Résumé

Cette thèse de doctorat s'inscrit dans le cadre de l'étude *in silico* des relations qui lient la séquence d'une protéine à sa structure, sa stabilité et sa fonction. Elle a pour objectif de permettre à terme la conception rationnelle de protéines modifiées qui restent actives dans des conditions physico-chimiques non physiologiques. Nous nous sommes plus particulièrement penchés sur la stabilité thermique des protéines, qui est définie par leur température de fusion T_m au-delà de laquelle leur structure n'est thermodynamiquement plus stable. Notre travail s'articule en trois grandes parties : la recherche de facteurs favorisant la thermostabilité des protéines parmi des familles de protéines homologues, la mise sur pied d'une base de données de protéines de structure et de T_m déterminées expérimentalement, de laquelle sont dérivés des potentiels statistiques dépendant de la température, et enfin la mise au point de deux outils bioinformatiques visant à prédire d'une part la T_m d'une protéine à partir de la T_m de protéines homologues et d'autre part les changements de thermostabilité d'une protéine (ΔT_m) engendrés par l'introduction d'une mutation ponctuelle.

La première partie a pour objectif l'identification des facteurs de séquence et de structure (*e.g.* fréquence de ponts salins, d'interactions cation-π ...) responsables des différentes stabilités thermiques de protéines homologues au sein de huit familles (chapitre 2). La spécificité de chaque famille ne nous a pas permis de généraliser l'impact de ces différents facteurs sur la stabilité thermique des protéines. Cependant, cette approche nous a permis de constater la multitude de stratégies différentes suivies par les protéines pour atteindre une plus grande thermostabilité.

La deuxième partie concerne le développement d'une approche originale pour évaluer l'influence de la température sur la contribution de différents types d'interactions à l'énergie libre de repliement des protéines (chapitres 3 et 4). Cette approche repose sur la dérivation de potentiels statistiques à partir d'ensembles de protéines de thermostabilité moyenne distincte. Nous avons d'une part collecté le plus grand nombre possible de protéines de structure et de T_m déterminées expérimentalement, et d'autre part développé des potentiels tenant compte de l'adaptation des protéines aux températures extrêmes au cours de leur évolution. Cette méthode originale a mis en évidence la dépendance en la température d'interactions protéiques tels les ponts salins, les interactions cation-π, certains empilements hydrophobes … Elle nous a en outre permis de mettre le doigt sur l'importance de considérer la dépendance en la température non seulement des interactions attractives mais également des interactions répulsives, ainsi que sur l'importance de décrire la résistance thermique par la T_m plutôt que la T_{env}, température de l'environnement de l'organisme dont elle provient (chapitre 5).

La dernière partie de cette thèse concerne l'utilisation des profils énergétiques dans un but prédictif. Tout d'abord, nous avons développé un logiciel bioinformatique pour prédire la thermostabilité d'une protéine sur la base de la thermostabilité de protéines homologues. Cet outil s'est avéré prometteur après l'avoir testé sur huit familles de protéines homologues. Nous avons également développé un deuxième outil bioinformatique pour prédire les changements de thermostabilité d'une protéine engendrés par l'introduction d'une mutation ponctuelle, en s'inspirant d'un logiciel de prédiction des changements de stabilité

thermodynamique des protéines développé au sein de notre équipe de recherche. Ce deuxième algorithme de prédiction repose sur le développement d'une grande base de données de mutants caractérisés expérimentalement, d'une combinaison linéaire de potentiels pour évaluer la ΔT_m, et d'un réseau de neurones pour identifier les coefficients de la combinaison. Les prédictions générées par notre logiciel ont été comparées à celles obtenues via la corrélation qui existe entre stabilités thermique et thermodynamique, et se sont avérées plus fiables.

Les travaux décrits dans notre thèse, et en particulier le développement de potentiels statistiques dépendant de la température, constituent une nouvelle approche très prometteuse pour comprendre et prédire la thermostabilité des protéines. En outre, nos travaux de recherche ont permis de développer une méthodologie qui pourra être adaptée à l'étude et à la prédiction d'autres propriétés physico-chimiques des protéines comme leur solubilité, leur stabilité vis-à-vis de l'acidité, de la pression, de la salinité … lorsque suffisamment de données expérimentales seront disponibles.