

RECONSTRUCTION DE STRUCTURES PROTÉIQUES À PARTIR D'UN ALPHABET STRUCTURAL ET D'UN ALGORITHME GLOUTON.

Maupetit Julien¹, Derreumaux Philippe², Tufféry Pierre¹

¹Equipe de Bioinformatique Génomique et Moléculaire, INSERM U726, Université Paris 7, case courrier 7113, 2 Place Jussieu, 75251 Paris Cedex 05

²Laboratoire de Biochimie Théorique, UPR 9080 CNRS, Institut de Biologie Physico-Chimique, 13 rue Pierre et Marie Curie, 75005 Paris



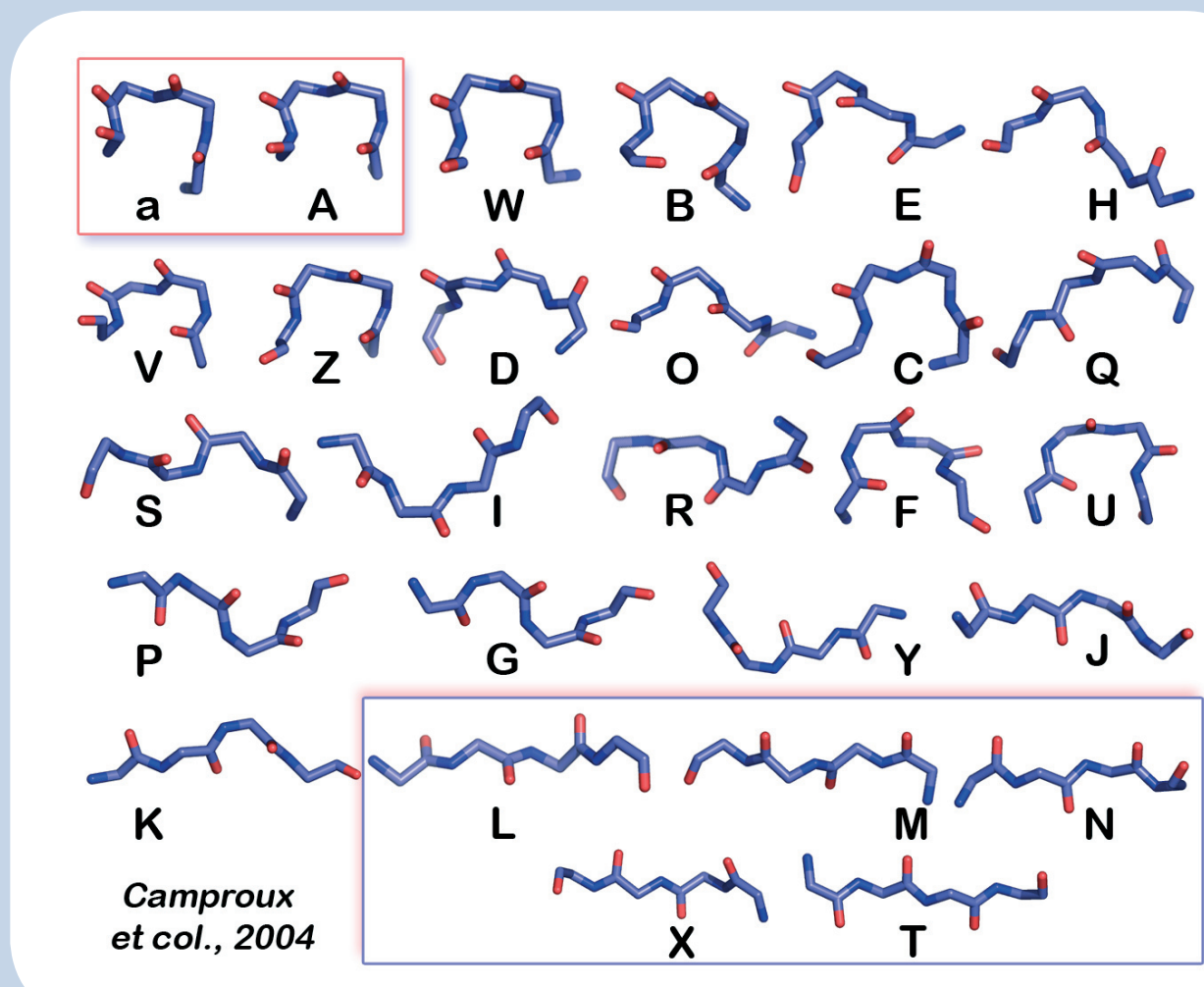
OBJECTIF

Nous présentons ici une méthode de reconstruction de modèles protéiques basée sur la notion d'alphabet structural. Cette méthode de reconstruction s'inscrit dans le cadre d'un projet de prédiction *ab initio* de la structure des protéines. Notre approche se déroule en 2 étapes : (i) prédiction de la séquence HMM-SA à partir de la séquence en acides aminés, et (ii) reconstruction tridimensionnelle de la séquence prédite.

HMM-SA^[1]

HMM Structural Alphabet

- Développé à partir de chaînes de Markov cachées sur un ensemble de 1429 structures extraites de la PDB
- Descripteurs basés sur les positions des carbones alpha
- 1 lettre = 1 fragment protéique de 4 résidus
- Chevauchant sur 3 résidus
- 27 lettres (prototypes moyens)
- 155 prototypes

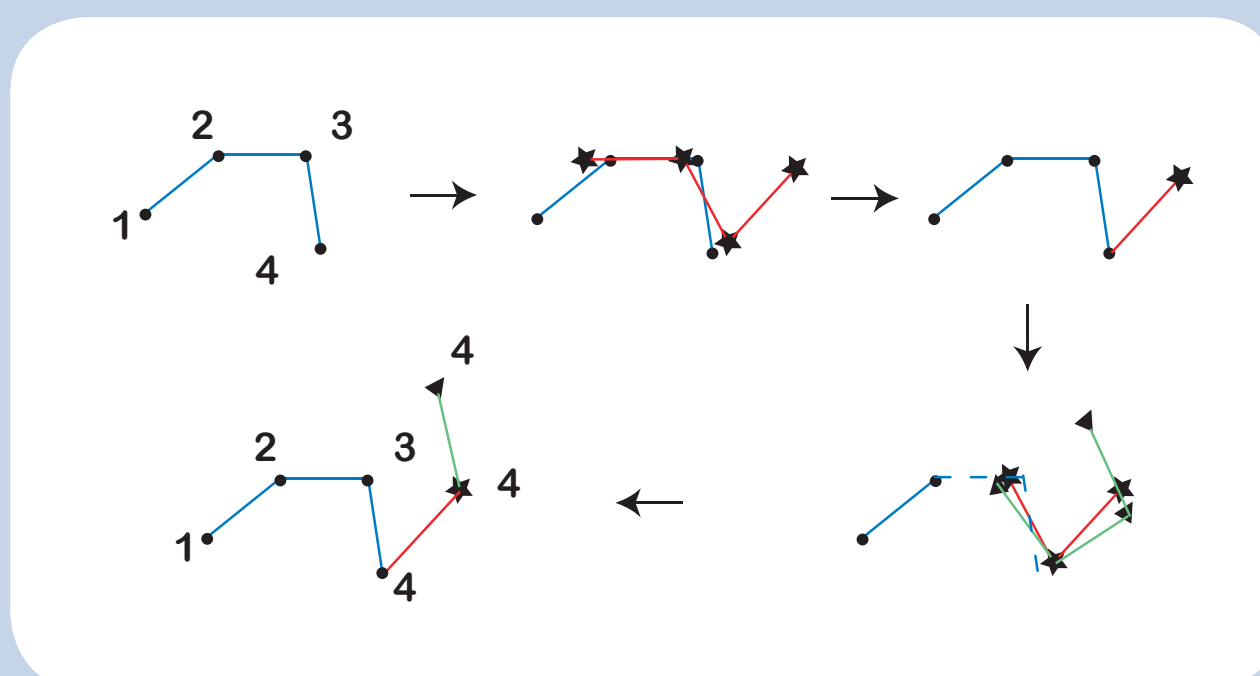


RECONSTRUCTION^[2]

1. Méthode :

• Assemblage :

- Concaténation itérative des fragments
- Superposition des 3 premiers carbones alpha du fragment à ajouter sur les 3 derniers carbones alpha du fragment précédemment ajouté (cas d'une croissance du N vers le C terminal).



• Algorithme glouton :

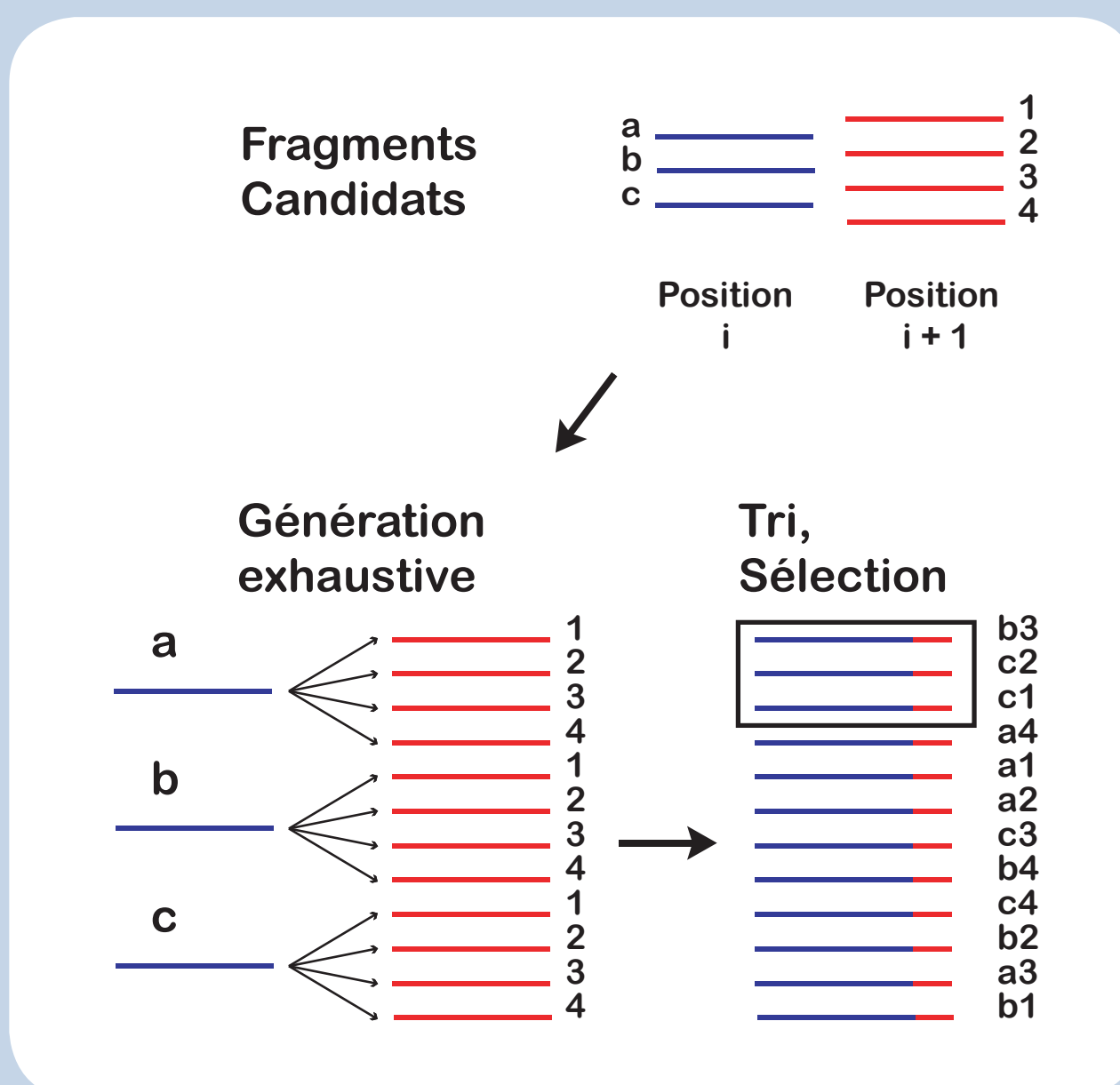
- Génération exhaustive des combinaisons de prototypes entre les positions *i* et *i+1*.
- Sélection des *k* meilleures (formant la Pile) triées selon le critère de recherche.

• Améliorations :

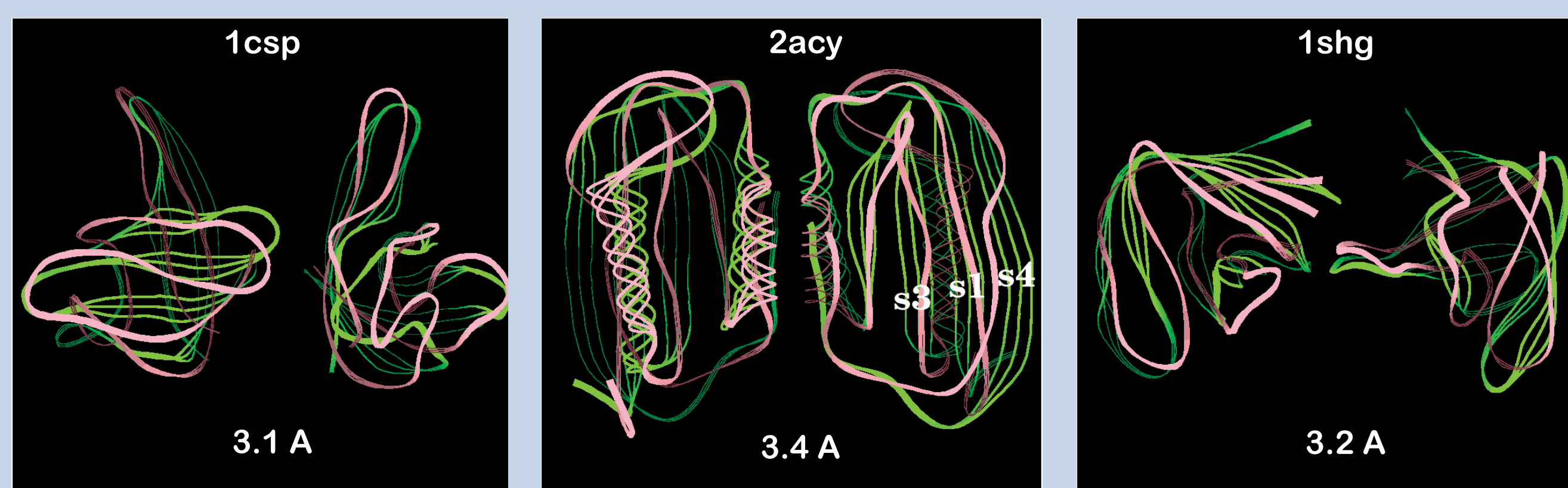
- Stochastique
- Itéré
- Préfiltré

• Critère(s) de recherche :

- Validation de l'algorithme : RMSd, puis un potentiel de type Go
- Reconstruction *ab initio* : potentiel moléculaire simplifié, sOPEP



2. Reconstruction en potentiel de type Go :

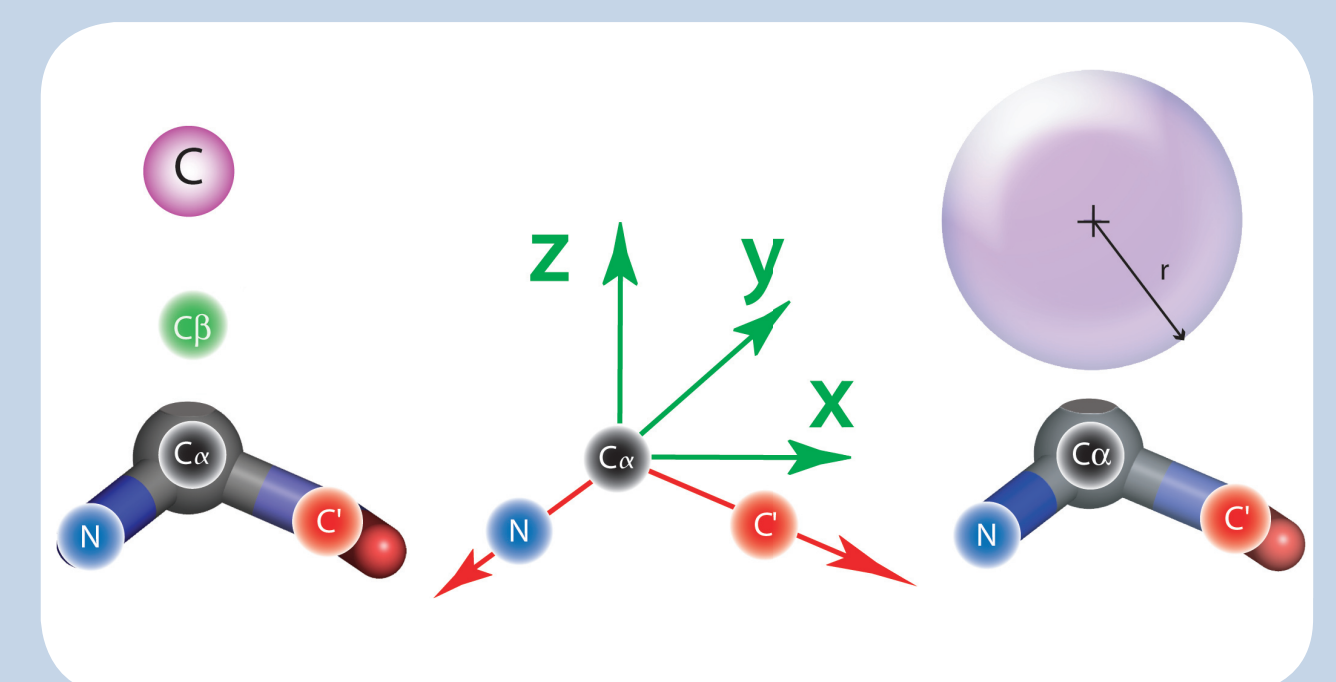
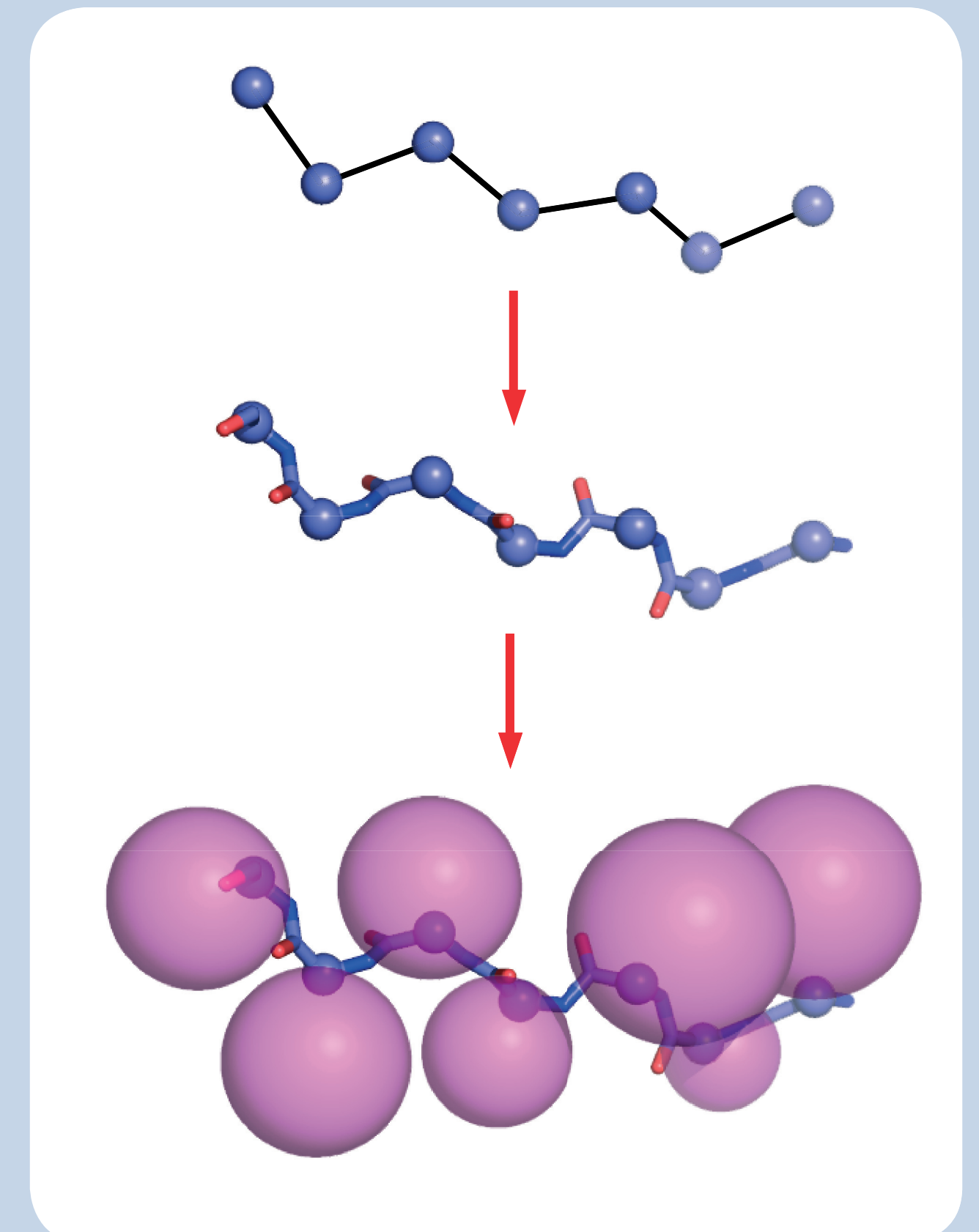


ID	E_{ref}	SF-R2-p	E	SB-R2-p	E	Best	E
1csp	-78			3.14	-75	3.14	-75
1ctf	-79			2.02	-79	1.16	-78
1pgb	-62			2.19	-62	2.18	-62
1shg	-73			3.21	-72	3.21	-72
1ubq	-95	1.86	-95			1.86	-95
2acy	-130					3.41	-117
2bby	-65			2.04	-65	2.04	-65
2ci2	-70					1.51	-70
2ife	-112	2.68	-111			2.68	-111
6pti	-53					1.93	-53

sOPEP

Le champ de forces, que nous utilisons pour évaluer les structures reconstruites, est basé sur le potentiel moléculaire OPEP (*Optimized Potential for Efficient peptide structure Prediction*) développé par Philippe Derreumaux [3]. Dans ce potentiel, la chaîne principale est représentée explicitement, tandis que les chaînes latérales sont représentées par une sphère dont le rayon varie selon le rayon considéré. L'implémentation d'un tel champ de forces dans notre méthode de reconstruction (basée sur les carbones alpha) a donc nécessité une adaptation du mode de représentation des structures reconstruites.

Pour ajouter les atomes lourds de la chaîne principale, nous utilisons l'information contenue dans les prototypes de HMM-SA. Une fois ces derniers mis en place, nous plaçons le centroïde (barycentre des atomes lourds de la chaîne latérale) dans un repère local défini pour chaque résidu (voir ci-contre). Le rayon de la sphère correspond à la distance moyenne du centroïde aux atomes lourds de la chaîne latérale du type de résidu considéré.



TERMES ÉNERGÉTIQUES

L'expression de sOPEP est la suivante : $E_{sOPEP} = E_{vdW} + E_{S_2} + E_{H_b} + E_{CL,CL}$. Sont prises en compte les interactions de type Van Der Waals (E_{vdW}), la structuration des modèles (E_{S_2}), l'énergie associée aux structures secondaires (E_{H_b}), et les interactions entre chaînes latérales ($E_{CL,CL}$), via un potentiel de forces moyennes.

AB INITIO

Pour tester la pertinence de l'implémentation d'un tel potentiel dans la méthode de reconstruction, nous avons réalisé les premiers tests de reconstruction *ab initio* de la cible ayant pour code PDB 1PGB.

Cette reconstruction avec une version très préliminaire de sOPEP présente un RMSd de 4.3A (sur les carbones alpha) avec la structure native. La topologie est similaire à une inversion près au niveau des 2 brins terminaux.



CONCLUSIONS - PERSPECTIVES

Notre méthode de reconstruction, basée sur la notion d'alphabet structural et un algorithme glouton amélioré, nous permet de générer des modèles protéiques à partir d'une information plus ou moins floue de la structure à reconstruire. Nous avons mis en place une version préliminaire d'un potentiel de champ force simplifié, nous permettant de nous absoudre de toute information concernant la structure à prédire, hormis sa séquence en acides aminés. Dans le but d'améliorer le pouvoir discriminant du potentiel, nous avons entrepris d'optimiser ce dernier par l'intermédiaire d'un algorithme génétique, dans lequel, chaque chromosome constitue un vecteur de poids ; chaque poids étant associé à un terme de la fonction d'énergie. Par ailleurs, cette optimisation nécessite la génération de leurres (*decoys*), problème sur lequel nous sommes actuellement en train de nous pencher.

A plus long terme, en dehors du problème de la reconstruction *ab initio*, notre méthode nous permet d'envisager d'autres applications telles que la reconstruction de boucles ou la reconstruction d'un modèle protéique complet à partir d'une trace en carbones alpha.

RÉFÉRENCES

- [1] Camproux AC, Gautier R, Tufféry P, J. Mol. Biol. (2004), 339(3):591-605
- [2] Tufféry P, Guyon F, Derreumaux P, J. Comput. Chem. (2005), 26(5):506-13
- [3] Derreumaux P, Phys. Rev. Lett. (2000), 85: 206-209