

Kwasy nukleinowe RNA i DNA są jedynymi nośnikami informacji genetycznej dla wszystkich form życia na Ziemi, a analiza związku między ich strukturą/dynamiką a funkcjami jest kluczem do zrozumienia mechanizmów biochemicznych, znalezienia celów terapeutycznych i opracowania nowych leków. Obecnie jednym z najtrudniejszych wyzwań związanych z badaniami pochodzenia życia na Ziemi jest zrozumienie w jaki sposób te podstawowe nośniki informacji genetycznej mogą podlegać samoreplikacji przy całkowitym braku enzymów. Co więcej, nie jest jasne, w jaki sposób RNA i DNA zostały wybrane jako podstawowe polimery informacyjne dla biologii z bogatej puli prebiotycznie utworzonych nukleotydów, w których skład również wchodziły alternatywne cukry i różne analogi zasad nukleinowych. Chociaż ostatnie postępy eksperymentalne w chemii prebiotycznej rzuciły pewne światło na możliwe mechanizmy nieenzymatycznej samoreplikacji RNA, postęp w tej dziedzinie został zahamowany przez brak zrozumienia struktury i dynamiki samoreplikujących się układów oraz bardzo fragmentaryczną wiedzę na temat leżących u ich podstaw molekularnych mechanizmów.

W ramach tego projektu zamierzam zastosować dokładne metody chemii obliczeniowej, aby sprostać tym wyzwaniom i zrozumieć pochodzenie DNA i RNA na ziemi. Chemia obliczeniowa posłuży tutaj jako mikroskop molekularny i pozwoli zrozumieć strukturę takich samoreplikujących się biomolekuł, a także opisać mechanizmy i energetykę podstawowych reakcji chemicznych. W szczególności, wielkoskalowe symulacje dynamiki molekularnej pozwolą nam scharakteryzować dynamiczne zachowanie samoreplikujących się fragmentów RNA, podczas gdy obliczenia chemii kwantowej dostarczą opisu zdarzeń zrywania i tworzenia kluczowych wiązań prowadzących do wydłużenia kopiowanej nici kwasu nukleinowego. Różnice energetyczne i strukturalne zaobserwowane dla różnych samoreplikujących się polimerów pozwolą nam zidentyfikować najbardziej wydajne warianty procesu samoreplikacji i bezpośrednio wesprzeć bieżące eksperymentalne wysiłki poprzez wyjaśnienia i przewidywania reaktywności. Aby przesunąć obecne granice metodologiczne, zamierzam również wytrenować potencjały uczenia maszynowego (ML) dla kwasów nukleinowych w oparciu o dokładne obliczenia chemii kwantowej. Obecne techniki uczenia maszynowego umożliwiają bardzo wydajne fitowanie funkcji wielu zmiennych, takich jak powierzchnie energii potencjalnej dużych cząsteczek. W rezultacie można uzyskać bardzo dokładny potencjał empiryczny, który może odtworzyć właściwości fizyczne i chemiczne układu molekularnego przy nieznacznych kosztach w porównaniu ze standardowymi metodami chemii kwantowej. W tym celu posłużę się najnowszymi modelami opartymi na sieciach neuronowych (np. PhysNet czy SpookyNet) aby uwzględnić kluczowe oddziaływania elektrostatyczne, van der Waalla i dalekiego zasięgu kształtujące struktury kwasów nukleinowych. Taki potencjał ML zostanie następnie zastosowany do najtrudniejszych fragmentów kanonicznych i alternatywnych kwasów nukleinowych, które mogą podlegać nieenzymatycznej samoreplikacji.

Podsumowując, projekt ten pozwoli nam zrozumieć, dlaczego RNA i DNA zostały wybrane z bogatej puli różnych bloków budujących kwasy nukleinowe i jak zaczęły się one kopiować, aby zainicjować proliferację pierwszych protokomórek. Wyniki tego badania pozwolą również lepiej ukierunkować bieżące poszukiwania biosygnatur, co jest jedną z głównych misji niedawno wystrzelonego Kosmicznego Teleskopu Jamesa Webba. Wreszcie, wyszkolony potencjał ML ukierunkowany na szerokie spektrum kwasów nukleinowych może również stać się dokładnym i opłacalnym narzędziem do odkrywania i projektowania terapii farmakologicznych ukierunkowanych na RNA i DNA.