

蛋白質立体構造予測システム (PE-T)

(PE-T: Protein Engineering Support System)

ホモロジーモデリングの手法に基づいて独自に開発した「距離束縛法 (Distance - Constraint Approach)」(1)によりC α 原子座標を予測します。

● PE-Tの3つの特徴

計算手法がとてもシンプルにして且つ迅速な解析が可能

- ・参照蛋白質とホモロジーのない領域の座標も自動計算するので煩雑な手続きが不要

予測精度の向上

- ・参照蛋白質とのホモロジーが30%~40%のもので~3台の精度が得られる(図1参照)。

予測精度も推定が可能

- ・X線結晶解析による構造が未知のままでもアミノ酸残基ごとの予測精度の推定が可能(図2参照)。

● 距離束縛法とは?

今、目的とする(つまり、アミノ酸配列は既知であるがそのC α 座標は未知の)n残基からなる蛋白質のアミノ酸残基C α の3n個の座標、

$$(x, y, z) = \{(x_i, y_i, z_i); i = 1, 2, \dots, n\} \quad (1)$$

に関する目的関数を

$$F(x, y, z) = \sum_{i < j} w_{ij} (d_{ij} - \langle d_{ij} \rangle)^2 / m \quad (2)$$

とします。ここで、 d_{ij} は計算されるべきコンフォメーションでのi番目とj番目のアミノ酸残基C α 間の相対距離、 $\langle d_{ij} \rangle$ はPDBデータから得られた20種のアミノ酸間の相対距離に関する統計データ(2, 3)、 w_{ij} は統計的重みで、 m は全残基ペアの数とします。アミノ酸配列が与えられるとこの関数Fは一意的に決まるので、乱数で生成された初期構造からこれを最小にするような座標 $(x, y, z)_{\min}$ を求めることにより目的とする蛋白質のC α 座標を予測します。

しかし、統計データ $\langle d_{ij} \rangle$ では余りに粗い過ぎて得られた構造はnativeなものとは程遠いものとなります。

そこで、本手法では式(2)において $\langle d_{ij} \rangle$ として参照蛋白質の(目的蛋白質の残基に)対応する残基間の"相対距離"を代入するという方法を採用しています(併置配列において、参照蛋白質側に目的蛋白質の残基に対応する残基がない場合は統計データ $\langle d_{ij} \rangle$ が使われることになります)。

図1はこうしてchymotrypsinogen (PDB code: 2cga)を参照蛋白質として得られた elastase (PDB code : 1est)の予測構造です。X線結晶解析と予測構造を最適に重ね合わせた時の偏差は2.95 Å です。図2は予測精度推定プロフィールを示しています。実線は予測精度を推定するプロットで アミノ酸配列データのみから得られた量 $\langle C(i) \rangle$ であることに注意して下さい〔1〕。破線はX線結晶解析データと予測構造を最適に重ね合わせた時の i 番目の残基のC_α の位置の偏差 l_{ii} (Å)です。 $\langle C(i) \rangle$ が大きい領域ほど予測精度が高いことが推定されますが、実際偏差 l_{ii} は小さくなっておりこのことが示めされています。

● 2件の特許を取得！！

特筆すべきことに PEPT の独創性に対して次の2件の特許が与えられており、その道の専門家の注目を集めています：

1. 発明の名称:蛋白質の立体構造の予測演算方法及び予測演算装置
(特許番号第2856306号)
2. 発明の名称:蛋白質の立体構造の予測精度演算方法及び予測精度演算装置
(特許番号第2930851号)

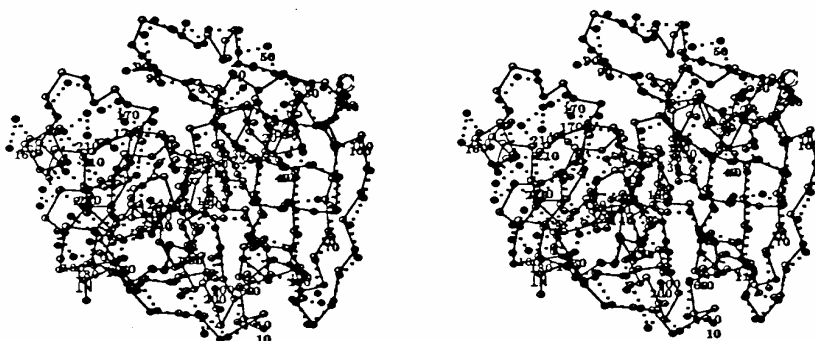


Fig.1 : The stereoviews of the structures predicted (solid line) and determined by X-ray (dotted line) of porcine elastase after optimal superposition of the two conformations. The rmsd between the two structures is 2.95 Å .

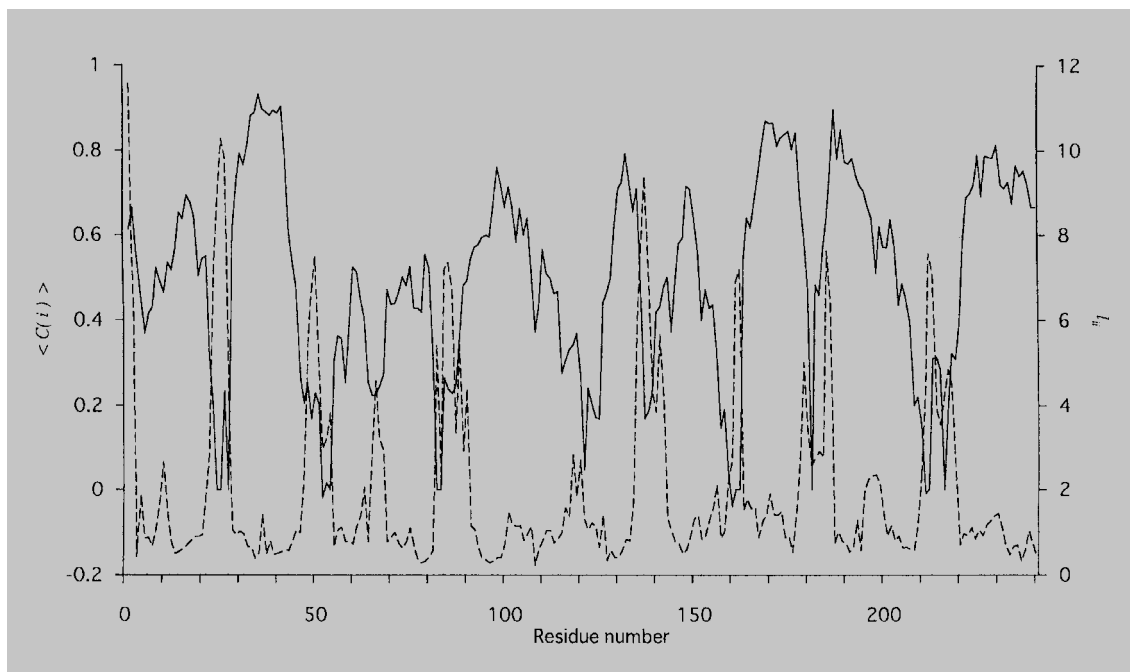


Fig.2 : The correlation profile (solid line) and plots of deviation in position of C atoms of i th residue, l_{ii} , (dashed line).

● C 予測から更なる予測へ向けて！

現在引き続きC など主鎖を予測次いで側鎖を予測することで全原子座標予測システムを構築中です。

参考文献

- [1] Kubota, Y., Higo, J., Sarai, A. and Wako, H. “*Prediction of Main-Chain Conformation of Proteins Based on the Distance-Constraint Approach with the Accuracy Estimation by the Correlation Profile*” (unpublished work)
- [2] Wako, H. and Scheraga, H.A. “*Distance-Constraint Approach to Protein Folding. I. Statistical Analysis of Protein Conformations in Terms of Distances Between Residues*” *J. Protein Chem.* 1, 5. (1982)
- [3] Wako, H. and Scheraga, H.A. “*Distance-Constraint Approach to Protein Folding. II. Prediction of Three-Dimensional Structure of Bovine Pancreatic Trypsin Inhibitor*” *ibid.*, 85. (1982)