

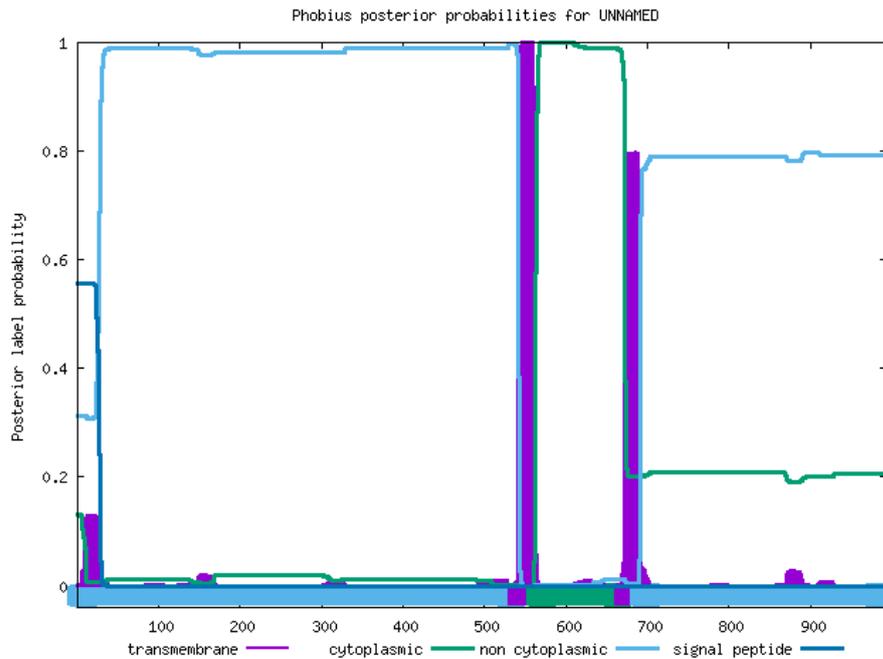
例題 3.8

UniProtKB を検索し、ヒトの受容体型チロシンキナーゼの配列を取得せよ。「Receptor-type tyrosine-protein kinase FLT3」で検索せよ。

1. 適当な予測システムを用いて、このタンパク質がシグナルペプチドをもつかもたないかを予測せよ。もつと予測される場合、その位置（残基番号）を求めよ。
2. 適当な予測システムを用いて、このタンパク質が膜貫通領域をもつかもたないかを予測せよ。もつと予測される場合、その位置（残基番号）を求めよ。
3. このタンパク質の UniProtKB のアノテーション情報と、(1), (2)の予測結果を比較せよ。

Prediction of UNNAMED

```
ID UNNAMED
FT TOPO_DOM 1 541 NON CYTOPLASMIC.
FT TRANSMEM 542 563
FT TOPO_DOM 564 672 CYTOPLASMIC.
FT TRANSMEM 673 691
FT TOPO_DOM 692 993 NON CYTOPLASMIC.
//
```



The probability data used in the plot is found [here](#), and the gnuplot script is [here](#).

膜貫通領域は 542-563, 673-691 の 2 箇所である。TMHMM では 542-564, 674-696、SOSUI では 10-32, 543-565 (2 つの膜貫通領域を予測できていない) という予測結果を示す。InterPro では、Phobius と TMHMM の 2 つのツールの結果が表示され、ドメイン情報、モチーフ情報などと合わせて、一度に予測結果を見ることができるようになっている。ただし、Phobius の M スコア (膜貫通領域のスコア) を見ると、1 つめの予測領域 542-563 は平均 0.94273 (8 残基は 1) であるのに対し、2 つめの予測領域 673-691 は平均 0.77142 で低くなっている。

(3) UniProtKB のアノテーション情報によれば、**シグナル配列は 1-26、膜貫通領域は 544-563 (1 つのみ)** である。NCBI のサイトにも、NCBI のフォーマットでこれらの情報が記載されている。なお、膜貫通領域が 1 つというアノテーション情報についてであるが、UniProtKB P36888 の記述 (「Structure」のセクション) によれば、このタンパク質の構造がいくつか決定され PDB に登録されている。そのうち、2 つめの予測膜貫通領域の部分を含む構造が 5 個あり、どれも膜貫通領域のアノテーションは付いていない。膜貫通領域は重要なアノテーションで、構造決定者による指定のほか、類似タンパク質のチェックや構造をもとにした膜貫通領域の同定ソフトウェアの解析結果などが反映されるので、この情報の信頼度は高いと考えられる。しかも、その部分の二次構造は、ヘリックスではなく、シートとループから構成されている。興味深いのは、PSIPRED や Jpred4 のよ

うな二次構造予測でも正確な予測をしていてヘリックスとは予測していない点である。従って、Phobius も TMHMM も、どちらもヘリックス予測（ヘリックス出ないという判定）ができていないことになる。このことは、予測する立場では、いくつかのツールを使用する余地があることを意味している。ちなみに1つめの予測膜貫通領域（実際の膜貫通領域）に相当する部分の構造はまだ決定されていない。

下の図は、このタンパク質の構造を AlphaFold2 で予測した結果である。シグナル配列と考えられる部位（1-26）を水色、膜貫通領域と考えられる部位（542-563）を赤色、Phobius や TMHMM で膜貫通領域と誤って予測した部位（673-691）を黄色で示している。膜貫通領域の位置からわかるように、ドメインリンカ部分の予測精度が低く、全体的なドメインの位置関係は正しくないと考えられる。

