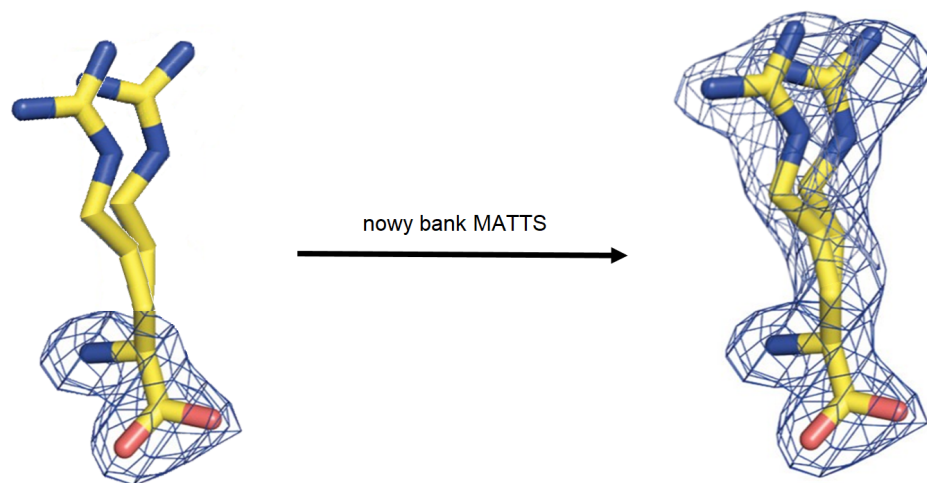


Ustalenie struktury makromolekuł, białek i ich kompleksów to jedno z podstawowych pytań badawczych stawianych przez biologię i chemię strukturalną. Białka to złożone cząsteczki składające się z dużej liczby aminokwasów połączonych ze sobą w łańcuch wiązaniami peptydowymi. Odpowiadają za budowę, funkcję i regulację tkanek i narządów wewnątrz organizmu oraz wykonują większość pracy w komórkach. Ich strukturę można określić na podstawie rozkładu gęstości elektronowej uzyskanej z eksperymentu dyfrakcji rentgenowskiej. Jedną z wartości opisujących jakość danych jest rozdzielczość – im jest ona lepsza, tym więcej szczegółów struktury można zaobserwować.

Uzyskanie danych dyfrakcji rentgenowskiej o wysokiej rozdzielczości dla makromolekuł jest trudne, ale konieczne do rekonstrukcji gęstości elektronowej używając asferycznego modelu multipolowego. Taki model przewyższa powszechnie używany sferyczny model pod względem opisu deformacji gęstości elektronowej z wiązań chemicznych i wolnych par elektronowych. Niestety, przeważająca większość dostępnych danych ma zbyt niską rozdzielczość i aby je wykorzystać stosuje się banki asferycznych typów atomowych. Takie banki pozwalają na wykorzystanie uśrednionego opisu podobnych do siebie atomów do odtworzenia gęstości elektronowej, zamiast opisywania każdego atomu osobno. Jednakże, obecnie dostępne banki nie są w stanie odtworzyć gęstości elektronowej większości makromolekuł bezpośrednio z pliku zawierającego ich strukturę – mmCIF.

W tym projekcie proponujemy rozwiązanie tego problemu. W tym celu wykorzystamy nowe podejście do rozpoznawania typów atomów dostosowane specjalnie do białek, RNA, DNA i najbardziej popularnych ligandów. Wykorzystując metody stworzone wcześniej przez naszą grupę badawczą, stworzymy nowy bank oparty o dane z największej na świecie krystalograficznej białkowej bazy danych – the Protein Data Bank. Proponujemy wprowadzenie algorytmu do rozpoznawania typów atomowych bezpośrednio na podstawie nazw atomów ściśle zdefiniowanych w słownikach i obecnych w cząsteczkach białek, RNA, DNA i najbardziej popularnych ligandów. Będzie to projekt wykorzystujący znaczne ilości danych i łączący ze sobą wiedzę z zakresu chemii i biologii strukturalnej oraz programowania.

Realizacja tego projektu poprawi jakość określania struktury molekularnej i rekonstrukcji rozkładu gęstości elektronowej białek, RNA, DNA i najpowszechniejszych ligandów. Ma to kluczowe znaczenie dla biologii strukturalnej i chemii, ponieważ analiza rozkładu ładunku pozwala określić funkcje makrocząsteczek, przewidzieć, jak mogą na nie wpływać zmiany w strukturze lub zrozumieć interakcje występujące w kompleksach molekularnych. Przewidujemy, że ta koncepcja powinna też rozwiązać problemy z atomami wodoru, które są trudne do zlokalizowania nawet w wysokiej rozdzielczości, ponieważ rozkład gęstości elektronowej z ich pojedynczego elektronu jest przesunięty w kierunku wiązania. Wprowadzenie naszego banku rozpoznającego typy atomowe na podstawie nazw atomów obecnych w cząsteczkach białek, RNA, DNA i najbardziej popularnych ligandów, znacząco wpłynie na dziedziny takie jak: farmakologia, biologia strukturalna, czy nanotechnologia.



Rysunek 1. Struktura argininy, będąca częścią białka. Atomy pokolorowane są według schematu: kolor niebieski - azot, czerwony - tlen, żółty - węgiel. Wodory nie są pokazane. Lewy panel: rekonstruowana gęstość elektronowa (kontur pokazany jako siatka) jest niekompletna z powodu braku poprawnego rozpoznania typów atomowych, które występują w dwóch konformacjach. Prawy panel: wszystkie typy atomowe są poprawnie rozpoznane i gęstość elektronowa może być poprawnie zrekonstruowana.