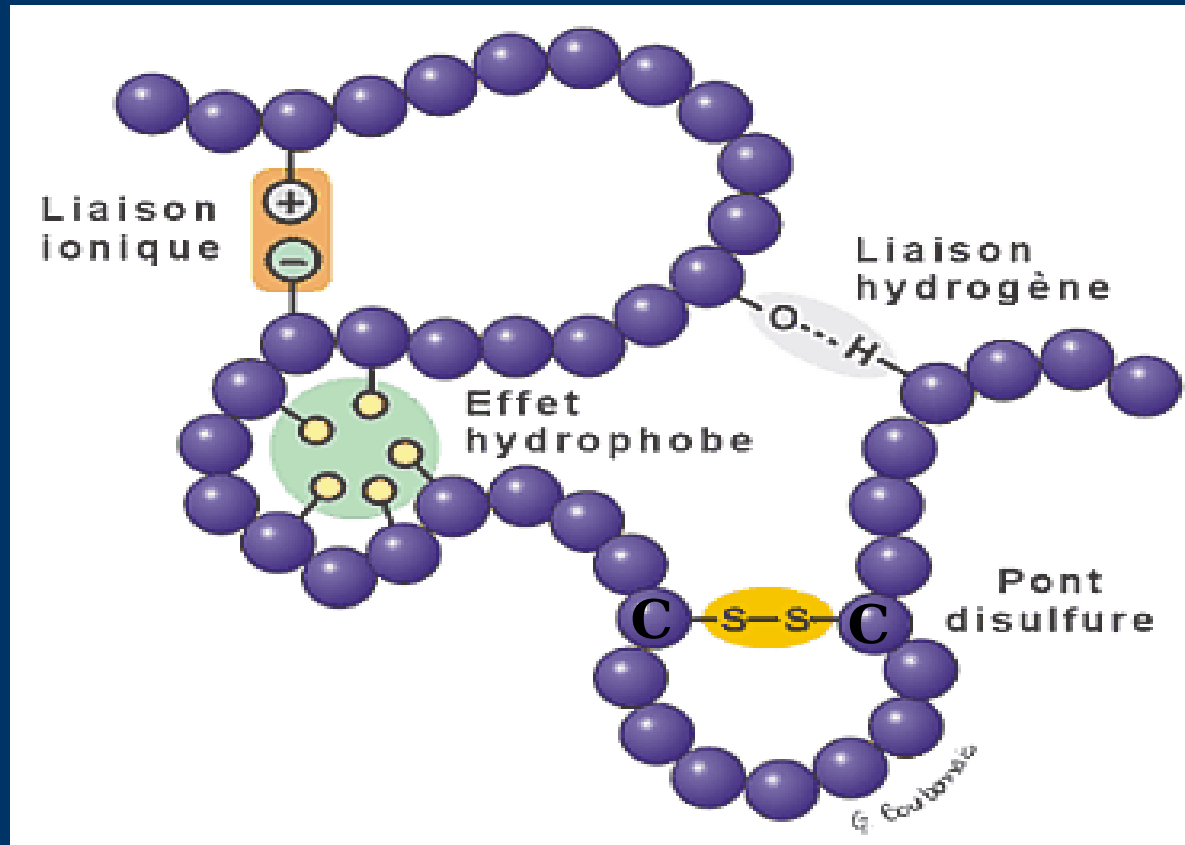


Synthèse de quelques résultats sur la prédiction des ponts disulfures dans les protéines

Objectifs de la présentation

- tour d'horizon des différents résultats connus
- description des méthodes utilisées et des résultats obtenus
- discussion sur :
 - la validité/pertinence des résultats
 - l'existence ou non d'une information locale
 - la faisabilité de la tâche

Rappels sur les ponts disulfures



Rappels sur les ponts disulfures

- Décomposition du problème en deux étapes :
 - prédiction des cystéines oxydées
 - nombreux travaux menés, articles,
 - dernières méthodes mises au point : >80% de bonne prédiction,
 - depuis un an: baisse des publications sur le sujet
 - prédiction des ponts eux-mêmes
 - peu d'articles
 - récent (2001 pour les premiers articles)
 - questions sur la faisabilité de la tâche

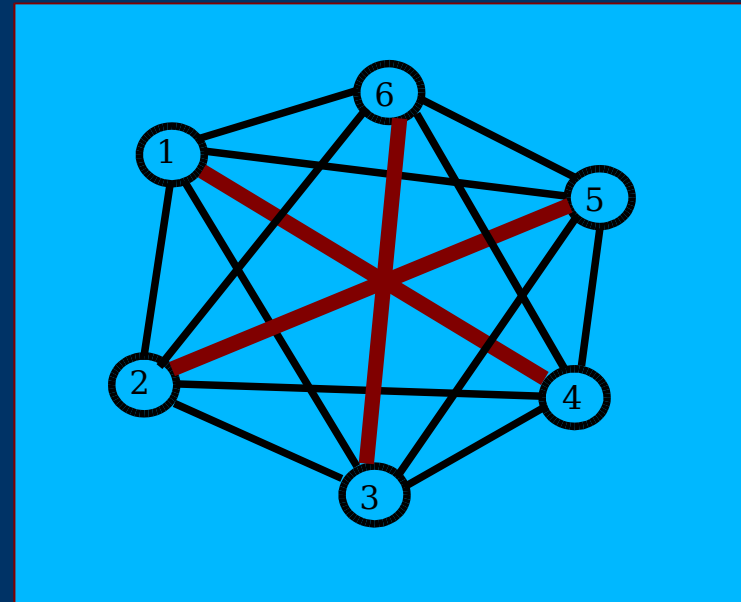
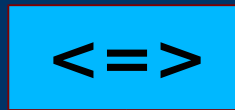
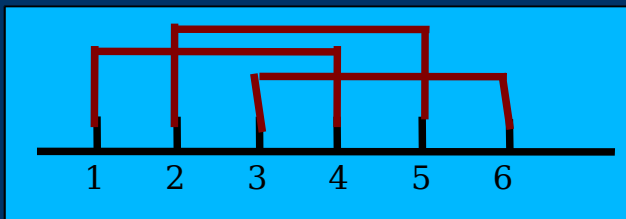
Articles existants

- Piero Fariselli et Rita Casadio
 - Prediction of disulfide connectivity in proteins (*Bioinformatics* 17:957-964, 2001)
 - *A neural network-based method for predicting the disulfide connectivity in proteins (Proceedings of the 6th International Conference on Knowledge Engineering Systems, 2002)*
- Alessandro Vullo et Paolo Frasconi
 - A recursive connectionist approach for predicting disulfide connectivity in proteins (*proceedings of the 18th ACM symposium on applied computing, 2003*)
 - Disulfide connectivity prediction using recursive neural networks and evolutionary information (*Bioinformatics, 2004*)

Piero Fariselli et Rita Casadio

Prediction of disulfide connectivity in proteins (2001)

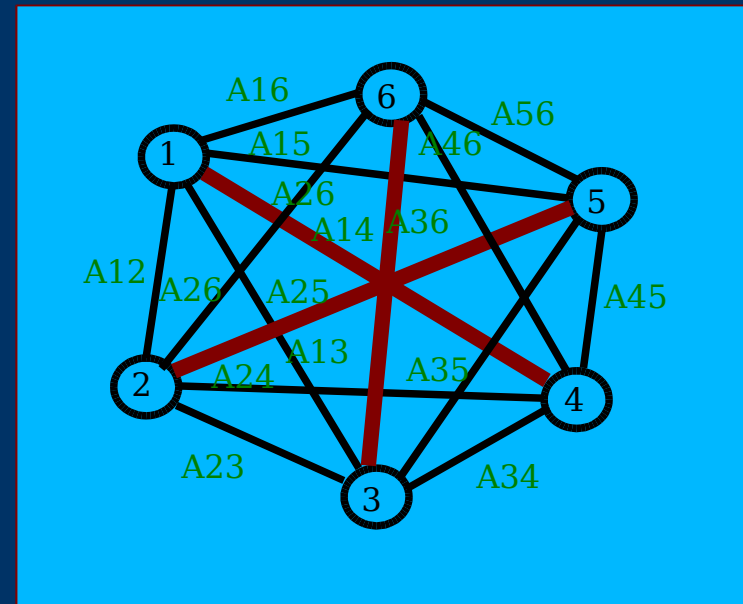
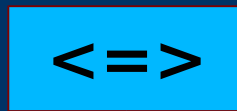
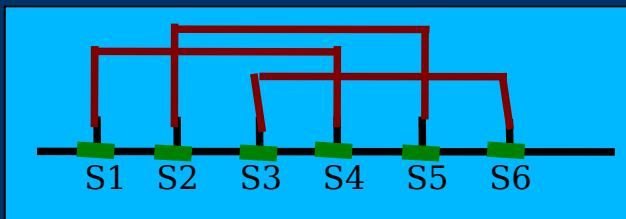
- Appariement des cystéines \Leftrightarrow couplage parfait de poids maximal



Piero Fariselli et Rita Casadio

Prediction of disulfide connectivity in proteins (2001)

- Appariement des cystéines \Leftrightarrow couplage parfait de poids maximal



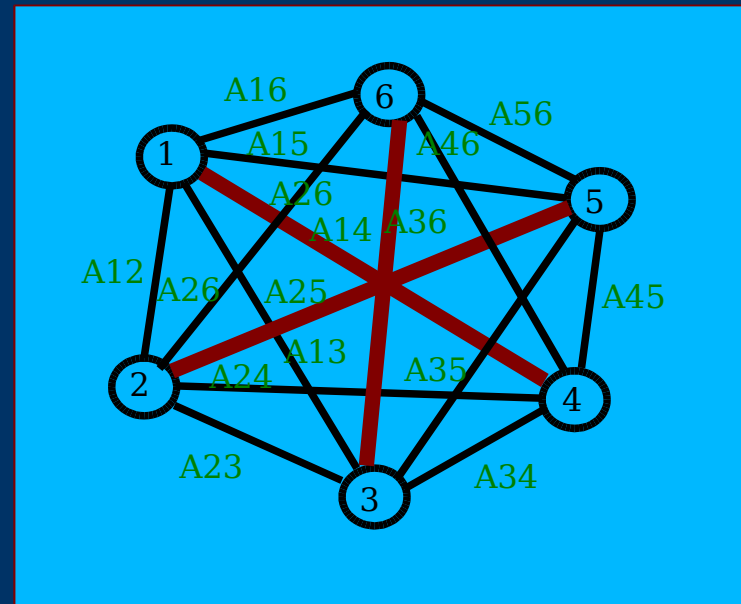
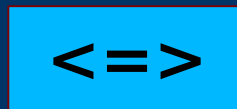
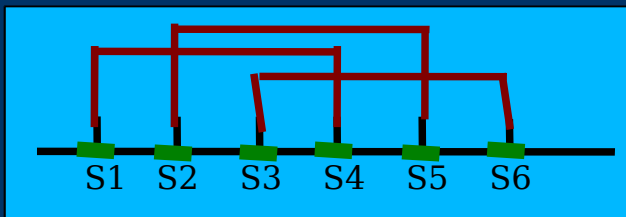
■ = portion de 11 résidus (ak) centrée sur une cystéine

A_{ij} = affinité entre la portion S_i et la portion S_j
 $= \sum U(a_k, a_l)$ a_k dans S_i , a_l dans S_j

Piero Fariselli et Rita Casadio

Prediction of disulfide connectivity in proteins (2001)

- Appariement des cystéines \Leftrightarrow couplage parfait de poids maximal
 - \Rightarrow Résolution en temps polynomial (Edmonds, 1965)



■ = portion de 11 résidus (ak) centrée sur une cystéine

A_{ij} = affinité entre la portion S_i et la portion S_j
 $= \sum U(a_k, a_l)$ a_k dans S_i , a_l dans S_j

Piero Fariselli et Rita Casadio

Prediction of disulfide connectivity in proteins (2001)

- Différentes affinités testées :
 - Mirny and Shakhnovich contact potentiel (1996)
 - optimisé pour des problèmes de prédiction d'agrégation de protéines
 - résultats de l'ordre de l'aléa

Nb ponts	Nb prot	QP	QP/QPA	QC	QC/QCA
2	158	0,34	1	0,35	1
3	153	0,05	0,9	0,19	1
4	103	0,01	1	0,14	1
5	44	0	0	0,12	1

Piero Fariselli et Rita Casadio

Prediction of disulfide connectivity in proteins (2001)

- contact potentiel dérivé de contraintes d'optimisation
 - aucun apport d'éléments biologiques supplémentaire à l'information locale
 - les potentiels $U(a_k, a_l)$ sont les variables d'un système linéaire
 - maximiser $Z = S(U, C_p^*) - \langle S(U, C_p) \rangle$
 - résultats légèrement meilleurs que l'aléa

Nb ponts	Nb prot	QP	QP/QPA	QC	QC/QCA
2	158	0,42	1,2	0,42	1,2
3	153	0,09	1,3	0,21	1
4	103	0,05	5	0,21	1,5
5	44	0	0	0,14	1,3

Piero Fariselli et Rita Casadio

Prediction of disulfide connectivity in proteins (2001)

- odd-ratio contact potentiel (Durbin et al, 1998)
- pour un couple d'acides aminés donné : le contact potentiel est défini par le rapport de la fréquence d'apparition de ce couple dans les portions appariées et non appariées
- Résultats :

Nb ponts	Nb prot	QP	QP/QPA	QC	QC/QCA
2	158	0,46	1,4	0,46	1,4
3	153	0,17	2,4	0,29	1,4
4	103	0,11	11,5	0,31	2,2
5	44	0	0	0,17	1,5

Piero Fariselli et Rita Casadio

Prediction of disulfide connectivity in proteins (2001)

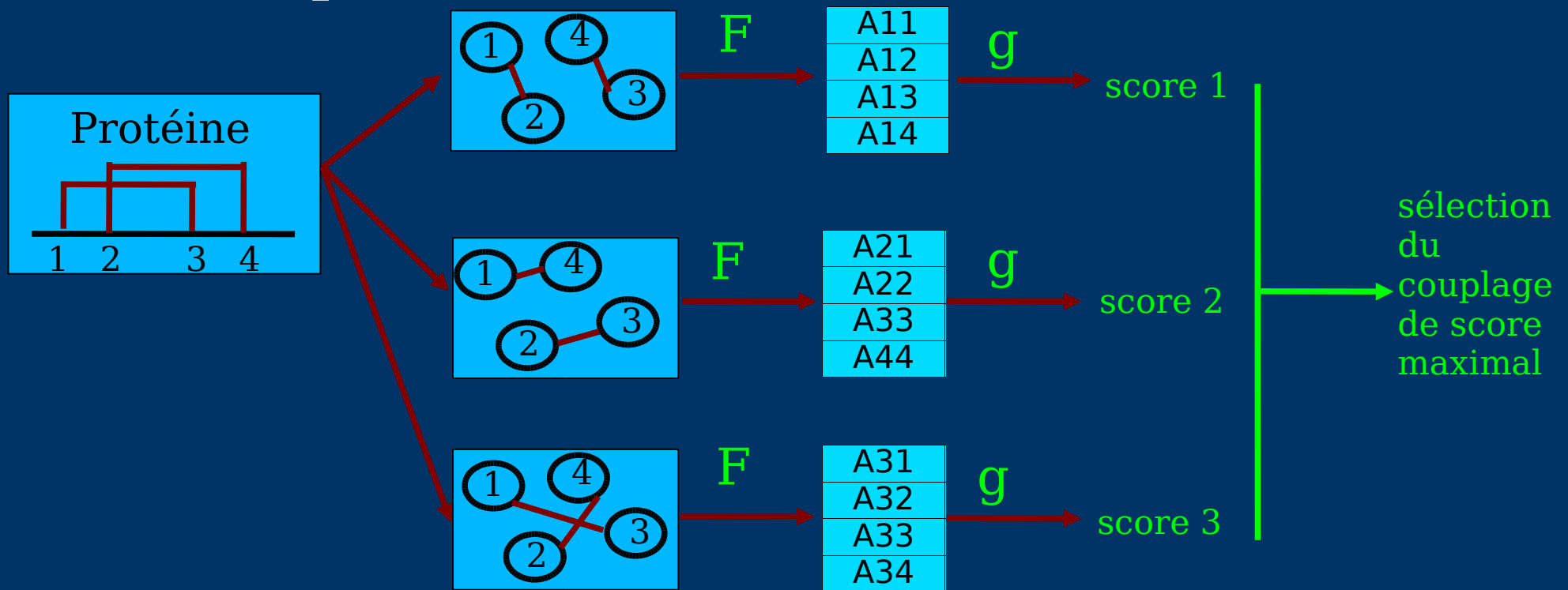
- Contact potentiel Odd-ratio optimisé
 - estimation d'un score $F(U)$ pour un contact potentiel U
 - > proportion d'appariements bien prédits
 - calcul de nouveaux contacts potentiels U' par la méthode du recuit simulé avec différentes température initiales et finales
 - calcul de la probabilité de garder ou rejeter U'
 - > $p = \min(1, \exp(F(U') - F(U))/T)$
 - Résultats
 - ♦ mise en avant d'acides aminés favorisant et défavorisant l'appariement
 - ♦ meilleurs résultats de l'étude

Nb ponts	Nb prot	QP	QP/QPA	QC	QC/QCA
2	158	0,56	1,7	0,56	1,7
3	153	0,21	3,1	0,36	1,7
4	103	0,17	17,8	0,37	2,6
5	44	0,02	22,7	0,21	1,9

Alessandro Vullo et Paolo Frasconi

A recursive connectionist approach for predicting disulfide connectivity in proteins (2003)

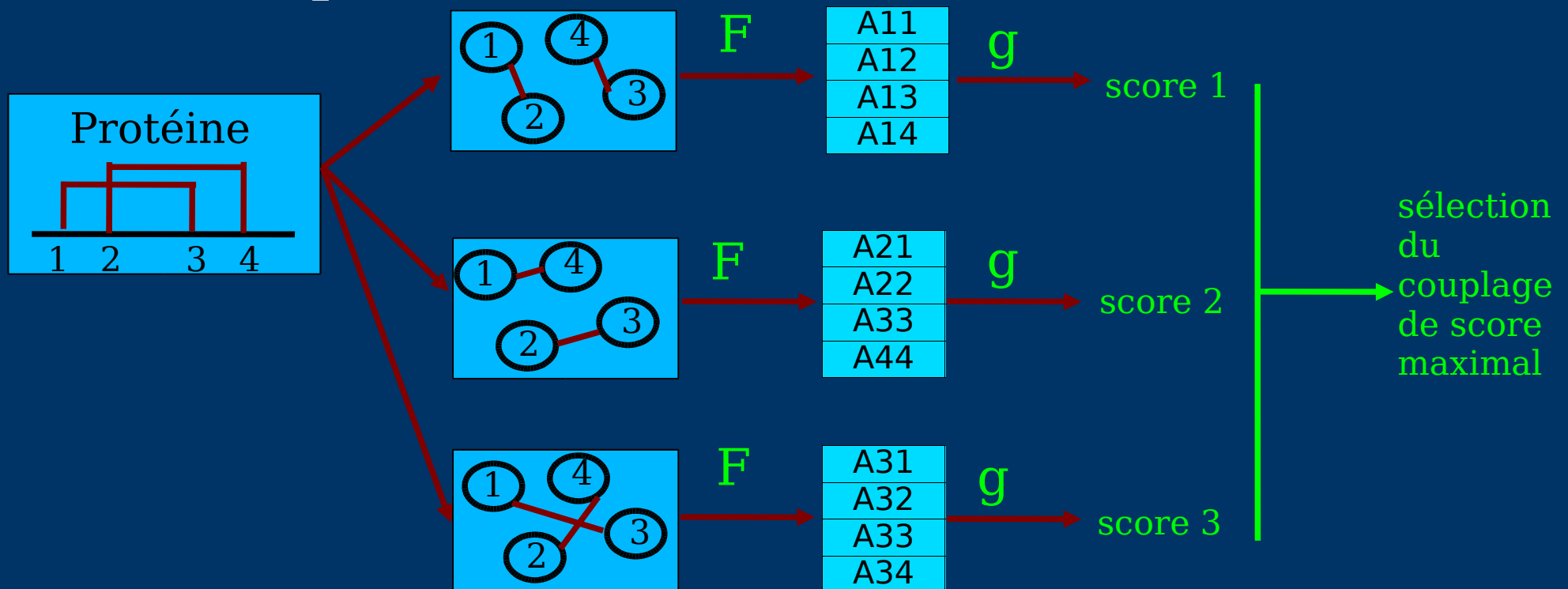
- Principe



Alessandro Vullo et Paolo Frasconi

A recursive connectionist approach for predicting disulfide connectivity in proteins (2003)

- Principe

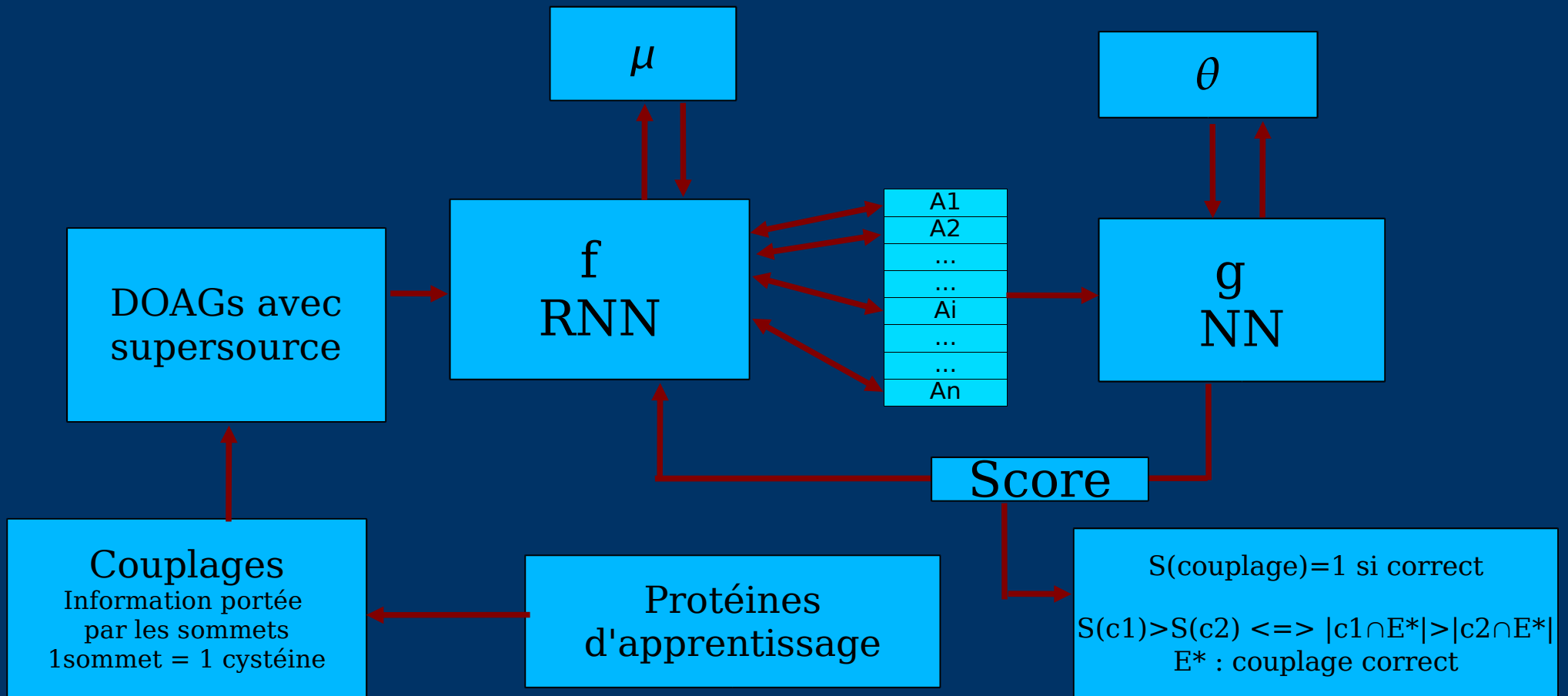


Objectif : trouver F et g tel que l'erreur de prédiction soit minimale

Alessandro Vullo et Paolo Frasconi

A recursive connectionist approach for predicting disulfide connectivity in proteins (2003)

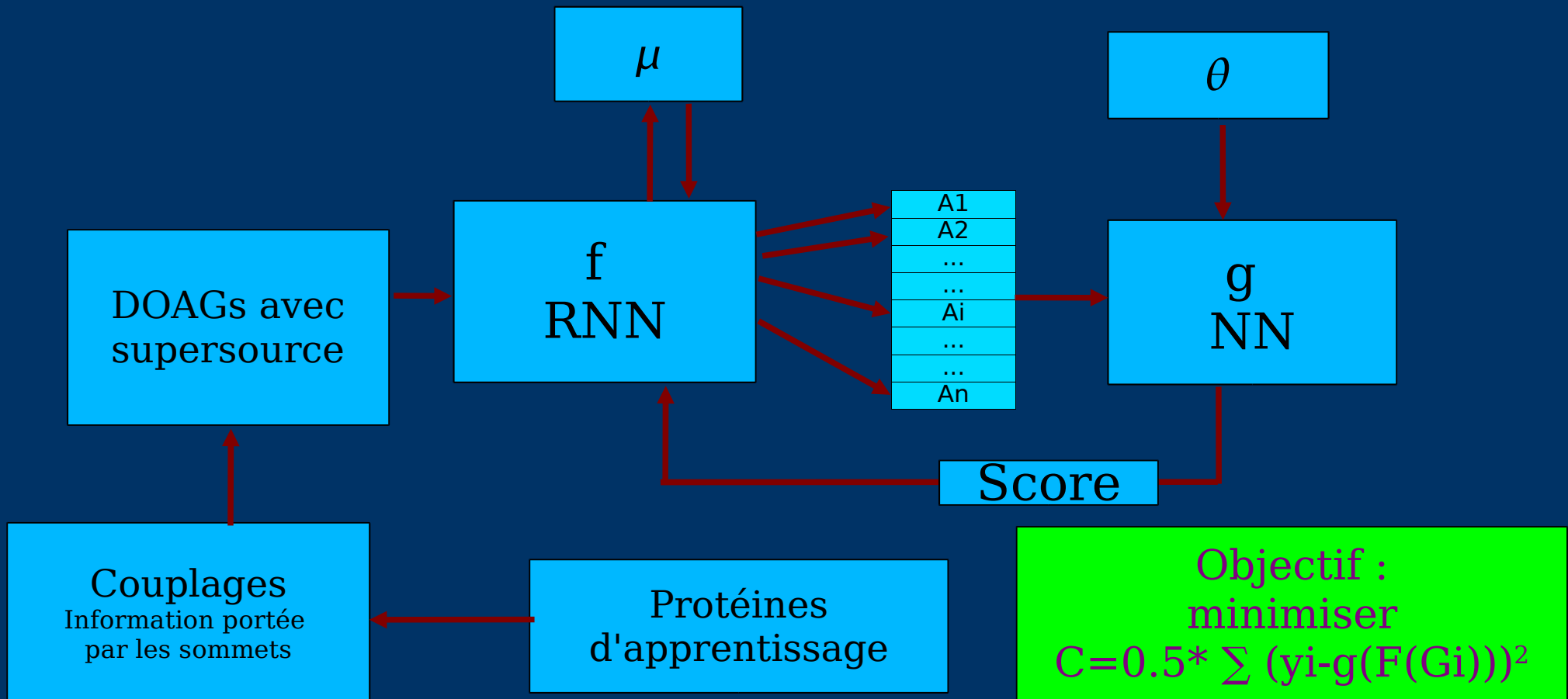
- Apprentissage des fonctions F et g



Alessandro Vullo et Paolo Frasconi

A recursive connectionist approach for predicting disulfide connectivity in proteins (2003)

- Apprentissage des fonctions F et g



Alessandro Vullo et Paolo Frasconi

A recursive connectionist approach for predicting disulfide connectivity in proteins (2003)

- Application :
 - information portée par les noeuds :
 - information locale : 20 acides aminés autour de la cystéine
 - information globale : position relative de la cystéine sur la chaîne
 - Protéines : Swiss-Prot 39
 - Nombre de ponts : 2, 3, 4 et 5
 - Pas d'informations sur le choix des paramètres initiaux

Alessandro Vullo et Paolo Frasconi

A recursive connectionist approach for predicting disulfide connectivity in proteins (2003)

- Meilleurs résultats :

Nb ponts	Nb prot	QP	QP/QPA	QC	QC/QCA
2	39	0,71	2,1	0,71	2,1
3	36	0,33	4,7	0,47	2,3
4	25	0,1	10,5	0,29	2
5	11	0,11	124,8	0,33	2,9

Alessandro Vullo et Paolo Frasconi

Disulfide connectivity prediction using recursive neural networks and evolutionary information (2004)

- Même méthode que l'article précédent
- Ajout d'informations :
 - rapport de la taille de la protéine sur la taille maximale des protéines dans l'ensemble d'apprentissage
 - utilisation de profils (PSI-BLAST)
- Changement de taille de fenêtre :
 - codage selon séquence : taille 4 (rayon 2)
 - codage selon profils : taille 5 (rayon 2)
- Résultats :

Nb ponts	2	3	4	5
QC Freq	0,58	0,37	0,1	0,23
Bi-RNN seq	0,59	0,34	0,3	0,24
Bi-RNN prof	0,73	0,51	0,37	0,3

Discussion

- L'information locale contribue-t-elle à ces résultats?
- A-t-on vraiment apprentissage?
- Faut-il continuer à chercher de l'information locale?
- Est-il possible de pouvoir prédire les ponts disulfures?