

課題4：蛋白質の構造解析

人にはどうしても
解かねばならない問題がある。

— Osamu Watanabe

課題のねらい：

ある1つの具体的な問題例をコンピュータを駆使して解く。

講義ノート

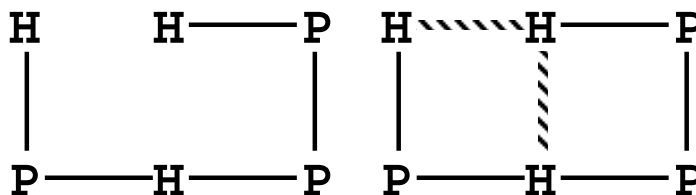
1. 問題の説明

今回の課題は、タンパク質を構成するアミノ酸の列から、タンパク質の構造を求める問題、タンパク質の構造決定問題である。

本格的な構造決定問題はかなり手ごわい。今回は、以下のようにぐっと簡略化したものを考える。

- (1) 20種類のアミノ酸で考えるのではなく、アミノ酸を親水基と疎水基に大別し、親水基か疎水基かで考える。以下、親水基をP、疎水基をHと略記する。つまり、データとしては、PとHの列をもらう。
- (2) 各アミノ酸は二次限の格子上に配置されると考える。
- (3) 構造は、格子上で隣り合った2つのHの組（これをHHコンタクトと呼ぶ）がいくつ存在するか、だけで決まる。つまり、HHコンタクトの数が最大となる配置を「答え」と考える。

たとえば、HPHPPPHという入力列—HP列と呼ぶ—に対し、下図（左）のような配置を求める問題である。ちなみに、この配置のHHコンタクトの数は、下図（右）のように2と数えられる。



配置方法

斜線がHHコンタクト

この問題をコンピュータと協力しながら解くことを今回の課題としよう。最終的には、

問題例1：P3H2P2H2P5H7P2H2P4H2P2HP2

を解くことを我々の目標としてみたい（注：Pが3つ続く場合、PPPをP3と略記する。）

もうまくいったら、

問題例2：P3H2P2H4P2H3PH23H2P8H6P2H6P9HPH2PH11P2H3PH2PHP2HPH3P6H3

にも挑戦したい．ちなみに，問題例 2 に対しては，現在，50 個の HH コタクトを持つ解が知られているが，それが最大であるかどうかはわかっていない．

2. 利用するプログラム

今回は，基本的なバックトラック枝狩り探索法を用いて，最適解を探索するプログラム¹を使用する．

プログラムの使用法

今回使うプログラム prot1 は，与えられる HP 列に対し，そのすべてを一括して解析するのではない．人間がガイドしながら，列の先頭から十数個～二，三十個ずつの解析を行うプログラムである．

具体的には，

s : 現時点から何ステップ先までの HP 列までを解析するかの指定

(x, y) : 二次元格子座標のどの地点まで進むかの指定

を与えると，その範囲の HP 部分列に対し，よいと思われる埋め込み（最適とは限らない）を答えてくれる．たとえば，下図の HP 列に対し， $s = 4$, $(x, y) = (10, 6)$ と指定すると，下図左のように答えてくれる（この場合，答えは 1 通りしかない）．

HP 列 : H P H P P P H P P H P P H H ...

0.1.2.3.4.5.6.7.8.9.0.1.2.3.4.5.7.8.9.0.*	0.1.2.3.4.5.6.7.8.9.0.1.2.3.4.5.7.8.9.0.*
1.....1	1.....1
2.....2	2.....2
3.....3	3.....3
4.....4	4.....4
5.....5	5.....5
6.....P.....6	6.....P-P-P.....6
7.....H.....7	7.....H...H-P.....7
8.....P.....8	8.....P...H-P.H.....8
9.....H.....9	9.....H...P-P-H.....9
0.....0	0.....0
⋮	⋮

その答えに対し， $s = 10$, $(x, y) = (14, 8)$ とやると，上図右のような解を答えてくれる（この場合，1 個の HH コンタクトが得られる）．

これをさらに先に進めてもよいが，この解はやめて別の方向へ進むよう指定してもよい．たとえば， $s = 10$ のままで $(x, y) = (11, 9)$ にすると，次ページの図のような解が得られる（この場合，HH コンタクトは 2 個）．

¹ 渡辺研の新倉氏 (TA) と加藤氏の力作です．

(2) 枝刈りの各種条件

枝刈り条件を決めるパラメータ（これを search parameter と呼ぶことにする）がいくつかある．このパラメータを上手に選んでやることで，プログラムの性能を引き出せる．詳しくは次回の演習資料で．

(3) 目標地点

埋め込みは，埋め込み領域（HP 列の長さにより自動的に決まる）の中心の座標から始まる．それに対し，今回のプログラムは，目標地点 (x, y) （ただし x, y は整数）を指定しなければならない．

長さが足りなくて，スターとの座標から，その目標地点に到達できない場合にはエラーとなり，再度入力しなければならない（長さが足りても，偶奇性により目標地点に到達できない場合もある．その場合には， ± 1 の座標に到達するようになっている．）

今回の宿題（ \times 切：1月16日，演習開始前）

- (1) 次の HP 列に対して，最適な配置を求めよ（注：学籍番号の最後の数字を 4 で割った余りに応じた列について考えること）

0 の人：HPHHPHPHHP

1 の人：PHPPHHPHPHP

2 の人：PHHPHPHPHP

3 の人：HPHHHPHPHP

- (2) 問題例 1 の最初の 15 個分の列に対して，よさそうな（最適でなくてもよい）配置を求めよ．
- (3) 適当なチューンアップを施した後実行してみたところ， $s = 10$ 程度であれば，10 秒くらいで探索ができた．その場合， $s = 12$ にはどの程度の時間がかかりそうか？推定せよ（枝刈りのことは考えないでよい．）