

DNA 構造体の体積予測システムの構築

百塚真弥 (指導教員: オベル加藤ナタナエル)

1 はじめに

本研究は、群れ行動を発生させる DNA 構造体から成るマイクロロボットの形状最適化を目的としている。この研究は以下に示す 3 段階の手順を踏んで行われる。1. DNA 鎖を水溶液に入れた際にできる DNA 構造体の体積を予測する。2.1 で作成した DNA 構造体が相互作用を生じるかどうか判定する。3. 以上の 2 段階を踏まえた上で、群れ行動を起こす DNA 配列を探索する。本論文では 1 のサイズ予測のみを行なった。群れ行動の創発には、個々の構造体の大きさを知ることが必要である。[1] 小さすぎると相互作用が生じず、大きすぎると動きが鈍くなる。そこで、DNA 構造体の大きさを予測する判定システムを作成した。また、これらは全て oxDNA 上で行った。[2]



図 1: DNA 鎖の集合体
: oxDNA の出力結果一例。各構造体の体積の平均を予測する対象とする。

2 体積予測システムの詳細

DNA 鎖の制約シミュレーションに DNA 鎖の配列、温度、容器の大きさ、タイムステップなどを入力し、約 3 時間後に仮想容器内の DNA 構造体の状態が出力される。本研究は DNA 鎖の組み合わせを入力し、DNA 構造体の体積の平均をニューラルネットワークを用いて予測するシステムを作成した。

2.1 oxDNA の概要

oxDNA とは粗視化分子動力学パッケージである。本研究と同様に oxDNA を用いた計測や予測をする研究も存在する。[3]

2.2 システムへの入力

入力値の制限として、入力する DNA 鎖の配列を 16 本から 1728 本までの DNA 鎖の 3 つのライブラリに注目し以下に分類した (表 1)。[4]

このシステムでは DNA 配列を考慮せず、DNA 鎖組み合わせの種類の有無のみを入力する。

2.3 体積計測

分子構造体の厳密な体積の方法として、Voronoi method や、Radical Planes method[5] が挙げられる。しかし、今回は相対的な体積を知ることが目的であるため、より計測法が単純である convex-hull 法を用い

た。これは、分子の座標を凸包で囲み、その内部の大きさを測る測定手法である。

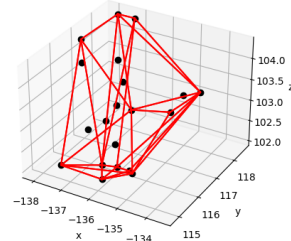


図 2: DNA 鎖
: 点は各ヌクレオチドの中心を示している。また、黒い点の集合は DNA 鎖を示している。長さ 1 単位を 8.518x10nm とする。

このシステムでは、出力を DNA 鎖の体積の平均とした。

3 実験

3.1 実験結果

L1, L2, L3 に対しそれぞれ 1222, 1500, 903 個のデータを用意した (図 3)。図 4 は、L1 を用いて回帰や学習を行なった結果である。L2, L3 を用いた結果も score の順序に変化はなかった。score が最も高いため、以降の実験にはニューラルネットワークを用いることにした。また、モデルは tensorflow の sequential model を用いた。L1, L2, L3 に対しニューラルネットワークのモデルで予測した結果は図 5 により示した。L3, L2, L1 の順に予測の精度が低い。しかし、この順序はデータ数を変化させた際に変化した。また、図 ?? について各ドメインの組み合わせについて、ドメイン配列を 10 個平均化したモデルとそうでないものを比較したものである。どちらも 384 個の L1 の結果を用いた。

3.2 考察

図 3 より、L1, L2 の体積分布範囲は類似しており L3 の分布範囲の偏差は大きい。図 5 から、L3 が L1, L2 よりも予測の精度が高い。分布範囲が高域であることと DNA 鎖 1 本あたりのヌクレオチド数が関係している可能性がある。図 5 下段について、ドメイン組み合わせが等しい DNA 鎖の内、10 個の異なるドメイン配列を平均した体積に変更したものが左図である。score はそれぞれ 0.009411783, 0.02635952 である。ドメイン配列は予測に大きく影響することが伺える。

4 まとめ

本研究では DNA 鎖の集合から DNA 構造体の体積予測を行なったが、現在のシステムの予測の精度は低

ライブラリ	種類数	DNA 結合ドメインの ヌクレオチド数	配列種類数	DNA 鎖を構成する DNA 結合ドメインの種類数
L1	2	10	2-6	2
L2	2	5	3-6	4
L3	6	15, 17	3-6	3

表 1: 入力 DNA 鎖の分類

い. 予想される 4 つの要因を以下に挙げる. 1) 図 5 下段より, ドメイン配列が体積に影響していると考えられる. これより考えられる対策はドメイン配列を固定したデータを入力に用いることである. また, 同じドメインの組み合わせの DNA 鎖で異なるドメインの配列の場合どの程度体積に変化を及ぼすか調べる必要がある. 2) ニューラルネットワークに用いた sequential model の設定が影響している可能性がある. 層数や, モデル自体を変更して実験する必要がある. 3) oxDNA への入力が最適でない可能性がある. 本研究で用いたライブラリ L1, L2, L3 は先行研究 [4] から得たものである. これは NUPACK[6] をもちいてシミュレートしていた. そのため同じ DNA 鎖の集合を入力しても oxDNA とは異なる結果を示していた. 本研究の出力結果は DNA 鎖同士が分子交雑することが稀であったため, DNA ドメインの組み合わせが DNA 構造体の体積への影響が小さかった可能性がある. 4) 入力の形式や, 予測因子の不足が考えられる. より良い予測因子を探すことを目標とする. また, そのためには oxDNA への理解を深める必要があると考える.

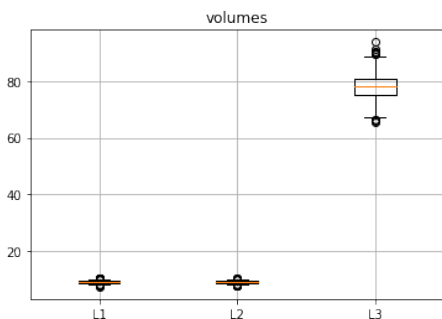


図 3: 体積分布

: 縦軸の単位は $(8.518 \times 10)^2 [nm^2]$ である.

	regressor	score
0	extratrees	-1.100624
1	randomforest	-0.024380
2	gradientboosting	-0.058813
3	stacking	-0.330919
4	voting	-0.157241
5	histgradientboosting	-0.202349
6	neural network	0.012746

図 4: 予測精度比較

: 各 regressor の精度比較. ここでは, L1 を用いた. score は平均二乗誤差を表している.

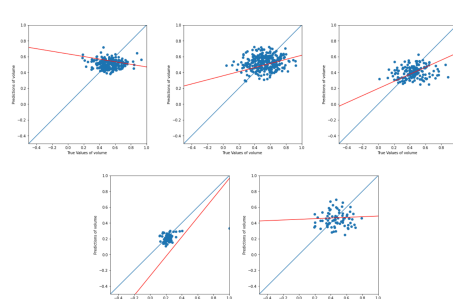


図 5: ニューラルネットワークを用いた L1, L2, L3 予測結果

: 上段: 横軸は体積の真値, 縦軸は予測値である. 赤い直線と青い直線の傾きが近いほど予測の精度が高い.

下段: 左図は各構造のドメイン配列を平均化した
右図: ドメイン配列はランダムである

参考文献

- [1] 分子ロボティクス研究会.
- [2] Erik Poppleton, Roger Romero, Aatmik Mallya, Lorenzo Rovigatti, and Petr Šulc. Oxdna.org: a public webserver for coarse-grained simulations of dna and rna nanostructures. *Nucleic acids research*, Vol. 49, No. W1, pp. W491–W498, 2021.
- [3] Erik Benson, Marco Loloico, Yevgen Tarasov, Andreas Gadin, and Bjorn Hogberg. Evolutionary refinement of dna nanostructures using coarse-grained molecular dynamics simulations. *ACS nano*, Vol. 13, No. 11, pp. 12591–12598, 2019.
- [4] L Cazenille, A Baccouche, and N Aubert-Kato. Automated exploration of dna-based structure self-assembly networks. *Royal Society open science*, Vol. 8, No. 10, p. 210848, 2021.
- [5] Katalin Nadassy, Isabel Tomás-Oliveira, Ian Alberts, Joël Janin, and Shoshana J Wodak. Standard atomic volumes in double-stranded dna and packing in protein–dna interfaces. *Nucleic acids research*, Vol. 29, No. 16, pp. 3362–3376, 2001.
- [6] Mark E Fornace, Jining Huang, Cody T Newman, Nicholas J Porubsky, Marshall B Pierce, and Niles A Pierce. Nupack: Analysis and design of nucleic acid structures, devices, and systems. 2022.