

Title	Unveiling Small Molecules Targeting CAG Trinucleotide Repeats: from Molecular Design to in-silico Approaches
Author(s)	Chen, Qingwen
Citation	大阪大学, 2023, 博士論文
Version Type	
URL	<a href="https://hdl.handle.net/11094/93023">https://hdl.handle.net/11094/93023</a>
rights	
Note	やむを得ない事由があると学位審査研究科が承認したため、全文に代えてその内容の要約を公開しています。全文のご利用をご希望の場合は、 <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈a href="https://www.library.osaka-u.ac.jp/thesis/#closed"〉</a> 大阪大学の博士論文について <a href="https://www.library.osaka-u.ac.jp/thesis/#closed">〈/a〉</a> をご参照ください。

***Osaka University Knowledge Archive : OUKA***

<https://ir.library.osaka-u.ac.jp/>

Osaka University

## 論 文 内 容 の 要 旨

氏 名 ( QINGWEN CHEN )

## 論文題名

Unveiling Small Molecules Targeting CAG Trinucleotide Repeats:  
 from Molecular Design to *in-silico* Approaches  
 (CAGトリプレットリピートを標的とした低分子の探索：  
 分子設計からインシリコアプローチへ)

## 論文内容の要旨

CAG repeats are the cause in a certain neurodegenerative diseases. This thesis centers around investigating small molecules targeting the CAG repeats, offering a potential avenue for therapeutic intervention. Chapter I starts from the molecular design, and Chapters II-IV are collaborate with *in-silico* machine learning techniques, including a classification task on the random forest algorithm (Chapter II), an analysis using clustering methodologies (Chapter III), and the classification task based on deep learning (Chapter IV).

**Chapter I**

In this chapter, a dimer of NA, DoNA was developed. Affinity assays including surface plasmon resonance (SPR), CSI-TOF MS, and UV melting were used to show DoNA's selective binding to CAG repeat RNAs. We further figured out that the binding of DoNA to A-A mismatch depends on the positions where the A-A mismatch locates. *In-silico* analysis including MD simulation and conformational search suggested that the positions involving a higher flexibility in conformation is likely essential. This study will give a new potential molecular probe for CAG repeat diseases and an insight into the molecular design of RNA-binding small molecule ligands.

**Chapter II**

In this chapter, we used a machine learning algorithm called random forest to classify a set of molecules and obtain a prediction accuracy above 80%. Those predictions were done under the given over 5000 molecule descriptors and hope to be applied to a broader field of drug discovery, such as the focused library generation.

**Chapter III**

In this chapter, we have developed a clustering-based approach for grouping SPR sensorgrams to identify potential hit compounds. Our method successfully distinguished 220 hits from all 2000 samples targeting CAG repeat DNA. The separation of the hit cluster could be visualized in low-dimension spaces.

**Chapter IV**

In this chapter, we use graph convolutional neural networks (GCNNs), a deep learning technique, for predicting small molecules that target CAG repeat DNA. In the case of deep learning, confronted with a common problem of a limited dataset in the field of biochemistry, an augmentation strategy which amplify the molecular dataset by generating additional conformations was included. Those results could be inspiring for future studies.

## 論文審査の結果の要旨及び担当者

氏 名 ( QINGWEN CHEN )	
論文審査担当者	(職) 氏 名
	主 査 教授 中谷 和彦
	副 査 教授 梶原 康宏
	副 査 教授 久保 孝史
副 査 教授 (情報科学研究科)	松下 康之

## 論文審査の結果の要旨

本論文は、遺伝性神経変性疾患であるハンチントン病の原因となる、異常伸長した **CAG** リピートを標的とした低分子創成と低分子創成への機械学習の新たな活用方法に関する研究を取りまとめたものである。第一章は **CAG** リピート **RNA** に結合する新たな分子創成に関する研究、第二章はランダムフォレスト・アルゴリズムによる分類、第三章はクラスタリング手法を用いた分析、第四章はディープラーニングに基づく分類課題について纏められている。

第一章では、**CAG** リピート **RNA** に結合する新たな分子の開発を、既知の化合物ナフチリジン-アザキノロン (**NA**) を基にして、その二量体 (**DoNA**) を合成した。表面プラズモン共鳴 (**SPR**) 法を使って **CAG** リピート **RNA** への結合を評価したところ、**NA** に比べて格段の結合が観察された。さらに、**CAG** リピートが形成する多様な構造のうち、どの構造に結合するのかを調べたところ、期待したインターナルループやヘアピンループではなく、ヘアピンステムの末端やオーバーハングなどの、構造の自由度の高い **CAG RNA** に結合することを明らかにした。この研究結果は、より強力に **CAG** リピート **RNA** に結合する化合物の創成につながることを期待される。

博士論文の後半部は、バインディングに適した物質を機械学習を用いて予測することを試みた研究であり、具体的には、(1) 教師あり学習による分子特徴量を利用したバインディングの予測、(2) 教師なし学習 (クラスタリング) による **SPR** プロファイルを利用した分子のグルーピングである。(1) の教師あり学習によるバインディングの予測では、テーブルデータに対して有効性が知られているランダムフォレストを用いて、分子の特徴量から **hit/non-hit** の分類を試みている。本研究で対象とするタスクにおいて一般的である「データのアンバランスさ」の問題を指摘し、前処理によってバランスの取れた学習データセットを作成することにより、高い精度の認識器を設計した。また、(2) 教師なし学習による **SPR** プロファイルを利用した分子のクラスタリングでは、クラスタリング結果が **SPR** プロファイルに含まれる高周波ノイズに強く影響されることを指摘し、周波数分析に基づくノイズ除去を行うことで、直感的にわかりやすい4つのクラスタが生成されることを示した。このクラスタリング手法を用いることにより、より客観的なデータのグルーピングが可能となった。クラスタリングでは、クラスタを作るところまでがゴールであり、あるクラスタが **hit** に属するのか、**non-hit** に属するのかという判断は未だに人間に依存するものであるが、プロファイルのグルーピングを自動化することで真に客観的なデータ分析に向けて前進したと言える。

以上のように、本論文は分子創成と機械学習を用いたバインディング予測に関する先駆的な研究として、理学に寄与するところが大きい。

よって、本論文は博士 (理学) の学位論文として十分価値あるものと認める。