**１．本書の位置づけ**

本書は、人工酵素設計PJT（2020年3月～2022年10月）の要素技術の一つである「AlphaFoldによる変異体評価」の調査研究の概要・実施結果について報告を行うものである。



**Confidential**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Date | 2022/10/24 | Description | 新規作成 | | | | | | |
| Issue | 横河電機 MKHQ INVC | | | Author | 熊谷 | Check | 生田目 | Approval | 生田目 |
|  |  | | |  |  |  |  |  |  |
|  |  | | |  |  |  |  |  |  |

Form Y-E31-2　Size A4

**３．２　要素技術の現状**

**３．２．２　変異体の評価**

* AlphaFold2の信頼度スコアによる変異体評価

AutoDock Vinaや分子動力学シミュレーションなどを用いて、CBD変異体を机上評価しているが、これらはセルロース結晶とのドッキングスコアやセルロース結晶表面上のダイナミクスを評価しており、立体構造へ折り畳む妥当性自体を評価しているものではない。一方、AlphaFold2[1]は、2021年に登場した、アミノ酸配列からその立体構造を高精度で予測する技術で、予測構造の信頼度スコア（pLDDT）を計算し、それを高めるように学習する。AlphaFold2の信頼度スコアは、熱力学的なスコアに基づくものではなく、学習データに基づいて統計的な観点で計算されるため、他のスコアとはある程度独立的な観点でアミノ酸配列を評価できる。しかしながら、AlphaFold2はPDBの天然タンパク質を中心に適用されており、設計した変異体への適用例は未だに少ない。ペプチド配列設計に関する研究成果において、アミノ酸の溶解度指標以外に、AlphaFold2の信頼度スコアも同時に最大化することで、標的タンパク質に対する結合親和性が高く、かつ脂溶性が高すぎない候補を効率良く生成できた事例[3]があり、候補設計における信頼度スコアの有用性が期待できるが、CBD変異体に対する事例は確認していない。このため、本項ではトライアルとして、本テーマで設計したCBD変異体の配列について、信頼度スコアを用いて評価することの可能性について簡易的に検証した。

〇〇項で検討した、結合能簡易評価系によって評価した変異体100種類（3点変異、4点変異、5点変異）のアミノ酸配列について、AlphaFold2（AF2）を適用し、その予測構造と各原子位置のpLDDTを計算した。AlphaFold2は、Webブラウザの計算環境（Google Colaboratory）で実行可能なColabFold[2]を使用した。

図3に、4つの3点変異体の予測構造とpLDDTを示す。緑色の構造がRosettaで生成した構造、白色の構造がAF2で予測した構造、水色が結合部位、紫色がWild-Typeから変異した部分である。予測構造の下の図は、各原子位置のpLDDTを色付けしており、青色が濃いほど信頼度が高いことを表す。図3から下記のことが確認できる。

* global pLDDT（全原子のpLDDTの平均値）は90%以上であるため、CBD変異体についてもAF2の予測信頼度は極めて高い。
* 各変異体のC-α原子のRMSDは0.5Åであるため、AF2予測構造は、Wild-Typeの構造やエネルギースコアを使用せずとも、Rosetta構造と似た結果を示している。
* 一方、変異体によっては、構造の一部のpLDDTが薄い水色（80%代）であるため、予測信頼度が低下する部位もある。

さらに、図4に、100個の変異体のスコアとglobal pLDDTの散布図を示す。左図の横軸がRosettaのエネルギー変位スコア（ΔΔGスコア）、右図の横軸がAutoDock VinaのAffinity（結合親和性）である。各データの色は、〇〇項で評価した、簡易評価によるセルロース結合能の強さで、赤色が強い、緑色が弱い、青色が無いことを示している。図4から下記のことが確認できる。

* 左図から、ΔΔGとpLDDTの間には関係性がみられない。
* 一方、右図から、Affinityが低く、pLDDTが高い領域に、結合能が強い変異体が多く位置する。

よって、Wet実験に移行する前に、机上でAffinity以外にもpLDDTを使用することで、CBD変異体候補のスクリーニング効率が改善する可能性を示唆している。これは、文献[3]の事例と通じる結果である。一方、今回の変異体の中で、結合性が強かった変異体は全て三点変異であり、四点・五点変異体は全て結合性が弱い、つまり変異数の影響を受けて、結果が偏っている可能性がある。したがって、CBD候補設計における信頼度スコアの有用性を詳細に検証するためには、より多様なサンプルに対して評価する必要がある。

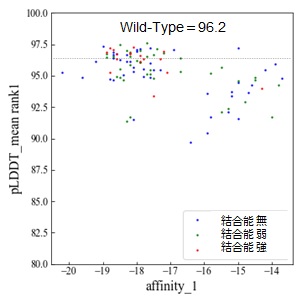
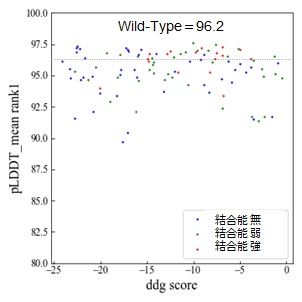


図4：各変異体のスコアと信頼度スコアの散布図（左図：ΔΔGスコア、右図：結合親和性）



図3：変異体の予測構造と信頼度スコア（三点変異体の例）

**参考文献**

1. J. Jumper et al.: “Highly accurate protein structure prediction with AlphaFold”, Nature (2021)
2. M. Mirdita et al.:“ColabFold: Making protein folding accessible to all”, Nature (2021)
3. T. Kosugi et al.:“Solubility-Aware Protein Binding Peptide Design Using AlphaFold”, Biomedicines, Vol.10, No.1626 (2022)