

Introduction d'une contrainte longue distance pour la prédiction des brin β dans les prédictions des structures secondaires par HMM

Tentatives

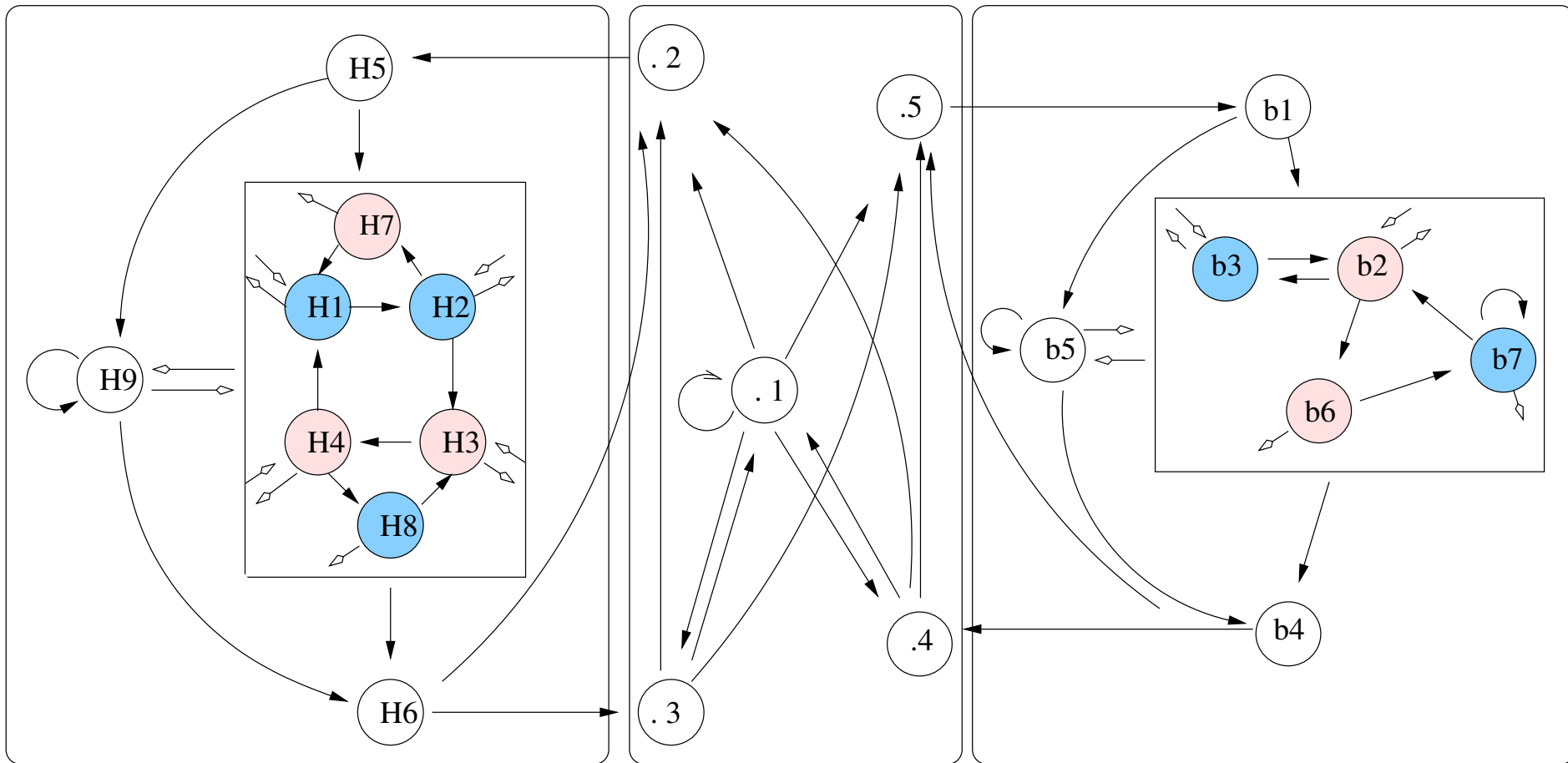
Juliette Martin, Jean-François Gibrat, François Rodolphe

HMM

HELICE

APERIODIQUE

BRIN



Qualité de prédiction

- 65.3% de bonne prédiction
- la prédiction des brins β est moins bonne que celle des hélices α
- prédit trop peu de β

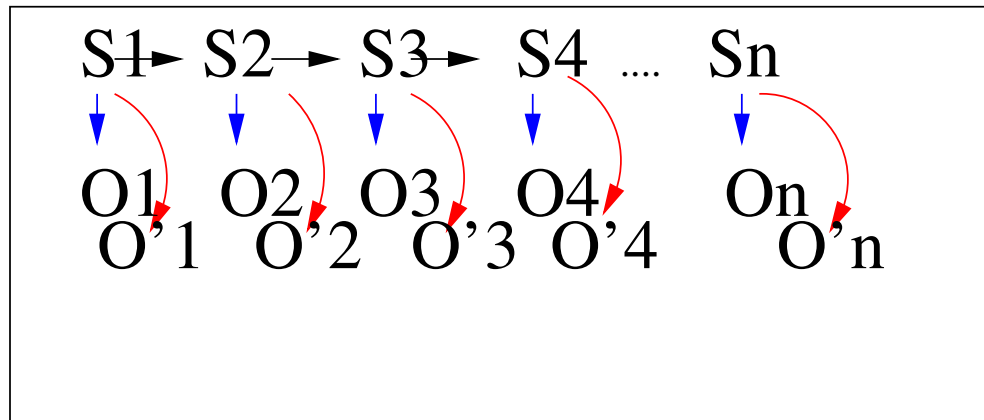
	α réel	β réel	coil réel
α prédit	27979(75%)	2054(5%)	7371(20%)
β prédit	4667(22%)	9534(45%)	7011(33%)
coil prédit	8756(23%)	3523(10%)	25142(67%)

	sensibilité	spécificité	CCM	SOV
α	75.8%	67.6%	0.57	64.0%
β	45.0%	63.1%	0.47	47.4%
autre	67.2%	63.6%	0.50	57.7%

	Proportions réelles	Proportions prédites
Hélice α	39 %	43 %
Brin β	22 %	16 %
Autre	39 %	41%

Caractéristique des brins β

- L'hélice α est stabilisée par des liaisons H locales (résidus i et $i+4$), le feuillet β , par des liaisons H entre résidus éloignés.
- Donner une information supplémentaire aun HMM, une fonction de score qui intègre la contrainte longue distance.



Fonction de score simple

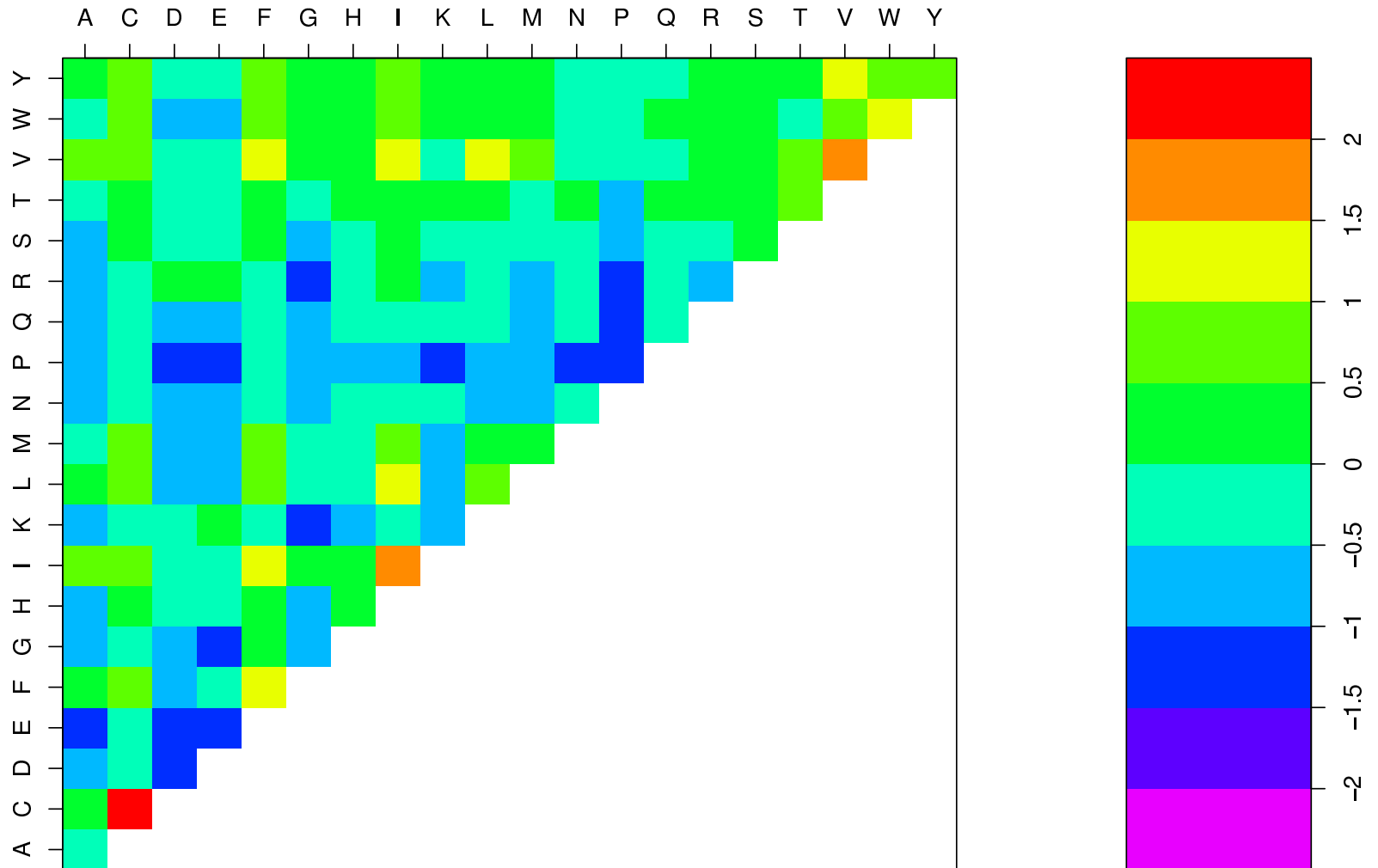
- Il s'agit de construire une matrice de scores pour trouver des appariements potentiels dans la séquence.



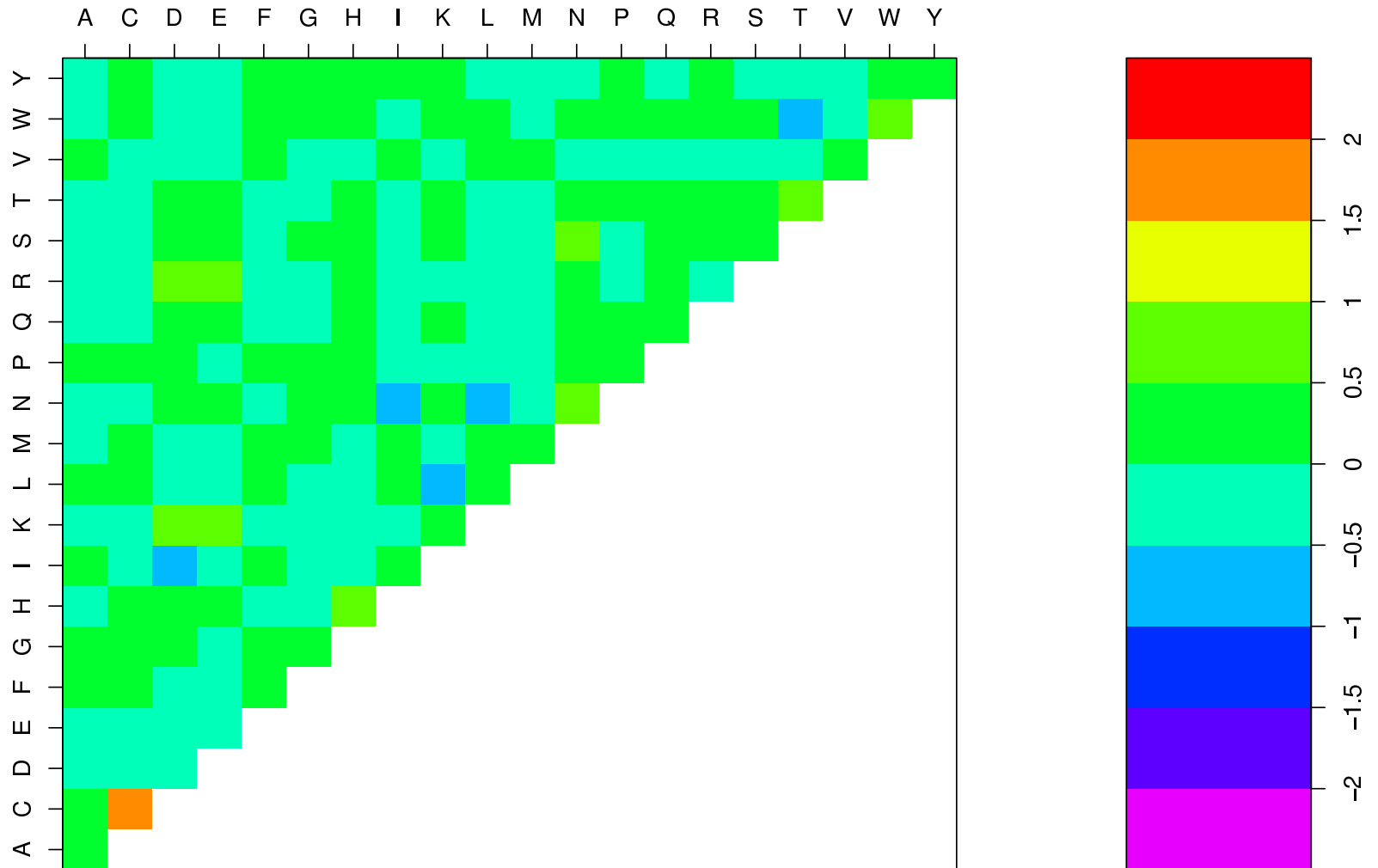
$$score = \log \frac{P_{obs}(paire)}{P_{at}(paire)}$$

avec $P_{obs}(paire)$ la probabilité observée de la paire dans des feuillet β , et $P_{at}(paire)$ la probabilité attendue aléatoirement dans une protéine.

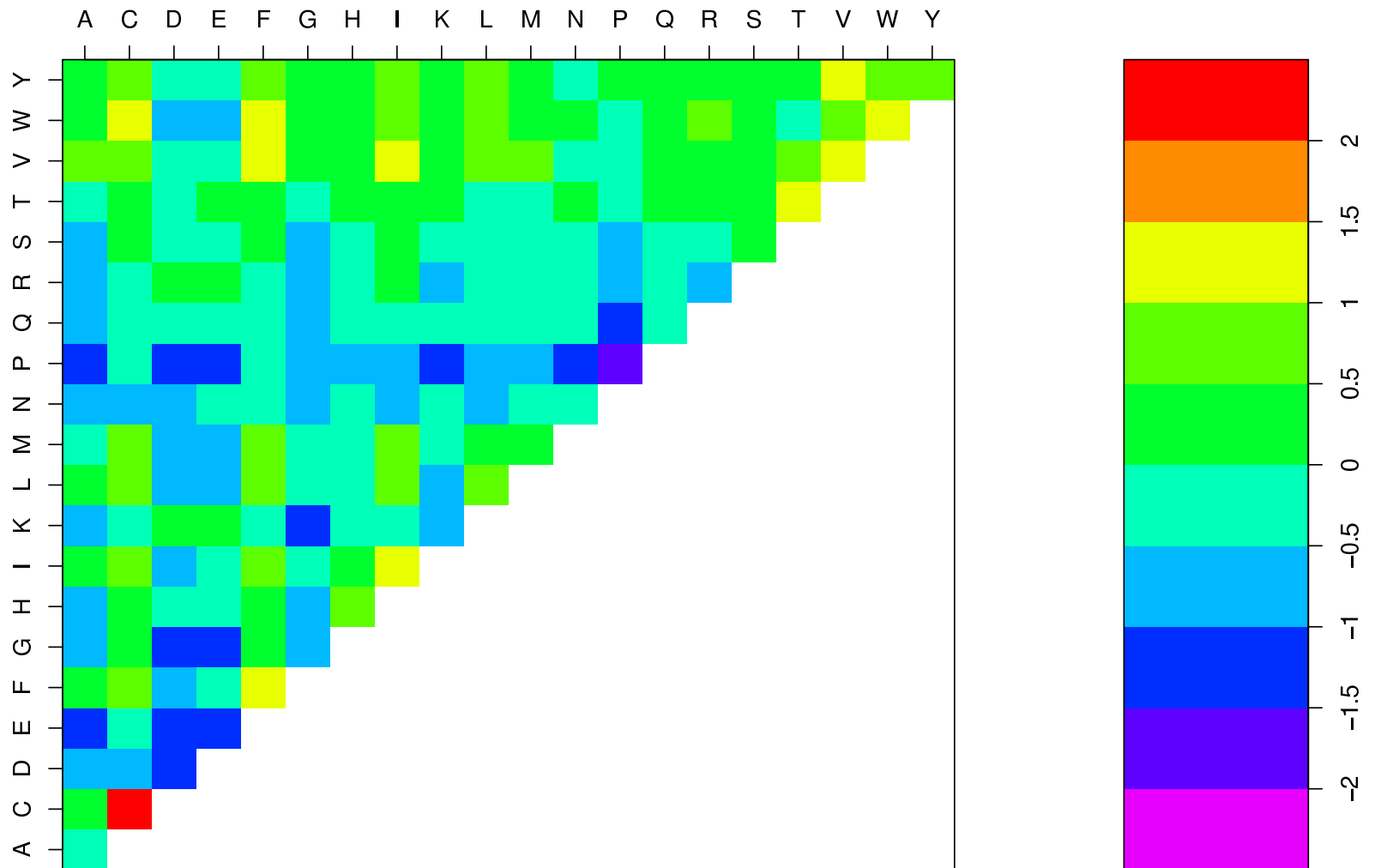
Scores Anti+Para



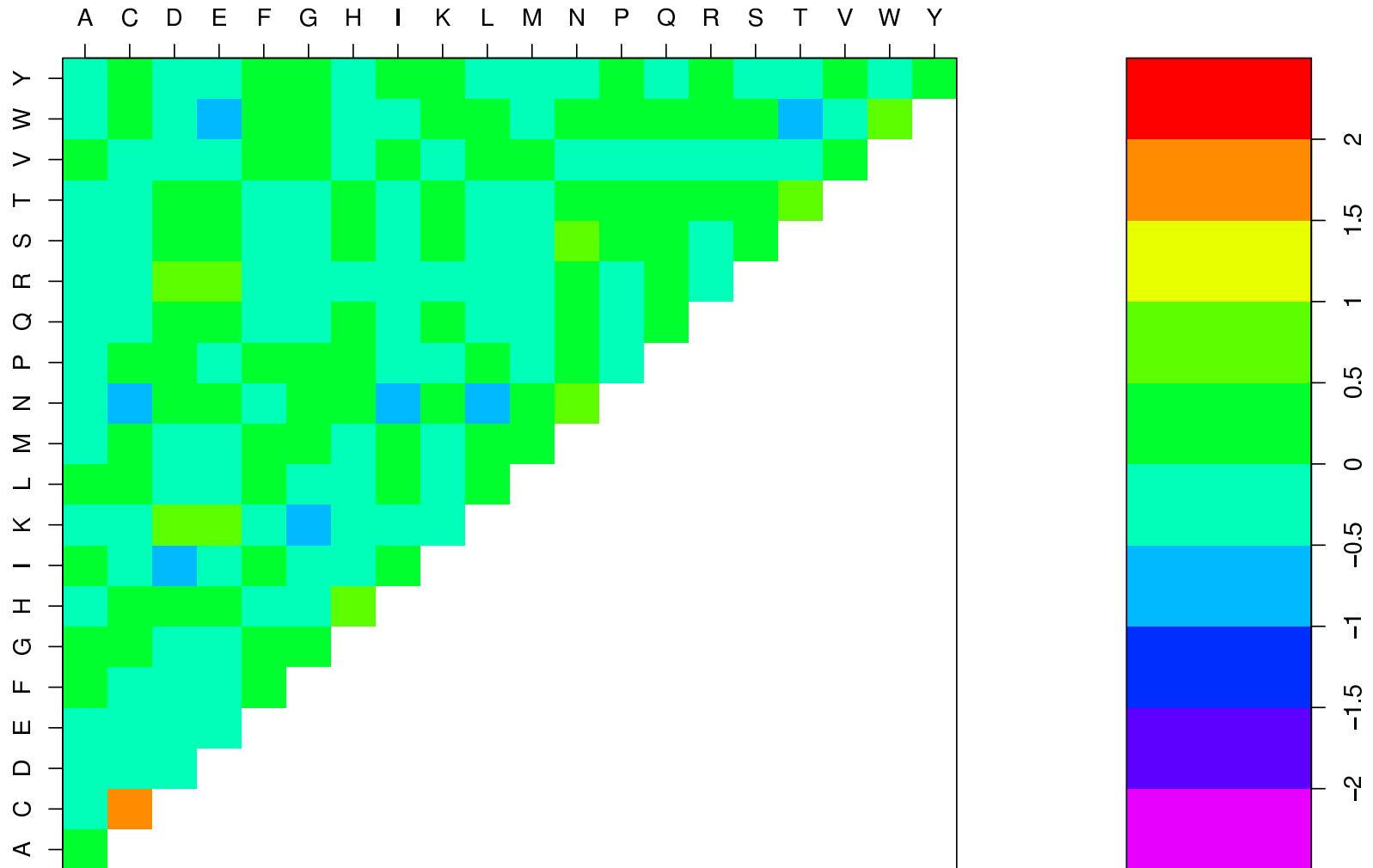
Spécificité de paires



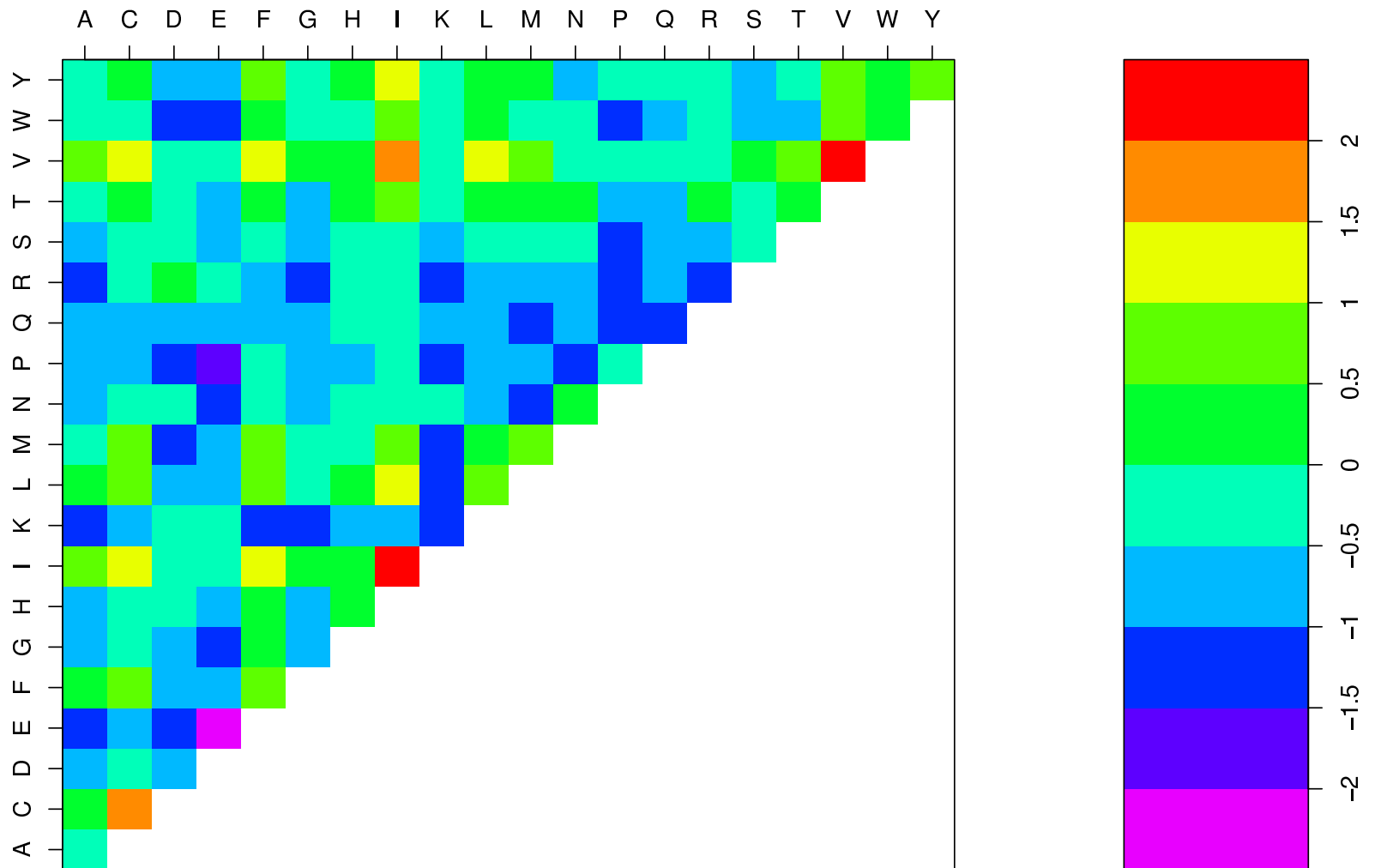
Scores Anti



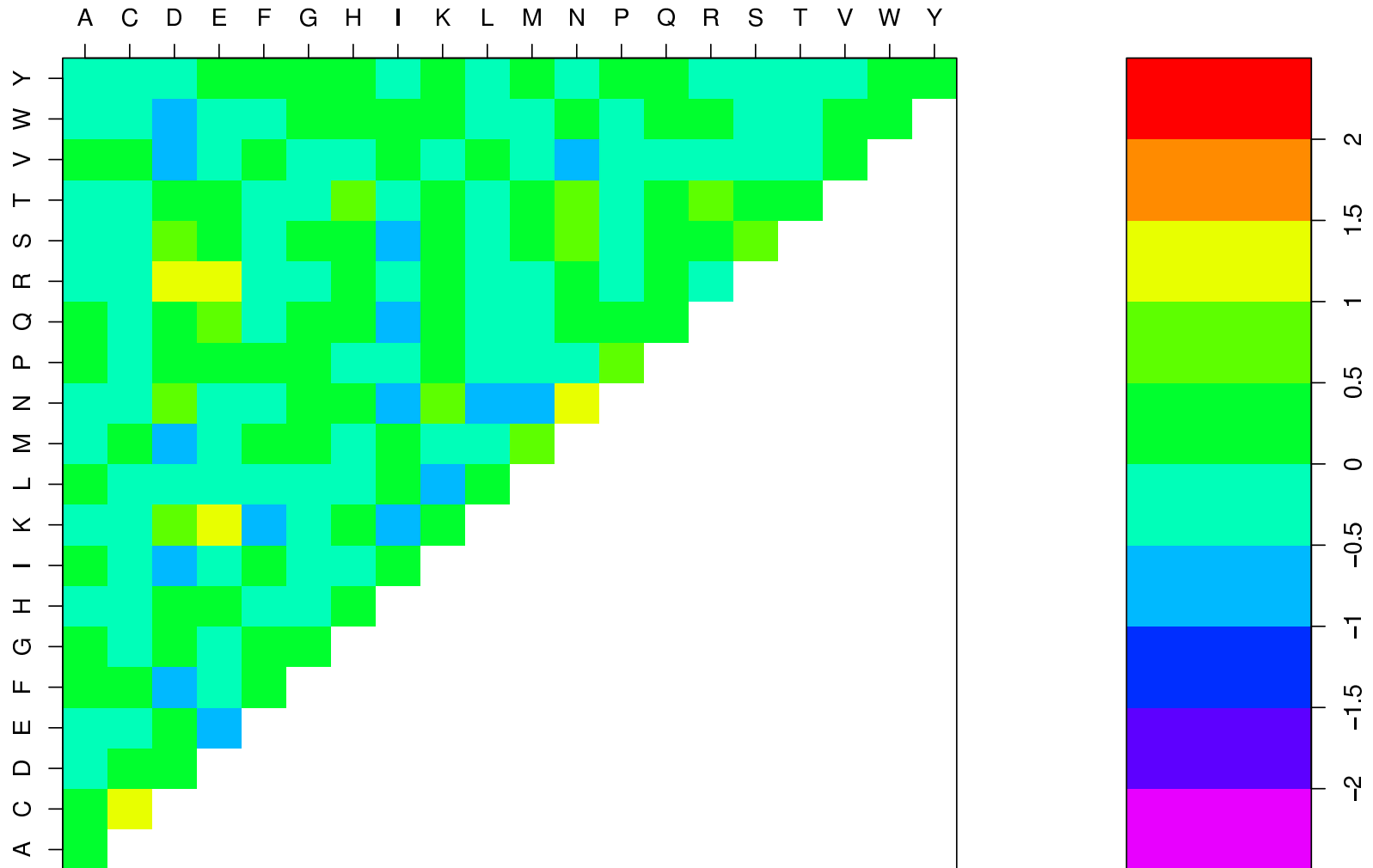
Spécificité de paires anti



Scores Para



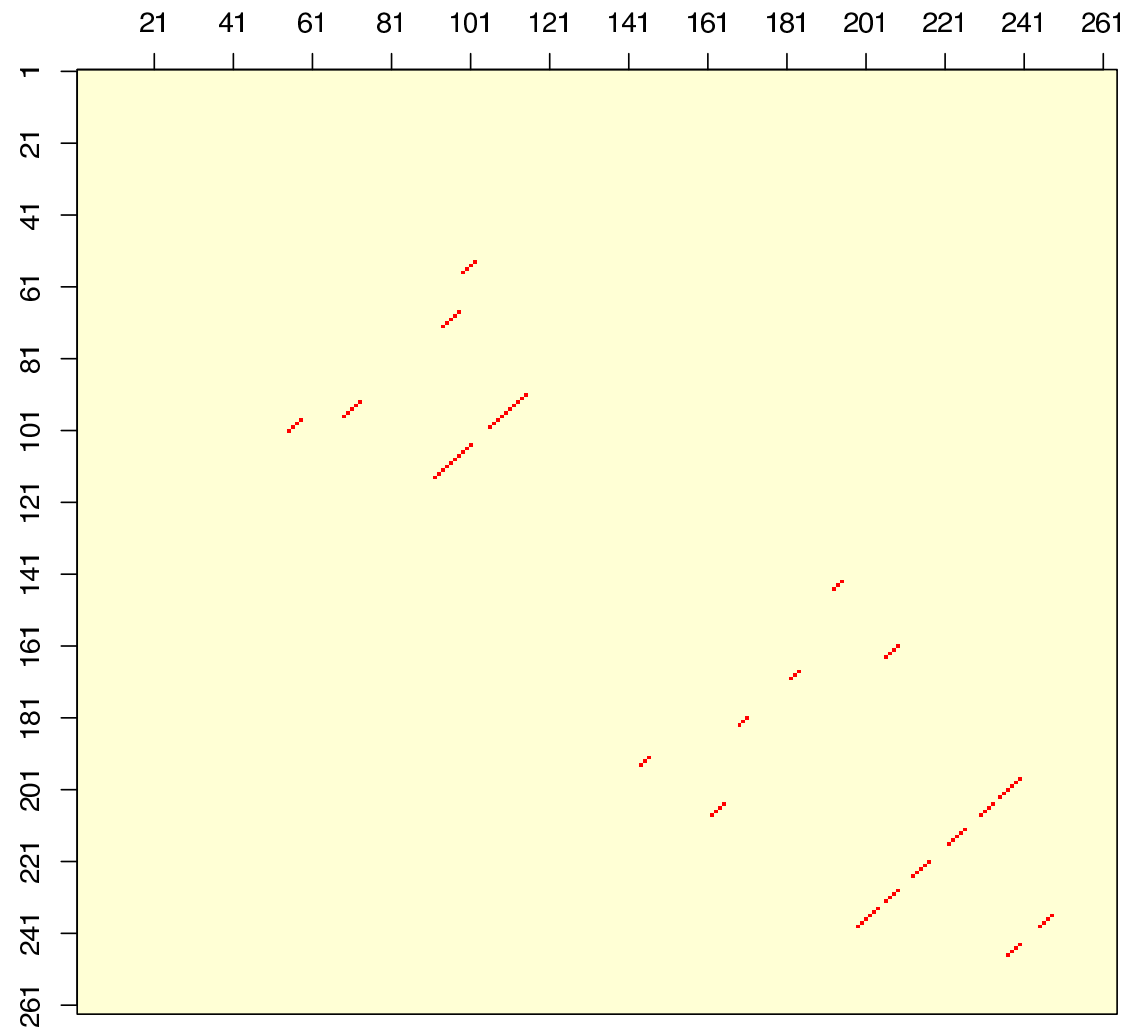
Spécificité de paires para



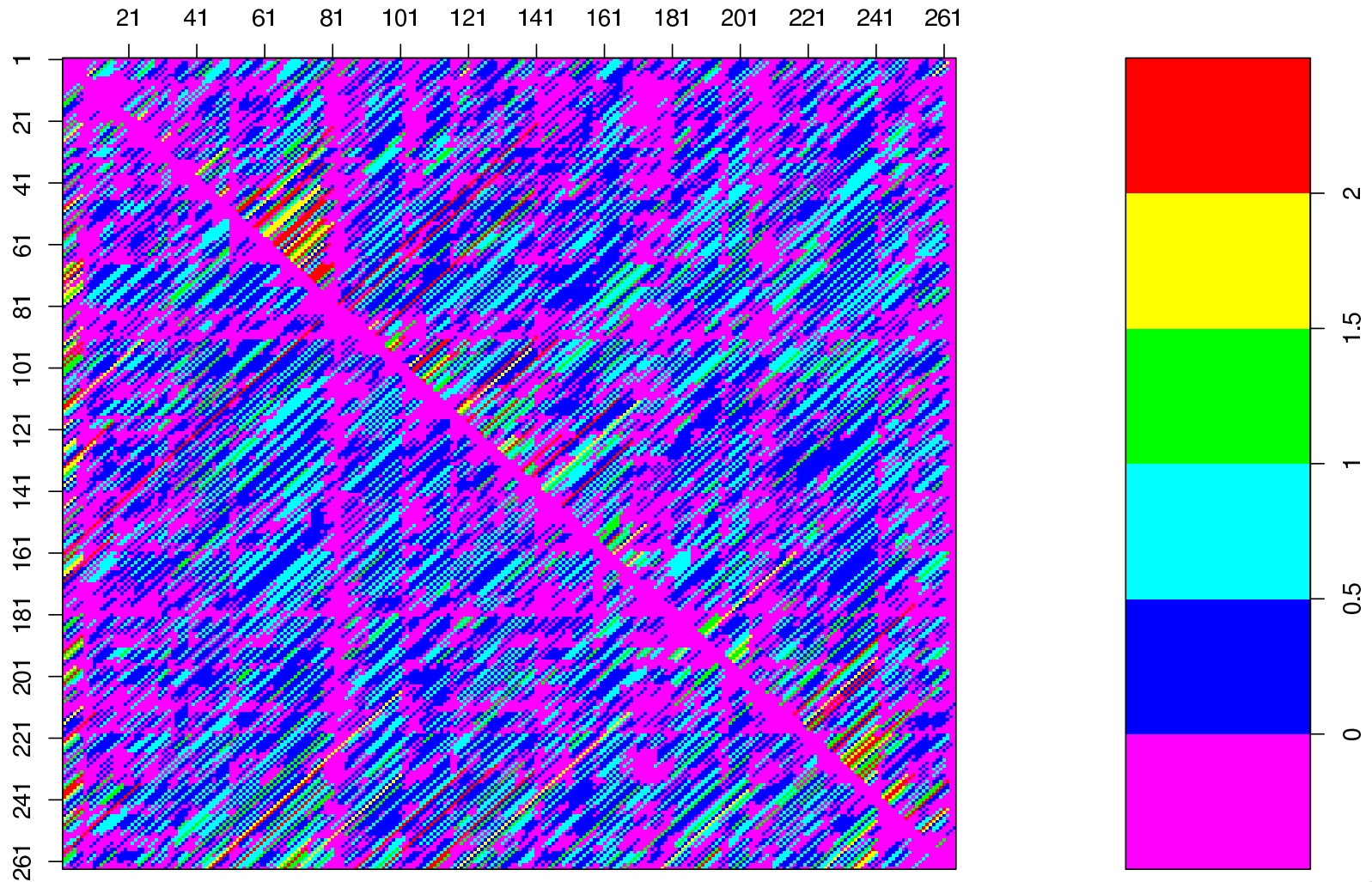
Recherche d'appariements

- Utilisation comme en programmation dynamique, sans gaps
- garder les diagonales >3 .

Exemple 1c22A : contacts

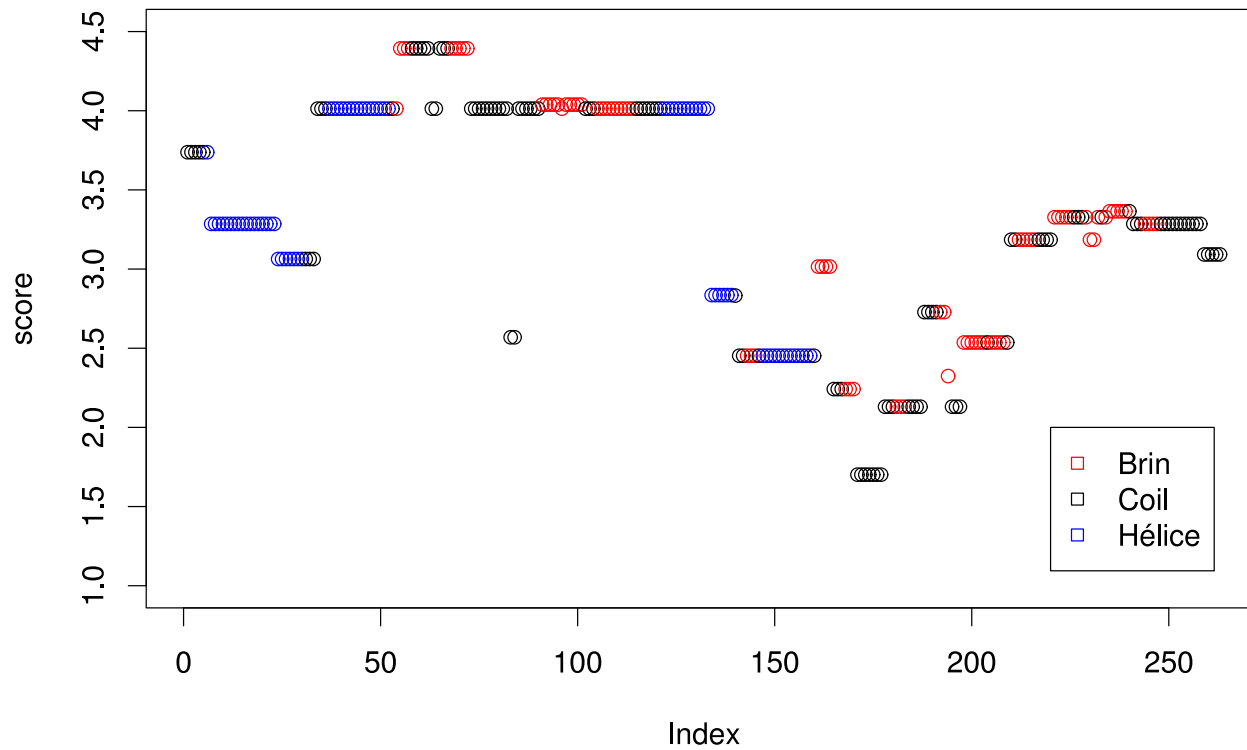


Exemple 1c22a : diagonales détectées

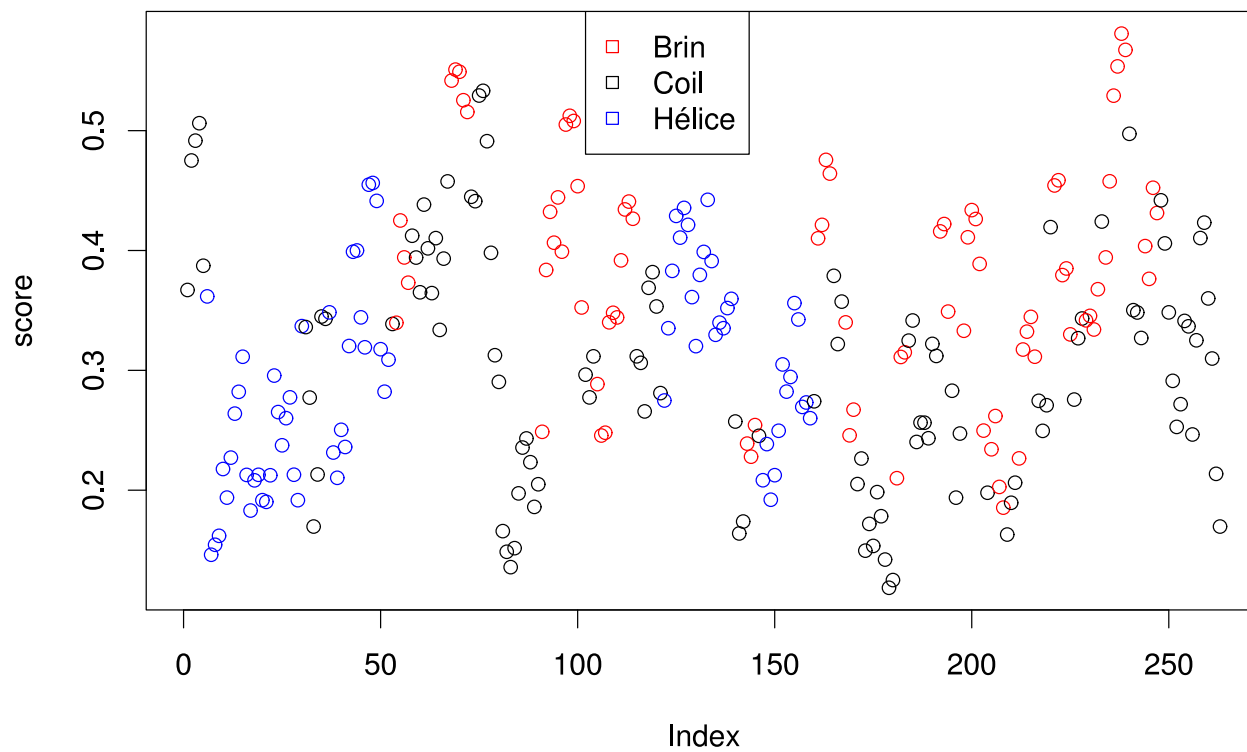


Réduction de la matrice en séquence 1D

Deux schémas envisagés
1 Score max

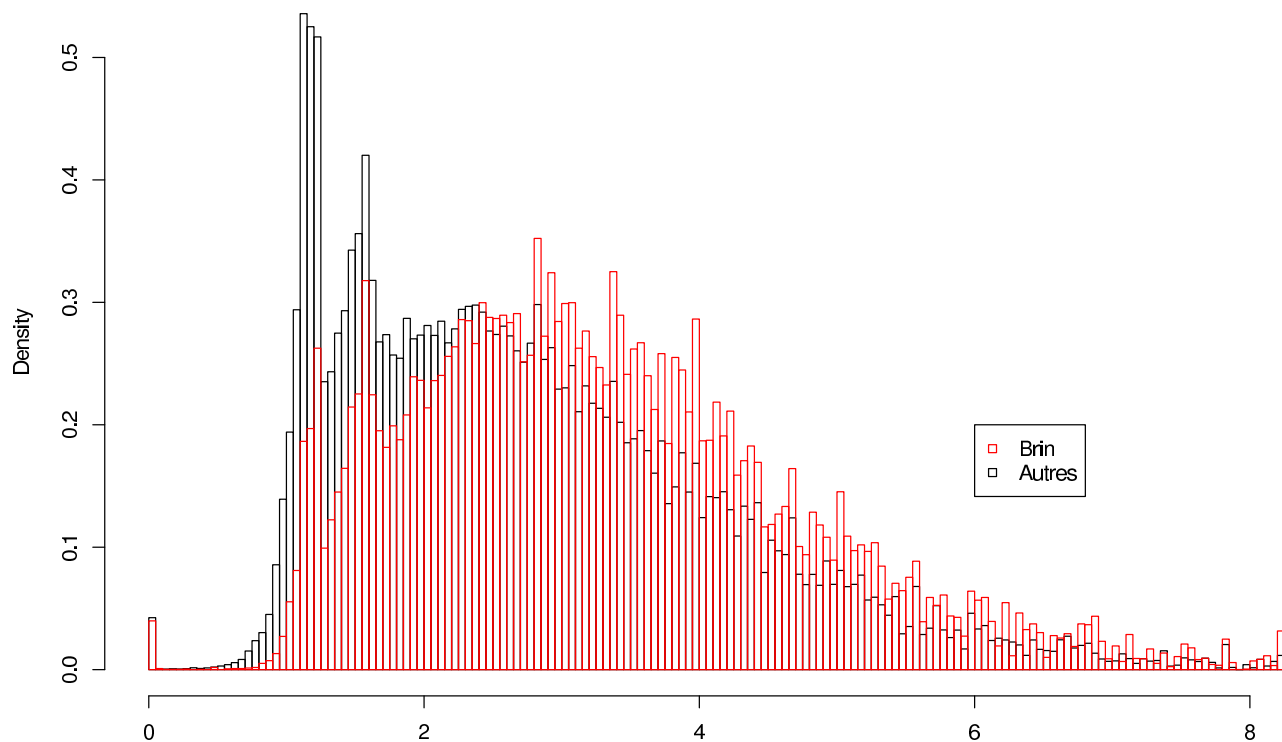


2 Score moyen

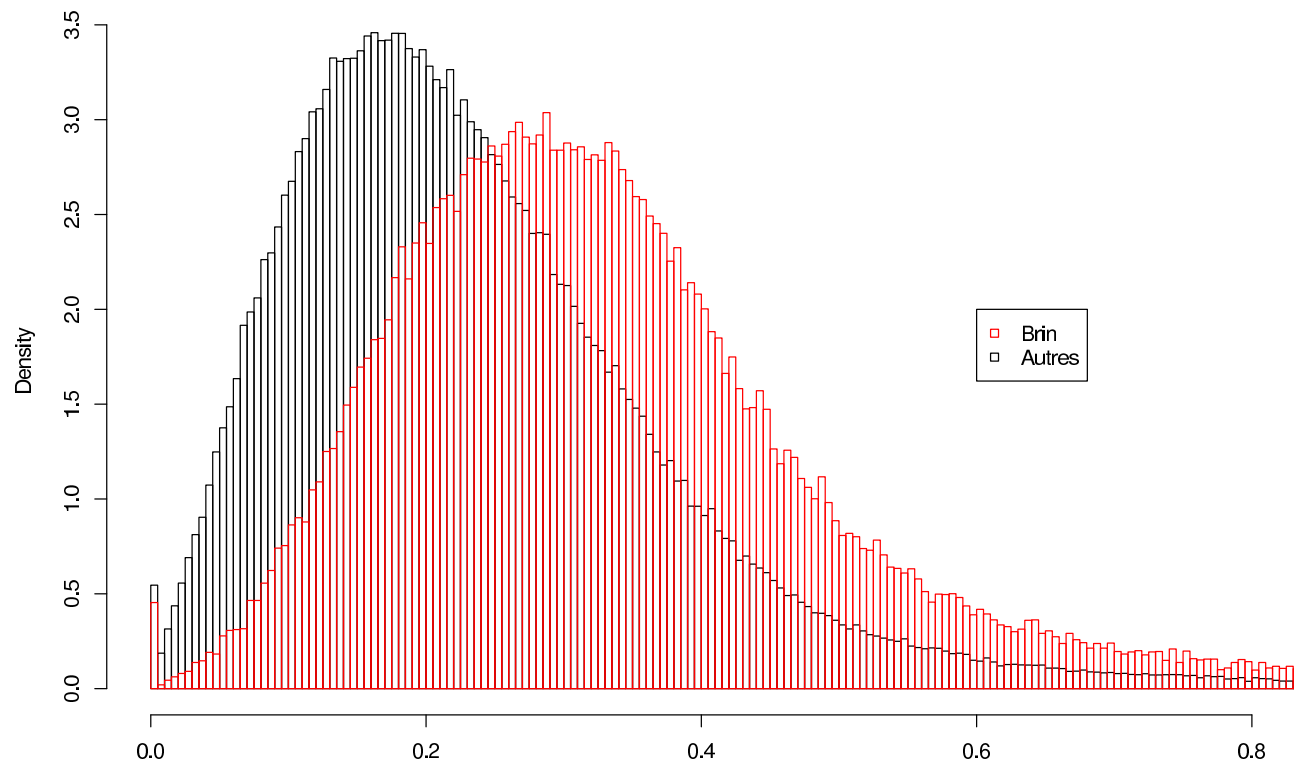


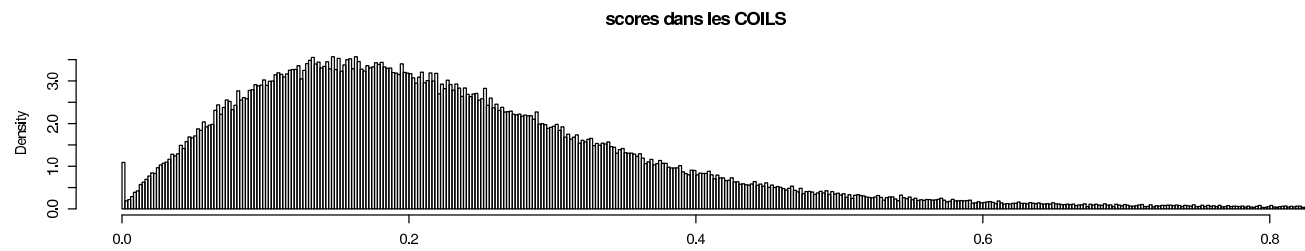
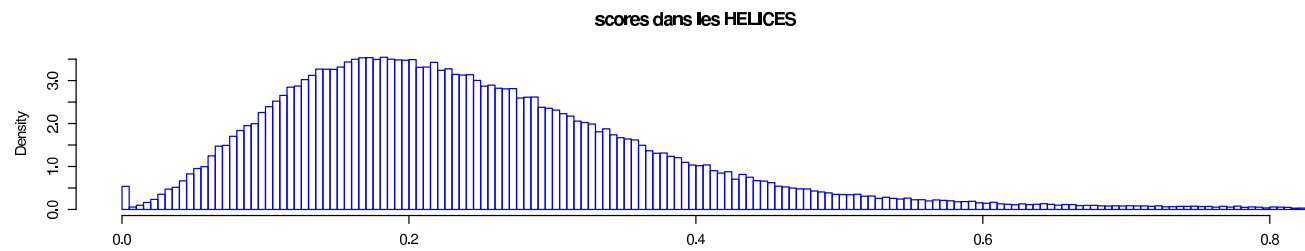
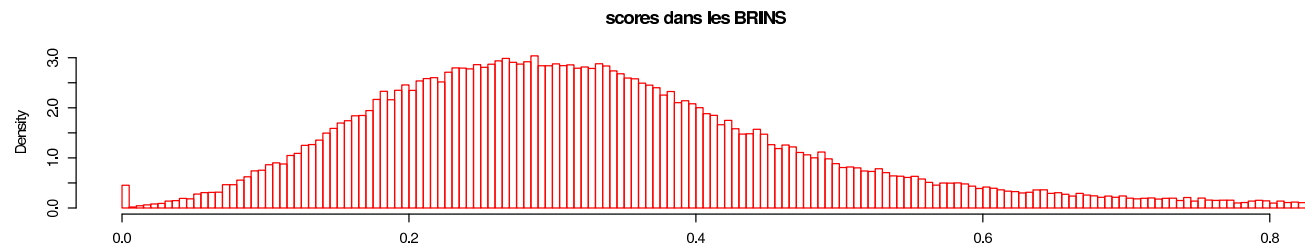
Discrimination des score 1D

Scores Max : séparation : 0.3 sd



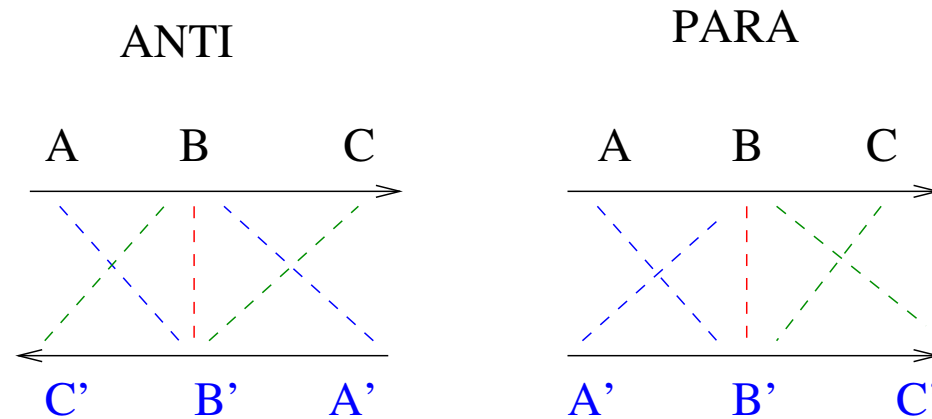
2 Scores Moyens : séparation : 0.5 sd





Fonction de score plus compliquée

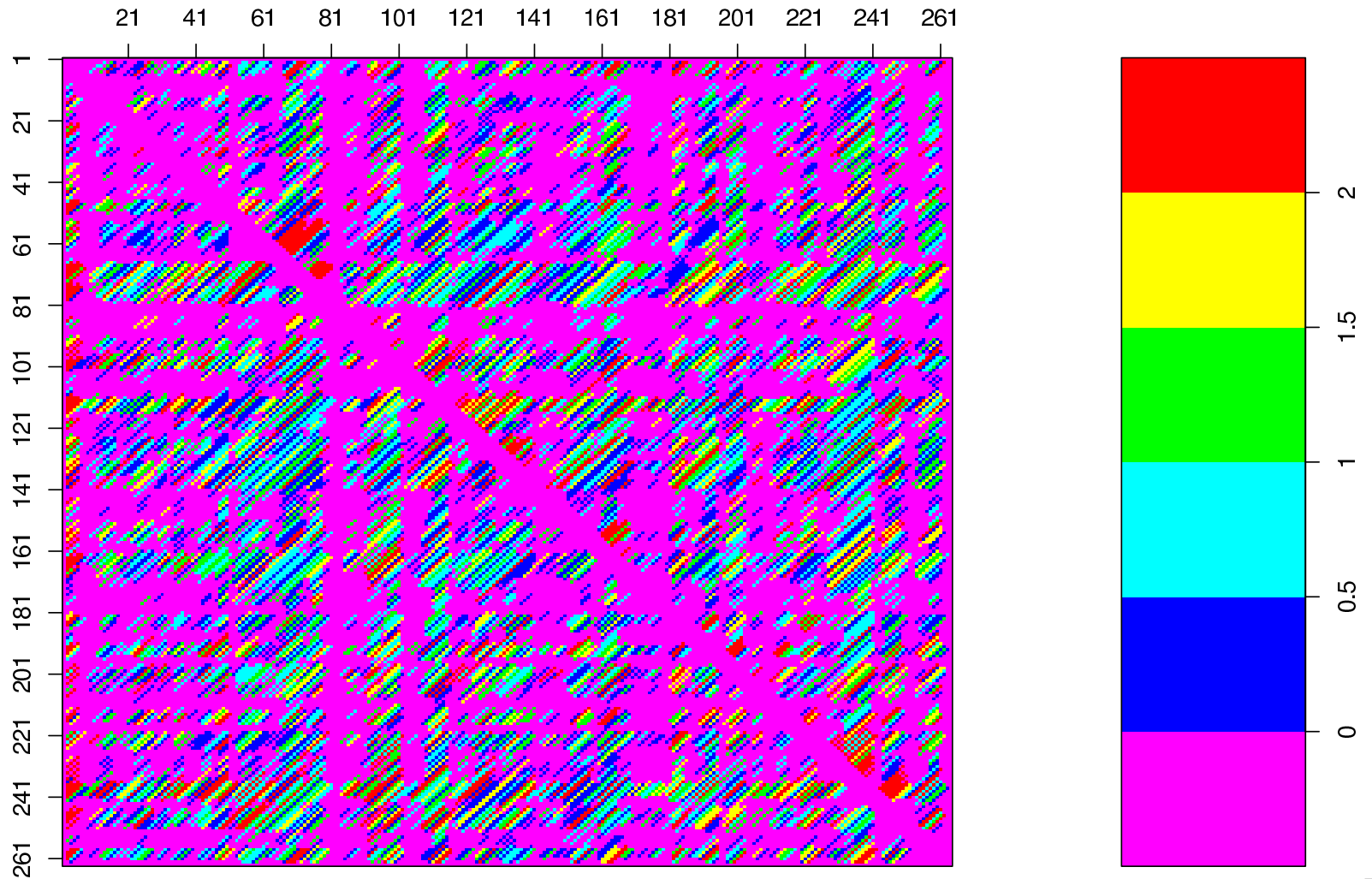
On cherche à incorporer l'information apportée par les résidus voisins, en calculant la probabilité de triplets appariés dans des brins β .



$$P(abc - a'b'c') = P(aa') \times P(a/b') \times P(a'/b) \times P(c/b') \times P(c'/b)$$

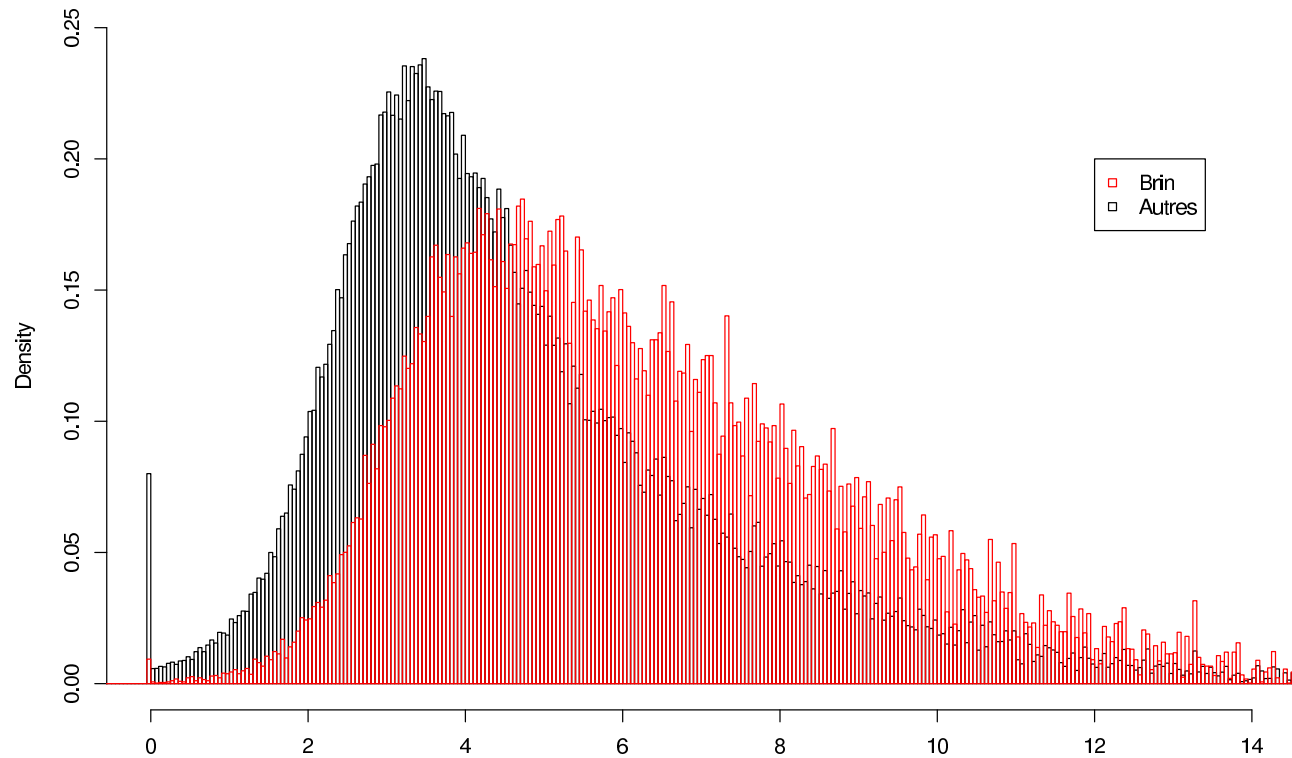
P attendue : Markov d'ordre 1.

Exemple



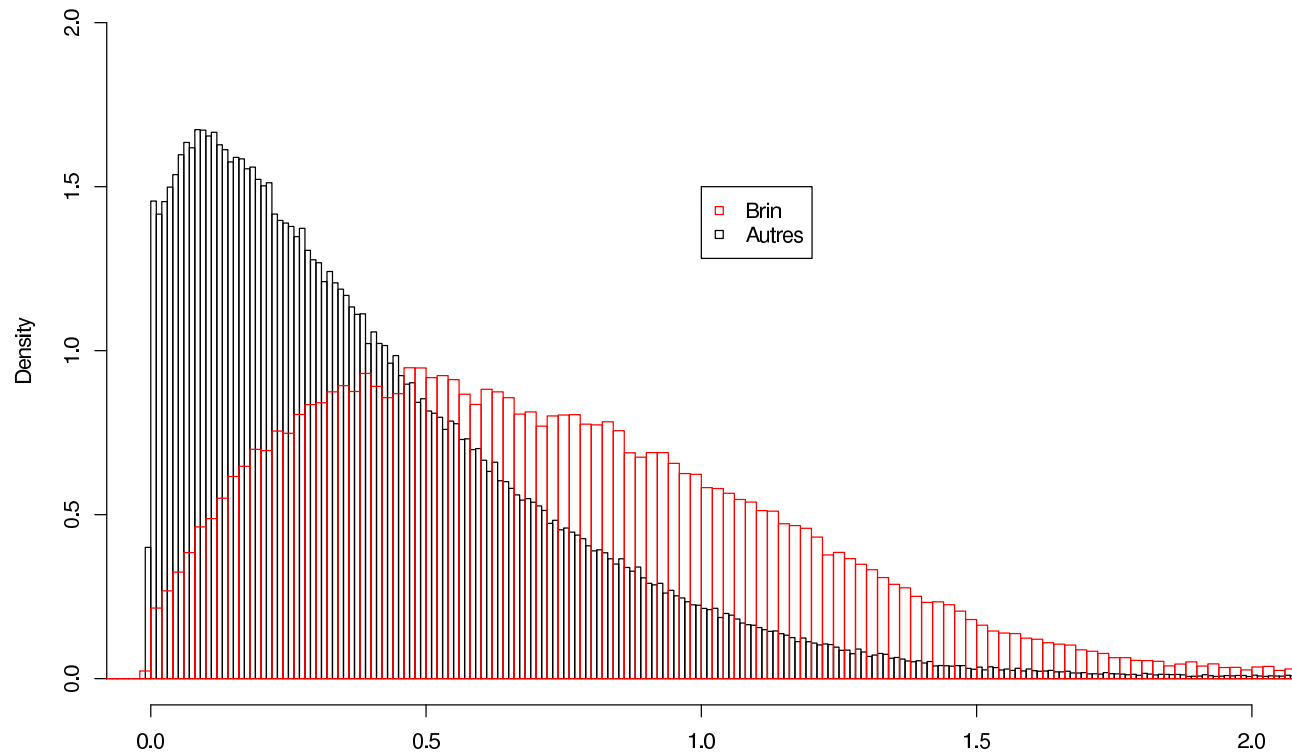
Séparation des scores

Scores MAX : séparation : 0.55 sd



Séparation des scores

Scores Moyen : séparation : 0.66 sd



Utilisation en prédiction

Deux options

1. soit on ré-estime l'ensemble du modèle (c'est à dire les observations d'acides aminés et de score)
2. soit on garde la partie "acide aminés" fixe et on estime seulement la partie score du modèle.

Dans les deux cas, dans la phase de prédiction, on donnera au modèle la séquence et les scores.

- Option 1 : 48.5% de bonne prédiction, proportions : 75% α /15% β /10% coil
- Option 2 : 55.6% de bonne prédiction, proportions :30/33/37.

Normalisation des scores

- normalisation 1

$$score_{norm} = \frac{score - \mu(AA)}{sd(AA)}$$

- normalisation 2

$$score_{norm} = \frac{score - \mu(AA, H)}{\mu(AA, b) - \mu(AA, coil)}$$

- normalisation 3

$$score_{norm} = \frac{score - \mu(AA, H)}{\mu(AA, b) - \mu(AA, H)}$$

équivalents en terme de séparation.

Essai avec la normalisation 3

- 54.4% de prédiction correcte si estimation totale, proportions 60/20/20.
- 56.9% de prédiction correcte si estimation partielle, proportions 36/30/33.

Correction des probabilités a posteriori

Formule de Bayes :

$$P(\textit{structure}/\textit{score}) = \frac{P(\textit{score}/\textit{structure}) \times P(\textit{structure})}{P(\textit{score})}$$

ou $P(\textit{structure})$ est la probabilité fournie par le modèle, et $P(\textit{score}/\textit{structure})$ et $P(\textit{score})$ facilement accessibles d'après les distributions calculées.

Performance de 64% et des proportions 40/23/37.