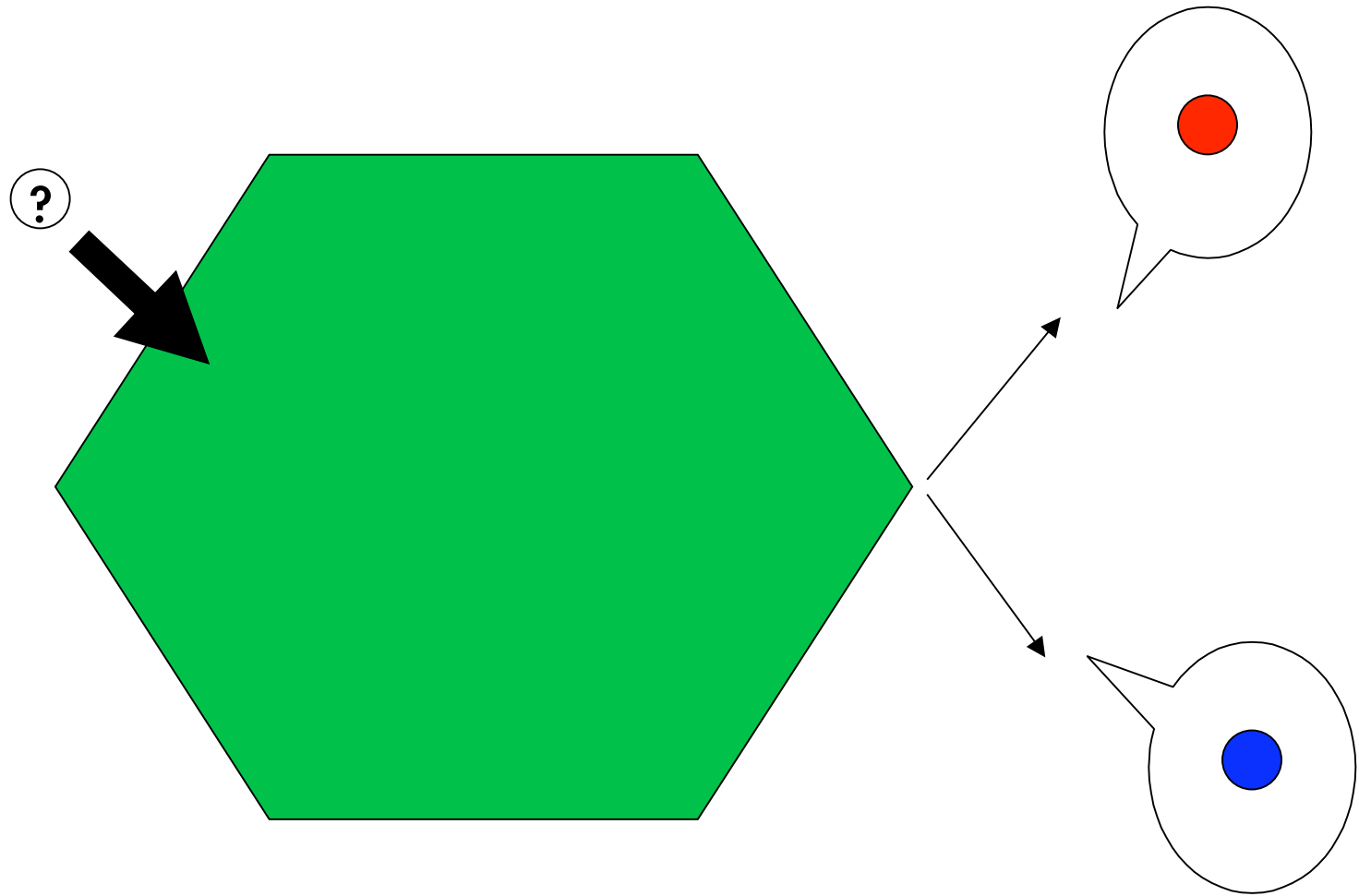


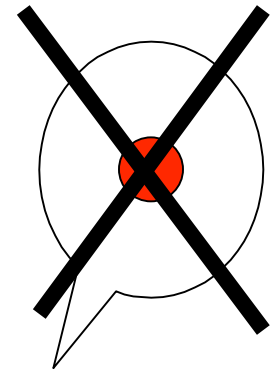
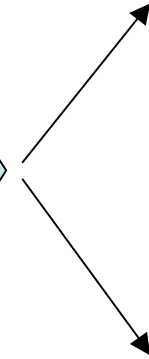
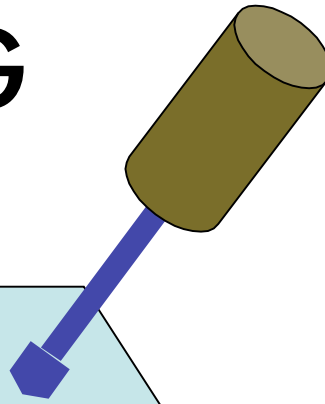
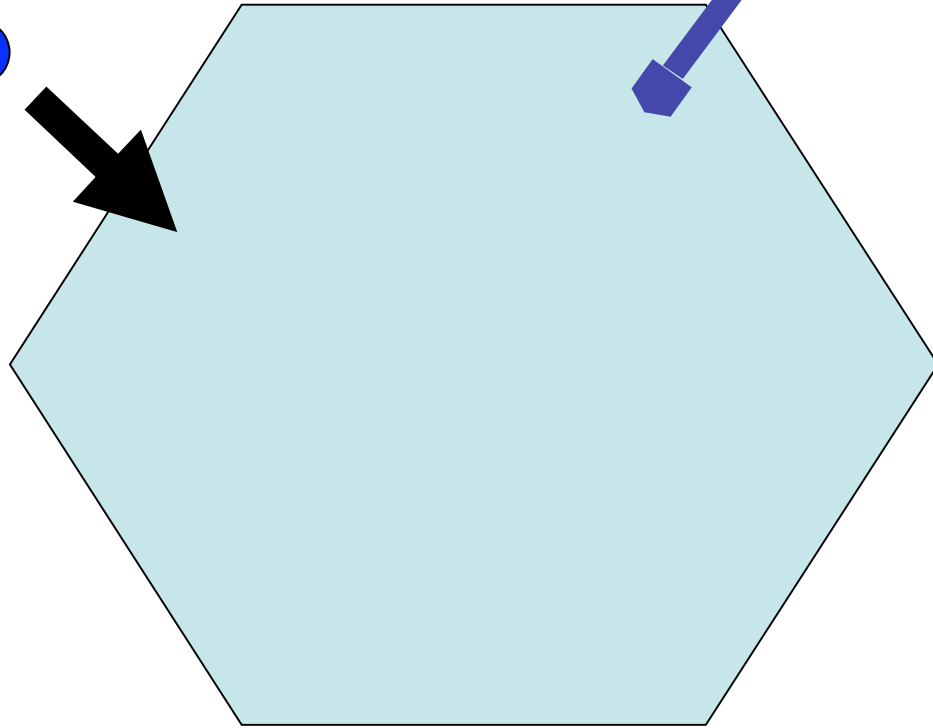
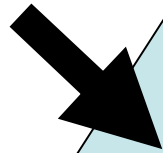
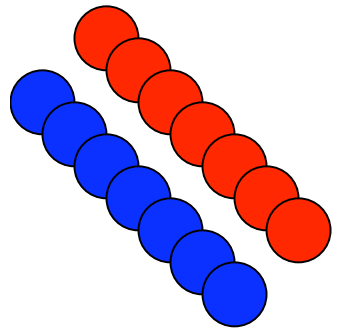
Prediction of the transmembrane regions of beta-barrel membrane proteins with a neural- network based predictor

I.Jacoboni, P.L. Martelli, P.Fariselli, V.De Pinto, R. Casadio

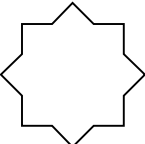
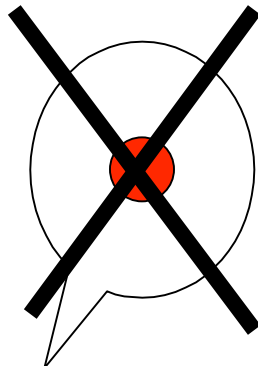
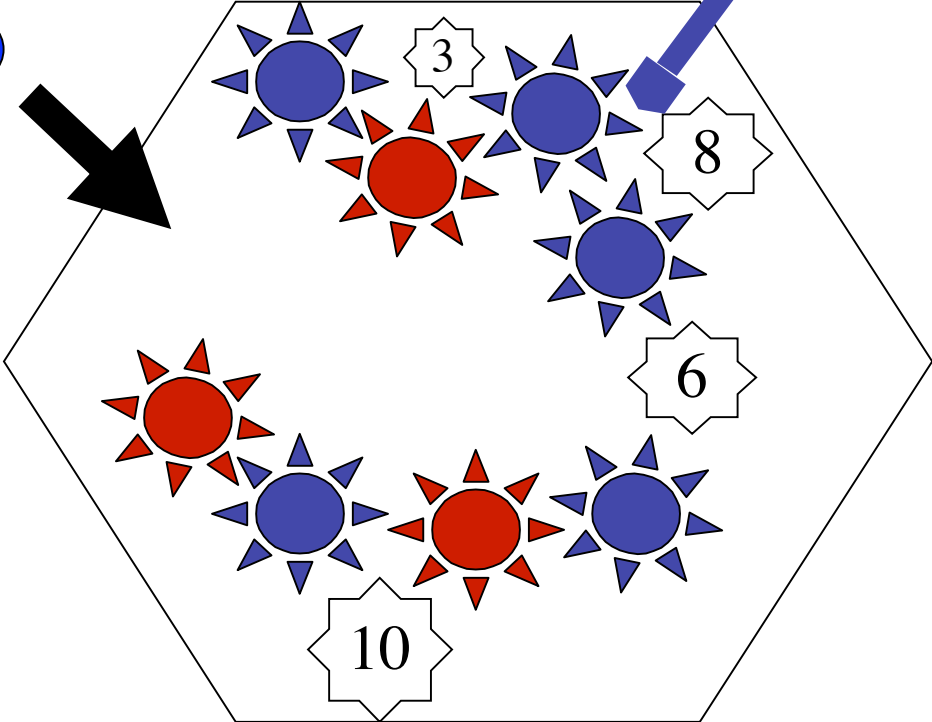
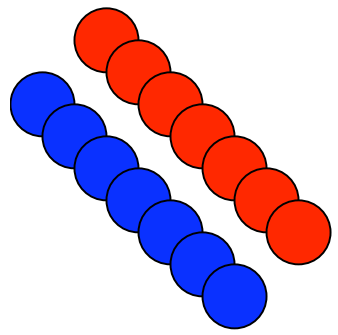
IL PREDITTORE



TRAINING

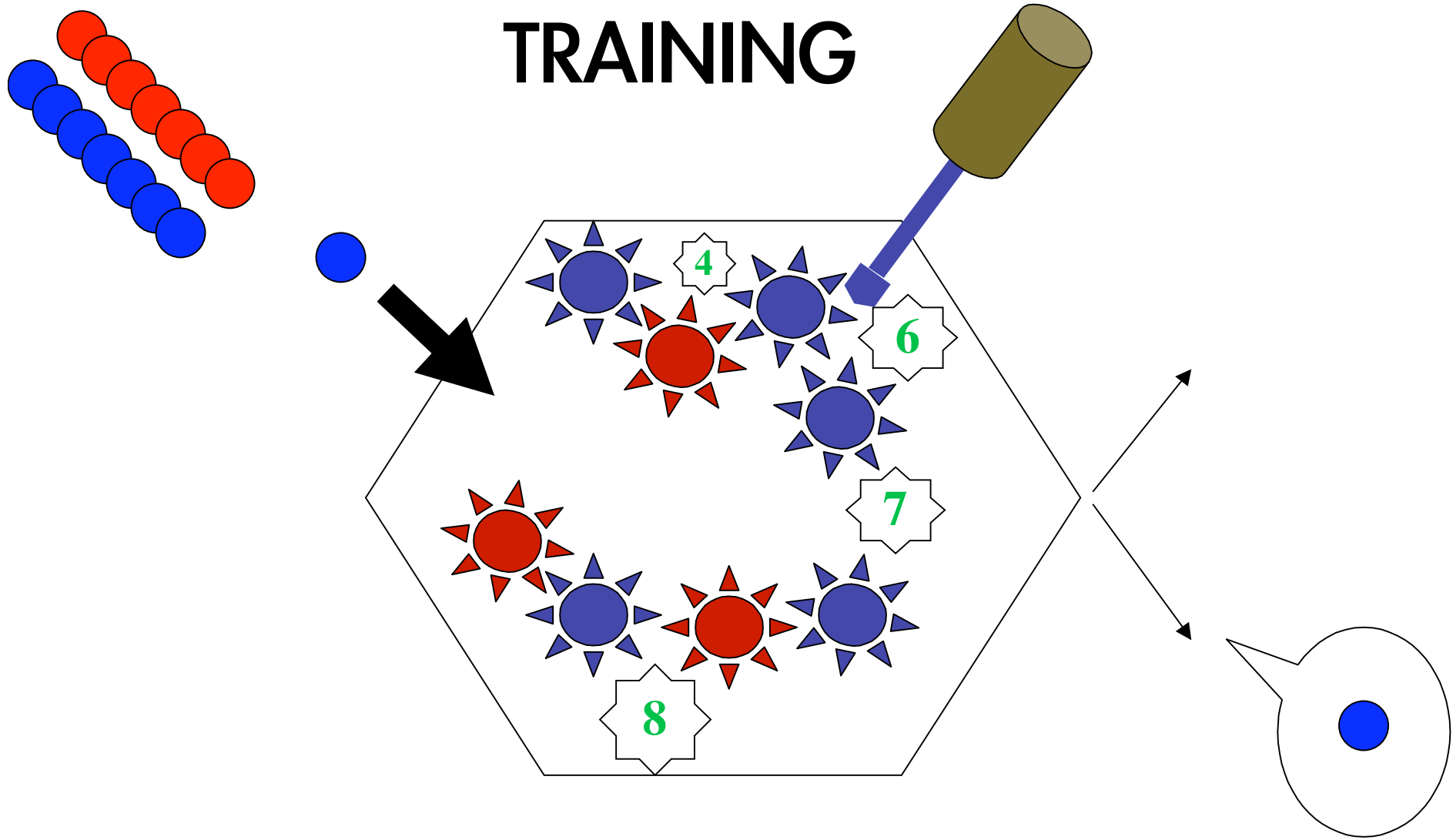


TRAINING



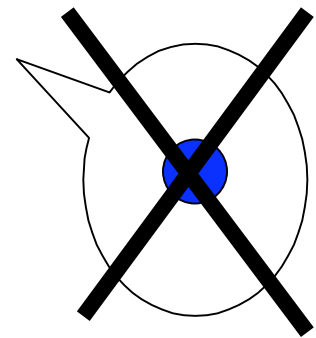
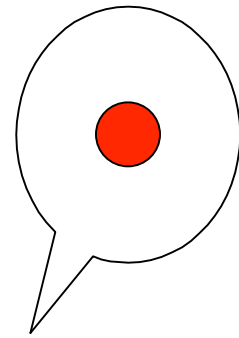
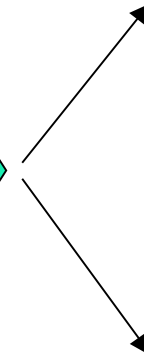
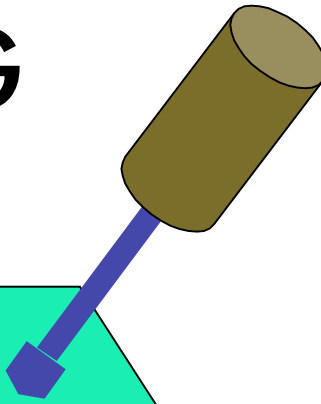
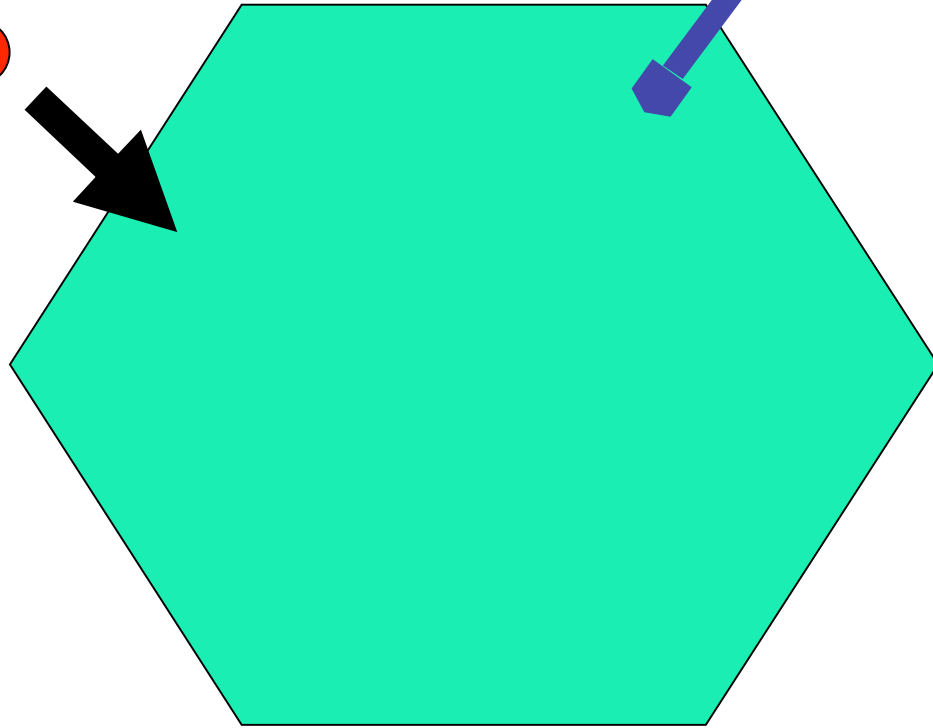
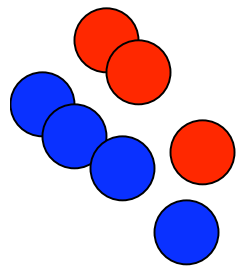
PARAMETRI

TRAINING

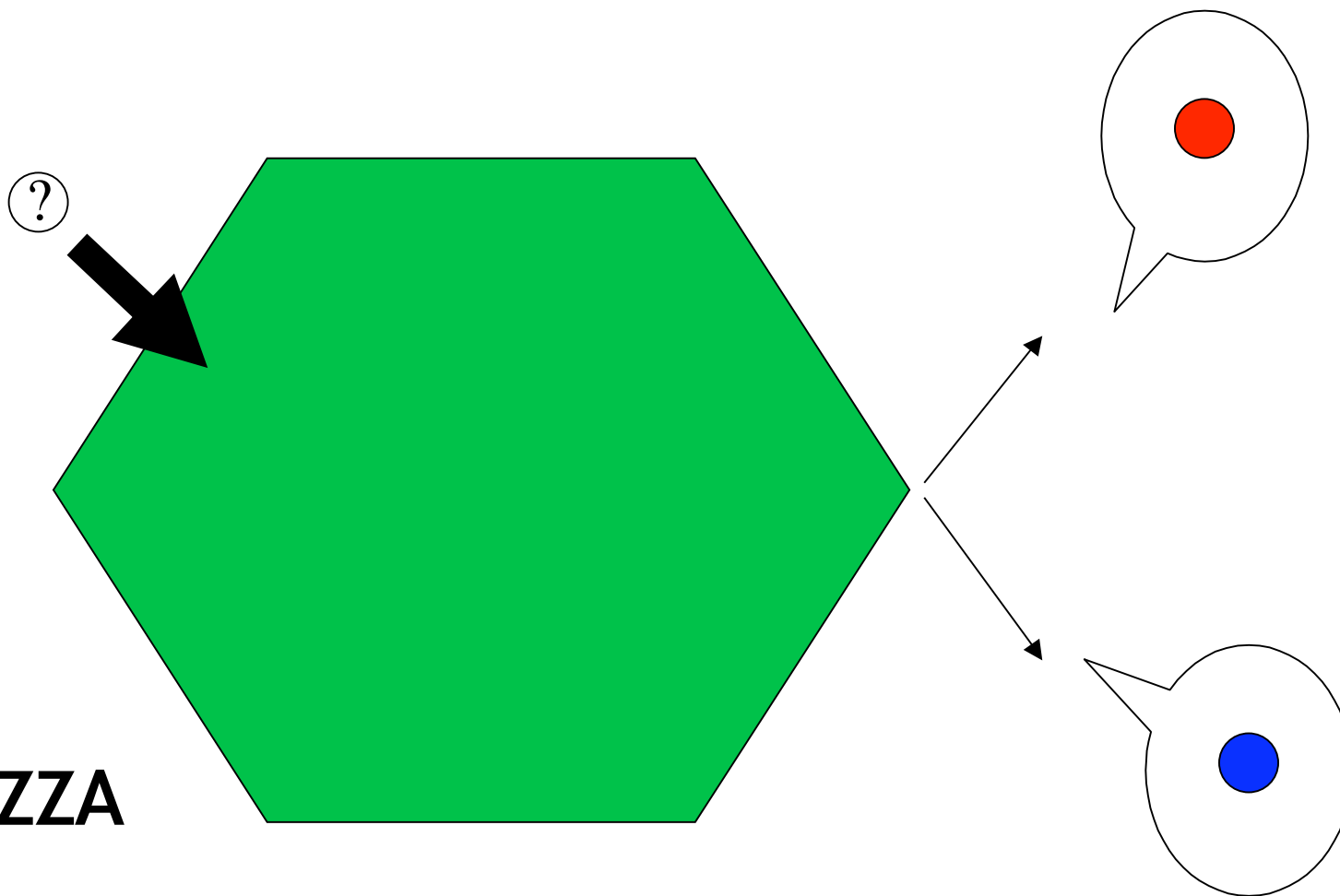


 PARAMETRI

TRAINING



IL PREDITTORE



ACCURATEZZA

SPElip: 96%

PSIPRED: 70-80%

BaCelLo: 70-75%

Machine Learning:

- L'Apprendimento automatico (noto in letteratura come Machine Learning) rappresenta una delle aree fondamentali dell'Intelligenza Artificiale e si occupa della realizzazione di sistemi che si basano su osservazioni o esempi come dati per la sintesi di nuova conoscenza (classificazioni, generalizzazioni, riformulazioni).
- L'Apprendimento supervisionato e' una tecnica di apprendimento automatico che mira a istruire un sistema informatico in modo da consentirgli di risolvere dei compiti in automatico

(da it.Wikipedia.org)

Machine Learning:

Reti neurali

PSIPRED: predittore di struttura secondaria

<http://bioinf.cs.ucl.ac.uk/psipred/>

Hidden Markov Models

CCHMM: predittore di regioni coiled-coil

<http://gpcr.biocomp.unibo.it/>

Support Vector Machines

BaCeLlo: predittore di localizzazione subcellulare

<http://gpcr.biocomp.unibo.it/bacello/>

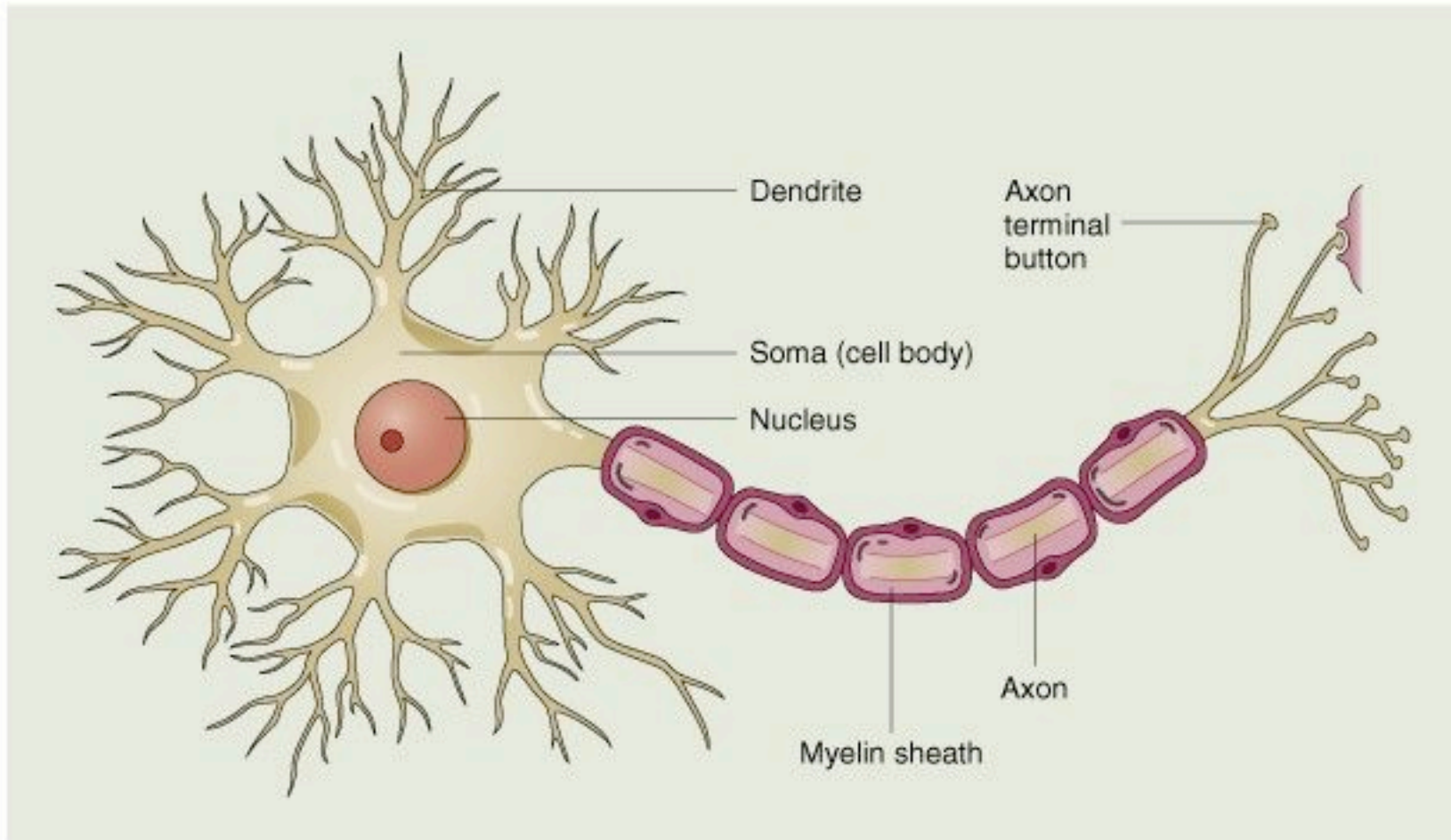
Reti neurali

Complesse organizzazioni di cellule nervose, con compiti di riconoscimento delle configurazioni assunte dall'ambiente esterno, memorizzazione e reazione agli stimoli provenienti dallo stesso.

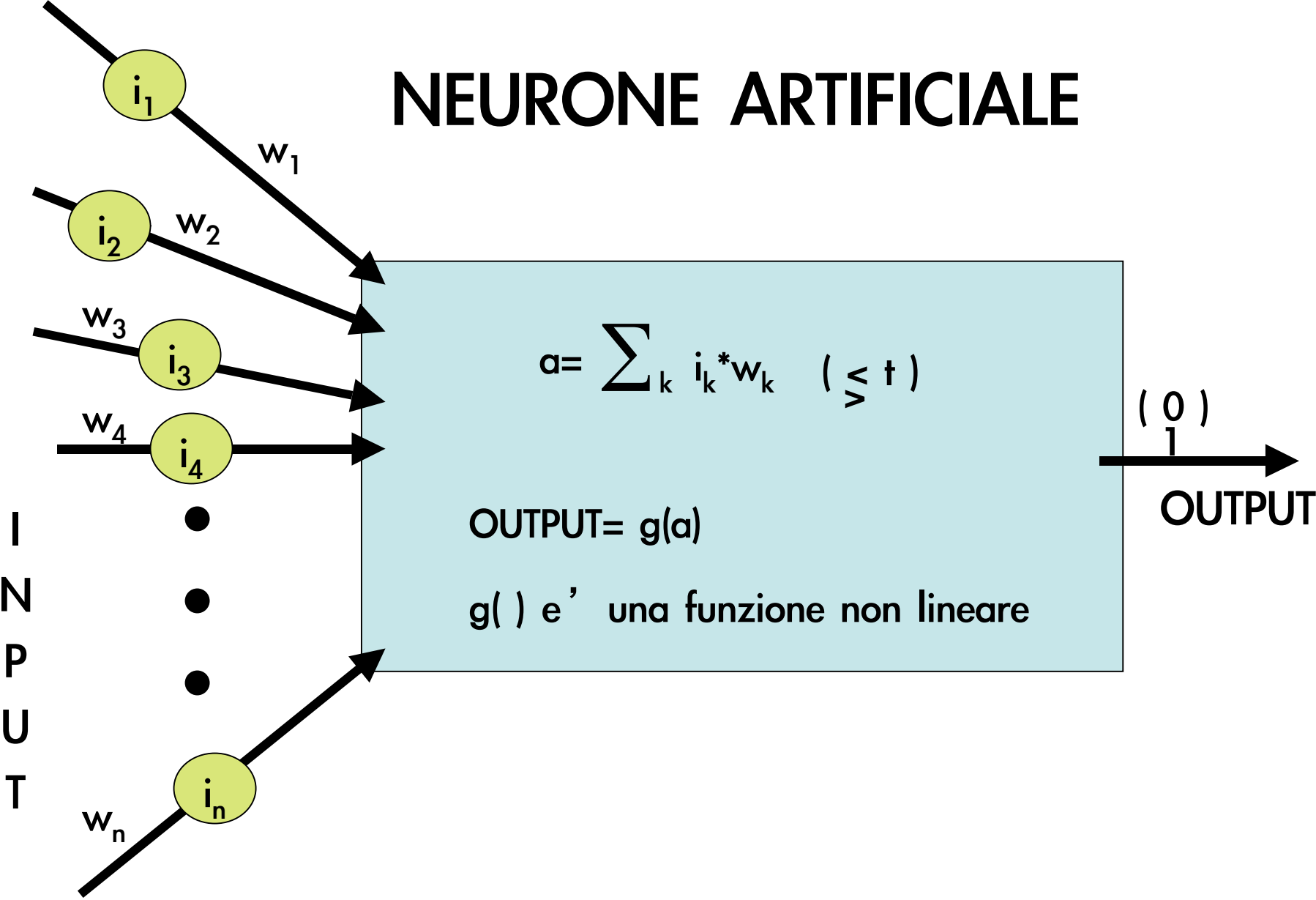
Il cervello umano rappresenta probabilmente il più mirabile frutto dell'evoluzione per le sue capacità di elaborare informazioni. Al fine di compiere tali operazioni, le reti biologiche si servono di un numero imponente di semplici elementi computazionali (neuroni) fittamente interconnessi in modo da variare la loro configurazione in risposta agli stimoli esterni: in questo senso può parlarsi di apprendimento.

(da [it.Wikipedia.org](http://it.wikipedia.org))

NEURONE

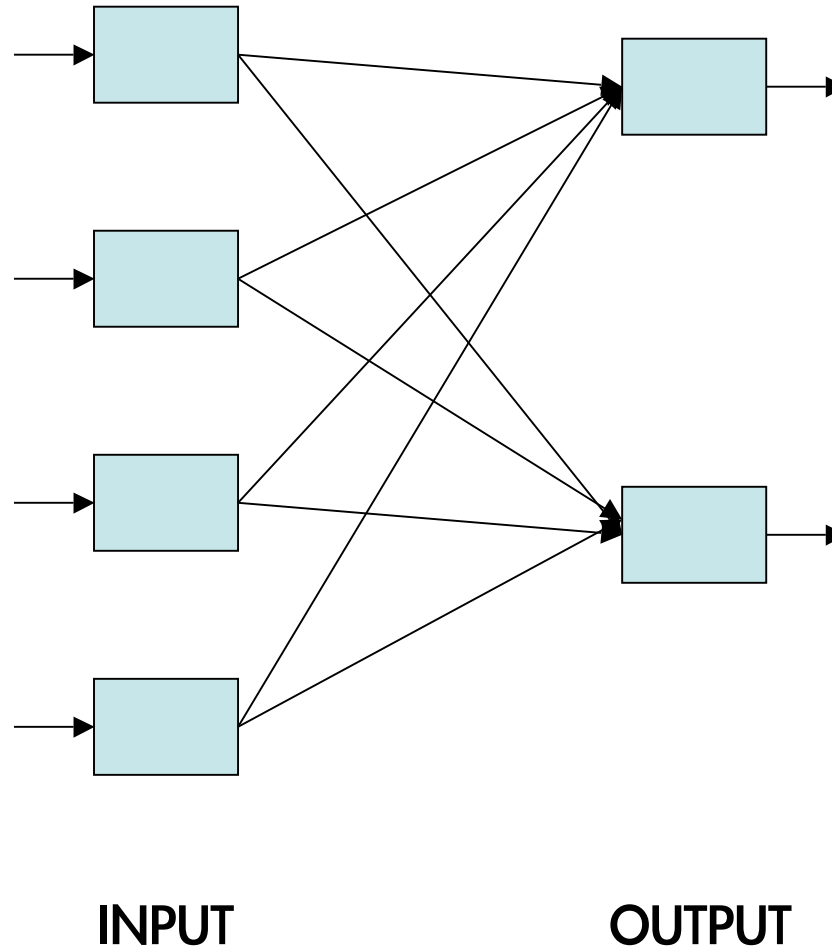


NEURONE ARTIFICIALE

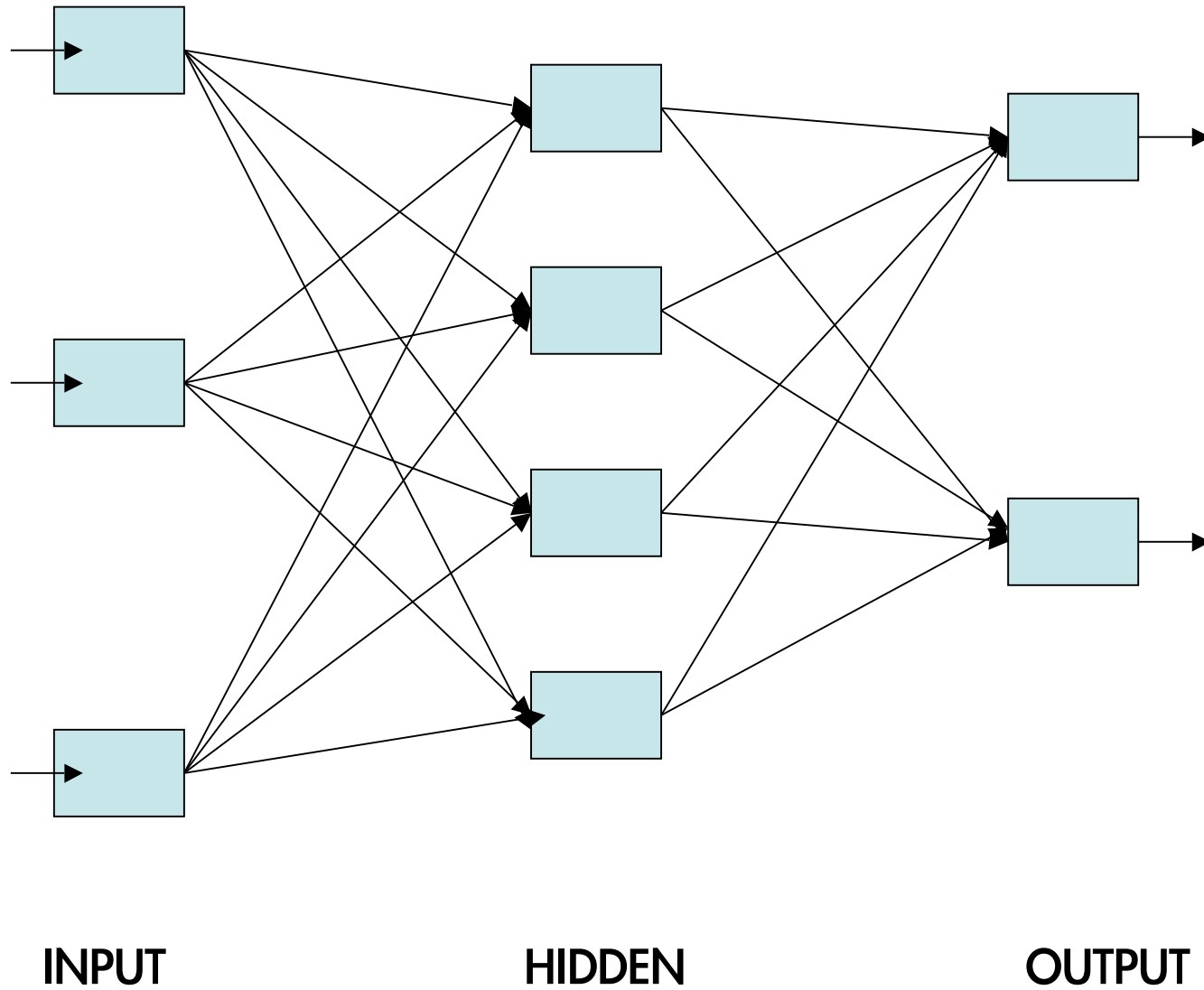


PERCEPTRON

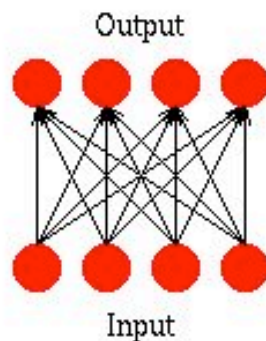
Rete feedforward con
un singolo strato di
pesi



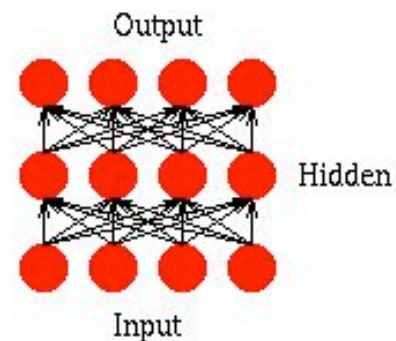
MULTI-LAYER PERCEPTRON



Feedforward networks

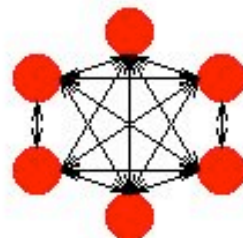


Single Layer Feedforward

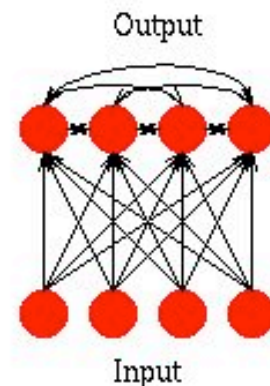


Multi Layer Feedforward

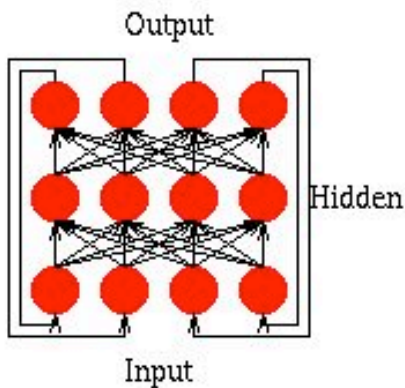
Recurrent networks



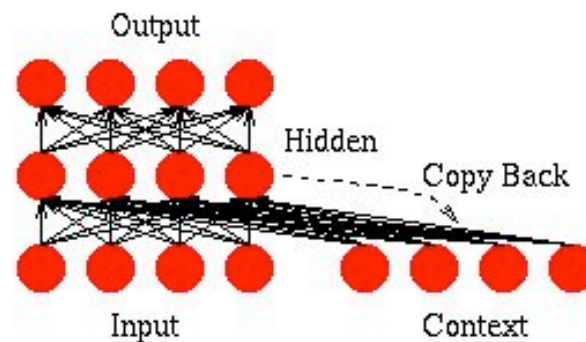
Fully Recurrent Network



Competitive Network



Jordan Network

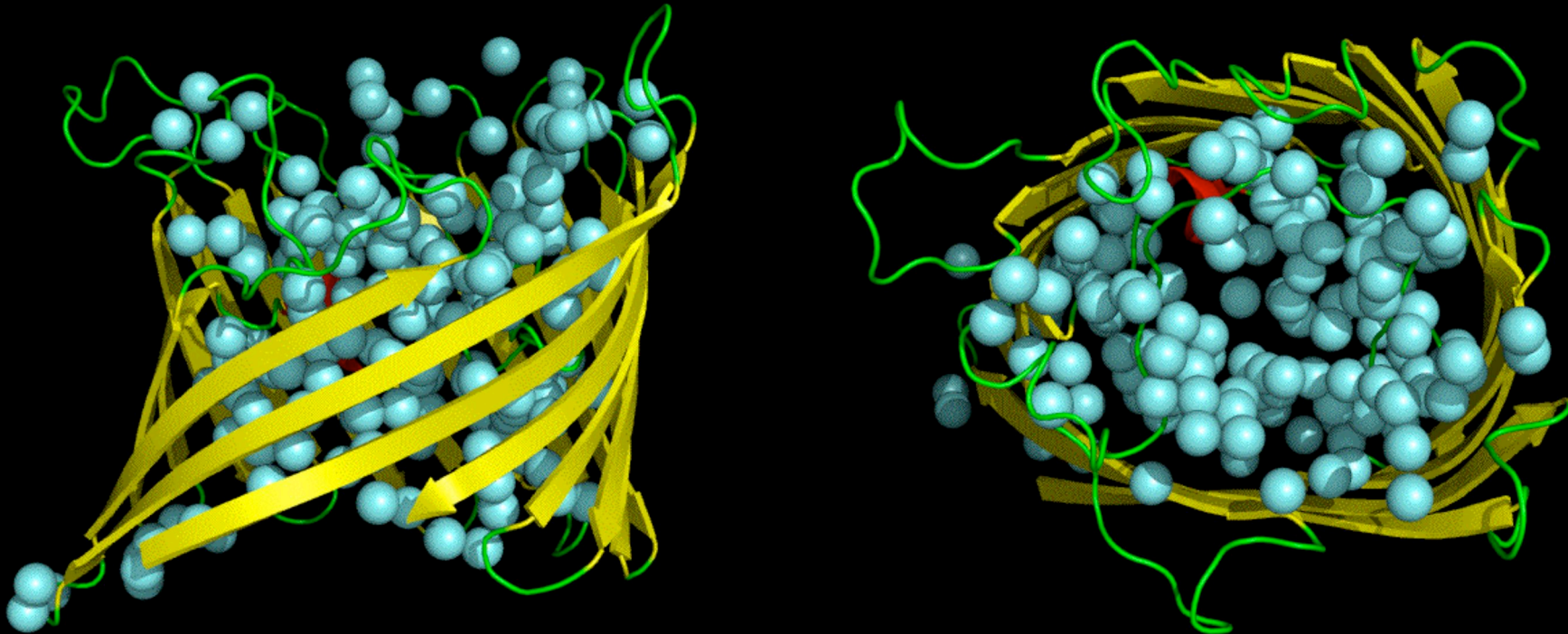


Simple Recurrent Network

La struttura a barile beta

La struttura a barile beta e' formata da un numero pari di filamenti beta che attraversano la membrana con un' inclinazione di circa 45° e si stabilizzano grazie a legami idrogeno con i filamenti adiacenti, formando un foglietto beta antiparallelo continuo a forma di cilindro.

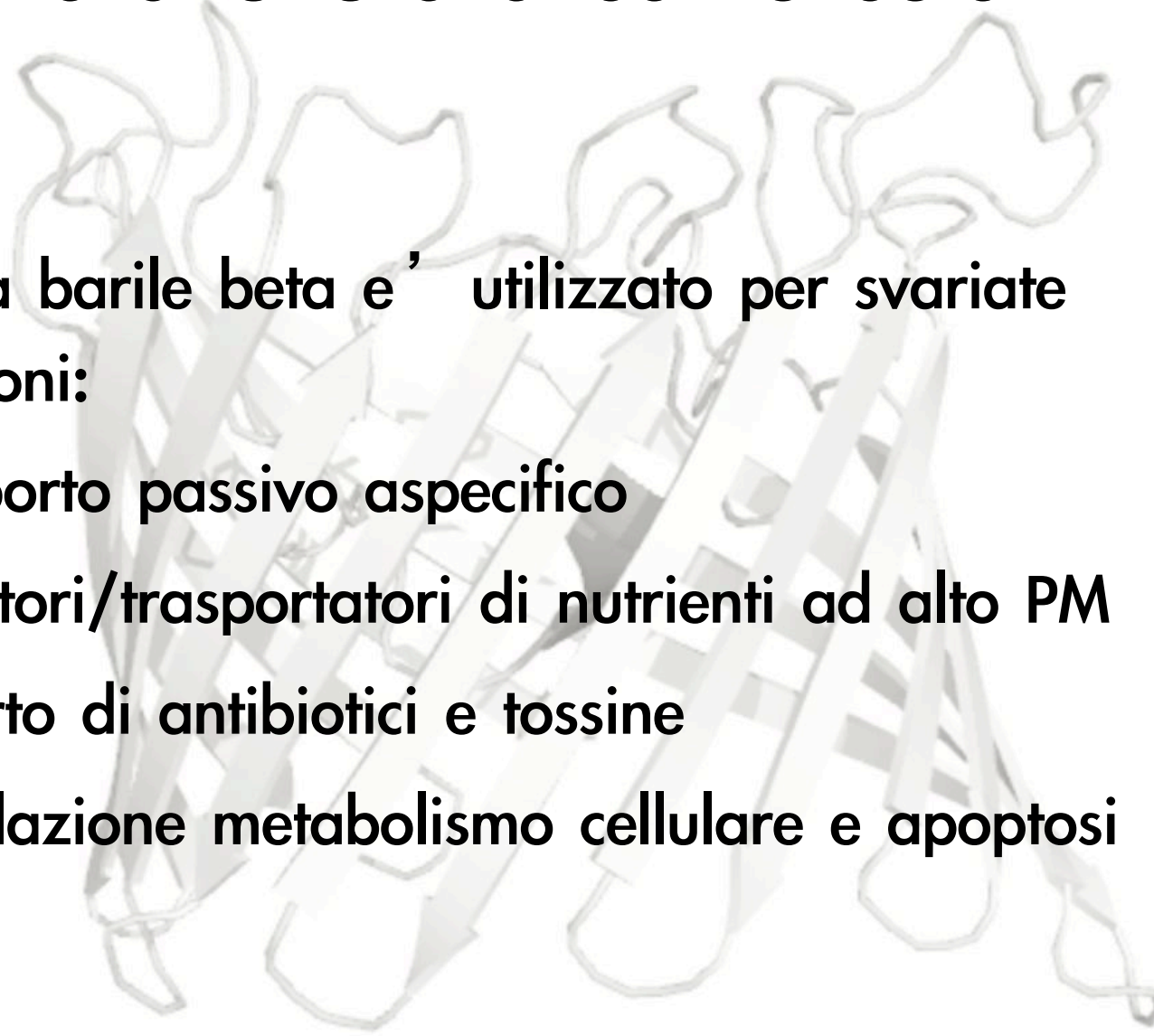
1PRN, porina aspecifica di Rhodospirillum rubrum (barile up-and-down):



La struttura a barile beta

Il fold a barile beta e' utilizzato per svariate funzioni:

- _ Trasporto passivo aspecifico**
- _ Recettori/trasportatori di nutrienti ad alto PM**
- _ Esporto di antibiotici e tossine**
- _ Regolazione metabolismo cellulare e apoptosi**
- _ ...**



Predittore di segmenti beta TM:

La predizione dei segmenti beta TM e' difficoltosa:

- _ contesto-dipendente
- _ scarsa identita' di sequenza fra proteine beta-barrel
- _ poche strutture cristallizzate, tutte batteriche

Perche' predire i segmenti TM?

THREADING

TARGET: IVDAC



SEQUENZA

VHSPQLVTDKHK6PDLLVYKQKNSGDRSMTSP6QK6TSTQTRDILLQGV6RSGPPKMTDLYVCTGSTRLE7NDE66PQPS6RVRVFDGSDKVELQVLFY6LQSGTSHATQPTVNFSDVSGH

PREDIZIONE
SEGMENTI
TRANSMEMBRANA

VHSPQLVTDKHK6PDLLVYKQKNSGDRSMTSP6QK6TSTQTRDILLQGV6RSGPPKMTDLYVCTGSTRLE7NDE66PQPS6RVRVFDGSDKVELQVLFY6LQSGTSHATQPTVNFSDVSGH

TEMPLATI: PORINE BATTERICHE

SEQUENZA

STRUTTURA

A green-bordered box containing the template information. It includes the title "TEMPLATI: PORINE BATTERICHE", the sequence "VHSPQLVTDKHK6PDLLVYKQKNSGDRSMTSP6QK6TSTQTRDILLQGV6RSGPPKMTDLYVCTGSTRLE7NDE66PQPS6RVRVFDGSDKVELQVLFY6LQSGTSHATQPTVNFSDVSGH", and a 3D ribbon structure of a bacterial porin template.

ALLINEAMENTO

VHSPQLVTDKHK6PDLLVYKQKNSGDRSMTSP6QK6TSTQTRDILLQGV6RSGPPKMTDLYVCTGSTRLE7NDE66PQPS6RVRVFDGSDKVELQVLFY6LQSGTSHATQPTVNFSDVSGH

VHSPQLVTDKHK6PDLLVYKQKNSGDRSMTSP6QK6TSTQTRDILLQGV6RSGPPKMTDLYVCTGSTRLE7NDE66PQPS6RVRVFDGSDKVELQVLFY6LQSGTSHATQPTVNFSDVSGH

A blue-bordered box containing the alignment step. It shows the target sequence and the template sequence aligned, with a red line indicating the alignment path.

MODELLING



STRUTTURA DEL TARGET

Costruzione di un predittore

- _ Architettura (NN, HMM, SVM)
- _ Caratteristiche da usare in input
- _ Dataset
- _ Metodo di training (backpropagation algorithm)
- _ Procedure di validazione

Predittore di segmenti beta transmembrana basato su reti neurali:

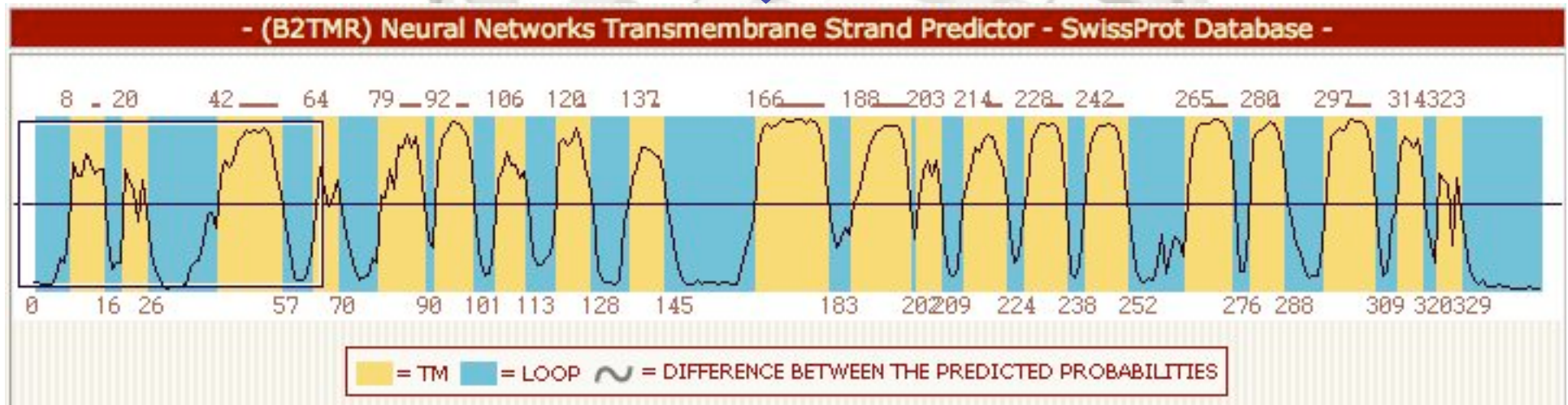
B2TMR

prende in input una sequenza proteica e predice la topografia e la topologia dei segmenti TM.

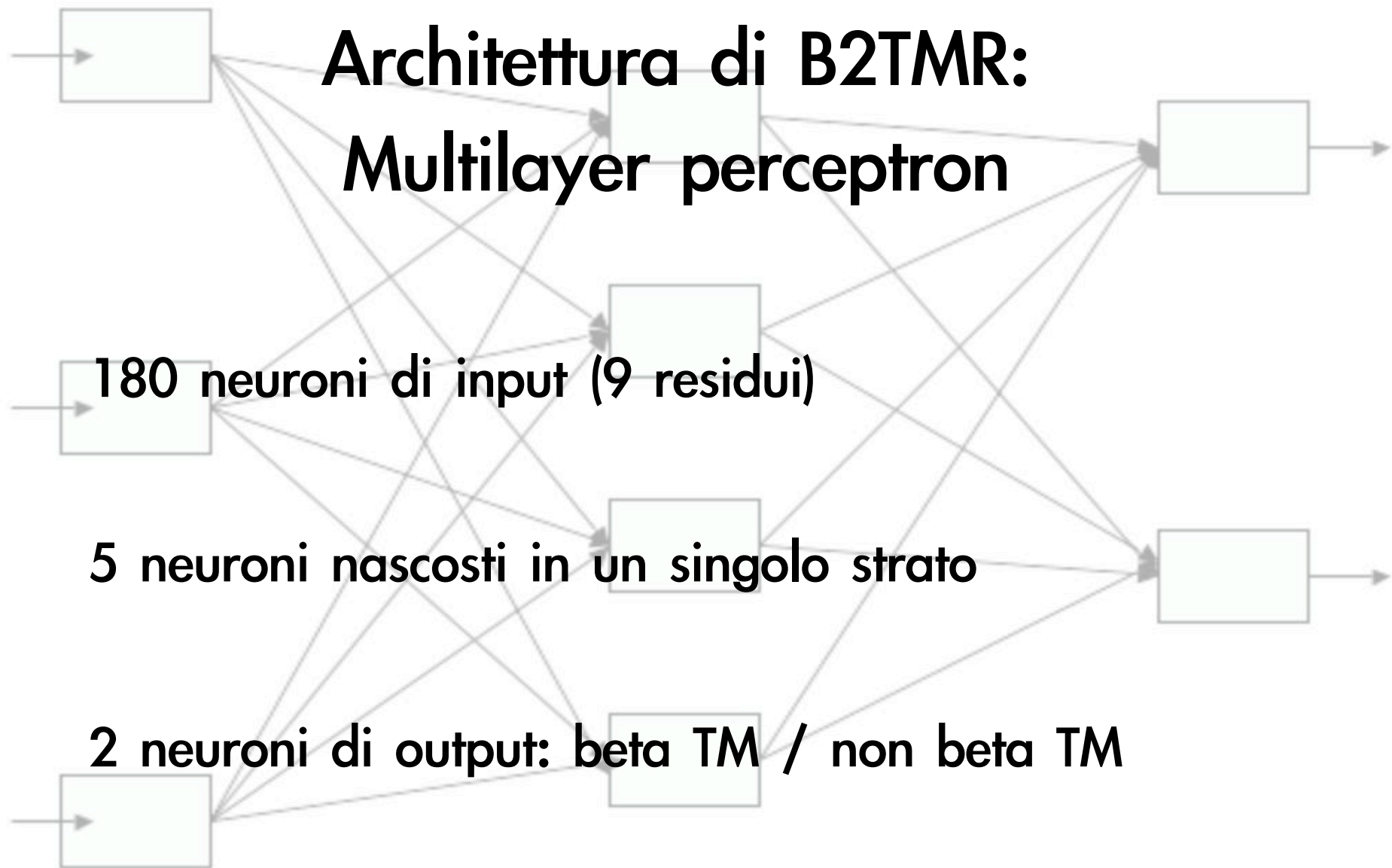
AVPPTYADLGKSARDVFTKGYGFLIKLDLTKSENGLEFTSSGSANTETTKVTGSLETKYRWTEYGLTFTEKWNTDNLGTEITVEDQLARGLKLTDFDSSFPNTGKKNAKIKTYKREHNLGCDMDFDIAGPSIRGALVLGYEGWLAGYQMFETAKSRVTSQSNFAVGYKTDEFQLHTNVNDGTEFGGSYQKVNKKLETAVNLAWTAG



<http://gpcr.biocomp.unibo.it/predictors/>



Architettura di B2TMR: Multilayer perceptron

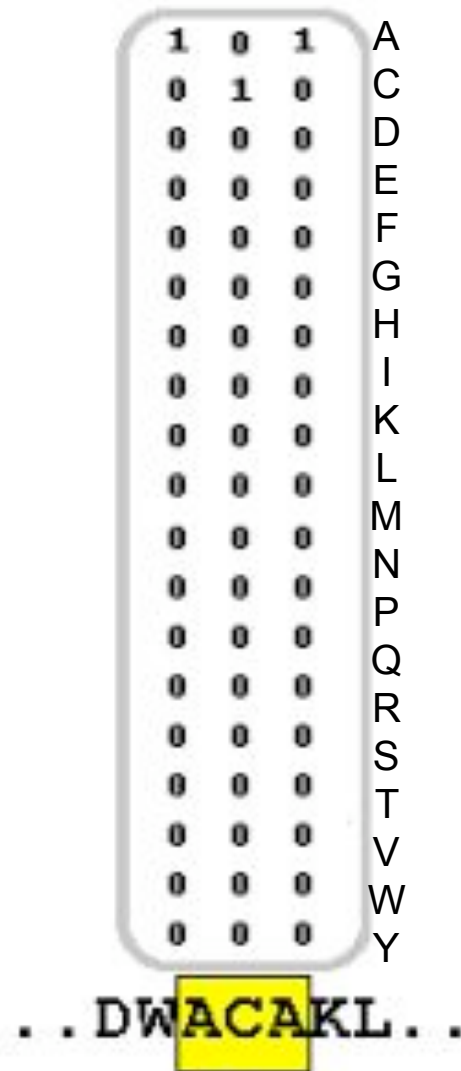


INPUT

HIDDEN

OUTPUT

INPUT



L' input e' una finestra di 9 residui che scorre la sequenza proteica.

Ogni residuo e' codificato in input da 20 neuroni corrispondenti ai 20 aminoacidi; uno solo presenta il valore 1, gli altri il valore 0.

I neuroni input di b2tmr sono dunque 180.

DATABASE

L' addestramento richiede una database di
“esempi” per la caratteristica da predire.

Deve essere:

- _ Ben annotato
- _ Rappresentativo
- _ Non ridondante

DATABASE - B2TMR

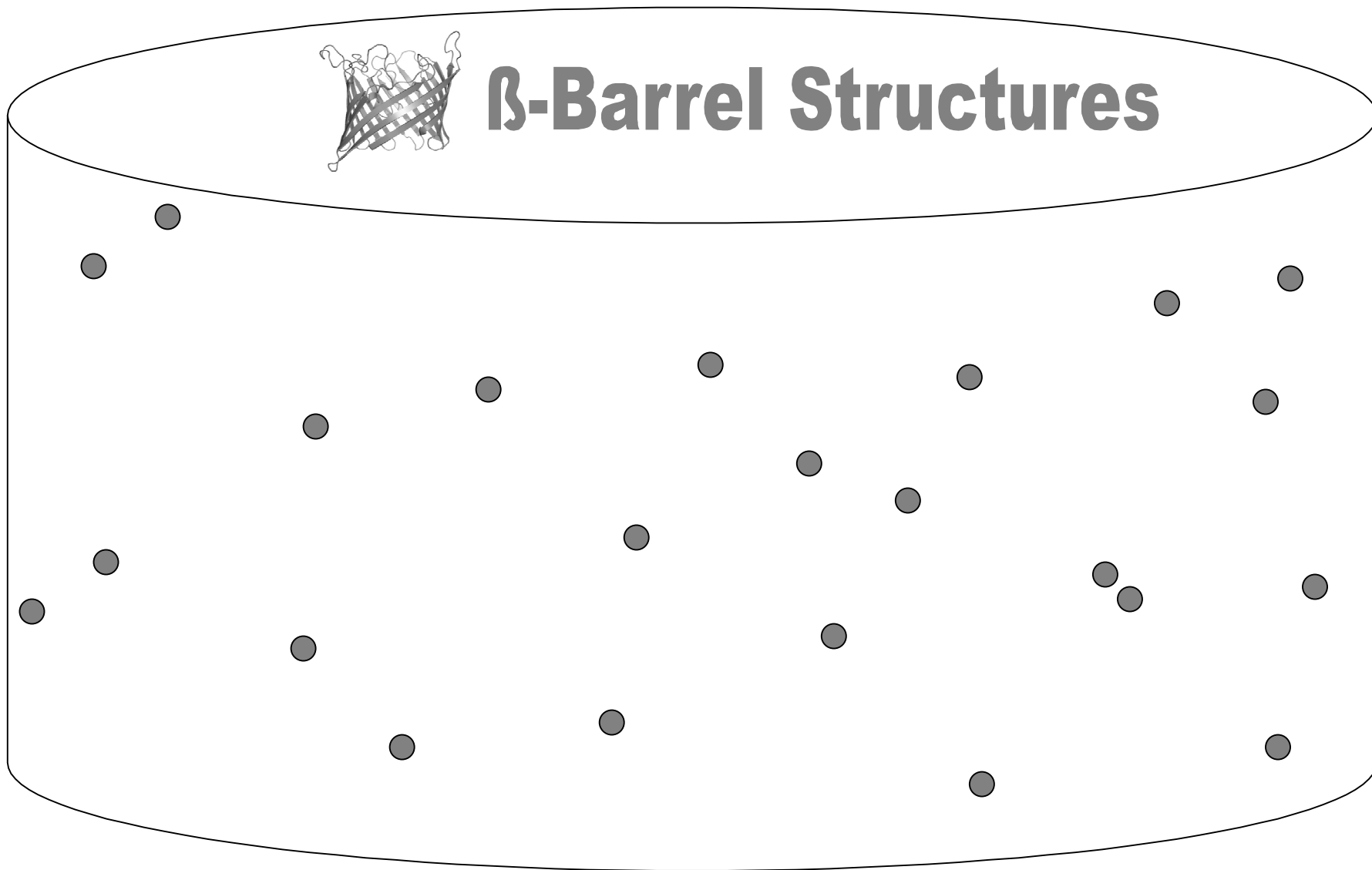
PDB

A 3D black cylinder representing a database. The top surface is a dark gray circle. The text 'PDB' is written in a bold, yellow and gold 3D font on the top surface. The cylinder has a slight shadow on the bottom edge.

DATABASE - B2TMR

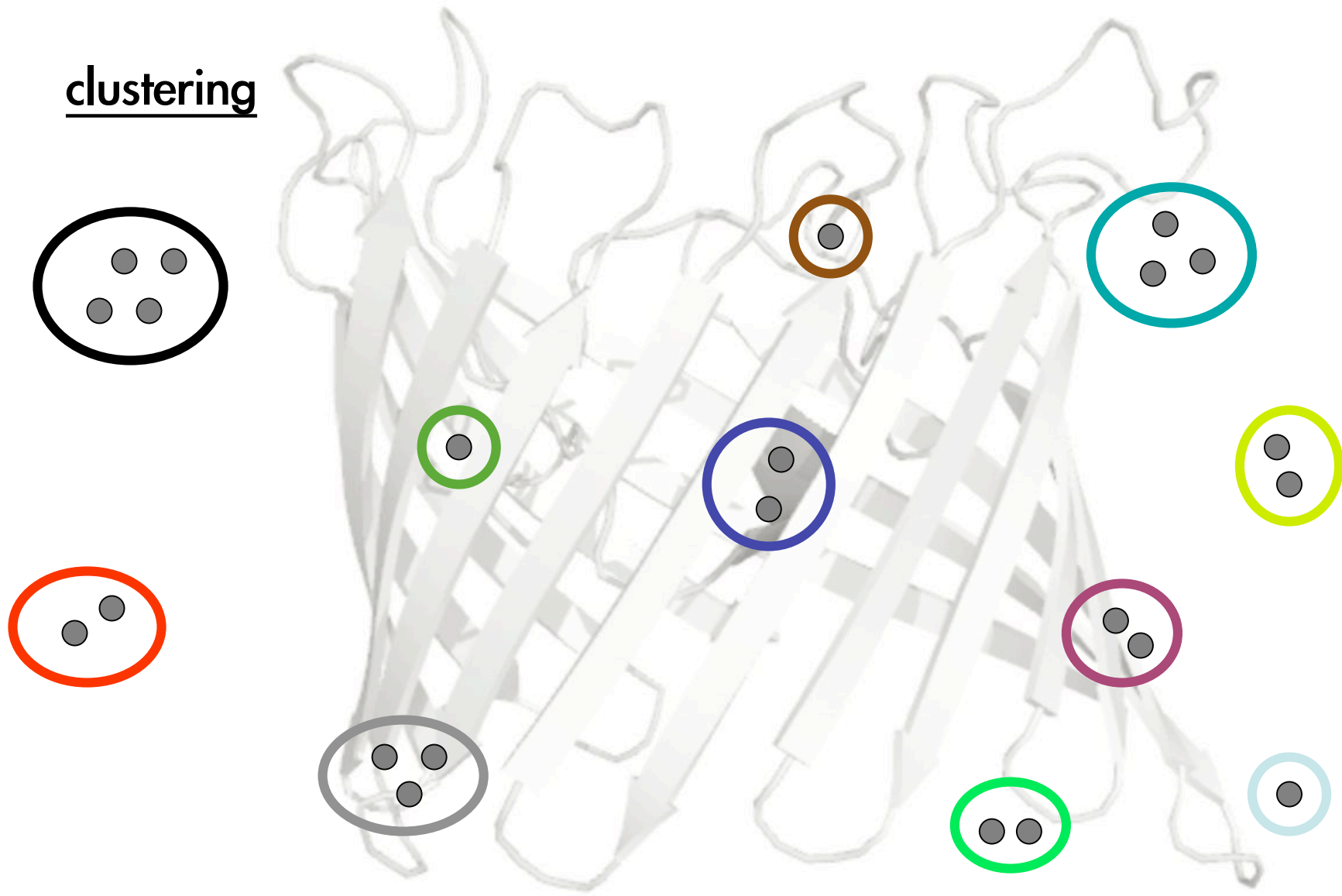


β -Barrel Structures



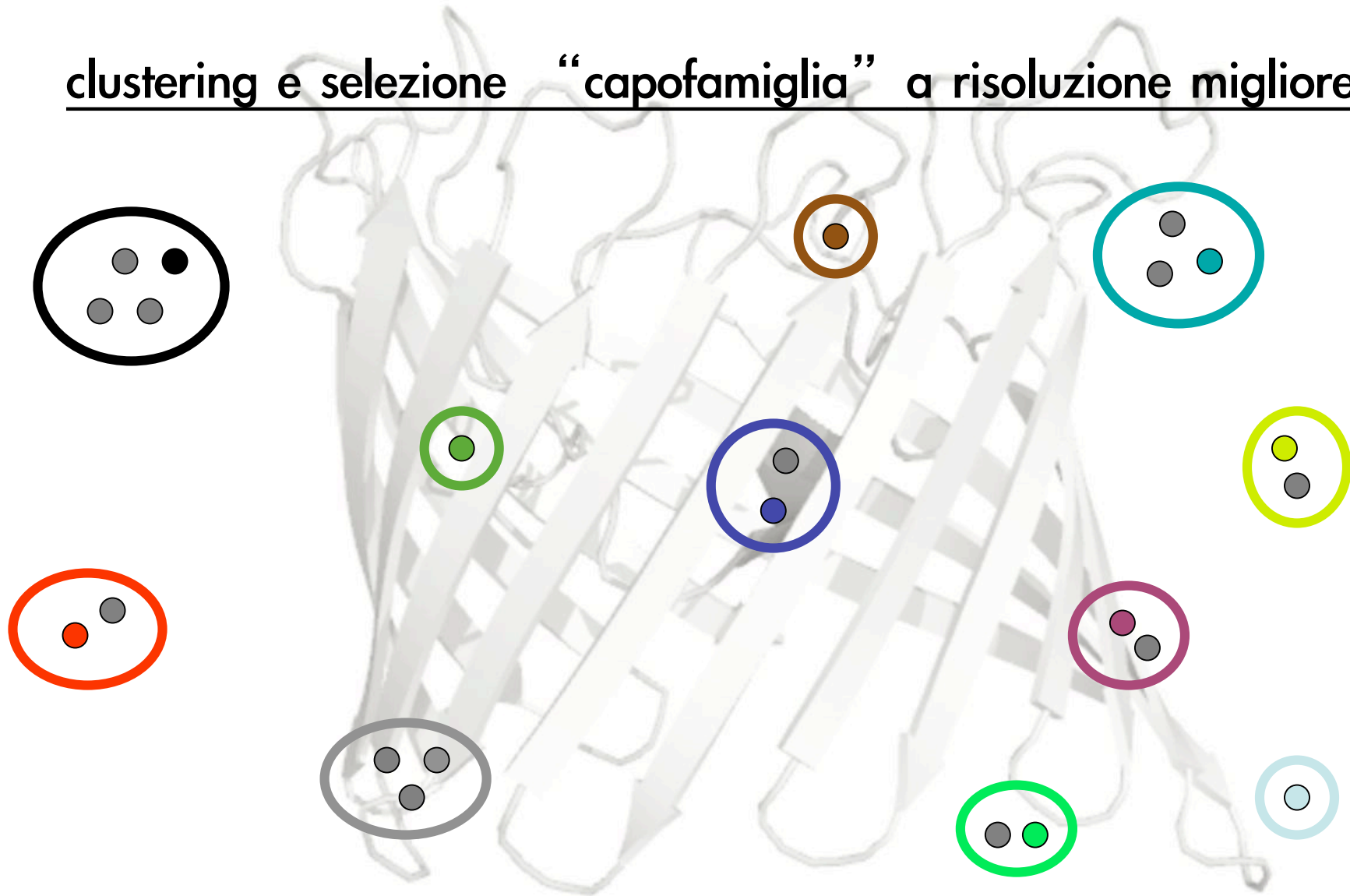
DATABASE - B2TMR

clustering



DATABASE - B2TMR

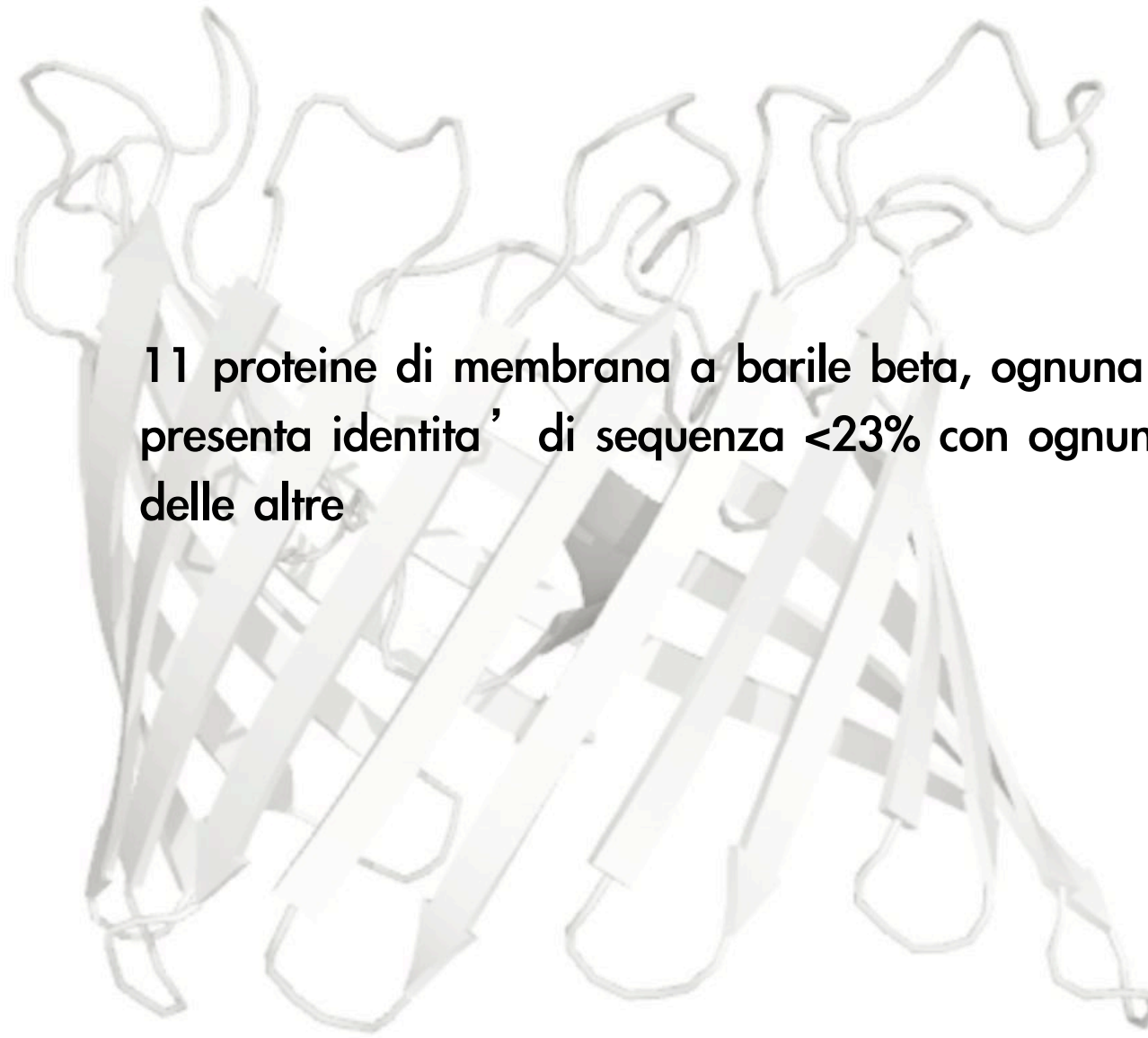
clustering e selezione “capofamiglia” a risoluzione migliore



DATABASE - B2TMR

- 2POR
- 1PRN
- 2OMF
- 2MPR
- 1A0S
- 1FEP
- 1FCP
- 1QJ8
- 1BXW
- 1QD5
- 7AHL

11 proteine di membrana a barile beta, ognuna presenta identita' di sequenza <23% con ognuna delle altre

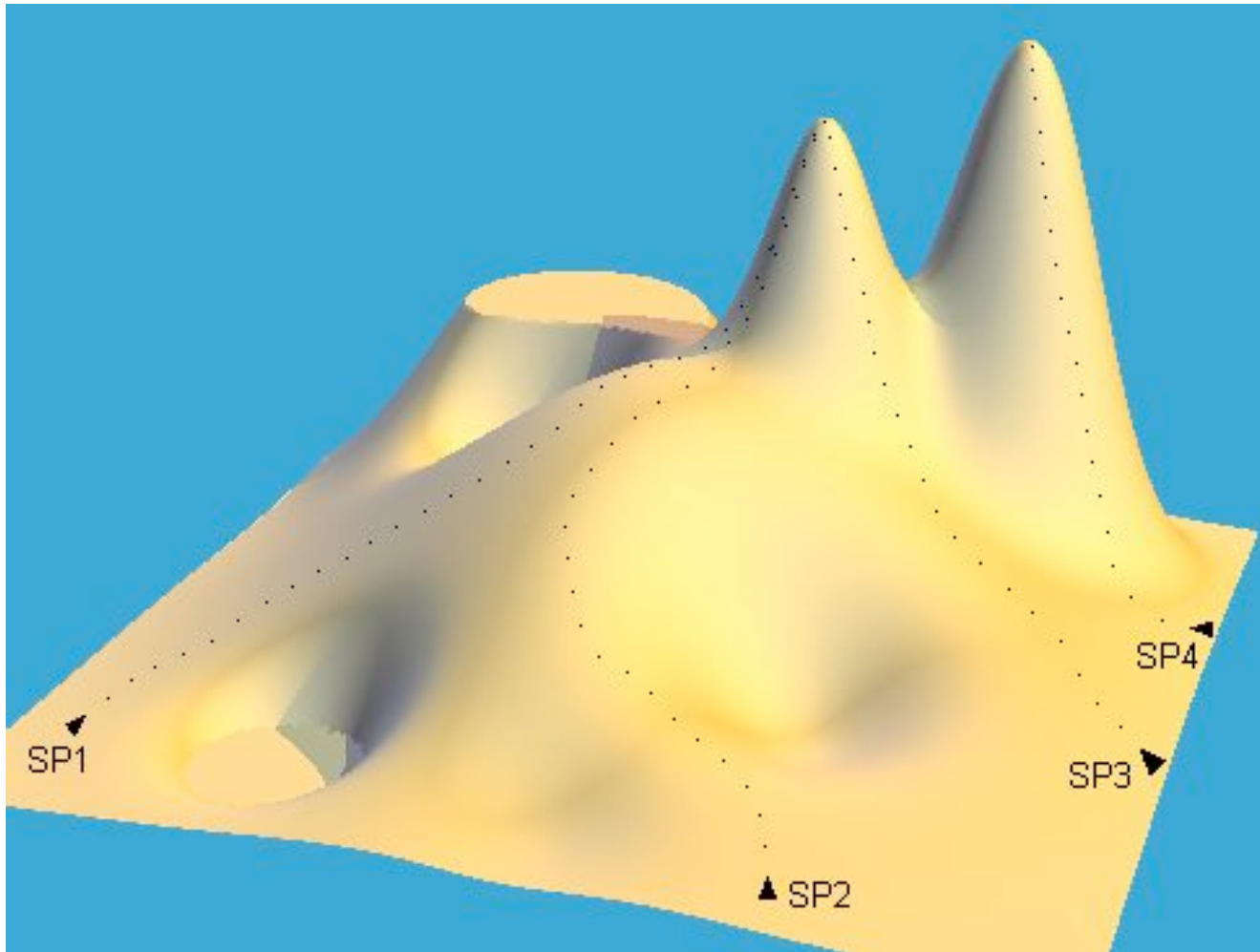


TRAINING

Il training e' essenzialmente una ricerca dei parametri ottimali della rete, e puo' essere svolto tramite un qualsiasi algoritmo di ottimizzazione; tuttavia l' elevato costo computazionale richiesto ha imposto come metodo dominante il Backpropagation algorithm

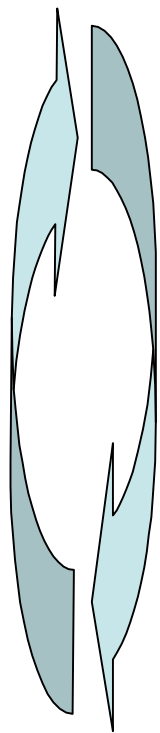
Metodi gradient descent

(steepest descent - hill climbing)

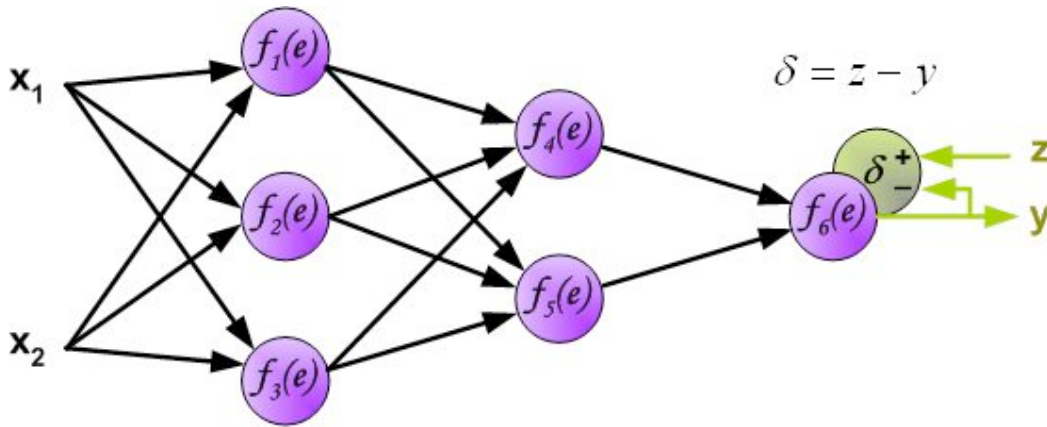


Viene valutata la derivata della funzione costo (errore) rispetto ai parametri da migliorare, e ci si muove verso il minimo locale

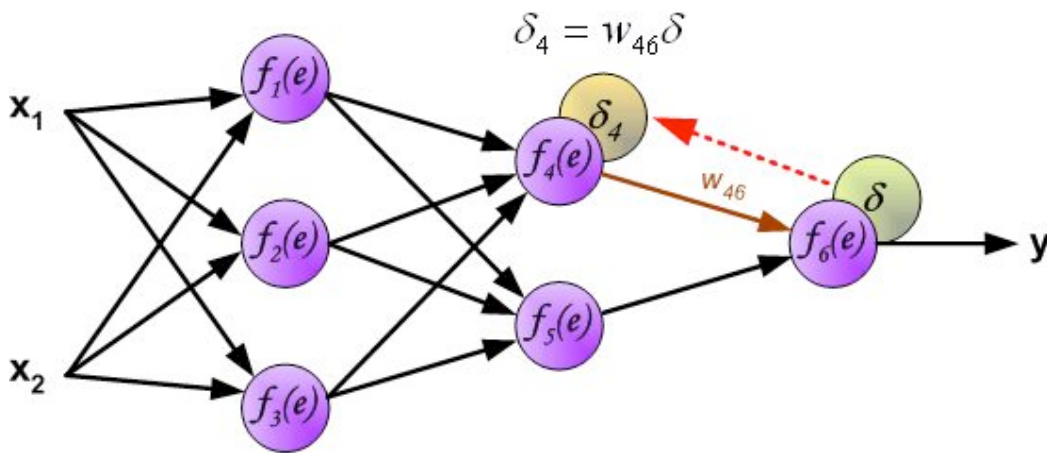
I pesi vengono inizializzati con piccoli valori random
(positivi e negativi)



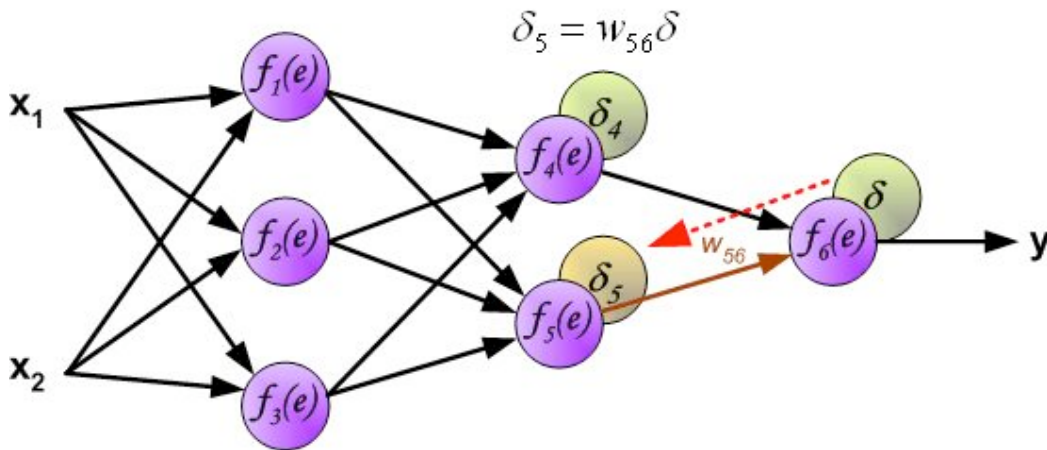
- Passo in avanti: uno o più input del training set vengono dati alla rete neurale
- viene calcolato lo scarto sull' output
- Passo indietro: viene calcolato lo scarto sui neuroni nascosti
- Vengono aggiornati i pesi (aumentati o diminuiti) in base agli scarti e alle derivate degli errori rispetto ai pesi



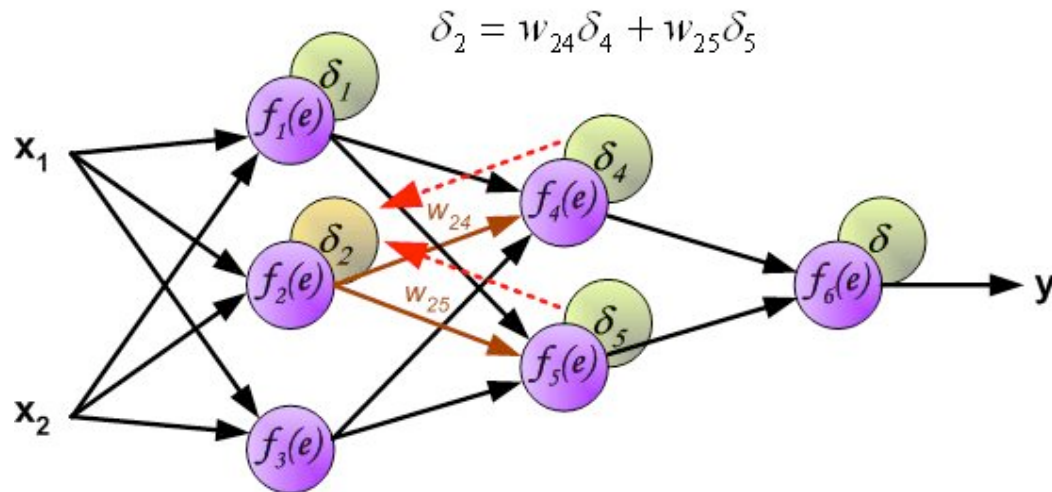
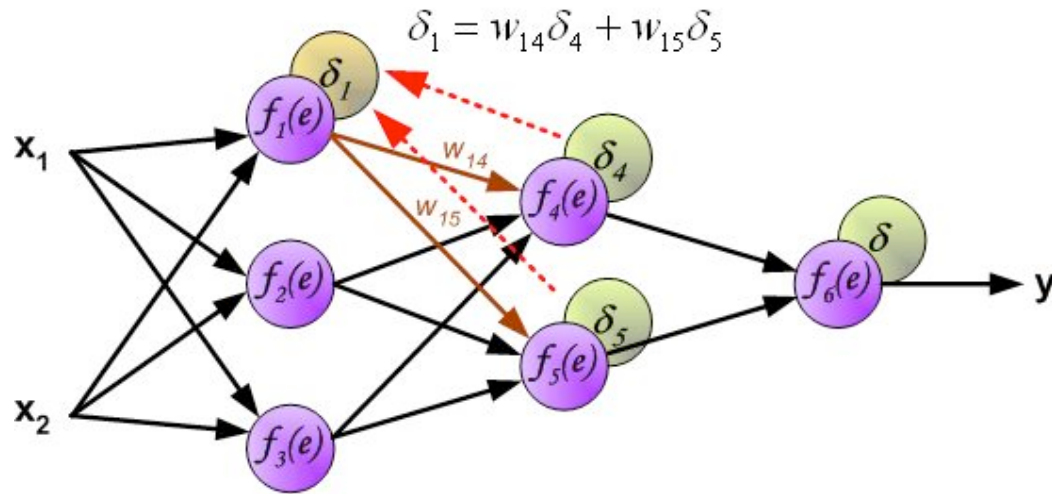
Passo in avanti:
 Calcolo dello scarto
 sugli output

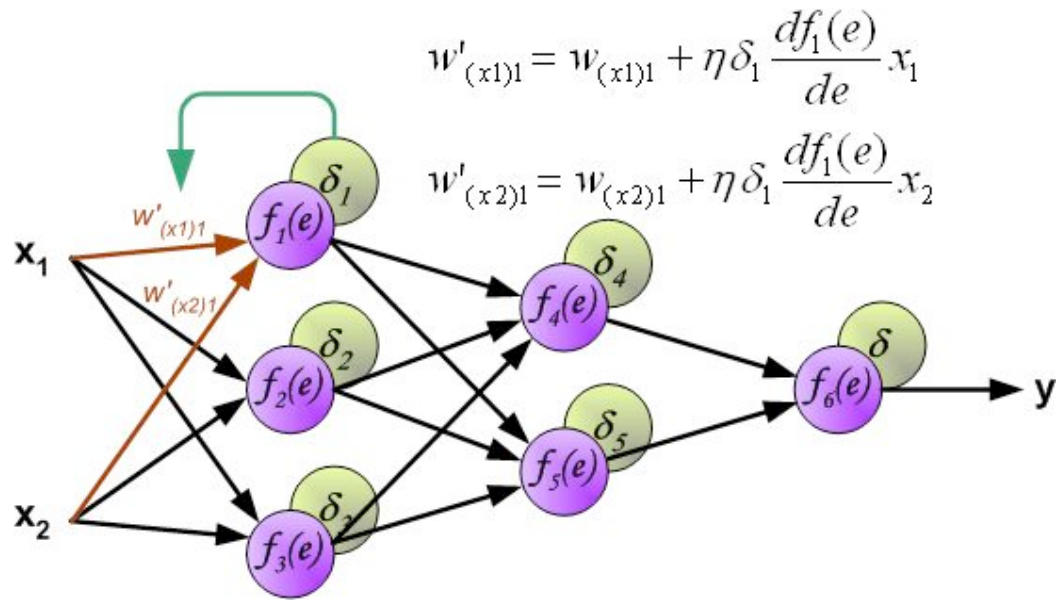


Retropropagazione dello scarto:
 Calcolo degli scarti sui
 neuroni degli strati precedenti



Retroprogazione dello scarto:
Calcolo degli scarti sui
neuroni degli strati precedenti

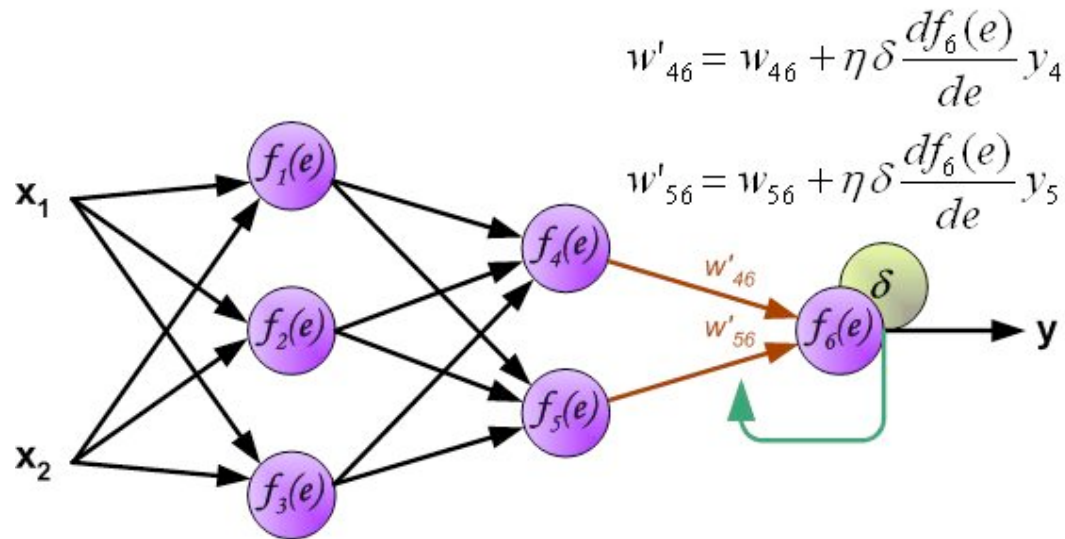




Aggiornamento dei pesi:

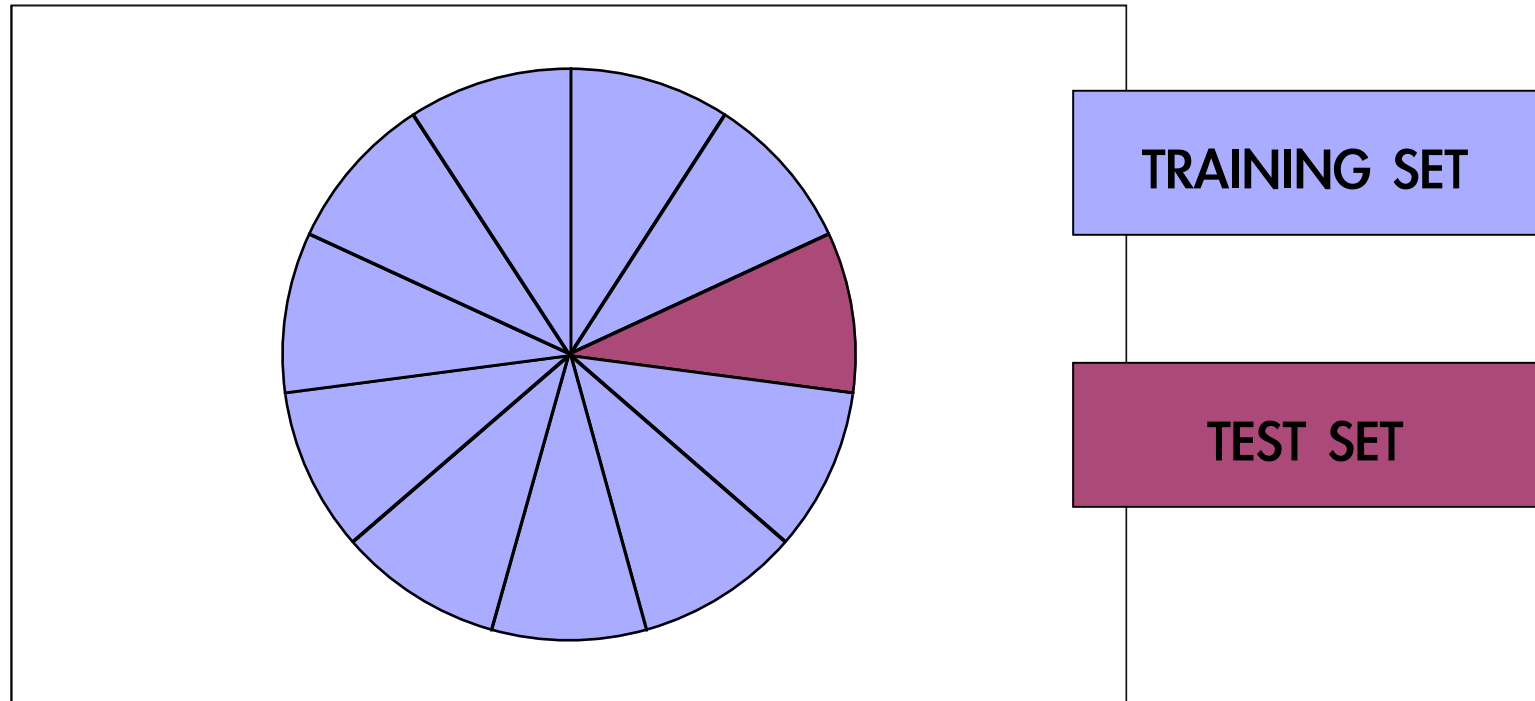
Ogni peso w è aggiornato in base allo scarto del neurone in cui si trova, alla derivata rispetto all'errore e ad un parametro η che definisce la velocità dell'addestramento

...

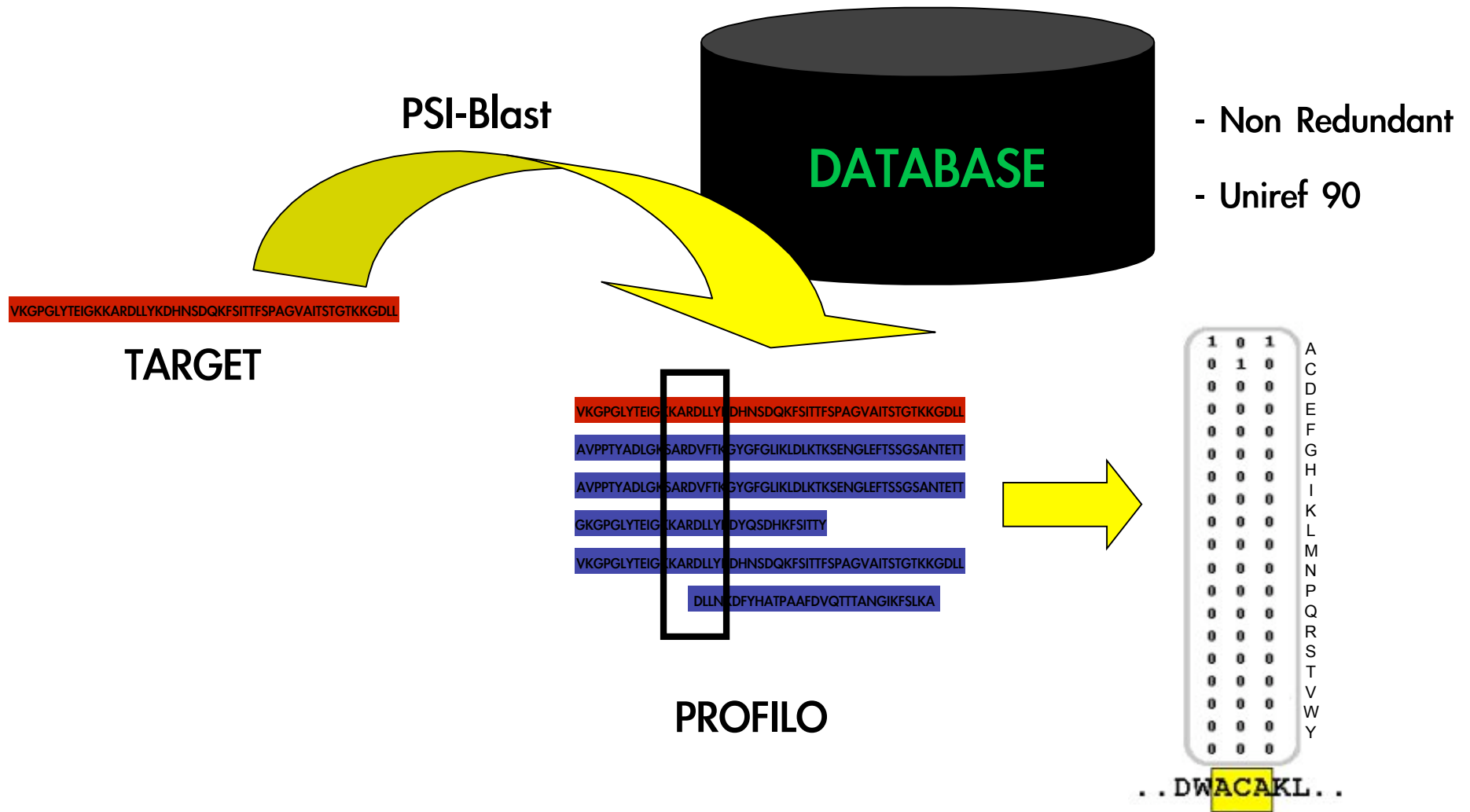


TRAINING AND TESTING

_ Jackknife procedure:



UTILIZZARE PROFILI IN INPUT

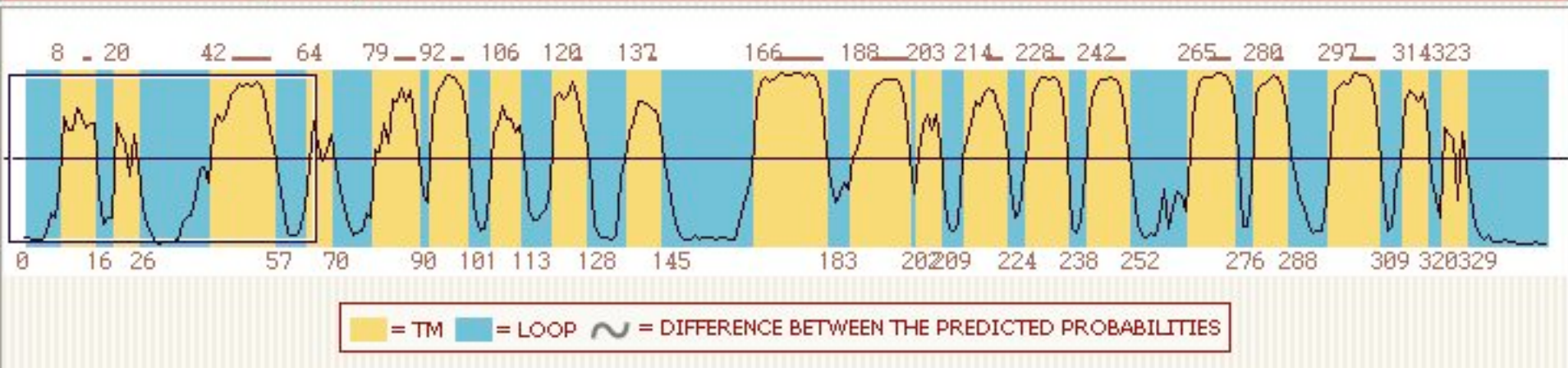


CODIFICA

TTTTTTTTT00000TTTTTTTTT0TT0000T0TTT...



- (B2TMR) Neural Networks Transmembrane Strand Predictor - SwissProt Database -



CODIFICA

Un algoritmo di programmazione dinamica sceglie il modello (la topografia) migliore fra quelli che soddisfano i vincoli imposti:

- _ Lunghezza segmenti TM (6-22 residui)
- _ Numero di segmenti pari, tra 2 e 22

La topologia e' scelta sulla base dell'osservazione che nelle proteine a barile beta i loop extracellulari sono piu' lunghi di quelli citoplasmatici

PRESTAZIONI

Table 1. Statistical analysis of the predictive performance

	Q2	Q(β)	Q(c)	P(β)	P(c)	C(β)	Sov(β)
<i>Training^a</i>							
Single sequence	0.95	0.94	0.95	0.95	0.94	0.89	0.97
HSSP	0.85	0.86	0.82	0.83	0.85	0.69	0.87
PSI-BLAST	0.89	0.84	0.93	0.92	0.85	0.77	0.91
PHD on β -barrel TM proteins ^b	0.71	0.55	0.88	0.82	0.65	0.45	0.61
PSIpred on β -barrel TM proteins ^b	0.77	0.73	0.83	0.81	0.75	0.56	0.72
<i>Testing^a</i>							
Single sequence	0.69	0.74	0.64	0.68	0.71	0.38	0.71
HSSP	0.73	0.76	0.70	0.72	0.74	0.46	0.75
PSI-BLAST	0.78	0.74	0.82	0.81	0.76	0.56	0.79

Il metodo raggiunge un accuratezza per residuo pari a 78%;
Il 93% dei filamenti beta e' localizzato correttamente.

Predizioni di topologia TM sui VDAC in Swiss-Prot

B2TMR <http://gpcr.biocomp.unibo.it/predictors/>

The screenshot displays a grid of protein sequences with their predicted transmembrane (TM) regions highlighted in red. Each row represents a different protein, and the columns show the sequence with red boxes indicating the predicted TM segments. The sequences are listed on the right side of the grid, including identifiers like 'vGPCR_1', 'vGPCR_2', etc.

PredTMBB <http://biophysics.biol.uoa.gr/PRED-TMBB/>

The screenshot shows a grid of protein sequences with predicted TM regions highlighted in red. The layout is similar to the B2TMR screenshot, with protein identifiers listed on the right and their corresponding sequences with red boxes indicating predicted TM segments.

profTMB <http://rostlab.org/cgi-bin/var/bigelow/proftmb/>

The screenshot displays a grid of protein sequences with predicted TM regions highlighted in red. The protein identifiers are listed on the right, and the sequences are shown with red boxes indicating the predicted TM segments.

THE END

