



10. ÜBUNG ZUR BIOINFORMATIK II, SS 2012

Aufgabe 1: Monte Carlo Docking II (40 Punkte)

Auf dem letzten Übungsblatt haben Sie einen einfachen Docking-Algorithmus basierend auf Simulated-Annealing und Kraftfeldevaluationen implementiert. Da Kraftfelder sehr sensitiv gegenüber Überlappungen sind, eignen Sie sich nicht unbedingt zur ersten Bewertung beim Docking. Auf diesem Übungsblatt sollen Sie das Kraftfeld durch eine einfache, aber robuste Scoring-Funktion austauschen. Seien $u, l, k_1, k_2 \in \mathbb{R}^+$ und A, B Mengen von Atomen, dann sind $\text{overlap}(A, B)$ und $\text{contact}(A, B)$ definiert als:

$$\begin{aligned}\text{contact}(A, B) &= |\{(a, b) \in A \times B \mid l \leq \|a - b\| < u\}| \\ \text{overlap}(A, B) &= |\{(a, b) \in A \times B \mid \|a - b\| < l\}|\end{aligned}$$

und schließlich:

$$\text{geom_score}(A, B) = k_1 \cdot \text{contact}(A, B) - k_2 \cdot \text{overlap}(A, B)$$

- (a) Implementieren Sie $\text{geom_score}(A, B)$. Achten Sie darauf, die benachbarten Atome möglichst effizient bestimmen zu können (*Hinweis*: Hashgrid!).
- (b) Vergleichen Sie die Scoring-Funktion mit den Ergebnissen des vorherigen Übungsblattes (ohne lokale Nachoptimierung). Verwenden Sie hierzu die Parameter $u := 4.0$, $l := 2.75$, $k_1 := 1.0$, $k_2 := 4.0$. Passen Sie die Akzeptanzwahrscheinlichkeit p wie folgt an:

$$p = \exp\left(\frac{-\Delta E}{0.01 \cdot T}\right)$$

Erhöhen Sie außerdem die Anzahl an inneren Schleifendurchläufen auf 500. Fügen Sie keine Wasserstoffe zu den beiden Proteinen hinzu.

- (c) Berechnen Sie die RMSD zwischen gedocktem und experimentellem Komplex. Berechnen Sie den Mittelwert und die Varianz der RMSD über mehrere Läufe. Beachten Sie, dass Sie für jeden Lauf einen neuen Random-Seed verwenden. Optimieren Sie Ihre Parameter, sodass Sie eine möglichst kleine RMSD erhalten. Dokumentieren Sie die Ergebnisse, die Sie für die unterschiedlichen Parameter erhalten.

Abgabe: Donnerstag, 28.06.2012, 23:59 Uhr