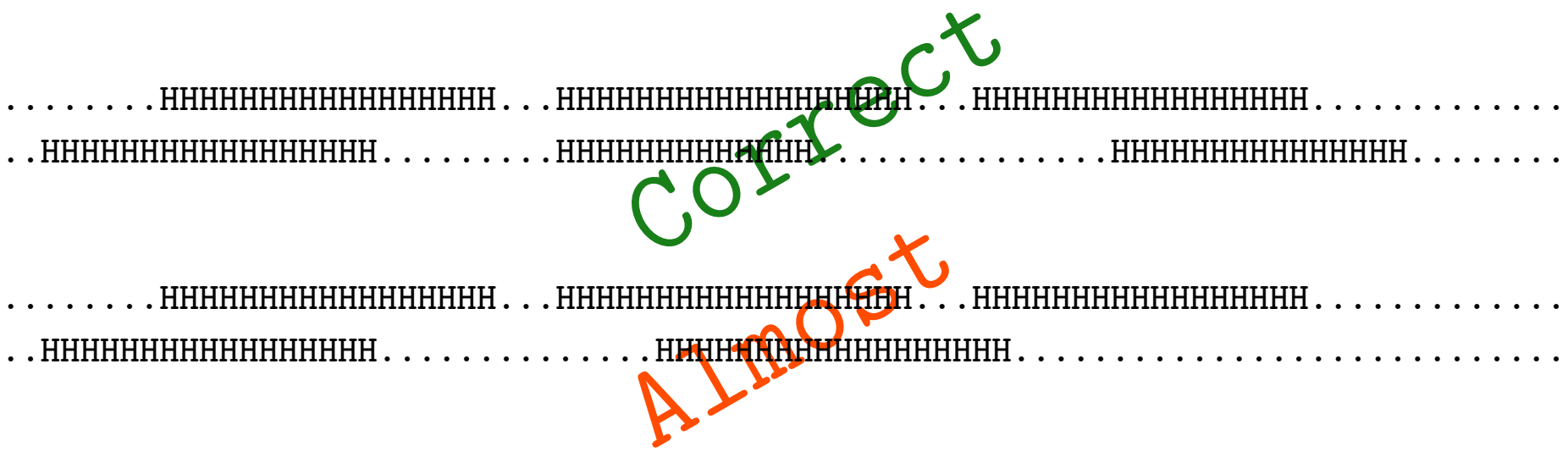


How to evaluate a prediction ?

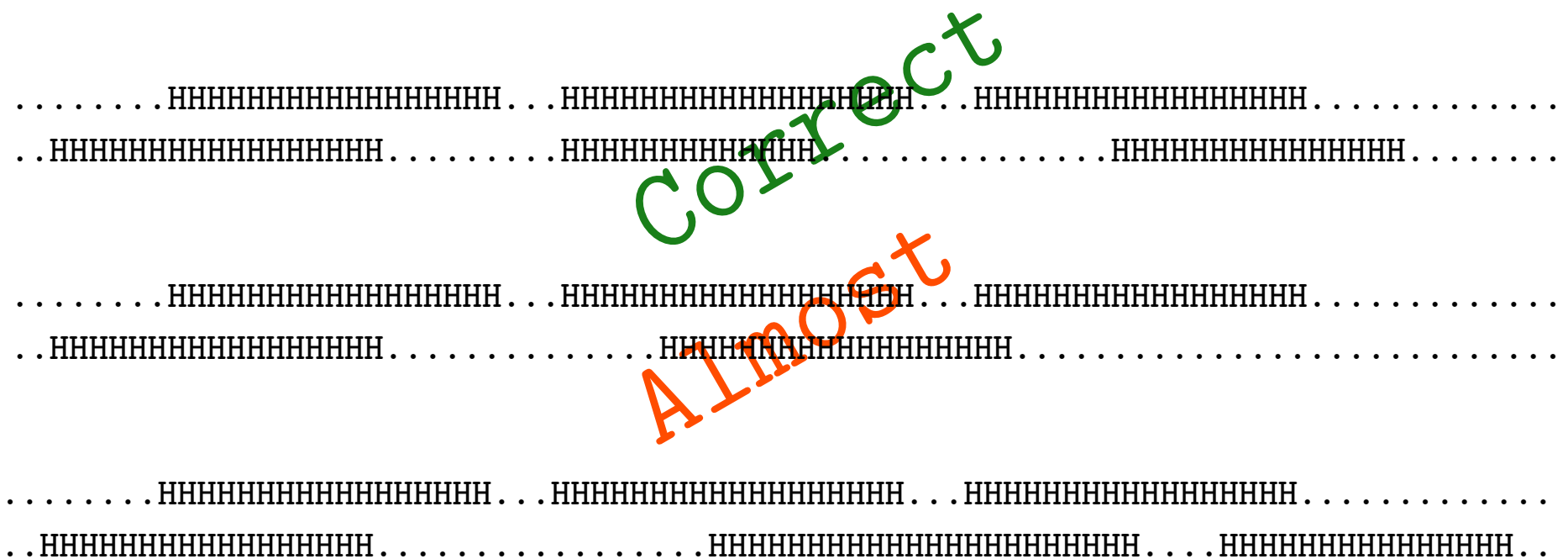
How to evaluate a prediction ?



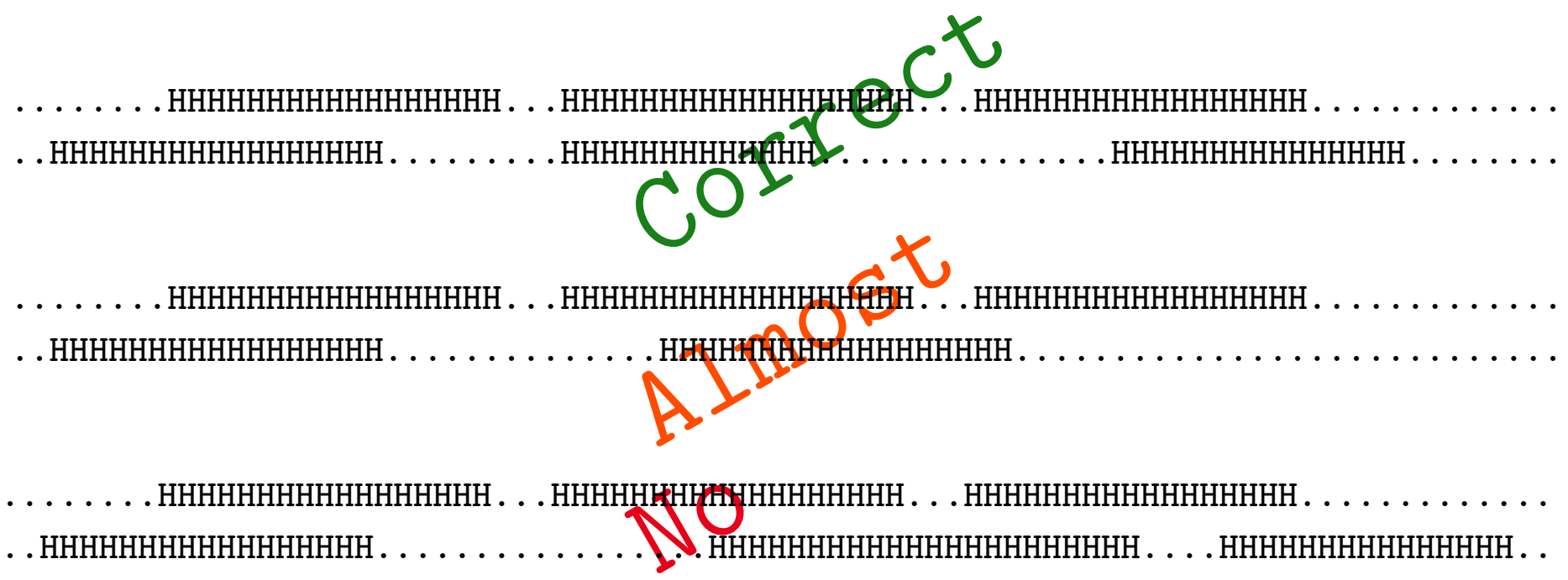
How to evaluate a prediction ?



How to evaluate a prediction ?



How to evaluate a prediction ?



How to evaluate a prediction ?

Estimator for the secondary structure element prediction

$$Q_{ok} = 100 \cdot \frac{\text{number of correctly predicted structures}}{\text{number of proteins}}$$

$$Q_{stm}^{\%obs} = 100 \cdot \frac{\text{number of TM segments correctly predited}}{\text{number of TM segment observed}}$$

$$Q_{stm}^{\%pred} = 100 \cdot \frac{\text{number of TM segments correctly predited}}{\text{number of TM segment predicted}}$$

How to evaluate a prediction ?

Estimator for the secondary structure assignment prediction

$$Q_2 = 100 \cdot \frac{\text{number of correctly predicted residus}}{\text{number of residus}}$$

$$Q_{2T}^{\%obs} = 100 \cdot \frac{\text{number of correctly predicted residus in TM segment}}{\text{number of residus observed in TM segments}}$$

$$Q_{2T}^{\%pred} = 100 \cdot \frac{\text{number of correctly predicted residus in TM segment}}{\text{number of residus predicted in TM segments}}$$

$$Q_{2N}^{\%obs} = 100 \cdot \frac{\text{number of correctly predicted residus in non-TM segment}}{\text{number of residus observed in non-TM segments}}$$

$$Q_{2N}^{\%pred} = 100 \cdot \frac{\text{number of correctly predicted residus in non-TM segment}}{\text{number of residus predicted in non-TM segments}}$$

Procedure

- 28 α-TM proteins known at high resolution level,
- 82 α-TM proteins known at low resolution level,
- 14 β-TM proteins known at high resolution level,
- 567 globular proteins,
- computation of sub-optimals structures,
- comparison with 8 other software.

Results

α-TM proteins known at high resolution level

Method	topology	helices		2-states	TM residus		non-TM residus	
	Q_{ok}	$Q_{stm}^{\%obs}$	$Q_{stm}^{\%pred}$	Q_2	$Q_{2T}^{\%obs}$	$Q_{2T}^{\%pred}$	$Q_{2N}^{\%obs}$	$Q_{2N}^{\%pred}$
tmmtsag-basic	75.00(92.86)	97.18	93.88	78.24	87.56	78.77	64.04	77.16
tmmtsag-opt	92.86(100.00)	99.30	98.60	80.06	89.17	80.07	66.17	80.04
hmmtop2	71.43(100.00)	97.18	96.50	79.50	71.84	92.55	91.18	68.00
memsat	60.00(100.00)	94.00	97.92	77.43	67.63	94.34	93.39	63.91
phd-psihtm	71.43(96.43)	88.73	96.92	76.29	68.76	89.55	87.77	64.83
pred-tmr	53.57(100.00)	88.73	100.00	74.41	59.12	97.52	97.71	61.07
sosui	82.14(100.00)	97.18	99.28	80.70	72.71	93.95	92.87	69.10
tmhmm1	71.43(100.00)	96.48	97.16	79.80	72.76	92.13	90.52	68.56
toppred2	71.43(100.00)	95.07	94.41	75.69	66.31	90.98	89.98	63.67

Results

α-TM proteins known at low resolution level

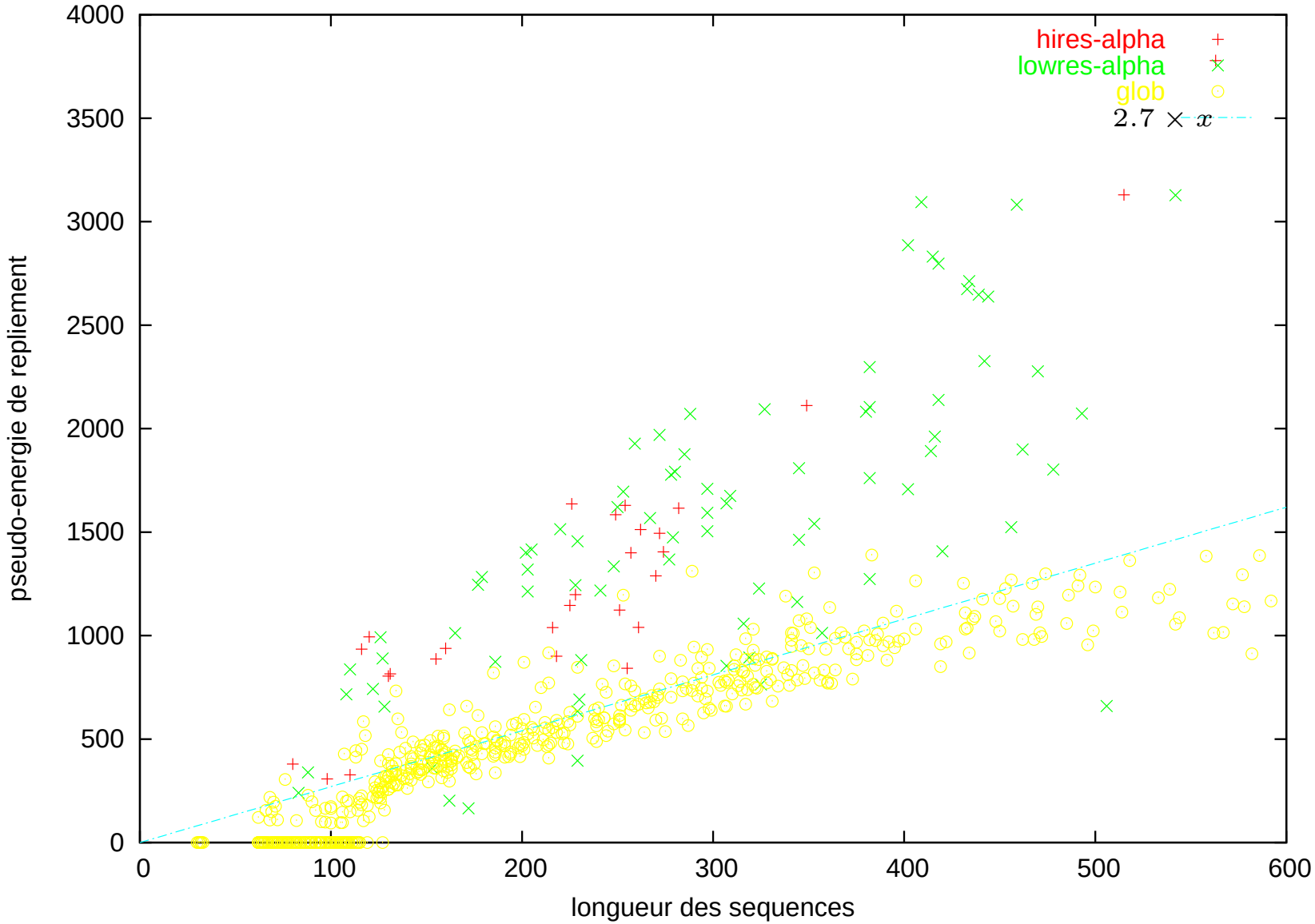
Method	topology	helices		2-states	TM residus		non-TM residus	
	Q_{ok}	$Q_{stm}^{\%obs}$	$Q_{stm}^{\%pred}$	Q_2	$Q_{2T}^{\%obs}$	$Q_{2T}^{\%pred}$	$Q_{2N}^{\%obs}$	$Q_{2N}^{\%pred}$
tmmtsag-basic	37.80(73.17)	87.11	81.68	69.30	87.06	60.31	55.50	84.67
tmmtsag-opt	60.98(95.12)	95.70	91.42	75.25	90.42	65.79	63.47	89.51
hmmtop2	60.98(86.59)	87.11	93.31	84.39	81.69	82.44	86.48	85.88
memsat	54.88(92.68)	91.02	93.20	85.26	81.40	84.80	88.35	85.60
phd-psihtm	24.36(43.59)	60.04	72.02	83.99	91.08	76.60	78.51	91.93
pred-tmr	50.00(96.34)	88.09	96.57	85.19	75.86	88.63	92.44	83.14
sosui	48.78(90.24)	86.33	94.85	82.36	79.17	80.21	84.83	83.98
tmhmm1	69.51(89.02)	90.23	95.45	85.32	83.02	83.34	87.11	86.85
toppred2	53.66(86.59)	83.59	95.75	83.46	74.48	85.81	90.43	82.02

Results

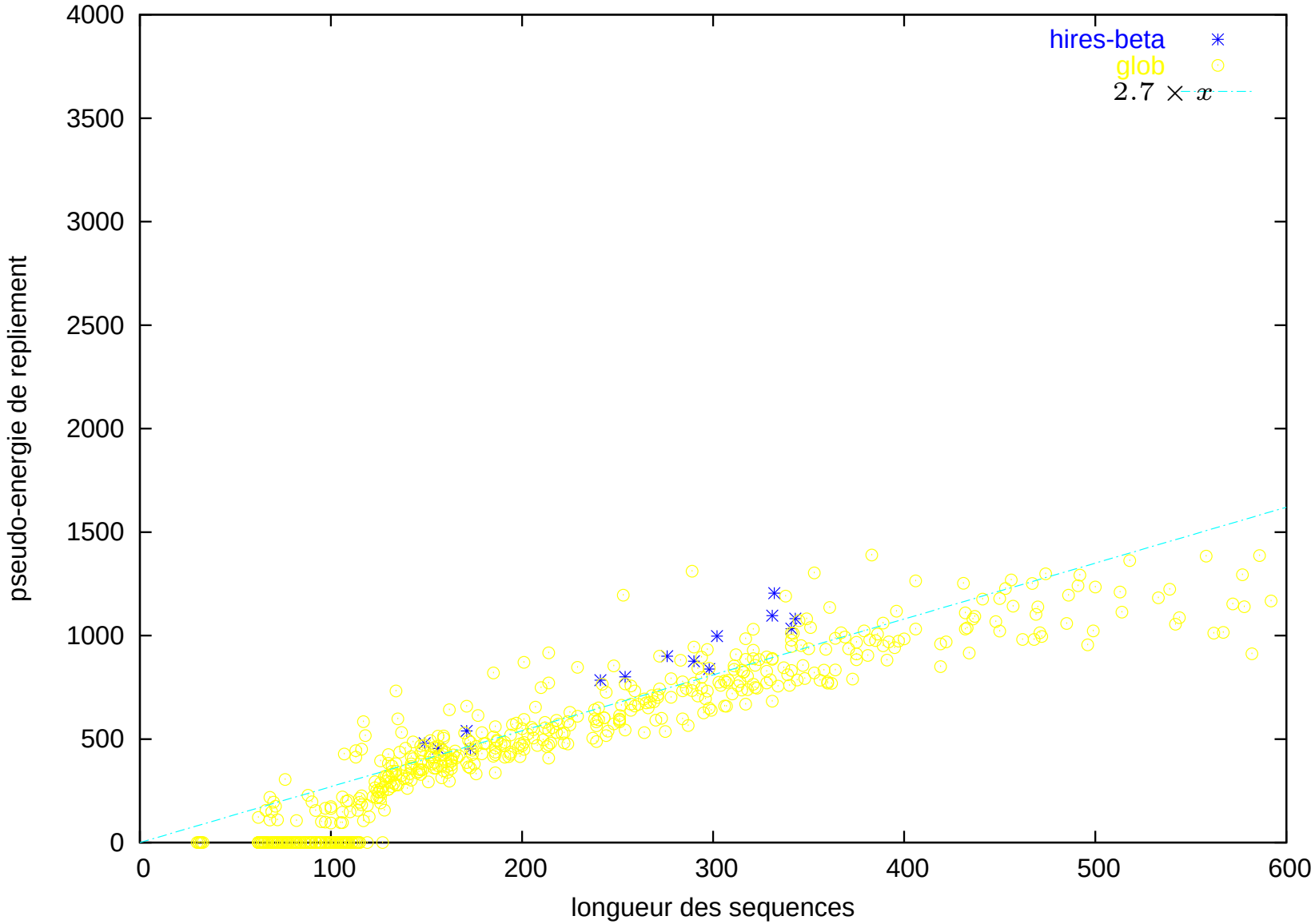
β-TM proteins known at high resolution level

Method	topology	strands		2-states	TM residus		non-TM residus	
	Q_{ok}	$Q_{stm}^{\%obs}$	$Q_{stm}^{\%pred}$	Q_2	$Q_{2T}^{\%obs}$	$Q_{2T}^{\%pred}$	$Q_{2N}^{\%obs}$	$Q_{2N}^{\%pred}$
total								
tmmtsag-basic	7.14(78.57)	83.14	66.82	64.87	72.71	69.99	53.07	56.37
tmmtsag-opt	21.43(92.86)	90.70	82.98	66.80	74.03	71.66	55.93	58.85
tmb-hmm	64.29(100.00)	97.09	97.09	84.04	85.75	87.45	81.47	79.15
small proteins								
tmmtsag-basic	50.00(75.00)	90.62	80.56	70.23	74.30	79.50	62.21	55.10
tmmtsag-opt	50.00(100.00)	93.75	93.75	76.12	78.97	84.08	70.51	62.96
tmb-hmm	75.00(100.00)	96.88	96.88	82.33	83.64	89.05	79.72	71.19

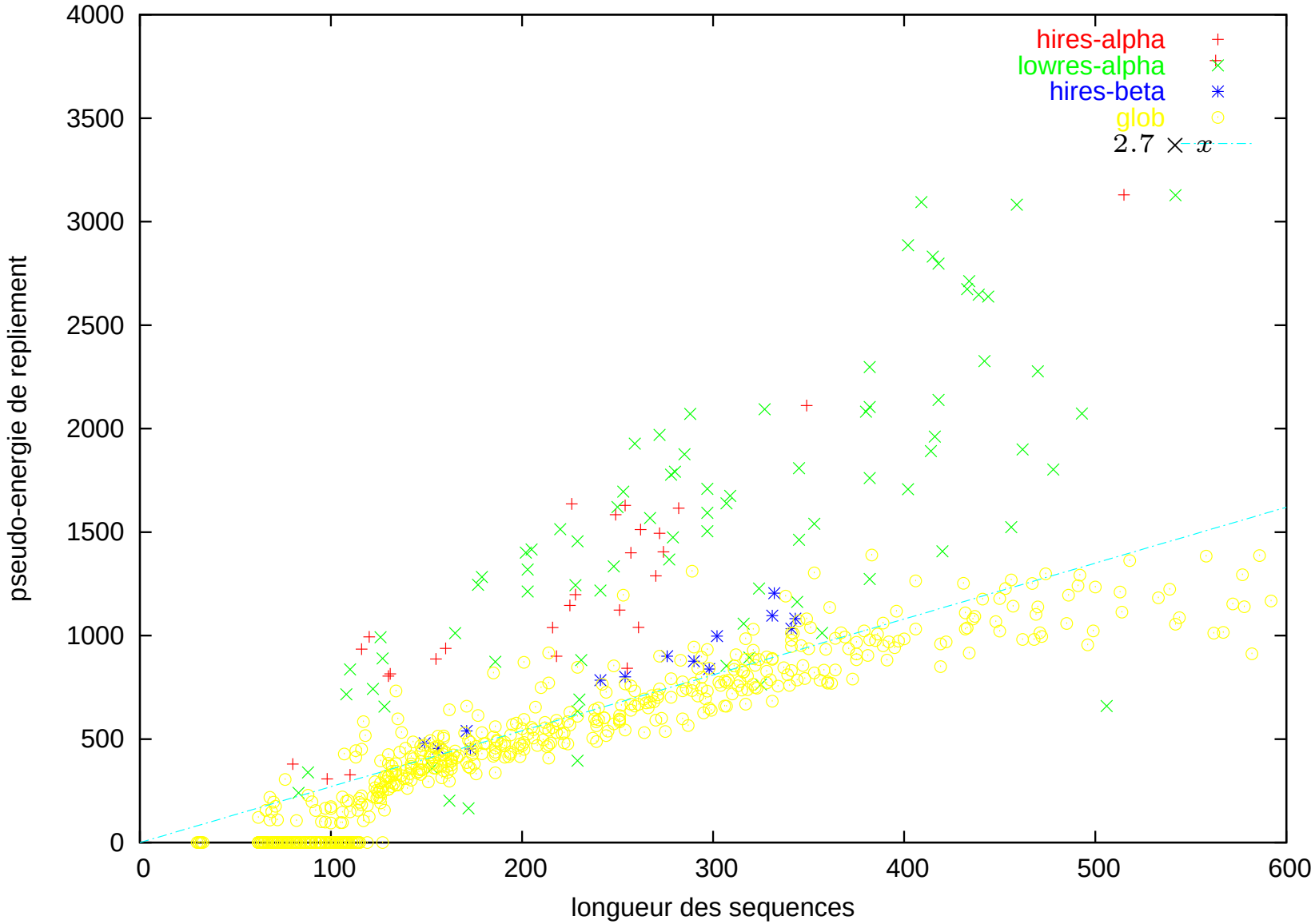
Results


hires-alpha
lowres-alpha
glob
2.7 × x
pseudo-energie de repliement
longueur des sequences
0
500
1000
1500
2000
2500
3000
3500
4000
0
100
200
300
400
500
600

Results


hires-beta
glob
2.7 × x
pseudo-energie de repliement
longueur des sequences
0
500
1000
1500
2000
2500
3000
3500
4000
0
100
200
300
400
500
600

Results


hires-alpha
lowres-alpha
hires-beta
glob
2.7 × x
pseudo-energie de repliement
longueur des sequences
0
500
1000
1500
2000
2500
3000
3500
4000
0
100
200
300
400
500
600

Results

Method	false negatives (%)	false positives (%)			
		$\Delta_{hires}^{tm,\alpha}$	$\Delta_{lowres}^{tm,\alpha}$	$\Delta_{hires}^{tm,\beta}$	$\Delta_{\alpha+\beta}^{tm}$
tmmtsag-basic	14.8	0	8.5	7.1	6.2
hmmtop2	6	0	1	-	-
phd-psihtm	2	3	8	-	-
pred-tmr	4	8	1	-	-
sosui	1	8	4	-	-
tmhmm1	1	8	4	-	-
toppred2	10	8	11	-	-
tmb-hmm	10	-	-	0	-

Conclusion

What has been done :

tmmtsag is the first software able to :

- use efficiently and without restrictions, long range interactions,
- unify α-channel and β-channel in the same model,
- discriminate the 3 categories α, β and globular.

Advantages :

- simple, versatile and efficient,
- no learning method used (more able to detect new structures, with lack of experimental data's).

Conclusion

What has to be done :

- refining the physical approximate model (structure and energy),
- integrating existing methods,
- modeling the quaternary structure,
- extension to globular proteins.

Conclusion

What we are doing :

- model refinement (joint work with T. Simonson),
- study of single point mutations in human rhodopsin (joint work with P. Clote),
- reconstructing the tertiary structure from contact predictions,
- screening of a complete genome,
- web interface : ASTRiD
  `http ://www.lix.polytechnique.fr/Labo/Jerome.Waldispuhl/astrid/`.

References

This work :

- J. Waldispühl and J.-M. Steyaert,
  Modeling and Predicting All-α Transmembrane Proteins Including Helix-helix Pairing, *to appear in Theoretical Computer Science, special issue on pattern discovery in the post genom*, 26 pages, 2005.
  available online : `http ://www.elsevier.com/locate/tcs/`

Multi-tape S-attributed grammars :

- F. Lefebvre,
  A Grammar-Based Unification of Several Alignment and Folding Algorithms, *Proceedings of the Fourth International Conference on Intelligent Systems for Molecular Biology*, pp 143-154, 1996.

Slides :

These slides have been realized with the **Youpla** LaTeXpackage, provided by E. Thome. Download it, at : `http ://www.loria.fr/~thome/`