

# バイオフィーマティクスへの機械学習の取り込みとオーキシシン阻害剤開発

|       |               |
|-------|---------------|
| 誌名    | 日本農薬学会誌       |
| ISSN  | 21870365      |
| 著者名   | 笥,雄介<br>嶋田,幸久 |
| 発行元   | 日本農薬学会        |
| 巻/号   | 44巻1号         |
| 巻号補足  |               |
| 掲載ページ | p. 40-43      |
| 発行年月  | 2019年2月       |

農林水産省 農林水産技術会議事務局筑波産学連携支援センター  
Tsukuba Business-Academia Cooperation Support Center, Agriculture, Forestry and Fisheries Research Council  
Secretariat



ミニレビュー

# バイオインフォマティクスへの機械学習の取り込みと オーキシン阻害剤開発<sup>#</sup>

寛 雄介<sup>1,2,\*</sup>, 嶋田 幸久<sup>1</sup>

<sup>1</sup>横浜市立大学・木原生物学研究所

<sup>2</sup>農研機構・野菜花き研究部門

(2018年12月29日受理)

## Application of machine learning technique to bioinformatics and development of auxin inhibitor

Yusuke KAKEI<sup>1,2</sup> and Yukihiisa SHIMADA<sup>1</sup>

<sup>1</sup>Yokohama-city University, KIBR, Yokohama, Japan

<sup>2</sup>NARO, Tsu, Japan

Keywords: machine learning, auxin inhibitor, bioinformatics.

### はじめに

生物学にかかわる解析やモデリングは早くから機械学習が取り入れられてきた分野の一つとして知られる。一方でもっと幅広く情報学、統計学を含め、コンピューターを用いた生物を対象にした解析としてDNA配列、アミノ酸配列の解析やタンパク質の構造予測、分子間相互作用のモデリング、遺伝子発現解析、生態学のモデルの作成などの情報解析技術が、総称してバイオインフォマティクスと呼ばれる。本稿では我々がオーキシン阻害剤開発に用いてきたバイオインフォマティクス手法を中心に農薬開発に関わるとされる話題を提供したい。

### 1. オーキシン阻害剤開発

#### 1.1. バイオインフォマティクスツールAtCASTの開発からオーキシン生合成阻害剤開発へ

まずは遺伝子発現全体（トランスクリプトーム）を利用して化合物の標的を明らかにした我々の研究について紹介する。もともと我々は、AtGenExpressと名付けられた大規模な遺伝子発現データ収集プロジェクトのメンバーとして、シロイヌナズナに主に植物ホルモンやその阻害剤、その他の化合物を処理した時に起こるトランスクリプトームの変化のデータを収集していた。<sup>1)</sup> その解析の中で、植物ホルモン阻害剤処理に対する遺伝子発現応答は一般に、植物ホルモンによって引き起こされるものとおおよそ負の相関を示すという情報を蓄積してきていた。阻害剤の中でもホルモンレセプターの阻害剤は負の相関係数が高く、生合成阻害剤、輸送阻害剤の順に低くなってゆく。この関係性をうまく取り出せるように解析し、わかりやすく示すために遺伝子発現のプロファイルを大規模に実験間で比較して、実験間の関連性をネットワーク解析するデータベースツールAtCAST (Arabidopsis thaliana transcriptome correlation analysis tool) を

<sup>#</sup> 第43回大会シンポジウムを取りまとめた解説。

\* 〒514-2392 三重県津市安濃町草生360

E-mail: kakeiy197@affrc.go.jp

© 日本農薬学会

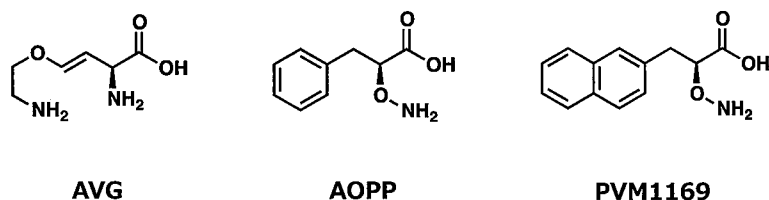


図1. オーキシン生合成酵素TAA1の阻害剤AVG, AOPP, PVM1169の構造.

開発した.<sup>2,3)</sup> AtCASTは薬剤が植物に与える影響を俯瞰する目的に使うことができる. このデータの中で, エチレン合成阻害剤aminoethoxyvinylglycine (AVG) を処理したシロイヌナズナのトランスクリプトームプロファイル (代表的な遺伝子の発現) は, エチレンの前駆体1-aminocyclopropane-1-carboxylic acid (ACC) を処理したシロイヌナズナのトランスクリプトームプロファイルと負の相関を示す (表1). 例えばACC 3hの行Correlation from AVG 3hの列データはAVG処理で代表的に変動していた遺伝子の発現をACC処理実験における変動と比較した時にどんな相関があるかを示し, Correlation to AVG 3hの列はACC処理で代表的に変動していた遺伝子の発現をAVG処理試験における変動と比較した相関を示す. 同様にAVG処理遺伝子発現プロファイルはエチレン阻害剤Daminozide, エチレンシグナル伝達タンパク質EBF1/2の安定化効果のあるMG132処理, エチレン非感受性の変異体 (ein2-1) に引き起こされる遺伝子発現プロファイルとは正の相関を示すので, エチレン阻害の結果似たような遺伝子発現変化が引き起こされていることが観察される. AVGは代表的なオーキシンであるindole-3-acetic acid (IAA) 処理下の遺伝子発現プロファイルに対しても負の相関を強く示したことから, AVGがオーキシンに対しても何らかの阻害の効果のある可能性が見出された.<sup>4)</sup> AVGの分子構造をもとに, IAAに対してさらに構造が近いと思われる化合物をスクリーニングし, その後の分子生物学実験と合わせることでオーキシンの生合成阻害剤L-

amino-oxyphenylpropionic acid (AOPP) を世界で初めて開発することができた (図1).<sup>4)</sup> AtCASTで使われている解析はModule-based correlation network analysisと名付けた遺伝子発現ネットワーク解析である. 詳しくはAtGenExpress JPNのWebページで公開されているAtCASTのMethodページを参照されたい.<sup>5)</sup> トランスクリプトームを用いたネットワーク解析も数多く開発されているが, Savoie<sup>6)</sup> は機械学習を用いたモデルであるBayesian network modelを用いて遺伝子発現制御ネットワークから抗生物質Griseofulvinのターゲットサイトを探し出すことに成功している.

## 2. 機械学習を用いた我々の研究

### 2.1. タンパク質—化合物ドッキングシミュレーションと機械学習

AOPPはトリプトファンからのIAA合成の第1段階の反応を担う酵素TAA1の阻害剤である (図2A). AOPPの他にキヌレニン (L-Kyn) がTAA1の活性を阻害することがわかっている.<sup>7)</sup> 我々はAOPPの開発の後, AOPPの構造に変化を加えてPyruvamine (PVM) と名付けた一群のオーキシン生合成阻害剤を報告した.<sup>8)</sup> これらのTAA1阻害剤とTAA1の*in vitro*活性阻害試験の結果を化合物—タンパク質ドッキングシミュレーション結果 (図2B) と比較したところ大まかに相関があることがわかった (図2C). ドッキングシミュレーション結果, 既知の阻害結果情報, タンパク質や化合物の分子量などの性質情報を大量のタンパク質と化合物の組み合

表1. AtCASTにおけるAVG処理シロイヌナズナのトランスクリプトームと正や負の相関を示す試験

| Experiment                         | Correlation from AVG 3h | Correlation to AVG 3h | Control experiment | Experimental category                      |
|------------------------------------|-------------------------|-----------------------|--------------------|--------------------------------------------|
| AVG 3h                             | 1.00                    | 1.00                  | Mock 3h            | Effect of ethylene inhibitors on seedlings |
| Daminozide 3h                      | 0.60                    | 0.79                  | Mock 3h            | Effect of other inhibitors on seedlings    |
| MG132 3h                           | 0.72                    | 0.66                  | Mock 3h            | Effect of other inhibitors on seedlings    |
| ein2-1+elf18 2h                    | 0.61                    | 0.74                  | WT+elf18 2h        | GEO bulk                                   |
| ein2-1+Pep2 2h                     | 0.61                    | 0.73                  | WT+Pep2 2h         | GEO bulk                                   |
| P-chlorophenoxy-isobutyric acid 3h | 0.53                    | 0.78                  | Mock 3h            | Effect of auxin inhibitors on seedlings    |
| ein2-1 untreated                   | 0.52                    | 0.75                  | WT untreated       | GEO bulk                                   |
| IAA 3h                             | -0.60                   | -0.64                 | Mock 3h            | Basic hormone treatment of seedlings       |
| ACC 3h                             | -0.56                   | -0.5                  | Mock 3h            | Basic hormone treatment of seedlings       |

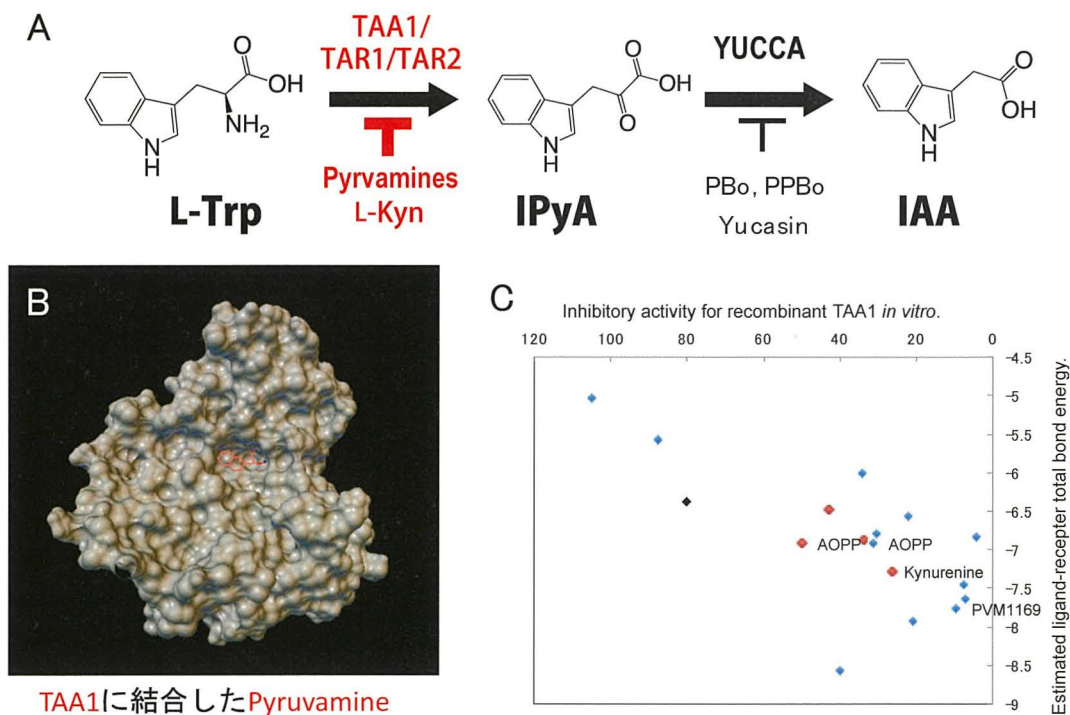


図2. シロイヌナズナにおける IAA 合成経路と Pyruvamine による TAA1 活性阻害. IAA 合成経路. 酵素 (矢印の上) とその阻害剤 (矢印の下) (A), TAA1 と Pyruvamine1169 のドッキングシミュレーションの様子 (B), ドッキングシミュレーションにおける結合力 (縦軸) と *in vitro* 阻害試験における TAA1 の比活性 (横軸), 赤と青の点はそれぞれ異なる大腸菌株から精製された TAA1 のロットを示す (C).

わせで用意することにより, 目的のタンパク質の阻害剤候補をインシリコスクリーニングし, トップ6化合物の *in vitro* 阻害活性試験を行った (図3). PVM 中でもナフチル基を

有する PVM1169 は阻害効果が高く TAA1 特異性の高い阻害剤である. これよりも阻害効果の高い阻害剤を選抜するため, PVM1169 が 50% 活性阻害を示す 1  $\mu$ M でトップ6化合物

### 1. Multiple Target Screening (トポロジー生成>ラフドッキング>MD計算)

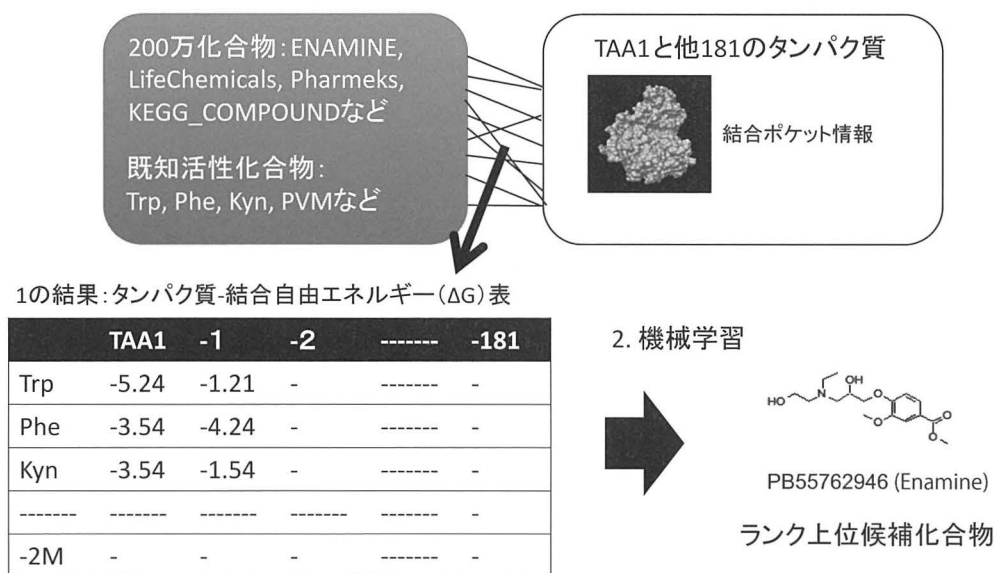


図3. myPresto による TAA1 阻害剤インシリコスクリーニングの様子. myPresto から提供されている解析済みの 200 万化合物と 181 の酵素のドッキング・阻害効果情報に TAA1 と既知の基質・阻害剤の情報を加え機械学習・マルチターゲットスクリーニング法により TAA1 の阻害剤候補のスクリーニングを行った.

を試験した。残念ながら阻害効果のある化合物は得られなかった。このインシリコスクリーニングにはmyPresto4.2<sup>9)</sup>を用いた。myPrestoはAMED・経済産業省・NEDOの支援によって産業技術総合研究所などによって開発され、高速かつ高精度な静電相互作用計算理論および高効率の構造サンプリング手法を高効率に実行可能な分子動力学計算プログラムで、GPGPUを用いることで高速な計算が可能になっている。既知の阻害剤とタンパク質結合結晶構造があれば利用できる「機械学習法とMTS & DSI法を組み合わせたインシリコスクリーニング」では10～数十パーセントと通常の0.1～1%オーダーのヒット率と比較して高いヒット率の結果が報告されている。阻害剤候補選抜の最終ステップには機械学習ニアレストネイバー (kNN) が用いられていた。

## 2.2. その他の機械学習を用いた筆者らの解析

遺伝子発現シミュレーションを最適化できる遺伝子転写開始点近傍の配列の抽出から特定の条件下での遺伝子発現を制御するシスエレメントを予測するプログラムMAMA (Microarray associated motif analyzer) は遺伝子発現シミュレーションのステップでサポートベクターマシンという機械学習を用いている。<sup>10)</sup> 最新の研究では、植物生長調節剤の処理と環境条件の組み合わせからトマトの生育シミュレーションモデルを、オミクス解析、機械学習 (主にXGB) などと組み合わせで作成する試みをおこなっている。例えばmyPrestoの最終ステップの機械学習部分を近年開発されているXGBに置き換えることによってヒット率は上がるかもしれないが筆者らは試していない。ディープラーニングは画像解析や波形解析に用いることのできる簡単で強力なツールとなったが、そうではないデータが多い生物系の分野ではXGBのようなその他の機械学習も含め、それぞれの特徴を生かして研究に用いてもらいたい。

## 引用文献

- 1) H. Goda, E. Sasaki, K. Akiyama, A. Maruyama-Nakashita, K. Nakabayashi, W. Li, M. Ogawa, Y. Yamauchi, J. Preston, K. Aoki, T. Kiba, S. Takatsuto, S. Fujioka, T. Asami, T. Nakano, H. Kato, T. Mizuno, H. Sakakibara, S. Yamaguchi, E. Nambara, Y. Kamiya, H. Takahashi, M. Y. Hirai, T. Sakurai, K. Shinozaki, K. Saito, S. Yoshida and Y. Shimada: *Plant J.* **55**, 526–542 (2008).
- 2) E. Sasaki, C. Takahashi, T. Asami and Y. Shimada: *Plant Cell Physiol.* **52**, 169–180 (2011).
- 3) Y. Kakei and Y. Shimada: *Plant Cell Physiol.* **56**, e7 (2015).
- 4) K. Soeno, H. Goda, T. Ishii, T. Ogura, T. Tachikawa, E. Sasaki, S. Yoshida, S. Fujioka, T. Asami and Y. Shimada: *Plant Cell Physiol.* **51**, 524–536 (2010).
- 5) <http://atpbsmd.yokohama-cu.ac.jp/> (2018年12月9日閲覧)
- 6) C. J. Savoie, S. Aburatani, S. Watanabe, Y. Eguchi, S. Muta, S. Imoto, S. Miyano, S. Kuhara and K. Tashiro: *DNA Res.* **10**, 19–25 (2003).
- 7) W. He, J. Brumos, H. Li, Y. Ji, M. Ke, X. Gong, Q. Zeng, W. Li, X. Zhang, F. An, X. Wen, P. Li, J. Chu, X. Sun, C. Yan, N. Yan, D. Y. Xie, N. Raikhel, Z. Yang, A. N. Stepanova, J. M. Alonso and H. Guo: *Plant Cell* **23**, 3944–3960 (2011).
- 8) M. Narukawa-Nara, A. Nakamura, K. Kikuzato, Y. Kakei, A. Sato, Y. Mitani, Y. Yamasaki-Kokudo, T. Ishii, K. Hayashi, T. Asami, T. Ogura, S. Yoshida, S. Fujioka, T. Kamakura, T. Kawatsu, M. Tachikawa, K. Soeno and Y. Shimada: *Plant J.* **87**, 245–257 (2016).
- 9) <https://www.mypresto5.jp/> (最新版: 2018年12月9日閲覧).
- 10) Y. Kakei, Y. Ogo, R. N. Itai, T. Kobayashi, T. Yamakawa, H. Nakanishi and N. K. Nishizawa: *Rice (N.Y.)* **6**, 22 (2013).