

革新的AI創薬

医療ビッグデータ、人工知能がもたらす創薬研究の未来像

監修 | 小長谷明彦

NTS

FMO DB ID: 1JL3Z

Calculation Name: 7BV2-ABCPT-EM123

Preferred Name:

Target Type:

Ligand Name: [(2-[r],3-[s],4-[r],5-[r])-5-(4-azanylpyrrolo[2,1-f][1,2,4]triazin-7-yl)-5-cyano-3,4-bis(oxidanyl)oxolan-2-yl)methyl dihydrogen phosphate

ligand 3-letter code: F86

PDB ID: 7BV2

Chain ID: ABCPT

ChEMBL ID:

UniProt ID: P0DT01

Base Structure: ElectronMicroscopy

Registration Date: 2020-05-27

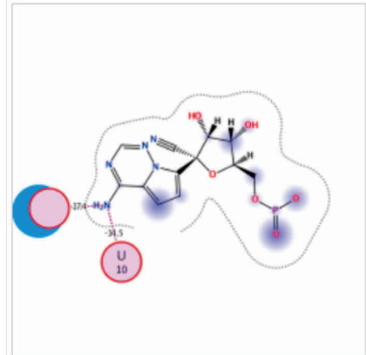
Reference: Kato, Koichiro; Honma, Teruki; Fukuzawa, Kaori (2020): Intermolecular interaction among Remdesivir, RNA and RNA-Dependent RNA polymerase of SARS-CoV-2 analyzed by fragment molecular orbital calculation. J. Mol. Graph. Model. 100 (2020) 107695-107699.

DOI: [10.1016/j.jmglm.2020.107695](https://doi.org/10.1016/j.jmglm.2020.107695)

 IFIE MAP

 Download Files

Ligand Interaction



Interactive mode: IFIE and PIEDA for fragment #1057(P:101:F86)

Summations of interaction energy for fragment #1057(P:101:F86)

IFIE [kcal/mol]	PIEDA [kcal/mol]				Charge transfer value [e]
	ES	EX	CT+mix	DI(MP2)	
-102.993	-89.141	52.515	-22.015	-44.353	0.002

Interaction energy analysis for fragmet #1057(P:101:F86)

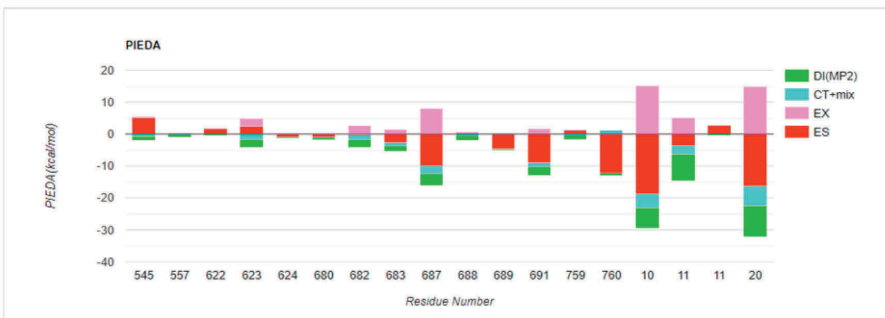


図3 RNA 依存性 RNA ポリメラーゼ, RNA, レムデシビル複合体の FMO 計算結果 (p.50)
(FMO DB より抜粋)

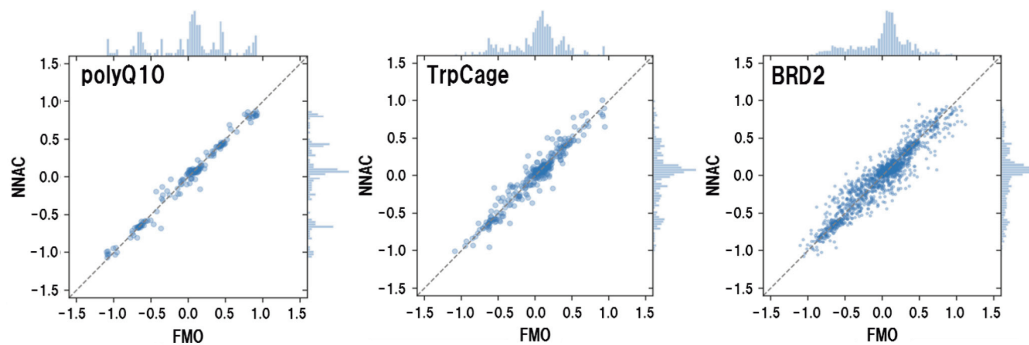


図4 原子電荷予測モデルによる未知コンフォメーションに対する検証結果 (p.52)
 (polyQ10, TrpCage, BRD2の順に R^2 は 0.982, 0.928, 0.896)

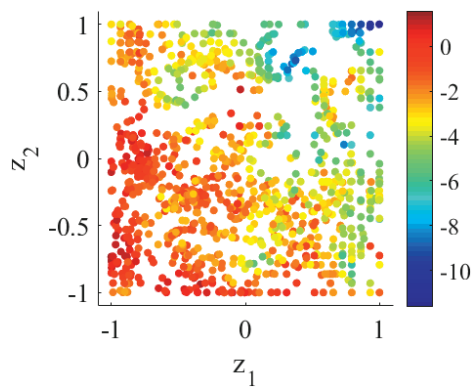


図2 GTMRによるデータの可視化結果。
 カラーバーは $\log S$ の値に対応する (p.60)
 (H. Kaneko: *Mol. Inf.*, 38, 1800088 (2019))

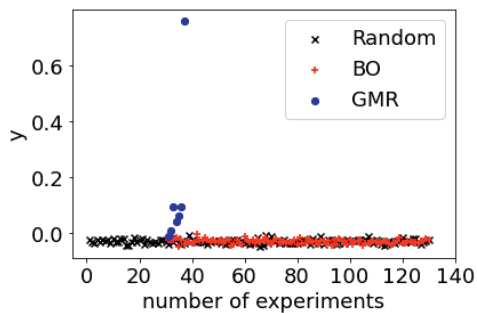


図5 手法ごとの実験回数ごとの y の値の推移
 の例 ($d = 15$) (p.62)
 (H. Kaneko: *Chemom. Intell. Lab. Syst.*,
 208, 104226 (2021))

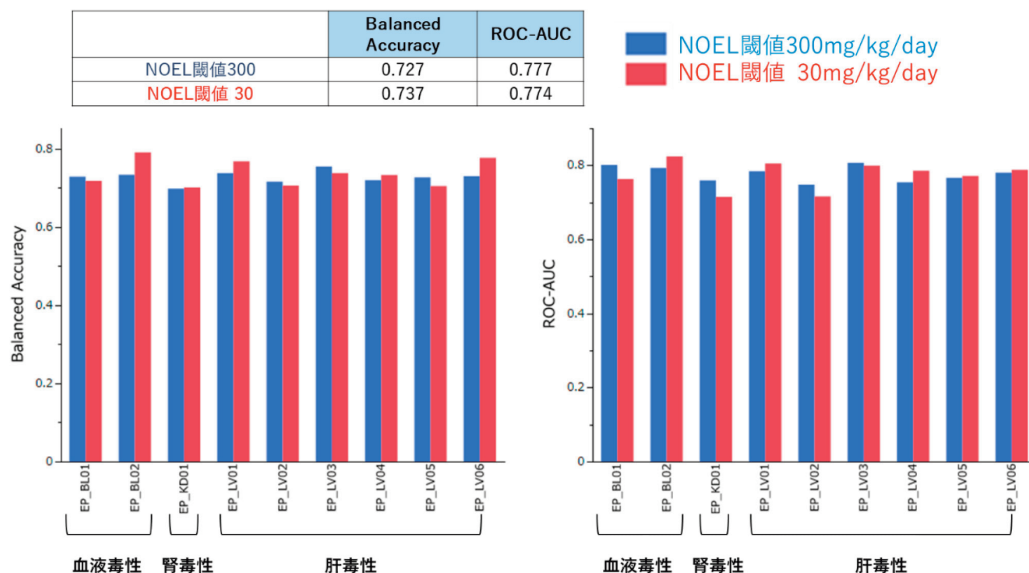


図3 外部検証における9種類の毒性エンドポイント予測モデルの精度評価 (p.81)

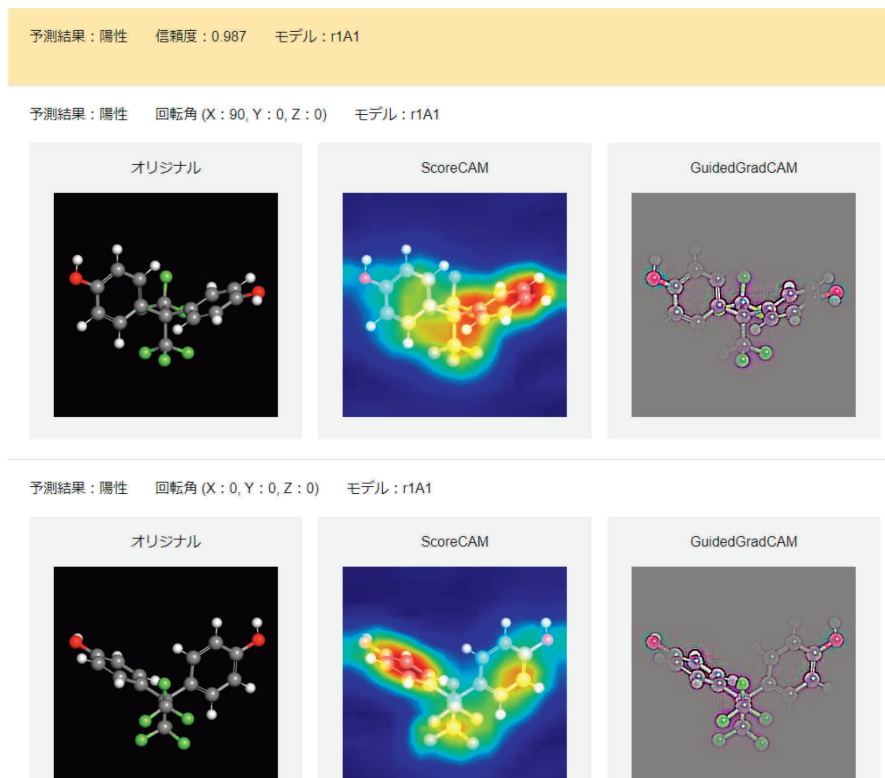


図5 DeepSnapによる特徴領域可視化の具体例 (p.82)

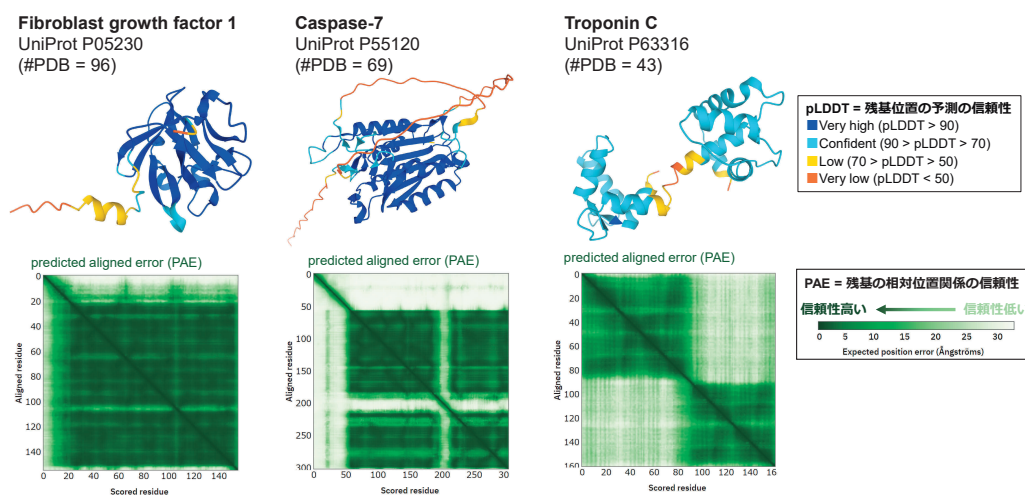


図2 AlphaFold2 によって予測された構造 (p.168)

Fibroblast growth factor 1, Caspase-7, Troponin C についての予測結果を表示した。各予測結果は残基位置の信頼性 (pLDDT) および残基間の相対的な位置関係の信頼性 (predicted aligned error; PAE) によって評価されている。(AlphaFold Protein Structure Database より取得, Licensed under CC BY 4.0)

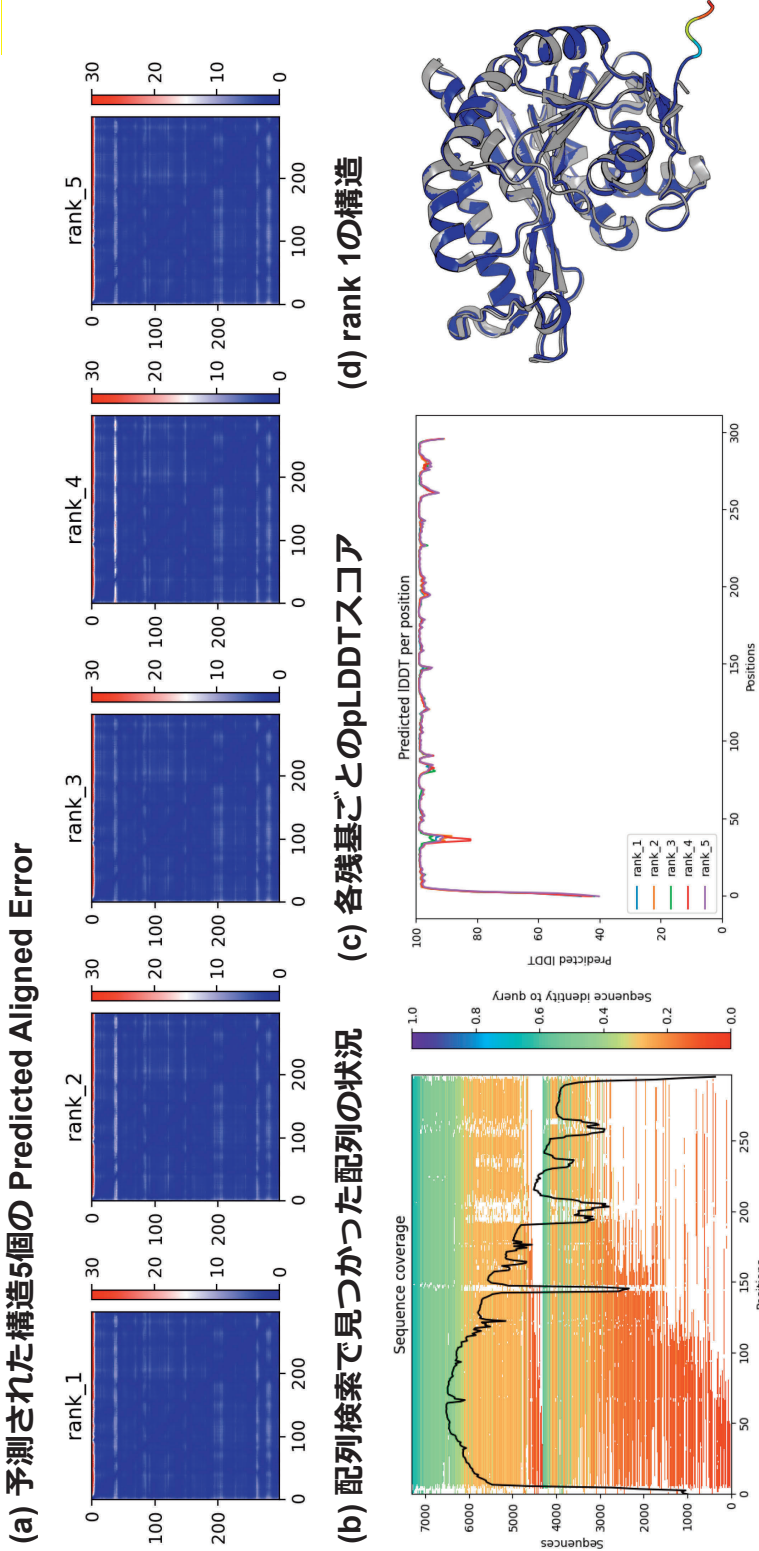


図 4 MEMO1 タンパク質の ColabFold の予測結果の確認 (p.172)
 (a) Predicted Aligned Error (PAE) によって残基と残基の間の位置関係が正しそくを見積もることができる。青い (低い) はど良い、(b)配列検索で見つかった配列のカバー度合いを示している。ここで見つかる配列の本数が多いほど構造予測が高精度になりやすい、(c)残基ごとの pLDDT スコア。残基の位置の予測の信頼性を表し、高いほど信頼性が高い (d)ColabFold が出力した rank 1 の構造 (カラー) と、真の構造 (灰色、PDB ID: 3bcz)。ColabFold の予測構造は pLDDT 値で色分けされている。

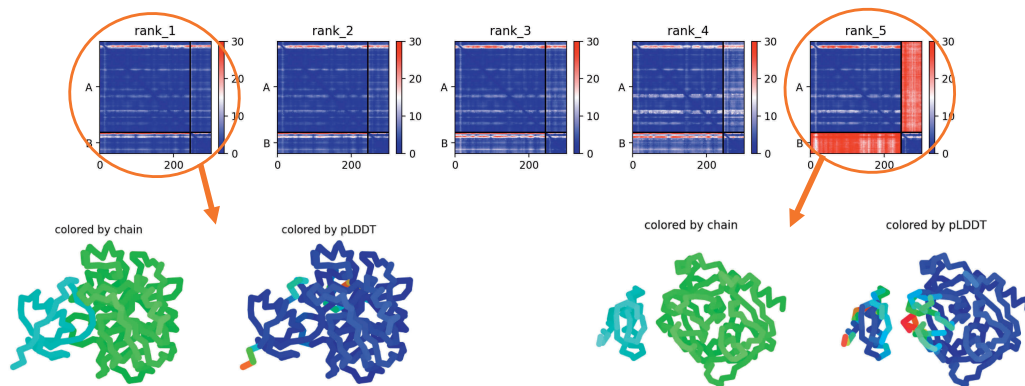


図5 ColabFoldによるヘテロ2量体の構造予測の例 (p.173)

rank 1～4は比較的正しい構造が予測できていたが、rank 5は結合位置を大きく外した予測結果となった。rank 5のPAEやpLDDTから、予測の信頼性が低い可能性が示唆されている。

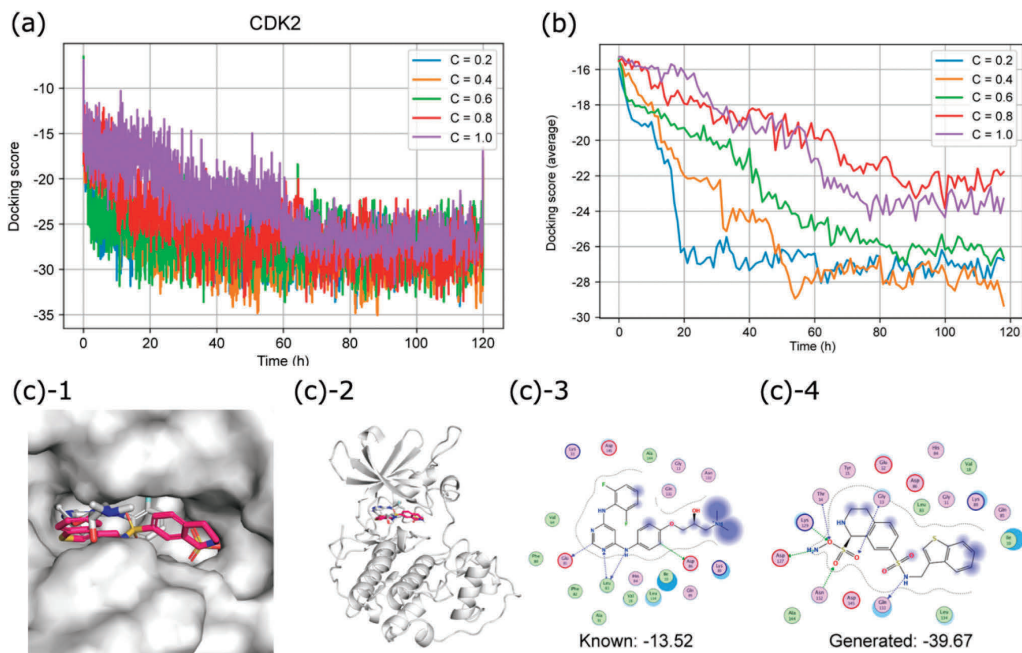


図2 SBMolGenによる分子の生成例 (p.179)

(a) CDK2をターゲットとし探索パラメータCを変化させたときのドッキングスコアの変化 (b) 4つのターゲットに対し探索パラメータを変化させた際の探索の様子。Cが小さいほど急激に探索が進むことがわかる (c) CDK2に対して生成した分子の(c)-4と既知阻害剤(c)-3のドッキングポーズ例。数値はドッキングスコアを表している。

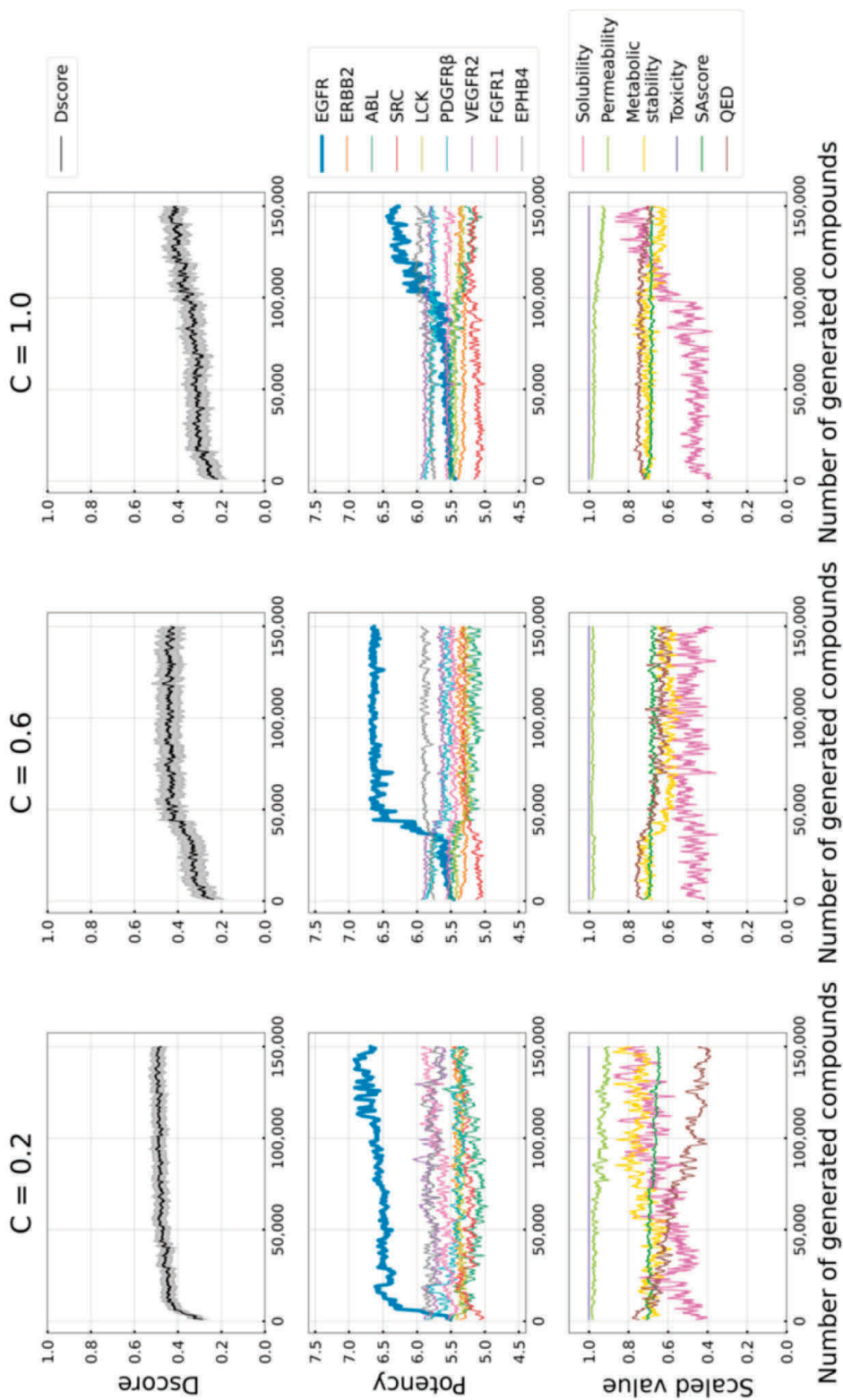


図 3 Dscore による多目的最適化例 (p.181)

探索が進むにつれてEGFRの活性予測値が上昇し、類縁キナーゼの活性予測値はさほど上昇していないことが確認できる。
各スコアは1000化合物ごとの移動平均を表しており、活性予測値が10倍以上離れた分子も多数生成されている。

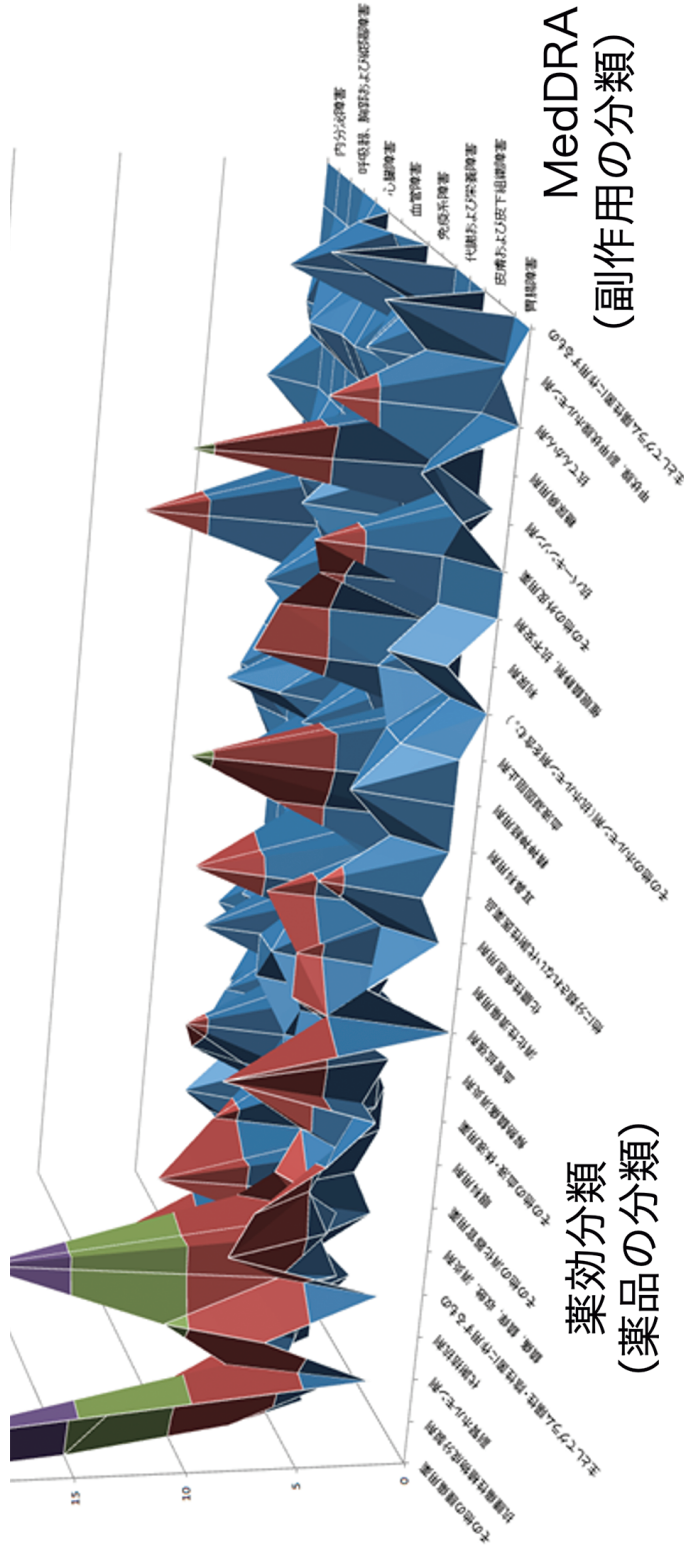
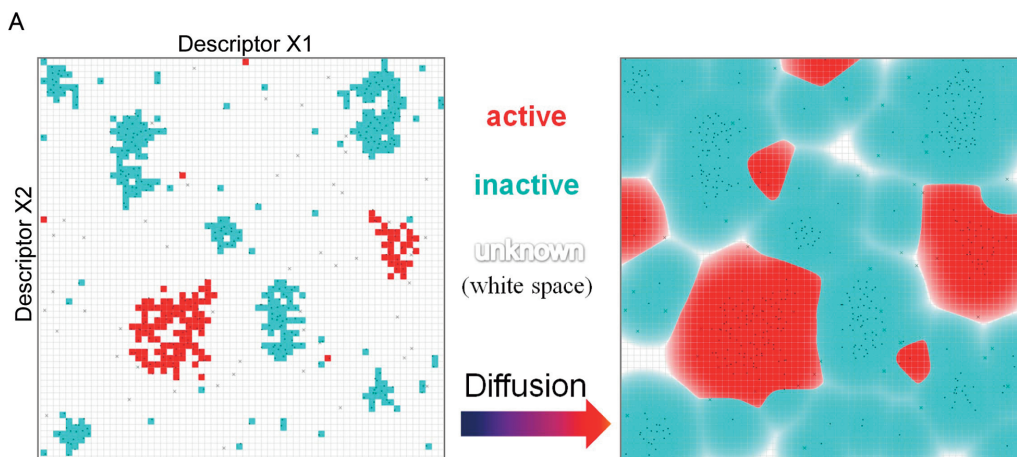


図2 言語処理のバイブライン (p.190)



B

m_1 : 活性データの個数、 m_2 : 不活性データの個数、 n : 記述子の個数
 $b_i^+(k)$: k 番目の活性データの i 番目の記述子の値
 $b_i^-(k)$: k 番目の不活性データの i 番目の記述子の値
 k 番目のデータのメッシュからの寄与度 (+: 活性データ、 -: 不活性データ)

$$u^\pm(\vec{x}, t, k) := \prod_{i=1}^n \exp\left\{-\frac{(x_i - b_i^\pm(k))^2}{4t}\right\} = \exp\left\{-\sum_{i=1}^n \frac{(x_i - b_i^\pm(k))^2}{4t}\right\}$$

v_k^+ : k 番目の活性データのスコア、 v_k^- : k 番目の不活性データのスコア
 z : 判定値、 h : パラメータ ($0 \leq h \leq \infty$)

V-score

$$f(\vec{x}, t, h) := \sum_{k=1}^{m_1} u^+(\vec{x}, t, k) \cdot \tanh(h(v_k^+ - z)) - C \sum_{k=1}^{m_2} u^-(\vec{x}, t, k) \cdot \tanh(h(z - v_k^-))$$

重み調整の定数

$$C := \frac{\sum_{k=1}^{m_1} \tanh(h(v_k^+ - z))}{\sum_{k=1}^{m_2} \tanh(h(z - v_k^-))}$$

$\tanh w = (e^w - e^{-w}) / (e^w + e^{-w})$ (双曲正接関数)
 $e^w = \exp w$ (指数関数)

図 1 HDE モデル (p.195)

- A. 熱拡散の active 化合物, inactive 化合物への適応例。
 B. 熱拡散方程式を応用した化合物のメッシュにおける寄与度と V-Score。
 (T. Hidaka et al.: *Patterns*, 1 (9), 100140 (2020))

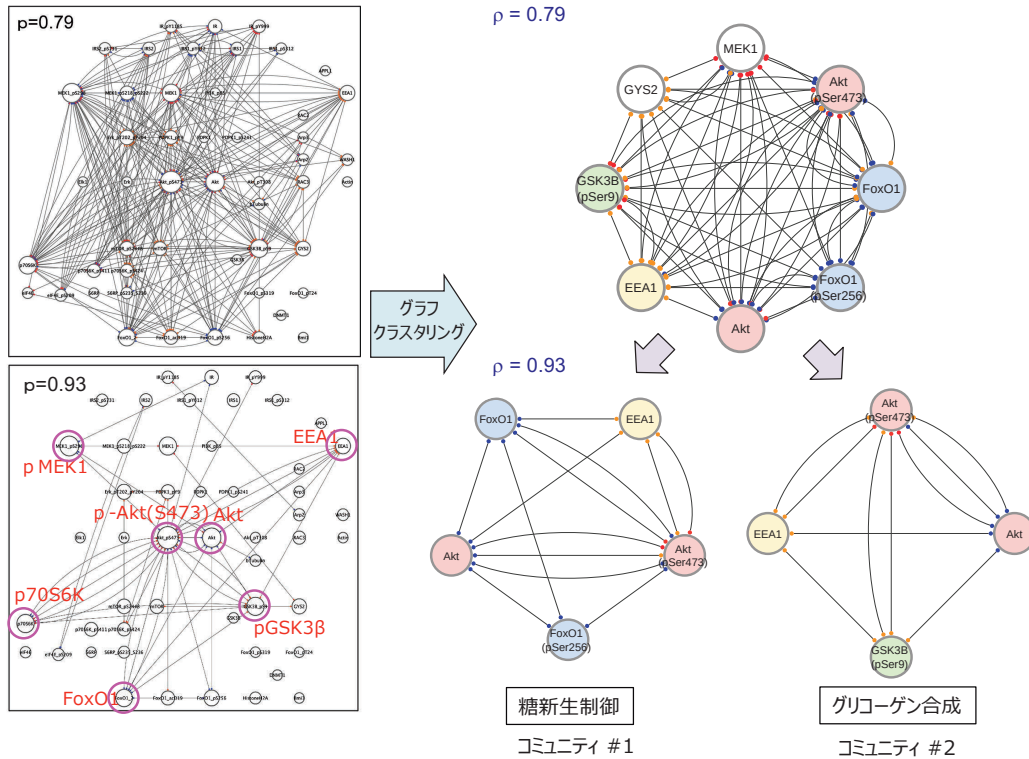


図2 インスリン刺激した H4IIEC3 細胞の共変動ネットワークとグラフクラスタリング (p.203)
 (A) 正則化パラメーター ρ が 0.79 あるいは 0.93 の時の共変動ネットワーク
 (B) ρ が 0.79 あるいは 0.93 の共変動ネットワークからグラフクラスタリングをもとに得られたコミュニティ。
 ノードの周りの色付きの小円は、タンパク質の「局在」情報を示す。赤：細胞質、青：核、黄：アクチンドメイン。

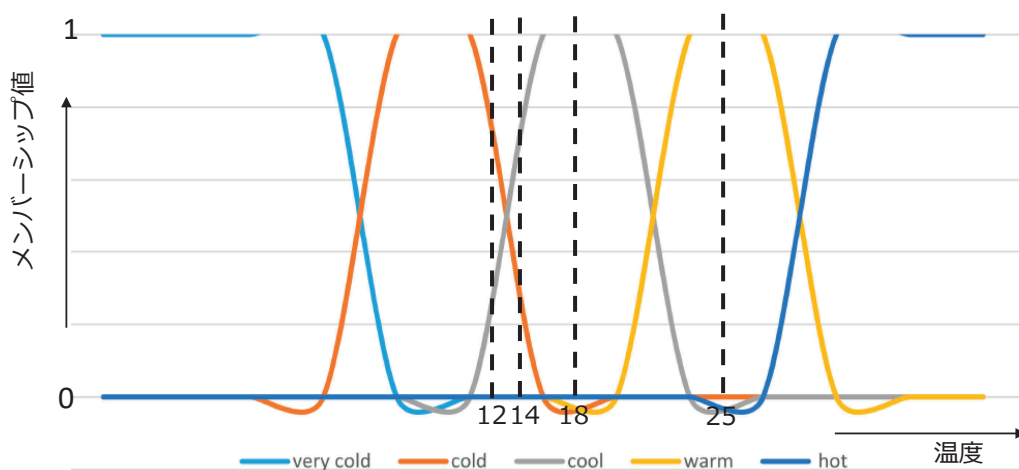


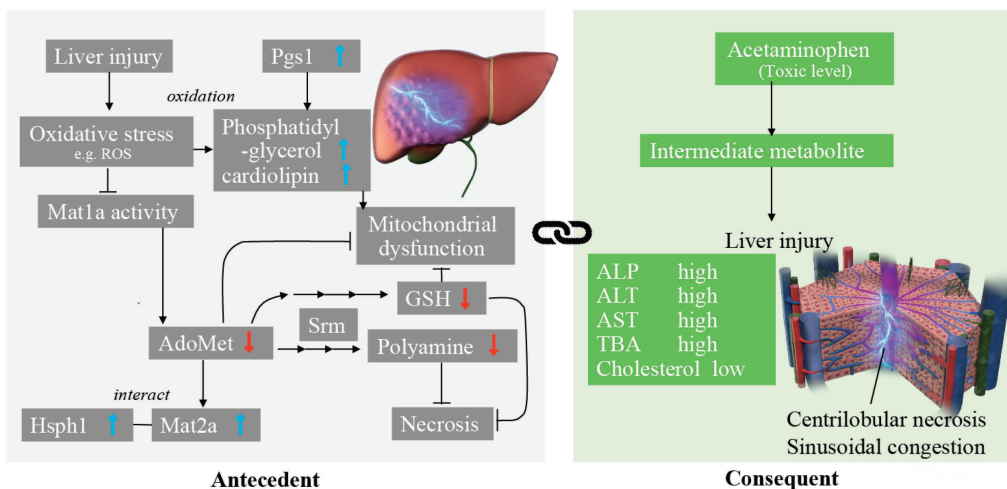
図1 温度属性のメンバーシップ関数の例 (p.238)
 (M. Abdel-Basset et al.: *Symmetry*, 10 (4), 106 (2018))

Probe	Gene (Entrez)	Gene name	High / Low
A_43_P10003	288444	heat shock protein family H (Hsp110) member 1(Hsph1)	High
A_42_P717602	171347	methionine adenosyltransferase 2A(Mat2a)	High
A_42_P484423	303698	phosphatidylglycerophosphate synthase 1(Pgs1)	High
A_43_P17455	117251	dynein, axonemal, heavy chain 9(Dnah9)	Low
A_43_P14864	58945	dynein light chain LC8-type 1(Dynll1)	High
A_43_P16523	361578	nodal modulator 1(Nomo1)	High
A_43_P19279	363168	lysozyme-like 4(Lyzl4)	Low
A_43_P12811	84596	spermidine synthase(Srm)	High
A_42_P655825	365215	SMG9 nonsense mediated mRNA decay factor(Smg9)	High
A_42_P804499	n.d.	n.d.	High

(a)データ 1 から抽出された関連遺伝子

Clinical measurement	Description	High / Low
ALP (alkaline phosphatase)	positive correlation with bile flow interruption	High
ALT (alanine aminotransferase)	positive correlation with liver injury	High
AST (aspartate aminotransferase)	positive correlation with liver injury	High
cholesterol	negative correlation with liver injury	Low
TBA (total bile acid)	positive correlation with bile flow interruption	High
Histopathological observations	Description	High / Low
LLL_Centrilob_Necrosis_High	Necrosis of the centrilobular tissue of the liver lobule	High
LLL_Hepat_Hypertrophy_High	Enlargement of the liver	High
LML_Centrilob_Necrosis_High	Necrosis of the centrilobular tissue of the liver lobule	High
LML_Sinusoid_Cogestion_High	Enlargement of the hepatic capillaries	High
Experimental conditions	Description	High / Low
Toxic_Dose	50, 150 mg/kg BW : subtoxic, 1500, 2000 mg/kg BW: severely toxic	High
Toxic_Time	Peak of the severe toxicity: 24hrs after exposure	High

(b)デー2から抽出された関連遺伝子



(c) 出力例の解釈

図 6 アセトアミノフェンの肝毒性データでの提供結果 (p.242)

- (a)データ 1 (遺伝子) から抽出された関連遺伝子。
- (b)データ 2 (組織病学的の観察データおよび臨床的の測定データでのアセトアミノフェンの毒性の度合を示す根拠。
- (c)出力例の解釈。提示している遺伝子 (青→量が増える、赤→量が減る) はアセトアミノフェンの毒性と紐づいていることが論文で示されていた。(Y. Natsume-kitatani; et al: *Research Square*, doi.org/10.21203/rs.3.rs-405195/v1)

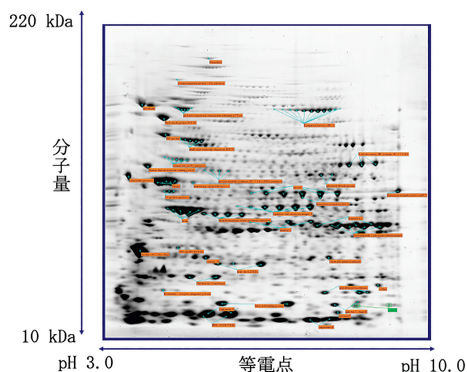


図3 二次元電気泳動画像の一例 (p.248)

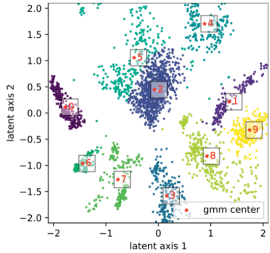



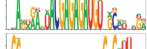


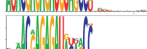
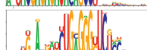

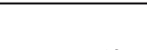
Latent embeddings	ID (dim1, dim2)	Sequence Logo	Maximum probable sequence
	A-GMM-0 (-1.77, 0.11)		AACGAGAGAUGGUAGACCUAUCUUUUAGCC
	A-GMM-1 (1.36, 0.22)		GUAGAGAUCUGAGGGUUCUCCUGCUAUA
	A-GMM-2 (-0.07, 0.45)		UUUUUAAAAAAGUGUUUUAAAAAGAUCUA
	A-GMM-3 (0.16, -1.58)		GUAGAAAUAACGAGAGAUGUCGCCUUUGA
	A-GMM-4 (0.88, 1.70)		GGGGUGCAGUAGAAUUGUCGAGUUUCUG
	A-GMM-5 (-0.46, 1.06)		AAUACCCGGGUUUUCACACAUUAAUUCA
	A-GMM-6 (-1.45, -0.96)		AUACGAGAGAUGUAGCCUUUUUCUGACUU
	A-GMM-7 (-0.77, -1.27)		AGUACGAGAGAUACAGCCUUUUUCUGCUU
	A-GMM-8 (0.93, -0.82)		GGUAGCAGAUGCUGAGGGGUCUCCUGAUGC
	A-GMM-9 (1.74, -0.33)		GUCGAGAUCUGAGGGUUCUCCUGUUAACC

図4 RaptGen の適用例 (p.260)

実際の HT-SELEX データの潜在空間 (第1列), GMM のクラスタリングの代表点 (第2列), プロファイル HMM から作成した代表点のシーケンスロゴ (第3列), 最大確率の配列 (第4列) を示している。
(N. Iwano et al.: *Nat. Comput. Sci.*, doi.org/10.1038/s43588-022-00249-6 (2022))

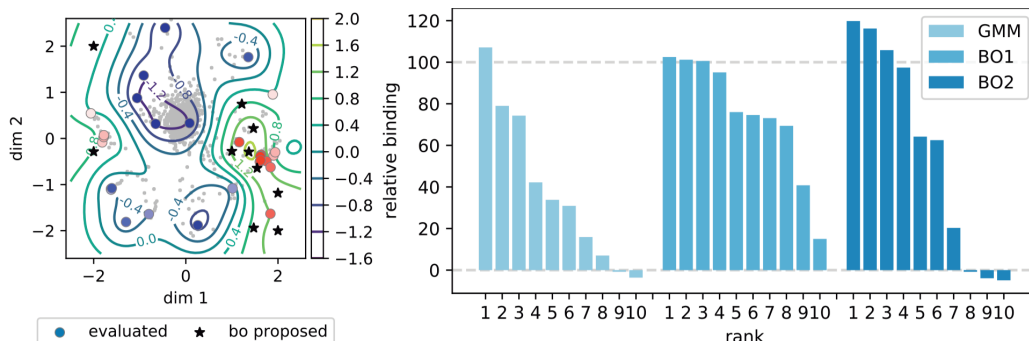


図5 ベイズ最適化による配列デザイン (p.261)

(左図) 学習された潜在空間。●の点は結合活性が取得されている配列 (赤いほど活性が強い) を★がベイズ最適化により新たに提案された配列である。

(右図) GMMにより選択された配列, 1回目のベイズ最適化により選択された配列 (BO1), 2回目のベイズ最適化により選択された配列 (BO2) の相対結合活性 (ポジティブコントロールを 100 とした場合の相対値)。

(N. Iwano et al: *Nat Comput. Sci.*, doi.org/10.1038/s43588-022-00249-6 (2022))

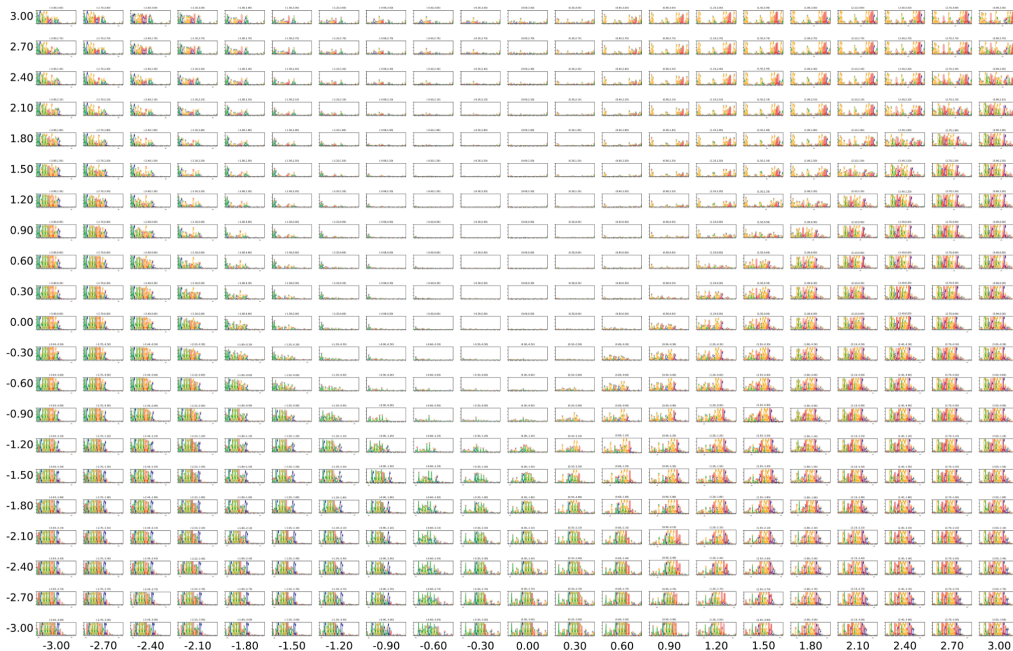


図6 潜在空間の可視化 (p.261)

実際の SELEX データで学習した潜在空間。なめらかな潜在空間が学習されていることがわかる。



図 1 武田における創薬研究でのデジタル・AI ツール (p.288)

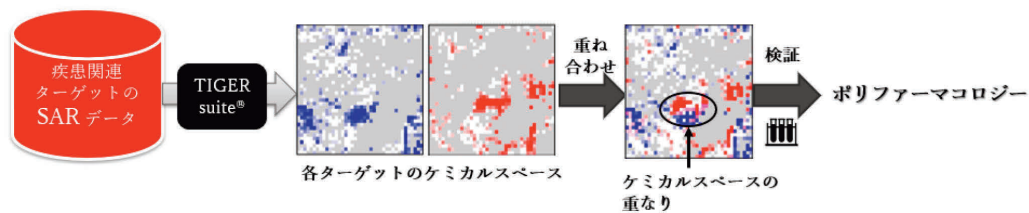


図 6 ポリファーマコロジーのターゲット予測 (p.291)

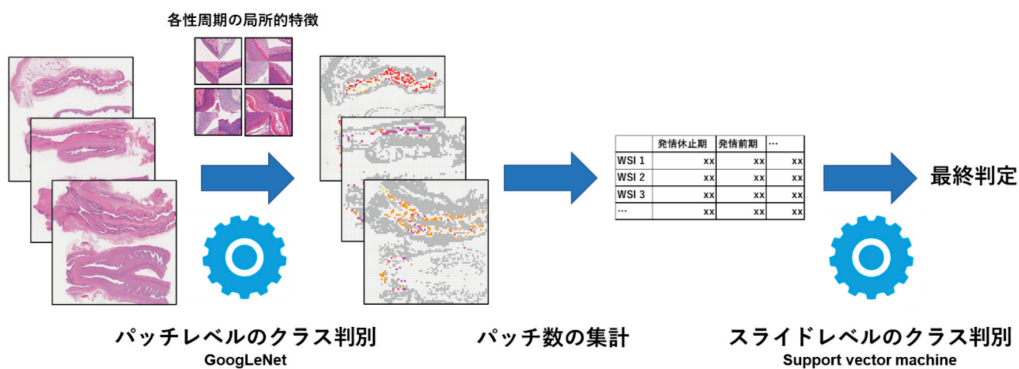


図 6 膺の病理画像からの性周期判定ワークフロー (p.319)

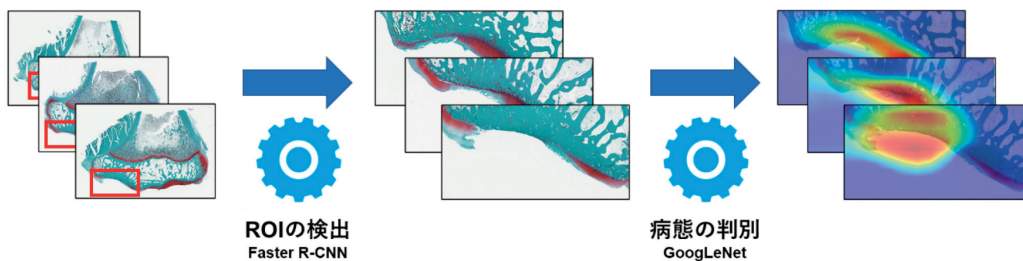


図 7 関節の病理画像からの病態評価ワークフロー (p.320)

まえがき

近年、深層学習などの人工知能技術を活用した創薬，すなわち「AI創薬」が注目を集めている。本書『革新的 AI創薬』では，AI創薬の可能性に向けて，2022年時点における日本のAI創薬の現状を俯瞰し，AI創薬普及への課題を整理し，今後のAI創薬の方向性を示すことを目標に企画した。AI創薬技術は日々刻々と変化し，また，AI創薬の対象は広大である。刊行スケジュールの都合上執筆していただけなかったAI創薬関係者も多く，日本の全てのAI創薬活動を網羅できたわけではないが，わずか半年たらずの執筆期間の中で，幸いにも38本の玉稿を拜受することができた。

本書では，第1章において，日本のAI創薬研究の現状を俯瞰し，第2章において化学情報学からアプローチするAI創薬技術について，第3章では生命情報学および医療情報学からのAI創薬技術について紹介する。また，第4章ではAI創薬を推進するための情報処理技術および数理科学について，第5章ではAI創薬技術の活用事例について，第6章では製薬会社におけるAI創薬への取り組みを，第7章ではAIベンチャー企業におけるAI創薬への取り組みを紹介する。

日本におけるAI創薬は始まったばかりだが，AI創薬の可能性に向けて，本書が一助となれば無上の喜びである。激務中にご執筆いただいたAI創薬関係者ならびに編集者の皆さまにこの場を借りて御礼申しあげる。

2022年6月 東京工業大学名誉教授 小長谷 明彦

監修者・執筆者一覧 (敬称略)

監修者

小長谷 明彦 東京工業大学名誉教授

執筆者

小長谷 明彦 東京工業大学名誉教授

夏目 やよい 国立研究開発法人医薬基盤・健康・栄養研究所 AI健康・医薬研究センター
プロジェクトリーダー

水口 賢司 国立研究開発法人医薬基盤・健康・栄養研究所 AI健康・医薬研究センター センター長/
大阪大学 蛋白質研究所 教授

萩島 創一 東北大学 高等研究機構未来型医療創成センター / 東北メディカル・メガバンク機構 教授

池口 満徳 横浜市立大学 大学院生命医科学研究科 教授 / 国立研究開発法人理化学研究所
計算科学研究センター ユニットリーダー

福澤 薫 大阪大学 大学院薬学研究科 教授

加藤 幸一郎 九州大学 大学院工学研究院 / 分子システム科学センター 准教授

金子 弘昌 明治大学 理工学部 准教授

船津 公人 奈良先端科学技術大学院大学 データ駆動型サイエンス創造センター 特任教授 /
センター長

山崎 浩史 昭和薬科大学 薬学部 教授

植沢 芳広 明治薬科大学 薬学部 教授

吉成 浩一 静岡県立大学 薬学部 教授

庄野 文章 奈良先端科学技術大学院大学 データ駆動型サイエンス創造センター 非常勤講師

清水 祐吾 慶應義塾大学 薬学部 特任助教

米澤 朋起 慶應義塾大学 薬学部 特任助教

池田 和由 慶應義塾大学 薬学部 特任准教授 / 国立研究開発法人理化学研究所 計算科学研究センター
HPC/AI 駆動型医薬プラットフォーム部門創薬化学 AI アプリケーションユニット 上級研究員

飯田 緑 九州工業大学 大学院情報工学研究院 日本学術振興会特別研究員 (RPD)

山西 芳裕 九州工業大学 大学院情報工学研究院 教授

小島 諒介 京都大学 大学院医学研究科 講師

水野 聖士 東北大学 東北メディカル・メガバンク機構 助教

佐藤 淳平 東京大学 生産技術研究所 協力研究員

合田 和生 東京大学 生産技術研究所 准教授

喜連川 優 東京大学 生産技術研究所 特別教授

石川 智基 一般財団法人医療経済研究・社会保険福祉協会医療経済研究機構 主任研究員

満武 巨裕 一般財団法人医療経済研究・社会保険福祉協会医療経済研究機構 研究副部長

関嶋 政和 東京工業大学 情報理工学院 准教授

大上 雅史 東京工業大学 情報理工学院 助教

寺山 慧 横浜市立大学 大学院生命医科学研究科 准教授

荒牧 英治 奈良先端科学技術大学院大学 先端科学技術研究科 教授

今村 恵子	京都大学 iPS 細胞研究所 特定拠点講師 / 国立研究開発法人理化学研究所 革新知能統合研究センター, バイオリソース研究センター / T-CiRA
日高 中	武田薬品工業株式会社 リサーチ・CNSDDU・ANCEL
儀我 美保	東京大学 大学院数理科学研究科 特任研究員 / 一般財団法人学際数学研究所
儀我 美一	東京大学 大学院数理科学研究科 特任教授 / 一般財団法人学際数学研究所
井上 治久	京都大学 iPS 細胞研究所 教授 / 国立研究開発法人理化学研究所 革新知能統合研究センター, バイオリソース研究センター / T-CiRA
加納 ふみ	東京工業大学 科学技術創成研究院 細胞制御工学研究センター 准教授
國重 莉奈	東京工業大学 科学技術創成研究院 細胞制御工学研究センター 特任助教
山下 慶子	日本電気株式会社 AI 創薬統括部 シニアエキスパート
尾上 広祐	日本電気株式会社 AI 創薬統括部 主任
田中 雄希	日本電気株式会社 AI 創薬統括部 主任
小野口 和英	日本電気株式会社 AI 創薬統括部 プロフェッショナル
北村 哲	日本電気株式会社 AI 創薬統括部 統括部長
興野 悠太郎	株式会社 MOLCURE 最高技術責任者
上田 修功	国立研究開発法人理化学研究所 革新知能統合研究センター 副センター長
林 宣宏	東京工業大学 生命理工学院 教授 / 副学長
浜田 道昭	早稲田大学 理工学術院 教授 / 国立研究開発法人産業技術総合研究所 / 日本医科大学
宮本 大輔	アマゾン ウェブ サービス ジャパン合同会社 技術統括本部 レディネスソリューション部 HPC スペシャリストソリューションアーキテクト
宇都宮 聖子	アマゾン ウェブ サービス ジャパン合同会社 技術統括本部 機械学習ソリューション部 シニア機械学習・量子スペシャリストソリューションアーキテクト
大島 貴弘	インフォコム株式会社 ヘルスケアサービス部 ライフサイエンスグループ
森下 奈央	武田薬品工業株式会社 リサーチ ニューロサイエンス創薬ユニット ドラッグディスカバリーケミストリー研究所 主任研究員
谷口 孝彦	武田薬品工業株式会社 リサーチ ニューロサイエンス創薬ユニット ドラッグディスカバリーケミストリー研究所 ディレクター
齊藤 隆太	田辺三菱製薬株式会社 創薬本部 創薬基盤研究所 主席研究員
市川 治	住友ファーマ株式会社 基盤技術研究ユニット第4グループ グループマネージャー
吉永 英史	住友ファーマ株式会社 化学研究ユニット第2グループ 主席研究員
寺本 礼仁	中外製薬株式会社 創薬基盤研究部 機械学習チームリーダー
水野 英明	中外製薬株式会社 創薬基盤研究部 バイオインフォマティクスチームリーダー
Casey Galvin	株式会社 Elix リサーチエンジニア
力丸 健太郎	株式会社 Preferred Networks ライフ&マテリアルサイエンス リサーチャー
木村 敏郎	株式会社ワールドフュージョン 技術営業部 マネージャー
櫻井 祐樹	株式会社ワールドフュージョン 技術営業部
神谷 幸太郎	SyntheticGestalt 株式会社 CTO

目次

第1章 日本の AI 創薬研究の現状

第1節 AI 創薬の可能性と普及への課題	小長谷 明彦
1. はじめに	3
2. AI 創薬へのアプローチ	4
3. AI 創薬技術の課題	7
4. AI 創薬技術の社会実装	10
5. まとめ	13
第2節 生命情報科学からの AI 創薬	夏目 やよい, 水口 賢司
1. はじめに	15
2. オミックスデータは「ビッグデータ」か	16
3. オミックスデータと既存知識の融合によるドラッグリポジショニング	17
4. ネットワークベースのオミックス解析による創薬標的探索	17
5. マルチオミックスデータ統合解析による創薬標的探索	18
6. 自然言語処理による知識自動抽出と知識グラフ	19
7. おわりに	20
第3節 医療ビッグデータからの AI 創薬と個別化ヘルスケア	荻島 創一
1. はじめに	23
2. 医療情報のビッグデータの形成	23
3. ゲノム情報のビッグデータの形成	27
4. 医療ビッグデータによる新しい医療	29
5. 医療ビッグデータによる AI 創薬	30
6. おわりに	32

第2章 化学情報学からの AI 創薬の最前線

第1節 AI 創薬のための分子動力学シミュレーション法の開発	池口 満徳
1. はじめに	37
2. スーパーコンピュータ, GPU と分子動力学ソフトウェア	38
3. リガンド複合体の分子動力学シミュレーション	39
4. リガンドの結合自由エネルギー計算	39
5. リガンド結合の速度論パラメータ	42
6. リガンド結合に伴うタンパク質の立体構造変化	42
7. 分子動力学シミュレーションと AI の連携	43
8. おわりに	44
第2節 FMO 法を活用した AI 創薬	福澤 薫, 加藤 幸一郎
1. はじめに	47
2. フラグメント分子軌道 (FMO) 法	47

3. FMO データベース (FMO DB)	49
4. 古典分子動力学法と FMO 法の融合	51
5. 機械学習による電荷予測モデルの構築	51
6. FMO-AI 力場の構築に向けて	52
第 3 節 人工知能を活用した分子設計・材料設計	金子 弘昌
1. はじめに	57
2. 新しいサンプルにおける予測誤差の推定	59
3. 直接的逆解析法による分子設計	60
4. 分子設計・材料設計・プロセス設計のための直接的逆解析法による適応的実験計画	62
5. おわりに	63
第 4 節 毒性関連ビッグデータを用いた人工知能による 次世代型安全性予測手法開発プロジェクト	
第 1 項 AI-SHIPS プロジェクトの意義 ～開発背景, 設計思想, および今後の展開について～	船津 公人
1. はじめに	65
2. プロジェクトスタートの背景	65
3. 毒性予測の仕組み (3 層モデル)	66
4. 毒性予測システムの構成	68
5. データ駆動型化学の全体像の中での毒性予測システムの位置づけと今後への期待	69
第 2 項 一般化学物質の吸収および体内動態予測手法開発	山崎 浩史
1. はじめに	71
2. 化合物の空間情報	74
3. 化合物の膜透過係数	74
4. 化合物の生理学的薬物動態モデル	74
5. おわりに	75
第 3 項 AI-SHIPS における一般化学物質の毒性予測モデル構築	植沢 芳広
1. はじめに	77
2. 説明性を重視した毒性予測モデルの構築戦略	77
3. MIE/KE 予測モデルと毒性エンドポイント予測モデル	78
4. 機械学習による予測モデルの構築	78
5. 予測性能の検証	79
6. 特徴領域可視化システム	81
7. おわりに	83
第 4 項 AI-SHIPS 統合的毒性予測システムの開発	吉成 浩一
1. はじめに	85
2. ユーザーシステム	85
3. モデル・データ管理システム	88
4. おわりに	89
第 5 項 AI-SHIPS に関連する最近の国内外の <i>in silico</i> 予測手法の開発状況	庄野 文章
1. 緒言	91
2. 国際的な計算科学的毒性予測手法開発の現状	91

3. 国内外の利用可能な毒性予測システムと研究開発の現状	93
4. おわりに	93

第5節 AI創薬における適切な学習データとは？ 清水 祐吾, 米澤 朋起, 池田 和由

1. はじめに	95
2. データの取得源	95
3. データの取得と選別	97
4. 学習時のデータの取り扱い	99

第3章 生命情報学および医療情報学からの AI 創薬

第1節 新薬創出を加速する人工知能の開発—臨床情報を活用した創薬標的探索

夏目 やよい

1. はじめに	107
2. 事業背景	108
3. 事業概要	108
4. 臨床情報データベースの構築	109
5. データ駆動的創薬標的探索に資する AI の開発と創薬標的候補の提示	110
6. 事業成果公開のためのオーブンプラットフォーム構築	111
7. おわりに	112

第2節 分子ネットワークを有効活用した AI 創薬手法

飯田 緑, 山西 芳裕

1. 研究の背景	113
2. 研究内容	115
3. まとめ	123

第3節 グラフニューラルネットワークを用いた医療関連データ処理

小島 諒介

1. はじめに	125
2. 医療 AI 分野で扱うデータの特徴	125
3. 複数の専門家が連携して取り組む医療 AI プロジェクトの課題	127
4. 医療 AI とグラフニューラルネットワーク	128
5. 創薬応用における GNN	129
6. GNN の知識グラフやその他データへの応用	130
7. まとめ	131

第4節 大規模コホートデータに基づく早期疾患予測モデル

水野 聖士, 荻島 創一

1. はじめに	133
2. 大規模コホート研究	134
3. 東北メディカル・メガバンク計画	135
4. 早期疾患予測モデル	137
5. おわりに	142

第5節 地域のレセプト情報に基づく介護需要将来推定

佐藤 淳平, 合田 和生, 喜連川 優, 石川 智基, 満武 巨裕

1. はじめに	145
---------------	-----

2. 医療レセプト情報・介護レセプト情報の概要	146
3. 要介護度と要介護状態の定義	147
4. レセプトの診療情報・介護情報に基づく将来の要介護状態の推定方法	148
5. 評価実験	150
6. 課題と今後の展望	152

第4章 AI創薬を推進する情報処理技術

第1節 AI創薬アルゴリズム	関嶋 政和
1. はじめに	157
2. バーチャルスクリーニング	158
3. メディシナルケミストの暗黙知を代替する機械学習	161
4. スマート創薬でより良いヒット化合物を探索する為に	162
第2節 AlphaFold2の登場と創薬への影響	大上 雅史
1. はじめに	165
2. AlphaFold2の登場	166
3. AlphaFold2の予測性能	167
4. AlphaFold2の利用	169
5. AlphaFold2の複合体構造予測への応用	173
第3節 AI創薬向け機械学習・最適化手法の開発	寺山 慧
1. 概要	175
2. はじめに	175
3. 強化学習に基づく分子生成手法 ChemTS とその応用例	176
4. ドッキングシミュレーションを利用した分子構造生成 SBMolGen	177
5. 多目的最適化への試み	179
6. まとめ	180
第4節 医療言語処理の現状と今後の展望	荒牧 英治
1. はじめに	185
2. 医療言語処理の入出力	185
3. 医療言語処理のながれ	189
4. 今後の展望	191
第5節 熱拡散方程式 (HDE) モデルによる新薬探索支援 AI ツールの開発	今村 恵子, 日高 中, 儀我 美保, 儀我 美一, 井上 治久
1. はじめに	193
2. 熱拡散方程式 (HDE) モデルによる予測	193
3. 実際の課題への応用	196
第6節 共変動ネットワーク解析法とリシール細胞技術を基盤とする AI主導の創薬・細胞医薬支援プラットフォーム創成	加納 ふみ, 國重 莉奈
1. 序論	199
2. PLOM-CON 解析法	200

3. リシール細胞技術 206

第5章 活用事例

**第1節 AIを用いたネオアンチゲン予測システムによる
個別化がんワクチンTG4050の開発**

山下 慶子, 尾上 広祐, 田中 雄希, 小野口 和英, 北村 哲

1. はじめに 213
 2. がん免疫の仕組みとがん免疫療法 214
 3. 個別化がんワクチン TG4050 の第 I 相臨床試験の概要 215
 4. NEC の AI を用いたネオアンチゲン予測方法 216
 5. おわりに 219

第2節 AI-ロボット駆動型科学による中・高分子医薬品設計

興野 悠太郎

1. はじめに 221
 2. AI-ロボット駆動型科学による中・高分子医薬品設計 222
 3. AI フレンドリーなスクリーニング自動ロボット群「HAIVE」 226
 4. おわりに 232

第3節 サブセット・バインディングによる患者層別化 AI の開発

上田 修功, 夏目 やよい

1. 背景 235
 2. ARM 236
 3. SB 法 237
 4. 実データでの検証 241
 5. まとめ 243

第4節 AI プロテオミクスによる診断, 治療, 創薬支援

林 宣宏

1. はじめに 245
 2. プロテオミクス 245
 3. AI プロテオミクス 248
 4. AI プロテオミクスによる診断 249
 5. AI プロテオミクスによる治療支援 251
 6. AI プロテオミクスによる創薬 252
 7. 実施例: 敗血症 252
 8. 将来の展望 254

第5節 RNA 情報科学・AI 技術を融合した AI アプタマー創薬技術の開発

浜田 道昭

1. 導入 255
 2. AI アプタマー創薬実現に向けた基盤技術 256
 3. AI アプタマー創薬の実例 259
 4. AI アプタマー創薬の今後 262
 5. まとめ 262

第6節 創薬における HPC, 機械学習, 量子コンピュータでのクラウド活用

宮本 大輔, 宇都宮 聖子

1. はじめに	265
2. 創薬研究におけるクラウド活用の特徴・利点	265
3. Amazon Web Services における High Performance Computing	267
4. Amazon Web Services における AI・機械学習	269
5. Amazon Web Services における量子コンピュータ	271
6. まとめ	272

第7節 機械学習自動化パッケージの開発と AI 創薬への適用事例

大島 貴弘

1. はじめに	275
2. KNIME とは	275
3. 機械学習を自動化する	278
4. 機械学習自動化パッケージ	278
5. 機械学習自動化パッケージのカスタマイズ	281
6. 今後の展望	283

第6章 製薬各社における AI 創薬の取り組み

第1節 武田薬品工業における AI 創薬の取り組み

森下 奈央, 谷口 孝彦

1. はじめに	287
2. 武田の創薬研究デジタルエコシステム	288
3. ポリファーマコロジー創薬への AI 活用	291
4. おわりに	292

第2節 田辺三菱製薬におけるデータ駆動型創薬の取り組み

齊藤 隆太

1. データ駆動型創薬によるプロセス変革	293
2. バイオインフォマティクスと自然言語処理 AI による創薬標的探索・適応症探索	295
3. AI ドラッグデザインと分子シミュレーション	296
4. 画像解析 AI を活用した薬効評価・スクリーニング	298
5. 仮想患者モデルを用いた臨床試験シミュレーション	299
6. 投資対効果の評価	300
7. まとめ	301

第3節 住友ファーマ株式会社における AI 創薬の取り組み

市川 治, 吉永 英史

1. はじめに	305
2. 当社における AI の実用	306
3. 今後の展望; インシリコ創薬の未来	310

第4節 中外製薬における AI 創薬の取り組み

寺本 礼仁, 水野 英明

1. 中外製薬のデジタル戦略～ AI 技術を活用した新薬創出～	313
2. 機械学習技術を用いた抗体医薬品設計	313
3. 機械学習技術を用いた病理組織学的評価	318

第7章 AIベンチャー企業におけるAI創薬の取り組み

第1節 ElixにおけるAI創薬	Casey Galvin
1. はじめに	323
2. AI創薬プラットフォーム「Elix Discovery™」	324
3. すべての人にAIを	326
4. 生成モデルと予測モデルの強化	326
5. データセットのサイズとアクセスの拡大	327
6. おわりに	328
第2節 Preferred NetworksにおけるAI創薬	力丸 健太郎
1. Introduction	329
2. PFNの創薬プラットフォーム	330
3. スコア計算モジュール・化合物生成モジュール	331
4. 化合物生成結果の登録・確認	332
5. 実創薬プロジェクトにおける活用例	333
6. おわりに	334
第3節 ワールドフュージョン社におけるAI創薬	木村 敏郎, 櫻井 祐樹
1. はじめに	337
2. 辞書(オントロジー)について	338
3. BERT-NER「深層学習手法」を使った作用機序(MoA)新規語彙の収集	338
4. BERT-MNLI「深層学習手法」を使った遺伝子-疾患関連性の検出	339
5. まとめ	341
第4節 SyntheticGestalt社におけるAI創薬	神谷 幸太郎
1. はじめに	343
2. 技術開発における設計原則と開発方針	343
3. 開発方針とAI創薬技術	344
4. おわりに	350

※本書に記載されている会社名, 製品名, サービス名は各社の登録商標または商標です。
 なお, 必ずしも商標表示(®, TM)を付記していません。

第1章

日本のAI創薬研究の現状

第1章 | 日本のAI創薬研究の現状

第1節 AI創薬の可能性と普及への課題

東京工業大学名誉教授 小長谷 明彦

1. はじめに

2017年に、日本化学会情報化学部会誌において、「AI創薬の可能性について」という特集を企画させていただいた¹⁾²⁾。当時、日本では、2015年に産業技術総合研究所（産総研）において人工知能研究センターが、2016年に理化学研究所（理研）において革新知能統合研究センターが発足したばかりであり、「AI創薬」関連は、大学および国立研究所を中心に化合物データやオミックスデータへの機械学習の適用事例がいくつかあるのみであった。2019年に、エヌ・ティー・エス社より刊行した『人と共生するAI革命（栗原聡監修）』に拙稿「AI活用による創薬の可能性³⁾」を寄稿させていただいた。この時点でも日本での「AI創薬」の活動はほとんど報告されておらず、大学および国の研究機関での研究事例および医療用QAシステム Watson による文献からの医療用コーパス抽出⁴⁾や Google 社の深層学習を用いた肺がん病理画像解析⁵⁾などの先駆的な海外事例を中心に報告した。日本では、2017年にアカデミアとライフサイエンス・IT系企業との連携によるライフインテリジェンスコンソーシアム（LINC）⁶⁾が発足し、日本のAI創薬の発展に貢献してきた。これにより、2021年度のCBI学会年次大会では「デジタルトランスフォーメーションで目指すライフサイエンス革命～AI駆動型の創薬戦略～」を大会テーマとし、AI創薬に関連した研究が多数報告された⁷⁾。本書には、このCBI学会2021年度大会での講演に関連したAI創薬研究が多数報告されていることをあらかじめ申し上げておく。

AI創薬において、深層学習などのAI手法は、創薬標的分子探索、化合物デザイン、薬物作用機序予測、薬効・毒性予測、薬物動態予測、薬物スクリーニング、疾患層別化、ドラッグリポジショニングといった一連の創薬プロセスほぼ全般に適用されている⁸⁾⁻¹⁰⁾。このように、AI手法は、もはや、創薬プロセス全般において普通に使われるIT技術の1つになりつつあると言っても過言ではない。一方、さまざまなAI創薬の事例研究が報告されるにつれ、AIモデルのブラックボックス化や予測結果の解釈の問題を中心に、AI創薬の問題点が議論され始めている¹¹⁾。AI創薬が今後インシリコ創薬の主流となるのか、従来型インシリコ創薬技術の巻き返しが起こるのか、はたまた、両者の折衷がおきるのか、いろいろな意味で2022年が分岐点にあることは否めない。

第1章 日本のAI創薬研究の現状

本稿の構成は以下のとおりである。はじめに、本節第2項において、AI創薬へのアプローチを化学情報学、生命情報学および医療情報学の観点から俯瞰する。次に、第3項において、AI創薬の課題を、「モデル」、「データセット」および「解釈」の観点から整理する。最後に、第4項において、AI創薬の事例を可能性実証、実用性実証および実運用の段階にわけて紹介する。第5項において、今後のAI創薬の可能性についてまとめる。

2. AI創薬へのアプローチ

AI創薬へのアプローチは多様であるが、大別すると、**図1**に示すように、元となる技術の出自に注目すると化学情報学、生物情報学および医学情報学からのアプローチに分類できる。

2.1 化学情報学からのアプローチ

化学情報学でのAI活用の歴史は古く、2000年代より support vector machine (SVM) や random forest (RF) などの機械学習を用いたタンパク質-リガンド予測は盛んに試みられていた¹²⁾。しかしながら、当時は機械学習手法の予測精度が必ずしも高かったわけでもなく、その適用範囲も限られていた。AIが実用的な創薬手法になる可能性があるとして注目されはじめたのは、Ragozaらが2017年に畳み込みニューラルネットワーク (CNN) を用いて、結合ポケットにおける正しい3次元成的リガンド結合ポーズと正しくない結合ポーズの違いが判別できることを示してからである¹³⁾。また、古典的なりカレントニューラルネットワーク (RNN) や long short-term memory (LSTM) の見直しに加えて、グラフニューラルネットワーク (GNN)、変分オートエンコーダー (VAE)、敵対的生成ネットワーク (GAN)、アテンション機構 (attention)、トランスフォーマ (transformer) などの洗練された深層学習手法の出現と

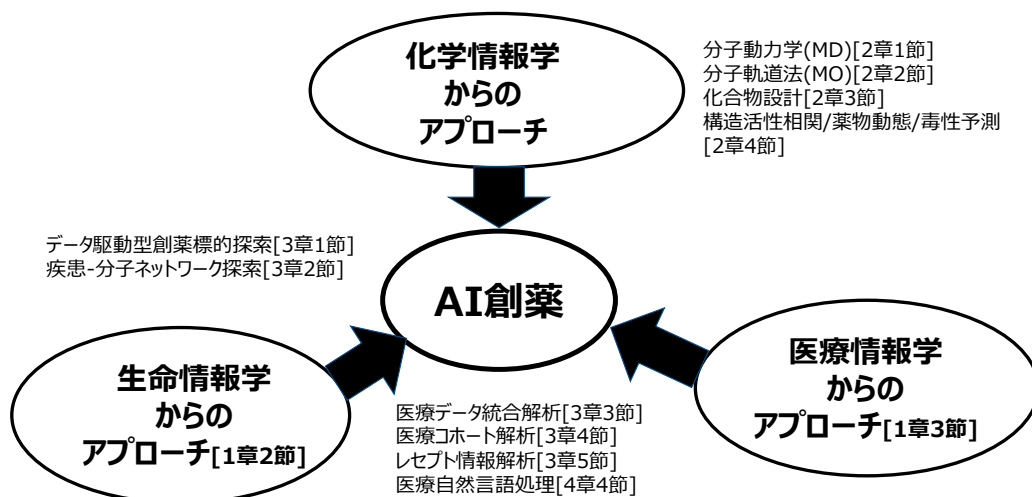


図1 AI創薬へのアプローチ

第1節 AI創薬の可能性と普及への課題

ともに、その適用範囲もタンパク質-リガンド予測だけでなく、創薬プロセス全体に広がっていった¹⁴⁾。本書では下記の先駆的事例を収録している。

池口満徳（横浜市立大学）は、創薬において、分子動力学シミュレーションを熱揺らぎの中で動いている標的タンパク質と結合するリガンド（薬の候補となる小分子）を探索する原理駆動型アプローチとして位置づけ、分子動力学の力場とソフトウェア、標的タンパク質とリガンドとの結合自由エネルギー計算法について紹介するとともに、AI力場など今後の分子動力学シミュレーションとAIとの連携の最新動向について言及している [2章1節]。

福澤薫（大阪大学）と加藤幸一郎（九州大学）は、タンパク質をアミノ酸残基単位に分割して分子の電子状態を計算するフラグメント分子軌道（FMO）法において、個々の原子を中心とする対称関数を用いて周辺原子との相互作用の記述法（ACFSSs）と階層型の深層学習モデルを用いた原子電化予測モデル（HDNNP）を用いることで、コンフォメーション変化に伴う分極や電荷移動の予測を行っている [2章2節]。

金子弘昌（明治大学）は、材料の物性・活性と材料の合成条件や製造条件の関係を記述した数理モデルと機械学習手法を組み合わせた逆解析法に基づく化合物設計法を提案し、並列生命情報処理イニシアティブ（IPAB）で編集した250万化合物データベースでの探索経験について述べている [2章3節]。

船津公人（奈良先端科学技術大学院大学）は、従来の構造活性相関（QSAR）による毒性予測の課題を解決するために、毒性予測を化合物層、細胞イベント層、生体毒性層の3層に分け、動物実験で観測可能な化合物と生体毒性の関係、試験管実験で観測可能な化合物と細胞イベントの関係に対して、化合物の吸収、分布、蓄積、代謝を生物学的薬物動態（PBPKモデル）で予測し、さらに、機械学習モデルを用いて毒性発現のキーとなる要因を評価する、次世代型安全性予測手法の開発（AI-SHIPS）プロジェクトの概要について述べている [2章4節1項]。

山崎浩史（昭和薬科大学）はAI-SHIPSにおいて、結腸がん由来Caco-2細胞を用いて多様な一般化学物質200種の生体吸収に関して消化管透過性を評価し、分子量1000以下の300種の有機化学物質のラット血中濃度推移データからPBPKモデルのパラメータを推定し、重要薬物動態パラメータを検討対象物質の構造データから予測するシステムについて述べている [2章4節2項]。

植沢芳広（明治薬科大学）は、AI-SHIPSにおいて、毒性発現経路の概念から、毒性を分子開始反応に続く、核内受容体、ストレス応答パスウェイ、薬物代謝酵素反応、トランスポーターなどの一連の生理的生化学的過程を経由する複雑な生理現象としてとらえ、ランダムフォレストやブースティングなどの機械学習を用いたQSAR解析で予測し、予測結果から肝毒性、腎毒性および血液毒性を予測モデルならびに深層学習を用いたCYP阻害活性に対する化学構造上の特徴領域を可視化するシステムについて述べている [2章4節3項]。

吉成浩一（静岡県立大学）は、AI-SHIPSの毒性予測、体内動態予測および疑似物質検索を行う「ユーザーシステム」と各種実験・計算データおよび予測モデル情報の管理・閲覧を行う「モデル・データ管理システム」について述べている [2章4節4項]。

第 1 章 日本の AI 創薬研究の現状

庄野文章（奈良先端科学技術大学院大学）は、欧米では、動物愛護の精神 3R（Replacement（代替）、Reduction（削減）、Refinement（改善））の観点から、AI を利用した化学物質の構造等から計算科学的手法により毒性予測する手法の開発が精力的に進められており、国際的な計算科学的毒性予測手法開発（OECD）による取り組み、米国における取り組み、欧州における取り組み、日本での取り組みについて報告している [2 章 4 節 5 項]。

2. 2 生命情報学からのアプローチ

ゲノム学 (genomics)、遺伝子発現解析学 (transcriptomics)、タンパク質発現解析学 (proteomics)、代謝解析学 (metabolomics) などは、遺伝子およびタンパク質の網羅的データを扱う学問であり、末尾に -omics が付くことから、このような網羅的データはオミックスデータと呼ばれる。オミックスデータは世界中から収集され、公共データベースとして無料で公開されているため、さまざまな解析手法やツールが生命情報学で研究されてきた。近年、疾患関連のオミックスデータが蓄積されてきたことから、オミックスデータから創薬標的を探索し、ドラッグリポジショニングなどの創薬プロセスにつなげる統合解析法や自然言語処理などの AI アプローチが注目を集めている [1 章 2 節]。

生命情報学からのアプローチは創薬プロセスの上流となる創薬標的探索が主目的となるため、具体的な創薬に結び付く事例はまだまだ少ないが、本書では、このようなアプローチとして、下記の先駆的事例を収録している。

夏目やよいは、国立研究開発法人医薬基盤・健康・栄養研究所を中心に開発が進められている AI による創薬標的探索を目的とした PRISM「新薬創出を加速する人工知能の開発」プロジェクト（平成 30 年度採択）での事業内容を中心に、肺がんおよび特発性肺繊維症（IPF）の臨床情報データベースの構築、データ駆動型創薬標的探索に資する AI の開発と創薬標的候補の提示、事業成果公開のためのオープンプラットフォームについて述べている [3 章 1 節]。

飯田緑および山西芳裕（九州工業大学）は、表現型が異なる疾患間において分子学的な繋がりがあがるタンパク質間相互作用、基質酵素相互作用、代謝物酵素相互作用、シグナル伝達相互作用などさまざまな相互作用をネットワーク化した生体分子ネットワークにおいて、治療標的分子や治療薬の共有性を予測する機械学習手法について述べている [3 章 2 節]。

2. 3 医療情報学からのアプローチ

医療情報学からの AI 創薬のアプローチはデジタル化が進んだ電子カルテを中心に、臨床検査・処方・医用画像・健康診断等の医療情報、起床疾患やがんの遺伝子検査のためのゲノム情報、ウェアラブルデバイスによる健康情報を加えた医療ビッグデータから、患者ごと病型ごとに最適な個別化医療および個別化ヘルスケアを実現するための創薬が中心となる [1 章 3 節]。本書では、このようなアプローチとして、下記の先駆的事例を収録している。

小島諒介（京都大学）は、医療 AI プロジェクトの現場で必須となる、医療画像や電子カルテなど多様なモダリティを持つ医療ビッグデータを統合して解析するために、グラフニューラルネットワークを活用した医療 AI 基盤の構築について述べている [3 章 3 節]。

第 1 節 AI 創薬の可能性と普及への課題

水野聖士（東北大学）および萩島創一（東北大学）は、世界的な大規模コホートの 1 つである、宮城県・岩手県の地域住民 15 万人をリクルートした東北メディカル・メガバンク計画において、三世代コホート妊婦 2 万 3 千人を対象にした、妊娠高血圧症の早期疾患予測モデルの有用性について述べている [3 章 4 節]。

佐藤淳平（東京大学）らは、日本が蓄積している医療レセプト情報（206 億件）と介護レセプト情報（推定 14 億件）に焦点を当て、これらの医療・介護データベースを利活用して将来の医療・介護サービスの需要を予測するために、将来の患者の要介護状態を推定するための予測モデルの構築・学習および推定フレームワークについて紹介している [3 章 5 節]。

荒牧英治（奈良先端科学技術大学院大学）は、自然言語処理技術を医療に応用した医療言語処理について、電子カルテや患者主導記録などの入力情報、薬効と副作用の頻度情報などの出力情報、形態素解析や固有表現認識などの自然言語パイプライン処理を概説し、深層学習などを用いた大規模言語処理や画像とテキストを組み合わせたマルチ・モダリティについて言及している [4 章 4 節]。

3. AI 創薬技術の課題

AI 創薬は、AlphaFold2¹⁵⁾のようなブレークスルーとなる深層学習モデルの出現が強い牽引力となっている。また、深層学習モデルの予測性能は、訓練に用いたデータセットの量と質に大きく影響される。さらに、深層学習モデルが予測した結果の妥当性を解釈するためにはドメイン専門家の知識や実験が必要となる。本節では、AI 創薬の課題を学習モデル、訓練データセット、結果の解釈の観点から整理する (図 2)。

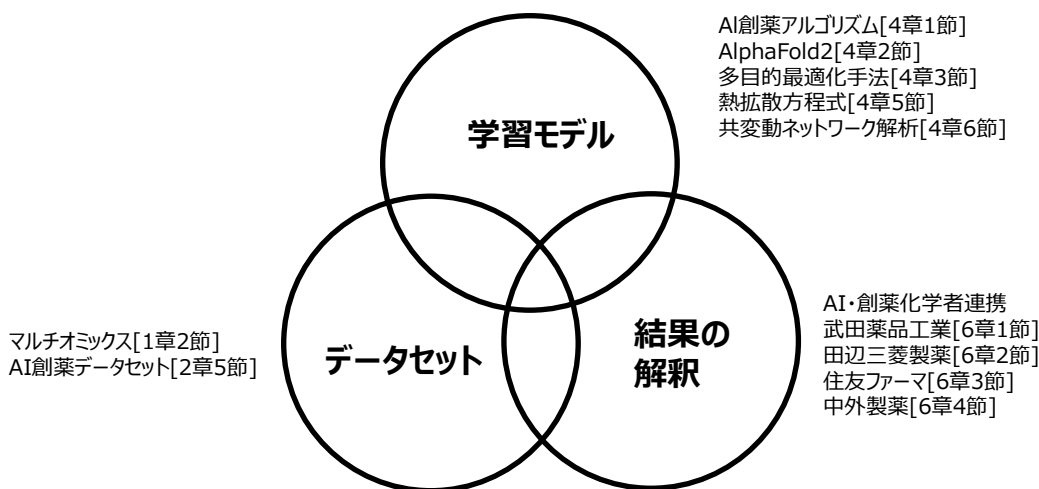


図 2 AI 創薬技術の課題

第 1 章 日本の AI 創薬研究の現状**3.1 AI 創薬における「学習モデル」の課題**

AI 創薬においては、学習モデルの定番が確立したとは言いがたいが、いくつかのトレンドは見えてきている。1つは、対象物の性質を表す化合物の記述子 (descriptor) や官能基のような特徴を積み上げて、性質の良い化合物を探索するルールベースの学習モデルから、前提知識をなるべく用いずに学習のために与えられたデータセットから入出力関係を直接学習する end-to-end 深層学習モデルへの流れである¹⁶⁾。end-to-end 深層学習が注目されている理由の1つとして、近年、顔画像認識や機械翻訳の精度がドメイン固有の知識を使わずに、有用な特徴そのものを入出力関係から直接学習することで大幅に向上したことがあげられる。

AI 創薬の場合、化合物の表現法としては、SMILES のような文字列表現や、分子の結合関係を表すグラフ表現および化合物の原子の配座を示すボクセル表現がよく用いられている。理論的には、このような化合物表現を入力データとして与えたときに、目的となる指標を出力する end-to-end 深層学習モデルを得ることができれば創薬プロセスの大幅な加速化が期待できる。本書では、AI 創薬の学習モデルに関連して、下記の先駆的事例を収録している。

関嶋政和 (東京工業大学) は NPO 並列生物情報処理イニシアティブ (IPAB) 主催の阻害化合物探索コンテスト「コンピュータで薬のタネを創る」でのバーチャルスクリーニング、およびアステラス製薬、長崎大学、産業技術総合研究所との共同研究におけるシャーガス病治療薬探索でのドッキングシミュレーションの経験に基づいて開発した、メディシナルケミストの暗黙知を代替する機械学習モデル VisINet (VISual Inspection NETwork) について述べている。

大上雅史 (東京工業大学) は 2020 年のタンパク質立体構造予測国際コンテスト CASP14 において、驚異的な予測性能を示し、タンパク質構造予測の標準的手法となった AlphaFold2 の背景、予測性能、利用法、複合体構造予測への応用について、利用経験を含めて紹介している [4 章 2 節]。

寺山慧 (横浜市立大学) は、量子化学計算と深層学習やモンテカルロ木探索などの機械学習・最適化手法を活用して現実的な創薬応用を目指した分子生成手法 ChemTS の開発経験に基づき、ターゲットタンパク質に対する主活性だけでなく、合成可能性やオフターゲットタンパク質に対する活性が少ないことや体内動態において望ましい特性を持つ分子を生成するための多目的最適化の重要性について言及している [4 章 3 節]。

今村恵子 (京都大学)、日高中 (武田薬品工業)、儀我美保、儀我美一 (東京大学)、井上治久 (京都大学) らは、病態機序や治療ターゲットがあきらかでなく、タンパク質立体構造やアミノ酸配列からは治療候補薬の探索が難しい難治性疾患の iPS 細胞を用いた表現型スクリーニングにおいて、数百万個の化合物スクリーニングを効率化するために、市販の 1273 個の記述子から選抜した 281 個の記述子からなる探索空間において active 化合物と inactive 化合物を分類する熱拡散方程式を応用した新薬探索支援ツールを開発し、ALS 患者 iPS 細胞から作成した脊髄運動神経細胞を用いて、教師データに含まれていないケモタイプを含むヒット化合物の探索に成功したという [4 章 5 節]。

加納ふみおよび國重莉奈 (東京工業大学) は、創薬の精緻化を目指して、細胞内のタンパク質の量と翻訳後修飾などの質的な変化と局在化情報を含む蛍光抗体染色画像ビッグデータの特

微量の時間的変化から共変動ネットワークを抽出する PLOM-COM 法と正常・病態モデル細胞の作成が可能な再封入（リシール）細胞技術を提案し、その有効性を糖尿病モデル細胞の化合物スクリーニングした結果について報告している [4 章 6 節]。

3.2 AI 創薬における「データセット」の課題

深層学習を含む機械学習では、データセットが学習モデルとともに重要な役割を果たす。データセットはトレーニングセット、バリデーションセットおよびテストセットから構成される。トレーニングセットは学習モデルを訓練するために使用するデータセットであり、通常は大量のデータが必要となる。バリデーションセットは正例か負例かの判定が明確なデータセットであり、学習モデルが正しく学習できたかどうかの検証ならびに学習アルゴリズムの動きを制御するハイパーパラメータのチューニングに用いる。テストセットは最終的に学習モデルで予測したいデータセットおよび学習モデルのパフォーマンスを評価するためのデータセットである。

近年、機械学習による顔画像認識や機械翻訳の精度が向上した理由の 1 つとして、手作業でアノテーションされた大規模データセットが公開され、さまざまな機械学習モデルおよび学習アルゴリズムの研究に利用されたことがあげられる。画像処理分野での著名な公開データセットとして、大規模画像データベース (ImageNet) や手書き数字の画像データベース (MNIST) が知られている¹⁷⁾。また、機械翻訳では、国連で作成された多数の対訳文書などがデータセットとして利用された¹⁸⁾¹⁹⁾。AI 創薬が顔認識や機械翻訳と大きく異なる点として、化合物の薬効を判断する基準が自明ではなく、関連するデータセットも多種多様であり、機械学習を適用する上で考慮すべき点が多いことがあげられる²⁰⁾。本書では、清水祐吾、米澤朋起、池田和由 (慶應義塾大学) らが、良質な化合物構造データを集めた公共データベースである ChEMBL を中心に、データの取得法、選別法、適用範囲の重要性、ラベルの取り扱い方、閾値の取り扱い方、データの量と質の問題、ネガティブデータの生成法など、AI 創薬において考慮すべき点をわかりやすく説明している [2 章 5 節]。

3.3 AI 創薬における「結果の解釈」の課題

一般に、深層学習モデルはブラックボックスであり、なぜそのような推論をしたかの根拠を示すのは困難と言われており、推論の根拠の可視化や説明可能な AI の研究が注目を集めている¹¹⁾。AI 創薬において、結果の解釈が特に重要となるのは、AI が推論した結果が正しいかどうかの判定が自明でなく、高度な専門知識を必要とするからである。例えば、顔認識であれば、人間ならだれでも人の顔とゴリラの顔を間違えることはないし、機械翻訳においても、それなりの語学の知識があれば翻訳結果が正しかったかどうかの判断は瞬時にできる。これに対し、AI 創薬の場合は AI が提示した化合物が本当に薬になるかどうかの判定は AI 研究者では不可能であり、創薬のドメイン知識を持つ創薬化学研究者の判断および実験による評価が必要となる。逆に言えば、AI 研究者と創薬化学研究者がいかにして連携できるかが AI 創薬を実現するための鍵と言えよう。

本書でも、多くの著者が、AI 研究者、データ分析者、創薬化学研究者および医療関係者との連携の重要性について言及している。特に、製薬会社の AI 創薬の事例においては、下記に

第1章 日本のAI創薬研究の現状

示すような課題が提起されている。

武田薬品工業におけるAI創薬では、インシリコサイエンティストとメディシナルケミストが協働し、Design-Make-Test-Analysis (DMTA) サイクルを加速化することでAIが提案する化合物デザインが起点となり、リード化合物の創出や化合物最適化が行えるようになったという [6章1節]。

田辺三菱製薬におけるAI創薬では、多様な指標に対する最適化が求められる創薬研究においては能動学習等による一部の特性予測だけでは創薬化学研究者を上回る効果は期待できないため、構造発生AIと構造未知の分子シミュレーションを組み合わせた化合物デザインに期待しているという [6章2節]。

住友ファーマのAI創薬では、AIを実用して成果を創出するためのポイントとして、解決したい課題と目的の明確化、実験研究者と計算研究者の連携、実験研究者と計算研究者の双方がAIの長所と短所を十分理解していることが重要という [6章3節]。

中外製薬のAI創薬では、研究者が合理的に配列を設計することが困難な多数の部位に変異を導入した抗体医薬品の配列設計には、抗体配列の特徴を深層生成モデルで学習した配列探索技術が有用であり、画像認識技術を用いた薬剤安全性評価および薬効評価においてはAI技術が専門家と比較可能な精度を示したという [6章4節]。

4. AI創薬技術の社会実装

AI創薬技術の進展は著しく、2022年の時点で、約160種の小分子が前臨床および臨床研究に、15種の小分子が治験フェーズ1および2があると報告されている²¹⁾。本節では、AI創薬技術を可能性実証、実用性実証、実運用の3段階に分け、それぞれでの事例について紹介する(図3)。可能性実証は論文として公開された先端的技术であり、実用性実証は社内利用および共同研究での成果報告である。実運用は、他社が利用可能なサービスとして展開されている技術の紹介である。

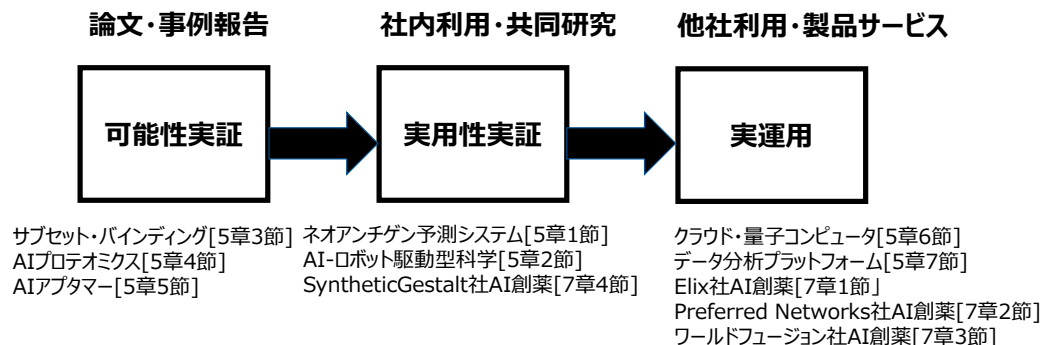


図3 AI創薬技術の社会実装

4.1 可能性実証

AI 創薬において、独自に測定した実験データセットから創薬に関連する情報を抽出するためには多くの困難を乗り越える必要があり、その経験から得られる知見は貴重であり、かつ、他の研究にも有用である。本書では、このような先端的 AI 創薬研究として、大学および国の研究所を中心とした下記の事例を収録している。

上田修功（理研）および夏目やよい（医薬基盤・健康・栄養研究所）は、質的データと量的データが混在し、精密医療の実現に必要な患者層別化が困難な状況において、患者群をグルーピングする代わりに、同種データ間および異種データ間の属性の共起情報に着目し、相関ルールマイニング技術とファジー論理を組み合わせたサブセット・バインディング (SB) 法による層別化を提案し、その有効性をアセトアミノフェンの肝障害に関する遺伝子発現データで確認した結果について報告している [5 章 3 節]。

林宜宏（東工大）は、血液（血清）を用いた二次元電気泳動画像データセットと転移学習アルゴリズムを用いて目視では判別しにくい糖鎖修飾パターンを用いて 98% 以上の精度で診断することに成功した結果について報告している [5 章 4 節]。

浜田道昭（早稲田大学）は、RNA アプタマーを進化的に創成する SELEX 法をベースに、局所的な構造と配列情報に基づくアプタマー配列のランキング法と入力データに含まれない配列を変分オートエンコーダと隠れマルコフモデルを用いて生成する確率的生成モデルについて述べている [5 章 5 節]。

4.2 実用性実証

AI 創薬により本当に薬の開発ができるのかどうかは、製薬企業や創薬研究者のみならず、IT 企業ならびに AI 研究者にとっても大きな関心事である。さらに言えば、AI 創薬ベンチャーにとっては死活問題と言っても過言ではない。本書では、実用性実証段階にある AI 創薬として下記の事例を収録している。

北村哲，山下慶子，小野口和英（日本電気株式会社）らは、個別化医療の観点から、患者の腫瘍組織および正常組織を採取して、全エクソームシーケンスおよび RNA シーケンスを解析し、グラフ関係性学習および能動学習を活用して、精密かつ効率的に候補エピトープの抽出、免疫応答に関するスコアリング、エピトープのランキングを行うネオアンチゲン予測法を用いて開発した個別化がんワクチン TG4050（米国およびフランスで卵巣がん第 1 相臨床試験中、イギリスおよびフランスで HPV 陰性頭頸部がん第 1 相臨床試験中）について述べている [5 章 1 節]。

興野悠太郎（株式会社 MOLCURE）は、AI が仮説生成し、ロボットが実験し、得られたデータを元に仮説を更新し、というサイクルを繰り返す医薬品探索システム“BIO LAB DX LOOP”と AI フレンドリーなスクリーニング自動ロボット群“HAIVE”を創薬の Target-to-hit および Hit-to-lead のステージに適用し、海外製薬企業との中・高分子医薬品開発の提携プロジェクトにおいて、提携先の保有技術で探索した抗体を凌駕する抗体の探索に成功した

第1章 日本のAI創薬研究の現状

という [5章2節]。

神谷幸太郎 (SyntheticGestalt 株式会社) は、「発明を量産する人工知能の開発」をビジョンとして、創薬活動の支援ではなく、候補化合物の発見を目指した機械学習技術の開発において、技術開発における設計原則と開発方針について論じている [7章4節]。

4.3 実運用

AI創薬は技術としては未だ発展途上であり、現在、実運用段階にあるのは、AI創薬に必要な大規模計算サービスの提供およびインシリコ創薬のための機械学習自動化パッケージである。一方、AIベンチャーを中心に、自社のAI活用サービスに創薬向けのサービスを組み込む動きも本格化しており、今後の発展が期待される。本書では、下記のサービスを収録している。

宮本大輔と宇都宮聖子 (アマゾン ウェブサービス ジャパン合同会社) は、創薬分野でのクラウドの活用例として、228万仮想CPUを用いた10億化合物を超えるドッキングシミュレーションや2万件の分子シミュレーションなどの創薬関連高性能計算が可能な Amazon Elastic Compute Cloud や量子計算機と古典計算機をハイブリッドに組み合わせた量子化学計算などが可能な Amazon Braket について紹介している [5章6節]。

大島貴弘 (インフォコム株式会社) は、創薬分野において10年以上の歴史を持ち、さまざまな機能モジュールをワークフローとして組み合わせることで医薬品設計パイプラインを構築する、データ分析プラットフォーム KNIME (Konstanz Information Miner) において、完全自動化によるブラックボックス化を避けるために、ウェブブラウザを介してインタラクティブにシステム構築可能な機械学習自動パッケージおよびAI創薬適用事例について紹介している [5章7節]。

Casey Galvin (株式会社 Elix) は、データや実験の管理、モデル開発、意思決定に役立つ分析・可視化ツールのためのAI創薬プラットフォーム「Elix Discovery」について、そのモジュール構造の考え方、ならびに、ノーコードプラットフォームや Automated Machine Learning など、AIの非専門家でもAIを使いこなせるためのシステム構築について述べている [7章1節]。

力丸健太郎 (株式会社 Preferred Networks) は、AI創薬が実活用のフェーズに入ったという観点から、Python や Python ライブラリのバージョンに依存せずに、自社開発したAIを含むさまざまなAI手法を使って創薬を支援する化合物生成モジュールとスコア計算モジュールとからなる分子設計プラットフォームについて述べている [7章2節]。

木村敏郎 (株式会社 ワールドフュージョン) は、ワールドフュージョン社が開発している、「遺伝子」、「タンパク質」、「化合物」、「疾患」、「作用機序」などのオントロジーを軸に、さまざまなライフサイエンス公共データベースから作用機序新規語彙情報を、深層学習手法 BERT-NER により収集した創薬支援情報検索システム LSKB について述べている [7章3節]。

5. まとめ

日本においても、AI 創薬は急速に発展し、大学および国の研究所のみならず、製薬企業、IT 企業ならびにベンチャー企業においても、AI を活用した創薬の研究開発が精力的に進められている。AlphaFold2 のような AI 技術の登場は、AI 創薬の発展をさらに後押しするものであり、AI 創薬を活用した COVID-19 のための薬の開発も報告されている²²⁾。一方、IBM は創薬向けの Watson のサービスを 2020 年に停止した²³⁾。このことは、専門家の知識を知識ベースとして形式化するアプローチについては、多くの研究課題が残されていることを意味している。近年、機械学習法として end-to-end learning が注目されているが、end-to-end learning は入力と出力の関係は学習しても、その「意味」まで「理解」しているわけでないことに留意が必要である²⁴⁾。機械学習の効用と限界を理解しつつ、AI が抽出した形式知と専門家の暗黙知の循環から新たな知識を創発する仕組みが AI 創薬に不可欠と考えている。今後の AI 創薬の可能性に向けて、さらなる発展を期待したい。

文 献

- 1) 小長谷明彦: 日本化学情報化学部会誌, **35** (2), 156 (2017).
- 2) 小長谷明彦: 日本化学情報化学部会誌, **35** (2), 198 (2017).
- 3) 小長谷明彦, 栗原聡監修: 人と共生する AI 革命, エヌ・ティー・エス, 221-235 (2019).
- 4) C. Wang and J. Fan: Proc. of the 52nd Computational Linguistics, 828-838, Baltimore, Maryland, USA, June 23-25 (2014).
- 5) N. Coudray et al.: *Nat. Med.*, **24**, 1559 (2018).
- 6) ライフインテリジェンスコンソーシアム (LINC): <https://linc-ai.jp/overview/objective.html>
- 7) 中嶋久, 嶋田朋嘉: CBI 学会誌, **9** (4), 10 (2020).
- 8) X. Yang et al.: *Chem. Rev.*, **119**, 10520 (2019). AI 創薬全般のレビュー
- 9) R. Gupta et al.: *Mol. Divers.*, **25**, 1315 (2021). AI 創薬関連ソフトウェアのレビュー
- 10) M. K. Tripathi et al.: *Mol. Divers.* (2021). <https://doi.org/10.1007/s11030-021-10256-w> AI 創薬関連のデータソース, データセットのレビュー
- 11) J. Jimenez-Luna et al.: *Expert Opin. Drug Discov.*, **16** (9), 949 (2021). AI 創薬の現在の課題と今後に対するオピニオン
- 12) A. Dhakal et al.: *Brief. Bioinform.*, **23**, 1 (2022). 蛋白質-リガンド結合に関する AI 手法のレビュー
- 13) M. Ragoza et al.: *J. Chem. Inf. Model*, **57**, 942 (2017).
- 14) C. Choudhury et al.: *Drug Discov. Today* (2022). <https://doi.org/10.1016/j.drudis.2022.03.006> AI 手法を用いたドラッグリポジショニングの報告
- 15) J. Jumper et al.: *Nature*, **596**, 583 (2021). AlphaFold2 の論文
- 16) F. Boniolo et al.: *Expert Opin. Drug Discov.*, **16**, 991 (2021). 生成モデルや end-to-end への流れを紹介している論文
- 17) C. G. Northcutt et al.: *arXiv*, arXiv:2103.14749v4 [stat.ML] (7Nov2021). 著名な公開データセットにエラーが多いという論文
- 18) M. Johnson et al.: *arXiv*, arXiv:1611.04558v2 [cs.CL], (21Aug2017). Google の深層学習機械翻訳の論文
- 19) O. Bojar et al.: Proc. of the Ninth Workshop on Statistical Machine Translation, 12 (2014). 統計機械翻訳の評価に使われたデー

第 1 章 日本の AI 創薬研究の現状

- タセット
- 20) H. Zhu: *Annu. Rev. Pharmacol. Toxicol.*, **60**, 573 (2020). AI 創薬のためのビッグデータ
- 21) M. K. P. Jayatunga et al.: *Nat. Rev. Drug Discov.*, **21**, 175 (2022). 創薬パイプラインに登録されている AI 創薬の速報
- 22) A. K. Arshadi et al.: *Front. Artif. Intell.*, **3** (65), 1 (2020). COVID-19 に使われた AI 創薬のレビュー
- 23) 5737-B19 IBM Watson for Drug Discovery, IBM United States Sales Manual Revised; (2020). IBM Watson 事業中止のレポート
- 24) J. Jimenez-Luna et al.: *Nat. Mach. Intell.*, **2**, 573 (2020). 創薬における Explainable artificial intelligence のレビュー論文

この先をご覧いただくには、パスワードが必要です。

制限つきPDFで全ページをご覧いただけます。
(制限内容：閲覧期間の設定、コピーやプリントの禁止など)

- ・ PDFの閲覧

「パスワード」と「専用のビューア」（無料）が必要です。
費用は一切かかりません。

※WindowsのPCでのみご覧いただけます。予めご了承ください。

- ・ パスワード ※電子試読ページよりお申込みください

<https://www.nts-book.com/ntsの電子試読>

ページ下部にお申込みフォームがあります。

右のQRコードからも
電子試読ページにアクセス
いただけます。



- ・ ビューアのダウンロード

PDFは、株式会社スカイコムの SkyPDF Viewer（無償のPDFビューア）をダウンロードしてご覧いただけます。

※Adobe Acrobat Readerなど他のPDF閲覧アプリケーションではご覧になれません。

SkyPDF Viewer 無償ダウンロード：

<https://www.skycom.jp/free/>

索引

英数・記号	
28 日間反復投与毒性試験	65
2 次構造	258
3D-RISM	40
3 層モデル	66
AAE (敵対的オートエンコーダー: Adversarial Autoencoder)	332
ABINIT-MP	47
ADASYN	101
ADL スコア	148
ADME	67
ADMET	129, 323
ADMET 予測	289
AfDesign	173
AI	133, 245
AI-SHIPS	65
AI-SHIPS 統合的毒性予測システム	85
AI-SHIPS プロジェクト	91
AI- ロボット駆動型科学	221
AI 創薬	3, 30, 256
AI 創薬技術の社会実装	10
AI 創薬のプラットフォーム	31
AI ドラッグデザイン	296
All of Us	134
AlphaFold2	166, 311
AlphaFold-Multimer	173
AlphaFold Protein Structure Database	167
AMBER	38
amplification bias	224
AOP	67, 77, 93
= 毒性 (有害性) 発現経路	
ARCH-Tox	93
AUC	161
Autoencoder	19
Automated ML (AutoML)	326
BERT	191, 338
BERT-MNLI	339
BFD (big fantastic database)	166
CASP14	166
chance correlation-excluded MAE	59
CHARMM	38
ChEMBL	96, 159
ChEMBL-DB	338
ColabFold	169
CompTox	92
CompTox-PK-CvTdb	92
Computer Aided Drug Design (CADD)	267
Convolutional Neural Network (CNN)	298
Cross Validation	151
DAVID	16
DeepSnap	81
DESMOND	39
Directory of Useful Decoys, Enhanced (DUD-E) データセット	161
DMTA サイクル	296
DNA	246
DNA シーケンサ	166
Doc2Vec	317
DPC	187
Druggability 評価	289
Edge Contribution value (ECv)	18
EF1%	162
entity linking	19
EU-ToxRisk	92
F-value	151
faster region convolutional neural network	320
Feasibility Study; FS	250
FEP	41
FMO-AI 力場	52
FMODB	49
FMODBID	50
FMO 創薬コンソーシアム	47
FMO データベース	47
FMO 法	47
FPGA	267
Gaussian mixture regression	60
generative topographic mapping regression	60
GeneRIF	339
GENESIS	39
Genomics England	28
Github	111
GNN (graph neural network)	20, 270
Google Colaboratory	169
GoogLeNet	319

GPU	37, 266	named entity recognition (NER)	19
Gradient Boosting Decision Tree	151	NCBI	338, 339
Graph Neural Network (GNN)	20, 270	neural network	52
gREST	43	New Approach Methodologies; NAM	92
GROMACS	38	Next Generation Sequencing (NGS)	265
GWAS	115	NGRA	93
HCS	298	NISQ	271
HDE	194, 288	NOEL	86
= 熱拡散方程式		OECD	66, 91
HESS	93	OECD (Q) SAR ツールボックス	92
HESS DB	93	OMIM	115
HHSearch	165	OpenMM	38
High Performance Computing (HPC)	265	OPLS	38
High Throughput Screening (HTS)	160	over-sampling	101
Hit-to-lead	223	PAINS	98
HL7 FHIR	27	PBPK	67
HL7 version 2.5	24	PBPK モデル	71, 74
HPCI	49	= 生理学的薬物動態モデル	
HT-SELEX	256	pChEMBL 値	98
Infrastructure as Code (IaC)	266	pLDDT	167
<i>in silico</i> 技術	83	Precision	151
<i>in vitro</i> 試験	87	predicted aligned error (PAE)	167
<i>in vivo</i> 毒性試験	88	PRISM: Public/Private R&D Investment Strategic Expansion Program	107
iPS 細胞	193	Protein Data Bank (PDB)	159, 165
Kaggle	329	= タンパク質構造データバンク	
Kernel matrix regression	119	Proximity	116
KGE (knowledge graph embedding)	20	PubChem	96
KMR	119	PubMed	338
LINC	51	Python	330
LSKB	337	QSAR	129
LSTM	314	QSP	299
MD-SAXS	43	Quantitative Systems Pharmacology	299
MDD	109	R&D productivity モデル	300
MID3	299	R&D 期間の短縮	301
MLOps	269	R&D コストの削減	301
MM-GBSA	40	Random Forest	161
MM-PBSA	40	RDF	130
MMseqs2	170	REACH	91
MNIST	252	Recall	151
Model-informed drug discovery and development	299	relation detection (RD)	19
Molecular Activity チャレンジ	30	RiskHUNT3R	93
MP-CAFEE	41	RNA アプタマー	255
mRNA	246	RNN	314
NAMD	38	ROC 曲線下面積 (ROC-AUC)	79

SARS-CoV-2	51	遺伝子発現変動プロファイル	295
SCARA アーム	228	遺伝的アルゴリズム	327
SELEX	255	移動経路の最適化	228
SEURAT-1	92	医薬品探索システム	223
SHAP 値	16	医薬品特性	98
SIEVE-Score	161	医薬品分子	221
SMOTE	101	医療 AI	125
SNP (一塩基多型)	340	医療言語処理	185
sparse Gaussian graphical models	202	医療情報	23
SS-MIX2 ストレージ	24	医療情報学からのアプローチ	6
Structure-Based Drug Design (SBDD)	37	医療情報と介護情報に基づく将来の要介護状態の 推定	146
support vector machine	319	医療ビッグデータ	23, 125
SVHN	252	インシリコスクリーニング	39
Target-to-hit	223	インシリコ創薬	306
TargetMine	16	インターフェース	127
Tox21 プログラム	78	インフォマティクス	306
TSCA	91	ウェットラボワーク	229
UK Biobank	28, 134	エクソソーム	110
under-sampling	101	エネルギー分割法	48
VAE (変分オートエンコーダー: Variational Autoencoder)	258, 332	エンベディングプロパゲーション	217
VisINet (VISual Inspection NETwork)	161	オープンプラットフォーム	111
word2vec	296	オブジェクトストレージ	267
XY プロッタ	227	オミックスデータ	15, 235
X 線結晶構造解析	160, 165		
y-randomization	59		
ZINC	96		

和文

あ 行

アイテムセット	236	カーネル行列回帰	119
アクティブラーニング	324	介護情報	146
アゴニスト	42	外部検証	78
アセトアミノフェン	241	解離定数	315
アプリアリアルゴリズム	237	化学情報学からのアプローチ	4
アミノ酸	314	学習データ	308
アラート構造	98	学習モデル	8
安全性	307	確信度	236
アンタゴニスト	42	拡張アンサンブル	43
暗黙知	13, 126, 226	隠れ範疇	250
異種データ	235	隠れマルコフモデル	258
異常検知	102	化合物空間	98
一般化学物質の毒性予測	65	化合物空間情報	72
遺伝子	245	化合物構造発生	289
遺伝子工学	222	化合物構造発生 AI	296
		化合物作用機序	210
		化合物スクリーニング	193, 209
		化合物生成モジュール	330

化合物の標準化	102	蛍光抗体法	202
化審法	65, 71	計算機クラスター	331
画像解析	202, 298	計算コスト	162
画像解析 AI	298	形式知	226
仮想患者モデル	299	形態素解析	189
画像診断	250	血液毒性	66
仮想配列ライブラリ	317	結果の解釈	9
仮想ライブラリー	96	結合自由エネルギー	39
カテゴリカル分布	315	結合親和性	96
がん遺伝子パネル検査	29	血清	251
関係抽出	189	欠損値補填	110
患者主導記録	186	ゲノム解析	266
患者層別化	108, 235	ケミカルスペース	71
関心領域	319	ケモインフォマティクス	57, 129
感度	79	公共データベース	96
肝毒性	66, 243	高次蛋白質凝集体	201
官能基	327	合成アクセシビリティ	334
官民研究開発投資拡大プログラム	107	構造活性相関	96
がん免疫療法	213	構造化データ	110
機械学習	74, 77, 119, 149, 157, 245, 269, 313, 323	構造ベースバーチャルスクリーニング (SBVS)	158
機械学習力場	52	抗体医薬品	313
記述子	96	抗体スクリーニング実験	225
忌避部分構造	332	抗体分子設計	221
吸収速度	71	抗体ライブラリ	314
強化学習 (Reinforcement Learning, RL)	176, 332	公的医療保険制度	145
教師あり学習	120	公的介護保険制度	145
教師データ	245	骨格構造	327
教師なし学習	119	個別化がんワクチン	213
共変動ネットワーク	200	個別化ヘルスケア	29
共有性	120	コホート研究	134
局在	204	固有表現認識	189
近接度	116	コンタクト予測	166
クーロン力	161	コンテナ	266
偶然の相関	59	根本的治療	252
組み合わせ爆発	317		
クライオ電子顕微鏡	265	さ 行	
クラウド	265	再現率	120
クラス活性化マッピング (Class Activation Mapping; CAM)	81	再生医療	252
クラス別分類	149	細胞画像診断システム	206
グラフ関係性学習	217	材料設計	57
グラフニューラルネットワーク	128	サブセットバインディング	235
京	49	作用機序 (MoA)	100, 338
蛍光抗体染色画像ビッグデータ	200	サンプル輸送	226

シグナル伝達	199	正規化	189
時系列可視化	189	生成モデル	325
次元の呪い	103	生体	245
試行回数の削減	301	生体情報	246
自己教師あり学習	327	生体内分子間の相互作用	123
自己符号化器	320	生体分子	113
事実性判定	189	製品パイプラインの拡充	301
支持度	237	精密医療	143
次世代医療基盤法	27	生命情報学からのアプローチ	6
次世代シーケンサー	214, 221, 314	生理学的薬物動態 (PBPK) モデル	71, 74
次世代創薬 AI 開発 (DAIIA)	292	セミンタクト細胞リシール技術	207
自然言語処理	110, 185, 295, 323	選択的阻害剤	179
疾患原因遺伝子	114	相関	202
疾患の類似性	115	相関ルール	236
疾患マーカー	206, 208	早期疾患予測モデル	137, 141
疾患モジュール	114	早期診断	251
実験自動ロボット	226	早期リスク予測モデル	133
質量分析	246	相互作用エネルギー	159
シミュレーション	306	相互作用フィンガープリント	160
受信者動作特性曲線下面積 (ROC-AUC)	78	創薬化学フィルター	98
主成分分析	99	創薬研究	123
手段的日常生活動作	147	創薬標的探索	108
進化分子工学	222	創薬プロセス	222
シングルオミックス	15	速度論パラメータ	42
人工知能	323	組織病理学的観察	243
人工的なネガティブデータの生成	102	ソフトウエア階層モデル	230
審査支払機関	146	ソフトセンサー	63
深層学習	126, 166, 175, 249, 329	損失関数	101
深層学習用フレームワーク	330		
深層生成モデル	314	た 行	
深層ニューラルネットワーク	30	退院サマリ	186
腎毒性	66	大規模スクリーニング	231
新薬創出を加速する人工知能の開発	107	ダイクストラ	229
診療記録	186	代謝消失速度	71
診療情報	146	対数尤度	315
診療データ	235	体内動態予測	85
スーパーコンピュータ	169	多因子疾患	133
水素結合	161	多クラス分類	149
スクリーニング実験	222	多重配列	171
スケールアウト	233	多重配列アラインメント (multiple sequence alignment; MSA)	165
スコア計算モジュール	331	畳み込みニューラルネットワーク	319
スパースモデリング	299	多目的最適化	179
スペクトル解析	63	タンパク質	245
スポット	246		
正解ラベル	149		

タンパク質-ペプチド複合体	173	毒性発現最小量	75
タンパク質構造データバンク (protein data bank; PDB)	159, 165	毒性予測	85
タンパク質デザイン	173	特発性間質性肺炎	109
蛋白質ネットワーク	199	特発性肺線維症 (IPF)	108
知識グラフ	20, 130	特徴量	58, 102, 149
知識グラフデータベース	295	特徴領域可視化システム	77
直接的逆解析法	58	特徴量選択	16
中分子医薬品	255	ドッキング	331
治療標的分子	113	ドッキングシミュレーション	158, 177, 266
治療薬	113	トピックモデル	18
データ駆動型創薬	293	ドメイン知識	127
データサイエンティスト	127	ドラッグライクネス	98
データ数の釣り合い	101	ドラッグリポジショニング	17, 130, 295, 342
データセット	9, 151		
データの取得源	95	な 行	
ディープニューラルネットワーク	217	二次元電気泳動	246
ディープラーニング	81, 157, 293, 313	ニューラルネットワーク	324
低分子化合物情報	157	ネオアンチゲン	213
定量的構造活性相関	77	ネガティブデータ	162
適応的実験計画法	58	熱拡散方程式 (HDE; Heat Diffusion Equation)	194, 288, 290
適合率	120	ネットワーク生物学	113
適用範囲	99	ノーコードプラットフォーム	326
デジタルイノベーション	287	濃縮率	258
デジタルエコシステム	288	能動学習	297
デジタルツイン	232	能動学習法	217
デプロイ	270		
デルタリンク機構	227	は 行	
転移学習	249, 299	バーチャルスクリーニング	129, 158, 266, 331
電荷	246	バイオインフォマティクス	214, 295
電子温度制御機能	229	バイオパニング	314
電子カルテ	185	バイオバンク	251, 309
電子顕微鏡	165	バイオバンク・ジャパン	28
電子レセプト情報	145	バイオマーカー	245, 309, 340
テンプレートモデリング	165	肺がん	108
糖鎖修飾	253	背景知識	126
投資対効果の定量化	300	敗血症	250
東北メディカル・メガバンク	135	ハイコンテンツスクリーニング	298
東北メディカル・メガバンク計画	28	ハイスループットスクリーニング (HTS)	96, 323
特異度	79	配列情報	157
読影レポート	186	パブリッククラウド	169
毒性 (有害性) 発現経路 (Adverse Outcome Pathway; AOP)	67, 77, 93	バランス精度 (Balanced Accuracy)	78
毒性試験関連データベース	88	汎化性	308
毒性発現機序	87		

反復投与毒性	86
汎用人工知能	311
非構造化データ	110
ヒット化合物探索	157
ヒット率	160
評価関数	159
病型分類	138
標準化	23
病態モデル細胞	208
標的ターゲット同定	289
標的タンパク質同定	157
標的の定義範囲	97
標的分子の同定	323
頻出パターンマイニング	235
ブースティング	78
ファージディスプレイ	223, 314
ファースト・イン・クラス	305
ファジー相関ルールマイニング	237
ファンデルワールス力	161
フィンガープリント	102
フェノタイピング	138
富岳	38, 47
不均衡	149
複合体構造予測	173
服薬指導	186
物性値	74
物性予測	326
部分構造フィルター	331
フラグメント	327
フラグメント間の相互作用エネルギー	48
フラグメント分子軌道	47
プレジジョン・メディシン	293, 305
プロセスインフォマティクス	57
プロセス管理	57
プロセス設計	57
プロテオミクス	245
分子生成	326
分子生物学	254
分子設計	57, 329
分子設計プラットフォーム	330
分子動力学シミュレーション	37
分子ネットワーク	113
分子量	246
分布容積	71
ベイジアンネットワーク	17

ベイズ最適化	58, 260
変分オートエンコーダ	258, 332
ホモロジーモデリング	165
ポリファーマコロジー	291
翻訳後修飾	248
翻訳後修飾情報	204

ま 行

マイクロサービス	331
膜タンパク	225
膜透過係数	74
マテリアルズ・インフォマティクス	57, 66
マネージドサービス	265
マルコフ状態モデル	42
マルチオミックス	15
マルチモーダル共変動ネットワーク	205
マルチモダリティ	126
ミニマム項目セット	26
無細胞ディスプレイ	223
無毒性	74
メタゲノム	166
メタデータ	251
メンバーシップ関数	237
モダリティ	222
モデル・データ管理システム	85
モデルの逆解析	58
モデルの適用範囲	58
モンテカルロ木探索	175

や 行

薬物動態	307
薬物標的遺伝子	114
薬効評価	199
ユーザーシステム	85
有害性評価支援システム統合プラットフォーム 「HESS」	68
要介護状態	145
要介護状態の推定方式	145
要介護度	145
要介護認定	147
要支援	147
予測誤差	59
予測モデル	149, 317, 325

ら 行

ライブラリ構築	222
ライブラリデザイン	223
ランク学習	99
ランダムフォレスト	78
リード化合物	307
リード抗体	313
罹患	246
リガンドスペースの可視化	291
リガンドベースバーチャルスクリーニング (LBVS)	158
力場	37
リシール細胞	207
立体構造	159, 165
立体構造情報	157
立体構造変化	42
立体構造予測	165

リバース・トランスレーショナル	309
リフト値	237
量子コンピュータ	265
リレーショナルデータベース	97
リン酸化	248
臨床検査	251
臨床試験シミュレーション	299
臨床情報	309
ルール・オブ・ファイブ	98
類似化合物検索	289
類似物質検索	85, 88
レプリカ交換法	43
連合学習	328

わ 行

ワークフロー	267, 317
--------	----------

革新的 AI 創薬

医療ビッグデータ、人工知能がもたらす創薬研究の未来像

発行日 2022年7月22日 初版第一刷発行
監修者 小長谷 明彦
発行者 吉田 隆
発行所 株式会社エヌ・ティー・エス
〒102-0091 東京都千代田区北の丸公園 2-1 科学技術館 2階
TEL.03-5224-5430 <http://www.nts-book.co.jp>
印刷・製本 藤原印刷株式会社

ISBN978-4-86043-788-6

© 2022 小長谷 明彦, 他

落丁・乱丁本はお取り替えいたします。無断複写・転写を禁じます。定価はケースに表示しております。本書の内容に関し追加・訂正情報が生じた場合は、(株)エヌ・ティー・エスホームページにて掲載いたします。

*ホームページを閲覧する環境のない方は、当社営業部(03-5224-5430)へお問い合わせください。



関連図書

	図書名	発刊年	体 裁	定価(10%税込)
1	快眠研究と製品開発、社会実装 ～生体計測から睡眠教育、スリープテック、ウェルネス、地域創生まで～	2022年	B5 812頁	55,000円
2	プレジジョン・メディシン ～ビッグデータの構築・分析から臨床応用・課題まで～	2018年	B5 406頁	50,600円
3	生物の寿命延長 ～老化・長寿命の基盤研究最前線～	2022年	B5 460頁	59,400円
4	ニューロモルフィックコンピューティング ～省エネルギーな機械学習のハードウェア実装に向けて～	2022年	B5 334頁	33,000円
5	医用工学ハンドブック	2022年	B5 544頁	30,800円
6	糖質・甘味のおいしさ評価と健康・調理・加工	2022年	B5 360頁	39,600円
7	生体ひかりイメージング 基礎と応用	2021年	B5 628頁	59,400円
8	アロマプロフィール解析による香りの科学 ～商品開発に向けたニオイ受容のしくみが導く香気複合臭解析～	2021年	B5 292頁	39,600円
9	ミトコンドリアダイナミクス ～機能研究から疾患・老化まで～	2021年	B5 458頁	53,900円
10	食品分野におけるメタボリックプロファイリング ～成分、産地、品質評価・向上～	2021年	B5 326頁	44,000円
11	進化する皮膚科学 ～機能研究・臨床・評価・製品開発の最前線～	2021年	B5 426頁	55,000円
12	代替プロテインによる食品素材開発 ～植物肉・昆虫食・藻類利用食・培養肉が導く食のイノベーション～	2021年	B5 322頁	46,200円
13	不定愁訴の統合生理学と商品開発 ～美・健康・老化予防と有効性評価試験～	2021年	B5 222頁	38,500円
14	食品テクスチャーの測定とおいしさ評価 ～食品構造とレオロジー、咀嚼・嚥下感覚、機器測定・官能検査、調理・加工～	2021年	B5 258頁	39,600円
15	抗ウイルス・抗菌製品開発 ～基礎、作用メカニズムから評価、認証、商品化まで～	2021年	B5 388頁	46,200円
16	データ駆動型材料開発 ～オントロジーとマイニング、計測と実験装置の自動制御～	2021年	B5 290頁	57,200円
17	生命金属ダイナミクス ～生体内における金属の挙動と制御～	2021年	B5 564頁	59,400円
18	ダイレクトリプログラミング ～再生医療の新展開～	2020年	B5 316頁	49,500円
19	Excelによる生体信号解析 ～心電図、脈波、血圧～	2020年	B5 120頁	19,800円
20	ヒトマイクロバイーム Vol.2 ～解析技術の進展とデータ駆動型・ターゲット機能型研究最前線～	2020年	B5 624頁	53,900円
21	スマート医療テクノロジー ～AI、ビッグデータの利活用による次世代手術システムと医療経営～	2019年	B5 368頁	58,300円
22	人と共生するAI革命 ～活用事例からみる生活・産業・社会の未来展望～	2019年	B5 480頁	52,800円