

令和2年度 博士論文

生体機能分子に関する  
電子状態インフォマティクス研究

熊本大学大学院自然科学研究科

井手尾 俊宏

## 目次

第1章 序論 .....	1
第2章 電子状態情報を用いた化合物の分類.....	8
第3章 脂肪酸合成酵素に対する天然物の阻害活性に関する回帰モデルの作成	25
第4章 抗菌活性を示す天然物に関する回帰モデルの作成およびバーチャルスクリーニング .....	61
第5章 結論 .....	109

## 第1章 序論

### 1-1 はじめに

人類の歴史において「生存する」ということは、生物としての人間が追い求める命題の一つである。有史以来、文明の発展に伴い、多くの人間は健康で長生きすること、すなわち「不老長寿」を理想としてきた。そのことは、メソポタミア文明のギルガメシュ叙事詩で語られる不死希求の旅や、不老不死の仙薬を求めた秦の始皇帝にまつわる徐福伝説など、世界各地の文明に共通して見られる。紀元1世紀ごろのエジプトに端を発し、のちにヨーロッパに広がった錬金術は、元々は卑金属から貴金属を生み出すことを目的とした技術であったが、不完全なものを完全なものにする、すなわち有限なものである肉体や魂を永遠のものにすることも目的の一つとして発展していった。東洋においても同様に、不老不死の霊薬を作る錬丹術が起こっている。その過程で霊薬として作られた水銀化合物を服用したことが、始皇帝などの複数の皇帝の死亡要因になったと考えられている。現代において、錬金術や錬丹術が目的としていた不老不死は達成されてはいないが、これらの技術の追究が今日の化学、医学、薬学の発展に繋がっていると考えられる。人間の寿命は徐々に伸びてきており、特に直近の数世紀においては科学技術の発展に伴い、目に見えて延伸した。その要因は栄養問題や生活環境の改善、疾病の治療方法の確立など多岐にわたるが、現代における最大の技術革新は、コンピューターの発明にあると言えよう。

薬という分野においていえば、古来より人々は木や草花などの植物や、虫、鉱物などの自然物を薬として使用してきた。近代における薬に関する研究は、それらの伝統的な有効成分の同定、あるいは全くの偶然の産物など様々な要因によって発展してきた。これに対して、現代の創薬は、分子生物学や生理学の見地からの疾病やアレルギー

一反応の機構の解明や、その見地で発見された薬の特性を理解することで行われている。このように科学の発展によって創薬の技術は進歩したが、コンピューターを使用したシミュレーションなどの計算科学的手法の導入により、対象とする物質の性質や反応性の予測などの多くのデータが得られるようになった。

データに対して AI や統計解析などの情報学的手法を適用し、膨大な数のデータから価値のある情報を探索する技術のことをデータマイニングという。アメリカの計算機科学者である James Nicholas Gray は、ビッグデータに対するデータ検索やデータマイニング、機械学習の手法などを用いて行う科学的研究手法のことを「データ集約型科学」と定義し、「経験科学」、「理論科学」、「計算科学」に継ぐ、「第 4 の科学」と称した[1]。この分野は科学全体の歴史から見れば新しいものであるが、近年急速に発展している分野であり、これらの技術に基づいた研究開発が盛んに行われている。特に化学の分野においては、コンピューターと情報化技術を用いて行う「ケモインフォマティクス」(情報化学、化学情報学)という研究手法が起っている。1999 年に計算化学者である Frank Kenneth Brown はケモインフォマティクスを以下のように定義している [2]。

“Chemoinformatics is the mixing of those information resources to transform data into information and information into knowledge for the intended purpose of making better decisions faster in the area of drug lead identification and optimization”

(ケモインフォマティクスとは、薬剤リードの識別および最適化の分野においてより良い決定をより迅速に行うことを目的として、データ

を情報に、情報を知識に変換するためにそれらの情報源を混合  
することである。)

現在ではケモインフォマティクスの技術は機能性材料の開発などの分野でも活用されているが、ケモインフォマティクスと創薬化学との結びつきは依然強く、様々な技術が活用されている。

## 1-2 情報化学手法の例

創薬において使用されている情報化学手法をいくつか示す。

### (a) データベース

実験等で得られた分子に関する情報を格納する目的で設計されており、化学データベースとしてはそのデータ構成により「化学物質データベース」と「化学反応データベース」の二つに分けることが出来る。データベースの例としては、タンパク質や核酸などの構造座標を蓄積する国際的なデータベースである PDB[3]や、アメリカの国立生物工学情報センターの PubChem[4]、欧州バイオインフォマティクス研究所の ChEMBL[5]、カリフォルニア大学の ZINC[6]などが挙げられる。

### (b) 記述子

情報化学における記述子とは、化合物の化学情報に基づいて、その化合物の構造特性や化学特性などの各要素を数値化したものである。広く使用されている記述子としては fingerprint がある。これは、対象とする分子中に、予め定義されている構造的特徴を含む場合は“1”、含まない場合は“0”として、その分子の化学構造を 1/0 のビット列で表現する方法である。この他にもケモインフォマティクスの分野では数多くの記述

子が開発されている。

### (c) 類似性評価手法

二つの分子の類似性を算出するための手法としては、Tanimoto 係数を使用した二つの集合間の類似性評価法[6]がある。分子 A・B 間の Tanimoto 係数 ( $T_{A,B}$ ) の計算式は以下の式で表される。

$$T_{A,B} = \frac{N_{A,B}}{N_A + N_B - N_{A,B}}$$

この場合、二つの集合を表す値は、(b)で説明した fingerprint のビット列を使用する。 $N_A$ と $N_B$ はそれぞれの分子のビット列で“ 1 ”となっているビットの個数で、 $N_{A,B}$ は分子 A・B で共に“ 1 ”となっているビットの個数である。Tanimoto 係数は、分子 A・B が全く同じ分子であれば 1 となり、分子 A・B の間に共通して 1 となるビットが一つもなければ 0 となる。

### (d) バーチャルスクリーニング手法

バーチャルスクリーニングとは、タンパク質受容体や酵素などの医薬品ターゲットと最もよく結合する化学構造を特定するために、コンピューターを用いて高速に多数の物質を評価してターゲットを絞り込む技術である。バーチャルスクリーニングには大きく分けてリガンド法と構造法の2つの手法がある。リガンド法は既知リガンドをクエリーとして類似性検索を行い、活性をもつと考えられる化合物を探索する方法で、構造法はターゲットタンパク質のモデル構造を構築し、リガンド結合サイトなどの薬剤を結合させたい部位に対して化合物ライブラリ中の化合物を一つ一つあてはめ、うまく適合する化合物を探索する方法である。リガンド法はリガンドの情報のみで適用することができ、構造法は必要な計算量が多いが、既知活性化合物情報に依存しない新しい化学種を

発見しやすいという特徴がある。

#### (e) 定量的構造活性相関 ( Quantitative Structure-Activity Relationship: QSAR )

化学物質の、構造と生物学的 (薬理的あるいは毒性学的) な活性との間に成り立つ相関関係のことである。これを解析することで構造的に類似した化合物の薬効予測を行うことができる。また、化学構造と物理的性質との関係は定量的構造物性相関 ( Quantitative Structure-Property Relationship: QSPR ) と呼ばれる。これらは、「構造－機能相関」という考え方による[7]。

### 1-3 情報化学手法の有用性

前述の様に、近代までの創薬では偶然の発見を除けば全てが「人の手」で行われてきた。この方法では、目的の薬剤を造るまでに多くの金銭的・時間的コストがかかる。これに対して、コンピューターを使用することで、膨大な数の化合物の中から薬として有望なものを短時間で見つけることができる。また、蓄えられた膨大な情報から、それらの化合物と標的との間の反応性の予測も行うことができる。このように、「コンピューターを用いて」創薬を行うことを「インシリコ創薬」と呼ぶ。インシリコ創薬は、従来の実験主体の手法に比べて理論的な解釈ができる研究成果を出すことができ、情報化学手法は創薬分野において非常に重要な技術であるということができる。

### 1-4 本研究の目的

物質の機能は分子の構造に依存するという「構造－機能相関」に基づいたインフォマティクス解析が広く行われている。しかし、物質の構造と機能はいずれもその物質の電子状態に起因するものでもありと考えられるため、電子状態情報に基づいた解析を

行えば、構造的には似ていないが似た機能をもつ物質を提案することができる可能性があると考えられる。本研究は、我々のグループで開発している電子的記述子を用いて、医薬品などの機能性分子がそれらの性質を示す電子的な要因を体系的に明らかにし、より高性能、あるいは、より安全な機能性分子を提示するための方法を開発することを目的として行われた。本論文の構成を以下に示す。第2章では機能性分子などの化合物の情報を蓄積する電子状態データベースの作成と、それらの電子状態情報を用いた化合物の分類について述べた。第3章では抗ガン剤の開発を念頭に置いて FASN 阻害剤についての定量的回帰モデルを作成し、FASN 阻害能が高い分子の特徴の解明を試みた結果について述べた。第4章ではより高活性な抗菌作用がある分子の探索を目的として、精油に含まれる分子の活性についての定量的回帰モデルを作成し、その他の天然成分の情報を適用して行なった活性予測の結果について述べた。そして、第5章では一連の研究で得られた成果を総括した。

#### 1-5 参考文献

- [1] A. J. G. Hey, D. S. W. Tansley, K. M. Tolle, Eds. *The Fourth Paradigm: Data-Intensive Scientific Discovery*, Microsoft research, Redmond, WA, (2009).
- [2] F. K. Brown, *Annu. Rep. Med. Chem.*, **33**, 375 (1998).
- [3] J. L. Sussman, D. Lin, J. Jiang, N. O. Manning, J. Prilusky, O. Ritter, E. E. Abola, *Acta Crystallogr. D*, **54**, 1078 (2009).
- [4] Y. Wang, J. Xiao, T. O. Suzek, J. Zhang, J. Wang, S. H. Bryant, *Nucleic Acids Res.*, **37**, W623 (2009).
- [5] A. Gaulton, L. J. Bellis, A. P. Bento, J. Chambers, M. Davies, A. Hersey, Y. Light,



S. McGlinchey, D. Michalovich, B. Al-Lazikani, J. P. Overington, *Nucleic Acids Res.*, **40**, D1100 (2012).

[6] J. J. Irwin, B. K. Shoichet, *J. Chem. Inf. Model.*, **45**, 177 (2005).

[7] A. Cherkasov, E. N. Muratov, D. Fourches, A. Varnek, I. I. Baskin, M. Cronin, J. Dearden, P. Gramatica, Y. C. Martin, R. Todeschini, V. Consonni, V. E. Kuz'min, R. Cramer, R. Benigni, C. Yang, J. Rathman, L. Terfloth, J. Gasteiger, A. Richard and A. Tropsha, *J. Med. Chem.*, **57**, 4977 (2014).

## 第2章 電子状態情報を用いた化合物の分類

### 2-1 はじめに

化学の分野において計算化学の手法は今やなくてはならない重要な道具となっている。計算結果は近似的な値ではあるが、それらにより我々は、分子の構造、反応性、及び特性を解析、理解、及び予測するための非常に豊富な情報を得ることができる。電子は結合の形成・開裂、及び外部への摂動の応答において本質的に重要であるため、電子状態を見積もる手法は計算化学における最も重要な方法の一つである。

電子状態計算を実行すると、それらの分子を記述する多数の数値データを取得できる。このようなデータが分子セットについて蓄積されると、化合物に関する有用な情報や重要な知見に繋がりを抽出することが期待される。これを実現するためには、いくつかの計算ツールがケモインフォマティクス(化学情報学)手法と効果的に組み合わせることが必要である。科学、あるいは工学の分野で研究を行う際には、広範な情報検索を必要とするため、この見方は妥当であると思われる。計算化学及び化学情報学を用いた解析において、一連の計算手順をただ一つに限定して確立することは難しい。妥当かつ合理的な方法としては、次の3つの技術を組み合わせることであると考えられる。

(i) 電子状態計算、

(ii) 分子データベースおよびデータベース検索技術、

(iii) データ科学に基づく解析(有用な情報を見つけるための統計データ分析)。

この組み合わせでは、蓄積された分子に関する知識を活用して、新しい知識を得ることができる。計算化学結果から得られる数値データは、その品質という観点では均一であるため、実験から得られる数値と比較して、データの比較などの取り扱いが容易で

あることが利点である。

本研究では、上記の(i)-(iii)のコンピューターを用いた技術の組み合わせによる、知識情報に基づいた分子探索が有用な情報を提供できるかどうかの確認を目的としている。その例として、分子データベースに登録された「ガン」関連分子のセットに焦点を当てて解析を行なった。

## 2-2 研究手法

### 2-2-1 計算化学的手法

データサイエンスに基づく解析を行うことを目的として、分子データベースを構築、及び分子間の類似度を算出するために、各分子の電子状態計算を実行した。これらの計算は Gaussian09 [1]を使用して行なった。計算には半経験的手法の一つである PM6 法、及び密度汎関数理論法 (DFT) を適用した。DFT 計算では、M06-2X 汎関数と 6-31G(d,p)基底関数を使用した。構造最適化計算を行なったのち、振動モードと赤外線 (IR) スペクトルに関する情報を取得するために、基準振動解析の計算を実行した。また、スピナー重項励起状態と励起エネルギースペクトルに関する情報を取得するために時間依存密度汎関数法 (TDDFT) による計算を同じジオメトリに対して行なった。

### 2-2-2 分子データベース

American Chemical Society (ACS) が提供する「Molecule of the Week (MOTW)」[2] のアーカイブの情報を参照した。MOTW の Web ページのスナップショットを Fig. 2-1 に示す。本研究内容を投稿した時点 (2015 年) で 377 種の分子が掲載されていた。我々がこれらのデータに着目した理由は、MOTW が「化学データ (分子構造)」と「知

識データ(それらの化合物に関する数行の解説文)」の両方を含んでいたからである。知識データの解説文に含まれる単語は、その分子に関する重要なキーワードであると見なすことができる。したがって、分子構造と電子状態、及び関連する知識情報の間の相関を見つけることができると考えられる。

MOTW の web サイトでは各分子の分子構造は画像形式で掲載されており、それらの画像ファイルをコンピューター上に保存した。印刷物や画像ファイルなどから分子の構造式を抽出するソフトウェア「CLiDE」[3]を使用して、画像ファイル中の化学式の図から分子構造を抽出して認識してファイル変換を行い、それらの分子を SMILES 記法[4]で表現した文字列を取得した。次に、化学構造ファイルのフォーマット変換を行うためのソフトウェア「Open Babel」[5-7]を使用して、SMILES のデータを 3 次元の分子構造のファイルに変換した。それらのファイルを電子状態計算の構造最適化の初期構造として使用した。本研究の初期段階では PM6 法で行なった計算結果を取り扱った。その理由は PM6 法が半経験的手法であるため、短時間で結果が得られるためである。しかし、実際に計算結果をデータベースに登録する段階では DFT (M06-2X) 法の結果を用いた。この理由は、後者の手法の結果の方が定量的でより信頼性が高いから、及び MOTW の分子を計算するのに必要な時間が十分に少ないと考えられるからである。

MOTW の分子の電子状態計算結果の情報と知識情報を、我々の研究室で開発しているデータベースである“Intelligent Molecular Informatics System (IMIS)”[8]に登録した。IMIS はデータベース、検索エンジン、分子構造の閲覧、モデリングツールを備えた Web ベースのシステムであり、本報告の時点では IMIS データベースに MOTW のアーカイブの分子セットに登録した。



Search

- Publications
- Meetings
- Careers
- Membership & Networks
- Education
- Advocacy
- Funding & Awards
- Press Room

American Chemical Society » Molecule of the Week » Archive » I » Tanshinone IIA

## Molecule of the Week

Archive

### Molecule of the Week

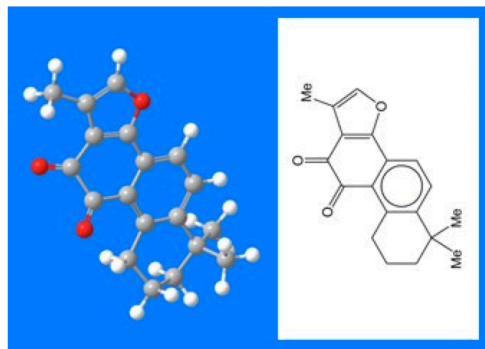
#### Tanshinone IIA

Tanshinone IIA is one of the most abundant constituents of Chinese sage (*Salvia miltiorrhiza*) roots, which are highly prized in traditional medicine for their ability to treat cardiovascular disease and stroke. In addition to its traditional uses, tanshinone IIA is under study for anti-cancer treatments.

Yet another use for tanshinone IIA is being investigated. S. A. Renshaw and colleagues at the University of Sheffield found that it causes neutrophils to move away from animals' injury sites, allowing the wounds to heal.

Neutrophils are sent by the immune system to reduce inflammation, but if they stay around, healing can be delayed. Tanshinone IIA's mechanism for removing neutrophils is under study.

 [More about this molecule](#) from CAS, the most authoritative and comprehensive source for chemical information.



Follow @MOTW\_ACS

- Bookmark
- Facebook
- LinkedIn
- Google+
- Twitter
- Print
- Email
- Share...
- ACS Network

Figure 2-1. ACS が公開している MOTW の Web ページの図[2]。今回の解析の際に基準の化合物とした tanshinonene IIA に関する情報が記載されている。

ID	Compound Name	Stoichiometry	Canonical SMILES	Charge	Spin	Method	Basis Set	Solvent Effect	Nuclear Repulsion Energy (hartree)	SCF Energy (hartree)	Dipole Moment (D)	HOMO Level (hartree)	LUMO Level (hartree)	Keywords
1	2,2-Azobis(isobutyronitrile)	C8H12N4	<chem>N#CC(N=N)C(C)C</chem>	0	1	RPM6	VSTO-6G	none	382.99906	0.12387	0.0041	-0.3594	-0.0206	2,2'-Azobis(isobutyronitrile), or AIBN, is a commonly used initiator for free-radical reactions, including many polymerizations. Its preparation was first reported by Thiele and Heuser in 1896, and it has been used in polymerization since at least the 1940s. Its downsides are high toxicity (it decomposes to HCN in vivo), flammability, and the explosivity of its acetone solutions.
2	6-Aminopenicillanic acid	C8H12N2O3S	<chem>OC(=O)C[C@@H]1N2C(=O)C[C@H]1C(=O)N2C</chem>	0	1	RPM6	VSTO-6G	none	559.76843	-0.18274	4.0723	-0.3502	-0.0153	A broad range of affordable semisynthetic penicillins and cephalosporins has been made possible by the use of the intermediate 6-aminopenicillanic acid (6-APA).
3	Abacavir	C14H18N6O	<chem>CC1=NC(=O)N2C(=O)N(C)C(=O)N2C1</chem>	0	1	RPM6	VSTO-6G	none	883.54345	0.07824	0.675	-0.31	-0.0016	Abacavir, is an antiretroviral drug. When a virus (such as HIV) tries to manufacture DNA from the viral RNA, it unknowingly incorporates abacavir instead of a natural component of DNA, guanosine, which stops the virus from reproducing.
4	Abscisic acid	C15H20O4	<chem>OC(=O)C=C(C)C(=O)C1=CC(=O)C=C1C</chem>	0	1	RPM6	VSTO-6G	none	866.97919	-0.26143	7.7042	-0.3619	-0.0184	Abscisic acid is a hormone that can cause plants to lose their leaves, a process known as abscission. In the 1960s, researchers determined its structure and showed that it is found in many plants. Subsequent research showed that it is involved with abscission in only a few plants but that it is an important signalling molecule for responses to environmental stresses. It was once thought that applying abscisic acid to crop plants would help them resist drought. But the hormone is expensive to produce, has a short half-life, and is photodegradable.
5	Abyssinone_II	C20H20O4	<chem>CC(=CC)C1C(=O)C(=O)C1C(=O)C(=O)C</chem>	0	1	RPM6	VSTO-6G	none	1044.75053	-0.21007	3.0698	-0.3314	-0.0176	Recently, S. R. Cutler and colleagues at the University of California, Riverside, discovered that the sulfonamide quinabactin binds and activates abscisic acid receptors. This compound may eventually be used to help plants retain water and resist drought, but much developmental and regulatory work needs to be done. (2S)-Abyssinone II is a prenylated flavonoid from the Chinese medicinal plant Broussonetia papyrifera. The racemic compound was synthesized by M. Cushman and co-workers, who also made several derivatives of it. One of them showed 20 times more activity against human aromatase than the base compound and is being tested as a breast cancer chemopreventive.

Figure 2-2. IMIS のデータベースに登録した化合物のうち、最初の 5 つの分子に関する情報のスクリーンショットの図。「Keywords」の列には、ACS の MOTW から取得した情報が表示されている。

### 2-2-3 データ科学的手法

データ科学に基づく解析として、分子の類似度を評価するための計算アルゴリズムを適用した。このアルゴリズムは我々の研究室で開発したプログラムを使用した[9]。この方法では、電子状態計算結果から得られたエネルギースペクトルに基づく分子の類似性を評価した。本研究の時点では、類似性の記述子として 8 つの記述子を定義した。8 つの記述子はそれぞれ、(a) UV/Vis スペクトルに関する類似度、(b) 励起エネルギースペクトルに関する類似度、(c) 軌道エネルギースペクトルに関する類似度、(d) 最高被占軌道 (HOMO) のエネルギー準位に関する類似度、(e) 最低空軌道 (LUMO) のエネルギー準位に関する類似度、(f) HOMO-LUMO 間のエネルギーギャップに関する類似度、(g) IR スペクトルに関する類似度、(h) 振動エネルギースペクトルに関する類似度、である。

UV/vis スペクトル、励起エネルギースペクトル、軌道エネルギースペクトル、IR スペクトル、振動数スペクトルの類似度は計算結果を状態密度関数のような形で表現し、グラフ同士の重なり具合によって類似度を求めた。式(1.1)に電子状態計算の結果をグラフ化するための式を示す。

$$f(x) = \sum \left( \beta \cdot \exp \left[ -\frac{\ln 0.5}{\alpha^2} (x - x_i)^2 \right] \right) \quad (1.1)$$

式(1.1)の  $f(x)$  はガウス関数の和を表している。 $\alpha$  はガウス関数におけるピークの幅であり、 $\beta$  はピークの高さである。 $x_i$  はピークの位置を決定するためのパラメータである。 $\alpha$ 、 $\beta$ 、 $x_i$  は記述子によって代入する値が異なっている。記述子の種類と代入したパラメータの組み合わせを Table 2-1 に示す。 $\alpha$  は代入する値を任意に決めた後、分子セットを用いて類似性の評価を行う操作を繰り返し、経験的に決定した。 $\beta$  の値は、UV/vis

スペクトル、IR スペクトルにおいては、それぞれのピークの強度が重要になるため、強度がピークの高さに反映されるようにしている。一方で、軌道エネルギースペクトル、励起エネルギースペクトル、振動数スペクトルの場合は、特定のエネルギー、もしくは波数における状態数を求めることが重要であると考えたため、 $\beta$  は全てのピークで一律である。次に式(1.1)により得られたスペクトルを積分値によって規格化した。最後に規格化したグラフ同士を重ね合わせ、積分によって重複部分の面積を求め、これを類似度とした。類似度計算の式を式(1.2)に示す。

$$r = \int \frac{f(x)}{\sqrt{\int (f(x))^2 dx}} \cdot \frac{g(x)}{\sqrt{\int (g(x))^2 dx}} dx \cdot 100 \quad (1.2)$$

HOMO のエネルギー準位、LUMO のエネルギー準位、HOMO-LUMO エネルギーギャップの類似度はそれぞれ式(1.3)によって、計算した。

$$r = \exp \left[ - \left| \frac{\varepsilon - \varepsilon_{ref}}{(\varepsilon + \varepsilon_{ref})/2} \right| \right] \cdot 100 \quad (1.3)$$

式(1.3)の $\varepsilon$ には HOMO のエネルギー準位、LUMO のエネルギー準位、HOMO-LUMO エネルギーギャップの値が入る。 $\varepsilon_{ref}$ は参照物質の値を代入する。

Table 2-1. スペクトルの作図に使用したパラメータ

descriptor	$x_i$	$\alpha$	$\beta$
$r_{UV}$	励起エネルギー[eV]	0.30	1
$r_{ex}$	励起エネルギー[eV]	0.15	$\sqrt{\alpha}/\sqrt{\pi}$
$r_{orb}$	軌道エネルギー[hartree]	0.02	1
$r_{IR}$	振動数[ $\text{cm}^{-1}$ ]	20	$\sqrt{\alpha}/\sqrt{\pi}$
$r_{freq}$	振動数[ $\text{cm}^{-1}$ ]	20	1

(a) UV/Vis スペクトル及び(b) 励起エネルギーのスペクトルは、それぞれ 0.3 及び



0.15 eV の幅(半値全幅:FWHM)のガウスバンドを重ね合わせることによって定義した。(c)軌道エネルギースペクトルについての FWHM は、0.02 hartree(0.54 eV)のガウスバンドとした。これらのバンドの幅は、分子の類似性のスコアリングとランキングにおける幾つかの試行の結果に基づいて決定した。(g)IR スペクトル及び(h)振動エネルギースペクトルについては、 $20\text{cm}^{-1}$  の FWHM で計算した。UV/Vis スペクトルと IR スペクトルについては、定義したバンドの領域は、DFT 計算における UV/Vis の振動子強度と IR の吸収強度にそれぞれ等しい形状になった。スペクトルデータによる分子の類似性は、参照分子とそれ以外の分子とのスペクトルをそれぞれ正規化し、それぞれのスペクトルを重ね合わせて、一致した面積の割合で評価した。(d)から(f)の非スペクトル系の記述子に関する類似度は、参照分子の数値との差を指数関数による式を用いて重みをつけた相対値に変換して算出した。

Ward 法を用いたクラスター解析を行なった[10]。解析には数式処理システム「Mathematica」を使用した[11]。

解析対象とした化合物は、テキスト検索をした結果、キーワードとして「ガン(cancer)」を含んだ化合物であり、内訳は、発ガン性物質(2化合物、Formaldehydeと *p*-Dichlorobenzene)、乳ガンの抗ガン剤(6化合物)、前立腺ガンの抗ガン剤(3化合物)、その他のガンの抗ガン剤(14化合物)であるため、これらの化合物群は4つのグループに分けられることが想定される。解析を行うにあたり、類似度を算出する際の参照分子として“Tanshinone IIA”を選択した。この化合物は乳ガンと前立腺ガンの両方の抗ガン剤として機能する。量子化学的類似性を使用して、発ガン性物質と抗ガン剤の合理的な分類が得られるまで、分類を繰り返し行なった。

## 2-3 結果と考察

まず、MOTW の分子のうち、キーワードとして「ガン(cancer)」を含む分子がいくつあるかを検索した。検索対象は IMIS データベースに登録されている MOTW の分子である。キーワード検索の結果、Fig. 2-3 に示す 24 個の分子が見つかった。量子化学的類似度を評価するために、M06-2X 汎関数による DFT および TDDFT 計算をこれらの分子に対して実行した。

Table 2-2 は Tanshinone IIA を参照分子として類似度を計算した場合の各分子の電子的類似度の値である。各行は一つの分子と参照分子との類似度の値であり、各列は類似度の評価に使用した指標とそれに対応する値である。評価に使用した類似度は 8 種類あり、それぞれの指標は、軌道エネルギー ( $r_{orb}$ )、HOMO のエネルギー準位 ( $r_{HOMO}$ )、LUMO のエネルギー準位 ( $r_{LUMO}$ )、HOMO-LUMO 間のエネルギーギャップ ( $r_D$ )、振動数 ( $r_{freq}$ )、IR スペクトル ( $r_{IR}$ )、励起エネルギー ( $r_{ex}$ )、UV/vis スペクトル ( $r_{UV}$ ) である。それら 8 種の類似度の平均値を算出し、各分子はその値が大きい方から順番に並んでいる。

各指標の値の範囲に着目すると、最も低い分子の値は、 $r_{orb}$  では 40%前後であり、 $r_{HOMO}$  では 70%前後、 $r_{LUMO}$  では 0%前後、 $r_D$  では 60%前後、 $r_{freq}$  では 10%前後、 $r_{IR}$  では 0%前後、 $r_{ex}$  では 0%前後、 $r_{UV}$  では 0%前後である。

算出した類似度の数値を記述子として、Ward 法を用いたクラスター分析を行なった。まず、定義した 8 種類の記述子を全て使用して行なったクラスタリングの結果のデンドログラムを Fig. 2-4 に示す。この結果では発ガン性物質の 2 つが別々の、かなり離れたクラスターに分類されているため、発ガン性物質と抗ガン剤をうまく区別できていないことがわかる。これを踏まえて、分類の精度をあげるために、化合物の性質をよく表

Table 2-2. Tanshinone IIA を参照分子とした場合の各分子の類似度の値とその分子の物性。各分子の物性について、MOTW の web サイト上で発ガン性や抗ガン剤などの記述があるものは該当する欄に丸証を、それ自体は抗ガン剤ではないが前駆体となりえるなどとしてガンについて言及しているものは抗ガン剤の欄に三角証をつけている。

	Quantum Chemical Similarity									物性	
	opt				freq		td		avg.	発ガン性物質	抗ガン剤
	orbital energy	homo	lumo	homo-lumo gap	frequencies	IR spectrum	excitation energy	UV/vis spectrum			
Tanshinone IIA	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	100.00	×	○
Mifepristone	92.20	89.05	33.96	92.64	87.21	68.62	80.68	34.39	72.34	×	○
Bortezomib	90.28	92.31	49.41	78.43	80.89	44.17	78.24	61.64	71.92	×	○
Abyssinone II	90.84	99.96	21.41	78.36	81.62	39.72	84.95	60.24	69.64	×	○
Folic acid	88.48	94.52	85.98	89.19	73.23	13.45	51.63	55.81	69.04	×	△
Curcumin	88.67	95.81	67.12	95.16	69.97	21.46	59.68	34.70	66.57	×	○
Methotrexate	90.64	98.69	60.93	90.13	80.03	18.93	42.08	45.84	65.91	×	○
Quercetin	85.34	94.48	26.59	84.64	60.20	26.47	84.65	60.67	65.38	×	○
Kaempferol	85.11	95.01	27.04	84.29	59.03	26.86	87.64	56.51	65.19	×	○
Pemetrexed	89.92	92.24	15.68	83.04	71.79	26.16	76.37	60.03	64.40	×	○
Tamoxifen	87.29	89.11	13.37	85.02	73.62	21.91	70.26	60.10	62.58	×	○
Proflavine	79.53	86.62	42.33	98.49	50.03	8.20	78.64	54.17	62.25	×	○
Amentoflavone	87.28	98.90	42.65	82.99	65.81	28.96	39.89	45.69	61.52	×	○
Crizotinib	89.40	86.47	29.77	94.13	77.87	22.89	47.78	41.85	61.27	×	○
Oroidin	81.06	90.10	12.81	83.85	64.44	22.89	75.44	50.25	60.10	×	△
Finasteride	92.30	87.81	8.35	65.47	84.98	76.71	31.54	31.63	59.85	×	○
Atorvastatin	91.27	93.69	10.10	79.56	77.80	18.83	38.15	56.90	58.29	×	○
delta-Tocotrienol	91.21	90.85	0.05	72.46	81.68	33.73	53.36	30.21	56.69	×	○
Allicin	76.26	92.28	19.03	71.64	54.73	8.76	47.64	51.20	52.69	×	○
Propofol	86.05	98.14	0.00	66.51	81.00	35.78	19.20	22.82	51.19	×	△
Betulinic acid	91.09	91.80	0.00	59.47	85.70	45.67	14.60	18.69	50.88	×	○
Nepetalactone	86.61	93.08	0.20	63.73	81.14	22.16	8.58	13.39	46.11	×	△
<i>p</i> -Dichlorobenzene	66.72	89.61	7.51	66.45	39.94	7.25	39.38	49.93	45.85	○	×
Formaldehyde	46.05	78.91	4.56	58.19	9.18	3.82	2.33	0.00	25.38	○	×

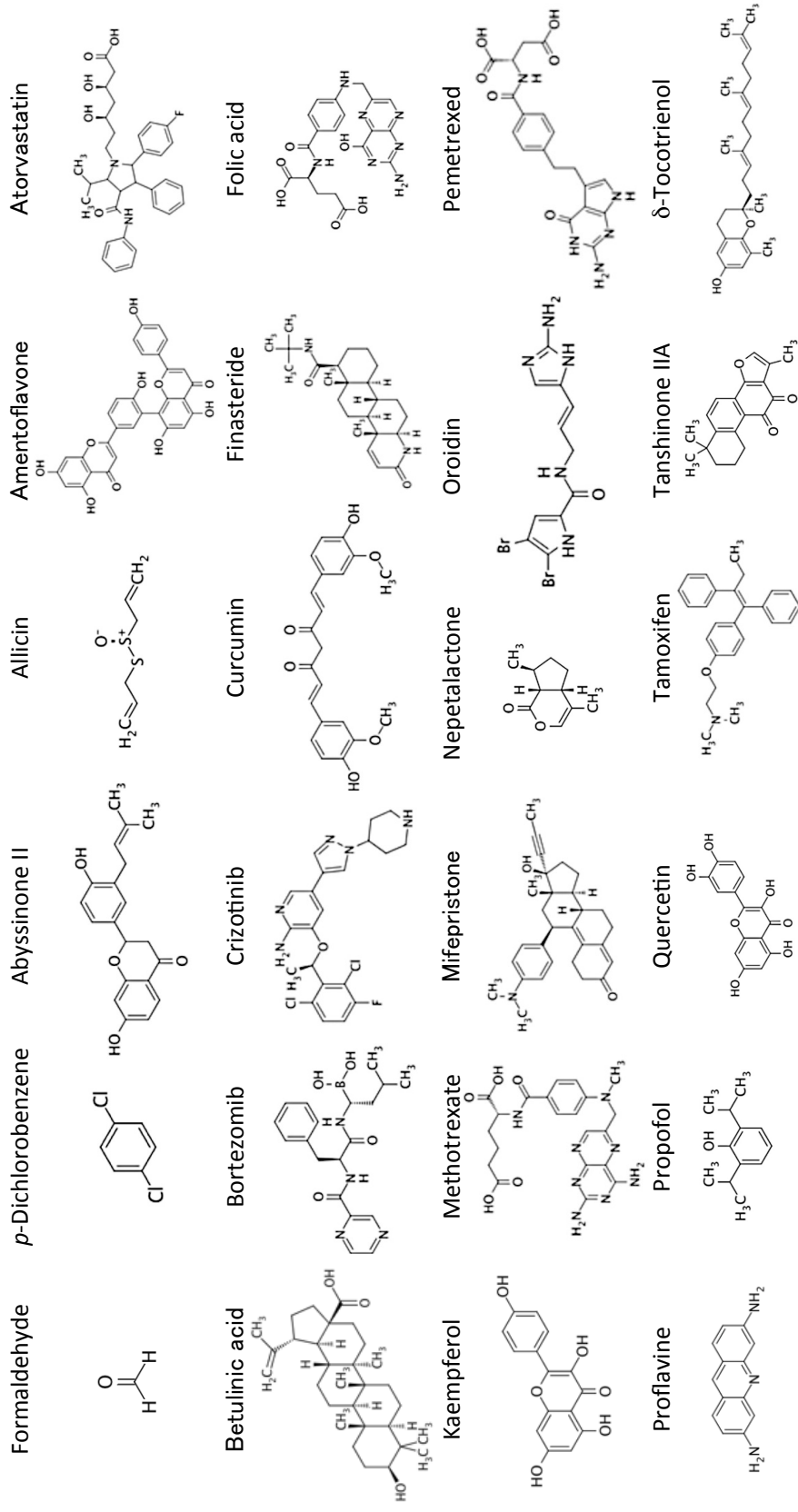


Figure 2-3. IMIS のデータベースに登録されている、「がん」に関連する化合物の構造式。

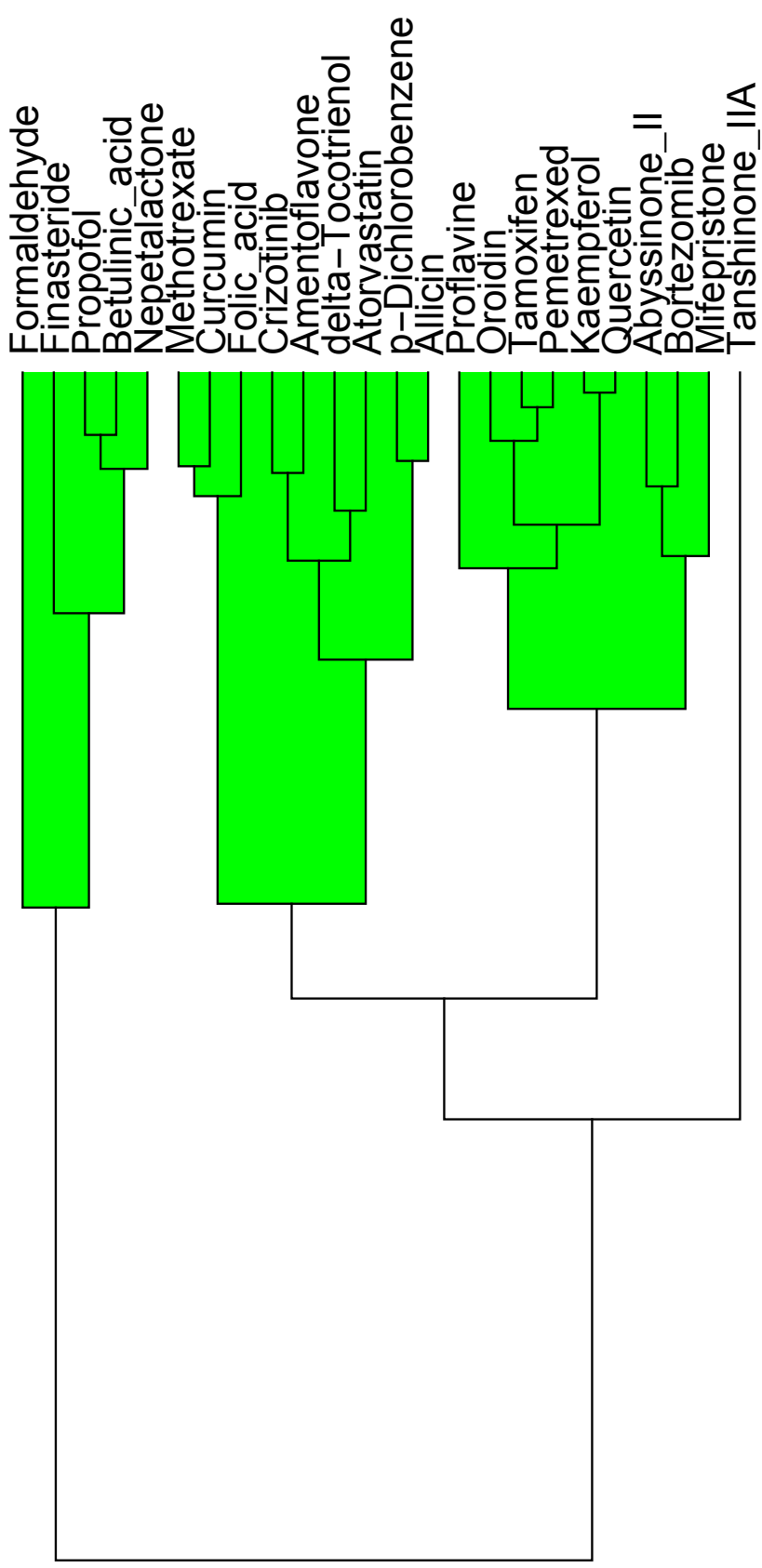


Figure 2-4. MOTW のデータベースに登録された、「がん」に関連する化合物についてのクラスタリングの結果の図。分類のために、8 種類の類似性に関する記述子を使用した (UV-Vis、励起エネルギー、軌道エネルギー、IR スペクトル、振動エネルギー、HOMO のエネルギー準位、LUMO のエネルギー準位、HOMO-LUMO ギャップ)。このデータセットのうち、formaldehyde と p-dichlorobenzene は発がん性物質で、それ以外の分子は抗がん剤である。乳がんと前立腺がんの抗がん剤である Tanshinone IIA は、分子の類似性を評価する際の参照分子である。

現していると考えられる記述子をいくつか選択し、それらの記述子のみを使用してクラスタリングを行うことにした。Fig. 2-5 は記述子ごとの類似度のランキングであり、対象とした分子群の各数値は、本質的には滑らかな変化を示している。これらのうち、HOMO のエネルギー準位、LUMO のエネルギー準位、HOMO-LUMO ギャップは、各点をつないだ線が類似度のランキングに応じてスムーズに変化している。それらと比較すると、Fig. 2-5 のうちの残りの5つのプロットは、ランキングに対する各点をつなぐ線の勾配が突然変化している箇所があることがわかる。このような勾配の変化は、分子のグループ化に関連していると考えられる。

この観察に基づき、Ward 法によるクラスタリングを行う際に、記述子を数パターン組み合わせで変更して解析を行なった。数回の試行の後、最終的に、(b) 励起エネルギースペクトル、(c) 軌道エネルギースペクトル、(g) IR スペクトル、(h) 振動スペクトル、の4つの記述子を使用した Ward 法による解析で、Fig. 2-6 に示しているように2つの発ガン性物質 (Formaldehyde と *p*-Dichlorobenzene) が含まれるクラスタリングの結果を得ることができた。抗ガン剤の一つである“Allicin”が「発ガン性物質」のクラスターに含まれているが、24 個の「ガン」関連の分子群から、3つの分子を選択した結果、2分子のみ含まれていた「発ガン性物質」を両方選択されており、他の21個の分子とは異なるカテゴリーに分類できたため、全体的に見ればかなり良好な分類ができていると言える。また、(c)と(g)と(h)のセット、(g)と(h)のセットなど、様々な記述子の組み合わせについて解析を行なった。その結果、(g)と(h)の記述子を含めた Ward 法による解析では、2つの発ガン性物質が同じクラスターに属するという結果が得られた。これらの結果は、振動に関連する特性の特徴的な違いが、抗ガン剤と発ガン性物質を区別するための良い指標であることを意味している。

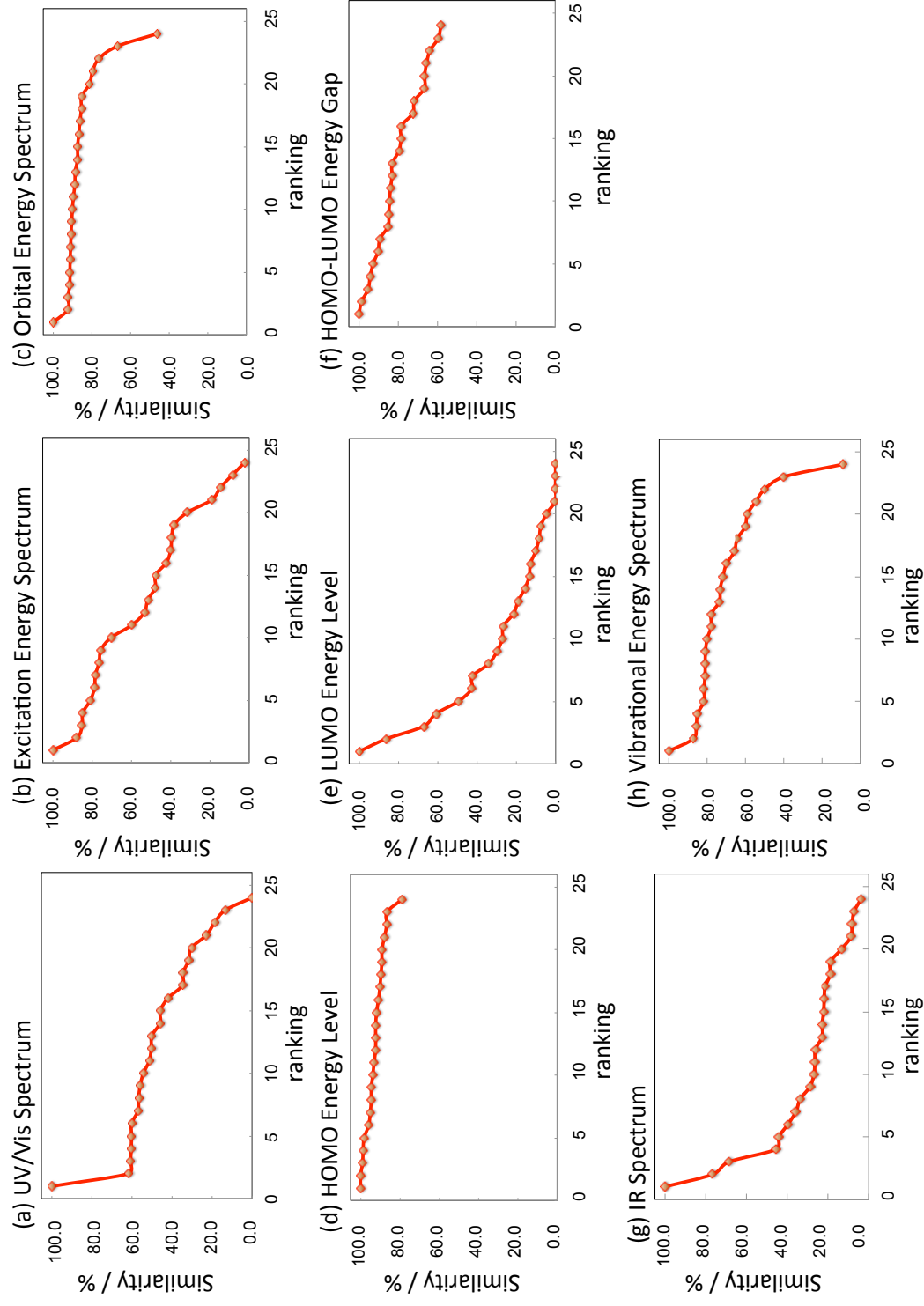


Figure 2-5. 記述子ごとの類似度のランキング。(a)UV-Vis スペクトル、(b)励起エネルギースペクトル、(c)軌道エネルギースペクトル、(d)HOMO のエネルギー準位、(e)LUMO のエネルギー準位、(f)HOMO-LUMO ギャップ、(g)IR スペクトル、および(h)振動エネルギースペクトル。

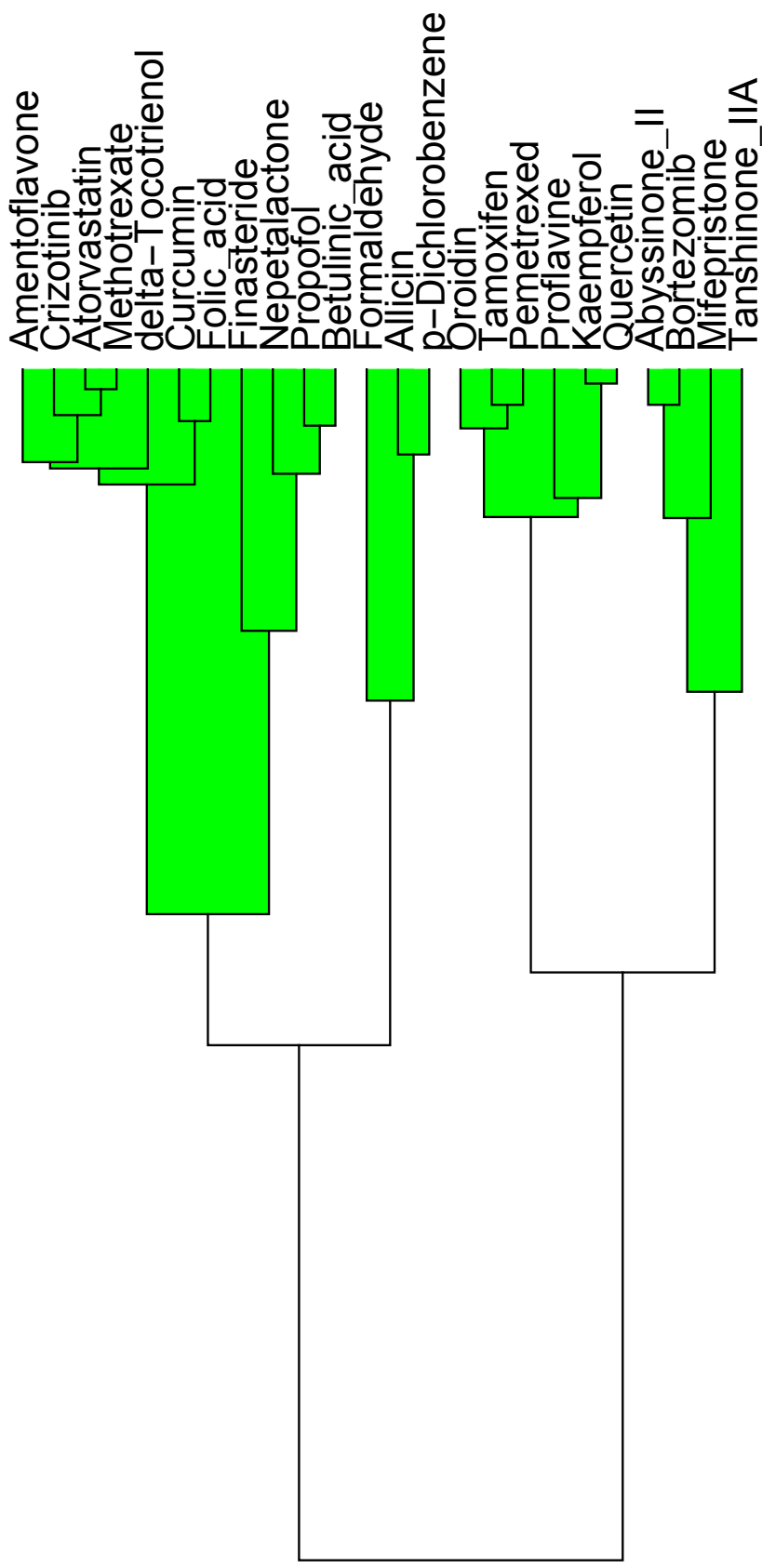


Figure 2-6. MOTW のデータベースに登録された、「がん」に関連する化合物についてのクラスタリングの結果の図。分類のために、4 種類の類似性に関する記述子を使用した (励起エネルギー、軌道エネルギー、IR スペクトル、振動エネルギー)。このデータセットのうち、formaldehyde と p-dichlorobenzene は発がん性物質で、それ以外の分子は抗がん剤である。乳がんと前立腺がんの抗がん剤である Tanshinone IIA は、分子の類似性を評価する際の参照分子である。



## 2-4 まとめ

本章では、計算化学とケモインフォマティクス的手法に基づいて知識情報を得るためにデータ分析を実施した結果について論じた。

まず、データベースに登録に使用するデータを得る手段として量子化学計算ソフトウェア“Gaussian 09”を使用した。登録したい分子の電子状態計算を行い、その計算結果のファイルから全エネルギーなどの必要な値のみを抽出してデータベースに登録できるようにすることで、分子に関する電子状態のデータベースを作成した。データベースに登録する化合物データセットとしては“Molecule of the Week (MOTW)”のデータセットを使用した。データセットの情報を基にそれらの分子の電子状態計算用のインプットファイルを作成して計算を行い、計算結果を分子情報データベースに登録した。

次に、キーワード探索によって分子を絞り込み、量子化学的類似度を用いて分子機能の解析を行った。手順としては MOTW の分子情報を登録したデータベースに対して、“cancer”をキーワードとして登録している分子の探索を行って対象とする分子を絞り込んだ。絞り込んだ分子について、それらの情報を基に量子化学的類似度の評価を行った。更に、その類似度を基に分子のクラスター解析を行った。

その結果、分子データベース検索で見つかったキーワード関連化合物について、電子的記述子を使用して分類を行うことができることが示された。これは計算化学とケモインフォマティクスを組み合わせると、研究者の関心がある一連の分子について実用的に有用な知識を提供できる可能性があることを示している。

## 2-5 参考文献

- [1] M. J. Frisch et al., Gaussian 09, Revision D.01, *Gaussian, Inc.*, Wallingford CT, 2009 ([http://www.gaussian.com/g\\_prod/g09.htm](http://www.gaussian.com/g_prod/g09.htm)).
- [2] “Molecule of the Week”, *American Chemical Society*, 2005-2015, (<https://www.acs.org/content/acs/en/molecule-of-the-week.html/>).
- [3] CLiDE standard, *Keyodule Ltd.*, Leeds, UK, 2014, (<http://www.keymodule.co.uk/products/clide/index.html>).
- [4] D. Weininger, *J. Chem. Info. Model.*, **28**, 31. (1988).
- [5] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open Babel: An open chemical toolbox." *J. Cheminf.* **3**, 33 (2011).
- [6] The Open Babel Package, version 2.3.1 (<http://openbabel.org>).
- [7] iBabel, version 2.6 (<http://www.macinchem.org/ibabel/>).
- [8] M. Sugimoto, N. Yoshida, T. Ideo, R. Iwane, E. Makiyama, H. Nagaya, T. Doi, T. Kurogi, and K. Kuroiwa, “Intelligent Molecular Informatics System (IMIS)”, *Kumamoto University* (2014).
- [9] M. Sugimoto, T. Ideo, A. B. Manggara, K. Yoshida, and T. Inoue, *J. Comput. Aided Chem.*, **20**, 65 (2019).
- [10] J. H. Ward, *J. Am. Stat. Assoc.*, **58**, 236 (1963).
- [11] Mathematica Ver. 10, *Wolfram Research Ltd.*, Champaign, IL, USA, 2014, (<http://www.wolfram.com/mathematica/?source=nav>).

## 第 3 章 脂肪酸合成酵素に対する天然物の阻害活性に関する回帰モデルの作成

### 3-1 はじめに

脂肪酸シンターゼ (FASN) は、エネルギー貯蔵のために、生体内の過剰な炭水化物を脂肪酸 (FA) に変換する酵素である[1]。ガン細胞の周囲に大量の FASN が発現していることが知られており、FASN による FA がガン細胞分裂のエネルギー源であることを示唆している[2]。したがって、FASN 阻害剤は抗ガン剤として働くことが期待されている[3]。

文献には、二次植物代謝物が FASN 阻害剤として働くことを立証する報告がある[4-15]。これらの化合物は、乳ガン、前立腺ガン、卵巣ガンなどの様々なガンに対する有効性から注目されている[16-19]。興味深いことに、FASN 阻害剤とシスプラチンの並行投与は、シスプラチンのみの投与よりも効果的であることが報告されている[20-21]。これは、FASN 阻害剤がガン治療の補完的な薬剤となりうることを意味する。免疫チェックポイント阻害剤などの高価な抗ガン剤[22]との併用は、治療の効率を高め、コストを削減したりできるため、興味深い課題である。

天然物中の非常に高活性な FASN 阻害剤を探すことは重要な研究分野であると考えられるが、この方向性の研究や Tian らによる研究[23-26]を除いて、あまり盛んに行われていない。有用な FASN 阻害剤を見つけるもっともらしい方法の一つは、天然物データベースからインシリコスクリーニングを実施することである[27]。Tian らは、既知の FASN 阻害剤の構造的特徴と量子化学計算に基づいて、定量的構造活性相関 (QSAR) 研究を実施した。彼らの研究はいくつかのユニークな特性を指摘することに成功したが、データベース検索の定量的モデルは今のところ提案されていない。

FASN は、サイズが大きく構造的に複雑な酵素、具体的には 7 つのドメインで構成さ

れるホモダイマーである[28-29]。従って、様々なリガンド分子に対して取りうる全てのドッキングパターンについて計算して調査する必要がある。この理由により、定量的な構造活性相関を理解し、特定の分子リストから高活性な FASN 阻害剤を発見するために、リガンドベースの分析と薬物設計を実行することが実用的であると思われる。

本研究では、定量分析のための回帰モデルを導き出し、FASN 阻害剤の半数阻害濃度 (IC<sub>50</sub>) 値を予測した。天然物に関する実験的研究[23-26]を参照することにより、量子化学計算から得られた数値情報を各分子に関する記述子(説明変数)として、我々の研究室で開発したケモインフォマティクス(化学情報学)手法に適用して解析を行なった。我々は本稿で言及する計算手順を「電子状態インフォマティクス」と呼んでいる。何故なら、この解析手法の背景となっている考え方は、電子状態に関する研究で培われた理論と本質的に共通しているからである。以下で説明するように、化学相互作用のエネルギー変化に関連する記述子を定義している。FASN の酵素活性は、リガンドの化学反応および分子間相互作用に関連している可能性があるため、これらの記述子は合理的である。このような反応プロセスでは、核運動に関する電子エネルギーの変化、つまり断熱近似で定義されたポテンシャルエネルギー曲線が極めて重要な役割を果たすと考えられる。したがって、エネルギーの変化に関連する記述子(「エネルギー記述子」)は重要であるといえることができる。

本研究では、エネルギー記述子を含む分子の記述子を提案し、その重要性を詳細に説明する。次に、これらを適用して、FASN 阻害剤として機能する天然物の IC<sub>50</sub> 値の回帰モデルを作成した。回帰モデルが実験との良好な相関関係を示すことに成功しているという結果が得られている。これは、電子状態計算を使用して評価された記述子が有意なものであり、化合物の物性の評価に適用できる可能性があることを示して

いる。

## 3-2 計算手法

### 3-2-1 電子的記述子

本研究で提案された全ての記述子を Table 3-1に示す。以下では、それぞれの記述子が FASN 阻害剤の性質を示すために必要なものである理由を説明する。

各分子の阻害活性は、 $IC_{50}$  値を使用して実験的に評価されている。Cheng と Prusoff は、次の式によって  $IC_{50}$  が結合親和性 ( $K_i$ ) に関連していることを示した。

$$K_i = IC_{50}/([A]/EC_{50} + 1) \quad (1)$$

式中の  $[A]$  は基質の濃度であり、 $EC_{50}$  は半数効果濃度のことを表す。本研究では、 $-\log IC_{50}$  を統計分析の際の目的変数とした。 $-\log IC_{50}$  の数値は式 (2) より得られる。

$$-\log IC_{50} = -\log K_i + const \quad (2)$$

よく知られているように、平衡定数  $K$  は、化学プロセスにおける自由エネルギーの変化に関連する可能性がある。

$$\Delta G = -RT \ln K \quad (3)$$

$K = K_i$  の場合、 $-\log IC_{50}$  は  $\Delta G$  に直線的に比例することがわかる。したがって、リガンド-受容体相互作用のエネルギー変化とそれに続くリガンド結合部位での化学変換に直接関連する記述子を使用して、 $-\log IC_{50}$  を記述することは妥当であるといえることができる。

Table 3-1. 記述子の記号および定義。 $\mathbf{R}_C$ 、 $\mathbf{R}_A$ 、 $\mathbf{R}_{T1}$ はそれぞれ、カチオン (C)、アニオン(A)、最低スピン三重項状態( $T_1$ )で構造最適化された構造である。構造の記載がない記述子はスピン一重項の基底状態( $S_0$ )で最適化された構造である。 $\rho$ はスペクトルデータであることを示す。 $S$ は、参照分子(HA-50)のスペクトルに対する対象分子のスペクトル類似性であることを示す。

Descriptor	Definition
1 $S(\rho_{UV})$	Similarity in UV/vis spectra
2 $S(\rho_{ex})$	Similarity in density of electronically excited states (ex)
3 $S(\rho_{orb})$	Similarity in density of molecular orbitals (orb)
4 $S(\rho_{IR})$	Similarity in IR spectra
5 $S(\rho_{vib})$	Similarity in density of vibrational states (vib)
6 $IP_v$	Vertical ionization potential from $S_0$
7 $\lambda_{S_0 \rightarrow C}$	Structural relaxation energy in C upon ionization
8 $\lambda_{C \rightarrow S_0}$	Structural relaxation energy in $S_0$ upon de-ionization at $\mathbf{R}_C$
9 $EA_v$	Vertical electron affinity from $S_0$
10 $\lambda_{S_0 \rightarrow A}$	Structural relaxation energy in A upon anionization
11 $\lambda_{A \rightarrow S_0}$	Structural relaxation energy in $S_0$ upon de-anionization at $\mathbf{R}_A$
12 $\Delta ST_v$	Energy of vertical transition from $S_0$ to $T_1$
13 $\lambda_{S_0 \rightarrow T_1}$	Structural relaxation energy in $T_1$ after transition from $S_0$
14 $\lambda_{T_1 \rightarrow S_0}$	Structural relaxation energy in $S_0$ after transition from $T_1$ at $\mathbf{R}_{T1}$
15 $V_{mol}$	Molecular volume
16 $B_1$	Length of the longest edge of the cuboid box enclosing a molecule
17 $B_2$	Length of the second longest edge of the cuboid box enclosing a molecule
18 $B_3$	Length of the shortest edge of the cuboid box enclosing a molecule
19 $\mu$	Dipole moment
20 $\Delta E_{solv}$	Solvation energy at the optimized geometry in vacuum of $S_0$

### 3-2-1-1 電荷移動プロセスに関するエネルギー記述子

生成物が既知あるいは予測可能であれば電子状態計算を使用して、その系のエネルギー変化を数値的に評価することが可能であるが、特定の分子セットでは必ずしも可能ではない。本章では、電子状態の遷移に関する3種類のエネルギー変化を、 $-\log IC_{50}$  の記述子として用いることとした。それらは、イオン化ポテンシャル ( $IP$ )、電子親和力 ( $EA$ )、およびS-Tギャップ ( $\Delta E_{ST}$ ) である。最後の1つは、スピン一重項基底状態 ( $S_0$  状態) と最低のスピン三重項 ( $T_1$  状態) のエネルギー差である。

これらのエネルギー記述子が重要であると考えられる理由を以下に示す。よく知られているように、有機化学における多くの反応は、求核攻撃または求電子攻撃によって怒ると考えられている。これらの概念は、局所的な電子分布、つまり基質間の局所的な酸化還元反応を考慮に入れている。したがって、分子の化学反応性は、酸化還元プロセスに関連する物理的性質に強く依存すると予想される。分子のこれらの酸化還元特性を反映する典型的な特性は、 $IP$  と  $EA$  である。これらの量は、それぞれ最高被占軌道 (HOMO) および最低空軌道 (LUMO) のエネルギーレベルと強い相関関係があることが知られている。したがって、それらは福井のフロンティア軌道理論[31]の観点からも重要であると考えられている。

$\Delta E_{ST}$  の重要性は、 $T_1$  状態が最低のスピン一重項励起状態 ( $S_1$  状態) と同じ電子配置をもっていることから理解できる。したがって、 $\Delta E_{ST}$  は  $S_0 \rightarrow S_1$  への励起のエネルギー変化を反映することが期待される。このエネルギー差は、HOMO-LUMO ギャップにも対応している。化学反応のポテンシャルエネルギー曲面 (PSE) の遷移状態において、基底状態と励起状態の PSE の交差を避けるため、励起エネルギーは小さくなる。

共有結合を切断すると、2つのラジカル中心が生成され、準縮退により、基底状態にある系は開殻特性をもつ。そのような場合、 $\Delta E_{ST}$  ははるかに小さくなり、均一結合開裂における化学反応性の尺度になる。

$IP$ 、 $EA$ 、 $\Delta E_{ST}$  のそれぞれについて、2つの定義が文献で知られている。1つは分子構造が維持される「垂直」遷移エネルギーで、もう1つは「断熱」遷移である。後者は、初期状態と最終状態の潜在的な表面上の2つのエネルギー最小値間の状態遷移である。垂直遷移におけるエネルギー変化は、 $IP_v$ 、 $EA_v$ 、および $\Delta E_{ST,v}$  などのように「v」を下付き文字で追加することで表現した。これらを算出する際には、構造最適化されたジオメトリの  $S_0$  状態からの状態遷移に対して定義している。同様に、断熱エネルギーの変化は、 $IP_a$ 、 $EA_a$ 、および $\Delta E_{ST,a}$  などのように「a」を添えることで示した。

それらの定義により、垂直遷移エネルギーと断熱遷移エネルギーは互いに等しくない。これらのエネルギーを使用して、例えば式(4)によって得られるエネルギー差( $\lambda$ )を定義できる。

$$\lambda = IP_v - IP_a \quad (4)$$

この量は、マーカスの電子移動の理論[32]に基づいて、再配向エネルギーと呼ばれる。また、脱励起時にも同様のエネルギー変化を定義することが可能である。

再配向エネルギー( $\lambda$ )は、陽イオンへの励起と陽イオンからの脱励起の両方に対して以下のように定義可能である。

$$\lambda_{S_0 \rightarrow c} = E_c(\mathbf{R}_{S_0}) - E_c(\mathbf{R}_c) \quad (5)$$

$$\lambda_{c \rightarrow S_0} = E_{S_0}(\mathbf{R}_c) - E_{S_0}(\mathbf{R}_{S_0}) \quad (6)$$

式中の、 $E_{S_0}$  および  $E_c$  は、それぞれ  $S_0$  およびイオン化(カチオン)状態の電子エネルギーを意味する。 $\mathbf{R}_{S_0}$  および  $\mathbf{R}_c$  は、それぞれ  $S_0$  およびイオン化状態のポテンシャルエ



エネルギー曲面上のエネルギー最小値に対応する各座標を意味する。同様に、式(7)と式(8)で $\lambda_{S_0 \rightarrow X}$ と $\lambda_{X \rightarrow S_0}$ を定義した。

$$\lambda_{S_0 \rightarrow X} = E_X(\mathbf{R}_{S_0}) - E_X(\mathbf{R}_X) \quad (7)$$

$$\lambda_{X \rightarrow S_0} = E_{S_0}(\mathbf{R}_X) - E_{S_0}(\mathbf{R}_{S_0}) \quad (8)$$

式中の X は A および T<sub>1</sub> のいずれかであり、それぞれ電子結合状態および T<sub>1</sub> 状態に対応する。これらの再配向エネルギーの重要性は、有機正孔輸送材料[33]、有機金属反応[34]、および半導体の欠陥物理学[35]に関する論文で提案されている。豊沢によって提案された配位座標モデル[36]によって、それらのような再配向エネルギーは、さまざまな分子および材料の重要な特性である双線形に結合した、電子-フォノン(あるいは電子-振動)相互作用との間に相関があることが示される。

軌道エネルギーではなく、 $IP_v$ 、 $EA_v$ 、および $\Delta E_{ST,v}$ を使用する理由として2つの理由が挙げられる。1つ目は、後者の数値が実験から定量的に観測可能であること、2つ目は、自己無撞着場計算における軌道緩和[37]を考慮に入れるためである。これらの量を使用するうえでの欠点は、それら算出する為の計算コストの高さである。本研究では、多くの電子系のエネルギー変化と対応する数値である、再配向エネルギーを算出する際に同時に定義できる値であるため、 $IP_v$ と $EA_v$ を用いて解析を行なった。

### 3-2-1-2 分子間相互作用に関するエネルギー記述子

電子状態遷移に関連するエネルギー記述子は、分子の化学反応性を記述するのに適切であると考えられているが、生物系の分子間相互作用と直接的な関係はない。この欠陥に対処するために、分子間相互作用に関連する記述子を追加した。例えば、静電相互作用に関しては、電子状態計算を行なった際に算出された、分子の双極子

モーメントが含まれる。これは、双極子-双極子間相互作用が極性溶媒において重要な役割を果たすからである。

ファンデルワールス相互作用も分子間相互作用において重要と考えられる。このタイプの相互作用は、Lennard-Jones ポテンシャルを使用することで十分に説明されることが知られている。この場合、分子の粗視化モデルでのサイズ(つまり、球体近似での分子の半径)が重要なパラメータになる。このような形状に関する記述子について考慮するために、以下に示すようないくつかの記述子を定義した。

リガンドと受容体の相互作用の「鍵と鍵穴モデル」の観点から、たんぱく質のリガンド結合部位のトポロジカルな特徴が重要であると予想される。分子認識プロセスでは、主にヘテロ原子によるマルチサイト相互作用が重要になる。そのような局所的な相互作用を考慮するために、エネルギー記述子よりもトポロジー記述子が機械学習をするうえでより有用であると考えられる。

このようなトポロジカルな特徴は、水素結合やハロゲン結合などの局所的な静電相互作用から生じると考えるのが自然であると考えられる。それらは、ホスト-ゲスト化学において重要な役割を果たす。エネルギー記述子としては、静電相互作用による溶媒和エネルギー( $\Delta E_{\text{solv}}$ )が妥当であると考えられる。この値は、電子状態計算で分極連続体モデル(Polarizable continuum model: PCM)計算を行うことで算出できる。

$\Delta E_{\text{solv}}$  は別の観点からも重要であると考えられている。生体系内のリガンドとタンパク質のドッキングプロセスでは、リガンド周辺の水分子が除去される。この「脱溶媒和」プロセスは、リガンドが多くの水分子と強く結合している場合には、より大きなエネルギーを要する。したがって、ドッキングプロセスは $\Delta E_{\text{solv}}$  と強く関係している可能性がある。

分子の形状に関連する記述子として知られているものとしては分子体積が挙げられ

る。この記述子は、ケモインフォマティクスに関する多くの研究で使用されている[38]。本研究では、電子状態計算によって算出された分子体積の数値を使用している。この記述子はスカラー量であり、分子サイズの尺度となるため、ドッキング即ち分子認識のプロセスにおいて重要であると考えられる。分子量は有用で興味深いトポロジカルな記述子であるが、分子の特性の等方的な記述にすぎない。言い換えると、トポロジーはスカラー量で記述される。分子構造の異方性をパラメータ化するために、我々は分子を囲む直方体のサイズを記述子として活用することを試みた。この手順を「ボックスパッキング(あるいはボクシング)メソッド」と呼ぶ。この方法では、分子をちょうどぴったり覆う直方体を定義して、直方体の各辺の長さを降順に  $B_1$ 、 $B_2$ 、 $B_3$  のようにラベリングして定義している。ジオメトリを最適化した分子を囲む直方体のサイズの計算は、Mathematica 11.0[39]を使用して行なった。

### 3-2-1-3 電子応答と分子振動を説明するエネルギー記述子

上記の記述子に加えて、分子の電子的特徴を説明する追加の 5 つの記述子をさらに導入した。1 つ目は、電子構造法を使用して計算された紫外/可視 (UV / Vis) スペクトルである。スペクトルデータを計算するために、励起エネルギーだけでなく遷移双極子モーメントも計算される。二次摂動論は、静電界に対する電子の柔らかさを反映し、電気分極率に関連する可能性がある電気分極率が、次の式で示すように、これら 2 つの物理的特性に依存することを示している。

$$\alpha_{tu} \propto \sum_I \frac{\langle \Psi_0 | \sum_i t_i | \Psi_I \rangle \langle \Psi_I | \sum_i u_i | \Psi_0 \rangle}{E_I - E_0} \quad (9)$$

ここで、 $\Psi_I$  と  $E_I$  は、それぞれ I 番目の状態の電子波動関数とエネルギーであり、 $I =$

0 は基底状態を意味する。 $t_i$  と  $u_i$  は、光による電場のテイラー展開の主要項から生じる  $i$  番目の電子のデカルト座標の成分であり、電子と光(光子)の相互作用を表す。これは、記述子としての UV / Vis スペクトルの考慮が、ファンデルワールス相互作用の起源である分散相互作用を含むことに暗黙的に対応することを意味している。

UV / Vis スペクトルは、電気分極率に関連する興味深い物理的特性であるが、ゼロ以外の遷移双極子モーメントを持つ励起状態のみを反映する。言い換えれば、限られた電子励起状態のみがスペクトルデータに含まれる。全体的な励起状態考慮できるようにするために、励起状態の密度を記述子として定義することにした。このスペクトルデータは、「励起状態の状態密度」とも言い換えることができるため、ex-DOS として表す。「状態密度」の DOS の略語は、固体物理学で凝縮物質の電子特性の議論においてよく知られている。このスペクトルデータは、エネルギー単位あたりの励起状態の数を数えることで得られる。実際にはこのスペクトルデータは、ガウスバンドを狭いバンド幅で重ね合わせることで取得できる。なお、この際の各バンド幅は、それぞれの分子について計算された励起エネルギーの数値を反映している。

ここで注意すべき点は、固体物理学において「状態」とは、電子的に励起された状態ではなく、独立粒子近似の軌道のことを意味するという点である。軌道は多電子波動関数ではなく、一電子波動関数であるが、固体物理学における DOS も記述子として含めているため、区別のためにそれを orb-DOS と表している。orb-DOS は ex-DOS と同等であるが、ex-DOS では、計算される軌道の数に制限され、本研究では 30 個の軌道が計算される。ex-DOS に対応するエネルギーの幅は狭い範囲であるため、より広い範囲をカバーするために orb-DOS もエネルギー記述子として使用している。

現在の統計分析の 4 番目と 5 番目の記述子は、赤外線 (IR) スペクトルと振動状態

の密度 (vib-DOS) である。後者は、ex-DOS および orb-DOS と同様の方法で定義される。IR スペクトルは、IR 光の吸収に寄与する振動モードを反映している。この領域のモードは分子の官能基を反映しているため、この記述子の使用は、ケモインフォマティクス研究における分子の構造フィンガープリントと同様の意味を持つ[40]。Holstein モデル[41]は、振動周波数と電子-フォノン相互作用の強さのバランスが分子と凝縮物質の特性にとって重要であることを示唆しているため、振動モードの情報を含めることも意味があると考えられる。記述子として vib-DOS を導入した理由は、IR スペクトルに振動レベルの選択規則があるからである。

電子状態計算で得られるのは、線状のスペクトル群だけである。全体的なスペクトルデータを生成するための畳み込みにおいては、次の式を使用して、各ピークにガウシアンブロードニング法を適用した。

$$\rho = \sum_i \beta_i \exp[-\alpha(\epsilon - \epsilon_i)^2] \quad (10)$$

$$\alpha = \ln 0.5 / (\Delta\epsilon)^2 \quad (11)$$

式中の $\Delta\epsilon$ はガウスバンドの半値全幅であり、それらの値は Table 3-2 にまとめてある。なお、UV/Vis や IR の各ピークにおける頻度因子 ( $\beta_i$ ) は、それらの数値の振動子強度や IR 強度の値をバンド面積として再現するよう経験的に決定した。ex-DOS、orb-DOS、vib-DOS については、 $\beta_i$  はいずれも 1.0 に設定した。

スペクトルデータは二次元的な特性をもっており記述子として直接使用することはできない。本研究では、これらをスカラー変数に変換するために、以下の式で定義されるターゲット分子と参照分子のスペクトルの重なりを評価し、記述子として使用した。

$$S = \int_0^W \rho(\epsilon) \rho_r(\epsilon) d\epsilon \quad (12)$$

**Table 3–2.** 化合物の分光学的特徴に由来する電子的記述子の計算に用いる定数の定義。 $f_i$ と $g_i$ はそれぞれ、 $i$ 番目の状態の振動子強度とIR強度である。

Spectral data	$x_i$	$\Delta x$	$\beta_i$
$\rho_{UV}$	electronic excitation energy	0.30 [eV]	$f_i\sqrt{\alpha/\pi}$
$\rho_{ex}$	electronic excitation energy	0.15 [eV]	1
$\rho_{orb}$	molecular orbital energy	0.02 [hartree]	1
$\rho_{IR}$	vibrational energy	20 [cm <sup>-1</sup> ]	$g_i\sqrt{\alpha/\pi}$
$\rho_{vib}$	vibrational energy	20 [cm <sup>-1</sup> ]	1

式中の $\rho$ と $\rho_r$ はそれぞれ、ターゲット分子と参照分子の正規化されたスペクトル関数である。インテグラルの上端にあたる「 $w$ 」は、類似度を評価するための範囲のサイズを表している。理想的には無限大である必要があるが、実際の計算では有限に設定している。以降、計算された UV/Vis、ex-DOS、orb-DOS、IR、および vib-DOS スペクトルは、それぞれ $\rho_{UV}$ 、 $\rho_{ex}$ 、 $\rho_{orb}$ 、 $\rho_{IR}$ 、および $\rho_{vib}$ と表現する。

### 3-2-2 統計解析、および機械学習

本研究では、Fig. 3-1 に示す植物由来の 45 種の化合物に着目した。これらの化合物は Tian らによって報告されたものである[23-26]。Table 3-3 にある 50%阻害濃度 (IC<sub>50</sub>) は、実験によって得られた、各分子の薬理学的効果の指標となる数値である。これらの IC<sub>50</sub> の実験値を再現する予測式を導出するために、重回帰分析を実行した。解析には後述する電子的記述子を説明変数として使用した。重回帰分析では、まずすべての記述子を含めて分析を使用して分析を行い、すべての説明変数で得られる有意確率 ( $p$  値) が 0.05 以下になるまで、説明変数を1つずつ削除する手順を繰り返した。

また、ランダムフォレスト回帰 (RFR) [42] も実行した。実行に際して Python3.7.1 の Scikit-learn ライブラリを使用した。今回実行した解析では、重要度が低い記述子の一つを体系的に削除することにより、重要な記述子が選択した。このプロセスを 100 回繰り返し、もっとも重要でないと考えられる記述子を削除した。予測の精度が大幅に低下するまで、この手順を繰り返し行なった。

また、同様のデータセットを使用してサポートベクター回帰 (SVR) [43] も実行した。SVR でも記述子選択を行い、記述子を一つずつ削除してその前後で R<sup>2</sup> 値が大きく変

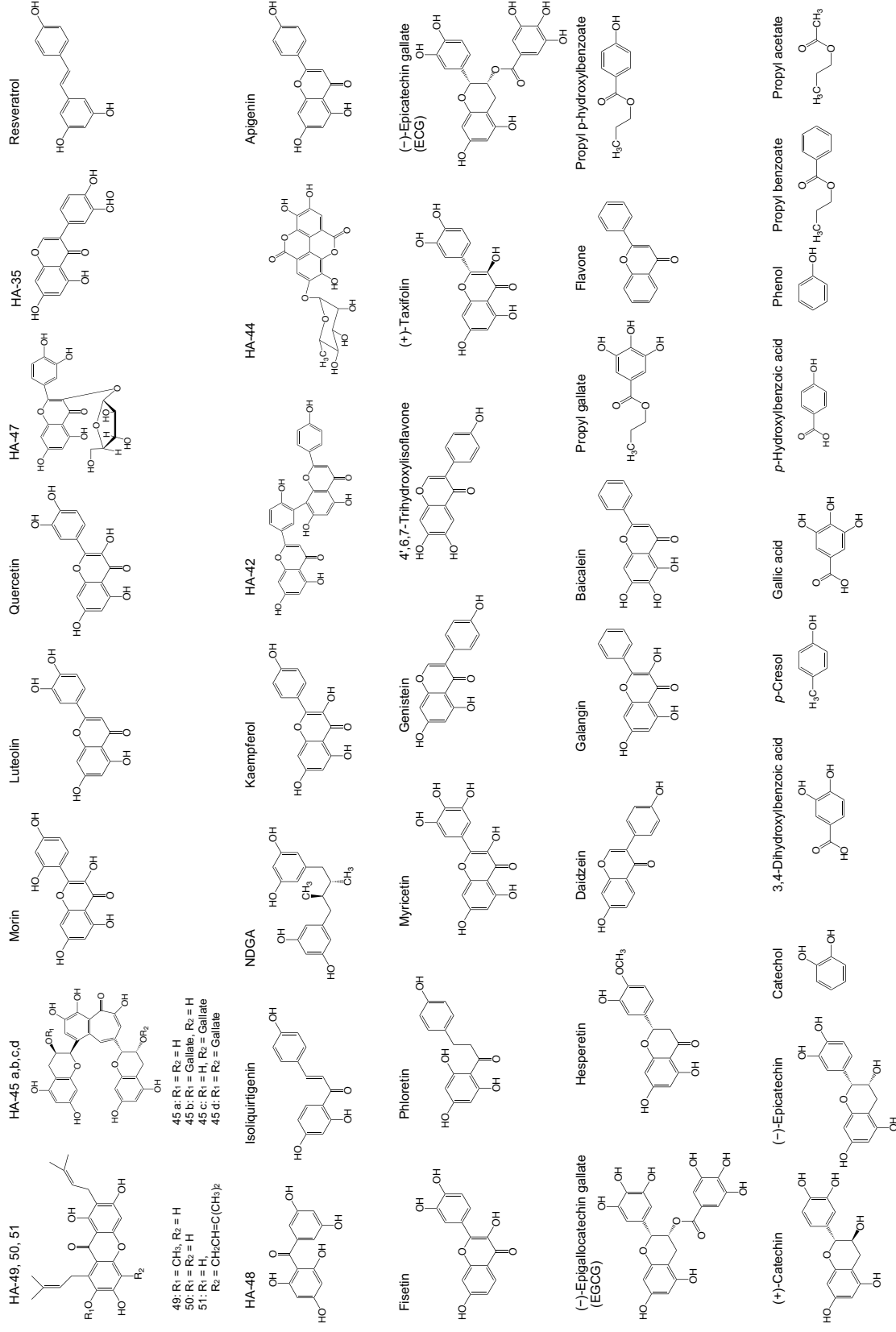


Figure 3-1. 本研究で解析対象とした FASN 阻害剤の構造式。これらの化合物の IC<sub>50</sub> の数値については、Table 3-3 を参照。



**Table 3–3.** FASN 阻害剤の 50%阻害濃度(IC<sub>50</sub>)の値。IC<sub>50</sub> の単位を M にし、-log(IC<sub>50</sub>)に変換した値を統計解析に用いた。

Compound name	Ref.	Experimental value		Predicted value	
		IC <sub>50</sub> ( $\mu$ M)	-log(IC <sub>50</sub> )	IC <sub>50</sub> ( $\mu$ M)	-log(IC <sub>50</sub> )
HA-50	[26]	1.24	5.9066	1.89	5.7235
HA-45a	[26]	1.77	5.7520	2.70	5.5688
HA-45c	[26]	1.77	5.7520	1.01	5.9938
HA-45b	[26]	1.77	5.7520	1.47	5.8327
HA-45d	[26]	1.77	5.7520	5.50	5.2599
Morin	[24]	2.33	5.6326	1.31	5.8814
Luteolin	[24]	2.52	5.5986	18.21	4.7398
HA-51	[26]	3.3	5.4815	4.82	5.3170
Quercetin	[24]	4.29	5.3675	6.73	5.1718
HA-49	[26]	5.54	5.2565	3.90	5.4095
HA-47	[26]	6.15	5.2111	9.78	5.0097
HA-35	[26]	7.7	5.1135	19.41	4.7119
Resveratrol	[25]	8.5	5.0706	22.68	4.6444
HA-48	[26]	8.59	5.0660	56.72	4.2463
Isoliquirtigenin	[25]	8.8	5.0555	19.15	4.7179
NDGA	[25]	9.3	5.0315	75.49	4.1221
Kaempferol	[24]	10.38	4.9838	9.02	5.0449
HA-42	[26]	13	4.8861	5.18	5.2853
HA-44	[26]	17	4.7696	23.61	4.6268
Apigenin	[25]	17.6	4.7545	24.69	4.6075
Fisetin	[24]	18.78	4.7263	35.41	4.4509
Phloretin	[25]	26.1	4.5834	22.01	4.6573
Myricetin	[24]	27.18	4.5658	21.48	4.6680
Genistein	[25]	28.7	4.5421	16.29	4.7881
4,6,7-Trihydroxyisoflavone	[25]	29.5	4.5302	38.97	4.4092
(+)-Taxifolin	[24]	41.16	4.3855	16.21	4.7901
ECG	[23]	42	4.3768	5.69	5.2447
EGCG	[23]	52	4.2840	6.37	5.1958
Hesperetin	[24]	68.86	4.1620	16.17	4.7913
Daidzein	[25]	73.2	4.1355	80.91	4.0920
Galangin	[24]	100	4.0000	37.00	4.4318
Baicalein	[24]	111.69	3.9520	66.38	4.1779
Propyl gallate	[23]	500	3.3010	1375.11	2.8617
Flavone	[25]	1000	3.0000	1174.43	2.9302
Propyl <i>p</i> -hydroxybenzoate	[23]	1100	2.9586	2448.51	2.6111
(+)-Catechin	[23]	1600	2.7959	2798.41	2.5531
(-)-Epicatechin	[23]	3800	2.4202	1686.12	2.7731
Catechol	[23]	7400	2.1308	12618.40	1.8990
3,4-Dihydroxybenzoic acid	[23]	9000	2.0458	6296.97	2.2009
<i>p</i> -Cresol	[23]	16000	1.7959	3063.66	2.5138
Gallic acid	[23]	21000	1.6778	3944.98	2.4040
<i>p</i> -Hydroxybenzoic acid	[23]	22000	1.6576	13254.10	1.8776
Phenol	[23]	26000	1.5850	31610.30	1.5002
Propyl benzoate	[23]	30000	1.5229	36651.47	1.4359
Propyl acetate	[23]	30000	1.5229	20952.20	1.6788

化するまで、記述子の削除と  $R^2$  値の評価を行うことで記述子の絞り込みを行った。手順は以下の(i)-(iv)である。

- (i) データセットを 4:1 の比率でトレーニングデータとテストデータに分割した。この際の分割はランダムに実行した。
- (ii) トレーニングデータに SVR を適用し、回帰モデルを導出した。そしてそのモデルの  $R^2$  値を評価した。
- (iii) 同じ回帰モデルを使用してテストデータの相関を確認し、その  $R^2$  値を評価した。
- (iv) トレーニングデータとテストデータの  $R^2$  値の平均値を計算した。

このプロセスを 100 回繰り返し、(iv)で産出された  $R^2$  の平均値を参照して、トレーニングデータとテストデータの最適な分割を選択した。今回実行した SVR の解析では、RBF カーネル (radial basis function kernel; 動径基底関数カーネル)と、過学習を防ぐためにグリッドサーチを使用して手動で最適化したハイパーパラメータを合わせて解析を行った。今回の SVR のモデルで最適なハイパーパラメータは、 $C = 100.0$ 、 $\varepsilon = 0.00100$ 、 $\gamma = 0.03125$  であった。

### 3-2-3 電子状態計算

FASN 阻害能があると報告されている分子を対象として、電子状態計算を実行した。全ての電子状態計算は Gaussian 09 (Rev. D01) [44]を使用して実行した。まず、各分子の構造最適化を DFT(M06-2X)/6-31G(d,p)法で行った。中性状態に加えて、カチオン状態、アニオン状態、スピン三重項状態の構造最適化も実行した。分子振動に関する情報を得るために、中性状態での基準振動解析を行った。各状態への遷移に伴う構造緩和エネルギーの計算は、遷移前の平衡構造のエネルギーと遷移後の平衡構

造のエネルギーとのエネルギー差を計算することによって評価した。中性状態の分子については、分子体積と溶媒和エネルギーも計算した。これは阻害剤(リガンド)がタンパク質に組み込まれて結合したときに重要な要素と見なされるためである。分子量の計算はモンテカルロ積分で行うため、5 回繰り返し行い、それらの平均値を分子量の最終的な数値とした。水中の溶媒和エネルギーは、分極連続体モデル (polarizable continuum model, PCM) 法を使用して計算した。さらに、電子励起状態は、TDDFT(M06-2X)/6-31G(d,p)法を使用して計算した。この計算では、エネルギー的に低い 30 の励起状態を計算した。

### 3-3 結果と考察

Fig. 3-1 に示す 45 種の天然物について統計解析を実行した。これらの分子は FASN 阻害剤として報告されており、 $IC_{50}$  の数値は文献に記載されている[23-26]。Fig. 3-1 の分子には、フラボン、イソフラボン、フラボノール、フラバンなどのフラボノイドが多く含まれており、六員環が多数含まれている。ここに記載されている化合物の多くは、複数のヒドロキシル基をもつポリフェノールに分類できる。

DFT 計算を適用して、分子構造の最適化を実行した。また、分子のポテンシャルエネルギー面のエネルギー最小値を特定するために、固有振動解析も行った。計算されたイオン化ポテンシャル、例としてフェノールと酢酸プロピルの値はそれぞれ 8.49 と 10.27 eV であった。これらは実験値 (8.50 および 10.04 eV) とよく一致している[45,46]。したがって、DFT 法は、この報告で提案された電子的記述子を評価するために適用可能であると考えられる。電子的記述子の数値は、Table 3-4 から 3-10 にまとめてある。

**Table 3-4.** HOMO と LUMO の軌道エネルギー準位、およびそれらの間のエネルギーギャップ。

Compound Name	HOMO (hartree)	LUMO (hartree)	HOMO- LUMO gap (hartree)
(-)-Epicatechin	-0.25424	0.04438	0.29862
(+)-Catechin	-0.25103	0.04566	0.29669
(+)-Taxifolin	-0.25404	-0.01245	0.24159
3,4-Dihydroxybenzoic acid	-0.27380	-0.00142	0.27238
4,6,7-Trihydroxyisoflavone	-0.25549	-0.01392	0.24157
Apigenin	-0.26410	-0.01638	0.24772
Baicalin	-0.27252	-0.02283	0.24969
Catechol	-0.25915	0.04358	0.30273
HA-35	-0.26642	-0.01850	0.24792
HA-42	-0.25931	-0.02518	0.23413
HA-44	-0.28147	-0.04401	0.23746
HA-45a	-0.24113	-0.00783	0.23330
HA-45b	-0.23866	-0.01404	0.22462
HA-45c	-0.25009	-0.02921	0.22088
HA-45d	-0.25578	-0.03817	0.21761
HA-47	-0.25854	-0.01761	0.24093
HA-48	-0.26601	-0.02317	0.24284
HA-49	-0.25870	-0.00795	0.25075
HA-50	-0.25103	-0.00784	0.24319
HA-51	-0.24924	-0.00637	0.24287
Daidzein	-0.25674	-0.01519	0.24155
ECG	-0.25301	-0.00645	0.24656
EGCG	-0.25243	-0.00597	0.24646
Fisetin	-0.24669	-0.02974	0.21695
Flavone	-0.28428	-0.03219	0.25209
Galangin	-0.25237	-0.02589	0.22648
Gallic acid	-0.27236	-0.00134	0.27102
Genistein	-0.25209	-0.00396	0.24813
Hesperetin	-0.25927	0.00089	0.26016
Isoliquirtigenin	-0.26212	-0.02374	0.23838
Kaempferol	-0.24311	-0.02191	0.22120
Luteolin	-0.26015	-0.01580	0.24435
Morin	-0.24855	-0.00413	0.24442
Myricetin	-0.24960	-0.02490	0.22470
NDGA	-0.25814	0.04383	0.30197
<i>p</i> -Cresol	-0.26239	0.03744	0.29983
<i>p</i> -Hydroxybenzoic acid	-0.28643	-0.00003	0.28640
Phenol	-0.27159	0.03662	0.30821
Phloretin	-0.26120	0.01846	0.27966
Propyl acetate	-0.34190	0.06606	0.40796
Propyl benzoate	-0.30905	-0.00582	0.30323
Propyl gallate	-0.27484	-0.00299	0.27185
Propyl <i>p</i> -hydroxybenzoate	-0.28271	0.00392	0.28663
Quercetin	-0.24150	-0.02125	0.22025
Resveratrol	-0.23969	-0.01137	0.22832

**Table 3–5.** 分子体積 ( $V_{\text{mol}}$ ) (5 回の計算の平均値)、双極子モーメント( $\mu$ )、溶媒和エネルギー ( $\Delta E_{\text{solv}}$ )。それぞれ DFT 法を使用して計算した。溶媒効果の計算には PCM 法を使用した。

Compound Name	$V_{\text{mol}}$ (cm <sup>3</sup> /mol)	$\mu$ (debye)	$\Delta E_{\text{solv}}$ (eV)
(-)-Epicatechin	178.34200	2.8899	0.590
(+)-Catechin	204.62860	3.6427	0.577
(+)-Taxifolin	197.20120	2.6374	0.669
3,4-Dihydroxybenzoic acid	99.83460	0.9588	0.329
4,6,7-Trihydroxyisoflavone	176.59680	1.7883	0.482
Apigenin	166.28440	4.8389	0.667
Baicalein	178.31520	3.5513	0.552
Catechol	77.37700	2.5523	0.230
HA-35	202.31740	4.3781	0.591
HA-42	322.95600	0.8751	1.235
HA-44	238.47900	0.4767	0.782
HA-45a	347.76580	5.9176	1.223
HA-45b	445.89420	0.7095	1.330
HA-45c	418.27680	2.9978	1.237
HA-45d	487.70920	5.0510	1.240
HA-47	270.71100	4.8633	1.057
HA-48	170.93180	4.6158	0.627
HA-49	307.40700	3.6335	0.467
HA-50	304.71000	4.5303	0.483
HA-51	359.15700	4.6605	0.494
Daidzein	165.02740	2.9471	0.463
ECG	285.49080	3.3727	0.716
EGCG	282.32580	4.3058	0.744
Fisetin	172.72140	0.9525	0.476
Flavone	155.61920	4.0518	0.264
Galangin	170.76560	4.4068	0.500
Gallic acid	101.26080	4.5855	0.383
Genistein	191.11600	3.7046	0.644
Hesperetin	194.72220	2.2655	0.627
Isoliquirtigenin	191.89100	2.5692	0.642
Kaempferol	178.03200	5.3568	0.608
Luteolin	185.06960	2.1563	0.714
Morin	197.34560	6.1364	0.912
Myricetin	198.33160	3.7121	0.728
NDGA	224.48920	3.8477	0.596
<i>p</i> -Cresol	83.44160	1.3400	0.184
<i>p</i> -Hydroxybenzoic acid	89.91100	1.6223	0.306
Phenol	78.16560	1.3697	0.186
Phloretin	194.08620	6.4322	0.710
Propyl acetate	85.54180	1.6606	0.139
Propyl benzoate	120.20480	1.6872	0.155
Propyl gallate	157.48360	3.0702	0.386
Propyl <i>p</i> -hydroxybenzoate	131.13140	1.0477	0.261
Quercetin	173.37940	1.7535	0.652
Resveratrol	163.45540	2.2714	0.503

**Table 3-6.**  $S_0$  状態で最適化した分子構造での垂直イオン化ポテンシャル( $IP_v$ )。  $\lambda_{S_0 \rightarrow C}$  と  $\lambda_{C \rightarrow S_0}$  は、それぞれ陽イオン状態および  $S_0$  状態への垂直遷移後の構造緩和エネルギーの数値を示す。

Compound Name	$IP_v$ (eV)	$\lambda_{S_0 \rightarrow C}$ (eV)	$\lambda_{C \rightarrow S_0}$ (eV)
(-)-Epicatechin	7.707	0.318	0.282
(+)-Catechin	7.676	0.323	0.320
(+)-Taxifolin	7.771	0.328	0.321
3,4-Dihydroxybenzoic acid	8.438	0.302	0.310
4,6,7-Trihydroxyisoflavone	7.697	0.320	0.327
Apigenin	7.916	0.252	0.271
Baicalein	8.182	0.188	0.241
Catechol	8.132	0.288	0.292
HA-35	7.943	0.302	0.329
HA-42	7.614	0.255	0.264
HA-44	8.340	0.464	0.423
HA-45a	7.174	0.668	0.103
HA-45b	7.075	0.408	0.387
HA-45c	7.348	0.351	0.365
HA-45d	7.469	0.490	0.273
HA-47	7.660	0.961	0.446
HA-48	8.063	0.312	0.347
HA-49	7.662	0.279	0.542
HA-50	7.471	0.272	0.311
HA-51	7.352	0.295	0.313
Daidzein	7.747	0.341	0.333
ECG	7.637	0.351	0.402
EGCG	7.624	0.357	0.386
Fisetin	7.415	0.279	0.287
Flavone	8.461	0.220	0.252
Galangin	7.564	0.286	0.300
Gallic acid	8.373	0.330	0.337
Genistein	7.609	0.331	0.334
Hesperetin	7.849	0.360	0.325
Isoliquirtigenin	7.898	0.307	0.457
Kaempferol	7.303	0.299	0.311
Luteolin	7.827	0.241	0.236
Morin	7.456	0.370	0.405
Myricetin	7.474	0.638	0.168
NDGA	7.760	0.153	0.155
<i>p</i> -Cresol	8.159	0.245	0.248
<i>p</i> -Hydroxybenzoic acid	8.799	0.249	0.253
Phenol	8.489	0.234	0.235
Phloretin	7.957	0.439	0.244
Propyl acetate	10.267	1.182	4.086
Propyl benzoate	9.355	0.226	0.229
Propyl gallate	8.414	0.317	0.326
Propyl <i>p</i> -hydroxybenzoate	8.635	0.253	0.259
Quercetin	7.267	0.285	0.296
Resveratrol	7.244	0.234	0.234

**Table 3-7.** S<sub>0</sub> 状態以最適化した分子構造での垂直電子親和力( $EA_v$ )。  $\lambda_{S_0 \rightarrow A}$  と  $\lambda_{A \rightarrow S_0}$  は、それぞれ陰イオン状態および S<sub>0</sub> 状態への垂直遷移後の構造緩和エネルギーの数値を示す。

Compound Name	$EA_v$ (eV)	$\lambda_{S_0 \rightarrow A}$ (eV)	$\lambda_{A \rightarrow S_0}$ (eV)
(-)-Epicatechin	-2.024	2.166	2.705
(+)-Catechin	-2.016	1.682	2.772
(+)-Taxifolin	-0.446	0.349	0.372
3,4-Dihydroxybenzoic acid	-0.968	0.374	0.371
4,6,7-Trihydroxyisoflavone	-0.397	0.278	0.476
Apigenin	-0.268	0.333	0.306
Baicalein	-0.105	0.358	0.308
Catechol	-2.348	0.475	0.620
HA-35	-0.231	0.277	0.507
HA-42	0.062	0.351	0.360
HA-44	0.476	0.295	0.298
HA-45a	-0.436	1.148	0.403
HA-45b	-0.186	1.003	0.683
HA-45c	0.224	0.456	0.477
HA-45d	0.499	1.170	0.595
HA-47	-0.147	0.435	0.493
HA-48	-0.082	0.462	0.461
HA-49	-0.393	0.236	0.257
HA-50	-0.411	0.293	0.281
HA-51	-0.401	0.291	0.283
Daidzein	-0.373	0.278	0.497
ECG	-0.596	0.514	0.518
EGCG	-0.609	0.522	0.527
Fisetin	0.097	0.246	0.309
Flavone	0.127	0.300	0.270
Galangin	-0.003	0.260	0.257
Gallic acid	-0.952	0.405	0.410
Genistein	-0.649	0.346	0.623
Hesperetin	-0.815	0.283	0.360
Isoliquirtigenin	-0.074	0.327	0.313
Kaempferol	-0.102	0.241	0.251
Luteolin	-0.274	0.392	0.407
Morin	-0.595	0.382	0.405
Myricetin	-0.007	0.272	0.286
NDGA	-1.927	0.272	0.649
<i>p</i> -Cresol	-2.140	0.262	0.434
<i>p</i> -Hydroxybenzoic acid	-1.022	0.352	0.351
Phenol	-2.190	0.257	0.460
Phloretin	-1.292	0.622	0.375
Propyl acetate	-3.010	1.018	1.569
Propyl benzoate	-0.790	0.365	0.362
Propyl gallate	-0.804	0.428	0.440
Propyl <i>p</i> -hydroxybenzoate	-1.023	0.380	0.377
Quercetin	-0.114	0.277	0.337
Resveratrol	-0.417	0.269	0.265

**Table 3–8.** S<sub>0</sub> 状態で最適化した分子構造での S<sub>0</sub> 状態から T<sub>1</sub> 状態への垂直遷移エネルギー ( $\Delta E_{ST,v}$ )。  $\lambda_{S_0 \rightarrow T_1}$  と  $\lambda_{T_1 \rightarrow S_0}$  は、それぞれ T<sub>1</sub> 状態および S<sub>0</sub> 状態への垂直遷移後の構造緩和エネルギーの数値を示す。

Compound Name	$\Delta E_{ST,v}$ (eV)	$\lambda_{S_0 \rightarrow T_1}$ (eV)	$\lambda_{T_1 \rightarrow S_0}$ (eV)
(-)-Epicatechin	4.326	0.672	1.818
(+)-Catechin	4.304	0.666	1.811
(+)-Taxifolin	3.953	0.399	0.551
3,4-Dihydroxybenzoic acid	3.818	0.449	0.439
4,6,7-Trihydroxyisoflavone	3.598	0.747	1.496
Apigenin	3.529	0.639	0.703
Baicalein	3.563	0.677	1.004
Catechol	4.600	0.901	1.984
HA-35	3.900	0.979	1.930
HA-42	3.438	0.565	0.740
HA-44	3.518	0.357	0.350
HA-45a	3.237	1.546	1.515
HA-45b	2.969	1.648	0.708
HA-45c	2.821	0.891	1.118
HA-45d	2.780	1.034	0.974
HA-47	3.386	0.989	1.279
HA-48	3.767	0.627	0.713
HA-49	3.931	0.490	0.734
HA-50	3.735	0.370	0.391
HA-51	3.747	0.379	0.387
Daidzein	3.595	0.786	1.405
ECG	4.848	1.619	0.782
EGCG	4.872	1.357	2.050
Fisetin	2.891	0.483	0.657
Flavone	3.526	0.614	0.592
Galangin	2.961	0.547	0.855
Gallic acid	3.814	0.509	0.718
Genistein	3.667	0.851	1.966
Hesperetin	4.060	0.444	0.489
Isoliquirtigenin	3.204	0.910	2.211
Kaempferol	2.928	0.518	0.856
Luteolin	3.526	0.659	0.592
Morin	3.349	0.850	1.388
Myricetin	2.948	0.538	0.900
NDGA	5.875	2.055	2.015
<i>p</i> -Cresol	4.550	0.701	1.781
<i>p</i> -Hydroxybenzoic acid	4.065	0.469	0.465
Phenol	4.727	0.240	0.238
Phloretin	3.752	0.722	0.668
Propyl acetate	5.596	1.217	2.143
Propyl benzoate	4.301	0.594	0.589
Propyl gallate	4.055	0.556	0.545
Propyl <i>p</i> -hydroxybenzoate	4.080	0.477	0.476
Quercetin	2.936	0.528	0.870
Resveratrol	3.111	0.929	2.794



**Table 3-9.** 中性基底状態で最適化された分子を囲む直方体の、各辺の長さ(Å 単位)。

Compound Name	$B_1$	$B_2$	$B_3$
(-)-Epicatechin	13.8559	8.1845	6.4718
(+)-Catechin	14.1309	8.6840	5.3985
(+)-Taxifolin	14.4014	8.5761	5.4348
3,4-Dihydroxybenzoic acid	10.0364	7.4114	3.3968
4,6,7-Trihydroxyisoflavone	15.2980	7.3140	4.3665
Apigenin	14.1347	8.7515	3.8601
Baicalein	14.1654	8.6908	4.0166
Catechol	7.6456	7.2869	3.3924
HA-35	15.0784	8.1836	4.6230
HA-42	17.0230	14.2738	7.1366
HA-44	17.0141	9.3198	5.7182
HA-45a	15.6477	12.2098	7.5901
HA-45b	14.4828	13.5498	10.0704
HA-45c	15.3198	12.1233	8.2901
HA-45d	15.3778	11.2011	10.4080
HA-47	14.2369	13.3016	5.9819
HA-48	12.5254	8.1087	5.6129
HA-49	16.7582	11.3213	6.6068
HA-50	15.7289	10.8868	6.2997
HA-51	15.6006	12.5251	7.0625
Daidzein	15.3379	7.0368	4.3531
ECG	14.3376	9.5411	8.5444
EGCG	14.3868	10.1388	8.5115
Fisetin	14.4336	9.1358	3.6463
Flavone	13.5786	8.3800	3.9514
Galangin	13.5316	8.8311	3.6955
Gallic acid	9.5733	8.0244	3.3982
Genistein	15.2854	7.7784	4.4744
Hesperetin	15.7595	8.3532	5.7069
Isoliquirtigenin	14.9956	8.0289	5.4546
Kaempferol	14.1266	8.8561	3.5135
Luteolin	14.4053	8.9263	3.9394
Morin	14.1889	8.4235	5.2428
Myricetin	14.1915	9.1587	3.7164
NDGA	11.1430	10.5504	7.7159
<i>p</i> -Cresol	8.9249	6.5440	4.0556
<i>p</i> -Hydroxybenzoic acid	9.9136	6.6733	3.3967
Phenol	7.3279	7.2459	3.3946
Phloretin	14.4200	7.5650	6.4525
Propyl acetate	9.4919	5.1767	4.8530
Propyl benzoate	12.4641	5.8666	5.4411
Propyl gallate	13.1102	7.6950	4.9474
Propyl <i>p</i> -hydroxybenzoate	13.2310	5.8472	5.4840
Quercetin	14.4187	9.1708	3.6403
Resveratrol	14.6514	7.8360	3.4567

**Table 3–10.** スペクトルデータの類似度。最も活性が HA-50 を参照分子とした。

Compound Name	$S(\rho_{UV})$ (%)	$S(\rho_{ex})$ (%)	$S(\rho_{orb})$ (%)	$S(\rho_{IR})$ (%)	$S(\rho_{vib})$ (%)
(-)-Epicatechin	43.7	60.0	89.2	62.8	73.6
(+)-Catechin	45.4	56.5	89.9	57.5	71.5
(+)-Taxifolin	53.6	54.6	91.3	67.8	77.6
3,4-Dihydroxybenzoic acid	62.9	18.1	80.7	35.2	54.0
4,6,7-Trihydroxyisoflavone	60.5	72.0	89.5	47.7	67.1
Apigenin	60.9	54.9	90.1	46.1	73.2
Baicalin	82.0	76.6	89.0	42.2	64.7
Catechol	46.9	20.7	72.9	26.1	47.8
HA-35	77.9	89.1	89.0	54.7	71.0
HA-42	60.1	37.0	91.5	47.8	71.8
HA-44	80.5	80.1	91.1	26.2	72.1
HA-45a	46.1	47.0	93.5	63.4	78.2
HA-45b	46.8	31.1	93.7	59.1	81.4
HA-45c	60.9	28.5	93.5	60.2	80.1
HA-45d	28.1	17.9	93.9	56.1	77.4
HA-47	72.2	92.8	92.4	55.4	78.1
HA-48	55.7	67.0	85.3	61.4	65.0
HA-49	96.2	93.1	97.5	88.0	95.3
HA-50	100.0	100.0	100.0	100.0	100.0
HA-51	92.6	89.3	98.2	84.8	95.6
Daidzein	63.0	66.1	88.2	62.5	62.8
ECG	45.9	87.8	92.3	54.9	75.3
EGCG	42.9	89.3	91.9	52.2	73.8
Fisetin	71.8	71.6	88.3	48.1	63.8
Flavone	79.2	66.7	84.3	40.0	61.1
Galangin	72.9	66.2	88.1	61.5	65.5
Gallic acid	43.1	21.9	81.6	23.9	56.5
Genistein	72.8	75.9	88.6	60.1	67.0
Hesperetin	54.5	65.0	91.2	65.4	78.7
Isoliquirtigenin	51.0	66.7	87.0	57.1	66.7
Kaempferol	69.6	72.6	88.1	60.3	65.2
Luteolin	72.2	71.7	89.3	49.8	66.4
Morin	82.5	81.7	89.7	70.8	68.8
Myricetin	68.7	79.5	88.9	63.0	65.5
NDGA	38.6	47.0	86.6	52.2	78.7
<i>p</i> -Cresol	48.9	17.7	80.2	22.7	65.7
<i>p</i> -Hydroxybenzoic acid	80.7	15.8	79.2	27.5	53.0
Phenol	45.0	18.7	71.7	34.0	45.5
Phloretin	59.8	59.4	89.0	58.8	75.5
Propyl acetate	0.1	1.4	81.7	16.1	63.7
Propyl benzoate	49.3	25.8	84.1	12.6	68.7
Propyl gallate	52.5	30.9	88.4	36.8	67.4
Propyl <i>p</i> -hydroxybenzoate	82.7	19.9	86.9	25.7	71.1
Quercetin	74.1	74.8	88.4	58.0	66.8
Resveratrol	31.3	55.9	85.5	44.9	62.4

### 3-3-1 重回帰分析

FASN 阻害剤の  $IC_{50}$  値を再現するための定量的なモデル式を導出するために、重回帰 (multiple linear regression、MLR) モデルの適用を試みた。今回の報告で行った解析では、 $-\log IC_{50}$  を目的変数として使用し、3-2-2 で説明されているように線形モデリングを適用した。ここで、いくつかの記述子はエネルギーの変化には関係がないため、線形モデリングに先験的な正当性がないということに注意が必要である。

MLR での最初の解析では、全て (20 種) の記述子を使用した。この解析の際の予測値の  $R^2$  値は 0.933 であった。この解析は成功しているように見えるが、ほとんどの記述子 (説明変数) の  $p$  値は 0.05 を超えていた。MLR 解析の制度を向上させるために、すべての  $p$  値が 0.05 未満になるまで、最大の  $p$  値をもつ記述子を削除して逐次解析を行った。

変数選択を行った結果、次に示す回帰モデルが得られた。

$$\begin{aligned} -\log(IC_{50})_{pred.} &= 4.15225 + (0.387532)S(\rho_{ex}) + (0.327216)S(\rho_{vib}) \\ &+ (-0.291392)B_2 + (0.814166)\Delta E_{solv} + (-0.560680)IP_v \\ &+ (-0.249938)\lambda_{S_0 \rightarrow C} + (0.488313)\lambda_{C \rightarrow S_0} + (-0.417311)\lambda_{A \rightarrow S_0} \quad (13) \end{aligned}$$

この際の  $R^2$  値は 0.910 であった。Fig. 3-2 に散布図を示す。本研究での解析では、2 つの記述子を任意に組み合わせた際の相関 (決定係数) は 0.8 未満であり、MLR の解析結果はかなり良好であることを示している。これは、本章で提案されている記述子を使用した線形モデリングの手法が有用であることを示している。分子の形状に関連

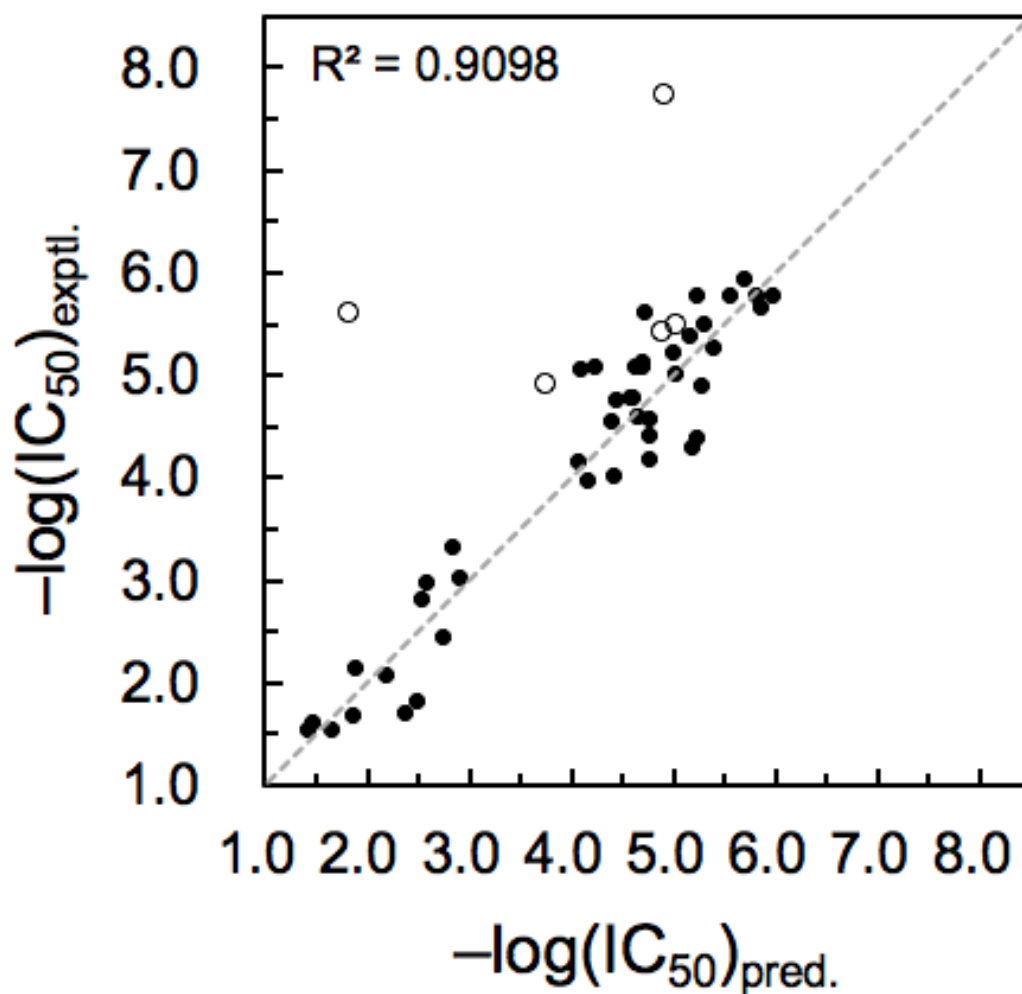
する記述子だけでなく、分子のエネルギーに関係する記述子も統計的に重要な変数として含まれていることを示唆しているのは興味深いことである。これは、FASN への結合だけでなく、阻害剤の化学的変換も  $IC_{50}$  にとって重要であることを意味している。前者は医薬品科学の基本概念である鍵と鍵穴モデルと一致している。一方で後者は不可逆的阻害が天然物における FASN 阻害剤の阻害作用の潜在的なメカニズムであることが知られているため、合理的である。

式(13)について解釈すると、分子の  $IC_{50}$  がより小さな値となる条件を示唆している。

- (i) 候補分子の ex-DOS および vib-DOS は、参照分子のものと類似している必要がある。
- (ii) B2 値が小さい、すなわち薄い分子が優れている。
- (iii) 溶媒と強く静電的に相互作用する ( $\Delta E_{\text{solv}}$  が大きい) 分子が優れている。
- (iv)  $IP_v$  が小さい分子の方が優れている。
- (v) 再配向エネルギーも重要であり、各要素の寄与はそれぞれの状態遷移の性質に依存する。すなわちイオン化と脱アニオン化には小さい再配向がより好ましく、脱イオン化にはより大きな再配向がより良い。

式(13)の 8 つの記述子のうち 3 つは、分子の陽イオン状態に関連したものであり、酸化プロセスが阻害反応に重要であることを意味している。そのプロセスに荷電種が含まれる場合、溶媒和 ( $\Delta E_{\text{solv}}$ ) が重要だとする条件 (iii) は妥当なものであると考えられる。

条件 (iv) と (v) は、化学反応に関連するパラメータが重要であることを示唆しており、 $IC_{50}$  値が FASN によって引き起こされた酸化反応に影響を受ける可能性があることを示唆している。FASN は脂肪酸を生成する化学変換のための酵素であるため、これ



**Figure 3-2.**  $-\log \text{IC}_{50}$  の実験値と予測値の相関図。予測値は、変数削減法で選択された記述子を使用した回帰モデルを使用して算出した。黒い丸は MLR のトレーニングデータを示し、白い丸はテスト分子のデータである(分子構造は Figure 3-3 を参照)。

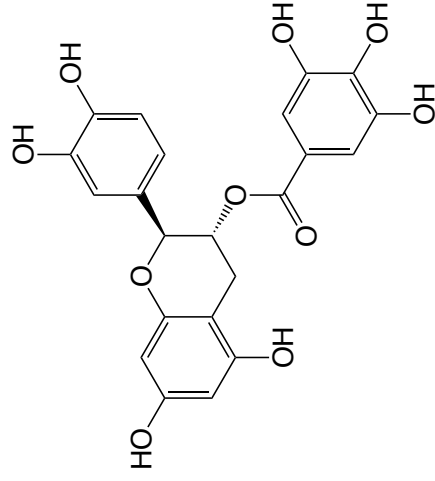
らの推測は合理的なものであると考えられる。条件(v)に関しては、 $\lambda_{so-c}$  と  $\lambda_{c-so}$  の符号は互いに反対のものとなっているが、これら 2 つのパラメータの大きさは、どちらも中性基底状態と最低のイオン化状態の最適化構造の形状を反映しているため、ある程度相関があるように考えられる。

回帰モデルの適用性を調査するために、Fig. 3-3 に挙げられている 5 つの化合物の記述子を計算した[26,47]。これら化合物に関する全ての数値データは Table 3-11 にまとめてある。それらの分子の中で、CG と Mullen-01 の予測は実験と合理的に一致したが、2 つの化合物 (Mullen-02 と Mullen04) の予測値は実験値からかなりずれたものであった。Mullen-03 の予測値の偏差も顕著なものであった。これらの偏差は、今回の回帰モデルが主に OH 基をもつ化合物に適していることを意味している。それらは、リガンドの構造の違いにより阻害活性のメカニズムの違いから生じる可能性もある。偏差の理由を明確にするために、より多くの実験データを MLR 解析のトレーニングデータに使用する必要がある。

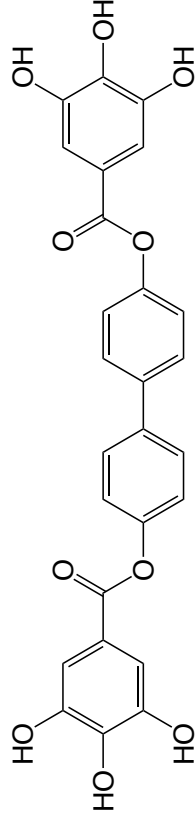
### 3-3-2 ランダムフォレストとサポートベクターマシンアルゴリズムを使用した回帰分析

Fig. 3-2 に示される MLR の解析結果 ( $IC_{50}$  の予測値とそれらの分子の  $IC_{50}$  の実験値についての散布図) について、全体的な相関は  $R^2$  値 (0.910) の観点からは良好なものであるが、より高性能な FASN 阻害剤を見つけたり設計したりするには、はるかに高い精度が要求されると思われる。Fig. 3-2 には各データの点からなる 2 つのクラスターが確認できる。一方は  $-\log(IC_{50})$  が大きい、すなわち  $IC_{50}$  が小さい高性能な分子のクラスターであるが、依然として各点は広い範囲に分布している。より定量的な予測の可能性を検討するために、RF と SVM 回帰分析について、MLR 解析で使用したのと

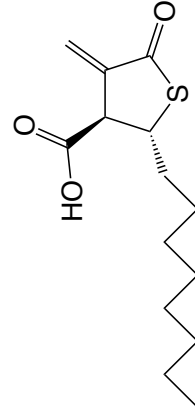
(-)-Catechin gallate (CG)



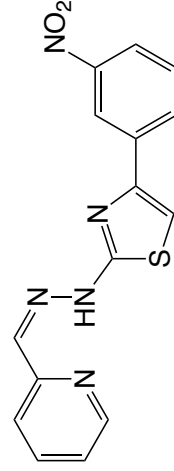
Mullen-01



Mullen-02



Mullen-03



Mullen-04

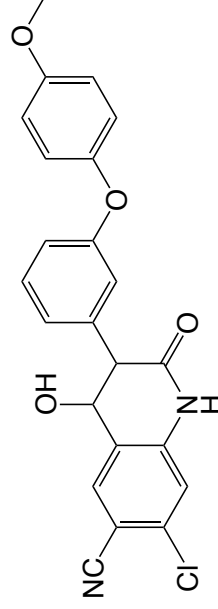


Figure 3-3. 重回帰分析の結果を検証するためのテストデータとした分子の構造式。

**Table 3-11.** Figure 3-3 に示すテストデータの分子について、計算された記述子の値。IC<sub>50</sub>の単位を M にし、 $-\log(\text{IC}_{50})$ に変換して統計分析で使用した。

Descriptor	CG	Mullen-01	Mullen-02	Mullen-03	Mullen-04
$IP_v$ (eV)	7.622	7.72	9.363	7.659	7.507
$\lambda_{S0 \rightarrow C}$ (eV)	0.282	0.423	0.379	0.351	0.293
$\lambda_{C \rightarrow S0}$ (eV)	0.352	0.444	0.387	0.304	0.286
$EA_v$ (eV)	-0.886	-0.22	-0.131	0.424	0.728
$\lambda_{S0 \rightarrow A}$ (eV)	0.457	0.188	0.415	0.455	0.271
$\lambda_{A \rightarrow S0}$ (eV)	0.458	0.293	0.43	0.443	0.27
$B_1$ (Å)	13.5708	25.7915	15.5893	11.5795	19.8677
$B_2$ (Å)	12.9831	7.4996	7.3673	8.9074	9.269
$B_3$ (Å)	8.0973	6.5011	6.9626	7.4236	6.2357
$S(\rho_{\text{ex}})$ (%)	89.5	66.7	29.9	53.3	61.3
$S(\rho_{\text{vib}})$ (%)	73.6	65.4	74.8	68.6	73.3
		(Predicted Value)			
IC <sub>50</sub> (μM)	9.0	12.9	14660.8	171.9	11.8
$-\log(\text{IC}_{50})$	5.04741670	4.88931810	1.83384320	3.76473494	4.92863120
Reference	[26]	(Experimental Value)	[44]	[44]	[44]
IC <sub>50</sub> (μM)	3.4	4.0	2.56	12.7	0.019
$-\log(\text{IC}_{50})$	5.46852108	5.39794001	5.59176003	4.89619628	7.7212464

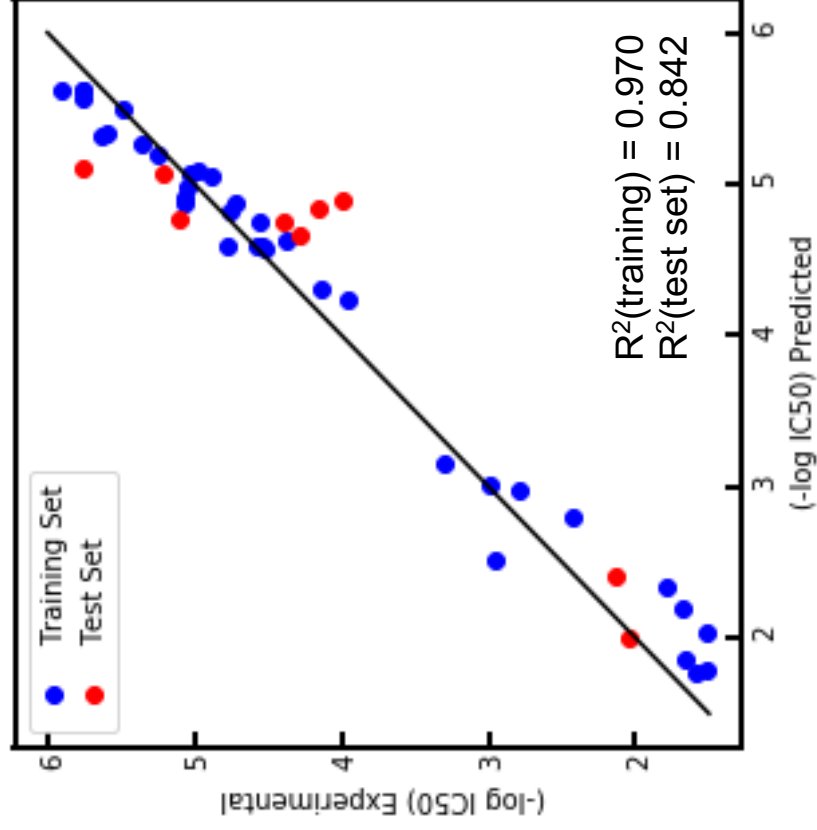


同じ記述子を使用して解析を行なった。これらの解析では、Kennard–Stone アルゴリズム[48]を適用して、80:20 の比率に設定したトレーニングセットとテストセットを定義した。予測値と実験値の相関関係を Fig. 3-4 に示す。トレーニングセットの全体的な相関は、MLR 解析と比較して改善された。今回の場合、ランダムフォレストアルゴリズムがサポートベクターマシンアルゴリズムよりも優れていることがわかる。

### 3-4 まとめ

我々の研究室では、化学相互作用のエネルギー変化に関連する記述子を使用した電子状態インフォーマティクスを提案している。この手法を適用して、抗ガン剤の開発を念頭に置いて FASN 阻害剤の定量的回帰モデルを導出した。今回の解析では、20 種の記述子を使用して、阻害活性に線形モデリングを適用した。これらの記述子のほとんどは、電子状態計算を使用して算出した。MLR 分析の  $p$  値を参照して説明変数を体系的に排除することにより、8 つの記述子が選択された。決定係数 ( $R^2$ ) が十分に大きいため、この統計モデリングは妥当なものであるといえる。モデルに適用した 2 つのテスト分子の予測値も実験値とよく一致していた。選択された記述子のほとんどは、分子認識に関連するものではなく、レドックス反応または状態変化を伴う化学反応に関連するものであった。MLR 分析に加えて、ランダムフォレストを用いた回帰分析を適用した。この場合、予測モデルの精度が向上した。これらの統計分析は、この研究で提案された記述子のセットが、FASN 阻害剤、特に OH 基を有するものの  $IC_{50}$  値の定量的記述に有用であることを示している。

(a) Random Forest



(b) Support Vector Machine

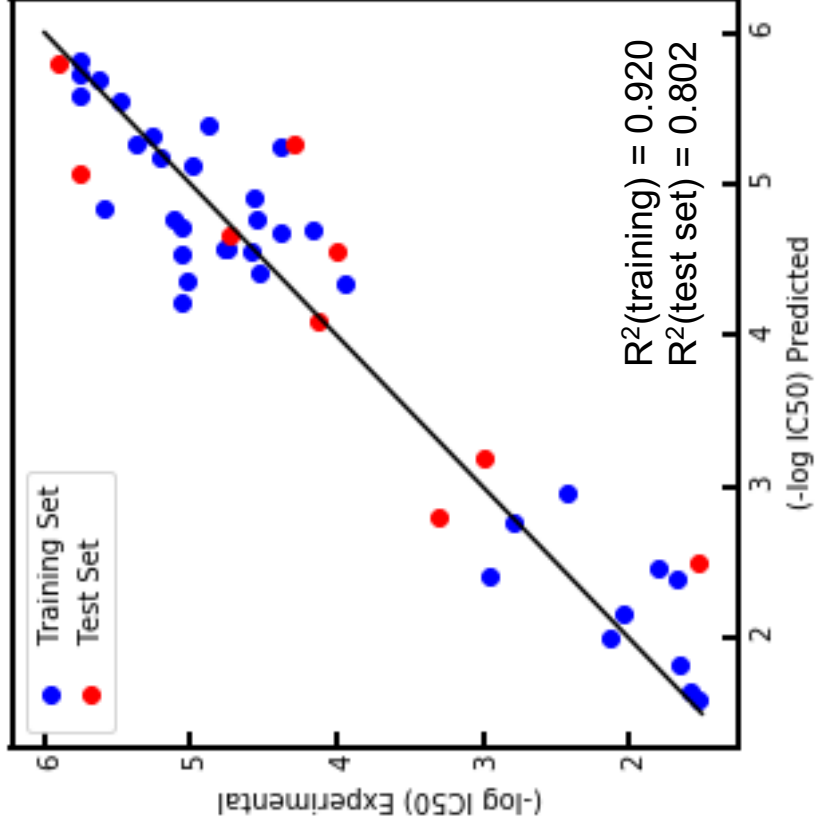


Figure 3-4. ランダムフォレスト回帰とサポートベクターマシン回帰の結果。青と赤の丸は、それぞれトレーニングセットとテストセットに対応する。

### 3-5 参考文献

- [1] T. Maier, S. Jenni, and N. Ban, *Science*, **311**, 1258-1262 (2006).
- [2] E. Currie, A. Schulze, R. Zechner, T. C. Walther, and R. V. Farese Jr., *Cell Metab.*, **18**, 153-161 (2013).
- [3] J. A. Menendez, and R. Lupu, *Nat. Rev. Cancer*, **7**, 763-777 (2007).
- [4] X. Wang, and W. Tian, *Biochem. Biophys. Res. Commun.*, **288**, 1200-1206 (2001).
- [5] Y.-X. Zhao, W.-J. Liang, H.-J. Fan, Q.-Y. Ma, W.-X. Tian, H.-F. Dai, H.-Z. Jiang, N. Li, and X.-F. Ma, *Carbohydr. Res.*, **346**, 1302-1306 (2011).
- [6] Y. Wang, S.-Y. Zhang, X.-F. Ma, and W.-X. Tian, *J. Enzyme Inhib. Med. Chem.*, **21**, 87-93 (2006).
- [7] S. Shukla, S. Gupta, *Pharm. Res.*, **27**, 962-978 (2010).
- [8] W.-X. Tian, L.-C. Li, X.-D. Wu, and C.-C. Chen, *Life Sci.*, **74**, 2389-2399 (2004).
- [9] H. Fan, D. Wu, W. Tian, and X. Ma, *Biochim. Biophys. Acta*, **1831**, 1260-1266 (2013).
- [10] Y. Liang, D. Luo, X. Gao, and H. Wu, *RSC Adv.*, **8**, 8112-8117 (2018).
- [11] H. Fan, W. Tian, and X. Ma, *Targ. Oncol.*, **9**, 279-286 (2014).
- [12] P. Zhao, J.-M. Mao, S.-Y. Zhang, Z.-Q. Zhou, Y. Tan, and Y. Zhang, *Oncol. Lett.*, **8**, 765-769 (2014).
- [13] W.-H. Zhao, C. Gao, Y.-X. Zhang, and W.-X. Tian, *J. Enzyme Inhib. Med. Chem.*, **22**, 501-510 (2007).
- [14] Y. Liang, W. Tian, and X. Ma, *BMC Complement. Altern. Med.*, **13** (2013).
- [15] B.-H. Li, X.-F. Ma, X.-D. Wu, and W.-X. Tian, *IUBMB Life*, **58**, 39-46 (2006).

- [16] J. A. Menendez, R. Lupu, *Expert Opin. Ther. Targets*, **21**, 1001-1016 (2017).
- [17] J. Li, L. Dong, D. Wei, X. Wang, S. Zhang, and H. Li, *Int. J. Biol. Sci.*, **10**, 171-180 (2014).
- [18] D. Veigel, R. Wagner, G. Stübiger, M. Wuczkowski, M. Filipits, R. Horvat, B. Benhamú, M. L. López-Rodríguez, A. Leisser, P. Valent, M. Grusch, F. G. Hegardt, J. García, D. Serra, N. Auersperg, R. Colomer, and T. W. Grunt, *Int. J. Cancer*, **136**, 2078-2090 (2015).
- [19] G. Zadra, C. Photopoulos, and M. Loda, *Biochim. Biophys. Acta*, **1831**, 1518-1532 (2013).
- [20] S. Kant, A. Kumar, and S. M. Singh, *Biochim. Biophys. Acta*, **1840**, 294-302 (2014).
- [21] S. Al-Bahlani, H. Al-Lawati, M. Al-Adawi, N. Al-Abri, B. Al-Dhahli, and K. Al-Adawi, *Apoptosis*, **22**, 865-876 (2017).
- [22] P. Sharma and J. P. Allison, *Science*, **348**, 56-61 (2015).
- [23] X. Wang, K.-S. Song, Q.-X. Guo, and W.-X. Tian, *Biochem. Pharmacol.* **66**, 2039-2047 (2003).
- [24] B.-H. Li and W.-X. Tian, *J. Biochem.* **135**, 85-91 (2004).
- [25] B.-H. Li, X.-F. Ma, Y. Wang, and W.-X. Tian, *J. Biochem.* **138**, 679-685 (2005).
- [26] H. N. Abramson, *J. Med Chem.* **54**, 5615-5638 (2011).
- [27] C. Turrado, T. Puig, J. García-Cárceles, M. Artola, B. Benhamú, S. Ortega-Gutiérrez, J. Relat, G. Oliveras, A. Blancafort, D. Haro, P. F. Marrero, R.

- Colomer, and M. L. López-Rodríguez, *J. Med. Chem.*, **55**, 5013-5023 (2012).
- [28] F. J. Asturias<sup>1</sup>, J. Z. Chadick, I. K. Cheung, H. Stark, A. Witkowski, A. K. Joshi, and S. Smith, *Nat. Struct. Mol. Biol.*, **12**, 225-232 (2005).
- [29] T. Maier, M. Leibundgut, and N. Ban, *Science*, **321**, 1315-1322 (2008).
- [30] Y. Cheng and W. H. Prusoff, *Biochem. Pharmacol.*, **1**, 3099-3108 (1973).
- [31] K. Fukui, *Angew. Chem. Int. Ed. Engl.*, **21**, 801-876 (1982).
- [32] R. A. Marcus, *Rev. Mod. Phys.*, **65**, 599-610 (1993).
- [33] K. Sakanoue, M. Motoda, M. Sugimoto, and S. Sakaki, *J. Phys. Chem. A*, **103**, 5551-5556 (1999).
- [34] M. Sugimoto, I. Yamasaki, N. Mizoe, M. Anzai, S. Sakaki, *Theor. Chem. Acc.*, **102**, 377-384 (1999).
- [35] W. Hayes and A. M. Stoneham, "Defects and Defect Processes in Nonmetallic Solids", John Wiley & Sons, Inc., New York, p. 93 (1985).
- [36] Y. Toyozawa, "Optical Processes in Solids", *Cambridge University Press* (2003).
- [37] M. C. Bohm, *J. Phys. B*, **17**, 3103-3116 (1984).
- [38] J. Bajorath, *J. Chem. Inf. Model.*, **41**, 233-245 (2001).
- [39] Mathematica Ver. 11.0, *Wolfram Research Inc.*, Champaign, Illinois (2016).
- [40] Z. Movasaghi, S. Rehman, and I. U. Rehman, *Appl. Spectrosc. Rev.*, **42**, 493-541 (2007).
- [41] T. Holstein, *Ann. Phys.*, **8**, 325 (1959).
- [42] L. Breiman, *Mach. Learn.*, **45**, 5-32, (2001).
- [43] C. J. C. Burges, *Data Min. Knowl. Discov.*, **2**, 121-167 (1998).

- [44] M. J. Frisch et al., Gaussian 09, Revision D. 01, *Gaussian, Inc.*, Wallingford CT (2016).
- [45] <https://pubchem.ncbi.nlm.nih.gov/compound/Phenol>
- [46] <https://pubchem.ncbi.nlm.nih.gov/compound/Propyl-acetate>
- [47] G. E. Mullen and L. Yet, *Bioorg. Med. Chem. Lett.*, **25**, 4363-4369 (2015).
- [48] R. W. Kennard and L. A. Stone, *Technometrics*, **11**, 137-148 (1969).

## 第 4 章 抗菌活性を示す天然物に関する回帰モデルの作成およびバーチャルスクリーニング

### 4-1 はじめに

機械学習の研究は、化学のさまざまな分野で精巧に行われてきた[1]。機械学習モデリングにおいて、重要な問題の一つは、記述子(説明変数)の選択である[2]。これは、いわゆる「ビッグデータ」と呼称されるような膨大な数のデータを利用することができない場合に当てはまる問題である。実際には、利用可能な化学データの数は、ほとんどの実験で通常 100 未満である。このようなデータは「スモールデータ」に分類できる。

ケモインフォマティクスの分野では、さまざまな記述子が開発され、定量的構造活性相関(QSAR)研究に適用されている。その代表的な例として、500 を超える記述子を含む、市販の「alvaDesc」という記述子セットが挙げられる[3]。このような大規模な記述子のセットは、薬理活性のような「ビッグデータ」に対する QSAR 研究においては非常に強力な柔軟性があるが、実際に利用するような「スモールデータ」での機械学習に適用すると、過剰適合(過学習:オーバーフィッティング)の問題が発生する可能性がある[4]。

我々のグループでは化学的な相互作用や反応に関連する分子特性に関連する一連の記述子を提案している。我々の手法では、それらの数値は電子状態計算を使用して評価している[5-8]。この手法は、取り扱う記述子がさまざまな化学的性質[9-11]と反応性[12]に関する量子化学分析で頻繁に議論される重要なパラメータであるため、「電子状態インフォマティクス」(Electronic-structure informatics: ESI)[6-8]と呼ぶことができる。

本章では、天然物の抗菌活性を予測するための一連の ESI 記述子の適用性を調査することを目的としている。また、生理活性を示す化合物のバーチャルスクリーニングの実現可能性についても調査を行なった。我々の知る限りでは、このような抗菌活性の定量的予測に関する研究は報告されていない。これはおそらく、既知の抗菌性化合物が分子構造に幅広い多様性をもっており、そのために構造ベースの記述子セットが活性予測において効率が低い、あるいは信頼性が低いということに起因するものと思われる。我々が提案する ESI 記述子は、構造的特性に焦点を当てるのではなく、化学相互作用時のエネルギー変化や相互作用エネルギーを記述するものであり、データセットのサイズが小さい場合でも生物学的特性をコンパクトに記述するのに役立つことが期待される。

教師あり機械学習モデルの一つであるランダムフォレスト法を適用して、抗菌活性の回帰モデルを作成した。この活性は抗微生物活性の一種である。抗生物質、抗細菌剤、抗ウイルス剤を含む抗菌試薬の発見は、医学における最も困難な問題の一つである。多剤耐性をもつ新しい細菌[13]や、COVID-19[14]を引き起こす SARS-CoV-2 などの新しいウイルスの出現などがその一例であり、現在において我々の健康問題を脅かしている非常に深刻な問題である。食品科学においても、病原体に対する抗菌活性のインシリコ予測は非常に困難であり、需要がとても大きい[15]。

今回、抗菌作用を示す化学物質として精油に着目した。精油は沸点が低く揮発性が高い非水溶性の混合物であり、芳香性植物の様々な部分(花卉、葉、果実、心材、根、種子、樹皮、樹脂など)から抽出された天然成分である。様々な植物から採取された精油には、古くから殺菌・抗菌作用があることが経験的に知られており、有史以来、感染症に対する予防薬、あるいは制菌作用を目的として防腐剤として用いられてきた。



具体例としては、古代エジプトでは複数種類のハーブを練り合わせた香料・香水であるキフィーが歴代ファラオや神官などによって、場の浄化や病気の治療など、医薬的な使われ方をしていた。また、ミルラ、シナモン、クローブ、カシア、ローズマリー、などの様々な精油がミイラの保存料、即ち防腐剤として用いられていたという記録がある。中世ヨーロッパではペストが流行した際にクローブを予防薬として用いるなど長く薬として用いられていた。そして近代以降、ペニシリンに始まる抗菌剤の開発によって多くの感染症が克服され、精油は感染症の予防や治療の目的で使用されることは少なくなるというような流れにあった。

しかし、近年抗菌剤を乱用したことで、遺伝子の変異や、別の細菌から抗菌剤に対する耐性遺伝子を獲得した薬剤耐性菌が出現し、それらによる感染症が問題となったことで、薬としての精油が再び注目されるようになってきている[13]。抗菌剤の主な作用機序は、細胞壁合成阻害作用、DNA 合成阻害作用、タンパク質合成阻害作用であり、特定の受容体を持つため、菌類はそれらのもつアミノ酸をひとつ変異させるだけで容易に抗菌剤の耐性を得てしまう可能性がある。そのため、新たな抗菌剤を開発してもすぐにその薬に耐性を持つ耐性菌が出現するイタチゴッコのような形になっている。一方、精油は細菌の細胞膜に非特異的に吸着することで抗菌作用を発揮するため、耐性化しにくく、薬剤耐性菌にも効果を発揮する[16]。このことから、精油を用いた薬剤耐性菌による感染症の防止が期待されている。感染症の予防に使われるのは消毒薬や除菌スプレーなどがあり、より効果の高いものの開発のためには抗菌活性の高い精油成分分子の探索を行う必要がある。この探索を行うためには、精油のどのような化学的性質が抗菌活性に影響を与えているかを調べる必要があると考えられる。そこで本研究では、電子状態計算から得られる種々の数値情報を記述子として定義し抗

菌活性との関係を調査すること、およびそこから得られた知見をもとに抗菌活性の高い分子を予測することを目的として研究を行なった。

## 4-2 計算方法

抗菌剤の抗菌力および感受性の指標として、実験によって測定される「最小発育阻止濃度 (MIC)」に注目した。これは、細菌の増殖を阻止するのに必要な生体外での抗菌剤の最小濃度であり、抗菌活性が高いほど小さな値をとる。本研究では、電子的記述子を使用して、*E. coli* に対する MIC 値が報告されている分子に注目して、それらについての解析を行なった。なお、本章では、化学反応速度理論や化学平衡理論が示唆するように、化合物の活性が自由エネルギーの変化に関連している可能性があるため、元の MIC ではなく MIC の対数 ( $\log \text{MIC}$ ) を目的変数とした[17]。この変換は、ESI 記述子について以前行なった別の解析で実際に有益な変換である[6,7]。

### 4-2-1 電子的記述子

我々の研究室では分子論的な考察から 21 種類の電子的記述子を提案しており、これらは、下記の考察より精油の作用機序に関して重要な記述子となりうる。

精油の抗菌作用は以下の様な過程である。まず細菌の細胞膜への吸着によって膜機能が低下して細胞の増殖が抑制されるため、細菌の発育・増殖を阻止する作用(制菌作用)が現れる。さらに精油成分が吸着すると細胞膜に綻びが生じ、高圧な内部を保護していた細胞壁の合成が止まって隙間ができることによって細胞質成分が流出し、菌が死滅する(殺菌作用)。これらの精油の抗菌作用機序から分子間相互作用が重要と考えられる。この観点から、抗菌作用の分子間相互作用に関係する記述子として、

溶媒和エネルギー ( $\Delta E_{\text{solv}}$ )、双極子モーメント ( $\mu$ )、全ての原子球を包括する直方体の大きさ  $B_n$  ( $n = 1-3$ ) に注目することとした。また、官能基の種類や分子振動の特徴を反映し、電子分局の大きさとも関係づけられる記述子として各種スペクトルデータを記述子として定義した。解析に用いたスペクトルデータは、軌道エネルギーの状態密度スペクトル、UV/Vis スペクトル、励起エネルギーの状態密度スペクトル、IR スペクトル、振動数の状態密度スペクトルである。ただし、これらのスペクトルデータは、活性の高い分子を参照分子とした、スペクトル間の類似度を実際の記述子とした。

今回の解析に使用した記述子の一覧を Table 4-1 に示す。スペクトル系の記述子が 5 種、スカラー系の記述子が 16 種、電子状態計算以外のデータから得られる記述子が 1 種類ある。スカラー系のうち、9 種が電荷移動に関連するもので、7 種が分子間相互作用に関連するものである。スペクトル系の記述子に関しては、分子ごとに単一の値を定義することができないため、分子ごとにグラフを作成して参照分子のグラフと重ね合わせて数値積分を行い、重なり合う部分の面積を算出して、それらを各分子の参照分子に対する類似度として定義した。参照分子には解析に使用する分子の中で最も活性が高い分子を使用した。これらの数学的処理には Wolfram 社の “Mathematica11” [18] を使用した。スペクトルの記述子の算出の際には記述子ごとに定数を使用している。これらの値はどのような半値幅に設定すれば、GaussView で表示した時のスペクトルの形状に近い形で各分子のスペクトルのグラフを作図できるかという観点から経験的に決定した。スカラー系の記述子のうち、イオン化エネルギー・電子親和力・ $S_0-T_1$  スプリッティングと 6 種類の構造緩和エネルギーについては、中性基底状態で構造最適化した全エネルギーの値や、それを初期構造としたカチオン状態

Table 4-1. 解析に使用した記述子の一覧

記号	定義
$S(\rho_{UV})$	類似度 (UV/vis スペクトルの重なり)
$S(\rho_{ex})$	類似度 (電子励起状態の状態密度スペクトルの重なり)
$S(\rho_{orb})$	類似度 (分子軌道の状態密度スペクトルの重なり)
$S(\rho_{IR})$	類似度 (IR スペクトルの重なり)
$S(\rho_{vib})$	類似度 (振動状態密度スペクトルの重なり)
$IP_v$	中性基底状態からカチオン状態への垂直イオン化のエネルギー
$\lambda_{S_0 \rightarrow C}$	中性基底状態からカチオン状態への遷移後の構造緩和エネルギー
$\lambda_{C \rightarrow S_0}$	カチオン状態から中性基底状態への遷移後の構造緩和エネルギー
$EA_v$	中性基底状態からアニオン状態への垂直イオン化のエネルギー
$\lambda_{S_0 \rightarrow A}$	中性基底状態からアニオン状態への遷移後の構造緩和エネルギー
$\lambda_{A \rightarrow S_0}$	アニオン状態から中性基底状態への遷移後の構造緩和エネルギー
$\Delta ST_v$	中性基底状態から $T_1$ 状態への垂直遷移のエネルギー
$\lambda_{S_0 \rightarrow T_1}$	中性基底状態から $T_1$ 状態への遷移後の構造緩和エネルギー
$\lambda_{T_1 \rightarrow S_0}$	$T_1$ 状態から中性基底状態への遷移後の構造緩和エネルギー
$V_{mol}$	分子体積
$B_1$	分子を囲む直方体の最も長い辺の長さ
$B_2$	分子を囲む直方体の2番目に長い辺の長さ
$B_3$	分子を囲む直方体の最も短い辺の長さ
$\mu$	双極子モーメント
$\Delta E_{solv}$	溶媒和エネルギー
$MW$	分子量
$XLogP3-AA$	オクタノール/水 分配係数

での一点計算・構造最適化計算などから得られた全エネルギーの値などのエネルギー差から得られる。 $B_1$ 、 $B_2$ 、 $B_3$  は中性基底状態で構造最適化された分子を CPK モデルで表現した際に、分子がぴったり収まるような直方体を定義した際の各辺の長さを降順に並べたものである。これらの記述子はタンパク質とリガンド間の鍵と鍵穴、即ちそれぞれの大きさが化学反応において重要な要素になっていることが知られているため、それらを考慮できるように記述子として採用している。また、今回の対象とした分子群である精油に関しては制菌作用や殺菌作用などの作用機序の観点から、細胞膜への吸着性・透過性が重要な要素であると考えられるため、生理活性物質の QSAR 解析で用いられることが多い XLogP を記述子として追加した。XLogP は水/オクタノール分配係数の計算値であり[19]、PubChem[20]に記載してある情報を参照して記述子として使用した。なお、PubChem に記載してある XLogP は分子によって複数の種類があるが、今回取り扱った分子に関しては、いずれの分子でも XLogP3-AA が記載されていた。

#### 4-2-2 電子状態計算

電子状態計算には汎用量子化学計算プログラムである“Gaussian 16”[21]を使用した。計算条件は、構造最適化・基準振動解析・各種一点計算では、DFT (M06-2X)/6-31G (d,p)とした。励起状態計算では、TDDFT (M06-2X)/6-31G (d,p)とし、求める励起状態の解の数は 30 個とした。計算手順としては、まず、中性状態での構造最適化を行い、そのジオメトリを使用して基準振動解析を行った。解析結果に虚数振動がないことを確認し、励起状態計算を行なった。また、同様のジオメトリを使用して電荷やスピン多重度などを変更し、カチオン・アニオン・ $T_1$  状態でそれぞれ一点計算を行なっ

た。そこからさらに各状態で構造最適化を行い、その構造を用いて、先ほど変更した電荷やスピン多重度を元に戻した状態での一点計算を行なった。これらの計算を行なった理由としては、実際の化学反応を想定した場合に反応性に関わってくると考えられる値をこれらの計算結果から算出できるようにするためである。また、溶媒として水を想定した PCM 計算、および各分子の分子体積の計算を行なった。Gaussian の体積計算にはモンテカルロ積分が使用されているため、同様の計算を5回行い、その平均値を分子体積とした。これらの DFT 関数およびそのほかの計算条件をこのように選択した理由は、我々のグループが有する、有機化合物[5-7]と配位化合物[8]の両方を含む電子状態データベースとの整合性をとるためである。以上の各種計算結果から得られたデータから 21 種類の数値情報を選択し、各分子の電子的記述子として定義した。なお、後述の機械学習では、4-2-1 で示した通りこれらの電子的記述子に XLogP3-AA を加えた 22 種の記述子を使用した。

#### 4-2-3 機械学習手法

ランダムフォレストによる回帰分析の流れを以下に示す。なお、回帰分析に使用したプログラムは Python [22]の機械学習のライブラリである scikit-learn[23]を用いて実装した。まず、目的変数を説明するための有意な説明変数を選出するために特徴量の重要度に基づいた変数選択を行なった。手順としては、すべての記述子を使用して目的変数を説明した際に、重要度が最も小さい記述子を削除して、再度各記述子の目的変数に関する重要度を確認した。各ステップで重要度が最も小さかった記述子を除外して各記述子の重要度が 0.1 以上になるまで繰り返し行い、残った記述子を説明力が高い記述子として選択した。次に、データの分割を行なった。全データを 4:1 にラ

ランダムに分割した。そのうち前者をトレーニングデータ、後者をバリデーションデータとし、これを合計 10 パターン作成した。そして、ハイパーパラメータの決定を行なった。各セットのトレーニングデータについてさらに 5 分割クロスバリデーションを行い、各分割での平均二乗誤差 (MSE: Mean square error) を算出した。ベイズ最適化の手法を用いて MSE の平均値が最小となるようにハイパーパラメータを決定した。この手法におけるハイパーパラメータは、決定木の数と最大の分岐数である。続いて、選択された記述子とハイパーパラメータを用いてランダムフォレストの回帰モデルを作成し、各セットで最初に分割しておいたバリデーションデータを回帰モデルに適用し、モデルの性能を評価した。各モデルの予測性能を比較し、バリデーションデータの予測値の決定係数が最大となったモデルを、予測モデルとして決定した。

#### 4-3 結果と考察

まず解析対象とした分子の分子構造の Fig. 4-1 に示す。これらの分子は Griffin らが投稿した論文に記載されている分子で、文献中には全部で 60 分子の記載があった [24]。その文献中には複数の対象に対しての最小発育阻止濃度 (MIC) の数値がまとめられており、今回の発表ではそれらの内、大腸菌 (*E. coli*) に対する数値をこれらの分子の抗菌活性を表す目的変数として使用した。図中の括弧内の数値が各分子の *E. coli* に対する MIC の値である。解析の際には MIC の対数をとった  $\log(\text{MIC})$  を目的変数として使用している。60 個の分子の構造式を作成して並べてみたところ、大きく分けて 3 種類の構造に分けられそうだと考えられたので、それぞれ構造ごとに分けて表示した。

1 グループ目は分子内に環を複数もつ複環式化合物で、全部で 21 分子ある。左上

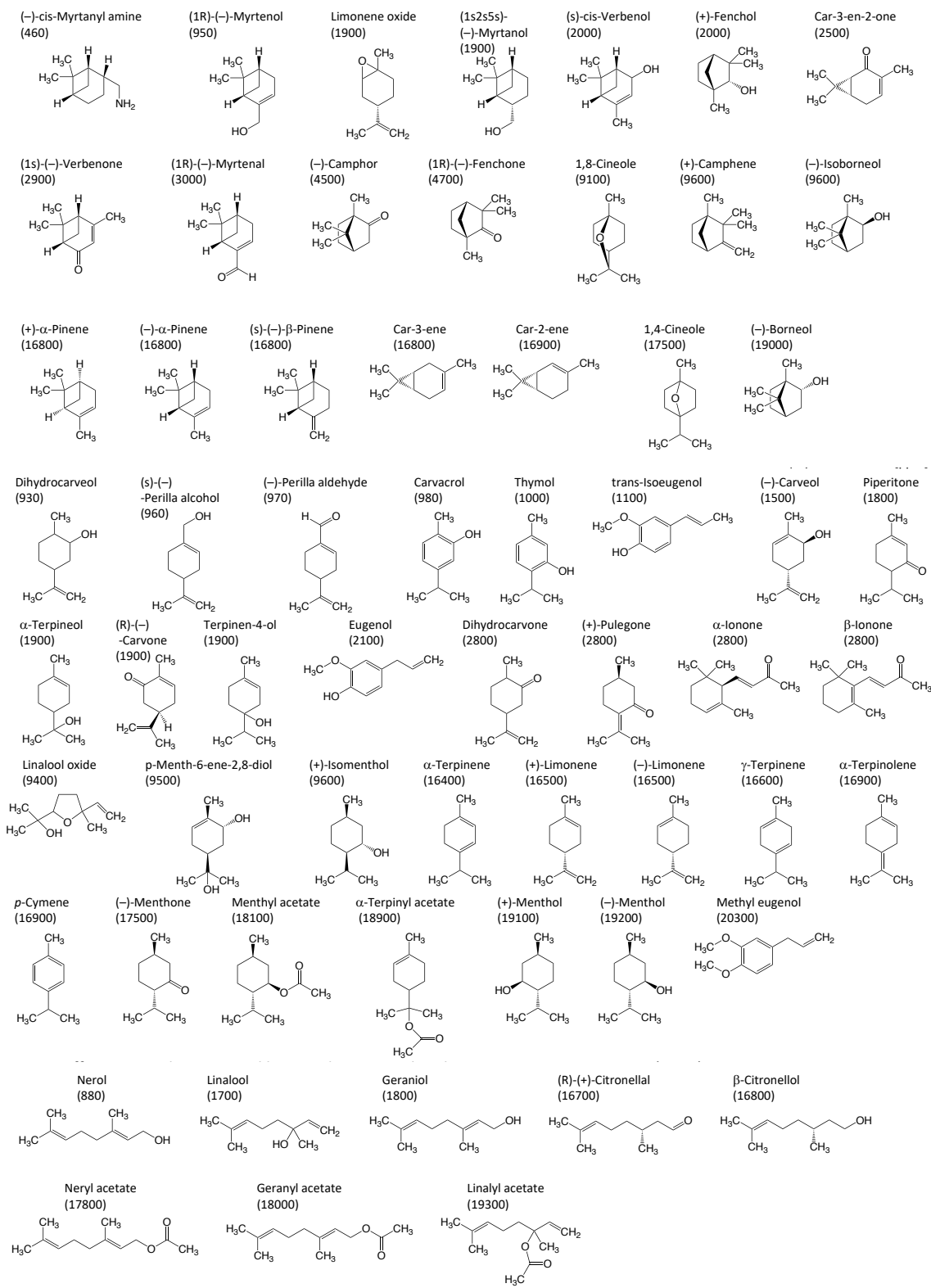


Figure 4-1. 解析対象とした精油成分の分子構造。括弧の中の数値はそれぞれの分子の MIC の値である(単位: ppm)。



の分子の“(–)-cis-Myrtanyl amine”が今回のデータセットの中で最も活性が高い分子であり、そこから右に向かってに MIC の値が大きくなる順に並べている。

2つ目のグループは、分子内に環をひとつだけもつ単環式化合物で、全部で 31 分子ある。

3つめのグループは、分子内に環をひとつもたない非環状化合物であり、全部で 8分子あった。以上の合計 60 種の分子について各分子の三次元構造を既存の化合物データベース(PubChem)より取得、あるいは作成し、電子状態計算を実行した。

電子状態計算の結果から各記述子を抽出した。それぞれの分子について求めた記述子の数値を Table 4-2 から Table 4-4 に示す。上から複環式、単環式、非環状化合物の数値の順番に並んでいる。1 行目の分子が参照分子とした“(–)-cis-Myrtanyl amine”であり、その記述子の類似度の数値は 100 となっている。これらの数値を使用して、ランダムフォレストによる解析を行なった。

まず目的変数に対する各記述子の特徴量の重要度の評価を繰り返し行った。重要度の選択の過程を Table 4-5 に示す。初めてすべての記述子で重要度が 0.1 を上回ったのは 17 ステップ目で、その際に残っていた記述子は 6 種類で、特徴量の重要度 (feature importance: FI) の降順でそれぞれ、 $\Delta E_{\text{solv}}$  (PCM 法による溶媒和エネルギー)、XLogP3-AA (オクタノール/水分配係数)、 $\lambda_{\text{A} \rightarrow \text{S0}}$  (アニオン状態から中性状態への脱励起時の構造緩和エネルギー)、 $MW$  (分子体積)、 $B_1$  (対象の分子を囲む直方体の最も長い辺の長さ)、 $S(\rho_{\text{UV}})$  (UV/Vis スペクトルの類似度) の 6 種類の記述子であった。主な特徴量は  $\Delta E_{\text{solv}}$  であり、特徴量の重要度は、 $\text{FI} = 0.322$  であった。この記述子は XLogP3-AA ( $\text{FI} = 0.154$ ) とともに、化合物の細胞透過性や水素結合の容量が抗菌活性を決定する重要な要因であることを意味している。これは文献で提案さ

Table 4-2. 解析対象の分子の実験値と記述子の数値(1/3)

Compound name	MIC (ppm)	Descriptors									
		S( $\rho_{UV}$ ) (%)	S( $\rho_{ex}$ ) (%)	S( $\rho_{orb}$ ) (%)	S( $\rho_{IR}$ ) (%)	S( $\rho_{vib}$ ) (%)	$IP_V$ (eV)	$\lambda_{S0 \rightarrow C}$ (eV)	$\lambda_{C \rightarrow S0}$ (eV)		
(-)-cis-Myrtanyl amine	460	100.0	100.0	100.0	100.0	100.0	9.124	0.994	0.787		
(1R)-(-)-Myrtenol	950	68.9	79.1	88.2	72.0	85.8	8.194	0.322	0.395		
Limonene oxide	1900	58.6	79.5	90.9	75.2	90.4	9.374	0.899	1.360		
(1s2s5s)-(-)-Myrtanol	1900	42.2	92.1	87.9	72.1	78.0	8.967	0.375	0.458		
(s)-cis-Verbenol	2000	81.0	86.8	90.0	67.2	86.2	8.480	0.453	0.442		
(+)-Fenchol	2000	78.5	71.7	88.1	69.6	82.8	9.464	0.984	2.000		
Car-3-en-2-one	2500	32.7	62.9	85.3	32.1	73.9	8.781	0.108	0.143		
(1s)-(-)-Verbenone	2900	52.7	59.5	84.7	28.2	83.1	8.715	0.128	0.211		
(1R)-(-)-Myrtenal	3000	41.1	62.4	86.5	33.6	84.1	9.101	0.502	1.125		
(-)-Camphor	4500	79.8	80.3	87.3	41.7	81.0	8.752	0.240	0.562		
(1R)-(-)-Fenchone	4700	72.6	67.7	88.1	50.6	78.7	8.623	0.245	0.349		
1,8-Cineole	9100	51.4	58.8	85.9	67.4	81.1	8.634	0.200	0.256		
(+)-Camphene	9600	33.5	78.3	85.8	66.3	80.5	8.711	0.374	0.554		
(-)-Isoborneol	9600	51.8	63.7	88.2	70.1	84.5	9.634	0.963	2.057		
(+)- $\alpha$ -Pinene	16800	76.4	89.1	89.1	78.6	87.4	8.261	0.312	0.392		

bicyclic

Table 4-2. (Continued)

Compound name	MIC (ppm)	Descriptors									
		S( $\rho_{UV}$ ) (%)	S( $\rho_{ex}$ ) (%)	S( $\rho_{orb}$ ) (%)	S( $\rho_{IR}$ ) (%)	S( $\rho_{vib}$ ) (%)	$IP_V$ (eV)	$\lambda_{S0 \rightarrow C}$ (eV)	$\lambda_{C \rightarrow S0}$ (eV)		
(-)- $\alpha$ -Pinene	16800	76.4	89.1	89.1	78.7	87.5	8.260	0.312	0.391		
(s)-(-)- $\beta$ -Pinene	16800	48.0	83.7	88.4	79.9	88.1	8.588	0.473	0.906		
Car-3-ene	16800	75.4	84.5	86.3	65.6	77.9	8.475	0.311	0.365		
Car-2-ene	16900	71.4	89.0	86.8	74.9	80.5	8.038	0.523	0.695		
1,4-Cineole	17500	26.1	55.4	87.0	63.6	81.7	9.264	0.746	3.965		
(-)-Borneol	19000	51.5	71.0	87.9	69.3	85.8	9.616	0.645	1.218		
Dihydrocarveol	930	42.2	87.9	87.3	56.2	78.0	8.968	0.426	0.533		
(s)-(-)-Perilla alcohol	960	11.8	56.6	84.9	57.7	72.7	8.464	0.294	0.337		
(-)-Perilla aldehyde	970	10.5	44.0	84.7	24.8	69.6	9.127	0.303	0.384		
Carvacrol	980	2.1	50.6	79.3	45.8	74.8	7.984	0.230	0.232		
Thymol	1000	2.5	49.2	82.3	44.3	72.5	8.016	0.268	0.243		
trans-Isoeugenol	1100	4.1	24.0	80.1	16.8	68.9	7.371	0.283	0.262		
(-)-Carveol	1500	21.5	54.6	85.5	60.2	76.0	8.498	0.331	0.420		
Piperitone	1800	48.6	76.2	86.1	38.4	81.7	8.767	0.145	0.252		
$\alpha$ -Terpineol	1900	57.2	85.4	86.0	64.7	81.4	8.274	0.641	0.490		

Table 4-2. (Continued)

Compound name	MIC (ppm)	Descriptors									
		S( $\rho_{UV}$ ) (%)	S( $\rho_{ex}$ ) (%)	S( $\rho_{orb}$ ) (%)	S( $\rho_{IR}$ ) (%)	S( $\rho_{vib}$ ) (%)	$IP_V$ (eV)	$\lambda_{S0 \rightarrow C}$ (eV)	$\lambda_{C \rightarrow S0}$ (eV)		
(R)-(-)-Carvone	1900	13.2	44.6	84.5	25.6	68.0	8.971	0.254	0.403		
Terpinen-4-ol	1900	42.5	86.8	85.9	66.4	78.9	8.509	0.373	0.418		
Eugenol	2100	3.4	28.4	80.3	15.0	64.8	7.682	0.289	0.296		
Dihydrocarvone	2800	29.2	69.2	84.8	40.8	82.6	8.902	0.277	0.396		
(+)-Pulegone	2800	35.4	69.7	85.4	42.8	74.1	8.713	0.356	0.433		
$\alpha$ -Ionone	2800	6.1	24.3	87.2	42.6	79.7	8.631	0.380	1.036		
$\beta$ -Ionone	2800	6.1	25.2	88.0	41.3	80.1	8.205	0.327	0.432		
Linalool oxide	9400	36.8	88.0	86.9	33.8	74.3	9.189	0.760	1.782		
p-Menth-6-ene-2,8-diol	9500	54.8	85.2	87.7	46.6	80.2	8.874	0.711	0.553		
(+)-Isomenthol	9600	55.9	64.5	86.9	67.3	84.1	9.586	0.629	0.829		
$\alpha$ -Terpinene	16400	40.0	76.5	85.0	60.1	75.9	7.512	0.256	0.263		
(+)-Limonene	16500	22.9	79.8	84.1	67.7	68.8	8.526	0.285	0.339		
(-)-Limonene	16500	22.9	79.8	84.1	68.4	69.0	8.527	0.285	0.340		
$\gamma$ -Terpinene	16600	11.6	84.8	85.5	50.9	69.1	8.318	0.210	0.283		
$\alpha$ -Terpinolene	16900	20.7	71.2	84.0	64.4	76.1	8.148	0.345	0.362		

monocyclic

Table 4-2. (Continued)

Compound name	MIC (ppm)	Descriptors									
		S( $\rho_{UV}$ ) (%)	S( $\rho_{ex}$ ) (%)	S( $\rho_{orb}$ ) (%)	S( $\rho_{IR}$ ) (%)	S( $\rho_{vib}$ ) (%)	$IP_V$ (eV)	$\lambda_{S0 \rightarrow C}$ (eV)	$\lambda_{C \rightarrow S0}$ (eV)		
<i>p</i> -Cymene	16900	2.5	55.7	77.9	61.0	72.4	8.394	0.208	0.232		
(-)-Menthone	17500	86.0	69.1	87.1	53.5	82.9	8.901	0.188	0.278		
Menthyl acetate	18100	82.2	77.2	88.4	40.4	84.3	9.814	0.573	0.956		
$\alpha$ -Terpinyl acetate	18900	55.5	75.4	86.3	26.2	75.3	8.614	0.317	0.372		
(+)-Menthol	19100	37.2	62.3	87.0	64.1	80.1	9.649	0.644	0.814		
(-)-Menthol	19200	37.2	62.3	87.0	64.3	80.4	9.649	0.644	0.812		
Methyl eugenol	20300	2.0	23.1	80.7	25.8	68.4	7.918	0.660	0.409		
Nerol	880	16.7	56.9	85.6	59.3	78.5	8.368	0.511	0.387		
Linalool	1700	10.4	67.2	86.0	47.1	76.2	8.490	0.348	0.412		
Geraniol	1800	13.8	61.6	85.3	59.1	77.7	8.326	0.334	0.361		
(R)-(+)-Citronellal	16700	28.6	75.3	87.0	46.2	79.1	8.649	0.323	0.390		
$\beta$ -Citronellol	16800	23.4	76.8	87.6	64.0	80.2	8.496	0.325	0.393		
Neryl acetate	17800	5.7	38.0	85.1	18.9	76.7	8.577	0.333	0.386		
Geranyl acetate	18000	8.3	45.1	85.6	23.1	75.5	8.483	0.353	0.382		
Linalyl acetate	19300	13.2	43.0	86.9	24.7	76.6	8.487	0.333	0.408		

monocyclic

acyclic

Table 4-3. 解析対象の分子の実験値と記述子の数値(2/3)

Compound name	MIC (ppm)	$E_{A_v}$ (eV)	$\lambda_{S0 \rightarrow A}$ (eV)	$\lambda_{A \rightarrow S0}$ (eV)	$\Delta ST_v$ (eV)	$\lambda_{S0 \rightarrow T1}$ (eV)	$\lambda_{T1 \rightarrow S0}$ (eV)	Descriptors	
								$V_{mol}$ (cm <sup>3</sup> /mol)	
(-)-cis-Myrtanyl amine	460	-3.908	0.216	0.354	7.217	2.613	7.077	148.195	
(1R)-(-)-Myrtenol	950	-2.904	0.890	0.577	4.321	1.275	2.318	132.038	
Limonene oxide	1900	-3.988	0.211	0.211	9.036	6.391	2.647	124.258	
(1s2s5s)-(-)-Myrtanol	1900	-2.954	0.589	1.031	4.579	1.664	4.083	133.846	
(s)-cis-Verbenol	2000	-2.477	0.726	0.720	4.240	1.197	2.196	136.311	
(+)-Fenchol	2000	-3.548	0.627	4.292	7.627	2.957	6.488	132.904	
Car-3-en-2-one	2500	-0.924	0.369	0.408	3.488	0.407	0.549	108.268	
(1s)-(-)-Verbenone	2900	-0.892	0.336	0.395	3.481	0.372	0.523	132.350	
(1R)-(-)-Myrtenal	3000	-0.672	0.359	0.362	3.510	0.836	1.997	132.276	
(-)-Camphor	4500	-2.015	0.531	0.861	3.969	0.568	1.162	132.634	
(1R)-(-)-Fenchone	4700	-1.977	0.555	0.910	3.933	0.561	1.154	126.933	
1,8-Cineole	9100	-3.598	0.136	0.178	7.909	4.177	3.756	129.782	
(+)-Camphene	9600	-2.841	0.524	0.955	4.521	1.673	2.826	124.393	
(-)-Isoborneol	9600	-3.877	0.940	4.260	7.828	3.179	4.656	137.572	
(+)- $\alpha$ -Pinene	16800	-2.840	0.508	0.674	4.288	1.245	2.418	128.275	

bicyclic

Table 4-3. (Continued)

Compound name	MIC (ppm)	$E_{A_v}$ (eV)	$\lambda_{S0 \rightarrow A}$ (eV)	$\lambda_{A \rightarrow S0}$ (eV)	$\Delta ST_v$ (eV)	Descriptors			
						$\lambda_{S0 \rightarrow T1}$ (eV)	$\lambda_{T1 \rightarrow S0}$ (eV)	$V_{mol}$ (cm <sup>3</sup> /mol)	
(-)- $\alpha$ -Pinene	16800	-2.840	0.508	0.676	4.289	1.245	2.414	122.164	
(s)-(-)- $\beta$ -Pinene	16800	-2.928	0.419	0.772	4.495	1.606	2.867	118.548	
Car-3-ene	16800	-3.134	0.659	1.099	4.621	1.552	3.429	128.509	
Car-2-ene	16900	-2.816	0.491	0.757	4.380	1.433	2.939	124.717	
1,4-Cineole	17500	-3.847	0.228	0.297	8.363	4.974	4.713	136.669	
(-)-Borneol	19000	-3.685	0.424	0.330	7.812	3.136	4.678	139.412	
Dihydrocarveol	930	-2.930	0.621	1.071	4.583	1.669	4.067	142.683	
(s)-(-)-Perilla alcohol	960	-2.852	0.746	0.817	4.588	1.675	3.827	132.500	
(-)-Perilla aldehyde	970	-0.827	0.373	0.415	3.923	1.264	2.534	137.300	
Carvacrol	980	-2.170	0.390	0.482	4.425	0.617	1.798	128.157	
Thymol	1000	-2.057	0.279	0.443	4.694	0.254	0.373	133.189	
trans-Isoeugenol	1100	-1.464	0.322	0.278	3.618	1.120	2.485	132.752	
(-)-Carveol	1500	-2.635	0.474	0.816	4.469	1.514	3.581	133.963	
Piperitone	1800	-0.910	0.372	0.379	3.516	0.398	0.556	135.265	
$\alpha$ -Terpineol	1900	-3.151	0.548	0.965	4.449	1.521	3.370	139.073	

Table 4-3. (Continued)

Compound name	MIC (ppm)	$E_{A_v}$ (eV)	Descriptors						
			$\lambda_{S0 \rightarrow A}$ (eV)	$\lambda_{A \rightarrow S0}$ (eV)	$\Delta ST_v$ (eV)	$\lambda_{S0 \rightarrow T1}$ (eV)	$\lambda_{T1 \rightarrow S0}$ (eV)	$V_{mol}$ (cm <sup>3</sup> /mol)	
(R)-(-)-Carvone	1900	-0.869	0.364	0.413	3.512	0.405	0.576	124.382	
Terpinen-4-ol	1900	-3.089	1.217	1.073	4.578	1.601	3.823	136.743	
Eugenol	2100	-2.092	0.436	0.598	4.493	0.879	1.827	124.764	
Dihydrocarvone	2800	-1.856	0.438	0.773	3.879	0.543	1.085	139.522	
(+)-Pulegone	2800	-0.996	0.457	0.365	3.626	1.350	3.118	134.797	
$\alpha$ -Ionone	2800	-0.702	0.394	0.358	4.821	2.277	2.429	172.618	
$\beta$ -Ionone	2800	-0.457	0.481	0.397	3.461	1.421	1.806	171.397	
Linalool oxide	9400	-2.917	1.811	3.718	4.657	1.771	2.858	149.837	
p-Menth-6-ene-2,8-diol	9500	-2.584	0.867	1.144	4.506	1.625	2.846	140.688	
(+)-Isomenthol	9600	-3.859	0.828	4.073	7.866	3.138	4.736	140.353	
$\alpha$ -Terpinene	16400	-1.687	0.303	0.322	3.323	0.894	0.936	129.676	
(+)-Limonene	16500	-2.814	0.476	0.923	4.562	1.583	3.795	128.721	
(-)-Limonene	16500	-2.814	0.476	0.924	6.232	3.317	4.088	127.715	
$\gamma$ -Terpinene	16600	-2.908	0.480	1.026	5.647	2.688	3.618	128.362	
$\alpha$ -Terpinolene	16900	-2.886	0.567	0.874	4.421	1.730	3.873	134.224	

monocyclic



Table 4-3. (Continued)

Compound name	MIC (ppm)	$E_{A_v}$ (eV)	$\lambda_{S0 \rightarrow A}$ (eV)	$\lambda_{A \rightarrow S0}$ (eV)	$\Delta ST_v$ (eV)	Descriptors			
						$\lambda_{S0 \rightarrow T1}$ (eV)	$\lambda_{T1 \rightarrow S0}$ (eV)	$V_{mol}$ (cm <sup>3</sup> /mol)	
<i>p</i> -Cymene	16900	-2.118	0.266	0.238	4.994	0.221	0.218	125.907	
(-)-Menthone	17500	-1.894	0.523	0.722	3.929	0.554	1.111	134.994	
Menthyl acetate	18100	-2.782	1.103	1.607	5.556	1.238	2.106	167.682	
$\alpha$ -Terpinyl acetate	18900	-2.705	0.886	1.619	4.572	1.564	3.804	167.259	
(+)-Menthol	19100	-3.957	0.934	4.090	7.887	3.145	4.752	136.536	
(-)-Menthol	19200	-3.956	0.933	4.090	7.887	3.144	4.753	142.661	
Methyl eugenol	20300	-1.966	0.329	0.342	4.369	0.916	2.346	143.842	
Nerol	880	-2.369	0.428	0.536	5.791	2.996	4.193	149.636	
Linalool	1700	-2.584	0.659	1.385	4.521	1.686	2.824	144.781	
Geraniol	1800	-2.587	0.697	0.667	4.503	1.712	2.778	153.772	
(R)-(+)-Citronellal	16700	-1.988	0.529	0.732	3.847	0.542	1.015	148.359	
$\beta$ -Citronellol	16800	-2.983	0.557	0.961	4.505	1.667	2.827	148.051	
Neryl acetate	17800	-2.149	2.172	3.305	5.882	3.052	3.955	167.368	
Geranyl acetate	18000	-2.320	2.073	3.125	5.921	3.249	3.886	178.348	
Linalyl acetate	19300	-2.317	2.277	3.148	4.444	1.593	3.970	180.522	

monocyclic

acyclic

Table 4-4. 解析対象の分子の実験値と記述子の数値(3/3)

Compound name	MIC (ppm)	Descriptors							$X_{\log P3-AA}$
		$B_1$ (Å)	$B_2$ (Å)	$B_3$ (Å)	$\mu$ (debye)	$\Delta E_{solv}$ (eV)	MW (g/mol)		
(-)-cis-Myrtanyl amine	460	9.157	7.222	6.565	1.239	0.108	153.26	2.2	
(1R)-(-)-Myrtenol	950	8.986	7.055	6.726	1.510	0.147	152.23	1.6	
Limonene oxide	1900	8.994	6.827	6.521	1.430	0.129	154.25	2.5	
(1s2s5s)-(-)-Myrtanol	1900	10.662	6.269	5.997	2.091	0.128	152.23	2.5	
(s)-cis-Verbenol	2000	7.956	7.165	7.016	1.519	0.135	152.23	1.6	
(+)-Fenchol	2000	8.148	7.245	6.578	1.276	0.106	154.25	2.5	
Car-3-en-2-one	2500	8.641	7.393	6.227	3.488	0.177	150.22	2.2	
(1s)-(-)-Verbenone	2900	7.989	6.991	6.928	4.060	0.193	150.22	1.6	
(1R)-(-)-Myrtenal	3000	7.774	7.079	6.869	3.381	0.152	150.22	2.1	
(-)-Camphor	4500	7.762	6.939	6.769	2.941	0.148	152.23	2.2	
(1R)-(-)-Fenchone	4700	8.579	7.246	6.073	2.725	0.139	152.23	2.3	
1,8-Cineole	9100	7.877	7.407	6.706	1.2954	0.082	154.25	2.5	
(+)-Camphene	9600	7.578	7.282	6.340	0.540	0.038	136.23	3.3	
(-)-Isoborneol	9600	7.341	7.252	6.990	1.449	0.109	154.25	2.7	
(+)- $\alpha$ -Pinene	16800	7.582	7.019	6.961	0.145	0.026	136.23	2.8	

bicyclic

Table 4-4. (Continued)

Compound name	MIC (ppm)	Descriptors							$X_{\log P3-AA}$
		$B_1$ (Å)	$B_2$ (Å)	$B_3$ (Å)	$\mu$ (debye)	$\Delta E_{solv}$ (eV)	MW (g/mol)		
(-)- $\alpha$ -Pinene	16800	7.449	7.081	7.015	0.145	0.026	136.23	2.8	
(s)-(-)- $\beta$ -Pinene	16800	8.124	6.644	6.632	0.600	0.040	136.23	3.1	
Car-3-ene	16800	9.431	6.761	6.095	0.161	0.037	136.23	2.8	
Car-2-ene	16900	9.378	6.880	6.289	0.290	0.040	136.23	2.8	
1,4-Cineole	17500	10.040	6.501	6.207	1.2418	0.081	154.25	2.5	
(-)-Borneol	19000	7.489	7.213	6.726	1.410	0.113	154.25	2.7	
Dihydrocarveol	930	10.564	6.531	6.223	1.028	0.142	154.25	3.0	
(s)-(-)-Perilla alcohol	960	11.376	6.225	5.943	1.878	0.177	152.23	2.1	
(-)-Perilla aldehyde	970	10.987	6.197	6.014	3.218	0.172	150.22	2.6	
Carvacrol	980	10.322	6.072	6.024	1.333	0.175	150.22	3.1	
Thymol	1000	10.448	7.058	5.384	1.523	0.168	150.22	3.3	
trans-Isoeugenol	1100	11.649	7.925	4.169	2.624	0.198	164.20	2.6	
(-)-Carveol	1500	8.066	7.580	6.757	1.663	0.159	152.23	2.1	
Piperitone	1800	10.524	6.732	5.554	3.725	0.174	152.23	2.2	
$\alpha$ -Terpineol	1900	10.501	6.323	5.992	1.529	0.125	154.25	1.8	

Table 4-4. (Continued)

Compound name	MIC (ppm)	Descriptors							$X_{\log P3-AA}$
		$B_1$ (Å)	$B_2$ (Å)	$B_3$ (Å)	$\mu$ (debye)	$\Delta E_{solv}$ (eV)	MW (g/mol)		
(R)-(-)-Carvone	1900	10.490	6.146	5.890	3.160	0.178	150.22	2.4	
Terpinen-4-ol	1900	10.583	6.257	6.102	1.111	0.123	154.25	2.2	
Eugenol	2100	11.842	7.521	5.184	2.801	0.190	164.20	2.0	
Dihydrocarvone	2800	9.774	6.425	6.326	3.011	0.175	152.23	2.7	
(+)-Pulegone	2800	10.442	7.091	5.166	2.758	0.147	152.23	2.8	
$\alpha$ -Ionone	2800	10.576	7.894	7.611	2.682	0.172	192.30	3.0	
$\beta$ -Ionone	2800	11.579	7.951	6.142	3.196	0.168	192.30	2.9	
Linalool oxide	9400	10.038	6.958	6.160	1.642	0.145	170.25	1.4	
p-Menth-6-ene-2,8-diol	9500	10.387	6.694	5.876	2.471	0.244	170.25	0.7	
(+)-Isomenthol	9600	9.664	6.646	6.516	1.502	0.116	156.26	3.0	
$\alpha$ -Terpinene	16400	10.564	6.436	5.344	0.554	0.057	136.23	2.8	
(+)-Limonene	16500	10.457	6.157	6.006	0.538	0.056	136.23	3.4	
(-)-Limonene	16500	10.443	6.147	6.033	0.537	0.056	136.23	3.4	
$\gamma$ -Terpinene	16600	10.494	6.482	5.372	0.041	0.054	136.23	2.8	
$\alpha$ -Terpinolene	16900	10.301	6.699	5.638	0.203	0.053	136.23	2.8	

monocyclic

Table 4-4. (Continued)

Compound name	MIC (ppm)	Descriptors							$X_{\log P3-AA}$
		$B_1$ (Å)	$B_2$ (Å)	$B_3$ (Å)	$\mu$ (debye)	$\Delta E_{solv}$ (eV)	MW (g/mol)		
<i>p</i> -Cymene	16900	10.473	5.910	5.871	0.093	0.081	134.22	4.1	
(-)-Menthone	17500	10.653	6.697	5.658	2.750	0.137	154.25	2.7	
Menthyl acetate	18100	9.596	8.637	6.145	1.902	0.131	198.30	3.6	
$\alpha$ -Terpinyl acetate	18900	12.099	7.089	6.566	2.000	0.142	196.29	2.4	
(+)-Menthol	19100	10.407	7.500	5.575	1.484	0.116	156.26	3.0	
(-)-Menthol	19200	10.403	7.467	5.614	1.484	0.116	156.26	3.0	
Methyl eugenol	20300	9.348	9.141	5.439	2.167	0.186	178.23	2.5	
Nerol	880	11.013	7.800	5.292	1.315	0.186	154.25	2.9	
Linalool	1700	12.092	6.158	5.999	1.497	0.149	154.25	2.7	
Geraniol	1800	12.686	7.157	5.111	1.568	0.190	154.25	2.9	
(R)-(+)-Citronellal	16700	12.867	6.247	5.830	2.885	0.163	154.25	3.0	
$\beta$ -Citronellol	16800	13.491	6.162	5.962	1.435	0.164	156.26	3.2	
Neryl acetate	17800	15.570	6.385	5.713	4.940	0.316	196.29	3.5	
Geranyl acetate	18000	14.323	8.575	5.385	2.349	0.192	196.29	3.5	
Linalyl acetate	19300	10.932	8.147	6.973	1.838	0.184	196.29	3.3	

Table 4-5. ランダムフォレストの予測モデルにおける各記述子の重要度に基づいた記述子の選択。重要度が最も小さい記述子を逐次取り除いて重要度を算出した。全ての記述子の重要度が0.1以上になるまで繰り返し行った。

	No. 1	No. 2	No. 3	No. 4	No. 5	No. 6	No. 7	No. 8	No. 9	No. 10	No. 11	No. 12	No. 13	No. 14	No. 15	No. 16	No. 17
$S(\rho_{UV})$	0.030	0.029	0.036	0.036	0.033	0.035	0.035	0.041	0.047	0.041	0.052	0.053	0.059	0.062	0.082	0.085	0.102
$S(\rho_{ex})$	0.023	0.027	0.024	0.026	0.033	0.030	0.030	0.033	0.028	0.031	-	-	-	-	-	-	-
$S(\rho_{orb})$	0.021	0.019	0.021	0.026	0.021	0.023	0.026	0.024	-	-	-	-	-	-	-	-	-
$S(\rho_R)$	0.015	0.017	0.018	0.017	0.020	0.020	0.019	-	-	-	-	-	-	-	-	-	-
$S(\rho_{vb})$	0.011	0.012	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
$IP_v$	0.019	0.017	0.014	-	-	-	-	-	-	-	-	-	-	-	-	-	-
$\lambda_{SO-C}$	0.028	0.024	0.025	0.026	0.034	0.033	0.031	0.032	0.040	0.036	0.036	0.039	0.048	0.049	-	-	-
$\lambda_{C-SO}$	0.018	0.016	0.022	0.019	0.020	0.022	0.025	0.026	0.027	-	-	-	-	-	-	-	-
$EA_v$	0.019	0.022	0.022	0.020	0.022	0.020	0.024	0.026	0.031	0.032	0.032	-	-	-	-	-	-
$\lambda_{SO-A}$	0.032	0.033	0.037	0.031	0.031	0.033	0.035	0.035	0.036	0.041	0.038	0.040	0.041	-	-	-	-
$\lambda_{A-SO}$	0.096	0.101	0.096	0.102	0.102	0.104	0.104	0.101	0.102	0.110	0.109	0.113	0.114	0.126	0.132	0.145	0.154
$\Delta ST_v$	0.014	0.016	0.019	0.013	-	-	-	-	-	-	-	-	-	-	-	-	-
$\lambda_{SO-T1}$	0.009	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
$\lambda_{T1-SO}$	0.025	0.027	0.023	0.028	0.024	0.027	0.035	0.031	0.029	0.034	0.036	0.036	-	-	-	-	-
$V_{mol}$	0.038	0.043	0.040	0.043	0.041	0.042	0.045	0.044	0.042	0.052	0.048	0.048	0.051	0.049	0.050	-	-
$B_1$	0.045	0.049	0.050	0.058	0.057	0.055	0.059	0.057	0.063	0.066	0.073	0.074	0.078	0.090	0.087	0.100	0.116
$B_2$	0.017	0.020	0.021	0.016	0.017	0.019	-	-	-	-	-	-	-	-	-	-	-
$B_3$	0.016	0.017	0.015	0.019	0.016	-	-	-	-	-	-	-	-	-	-	-	-
$\mu$	0.033	0.029	0.033	0.031	0.039	0.043	0.034	0.043	0.047	0.036	0.049	0.058	0.066	0.062	0.079	0.081	-
$\Delta E_{solv}$	0.264	0.263	0.267	0.264	0.263	0.270	0.277	0.270	0.267	0.284	0.285	0.292	0.298	0.302	0.296	0.297	0.322
$MW$	0.103	0.098	0.094	0.102	0.101	0.097	0.102	0.110	0.116	0.109	0.114	0.111	0.112	0.122	0.130	0.151	0.152
$X_{LogP3-AA}$	0.125	0.122	0.124	0.122	0.128	0.125	0.119	0.126	0.123	0.128	0.128	0.136	0.133	0.138	0.144	0.141	0.154

れている「作用様式」(Modes of action: MoA)の一つと一致している[15,24]。

次に、データを 4:1 の比率で 2 つのセットに分割し、前者をトレーニングセット、後者をバリデーションセットとした。このデータ分割は 10 パターン繰り返しおこなった。各セットのトレーニングデータに対して 5 分割クロスバリデーションを行い、各分割の平均二乗誤差 (MSE) を計算した。決定木の数と分岐の最大数であるハイパーパラメータは、ベイズ最適化によって、最小の MSE を与えるように決定した。10 個のデータセットの全ての  $R^2$  値を表 Table 4-6 に示す。この Table で DS<sub>4</sub> としているセットを最良の回帰モデルとして最終的に決定した。DS<sub>4</sub> の  $\log(\text{MIC})$  の予測値と  $\log(\text{MIC})$  の実験値の相関関係を Fig. 4-2 に示す。ここでトレーニングセットとバリデーションセットの  $R^2$  値はそれぞれ、0.9111 と 0.9073 である。10 個のデータセットの相関関係を Fig. 4-3 に示す。DS<sub>4</sub> の回帰モデルを使用したバリデーションセットの  $\log(\text{MIC})$  の値を表 Table 4-7 に示す。また、参考としてトレーニングセットのデータを Table 4-8 に示す。Table 4-7 が示すように予測は合理的に一致している。この「スモールデータ」を使用した機械学習でも最大偏差は 10%未満であり、平均絶対誤差は 3.2%と推定される。ただし、現状では MIC の平均絶対誤差は大きく、トレーニングセットとバリデーションセットでそれぞれ 27.9%と 28.2%であるため、予測精度を大幅に改善する必要があると言っても過言ではない。さらなる改善は確実に必要ではあるが、Fig. 4-2 における正の相関は良好であるため、現状の回帰モデルは化合物のスクリーニングに実際に役立つと結論づけることができる。これは上述の重要な記述子の選択からも正当なものであると言える。

回帰モデルの作成は半定量的に成功しているように結論づけられるため、その予測モデルに対して、新たな分子について各記述子の数値を適用して抗菌剤の探索を

Table 4-6. 各データセットにおいて 4:1 の比率で分割されたトレーニングおよびバリデーションデータを使用して作成した、ランダムフォレスト回帰モデルについての実験値と予測値の相関 ( $R^2$ )。この際、10 個のデータセットは元のすべてのデータの中からのランダムにサンプリングすることで生成された。

Data set	$R^2$	
	Training data	Validation data
DS <sub>1</sub>	0.7943	0.5186
DS <sub>2</sub>	0.9058	0.4007
DS <sub>3</sub>	0.8428	0.5471
DS <sub>4</sub>	0.9111	0.9073
DS <sub>5</sub>	0.9137	0.3073
DS <sub>6</sub>	0.9122	0.3363
DS <sub>7</sub>	0.7973	-0.2626
DS <sub>8</sub>	0.7993	0.5380
DS <sub>9</sub>	0.9143	0.3421
DS <sub>10</sub>	0.8151	0.3434



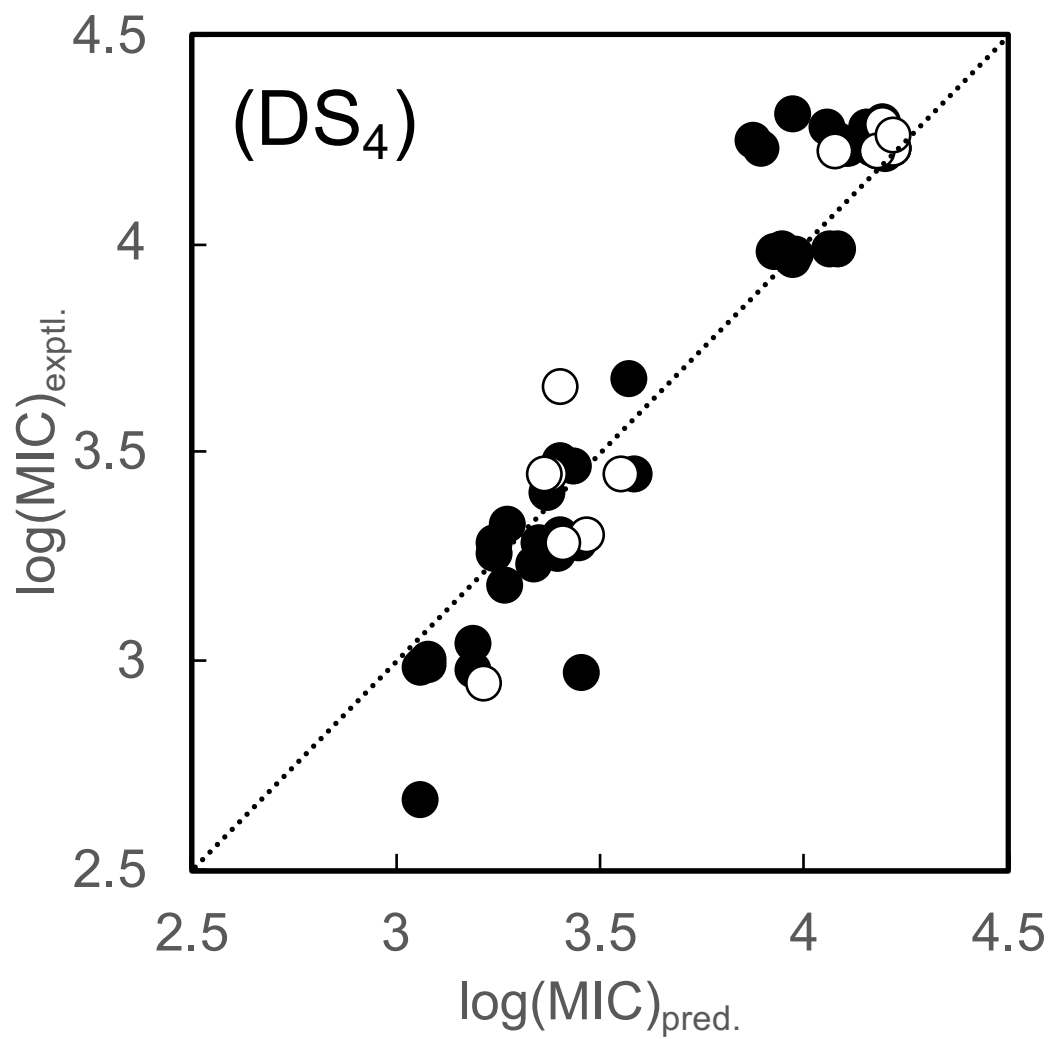


Figure 4-2.  $\log(\text{MIC})$ についての予測値と実験値との相関図。予測値は、最良のモデルとした DS<sub>4</sub> の回帰モデルを使用して評価した数値である。黒丸と白丸は、それぞれトレーニングセットとバリデーションセットの分子を示している。

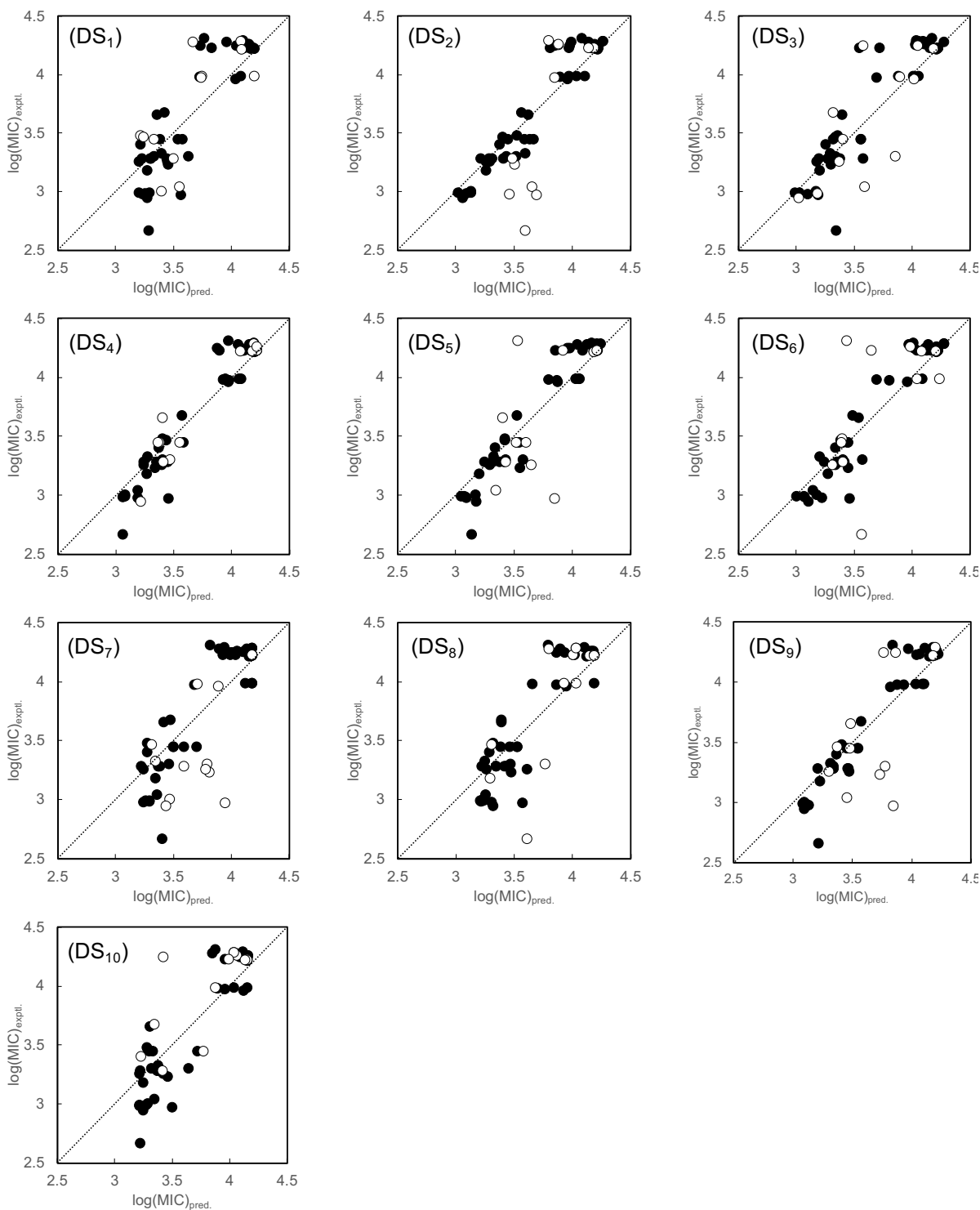


Figure 4-3. 全データセットについての MIC 値の予測値および実験値の散布図。黒丸と白丸は、それぞれトレーニングデータとバリデーションデータを示している。60 分子を含む全データは、トレーニング(48 データ)とバリデーション(12 データ)セットに 10 回分割されました。「DS<sub>n</sub>」(n = 1-10)は、n 番目の試行でのデータ分割を示している。R<sup>2</sup> 値を参照して、予測に最適な回帰モデルを選択した。

Table 4-7. バリデーションデータについての log(MIC)の予測値と実験値の比較。MIC の単位は ppm である。

Compound	Log (MIC <sub>pred.</sub> )	Log (MIC <sub>exptl.</sub> )	Absolute Error
Nerol	3.213	2.944	9.1%
Terpinen-4-ol	3.412	3.279	4.1%
(s)-cis-Verbenol	3.468	3.301	5.1%
Dihydrocarvone	3.361	3.447	2.5%
(+)-Pulegone	3.375	3.447	2.1%
$\alpha$ -Ionone	3.555	3.447	3.1%
( $\pm$ )-Camphor	3.402	3.653	6.9%
(+)-Limonene	4.182	4.217	0.8%
$\gamma$ -Terpinene	4.075	4.220	3.4%
Car-2-ene	4.220	4.228	0.2%
Geranyl acetate	4.220	4.255	0.8%
(-)-Menthol	4.194	4.283	2.1%

Table 4-8. データセット DS<sub>4</sub> に関する、トレーニングセットのすべての分子の MIC の予測値及び実験値 (ppm 単位)。

Compound Name	MIC <sub>pred.</sub>	MIC <sub>exptl.</sub>
(-)-cis-Myrtanyl amine	1146	460
Dihydrocarveol	2836	930
(1R)-(-)-Myrtenol	1542	950
(s)-(-)-Perilla alcohol	1147	960
(-)-Perilla aldehyde	1160	970
Carvacrol	1203	980
Thymol	1205	1000
trans-Isoeugenol	1541	1100
(-)-Carveol	1839	1500
Linalol	2188	1700
Piperitone	1738	1800
Geraniol	2478	1800
(R)-(-)-Carvone	1737	1900
$\alpha$ -Terpineol	2229	1900
Limonene oxide	2455	1900
(1s2s5s)-(-)-Myrtanol	2822	1900
(+)-Fenchol	2524	2000
Eugenol	1875	2100
Car-3-en-2-one	2354	2500
$\beta$ -Ionone	3857	2800
(1s)-(-)-Verbenone	2708	2900
(1R)-(-)-Myrtenal	2530	3000
(1R)-(-)-Fenchone	3721	4700
1,8-Cineole	9446	9100
Linalol oxide	9510	9400

Table 4–8. (Continued)

Compound Name	MIC <sub>pred.</sub>	MIC <sub>exptl.</sub>
p-Menth-6-ene-2,8-diol	8424	9500
(-)-Isoborneol	8796	9600
(+)-Isomenthol	11492	9600
(+)-Camphene	12168	9600
$\alpha$ -Terpinene	15890	16400
(-)-Limonene	15197	16500
(R)-(+)-Citronellal	7866	16700
$\beta$ -Citronellol	12724	16800
(s)-(-)- $\beta$ -Pinene	14661	16800
Car-3-ene	15610	16800
(+)- $\alpha$ -Pinene	16413	16800
(-)- $\alpha$ -Pinene	16503	16800
<i>p</i> -Cymene	12806	16900
$\alpha$ -Terpinolene	15552	16900
(-)-Menthone	7535	17500
1,4-Cineole	12027	17500
Neryl acetate	16254	17800
Menthyl acetate	15797	18100
$\alpha$ -Terpinylacetate	14305	18900
(-)-Borneol	11423	19000
(+)-Menthol	15628	19100
Linalyl acetate	15496	19300
Methyleugenol	9381	20300

Table 4-9. 特徴量重要度の分析によって選択された ESI 記述子を使用した、ランダムフォレスト回帰モデルから得られた MIC の予測値。予測は、精油の成分として知られている 175 の天然物に対して実行した。

Ranking	Compound Name	MIC <sub>pred.</sub>	Ranking	Compound Name	MIC <sub>pred.</sub>
1	Methyl salicylate	1854	34	2,6-Dimethylpyridine	3594
2	Ethyl benzoate	1859	35	Benzyl alcohol	3610
3	Methyl anthranilate	1884	36	Maltol	3670
4	L-Carvone	1918	37	<i>trans</i> -2-Methyl-2-butenal	3690
5	Piperitenone	1979	38	5-Ethyl-2-methylpyridine	3696
6	Cinnamic acid	2191	39	3-Ethylpyridine	3788
7	5-Methylquinoxaline	2263	40	Butylamine	3790
8	Piperonal	2302	41	Methyl angelate	3793
9	Geranial	2311	42	1-Penten-3-ol	3795
10	Vanillin	2325	43	Benzaldehyde	3801
11	Pinocarvone	2439	44	Acetaldehyde	3802
12	Neral	2553	45	2,3-Dimethylpyrazine	3805
13	Lavandulol	2644	46	Cinnamaldehyde	3872
14	2,3-Diethyl-5-methylpyrazine	2658	47	2,3,5-Trimethylpyrazine	3873
15	Methyl <i>cis</i> -cinnamate	2721	48	Pyrazine	3944
16	Methyl <i>trans</i> -cinnamate	2850	49	Hexanoic acid	3953
17	Hinokitiol	2892	50	Valeraldehyde	3956
18	Benzyl acetate	2917	51	Butyraldehyde	3957
19	$\delta$ -Terpineol	2986	52	Furfural	3971
20	Mequinol	3014	53	3,4-Dihydroxycinnamic acid	3979
21	$\beta$ -terpineol	3153	54	5,6,7,8-Tetrahydroquinoxaline	4001
22	Acetophenone	3157	55	2-Ethyl-3-methylpyrazine	4003
23	Isoquinoline	3210	56	Methylparaben	4053
24	Methyl chavicol	3300	57	2-Methylpyrazine	4068
25	6-Methyl-5-hepten-2-ol	3328	58	Propionic acid	4119
26	Phenylethylamine	3369	59	Butyric acid	4124
27	Phenethyl alcohol	3386	60	5-Methylfurfural	4128
28	Phenylacetaldehyde	3400	61	Isopentylamine	4132
29	Safrole	3415	62	2,5-Dimethylpyrazine	4135
30	3-Methyl-2-buten-1-ol	3418	63	Ethyl acetoacetate	4146
31	2,3,5,6-Tetramethylpyrazine	3495	64	2,6-Dimethylpyrazine	4157
32	<i>p</i> -Anisaldehyde	3546	65	Pyrrolidine	4167
33	Pyrrole	3567	66	2,3-Diethylpyrazine	4175

Table 4–9. (Continued: Ranking 67-133)

Ranking	Compound Name	MIC <sub>pred.</sub>	Ranking	Compound Name	MIC <sub>pred.</sub>
67	Ethylisovalerate	4228	100	Benzoic acid	6727
68	Isovaleraldehyde	4240	101	Cinnamyl alcohol	6733
69	2-Ethylpyrazine	4267	102	Bisabolol oxide A	6734
70	Isoamyl alcohol	4273	103	Nonanal	6772
71	Isopentyl formate	4306	104	Sorbic acid	6857
72	2-Ethyl-5-methylpyrazine	4309	105	Isomenthone	7014
73	Ethylcis-cinnamate	4347	106	Patchoulol	7016
74	2-Pentanol	4359	107	Decanal	7067
75	2-Ethyl-6-methylpyrazine	4437	108	$\beta$ -Atlantone	7070
76	Farnesol	4576	109	<i>cis</i> - $\beta$ -epi-Santalol	7106
77	<i>S</i> -Methylmethionine	4677	110	Isobutyl angelate	7192
78	3-Methyl-2-butanol	4734	111	Citropten	7204
79	Isopulegol	4736	112	$\beta$ -Santalol	7225
80	Isobutyl alcohol	4793	113	$\alpha$ -Santalol	7365
81	Ethyl propionate	4836	114	$\beta$ -Farnesene	7399
82	Ethyl <i>trans</i> -cinnamate	4843	115	$\alpha$ -Curcumene	7470
83	Benzyl benzoate	5214	116	Trimethylamine	7488
84	<i>trans</i> -Cinnamicanhydride	5483	117	Osthol	7602
85	Nootkatone	5558	118	<i>trans</i> -Nerolidol	7603
86	Columbianetin	5676	119	Bergamottin	7682
87	Piperidine	5743	120	Bergapten	7760
88	4-Methylanisole	5767	121	Eugenol acetate	7842
89	Vetiverol	5779	122	Ethyl acetate	7860
90	Octanal	5829	123	$\beta$ -Bisabolene	7868
91	Hydroquinone	5882	124	Propylparaben	7909
92	Hydroxycitronellal	6162	125	$\beta$ -Elemene	7981
93	Curzerene	6191	126	Leptospermone	8093
94	Bisabolol oxide B	6255	127	$\delta$ -Elemene	8166
95	Acetylsalicylic acid	6399	128	Chamazulene	8332
96	$\alpha$ -Atlantone	6504	129	Lindestrene	8390
97	Isoamyl angelate	6506	130	Cedrol	8416
98	1-Octanol	6622	131	$\gamma$ -Elemene	8434
99	Salicylic acid	6712	132	Zingiberene	8593

Table 4–9. (Continued: Ranking 133-175)

Ranking	Compound Name	MIC <sub>pred.</sub>	Ranking	Compound Name	MIC <sub>pred.</sub>
133	Germacrene D	8642	155	Caryophyllene	10442
134	$\alpha$ -Farnesene	8703	156	$\alpha$ -Selinene	10481
135	Caryophyllene oxide	8750	157	$\delta$ -Cadinene	10540
136	Rose oxide	8940	158	$\gamma$ -Cadinene	10540
137	Isoamyl acetate	8951	159	$\epsilon$ -Muurolene	10554
138	$\gamma$ -Bisabolene	9005	160	$\beta$ -Cadinene	10644
139	Globulol	9395	161	$\beta$ -trans-Ocimene	10681
140	Myrcene	9405	162	$\beta$ -cis-Ocimene	10719
141	$\alpha$ -Himachalene	9629	163	Fernene	10729
142	Bisabolene oxide	9673	164	$\beta$ -Citronellene	11239
143	Spathulenol	9675	165	$\alpha$ -Cadinene	11366
144	Citronellyl formate	9681	166	$\alpha$ -Copaene	12258
145	Humulene	9718	167	Alloaromadendrene	12363
146	$\beta$ -Selinene	9765	168	$\beta$ -Phellandrene	12415
147	$\beta$ -Myrcene	9960	169	Thujopsene	12427
148	Phytol	10000	170	Bornyl acetate	12430
149	$\beta$ -Himachalene	10196	171	$\alpha$ -Thujene	12548
150	$\gamma$ -Himachalene	10302	172	$\alpha$ -Cubebene	12582
151	$\alpha$ -Phellandrene	10322	173	$\beta$ -Cubebene	12707
152	Sclareol	10341	174	$\alpha$ -Gurjunene	13095
153	$\beta$ -Cymene	10427	175	Sabinene	14517
154	$\beta$ -Caryophyllene	10442			



試みた。研究室で計算していた化合物データベースのうち、175 種の分子のデータについて予測モデルを用いて、抗菌活性の予測を試みた。これらの分子について、予測モデルの作成に使用した記述子の計算を行い、それらの数値を予測モデルに適用して予測値を算出した。予測された各分子の MIC の値を Table 4-9 に示す。続いて、予測された MIC の値を参照して化合物のランキングを行なった。

高活性である、すなわち MIC 値が小さいと予測される分子を Table 4-10 にまとめた。Fig.4-4 にそれらの分子構造を示す。これらの化合物選択の妥当性を確認するために、Scopus データベースを使用して文献検索を行なった[25]。化合物名を使用した検索は、2020年8月末まで繰り返し行い、最終的に274件の参考文献が収集された。各分子の詳細については後述するが、この調査から2種類の説明によって予測の正当性が表現できることがわかった。1つは化合物の抗菌活性を裏付ける直接的な証拠であり、もう1つは予測された化合物を含む精油の抗菌活性を示す証拠である。

Table 4-10 にまとめた化合物について調査した内容を列挙する。

予測対象とした分子の中で、最も活性が高かったのはサリチル酸メチル (Methyl salicylate) で、MIC の予測値は 1853.96 ppm である。この分子が多く含まれる植物としては、ツツジ科の小低木で、ネパール原産の *Gaultheria fragrantissima* や北アメリカ原産の *Gaultheria procumbens* L. が挙げられる。これらの植物は伝統的な薬用植物として知られており、香料や鎮痛用の塗布剤として使用されている。これらの植物から抽出される精油に占めるサリチル酸メチルの割合は 95% を超えているため[26,27]、これらの精油の薬効はこの成分に起因しているものと考えられる。

次に活性が高いのは安息香酸エチル (Ethyl benzoate) で、MIC の予測値は 1858.92

Table 4-10. 高い抗菌活性を示すと予測された上位 20 位の分子と、それらの MIC の予測値 (単位は ppm)。分子名の右側にある\*は、化合物またはそれを含む精油のいずれかが抗菌活性を有することが文献で報告されている分子であることを示す。

Compound name	MIC <sub>pred.</sub>	Log(MIC <sub>pred.</sub> )
Methyl salicylate	1854	3.268
Ethyl benzoate	1859	3.269
Methyl anthranilate	1884	3.275
L-Carvone*	1918	3.283
Piperitenone*	1979	3.296
Cinnamic acid*	2191	3.341
5-Methylquinoxaline	2263	3.355
Piperonal*	2302	3.362
Geranial*	2311	3.364
Vanillin*	2325	3.366
Pinocarvone*	2439	3.387
Neral*	2553	3.407
Lavandulol*	2644	3.422
2,3-Diethyl-5-methylpyrazine	2658	3.425
Methyl <i>cis</i> -cinnamate*	2721	3.435
Methyl <i>trans</i> -cinnamate*	2850	3.455
Hinokitiol*	2892	3.461
Benzyl acetate	2917	3.465
$\delta$ -Terpineol	2986	3.475
Mequinol*	3014	3.479

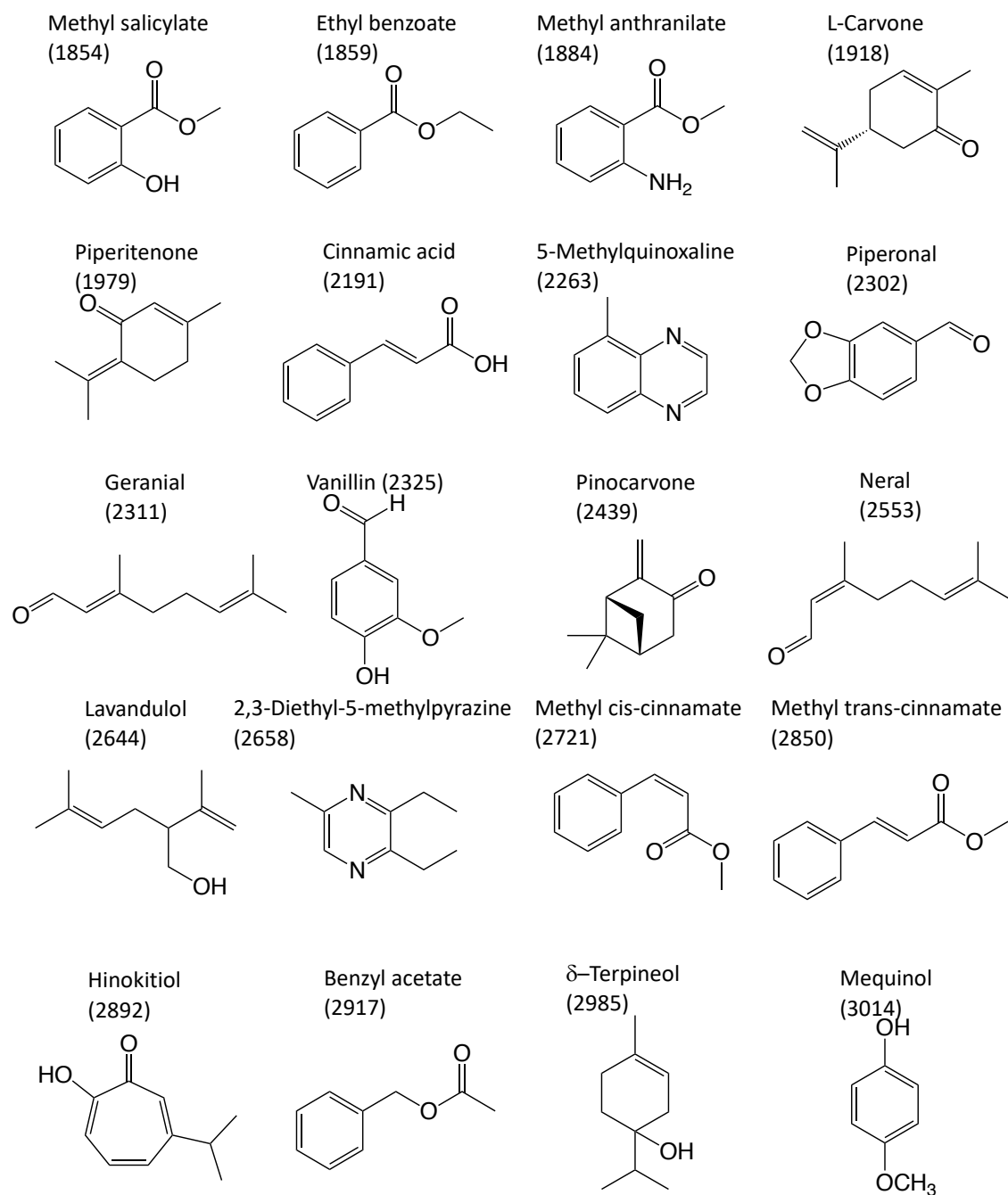


Figure 4-4. 予測モデルによる評価で、活性が高い(MIC が小さい)と予測された分子。括弧の中の数値はそれぞれの分子の MIC の予測値である(単位:ppm)。

ppm である。この分子は香水の成分や果実の香りの人工香料として用いられている。

次に活性が高いのはアントラニル酸メチル(Methyl anthranilate)で、MIC の予測値は 1884.00 ppm である。この分子はジャスミンやブドウなどに含まれており、ブドウ風味の香料として用いられている。

次に活性が高いのはL-カルボン(L-Carvone)で、MIC の予測値は 1918.23 ppm である。この分子が多く含まれる植物としてはハッカ属のスペアミント(Mentha spicata L.)が挙げられる。この植物は古くからハーブの一種として用いられており、主に香料として用いられている。その精油に占める L-カルボンの割合は 50%を超えている。また、精油から単離されたL-カルボンの抗菌活性は抽出元の精油の抗菌活性に匹敵することが報告されており[28]、この精油の薬効はこの成分に起因しているものと考えられる。

次に活性が高いのはピペリテノン(Piperitenone)で、MIC の予測値は 1979.06 ppm である。この分子が多く含まれる植物としては、シソ科の垂低木でハーブとして使用されているジジフォラ属が挙げられる。これらの植物の精油自体とメタノール抽出物に抗菌活性があることが報告されている[29–31]。

次に活性が高いのはケイ皮酸(桂皮酸、Cinnamic acid)で、MIC の予測値は 2191.09 ppm である。この分子は植物界に広く存在し、ケイ皮酸メチルやケイ皮酸エチル、3,4-ジヒドロキシケイ皮酸(コーヒー酸、Caffeic acid)など、多くの誘導体が知られている。桂皮酸及びその誘導体(エステル・置換誘導体・アミド)についての研究で、いずれの分子でも一般的な抗菌分子の標準値に匹敵する抗菌活性を持つことが報告されている[32–35]。

次に活性が高いのは 5-メチルキノキサリン(5-Methylquinoxaline)で、MIC の予測値は 2262.77 ppm である。この分子はロースト臭、あるいはコーヒーのような加熱香気

を有している。食品中に天然に存在するが加熱により生成する場合もある。欧米では、肉製品やソフトキャンディなど、様々な加工食品において香りを再現するために添加されている。

次に活性が高いのはピペロナル(Piperonal)で、MICの予測値は2302.24 ppmである。この分子はスミレやコショウなど、様々な植物に含まれており、食品を媒介する細菌性病原体(*Yersinia enterocolitica*, *Listeria monocytogenes* and *Salmonella typhimurium*)の活性を阻害することが報告されている[36]。また、コショウ科のハイゴシヨウ由来の精油から単離されたピペロナルについて調査した結果、イネ条斑細菌病菌に対して強い阻害能をもつことが報告されている[37]。

次に活性が高いのはゲラニール(Geraniol)で、MICの予測値は2310.84 ppmである。この分子は一对のシス-トランス異性体をもつアルデヒドであるシトラール(Citral)のうち、trans(*E*)体に当たる分子である。レモングラスなどから採れる精油の主成分であり、強いレモン臭をもつ。レモンなどの柑橘類にも含まれ、香水や香料などに用いられている。レモンバームというシソ科のハーブについての研究で、抽出される精油が大腸菌に対する阻害活性をもつことと、精油に占める成分の中で最も多いのがゲラニールで、次に多いのがcis(*Z*)体に当たるネラールであることが報告されている[38]。

次に活性が高いのはバニリン(Vanillin)で、MICの予測値は2324.72 ppmである。この分子はバニラの香りの主要な成分であり、アイスクリームやチョコレート、タバコや香水などに香料として使用される。バニリンの抗菌活性に関する研究で、バニリンは主に膜に対して活性を持つ化合物であり、大腸菌やリステリア菌などの呼吸経路を阻害することが報告されている[39]。

次に活性が高いのはピノカルボン(Pinocarvone)で、MICの予測値は2438.55 ppm

である。この分子はカルボンの構造異性体であり、ピネンに類似した構造をもつ。この分子を含む植物の一例として、ミロタムヌス・フラベリフォリア (*Myrothamnus flabellifolius*) が挙げられる。この植物はアフリカの砂漠地帯に自生していて、その優れた吸水性と保湿性から「復活の木」と呼ばれ、アフリカでは伝統的にお茶やスパイス、薬などとして使用されている。この植物の精油に関する研究では、精油に抗菌活性があり、精油に含まれる 85 種の化合物のうち、ピノカルボンが占める割合は二番目に多い 10%であることが報告されている[40]。

次に活性が高いのはネラール(Neral)で、MIC の予測値は 2553.29 ppm である。この分子は先述のシトラールのシス-トランス異性体の cis (Z) 体に当たる分子である。ブロードリーフタイムやコウスイガヤなどの精油に含まれ、抗菌作用があることが報告されている[41,42]。シトラールが含まれる植物には同程度の割合でゲラニールが含まれていることが多い。

次に活性が高いのはラバンジュロール(Lavandulol)で、MIC の予測値は 2643.69 ppm である。この分子はラベンダー油に含まれる。ラベンダー油には抗菌作用があることが報告されている[43]。精油に占める最も多い成分はリナロールで、ラバンジュロールは 3 番目に多い成分である[44]。また、リナロールとラバンジュロールは構造異性体である。

次に活性が高いのは 2,3-ジエチル-5-メチルピラジン (2,3-Diethyl-5-methylpyrazine)で、MIC の予測値は 2658.28 ppm である。この分子は別名ヘーゼルナッツピラジンと呼ばれ、ロースト臭を有する物質である。

次に活性が高いのは cis-ケイ皮酸メチル (Methyl cis-cinnamate)で、MIC の予測値は 2720.57 ppm である。この分子は先述のケイ皮酸をエステル化した誘導体の一種で

ある。*Ocimum americanum* というバジルの一種から採取される精油の主成分であり、この精油には複数種類のカンジダ菌に対する抗真菌作用があることが報告されている[45]。

次に活性が高いのは trans-ケイ皮酸メチル (Methyl trans-cinnamate) で、MIC の予測値は 2849.71 ppm である。この分子も先述のケイ皮酸をエステル化した誘導体の一種である。*Senecio nutans* というキク科の多肉植物の主成分であり、コレラ菌に対して抗菌作用があることが報告されている[46]。

次に活性が高いのはヒノキチオール (Hinokitiol) で、MIC の予測値は 2891.94 ppm である。この分子はヒノキ科の植物に特有であり、不飽和七員環構造を有するトロポロンの誘導体の一つである。香料として歯磨き粉やヘアトニックなどに使用されている。青森ヒバから採取されたヒノキチオールとその類縁種について抗真菌・抗菌作用があることが報告されている[47, 48]。

次に活性が高いのは酢酸ベンジル (Benzyl acetate) で、MIC の予測値は 2917.36 ppm である。この分子は多くの植物の花に含まれており、ジャスミンやイランイランなどの精油の主成分として知られている。この分子自体がジャスミンの香りを有しており、香水や化粧水に用いられている。

次に活性が高いのは $\delta$ -テルピネオール ( $\delta$ -Terpineol) で、MIC の予測値は 2985.86 ppm である。この分子は松根油や月桂樹の精油などに含まれる。異性体として他に $\alpha$ 、 $\beta$ 、 $\gamma$ -テルピネオールが見られ、主成分は $\alpha$ -テルピネオールである。 $\alpha$ 、および $\delta$ -テルピネオールはショウガ科のブラックカルダモン (*Amomum subulatum*) から採取される精油に含まれ、精油自体に抗真菌活性、殺虫活性があることが報告されている[49]。

次に活性が高いのはメキノール (Mequinol) で、MIC の予測値は 3013.98 ppm であ

る。この分子が含まれる植物の一例にクワ科の *Dorstenia turbinata* が挙げられる。この植物の精油から単離された分子についての抗菌活性について研究されており、メキノールにも中程度の選択的活性があると報告されている[50]。

以上のように、Table 4-10 に記載されている化合物のうち、これまでに抗菌活性が実験的に報告されているのは、我々の知る限りでは、L-カルボン、ピペリテノン、ケイ皮酸、ピペロナール、ヒノキチオール、およびメキノールである。また、ゲラニアール、バニリン、ピノカルボン、ネラール、ラバンジュロール、cis-ケイ皮酸メチル、trans-ケイ皮酸メチルを含む精油には抗菌活性があることが知られている。これらの実験情報は、今回行なった解析手法による化合物スクリーニングの正当性を示していると考えられる。これは、ESI 記述子が抗菌活性を揺する化合物の仮想スクリーニングにおいても有用であることを意味している。

#### 4-4 まとめ

電子状態計算から得られる数値を用いて、大腸菌の阻害活性の予測モデルを作成するためにランダムフォレストによる回帰分析を行なった。

まず、抗菌作用をもつ分子について電子状態計算を行った。それらの計算結果から得られる数値情報を電子的記述子として定義した。各分子の大腸菌に対する最小発育阻止濃度 (MIC) の値を目的変数、各記述子の値を説明変数として、ランダムフォレストによる回帰分析を行なった。解析に際して、まず 60 個の分子をランダムにトレーニングデータとバリデーションデータで 4:1 に分割し、それを 10 セット作成した。次に、目的変数に関する重要度が 0.1 未満の記述子を逐次削除して変数選択を行なった。



そして、各セットでトレーニングデータをさらに 5 分割し、5 分割考査検証を用いて、各分割での平均二乗誤差(MSE)を算出した。ベイズ最適化を使用して、MSE の平均値が最小となるハイパーパラメータ(決定木の数、最大の分岐数)を決定した。選択された記述子とハイパーパラメータを用いてランダムフォレストの回帰モデルを作成した。各セットのバリデーションデータをそれぞれの回帰モデルに適用し、各モデルの性能を評価した。各セットで作成した回帰モデルのうち、バリデーションデータの予測値の  $R^2$  が最大のもので、予測モデルとして決定した。そのモデルでのトレーニングデータとバリデーションデータの  $R^2$  値はそれぞれ 0.911 と 0.907 であった。選択された記述子は 6 種 ( $S(\rho_{UV})$ 、 $\lambda_{A \rightarrow S_0}$ 、 $B_1$ 、 $\Delta E_{solv}$ 、 $MW$ 、 $XLogP3-AA$ ) であった。選択された記述子は、極性溶媒中での溶媒和の強さと、分子の大きさに関するものである。これらは、脂質二重層からなる細胞膜への取り込みに直接関係すると思われる記述子であり、精油成分が細胞膜に取り込まれて抗菌作用を発現するとの知見と合致するものである。

得られたランダムフォレストによる回帰モデルを用いて、精油成分および食品香料からなる合計 175 分子の MIC の予測を行った。その中で抗菌活性の高い(MIC の小さい)分子として、予測された 20 個の分子について論文検索を行なった。その結果、各分子、およびそれらを含む精油が実際に抗菌活性をもつという情報が記載されていた。以上のことから、この研究で用いた記述子のセットが、精油の抗菌活性に関する定量的記述に有用であるといえることができる。

#### 4-5 参考文献

- [1] E. O. Pyzer-Knapp, T. Laino, Eds. *Machine Learning in Chemistry: Data-Driven Algorithms, Learning Systems, and Predictions*, American Chemical Society, Washington, DC, (2020).
- [2] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, Wiley-VCH, Weinheim, (2009).
- [3] A. Mauri, *alvaDesc: A Tool to Calculate and Analyze Molecular Descriptors and Fingerprints*. In: Roy K. (eds) *Ecotoxicological QSARs. Methods in Pharmacology and Toxicology*. Humana, New York, NY, (2020).
- [4] T. Engel, J. Gasteiger, Eds. *Chemoinformatics: Basic Concepts and Methods*, Wiley-VCH, Weinheim, (2018).
- [5] M. Sugimoto, R. Iwane, T. Ideo, *AIP Conf. Proc.*, 1702, 090038, (2015).
- [6] M. Sugimoto, T. Ideo, A. B. Manggara, K. Yoshida, T. Inoue, *J. Comp. Aid. Chem.*, **20**, 65 (2019).
- [7] M. Sugimoto, A. B. Manggara, K. Yoshida, T. Inoue, T. Ideo, *Mol. Inf.*, **39**, 1900121 (2020).
- [8] A. A. Buglak, M. A. Filatov, M. A. Hussain, M. Sugimoto, *J. Photochem. Photobiol. A*, **403**, 112833 (2020).
- [9] K. Sakanoue, M. Motoda, M. Sugimoto, S. Sakaki, *J. Phys. Chem.*, **103**, 5551 (1999).
- [10] M. Sugimoto, S. Sakaki, K. Sakanoue, M. D. Newton, *J. Appl. Phys.*, **90**, 6092 (2001).

- [11] I. Takashima, M. Kinoshita, R. Kawagoe, S. Nakagawa, M. Sugimoto, I. Hamachi, A. Ojida, *Chem. Euro. J.*, **20**, 2184 (2014).
- [12] M. Sugimoto, I. Yamasaki, N. Mizoe, M. Anzai, S. Sakaki, *Theor. Chem. Acc.*, **102**, 377 (1999).
- [13] M. K. Rai, K. V. Kon, Eds. *Fighting Multidrug Resistance with Herbal Extracts, Essential Oils and their Components*, Elsevier, Amsterdam (2013).
- [14] A. K. Ghosh, M. Brindisi, D. Shahabi, M. E. Chapman, A. D. Mesecar, *ChemMedChem*, **15**, 907 (2020).
- [15] S. Burt, *Intern. J. Food Microbiol.*, **94**, 223 (2004).
- [16] V. Moleyar, P. Narasimham, *Indian J. Exp. Biol.*, **25**, 781 (1987).
- [17] D. A. McQuarrie, J. D. Simon, *Physical Chemistry. A Molecular Approach*, Univ. Sci. Books, Sausalito, (1997).
- [18] Wolfram Research Inc., *Mathematica ver. 11.0*, Wolfram Research Inc., Champaign, Illinois (2018).
- [19] C. Hansch and A. Leo, *Exploring QSAR*, American Chemical Society, Washington, DC, (1995).
- [20] <https://pubchem.ncbi.nlm.nih.gov/>
- [21] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D.

Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery, Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox, *Gaussian 16, Revision B.01*, Gaussian Inc., Wallingford CT (2016).

[22] <https://www.python.org/>

[23] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, *J. Mach. Learn. Res.*, **12**, 2825 (2011).

[24] S. G. Griffin, S. G. Wyllie, J. L. Markham, and D. N. Leach, *Flavour Fragr. J.*, **14** (5), 322 (1999).

[25] <https://www.scopus.com/>

[26] B. P. Pandey, R. Thapa, and A. Upreti, *Asian Pac. J. Trop. Med.*, **10** (10), 952 (2017).

[27] A. Magiera, M. Sienkiewicz, M. A. Olszewska, A. Kicel, and P. Michel, *Acta Poloniae Pharmaceutica - Drug Research*, **76** (1), 93 (2019).

[28] K. K. Aggarwal, S. P. S. Khanuja, A. Ahmad, T. R. Santha Kumar, V. K. Gupta,

- and S. Kumar, *Flavour Fragr. J.*, **17** (1), 59 (2002).
- [29] S. Ozturk and S. Ercisli, *Food Control*, **18** (5), 535 (2007).
- [30] S. Ozturk and S. Ercisli, *J. Ethnopharmacol.*, **106** (3), 372 (2006).
- [31] M. Masrournia and A. Shams, *Asian J. Chem.*, **25** (12), 6553 (2013).
- [32] G. G. F. Nascimento, J. Locatelli, P. C. Freitas, and G. L. Silva, *Braz. J. Microbiol.*, **31** (4), 247 (2000).
- [33] A. O. Gill and R. A. Holley, *Appl. Environ. Microbiol.*, **70** (10), 5750 (2004)
- [34] B. Narasimhan, D. Belsare, D. Pharande, V. Mourya, and A. Dhake, *Eur. J. Med. Chem.*, **39** (10), 827 (2004).
- [35] M. Unlu, E. Ergene, G. V. Unlu, H. S. Zeytinoglu, and N. Vural, *Food Chem. Toxicol.*, **48** (11), 3274 (2010).
- [36] B. L. Bowles and V. K. Juneja, *J. Food Saf.*, **18** (2), 101 (1998).
- [37] P. Chanprapai and W. Chavasiri, *J. Integr. Agric.*, **16** (11), 2513 (2017).
- [38] N. Mimica-Dukic, B. Bozin, M. Sokovic, and N. Simin, *J. Agric. Food Chem.*, **52** (9), 2485 (2004).
- [39] D. J. Fitzgerald, M. Stratford, M. J. Gasson, J. Ueckert, A. Bos, and A. Narbad, *J. Appl. Microbiol.*, **97** (1), 104 (2004).
- [40] A. M. Viljoen, M. E. Klepser, E. J. Ernst, D. Keele, E. Roling, S. Van Vuuren, B. Demirci, K. H. C. Başer, and B.-E. van Wyk, *S. Afr. J. Bot.*, **68** (1), 100 (2002).
- [41] K. Ložiene, P. R. Venskutonis, A. Šipailiene, and J. Labokas, *Food Chem.*, **103** (2), 546 (2007).
- [42] L. G. de Toledo, M. A. dos Santos Ramos, L. Spósito, E. M. Castilho, F. R. Pavan,

- É. De Oliveira Lopes, G. J. Zocolo, F. A. N. Silva, T. H. Soares, A. G. dos Santos, T. M. Bauab, and M. T. G. de Almeida, *Int. J. Mol. Sci.*, **17** (8), (2016).
- [43] H. M. A. Cavanagh and J. M. Wilkinson, *Phytother. Res.*, **16** (4), 301 (2002).
- [44] B. Blažeković, W. Yang, Y. Wang, C. Li, M. Kindl, S. Pepeljnjak, and S. Vladimir-Knežević, *Ind. Crops Prod.*, **123**, 173 (2018).
- [45] P. R. N. Vieira, S. M. de Moraes, F. H. Q. Bezerra, P. A. Travassos, T. R. Oliveira, and M. G. V. Silva, *Ind. Crops Prod.* **55**, 267 (2014).
- [46] A. Paredes, Y. Leyton, C. Riquelme, and G. Morales, *SpringerPlus*, **5** (1), (2016).
- [47] Y. Inamori, S. Shinohara, H. Tsujibo, T. Okabe, Y. Morita, Y. Sakagami, Y. Kumeda, and N. Ishida, *Biol. Pharm. Bull.*, **22** (9), 990 (1999).
- [48] Y. Morita, E. Matsumura, T. Okabe, T. Fukui, T. Ohe, N. Ishida, and Y. Inamori, *Biol. Pharm. Bull.*, **27** (10), 1666 (2004).
- [49] G. Boatto, G. Pintore, M. Palomba, F. De Simone, E. Ramundo, and C. Iodice, *Fitoterapia*, **65** (3), 279 (1994).
- [50] B. Ngameni, V. Kuete, I. K. Simo, A. T. Mbaveng, P. K. Awoussong, R. Patnam, R. Roy, and B. T. Ngadjui, *S. Afr. J. Bot.*, **75** (2), 256 (2009).

## 第5章 結論

より高性能でより安全な医薬品の開発は、人々が健やかに暮らしていくために有益なことである。物質の性質はその分子の電子状態に由来するものであるため、より良い薬を生み出すためには、それらを構成する成分のどのような電子的特徴が、目的の薬効あるいは副作用を示す要因になっているかを解明する必要がある。本研究では、それらの方法を開発することを目的として、数種類の異なる機能に関する機能性分子の組について、我々のグループで開発している電子的記述子を使用して情報化学的手法を用いて解析を行い、手法の有用性を実証することができた。

第2章では電子状態情報を蓄積する分子情報データベースを作成し、“cancer”というキーワードで検索をして該当した分子について、クラスター解析を行い、物性ごとに分類できるかを試みた。いくつかの電子的記述子を組み合わせることで、物性が異なる分子を分類できることがわかった。

第3章ではガンの治療薬として期待される FASN 阻害剤について情報化学的手法を適用し、それらの物性の定量的回帰モデルを作成した。モデル作成にあたり目的変数とする活性値を説明する記述子(説明変数)の絞り込みを行なった結果、選択された記述子は酸化還元反応や状態変化を伴う化学反応に関連するものであった。

第4章では抗菌作用を示す精油に含まれる化学物質について、電子状態情報を用いて回帰モデルを作成した。作成した回帰モデルに対して、モデル作成に使用したのとは別の175種の天然成分に関する電子的記述子の数値を適用した。適用した結果、高活性であると予測された分子について、Scopus データベースを用いて文献検索を行なった。その結果、それらの分子、あるいはそれらを含む精油が実際に抗菌活性をもつという記載がある文献が確認できた。

以上のことから、電子的記述子を使用する現在の解析手法は化学、及び創薬化学における様々な分野で有用であるはずであるといえる。我々のグループ内では、電子的記述子のセットの適用性、精度の向上、および抗菌化合物スクリーニングへの広範な適用に関する調査が進行中であり、今後別の場で報告される予定である。



## 謝辞

本論文は平成 25 年 4 月から令和 3 年 3 月までの熊本大学大学院自然科学研究科博士前期課程および博士後期課程在籍時の 8 年間に遂行した研究成果に基づいて構成されたものであります。

本研究の遂行ならびに本論文の作成に至ったことは、指導教員であります杉本学准教授の始終変わらぬ愛情ある叱咤激励、ご指導ご鞭撻の賜物であり、ここに心より感謝し、厚く御礼申し上げます。

本研究の遂行にあたり、研究室の研究員、大学院生、学部生の方々には、日頃より私の研究についてご討論していただきました。深く感謝いたします。特に、非線形回帰学習のプログラム構築および実行を担当して頂きました吉田一輝氏、Algafari Bakti Manggara 氏に心より感謝いたします。

最後に、長い大学生活を広い心で暖かく見守ってくださった両親を始めとする家族に深い感謝と愛情の意を表して謝辞といたします。