課題４：蛋白質の構造解析

実験ノート

今回の課題では，与えられたソフトウェアを上手に使い，解析したいアミノ酸列からタンパク質の構造（の候補）を求める．

このような場合，与えられたソフトをただ闇雲に使うだけでは思うような効果は得られない．その仕組みをある程度知った上で，どのように使えば最も効率よく使えるか，適度なチューンアップとうまい使い方の工夫が必要である．今回の課題を通して，それを学んでほしい．

１．本日やること（宿題の説明）

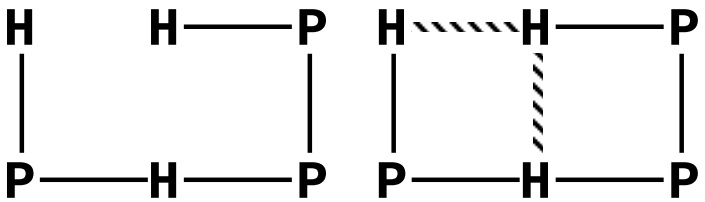
プログラム prot1 を使ってみて，各パラメータの意味を調べる．つまり，探索の際に変えることのできるパラメータ（search parameter と呼ぶことにする）に対し，どのように設定するとどのような効果が得られるか（あるいは問題が生じるか）を調べる．

具体的には，各パラメータを設定したとき，どんな性能向上が得られるか，あるいは，まずいことが生じる例を調べる．宿題は，それをレポートにして提出する（来週の演習始めに提出）

２．プログラムの説明

概要

プログラム prot1 は，与えられた HP 列から，その二次限格子上への配置を求めるプログラムである．たとえば，H P H P P P H 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．



配置方法　　斜線が HH コンタクト

その際，次の基本制約を満たすようなものを求める．

(1) 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）がなるべく多く出るようにする（たとえば，上の配置の HH コンタクトの数は図（右）のように 2 と数えられる．）

(2) 配置は，HP 列の長さに応じて，あらかじめ決められた大きさの配置画面に行われる．その際，出発点は配置画面の中央．一方，終りは指定された点（目標点）になるように配置する．（ただし，長さ（の奇偶）の都合上，目標点にピッタリで終われない場合もある．その場合には目標点の周囲距離 1 以内で終わるようにする．）

探索パラメータ (Search Parameter)

プログラム prot1 のアルゴリズムは，先週説明したような基本的なバックトラック枝刈り探索法である．配置のやり方の場合の数は，HP 列の長さを n とすると，大雑把にいって 3^n 通りである．これをすべて探索できるのは $n = 10$ 程度まで．そこで，可能性のない選択を早めにやめて，大切なところだけを中心に探索するようにしたい．それを枝刈りという．今回調べる search parameter は，この枝刈りを調整するパラメータである．

具体的には，以下のような search parameter を用いる（なお，特に断わりの無い場合には，パラメータ値 0 は，その枝刈りを使わないことを意味する．）

1. H_Weight: (0 または $0 <$ H_Weight ≤ 2)
   残りの HP 列の長さから，どの程度の HH コンタクトを期待できるかを判定するときの重み．この期待値を使うことで，現在得られている最大値を超えられるかを判定し，見込みがない場合には枝刈りをする．
2. ST_ Length: (0 または 1 以上の整数)
   直線に進む限度．まっすぐ進めるステップ数の上限．
3. Dist_Limit: (0 または $0 <$ Dist_Limit ≤ 1)
   出発点からの距離の上限 (ステップ数に対する割合で表わす)．
4. Close_Region_SW: (0 または 1)
   閉領域に迷い込むことを未然に防ぐかどうか ?(注：SW とはスイッチという意味．この場合は 1 でオン (有効)，0 でオフ (無効))
5. Tight_SW: (0 または 1)
   目標地点に対して残り距離が少なくなった場合，一通りの進み方にしてしまうか (オン 1)，すべての可能性を調べるか (オフ 0) の指定．

以上のパラメータを安全目にとると，高い得点が得られる可能性が高くなるが計算時間はかかるようになる．計算時間を抑えるには，適度にパラメータを設定しなくてはならない．その適度な調整のためには，パラメータを軽く設定すると，どのような問題が生じる可能性があるかを知っておく必要がある．本日のテーマは，それを探ることである．

3．プログラムの使い方

- 準備

1. kadai4 というディレクトリを作り，そこに移動．
2. ~owatanab/pub/kadai4 の下から prot1 をコピー．

- 実行方法

まず，適当な HP 列を入れたファイル hp.dat を作る．あとは，単に

```
[自分のプロンプト] ./prot1
```

とやればパラメータ等を聞いてくるので，それを入力すれば探索を開始する．