

文献調査及び PSA/GAc における選択・交叉手法の検討
青井 桂子

1 文献調査

実践バイオインフォマティクスのシーケンス解析、ペアワイズアラインメント、データベースサーチの文献調査を行った。DNA やタンパク質の配列を解析するシーケンス解析技術はバイオインフォマティクス分野の最も重要な技術の 1 つであり、シーケンスの文字間の最適な対応関係を求めることで解析できる。シーケンスの配列比較は、2 つのシーケンスのうち、完全に一致している部分、類似の塩基またはアミノ残基である部分、ギャップ部分をスコア化することで行われる。

2 PSA/GAc の交叉後の探索点の選択手法

これまで、PSA/GAc の交叉後の探索点は最良 2 個体を選択していたが、この選択方法の有効性は検証されていなかった。このため、本実験では交叉後に選択する個体をそれぞれ”ベスト 2 個体 (従来手法)”、”ベスト 1 個体とランダム 1 個体”、”ランダム 2 個体”、”子 2 個体”、”PSA”の 5 パターンの実験を行い、有効な選択手法を検討した。実験で用いたパラメータは、16 個体 × 6000MCsweep とし、各々のタンパク質で最適な交叉間隔、温度パラメータを用いた。Fig. 1 に各タンパク質に対して 5 パターンの選択手法を用いた時の最小エネルギーの平均値と標準偏差を示す。

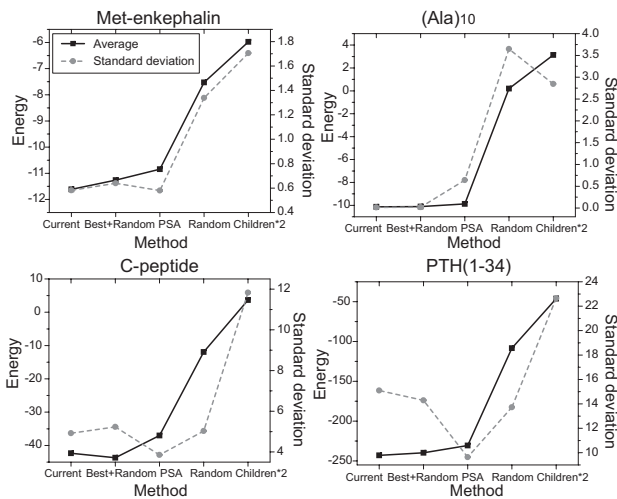


Fig. 1 エネルギーの平均値と標準偏差

Fig. 1 より、従来の”ベスト 2 個体”と、”ベスト 1 個体とランダム 1 個体”の場合に、いずれのタンパク質においても良好な結果が得られている。

3 PSA/GAc の交叉手法

これまで、PSA/GAc の交叉手法は設計変数間でのランダムな 1 点交叉を用いてきた。本実験で提案する交叉手法は Fig. 2 に示すように、主鎖の設計変数間で交叉点を決め、側鎖の部分は主鎖に付随した形で交叉を行う。

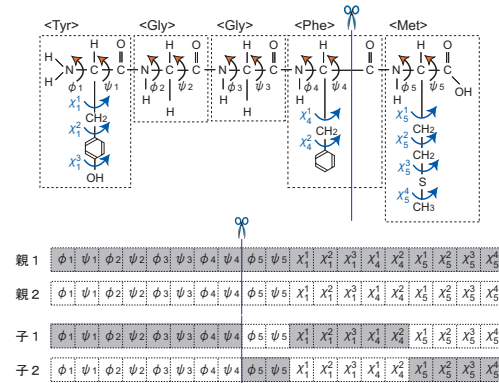


Fig. 2 提案する交叉手法

実験には、従来の全設計変数間で 1 点交叉 (Current)、主鎖の設計変数間で 1 点交叉 (ver.1)、主鎖の設計変数間で 1 点交叉し、側鎖は付随させる (ver.2)、主鎖の設計変数間で 2 点交叉し、側鎖は付随させる (ver.3)、PSA の 5 パターンで比較を行った。また、前節の選択手法のうち、良好な結果を得た 2 パターンについて実験を行った。実験に用いたパラメータは前節の実験と同様である。Fig. 3 に最小エネルギーの平均値を示す。

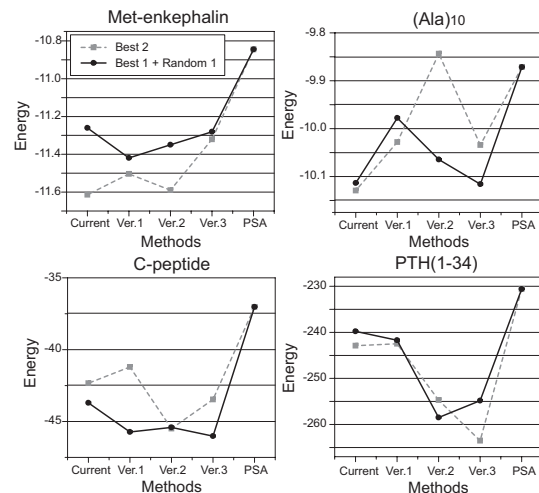


Fig. 3 エネルギーの平均値

Fig. 3 より、提案した交叉手法はより大規模なタンパク質に対して、良好な結果を示すことがわかった。また、側鎖を付随させた交叉手法は通常交叉手法に比べて良好な結果が得られることがわかった。