

2020 年度卒業論文

蛋白質の滴定曲線描画ツールの作成

指導教官 黒田裕

東京農工大学 工学部 生命工学科

黒田研究室

2015 年度入学

学籍番号 15251052

仁禮 純人

指導教官印	学科主任印

内容

1 章 緒言	4
1-1 本研究の背景	4
1-1-1 滴定曲線について	4
1-1-2 Python について	4
1-2 本研究の目的	5
2 章 研究方法	6
2-1 滴定曲線描画ツール作成の流れ	6
2-2 プログラムの作成	6
2-2-1 滴定曲線描画のプログラム	6
2-2-2 分子量、吸光係数、等電点計算のプログラム	6
2-3 ツールの評価	7
2-3-1 滴定曲線の評価	7
2-3-2 その他の出力情報の評価	7
2-3-3 評価に使用した 10 種類の配列	7
3 章 結果と考察	10

3-1 完成した蛋白質の滴定曲線描画ツールについて	10
3-1-1 完成したプログラム	10
3-1-2 完成したツール	10
3-2 ツールの評価	10
3-2-1 滴定曲線の評価	10
3-2-2 その他の出力情報の評価	11
3-3 当研究室のウェブサイト上への公開	14
第4章 まとめ	15
第5章 図面集	16
第6章 参考文献	25
第7章 謝辞	25

1 章 緒言

1-1 本研究の背景

1-1-1 滴定曲線について

蛋白質の電荷は pH に依存し、各官能基の解離状態によって決まる。蛋白質の解離基には、N 末端のアミノ基、C 末端のカルボキシル基のほか、側鎖の官能基としてアスパラギン酸の β -カルボキシル基、グルタミン酸の γ -カルボキシル基、ヒスチジンのイミダゾール基、システインの SH 基、リシンの ϵ -アミノ基、チロシンの OH 基、アルギニンのグアニジノ基などがある。今、解離基を RH とし、解離定数を K_a とすると、

$$K_a = \frac{[R^-][H^+]}{[RH]} \quad (1)$$

と書ける。両辺の常用対数を取り整理すると、

$$-\log[H^+] = -\log K_a + \log \frac{[R^-]}{[RH]} \quad (2)$$

ここで、 $p = -\log$ とおいて、

$$pH = pK_a + \log \frac{[R^-]}{[RH]} \quad (3)$$

が得られる。この式はヘンダーソン-ハッセルバルヒの式と呼ばれる。

この式は、水素原子（プロトン）の解離度を α とすると、

$$pH = pK_a + \log \frac{\alpha}{1-\alpha} \quad (4)$$

と書ける。

総電荷を Q と置くと、アミノ基のように、プロトンが結合すると電荷が $0 \rightarrow 1$ に変わるものは、(4)より、

$$Q = \frac{1}{1+10^{pH-pK_a}} \quad (5)$$

カルボキシル基のように結合すると、電荷が -1 から 0 に変わるものは、

$$Q = -\frac{10^{pH-pK_a}}{1+10^{pH-pK_a}} \quad (6)$$

によって、電荷を求めることができる。

N 末端のアミノ基、C 末端のカルボキシル基、また側鎖の解離基における電荷の総和を、一定の pH 間隔ごとにプロットしたものが、滴定曲線となる。

1-1-2 Python について

Python とは、組み込み開発、ウェブアプリケーション、デスクトップアプリケーションなどで利用されるプログラミング言語のことである。言語の特徴としては、文法が単純であるため、書く人物に関係なく同様のプログラムになるため、可読性が高く、初心者が学習しやすいことが挙げられる。また、Python では、豊富なライブラリ（多くのアプリケーション

ョンで利用するためにまとめられたプログラム集) が公開されているため、それを用いることで作りたいプログラムが容易に作成できる。現在数万個以上のライブラリが公開されている。また、機械学習についてのライブラリも公開されているため、近年注目を集めている。

このような特徴を持つことから、当研究室でも Python が使用されている。

1-2 本研究の目的

蛋白質は、生命現象解析の主要な対象となっている。蛋白質は、20 種類のアミノ酸が重合してできた生体高分子化合物である。この蛋白質は、そのアミノ酸の配列によって立体構造を形成し、その立体構造により様々な機能を有する。電氣的性質もその一つである。蛋白質の性質は分子表面に多く存在するアミノ酸側鎖の解離基の状態、すなわち解離基の電荷の状態に大きく依存する。特に、酵素の活性は活性部位の解離基の状態に大きく依存する。蛋白質が持つ性質や機能を応用し、広い範囲で蛋白質の活用が進められている。また、蛋白質を単離・精製する際、電荷の状態は、目的の蛋白質を他の蛋白質から分離する重要な指標となる。そのため、インターネット上には、蛋白質の配列に基づき等電点を計算するツールが数多く存在している。しかし、一定の pH 間隔における電荷の状態を示す、滴定曲線を計算によって描画するツールは現在確認できない。滴定曲線を、事前に計算によって知ることができれば、より正確な実験を迅速に行うことが可能になる。本研究では、蛋白質の配列から、その滴定曲線を表示するツールを作成することを目的とする。

2章 研究方法

2-1 滴定曲線描画ツール作成の流れ

1. まず、Python というプログラミング言語を用いて、蛋白質の滴定曲線を描画するプログラムを作成した。
2. 次に、CGI (Common Gateway Interface) という仕組みを用いて、ウェブサーバー上でプログラムを動かせるようにした。
3. 既存の等電点計算ツールと比較し、計算方法等に問題がないかを検討した。
4. 最後に、当研究室のウェブサイト上に公開した。

2-2 プログラムの作成

蛋白質のアミノ酸配列を入力すると、1-1-1 の式を元に滴定曲線が出力されるプログラムを作成した。さらに、アミノ酸配列から計算できる情報として、組成、分子量、吸光係数、等電点も出力できるようにした。

また、滴定曲線を描画する際に必要となる pKa の値は文献によって様々であるため、利用者が任意に変更できるよう、配列と同じく pKa も入力によって得られるようにした。

2-2-1 滴定曲線描画のプログラム

まず、入力した蛋白質の配列から、各アミノ酸の個数、すなわちアミノ酸組成を調べる。そこから側鎖に解離基を持つアミノ酸（アルギニン、リシン、ヒスチジン、チロシン、アスパラギン酸、グルタミン酸、システイン）を抽出し、一定の pH 間隔ごとの総電荷を、式 (5)、(6) を用いて導き出す。このとき、pKa の値、プロットする pH 間隔も配列と同時に入力によって得る。それらをプロットして曲線を描画するプログラムを作成した。ここで得られた組成も出力できるようにした。

2-2-2 分子量、吸光係数、等電点計算のプログラム

2-1 で得られた組成を元に、各アミノ酸の分子量を掛け合わせ、最後に水分子 1mol の分子量を足した値を、蛋白質の分子量として出力する。

吸光係数は、

$$\varepsilon = \text{トリプトファンの個数} \times 5500 + \text{チロシンの個数} \times 1490 + \text{システインの個数} \times 125 \text{ [A280/mol/cm]} \quad (7)$$

によって求め、出力する。

等電点は、1-1 によってプロットした点の中で最も絶対値の小さい点、すなわち最も 0 に近い点の pH の値を出力する。

2-3 ツールの評価

2-3-1 滴定曲線の評価

滴定曲線を計算するツールはネット上には公開されていないため、曲線自体を比較することは不可能であった。そのため、既存の等電点計算ツールである IPC

(<http://isoelectric.org/index.html>) と本研究のツールによって導き出される等電点を比較することで、精度の評価を行った。IPC では、pKa の値を様々な文献から引用し、それぞれの pKa における等電点の結果が表示される。

方法としては、無作為に選んだ 10 種類の蛋白質（ペプチド含む）の配列を 15 種類の pKa 値を用いて二つのツールにかけ、その誤差を表にまとめた。表 1 が、使用した pKa のリストである。

2-3-2 その他の出力情報の評価

2-3-1 で使用した 10 種類の配列から、本ツールによって導かれた組成、分子量、吸光係数を、ExPASy ProtParam tool (<https://web.expasy.org/protparam/>) における結果と比較し、評価を行った。組成、分子量、吸光係数は、pKa の値が関係しないため、10 通りの比較のみ行った。

2-3-3 評価に使用した 10 種類の配列

・ 4X42

Crystal structure of DEN4 ED3 mutant with epitope two residues substituted from DEN3 ED3

GSGMSYTMCSGKFSIDKEMAETQHGTTVVKVKYEGAGAPCKVPIEIRDVNKEKVVGR
IISSTPLAENTNSVTNIELEPPFGDSYIVIGVGDKALKLNWFRKGSIGK

・ 2JLR

Dengue virus 4 NS3 helicase in complex with AMPPNP

GSAMGEPDYEVEDEDIFRKKRLTIMDLHPGAGKTKRILPSIVREALKRRRLTLILAPTRV
VAAEMEEALRGLPIRYQTPAVKSDHTGREIVDLMCHATFTTRLLSSTRVPNYNLIVM
DEAHFTDPCSVAARGYISTRVEMGEAAAIFMTATPPGSTDPFPQSNSPIEDIEREIPERS
WNTGFDWITDYQGKTWVWFVPSIKAGNDIANCLRKSGKRVIQLSRKTFDTEYPKTKLT
DWDFVVTTDISEMGANFRAGRVIDPRRCLKPVILTDGPERVILAGPIVTPASAAQRR
GRIGRNPAQEDDQYVFSGDPLRNDEDHAHWTEAKMLLDNIYTPEGIIPITLFGPEREK
TQAIDGEFRLKGEQRKTFVELMRRGDLPVWLSYKVASAGISYKDREWCFNGERNNQI
LEENMEVEIWTREGEKKKLRPKWLDARVYADPMALKDFKEFASGRK

• 7JGW

Crystal structure of BCL-XL in complex with COMPOUND 1620116, CRYSTAL FORM 1
GPLGSMQSNRELVVDFLSYKLSQKGYSSQMAAVKQALREAGDEFELRYRRAFSDL
TSQLHITPGTAYQSFEQVVNELFRDGVNWGRIVAFFSFGGALCVESVDKEMQVLVSRI
AAWMATYLNHLEPWIQENGGWDTFVELYGNNAEAESRKGQER

• 7KZC

Potent SARS-CoV-2 binding and neutralization through maturation of iconic SARS-CoV-2 antibodies

QVQLQQSGAEVKKPGSSVKVCKASGGTFSTYISWVRQAPGQGLEWMGGIAPSHGF
ANYAQKFQGRVTITTDSTAYMELSSLRSEDVAVYYCARDTATGGMDVWGQGT
VTVSSASTKGPSVFLAPSSKSTSGGTAALGCLVKDYFPEPVTVSWNSGALTSGVHTF
PAVLQSSGLYSLSSVVTVPSSSLGTQTYICNVNHKPSNTKVDKKEPKSC

• 6LZQ

Chitin-specific solute binding protein from *Vibrio harveyi* in complex with chitotriose.

AERSELTIHPEFTTFVRNFPFLGATNLHTTTDFIYEPLVVFNEMHGNTPVFRLAEN
FQMSDDLMSVTFDIRKGVKWSDEGAEFTADDVVYSFNLVKEKPELDQSGINSWVTGV
EKVNDYQVKFRLSEANSNVPYEIAKVPVVPKHVWSKVKDPSTFTNENPVGSGPFTVID
TFTPQLYIQCENPNYWDAANLDVDCLRVPQIANNDQFLGKVVNGEMDWTSSFPDI
DRTYAAASPKHHYWYPPAGTQAFVVFKNPDAAKNEALTNVDFRRAFSMALDRQTII
DIAFYGGGTVNDFASGLGYAFEAWSDEKTHDKFKAYNSYNAEGAKLLAKAGFKDV
NKDGFVDTPSGKSFELLIQSPNGWTFNNTVQLAVEQLAEVGIKARARTPDFSVYNQ
AMLEGTYDVAYTNYFHGADPYTYWNSAYNSALQSGDGMPRFAMHFYKNEKLDGLL
NSFYKTADKQEQLIAHQIIAQDQVTIPVLSGAYMYQYNTTRFTGWWNEENPKG
RPNIWAGIPERLLHVLDLKPVKLEHHHHHH

• 7BE1

X-ray structure of Hen Egg White Lysozyme with dirhodium tetraacetate (3)

KVFGRCELAAAMKRHGLDNYRGYSLGNWVCAAKFESNFNTQATNRNTDGSTDYGIL
QINSRWWCNDGRTPGSRNLCNIPCSALLSSDITASVNC AKKIVSDGNGMNAWVAWR
NRCKGTDVQAWIRGCR

• 1BHC

**BOVINE PANCREATIC TRYPSIN INHIBITOR CRYSTALLIZED FROM
THIOCYANATE**

**RPDFCLEPPYTGPCKARIIRYFYNAKAGLCQTFVYGGCRAKRNNFKSAEDCMRTCGG
A**

• 2IGR

Solution structure of CB1a, a novel anticancer peptide derived from natural antimicrobial peptide cecropin B

KWKVFKKIEKKWKVFKKIEKAGPKWKVFKKIEKX

• 1IRR

Solution structure of paralytic peptide of the silkworm, *Bombyx mori*

ENFVGGCATGFKRTADGRCKPTF

• 5IIT

Structure of SPX domain of the yeast inorganic polyphosphate polymerase Vtc4 crystallized by carrier-driven crystallization in fusion with the macro domain of human histone macroH2A1.1

**MKFGHEHLSKSLIRQYSYYYISYDDLKTELEDNLSKNNGQWTQELETDFLESLEIELDKV
YTFCKVKHSEVFRRVKEVQEQVQHTVRLDLSNNPPTQLDFEILEEELSDIADVHDLA
KFSRLNYTGFQKIIKKHDKKTGFILKPVFQVRLDSKPFKENVDELVVKISQLYDIARTS
GAGSDGFTVLSTKSLFLGQKLQVVQADIASIDSDAVVHPTNTDFYIGGEVGNLTLEKK
GGKEFVEAVLELRKKNNGPLEVAGAAVSAGHGLPAKFVIHCNSPVWGADKCEELLEKT
VKNCLALADDKLKSIAFPSIGSGRNGFPKQTAAQLILKAISSYFVSTMSSSIKTVYFVLF
DSESIGIYVQEMAKLEHHHHHH**

3 章 結果と考察

3-1 完成した蛋白質の滴定曲線描画ツールについて

3-1-1 完成したプログラム

図 1~8 に、完成したプログラムを載せた。左端の数字は行数である。

図 1~5 が計算と出力画面、図 6~7 が入力画面、図 8 がサーバーを立ち上げるためのプログラムである。

出力画面のプログラムの説明をする。図 1 の 21~27 行目でアミノ酸の組成、図 2 の 58~59 行目で吸光係数、図 2~3 の 61~68 行目で分子量が計算される。また、図 3~4 の 90~129 行目で、滴定曲線の描画及び、そこから等電点を導き出す。図 4~5 の 131 行目以降は、ウェブサイトに表示される出力画面の構成を HTML 形式で書いたものである。

3-1-2 完成したツール

図 9~10 に、完成したサイトの入力画面、出力画面を載せた。

図 9 の入力画面について、まず、pH space の欄で総電荷をプロットする pH 間隔の調節ができる。次に、NH₂'s pKa から Y's pKa の欄に任意の pKa の値を入力し、sequence の大きな入力欄に配列を入力し、submit ボタンを押せば、結果が表示される。

初期値として、2-3 で使用した IPC の等電点計算に用いられている値が入力されるよう設定した。変更したい場合は、値を書き換えればよい。

図 10 の入力画面について、今回は、例として、PDB ID 『4X42』の結果を表示した。まず、一番上に、縦軸が総電荷、横軸が pH の滴定曲線が表示される。その下に、組成、分子量、吸光係数、等電点が表示されている。

3-2 ツールの評価

3-2-1 滴定曲線の評価

表 2、表 3 に各文献値の各配列における 2 つのツールの誤差（絶対値）を示した。評価する際、IPC で計算される等電点が小数点以下三桁までであったため、滴定曲線描画ツールの pH space も 0.001 とし、同様の桁数で比較した。

これら 150 通りの試行の平均誤差は 0.000827 であり、ほとんど無視できる程度のものであったため、計算方法に問題はないと判断した。

3-2-2 その他の出力情報の評価

以下、各配列、上段（太字）に今回作成したツールにおける結果、下段に Expsy ProtPram tool の結果を示した。

2IGR を除いた 9 種類の配列において、組成はすべて一致し、分子量も無視できる程度の誤差しかない。この僅かな分子量の誤差は、計算に使用した各アミノ酸の分子量の文献値の違いによるものだと考えられる。例外の 2IGR に関しては、未知のアミノ酸 X が含まれた配列であり、作成したツールではこれを無視するようプログラムしていたが、Expsy ProtPram tool においてはこれも含めて計算をするため、誤差が大きくなったと考えられる。したがって、組成、分子量の計算結果に問題は見られなかった。

吸光係数に大きな誤差がみられるのは、作成したツールで求めた値は理論値であるが、Expsy ProtPram tool では、その計算方法を改良し、より実測値に近いものが得られるような式を使用しているためである。こちらは本ツールへの導入も検討する。しかし、本ツールで使用した計算式の結果としては、こちらも正しい結果が得られた。

・ 4X42

composition {'A': 5, 'R': 3, 'N': 5, 'D': 4, 'C': 2, 'Q': 1, 'E': 8, 'G': 12, 'H': 1, 'T': 9, 'L': 4, 'K': 11, 'M': 3, 'F': 3, 'P': 5, 'S': 10, 'V': 7, 'W': 1, 'Y': 3, 'V': 10}

composition {'A': 5, 'R': 3, 'N': 5, 'D': 4, 'C': 2, 'Q': 1, 'E': 8, 'G': 12, 'H': 1, 'T': 9, 'L': 4, 'K': 11, 'M': 3, 'F': 3, 'P': 5, 'S': 10, 'T': 7, 'W': 1, 'Y': 3, 'V': 10}

molecular weight = 11508.99 [Da]

molecular weight = 11509.20[Da]

Absorption coefficient= 10220 [A280/mol/cm]

Absorption coefficient= 10095[A280/mol/cm]

・ 2JLR

composition {'A': 34, 'R': 41, 'N': 13, 'D': 31, 'C': 5, 'Q': 10, 'E': 36, 'G': 29, 'H': 6, 'T': 30, 'L': 31, 'K': 27, 'M': 12, 'F': 17, 'P': 30, 'S': 21, 'T': 33, 'W': 9, 'Y': 11, 'V': 25}

composition {'A': 34, 'R': 41, 'N': 13, 'D': 31, 'C': 5, 'Q': 10, 'E': 36, 'G': 29, 'H': 6, 'T': 30, 'L': 31, 'K': 27, 'M': 12, 'F': 17, 'P': 30, 'S': 21, 'T': 33, 'W': 9, 'Y': 11, 'V': 25}

molecular weight = 51278.7 [Da]

molecular weight = 51279.51[Da]

Absorption coefficient= 66515 [A280/mol/cm]

Absorption coefficient= 66140 [A280/mol/cm]

· 7JGW

composition {'A': 14, 'R': 10, 'N': 7, 'D': 7, 'C': 1, 'Q': 10, 'E': 13, 'G': 13, 'H': 2, 'T': 4, 'L': 14, 'K': 5, 'M': 4, 'F': 9, 'P': 3, 'S': 14, 'V': 12}

composition {'A': 14, 'R': 10, 'N': 7, 'D': 7, 'C': 1, 'Q': 10, 'E': 13, 'G': 13, 'H': 2, 'T': 4, 'L': 14, 'K': 5, 'M': 4, 'F': 9, 'P': 3, 'S': 14, 'V': 12}

molecular weight = 17900.75 [Da]

molecular weight = 17901.03 [Da]

Absorption coefficient= 36565 [A280/mol/cm]

Absorption coefficient= 36440 [A280/mol/cm]

· 7KZC

composition {'A': 16, 'R': 4, 'N': 5, 'D': 6, 'C': 5, 'Q': 11, 'E': 7, 'G': 22, 'H': 3, 'T': 4, 'L': 12, 'K': 13, 'M': 3, 'F': 6, 'P': 12, 'S': 35, 'V': 22}

composition {'A': 16, 'R': 4, 'N': 5, 'D': 6, 'C': 5, 'Q': 11, 'E': 7, 'G': 22, 'H': 3, 'T': 4, 'L': 12, 'K': 13, 'M': 3, 'F': 6, 'P': 12, 'S': 35, 'V': 22}

molecular weight = 23059.34 [Da]

molecular weight = 23059.77 [Da]

Absorption coefficient= 34545 [A280/mol/cm]

Absorption coefficient= 34170 [A280/mol/cm]

· 6LZQ

composition {'A': 44, 'R': 16, 'N': 40, 'D': 37, 'C': 2, 'Q': 21, 'E': 30, 'G': 32, 'H': 17, 'T': 21, 'L': 34, 'K': 32, 'M': 9, 'F': 35, 'P': 29, 'S': 28, 'V': 41}

composition {'A': 44, 'R': 16, 'N': 40, 'D': 37, 'C': 2, 'Q': 21, 'E': 30, 'G': 32, 'H': 17, 'T': 21, 'L': 34, 'K': 32, 'M': 9, 'F': 35, 'P': 29, 'S': 28, 'V': 41}

molecular weight = 61222.09 [Da]

molecular weight = 61223.36 [Da]

Absorption coefficient= 103500 [A280/mol/cm]

Absorption coefficient= 103375 [A280/mol/cm]

· 7BE1

composition {'A': 12, 'R': 11, 'N': 14, 'D': 7, 'C': 8, 'Q': 3, 'E': 2, 'G': 12, 'H': 1, 'T': 6, 'L': 8, 'K': 6, 'M': 2, 'F': 3, 'P': 2, 'S': 10, 'W': 6, 'Y': 3, 'V': 6}

composition {'A': 12, 'R': 11, 'N': 14, 'D': 7, 'C': 8, 'Q': 3, 'E': 2, 'G': 12, 'H': 1, 'T': 6, 'L': 8, 'K': 6, 'M': 2, 'F': 3, 'P': 2, 'S': 10, 'W': 6, 'Y': 3, 'V': 6}

molecular weight = 14312.94 [Da]

molecular weight = 14313.14 [Da]

Absorption coefficient= 38470 [A280/mol/cm]

Absorption coefficient= 37970 [A280/mol/cm]

· 1BHC

composition {'A': 6, 'R': 6, 'N': 3, 'D': 2, 'C': 6, 'Q': 1, 'E': 2, 'G': 6, 'H': 0, 'T': 2, 'L': 2, 'K': 4, 'M': 1, 'F': 4, 'P': 4, 'S': 1, 'W': 0, 'Y': 4, 'V': 1}

composition {'A': 6, 'R': 6, 'N': 3, 'D': 2, 'C': 6, 'Q': 1, 'E': 2, 'G': 6, 'H': 0, 'T': 2, 'L': 2, 'K': 4, 'M': 1, 'F': 4, 'P': 4, 'S': 1, 'W': 0, 'Y': 4, 'V': 1}

molecular weight = 6517.43 [Da]

molecular weight = 6517.54 [Da]

Absorption coefficient= 6710 [A280/mol/cm]

Absorption coefficient= 6335 [A280/mol/cm]

· 2IGR

composition {'A': 1, 'R': 0, 'N': 0, 'D': 0, 'C': 0, 'Q': 0, 'E': 3, 'G': 1, 'H': 0, 'T': 3, 'L': 0, 'K': 15, 'M': 0, 'F': 3, 'P': 1, 'S': 0, 'W': 3, 'Y': 0, 'V': 3}

composition {'A': 1, 'R': 0, 'N': 0, 'D': 0, 'C': 0, 'Q': 0, 'E': 3, 'G': 1, 'H': 0, 'T': 3, 'L': 0, 'K': 15, 'M': 0, 'F': 3, 'P': 1, 'S': 0, 'W': 3, 'Y': 0, 'V': 3, 'X': 1}

molecular weight = 4190.11 [Da]

molecular weight = 4301.60 [Da]

Absorption coefficient= 16500 [A280/mol/cm]

Absorption coefficient= 16500 [A280/mol/cm]

・ 1IRR

composition {'A': 2, 'R': 2, 'N': 1, 'D': 1, 'C': 2, 'Q': 0, 'E': 1, 'G': 4, 'H': 0, 'T': 0, 'L': 0, 'K': 2, 'M': 0, 'F': 3, 'P': 1, 'S': 0, 'T': 3, 'W': 0, 'Y': 0, 'V': 1}

composition {'A': 2, 'R': 2, 'N': 1, 'D': 1, 'C': 2, 'Q': 0, 'E': 1, 'G': 4, 'H': 0, 'T': 0, 'L': 0, 'K': 2, 'M': 0, 'F': 3, 'P': 1, 'S': 0, 'T': 3, 'W': 0, 'Y': 0, 'V': 1}

molecular weight = 2462.72 [Da]

molecular weight = 2462.78 [Da]

Absorption coefficient= 250 [A280/mol/cm]

Absorption coefficient= 125 [A280/mol/cm]

・ 5IIT

composition {'A': 21, 'R': 9, 'N': 13, 'D': 23, 'C': 4, 'Q': 16, 'E': 29, 'G': 22, 'H': 14, 'T': 22, 'L': 38, 'K': 36, 'M': 3, 'F': 21, 'P': 10, 'S': 31, 'T': 18, 'W': 2, 'Y': 13, 'V': 29}

composition {'A': 21, 'R': 9, 'N': 13, 'D': 23, 'C': 4, 'Q': 16, 'E': 29, 'G': 22, 'H': 14, 'T': 22, 'L': 38, 'K': 36, 'M': 3, 'F': 21, 'P': 10, 'S': 31, 'T': 18, 'W': 2, 'Y': 13, 'V': 29}

molecular weight = 42175.31 [Da]

molecular weight = 42176.02 [Da]

Absorption coefficient= 30870 [A280/mol/cm]

Absorption coefficient= 30620 [A280/mol/cm]

3-3 当研究室のウェブサイト上への公開

以下の URL にて、完成したツールを公開した。

<http://domserv.lab.tuat.ac.jp/Tcurve.html>

第4章 まとめ

本研究では、蛋白質の配列を入力するだけで滴定曲線を描画できるツールを作成することを目指した。

評価の方法として、先行研究に滴定曲線を計算によって描画するツールというものが見当たらなかったために、やむを得ず等電点のみの比較という方法をとった。しかし、この比較だけでは、曲線自体が正しいかどうかについては断言できないのが事実である。

また、等電点も滴定曲線も、本ツールから得られる結果は理論値であるため、実測値とはどうしてもずれが生じてしまう。これは、官能基の pKa が置かれた環境によって異なることと、蛋白質の電荷同士の反発が原因である。前者は、pKa の値を可変にしたことにより、対応が可能であると考えられる。後者については、現在多くの蛋白質において、高分解能の構造が出ているため、今後の改良次第ではより正確な滴定曲線の予測が原理的には可能になってきている。

しかし、計算によって滴定曲線を描画するという、これまでに無いツールを作成したことに意義がある。今後、これを利用して実験的手法と組み合わせる事で電荷に伴う蛋白質精製の一助になると考えられる。

第5章 図面集

表 1. IPC で使用される pKa (縦：引用先、横：解離基)

	NH2	COOH	C	D	E	H	K	R	Y
IPC_protein	9.094	2.869	7.555	3.872	4.412	5.637	9.052	11.84	10.85
IPC_peptide	9.564	2.383	8.297	3.887	4.317	6.018	10.517	12.503	10.071
Toseland	8.71	3.19	6.87	3.6	4.29	6.33	10.45	12	9.61
Thurlkill	8	3.67	8.55	3.67	4.25	6.54	10.4	12	9.84
Nozaki_ Tanford	7.5	3.8	9.5	4	4.4	6.3	10.4	12	9.6
Dawson	8.2	3.2	8.3	3.9	4.3	6	10.5	12	10
DTASelect	8	3.1	8.5	4.4	4.4	6.5	10	12	10
EMBOSS	8.6	3.6	8.5	3.9	4.1	6.5	10.8	12.5	10.1
Grimsley	7.7	3.3	6.8	3.5	4.2	6.6	10.5	12.04	10.3
Patrickios	11.2	4.2	-	4.2	4.2	-	11.2	11.2	-
Rodwell	8	3.1	8.33	3.68	4.25	6	11.5	11.5	10.07
Sillero	8.2	3.2	9	4	4.5	6.4	10.4	12	10
Solomon	9.6	2.4	8.3	3.9	4.3	6	10.5	12.5	10.1
Lehninger	9.69	2.34	8.33	3.86	4.25	6	10.5	12.4	10
Wikipedia	8.2	3.65	8.18	3.9	4.07	6.04	10.54	12.48	10.46

表 2. 誤差①

	4X42	2JLR	7JGW	7KZC	6LZQ
IPC_protein	0.001	0	0	0	0
IPC_peptide	0	0	0	0	0
Toseland	0	0.001	0	0	0
Thurkill	0.001	0.001	0.001	0	0
Nozaki_Tanford	0	0.001	0	0	0
Dawson	0.008	0.004	0.001	0.012	0
DTASelect	0.001	0.001	0.001	0	0
EMBOSS	0.001	0.001	0.001	0.001	0
Grimsley	0	0	0.001	0	0
Patrickios	0	0.001	0.001	0.001	0.001
Rodwell	0.001	0.001	0.001	0	0
Sillero	0	0	0	0.001	0.001
Solomon	0	0.001	0.001	0.001	0
Lehninger	0.001	0.001	0.001	0.001	0
Wikipedia	0.001	0	0.001	0.001	0.001

表 3. 誤差②

	7BE1	1BHC	2IGR	1IRR	5IIT
IPC_protein	0.001	0	0.001	0.001	0.001
IPC_peptide	0.001	0	0.001	0.001	0.001
Toseland	0.001	0	0	0	0
Thurkill	0.001	0	0.001	0	0.001
Nozaki_Tanford	0	0.001	0	0	0
Dawson	0.015	0.017	0.001	0	0.001
DTASelect	0	0	0	0	0.001
EMBOSS	0.001	0.001	0	0	0
Grimsley	0	0.001	0	0.001	1.001
Patrickios	0	0.001	0	0	0
Rodwell	0	0	0	0	1.001
Sillero	0.001	0.001	0.001	0	0.001
Solomon	0.001	0	0.001	0	0.001
Lehninger	0.001	0	0	0.001	0
Wikipedia	0	0.001	0	0	0

```

1  #!/usr/bin/python3
2  # -*- coding:utf-8 -
3
4  import cgi
5  import sys
6  import numpy as np
7  import matplotlib.pyplot
8  matplotlib.use('Agg')
9  import matplotlib.pyplot as pl
10 import io
11 import base64
12
13 form = cgi.FieldStorage()
14 seq = form.getvalue('seq','')
15 pH_space = form.getvalue('pH_space','')
16
17 aa=["A","R","N","D","C","Q","E","G","H","I","L","K","M","F","P","S","T","W","Y","V"]
18 sa=["R","H","K"]
19 sb=["D","E","C","Y"]
20
21 #組成のディクショナリ作成
22 ns={}
23 N=0
24 for aax in aa:
25     n=seq.count(aax)
26     ns[aax] = n
27     N=N+n
28
29 #任意のpKaを取得し、ディクショナリにまとめる
30 NH2 = form.getvalue('NH2','')
31 pKa1 = float(NH2)
32
33 COOH = form.getvalue('COOH','')
34 pKa2 = float(COOH)
35
36 pKa3={}
37

```

図 1. プログラム①

```

38 r = form.getvalue('r','')
39 pKa3["R"]=float(r)
40 d = form.getvalue('d','')
41 pKa3["D"]=float(d)
42 c = form.getvalue('c','')
43 pKa3["C"]=float(c)
44 e = form.getvalue('e','')
45 pKa3["E"]=float(e)
46 h = form.getvalue('h','')
47 pKa3["H"]=float(h)
48 k = form.getvalue('k','')
49 pKa3["K"]=float(k)
50 y = form.getvalue('y','')
51 pKa3["Y"]=float(y)
52
53 #任意のpH間隔を取得
54 p_s=float(pH_space)
55 p=(14/p_s)+1
56 p=int(p)
57
58 #吸光係数の計算
59 e=ns["W"]*5500+ns["Y"]*1490+ns["C"]*125
60
61 #分子量の計算
62 S=18.01
63 gw=seq.count("G")*57.05
64 aw=seq.count("A")*71.08
65 sw=seq.count("S")*87.08
66 pw=seq.count("P")*97.11
67 vw=seq.count("V")*99.13
68 tw=seq.count("T")*101.10
69 cw=seq.count("C")*103.14
70 iw=seq.count("I")*113.16
71 lw=seq.count("L")*113.16
72 nw=seq.count("N")*114.10
73 dw=seq.count("D")*115.09
74 qw=seq.count("Q")*128.13
75 kw=seq.count("K")*128.17
76 ew=seq.count("E")*129.11

```

図2. プログラム②

```

77 mw=seq.count("M")*131.20
78 hw=seq.count("H")*137.14
79 fw=seq.count("F")*147.17
80 rw=seq.count("R")*156.19
81 yw=seq.count("Y")*163.17
82 ww=seq.count("W")*186.20
83
84 residue=[gw,aw,sw,pw,vw,tw,cw,iw,lw,nw,dw,qw,kw,ew,mw,hw,fw,rw,yw,ww]
85
86 for mw in residue:
87     S=S+round(mw,2)
88     Mw=round(S,2)
89
90 #滴定曲線の作成
91 Q={}
92 Q1=0
93 ph=[]
94 num=[]
95 for cnt in range(p):
96     pH=p_s*cnt
97     Q1=0
98     ph.append(pH)
99     nh=1/(1+10**(pH-pKa1))
100     cooh=-(10**(pH-pKa2)/(1+10**(pH-pKa2)))
101     Q1=cooh+nh
102     for aax in aa:
103         if aax in sa:
104             a=1/(1+10**(pH-pKa3[aax]))
105             Q1=Q1+a*ns[aax]
106         if aax in sb:
107             b=-(10**(pH-pKa3[aax]))/(1+10**(pH-pKa3[aax]))
108             Q1=Q1+b*ns[aax]
109     Q[pH]=Q1
110     num.append(Q1)
111
112
113 q=Q.values()
114
115 q_a=list(map(float,q))

```

図 3. プログラム③

```

116
117     q_b=list(Q.keys())
118
119     idx=np.abs(np.asarray(q_a)-0).argmin()
120
121     IP=q_b[idx]
122
123     pl.axhline(y=0, color='k')
124     pl.plot(ph,num)
125     pl.title('Titration curve')
126     pl.xlabel('pH')
127     pl.ylabel('charge')
128     pl.grid()
129     pl.xticks(np.arange(0,14,1.0))
130
131     def get_base64():
132         buffer = io.BytesIO()
133         pl.savefig(buffer,format='PNG')
134         b64 = base64.b64encode(buffer.getvalue()).decode()
135         return b64
136
137
138     sys.stdout = io.TextIOWrapper(sys.stdout.buffer, encoding="utf-8")
139     print("Content-Type: text/html; charset=utf-8")
140     print("")
141     print("<html>")
142     print("<body>")
143
144     b64 = get_base64()
145     print("<img src='data:image/png;base64," + b64 + "'>")
146
147     print("<br>")
148
149     print("composition",ns)
150
151     print("<br>")
152
153     print("molecular weight =",MW,"[Da]")
154

```

図 4. プログラム④

```

155     print("<br>")
156
157     print("Absorption coefficient=",e,"[A280/mol/cm]")
158
159     print("<br>")
160
161     print("Isoelectric point : ",IP)
162
163
164     print("</body>")
165     print("</html>")

```

図 5. プログラム⑤

```

1 <!DOCTYPE html>
2 <html>
3 <head>
4 <title>cgitest.py</title>
5 </head>
6 <body>
7 <p class="lead-form">Drawing a protein titration curve.</p>
8 <form action="/cgi-bin/Tcurve2.py" method="POST">
9
10 <div class="item">
11   <label for="pH_space">pH space</label>
12   <input type="text" value="0.5" name="pH_space" size="3">
13 </div>
14
15 <div class="item">
16   <label for="NH2">NH2's pKa</label>
17   <input type="text" value="9.094" name="NH2" size="3">
18 </div>
19
20 <div class="item">
21   <label for="COOH">COOH's pKa</label>
22   <input type="text" value="2.869" name="COOH" size="3">
23 </div>
24
25 <div class="item">
26   <label for="c">C's pKa</label>
27   <input type="text" value="7.555" name="c" size="3">
28 </div>
29
30 <div class="item">
31   <label for="d">D's pKa</label>
32   <input type="text" value="3.872" name="d" size="3">
33 </div>
34
35 <div class="item">
36   <label for="e">E's pKa</label>
37   <input type="text" value="4.412" name="e" size="3">
38 </div>
39

```

図6. プログラム⑥

```

40 <div class ="item">
41 | <label for="h">H's pKa</label>
42 | <input type="text" value="5.637" name="h" size="3">
43 </div>
44
45 <div class ="item">
46 | <label for="k">K's pKa</label>
47 | <input type="text" value="9.052" name="k" size="3">
48 </div>
49
50 <div class ="item">
51 | <label for="r">R's pKa</label>
52 | <input type="text" value="11.84" name="r" size="3">
53 </div>
54
55 <div class ="item">
56 | <label for="y">Y's pKa</label>
57 | <input type="text" value="10.85" name="y" size="3">
58 </div>
59
60 <div class="item">
61 | <label for="seq">Sequence</label>
62 | <textarea id="seq" name="seq" cols="40" rows="10"></textarea>
63 </div>
64
65 <div class="btn-area">
66 | <input type="submit" value="submit" />
67 </div>
68
69 </form>
70 </body>
71 </html>

```

図 7. プログラム⑦

```

1 # -*- coding:utf-8 -
2 import http.server
3 http.server.test([HandlerClass=http.server.CGIHTTPRequestHandler])

```

図 8. プログラム⑧

Drawing a protein titration curve.

pH space

NH2's pKa

COOH's pKa

C's pKa

D's pKa

E's pKa

H's pKa

K's pKa

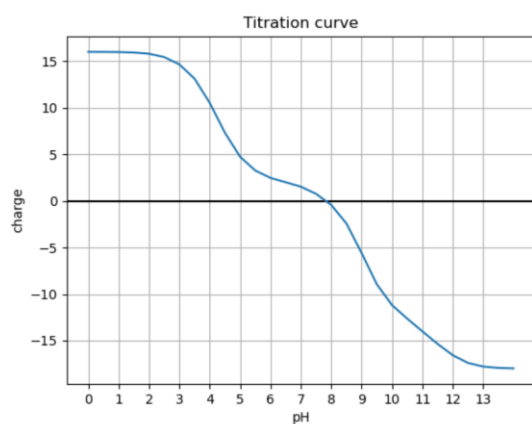
R's pKa

Y's pKa

Sequence



图 9. 入力画面



composition {'A': 5, 'R': 3, 'N': 5, 'D': 4, 'C': 2, 'Q': 1, 'E': 8, 'G': 12, 'H': 1, 'I': 9, 'L': 4, 'K': 11, 'M': 3, 'F': 3, 'P': 5, 'S': 10, 'T': 7, 'W': 1, 'Y': 3, 'V': 10}
molecular weight = 11508.99 [Da]
Absorption coefficient= 10220 [A280/mol/cm]
Isoelectric point : 8.0

图 10. 出力画面

第6章 参考文献

- Lukasz P. Kozlowski (2016) “IPC - Isoelectric Point Calculator” Biology Direct, 11(1), Additional file1 – table S3
- 有坂文雄 (2004) バイオサイエンスのための蛋白質科学入門, 裳華房
- 柴田淳 (2016) みんなのPython 第4版, SBクリエイティブ

第7章 謝辞

本研究を進めていくにあたり、指導教官の黒田裕教授にはたくさんの助言を賜りました。深く感謝を申し上げます。また、プログラミングに関しまして親切にご指導くださいました松沢さんにお礼申し上げます。さらに、日頃から多くのご指摘を下さいました黒田研究室の同期、先輩の皆様そして留学生の皆さまに感謝いたします。

最後になりましたが、留年してしまい、3年以上もの期間、大変お世話になり、多くのことを学ばせていただきました黒田研究室に感謝の言葉を述べさせていただくとともに、今後の益々のご発展をお祈りしています。本当にありがとうございました。

令和3年3月10日