



Hidden Markov Models

$$A = a_1 a_2 a_3 a_4 a_5$$

$$B = b_1 b_2 b_3 b_4 b_5$$

$$C = c_1 c_2 c_3 c_4 c_5 \quad P_1(c_1)P_2(c_2)P_3(c_3)P_4(c_4)P_5(c_5) = P_{\text{total}}$$

$$D = d_1 d_2 d_3 d_4 d_5$$

$$E = e_1 e_2 e_3 e_4 e_5$$

$$P_1 P_2 P_3 P_4 P_5$$


Bayesian Inference

☞ $M = \{ \text{set of events derived from a model} \}$

☞ $D = \{ \text{corpus of known data} \}$

Bayes theorem:

$$P(M | D)P(D) = P(M \cap D)$$

$$P(D | M)P(M) = P(D \cap M)$$

$$P(M|D) = \frac{P(D|M) P(M)}{P(D)}$$



$$-\log(P(M|D)) = -\log(P(D|M)) - \log(P(M)) + \log(P(D))$$

posterior

likelihood

Maximum a posteriori MAP Minimizar $F = -\log(P(D|M)) - \log(P(M))$

Maximum likelihood ML Minimizar $F = -\log(P(D|M))$

Bajo condiciones de contorno $\Psi_i(M) = 0$

$$\Gamma(M) = F(M) - \sum_i \lambda_i \Psi_i(M)$$



$$\Gamma(M) = F(M) - \sum_i \lambda_i \Psi_i(M)$$

$$M = (w_0, w_1, w_2, w_3, \dots, w_n)$$

$$0 = \vec{\nabla} \Gamma = \vec{\nabla} F - \sum_i \lambda_i \vec{\nabla} \Psi_i$$



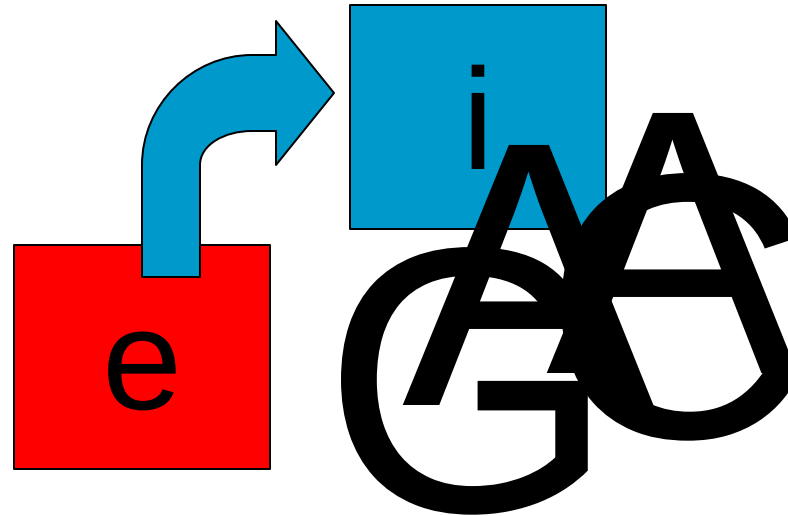
HMM

AGG
ACCG

A G C G T

$p_e(A)$ $p_e(G)$ ■■■■■

HMM

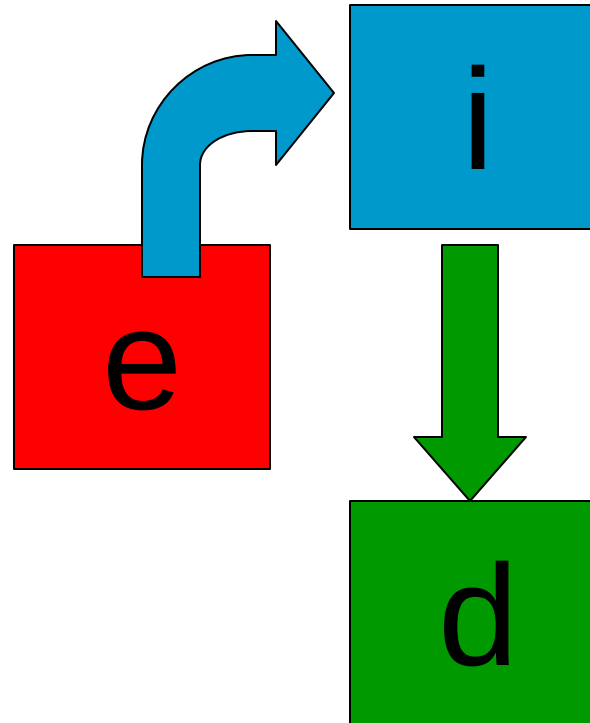


A G C G T A A C G

$p_e(A) p_e(G) \dots T_{ei} p_i(A) p_i(A) \dots$



HMM

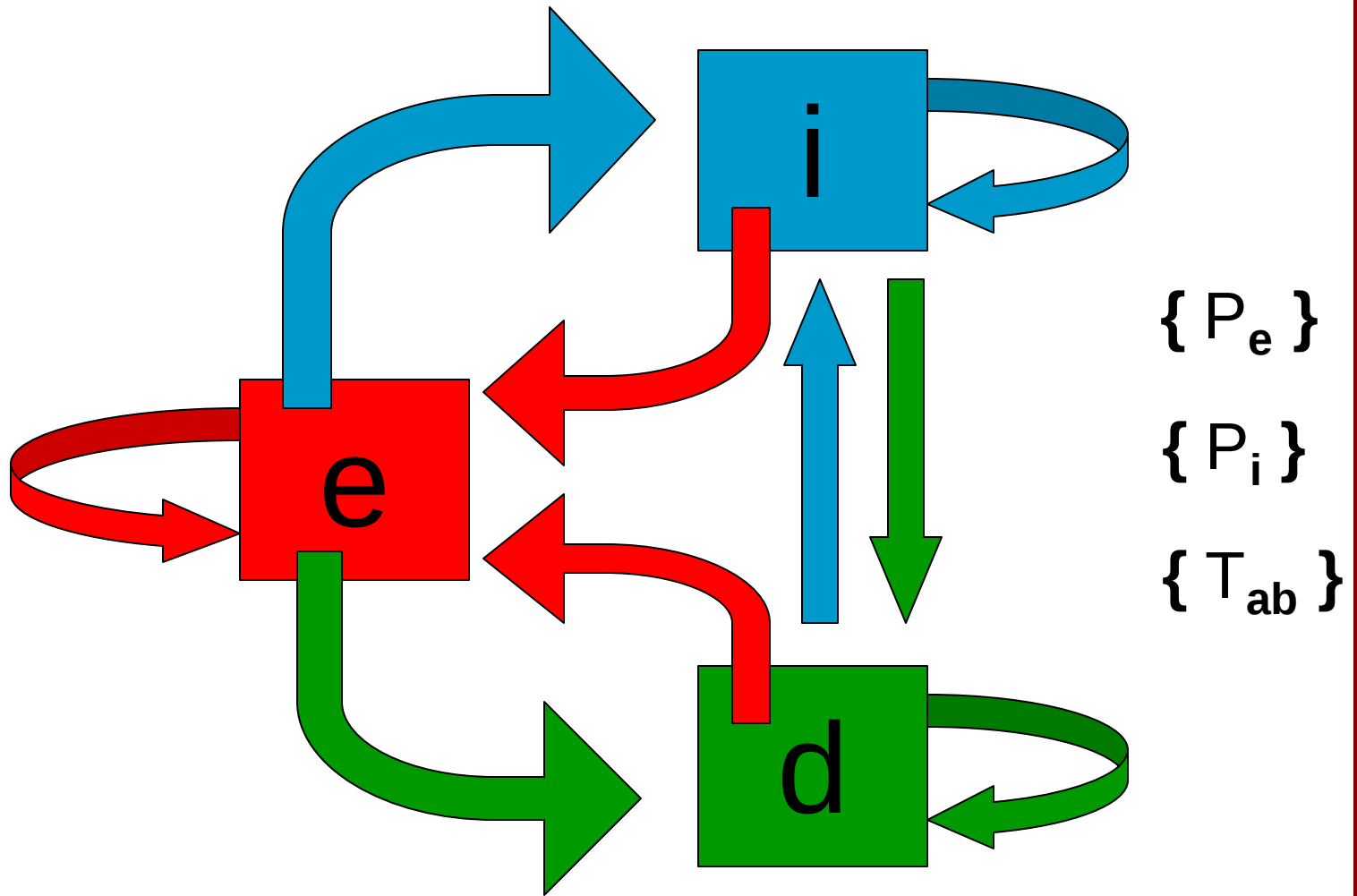


A G C G T

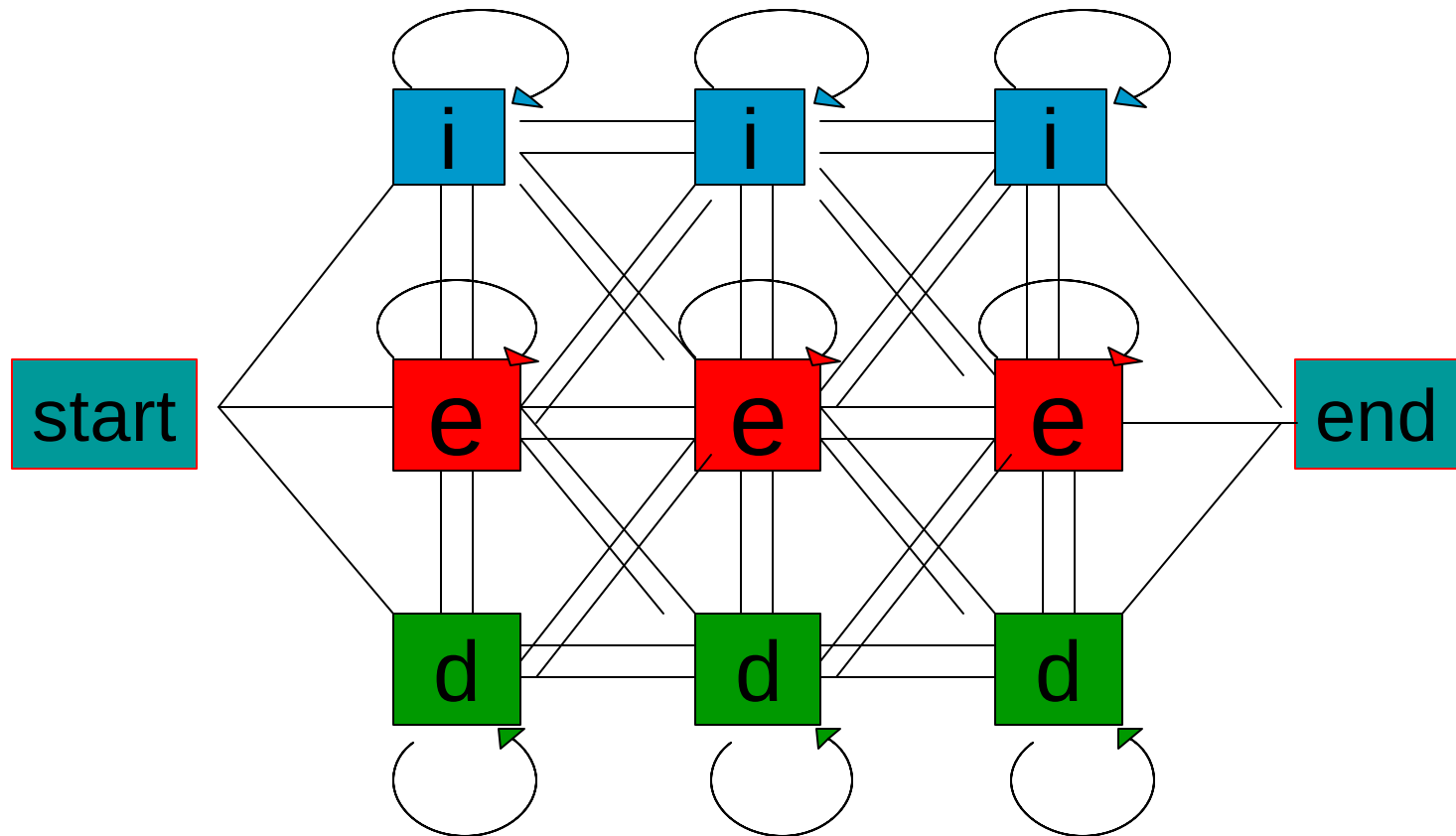
A A C G

- -

$p_e(A) p_e(G) \blacksquare \blacksquare \blacksquare \blacksquare T_{ei}$ $p_i(A) p_i(A) \blacksquare \blacksquare \blacksquare \blacksquare T_{id}$



Architecture HMM



$$P(\text{Sequence, path}|\text{model}) = \text{Product}$$



A posteriori

DATA

sequence Θ

path π

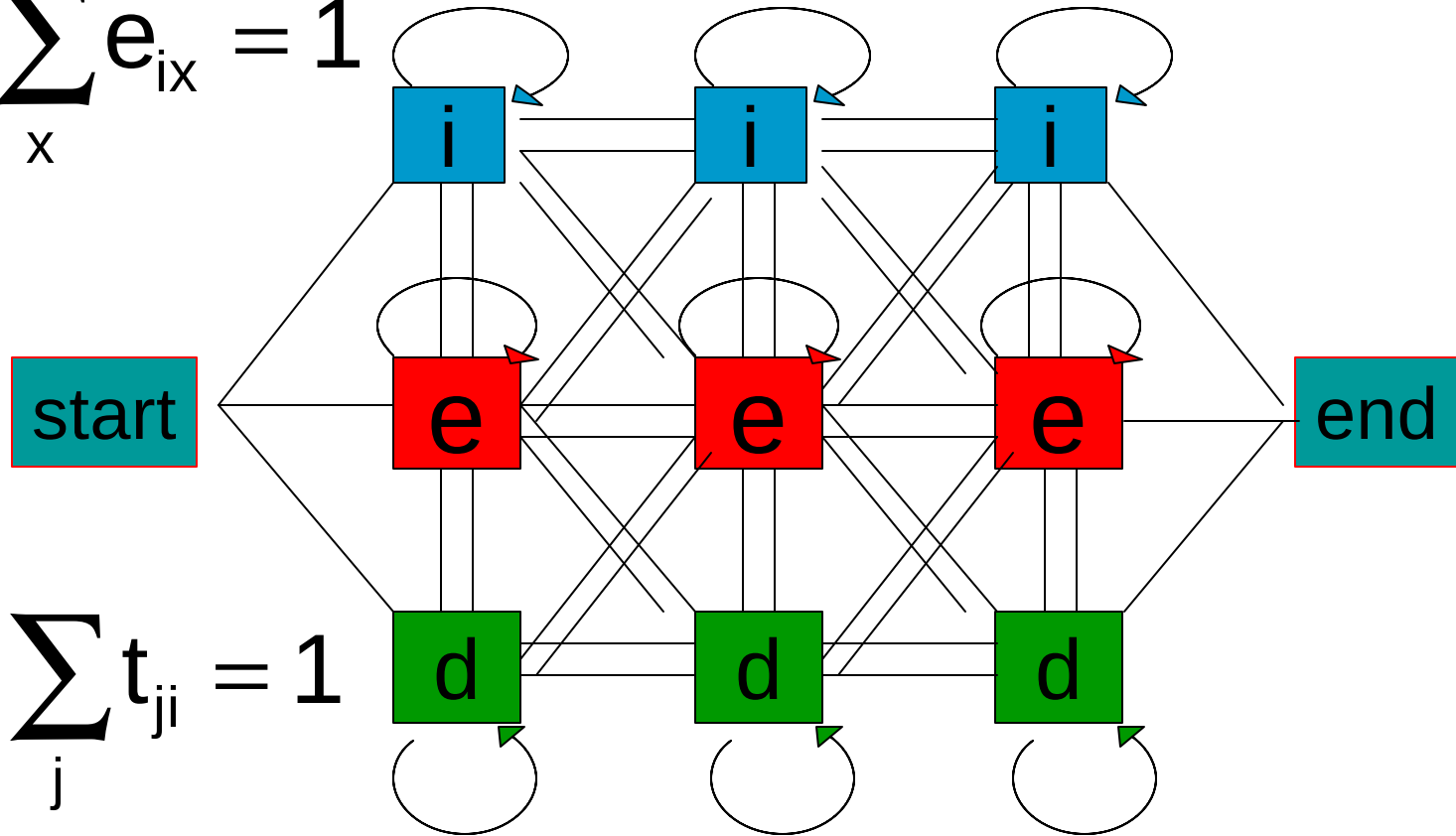
$\{ P_e; P_i \}$ state i

$$P(\Theta, \pi \mid M) = \prod_{j,i} t_{ji} \prod_{i|X^t} e_{i|X^t}$$

$$P(\Theta \mid M) = \sum_{\pi} P(\Theta, \pi \mid M)$$

Boundary conditions

$$\sum_x e_{ix} = 1$$



$$\sum_j t_{ji} = 1$$



$$\frac{\partial \log(P(\Theta | M))}{\partial e_{kX}} = \frac{1}{P(\Theta | M)} \frac{\partial P(\Theta | M)}{\partial e_{kX}}$$

$$\frac{\partial \log(P(\Theta | M))}{\partial e_{kX}} = \frac{1}{P(\Theta | M)} \sum_{\pi} \frac{\partial P(\Theta, \pi | M)}{\partial e_{kX}}$$

$$\frac{\partial P(\Theta, \pi | M)}{\partial e_{kX}} = n(k, X, \pi, \Theta) * e_{kX}^{n(k, X, \pi, \Theta) - 1} \prod_{ji} t_{ji} \prod_{i \neq k} e_{iX^t}$$

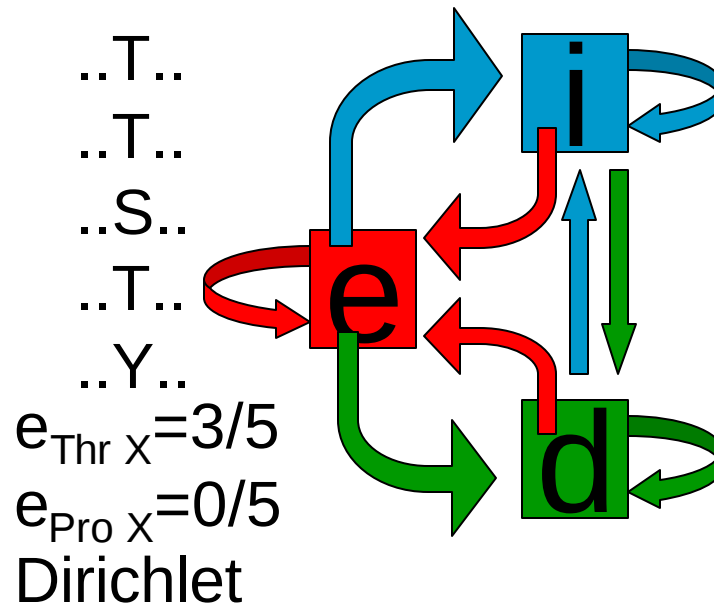
$$\frac{\partial P(\Theta, \pi | M)}{\partial e_{kX}} = n(k, X, \pi, \Theta) * \frac{P(\Theta, \pi | M)}{e_{kX}}$$

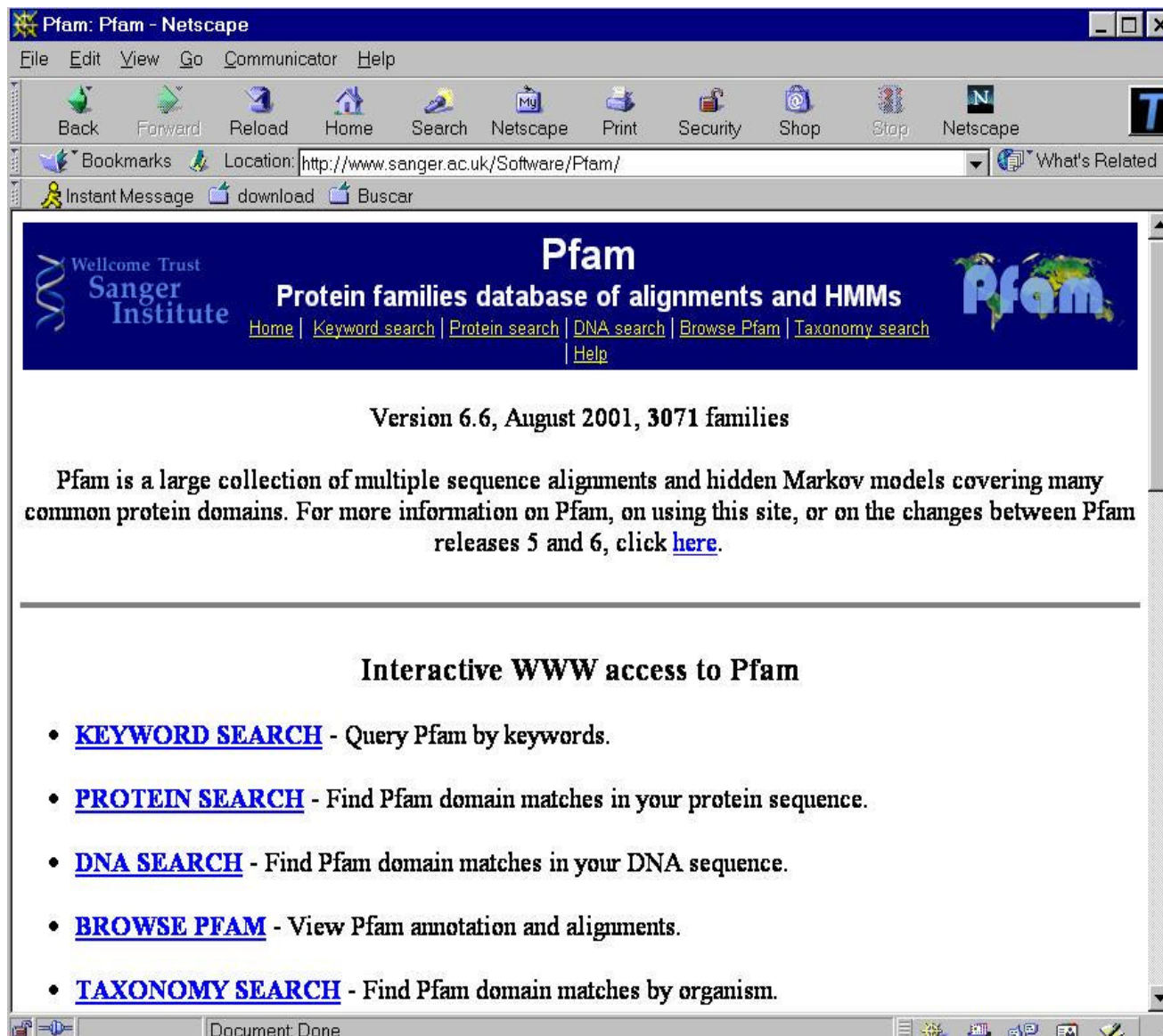
$$\lambda_k = \frac{1}{P(\Theta | M)} \sum_{\pi} \frac{\partial P(\Theta, \pi | M)}{\partial e_{kX}}$$

$$\lambda_k e_{kX} = \sum_{\pi} n(k, X, \pi, \Theta) * \frac{P(\Theta, \pi | M)}{P(\Theta | M)} = n_{kX}$$

$$e_{kX} = \frac{n_{kX}}{\sum_X n_{kX}}$$

$$t_{ji} = \frac{n_{ji}}{\sum_j n_{ji}}$$





Pfam: Pfam - Netscape

File Edit View Go Communicator Help

Back Forward Reload Home Search Netscape Print Security Shop Stop Netscape

Bookmarks Location: <http://www.sanger.ac.uk/Software/Pfam/> What's Related

Instant Message download Buscar

Pfam
Wellcome Trust Sanger Institute Protein families database of alignments and HMMs

[Home](#) | [Keyword search](#) | [Protein search](#) | [DNA search](#) | [Browse Pfam](#) | [Taxonomy search](#) | [Help](#)

Version 6.6, August 2001, 3071 families

Pfam is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains. For more information on Pfam, on using this site, or on the changes between Pfam releases 5 and 6, click [here](#).

Interactive WWW access to Pfam

- [KEYWORD SEARCH](#) - Query Pfam by keywords.
- [PROTEIN SEARCH](#) - Find Pfam domain matches in your protein sequence.
- [DNA SEARCH](#) - Find Pfam domain matches in your DNA sequence.
- [BROWSE PFAM](#) - View Pfam annotation and alignments.
- [TAXONOMY SEARCH](#) - Find Pfam domain matches by organism.

Document Done

INFORMATION



Figure 1: 1bec
Receptor
Beta chain of a t cell antigen
receptor

Accession number: PF00047

Immunoglobulin domain

Members of the immunoglobulin superfamily are found in hundreds of proteins of different functions. Examples include antibodies, the giant muscle kinase titin and receptor tyrosine kinases.

Immunoglobulin-like domains may be involved in protein-protein and protein-ligand interactions. The Pfam alignments do not include the first and last strand of the immunoglobulin-like domain.

NEW! This family forms **structural complexes** with other Pfam families, to view them click [here](#)