

# 医薬品と健康食品の相互作用に 関する知識抽出技術の開発

令和6年度 博士論文

徳島大学 大学院先端技術科学教育部

物質生命システム工学専攻 生命テクノサイエンスコース

3年 吉崎万莉

## 要旨

「健康食品(HFPs)」は健康の維持・増進に関連する食品や製品を指し、正しく利用するためには正しい情報共有の基盤が必要である。例えば、HFPsの一部は医薬品の成分と相互作用する可能性があり、特にステビアやステビオサイドのような成分は注意が必要である。情報抽出に関する課題や HFPs と医薬品の相互作用情報の重要性に鑑み、本研究では情報科学の手法を用いた新しいワークフローの確立を目指している。酵素をハブとしての医薬品—酵素—HFPs のネットワーク分析により、効果的な情報抽出が期待される。

まず、様々なデータベース (KEGG, National Institutes of Health, National Institutes of Biomedical Innovation, Health and Nutrition, PubMed) や文献情報から、医薬品、酵素、HFP、の情報を収集した。それらの相互作用情報を収集するために BioGPT および共起検索を用いた。共起検索において有用性が見出され、条件検討を行った。KEGG から収集した医薬品—酵素ペアと共起検索を用いて収集した酵素—HFP ペアからネットワークを構築し、Drug-HFP interactions (DHIs)を得た。ネットワークの構築には3種類の方法を試し、Cytoscape が最適であった。Cytoscape を利用して可視化することにより、ハブとなる酵素で DHIs をグループ分けすることができた。この抽出方法およびグループ分けを用いた方法を方法1とした。

さらに、この共起検索の検討には評価方法が必要不可欠であるため、既知の DHIs 情報を収集した。収集方法としては3つの方法を試した。1つは医薬品と HFPs を共起検索して設定した閾値以上であれば既知の相互作用とする方法である。取得できた医薬品—HFP ペアが非常に少なく、閾値以下のペアにも相互作用が確認されたため、不相当と考えた。次は、医薬品と HFPs の相互作用情報を記載しているデータベースまたは書籍からの引用する方法である。複数のデータベースや書籍を吟味した結果、臨床においても使用可能とされている書籍に有用性を見出した。しかし、情報抽出に時間がかかるため、他の方法を検討した。3つ目は医薬品と HFPs の相互作用情報が記載されている可能性がある論文を読み、抽出する方法である。しかしながら、数多の論

文を読む必要があり、現実的ではなかった。そのため、書籍の情報を基にゴールドスタンダードデータを作成した。このゴールドスタンダードデータを既知の DHI 情報として、方法 1 で取得した DHI を評価した。その結果、方法 1 は一次スクリーニングとしては有用であるが、酵素で分けたグループをさらに関係性ごとに分ける必要があると考えられる。

方法 1 で抽出した DHI を関係性ごとに分類する方法として DeepWalk を用いた。DeepWalk は、ネットワーク埋め込みの一種で、グラフデータを対象としてノードの埋め込み表現を学習するためのアルゴリズムである。この方法ではネットワークにおけるトポロジーを基に関係性を導いており、実際の相互作用との関連性は低かった。そのため、他の二次スクリーニング方法を検討した。

さらに、DHI の解析を進めるための二次スクリーニングとして、自然言語処理技術を適用し、最適な方法を検討し、決定した。生物医学的な単語の定量化に広く用いられている BioWordVec を適用し、医薬品と HFPs に含まれる各単語の生物医学的な文脈を定量化した。医薬品と HFPs の差分ベクトルを使用する方法、酵素に関連する医薬品と HFPs を選抜して差分ベクトルを使用する方法を検討した。その結果、酵素に関連する医薬品と HFPs を選抜する方法を採用した。この酵素名について、異なる酵素の名称を使用した結果を比較し、最適な酵素の名称を決定した。ネットワーク解析の結果を統合することで、未知および既知の DHI を効率的に抽出するワークフローを構築できた。

## 目次

緒論 .....	1
第 1 章 医薬品, 酵素および HFPs のリスト化 .....	3
第 1 節 方法 .....	3
第 1 項 KEGG を利用した医薬品と代謝酵素の相互作用リストの作成 .....	3
第 2 項 HFPs のデータセットの準備 .....	3
第 2 節 結果 .....	4
第 3 節 考察 .....	45
第 2 章 BioGPT を用いた DHIs の取得 .....	48
第 1 節 方法 .....	48
第 2 節 結果 .....	49
第 3 節 考察 .....	57
第 3 章 方法 1 : PubMed に基づく代謝酵素と HFPs の共起検索とネットワーク分 析と酵素によるグルーピング .....	59
第 1 節 方法 .....	60
第 1 項 PubMed に基づく代謝酵素と HFPs の共起検索 .....	60
第 2 項 共起検索を用いたゴールドスタンダードデータの取得 .....	60
第 3 項 書籍情報を用いたゴールドスタンダードデータの取得 .....	61
第 4 項 PubMed の文献を用いたゴールドスタンダードデータの取得 .....	61
第 5 項 ネットワーク構築および酵素によるグルーピング .....	61
第 6 項 ゴールドスタンダードとの一致率 .....	61
第 2 節 結果 .....	62
第 1 項 共起検索を用いたゴールドスタンダードデータの有用性 .....	62
第 2 項 書籍情報を用いたゴールドスタンダードデータの有用性 .....	62
第 3 項 PubMed の文献を用いたゴールドスタンダードデータの有用性 .....	63
第 4 項 ゴールドスタンダードデータの比較 .....	63

第 5 項	共起検索における有用性および閾値 .....	64
第 6 項	閾値 20 とした共起検索により選抜された DHIs .....	66
第 3 節	考察 .....	72
第 4 章	Deepwalk を用いた DHIs の分類.....	75
第 1 節	方法 .....	75
第 1 項	KEGG を利用した医薬品と代謝酵素の相互作用リストの作成 .....	75
第 2 節	結果 .....	76
第 3 節	考察 .....	76
第 5 章	Biowordvec を用いて数値化した医薬品と HFPs の差分ベクトルにおけるクラスタリング.....	78
第 1 節	方法 .....	78
第 1 項	Biowordvec を用いた医薬品および HFPs の数値化(ベクトル化) .....	78
第 2 項	差分ベクトル計算と一次クラスタリング .....	79
第 2 節	結果 .....	79
第 3 節	考察 .....	80
第 6 章	方法 2-1：酵素と直接的に関連する医薬品および HFPs スクリーニング.81	
第 1 節	方法 .....	81
第 1 項	Biowordvec を用いた酵素, 医薬品および HFPs の数値化(ベクトル化) .....	81
第 2 項	酵素と直接的に関連する医薬品および HFPs スクリーニング .....	82
第 3 項	差分ベクトル計算とクラスタリングおよびハブとなる酵素による後処理.....	82
第 2 節	結果 .....	82
第 1 項	酵素の名称による数値化への影響.....	82
第 2 項	方法 2-1 の有用性 .....	87
第 3 節	考察 .....	89

第7章 方法 2-2：酵素と間接的に関連する医薬品および HFPs スクリーニング	92
第1節 方法	92
第1項 酵素と間接的に関連する医薬品および HFPs スクリーニング	92
第2節 結果	93
第1項 酵素の名称による数値化への影響	93
第2項 方法 2-2 の有用性	94
第3節 考察	99
第1項 方法 2-2 の有用性	99
第2項 方法 2-1 と方法 2-2 の比較	100
第3項 ゴールデンスタンダードとの比較	102
総括	106
謝辞	109
データの利用可能性に関する記述	110
参考文献	111

## 緒論

「健康食品(HFPs)」とは、健康の維持・増進に役立つことを期待して摂られている食品、製品、飲料全般を指す。これまで日常的に摂取されていなかった食品や成分が、研究や情報共有の進展によって「HFPs」として注目され、健康をサポートするために摂取されるようになった(Jagdale et al., 2021)。ただし、「HFPs」の中には、成分が濃縮されているものや医薬品の成分を含んでいるものも多く存在するため、これらの成分の摂取にあたっては正しい注意が必要である。例えば、ステビアと表示された食品にはステビアリーフエキスが高濃度で含まれているものがあり、その摂取量によっては高血圧に対して有効性が示唆されている(Meding, 1992; Hsieh et al., 2003)。一方で、ステビオの主成分であるステビオサイドは血圧を下げる作用があるため、ベラバミルなどのカルシウム拮抗剤(Hsieh et al., 2003; Meding, 1992; Melis et al., 1991)や血圧降下剤(Hsieh et al., 2003, Chan et al., 2000)などの医薬品と相互作用する恐れがあり、注意が必要であると言われている。このように、「HFPs」を摂取することで、一部の HFP 成分により体内で医薬品と相互作用することによって、医薬品の作用が変化する可能性がある(Asher et al., 2017; Roy et al., 2022)。このように医薬品と HFPs の組み合わせによっては、思わぬ健康被害が発生する可能性がある。

このような危険性を抱えている医薬品と HFP の相互作用についての情報を知るには、多くの課題がある。データベースや文献に記載されている相互作用もあるが、専門知識が必要で、数多の論文を読む必要があり、情報抽出が難しい(Ziani et al., 2022)。さらに、HFP 自体が多くの有効成分を含んでおり、その反応は非常に複雑である(Rein et al., 2013; Sorrenti et al., 2023)。また、HFP の成分を抽出する方法によって、HFP 成分の化学構造が変化することがあり、相互作用を正確に予測することをさらに困難にしている(Jiamjariyatam et al., 2023; Mounika et al., 2022)。

上述の問題で、最初に解決すべき問題は医薬品と HFP の相互作用情報を抽出することである。本研究では情報科学の手法を用いて、相互作用する可能性がある

医薬品と HFP のペアを抽出する新しいワークフローを確立することを目的としている。

情報が充実している医薬品についての研究から、一部の相互作用には医薬品の代謝される酵素が関与していることが示唆されている(Almazroo *et al.*, 2017). 一部の医薬品と代謝酵素の関係はすでによく研究されているが、HFPs と代謝酵素との関係はほとんど研究されていない。そこで、HFPs と代謝酵素の情報を収集することで、酵素をハブとして医薬品—酵素—HFPs のネットワークを解析し、この三者の相互作用から、相互作用する可能性がある医薬品と HFP のペアを抽出することが可能と考えた。この仮定に基づいて実験を行ったところ、取得した結果が既知ではないペアをかなり多く含んでいた。この問題を解決するために、異なるスクリーニング方法の組み合わせることでより正確にペアを抽出可能ではないかと考えた。ただし、それぞれのスクリーニングについて方法と評価方法を新たに確立する必要があるため、様々な検討を行った。

最初のスクリーニング (一次スクリーニング) では、医薬品、酵素および HFPs のリストから最も良かった共起検索の条件についてさらに検討を加えた。この結果得られた共起検索で取得した Drug-HFP interactions (DHIs) について、より詳細な分析結果を得るための二次スクリーニングでは、テキストマイニングを中心とした情報科学手法の適用を検討した。本研究で新しく確立した評価法を用いてこれらのスクリーニング結果について比較を実施した。その結果、相互作用する医薬品と HFPs のペアが抽出できた。以上により、医薬品と HFP の相互作用の情報を効率的に抽出する新しいワークフローを確立したと結論づけた。

相互作用する可能性がある医薬品と HFP のペアを抽出する新しいワークフローを確立したことにより、相互作用情報を効率的に抽出することが可能となり、医薬品と HFPs の相互作用に関する知識の拡充に寄与することが期待される。



## 第1章 医薬品、酵素およびHFPsのリスト化

医薬品とHFPの相互作用に関する知識の拡充には、医薬品とHFPの相互作用情報を抽出する方法を確立する必要がある。相互作用情報を抽出するには、はじめに医薬品、酵素およびHFPsのリストが必要である。医薬品および酵素の情報はKEGG (Kyoto Encyclopedia of Genes and Genomes) にも記載があるように充実している。一方でHFPsの情報を記載しているデータベースはほとんどない。リスト作成のために、医薬品と酵素はKEGGから情報抽出し、HFPs情報の探索および抽出を行なった。

### 第1節 方法

#### 第1項 KEGGを利用した医薬品と代謝酵素の相互作用リストの作成

KEGGの中から代謝酵素および医薬品の情報を得るために”Drug Metabolizing Enzymes and Transporters” (<https://www.genome.jp/kegg/drug/br08309.html>) を利用した。この中に記載されている医薬品および代謝酵素の情報を収集した。

#### 第2項 HFPsのデータセットの準備

健康食品の情報を得るために医薬基盤・健康・栄養研究所の素材情報データベース (<https://www.nibiohn.go.jp/eiken/info/hf2.html>) と National Institutes of Health (NIH)を利用した(Garcia-Cazarin *et al.*, 2014; Dwyer *et al.*, 2018)。前者のデータベースには「HFPs」に利用される素材や成分について、その概要、主成分の分析方法、有効性情報、安全性などの研究データが掲載されている。後者のデータベースの Dietary Supplement Fact Sheets にはビタミン、ミネラル、ハーブや植物、プロバイオティクスなどの栄養補助食品とその成分に関する情報が記載されている。この2種のデータベースからHFPs情報を取得した。

## 第 2 節 結果

KEGG の”Drug Metabolizing Enzymes and Transporters”から代謝酵素 59 種とその名称と略称, HSA 番号(または KO 番号), relation type と代謝に関連する医薬品 1270 種とその名称と D 番号, 関連酵素を得て, リスト化した(Table 1-1). 素材情報データベースと Dietary Supplement Fact Sheets から, HFPs 情報を両者のデータベースから取得した. その際に, HFP 以外のものを除外した. さらに両者のデータベースの重複を除き, 365 種取得した. 取得した HFPs365 種には HF1 などの独自の ID を振り, 区別した(Table 1-2).

Table 1-1 KEGG から取得した 59 種の酵素の略称とフルネーム一覧

Abbreviation	Full name
CYP3A4	cytochrome P450 family 3 subfamily A member 4
CYP2C19	cytochrome P450 family 2 subfamily C member 19
CYP2C8	cytochrome P450 family 2 subfamily C member 8
CYP3A5	cytochrome P450 family 3 subfamily A member 5
CYP3A7	cytochrome P450 family 3 subfamily A member 7
CYP27B1	cytochrome P450 family 27 subfamily B member 1
CYP24A1	cytochrome P450 family 24 subfamily A member 1
CYP2C9	cytochrome P450 family 2 subfamily C member 9
CYP2E1	cytochrome P450 family 2 subfamily E member 1
CYP1A2	cytochrome P450 family 1 subfamily A member 2
CYP2D6	cytochrome P450 family 2 subfamily D member 6
CYP2A6	cytochrome P450 family 2 subfamily A member 6
CYP2B6	cytochrome P450 family 2 subfamily B member 6

---

CYP1A1	cytochrome P450 family 1 subfamily A member 1
CYP2J2	cytochrome P450 family 2 subfamily J member 2
CYP4F2	cytochrome P450 family 4 subfamily F member 2
CYP2C18	cytochrome P450 family 2 subfamily C member 18
CYP4A11	cytochrome P450 family 4 subfamily A member 11
CYP4F3	cytochrome P450 family 4 subfamily F member 3
MAOA	monoamine oxidase A
MAOB	monoamine oxidase B
AOX1	aldehyde oxidase 1
DPYD	dihydropyrimidine dehydrogenase
FMO1	flavin containing dimethylaniline monooxygenase 1
FMO3	flavin containing dimethylaniline monooxygenase 3
CBR1	carbonyl reductase 1
CBR3	carbonyl reductase 3
FMO2	flavin containing dimethylaniline monooxygenase 2
FMO4	flavin containing dimethylaniline monooxygenase 4
FMO5	flavin containing dimethylaniline monooxygenase 5
XDH	xanthine dehydrogenase
UGT2B7	UDP glucuronosyltransferase family 2 member B7
UGT2B4	UDP glucuronosyltransferase family 2 member B4
UGT	glucuronosyltransferase
UGT1A4	UDP glucuronosyltransferase family 1 member A4
UGT2B15	UDP glucuronosyltransferase family 2 member B15
UGT1A1	UDP glucuronosyltransferase family 1 member A1
UGT1A9	UDP glucuronosyltransferase family 1 member A9

---

---

UGT1A3	UDP glucuronosyltransferase family 1 member A3
UGT1A7	UDP glucuronosyltransferase family 1 member A7
UGT1A8	UDP glucuronosyltransferase family 1 member A8
UGT2B17	UDP glucuronosyltransferase family 2 member B17
COMT	catechol-O-methyltransferase
NAT1	N-acetyltransferase 1
NAT2	N-acetyltransferase 2
SULT1A1	sulfotransferase family 1A member 1
SULT1A2	sulfotransferase family 1A member 2
SULT1A3	sulfotransferase family 1A member 3
SULT1E1	sulfotransferase family 1E member 1
GST	glutathione S-transferase
CES1	carboxylesterase 1
CES2	carboxylesterase 2
CES3	carboxylesterase 3
CES5A	carboxylesterase 5A
ACHE	acetylcholinesterase
ADA	adenosine deaminase
BCHE	butyrylcholinesterase
CTSA	cathepsin A
DDC	dopa decarboxylase

---

Table 1-2 収集した HFP リスト

HFP_id	Main name	Synonym name
HF1	4-hydroxyproline	4-hydroxyproline
HF2	5-Aminiolevulinic Acid	5-Aminiolevulinic Acid
HF2	6-Aminiolevulinic Acid	5-Amino-4-oxopentanoic acid
HF3	Docosahexaenoic acid	Docosahexaenoic acid
HF4	Eicosapentaenoic acid	Eicosapentaenoic acid
HF5	N-acetylglucosamine	N-acetylglucosamine
HF6	Eyebright	Euphrasia officinalis L.
HF6	Eyebright	Euphrasia rostkoviana F. Hayne
HF6	Eyebright	Euphrasia stricta J.P.Wolff ex J.F.Lehm
HF6	Eyebright	Eyebright
HF7	Butea Gum Tree	Butea Gum Tree
HF7	Butea Gum Tree	Butea Superba
HF7	Butea Gum Tree	Kwao Krua Dang
HF7	Butea Gum Tree	Red Kwao Krua
HF8	Turmeric	Curcuma domestica Val.
HF8	Turmeric	Curcuma longa
HF8	Turmeric	Curcuma longa L.
HF8	Turmeric	Turmeric
HF9	Hemerocallis fulva var. sempervirens	Hemerocallis fulva var. sempervirens
HF10	Astaxanthin	Astaxanthin
HF11	Aspartic acid	Aspartic acid
HF12	Acetyl-L-carnitine	Acetyl-L-carnitine

---

HF13	Amygdalin	Amygdalin
HF13		laetrile
HF14	Arachidonic acid	Arachidonic acid
HF15	Acacia gum	Acacia gum
HF15	Acacia gum	Acacia senegal (L.) willd
HF15	Acacia gum	Acacia senegal L. willd
HF15	Acacia gum	Acacia senegal willd
HF15	Acacia gum	Arabic gum
HF15	Acacia gum	Gum arabic Acacia
HF16	Alginic acid	Alginic acid
HF16	Alginic acid	alginate
HF17	Anserine	Anserine
HF18	Anthocyanin	Anthocyanin
HF19	Isoflavone	Isoflavone
HF20	Isomalto-oligosaccharide	Isomalto-oligosaccharide
HF21	Hu Zhang	Bambou Japonais
HF21	Hu Zhang	Banbou Mexican
HF21	Hu Zhang	Fallopia japonica
HF21	Hu Zhang	Fleece Flower
HF21	Hu Zhang	Hu Zhang
HF21	Hu Zhang	Polygonunm cuspidatum
HF21	Hu Zhang	Reynoutria japonica Houtt.
HF22	Ginkgo	Ginkgo
HF22	Ginkgo	Ginkgo biloba
HF22	Ginkgo	Ginkgo biloba L.

---

---

HF22	Ginkgo	Ginkgo biloba extract
HF22	Ginkgo	Ginkgo leaf extract
HF22	Ginkgo	Maidenhair tree
HF23	Inulin	beta (2-1) fructans
HF23	Inulin	Inulin
HF23	Inulin	Beta (2-1) fructans
HF24	Inositol	Hexahydroxycyclohexane
HF24	Inositol	Inositol
HF25	Imidazole dipeptide	Imidazole dipeptide
HF26	Phaseolus Vulgaris	Common bean
HF26	Phaseolus Vulgaris	Kidney Bean
HF26	Phaseolus Vulgaris	Phaseolus vulgaris L.
HF26	-	bayo
HF26	-	butter bean
HF27	Indian Long Pepper	Indian long pepper
HF27	Indian Long Pepper	Long pepper
HF27	Indian Long Pepper	Piper longum L.
HF27	Indian Long Pepper	jaborandi pepper
HF28	Fennel	Fennel
HF28	Fennel	Foeniculum vulgare Mill
HF29	Japanese Apricot	Armeniaca mume Sieb.
HF29	Japanese Apricot	Japanese apricot
HF29	Japanese Apricot	Prunus mume Sieb.et Zucc.
HF30	Eleuthero	Acanthopanax senticosus
HF30	Eleuthero	Eleuthero

---

---

HF30	Eleuthero	Eleutherococcus senticosus
HF30	Eleuthero	Eleutherococcus senticosus Maxim.
HF30	-	Rupr.et.Maxim.
HF30	-	Siberian ginseng
HF31	Oats	Avena sativa L.
HF31	Oats	Oats
HF32	Barley	Barley
HF32	Barley	Hordeum distichum L.
HF32	Barley	Hordeum vulgare L.
HF32	-	Malt
HF33	Olive	Olea europaea L.
HF33	-	Olea sativa Hoffmanns. Et Link
HF33	Olive	Olive
HF33	Olive	Olive leaf extract
HF34	Ornithine	Ornithine
HF35	Cocoa	Cacao
HF35	Cocoa	Theobroma cacao
HF36	-	Crassostrea gigas THUNB.
HF36	Oyster Extract	Ostrea gigas THUNB.
HF36	-	Ostreidae
HF36	Oyster Extract	Oyster
HF37	Black currant	Black currant
HF37	Black currant	Cassis
HF37	Black currant	Ribes nigrum L.
HF38	Green Tea	Catechin

---



---

HF38	Green Tea	Catechinic acid
HF39	Carrageenan	Carrageenan
HF40	Galacto-oligosaccharides	Galacto-oligosaccharide
HF41	Carnosine	Carnosine
HF42	Beta-Carotene	Carotene
HF43	Helianthus Tuberosus	Helianthus tuberosus L.
HF43	Artichoke	Jerusalem artichoke
HF44	Aloe	Aloe arborescens Mill.
HF44	Aloe	Kidachi Aloe
HF45	-	Chitin
HF46	Guava	Common guava
HF46	Guava	Guava
HF46	Guava	Psidium guajava L.
HF47	Citric acid	Citric acid
HF48	-	Chinese box thorn
HF48	Goji	Chinese wolfberry
HF48	Goji	Duke of Argyll's tea-tree
HF48	Goji	Goji
HF48	Goji	Lycium barbarum L.
HF48	Goji	Lycium barbarum Linne
HF48	Goji	Lycium chinense Mill.
HF48	Goji	Wolfberry
HF48	Goji	matrimony vine.
HF49	Kudzu	Japanese arrowroot
HF49	Kudzu	Kudzu vine

---

---

HF49	-	Ohwi
HF49	Kudzu	<i>Pueraria lobata</i>
HF49	Kudzu	<i>Pueraria montana</i> (Lour.) Merr.
HF49	Kudzu	<i>Pueraria montana</i> Lour. Merr.
HF49	Kudzu	<i>Pueraria montana</i> Merr.
HF50	Cumin	Cumin
HF50	Cumin	<i>Cuminum cyminum</i> L.
HF51	Glycine	Glycine
HF51	-	Glycocoll
HF52	Curcumin	Curcumin
HF53	Glutathione	5-L-Glutamyl-L-cysteinylglycine
HF53	Glutathione	Glutathione
HF54	Glutamic acid	Glutamic acid
HF55	-	Croctin
HF56	Blue-green algae	<i>Anabaena</i>
HF56	Blue-green algae	<i>Anabaenopsis</i>
HF56	Blue-green algae	Blue-green algae
HF56	Blue-green algae	<i>Chlamydomonas</i>
HF56	Blue-green algae	<i>Chlorella</i>
HF56	Blue-green algae	<i>Chlorella pyrenoidosa</i>
HF56	Blue-green algae	<i>Chlorella vulgaris</i>
HF56	Blue-green algae	<i>Microcystis</i>
HF56	Blue-green algae	<i>Scenedesmus</i>
HF57	Chlorophyll	Chlorophyll
HF58	White Mulberry	<i>Morus alba</i>

---

---

HF58	White Mulberry	Mulberry
HF59	-	Silicide
HF59	Silicon	Silicon
HF60	Kale	Brassica oleracea L. var. acephala DC.
HF60	Kale	Kale
HF61	Germanium	Germanium
HF62	Coffee	Coffee
HF63	Coconut Oil	Coconut Oil
HF63	Coconut Oil	Cocos nucifera L.
HF64	Burdock	Arctium lappa L.
HF64	Burdock	Burdock
HF64	Burdock	Lappa major Gaertn.
HF65	Sesame	Sesame
HF65	Sesame	Sesame seeds
HF65	Sesame	Sesamum indicum L
HF66	Collagen	Collagen
HF67	Chondroitin sulfate	Chondroitin 4- and 6-sulfate
HF67	Chondroitin sulfate	Chondroitin 4-sulfate
HF67	Chondroitin sulfate	Chondroitin sulfate
HF68	-	Sardine peptide
HF69	Saffron	Crocus sativus L.
HF69	Saffron	Saffron
HF69	Saffron	Saffron crocus
HF70	Salacia	Chundan
HF70	Salacia	Kathala Hibutu Tea

---

---

HF70	Salacia	Ponkoranti
HF70	Salacia	SO.
HF70	Salacia	Salacia chinensis
HF70	Salacia	Salacia oblonga
HF70	Salacia	Salacia reticulata
HF71	Sunflower Oil	Helianthus annuus L.
HF71	Sunflower Oil	Sunflower
HF72	Cysteine	Cysteine
HF73	Indian Long Pepper	Javanese long pepper
HF73	Indian Long Pepper	Jawa long pepper
HF73	Indian Long Pepper	Piper retrofractum Vahl.
HF74	-	Silk fibroin
HF74	-	Silk protein
HF74	-	sericin
HF75	Squalene	Squalene
HF76	Stevia	Paraguayan sweet herb
HF76	Stevia	Stevia
HF76	Stevia	Stevia rebaudiana (Bertoni) Hemsl.
HF76	Stevia	Stevia rebaudiana Bertoni Hemsl.
HF76	Stevia	Stevia rebaudiana Hemsl.
HF77	Blue-Green Algae	Arthrospira maxima
HF77	Blue-Green Algae	Arthrospira platensis
HF77	Blue-Green Algae	Spirulina
HF77	Blue-Green Algae	Spirulina maxima
HF77	Blue-Green Algae	Spirulina platensis

---

---

HF78	Sulforaphane	1-isothiocyanate-4-methylsulfonyl butane
HF78	Sulforaphane	Sulforaphane
HF79	Zeaxanthin	Zeaxanthin
HF80	Sweet clover	Melilotus
HF80	Sweet clover	Melilotus officinalis
HF80	Sweet clover	Melilotus suaveolens Ledeb.
HF80	Sweet clover	Sweet clover
HF81	Valerian	Valerian
HF81	Valerian	Valeriana officinalis L.
HF81	Valerian	Valeriana officinalis var. latifolia
HF81	Valerian	Valeriana wallichii
HF82	Hawthorn	Crataegus laevigata DC.
HF82	Hawthorn	Crataegus monogyna Jacq.
HF82	Hawthorn	English hawthorn
HF83	-	Ceramide
HF83	-	N-acylsphingosine
HF84	Senna	Alexandria Senna
HF84	Senna	Cassia acutifolia Delile.
HF84	Senna	Cassia angustifolia Vahl.
HF84	Senna	Senna
HF84	Senna	Tinnevelly Senna
HF85	Soy Isoflavone	Soy Isoflavone
HF86	Taurine	Taurine
HF87	Onion	Allium cepa L
HF87	Onion	Onion

---

---

HF88	-	Indian data
HF88	Tamarind	Tamarind
HF88	Tamarind	Tamarindus indica L.
HF89	Chia	Chia
HF89	Chia	Chia seed
HF89	Chia	Salvia hispanica.
HF90	Tyrosine	Tyrosine
HF91	FO-TI	Chinese Cornbind
HF91	FO-TI	Chinese Knotweed
HF91	FO-TI	FO-TI
HF91	-	Fallopia multiflora
HF91	FO-TI	FloweryKnotweed
HF91	-	Pleuropterus multiflorus Thunb.
HF91	FO-TI	Polygonum multiflorum
HF91	FO-TI	Polygonum multiflorum Turcz.
HF92	L-Theanine	L-Theanine
HF92	L-Theanine	$\gamma$ -Glutamylethylamide
HF93	-	Dextrin
HF94	Sea Buckthorn	Argasse
HF94	-	Argousier
HF94	-	Chharma
HF94	-	Dhar-Bu
HF94	Sea Buckthorn	Hippophae rhamnoides
HF94	Sea Buckthorn	Sea Buckthorn
HF95	-	Rubus suavissimus S. Lee

---

---

HF96	Capsicum	Capsicum
HF96	Capsicum	Capsicum annuum L.
HF96	Capsicum	Chili Pepper
HF96	Capsicum	Green Pepper
HF96	Capsicum	Paprika
HF96	Capsicum	Red Pepper
HF97	Cordyceps	Cordyceps
HF97	Cordyceps	Cordyceps sinensis
HF97	-	Plant worms
HF97	-	Vegetative wasps
HF98	-	Chinese gutta percha
HF98	Eucommia	Eucommia
HF98	-	Eucommia ulmoides
HF98	-	Hardy rubber tree
HF99	Tomato	Extrait de Tomate
HF99	Tomato	Love Apple
HF99	Tomato	Lycopersicon esculentum Mill.
HF99	Tomato	Tamatar
HF99	Tomato	Tomate
HF99	Tomato	Tomato
HF100	Dolomite	Dolomite
HF101	Nattokinase	Bacillus subtilis
HF101	Nattokinase	Fermented soybeans
HF101	-	Glycine max (L.) MERR.
HF101	-	Glycine max L. Merr.

---

---

HF101	-	Glycine max Merr.
HF101	-	Hay bacillus
HF101	-	natto
HF102	-	Narirutin
HF103	Naringin	Naringin
HF104	Nicotinamide mononucleotide	NMN
HF104	Nicotinamide mononucleotide	Nicotinamide mononucleotide
HF104	Nicotinamide mononucleotide	$\beta$ -nicotinamide mononucleotide
HF105	Nicotinamide Riboside	Nicotinamide Riboside
HF106	-	Ananas comosus L. Merr.
HF106	-	Ananas comosus L. Voss.
HF106	-	Ananas comosus Merr.
HF106	-	Ananas comosus Voss.
HF106	-	Ananas sativus Schult. f.
HF106	-	Bromelia ananas L.
HF106	-	Bromelia comosus L.
HF106	Pineapple	Pineapple
HF107	Parsley	Parsley
HF107	Parsley	Petroselinum crispum (Mill.) Nyman ex A.W.Hill
HF107	Parsley	Petroselinum crispum Mill. Nyman ex A.W.Hill
HF107	Parsley	Petroselinum crispum NYMAN var. angustifolium
		HARA

---



---

HF107	Parsley	Petroselinum crispum Nyman ex A.W.Hill
HF108	Honey	Honey
HF109	Job's Tears	Coix lacryma-jobi L. var. mayuen (Roman.) Stapf
HF109	Job's Tears	Coix lacryma-jobi L. var. mayuen Roman. Stapf
HF109	Job's Tears	Coix lacryma-jobi L. var. mayuen Stapf
HF109	Job's Tears	Coix seed
HF109	Job's Tears	Job's Tears
HF110	-	Candle bush
HF110	-	Cassia alata
HF110	-	Senna alata L.
HF111	-	Aromatic
HF111	-	Curcuma aromatica Salisb.
HF111	-	wild turmeric
HF111	-	yellow zadoary
HF112	Hyaluronic acid	Hyaluronan
HF112	Hyaluronic acid	Hyaluronic acid
HF113	Pinitol	(1S,2R,4R,5R)-6-methoxycyclohexane-1,2,3,4,5-pentol
HF113	Pinitol	(1S,2S,4S,5R)-6-methoxycyclohexane-1,2,3,4,5-pentol
HF113	Pinitol	D-pinitol
HF113	Pinitol	L-pinitol
HF113	Pinitol	Pinitol
HF114	Bifidobacteria	Bifidobacteria
HF114	Bifidobacterium bifidum	Bifidobacterium bifidum
HF114	Bifidobacterium breve	Bifidobacterium breve
HF114	Bifidobacterium longum	Bifidobacterium longum

---

---

HF115	Black pepper is piperine	Piperine
HF116	Periwinkle	Common periwinkle
HF116	Periwinkle	Lesser Periwinkle
HF116	Periwinkle	Periwinkle
HF116	Periwinkle	Small periwinkle
HF116	Periwinkle	Vinca minor.
HF117	Loquat	Eriobotrya japonica (THUNB.) LINDL.
HF117	Loquat	Eriobotrya japonica LINDL.
HF117	Loquat	Eriobotrya japonica THUNB. LINDL.
HF117	Loquat	Japanese medlars
HF117	Loquat	Loquat
HF118	Phenylalanine	Phenylalanine
HF119	-	Pueraria candollei var. mirifica
HF119	-	Pueraria mirifica
HF119	-	White Kwao keur
HF120	Fucoidan	Fucoidan
HF121	Fructo-oligosaccharides	FOS
HF121	Fructo-oligosaccharides	Fructo-oligosaccharides
HF121	Fructo-oligosaccharides	oligofructose
HF122	Placenta	Placenta
HF123	Kaempferia parviflora	Black Ginger
HF123	Kaempferia parviflora	Kaempferia parviflora
HF123	Kaempferia parviflora	Kura-Chai-Dum
HF125	Broccoli	Brassica oleracea L.var.italica Plencke
HF125	Broccoli	Broccoli

---

---

HF126	-	Proteoglycan
HF127	Proline	Proline
HF128	Blond psyllium	Blond psyllium
HF128	-	Plantago decumbens
HF128	Blond psyllium	Plantago fastigiata
HF128	Blond psyllium	Plantago insularis
HF128	-	Plantago isphaghula
HF128	Blond psyllium	Plantago ovata Forsk
HF128	Blond psyllium	Plantago ovata Forssk.
HF128	-	Psyllium
HF129	Hesperidin	Hesperidin
HF130	Red yeast rice	Monascus albidus K.Sato
HF130	Red yeast rice	Monascus anka K.Sato
HF130	Red yeast rice	Monascus purpureus Went.
HF130	Red yeast rice	Red yeast rice
HF130	-	red leaven
HF130	Red yeast rice	red rice
HF131	Phosphatidylserine	Phosphatidylserine
HF132	-	Leptospermum scoparium
HF132	-	Manuka
HF133	-	Euglena
HF133	-	Euglena gracilis
HF134	Methionine	Methionine
HF135	Methylsulfonylmethane	Dimethylsulfone
HF135	Methylsulfonylmethane	MSM

---

---

HF135	Methylsulfonylmethane	Methylsulfonylmethane
HF136	Lemon balm	Bee Balm
HF136	Lemon balm	Lemon balm
HF136	Lemon balm	Melissa
HF136	Lemon balm	Melissa officinalis L.
HF137	Moringa	Behen
HF137	Moringa	Ben nut tree
HF137	Moringa	Drumstick tree
HF137	Moringa	Horseradish tree
HF137	-	Malungai
HF137	Moringa	Moringa
HF137	Moringa	Moringa oleifera Lam.
HF137	Moringa	Moringa pterygosperma
HF137	-	Sanjanaa
HF138	German Ipecac	Asclepias vincetoxicum
HF138	German Ipecac	Cynanchum vincetoxicum
HF138	German Ipecac	Dompte-Venin
HF138	German Ipecac	German Ipecac
HF138	German Ipecac	Swallo Wort
HF138	German Ipecac	Vincetoxicum hirundinaria
HF138	German Ipecac	Vincetoxicum officinale
HF139	Momordica grosvenori Swingle	Momordica grosvenori Swingle
HF139	Momordica grosvenori Swingle	Siraitia grosvenori Swingle

---

---

HF140	Lactoferrin	Lactoferrin
HF141	-	Apocynum venetum L.
HF141	-	Luobuma
HF142	Lysine	Lysine
HF143	Conjugated Linoleic Acid	Linoleic acid
HF144	Rooibos	Aspalathus linearis
HF144	Rooibos	Kaffree tea
HF144	Rooibos	Red bush tea
HF144	Rooibos	Rooibos
HF145	Rutin	Rutin
HF146	Lecithin	Lecithin
HF146	Phosphatidylcholine	Phosphatidylcholine
HF147	Rosemary	Rosemary
HF147	Rosemary	Rosmarinus officinalis L.
HF148	Royal jelly	Royal jelly
HF149	-	Basidiomycetes-X
HF149	-	Echigoshirayukidake
HF150	Nucleic acid	Nucleic acid
HF151	-	Maltitol
HF152	Bear's Garlic	Ail des Bois
HF152	Bear's Garlic	Ail des Ours
HF152	Bear's Garlic	Allium ursinum L.
HF152	Bear's Garlic	Barlauch
HF152	Bear's Garlic	Bear's Garlic
HF152	Bear's Garlic	Ramson

---

---

HF153	Kombucha	Kombucha
HF153	Kombucha	Kombucha tea
HF154	black vinegar	Brewed rice vinegar
HF154	black vinegar	Husked rice vinegar
HF154	black vinegar	Unpolished rice vinegar
HF154	black vinegar	black vinegar
HF155	-	FLAVANGENOL
HF155	Maritime Pine	French marine pine bark extract
HF155	Maritime Pine	Pinus pinaster
HF155	Maritime Pine	Pycnogenol
HF156	Phytosterol	Phytosterol
HF157	Dietary fiber	Dietary fiber
HF158	-	Acetic acid
HF158	-	Vinegar
HF158	-	balsamico vinegar
HF158	-	coconut vinegar
HF158	-	cone vinegar
HF158	-	honey vinegar
HF158	-	mort vinegar
HF158	-	sherry vinegar
HF158	-	white vinegar
HF158	-	wine vinegar
HF159	-	Hydrogen water
HF160	-	Japanese sake yeast
HF160	-	Saccharomyces cerevisiae

---

---

HF161	-	Aojiru
HF162	-	Soya-oligosaccharide
HF163	Medium chain triglycerides	Medium chain triglycerides
HF164	-	Indigestible Dextrin
HF164	maltodextrin	Resistant maltodextrin
HF165	Rice bran	Rice bran
HF166	-	3,4,5-Trihydroxybenzoic acid
HF166	-	gallic acid
HF167	-	Choline glycerophosphate
HF167	-	L-alphaGlycerolphosphorylcholine
HF168	Alpha-Linolenic acid	Alpha-Linolenic acid
HF169	Alpha-lipoic acid	Alpha-lipoic acid
HF169	Alpha-lipoic acid	Thioctic acid
HF170	Beta-Cryptoxanthin	(3R)- $\beta$ , $\beta$ -caroten-3-ol
HF170	Beta-Cryptoxanthin	Beta-cryptoxanthin
HF170	-	alltrans-cryptoxanthin
HF170	Beta-Cryptoxanthin	caricaxanthin
HF170	Beta-Cryptoxanthin	cryptoxanthol
HF170	-	hydroxy- $\beta$ -carotene
HF170	-	$\beta$ -caroten-3-ol
HF171	Gamma-aminobutyric acid	GABA
HF171	Gamma-aminobutyric acid	Gamma-aminobutyric acid
HF172	-	GSAC
HF172	-	$\gamma$ -glutamyl-S-allyl-cysteine
HF173	Branched chain amino acid	BCAAs

---

---

HF173	Branched chain amino acid	Branched chain amino acid
HF173	Branched chain amino acid	branched-chain amino acids
HF174	L-tryptophan	2-amino-3- (3-indolyl) propionic acid
HF174	L-tryptophan	L-tryptophan
HF174	L-tryptophan	Tryptophan
HF175	Zinc	Zinc
HF176	Ashwagandha	Ajagandha
HF176	Ashwagandha	Amangura
HF176	Ashwagandha	Ashwagandha
HF176	Ashwagandha	Withania
HF176	Ashwagandha	Withania somnifera Dunal
HF177	Irvingia gabonensis	African mango
HF177	Irvingia gabonensis	Agbono
HF177	-	Aubry-Lecomte ex O'Rorke
HF177	Irvingia gabonensis	Bread Tree
HF177	Irvingia gabonensis	Dica Nut
HF177	Irvingia gabonensis	Irvingia barteri
HF177	Irvingia gabonensis	Irvingia gabonensis
HF177	Irvingia gabonensis	Irvingia gabonensis Baill.
HF177	Irvingia gabonensis	Mangifera gabonensis
HF178	Flaxseed	Flax
HF178	Flaxseed	Flax seed
HF178	Flaxseed	Flaxseed
HF178	Flaxseed oil	Flaxseed oil
HF178	Flaxseed	Linseed

---



---

HF178	Flaxseed oil	Linseed oil
HF178	Flaxseed	Linum
HF178	Flaxseed	Linum usitatissimum L.
HF179	Alanine	Alanine
HF179	Beta-alanine	Beta-alanine
HF180	L-arginine	Arginine
HF180	L-arginine	L-arginine
HF181	Alfalfa	Alfalfa
HF181	Alfalfa	Lucerne
HF181	Alfalfa	Medicago sativa L.
HF182	Aloe	Aloe
HF182	Aloe	Aloe africana Mill.
HF182	Aloe	Aloe barbadensis Mill.
HF182	Aloe	Aloe ferox Mill.
HF182	Aloe	Aloe vera
HF182	Aloe	Aloe vera (L.) Burm.f.
HF182	Aloe	Aloe vera Burm.f.
HF182	Aloe	Aloe vera L. Burm.f.
HF182	Aloe	Cape aloe
HF182	Aloe	Curacao Aloes
HF183	Isoleucine	Isoleucine
HF185	Echinacea	Echinacea
HF185	Echinacea	Echinacea purpurea (L.) Moench
HF185	Echinacea	Echinacea purpurea L. Moench
HF185	Echinacea	Echinacea purpurea Moench

---

---

HF185	Echinacea	angustifolia
HF185	Echinacea	pallida.
HF186	Blessed thistle	Blessed thistle
HF186	Milk thistle	Milk thistle
HF186	Milk thistle	Silybum marianum (L.) Gaertn.
HF186	Milk thistle	Silybum marianum Gaertn.
HF186	Milk thistle	Silybum marianum L. Gaertn.
HF187	-	N-3 unsaturated fatty acid
HF187	-	Omega-3 fatty acids
HF189	Bitter orange	Bitter orange
HF189	Bitter orange	Citrus aurantium L.
HF189	Bitter orange	Seville orange
HF189	Bitter orange	Sour orange
HF190	Black tea	Black tea
HF190	Black tea	Camelliasinensis (L.) Kuntze
HF190	Black tea	Camelliasinensis Kuntze
HF190	Black tea	Camelliasinensis L. Kuntze
HF190	Black tea	Chinese tea
HF190	Black tea	Tea
HF191	Maritime pine	Maritime pine
HF191	Seaweed	Seaweed
HF192		Caffeine
HF192	-	methyltheobromine
HF193	Roman Chamomile	Chamaemelum nobile
HF193	Roman Chamomile	Chamomile

---

---

HF193	German Chamomile	German chamomile
HF193	German Chamomile	Matricaria chamomilla L.
HF193	German Chamomile	Matricaria recutita
HF193	Roman Chamomile	Roman Chamomile
HF194	Guarana	Guarana
HF194	Guarana	Paullinia cupana Kunth
HF195	Potassium	Potassium
HF196	Calcium	Calcium
HF197	Garcinia	Brindal Berry
HF197	Gamboge	Gambooge
HF197	Garcinia	Garcinia
HF197	Garcinia	Garcinia cambogia
HF197	Garcinia	Malabar Tamarind
HF198	L-Carnitine	Carnitine
HF199	Licorice	Glycyrrhiza glabra
HF199	Licorice	Glycyrrhiza uralensis Fisch. Et DC.
HF199	Licorice	Licorice
HF199	Licorice	Licorice root
HF199	Licorice	Liquorice
HF200	Cannabidiol	2-[(1R,6R)-6-isopropenyl-3-methylcyclohex-2-en-1-yl]-5-pentylbenzene-1,3-diol
HF200	Cannabidiol	CBD
HF200	Cannabidiol	Cannabidiol
HF201	Chitosan	Chitosan
HF202	Agaricus Mushroom	Agaricus

---

---

HF202	Agaricus Mushroom	Agaricus blazei Murill
HF202	Reishi Mushroom	Ganoderma lucidum. Karst
HF202	Shiitake Mushroom	Lentinus edodes (Berk.) Sing.
HF202	Shiitake Mushroom	Lentinus edodes Berk. Sing.
HF202	Shiitake Mushroom	Lentinus edodes Sing.
HF202	Agaricus Mushroom	Mushrooms
HF202	Reishi Mushroom	Reishi
HF202	Shiitake Mushroom	Shiitake
HF203	Gymnema	Gymnema
HF203	Gymnema	Gymnemic acid
HF204	Conjugated linoleic acid	CLA
HF204	Conjugated linoleic acid	Conjugated linoleic acid
HF205	Guar gum	Guar gum
HF206	Glucosamine	Glucosamine
HF207	Glucomannan	Glucomannan
HF208	Glutamine	Glutamine
HF208	Glutamine	L-(+)-Aminoglutaramic acid
HF209	Creatine	Creatine
HF210	-	Citrus decumana
HF210	-	Citrus maxima
HF210	Grapefruit	Citrus paradisi
HF210	Grapefruit	Grapefruit
HF211	Quercetin	Quercetin
HF212	Coenzyme Q10	Coenzyme Q10
HF212	Coenzyme Q10	UQ

---

---

HF212	-	Ubiquinone
HF213	Choline	Choline
HF214	Coleus	Coleus
HF214	Coleus	Coleus Forskohlii
HF214	Coleus	Coleus barbatus
HF214	Coleus	forskolin
HF214	Coleus	Plectranthus barbatus
HF215	Fenugreek	Fenugreek
HF215	Fenugreek	Trigonella foenum-graecum L.
HF216	Fish oil	Fish oil
HF217	Pomegranate	Pomegranate
HF217	Pomegranate	Punica granatum L.
HF218	Citrulline	(2S)-2-amino-5- (carbamoylamino)pentanoic acid
HF218	Citrulline	Citrulline
HF219	Cassia Cinnamon	Cinnamomum aromaticum
HF219	Cassia Cinnamon	Cinnamomum cassia
HF219	-	Cinnamomum sieboldii
HF219	Ceylon Cinnamon	Cinnamomum verum
HF219	Ceylon Cinnamon	Cinnamomum zeylanicum
HF219		Cinnamon
HF220	Ginger	Ginger
HF220	Ginger	Zingiber officinale (Willd.) Roscoe
HF220	Ginger	Zingiber officinale Roscoe
HF220	Ginger	Zingiber officinale Willd. Roscoe
HF221	St. John's Wort	Hypericum perforatum L.

---

---

HF221	St. John's Wort	St. John's Wort
HF222	Selenium	Selenium
HF223	Soy	Soy
HF223	Soy	Soybean
HF224	Dehydroepiandrosterone	Dehydroepiandrosterone
HF224	Dehydroepiandrosterone	Prasterone
HF225	Copper	Copper
HF226	Niacin	Niacin
HF226	Niacin	Nicotinamide
HF226	Niacin	Nicotinic acid
HF226	Niacin	Vitamin B3
HF227	Shark cartilage	Cartilage
HF227	Shark cartilage	Shark cartilage
HF228	Lactobacillus acidophilus	Lactobacillus
HF228	Lactobacillus acidophilus	Lactobacillus acidophilus
HF228	Lactobacillus brevis	Lactobacillus brevis
HF228	Lactobacillus delbrueckii	Lactobacillus delbrueckii
HF229	Garlic	Ail
HF229	Garlic	Ajo
HF229	Garlic	Allium sativum
HF229	Garlic	Garlic
HF230	-	Sabal palm
HF230	Saw palmetto	Saw palmetto
HF230	Saw palmetto	Serenoa repens (W.Bartram) Small.
HF230	Saw palmetto	Serenoa repens Small.

---

---

HF230	Saw palmetto	<i>Serenoa repens</i> W.Bartram Small.
HF230	Saw palmetto	<i>Serenoa serrulata</i> (Michx.) G.Nichols
HF230	Saw palmetto	<i>Serenoa serrulata</i> (Michx.) Nutt. Ex Schult. Et Schult.
HF230	Saw palmetto	<i>Serenoa serrulata</i> G.Nichols
HF230	Saw palmetto	<i>Serenoa serrulata</i> Michx. G.Nichols
HF230	Saw palmetto	<i>Serenoa serrulata</i> Michx. Nutt. Ex Schult. Et Schult.
HF230	Saw palmetto	<i>Serenoa serrulata</i> Nutt. Ex Schult. Et Schult.
HF231	Ba Ji Tian	Morinda
HF231	Ba Ji Tian	<i>Morinda citrifolia</i> L.
HF231	Ba Ji Tian	Noni
HF232	Hibiscus sabdariffa	Hibiscus
HF232	Hibiscus sabdariffa	<i>Hibiscus sabdariffa</i> L.
HF232	Hibiscus sabdariffa	roselle
HF233	Bee pollen	Bee pollen
HF233	Bee pollen	Buckwheat Pollen
HF233	Bee pollen	Honeybee Pollen
HF234	Passion Flower	<i>Passiflora incarnata</i>
HF234	Passion Flower	Passionflower
HF235	Tribulus	Al-gutub
HF235	Tribulus	Cat's-Head
HF235	Tribulus	Gokhuru
HF235	Tribulus	Puncture vine
HF235	Tribulus	Tribulus
HF235	Tribulus	<i>Tribulus terrestris</i> L.
HF236	L-Valine	Valine

---

---

HF237	Pantothenic acid	Pantothenic acid
HF237	Pantothenic acid	Vitamin B5
HF239	VITAMIN A	Retinol
HF239	VITAMIN A	VITAMIN A
HF240	Thiamine	Thiamine
HF240	Thiamine	Vitamin B1
HF241	Vitamin B12	Cyanocobalamin
HF241	Vitamin B12	Vitamin B12
HF242	Riboflavin	Riboflavin
HF242	Riboflavin	Vitamin B2
HF243	Vitamin B6	Pyridoxine
HF243	Vitamin B6	Vitamin B6
HF244	Vitamin C	L-ascorbic acid
HF244	Vitamin C	Vitamin C
HF245	Vitamin D	Vitamin D
HF245	Vitamin D	cholecalciferol
HF245	Vitamin D	ergocalciferol
HF245	Vitamin D	vitamin D2
HF245	Vitamin D	vitamin D3
HF246	Vitamin E	Tocopherol
HF246	Vitamin E	Vitamin E
HF247	Vitamin K	Menadiol
HF247	Vitamin K	Menadione
HF247	Vitamin K	Menaquinone
HF247	Vitamin K	Phylloquinone

---



---

HF247	Vitamin K	Vitamin K
HF247	Vitamin K	vitamin K1
HF247	Vitamin K	vitamin K2
HF247	Vitamin K	vitamin K3
HF247	Vitamin K	vitamin K4
HF248	Bilberry	Bilberry
HF248	Bilberry	Huckleberry
HF248	Bilberry	<i>Vaccinium myrtillus</i> L.
HF248	Bilberry	Whortleberry
HF249	Grape	Grape
HF249	Grape	<i>Vitis</i> L.
HF250	Black cohosh	<i>Actaea racemosa</i> L.
HF250	Black cohosh	Black cohosh
HF250	Black cohosh	Black snakeroot
HF250	Black cohosh	<i>Cimicifuga racemosa</i> L.
HF250	Black cohosh	<i>Cimicifuga racemosa</i> Nutt.
HF251	Blueberry	Blueberry
HF251	Blueberry	Highbush blueberry
HF251	Blueberry	<i>Vaccinium angustifolium</i>
HF251	Blueberry	<i>Vaccinium ashei</i>
HF251	Blueberry	<i>Vaccinium corymbosum</i> L.
HF251	Blueberry	<i>Vaccinium pallidum</i>
HF252	Probiotic	Probiotic
HF252	Probiotic	Probiotics
HF252	Probiotique	Probiotique

---

---

HF253	Propolis	Propolis
HF254	-	Betaine
HF255	Whey protein	Whey
HF255	Whey protein	Whey protein
HF256	Hop	European Hops
HF256	Hop	Hop
HF256	Hop	Hop Strobile
HF256	Hop	Hopfenzapfen
HF256	Hop	Hops
HF256	Hop	Houblon
HF256	Hop	Humulus lupulus L
HF256	Hop	common hop
HF257	Maca	Lepidium meyenii
HF257	Maca	Maca
HF257	Maca	Peruvian ginseng
HF258	Magnesium	Magnesium
HF259	Yerba mate	Ilex paraguayensis St. Hill.
HF259	Yerba mate	Mate
HF259	Yerba mate	Yerba mate
HF260	Melatonin	Melatonin
HF261	Iodine	Iodine
HF262	Lycopene	Lycopene
HF263	Lutein	Lutein
HF264	Resveratrol	3,4',5-stilbenetriol
HF264	Resveratrol	3,4',5-trihydroxystilbene

---

---

HF264	Resveratrol	3,5,4'- trihydroxystilbene
HF264	Resveratrol	Resveratrol
HF265	Leucine	Leucine
HF266	Rose hip	Dog rose
HF266	Rose hip	Rosa canina
HF266	Rose hip	Rosa laevigata Michx.
HF266	Rose hip	Rosa species
HF266	Rose hip	Rose hip
HF266	-	Wild Dog Rose
HF267	Panax Ginseng	Asian Ginseng
HF267	Panax Ginseng	Chinese ginseng
HF267	Panax Ginseng	Ginseng root
HF267	Panax Ginseng	Korean ginseng
HF267	Panax Ginseng	Oriental ginseng
HF267	Panax Ginseng	Panax ginseng C.A.Mey.
HF268	folic acid	Folate
HF268	folic acid	Vitamin B9
HF268	folic acid	folic acid
HF269	Iron	Iron
HF270	Sodium bicarbonate	Sodium bicarbonate
HF271	beta-hydroxy-beta-methylbutyrate	beta-hydroxy-beta-methylbutyrate
HF271	beta-hydroxy-beta-methylbutyrate	HMB
HF273	N-acetylcysteine	N-acetylcysteine

---

---

HF273	N-acetylcysteine	NAC
HF274	-	Ganoderma lucidum, Karst
HF274	-	Georgi
HF274	-	PC-SPES
HF274	Rabdosia rubescens	Rabdosia rubescens, Hara
HF274	Scutellaria baicalensis	Scutellaria baicalensis
HF275	SAMe	S-adenosyl-L-methionine
HF275	SAMe	SAMe
HF277	-	Arctium lappa
HF277	-	Essiac
HF277	-	Flor-Essence
HF277	Sorrel	Rumex acetosella
HF277	Slippery elm	Slippery elm
HF277	Slippery elm	Ulmus fulva
HF277	Slippery elm	Ulmus rubra
HF278	Hemp	Cannabis sativa L.
HF278	Hemp	Hemp
HF279	Acai	Acai
HF279	Acai	Euterpe oleracea Mart.
HF280	Acai	Acai Berry
HF280	Acai	Euterpe oleracea
HF282	Andrographis	Andrographis
HF282	Andrographis	Andrographis paniculata
HF283	Nitric oxide	Nitric oxide
HF285	Ephedra	Ephedra

---

---

HF285	Ephedra	Ephedra equisetina Bunge
HF285	Ephedra	Ephedra intermedia Schrenk
HF285	Ephedra	Ephedra sinica Stapf
HF287	Elderberry	Elderberry
HF287	Elderberry	Sambucus nigra L.
HF289	Omega-6 fatty acids	Omega-6 fatty acids
HF290	-	Omega-7 fatty acids
HF291	Oregano	Oregano
HF292	Cascara sagrada	Cascara sagrada
HF292	Cascara sagrada	Rhamnus purshiana
HF293	Activated charcoal	Activated charcoal
HF294	Kava	Kava
HF294	Kava	Piper methysticum
HF295	-	Capsaicinoids
HF296	Capsaicin	Capsaicin
HF297	Calendula	Calendula
HF297	Calendula	Calendula officinalis
HF298	Cranberry	Cranberry
HF298	Cranberry	Vaccinium macrocarpon
HF299	Clove	Clove
HF299	Clove	Syzygium aromaticum
HF300	Chromium	Chromium
HF301	Diatomaceous earth	Diatomaceous earth
HF302	-	Ketones
HF303	Bitter melon	Bitter melon

---

---

HF303	Bitter melon	Momordica charantia L.
HF304	Cola nut	Cola nitida
HF304	Cola nut	Cola nitida (Vent.) Schott & Endl.
HF304	Cola nut	Cola nitida Schott & Endl.
HF304	Cola nut	Cola nitida Vent. Schott & Endl.
HF304	Cola nut	Cola nut
HF304	Cola nut	kola nut
HF305	Goldenseal	Goldenseal
HF305	Goldenseal	Hydrastis canadensis L.
HF307	Cobalamin	Cobalamin
HF308	Colloidal silver	Colloidal silver
HF309	Chondroitin	Chondroitin
HF310	Saccharomyces boulardii	Saccharomyces boulardii
HF311	Wild yam	Dioscorea spp.
HF311	Wild yam	Dioscorea villosa
HF311	Wild yam	Wild yam
HF312	Phaseolus vulgaris	Phaseolus vulgaris
HF312	Phaseolus vulgaris	White kidney bean
HF313	-	Steroids
HF314	Spearmint	Mentha spicata subsp. spicata
HF314	Spearmint	Spearmint
HF315	Cesium	Cesium
HF317	Gelatin	Gelatin
HF319	Diet pills	Diet pills
HF320	Cod liver oil	Cod liver oil

---

---

HF320	-	Gadinae
HF320	-	Gadus morhua
HF320	-	Lotinae
HF321	Tart cherry	Prunus cerasus L.
HF321	Tart cherry	Tart cherry
HF321	Tart cherry	sour cherry
HF322	Protein	Protein
HF323	Dandelion	Dandelion
HF323	Dandelion	Taraxacum officinale
HF324	Thiamin	Thiamin
HF325	Vitex agnus-castus	Chasteberry
HF325	Vitex agnus-castus	Vitex agnus-castus
HF326	Oenothera biennis	Evening primrose oil
HF326	Oenothera biennis	Oenothera biennis L.
HF327	Horsetail	Equisetum arvense
HF327	Horsetail	Horsetail
HF328	Deer velvet	Deer velvet
HF329	Tea Tree Oil	Melaleuca alternifolia
HF329	Tea Tree Oil	Tea Tree Oil
HF330	Devil's Claw	Devil's Claw
HF330	Devil's Claw	Harpagophytum procumbens
HF332	Horse chestnut	Aesculus hippocastanum L.
HF332	Horse chestnut	Horse chestnut
HF333	Dong quai	Angelica sinensis
HF333	Dong quai	Dong quai

---

---

HF334	Feverfew	Feverfew
HF334	-	anacetum parthenium L.
HF335	Coffee	Coffea arabica L.
HF335	Green coffee beans	Green coffee beans
HF337	Cat's Claw	Cat's Claw
HF337	Cat's Claw	Uncaria tomentosa
HF338	Bacopa	BACOPA MONNIERI
HF338	Bacopa	Bacopa monnieri (L.) Wettst.
HF338	Bacopa	Bacopa monnieri L. Wettst.
HF338	Bacopa	Bacopa monnieri Wettst.
HF339	Peppermint oil	Mentha piperita L.
HF339	Peppermint oil	Peppermint oil
HF340	Papaya	Carica papaya
HF340	Papaya	Papaya
HF342	beet	Beetroot
HF342	beet	Beta vulgaris
HF342	beet	beet juice
HF345	Biotin	Biotin
HF345	Biotin	Vitamin B7
HF347	Pyruvate	2-oxopropanoic acid
HF347	Pyruvate	Pyruvate
HF347	Pyruvate	pyroracemic acid
HF347	Pyruvate	$\alpha$ -ketopropionic acid
HF348	Vinpocetine	Vinpocetine
HF349	Hoodia	Hoodia

---



---

HF349	Hoodia	Hoodia gordonii
HF351	Butterbur	Butterbur
HF351	Butterbur	Petasites
HF351	Butterbur	Petasites L.
HF351	Butterbur	Petasites hybridus
HF352	Fucoxanthin	Fucoxanthin
HF353	Fluoride	Fluoride
HF354	Grape seed extract	Grape seed extract
HF354	Grape seed extract	Vitis vinifera
HF355	Huperzine A	Huperzine A
HF356	Bromelain	Bromelain
HF358	Berberine	Berberine
HF359	Hawthorn	Crataegus chungtienensis
HF359	Hawthorn	Hawthorn
HF360	Pau D'Arco	Pau D'Arco
HF360	-	Tabebuia
HF360	Pau D'Arco	Tabebuia avellanedae
HF361	Horny goat weed	Epimedium
HF361	Horny goat weed	Horny goat weed
HF362	Boron	Boron
HF363	Manganese	Manganese
HF364	Mangosteen	Garcinia mangostana
HF364	Mangosteen	Garcinia mangostana L
HF364	Mangosteen	Mangosteen
HF365	Nootropics	Nootropics

---

---

HF366	Molybdenum	Molybdenum
HF367	-	Arceuthobium pusillum
HF367	American Mistletoe	Mistletoe
HF367	European Mistletoe	Viscum album
HF367	European Mistletoe	Viscum album ssp. Abietis
HF368	Salix alba	Salix
HF368	Salix alba	Salix alba
HF368	Willow bark	Willow bark
HF369	Eucalyptus	Eucalyptus
HF370	European Mistletoe	European Mistletoe
HF370	European Mistletoe	Viscum album L.
HF371	Yohimbe	Yohimbe
HF372	-	Artemisia lactiflora Wall.
HF372	Mugwort	Mugwort
HF373	-	4-(4-hydroxyphenyl)-2-butanone
HF373	Raspberry ketones	Raspberry ketones
HF374	Lavender	Lavandula angustifolia Mill.
HF374	Lavender	Lavender
HF375	Ribose	Ribose
HF377	Green tea	Camellia sinensis
HF377	Green tea	Green tea
HF378	APPLE CIDER VINEGAR	APPLE CIDER VINEGAR
HF379	Phosphorus	Phosphorus
HF380	Thunder god vine	Thunder god vine
HF380	Thunder god vine	Tripterygium wilfordii

---

HF381	Red clover	Red clover
HF381	Red clover	Trifolium pratense L.
HF382	Astragalus	Astragalus
HF383	Sage	Sage
HF383	Sage	Salvia officinalis L.
HF384	Rhodiola	Rhodiola
HF386	Beta-carotene	Beta-carotene
HF387	Beta-glucans	Beta-glucans
HF388	5-HTP	5-HTP
HF388	5-HTP	5-hydroxytryptophan

素材データベースと Dietary Supplement Fact Sheets から取得した HFP のリスト。  
HFP\_id：今回割り当てた HFPs の ID である。同じ ID であれば、同一の HFP を指すとした。Main name：ゴールドスタンダードを検索する際に Main name となっていたものを示した。「-」は、対象の Synonym name を検索しても見つからなかった名称である。

### 第 3 節 考察

食品に関するデータベースは存在するが、サプリメントやハーブなどを含む HFPs に関するデータベースはほとんどなかった。そのような状況から 365 種の HFPs をリスト化できたが、医薬品数に比べて HFPs の方が少なく、HFPs のリスト化自体にも課題があると考えられる。この課題を解決すべく、さらなるデータベースの探索が必要である。さらに素材情報データベースと Dietary Supplement Fact Sheets の重複を確認すると、国の違いによる特徴的な HFPs が存在することがわかった。今回は日本とアメリカの HFPs リストを取得した。今後は他の国の HFPs にも目を向けることで、さらに多くの HFPs をリスト化できると考える。書籍 (Japan Health Food

and Supplement Information Center 2022)を使用し、リスト化した HFP を調査した。この書籍は日本の厚生労働省から信頼できる健康食品の情報源として認められており、医薬品と健康食品の相互作用について解説されている。原典は英語で WHO(世界保健機関)が信頼できると評したデータベースであり、学術誌に日々投稿されている世界中の科学論文を対象に、医療従事者をはじめ、薬学、医学、生化学など関連する学術分野の専門家たちが、一貫した編集方針に従い、情報を精査されている。その中で、今回は権利問題を考慮し、日本語の書籍を用いた。さらに他のツールも検討したが、この書籍に最も相互作用情報が記載されており、かつ公的機関にも認められているため、この書籍を用いた。しかし、調査の結果、一部の HFP を見つけることができなかった。これは、国によって異なる HFP が利用されている場合、また最新の情報が記載されていない場合や、各書籍やデータベースごとに HFP が異なる方法で分類されている可能性があるためである。今後の研究で HFP の分類について再評価し、新しい HFP 情報を継続的に収集する必要があると考えられる。さらに、取得した HFPs の中には、ブルーベリーなどの食物名やブルーベリーに含まれるアントシアニンのような成分名(サプリメントとして用いられる場合には製品名)が別の HFP としてリストアップされた。一方で、このリストではブルーベリーとアントシアニンの関係性は示されていない。成分とそれを含有する食物の関連性をリストに反映させることができれば、より詳細な相互作用情報を抽出できると考えられる。今回データベースを調査した結果、リストの作成やデータベースへの情報の集積がウェット研究とドライ研究の間に存在する障壁であると感じた。特に、ドライ研究において関係性や原因を解明するためのツールはある程度充実しているが、ドライの研究者が扱うことができる基礎データが不足している。この問題は、データはすでに論文で報告されているが、データのフォーマットが研究者ごとに異なることや、ネガティブデータが公開されないことが原因の一つとなっている。微生物研究者に対し、これらのデータを集積するために統一されたフォーマットでデータを登録し、ネガティブデータの提供が可能かどうかを尋ねたことがある。しかし、研究の視点が異なることや主張内容に合わせて変

化させる必要があるため完全に同じフォーマットに統一するのは困難であり、さらにネガティブデータは、その研究者が当該研究を実施した証拠としての役割を持ち、他者によるデータの盗用や研究成果の横取りを防ぐ場合があるため、公開には慎重にならざるを得ないとの回答を得た。このようなドライとウェットの研究の障壁を取り除き、データの集積と解析を進めることによって、日本の生物工学分野は今後さらに発展する可能性があると考えられる。

## 第 2 章 BioGPT を用いた DHIs の取得

現代の科学は、過去に蓄積された膨大な科学的知識に基づいて進化し続けている。特に医学と薬学の分野は人々の生命に関わる研究であり、日々の研究活動において多くの文献から特定の情報を早急に抽出することが不可欠である。医学や薬学の研究者は、新たな知識や発見を得るために、文献データベースや研究論文を探し、分析する必要がある。

医薬品と HFPs, 酵素と HFPs の相互作用情報の抽出においても、この精査された情報のデータベースを構築は極めて重要である。現在、このような情報を検索するためには、研究者や医学専門家が手動で大量の文献を読み、必要な情報を見つけるために膨大な時間と労力を費やす必要がある。

しかし、この課題を効率的に解決するために、BioGPT という新しいツールを用いた。BioGPT は、広範な生物医学文献でトレーニングされた言語モデルで、医療専門家や研究者が医学的情報を素早く検索し、研究に必要な情報を迅速にアクセスできるようになる貴重なツールとして注目されている。

我々は、この BioGPT が医薬品と HFPs, 酵素と HFPs の相互作用に関する情報の抽出に特に適していると考えた。このモデルを活用することで、医薬品と食品の相互作用に関する知識をより迅速かつ効果的に抽出し、研究の進展に寄与することができれば、その他の医療や栄養に関する研究もより効率的に進められ、健康と医学の分野に新たな洞察がもたらされることが期待される。

### 第 1 節 方法

Google Colab (<https://colab.research.google.com/?hl=ja>) を使用して、BioGPT(<https://github.com/microsoft/BioGPT>)を動かした。BioGPT を用い、医薬品と HFP の相互作用に関する情報を検索した。具体的には「A(HFP 名) and B

(Drug 名) interaction is」などのように、文の始まりを指定した後、BioGPT が自動的に文を生成する方法を採用した。生成される文の長さは最長 1,024 文字とし、複数の異なる文の始まりを試して研究を進行しました。

## 第 2 節 結果

一部の DHIs について、文頭の条件を検討するために 1000 字程度の文章を BioGPT から取得した。文頭には「A(HFP 名) and B (Drug 名) interaction is」、  
「A(HFP 名) and B (Drug 名) interactions are」、  
「A(HFP 名) and B (Drug 名) interactions in human are」、  
「Interaction between A(HFP 名) and B (Drug 名) is」、  
「Interactions between A(HFP 名) and B (Drug 名) are」、  
「All cases of interactions between A(HFP 名) and B (Drug 名) are」などを試した。出力結果を確認したところ、  
「Interactions between A(HFP 名) and B (Drug 名) are」から始まる文章に最も有用な情報が含まれていた (Table 2-1)。ゆえに「Interactions between A(HFP 名) and B (Drug 名) are」を用い、すべての DHIs について出力を試みた。しかし、Google Collaboratory は 90 分に経過するとランタイムがリセットされ、自動で出力していた文章が止まってしまうことがわかった。設定を変更して回避コードも行ったが、12 時間経つと強制的に初期化されてしまう。そこで PC にダウンロードして BioGPT を稼働させた。その結果、「Interactions between A(HFP 名) and B (Drug 名) are」を用い、すべての DHIs について出力できた (Table 2-2)。

Table2-1 検討した文頭ごとに出力された文章

文頭	出力された文章
A(HFP 名) and B	4-hydroxyproline and Progesterone interaction is discussed in relation to the mechanism of action of these compounds on the uterine cervix of the guinea-pig, and

---

(Drug 名)            the possible significance of these findings in relation to the mechanism of action of  
interaction            these compounds on the cervix of the human is considered. (ABSTRACT  
is                        TRUNCATED AT 250 WORDS) The significance of these findings is discussed in  
relation to the mechanism of action of these compounds on the cervix of the guinea-  
pig, and the possible significance of these findings in relation to the mechanism of  
action of these compounds on the human cervix is considered.

                          4-hydroxyproline and Progesterone interactions are discussed in terms of their  
A(HFP                possible role in the mechanism of action of the steroid hormones on the uterus and the  
名) and B             possible role of 4-hydroxyproline in the mechanism of action of the steroid hormones  
(Drug 名)             on the uterus and other tissues is also considered in relation to the possible role of 4-  
interactions         hydroxyproline in the mechanism of action of the steroid hormones on the uterus and  
are                     other tissues and the possible role of 4-hydroxyproline in the mechanism of action of  
the steroid hormones on the uterus and other tissues.

                          4-hydroxyproline acid and Progesterone interactions in human are similar to those  
A(HFP                observed in the rat, suggesting that the rat may be a useful model for the study of the  
名) and B             effects of 4-hydroxyproline acid and progesterone on collagen metabolism in the  
(Drug 名)             human, and that the rat may be a useful model for the study of the effects of 4-  
interactions         hydroxyproline acid and progesterone on collagen metabolism in the human, as well as  
in human             for the study of the effects of 4-hydroxyproline acid and progesterone on collagen  
are                     metabolism in the rat, as well as for the study of the effects of 4-hydroxyproline acid  
on collagen metabolism in the rat.

Interactio            Interaction between 4-hydroxyproline and Progesterone is studied by NMR  
n between            spectroscopy and molecular dynamics simulations in aqueous solution and in the  
A(HFP 名)            presence of micelles of sodium bis (2-ethylhexyl) sulfosuccinate (AOT) and sodium bis  
and B                 (2-ethylhexyl) sulfosuccinate / sodium dodecyl sulfate (AOT / SDS) in the presence of

---





Table2-2 文頭に「Interactions between A(HFP 名) and B (Drug 名) are」を用い  
BioGPT で出力した文章

drug_id	HFP_id	出力された文章
D00066	HF1	<p>Interactions between progesterone and 4-hydroxyproline are important in the regulation of collagen synthesis in the uterus of the guinea-pig during the oestrous cycle and early pregnancy. (ABSTRACT TRUNCATED AT 250 WORDS) The present study was designed to investigate the effects of progesterone and 4-hydroxyproline on collagen synthesis in the uterus of the guinea-pig during the oestrous cycle and early pregnancy, and to determine whether these effects are mediated by the progesterone receptor (PR) or the progesterone metabolite 4-hydroxypregnen-3-one (4-OHP).</p> <p>Interactions between estrone and 4-hydroxyproline are discussed in relation to the formation of cross-links between collagen and elastin and to the formation of elastin-collagen cross-links in the aorta of rats fed a high-fat diet containing 4-hydroxyestrone (4-OHE1), a metabolite of estrone, as the sole source of dietary fat for 4 weeks, followed by a 4-week period of a low-fat diet without 4-OHE1, and then a 4-week period of a high-fat diet without 4-OHE1.</p> <p>Interactions between vitamin a oil and 4-hydroxyproline are studied in the presence of 4-hydroxyproline-4-carboxylic acid (4-Hyp-4-COOH), a model compound for collagen, and vitamin A acid (VA-A), a model compound for vitamin A acid (VA-A), in aqueous solution at pH 7.4. The results show that 4-Hyp-4-COOH forms a 1: 1 complex with VA-A and VA-A forms a 1: 1 complex with 4-Hyp-4-COOH.</p>

---

D00076	HF1	<p>Interactions between noradrenaline and 4-hydroxyproline are mediated by alpha 2-adrenoceptors in the guinea-pig isolated vas deferens, but not in the rat isolated vas deferens or in the guinea-pig isolated trachea, nor by alpha 1-adrenoceptors in the guinea-pig isolated vas deferens, nor by alpha 2-adrenoceptors in the rat isolated vas deferens, nor by alpha 1-adrenoceptors in the guinea-pig isolated trachea, nor by alpha 2-adrenoceptors in the rat isolated trachea, nor by beta 2-adrenoceptors in the rat isolated trachea.</p>
D00088	HF1	<p>Interactions between hydrocortisone and 4-hydroxyproline are discussed in relation to the mechanism of action of hydrocortisone in the treatment of rheumatoid arthritis and other connective tissue diseases, and in relation to the mechanism of action of 4-hydroxyproline in the treatment of connective tissue diseases and in the prevention of collagen cross-linking in the skin of patients with systemic lupus erythematosus (SLE) and rheumatoid arthritis (RA), and in the prevention of collagen cross-linking in the skin of patients with SLE and in the prevention of collagen cross-linking in the skin of patients with RA.</p>
D00095	HF1	<p>Interactions between epinephrine and 4-hydroxyproline are important in the regulation of blood pressure and in the pathogenesis of hypertension in the spontaneously hypertensive rat (SHR) and the Milan hypertensive rat (MHS), two models of genetic hypertension characterized by elevated levels of plasma epinephrine and reduced levels of plasma 4-hydroxyproline (4-OH-Hyp), a collagen cross-linking amino acid in the extracellular matrix (ECM) of blood vessels and other</p>

---



---

transform infrared (FTIR) spectroscopy, differential scanning calorimetry (DSC), and X-ray powder diffraction (XRPD), and the results are compared with those obtained for ibuprofen in the presence of 4-hydroxyphenylpropionic acid (4-HPPA) and 4-hydroxyphenyllactic acid (4-HPLA), as well as for ibuprofen in the presence of 4-HPPA.

D00127

HF1

Interactions between piroxicam and 4-hydroxyproline are studied in aqueous solution by <sup>1</sup>H NMR spectroscopy and molecular dynamics simulations: evidence for the formation of a piroxicam-4-hydroxy-4-oxo-4-cyclohexene-1-carboxylic acid complex in solution and in the solid state, and for the formation of a piroxicam-4-hydroxy-4-oxo-4-cyclohexene-1-carboxylic acid complex in the solid state, as well as for the formation of a piroxicam-4-hydroxy-4-oxo-4-cyclohexene-1-carboxylic acid complex in solution.

D00131

HF1

Interactions between disulfiram and 4-hydroxyproline are discussed in relation to the mechanism of action of disulfiram as an inhibitor of aldehyde dehydrogenase and to the mechanism of action of 4-hydroxyproline as an inhibitor of collagen cross-linking and of the formation of cross-links between collagen molecules in the extracellular matrix of connective tissue. (ABSTRACT TRUNCATED AT 250 WORDS) The interaction between disulfiram and 4-hydroxyproline is discussed in relation to the mechanism of action of disulfiram as an inhibitor of aldehyde dehydrogenase and to the mechanism of action of 4-hydroxyproline as an inhibitor of collagen cross-linking and of the formation of cross-links between collagen molecules in the extracellular matrix of connective tissue.

---

---

D00136	HF1	<p>Interactions between haloperidol and 4-hydroxyproline are consistent with a model in which haloperidol and 4-hydroxyproline compete for a common binding site on the dopamine D2 receptor and that haloperidol and 4-hydroxyproline compete for a common binding site on the dopamine D3 receptor, which is distinct from the haloperidol-binding site on the dopamine D2 receptor and which is distinct from the 4-hydroxyproline-binding site on the dopamine D3 receptor and which is distinct from the 4-hydroxyproline-binding site on the dopamine D2 receptor and which is distinct from the haloperidol-binding site on the dopamine D3 receptor.</p>
D00139	HF1	<p>Interactions between methoxsalen and 4-hydroxyproline are studied by 1H NMR spectroscopy and molecular mechanics calculations. (ABSTRACT TRUNCATED AT 250 WORDS) Interactions between methoxsalen and 4-hydroxyproline have been studied by 1H NMR spectroscopy and molecular mechanics calculations, and the results are compared with those obtained for the interaction of methoxsalen with 4-hydroxyproline and 4-hydroxypicolinic acid, which have been studied previously by 1H NMR spectroscopy and X-ray diffraction methods, and for the interaction of methoxsalen with 4-hydroxypicolinic acid.</p>
D00140	HF1	<p>Interactions between erythromycin and 4-hydroxyproline are studied in aqueous solution by 1H NMR spectroscopy and molecular dynamics simulations: evidence for the formation of an intramolecular hydrogen bond between the carbonyl oxygen of the 4-hydroxy group and the amide nitrogen of the 4-hydroxyproline residue in the erythromycin-4-hydroxy-4-methyl-4'-phenyl-4''-(1H-tetrazol-5-yl)-biphenyl-4'''-</p>

---

※出力結果から最初の 15 行を抜粋して表記. Drug\_id: KEGG から取得した医薬品の ID. HFP\_id: 第 1 章で振り分けた HFP の ID. 出力された文章: 文頭の医薬品名と HFP 名に左記の医薬品名と HFP 名を入力し, 出力した結果の文章.

### 第 3 節 考察

BioGPT の有用性を確認するために, まずは医薬品と HFPs の相互作用について調べた. 文頭については「in human」のように限定的な表現をつけると人間における相互作用についての文章ではなく, 人間の研究とすでに行われている研究の比較の文章が出力される. さらに人間で研究されているものを探すため, 4-hydroxyproline から 4-hydroxyproline acids に自動的に変更された出力結果となっている. これにより, 相互作用の有無の議論が困難となり限定的な表現は避けるべきであると考えた. さらに「interaction」のように相互作用の単数表記では特定の相互作用を調べた研究にフォーカスされてしまい, これも適切ではないと考えた. 「Interaction between A and B」のような文頭では物質間の相互作用について議論され, 生体内の相互作用の議論は行われなかった. また, 表のように出力結果の異常も確認された. これらの結果から, 複数の研究を幅広く取得できる「Interactions between A(HFP 名) and B (Drug 名) are」が最適であると考え, 以降の出力に用いた.

結果を確認すると, 1,000 字程度の文章の効率的な判定方法についての課題が浮かび上がった. 現在, 44,965 件の文章を全て手動で判定する必要がある, その中には異なる相互作用パターンが含まれている. 具体的には, 理論上, 分子レベル, 細胞レベル, マウスレベル, マウス以外の哺乳類でのレベルなどがあり, これらのパターンによって判定の難易度が異なる. さらに, 手動判定には主観的な要素が含まれ, 判定の一貫性を確保することが困難であると考えられる. この主観性は, 研究の信頼性に潜在的

な問題を引き起こす可能性がある。一方、ファインチューニングを行うことで出力結果を調整する方法もある。しかしながらこの方法は非常に高い金銭的成本および計算コストがかかり、難しい。したがって、一次スクリーニング方法の検討のために他の自然言語処理の手法を探索する必要がある。

さらに、文章を読み判定をしたとしても、これらの情報の信頼性や潜在的な問題を確認するためにゴールドスタンダードデータが必要である。



### 第3章 方法1：PubMedに基づく代謝酵素とHFPsの共起検索とネットワーク分析と酵素によるグルーピング

HFPs と医薬品の相互作用についての情報は、健康をサポートするために摂取される「HFPs」が増加している中で非常に重要である。しかし、この相互作用に関する情報を獲得することは容易ではない。特に、HFPs と医薬品の相互作用情報を抽出するには、専門的な知識と多くの文献調査が必要である。この課題に対処するため、我々は共起検索というアプローチを提案した。

共起検索は、関連性の高い情報が互いに関連して出現する傾向があるという考えに基づいている(White, 2020; Motschall *et al.*, 2005)。つまり、医薬品とHFPsに関する文献で一緒に頻繁に現れるキーワードやフレーズを特定し、それらを分析することで、相互作用情報を効率的に抽出することができるという仮説である。

この方法を医薬品とHFPの相互作用情報を抽出する手法として検討しました。具体的には、関連する文献から共起するキーワードを抽出し、それに基づいて医薬品とHFPsの相互作用に関する候補情報を特定しました。このアプローチを用いることで、膨大な文献から有用な情報を効率的に抽出し、健康と医学の分野における研究に貢献する可能性がある。

さらに研究活動において、方法論はその信頼性と有用性に関する鍵となる要素である。緻密に計画された方法論は、問題の解決に向けた道筋を示し、データの収集や分析の際に方向性を提供する。しかし、方法論の妥当性と効果を評価するためには、評価方法が不可欠である。

方法論を議論する上で、評価方法は最も重要な項目の一つである。なぜなら、正確な結果や信頼性のあるデータを得るためには、適切な方法論の採用だけでなく、その方法論が実際に適用可能であるかどうかを確認する必要があるからである。研究者が提案する方法論が科学的な基準を満たし、適切に評価されることは、研究の信頼性を高め、他の研究者との情報共有を可能にする。

本章では、新しいワークフローを確立するために、共起検索を用いて医薬品と HFP の相互作用情報抽出を行い、決定した評価方法を用いて評価した。

## 第 1 節 方法

### 第 1 項 PubMed に基づく代謝酵素と HFPs の共起検索

医薬品と代謝酵素間の相互作用は一部の組についてはすでに広く認識されており、KEGG にもその情報が詳しく説明されているが、酵素と HFPs の相互作用について詳しく説明されているオープンデータベースはほとんど存在していない (DrugBank, Wishart et al., 2018; Drug.com). そこで HFPs と酵素の相互作用について情報を取得するために pubmed で酵素と HFPs の共起検索を行った。共起検索は、生物学の実体間の関連性を迅速かつ含めて評価する一般的な手法であり、その有効性は既存の文献でも確認されている (Fleuren *et al.*, 2015; Junge *et al.*, 2020; Pletscher-Frankild *et al.*, 2015; Szklarczyk *et al.*, 2019). 共起検索のために、2023 年 4 月 14 日に PubMed(<https://pubmed.ncbi.nlm.nih.gov/>)に 2022 年 12 月 31 日時点で記載されている全文献約 340,960,700 件のタイトルおよびアブストラクト、キーワード、PMID をダウンロードした。上記で集めたすべての酵素(フルネーム)と HFPs をペアで組み合わせ、タイトルおよび抄録内において共起検索し、相互作用の可能性のある酵素と HFPs のペアを抽出した。さらに関連性の高い酵素と HFPs のペアを選抜するために、PubMed における酵素・HFP ペアの検索ヒット件数 (PubMed Hits) が 1, 5, 10, 15, 20, 25, 30, 35, 40 件以上のペアをそれぞれ保持し、比較した。

### 第 2 項 共起検索を用いたゴールドスタンダードデータの取得

第 3 章第 1 項の PubMed に基づく酵素と HFPs の共起検索と同様の方法を用い、リスト化した医薬品と HFP をペアにし、PubMed で共起検索した。PubMed Hits 20

以上の医薬品と HFP のペアを相互作用ありと仮定した。

### 第 3 項 書籍情報を用いたゴールドスタンダードデータの取得

HFPs と医薬品の相互作用情報を確認するために日本医師会が出版している書籍を利用した (Japan Health Food and Supplement Information Center 2022)。この本で HFPs を調べ、HFP の情報として記載されている医薬品との相互作用を抜き出し、ゴールドスタンダードデータとした。

### 第 4 項 PubMed の文献を用いたゴールドスタンダードデータの取得

PubMed データベースを活用し、酵素に関連する医薬品と HFP の相互作用情報を収集するため、「酵素名 Interaction」をキーワードとして検索を行った。検索結果からヒットした文献から医薬品との相互作用に関連するとされる文献をタイトルとアブストラクトから判断し、選定した。さらに選定された文献を読み、文献内に記載されている医薬品と HFP の相互作用情報を取得した。

### 第 5 項 ネットワーク構築および酵素によるグルーピング

医薬品、酵素、HFPs をノードとし、KEGG および PubMed より取得した医薬品-酵素ペアの相互作用、HFPs-酵素ペアの相互作用候補をエッジとしてネットワークを構築した。PubMed Hits の閾値は 20 とした。また、可視化および解析には NetworkX (<https://networkx.org/>), pyVis (<https://pyvis.readthedocs.io/en/latest/>), Cytoscape 3.10 を使用した (Shannon *et al.*, 2003)。相互作用する可能性がある医薬品-酵素-HFPs の三者の関係から解析される医薬品と HFPs のペアを DHIs とする。さらに DHIs のハブとなった酵素ごとにグループ分けする。

### 第 6 項 ゴールドスタンダードとの一致率

HFPs と医薬品の相互作用情報を確認するために日本医師会が出版している書籍

を利用した (Japan Health Food and Supplement Information Center 2022). リスト化した 365 種の HFPs の医薬品との相互作用情報を検索し, 1,205 種の医薬品, 83 種の HFPs, 21 種類の酵素を含む 56,295 件の医薬品と HFPs の相互作用情報を取得した. 取得した 56,295 件の HFPs と医薬品との相互作用情報をゴールドスタンダードとした. 医薬品と HFPs の相互作用とゴールドスタンダードの一致率とした.

一致率(%) = Drug-HFP interactions (DHIs)中でゴールドスタンダードと一致した DHIs/ DHIs×100

## 第 2 節 結果

### 第 1 項 共起検索を用いたゴールドスタンダードデータの有用性

第 3 章第 1 節第 2 項のアプローチを用いて PubMed での文献ヒットを利用して医薬品と HFPs の共起検索を行い, 既知の医薬品と HFPs の相互作用情報を取得しようとした. しかし, このアプローチにはいくつかの課題が浮き彫りとなった. まず, PubMed の検索ヒット数の閾値を設定することが難しい. さらに, ヒット数が多いからといって, 実際には相互作用情報が含まれていないこともあった. 閾値を設定せずに全文献を調査すると, 詳細な精査が必要であり, 時間がかかることが明らかであった. 先行研究を元に閾値を 20 としたところ, 医薬品と HFP のペアは非常に少なく, 2,000 件程度であった(Wang *et al.*, 2023). これらは第 3 章第 5 項で取得した DHIs と比較して非常に少なく限られており, 評価として不適當である.

### 第 2 項 書籍情報を用いたゴールドスタンダードデータの有用性

次に HFPs と医薬品の相互作用情報を確認するために日本医師会が出版している書籍を利用した (Japan Health Food and Supplement Information Center 2022). この書籍には, 世界中の研究者が「健康食品」・サプリメントに関する文献を系統的にレビューした成果がまとめられている. この書籍を本研究で利用可能なデータ化し

ようとした。しかし、HFPs リストの 365 種中 83 種のみが記載されており、データが不足していた。さらに書籍のデータ化には膨大な時間がかかることが予測された。書籍に記載されている医薬品と HFP の相互作用情報は HFP の紹介ページ内に相互作用する Drug 名または CYP3A4 阻害剤のような集団名で記載されている。これらの医薬品名を取得したリストと照らし合わせ、一致した医薬品を手作業で抜き出さなければならない。このように非常に煩雑な作業になった。これらの要因から、データの収集と処理には時間と労力が膨大にかかることが明らかとなった。

膨大な時間をかけ、記載されている 83 種の HFP の医薬品との相互作用情報を抜き出し、ゴールドスタンダードデータとした。これらを第 3 章第 1 節第 5 項で酵素ごとに分けたグループと照合し、グループ内の正解率を既知率としてスコア付する方法を検討した。同一グループ内であれば、同じスコアとして既知率が高いほど相互作用が起こる可能性が高いとした。しかしながら、この既知率を用いた方法では著作権に抵触する恐れがあり、他の方法を検討した。

### **第 3 項 PubMed の文献を用いたゴールドスタンダードデータの有用性**

続いて PubMed において検索ワードを「酵素名 Interaction」などのように工夫し、医薬品と HFPs の相互作用を取得しようとした。検索ワードを検討して 1,943 本の論文を選抜した。この 1,943 本の論文の内容を吟味し、53 件の論文から医薬品と HFPs の相互作用情報を取得した。しかし、残り 50,000 件以上の論文を精査する必要があり、大学の契約上内容を確認できない文献や論文出版年数が古くタイトルや要約しか載っていない文献も多数あった。これらの要因からこの方法による正解データの取得は不可能であった。

### **第 4 項 ゴールドスタンダードデータの比較**

第 3 章第 2 節第 1 項から第 3 項の結果を比較した。その結果、正解データとして用いることができるのは書籍から取得したデータのみであった。そこでデータの二

次利用である「スコア付」から視点を変え、書籍から取得したデータを確認のために用いるゴールドスタンダードとしてのみ利用することとした。

## 第 5 項 共起検索における有用性および閾値

PubMed は医学，生物学，関連分野で重要な位置を占めている。酵素と HFP 間の相互作用に関する情報を収集するために共起検索を採用した。様々な PubMed Hits (閾値) (1, 5, 10, 15, 20, 25, 30, 35, 40) がゴールドスタンダードとの一致度を評価した。閾値 0 (リストに記載されている酵素と HFP を総当たりでペアにしたもの) では一致率は 13%，閾値 1 では 21% に上昇した (Fig. 3-1)。これらの結果から，共起検索は PubMed が非常に包括的であることが有用であることが示唆された。その結果，閾値 20 が 30% と最も高い一致率をもたらすことが確認された。この閾値は，一致率の最大化と過剰な情報の包含の最小化のために選択された (Wang *et al.*, 2023)。

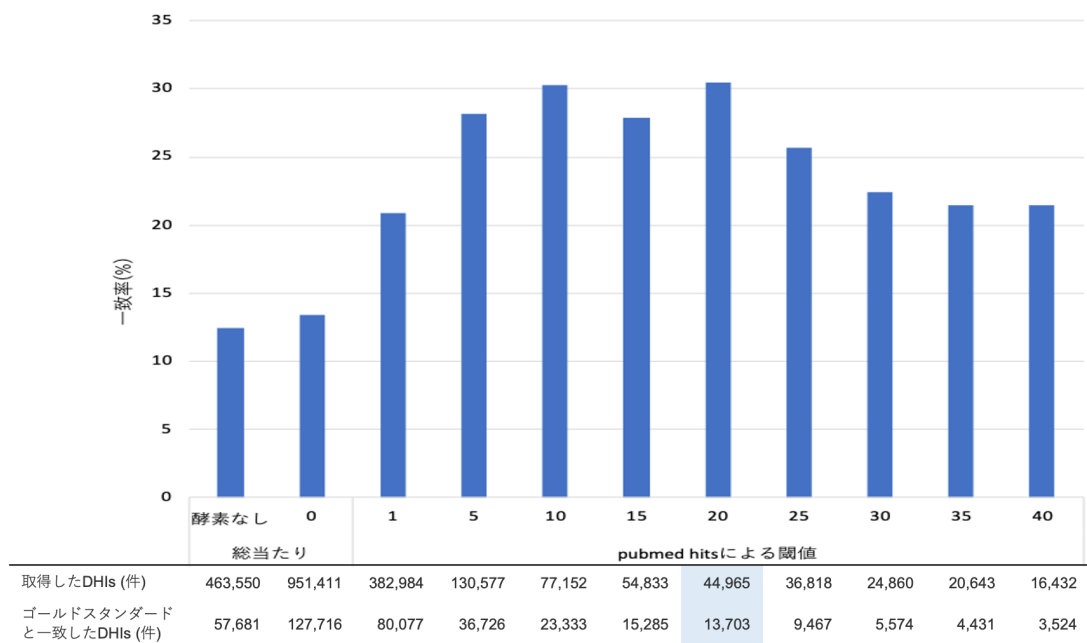


Fig. 3-1 PubMed Hits による閾値ごとの一致率

酵素なし：ハブとして酵素を用いず、医薬品と HFP を直接繋げた場合。

医薬品—酵素—HFP を繋げた複雑な結果と比較するために、医薬品と HFP を直接結びつける方法を用いた。この方法では、検索数が 20 以上の医薬品と HFP のペアが 2,764 組観察されたが、そのうち既知の相互作用は 89 組のみであった。この結果は、この方法では限られた範囲の相互作用、特に新規の未知の相互作用を捉えることに限界があることを示唆している。一方、医薬品、酵素、HFP をノードとするネットワーク構築法を採用することで、44,965 個の DHI が作成され、より広範なスペクトルを調べることができた。このアプローチでは、医薬品と HFP 間の直接的な相互作用を捉えるだけでなく、酵素を介したメカニズムに基づく相互作用の検出も可能である。特に注目すべきは、HFP と酵素との相互作用が、しばしばこれらの酵素に関連する医薬品との関連につながり、新たな相互作用の発見を容易にするとい

う事実である。これらの結果から、複雑さはあるものの、医薬品、酵素、HFP をノードとするネットワーク構築法を採用することで、より包括的な相互作用情報を収集し、新たな知見を得る機会が増えたといえる。

#### **第 6 項 閾値 20 とした共起検索により選抜された DHIs**

共起検索を行ったところ、100 万件以上の医薬品—酵素—HFPs の関係性が示唆された。取得した関係性は非常に多く、判別が困難であった。そこで PubMed Hits 20 以上を選抜するという条件を設定した。HFPs と酵素のペアについて共起検索を行い、PubMed 検索で 20 件以上ヒットするペアを選抜した。この結果、44,965 件の DHIs に絞ることができた。さらに条件を検討するために可視化した。可視化にはネットワーク図を作成する NetworkX, pyvis, Cytoscape を用いた。ネットワークの傾向ごとに配置できる Cytoscape が今回の可視化には最適であった (Fig. 3-2, Fig. 3-3, Fig. 3-4)。



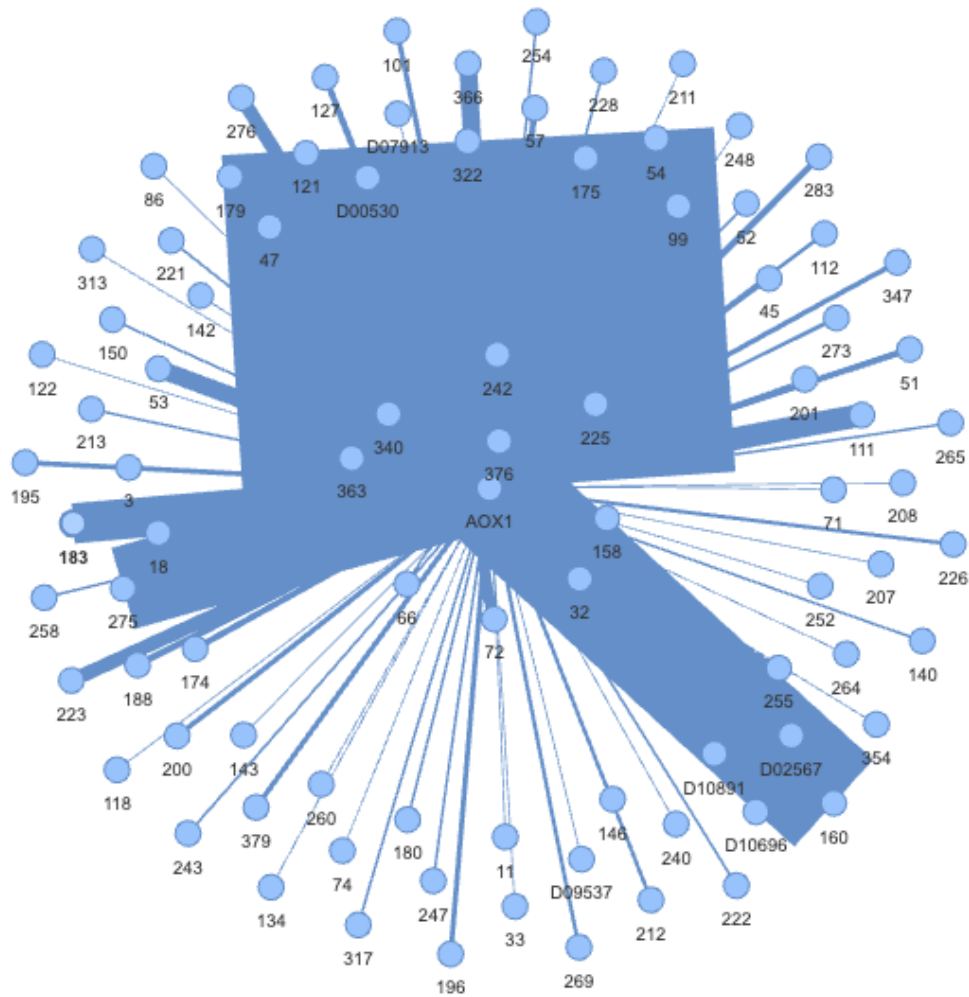


Fig. 3-2 Network X を用いテストデータで作成したネットワーク図

数字：HF を除いた HFP ID. エッジの太さは PubMed Hits の数に比例している.

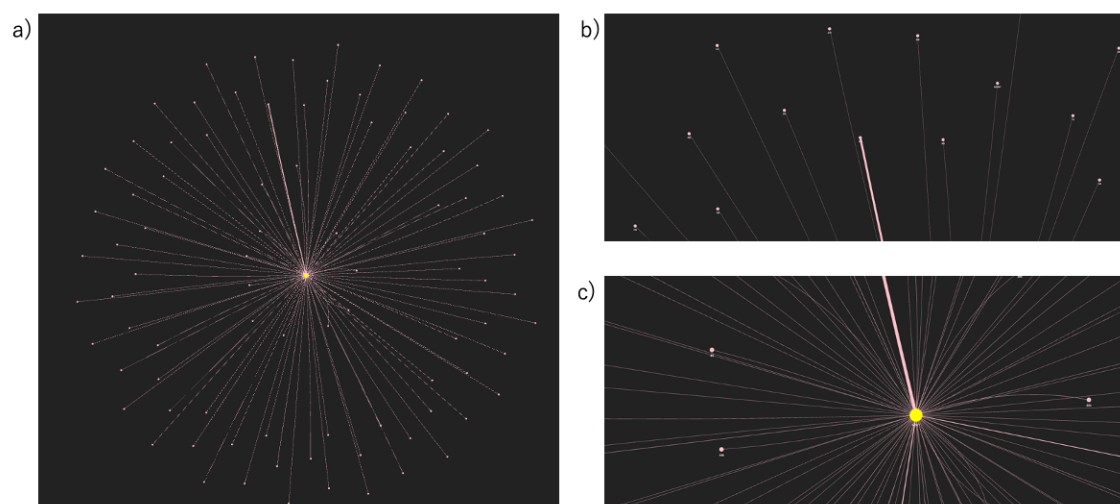


Fig. 3-3 Pyvis を用いて作成した HFP と医薬品のネットワーク図

a : 全体図, b : 末端付近の拡大図, c : 中心付近の拡大図

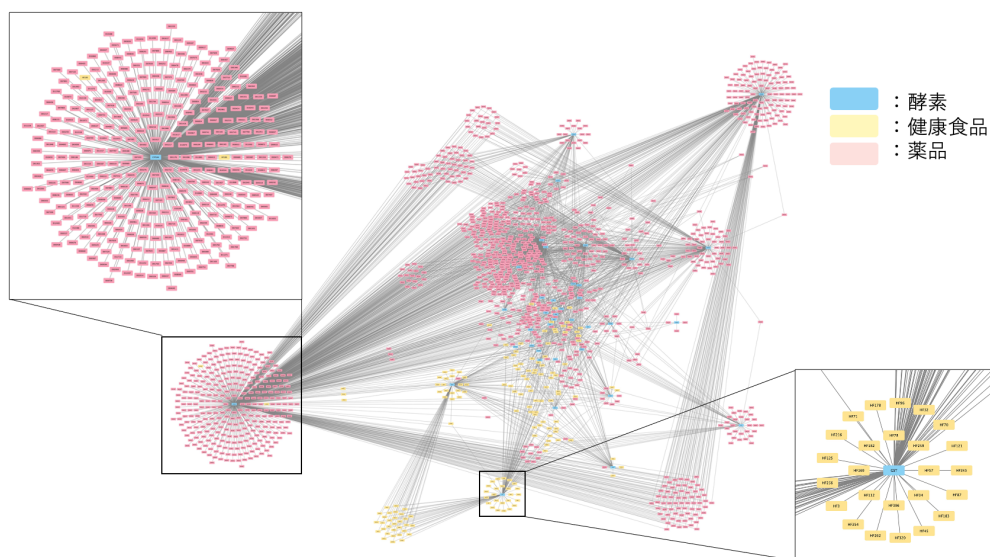


Fig. 3-4 酵素をハブとした医薬品と健康食品のネットワーク図

医薬品、酵素、HFPs をノードとし医薬品—酵素ペアの相互作用、HFPs—酵素ペアの相互作用候補をエッジとしてネットワークを構築し、Cytoscape 3.10 で可視化および解析した。相互作用する可能性がある医薬品—酵素—HFPs の三者の関係から解析される医薬品と HFPs のペアを DHIs とする。ピンクのノード：医薬品、黄色のノード：HFPs、青いノード：酵素、ネットワーク：医薬品と HFPs を関係ごとに配置した。拡大した図のようにハブとなる酵素ごとに医薬品および HFPs はおおよそ分類された。

Cytoscape を用いて、PubMed に代謝ベースの酵素と HFPs の共起検索と、KEGG による酵素—医薬品に基づいたネットワーク図を作成した(Fig. 3-4)。このネットワークに関しては、酵素が中心となり、複数のグループに分かれていた。ハブとなる酵素ごとに分類を行い、結果として 38 種の酵素、157 種の HFP、1256 種の医薬品及び 44,965 件の DHI が抽出された (Table 1)。共起検索で閾値 20 とし、Cytoscape を用い可視化、酵素グループごとに分類する方法を最適とし、方法 1 とした。

CYP2E1 グループについて、特定の 3 種類の薬との DHI を指摘し、それらの薬は 21 種類の HFP と関連が示されたが、一貫性はなかった。

Table 1 方法 1 で選抜された DHIs

Class	DHIs	Number of drugs	Number of HFPs
ACHE	117	1	117
ADA	45	1	45
AOX1	6	6	1
BCHE	30	1	30
CBR1	1	1	1
COMT	486	18	27
CYP1A1	322	14	23
CYP1A2	3586	163	22
CYP24A1	3	1	3
CYP27B1	4	1	4
CYP2A6	60	30	2
CYP2B6	124	62	2
CYP2C19	1015	145	7
CYP2C8	316	79	4
CYP2C9	1720	172	10
CYP2D6	3250	325	10
CYP2E1	441	21	21
CYP2J2	8	8	1
CYP3A4	31500	875	36
CYP3A5	669	223	3
CYP3A7	186	186	1
CYP4A11	2	2	1

CYP4F2	12	4	3
DPYD	8	2	4
FMO3	13	13	1
GST	120	1	120
MAOA	96	16	6
MAOB	66	11	6
NAT1	24	6	4
NAT2	24	6	4
SULT1A1	2	1	2
UGT	510	51	10
UGT1A1	145	29	5
UGT1A7	2	2	1
UGT2B15	4	4	1
UGT2B17	3	3	1
UGT2B7	31	31	1
XDH	13	1	13
<b>Total</b>	<b>44,965</b>	<b>1256</b>	<b>157</b>

バブとなる酵素ごとに分類した方法 1 で選抜された DHIs. Class : DHIs のハブ  
 となる酵素, DHIs : Class に含まれる DHIs 数, Number of drugs : Class に含まれる  
 DHIs の医薬品種類数, Number of HFPs : Class に含まれる DHIs の HFPs 種類数,  
 Total : 方法 1 で選抜された DHIs 数, 方法 1 で選抜された DHIs に含まれる医薬品  
 数および HFPs 数を示す. DHIs : 共起検索において検索件数 20 以上の DHIs 数

### 第3節 考察

第3章第2項のアプローチは、医薬品と HFPs の相互作用情報を効果的に獲得するためには適切ではなかった。共起検索におけるヒット数と二者の関係性の深さは比例することが知られている。しかし、二者の関係は相互作用以外にも多くの関係性があり、共起検索におけるヒット数と二者が相互作用する可能性は必ずしも比例するわけではないと考えられる。また今回用いた PubMed においても情報の限界があり、全ての文献情報を集積しているわけではないという点においても注意する必要がある。PubMed の共起検索は見落としの可能性がある一方で、既知の相互作用だけでなく、研究されていない可能性のある相互作用を見つける可能性があるという価値がある。一方で PubMed の文献ヒット数単独では、情報の不足や誤検出のリスクが高いと考えられる。そのため、新たなアプローチやデータソースの検討が必要である。例えば、専門的なデータベースや情報源を活用することで、より信頼性の高い相互作用情報を収集できる可能性がある。そこで新たな方法を検討した。オンライン上のデータベースもいくつか検討したが、オープンな情報が非常に少なく、情報の信頼性も低かった (DrugBank, Wishart *et al.*, 2018; Drugs.com, <https://www.drugs.com/>)。これはデータベース自体が教育用途には利用可能である一方で、臨床目的には使用しないよう推奨していることに起因する。そこで、臨床目的でも使用される情報を探した。その結果、Japan Health Food and Supplement Information Center 2022 が臨床目的でも使用でき、かつ情報量も最大であると考えられた。

文献から関連情報を抽出する方法として、多くの研究者は論文を個別に読むのが一般的である。しかし、複雑な混合物を含む分野では、化学構造から相互作用を予測することは困難である。共起検索は、PubMed のようなデータベースを利用して、文献内の単語の共起関係を調べる手法である。共起検索は既知の相互作用だけでなく、これまで研究されていなかった相互作用も発見できる可能性がある。

共起検索によって得られた DHIs の一致率が 30%程度であった原因として次の要因が考えられる。一つは相互作用情報に関係のない過剰な情報が含まれてしまっている

ことである。この問題に対しては閾値を 20 に設定することで、できる限りの過剰な情報を削除したが、依然として多く残っていると考えられる。2 つ目は PubMed のデータ保存形式が対応していない可能性があることである。PubMed は 1990 年頃の論文に関しては PDF やタイトルのみでデータが保存されていた。本研究に用いた形式では PDF は適合しておらず、PDF 内の情報を精査できなかつた可能性が高い。その上、CYP ファミリーなどの研究は 1990 年頃が最も活発であり、この年代の研究成果を反映させられていない可能性があり、一致率の低下につながつたと考えられる。3 つ目としてはゴールドスタンダードに含まれる HFP が限られていることである。特定の HFP に関する相互作用情報しか含まれておらず、他の HFP に関しては評価できないため、一致率が低くなる傾向がある。ゴールドスタンダードに含まれる HFP に関連する DHI のみについて一致率を計算したところ閾値 20 の場合、一致率は 80% 以上となつた。

ネットワークの可視化に使用した Network X は python のコード内でネットワークの表示方法を指定しなければならず、可視化後の修正が Cytoscape に比べて難しかった。図のように PubMed Hits をエッジに反映させたが、非常に見にくくなつた。ノードに PubMed Hits を反映させようとしたが、できなかつた。ノードを自由に移動させることもできないため、今回の可視化には不適當であると考えた。続いて、pyvis を用いて可視化を行つた。Pyvis は Network X と異なり、可視化後にネットワーク図を確認しながら変更を加えることができる。しかし、変更できる範囲が狭く、ノードを自由に移動させることもできない。重なり合っているノードを同時に確認することもできない。Network X および pyvis は少数のノードを持つネットワーク図の可視化には適している可能性があるが、多量のノードを持つネットワークの可視化には不適當であつた。ゆえに他の可視化方法を検討した。それが Cytoscape である。Cytoscape は可視化後にネットワーク図を自由に修正することが可能であり、ノードの傾向やグループごとに自動的に移動させることもできる。これにより、図のように傾向ごとにノードをまとめて可視化することができた。このような傾向を掴み可視化するには

Cytoscape が最適であると考えられる。

方法 1 で取得された酵素グループの中で関連する医薬品の種類が多いのは CYP や UGT ファミリーなどの広範な誘導・阻害の影響を受ける酵素グループが多い傾向にあった。これは酵素として医薬品のターゲットになりやすい酵素であり、医薬品との反応については活発に研究されているが、食品などとの関係性は複雑で研究があまりなされていないことが原因であると考えられる。さらに HFP の種類が多い酵素グループは ACHE や BCHE, COMT, GST などの比較的特異性がある酵素が多い傾向にあった。これらの酵素は基質特異性が高く、反応する薬の種類が比較的少ないこと、HFP との関係性が比較的シンプルであることが原因として考えられる。さらに神経系に影響があることが知られている酵素が多く、幻覚など重大な問題に影響があるため研究が進んでいる可能性も高いと考えられる。

DHIs が確認されなかった 21 酵素は、基準としたタンパク質との関係性も PubMed ヒットが 20 以下であり、これは研究が進んでいない酵素である可能性が高い。方法 1 では、グループ内の DHIs が多様でありながら、それ以上の区別ができない。グループ内の DHIs を区別し、解析を進めるためには、新しい方法や別の視点が必要とされる。そのため、DHI の関係を踏まえてクラスタリングする方法を検討した。



## 第 4 章 Deepwalk を用いた DHIs の分類

DeepWalk は、複雑なネットワーク構造を低次元のベクトル空間に埋め込むためのアルゴリズムである。この技術は、ネットワーク上のノードのトポロジーとその局所的な構造を捉えることを目的としている。アルゴリズムの核心は、ネットワーク上でのランダムウォークを利用することにある。各ノードから始まる複数のランダムウォークを生成し、これらのウォークをシーケンスとして扱う。このシーケンスは、自然言語処理の分野での文に似ており、各ノードは単語のように機能する。これらのシーケンスを、Word2Vec という分散表現学習技術を用いて学習する。学習の結果として、各ノードは低次元のベクトルとして表現される。このベクトル表現は、ノードのトポロジー的な近さやネットワーク上での役割を反映している。したがって、DeepWalk は、ネットワーク構造の情報を維持しつつ、ノードを数値的なベクトルに変換するための効果的な手法として広く認識されている。このベクトルは、後続の分析、例えばクラスタリングやノード間の関係性の予測などに使用されることが期待される。この方法を用いることで、方法 1 で選抜した DHIs をクラスタリングできると考えた。さらに新規の相互作用の予測も可能であると考えた。

### 第 1 節 方法

#### 第 1 項 KEGG を利用した医薬品と代謝酵素の相互作用リストの作成

第 3 章で決定した方法 1 で取得したネットワークに対して、DeepWalk を実装し、HFPs、酵素および医薬品をベクトル化した。計算コストも確認し全データでも問題なく計算できることを確認した。さらに以下の 4 パターンでモデルを構築し、Drug-HFood の新規関係性 top10 を抽出した。

1. Enzyme-Drug, Enzyme-HFood のネットワーク + 学習回数(epoch) 5
2. Enzyme-Drug, Enzyme-HFood のネットワーク + 学習回数(epoch)200

3. Enzyme-Drug, Enzyme-HFood, Drug-HFood のネットワーク + 学習回数 (epoch)5

4. Enzyme-Drug, Enzyme-HFood, Drug-HFood のネットワーク + 学習回数 (epoch)5

## 第 2 節 結果

結果を確認すると、解析から得られたベクトルは、トポロジーの類似性を示すものであり、有用性は低かった。偽陽性も多く存在した (data not shown)。

## 第 3 節 考察

Deepwalk は次の 4 つの視点から使用されることが多い。まずは類似性の評価である。DeepWalk によって学習された埋め込みを使用して、ノード間の類似性を評価できる。つまり、ネットワーク内の異なる要素同士がどれだけ関連しているかを定量的に評価することができる。たとえば、ある医薬品と別の医薬品の高い類似度がある場合、それらの医薬品は共通の相互作用を持つ可能性が高いと考えらる。

次に、ネットワーク構造の理解である。DeepWalk は、ネットワーク内の構造をキャプチャするために使用できる。類似した埋め込みを持つノードは、元のネットワーク内で近接していることが期待される。したがって、ネットワーク内のクラスタリングやコミュニティの発見など、ネットワークの構造に関する情報を提供する。

3 つ目は、予測と推論である。DeepWalk によって学習された埋め込みは、新しい相互作用を予測したり、既知の情報から新しい知識を推論したりするために使用できる。たとえば、特定の医薬品と酵素が類似している場合、これらの要素が新しい相互作用を持つ可能性が高いと考えられ、実験的に検証するための仮説を立てることができる。

最後にデータの可視化である。DeepWalk によって生成された埋め込みを可視化す

ることで、ネットワーク内の要素間の関係を視覚化することができる。これにより、データの理解や説明が容易になる。このような利点があるとされている。本章では 1 つ目および 3 つ目の利点に着目し、相互作用情報の抽出およびそれぞれの関連性を導きたいと考えた。しかしながら、今回の結果を確認すると、有用な情報は得られず、グループ分けを細分化することもできなかった。

これは、解析から得られたベクトルは、トポロジーの類似性を示し、実際の相互作用の類似性を反映してと考えられる。ネットワーク上の距離が遠くてもネットワークのエッジが似ていれば、トポロジー上では類似と判断され、関連のない要素同士が誤って類似と見なされる可能性がある。そのため、偽陽性が多く、判定も困難であった。方法 1 で選抜された DHIs をより効率的に分類するために他の方法を検討する必要がある。

## 第 5 章 Biowordvec を用いて数値化した医薬品と HFPs の差分ベクトルにおけるクラスタリング

近年、生物学的データの急速な増加により、データの解析と解釈を行うバイオインフォマティクスへの関心が高まっている (Le *et al.*, 2022) 多くの研究者は、生物学的または生物医学的情報を表現するために自然言語処理 (NLP) を用いている (Le *et al.*, 2023) 特にメタボロミクスで用いられる分子ネットワークでは、ケムインフォマティクスのベクトル化手法が広く知られている。一方、本研究で用いたベクトル化手法はテキストマイニングに基づくものであり、BioWordVec は、生物学的テキストデータの特徴や情報を数値的に表現するために用いられる (Gaudry *et al.*, 2022; Zhang *et al.*, 2019). このツールを用いることで、医薬品、健康食品のようにターゲットに関連する文献や情報から意味的なベクトルを抽出し、これらの要素の関係性を捉えることが可能となる。さらに、医薬品と HFPs の差分ベクトルを用いたクラスタリングアプローチを採用することで、これらの要素の関連性を新たな視点から探索し、関係性ごとにクラスタリングすることができる。この方法を方法 1 で選抜された DHI たちに適用することで、DHI たちの解析を進めることができると考えた。

### 第 1 節 方法

#### 第 1 項 Biowordvec を用いた医薬品および HFPs の数値化(ベクトル化)

Biowordvec は、生物学テキストデータや情報の特徴を数値的に表現する際に使用される手法の一つとして使われる。臨床データベースから事前学習されている BioWordVec モデルを用いて、上記で取得した医薬品名、酵素名、HFPs 名に対して、生命科学の用語との関連性などの特徴を反映した 200 次元のベクトルを生成した (Han *et al.*, 2012). 医薬品はシノニムが存在しなかったため、シノニム処理の必要はなかったが、HFPs はシノニムが存在したため、シノニム処理を行

った。HFPs のシノニムが 2 個の場合、Pubmed Hits 数が多い方の名称のベクトルを採用した。HFPs のシノニムが 3 個以上の場合、まず、HFPs のシノニムのベクトルを用い、全てのシノニムペア間でのコサイン類似度を算出した。次いで、任意のシノニムがペアの一方に含まれる全てのコサイン類似度を集計し、その平均値を算出した。そして、各シノニムについて得られたこのコサイン類似度の平均値を各シノニムの評価値とみなした。最終的に、平均値が最も高いシノニムのベクトルを HFPs の特徴を最も表現したベクトルとしてその代表ベクトルとして採用した。

## 第 2 項 差分ベクトル計算と一次クラスタリング

酵素グループを DHIs の関係性ごとに医薬品と HFPs のペアをクラスタリングするために、数値化した医薬品と HFPs の差分ベクトルを算出した。医薬品と HFPs の差分ベクトルは UMAP で 2 次元圧縮し、HDBscan を使用してノイズ除去クラスタリングを実施した(McInnes *et al.*, 2018; Campello *et al.*, 2013; Melvin *et al.*, 2016)。また、クラスタリング後の結果を可視化した。方法 1 選抜された DHIs と、クラスタリングされた医薬品と HFPs のペアで重複した DHIs を選抜した。方法 1 により酵素で分類されたグループをクラスとし、差分ベクトルでクラスタリングされたクラスをサブクラスとした。

## 第 2 節 結果

方法 1 で選抜された DHIs と本章で選抜された DHIs の重複を確認したところ、方法 1 で選抜された 44,965 件、全ての DHIs と重複し、本章の方法は選抜圧にはならなかった。さらにクラスタリング結果を可視化したところ、ノイズが多く確認された (Fig. 5-1)。全体の 21%がノイズであった。

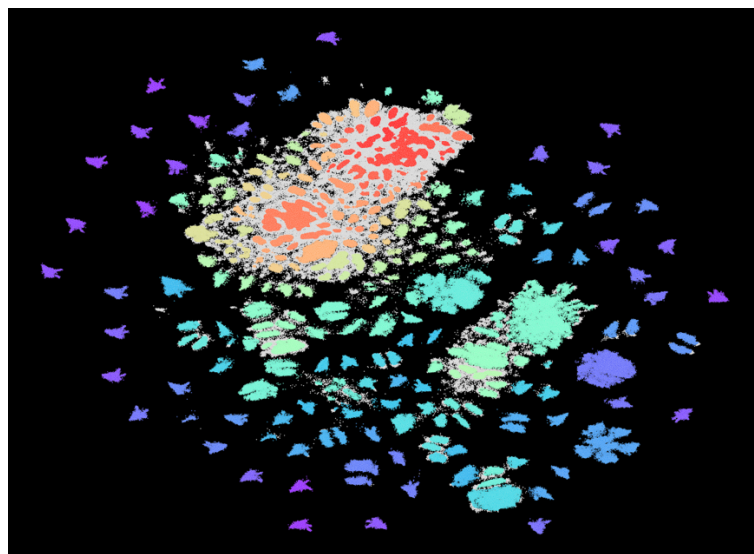


Fig. 5-1 クラスタリングされた医薬品と HFPs の差分ベクトルの可視化した図

### 第 3 節 考察

UMAP を使用したのは、クラスタリング重視、スケーラビリティなどの特徴から最適と考えられたからである。HDBscan も同様に検討され、使用された。

酵素をハブとしない方法では DHI を医薬品と HFPs の関係性ごとにクラスタリングすることはできたが、選抜圧にはならず、ノイズも多く存在したため、新たな方法が必要であると考えられる。特に、HFPs と医薬品の関係性は区別できるが、酵素との関連が不明であるため、方法 1 でハブとしている酵素に着目した選抜圧を検討すべきである。

## 第 6 章 方法 2-1：酵素と直接的に関連する医薬品および HFPs スクリーニング

前章では、DHIs を医薬品と HFPs の関係性ごとにクラスタリングすることはできたが、選抜圧にはならず、ノイズも多く存在したため、前章に追加して前処理または後処理が必要であると考えた。そこで、本研究で着目した酵素を選抜圧として利用するために、医薬品と HFPs の差分ベクトルを計算する前に前処理として酵素と直接関連する医薬品と HFPs の選抜を行った。この前処理には酵素とのコサイン類似度を利用した。コサイン類似度はテキストやベクトルデータなど、異なるデータセットの間で類似性を測定するための一般的な方法の一つである。これは、ベクトル空間モデルを用いて、2つのベクトル間の角度を計算して類似性を評価する方法である。主に自然言語処理や情報検索の分野で使用されるが、さまざまな他の応用分野でも利用されている。コサイン類似度は、-1 から 1 の値を取り、1 に近いほど2つのベクトルが類似していることを示し、-1 に近いほど逆の向きを持つことを示す。0 は完全に無関連を表す。この指標は、テキスト間の類似性やクラスタリング、情報検索など、さまざまなタスクで使用される。酵素とのコサイン類似度が高い医薬品および HFPs を前処理として選抜することで、酵素と直接関係のある医薬品および HFPs を選抜できると考えた。

### 第 1 節 方法

#### 第 1 項 Biowordvec を用いた酵素、医薬品および HFPs の数値化(ベクトル化)

第 5 章 第 1 節 第 1 項で行なった方法と同様の方法で医薬品および HFPs は数値化した。酵素名に関しては全て full name, 全て略称, またはクリーニング後の 3 パターンのシノニム処理を行なった。クリーニング後のパターンでは、略称がクリーニング後となっている場合は略称を採用し、略称が他の略称と重複している場合は Full name を採用した。

## 第 2 項 酵素と直接的に関連する医薬品および HFPs スクリーニング

第 6 章第 1 節第 1 項で数値化された HFPs, 医薬品について, 酵素と直接的に関連する医薬品および HFPs を選抜するために, 酵素ごとに医薬品-酵素ペアおよび HFP-酵素ペアのコサイン類似度を算出した. コサイン類似度は二者間の類似性を示す. 酵素との関連性が高いものを選抜するためにそれぞれのコサイン類似度が高いものを, 酵素と関連する医薬品および HFPs として選抜した.

## 第 3 項 差分ベクトル計算とクラスタリングおよびハブとなる酵素による後処理

第 6 章第 1 節第 2 項で選抜された HFPs および医薬品を第 5 章 第 1 節 第 2 項と同様の方法でクラスタリングした.

さらに, 方法 1 によって選抜された DHIs と, 本項で選抜された DHIs を比較し, 一致した DHIs を抽出した. さらに方法 1 によって選抜された DHIs のハブとなった酵素と, 方法 2-1 によって選抜された DHIs と直接関係するとされた酵素を比較し, 一致した DHIs を抽出した. 方法 1 により酵素で分類されたグループをクラスとし, 差分ベクトルでクラスタリングされたクラスをサブクラスとした.

## 第 2 節 結果

### 第 1 項 酵素の名称による数値化への影響

まず, 酵素と直接的に関連する医薬品および HFPs スクリーニングにおいて酵素のフルネームを使用した. その結果, 前章よりも大幅にノイズが減少した (Fig. 6-1). 方法 1 で取得した DHIs と比較して一致した医薬品-HFP ペア数を計測した. この結果, 方法 1 で取得した 44,965 件中の 7,440 件が一致した. さらにハブとなる酵素が一致した DHI は 80 件であった. FMO ファミリーが直接関連する酵素と



して頻出した。

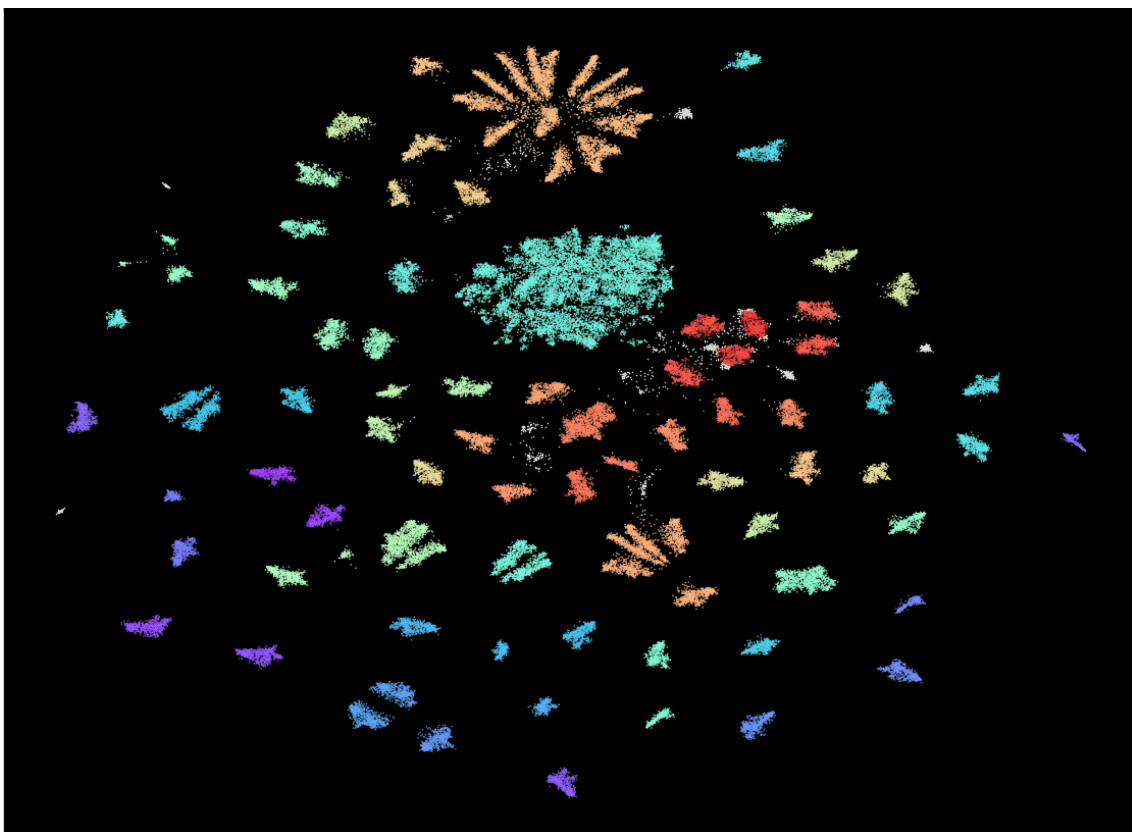


Fig. 6-1 酵素と直接的に関連する医薬品および HFPs スクリーニングにおいて酵素のフルネームを使用した場合にクラスタリングされた医薬品と HFPs の差分ベクトルの可視化した図

続いて、酵素名を略称にした場合、方法 1 で取得した 44,965 件中の 32,219 件が一致した。さらにハブとなる酵素が一致した DHI は 7561 件であった。フルネームを使う場合に比べ、非常に一致する DHIs が増加した。しかしながら、いまだに CYP や UGT ファミリー以外の ACHE, BCHE, SULT ファミリーなどの酵素の出現率が高い。

そこで、使用する酵素名のクリーニングを行った。酵素の略称およびフルネームを PubMed で検索し、ヒット数を確認した (Table 6-1)。ヒットした論文を確認す

ること、その単語が示す対象を確認できる。Biowordvec は単語の意味を数値化するため、複数の意味を持つ単語の場合、本研究には適さない。そのため、ヒット数で確認した。その結果、フルネームと略称でヒット数が多い方を採用した。略称が他の言葉を指している意味合いが強い場合はフルネームを採用した。フルネームと略称でほとんど差がない場合は両方を採用した。

Table 6-1 酵素の名称ごとの PubMed ヒット数

酵素 ID	略称	"フルネーム "	採用した酵素名
1	16,871	1	CYP3A4
2	6,198	1	CYP2C19
3	1,589	0	CYP2C8
7	12,417	0	CYP3A5
8	377	0	CYP3A7
9	1,828	4	CYP27B1
10	1,893	4	CYP24A1
11	5,710	1	CYP2C9
12	6,524	1	CYP2E1
13	7,000	5	CYP1A2
14	8,376	5	CYP2D6
15	1,861	0	CYP2A6
19	2,544	1	CYP2B6
30	10,607	9	CYP1A1
33	685	1	CYP2J2
38	524	1	CYP4F2

51	155	0	CYP2C18
52	215	0	CYP4A11
53	80	0	CYP4F3
5	3,652	2,471	MAOA
6	3,375	2,018	MAOB
24	785	54	AOX1
25	668	1,743	DPYD, dihydropyrimidine dehydrogenase
35	970	0	FMO1
36	1,076	0	FMO3
44	294	148	CBR1, carbonyl reductase 1
45	147	32	CBR3
46	946	0	FMO2
47	908	0	FMO4
48	106	0	FMO5
50	766	1,879	XDH, xanthine dehydrogenase
16	972	0	UGT2B7
17	178	0	UGT2B4
18	3,208	9,214	UGT
20	387	1	UGT1A4
21	319	3	UGT2B15
32	3,157	20	UGT1A1
34	766	0	UGT1A9
37	406	0	UGT1A3
54	340	0	UGT1A7
55	278	1	UGT1A8

58	309	0	UGT2B17
4	5,806	6,250	catechol-O-methyltransferase
22	877	560	NAT1
23	2,020	777	NAT2
40	758	1	sulfotransferase family 1A member 1
41	108	11	sulfotransferase family 1A member 2
42	184	0	sulfotransferase family 1A member 3
43	277	5	sulfotransferase family 1E member 1
57	26,315	31,666	GST, glutathione S-transferase
26	548	334	CES1
27	370	205	CES2
28	41	25	CES3
29	7	0	CES5A
31	1,003,543	37,941	acetylcholinesterase
39	15,329	10,940	adenosine deaminase
49	2,761	6,870	butyrylcholinesterase
56	1,108	864	cathepsin A
59	3,241	2,129	dopa decarboxylase

---

酵素 ID：取得した酵素に独自で振った酵素の ID. 略名：PubMed において酵素の略名を検索した場合の論文ヒット数. "フルネーム"：PubMed において酵素のフルネームを単語ごとに検索されないように"'"を活用し，検索した場合の論文ヒット数. 採用した酵素名：これらの結果から最終的に採用した酵素名.

ヒット数によりクリーニングした酵素名を使用した場合，方法 1 で取得した 44,965 件中の 43,904 件が一致した. さらにハブとなる酵素が一致した DHI は 7,401 件であった. クリーニング前に比べ，方法 1 で取得した医薬品—HFP ペアと一致す

るペアが増加した (Fig. 6-2). さらにノイズ判定により除去されたノイズ率は, フルネームの場合に 2.4%, 略称の場合に 3.4%, クリーニング後で 1.5%であった. これらの結果から, 酵素と直接的に関連する医薬品および HFPs スクリーニングとした方法 2-1 では, 酵素名をクリーニングする方法を採用した.

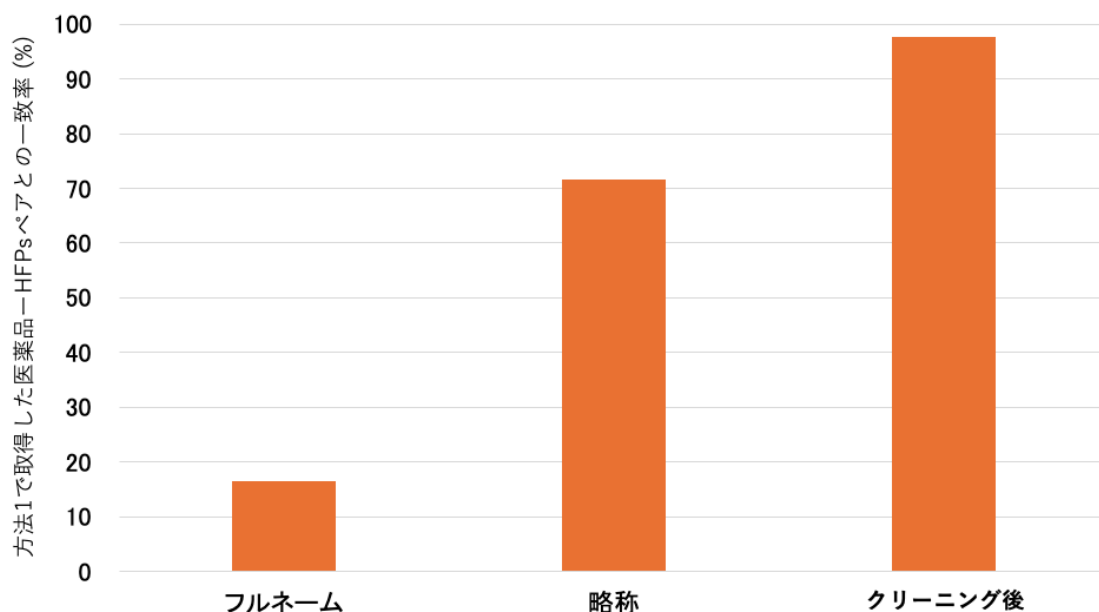


Fig. 6-2 方法 1 と方法 2-1 の医薬品-HFP ペアの一致率: 方法 1 で取得した DHIs の医薬品-HFP ペアと比較し, 一致した医薬品-HFP ペアの数/方法 1 で取得した DHIs 数

## 第 2 項 方法 2-1 の有用性

方法 1 により選抜された 44,965 の DHI に方法 2-1 を適用し, DHI タイプ別に 38 の酵素に分類された 44,965 の DHI を区別した. 方法 2-1 では酵素に直接関連する医薬品と HFP を主として選抜した. すべての医薬品, 酵素, HFP は BioWordVec を用いて数値化した. 数値化された医薬品-酵素と酵素-HFP 間のコサイン類似度を計算し, コサイン類似度が最も高い上位 25%が選抜された. 上位 25%が選ばれたのは, 計算資源と処理速度の制約の中で広範囲に収集できる限界だったからである.

コサイン類似度が設定された閾値より高い医薬品と HFP が選抜され、医薬品-HFP 差分ベクトルを用いてクラスタリングされた。さらに、PubMed を用いた酵素の共起検索で得られた 44,965 個の DHI と一致する DHI と、方法 1 で酵素をハブとして得られた DHI と一致する HFP を選抜した。方法 2-1 で行ったクラスタリングにより分割されたものを「サブクラス」とした。この方法により、7,401 DHIs, 1,081 drugs, 17 subclasses, 28 HFPs に絞られた (Table 6-2)。クラスタリングにより、17 subclasses に分けられたが、酵素のグループとしては 4 グループであった。さらに、一致率を計算した結果、7,401 件の DHIs の中で一致率は 17%であった。このサブクラスごとの一致率を計算したところ、17 サブクラス中 3 サブクラスが 80%以上の一致率を示した (Fig. 6-3)。

Table 2 方法 2-1 で選抜された DHIs

Class	Number of subclasses	DHIs	Number of drugs	Number of HFPs
CBR1	1	1	1	1
CYP2C9	1	383	166	10
CYP2D6	2	1051	325	5
CYP3A4	13	5,966	875	22
Total	17	7,401	1,081	28

ハブとなる酵素ごとに分類した方法 2-1 で選抜された DHIs. Class : DHIs のハブとなる酵素, Number of subclasses : 方法 2-1 でクラスタリングされた DHIs のクラス数, DHIs : Class に含まれる DHIs 数, Number of drugs : Class に含まれる DHIs の医薬品種類数, Number of HFPs : Class に含まれる DHIs の HFPs 種類数, Total : 方法 2-1 でクラスタリングされた DHIs のクラス数, 方法 2-1 で選抜された DHIs 数, 方法 2-1 で選抜された DHIs に含まれる医薬品数および HFPs 数を示す。

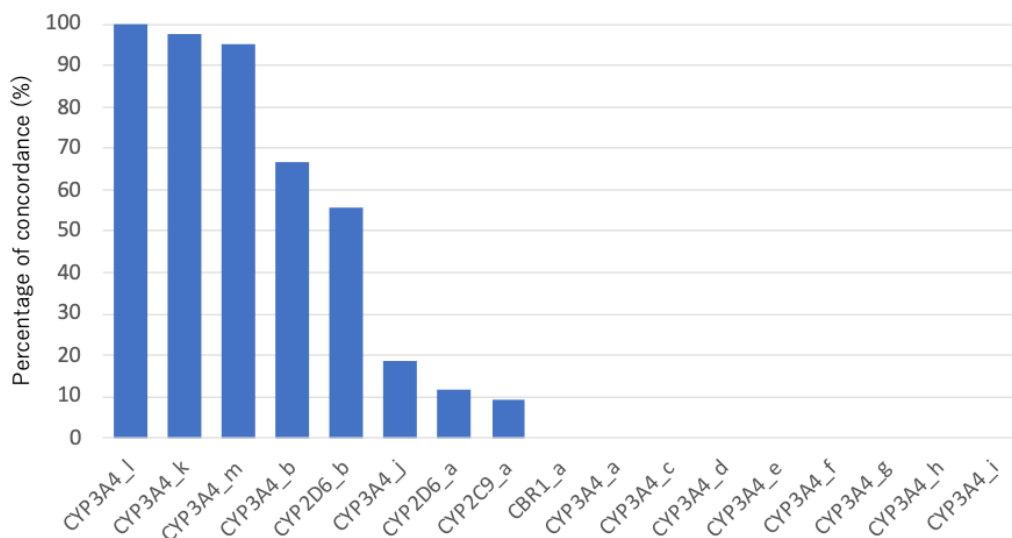


Fig. 6-3 方法 2-1 によるサブクラスの一一致率

方法 2-1 でクラスタリングされた DHIs の 17 クラスを示す。横軸：DHIs のハブとなる酵素名\_サブクラスを識別するためにサブクラスを a から順に文字で示す。これにより 17 サブクラスを識別。縦軸：サブクラスごとに計算したゴールデンスタンダードとの一致率を示す。

### 第 3 節 考察

まず、酵素と直接的に関連する医薬品および HFPs スクリーニングにおいて酵素のフルネームを使用した場合、方法 1 で取得した 44,965 件中の 7,440 件が一致した。さらにハブとなる酵素が一致した DHI は 80 件と非常に少なかった。特に注目したいのは本章の方法で取得した DHIs のハブとなる酵素はほとんどが FMO ファミリーであったことである。医薬品との相互作用で最も知られているのは CYP や UGT ファミリーであり、FMO ファミリーの相互作用は非常に少ない。それにもかかわらず、FMO フ

ファミリーばかりがハブとして出現したのは、ハブとなる酵素の名称に問題があるからではないかと考えた。そこで酵素名を略称に変更し、再び DHIs を確認した。その結果、一致する DHIs 数は大幅に増加した。しかし、CYP や UGT ファミリー以外の酵素の出現率が高かったため、フルネームで使用する酵素と略称を使用する酵素の仕分けが必要であると考えた。

Pubmed ヒット数を元に、単語の示す意味を確認した。ほとんどの酵素では略称が最適であると考えられた。一方で、他の略称と重複している酵素も確認された。それらを略称のまま使用すると他の意味がベクトルに反映されてしまい、本来頻出しない酵素がハブとして頻出する結果となったと考えられる。フルネームを使用した場合の FMO ファミリーが頻出したのは「flavin-containing monooxygenase」の「oxygenase」部分の意味が強く反映されてしまったからであると考えられる。これらの点を踏まえ、クリーニングした名称が最適であった。

さらに、クリーニング前後の医薬品—HFP ペアの一致率やノイズ率を比較したところ、クリーニング後が一致率は最も高く、ノイズは最も少なくなり、最適であると示唆された。このような名称のクリーニングがベクトル化には非常に重要であると考えられる。

方法 2-1 により、7,401 件の DHIs を選抜できたが、HFPs および酵素のバリエーションが非常に少ないため、更なる方法が必要だと考えられる。DHIs が絞られたのは主に HFPs および酵素の数が絞られたことが原因と考えられる。1 つだけ選抜された CBR1 の DHIs は Vitamin K と Tetrabenazine のペアであり、これらの相互作用は知られていないが、いずれも CBR1 に関与することは知られており、相互作用する可能性があると考えられる (Vanessa Gonzalez-Covarrubias et al. 2008; Joseph Jankovic et al. 2011)。この研究で利用したゴールドスタンダードデータには 1,205 種の医薬品、83 種の HFPs、21 種類の酵素含まれており、方法 2-1 で選抜された HFPs と重複した HFPs は 9 種であった。その中で一致率が 80%以上のサブクラスが存在したことから、方法 2-1 によって選抜された DHIs は相互作用する可能性がある



考えられる。一致率が低いサブクラスが出た原因として、相互作用が既知であるにもかかわらず、ゴールドスタンダードに含まれていないか、あるいはまだ十分に研究されていない相互作用、すなわち未知の相互作用が含まれているためである。まだ解明されていない相互作用が存在する可能性があるため、今後の研究で、これらのサブクラス内で選抜された DHI の信頼性をさらに調査する価値があると考えられる。

## 第 7 章 方法 2-2：酵素と間接的に関連する医薬品および HFPs スクリーニング

方法 2-1 では、方法 1 で選抜された DHIs を医薬品と HFPs の関係性ごとにクラスタリングできたが、バリエーションが低いことが課題であった。そのため、前処理の方法を検討した。方法 2-1 では酵素とのコサイン類似度が高い医薬品および HFPs を選抜した。これらは酵素と類似性が高く、直接関連するとされる医薬品と HFPs が選抜されている。そこで、方法 2-2 では間接的に関連するとされる医薬品および HFPs を選抜する前処理を考案した。「医薬品と酵素の差分ベクトル」と「HFPs と酵素の差分ベクトル」のコサイン類似度が高いものを前処理として選抜した。これにより、酵素との関わり方が似ている医薬品と HFPs を選抜することができると考えた。

### 第 1 節 方法

#### 第 1 項 酵素と間接的に関連する医薬品および HFPs スクリーニング

酵素と間接的に関連する医薬品および HFPs を選抜するために、第 5 章第 1 節第 4 項および第 5 項と異なる選抜方法を考案した。第 5 章第 1 節第 4 項および第 5 項とは異なる酵素との関連性を持つ医薬品および HFPs を選抜するために、医薬品－酵素の差分ベクトルおよび HFPs－酵素の差分ベクトルを構築した。これらの構築には第 5 章第 1 節第 3 項で数値化した HFPs、医薬品および酵素を使用した。構築した差分ベクトル間のコサイン類似度を算出し、コサイン類似度が高い上位 25 %を選抜されるようにその閾値を設定し、酵素と関連する医薬品および HFPs として選抜した。酵素との差分ベクトルは酵素との関係性を示す。選抜した医薬品と HFPs に対し、方法 2-1 と同様の方法を用い、DHIs を抽出した。方法 1 により酵素で分類されたグループをクラスとし、差分ベクトルでクラ

スタリングされたクラスをサブクラスとした。

## 第 2 節 結果

### 第 1 項 酵素の名称による数値化への影響

前章と同様に酵素名の違いによる影響を確かめた。まず、フルネームを使用した場合、前章よりもノイズが増加した (Fig. 7-1)。方法 1 で取得した DHIs と比較して一致した医薬品-HFP ペア数を計測した。この結果、方法 1 で取得した 44,965 件中の 44,965 件が一致した。さらにハブとなる酵素が一致した DHI は 2,846 件であった。前章と同様に FMO ファミリーが間接席に関連する酵素として頻出した。

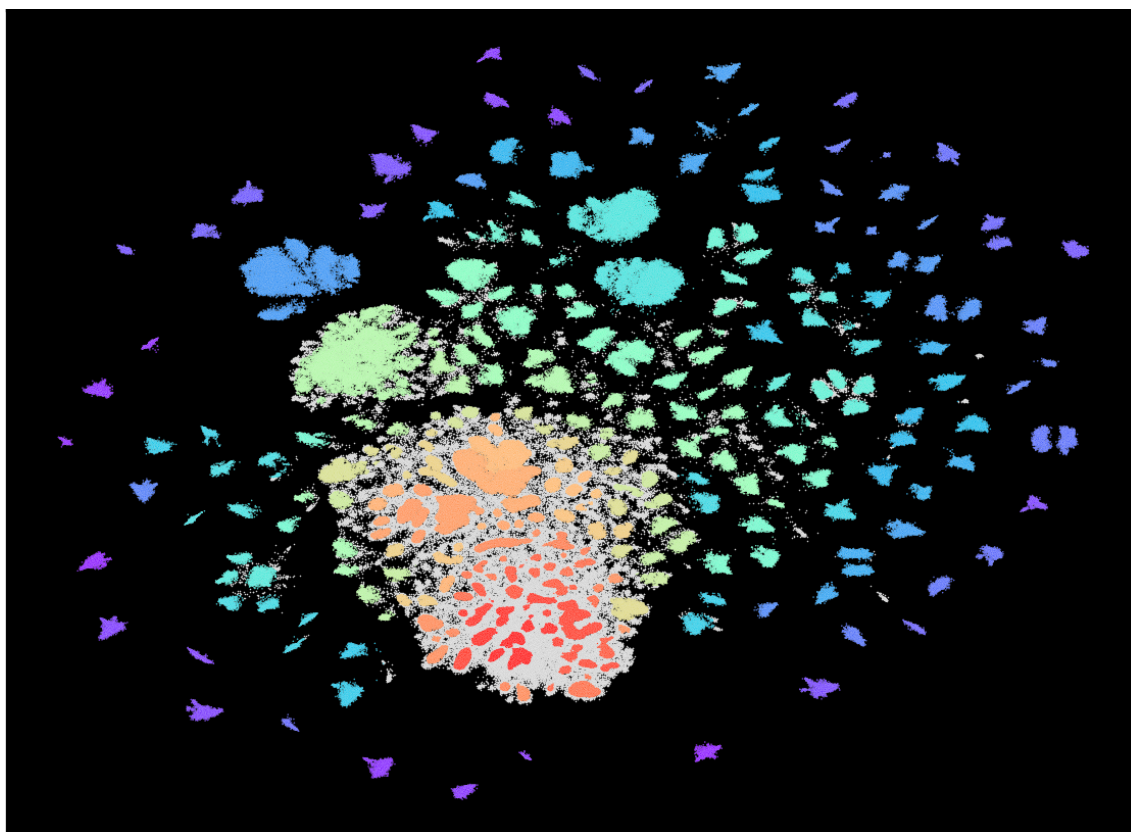


Fig. 7-1 酵素と間接的に関連する医薬品および HFPs スクリーニングにおいて酵素のフルネームを使用した場合にクラスタリングされた医薬品と HFPs の差分ベクトルの

## 可視化した図

続いて、酵素名を略称にした場合においても方法 1 で取得した 44,965 件中の 44,965 件が一致した。さらにハブとなる酵素が一致した DHI は 10,502 件であった。こちらも同様にフルネームを使う場合に比べ、非常に一致する DHIs が増加した。一方で、間接的に関連するとされる酵素の数が非常に多かった。1DHIs あたりに関連している酵素数は、44,965 件中フルネームの時には平均 7.2 個、略称で平均 20.5 個であった。ハブとなる酵素も一致している DHIs 中ではフルネームで平均 32.4 個、略称では平均 52.5 個であった。略称の方が関連酵素数は多くなった。

前章と同様に、使用する酵素名のクリーニングを行った。その結果、ノイズ判定により除去されたノイズ率は、フルネームの場合に 20.9%、略称の場合に 20.9%、クリーニング後で 15.9%であった。これらの結果から、酵素と間接的に関連する医薬品および HFPs スクリーニングとした方法 2-2 では、酵素名をクリーニングする方法を採用した。

## 第 2 項 方法 2-2 の有用性

方法 2-1 と同様に、方法 2-2 で行ったクラスタリングによって分けられたものを「サブクラス」とした。この方法により、方法 1 で選抜された 44,965DHIs が 2,819 DHIs, 958 drugs, 71 classes, 150 HFPs に絞られた (Table 7-1)。クラスタリングにより、71 subclasses に分けられたが、酵素のグループとしては 31 グループであった。DHIs が 1 件であった AOX1, CYP4A11, CYP4F2, UGT2B15 について、これらの相互作用の可能性を確認したところ、3 つは相互作用の情報は確認できなかったが、Lorazepam と Steroids の相互作用は示された ([https://www.kegg.jp/medicubin/japic\\_med?japic\\_code=00062316](https://www.kegg.jp/medicubin/japic_med?japic_code=00062316), Hui Shen et al. 2005, Table 7-2)。さらに、方法 1 と同様に一致率を算出したところ、2,819 件の DHIs のうち 30%であった。このサブクラスについても一致率を計算したところ、71 サブクラス中 10 サブクラス

が 80%以上であった (Fig. 7-2). 一致率が 80%以上であったサブクラスについて含まれる DHIs を確認した (Table 7-3).

Table 7-1 方法 2-2 で選抜された DHI

Class	Number of subclasses	DHIs	Number of drugs	Number of HFPs
ACHE	1	111	1	111
ADA	1	38	1	38
AOX1	1	1	1	1
BCHE	1	27	1	27
COMT	1	264	18	27
CYP1A1	1	27	13	18
CYP1A2	6	150	106	19
CYP2A6	1	4	4	2
CYP2B6	1	14	14	2
CYP2C19	3	45	41	7
CYP2C8	1	4	4	3
CYP2C9	3	106	82	10
CYP2D6	4	19	19	7
CYP2E1	1	6	6	5
CYP3A4	27	1,620	763	35
CYP3A5	2	93	81	3
CYP3A7	1	39	39	1
CYP4A11	1	1	1	1
CYP4F2	1	1	1	1
DPYD	1	2	2	2

FMO3	1	3	3	1
GST	1	106	1	106
MAOA	1	23	14	6
MAOB	1	14	10	5
NAT1	1	4	3	3
NAT2	1	4	4	3
UGT	2	79	44	10
UGT1A1	1	8	6	4
UGT2B15	1	1	1	1
UGT2B7	1	3	3	1
XDH	1	2	1	2
<b>Total</b>	<b>71</b>	<b>2,819</b>	<b>958</b>	<b>150</b>

方法 2-2 によって選抜された DHI を, DHI のハブである酵素によって分類. Class : DHI のハブである酵素. Number of subclasses : 方法 2-2 によってクラスタ化された DHI のクラス数. DHI : クラスに含まれる DHI の数. Number of drugs : クラス内の DHI に含まれる医薬品数. Number of HFPs : HFP の数 : クラス内の DHI に含まれる HFP の数. Total : 方法 2-2 でクラスタ化された DHI のクラス数, 方法 2-2 で選抜された DHI の数, 方法 2-2 で選抜された DHI に含まれる医薬品の数, HFP の数.

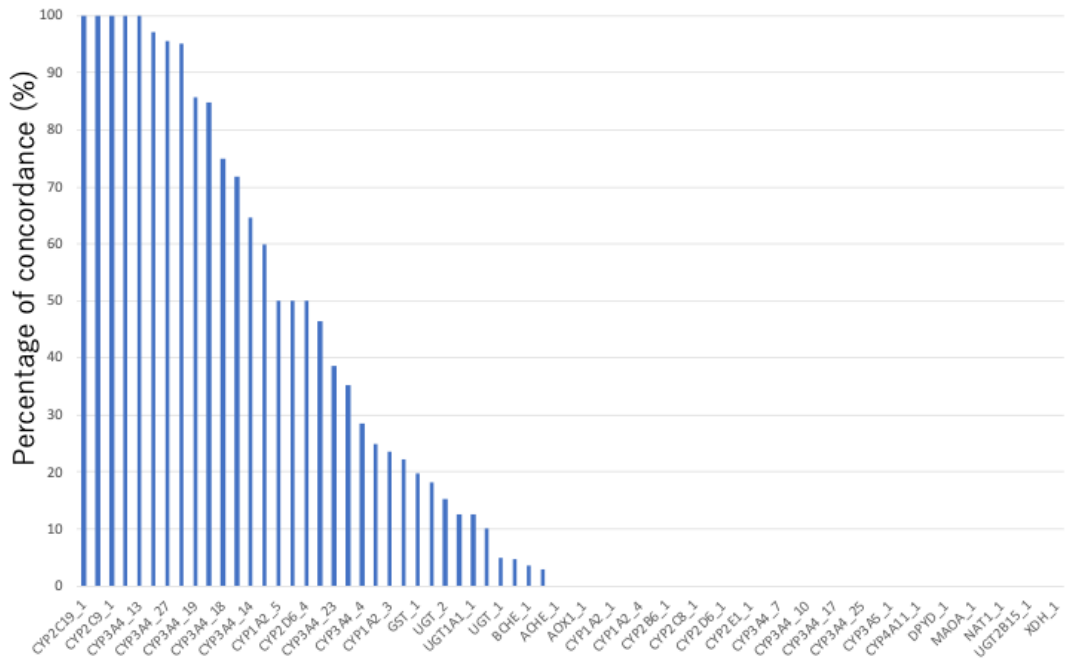


Fig. 7-2 方法 2-1 によるサブクラス的一致率

方法 2-1 でクラスタリングされた DHIs の 17 クラスを示す。横軸：DHIs のハブとなる酵素名\_サブクラスを識別するためにサブクラスを a から順に文字で示す。これにより 17 サブクラスを識別。縦軸：サブクラスごとに計算したゴールデンスタンダード

Table 4 方法 2-2 で選抜された DHI が最も少ない酵素クラス

Class	Subclass	Drug ID	drug name	HFPs ID	HFPs name
AOX1	1	D00530	Zaleplon	HF160	Saccharomyces cerevisiae, Japanese sake yeast
CYP4A11	1	D10476	Lusutrombopag	HF14	Arachidonic acid
CYP4F2	1	D10547	Brincidofovir	HF247	Vitamin K
UGT2B15	1	D00365	Lorazepam	HF313	Steroids

方法 2-2 で選抜された DHIs においてハブとなる酵素ごとのクラスで, DHIs 数が 1 であった酵素クラスを選抜して示す. Class : ハブとなる酵素名, Subclass : サブクラスを識別するためにサブクラスを 1 から順に数字, Drug ID : DHIs の医薬品の KEGG の識別 ID, drug name : DHIs の医薬品名, HFPs ID : 作成した HFPs リストに振った独自の HFPs の ID, HFPs name : DHIs の HFPs 名を示す.

Table 7-3 方法 2-2 のサブクラス CYP3A4\_19 に含まれる DHIs

drug number	drug name	HFP id	HFP name
D00210	Norfloxacin	HF224	Dehydroepiandrosterone
D00211	Rifampin	HF224	Dehydroepiandrosterone
D00255	Carvedilol	HF224	Dehydroepiandrosterone
D00295	Cimetidine	HF224	Dehydroepiandrosterone
D01293	Ethyl loflazepate	HF224	Dehydroepiandrosterone
D01886	Hydrocortisone probutate	HF224	Dehydroepiandrosterone
D06405	Nalfurafine hydrochloride	HF224	Dehydroepiandrosterone
D06876	Hydrocortisone aceponate	HF224	Dehydroepiandrosterone
D07520	Bepiridil	HF224	Dehydroepiandrosterone



D07667	Cevimeline	HF224	Dehydroepiandrosterone
D07868	Domperidone maleate	HF224	Dehydroepiandrosterone
D07913	Escitalopram	HF224	Dehydroepiandrosterone
D08046	Hydrocodone hydrochloride	HF224	Dehydroepiandrosterone
D08095	Ivabradine hydrochloride	HF224	Dehydroepiandrosterone
D08456	Quetiapine	HF224	Dehydroepiandrosterone
D10542	Alectinib	HF211	Quercetin
D10688	Abemaciclib	HF211	Quercetin
D10893	Acalabrutinib	HF211	Quercetin
D10994	Upadacitinib	HF211	Quercetin
D11040	Apalutamide	HF211	Quercetin
D11460	Siponimod	HF211	Quercetin

方法 2-2 で分類した CYP3A4\_19 サブクラスに含まれる DHI のうち、一致率が 80%を超えるもの。医薬品 ID : DHI に含まれる医薬品の KEGG の識別 (ID) , 医薬品名 : DHI に含まれる医薬品名, HFP 名 : DHI に含まれる HFP の ID, HFP 名 : DHI ID に含まれる HFP 名, 医薬品名 : DHI に含まれる医薬品名, HFP ID : HFPID : 作成された HFP リストに割り当てられたユニークな HFP の ID, HFP name : DHI の HFP の名前。

### 第 3 節 考察

#### 第 1 項 方法 2-2 の有用性

この研究で利用したゴールドスタンダードデータには 1,205 種の医薬品, 83 種の HFPs, 21 種類の酵素含まれており, 方法 2-2 で選抜された HFPs と重複した HFPs は 38 種であった。方法 2-1 と同様に, その中で一致率が 80%以上のサブクラスが存在したことから, 方法 2-2 によって選抜された DHIs も相互作用する可能性があ

ると考えられる。

## 第2項 方法 2-1 と方法 2-2 の比較

方法 1 は 44,965 個の DHIs を選抜し、方法 2-1 は 7,401 個の DHIs を選抜し、方法 2-2 は 2,819 個の DHIs を選抜した。方法 1 は、単一の記事内で同時に頻繁に発生する HFPs と酵素を抽出して、幅広い DHI を選抜した。しかし、方法 1 だけでは DHIs を区別することができず、さらなる分析は不可能であったため、追加の方法が必要だった。方法 2-1 と 2-2 は多くの種類の酵素と HFPs を残したが、方法 2-2 の方が方法 2-1 よりも DHIs が少なかった。方法 2-1 は、鉄やカルシウムなどの酵素と反応することが知られている HFPs や、医薬品に関連する特定の酵素との相互作用が既知である HFPs を特に選抜した。これにより、方法 2-1 で選抜された DHIs に対する酵素の一致が多くなり、全体的な DHI の数が増加した。これは以前の酵素-HFP の相互作用に関する研究を反映している可能性が高い。一方、方法 2-2 は、医薬品の代替物として研究されている HFPs、医薬品の構成要素や薬理学的な物質に分類される HFPs、HFP 内の成分が有効成分として研究されている HFPs を選抜した。医薬品と同様の作用をする可能性が高い HFPs が選抜されるが、それらが酵素との関係が不明である。したがって、方法 2-2 で選抜された DHIs は、共起検索の中心だった酵素との一致が少なく、全体的な DHI の数が減少した。一方、医薬品と同様の作用を研究しているため、多くの酵素と関連付けられる可能性が高く、酵素の種類が多くなる。この違いは方法 2-1 と方法 2-2 の選抜された DHIs の種類と数に影響を与える可能性があり、それが HFPs に影響を与えたかもしれない。

医薬品への影響が少なかったのは、多くの医薬品が CYP3A4 との相互作用が知られており、方法 2-1 および 2-2 の両方で CYP3A4 が残っていたからだと考えられる。さらに方法 2-1 と重複した DHIs は 238 件であり、合計 9982 件の DHIs が選抜された。重複した DHIs の件数から考えると酵素と医薬品、酵素と健康食品の差分ベクトルのコサイン類似度で選抜することにより、酵素との関係の類似性で選抜し、

方法 2-1 とは異なる DHI をクラスタリングすることができたと考えられる。このように方法 2-1 と方法 2-2 を組み合わせることで DHI のバリエーションを増やすことができると考えられる。

方法 1 により選抜された DHI に対して、追加して方法 2-1 および方法 2-2 を行っているが、方法 2-1 のみを行った場合の一致率は 14%であった。そして方法 2-2 のみを行った場合の一致率は 13%であった。いずれも方法 1 に追加して行う方が単独よりも一致率は高い。したがって、方法 1 に追加して方法 2-1 および方法 2-2 を行う必要があると考えられる。

方法 1 ではグループ内の平均 DHI は 1183 件、方法 2-1 のサブクラス内の平均 DHI は 435 件、方法 2-2 のサブクラス内の平均 DHI は 40 件であった。クラス内の DHI が少なければ、DHI の調査が容易になる。抽出した DHI を調査するには方法 2-2 が最も適していると考えられる。

さらに方法 1 で選抜された 44,965 件の DHI のうち、21,005 件(47%)の DHI が PubMed Hits 20 以上 30 未満であった。方法 2-1 で選抜された 7,401 件の DHI のうちの 3,822 件(52%)、方法 2-2 で選抜された 2,819 件の DHI のうちの 985 件(35%)の DHI が PubMed Hits 20 以上 30 未満であった。方法 2-1 および方法 2-2 を方法 1 に追加して行っても PubMed Hits 20 以上 30 未満の DHI を保持することができた。このように今回の手法では共起検索で共起ランクが低い DHI を選抜することが可能であると考えられる。

これらの結果から、まず方法 1 により DHI を選抜し、この DHI に対して方法 2-1 および方法 2-2 の 2 パターン行う方法が医薬品と HFPs の相互作用を抽出するワークフローとして最適であると考えられる (Fig. 7-3)。この複合アプローチの利点は、方法 1 の共起検索だけでは特定できなかった異なる関係の関連性や相互作用を包括的に考慮できることである。また、一致率の高いクラスターを優先することで、より信頼性の高い DHI を同定し、効率的に解析することができる。このワークフローは将来の研究においても応用が可能であり、他のデータセットに対して同様

の分析を適用することで、新たな関係性や相互作用を発見する可能性がある。

他のツールでは、医薬品や食品に関する情報から医薬品と食品の相互作用を抽出するのが一般的であるが、我々の研究では、酵素を中心ハブとして医薬品と食品の相互作用を抽出することに焦点を当てている。その結果、我々の結果を他のツールの結果と比較することは困難である。

### 第3項 ゴールドスタンダードとの比較

ゴールドスタンダードとして使用される文献には、一定レベルの網羅性と正確性が求められる。本研究で使用した資料は、他の資料と比較して網羅性と正確性に優れていることから選択された。網羅性と正確性の基準は、データの使用目的のレベル（臨床的な意思決定プロセスをサポートするかどうか）、同じDHIを検索した場合の情報の有無などの要素に基づいている。しかし、このような資料であっても、網羅性という点では限界があることは認識している。とはいえ、本研究におけるゴールドスタンダードの役割は検証目的であるため、これらの限界は許容範囲と考える。

DHIとゴールドスタンダードの一致率は低く、方法1では30%、方法2-1では17%、方法2-2では30%であった。この理由として考えられるのは、確定的でないDHIが含まれる可能性や、ゴールドスタンダードによる制限などである。特定のHFPに関する相互作用情報のみが含まれ、他のHFPに関して評価することができないため、一致率は低くなる傾向にある。ゴールドスタンダードに含まれるHFPに関連するDHIのみについて一致率を算出したところ、方法1、方法2-1、方法2-2で一致率は80%以上であった。

本研究において、ゴールドスタンダードのDHIsとワークフローで取得したDHIsに関する重要な観察がある。我々は、ゴールドスタンダードのDHIsとワー

クフローの DHIs の一部の DHIs が重複していることを確認した。例えば、医薬品との相互作用が広く知られているグレープフルーツ、緑茶、コーヒーとの相互作用情報を抽出した。このような実例を通して、我々の手法が実際の医薬品-HFP 相互作用情報を有用かつ信頼性の高い方法で提供できることを示した。しかし、ゴールドスタンダードに記載されているが、ワークフローで取得できなかった DHIs も存在する。この DHIs の中には方法 1 の共起検索で Pubmed hits 1 である HFPs と酵素から抽出される DHIs が含まれている。CYP ファミリーの研究は 1985 年ごろが盛んであったが、その年代の文献の中には pdf で保存されていたり、アブストが pubmed に記載されていない文献があった(Nebert and Gonzalez, 1987)。これらが PubMed における共起検索の弊害となり、PubMed hits が低くなったと考えられる。このように共起検索の結果、pubmed hits 20 以上の酵素と HFPs を選抜することにより、抽出できなかった DHIs も存在する。しかしながら、PubMed hits で選抜圧をかけなければ、非常に多くのノイズが生まれる。将来の研究では、これらのゴールドスタンダードに記載されているが、ワークフローで取得できなかった DHIs を取得するための新しい方法やアプローチを検討し、データセットの網羅性を向上させることが重要であると考えられる。

また、今回の研究で特に検出が困難であった DHI についても説明する。方法 1 で検出できなかった HFP の中には、有効性が未確定のまま研究段階にあるものが数多くある。例えば、ピニトール、パセリ、ニコチンアミドリボシドなどである。今後の研究で、これらの DHI もより効果的に検出できるようになるかもしれない。方法 1 では検出されたが方法 2-2 では除外された DHI については、これらの物質の中には、長い間研究され、安全性が確立され、多くの研究で使用されているものがある。例えば、ゼラチン、大麦、モリブデンなどである。これらの物質の多くは、研究の主眼ではなく、むしろ標準条件または培地成分として一般的に使用されているため、方法 1 に選択された可能性がある。

さらにゴールドスタンダードに記載されていないが、ワークフローで取得さ

れた DHIs も存在する。これらの DHIs においては、酵素と HFPs の研究は進んでいるが、医薬品と HFPs の研究は進んでいないことが原因として考えられる。これらの DHIs に関して予測モデルを作成すれば、新たな DHIs の予測が可能になると考えられるため、今後、研究を進めていく必要がある。一般化と頑健性については問題があるが、提案された手法はデータ不足を補うものである。新しいデータを生成し、それを拡張し、予測モデルを構築するというアプローチである。この研究では、予測モデルの基礎となるデータを生成する。将来的には、これらのデータを拡張し、汎化と頑健性に関する問題を考慮しながら、より強力な学習モデルを構築することを目指している。

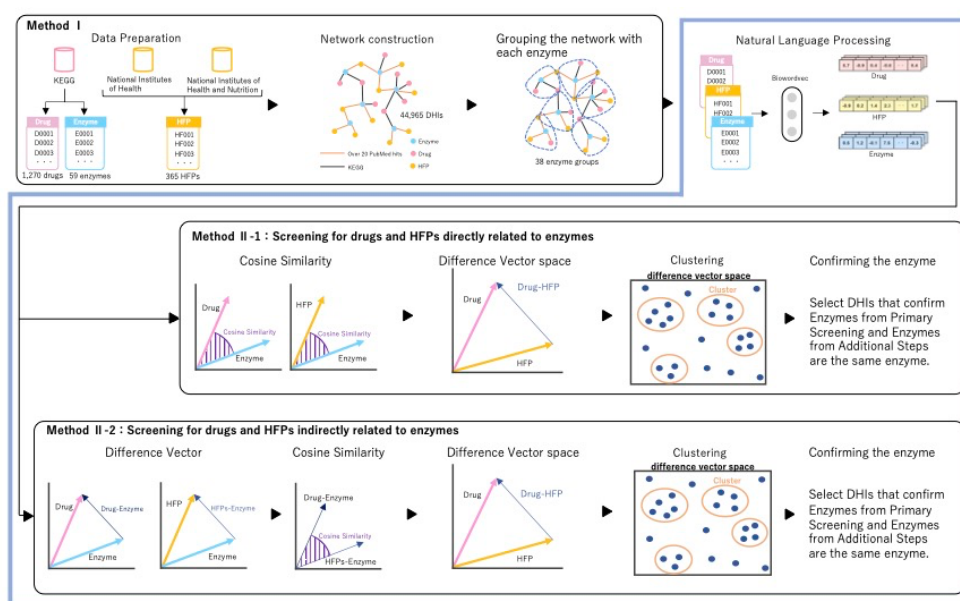


Fig. 7-3 最適化された方法のフレームワーク図

方法 1：まず、KEGG から医薬品リストと酵素リストを取得し、NIH の Dietary Supplement Fact Sheets と医薬基盤・健康・栄養研究所の素材情報データベースから HFPs のリストをそれぞれ取得する。KEGG から医薬品と酵素の相互作用情報を取得した。次に、取得したすべての酵素と HFPs をペアで組み合わせ、このペアを PubMed に入力して、出版物における共起の数を取得します。次に、データの品質

を向上させるために、見つかった酵素と HFPs のペアをさらにフィルタリングするため pubmed hits 20 以上の酵素と HFPs のペアを選抜した。選抜された酵素と HFPs のペアと KEGG から取得した医薬品と酵素のペアを使用してネットワークを構築する。医薬品—酵素—HFP の三者の関係から、相互作用すると考えられる医薬品と HFPs を DHIs とする。方法 2-1：方法 1 で取得した医薬品、酵素、HFPs のリストを biowordvec を用い、数値化した。取得した医薬品ベクトルと酵素ベクトルのコサイン類似度を算出し、取得した HFPs ベクトルと酵素ベクトルのコサイン類似度を算出した。酵素ベクトルとのコサイン類似度が高い医薬品および HFPs を選抜し、医薬品ベクトルと HFPs ベクトルの差分ベクトルを算出し、クラスタリングした。クラスタリングにより得られた DHIs と方法 1 で選抜された DHIs と比較し、重複した DHIs を選抜した。選抜された DHIs において、方法 1 でハブとなった酵素と高いコサイン類似度を示した酵素が重複した DHIs を取得した。方法 2-2：方法 2-2 と同様に方法 1 で取得したリストを数値化した。医薬品ベクトルと酵素ベクトルの差分ベクトルを算出し、HFPs ベクトルと酵素ベクトルの差分ベクトルを算出し、これら算出した差分ベクトルのコサイン類似度を算出した。コサイン類似度が高い差分ベクトルを示す HFPs および医薬品を選抜し、医薬品ベクトルと HFPs ベクトルの差分ベクトルを算出し、クラスタリングした。クラスタリングにより得られた DHIs と方法 1 で選抜された DHIs と比較し、重複した DHIs を選抜した。選抜された DHIs において、方法 1 でハブとなった酵素と高いコサイン類似度を示した酵素が重複した DHIs を取得した。

## 総括

HFPs と医薬品の相互作用は、消費者の健康に重大な影響を及ぼす可能性があるものの、情報の抽出は難しい。本研究では情報科学の技術を用い、この相互作用の効率的な抽出を目的とした新しいワークフローを開発した。特に代謝酵素を中心としたネットワーク分析を活用し、HFPs と医薬品の潜在的な相互作用を詳細に調査した。

医薬品と酵素については情報が充実しており、データベースも存在するが、HFPs のみが集約されているデータベースはほとんどない。本研究では二つの情報源からデータを精査して収集することで 365 種の HFPs リストを得た。

酵素と HFPs の相互作用情報抽出の有用性検証のために、BioGPT を用いて医薬品と HFPs の相互作用情報抽出を試みた。しかしながら、現時点の BioGPT および環境ではこれらの情報抽出には不適當であった。続いて、医薬品と HFPs、酵素と HFPs の相互作用情報抽出において、共起検索の有用性を検証した。この結果、医薬品と HFP の相互作用情報抽出には医薬品—酵素ペアおよび酵素—HFP ペアの関連情報を抽出する方法がより網羅的な情報を抽出することができると示唆された。さらに、PubMed Hits の閾値の条件検討を行った結果、閾値を 20 とすることで、関連性の高い酵素—HFPs ペアの抽出を可能になった。これにより、44,965 件の DHIs を取得した。

これらの結果の評価方法の検討を行ったところ、書籍情報を基にゴールドスタンダードを作成し、一致率により評価する方法が最適である。今後は、文献に記載されている内容から医薬品と HFP の相互作用情報を抽出するツールの作成を進めていきたい。

取得した DHIs の関係性を可視化するためにネットワーク図を作成するツールである NetworkX, pyvis, cytoscape を用いた。DHIs の関係性ごとの可視化について、それぞれのツールで条件検討を行った結果、今回の可視化には Cytoscape が最適なツールであった。この Cytoscape を用いて可視化したところ、DHIs はハブとなる酵素ごとにグループに分けられた。この一連の流れを方法 1 とし、最終的に 38 種の酵素、157 種の HFP、1256 種の医薬品及び 44,965 件の DHI が抽出された。



さらに DHIs の関係性ごとに分類および精査するために、二次スクリーニングを行った。はじめに、DeepWalk を用いて DHIs の分類および精査を試みた。ネットワークのトポロジーとして分類はできたが、相互作用としての分類は困難であった。そこで BioWordVec を用いて DHIs の分類および精査を試みた。医薬品と HFP の差分ベクトルのみの計算では有用性は見出されなかったが、酵素との関連性で医薬品および HFP をスクリーニングしたのちに医薬品と HFP の差分ベクトルを算出する方法では有用性が示唆された。また、このプロセスで使用する酵素名の条件を検討し、クリーニング後の名称が最適となった。酵素と直接的に関連する医薬品と HFP をスクリーニングする方法を方法 2-1 とし、方法 1 で取得した DHIs について方法 2-1 を行った結果、4 種の酵素グループ、17 サブクラス、28 種の HFPs、1,081 種の医薬品、7,401 件の DHIs に絞られた。このサブクラスについても一致率を計算したところ、17 サブクラス中 3 サブクラスが 80%以上であった。さらに間接的に関連している医薬品と HFPs をスクリーニングする方法である方法 2-2 を方法 1 に適用すると、31 種の酵素グループ、71 サブクラス、150 種の HFPs、958 種の医薬品、2,819 件の DHIs に絞られた。このサブクラスについても一致率を計算したところ、71 サブクラス中 10 サブクラスが 80%以上であった。方法 2-1 と方法 2-2 で取得した DHIs の重複は非常に少なく、いずれの手法も必要であると考えた。方法 2-1 では医薬品に関連する特定の酵素との相互作用が既知である HFPs を特に選抜した。方法 2-2 は、医薬品の代替物として研究されている HFPs、医薬品の構成要素や薬理学的な物質に分類される HFPs、および HFP 内の成分が有効成分として研究されている HFPs を選別した。方法 1 を行わずに、方法 2-1 および方法 2-2 のみで DHIs を選抜する方法では、ハブとなる酵素の検証を行うことができず、有用性は低くなった。データセットの網羅性やアプローチの有用性をゴールドスタンダードとの一致率で評価し、最適化されたフレームワークを作成した。このフレームワークを用いることで、これまで抽出が困難であった情報についても抽出可能性が見出され、他の情報抽出にも役立つと考えられる。さらに医薬品と HFPs の相互作用情報抽出を可能にしたため、取得した DHIs を基に予測モデルを構築し、

新たな DHIs の予測の可能性が見出された。今後は、このワークフローを他のデータセットに適用し、さらに広範な関係性の抽出を目指す。また、ゴールドスタンダードには記載されていない DHIs に関する予測モデルの開発も研究の方向性として検討していくことで、まだ見つかっていない医薬品と HFPs の相互作用に関する知識の拡充に寄与することが期待される。

本研究を通じて、筆者はウェット研究の経験を持ちながらドライ研究に挑戦した。ドライ研究を進める中で、ツールや解析手法は充実しているものの、基礎データが不足しているために、関係性や原因を十分に解明できないという課題に気づいた。さらに、データフォーマットの不統一や、ネガティブデータの共有が進まない現状も障壁となっている。特にネガティブデータについては、研究者が自身の研究活動を証明し、他者によるデータの盗用や成果の横取りを防ぐために公開が難しいという問題がある。このような課題があるものの、今後これらの障壁を克服し、データの集積と解析を進めることができれば、日本の生物工学分野のさらなる発展が期待される。ドライ研究ではツールは十分に整備されているが、情報の不足が課題であり、これを解決するためには、ウェット研究とドライ研究の双方が歩み寄り、データ共有の標準化と課題解決を目指した協力が不可欠である。互いの立場や視点を尊重し、連携を深めることが今後の課題を解決する鍵となるだろう。

## 謝辞

本研究の遂行にあたり、多くの方々からご支援とご協力を賜りました。心より感謝申し上げます。

まず、医薬基盤・健康・栄養研究所の荒木先生、研究員の厨さん、山本さん、渡邊さんをはじめとして皆様に深く感謝申し上げます。先生方の専門知識と助言は、この研究を成功に導く大きな要因でした。研究所の皆様の支援と協力に心から感謝いたします。

指導教員である松木先生には、貴重なアドバイスと熱心な指導をいただきました。先生のお陰で、本研究が成功裏に進行することができました。また、事務の皆様にも研究の手続きや資料整備において、大変お世話になりました。

また、阪大微研の阿部先生、医薬基盤・健康・栄養研究所の李さん、黒田さんや協力者の皆様にも感謝の意を表します。皆様の協力により、研究を円滑に進行し、有益な結果を得ることができました。

最後に、友人への感謝の意を述べるとともに、研究活動において支えてくれたすべての方々に心から感謝申し上げます。

皆様のご支援とご協力のおかげで、この研究が成功に導かれました。心より感謝申し上げます。

吉崎万莉

## データの利用可能性に関する記述

本研究で利用したソフトウェアとデータ資源は以下の通り：

- Python 3.9.16 (<https://www.python.org/>)
- BioWordVec (Code: <https://github.com/ncbi-nlp/BioSentVec>, Model: [https://ftp.ncbi.nlm.nih.gov/pub/lu/Suppl/BioSentVec/BioWordVec\\_PubMed\\_MI\\_MICIII\\_d200.bin](https://ftp.ncbi.nlm.nih.gov/pub/lu/Suppl/BioSentVec/BioWordVec_PubMed_MI_MICIII_d200.bin))
- fastText 0.9.2 (<https://fasttext.cc/>)
- UMAP (umap-learn) 0.5.3 (<https://umap-learn.readthedocs.io/>)
- HDBScan 0.8.30 (<https://hdbscan.readthedocs.io/>)

パフォーマンス評価 性能評価には BioWordVec の訓練済みモデルを採用した。性能評価指標の詳細については、以下のウェブサイトをご参照ください：

[[BioWordVec Performance Metrics](https://github.com/ncbi-nlp/BioSentVec#biowordvec)](<https://github.com/ncbi-nlp/BioSentVec#biowordvec>).

私たちの研究で使用された主要なパフォーマンス指標の1つは、一致率(%)であり、以下のように計算される： $\text{Concordance Rate (\%)} = (\text{Drug-HFP interactions (DHIs) concordance with the gold standard} / \text{Total DHIs}) \times 100$ .

また、以下の外部リソースを活用した：

- Kyoto Encyclopedia of Genes and Genomes (KEGG) URL: <https://www.genome.jp/kegg/drug/br08309.html>

References: Kanehisa et al., 2000; Kanehisa et al., 2017; Tanabe et al., 2012

- Dietary Supplement Fact Sheets URL: <https://ods.od.nih.gov/factsheets/list-all/>
- PubMed URL: <https://pubmed.ncbi.nlm.nih.gov/>

## 参考文献

- Almazroo, O. A., Miah, M. K., & Venkataramanan, R. (2017). Drug Metabolism in the Liver. *Clinics in Liver Disease*, 21(1), 1–20. <https://doi.org/10.1016/j.cld.2016.08.001>
- Asher, G. N., Corbett, A. H., & Hawke, R. L. (2017). Common Herbal Dietary Supplement-Drug Interactions. *American Family Physician*, 96(2), 101–107.
- Campello, R. J. G. B., Moulavi, D., & Sander, J. (2013). Density-Based Clustering Based on Hierarchical Density Estimates. In *Advances in Knowledge Discovery and Data Mining* (pp. 160–172).
- Chan, P., Tomlinson, B., Chen, Y. J., Liu, J. C., Hsieh, M. H., & Cheng, J. T. (2000). A double-blind placebo-controlled study of the effectiveness and tolerability of oral stevioside in human hypertension. *British Journal of Clinical Pharmacology*, 50(3), 215–220. <https://doi.org/10.1046/j.1365-2125.2000.00260.x>
- Dwyer, J. T., Coates, P. M., & Smith, M. J. (2018). Dietary Supplements: Regulatory Challenges and Research Resources. *Nutrients*, 10(1), 41. <https://doi.org/10.3390/nu10010041>
- Fleuren, W. W., & Alkema, W. (2015). Application of text mining in the biomedical domain. *Methods*, 74, 97–106. <https://doi.org/10.1016/j.ymeth.2015.01.015>
- Garcia-Cazarin, M. L., Wambogo, E. A., Regan, K. S., & Davis, C. D. (2014). Dietary supplement research portfolio at the NIH, 2009–2011. *The Journal of Nutrition*, 144(4), 414–418. <https://doi.org/10.3945/jn.113.189803>

Gaudry, A., Huber, F., Nothias, L.-F., Cretton, S., Kaiser, M., Wolfender, J.-L., & Allard, P.-M. (2022). MEMO: Mass Spectrometry-Based Sample Vectorization to Explore Chemodiverse Datasets. *Frontiers in Bioinformatics*, 2, 842964. <https://doi.org/10.3389/fbinf.2022.842964>

Gonzalez-Covarrubias, V., Kalabus, J. L., & Blanco, J. G. (2008). Inhibition of polymorphic human carbonyl reductase 1 (CBR1) by the cardioprotectant flavonoid 7-monohydroxyethyl rutoside (monoHER). *Pharmaceutical Research*, 25(7), 1730–1734. <https://doi.org/10.1007/s11095-008-9592-5>

Han, J., Kamber, M., Pei, J., & Wei, D. (2012). *Data Mining: Concepts and Techniques* (3rd ed.).

Hsieh, M. H., Chan, P., Sue, Y. M., Liu, J. C., Liang, T. H., Huang, T. Y., Tomlinson, B., Chow, M. S., Kao, P. F., & Chen, Y. J. (2003). Efficacy and tolerability of oral stevioside in patients with mild essential hypertension: a two-year, randomized, placebo-controlled study. *Clinical Therapeutics*, 25(11), 2797–2808. [https://doi.org/10.1016/s0149-2918\(03\)80334-x](https://doi.org/10.1016/s0149-2918(03)80334-x)

Hui Shen et al. (2005) Short-term steroid treatment increases  $\alpha$ 1 GABAA receptor subunit expression in rat CA1 hippocampus: Pharmacological and behavioral effects. *Neuropharmacology* 49 (5) 573-586. <https://doi.org/10.1016/j.neuropharm.2005.04.026>

Jankovic, J., & Clarence-Smith, K. (2011). Tetrabenazine for the treatment of chorea and

other hyperkinetic movement disorders. *Expert review of Neurotherapeutics*, 11(11), 1509–1523. <https://doi.org/10.1586/ern.11.149>

Jagdale, Y. D., Mahale, S. V., Zohra, B., Nayik, G. A., Dar, A. H., Khan, K. A., Abdi, G., & Karabagias, I. K. (2021). Nutritional Profile and Potential Health Benefits of Super Foods: A Review. *Sustainability*, 13(16), 9240. <https://doi.org/10.3390/su13169240>

Japan Health Food and Supplement Information Center. (2022). All About Health Foods and Supplements [Ingredients] (7th ed.).

Jiamjariyatam, R., Phucharoenrak, P., Samosorn, S., Dolsophon, K., Lorliam, W., Krajangsang, S., & Tantayotai, P. (2023). Influence of Different Extraction Methods on the Changes in Bioactive Compound Composition and Antioxidant Properties of Solid-State Fermented Coffee Husk Extracts. *The Scientific World Journal*, 2023, 6698056. <https://doi.org/10.1155/2023/6698056>

Junge, A., & Jensen, L. J. (2020). CoCoScore: context-aware co-occurrence scoring for text mining applications using distant supervision. *Bioinformatics*, 36(1), 264–271. <https://doi.org/10.1093/bioinformatics/btz490>

Kanehisa Laboratories. (2023). Ethical Drugs : Lorazepam. KEGG: URL : [https://www.kegg.jp/medicus-bin/japic\\_med?japic\\_code=00062316](https://www.kegg.jp/medicus-bin/japic_med?japic_code=00062316)

Kanehisa, M., & Goto, S. (2000). KEGG: kyoto encyclopedia of genes and genomes. *Nucleic Acids Research*, 28(1), 27–30. <https://doi.org/10.1093/nar/28.1.27>

- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y., & Morishima, K. (2017). KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Research*, 45(D1), D353–D361. <https://doi.org/10.1093/nar/gkw1092>
- Le, N.Q.K., Ho, Q.T., Nguyen, V.N., & Chang, J.S. (2022). BERT-Promoter: An improved sequence-based predictor of DNA promoter using BERT pre-trained model and SHAP feature selection. *Computational Biology and Chemistry*, 99, 107732. <https://doi.org/10.1016/j.compbiolchem.2022.107732>.
- Le, N.Q.K. (2023). Leveraging transformers-based language models in proteome bioinformatics. *Review Proteomics*. Advance online publication. <https://doi.org/10.1002/pmic.202300011>.
- McInnes, L., Healy, J., & Melville, J. (2018). UMAP: Uniform Manifold Approximation and Projection for Dimension Reduction. arXiv, arXiv:1802.03426.
- Meding, B. (1992). Normal standards for dermatological health screening at places of work. *Contact Dermatitis*, 27(4), 269–270. <https://doi.org/10.1111/j.1600-0536.1992.tb03270.x>
- Melis, M. S., & Sainati, A. R. (1991). Effect of calcium and verapamil on renal function of rats during treatment with stevioside. *Journal of Ethnopharmacology*, 33(3), 257–262. [https://doi.org/10.1016/0378-8741\(91\)90086-s](https://doi.org/10.1016/0378-8741(91)90086-s)
- Melvin, R. L., Godwin, R. C., Xiao, J., Thompson, W. G., Berenhaut, K. S., & Salsbury, F. R., Jr (2016). Uncovering Large-Scale Conformational Change in Molecular Dynamics



- without Prior Knowledge. *Journal of Chemical Theory and Computation*, 12(12), 6130–6146. <https://doi.org/10.1021/acs.jctc.6b00757>
- Motschall, E., & Falck-Ytter, Y. (2005). Searching the MEDLINE literature database through PubMed: a short guide. *Onkologie*, 28(10), 517–522. <https://doi.org/10.1159/000087186>
- Mounika, A., Ilangoan, B., Mandal, S., Shraddha Yashwant, W., Priya Gali, S., & Shanmugam, A. (2022). Prospects of ultrasonically extracted food bioactives in the field of non-invasive biomedical applications - A review. *Ultrasonics Sonochemistry*, 89, 106121. <https://doi.org/10.1016/j.ultsonch.2022.106121>
- Nebert, D. W., & Gonzalez, F. J. (1987). P450 genes: structure, evolution, and regulation. *Ann. Rev. Biochem.* 56, 945-993. <https://doi.org/10.1146/annurev.bi.56.070187.004501>
- Pletscher-Frankild, S., Pallejà, A., Tsafou, K., Binder, J. X., & Jensen, L. J. (2015). DISEASES: text mining and data integration of disease-gene associations. *Methods*, 74, 83–89. <https://doi.org/10.1016/j.ymeth.2014.11.020>
- Rein, M. J., Renouf, M., Cruz-Hernandez, C., Actis-Goretta, L., Thakkar, S. K., & da Silva Pinto, M. (2013). Bioavailability of bioactive food compounds: a challenging journey to bioefficacy. *British Journal of Clinical Pharmacology*, 75(3), 588–602. <https://doi.org/10.1111/j.1365-2125.2012.04425.x>
- Roy, R., Marakkar, S., Vayalil, M.P., Shahanaz, A., Anil, A.P., Kunnathpeedikayil, S.,

- Rawal, I., Shetty, K., Shameer, Z., Sathees, S., Prasannakumar, A.P., Mathew, O.K., Subramanian, L., Shameer, K., & Yadav, K.K. (2022). Drug-food Interactions in the Era of Molecular Big Data, Machine Intelligence, and Personalized Health. *Recent Advances in Food, Nutrition, and Agriculture*, 13(1), 27-50. <https://doi.org/10.2174/2212798412666220620104809>.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., Amin, N., Schwikowski, B., & Ideker, T. (2003). Cytoscape: A software environment for integrated models of biomolecular interaction networks. *Genome Research*, 13(11), 2498-2504. [10.1101/gr.1239303](https://doi.org/10.1101/gr.1239303)
- Sorrenti, V., Burò, I., Consoli, V., & Vanella, L. (2023). Recent Advances in Health Benefits of Bioactive Compounds from Food Wastes and By-Products: Biochemical Aspects. *International Journal of Molecular Sciences*, 24(3), 2019. <https://doi.org/10.3390/ijms24032019>
- Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., Simonovic, M., Doncheva, N. T., Morris, J. H., Bork, P., Jensen, L. J., & Mering, C. V. (2019). STRING v11: protein-protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, 47(D1), D607–D613. <https://doi.org/10.1093/nar/gky1131>
- Tanabe, M., & Kanehisa, M. (2012). Using the KEGG database resource. *Current Protocols in Bioinformatics*, Chapter 1, 1.12.1–1.12.43. <https://doi.org/10.1002/0471250953.bi0112s38>

- Wang, T., Yang, J., Xiao, Y., Wang, J., Wang, Y., Zeng, X., Wang, Y., & Peng, J. (2023). DFinder: a novel end-to-end graph embedding-based method to identify Drug-food interactions. *Bioinformatics*, 39(1), btac837. <https://doi.org/10.1093/bioinformatics/btac837>
- White J. (2020). PubMed 2.0. *Medical Reference Services Quarterly*, 39(4), 382–387. <https://doi.org/10.1080/02763869.2020.1826228>
- Wishart, D. S., Feunang, Y. D., Guo, A. C., Lo, E. J., Marcu, A., Grant, J. R., ... Wilson, M. (2018). DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Research*, 46(D1), D1074–D1082. <https://doi.org/10.1093/nar/gkx1037>
- Zhang, Y., Chen, Q., Yang, Z., Lin, H., & Lu, Z. (2019). BioWordVec, improving biomedical word embeddings with subword information and MeSH. *Scientific Data*, 6(1), 52. <https://doi.org/10.1038/s41597-019-0055-0>
- Ziani, K., Negrei, C., Ioniță-Mîndrican, C.-B., Musuc, A. M., Predoi, V. P., Udeanu, D. I., & Mititelu, M. (2022). Drug-food interactions: the influence on the patient's therapeutic plan. *Farmacia*, 70(5)

