

Celem projektu jest stworzenie programów komputerowych pozwalających na dokładniejsze przewidywanie punktu izoelektrycznego ( $pI$ ) białek i peptydów (punkt izoelektryczny to wartość pH przy którym suma ładunków ujemnych i dodatnich białka jest zrównoważona i całkowity ładunek cząsteczki jest neutralny). Parametr ten jest szeroko wykorzystywany w badaniach laboratoryjnych w celu identyfikacji określonego białka w mieszaninie celem jego dalszej analizy.

Precyzyjne przewidywanie punktu izoelektrycznego będzie opierać się na wykorzystaniu najnowszych technik sztucznej inteligencji a konkretnie użyta zostanie technologia firmy Google o nazwie Tensorflow, która jest szeroko stosowana w rozwiązywaniu złożonych problemów takich jak rozpoznawanie obrazów czy mowy. Należy podkreślić, że właśnie ta klasa sztucznych sieci neuronowych (ang. *deep learning*) została wykorzystana ostatnio do pokonania mistrzów szach, gry Go czy pokera, co jeszcze do niedawna było poza zasięgiem sztucznej inteligencji.

Punkt izoelektryczny jest obok masy cząsteczkowej podstawowym parametrem dzięki któremu można w łatwy i tani sposób rozdzielić mieszaninę białek. Przykładowo, w komórce człowieka w zależności od jej typu, ekspresji ulega około połowy białek (których jest około 20 tysięcy nie licząc izoform splicingowych) – inne białka aktywne są w komórce nerwowej a inne w komórce mięśniowej. Skład białek decyduje o tym co dana komórka jest w stanie robić. W przypadku większości chorób genetycznych na poziomie molekularnym dochodzi do zaburzenia w działaniu pojedynczego białka (rzadziej kilku białek). W celu opracowania leku należy zidentyfikować i zanalizować konkretne białko. Podobnie badając dowolny proces biologiczny z dużą dozą prawdopodobieństwa badacz będzie zainteresowany badaniem określonego białka lub jej grupy. W każdym z tych przypadków potrzebuje on jednorodnej próbki z oczyszczonym białkiem. W uproszczeniu można więc powiedzieć, że pierwszym etapem wielu analiz biologicznych jest izolacja konkretnego białka ze złożonej mieszaniny lizatu komórkowego. W tym celu można wykorzystać techniki takie jak elektroforeza dwuwymiarowa w żelu poliakrylamidowym (2D-PAGE) w której rozdziela się białka w dwóch wymiarach ze względu na ich masę cząsteczkową i punkt izoelektryczny. W skrócie po nałożeniu próbki pod wpływem działania prądu na żelu powstaje dwuwymiarowy obraz na którym odpowiednie plamy zawierają cząsteczki białek o określonej masie i ładunku. Na tym etapie można powiedzieć, że dany obszar zawiera białka o masie przykładowo około 50 kDa i  $pI$  wynoszącym około 5,5. Teraz należy określić jakie białka mogą posiadać takie parametry i w tym konkretnym momencie wkracza do działania oprogramowanie do przewidywania punktu izoelektrycznego. Obecnie programy tego typu pozwalają przewidzieć wartość  $pI$  z dokładnością do  $\pm 0,9$  jednostki czyli tak naprawdę dla naszego przykładowego białka będziemy musieli założyć, że dany fragment żelu zawiera dowolne białka o masie 45-55 kDa i  $pI$  4,6-6,4. Niestety takich białek może być kilkaset, może więc się okazać, że ciągle niezbędne będzie użycie innych, dokładniejszych (ale i droższych) metod biochemicznych (np. spektrometria mas), aby upewnić się co konkretnie zawiera dana próbka. Jeśli uda się stworzyć lepszy program do przewidywania  $pI$  zawężą to zakres niepewności i liczba potencjalnych białek które znajdują się w analizowanej próbce zostanie ograniczona. W efekcie skróci to cały proces izolacji białka.

Dodatkowo istnieje szereg innych technik molekularnych, które mogą skorzystać na znajomości teoretycznego punktu izoelektrycznego. Przykładowo wspomniana spektrometria mas często poprzedzona jest frakcjonowaniem próbki za pomocą  $pI$  (tzw. ogniskowanie izoelektryczne, IEF) w celu zmniejszenia złożoności dalszej analizy. Inną techniką, w której można wykorzystywać znajomość  $pI$  jest krystalografia rentgenowska białek. Dysponując oczyszczonym białkiem badacz musi zbadać działanie białka, a najlepszym tego sposobem jest określenie budowy tj. struktury białka. W tym celu hoduje się kryształy białka, które dzięki regularnemu ułożeniu identycznych cząsteczek można wykorzystać do zobrazowania wyglądu białka. Znając strukturę białka można wnioskować o sposobie jego działania lub w przypadku choroby zaproponować przyczynę dysfunkcji białka. Niestety, nawet wysoko oczyszczona próbka białkowa nie zawsze krystalizuje. Aby pobudzić białko do krystalizacji stosuje się szereg zabiegów umieszczając próbkę w złożonych mieszaninach związków chemicznych. Na tym etapie manipuluje się różnymi parametrami takimi jak skład, temperatura czy pH roztworu. Analiza tysięcy prób krystalizacji wykazała, że w większości przypadków białka nie krystalizują w pH równym ich punktowi izoelektrycznego, dlatego należy unikać takich roztworów. W tym momencie dokładniejsze teoretyczne przewidywanie punktu izoelektrycznego jedynie z sekwencji aminokwasowej może okazać się nieocenione. W przypadku wielu chorób wiemy jakie białko jest kluczowe, ale bez znajomości jego struktury nie jesteśmy w stanie zaprojektować odpowiedniego leku. Otrzymanie kryształów białka i w efekcie poznanie jego struktury często jest wynikiem wieloletnich prób obarczonych milionowymi kosztami. Dokładniejsza znajomość punktu izoelektrycznego pomoże skrócić ten proces.

Podsumowując, prezentowany projekt pozwoli na stworzenie programów i baz danych celem precyzyjniejszego określenia punktu izoelektrycznego. Ułatwi to analizę danych z całego szeregu metod biochemicznych stosowanych w praktycznie każdym laboratorium (już w tej chwili z wysoce uproszczonego prototypu proponowanej metody skorzystało ponad 100 tys. badaczy z całego świata; <http://isoelectric.org>). Pośrednio pozwoli to na szybszą i tańszą analizę budowy i funkcji białek związanych między innymi z chorobami genetycznymi, neurodegeneracyjnymi (Alzheimer, Parkinson) czy nowotworowymi.