

# Metody analizy aktywności katalitycznej i projektowania biokatalizatorów oparte na teorii oddziaływań międzycząsteczkowych

Wiktor Beker

Praca doktorska wykonana pod kierunkiem  
prof. dr hab. inż. W. Andrzeja Sokalskiego

Katedra Inżynierii i Modelowania Materiałów Zaawansowanych,  
Wydział Chemiczny,  
Politechnika Wrocławska

Wrocław, 2017



*Mojej Żonie, dziękując jej za ciągly i  $\infty$ -krotnie różniczkowalny strumień moralnego wsparcia*



# Podziękowania

W pierwszej kolejności pragnę podziękować mojej Żonie, Dorocie, za podtrzymywanie mnie na duchu w chwilach zwątpienia. Gorąco dziękuję także mojemu mentorowi, profesorowi Andrzejowi Sokalskiemu, za jego (nie tylko naukową) opiekę przez wszystkie lata naszej współpracy, niezachwianą wiarę w moje możliwości i zaufanie, jakim mnie darzył, powierzając mi różne zadania. Składam również wyrazy uznania moim współpracownikom, w szczególności:

- doktorowi Karolowi Langnerowi, po którym przejąłem oprogramowanie do obliczeń z udziałem CAMM,
- doktor Edycie Dygudzie-Kazimierowicz, za udostępnienie ścieżek reakcji enzymatycznych i pomoc w nauce sztuki ich obliczania,
- doktorowi Marcowi van der Kampowi, za udostępnienie ścieżek reakcji w izomerazie ketosteroidowej i współpracę w badaniach nad tym układem,
- profesorowi Adrianowi Mulhollandowi, za współpracę, opiekę merytoryczną i umożliwienie mi wizyty w jego zespole na Uniwersytecie Bristolskim.

Szczerze dziękuję także tym, którzy pomagali mi w edycji niniejszego tekstu: mojej Żonie, promotorowi, dr Edycie i mgr inż. Mateuszowi Jasikowi.

Na koniec dziękuję instytucjom, które finansowały prowadzone przeze mnie badania w ramach projektów:

- projekt Wrocławskiego Centrum Badawczego EIT+ *Biotechnologie i zaawansowane technologie medyczne* BIOMED (POIG.01.01.02-02-003/08) współfinansowany przez Europejski Fundusz Rozwoju Regionalnego (Program operacyjny Innowacyjna Gospodarka, 1.1.2),
- project Nr DI2012 016642 *Zastosowanie dynamicznych pól katalitycznych w projektowaniu enzymów* w ramach programu „Diamantowy Grant” Polskiego Ministerstwa Nauki i Szkolnictwa Wyższego,
- projekt *Zastosowanie modeli elektrostatycznych uwzględniających właściwości dielektryczne w projektowaniu biokatalizatorów* w ramach programu „Młoda Kadra”, współfinansowanego przez Unię Europejską w ramach Europejskiego Funduszu Społecznego.

Dziękuję także Wrocławskiemu i Poznańskiemu Centrum Sieciowo-Superkomputerowemu oraz Interdyscyplinarnemu Centrum Modelowania Matematycznego w Warszawie, w których prowadzone były obliczenia na potrzeby tej pracy.



# Spis treści

<b>Spis treści</b>	<b>7</b>
<b>1 Wstęp</b>	<b>9</b>
1.1 Perspektywa . . . . .	9
1.2 Źródła katalizy enzymatycznej . . . . .	10
1.2.1 Hipoteza dynamiczna . . . . .	11
1.2.2 Kataliza elektrostatyczna . . . . .	13
1.2.3 Efekty entropowe i inne . . . . .	14
1.2.4 Krytyczna ocena . . . . .	15
1.3 Historia projektowania <i>de novo</i> enzymów . . . . .	15
1.3.1 Początki . . . . .	15
1.3.2 Algorytm projektowania enzymów . . . . .	17
1.3.3 Eliminaza Kempa . . . . .	17
1.4 Cel pracy . . . . .	18
<b>2 Metody</b>	<b>21</b>
2.1 Różnicowa stabilizacja stanu przejściowego . . . . .	21
2.1.1 Reakcje wieloetapowe . . . . .	22
2.2 Oddziaływania międzycząsteczkowe . . . . .	23
2.2.1 Przybliżenie polaryzacyjne . . . . .	24
2.2.2 Rachunek zaburzeń o adaptowanej symetrii . . . . .	25
2.3 Składowe energii oddziaływań . . . . .	27
2.3.1 Addytywność oddziaływań . . . . .	27
2.3.2 Dekompozycja energii oddziaływań . . . . .	27
2.4 Pola katalityczne . . . . .	29
2.5 Kumulatywne atomowe rozwinięcie multipolowe . . . . .	30
2.5.1 Biblioteka CAMM rotamerów aminokwasów . . . . .	31
2.6 Obliczenia QM/MM . . . . .	31
<b>3 Wyniki</b>	<b>33</b>
3.1 Analiza modeli elektrostatycznych . . . . .	33
3.1.1 Reprezentacja rozkładu ładunku układu reagującego . . . . .	33
3.1.2 Reprezentacja otoczenia katalitycznego . . . . .	34
3.2 Wpływ mutacji na aktywność katalityczną . . . . .	39
3.2.1 Izomeraza ketosteroidowa . . . . .	39
3.2.2 Analiza eliminazy Kempa KE07 i jej mutantów . . . . .	43
<b>4 Podsumowanie</b>	<b>57</b>
<b>Suplement</b>	<b>59</b>
S.1 Przegląd pojęć związanych z rozwinięciem multipolowym . . . . .	59
S.2 SAPT– komentarze szczegółowe . . . . .	59
S.2.1 Deltę Murrella i Landshoffa . . . . .	59
S.2.2 Pierwsza poprawka SAPT w bazach MO i AO . . . . .	60
S.2.3 Całka elektrostatyczna w bazie AO . . . . .	61

S.3	Dynamiczne pole katalityczne . . . . .	62
S.3.1	Testy . . . . .	63
S.4	Powiązanie z metodą zaburzenia energii swobodnej (FEP) . . . . .	66
S.4.1	Korelacja pomiędzy DTSS a barierą własną . . . . .	66
S.5	Obliczenia ONIOM . . . . .	67
S.6	Reprezentacja $\Delta_S$ . . . . .	70
S.7	Model ciągłego rozpuszczalnika . . . . .	70
S.7.1	Walidacja: subtylizyna . . . . .	72
S.8	Biblioteka multipoli atomowych . . . . .	73
S.8.1	Motywacja utworzenia biblioteki . . . . .	73
S.8.2	Zależność konformacyjna CAMM . . . . .	74
S.8.3	Rodzaje interpolacji – testy 1D i 2D . . . . .	74
S.8.4	Testy 2D i 3D . . . . .	78
S.8.5	Uwzględnienie konformacji łańcucha głównego . . . . .	80
	<b>Bibliografia</b>	<b>83</b>
	<b>Spis rysunków</b>	<b>95</b>
	<b>Spis tabel</b>	<b>97</b>
	<b>Spis publikacji</b>	<b>99</b>
	<b>Spis wystąpień konferencyjnych</b>	<b>101</b>



# Rozdział 1

## Wstęp

### 1.1 Perspektywa

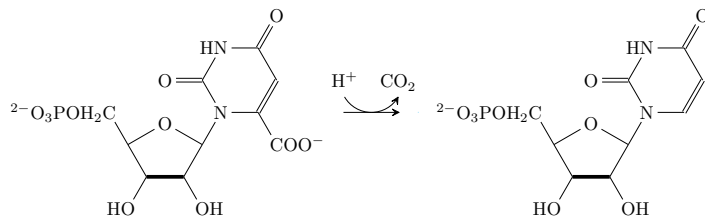
Choć różnie można zdefiniować kierunek rozwoju nauki i techniki, w kontekście prowadzonych badań rozsądnie jest przyjąć, że wiedzie on ku coraz mniejszym obiektom. Zmagania, jakie toczyliśmy próbując okiełznać Naturę, rozpoczynaliśmy od przedmiotów o rozmiarach porównywalnych z tymi, które możemy unieść we własnych rękach i zobaczyć gołym okiem, jak kamień, patyk czy piłka; stąd także o rozmiarach budowanych urządzeń był początkowo podobnego rzędu. Stopniowe ujarzmianie praw Przyrody pozwoliło zredukować tworzonych konstrukcji, czego doskonałym przykładem jest miniaturyzacja urządzeń elektronicznych w ciągu ostatnich kilku dekad. Pierwsze komputery — urządzenia zajmujące kilka szaf — miały moc obliczeniową zbliżoną do dzisiejszego kieszonkowego kalkulatora, zaś dzisiejsze telefony komórkowe mają możliwości większe niż komputery osobiste sprzed 20-30 lat. Kontynuacja rozwoju w tym kierunku — ku coraz mniejszym mechanizmom — wiedzie ku maszynom o rozmiarach molekuł, manipulujących materią na poziomie atomowym. Wizję tę przedstawił już w latach '50 słynny fizyk Richard Feynmann, w wykładzie pod tytułem *There's Plenty of room at the bottom* (ang. *Tam na dole jest dużo miejsca*).

O ile zsyntetyzowano już kilka cząsteczek mogących służyć za proste „nanomaszyny” [1], wciąż daleko do stworzenia układów w pełni odpowiadających wizji cytowanego Noblisty. Droga ku temu celowi jest jeszcze długa, chyba że zwróci się uwagę w stronę *istniejących nanomaszyn*. Każdy Czytelnik zaznajomiony z biologią molekularną i biochemią powinien wiedzieć już, do czego zmierza ten opis. Organizmy żywe, od bakterii po człowieka, wykonują dokładnie taką pracę: przekształcają otaczającą materię wedle swoich potrzeb, atom po atomie. Azot i dwutlenek węgla z atmosfery, dzięki energii światła słonecznego, w płataninie reakcji biochemicznych przetwarzane są na całą (molekularną) złożoność biosfery, od prostego cukru po podwójną helisę DNA. Nie byłoby to możliwe bez wyspecjalizowanych nanomaszyn, jakimi są enzymy.

Mikroorganizmy bądź też wyizolowane z nich białka o aktywności katalitycznej są chętnie stosowane w różnych gałęziach przemysłu chemicznego, przede wszystkim ze względu na swoją regio- i stereoselektywność. Pierwsza z wymienionych cech biokatalizatorów pozwala znacząco uprościć proces syntezy docelowego związku, która przeprowadzana tradycyjnymi technikami wymaga wielu dodatkowych kroków, np. przyłączania i odłączania grup protekcyjnych. Druga cecha jest szczególnie pożądana w przypadku produktów aktywnych biologicznie, gdzie zmiana chiralności substancji może wiązać się z wręcz drastycznymi konsekwencjami. Być może najdobitniejszym przykładem takiego zjawiska jest historia Talidomidu [2], stosowanego w latach '50 środka nasennego; stereoizomer *R* ma działanie uspokajające i przeciwbólowe, natomiast jego lustrzane odbicie jest silnym teratogenem, zaburzającym m.in. rozwój kończyn u płodu. Chociaż asymetryczna synteza organiczna jest już możliwa, enzymy wciąż stanowią konkurencyjne i eleganckie rozwiązanie tego problemu [3].

Uzyskanie takiej specyficzności jest możliwe dzięki odpowiedniemu dopasowaniu kieszeni aktywnej, faworyzującemu konkretne orientacje substratów, oraz stworzeniu wewnątrz białka środowiska obniżającego barierę aktywacji tak, by reakcja mogła zajść w rozsądnym czasie. Jednym z najbardziej ekstremalnych przykładów zdolności katalitycznej enzymów jest dekarboksylaza orotydylanowa, katalizująca jeden z etapów biosyntezy nukleotydów pirymidynowych (Rysunek 1.1). Pod nieobecność enzymu, reakcja ta przebiega skrajnie wolno- obliczono, że w obojętnym roztworze w temperaturze

pokojoyej okres półtrwania tego związku wynosi 78 *milionów* lat. Biorąc pod uwagę, że w obecności dekarboksylazy czas ten jest rzędu setnych części sekundy, wywnioskowano, że enzym zwiększa szybkość reakcji  $10^{17}$  razy [4].



Rysunek 1.1: Reakcja katalizowana przez dekarboksylazę orotydylanową.

Problemy stojące na przeszkodzie w zastosowaniu enzymów wynikają w znacznej mierze z tego, że, jako produkty ewolucji biologicznej, dostosowane są przede wszystkim do tych reakcji chemicznych, które służą przetrwaniu danego organizmu. Szereg reakcji o potencjalnym znaczeniu przemysłowym nie posiada swojego odpowiednika biochemicznego. Stąd wynika konieczność bądź to modyfikowania istniejących enzymów (w celu np. zmiany ich specyficzności substratowej), bądź też projektowanie ich *de novo*. Oba te zadania wymagają przeszukiwania olbrzymiej przestrzeni możliwych mutacji, rosnącej eksponencjalnie wraz z liczbą rozważanych pozycji. Dla przykładu, istnieje  $20^{10} \simeq 10^{13}$  deka-peptydów złożonych tylko i wyłącznie z aminokwasów biogennych; liczba możliwości w przypadku projektowania *de novo* jest jeszcze większa, jako że typowe białko składa się z kilkuset reszt aminokwasowych. Nawet ograniczywszy ten zbiór, jego rozmiar wciąż pozostaje poza zasięgiem eksperymentalnych metod badawczych. Dlatego cennym — wręcz niezbędnym — narzędziem są metody modelowania komputerowego; ponieważ są one rozwijane w oparciu o zdobytą wiedzę o źródłach efektów katalitycznych w enzymach, wyzwaniem to stanowi zarazem sprawdzian naszego rozumienia katalizy enzymatycznej.

## 1.2 Źródła katalizy enzymatycznej

Współczesne rozumienie katalizy enzymatycznej sięga korzeniami pierwszej połowy XX wieku. W 1921 roku Węgier Mihaly Polanyi jako pierwszy zasugerował istnienie związku pomiędzy efektami katalitycznymi a oddziaływaniami międzycząsteczkowymi [5]. Myśl tę rozwinął Linus Pauling w swojej słynnej hipotezie, sformułowanej w drugiej połowie lat '40 ubiegłego wieku [6–8]. Centralnym jej punktem jest twierdzenie, że enzymy katalizują reakcje chemiczne poprzez specyficzną stabilizację odpowiedniego stanu przejściowego, a tym samym obniżenie bariery aktywacji; podstawą do wysunięcia tej hipotezy była opracowana wcześniej przez Henry'ego Eyringa w 1935 r. teoria aktywowanego kompleksu, później nazwana teorią stanu przejściowego (ang. *Transition-State Theory*, TST) [9]. Wychodząc z termodynamiki statystycznej, stałą szybkości reakcji według TST można najogólniej wyrazić za pomocą równania (1.1).

$$k_{TST} = \frac{1}{2} \langle |\dot{x}| \rangle_{TS} \exp(-\Delta g^\ddagger \beta) \left/ \int_{-\infty}^{\ddagger} \exp(-\Delta g(x)\beta) dx \right. \quad (1.1)$$

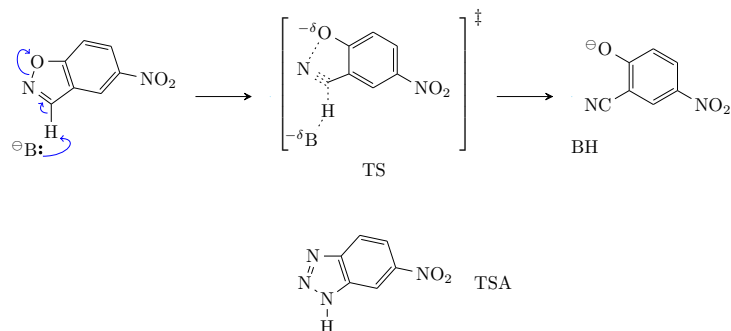
W powyższym równaniu,  $\beta$  symbolizuje czynnik  $(k_B T)^{-1}$  ( $k_B$  — stała Boltzmana),  $\Delta g(x)$  — energię swobodną względem substratów jako funkcję współrzędnej reakcji  $x$  (w szczególności energię aktywacji  $\Delta g^\ddagger = \Delta g(x^\ddagger)$ ), zaś  $\langle |\dot{x}| \rangle_{TS}$  średnią szybkość wzdłuż współrzędnej reakcji w rejonie stanu przejściowego. Przy założeniu ekwipartycji energii pomiędzy współrzędną reakcji a pozostałymi stopniami swobody oraz harmonicznosci funkcji  $\Delta g(x)$ , równanie to redukuje się do (1.2).

$$k_{TST} \approx (\beta h)^{-1} \exp(-\beta \Delta g^\ddagger) \quad (1.2)$$

W praktyce,  $k_{TST}$  mnoży się dodatkowo przez tzw. współczynnik transmisji  $\kappa$ , który zwykle uzależnia się od czynników takich jak powtórne przekraczanie bariery (ang. *barrier-recrossing*), oraz współczynnik tunelowania  $\gamma$ .

Hipoteza Paulinga była m.in. inspiracją do wykorzystania w badaniach tzw. analogów stanów przejściowych (ang. *Transition-State Analogues*, TSA) — związków chemicznych, których cząsteczki przypominają pod względem struktury i rozkładu ładunku stan przejściowy danej reakcji; zgodnie z hipotezą

Paulinga powinny one być znacznie lepiej wiązane przez enzymy, przez co stanowią potencjalnych kandydatów na ich inhibitory [10]. Przykładem może być 5-nitrobenzotriazol, TSA będący inhibitorem eliminazy Kempa omawianej w sekcji 1.3.3 (*vide* Rysunek 1.2) [11]. Zastosowanie metod pozwalają-



Rysunek 1.2: Reakcja eliminacji Kempa oraz jej analog stanu przejściowego.

cych na określenie struktury atomowej białka (jak techniki krystalografii rentgenowskiej i neutronowej czy magnetycznego rezonansu jądrowego) do kompleksów enzym-TSA dostarcza informacji na temat rzeczywistego stanu przejściowego; wraz z innymi technikami eksperymentalnymi, jak np. badanie kinetycznych efektów izotopowych [12,13]<sup>1</sup> czy kierunkowana mutageneza (ang. *Site-Directed Mutagenesis*, SDM) [17], stanowi potężny zestaw narzędzi służących weryfikacji mechanizmów reakcji i określaniu roli poszczególnych reszt aminokwasowych w katalizie. Dalszym owocem TST był rozwój badań nad przeciwciałami katalitycznymi (abzymami od ang. *Antibody enzyme*) zapoczątkowany w latach '80 [18,19]. Wykorzystano tutaj zdolność układu odpornościowego do wytwarzania przeciwciał komplementarnych do danego antygenu, w tym przypadku związanego z analogiem stanu przejściowego. Chociaż uzyskane w ten sposób abzymy rzeczywiście wykazywały aktywność katalityczną, była ona o kilka rzędów wielkości niższa od tej prezentowanej przez analogiczne enzymy. Zrodziło to przypuszczenia, że oprócz stabilizacji stanu przejściowego muszą istnieć inne czynniki kluczowe do tak efektywnej katalizy obserwowanej w naturze. Jednym z możliwych wytłumaczeń jest nieoptymalne oddziaływanie abzymów z substratami<sup>2</sup>; pośród innych koncepcji na szczególną uwagę zasługują te skoncentrowane wokół naturalnej dla enzymów labilności łańcucha białkowego.

### 1.2.1 Hipoteza dynamiczna

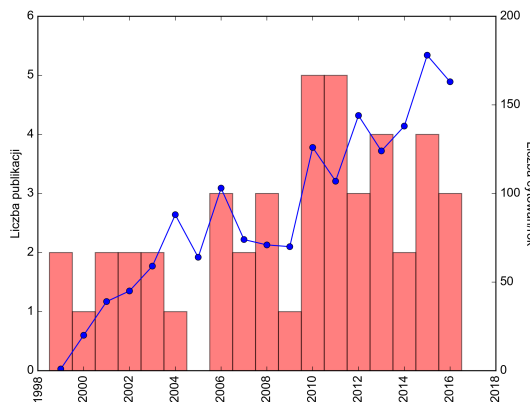
Problem znaczenia dynamiki enzymów dla ich zdolności katalitycznych od prawie dwóch dekad zyskuje na popularności, o czym świadczyć może analiza cytowań z bazy *Web of Science*, zaprezentowana na Rysunku 1.3. W wymienionym okresie pojawiło się zaledwie 45 publikacji odpowiadających frazom kluczowym *dynamical effects* i *enzyme catalysis*<sup>3</sup>, podczas gdy łączna liczba ich cytowań na dany rok kalendarzowy wzrastała prawie liniowo, osiągając maksymalną liczbę 178 cytowań w roku 2015. Umysły badaczy i dyskusje pomiędzy nimi rozpaliała hipoteza, że nie tylko struktura enzymów, ale także *dynamika ich części składowych* ewoluowały tak, by jak najbardziej najoptymalnie katalizować reakcje chemiczne [20]. Bazując na badaniach (zarówno eksperymentalnych jak i teoretycznych) nad reduktazą dihydrofolianową (DHFR od ang. *Dihydrofolate reductase*) [21,22] czy cyklofiliną A (CphA od ang. *Cyclophilin A*) [23,24], zaproponowano szereg modeli mających tłumaczyć wpływ dynamiki białka na aktywność katalityczną, np. teorię drgań promujących [25,26].

Podstawowym problemem ożywionego dyskursu na ten temat jest nie zawsze jednoznaczne definiowanie pojęć i układu odniesienia oraz mylenie przyczyny ze skutkiem, jak punktuje Warshel i współpracownicy [27,28]. Podążając za zaproponowanym przez nich tokiem rozumowania, właściwą podstawą

<sup>1</sup> Podwaliny pod tę technikę położył Jacob Bigeleisen w 1947 roku i latach następnych [14,15]; jednym z pierwszych jej zastosowań do reakcji enzymatycznych były prace Morrisona i Clelanda w latach '60 [16], lecz prawdziwy rozkwit metody nastąpił dopiero w następnej dekadzie [12].

<sup>2</sup> Brak tzw. destabilizacji substratu, proponowanego mechanizmu funkcjonowania enzymów; hipoteza ta będzie wspomniana przy okazji omawiania efektu Kirke w sekcji 1.2.3.

<sup>3</sup> Szczegóły wyszukiwania: frazy *"dynamic\* effect"* i *"enzym\* cataly\*"* w polu *Topic*, połączone spójnikiem logicznym AND



Rysunek 1.3: Analiza cytowań prac zawierających frazy *dynamic effect* i *enzyme catalysis* z ostatnich dwóch dekad. Słupki- liczba opublikowanych w danym roku artykułów, linia ciągła- liczba cytowań w danym roku.

wszelkich rozważań dotyczących efektów dynamicznych powinno być możliwie precyzyjne sformułowanie, co rozumiemy pod pojęciem *kataliza*, oraz w jaki sposób *dynamiczność* (tzn. ruchy otoczenia białkowego) ma ją wspomagać. Powszechnie katalizę rozumie się jako zjawisko polegające na przyśpieszeniu pewnego procesu, w omawianym kontekście — reakcji chemicznej. Ponieważ zaś *przyśpieszenie* jest pojęciem względnym, bardzo ważne jest określenie układu referencyjnego, względem którego ono następuje. Zwyczajowym punktem odniesienia jest reakcja zachodząca w tych samych warunkach i rozpuszczalniku, lecz pod nieobecność katalizatora; to samo podejście winno się też stosować wobec wszelkich zjawisk, które nazywać chcemy katalitycznymi, także dynamiki otoczenia reagentów. I tak, dla przykładu, sprzężenie konformacji reszty argininy 55 (tworzącej wiązanie wodorowe z substratem) [29] ze współrzedną reakcją w CphA nie jest samo w sobie niczym szczególnym, jako że cząsteczki rozpuszczalnika w bezpośrednio oddziałujące z reagentami także muszą dostosować swoją orientację do zmieniającej się struktury reagentów; istotna różnica polega na tym, że w przypadku roztworu wiąże się to ze znacznie większą zmianą energii swobodnej otoczenia niż w przypadku enzymu (tzw. energia reorganizacji rozpuszczalnika). Z tej perspektywy eksperymenty z mutantem Ser99Thr CphA [24], w których zauważono zmianę aktywności katalitycznej proporcjonalną do zmiany szybkości zmian konformacyjnych, nie dowodzą, że mają one znaczenie katalityczne *per se*, ale że wprowadzona mutacja upośledza optymalną eksplorację przestrzeni fazowej tego układu<sup>4</sup>. Innymi słowy, należy rozróżnić efekt katalityczny od funkcjonowania środowiska białkowego.

Dokładne przedstawienie dyskusji wokół różnych koncepcji efektów dynamicznych nie jest zasadne z punktu widzenia celu tej pracy; dlatego Autor ogranicza się tutaj do przedstawienia wyłącznie wybranych przykładów i ogólnego zarysowania aparatu pojęciowego, za pomocą którego problem ów będzie rozpatrywany. Bardziej zainteresowanemu tą kwestią Czytelnikowi warto polecić cytowane prace [27] i [28], na których wzorowany jest przedstawiony pogląd.

Zjawiska opisywane jako *dynamiczne efekty katalityczne* odpowiadają w istocie różnym procesom fizycznym, dlatego właśnie ich fizyczny opis przyjęto za podstawowe kryterium opisu. Podstawowym narzędziem w analizie fazy skondensowanej jest równowagowa (*boltzmannowska*) mechanika statystyczna i oparta na niej teoria stanu przejściowego (TST). W tym ujęciu najważniejsze z efektów niedających się opisać za pomocą jednej struktury, a przez to intuicyjnie określanymi jako *dynamiczne*, można sklasyfikować w sposób następujący.

### Zmiana dostępnej przestrzeni fazowej

Wpływ części mutacji można tłumaczyć poprzez wywołaną zmianę dystrybucji dostępnych konformacji; do tej kategorii zaliczać się będą efekty obserwowane w kierunkowanej ewolucji zaprojektowanych teoretycznie enzymów, o czym mowa w 3.2.2.

### Zmiana współczynnika transmisji lub tunelowania

<sup>4</sup> Właściwie można powiedzieć, że w zmutowanym enzymie położenie Arg55 nie jest optymalne dla katalizy [28].

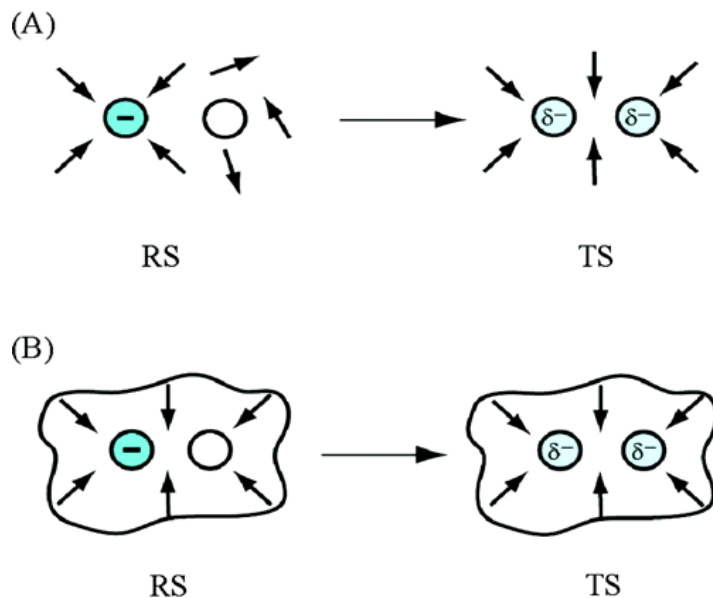
W reakcjach związanych z przeniesieniem protonu lub jonu wodorkowego, czynnikiem katalitycznym może być istotna zmiana współczynnika transmisji lub tunelowania względem reakcji przebiegającej w roztworze; w praktyce nie znaleziono przykładu reakcji enzymatycznej o współczynniku transmisji  $\kappa$  większym od ok. 3 [30], zaś efekt tunelowy raczej nie ma znaczenia katalitycznego [31].

### Odstępstwa od TST i równowagowej termodynamiki statystycznej

Warunek ekwipartycji energii leżący u podstaw TST nie musi być spełniony; w takich układach obserwowana szybkość reakcji ma prawo nie zgadzać się z modelem opartym na termodynamice boltzmannowskiej, może również występować egzotyczny efekt pamięciowy (w którym prawdopodobieństwo przejścia z substratu do produktu zależy od wcześniejszej historii układu), określane także jako koherencja dynamiczna; dotyczy przede wszystkim bardzo szybkich procesów (np. reakcje fotochemiczne); w przypadku innych reakcji, jak dotąd, niezauważony [32].

### 1.2.2 Kataliza elektrostatyczna

Jeśli by wskazać jeden uniwersalny czynnik charakteryzujący katalizę enzymatyczną, dobrym kandydatem jest koncepcja elektrostatycznej preorganizacji zaproponowana przez Warshela [33]. Polega ona na obserwacji, że o ile oddziaływania (np. wiązania wodorowe) stabilizujące stan przejściowy obecne są również w roztworze, to w przypadku rozpuszczalnika osiągnięcie takiej konfiguracji wiąże się z dodatkową porcją energii swobodnej potrzebnej, by dostosować orientację losowo zorientowanych cząsteczek do zmiennej konfiguracji reagentów; w przypadku enzymu, uczestniczące w procesie reszty aminokwasowe z reguły są już blisko wymaganego położenia i ich reorientacja nie wiąże się z takim kosztem energetycznym (Rysunek 1.4). Stąd nazywa się środowisko w kieszeni aktywnej enzymu *preorganizowanym*, czyli wstępnie uporządkowanym. Szczegółowe definicje wraz cyklami termodynamicznymi Czytelnik może znaleźć w pracach [34, 35] oraz pozycjach tam cytowanych.

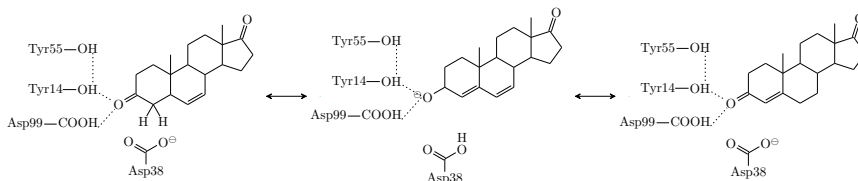


Warshel, *Biochemistry* 46, 1466-1476, 2007

Rysunek 1.4: Schemat przedstawiający ideę elektrostatycznej preorganizacji. Reakcja, w której zachodzi przeniesienie ładunku (np.  $S_N2$ ) wiąże się z dużą reorganizacją rozpuszczalnika (reprezentowanego schematycznie przez wektory momentów dipolowych) (A), lecz niewielką enzymu, którego rozkład ładunku jest rozmieszczony optymalnie – *preorganizowany* (B) [35] (tłumaczenie i przedruk za zgodą wydawnictwa Elsevier).

Efekt ten został zaobserwowany w wielu układach enzymatycznych [35], w tym także, co niespodziewane, w katalizujących reakcje rodnikowe enzymach zależnych od witaminy  $B_{12}$  [36]. Świadczy to nie

tylko o wiodącym znaczeniu oddziaływań elektrostatycznych w katalizie z udziałem enzymów, ale także o tym, że ważne jest odpowiednie *ukształtowanie* środowiska katalitycznego pod kątem danej reakcji. Dalszych argumentów dostarczają prace wykonane w naszej grupie, dotyczące takich enzymów jak mutaza choryzmianowa [37], kinaza białkowa A (PKA) [38] czy izomeraza ketosteroidowa (KSI) [39]. Ostatni z wymienionych enzymów jest szczególnie interesujący, ponieważ eksperymentalnie zmierzono natężenie pola elektrycznego w kieszeniach aktywnych szeregu jego mutantów i wykazano liniową relację pomiędzy tą wielkością a efektywną barierą aktywacji<sup>5</sup> [40]. Przedstawiony w pracy podział efektu katalitycznego na *katalizę elektrostatyczną* i *preorientację zasady* wywołał gorącą dyskusję w środowisku naukowym [41, 42]. Podobnie ożywiona wymiana opinii miała miejsce, kiedy wyniki eksperymentów wykorzystujących fenolany jako TSA sugerowały niewielkie znaczenie oddziaływań elektrostatycznych w katalizie zachodzącej w KSI [43, 44]. W obu przypadkach kluczowe okazało się nie tyle rozpatrzenie obecności tego rodzaju oddziaływań lub ich braku, lecz ich odpowiednia (pre)organizacja.



Rysunek 1.5: Schemat reakcji katalizowanej przez izomerazę ketosteroidową (KSI).

### 1.2.3 Efekty entropowe i inne

Przedstawiony w poprzednich sekcjach obraz należy uzupełnić o efekty entropowe, których badanie i modelowanie stanowi spore wyzwanie. Część entropowego wkładu do całkowitego efektu katalitycznego jest obecna już w różnicy pomiędzy energiami preorganizacji enzymu i roztworu wodnego ( $\langle \Delta U \rangle_0^P - \langle \Delta U \rangle_0^w$  z Rysunku 1.4), istnieje jednak kilka innych propozycji związanych z tą wielkością fizyczną; wybrane z nich są w skrócie opisane poniżej.

#### *Near Attack Conformation* (NAC) - konformacja bliska atakowi.

Jest to koncepcja wprowadzona przez Thomasa Bruce'a w roku dwutysięcznym, wedle której podstawowym mechanizmem działania enzymów jest wiązanie substratu w konformacji możliwie zbliżonej do stanu przejściowego [45]. Przykładowo, dehalogenaza haloalkanów (ang. *haloalkane dehalogenase*, DhIA<sup>6</sup>), katalizująca reakcję substytucji nukleofilowej  $S_N2$  wydaje się funkcjonować wedle tego schematu: faworyzowanie przez enzym konfiguracji odpowiadającej atakowi nukleofilowemu jest zgodne z „chemiczną intuicją” [46]. Jednakże bardziej szczegółowa analiza pokazuje, że z całkowitego obniżenia bariery przez enzym rzędu 10 kcal/mol<sup>7</sup> efekt ten stanowi wkład zaledwie ok. 2.8 kcal/mol; resztę przypisuje się efektom elektrostatycznym [47]. Podobna dyskusja dotyczyła mutazy choryzmianowej; w tym przypadku czołowy argument przeciwko NAC stanowi, że wiązanie tej konformacji nie jest warunkiem katalizy, a efektem stabilizacji stanu przejściowego przez enzym [47]. Kwestia ta jest dyskusyjna; istnieje możliwość, że rozstrzygnięcie pomiędzy tymi modelami zależy od arbitralnie przyjętego sposobu uszeregowania efektów katalitycznych.

#### Efekt Kirke<sup>8</sup>

W 1975 William Jencks zauważył podobieństwo pomiędzy energią swobodną wiązania substratu a zmianą entropowego przyczynku do energii aktywacji niektórych reakcji enzymatycznych

<sup>5</sup> Coraz popularniejsze pomiary tego rodzaju wykonuje się wykorzystując efekt Starka w widmie oscylacyjnym: badając przesunięcie spektralne pasma określonej grupy (tzw. sondy) określa się natężenie pola elektrycznego w badanym środowisku. Popularnymi sondami są grupy nitrylowe, gdyż mają silne pasmo absorpcyjne w regionie, w którym białka absorbują stosunkowo niewiele promieniowania. W cytowanej pracy badacze wykorzystali grupę karbonylową inhibitora, 19-nortestosteronu, co wymagało zastosowania bardziej złożonej procedury eksperymentalnej.

<sup>6</sup> Od nazwy kodującego ją genu z bakterii *Xanthobacter autotrophicus*.

<sup>7</sup> Względem bariery analogicznej reakcji w środowisku wodnym.

<sup>8</sup> W greckiej mitologii Kirke była wiedźmą, która zamieniła towarzyszy Odyseusza w świnie.

(względem referencyjnej reakcji w roztworze); w oparciu o to zaproponował hipotezę, że enzym wykorzystuje energię wiązania by zniwelować entropowy wkład do bariery reakcji [48]. Miało by to zachodzić poprzez silne wiązanie substratu w jednym rejonie, co umożliwiłoby z kolei wprowadzenie większych naprężeń w innym, tym samym wymuszając zajście reakcji chemicznej (efekt zwany czasem destabilizacją substratu, ang. *Ground State Destabilization*, GSD). Efekt ten wydawał się być szczególnie wyraźny w przypadku dekarboksylazy orotydylanowej i deaminazy cytydylowej. Szczegółowa analiza z wykorzystaniem symulacji komputerowych nie potwierdziła jednak hipotezy o „zwiększonych naprężeniach” w pierwszym z wymienionych enzymów [49]; natomiast w drugim przypadku zmianę entropii aktywacji przypisać należy raczej zmianie mechanizmu reakcji z jedno- na kilkietapowy [50].

Przedstawione przykłady pokazują, że chociaż efekty entropowe są istotnym elementem katalizy enzymatycznej (odpowiedzialne są np. za zdolność katalityczną enzymów mikroorganizmów psychrofilnych [51]), w wielu przypadkach dają się one wyrazić poprzez pojęcie elektrostatycznej preorganizacji. Jest to istotna konkluzja istotna z punktu widzenia dalszego rozwoju metod projektowania enzymów, jak zostanie to pokazane w następnych sekcjach.

### 1.2.4 Krytyczna ocena

Można argumentować, że przedstawiony schemat podziału i systematyzacji efektów katalitycznych (oparty w znacznym stopniu na pracach zespołu Warshela) jest do pewnego stopnia płynny i arbitralny; ma on jednak, pośród innych koncepcji, tę zaletę, że jest jednoznacznie i systematycznie zdefiniowany w ramach obowiązujących teorii fizycznych, co czyni z niego solidne narzędzie w analizie złożonych układów molekularnych. Z drugiej jednak strony, na tym rozumowaniu spoczywa cień uproszczeń, przyjętych, by możliwe były symulacje konieczne w celu poznania choćby wycinka przestrzeni fazowej skomplikowanego układu molekularnego, a tym samym ujęcie w ramach modelu energii swobodnej związanej np. z reorganizacją rozpuszczalnika. Do tych założeń należy przede wszystkim rozpatrywanie oddziaływań elektrostatycznych w modelu ładunków punktowych, co może prowadzić do znacznych błędów w związku z brakiem opisu anizotropowości rozkładu ładunku wokół atomów<sup>9</sup> [52]; przyjęcie reakcji w roztworze wodnym za układ referencyjny jest naturalne z chemicznego punktu widzenia, lecz ze względu na ruchliwość cząsteczek rozpuszczalnika (z czym wiąże się wspomniana energia reorganizacji) niemożliwe obecnie do zrealizowania inaczej, niż we wspomniany przed chwilą sposób. Problematyczne jest również to, że, jak pokazano, mechanizmy mogą być różne w roztworze wodnym i środowisku enzymatycznym. Na drugim biegunie są prace naszego zespołu [37, 38], które za stan referencyjny przyjmują fazę gazową lub, jak w ostatniej publikacji [39], dziką formę enzymu. Pozwala to wyeliminować te problemy (*vide* Rozdział 2), zaś wszystkie analizy są rygorystycznie wyprowadzone z pierwszych zasad (*ab initio*), ze szczególnym naciskiem położonym na niezależność wyniku od przyjętej bazy funkcyjnej i możliwość systematycznego wprowadzania dalszych przybliżeń. Należy podkreślić, że również tutaj eksponowane jest znaczenie oddziaływań elektrostatycznych, lecz w najprostszym ujęciu za pomocą rozwinięcia multipolowego (opisanego w Rozdziale 2) uwzględniającego anizotropię ładunku. Jak to zostanie wspomniane na końcu tego rozdziału, w pełni solidna pod względem matematycznym i fizycznym metoda powinna łączyć w sobie najlepsze cechy obu tych podejść.

W dziedzinie badań nad katalizą enzymatyczną pozostało jeszcze wiele do zrobienia. Mimo to trzeba przyznać, że wiele na tym polu już osiągnięto w ciągu ostatnich kilku dekad. Jak zatem przekłada się to na możliwości projektowania enzymów?

## 1.3 Historia projektowania *de novo* enzymów

### 1.3.1 Początki

Projektowanie enzymów jest blisko związane z racjonalnym projektowaniem białek jako takich. Jest to logiczną konsekwencją prostego faktu, że uzyskanie struktury o własnościach katalitycznych wymaga umiejętności uzyskania jakiegokolwiek zadanej z góry struktury łańcucha białkowego. Absolutnym pio-

<sup>9</sup> W niniejszej pracy *anizotropowość rozkładu ładunku wokół atomów* będzie często w skrócie określana po prostu jako *anizotropowość rozkładu ładunku*.

nierem w tej dziedzinie jest Bernd Gutte, który wraz z Robertem Bruce Merrifieldem<sup>10</sup> jako pierwszy zsyntetyzował *de novo* aktywny katalitycznie enzym (rybonukleazę A krowiej trzustki) [53, 54]<sup>11</sup>. W roku 1975 zaprojektował i zsyntetyzował aktywny katalitycznie wariant rybonukleazy liczący zaledwie 70 reszt (połowę długości naturalnego białka) [55]; w 1979 – peptyd wiążący specyficznie trinukleotyd GAA [56]; w 1983 – peptyd wiążący insektycyd DDT [57]. Te imponujące osiągnięcia nie przyczyniły się jednak do większego rozwoju metodologii projektowania białek; stały się za to inspiracją do dalszych badań nad związaniem łańcuchów polipeptydowych i ich samoorganizacją, co wraz ze wzrastającą liczbą danych strukturalnych dostarczyło pewnych „schematów i zasad konstrukcyjnych” o szerokich możliwościach [58–61].<sup>12</sup> Za przykład niech posłuży popularny motyw wiązki czterech helis (ang. *four-helix bundle*), po raz pierwszy zsyntetyzowany przez zespół Wiliama DeGrado w roku 1986 [63]. Ten eleganci i dobrze poznany motyw, wzorowany na naturalnie występujących strukturach, wykorzystywany jest obecnie np. w projektowaniu peptydów wiążących jony metali [64].

Wraz z rozwojem technik inżynierii genetycznej i biologii molekularnej, w tym kierunkowanej mutagenyzy [17], stało się możliwe badanie i modyfikowanie większych białek [65, 66]. Równocześnie metody komputerowe zaczęły nabierać coraz większego znaczenia; wraz ze wzrostem mocy obliczeniowej komputerów i postępów w rozwoju metod i oprogramowania, perspektywa projektowania kompletnych białek *in silico* stawała się coraz bliższa. W roku 1991 Homme Hellinga opublikował pierwszy program wspomagający projektowanie mutacji w białkach – *Dezzymer* [67]; kolejnym programem tego rodzaju był *Orbit*, z pomocą którego zespół Stephena Mayo w 1997 roku z powodzeniem zaprojektował kilka polipeptydów; rok później grupa Meneza po raz pierwszy zmieniła charakter enzymu, projektując endopeptydazę specyficzną względem reszt proliny (*cyproazę 1*) w oparciu o sekwencję cyklofiliny A [68]. Na początku XXI wieku zaczęto syntezować pierwsze „enzymopodobne” polipeptydy projektowane komputerowo za pomocą *Orbita* (tzw. *protozomy*) [69]. Zasadniczą wadą tych projektów była niska aktywność katalityczna w porównaniu z naturalnymi enzymami. W 2004 roku Hellinga wraz z Mary Dwyer opublikowali w *Nature* pracę, w której ogłaszali wielki sukces: izomerazę triozofosforanową (TIM) zaprojektowaną w oparciu o białko *E. Coli* wiążące rybozę, o aktywności przewyższającej dotychczas projektowane enzymy. Jednak po kilku miesiącach okazało się, że domniemany wynik był rezultatem zanieczyszczenia naturalną TIM i badacze musieli wycofać publikację [70].

Rzeczywistym sukcesem w dziedzinie komputerowego projektowania nowych enzymów (*teozymów*<sup>13</sup>, jak nazywane będą w niniejszej pracy) były dopiero retroaldolazy zaprojektowane w roku 2008 przez zespół Davida Bakera, twórców programu *Rosetta*<sup>14</sup> [71]. Jak na ironię, za szkielet dla tych enzymów posłużyła naturalna TIM; na tym samym wzorcu oparto projekt teozymu katalizującego reakcję eliminacji Kempa, która nie jest wykorzystywana w świecie ożywionym [72]. W kolejnych latach powstała w tym samym laboratorium pierwsza sztuczna diels-alderaza [73]<sup>15</sup>, aktywnością katalityczną przewyższającą odpowiednie przeciwiactła katalityczne, oraz fosfoesteraza [75]. Jednakże początkowy, pełen sukcesów postęp uległ stagnacji, co z kolei wzbudziło dyskusję na temat braków stosowanej metodologii [76–78].

<sup>10</sup> Twórca metody syntezy peptydów na podłożu stałym; uzyskał za to nagrodę Nobla z dziedziny chemii w roku 1984.

<sup>11</sup> Sam eksperyment miał miejsce w roku 1968; list do redakcji *Journal of American Chemical Society* został opublikowany w styczniu 1969 roku [53]. Sztuczna synteza pierwszego enzymu dokonana przez człowieka nastąpiła zatem wcześniej, niż postawił on nogę na Księżycu!

<sup>12</sup> Cytując powyższe przeglądy nie sposób nie wspomnieć pokrótce o autorce jednego z nich, Jane Richardson. Interesujący jest początek jej kariery naukowej; w roku 1958, mając 17 lat, na podstawie dwóch nocnych obserwacji obliczyła orbitę Sputnika, dzięki czemu zajęła trzecie miejsce w konkursie naukowym *Westinghouse Science Talent Search* [62]; swoją karierę poświęciła jednak badaniom krystalograficznym nad strukturą białek. Najbardziej znana jest z zaproponowania sposobu ich reprezentacji, tzw. modeli wstęgowych (ang. *ribbon diagrams*)

<sup>13</sup> W literaturze anglosaskiej przez termin *teozym* (od ang. *theozyme* – *theoretical enzyme*) rozumie się zarówno końcowy efekt teoretycznego (*in silico*) projektowania, jak i definicję podstawowych elementów centrum aktywnego, wspomnianą w sekcji 1.3.2; w celu rozróżnienia między tymi pojęciami, zdecydowano się tutaj na wykorzystanie terminu *teozym* w odniesieniu do teoretycznie zaprojektowanego enzymu, natomiast w znaczeniu definicji najważniejszych reszt katalitycznych – *rdzeń teozymu*.

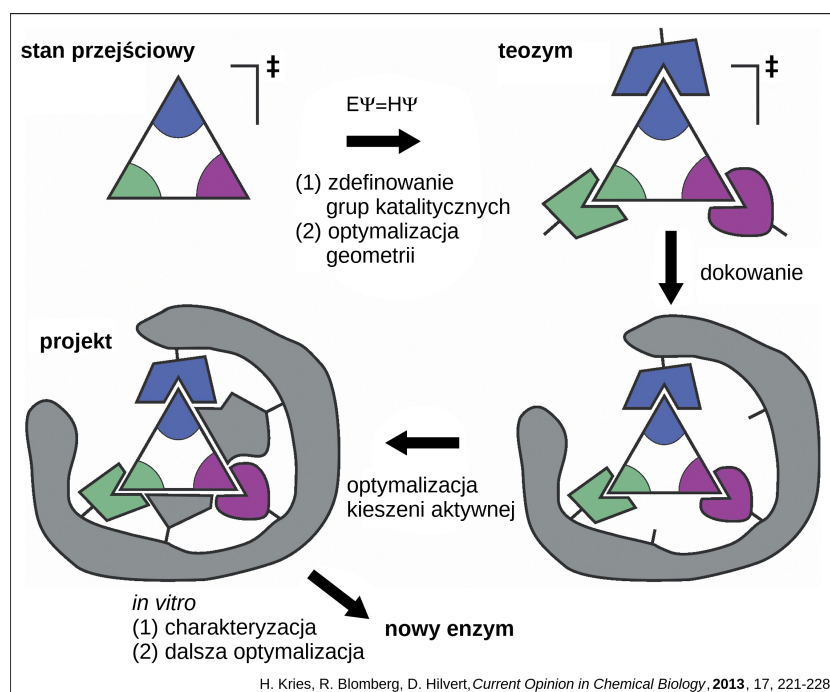
<sup>14</sup> Kamień z Rosetty, na którym ten sam tekst był napisany egipskimi hieroglifami i w języku greckim, pozwolił w XIX w. na rozszyfrowanie pisma starożytnych Egipcjan.

<sup>15</sup> Uważano to za kolejny przykład teozymu katalizującego reakcję niezachodzącą naturalnie w organizmach żywych, lecz niedawno okazało się to nieprawdą, gdyż odkryto w przyrodzie takie enzymy [74].



### 1.3.2 Algorytm projektowania enzymów

Kluczowym elementem algorytmów teoretycznego projektowania enzymów jest definicja „rdzenia” teozymu, będąca bezpośrednim zastosowaniem idei stabilizacji stanu przejściowego (Rysunek 1.6). Obejmuje on TS danej reakcji chemicznej wraz ze zbiorem reszt aminokwasowych zorientowanych tak, by w specyficzny sposób z nim oddziaływać. Mając definicję rdzenia teozymu, kolejnym krokiem jest znalezienie w bazie danych szkieletu białkowego, w którym da się go zrealizować [79,80]<sup>16</sup>. Następnie w kilku cyklach optymalizuje się oddziaływania z sąsiednimi resztami (ang. *packing*, „upakowanie”); fазie tej często towarzyszą symulacje Monte Carlo, mające na celu określenie stabilności struktury i ominięcie problemu lokalnego minimum [80,82]. W końcowej fazie warianty projektowe są szeregowane za pomocą zadanej funkcji oceny (ang. *scoring function*), a najlepsze z nich poddawane są weryfikacji eksperymentalnej. Jak dotąd, uzyskane w ten sposób teozymy przejawiały stosunkowo niewielką aktywność; dalsza optymalizacja eksperymentalna za pomocą metod kierunkowanej ewolucji (ang. *Laboratory Directed Evolution*, LDE) [83] pozwoliła uzyskać znacznie bardziej aktywne warianty [84].



Rysunek 1.6: Schemat algorytmu projektowania nowych enzymów [79] (tłumaczenie i przedruk za zgodą wydawnictwa Elsevier).

### 1.3.3 Eliminaza Kempa

Jednym z najlepiej przebadanych teozymów jest eliminaza Kempa (schemat zachodzącej w niej reakcji przedstawiono na Rysunku 1.2). Jak już wspomniano, najaktywniejszy z pierwszych teozymów KE07 [72] został znacznie ulepszony w eksperymentach kierunkowanej ewolucji: najlepszy enzym z siódmej rundy eksperymentu charakteryzował się stałą  $k_{cat}$  ok. tysiąc razy większą od pierwotnego enzymu (choć wiązało się to ze zmniejszeniem stabilności białka) [85]. W 2012 roku, opierając się szkielecie ksylanazy, grupa Mayo zaprojektowała eliminazy HG1, HG2 i HG3 [86]; osiągnięto to poprzez iteracyjną weryfikację eksperymentalną i symulacje dynamiki molekularnej białka (ang. *Molecular Dynamics*, MD). W toku kierunkowanej ewolucji udało się uzyskać wariant HG3.17, którego aktywność, największa z osiągniętych do tej pory, dorównuje naturalnym enzymom [11].

Obie z rodzin eliminaz były przedmiotem intensywnych badań. W roku 2008 William Jorgensen we współpracy z Davidem Bakerem, Anastazją Alexandrovą i Danielą Röthlisberger analizowali me-

<sup>16</sup> Tak zaprojektowane są programy *Rosetta3* i *Orbit*; inny program, *Saber* [81], służy identyfikacji białek już posiadających odpowiednio zorientowane reszty z definicji rdzenia teozymu.

chanizm reakcji w czterech zaprojektowanych eliminacjach Kempa wykorzystując metodę QM/MM<sup>17</sup> angażującą półempiryczną metodę PM3 i symulacje Monte Carlo [88]; w 2011, Anastazja Alexandrova i Manuel Sparta w podobny sposób analizowali dwa kolejne teozymy z tej rodziny [89]. W obu przypadkach wyniki były jakościowo zgodne z eksperymentalnymi, ale nie znaleziono znaczącej korelacji pomiędzy teoretycznymi i eksperymentalnymi barierami aktywacji. W pracy Kendalla Houka i współpracowników z 2010 roku wykorzystano metodę ONIOM<sup>18</sup> [90] i popularny funkcjonal DFT B3LYP, znajdując tylko jakościową zgodność z eksperymentem, lecz wystarczającą, by odróżnić aktywne i nieaktywne warianty projektowe [91].

Zespół Warshela, analizując kilka eliminaz z serii za pomocą empirycznego modelu wiązań walencyjnych (ang. *Empirical Valence Bond*, EVB), zwrócił uwagę na trudności w uzyskaniu preorganizowanego elektrostatycznie środowiska dla tej reakcji ze względu na małą zmianę rozkładu ładunku pomiędzy substratem a stanem przejściowym [92–94]. W innej pracy Monika Fuxreiter wraz ze współpracownikami wskazała na optymalizację energii reorganizacji podczas ewolucji KE07 [95]. W 2015 roku Katarzyna Świderek stwierdziła w HG3.17 obecność dwóch konformerów o różnych aktywnościach oraz podkreśliła znaczenie oddziaływań elektrostatycznych z drugiej sfery koordynacyjnej [96], zaś Sylvia Osuna pokazała, że centra aktywne enzymów uzyskanych drogą LDE są znacznie mniej labilne od oryginalnych teozymów [97]. Z kolei zespół Teresy Head-Gordon, powiązał wpływ mutacji ze zmianą entropii łańcuchów bocznych aminokwasów i informacji związanej ze sprzężeniami pomiędzy nimi [98].

Wymienione wyżej obserwacje wpisują się w stwierdzenie samego Bakera z 2010 roku, w którym dopatrywał się źródeł niepowodzeń teozymów w dynamice białka i braku opisu dalekozasięgowych oddziaływań elektrostatycznych [99]. Inną przyczyną wspomnianą w literaturze jest skupienie całej procedury na oddziaływaniach tylko ze stanem przejściowym, a nie z substratem [84]. Warto przy tym wspomnieć obserwację Warshela z jego najnowszej pracy, w której porównuje przebieg kierunkowanej ewolucji KE07 i HG3 [100]: głównym czynnikiem związanym z obniżeniem bariery aktywacji w pierwszym przypadku jest destabilizacja substratu, natomiast w drugim — stabilizacja stanu przejściowego. Oznacza to, że bardziej ogólny algorytm projektowania enzymów powinien uwzględniać i jedno, i drugie; jest to możliwe przy wykorzystaniu technik związanych z różnicową stabilizacją stanu przejściowego (ang. *Differential Transition State Stabilisation*, DTSS [37,101]), jak zostanie pokazane w Rozdziale 2.

## 1.4 Cel pracy

Jak wynika z zaprezentowanego opisu, w dotychczasowych poszukiwaniach dróg ulepszenia technik projektowania enzymów (jednego ze "Świętych Graali" chemii obliczeniowej w XXI wieku [102]) trzy główne czynniki wysuwają się na pierwszy plan:

- uwzględnienie dalekozasięgowych oddziaływań elektrostatycznych,
- równoczesne uwzględnienie oddziaływań zarówno z substratem, jak i ze stanem przejściowym,
- uwzględnienie wpływu dynamiki środowiska białkowego na aktywność katalityczną.

Równocześnie, każde podejście do tego problemu musi sprostać podstawowym wyzwaniom, jakimi są rozmiar układu i eksponencjalny wzrost liczby możliwych wariantów wraz z liczbą mutowanych miejsc w sekwencji aminokwasowej; oba te czynniki zwiększają koszt obliczeniowy związany z wykonywaniem potrzebnych obliczeń i symulacji. Konieczne jest szukanie kompromisu pomiędzy rozmiarem przestrzeni fazowej jaką ma objąć model, a dokładnością i poprawnością fizyczną zastosowanego hamiltonianu. W jednym ekstremum znajdują się długie symulacje dające wgląd w zmiany energii swobodnej układu, ale oparte na empirycznych parametrach pól siłowych, w drugim — obliczenia klasterowe układów enzymatycznych metodami kwantowo-mechanicznymi [103]; pośrodku znajdują się metody QM/MM, angażujące bądź to zaawansowane metody kwantowe, lecz w niewielkim stopniu próbujące przestrzeń fazową, bądź też metody półempiryczne (np. PM6, PM7 [104]), mogące być źródłem znacznych błędów (np. wskazywać w miejscu TS minimum energetyczne), lecz pozwalające na większą eksplorację różnych

<sup>17</sup> Od ang. *Quantum Mechanics/Molecular Mechanics*; są to metody hybrydowe, w których część układu (np. centrum reakcji) opisywane jest metodami chemii kwantowej (QM), zaś otoczenie — mechaniką molekularną (MM). Przykłady tego rodzaju metod oraz ich zastosowań znaleźć można w książce [87]

<sup>18</sup> Od ang. *our own N-layered integrated molecular orbital and molecular mechanics*; metoda QM/MM opracowana przez zespół Keiji Morokumy, zaimplementowana np. w programie *Gaussian*.

konfiguracji<sup>19</sup>. To zestawienie można przedstawić także jako podejścia *odgórne* (ang. *top-down*, „z góry do dołu”), w opozycji do *oddolnych* (ang. *bottom-up*, „z dołu do góry”). Zasadnicza różnica pomiędzy nimi polega na sposobie konstrukcji modelu i wprowadzania kolejnych przybliżeń; w pierwszym przypadku złożony układ systematycznie rozbiórany jest na części składowe, co wiąże się z wprowadzaniem kolejnych założeń (dotyczących stanów protonacji reszt aminokwasowych, wartości stałej dielektrycznej w modelu rozpuszczalnika ciągłego, podziału układu na region QM i MM itd.); w drugim natomiast punkt wyjścia stanowią fundamentalne zasady fizyczne, a każdy następny, rozbudowujący model krok jest poddawany możliwie szczegółowej analizie. Do tej drugiej kategorii zaliczyć można koncepcję pól katalitycznych (bazującą na elektrostatycznym przybliżeniu DTSS) [37, 101], opisaną w Rozdziale 2.

Ominięcie pułapek parametryzacji metod *odgórnych* i wstępne wyselekcjonowanie zestawu potencjalnie korzystnych mutacji, a także ich późniejsza weryfikacja (z uwzględnieniem wymienionych trzech czynników), możliwe jest w obrębie bazujących na obliczeniach *ab initio* strategii *oddolnych*, na co wskazują dotychczasowe prace naszego zespołu [37, 38, 106].

### Cel pracy

*Opracowanie sposobu zastosowania techniki pól katalitycznych do określania wpływu mutacji na aktywność katalityczną.*

Kluczowym elementem jest tutaj opis gęstości elektronowej cząsteczek; w tej pracy wiodącą metodą będzie atomowe rozwinięcie multipolowe [107], pozwalające niedużym kosztem obliczeniowym precyzyjnie, na podstawie pierwszych zasad, uwzględnić anizotropowość rozkładu ładunku wokół atomów [52], a dzięki temu uniknąć możliwych błędów związanych z tradycyjnie stosowanymi ładunkami atomowymi. Najważniejszym zadaniem jest pierwsze w dotychczas prowadzonych badaniach przetestowanie proponowanego podejścia na szeregu mutantów tego samego enzymu, aby wykazać jego przydatność w projektowaniu teozymów; kolejnym zadaniem jest wyjaśnienie choćby części mutacji wprowadzonych do białka drogą eksperymentów kierunkowanej ewolucji (LDE), co rodziłoby nadzieję na rozwinięcie metody obliczeniowej pozwalającej zastąpić część fazy eksperymentalnej.

Powyższe cele można przedstawić za pomocą następującego harmonogramu czy też listy zagadnień.

- **Reprezentacja oddziaływań elektrostatycznych.** W sekcji 3.1 podjęto próbę oceny wpływu wyboru modelu (ładunki *vs* multipole atomowe) na wyznaczone wartości potencjału elektrostatycznego; sekcja 3.1.1 skupia się na jakości opisu reorganizacji ładunku podczas reakcji chemicznej, natomiast w sekcji 3.1.2 badane jest oddziaływanie przez nie elektrostatycznych energii DTSS na przykładzie kinazy białkowej A. Przeprowadzone testy mają za zadanie określić jakość poszczególnych przybliżeń.
- **Ocena wpływu mutacji na aktywność katalityczną.** W sekcji 2.1 pokazano, że o ile energia DTSS dana przez różnicę energii oddziaływania odnosi się do wirtualnej *fazy gazowej*, to różnice energii DTSS pomiędzy mutantami a dziką formą enzymu można traktować jako zmiany bariery aktywacji wskutek mutacji. W sekcji 3.2.1 jest to eksplorowane na przykładzie mutantów izomerazy ketosteroidowej; badana jest tam również fizyczna natura oddziaływań w tym szeregu enzymów, a dotychczasowy model DTSS rozszerzany jest do postaci obejmującej reakcje wieloetapowe. Jest to pierwsza w historii dotychczasowych badań szczegółowa analiza kilku mutantów za pomocą technik rozwijanych w naszej grupie.
- **Analiza mutacji (oraz wpływu konformacji łańcuchów bocznych naładowanych aminokwasów na aktywność katalityczną) z wykorzystaniem biblioteki rotamerów.** Na przykładzie otrzymanych drogą LDE mutantów teozymu KE07 (sekcja 3.2.2), podjęta została próba wyjaśnienia mutacji z udziałem naładowanych reszt aminokwasowych oraz efektów nieaddytywnych związanych z wielokrotnymi mutacjami. Testowany na tej rodzinie *oddolny* (ang. *bottom-up*) model obliczeniowy może mieć potencjalne zastosowanie w projektowaniu enzymów, być może zastępując część eksperymentalnej fazy poszukiwań nowych mutantów, tym samym przyspieszając cały proces i obniżając jego koszty.

<sup>19</sup> Istnieją jeszcze specjalne pola siłowe projektowane do opisu reakcji chemicznych [105], ale nie zostały one ujęte w tym zestawieniu.



# Rozdział 2

## Metody

### 2.1 Różnicowa stabilizacja stanu przejściowego

Podstawą metodologii stosowanej w niniejszej pracy jest koncepcja różnicowej stabilizacji stanu przejściowego (DTSS) [101], którą można postrzegać jako rozszerzenie koncepcji stabilizacji stanu przejściowego (ang. *Transition State Stabilisation*, TSS) [8]. Centralnym jej punktem jest porównanie profilu reakcji w katalizatorze (np. enzymie) z tzw. *własnym profilem energetycznym reagentów* (ang. *intrinsic*); jest on tożsamy profilowi reakcji w fazie gazowej, jeżeli obecność katalizatora nie zmienia w sposób istotny mechanizmu reakcji (stąd, we wcześniejszych pracach, nazywany był *fazą gazową*). Z technicznego punktu widzenia oznacza to, że *energia własna* stanowi energię wyznaczoną dla samych tylko reagentów w geometrii takiej, jaką przyjęto przy obliczeniach w obecności katalizatora<sup>1</sup>; można to wykonać na dwa sposoby: albo przyjęc geometrię reagentów w katalizatorze (jeśli jest znana, np. z obliczeń QM/MM lub w oparciu o techniki eksperymentalne), albo pochodzącą z obliczeń w fazie gazowej (jest pierwsze przybliżenie zaniedbujące wpływ katalizatora na ścieżkę reakcji). Dla zachowania ciągłości z poprzednimi pracami, określenie *faza gazowa*<sup>2</sup> stosować się będzie w znaczeniu energii *własnej* zdefiniowanym powyżej, o ile nie zostanie nadmienione inaczej.

Przyjmując opisany układ odniesienia, różnica pomiędzy barierą w enzymie  $B_{WT}$  a *własną* barierą reakcji  $B_0$  sprowadza się do różnicy energii oddziaływania enzymu ze stanem przejściowym  $E_{IE}(TS)$  i substratami  $E_{IE}(S)$  (tzw. energii DTSS) oraz zmiany energii wewnętrznej enzymu  $\Delta E_{WT}$  (równanie 2.1)<sup>3</sup>. Jeżeli zmiana energii katalizatora pomiędzy substratem a stanem przejściowym jest zaniedbywalna w porównaniu z energią oddziaływań (środowisko *preorganizowane*<sup>4</sup>), sprowadza się to do energii DTSS  $E_{DTSS}$ ; schematycznie jest to przedstawione na Rysunku 2.1.

$$B_{WT} - B_0 = \underbrace{E_{IE}(TS) - E_{IE}(S)}_{E_{DTSS}} + \underbrace{E_{WT}(TS) - E_{WT}(S)}_{\Delta E_{WT} \approx 0} \quad (2.1)$$

$$B_{mut} - B_{WT} \approx E_{DTSS}(mut) - E_{DTSS}(WT) \quad (2.2)$$

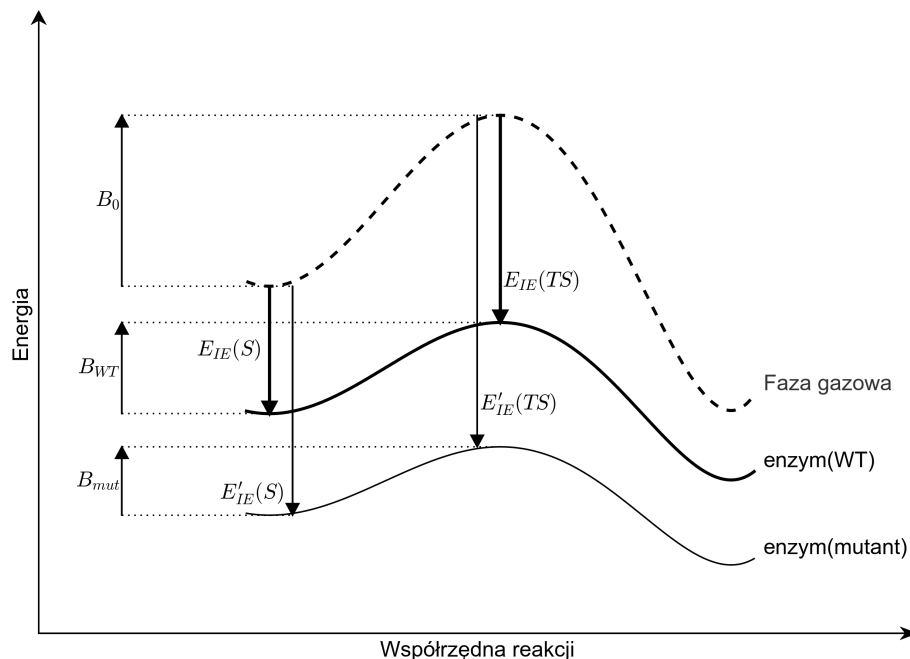
O ile wykorzystanie fazy gazowej jako stanu referencyjnego dla mierzenia efektów katalitycznych można kwestionować (powszechnie uznanym standardem jest tutaj reakcja w roztworze [108]), o tyle stanowi to podstawę prostych i wydajnych obliczeniowo metod analizy aktywności katalitycznej [39]. *Energia własna* może być w świetle Rysunku 2.1 postrzegana jako pewien w przybliżeniu stały wkład zależny od reakcji chemicznej, niezmienny pod wpływem mutacji (o ile nie wpływa ona na mechanizm

<sup>1</sup> Uwzględnione są więc tylko efekty elektronowe.

<sup>2</sup> W „żargonie” specyficznym dla tej dziedziny, nazywa się to czasem obliczeniami *in vacuo*, ze względu na to, że dookoła układu, dla którego obliczana jest funkcja falowa, nie ma *nic*. Określenie to nie jest jednak poprawnym terminem fizycznym, raczej ilustracją sposobu wykonywania rachunków.

<sup>3</sup> Wielkość odpowiadająca energii reorganizacji, lecz nie nazwana w ten sposób z powodu odmiennej definicji niż pracach Warshela.

<sup>4</sup> Oznaczać to może statyczność środowiska (brak wówczas efektów związanych z jego ruchami), ale także sytuację, w której koszt energetyczny związany z odkształceniem katalizatora pomiędzy substratem a stanem przejściowym jest zaniedbywalny w stosunku do pozostałych członów równania (2.2)



Rysunek 2.1: Schematyczne przedstawienie różnicowej stabilizacji Stanu Przejściowego (DTSS).

reakcji)<sup>5</sup>; wówczas różnica barier aktywacji sprowadza się do różnic energii DTSS (2.2). W ten sposób wirtualny stan fazy gazowej zostaje wyeliminowany z równania, a model poprzez równanie (2.2) można uważać za wyrażający zmiany  $B_{mut}$  względem  $B_{WT}$ , bez konieczności rozpatrywania reakcji w rozpuszczalniku lub fazie gazowej. Przykładowo, wyznaczenie wpływu mutacji aminokwasu  $X$  na  $Y$  wymaga obliczenia czterech energii oddziaływania:  $X \cdot TS$  i  $X \cdot S$  oraz  $Y \cdot TS$  i  $Y \cdot S$ , a różnica barier aktywacji  $B_{mut}(X \rightarrow Y) - B_{WT}$  wyrażona jest równaniem (2.3).

$$\begin{aligned}
 B_{mut}(X \rightarrow Y) - B_{WT} &= E_{DTSS}(Y) - E_{DTSS}(X) \\
 &= E_{IE}(Y \cdot TS) - E_{IE}(Y \cdot S) \\
 &\quad - (E_{IE}(X \cdot TS) - E_{IE}(X \cdot S))
 \end{aligned} \tag{2.3}$$

### 2.1.1 Reakcje wieloetapowe

Wspominane w tej pracy teozymy katalizują reakcje jednoetapowe, których stałą szybkości determinuje tylko jeden stan przejściowy. Jednakże przeważająca część reakcji chemicznych zachodzących w przyrodzie jest kilkuetapowa, co wymaga bardziej złożonego podejścia. W przypadku, gdy energia aktywacji któregoś ze stanów przejściowych jest dominująca (a odpowiadający jej etap stanowi wąskie gardło procesu), można pozostałe bariery zaniedbać i potraktować reakcję tak, jak jednoetapową. Pewien problem mogą stanowić przypadki, w których takie sytuacja nie zachodzi; wówczas jednym z możliwych rozwiązań jest rozważanie energii DTSS dla każdego etapu z osobna; innym podejściem do problemu jest skupienie się na intermedacie zamiast na TS. Istnieje szereg argumentów przemawiających za tym, że właśnie stabilizacja intermediatu reakcji jest jedną z kluczowych „strategii” katalizy enzymatycznej [109,110]. Aby lepiej to unaocznić, rozważmy fakt, że rozkład ładunku każdego ze stanów przejściowych można uważać za w pewnym stopniu „pośredni” pomiędzy sąsiednimi punktami stacjonarnymi na ścieżce reakcji; wówczas stabilizacja intermediatu wiąże się z częściową stabilizacją

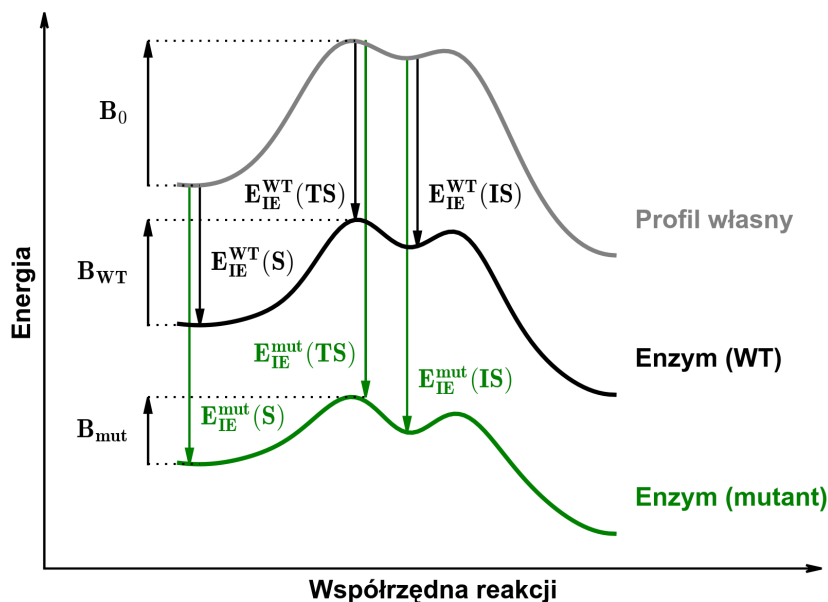
<sup>5</sup> W przypadku zaprojektowanego teoretycznie teozymu KE07 (vide sekcja 3.2.2), gdzie analizowane mutacje znajdowały się w drugiej sferze koordynacyjnej, założenie to jest spełnione

obu sąsiadujących z nim stanów przejściowych; można także zaobserwować, że rozkład ładunku intermediatu jest, w porównaniu z TS, bardziej zlokalizowany, co ułatwia znalezienie specyficznych dla niego oddziaływań.

W kontekście metody DTSS, oznacza to wprowadzenie różnicowej stabilizacji intermediatu (ang. *Differential Intermediate State Stabilisation*, DISS; *vide* równanie 2.4), co schematycznie pokazane jest na Rysunku 2.2. Podejście to zostało wykorzystane do porównywania mutantów izomerazy ketosteroidowej (w myśl równania 2.5, analogicznego do równania 2.2) w ostatniej pracy naszego zespołu [39] (*vide* sekcja 3.2.1).

$$E_{DISS}^{WT} = E_{IE}^{WT}(IS) - E_{IE}^{WT}(S) \quad (2.4)$$

$$B_{mut} - B_{WT} \sim E_{DISS}^{mut} - E_{DISS}^{WT} \quad (2.5)$$



Rysunek 2.2: Różnicowa stabilizacja intermediatu (DISS).

## 2.2 Oddziaływania międzycząsteczkowe

Wyznaczenie energii DTSS lub DISS wymaga obliczenia energii oddziaływań reagentów z otoczeniem; w tym celu można bezpośrednio zastosować definicję energii oddziaływania pomiędzy dwoma układami  $A$  i  $B$  jako różnicy pomiędzy energią kompleksu  $AB$  a energią izolowanych jego składowych (2.6); podejście to (zwane *supermolekularnym*) może być jednak pracochłonne i mało dokładne, gdyż energie tworzenia molekul<sup>6</sup> są rzędu dziesiątek i więcej jednostek atomowych (*hartree*, Ha)<sup>7,8</sup>, zaś

<sup>6</sup> Energie będące wynikiem obliczeń kwantowo-chemicznych są najczęściej podawane jako suma energii odpychania pomiędzy jądrami atomowymi układu  $V_{NN}$  oraz energii elektronowej, która z kolei jest sumą energii kinetycznych i potencjalnych każdego elektronu; można tę wielkość interpretować zatem jako energię tworzenia molekuły z izolowanych elektronów i jąder atomowych. Z tego właśnie powodu liczba ta zawsze jest ujemna — w przeciwnym wypadku cząsteczka nie byłaby stabilna.

<sup>7</sup> Od nazwiska Douglasa Raynera Hartree, znanego m.in. jako współtwórca metody Hartree-Focka.

<sup>8</sup> 1 hartree = 627.5 kcal/mol.

szukana różnica to zaledwie tysięczne lub setne części tej wielkości<sup>9</sup>. Jest to jednym z powodów, dla których warto poszukać innych sposobów obliczania energii oddziaływania.

$$E_{IE}(A, B) = E_{AB} - E_A - E_B \quad (2.6)$$

### 2.2.1 Przybliżenie polaryzacyjne

Inne wyrażenie na energię oddziaływania można uzyskać, jeżeli potraktować oddziaływanie między cząsteczkami jako małe *zaburzenie* w stosunku do sił utrzymujących każdą z nich w całości. Metody korzystające z tego przybliżenia bazują na rachunku zaburzeń Raighleya-Schrödingera; szczegółowe informacje na temat tych technik można znaleźć w [111, Rozdział 13] oraz [112]. Ze względu jednak na interdyscyplinarny charakter pracy, a tym samym potencjalnie zróżnicowane grono Czytelników, w bieżącej sekcji umieszczony zostaje podstawowy opis rachunku perturbacyjnego w zastosowaniu do energii oddziaływań.

W przybliżeniu polaryzacyjnym, jako niezaburzony hamiltonian traktujemy sumę hamiltonianów izolowanych molekuł (2.7), którego funkcją własną jest iloczyn funkcji falowych molekuł (2.8); rolę zaburzenia pełni operator oddziaływania między cząsteczkami, składający się z odpychania elektron-elektron, jądro-jądro oraz przyciągania jądra-elektron (2.9). W wymienianych równaniach indeksy  $i$  i  $j$  oznaczają elektrony odpowiednio molekuł A i B, zaś  $a$  i  $b$  przypisane są jądrům atomowym odpowiednich cząsteczek. Zgodnie z metodą Raighleya-Schrödingera, poprawki do energii, stanowiące w tym wypadku energię oddziaływania, wyznaczane są w postaci szeregu (2.10), przy czym najczęściej stosowane są jego pierwsze dwa wyrazy.

$$\hat{H}^{(0)} = \hat{H}_A + \hat{H}_B \quad (2.7)$$

$$\psi_0^{(0)} = \psi_{A,0}\psi_{B,0} \equiv \phi_0 \quad (2.8)$$

$$\hat{V} = \sum_{i \in A} \sum_{j \in B} \frac{1}{r_{ij}} + \sum_{a \in A} \sum_{b \in B} \frac{Z_a Z_b}{r_{ab}} - \sum_{i \in A} \sum_{b \in B} \frac{Z_b}{r_{ib}} - \sum_{a \in A} \sum_{j \in B} \frac{Z_a}{r_{aj}} \quad (2.9)$$

$$E_{int} = \sum_{i=1}^{\infty} E^{(i)} \approx E^{(1)} + E^{(2)} \quad (2.10)$$

Zwyczajowo nazywane są one stosownie do ich znaczenia fizycznego; pierwsza poprawka (2.11), nazywana energią elektrostatyczną, równa jest kulombowskiemu oddziaływaniu pomiędzy niezaburzonymi rozkładami ładunków molekuł A i B; w drugim przyczynku (2.12) wyróżnia się człon indukcyjny (który dalej można przedstawić jako sumę wkładów opisanych równaniami 2.13 i 2.14) oraz człon dyspersyjny (2.15).

$$E_0^{(1)} \equiv E_{EL}^{(1)} = \langle \psi_0^{(0)} | \hat{V} | \psi_0^{(0)} \rangle \quad (2.11)$$

$$\begin{aligned} E_0^{(2)} &= \sum_{\substack{i \in A, j \in B \\ i, j \neq 0, 0}} \frac{|\langle \psi_{A,i} \psi_{B,j} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle|^2}{(E_{A,0} - E_{A,i}) + (E_{B,0} - E_{B,j})} \\ &= \underbrace{E_{ind}^{(2)}(A \rightarrow B) + E_{ind}^{(2)}(B \rightarrow A)}_{E_{ind}^{(2)}} + E_{disp}^{(2)} \end{aligned} \quad (2.12)$$

Energię indukcyjną (2.13) rozumie się jako energię związaną z polaryzacją molekuły B przez oddziaływanie elektrostatyczne z niezaburzonym rozkładem ładunku drugiej molekuły<sup>10</sup>; analogicznie interpretowany jest człon (2.14). Człon dyspersyjny (2.15) opisuje oddziaływanie chwilowych momentów dipolowych obu molekuł; jest to bardzo słabe oddziaływanie, odpowiedzialne jednak np. za proces skraplania gazów szlachetnych; jest to także efekt silnie związany z korelacją elektronową, przez co

<sup>9</sup> Wyznaczanie energii oddziaływania jako różnicy dwóch liczb przypomina procedurę ważenia kota. Na wadze łazienkowej przeprowadza się dwa pomiary wagi: swój oraz siebie z kotem na rękach; różnica wyników jest szukaną masą zwierzęcia. Jednakże, ze względu na niską czułość urządzenia, metoda ta zawodzi w przypadku chomika.

<sup>10</sup> Odzwierciedla to sumowanie w równaniu (2.13) przebiegające po stanach wzbudzonych molekuły B, którym towarzyszy funkcja falowa stanu podstawowego molekuły A  $\psi_{A,0}$ .



oddziaływania dyspersyjne nie są dobrze opisywane przez metody, w których nie jest ona właściwie ujęta.

$$E_{ind}^{(2)}(A \rightarrow B) = \sum_{\substack{j \in B \\ j \neq 0}} \frac{|\langle \psi_{A,0} \psi_{B,j} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle|^2}{E_{B,0} - E_{B,j}} \quad (2.13)$$

$$E_{ind}^{(2)}(B \rightarrow A) = \sum_{\substack{i \in A \\ i \neq 0}} \frac{|\langle \psi_{A,i} \psi_{B,0} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle|^2}{E_{A,0} - E_{A,i}} \quad (2.14)$$

$$E_{disp}^{(2)} = \sum_{\substack{i \in A \\ i \neq 0}} \sum_{\substack{j \in B \\ j \neq 0}} \frac{|\langle \psi_{A,i} \psi_{B,j} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle|^2}{(E_{A,0} - E_{A,i}) + (E_{B,0} - E_{B,j})} \quad (2.15)$$

### 2.2.2 Rachunek zaburzeń o adaptowanej symetrii

Wyrażenia na energię elektrostatyczną  $E_{el}^{(1)}$ , indukcyjną  $E_{ind}^{(2)}$  i dyspersyjną  $E_{disp}^{(2)}$  (równania 2.11 i 2.12) są poprawne w granicy dużych odległości między molekułami  $A$  i  $B$ ; problemem jest także to, że oddziaływania indukcyjne i dyspersyjne zawsze są wiążące, co kłóci się z obserwowanym odpychaniem pomiędzy cząsteczkami, kiedy są blisko siebie, niezależnym od ich oddziaływania elektrostatycznego. Źródłem tych rozbieżności jest niepoprawna symetria wyjściowej funkcji falowej względem permutacji elektronów<sup>11</sup>. Nawet jeśli funkcje  $\psi_{A,0}$  i  $\psi_{B,0}$  są z osobna antysymetryczne, ich iloczyn nie spełnia tego warunku dla zamiany elektronów pomiędzy molekułami.

Aby nadać funkcji  $\psi_0$  (2.8) poprawną symetrię, należy podzielić na nią operatorem antysymetryzacji (2.16), będącym sumą operatorów  $\hat{P}$  „wymieniających” określoną ilość elektronów pomiędzy molekułami. Tak skonstruowana funkcja (2.17) spełnia już wymóg antysymetryczności, ale nie jest funkcją własną hamiltonianu niezaburzonego (2.7). Z tego powodu nie można zastosować rachunku zaburzeń Raighleya-Schrödingera; można jednak wyprowadzić analogiczną do niego metodę iteracyjną.

$$\hat{A} = \frac{N_A! N_B!}{(N_A + N_B)!} \sum_k (-1)^k \hat{P}_k \quad (2.16)$$

$$\psi_0^{(0)} = \hat{A} \psi_{A,0} \psi_{B,0} \equiv \hat{A} \phi_0 \quad (2.17)$$

Ogólny schemat iteracyjny, generalizujący różne możliwe wersje rachunku zaburzeń o adaptowanej symetrii (ang. *Symmetry-Adapted Perturbation Theory*, SAPT), sformułowany został przez Bogumiła Jeziorskiego i Włodzimierza Kołosa z Uniwersytetu Warszawskiego. W tej pracy stosowany będzie najpopularniejszy schemat Jeziorskiego-Kołosa [113], w którym funkcja startowa jest opisana przez (2.18), natomiast energia oddziaływania w pierwszym rzędzie (iteracji) rachunku dana jest przez (2.19).

$$\psi_0^{(0)} = \hat{B} \phi_0 = \frac{\hat{A} \phi_0}{\langle \phi_0 | \hat{A} | \phi_0 \rangle} \quad (2.18)$$

$$\mathcal{E}_0(1) = \frac{\langle \phi_0 | \hat{V} \hat{A} | \phi_0 \rangle}{\langle \phi_0 | \hat{A} | \phi_0 \rangle} = E_{el}^{(1)} + E_{ex}^{(1)} \quad (2.19)$$

Różnica pomiędzy pierwszą poprawką SAPT a członem elektrostatycznym zwana jest *energiją wymienną*; ma ona zawsze wartość dodatnią i zanika eksponencjalnie wraz z odległością między molekułami; jest też przyczyną ich odpychania pomiędzy nimi na małych odległościach (dla molekuł zamkniętopowłokowych nazywana *odpychaniem walencyjnym*, jako efekt zakazu Pauliego). Całkiem dobrze można ją przybliżyć analizując permutacje zawarte w antysymetryzatorze (2.16). W sumie operatorów permutacji wyróżnić można wkład od wymian pojedynczych  $\hat{P}^{AB}$  (równanie 2.21, w którym  $\hat{P}_{ij}$  oznaczają

<sup>11</sup> Problem tkwi w tym, że cząstki elementarne są *nirozróżnialne*, tzn. nie można, nawet wirtualnie, ich ponumerować; stosowany przez nas aparat matematyczny (jak np. operator odpychania pomiędzy elektronami) jednak to robi. Należy zatem tak skonstruować funkcję falową, by zmiana „numeracji” cząstek nie zmieniała jej modułu (ten bowiem jest wielkością mierzalną, zgodnie z interpretacją Maxa Born’a). W przypadku elektronów, które są *fermionami*, funkcja musi być *antysymetryczna* względem ich permutacji (zamiany indeksów). Konsekwencją tego warunku jest m. in. zakaz Pauliego.

operatory zamiany elektronów  $i$  i  $j$ ) oraz pozostałe  $\hat{P}'$  (podwójne, potrójne etc.) (2.20).

$$\sum_k (-1)^k \hat{P}_k = 1 - \hat{P}^{AB} + \hat{P}' \quad (2.20)$$

$$\hat{P}^{AB} = \sum_{i \in A} \sum_{j \in B} \hat{P}_{ij} \quad (2.21)$$

Takie rozpisanie antysymetryzatora  $\hat{A}$  pozwala na uzyskanie wzoru na energię pierwszego rzędu SAPT w postaci (2.22), gdzie  $O(S^4)$  oznacza czony zanikające z odległością co najwyżej jak czwarta potęga całki nakładania  $S$ . Wówczas energia wymienna  $E_{ex}^{(1)} = E^{(1)} - E_{el}^{(1)}$  przyjmuje postać (2.23). Rozpisanie tych wzorów na całki pomiędzy orbitalami molekularnymi cząsteczek Czytelnik znaleźć może w Suplemencie (S.2.2)

$$E^{(1)} = \frac{\langle \psi_{A,0} \psi_{B,0} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle + \langle \psi_{A,0} \psi_{B,0} | \hat{V} \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle + O(S^4)}{1 + \langle \psi_{A,0} \psi_{B,0} | \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle + O(S^4)} \quad (2.22)$$

$$E_{ex}^{(1)} = \langle \psi_{A,0} \psi_{B,0} | \hat{V} \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle - \langle \psi_{A,0} \psi_{B,0} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle \langle \psi_{A,0} \psi_{B,0} | \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle + O(S^4) \quad (2.23)$$

W analogiczny sposób, każdą kolejną poprawkę SAPT można przedstawić jako sumę wkładu z przybliżenia polaryzacyjnego oraz odpowiadającej poprawki wymiennej (2.24); tak więc w drugim rzędzie wyróżniamy czony polaryzacyjno-wymienne  $E_{ex-ind}^{(2)}$  oraz dyspersyjno-wymienne  $E_{ex-disp}^{(2)}$ <sup>12</sup>

$$E_{SAPT} = \sum_n \left( E_{pol}^{(n)} + E_{ex}^{(n)} \right) \approx E_{el}^{(1)} + E_{ex}^{(1)} + E_{ind}^{(2)} + E_{ex-ind}^{(2)} + E_{disp}^{(2)} + E_{ex-disp}^{(2)} \quad (2.24)$$

Dla uzupełnienia prezentowanego wywodu wspomnieć należy o „dwuwymiarowej” wersji rachunku zaburzeń o adaptowanej symetrii. Można ją uznać za połączenie perturbacyjnego podejścia do energii korelacji (tzw. wielociałowy rachunek zaburzeń, ang. *Many-Body Perturbation Theory*, MBPT<sup>13</sup> [114]) z SAPT. Występują wówczas dwa zaburzenia: potencjał oddziaływania międzycząsteczkowego  $\hat{V}$  oraz tzw. potencjał fluktuacyjny (ang. *fluctuation potential*)  $\hat{W} = \hat{W}_A + \hat{W}_B$ , wykorzystywany w MBPT (2.25); funkcja  $\phi_0$  zaś jest iloczynem funkcji Hartree-Focka izolowanych molekuł (2.26). Energia SAPT jest wówczas dana podwójną sumą przedstawioną w równaniu (2.27); indeks  $n$  oznacza tutaj rząd poprawki względem  $\hat{V}$ , natomiast  $m$  — względem  $\hat{W}$ . Przykładowo,  $E_{el}^{(10)}$  oznacza oddziaływanie elektrostatyczne pomiędzy nieskorelowanymi rozkładami ładunku<sup>14</sup>, natomiast  $E_{el}^{(12)}$  poprawkę do tego oddziaływania uwzględniającą wewnątrzcząsteczkową korelację elektronową na poziomie MP2<sup>15</sup>.

$$\hat{H} = \hat{H}^{(0)} + \hat{V} + \hat{W} \quad (2.25)$$

$$\phi_0 = \psi_{A,HF} \psi_{B,HF} \quad (2.26)$$

$$E_{SAPT} = \sum_m \sum_n \left( E_{pol}^{(nm)} + E_{ex}^{(nm)} \right) \quad (2.27)$$

<sup>12</sup> Wyższe czony określane są jako indukcyjne, dyspersyjne oraz indukcyjno-dyspersyjne określonego rzędu.

<sup>13</sup> Najprostszym przykładem jest rachunek Møllera-Plesseta; metoda angażująca drugą poprawkę tego rodzaju (pierwsza znika) nazywana jest MP2.

<sup>14</sup> Pochodzącymi z metody Hartree-Focka.

<sup>15</sup> Innymi słowy, energia elektrostatyczna bazująca na gęstościach ładunku pochodzących z metody MP2 wynosi  $E_{el}^{(10)} + E_{el}^{(12)}$ .

## 2.3 Znaczenie poszczególnych składowych energii oddziaływań i ich koszt obliczeniowy

### 2.3.1 Addytywność oddziaływań

Przedstawiony w poprzedniej sekcji schemat odnosi się do oddziaływania pomiędzy dwiema cząsteczkami. W sytuacji, gdy w tworzeniu kompleksu uczestniczy więcej indywiduów, energię oddziaływania definiujemy jako (2.28) (analogicznie do równania 2.6).

$$E_{IE} = E_{supermol} - \sum_{i=1}^{monomery} E_i = \sum_{i \neq j} E_{ij} + \sum_{i \neq j \neq k} E_{ijk} + \dots \quad (2.28)$$

Energię oddziaływania złożonego kompleksu możemy rozpisać jako sumę oddziaływań pomiędzy wszystkimi parami cząsteczek  $E_{ij}$  oraz oddziaływań wielociałowych (2.28). Powszechnym przybliżeniem w modelowaniu fazy skondensowanej jest skupienie się tylko na oddziaływaniach dwuciałowych ze względu na koszt obliczeniowy (ilość par w układzie  $N$  wynosi  $\frac{N(N-1)}{2} \sim N^2$ , natomiast trójek  $\frac{N(N-1)(N-2)}{6} \sim N^3$ ). Jednakże wykazano, że uwzględnienie efektów trójciałowych w niektórych wypadkach jest decydujące dla poprawnego opisu zachodzących zjawisk [115]<sup>16</sup>.

Z tej perspektywy warto poruszyć kwestię addytywności oddziaływań. Z szeregu (2.24) tylko dwa człony są addytywne w złożonych kompleksach: elektrostatyczny  $E_{el}^{(1)}$  i dyspersyjny  $E_{disp}^{(2)}$  (jednakże tylko w drugim rzędzie – wyższe człony dyspersyjne addytywne już nie są). Fakt ten stanowi podstawową zaletę elektrostatycznego opisu oddziaływań: w tym rzędzie obliczanie energii wiązania układu jako sumy oddziaływań dwuciałowych jest całkowicie uprawnione.

### 2.3.2 Dekompozycja energii oddziaływań

Przed dyskusją na temat szeregowania wkładów do energii oddziaływania, warto przyrównać szereg (2.27) do energii oddziaływania liczonej metodą supermolekularną na poziomie Hartree-Focka (2.29) i MP2 (2.30) [116, 117].

$$E_{IE}^{HF} = E_{el}^{(10)} + E_{ex}^{(10)} + E_{ind,resp}^{(20)} + E_{ex-ind,resp}^{(20)} + \delta_{int}^{HF} \quad (2.29)$$

$$E_{IE}^{MP2} = E_{IE}^{HF} + E_{el}^{(12)} + E_{ex}^{(11)} + E_{ex}^{(12)} + E_{ind,resp}^{(22)} + E_{ex-ind,resp}^{(22)} + E_{disp}^{(20)} + E_{ex-disp}^{(20)} + \delta_{int}^{MP2} \quad (2.30)$$

W powyższych równaniach indeks *resp* oznacza człony, w których uwzględniono relaksację orbitali, zaś  $\delta_{int}^{HF}$  (delta HF<sup>17</sup>) oraz  $\delta_{int}^{MP2}$  (delta MP2<sup>18</sup>) oznaczają różnicę pomiędzy odpowiednim szeregiem a energią supermolekularną, częściowo zawierającą poprawki wyższych rzędów. Przez porównanie obu wyrażeń widać, że energia dyspersyjna jest nieobecna w energii oddziaływania Hartree-Focka, w której nieuwzględniona jest korelacja elektronowa; widać także, że energia korelacji na poziomie MP2 zawiera, prócz wymienionej już dyspersji, korelacyjne poprawki do energii elektrostatycznej, wymiennej i indukcyjnej<sup>19</sup>.

Istnieje szereg sposobów kategoryzowania (*dekompozycji*) składowych oddziaływań, od opartych na opisanym rachunku SAPT [117] po inne, popularne schematy jak np. metoda Morokumy [118]. Na potrzeby tej pracy postanowiono odwołać się do hierarchii oddziaływań zdefiniowanej w ramach hybrydowego rachunku perturbacyjno-wariacyjnego (ang. *Hybrid Variation-Perturbation Theory*, HVPT) [119–121]. Wymaga to zdefiniowania kilku wielkości.

<sup>16</sup> Na szczęście, efekty czterociałowe i bardziej złożone mają sukcesywnie mniejszy udział w (2.28)

<sup>17</sup> Oprócz pewnych wyższych członów, wielkość ta uwzględnia też efekty związane z przybliżonym charakterem SAPT, np. tzw. blokadę Pauliego [111, Rozdział 13].

<sup>18</sup> Szczegółową dyskusję na temat różnicy pomiędzy energią korelacyjną na poziomie MP2 a wymienionymi członami SAPT można znaleźć w [112].

<sup>19</sup> Poprawki korelacyjne podzielić można na związane z korelacją *międzycząsteczkową* (dotyczy to oddziaływań dyspersyjnych) i *wewnątrzcząsteczkową* (związaną z oddziaływaniami elektrostatycznym i indukcyjnym).

### Składowa penetracyjna $E_{pen}^{(1)}$ i multipolowa $E_{mtp}^{(1)}$

Energię elektrostatyczną  $E_{el}^{(1)}$  można opisać za pomocą multipoli atomowych  $E_{mtp}^{(1)}$ ; ze względu jednak na ich punktowy charakter, na bliskich odległościach wielkość ta będzie odbiegać od pełnej całki kulombowskiej opisującej oddziaływanie *niepunktowych* rozkładów ładunku; owa „niepunktowość” ilościowo określona jest przez tzw. *składową penetracyjną*, zdefiniowanej jako  $E_{pen}^{(1)} = E_{el}^{(1)} - E_{mtp}^{(1)}$ .

### Energia Heitlera-Londona $E^{HL}$

Pierwsza poprawka SAPT (2.19) równa jest energii oddziaływania Heitlera-Londona (mającej duże znaczenie historyczne)  $E^{HL} = \frac{\langle \hat{A}\phi^{(0)} | \hat{H} | \hat{A}\phi^{(0)} \rangle}{\langle \hat{A}\phi^{(0)} | \hat{A}\phi^{(0)} \rangle} - E_{A,0} - E_{B,0}$ ; można ją uważać za spełniającą zakaz Pauliego energię oddziaływania dwóch „sptywnych” molekuł. Z tej perspektywy energię wymienną  $E_{ex}^{(1)}$  można uzyskać albo w przybliżony sposób ze wzoru (2.23), albo dokładnie jako różnicę  $E_{ex}^{(1)} = \Delta^{HL} - E_{el}^{(1)}$ <sup>20</sup>.

### Energia delokalizacyjna $E_{del}^{(R)}$

Efekty związane z polaryzacją cząsteczek (uwzględniające relaksację, lecz pomijające efekty korelacyjne) oraz międzycząsteczkowym przeniesieniem ładunku<sup>21</sup> można zawrzeć w tzw. *członie delokalizacyjnym*  $E_{del}^{(R)}$ ; na mocy równania (2.29) i definicji energii Heitlera-Londona jest on równy  $E_{del}^{(R)} = E_{int}^{HF} - E^{HL}$ .

### Energia korelacyjna $E_{corr}$

Efekty korelacyjne — zarówno wewnątrz- jak i międzycząsteczkowe — można połączyć razem w członie korelacyjnym  $E_{corr}$ ; zdefiniowany on jest jako różnica pomiędzy energią oddziaływania uzyskaną za pomocą metody uwzględniającej korelację elektronową (MBPT, teoria sprzężonych klastrów i inne; tutaj, jako przykład, wspomniana zostanie metoda MP2) a energią HF  $E_{corr} = E_{MP2} - E_{HF}$ ; interpretować ją można jako miarę wkładu efektów korelacyjnych do całkowitej energii oddziaływania.

Mając na uwadze powyższe definicje, energię DTSS (jak również każdą inną *różnicową* energię oddziaływania  $\Delta E$ ) można przedstawić jako sumę składowych daną przez równanie (2.31); posługując się notacją  $O(f(X))$  oznaczającą asymptotyczne skalowanie czasu obliczeniowego<sup>22</sup> oraz przyjmując  $A$  za liczbę atomów w układzie i  $N$  za liczbę funkcji bazy, ustalić można następującą hierarchię poziomów

<sup>20</sup> Ponieważ w praktyce dysponujemy tylko przybliżonymi funkcjami falowymi, tak wyrażona energia wymienna nie jest równa  $E_{ex}^{(10)}$  z rachunku SAPT przedstawionego wcześniej. Jeżeli obliczenia wykonywane są w bazie dimeru, różnią się one o tzw. deltę Murrella, co zostało opisane w Suplemencie, sekcja S.2.1. Szczegółową dyskusję na ten temat można znaleźć w książce [112].

<sup>21</sup> Człon opisujący przeniesienie ładunku (ang. *Charge Transfer*, CT) może być alternatywnie opisany przy pomocy wyznacznikowych funkcji jonowych. Niemniej jednak, zarówno człon indukcyjny jak i CT są silnie zależne od bazy i dopiero ich suma nie wykazuje takiej zależności (Tabela 5 w [121]).

<sup>22</sup> Notacja wykorzystywana w analizie wydajności algorytmów;  $X$  oznacza pewną wielkość opisującą rozmiar danych wejściowych; mówi się, że tempo wzrostu funkcji kosztu  $g(X)$  (np. czasu obliczeniowego, ale może się także to odnosić do innych zasobów) jest rzędu co najwyżej  $f(X)$  (czyli  $g(X) \in O(f(X))$ ), jeżeli istnieją takie stałe  $X_0$  i  $c$ , że dla każdego  $X > X_0$  zachodzi nierówność  $g(X) \leq f(X) \cdot c$  (czyli  $\exists X_0, c \forall X > X_0 g(X) \leq f(X) \cdot c$ ).

teoretycznego opisu  $\Delta E$  (2.31).

$$\begin{aligned}
 \Delta E &= \Delta E_{mtp}^{(1)} + \Delta E_{pen}^{(1)} + \Delta E_{ex}^{(1)} + \Delta E_{del}^{(R)} + \Delta E_{corr} & (2.31) \\
 O(N^5) &\underbrace{\hspace{10em}}_{\Delta E^{MP2}} \\
 O(N^4) &\underbrace{\hspace{10em}}_{\Delta E^{SCF}} \\
 O(N^4) &\underbrace{\hspace{10em}}_{\Delta E^{HL}} \\
 O(N^4) &\underbrace{\hspace{10em}}_{\Delta E_{el}^{(10)}} \\
 O(A^2) &\underbrace{\hspace{10em}}_{\Delta E_{mtp}^{(10)}} \\
 O(A) &\Delta E_{Vq} = \sum_i (V^{TS} - V^S) q_i^E & (2.32)
 \end{aligned}$$

W szeregu prac naszego zespołu [37, 38] wykorzystano powyższą hierarchię by pokazać, że decydującą składową energii różnicowej jest dalekozasięgowa składowa multipolowa oddziaływania elektrostatycznego, gdyż pozostałe składowe krótkozasięgowe wykazują tendencję do wzajemnego niwelowania się. Pozwala to uzasadnić — w oparciu o obliczenia *ab initio* — wybór modeli elektrostatycznych jako mających największe zastosowanie w racjonalnym projektowaniu katalizatorów. Co więcej, przybliżenie energii elektrostatycznej  $\Delta E_{el}^{(10)}$  jej składową multipolową  $\Delta E_{mtp}^{(10)}$  pozwala na znaczne obniżenie kosztu obliczeniowego ( $O(A^2)$  zamiast  $O(N^4)$ ); możliwe jest także dalsze uproszczenie modelu do poziomu oddziaływania ładunków katalizatora z potencjałami elektrostatycznymi reagentów  $\Delta E_{Vq}$  (skalującymi się liniowo  $O(A)$  — równanie 2.32).

## 2.4 Pola katalityczne

Przedstawiona w sekcji 2.3.2 dekompozycja energii oddziaływań pozwala w systematycznie upraszczać energię DTSS (2.1); przybliżając energię stabilizacji tylko składową oddziaływań elektrostatycznych (2.11), a następnie opisując je jako oddziaływanie ładunków otoczenia z potencjałem elektrostatycznym reagentów (2.33), zyskujemy bardzo wydajne przybliżenie (2.32). W równaniu (2.33) wykrzyknik oznacza miejsce, w którym konieczne jest założenie nieziennej (choćby w przybliżeniu) konfiguracji ładunków otoczenia na ścieżce reakcji.

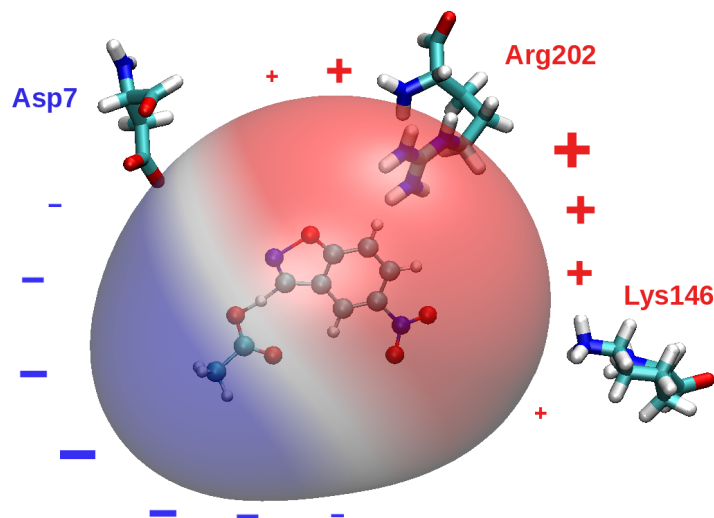
$$\begin{aligned}
 E_{IE}(TS) - E_{IE}(S) &\approx \sum_{i \in env} q_i V_{TS}(\vec{r}_i) - \sum_{i \in env} q_i V_S(\vec{r}_i) \\
 &\stackrel{!}{=} \sum_{i \in env} q_i (V_{TS} - V_S)(\vec{r}_i) \equiv \sum_{i \in env} q_i \Delta_S(\vec{r}_i) & (2.33)
 \end{aligned}$$

Najlepszą wartość energii DTSS (minimum energetyczne), a tym samym największe obniżenie bariery uzyskuje się, gdy ładunki otoczenia rozmieszczone są zgodnie z ekstremalnymi wartościami  $-\Delta_S$ , pola określanego jako *statyczne pole katalityczne*. Stąd, przy spełnieniu powyższych założeń, pole to zawiera informację o rozkładzie ładunku katalizatora idealnego dla danej reakcji. Warto zauważyć, że ta wiedza pochodzi wyłącznie ze znajomości struktur substratu i stanu przejściowego. Przykładowy schemat takiego pola przedstawia Rysunek 2.3.

Kolejne wiadomości o „idealnym otoczeniu katalitycznym” poznajemy w oparciu o gradient  $\Delta_S$ , zwany *dynamicznym polem katalitycznym* (2.34).

$$\vec{\Delta}_d \stackrel{def}{=} -grad \Delta_S = \vec{F}_{TS} - \vec{F}_S & (2.34)$$

W pierwotnej interpretacji opublikowanej w [122], pole wektorowe  $\vec{\Delta}_d$  mówi o takich ruchach ładunków otoczenia, które obniżają barierę reakcji „podażając” za postępowaniem reakcji, co jest bardziej efektywne od otoczenia statycznego. Jednakże bardzo trudno jest kontrolować takie właściwości w projektowaniu molekularnym, przez co zastosowanie tak rozumianego dynamicznego pola katalitycznego



Rysunek 2.3: Schematyczne przedstawienie przykładowego pola katalitycznego. Powierzchnia wokół stanu przejściowego reakcji eliminacji Kempa (*vide* sekcja 3.2.2) została pokolorowana zgodnie z wartością  $\Delta_S$ : kolor czerwony oznacza wartość dodatnią, niebieski – ujemną; dodatkowo umieszczono również reszty aminokwasowe ze zmutowanej wersji enzymu, których położenie zgodne jest z  $\Delta_S$ .

$\vec{\Delta}_d$  ma mniejsze znaczenie praktyczne. Podjęto próby alternatywnego wykorzystania  $\vec{\Delta}_d$ , jednakże uzyskane dotąd wyniki nie mają istotnego znaczenia praktycznego z punktu widzenia projektowania enzymów (*vide* sekcje S.3 i S.4.1 w Suplemencie).

## 2.5 Kumulatywne atomowe rozwinięcie multipolowe

Sprowadzenie energii DTSS do pól katalitycznych (sekcja 2.4) stanowi wygodne narzędzie, pozwalające w jednym kroku obliczeniowym (!) uzyskać informację o optymalnych położeniach ładunków otoczenia katalitycznego. Równocześnie jednak przybliżenie to jest bardzo zgrubne, ze względu na anizotropowy charakter rozkładu ładunku wokół atomów [52]; dokładniejsze przybliżenie  $E_{el}^{(1)}$  można uzyskać stosując atomowe rozwinięcie multipolowe. Takich rozwinięć, w związku z arbitralnością definicji ładunku (i dowolnego wyższego multipola) atomowego, istnieje wiele; popularne są, dla przykładu, rozproszone multipole zaproponowane przez Antony’ego Stone’a (ang. *Distributed Multipole Analysis*, DMA) [123] oraz multipole oparte na podziale gęstości Hirshfelda [124]. W niniejszej pracy wykorzystywane będą kumulatywne atomowe multipole (ang. *Cumulative Atomic Multipole Moments*, CAMM), będące uogólnieniem analizy populacyjnej Mullikena na wyższe multipole [107].

Addytywny wkład do kartezyjskiego molekularnego momentu multipolowego  $M_{klm}$  pochodzący od atomu  $i$  wyznaczany jest zgodnie z równaniem (2.35), gdzie  $I$  i  $J$  oznaczają orbitale bazy, natomiast  $P_{IJ}$  jest odpowiednim elementem macierzy gęstości; w przypadku, gdy  $k = l = m = 0$  sprowadza się to do ładunków Mullikena. Następnie tak otrzymane multipole poddane są translacji do układu współrzędnych związanego z danym atomem, co wiąże się z iteracyjną transformacją (2.36). W ten sposób otrzymane zostają kumulatywne multipole CAMM scentrowane na atomach cząsteczki. Procedurę tę można wykonać w oparciu o dowolną macierz gęstości, a zatem mogą to być także multipole skorelowane (tzn. wyznaczone z gęstości otrzymanej metodą uwzględniającą korelację elektronową). Podprogram wyznaczający momenty CAMM (autorstwa Karola M. Langnera [125]) dostępny jest w programie *GAMESS-US* [126], natomiast manipulację momentami multipolowymi (w tym wyznaczania energii oddziaływania) umożliwia moduł Pythona *pymolecule* (opracowany przez tego samego

autora).

$$M_{klm,i} = \langle x^k y^l z^m \rangle_i = Z_i x_i^k y_i^l z_i^m - \sum_{I \in i} \sum_J P_{IJ} \langle I | x^k y^l z^m | J \rangle \quad (2.35)$$

$$M_{klm,i}^{C\text{AMM}} = M_{klm,i} - \sum_{\substack{k' \geq 0 \\ k' l' m' \neq klm}} \sum_{l' \geq 0} \sum_{m' \geq 0} \binom{k}{k'} \binom{l}{l'} \binom{m}{m'} \times x_i^{k-k'} y_i^{l-l'} z_i^{m-m'} M_{k'l'm',i}^{C\text{AMM}} \quad (2.36)$$

Energie wyznaczone za pomocą multipoli CAMM odznaczają się zredukowaną zależnością od bazy funkcyjnej [107], i okazały się dobrym predyktorem względnej stabilności kompleksów z wiązaniem wodorowym [127] oraz energii wiązania inhibitorów [128]; w tym drugim przypadku, dzięki rozszerzeniu o potencjał dyspersyjny  $D_{as}$  opracowany przez Katarzynę Pernal [129] (fitowany do  $E_{disp}^{(2)} + E_{ex-disp}^{(2)}$  z SAPT)<sup>23</sup>, możliwe było osiągnięcie tego nawet dla inhibitorów o charakterze hydrofobowym, w których oddziaływania elektrostatyczne odgrywają mniejszą rolę.

Na koniec wspomnieć należy, że oprócz wyznaczania energii oddziaływania, CAMMy mogą służyć także obliczaniu wartości pól katalitycznych  $\Delta_s$  i  $\bar{\Delta}_d$ . Jest to podejście ogólniejsze od modelu bazującego na ładunkach punktowych i może mieć krytyczne znaczenie w przypadku reakcji, w których układ reagujący jest neutralny elektrycznie [130].

### 2.5.1 Biblioteka CAMM rotamerów aminokwasów

Skalowanie czasu obliczeniowego związanego z wyznaczaniem  $E_{mtp}^{(1)}$  opisane w (2.31) jako  $O(A^2)$  (gdzie  $A$  oznacza liczbę atomów w układzie) nie uwzględnia czasu obliczeń niezbędnych do wyznaczenia atomowych multipoli rozważanych podukładów. Ta część procesu obliczeniowego stanowi główne ograniczenie zastosowania rozważanego modelu, o ile bowiem czas wyliczenia energii oddziaływania pomiędzy dwoma indywiduami mieści się w granicach kilku milisekund (dla rozwinięcia rzędu  $R^{-5}$ ), o tyle czas potrzebny na obliczenie funkcji falowej każdego z nich liczy się najczęściej w minutach. Jednakże, ponieważ w białkach będących głównym obiektem zainteresowania w niniejszej pracy występuje ograniczona ilość aminokwasów, całość obliczeń można znacznie usprawnić dysponując przeliczonymi dla nich wcześniej tablicami multipoli atomowych. Postanowiono więc skorzystać z biblioteki rotamerów Dynameonics (aktualizowanej w 2015 roku) [131], w pierwszej kolejności skupiając się na rotamerach niezależnych od konformacji głównego łańcucha polipeptydowego. Dla każdego rotameru z bazy obliczono CAMMy z gęstości elektronowej Hartree-Focka w bazach 6-311G(3d,2p) oraz 6-31G(d)<sup>24</sup>; ze względu na ograniczoną przenaszalność CAMM pomiędzy konformacjami, przetestowano także kilka schematów interpolacji multipoli (Supplement, sekcja S.8.3). Biblioteka, jak również rozszerzony (m. in. o funkcje wyliczające potencjał i wektor pola elektrycznego) moduł *pymolecule* dostępne są na stronie <http://156.17.246.1/CAMM>.

## 2.6 Obliczenia QM/MM

W celu uzyskania stanu przejściowego w teozymie KE07 wykorzystano metodę ONIOM [90] zaimplementowaną w programie *Gaussian09* [133]. W tej metodzie część układu, np. centrum reakcji chemicznej (zwana regionem QM lub *układem modelowym*), opisywana jest metodą „wysoką” (najczęściej którąś z metod obliczeniowych chemii kwantowej), zaś jego otoczenie metodą „niską” (np. mechaniką molekularną — stąd region ten nazywany bywa regionem MM); cały układ określa się mianem *układu rzeczywistego*. Przytoczono tutaj bezpośrednie tłumaczenie terminów użytych przez autorów; rozróżnienie na metody „wysokie” i „niskie” ma znaczenie czysto praktyczne — przypisanie ich do poszczególnych rejonów zależy od rozmiaru (pod)układu i kosztu obliczeniowego metody<sup>25</sup>.

<sup>23</sup> Możliwe to jest dzięki wspomnianej w sekcji 2.3.1 addytywności energii dyspersyjnej drugiego rzędu; chociaż człon wymiennie-dyspersyjny nie jest addytywny, silny zanik poprawek wymiennych wraz z odległością sprawia (choć nie zawsze), że są one zdominowane przez pojedyncze kontakty atom-atom.

<sup>24</sup> W przygotowaniu, aczkolwiek niewykorzystane w tej pracy, są także multipole skorelowane obliczone metodą MP2. Uwzględnienie efektów korelacyjnych skutkuje zwykle zmniejszeniem wartości multipolowej składowej energii oddziaływan elektrostatycznych o ok 10% [132].

<sup>25</sup> Metoda ONIOM umożliwia nie tylko dosłowne połączenia QM/MM (np. B3LYP/Amber94), ale praktycznie dowolne kombinacje metod (np. MP2/PM3).

Hamiltonian w metodzie ONIOM jest *subtraktywny*, tzn. zdefiniowany jak w równaniu (2.37), gdzie indeks dolny oznacza metodę obliczeniową, zaś indeks górny — region.

$$\hat{H}_{ONIOM} = \hat{H}_{high}^{model} + \hat{H}_{low}^{real} - \hat{H}_{low}^{model} \quad (2.37)$$

W tak przedstawionym równaniu oddziaływanie pomiędzy regionem modelowym a otoczeniem opisywane jest na poziomie metody niskiej (w obrębie  $\hat{H}_{low}^{real}$ ; ponieważ zwykle jest to mechanika molekularna, taką konstrukcję nazywa się *osadzeniem mechanicznym* (ang. *mechanical embedding*). Metoda pozwala także na uwzględnienie oddziaływania elektrostatycznego poprzez całki jednoelektronowe w  $\hat{H}_{high}^{model}$ , dzięki czemu opisuje ona polaryzację rozkładu ładunku regionu QM wskutek oddziaływania ze środowiskiem; taki schemat nazywa się *osadzeniem elektronowym* (ang. *electronic embedding*). Osobnym zagadnieniem jest kwestia traktowania wiązań przecinanych przez granicę między regionami QM i MM; jednym z rozwiązań, stosowanym w przeprowadzonych na potrzeby tej pracy obliczeniach, są tzw. *atomy łącznikowe* (ang. *link-atoms*). Zabieg polega na „zaślepieniu” przeciętych wiązań atomami wodoru w trakcie obliczeń QM. Szczegółową dyskusję na temat metod hybrydowych można znaleźć w [87].

Za punkt startowy posłużyła struktura otrzymana w prywatnej korespondencji od Anastazji Alexandrowej; wstępne przygotowanie struktury do obliczeń zostało wykonane za pomocą zestawu skryptów autorstwa Peng Tao i H. Bernharda Schlegela [134] oraz własnego skryptu przeznaczonego do wyznaczenia poprawnej macierzy połączeń (ang. *connectivity matrix*). W obliczeniach metodą ONIOM [90] z wykorzystaniem funkcjonału B3LYP/6-311+G(d,p) oraz pola siłowego Amber94 (osadzenie mechaniczne), region modelowy składał się z reagenta (benzoizoksazol) oraz reszty octanowej reprezentującej Glu101 (graniczne wiązanie C<sub>γ</sub>-C<sub>δ</sub> Glu101 zostało zastąpione łącznikowym atomem wodoru). W ten sposób zoptymalizowano kompleks enzym-substrat oraz znaleziono stan przejściowy reakcji w enzymie. Podejście to jednak w sposób jawny pomija polaryzację reagentów wskutek oddziaływania otoczenia białkowego, którego uwzględnienie jest szczególnie zalecane w przypadku reakcji chemicznych; ponieważ napotkano problemy techniczne w optymalizacji z wykorzystaniem osadzenia elektrostatycznego, posłużono się procedurą refitowania ładunku regionu QM [135].



# Rozdział 3

## Wyniki

Zrealizowanie przedstawionego w Rozdziale 1 głównego celu badań, jakim jest rozwój metod projektowania enzymów opartych na koncepcji DTSS, wymaga rozstrzygnięcia szeregu szczegółowych zagadnień. Przedstawione one są poniżej w formie pytań.

- Jak reprezentować pole katalityczne?
- Jak opisać rozkład ładunku katalizatora?
- Jak duża jest korelacja pomiędzy energią DTSS a barierą własną?
- Czy model DTSS/DISS oparty na multipolach CAMM poprawnie opisuje proste mutacje?
- Czy model DTSS i pola katalityczne mogą być użyteczne w wyjaśnieniu roli mutacji uzyskanych doświadczalnie drogą kierowanej ewolucji (LDE)?

Pierwsze trzy pytania dotyczą natury wykorzystywanych przybliżeń, zakresu ich stosowalności i możliwego błędu, z jakim się wiążą. Dwa kolejne zagadnienia odnoszą się do konkretnych enzymów, na których mutantach testowany był model: izomerazy ketosteroidowej i teozymu KE07.

### 3.1 Analiza modeli elektrostatycznych

#### 3.1.1 Reprezentacja rozkładu ładunku układu reagującego

Ze względu na koszt obliczeniowy (*vide* równania 2.31 i 2.32), najbardziej efektywnym modelem rozkładu ładunku jest jego reprezentacja za pomocą ładunków punktowych, zazwyczaj osadzonych na atomach. Problem jednakże polega na tym, że nie ma jednoznacznej definicji ładunków atomowych. Każda z możliwych metod ich obliczania ma pewne wady; dla przykładu, ładunki oparte na analizie populacyjnej Mullikena silnie zależą od wykorzystanej bazy funkcyjnej, zaś ładunki dopasowywane do potencjału elektrostatycznego<sup>1</sup> mogą uzyskać нефizyczne wartości wskutek nadmiernego dopasowania (ang. *overfitting*)<sup>2</sup>. Ponadto, jak wspomniano w Rozdziale 1, w pewnych przypadkach modele oparte na ładunkach atomowych nie oddają poprawnie potencjału elektrostatycznego wokół cząsteczki, co wiąże się z anizotropowym charakterem jej rozkładu ładunku (np. obecnością wolnych par elektronowych).

Z punktu widzenia tej pracy, ważny jest nie tyle poprawny opis absolutnych wartości gęstości elektronowej cząsteczki, co jej zmian w czasie reakcji chemicznej. Dlatego na przykładzie kilku reakcji prześledzono zmiany potencjału elektrostatycznego reagentów (ang. *Molecular Electrostatic Potential*, MEP) wyznaczonego różnymi modelami; za miarę przyjęto średnie odchylenie kwadratowe (ang. *Root-mean-square deviation*, RMSD) potencjału obliczonego daną metodą względem wartości spodziewanej

---

<sup>1</sup> Podejście zaproponowane w 1978 roku przez Franka Momany [136]; popularną metodą tego rodzaju, wykorzystywaną w tej pracy, jest metoda Merza-Kollmana [137].

<sup>2</sup> Zjawisko to spotyka się np. gdy dane wejściowe, wbrew założeniom, nie są niezależne statystycznie; podobna sytuacja ma miejsce w przypadku *atomowych multipoli* dopasowywanych do potencjału elektrostatycznego. Zgodnie z ostatnimi pracami Jakobsen [138], ilość istotnych atomowych momentów multipolowych w układzie  $N_A$  atomów jest rzędu  $2N_A$ , tzn. dodatkowe parametry nie poprawiają istotnie opisu MEP.

na powierzchni dostępnej rozpuszczalnikowi (ang. *Solvent Accessible Surface*, SAS); obliczenia gęstości elektronowej i potencjału przeprowadzono metodą Hartree-Focka w bazie 6-311++G(d,p).

W pracy [130] przeanalizowano łącznie cztery reakcje chemiczne: alkaliczną hydrolizę demetonu-*S* i fosalonu, izomeryzację cyjanowodoru oraz syntezę kwasu węglowego. Tutaj zdecydowano się przedstawić tylko dwa reprezentatywne przykłady — syntezę  $\text{H}_2\text{CO}_3$  i hydrolizę fosalonu (Rysunek 3.1). Na przedstawionym rysunku  $L$  oznacza rząd rozwinięcia CAMM (zdefiniowany w równaniu S.1 i omówiony w sekcji S.1 Suplementu). W obu przypadkach odchylenie MEP uzyskanego na podstawie ładunków Mullikena ( $L = 0$ ) zmienia się znacznie na drodze reakcji; przy  $L = 3, 4$  staje się ono mniej więcej stałe, zaś przy  $L = 5$  można uznać, że wartość RMSD osiągnęła wartość graniczną. W przypadku ładunków Merza-Kollmana, odchylenie MEP osiąga niską i mniej więcej stałą wartość tylko w przypadku hydrolizy fosalonu; w przypadku syntezy  $\text{H}_2\text{CO}_3$  jest on nie tylko znaczny, ale także istotnie zmienny na ścieżce reakcji. Podobnie wygląda zestawienie hydrolizy demetonu-*S* z izomeryzacją cyjanowodoru. Różnice pomiędzy tymi reakcjami wynikają z różnego całkowitego ładunku reagentów: -1 w przypadku hydrolizy wspomnianych fosfoestrów i 0 w pozostałych dwóch reakcjach. Oznacza to, że gdy całkowity ładunek układu jest zerowy, prawidłowe opisanie MEP oraz jego zmian podczas reakcji chemicznej nie jest możliwe za pomocą ładunków punktowych. Jeżeli jednak układ jest obdarzony niezerowym ładunkiem elektrycznym, ładunki ESP mogą opisać MEP i jego zmiany lepiej od przedstawionego rozwinięcia CAMM.

Jednakże nawet wówczas energia oddziaływania liczona za pomocą ładunków fitowanych do potencjału elektrostatycznego jest obciążona pewnym błędem. Wynika to stąd, że potencjał, do którego są one dopasowywane, zawiera krótkozasięgową część penetracyjną, zanikającą na odległościach typowych w oddziaływaniach międzycząsteczkowych. Aby to zobrazować, przygotowano minimalny model imitujący oddziaływania z enolanową formą reagenta (obdarzoną ładunkiem  $-1$ ) w izomerii ketosteroidowej (*vide* 3.2.1), dla uproszczenia donor wiązania wodorowego zastępując cząsteczką fluorowodoru (Rysunek 3.2); zbadano także zależność analizowanych wielkości od odległości donor-akceptor mierzonej wzdłuż osi wiązania wodorowego. Momenty multipolowe CAMM (oraz DMA [123] dla porównania) oraz ładunki ESP fitowane różnymi metodami (Merza-Kollmana [137], Chelp [139] i ChelpG [140]) wyznaczono w oparciu o funkcje falowe monomerów uzyskane metodą HF/aug-cc-pVDZ. Wyniki, przedstawione w formie odchylenia względem wartości spodziewanej  $E_{elst}^{(10)}$ , zamieszczono na Rysunku 3.3. Można zauważyć, że owe błędy maleją z odległością mniej więcej eksponencjalnie, jak można się spodziewać po członie penetracyjnym energii oddziaływania. Widać również, że oba rozważane rozwinięcia multipolowe wiążą się z mniejszym błędem niż ładunki ESP; błąd energii oddziaływania już w pobliżu odległości równowagowej jest bliski 1 kcal/mol, podczas gdy rozważane ładunki atomowe dają podobny wynik dopiero po rozsunięciu monomerów o ok. 2Å względem położenia równowagowego.

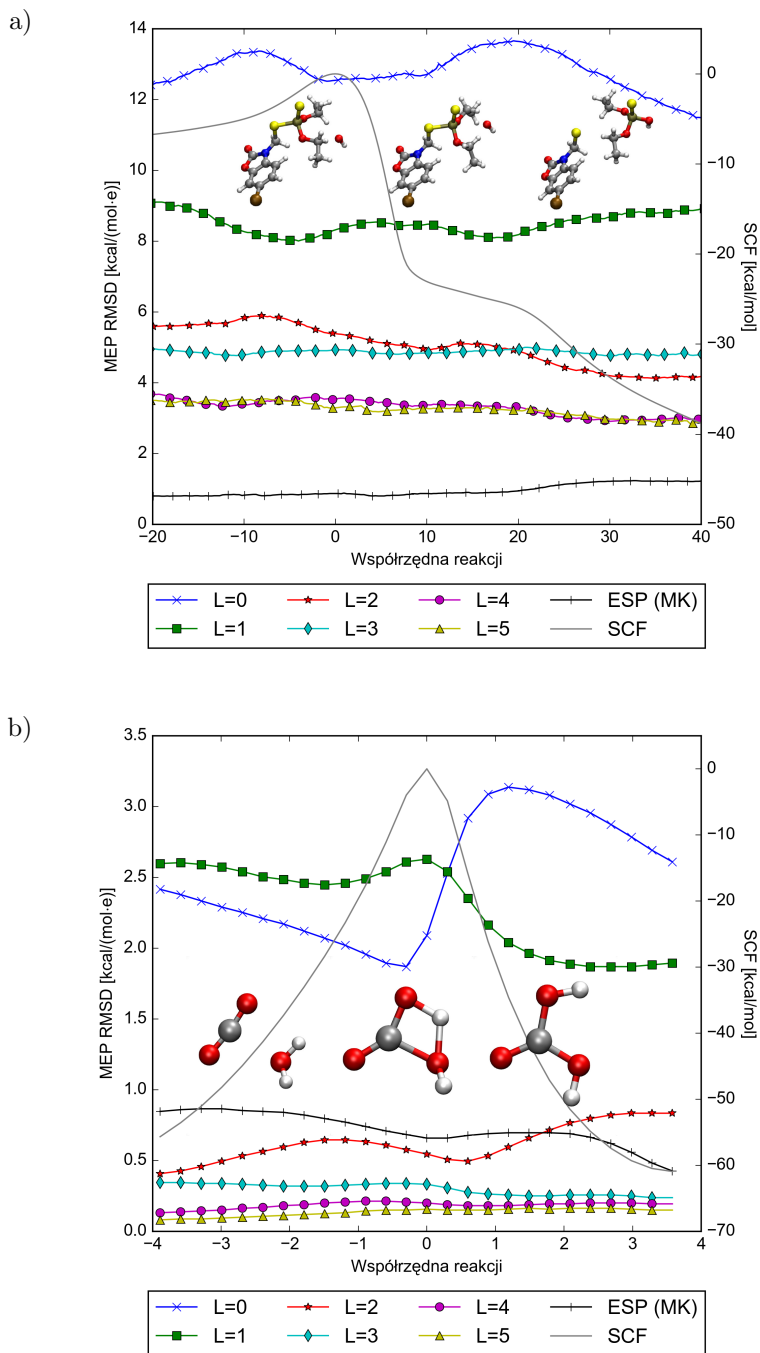
Na podstawie tych wyników można wysnuć hipotezę, że anizotropia rozkładu ładunku i jego zmian podczas reakcji chemicznej jest zanedbywalna w reakcjach, w których reagenty są obdarzone ładunkiem; oznacza to także, że w takich wypadkach reprezentacja  $\Delta_S$  za pomocą ładunków atomowych jest wystarczająco dokładna. Jednakże nie pozbawia to znaczenia bardziej uniwersalnego rozwinięcia CAMM; nawet jeżeli założenie to jest często spełnione w praktyce, wciąż pozostaje przypadkiem szczególnym; ponadto, wciąż należy mieć na uwadze zjawisko nadmiernego dopasowania, nieobecne w przypadku CAMM, jak również wpływ członu penetracyjnego obecnego w procedurze fitowania ESP na wartości energii oddziaływania elektrostatycznego (Rysunek 3.3).

### Wniosek 1

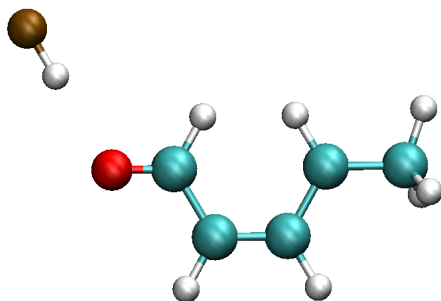
*Modele oparte na atomowym rozwinięciu multipolowym mają charakter uniwersalny, ponieważ mogą być stosowane zarówno do układów neutralnych jak i naładowanych. Niemniej jednak, w układach naładowanych elektrycznie potencjał elektrostatyczny jest zdominowany przez człon  $R^{-1}$ , przez co można opisać  $\Delta_S$  z dobrą dokładnością za pomocą ładunków Merza-Kollmana lub podobnych. Jednakże nawet wówczas energia oddziaływania uzyskana za pomocą tego typu ładunków może być obciążona błędem rzędu kilku kcal/mol i więcej na odległościach bliskich równowagowym.*

### 3.1.2 Reprezentacja otoczenia katalitycznego

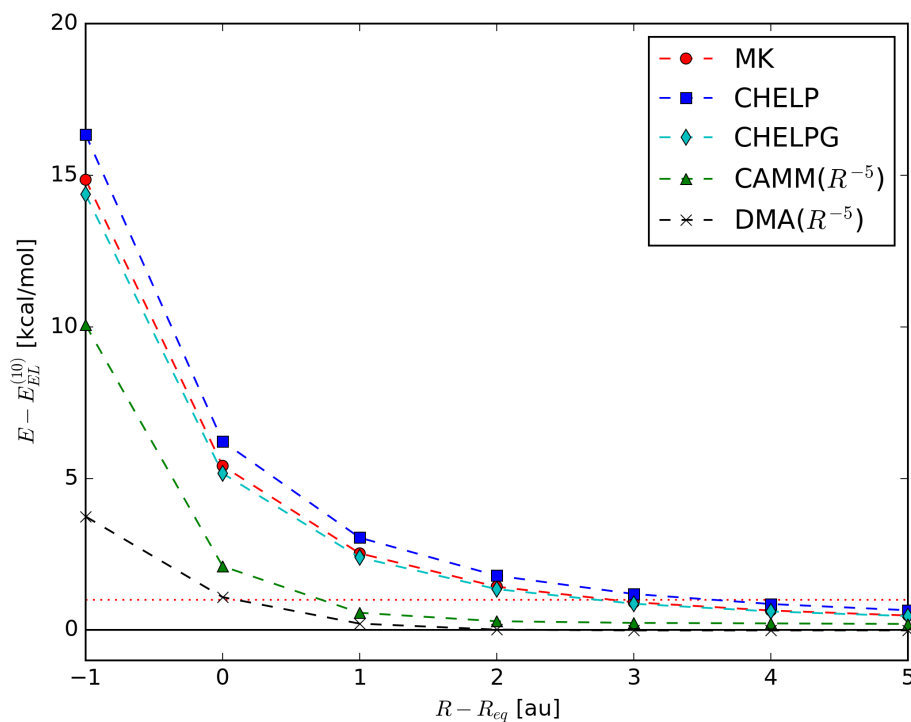
Obliczanie multipolowej składowej energii DTSS  $\Delta E_{mt}^{(10)}$  wymaga wyznaczenia atomowych momentów multipolowych zarówno dla reagentów, jak i otoczenia; jak wspomniano w sekcji 2.5.1, w przypadku białek, będących głównym obiektem zainteresowania w tej pracy, obliczenia można znacząco przyspieszyć korzystając z przygotowanej wcześniej biblioteki multipoli atomowych dla aminokwasów. Spośród



Rysunek 3.1: Porównanie RMSD potencjału elektrostatycznego (MEP) wyznaczonego za pomocą różnych rzędów rozwinięcia multipolowego CAMM oraz ładunków Merza-Kollmana (ESP). a) na przykładzie hydrolizy fosalamu, b) na przykładzie syntezy  $\text{H}_2\text{CO}_3$ . Modele kulowe reprezentują dla każdej reakcji odpowiednio substrat, TS i produkt; szara linia w tle przedstawia profil energetyczny reakcji (kcal/mol) liczony względem TS.



Rysunek 3.2: Układ wykorzystany do porównania ładunków fitowanych i multipoli atomowych. Zielone kule– atomy węgla, czerwone– tlenu, białe– wodoru, brązowa– fluoru.



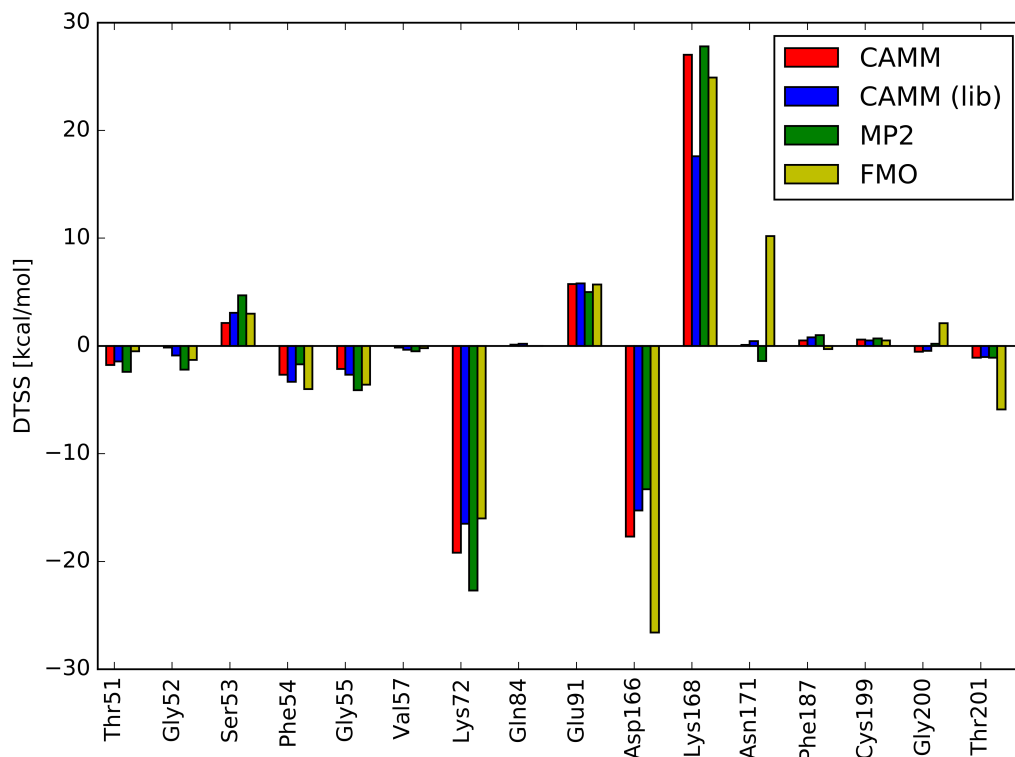
Rysunek 3.3: Błąd wyznaczenia wartości energii oddziaływania elektrostatycznego względem wartości spodziewanej  $E_{elst}^{(10)}$ . MK– ładunki Merza-Kollmana [137], CHELP– ładunki Chelp [139], CHELPG– ładunki ChelpG [140], DMA–multipole Stone'a [123]. Na osi  $OX$  odłożono odległość  $R$  mierzoną wzdłuż osi wiązania wodorowego względem wartości równowagowej. Czerwona przerywana linia oznacza poziom 1 kcal/mol.

podejść alternatywnych, wymienione tutaj zostaną metoda orbitali fragmentów molekularnych (ang. *Fragment Molecular Orbital*, FMO)<sup>3</sup> oraz wykorzystanie ładunków z popularnych pól siłowych.



Rysunek 3.4: Schemat reakcji katalizowanej przez kinazę białkową A. Zamiast reszty seryny może w niej uczestniczyć treonina.

Metoda FMO została niedawno zastosowana przez zespół Tore Brincka do kinazy białkowej A (PKA), badanej wcześniej przez pracowników naszej grupy, którzy wykazali dominujący wkład oddziaływań elektrostatycznych do efektu katalicznego tego enzymu [38]. Na tym przykładzie porównano wyniki FMO [141], MP2 [38] oraz CAMM, liczone *de novo* dla danych konformacji aminokwasów, oraz pobranych z biblioteki multipoli (Rysunek 3.5); multipole CAMM wyznaczone były z gęstości elektronowej obliczonej metodą Hartree-Focka w bazie 6-31G(d). Jak widać, człon  $\Delta E_{mtp}^{(10)}$  (CAMM), liczony zarówno *de novo*, jak i w oparciu o bibliotekę, pozostaje w dobrej zgodności z wynikami MP2; wartości uzyskane za pomocą FMO różnią się diametralnie w przypadku Asn171, Gly200 i Thr201, co może mieć związek z efektami polaryzacyjnymi i trójciałowymi, które stara się ona opisywać. Zaznaczyć należy, że pomimo liniowego skalowania FMO z rozmiarem układu [141], uzyskanie analogicznego wyniku za pomocą biblioteki CAMM jest wydajniejsze obliczeniowo; w tym konkretnym przypadku, przeprowadzając obliczenia jednowątkowe (procesor Intel<sup>®</sup> Xeon<sup>®</sup> CPU E5-2630 v2), czas wyznaczenia wkładu DTSS pochodzącego od 16 aminokwasów był rzędu 3 minut — tyle samo, ile średni czas wyznaczenia funkcji falowej Hartree-Focka (HF) dla *jednego* aminokwasu w tych samych warunkach.

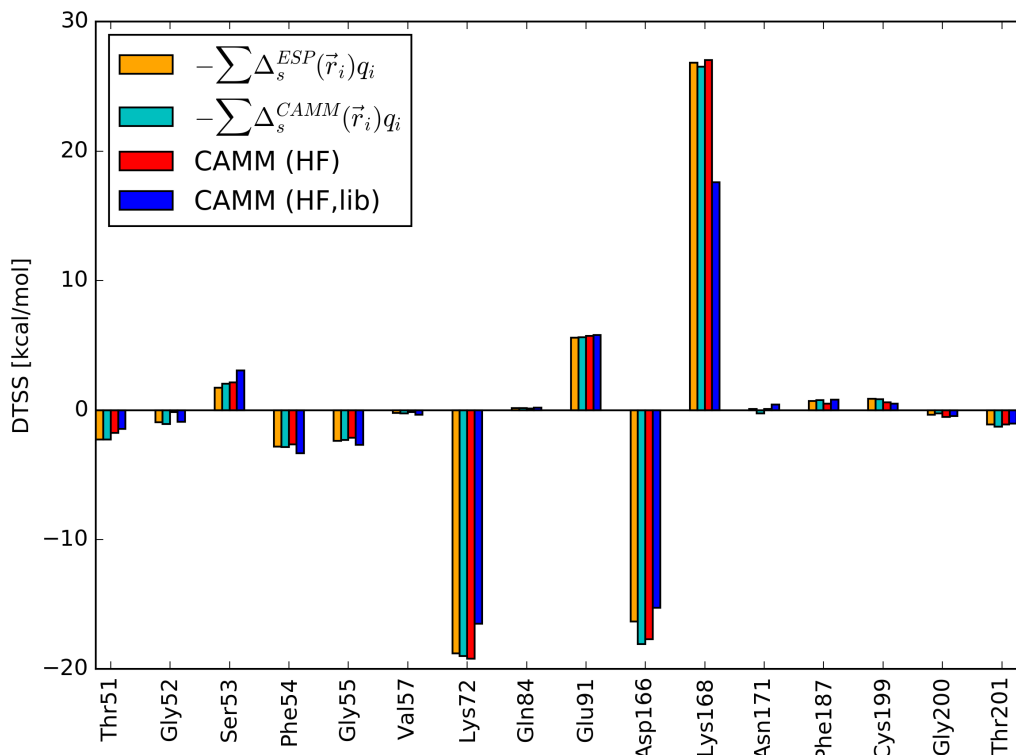


Rysunek 3.5: Porównanie biblioteki CAMM oraz metody FMO na przykładzie kinazy białkowej A.

Czas obliczeń można jeszcze bardziej skrócić (poniżej jednej sekundy we wspomnianych warun-

<sup>3</sup> W metodzie tej układ (molekuła) jest rozbijany na fragmenty, z których każdemu przypisany zostaje osobny układ orbitali molekularnych; dosłowne tłumaczenie tej nazwy powinno więc brzmieć *metoda orbitali molekularnych fragmentów*, jednakże nie brzmi to dobrze w języku polskim; z kolei uszczegółowienie *metoda orbitali molekularnych fragmentów cząsteczki* też nie jest zadowalającym terminem. Dlatego zdecydowano się na wymienione w głównym tekście tłumaczenie, przyjmując *implicite*, że występujące w nazwie *orbitale* to właśnie orbitale molekularne.

kach), przedstawiając multipolową składową DTSS  $\Delta E_{mtp}^{(10)}$  jako iloczyn wartości pola katalitycznego  $\Delta_S$  i ładunków środowiska  $\{q_i\}$ , przy czym wykonując obliczenia na przykładzie PKA nie znaleziono istotnej różnicy w czasie obliczeń pomiędzy reprezentacją CAMM pola katalitycznego  $\Delta_s^{CAMM}$  a jego reprezentacją za pomocą ładunków Merza-Kollmana  $\Delta_s^{ESP}$ . Jak widać na Rysunku 3.6, obie reprezentacje  $\Delta_S$  (wraz z ładunkami pola siłowego Amber94 osadzonymi na aminokwasach centrum aktywnego) prowadzą do wyników bliskich tym uzyskanym z wykorzystaniem CAMM zarówno na aminokwasach, jak i reagentach; pozostaje to w zgodności z Wnioskiem 1, jako że układ reagujący (ATP i reszta seryny) jest naładowany ujemnie.



Rysunek 3.6: Porównanie biblioteki CAMM oraz ładunków atomowych na przykładzie kinazy białkowej A. CAMM(HF) — DTSS z wykorzystaniem CAMM (gęstość Hartree-Focka); CAMM(HF,lib) — DTSS z wykorzystaniem biblioteki CAMM aminokwasów;  $-\sum \Delta_s^{CAMM}(\vec{r}_i)q_i$  — DTSS obliczony za pomocą pola katalitycznego (reprezentacja CAMM) i ładunków pola siłowego Amber94 na aminokwasach;  $-\sum \Delta_s^{ESP}(\vec{r}_i)q_i$  — DTSS obliczony za pomocą pola katalitycznego (reprezentacja za pomocą ładunków Merza-Kollmana) i ładunków pola siłowego Amber94 na aminokwasach.

W przypadku aminokwasów naładowanych (mających najsilniejszy wkład do DTSS) można pójść o krok dalej i przedstawić daną resztę jako punktowy ładunek równy jej ładunkowi formalnemu. W Tabeli 3.1 uzyskane w ten sposób energie porównano z wynikiem DTSS (metoda MP2) i sumą  $-\sum \Delta_s^{CAMM}(\vec{r}_i)q_i$  wykorzystującą ładunki pola siłowego Amber94; ładunki formalne zostały umieszczone na aminowym atomie azotu w przypadku lizyny oraz atomie węgla grupy karboksylowej w przypadku reszt kwasów asparaginowego i glutaminowego. Na tym przykładzie widać, że taka prosta reprezentacja daje przyzwoity wynik; największe odchylenie (ok. 9 kcal/mol) wystąpiło w przypadku Asp166, co jest spowodowane niewielką odległością od reagenta od reszty w porównaniu z jej rozmiarem<sup>4</sup> oraz samym kształtem tej grupy. Niemniej jednak wynik przedstawia bardzo dobrą zgodność jakościową.

Podsumowując, oba modele elektrostatyczne analizowane na wymienionym przykładzie — CAMM i ESP — dają wynik o bardzo dobrej zgodności (w przybliżeniu ilościowej) z bardziej zaawansowanymi

<sup>4</sup> Odległość w przypadku reszt Lys72 i Lys168 jest podobna, ale grupa aminowa jest mniejsza w porównaniu z karboksylową.

Tabela 3.1: Porównanie reprezentacji aminokwasów naładowanych (wszystkie energie w kcal/mol). Ładunki formalne (całkowite) zostały umieszczone na aminowym atomie azotu reszt lizyny oraz karboksylowych atomach węgla reszt kwasów glutaminowego i asparaginowego. Dodatkowo podano także najmniejszą odległość  $D_{min}$  pomiędzy reagentem a daną resztą.

Aminokwas	DTSS(MP2)	$q_i$ (Amber94)	$q_i$ (formalne)	$D_{min}$ [Å]
Lys 72	-22.7	-19.0	-18.4	1.8
Glu 91	5.0	5.6	5.1	4.8
Asp 166	-13.3	-18.1	-9.2	1.5
Lys 168	27.8	26.5	26.0	1.7

metodami, takimi jak MP2 czy FMO; jest to po części potwierdzenie Wniosku 1, a także konsekwencja faktu, że wkład do DTSS pochodzący od reszt naładowanych jest dominujący w porównaniu z innymi aminokwasami.

### Wniosek 2

*W przypadku, gdy układ reagujący jest naładowany, anizotropowość rozkładu ładunku katalizatora zdaje się mieć drugorzędne znaczenie.*

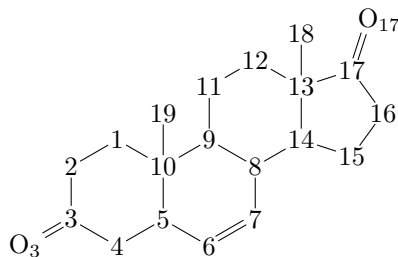
### Wniosek 3

*W przypadku aminokwasów naładowanych, dających najsilniejszy wkład do DTSS, ich reprezentacja za pomocą ładunku formalnego umieszczonego w miejscu grupy funkcyjnej (atomu azotu aminowego lub węgla karboksylowego, jak przyjęto wyżej) daje wynik bliski wartości  $\Delta E_{mtp}^{(10)}$ .*

## 3.2 Analiza wpływu mutacji na aktywność katalityczną

### 3.2.1 Izomeraza ketosteroidowa

Aby przetestować model DTSS(CAMM) na przykładzie szeregu mutantów jednego enzymu, wybrano w pierwszej kolejności izomerazę ketosteroidową (KSI); była on przedmiotem ważnej dyskusji na temat znaczenia oddziaływań elektrostatycznych w katalizie, jaka niedawno miała miejsce w literaturze [142–144] (*vide* sekcja 1.2.2 w Rozdziale 1). Punktem wyjścia dla prezentowanych w tej sekcji badań była ścieżka reakcji uzyskana metodą QM/MM przez Marca van der Kampa [145], udostępniona w ramach współpracy z Uniwersytetem w Bristolu. Jak wspomniano w Rozdziale 2 (sekcja 2.1.1), wprowadzono tutaj *różnicową stabilizację intermediatu* (DISS) [39].

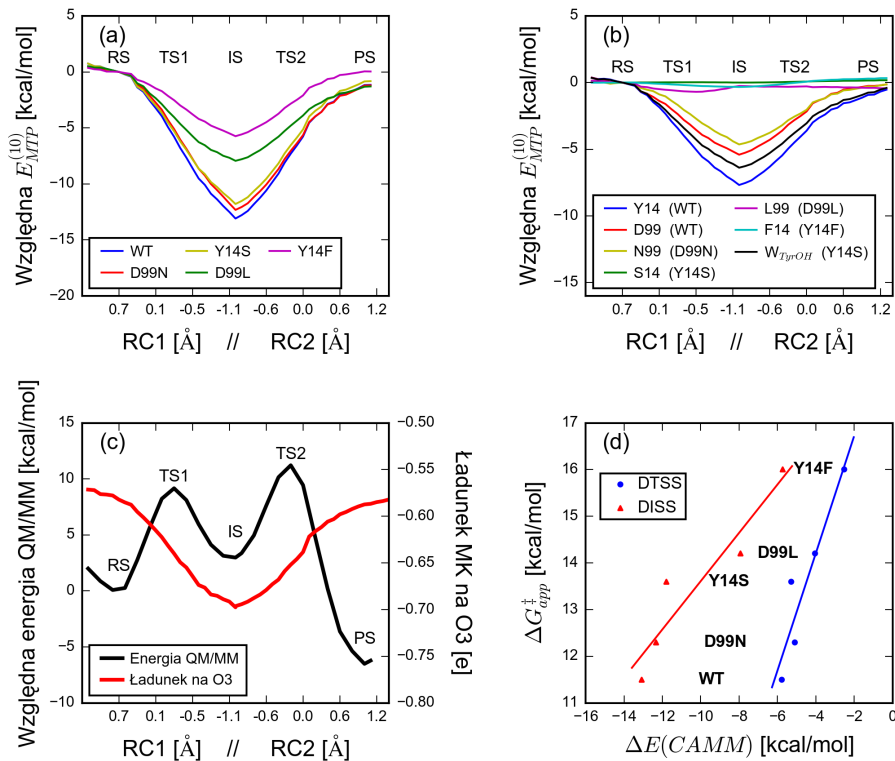


Rysunek 3.7: Numeracja atomów ciężkich w substracie izomerazy ketosteroidowej.

W KSI reszty Tyr14 i Asp99 (numeracja z enzymu pochodzącego z *Pseudomonas testosteroni*) tworzą wiązanie wodorowe z grupą karbonylową substratu (*vel* tlenem enolanowym intermediatu, *vide* Rysunek 1.5 w sekcji 1.2.2); analizie poddano, wraz z formą dziką, cztery mutanty: Y14F, Y14S, D99L, D99N. Ich struktury wygenerowano na dwa sposoby: w pierwszym przybliżeniu wprowadzono *minimalne zmiany* (MZ) do już posiadanej struktury, np. zamiana grupy hydroksylowej Tyr14 na atom wodoru czy zamiana grupy OH Asp99 na grupę aminową (w przypadku Y14S uwzględniono obecność cząsteczki wody w powstałej przestrzeni) [39]; drugi zbiór stanowiły centroidy głównych klastrów z symulacji dynamiki molekularnej (MD). Energie oddziaływania liczone tylko pomiędzy układem reagującym

(19-nortestosteron i Asp38) a dwiema wymienionymi resztami dziury oksyanionowej; ze względu na rozmiar układu reagującego, momenty CAMM wyznaczono w oparciu funkcję falową uzyskaną metodą HF/6-311G\*\*<sup>5</sup>; dla spójności wyników tą samą bazę zastosowano w obliczeniach SAPT0.

Na przykładzie pierwszego zestawu struktur (MZ) pokazano, że DISS i DTSS są wielkościami wzajemnie sobie odpowiadającymi (Rysunek 3.8). Śledząc profile energii oddziaływania (panele a i b na Rysunku 3.8), jak również całkowite energie DTSS/DISS (panel d), z dobrym przybliżeniem można przyjąć, że energia różnicowej stabilizacji pierwszego stanu przejściowego jest równa *połowie* energii DISS; patrząc na profil ładunku Merza-Kollmana na atomie tlenu O3 (panel c; numeracja atomów przedstawiona na Rysunku 3.7), można tłumaczyć to tym, że rozkład ładunku pierwszego TS tej reakcji odpowiada „połowicznemu” przekształceniu substratu w intermediat (w TS1 ładunek na O3 jest bliski średniej arytmetycznej ładunków w RS i IS).



Rysunek 3.8: Oddziaływania elektrostatyczne na ścieżce reakcji w KSI.

a– całkowite profile energii oddziaływania z reagentem, b– profile wkładów poszczególnych reszt do energii oddziaływania, c– profil energetyczny reakcji oraz ładunku (Merza-Kollmana [137], MK) na karbonylowym atomie tlenu reagenta, d– porównanie energii różnicowej stabilizacji stanu przejściowego (DTSS) i intermediatu (DISS) z eksperymentem. Współrzędne reakcji zdefiniowane jako różnice odległości transferowanego protonu od atomu docelowego i początkowego, *vide ref.* [145].

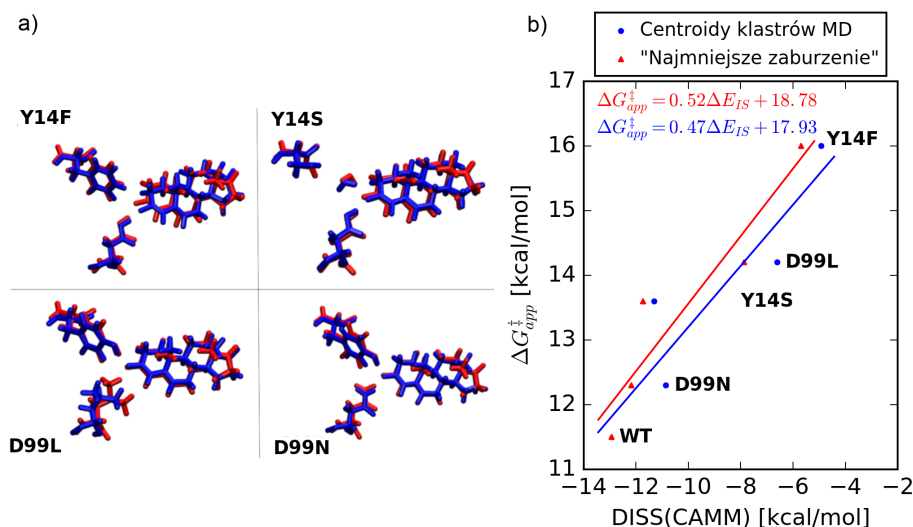
Drugi zestaw struktur (MD) pozwolił oszacować zmiany geometrii kompleksu enzym-substrat w stosunku do MZ; w przypadku mutantów Y14F substrat obrócił się bardziej w kierunku donora wiązania wodorowego Asp99, zaś w D99L łańcuch boczny leucyny przyjął odmienną konformację niż założona; pozostałe dwie struktury były dość bliskie MZ (Rysunek 3.9). Mimo to wynik  $\Delta E_{mtp}^{(10)}$  pozostaje zbliżony, łącznie z tym, że punkt przecięcia z osią rzędną (18.78 kcal/mol dla MZ i 17.93 kcal/mol dla MD) jest w dobrej zgodności z barierą<sup>5</sup> „zerowego pola” (18.8 kcal/mol) z pracy Frieda i Boxera [40].

Dla struktur MZ wykonano także dekompozycję oddziaływań wedle schematu SAPT0<sup>6</sup>, zdefiniowanego w równaniu 3.1 [117] (Rysunek 3.11). Potwierdza ona elektrostatyczny charakter stabilizacji

<sup>5</sup> Eksperymentalne bariery przyjęto jako związane z  $k_{cat}$  zależnością  $k_{cat} = \frac{k_b T}{h} \exp - \frac{\Delta G^\ddagger}{k_b T}$ .

<sup>6</sup> Ze względu na problemy techniczne związane z rozmiarem układu, nie udało się przeprowadzić obliczeń HVPT.





Rysunek 3.9: Zależność pomiędzy aktywnością katalityczną a DISS dla różnie wygenerowanych struktur.

a– porównanie geometrii „najmniejszego zaburzenia” (MZ) i głównych klastrów z symulacji dynamiki molekularnej (MD) centrum aktywnych mutantów, b– porównanie zależności pomiędzy DISS dwóch zestawów struktur a aktywnością eksperymentalną.

w KSI i jej mutantach; szczególnie na przykładzie cząsteczki wody (donora wiązania wodorowego w Y14S) widać wzajemne niwelowanie się składowych krótkozasięgowych. W przypadku Leu99 dodatkowo zaobserwowano dramatyczny wpływ tych składowych na finalny wynik; powtórzenie obliczeń dla konformacji tej reszty zgodnej ze strukturą MD prowadzi do wartości DISS bliskiej zeru (dotyczyło to także pojedynczych składowych). Warto zaznaczyć, że w przypadku innych reszt wpływ struktury nie był aż tak wyraźny; największą zmianę wartości wraz ze zmianą struktury wykazuje człon elektrostatyczny  $\Delta E_{elst}^{(10)}$ , gdzie niewielka odległość pomiędzy Leu99 a reagentem w strukturze MZ (1.51 Å) w porównaniu ze strukturą MD (2.57 Å) skutkuje znacznie większą (co do modułu) wartością  $\Delta E_{elst}^{(10)}$  ze względu na efekty penetracyjne. W przypadku kolejnych metod w hierarchii (HF i SAPT0) wzajemne znoszenie się oddziaływań krótkozasięgowych niweluje ten efekt. Co ciekawe, problem jest nieobecny w przypadku CAMM — wkład DISS pochodzący od Leu99 w obu przypadkach jest równy ok. -0.3 kcal/mol. Warto także zauważyć, że zasadniczo podobne do CAMM wyniki SAPT0 i HF uzyskane są znacznie większym nakładem obliczeniowym.

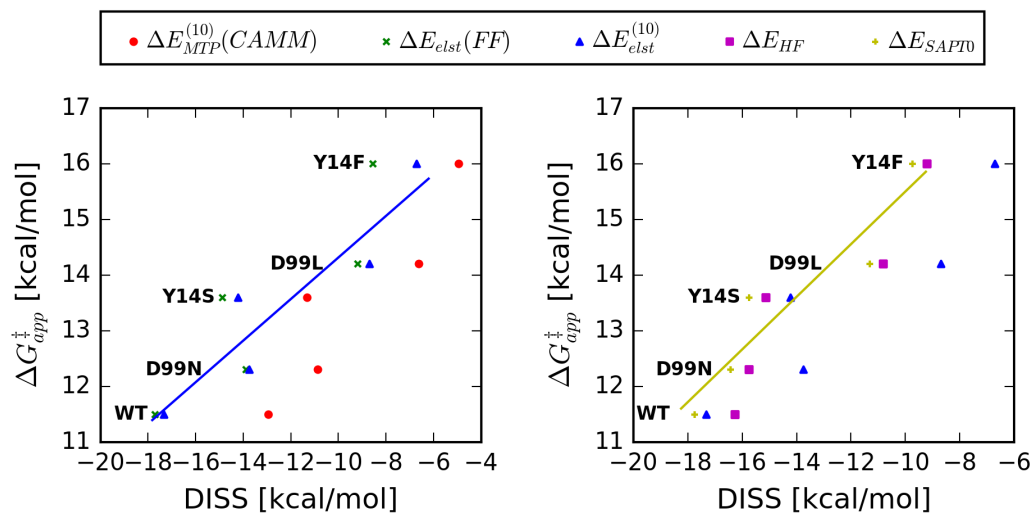
$$E_{SAPT0} = E_{el}^{(10)} + E_{ex}^{(10)} + E_{ind} + E_{disp} \quad (3.1)$$

$$E_{ind} = E_{ind,resp}^{(20)} + E_{ind-ex,resp}^{(20)} + \delta_{HF,resp} \quad (3.2)$$

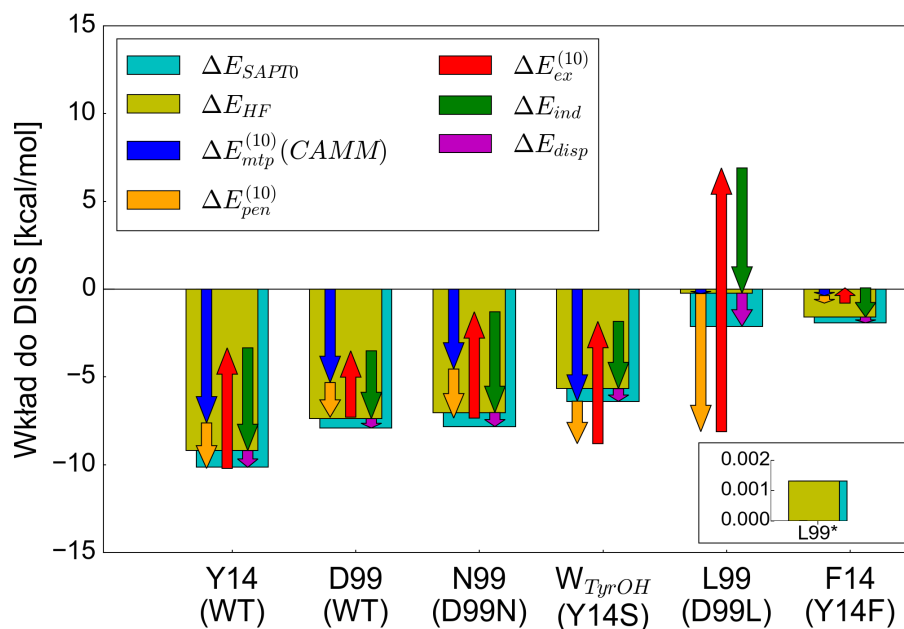
$$E_{disp} = E_{disp}^{(20)} + E_{disp-ex}^{(20)} \quad (3.3)$$

Rysunek 3.10 pokazuje relację pomiędzy DISS uzyskanym różnymi metodami a barierą eksperymentalną (struktury MD). Wspólną cechą wszystkich rozważonych metod jest jakościowa zgodność z eksperymentem oraz to, że od ogólnego trendu najbardziej odbiega mutant Y14S. Jak wspomniano, w centrum aktywnym tego mutantu donorem wiązania wodorowego w miejsce Tyr14 jest cząsteczka wody zajmująca to miejsce w kieszeni aktywnej; ze względu na pewną swobodę ruchu — większą niż w przypadku pozostałych analizowanych reszt — pochodzący od niej wkład entropowy do bariery reakcji może być istotnie większy. Wielkość tą można oszacować, biorąc pod uwagę, że  $T\Delta S^\ddagger$  w przypadku reakcji katalizowanych przez KSI i jon octanowy wynosi odpowiednio -4.2 and -7.5 kcal/mol [146]; w Y14S wartość ta jest zapewne bliższa tej pierwszej. W ten sposób można wyjaśnić, dlaczego bariera obliczona za pomocą DISS(CAMM) jest zaniżona o ok. 1 kcal/mol. Na wykresie można także zauważyć, że model oparty na ładunkach pola siłowego daje w przypadku Y14F wartość DISS znacznie odbiegającą zarówno od CAMM, jak i  $\Delta E_{elst}$ .

Pomimo zgodności przedstawionych wyników ze wspomnianą pracą [40], nie wnoszą one nic do dyskusji na temat udziału oddziaływań elektrostatycznych w całkowitym efekcie katalitycznym KSI



Rysunek 3.10: Porównanie wartości DISS otrzymanych różnymi metodami.



Rysunek 3.11: Dekompozycja energii oddziaływań w mutantach KSI (struktury „najmniejszego zaburzenia”). Człony energii oddziaływań zdefiniowane według schematu SAPT0 [117]; energię  $E_{pen}^{(10)}$  obliczono jako różnicę  $E_{el}^{(10)} - E_{mtp}^{(10)}(CAMM)$ . Wkład reszty seryny, bliski zeru, został pominięty na Rysunku; wstawka przedstawia analogiczny wykres dla zrelaksowanej konformacji Leu99.

[143, 144]; w cytowanej wymianie komentarzy głównym problemem jest definicja „katalizy elektrostatycznej”: Fried i Boxer definiują ją względem bariery zerowego pola, podczas gdy w roztworze wodnym wartość tych oddziaływań nie jest zerowa. Ponieważ jednak zasadniczym celem jest porównywanie mutantów z enzymem „macierzystym”<sup>7</sup>, problem układu odniesienia związanego z roztworem wodnym nie jest istotny — bariera zerowego pola odnosi się wówczas do teoretycznego enzymu nie wykazującego stabilizacji TS/IS.

Dwa najmniej aktywne z rozważanych tutaj mutantów KSI (D99L and Y14F) były analizowane również w pracach Chakravorty i Hammes-Schiffer [147] oraz Tore Brincka [148]. Pierwsza z nich, w której wykorzystano empiryczną metodę wiązań walencyjnych (ang. *Empirical Valence Bond*, EVB) [149], podkreśla znaczenie oddziaływań elektrostatycznych, jednakże wyniki te wykazują zaledwie jakościową zgodność z wartościami eksperymentalnymi; elektrostatyczna energia stabilizacji intermediatu (zarówno różnicowa, jak i „bezwzględna”<sup>8</sup>) nie odpowiada obserwowanemu trendowi. Przyczyną takiego stanu rzeczy jest inna definicja energii stabilizacji niż przyjęta tutaj: autorzy [147] brali pod uwagę sam reagent, podczas gdy przy obliczaniu DISS zarówno substrat, jak i katalityczna reszta Asp38 traktowane były łącznie (na wzór podziału układu na region QM i MM w metodach QM/MM). Drugą pracę, angażującą metodę FMO, trudniej porównać z zaprezentowanymi wynikami — autorzy wykorzystali w niej inną strukturę i substrat; niemniej wynik ten pozostaje jakościowo w zgodzie z DISS(CAMM): bariera aktywacji w D99L i Y14S, liczona względem dzikiej formy enzymu, wynosi, odpowiednio: 1.4 i 2.3 kcal/mol w przypadku CAMM i 4.3 i 5.1 kcal/mol w przypadku FMO, co wiąże się z różnicą pomiędzy metodami na poziomie ok. 3 kcal/mol.

Dla dopełnienia analizy oszacowano także wpływ reszt tyrozyny w pozycjach 55 i 30, uczestniczących w sieci wiązań wodorowych z Tyr14 [150, 151]. Korzystając z wyników dla struktur MZ (Rysunek 3.9), oszacowano bariery reakcji w mutancie Y55F; energia DISS(CAMM) wyniosła -11.6 kcal/mol (w porównaniu do -13.1 kcal/mol w dzikiej formie KSI), co przekłada się na barierę ok. 12.8 kcal/mol (wykorzystując równanie z panelu b Rysunku 3.9<sup>9</sup>), co znakomicie się zgadza z wartością eksperymentalną 12.7 kcal/mol [151]. W przypadku Tyr30, mutacja Y30F wiąże się z niewielką zmianą bariery (ok. 0.3 kcal/mol [151]), co znajduje odzwierciedlenie w znikomej wartości DISS(CAMM) dla tej reszty (ok. 0.03 kcal/mol).

#### Wniosek 4

*W przypadku KSI, model różnicowej stabilizacji intermediatu (DISS), dający wyniki równoważne DTSS, poprawnie opisuje wpływ analizowanych mutacji na aktywność katalityczną. Analiza składowej elektrostatycznej pozwala na ilościowe określenie wpływu mutacji na aktywność katalityczną, także mutantów nieujętych we wstępnych obliczeniach (Y55F i Y30F). Reprezentacja członu elektrostatycznego za pomocą multipoli CAMM jest konkurencyjna w stosunku do modelu opartego na ładunkach atomowych i metod uwzględniających inne człony energii oddziaływania, ze względu na względną dokładność (w analizowanym przykładzie wykazuje ona wręcz ilościową zgodność z eksperymentem), niski koszt obliczeniowy i mniejszą podatność na zaburzenia strukturalne.*

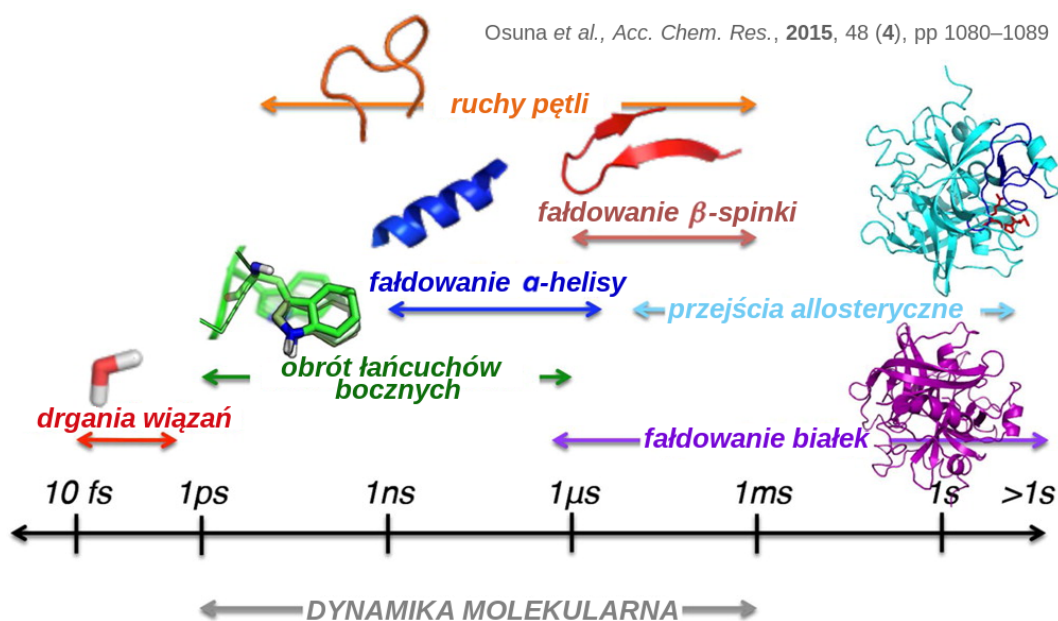
### 3.2.2 Analiza eliminazy Kempa KE07 i jej mutantów

Znacznie ciekawszym zagadnieniem jest wyjaśnienie mutacji wprowadzanych do teozymów w drodze eksperymentów kierunkowanej ewolucji (LDE); równocześnie problem jest istotny z punktu widzenia projektowania enzymów. Spośród różnych teozymów prawdopodobnie najlepiej zbadanym są eliminazy Kempa (katalizujące reakcję wspomnianą w Rozdziale 1 — *vide* Rysunek 1.2), a wśród nich szczególnie rodziny enzymów znalezione drogą LDE na podstawie teozymów KE07 [85] i HG3 [86]. Jak wspomniano we sekcji 1.3.3, podjęto dotąd szereg prób mających wyjaśnić rolę znalezionych mutacji, jednakże zastosowane w nich techniki wymagają nakładu obliczeniowego, który pozwala je zastosować co najwyżej do

<sup>7</sup> Tutaj jest to dzika forma KSI, jednakże w kontekście projektowania enzymów oznaczać to będzie np. wstępnie zaprojektowany teozym, który ma zostać zoptymalizowany.

<sup>8</sup> Tzn. wyliczona z prawa Coulomba.

<sup>9</sup> W podanym oszacowaniu uwzględniono fakt, że wkład od reszty Tyr55 nie był obecny w obliczeniach, na podstawie których wyznaczono to równanie. W modelu „najmniejszego zaburzenia” dodanie reszty wspólnej dla wszystkich wariantów enzymu oznacza przesunięcie wykresu o stałą wartość; dlatego DISS dla Y55F wyznaczono, dodając do wartości dla dzikiej formy enzymu różnicę wkładów Phe55 i Tyr55.



Rysunek 3.12: Schemat przedstawiający skale czasowe różnych ruchów występujących w białkach [97]. Czas na osi  $OX$  zaprezentowany jest w skali logarytmicznej; widać, że nawet symulacja długości 30 ns może nie uchwycić pewnych zmian konformacji łańcuchów bocznych (tłumaczenie i przedruk za zgodą Amerykańskiego Towarzystwa Chemicznego).

selekcji końcowego zbioru kandydatów. Alternatywą jest prezentowana tutaj metoda oparta na koncepcji DTSS, jak wspomniano już wcześniej; eliminaza Kempa stanowi dla niej dobry poligon testowy. Ze względu na dużą ilość mutantów, z wyszczególnieniem tych, w których zmieniano ilość i rozmieszczenie aminokwasów naładowanych (Tabela 3.2), wybrano rodzinę teozymu KE07. Dla macierzystego enzymu znaleziono stan przejściowy i ścieżkę reakcji metodą ONIOM (B3LYP/6-311+G\*\*:*Amber94*), bazując na strukturze otrzymanej w prywatnej korespondencji od Anastazji Alexandrowej. Jest to w zasadzie analogiczne podejście do pracy Houka i Bakera [91], z tą różnicą, że wykorzystano większą bazę funkcyjną i osadzenie mechaniczne<sup>10</sup>; dla stanu przejściowego dodatkowo zweryfikowano strukturę za pomocą procedury *refitowania ładunków* regionu QM, zgodnie z zaleceniami Tao i współpracowników [135], jednakże zmiany w strukturze były subtelne (*vide* sekcja S.5 w Suplemencie). Uzyskane geometrie substratu i TS zostały wykorzystane do obliczenia momentów multipolowych CAMM (HF/6-31Gd) dla obu struktur oraz analiz z wykorzystaniem biblioteki rotamerów aminokwasów. Podejście to jest motywowane kilkoma obserwacjami. Po pierwsze, dotychczasowe wyniki wskazują na to, że do zaprojektowania teozymu wystarczy znajomość *przybliżonej* struktury stanu przejściowego i jego rozkładu ładunku; przykładowo, fosfoesterazę PT3 [75] zaprojektowano w oparciu o stan przejściowy z pracy dr Edyty Dygudy-Kazimierowicz [152], w którym zmieniono tylko odpowiednie podstawniki; z kolei w symulacjach EVB [149] stanami referencyjnymi są substrat i produkt, zaś rozkład ładunku TS jest odpowiednią ich kombinacją liniową. Kolejnym czynnikiem jest skala czasowa różnych ruchów występujących w białku (Rysunek 3.12; uchwycenie właściwej orientacji łańcuchów bocznych aminokwasów może być trudne mimo długich symulacji dynamiki molekularnej (MD)). Stąd zaprezentowana w następnej sekcji analiza wkładów DTSS różnych konfiguracji rotamerów, zgodna z przyjętą w tej pracy strategią *oddolną*, może stanowić przydatne narzędzie.

### Zastosowanie biblioteki multipoli rotamerów aminokwasów

W obliczeniach analizowano trzynaście pozycji, w których wprowadzano mutacje w eksperymentach ukierunkowanej ewolucji (Tabela 3.2) [85]. W każdej z pozycji, w każdej mutancie, kolejno wczytywano poszczególne rotamery odpowiednich aminokwasów z bazy, eliminując te, których najkrótsza

<sup>10</sup> Ze względu na problemy ze zbieżnością modelu z osadzeniem elektronowym.

Tabela 3.2: Mutacje wprowadzone do teozymu KE07 w eksperymentach kierunkowanej ewolucji. Reszty naładowane zaznaczono kolorem w zależności od ich ładunku formalnego (czerwony – dodatni, niebieski – ujemny).

Enzym	Oznaczenie	$k_{cat}$	Reszta													
			7	12	19	77	84	86	102	123	146	202	207	224	229	
KE07	P <sub>0</sub>	0.018	Ile	Val	Lys	Phe	His	Phe	Ile	Gln	Lys	Gly	Met	Asn	Phe	
R2-11/10D	P <sub>1</sub>	0.0213	Ile	Val	Glu	Phe	His	Phe	Ile	Arg	Thr	Arg	Met	Asp	Phe	
R3-13/10A	P <sub>2</sub>	0.206	Gln	Val	Lys	Phe	His	Leu	Ile	Gln	Thr	Arg	Met	Asp	Ser	
R4-1E/11H	P <sub>3</sub>	0.699	Asp	Val	Lys	Phe	His	Phe	Ile	Gln	Glu	Arg	Met	Asp	Phe	
R5-10/3B	P <sub>4</sub>	0.49	Asp	Met	Lys	Phe	His	Phe	Ile	Gln	Lys	Arg	Met	Asp	Phe	
R6-3/7F	P <sub>5</sub>	0.6	Asp	Val	Glu	Phe	His	Phe	Ile	Gln	Thr	Arg	Met	Asp	Phe	
R7-1/3H	P <sub>6</sub>	0.76	Asp	Leu	Lys	Ile	Tyr	Phe	Ile	Gln	Lys	Arg	Thr	Asp	Phe	
R7-10/11G	P <sub>7</sub>	1.37	Asp	Met	Thr	Ile	His	Phe	Phe	Gln	Lys	Arg	Met	Asp	Phe	

odległość od reagenta wynosiła mniej niż 2Å; dla każdego rotameru wyznaczono energię oddziaływania ze stanem przejściowym (*energię stabilizacji stanu przejściowego*, TSS) oraz różnicowej stabilizacji stanu przejściowego (DTSS). Ostatecznie dla każdej mutacji wybrano rotamery o ekstremalnych wartościach tych energii; w dalszej części pracy procedura ta określana będzie mianem *skanu* lub *skanu po rotamerach reszt aminokwasowych*. Wyniki przedstawiają Tabele 3.3 i 3.4.

Tabela 3.3: DTSS- najważniejsze rotamery ( $|E_{DTSS}| > 0.5$  kcal/mol). W tabeli podano aminokwas, pozycję w sekwencji, skrajne wartości DTSS oraz odpowiadające im rotamery (zapis zgodny z definicją w bazie Dymameonics [131]).

		MIN	MAX	Zakres
Arg	202	-4.66 g-, g+, g-, t	-0.98 g+, g+, t, t	3.68
Lys	146	-2.33 g-, g+, t, t	-0.89 g+, g+, t, t	1.45
Asp	7	-1.70 t, t	0.02 g+, g+	1.72
Asn	224	-1.00 g-, Og-	0.83 g-, Ng-	1.83
Lys	19	-0.93 g+, g-, t, g+	-0.41 t, g+, g-, t	0.51
Gln	7	-0.93 t, t, Ot	0.69 t, g-, Nt	1.62
Phe	77	-0.58 t, g	-0.48 g-, t	0.10
Ile	77	-0.53 t, g+	-0.48 g-, g-	0.05
Asp	224	1.31 t, g+	2.88 g-, g-	1.57
Glu	19	0.40 t, g+, t	0.77 g+, g-, t	0.37
Arg	123	0.69 g-, t, t, t	1.77 t, g-, g+, t	1.08
Glu	146	0.80 t, g-, g-	1.91 g-, g+, g+	1.12

Można zauważyć, że spośród trzynastu pozycji, w najaktywniejszej wersji enzymu cztery aminokwasy są naładowane elektrycznie i zachowane od kilku rund eksperymentu. Trzy z nich (Arg202, Asp7, Lys146) wykazują wartości DTSS wskazujące na obniżenie bariery. Równocześnie są to reszty o najbardziej stabilizujących wkładach o module większym od 1.5 kcal/mol. Dwie pozycje charakteryzują się mniej sprzyjającym wynikiem. Wartość DTSS Lys19 zdaje się sprzyjać obniżeniu bariery, jednakże aminokwas ten występuje w tej pozycji tylko w kilku pośrednich mutantach, ostatecznie zastąpiony przez treoninę. Natomiast w przypadku Asp224 wartość DTSS sugeruje zwiększenie bariery aktywacji. Niezgodność tego cząstkowego wyniku z eksperymentem oznacza, że prawdopodobnie aminokwas ten wpływa na aktywność w inny sposób.

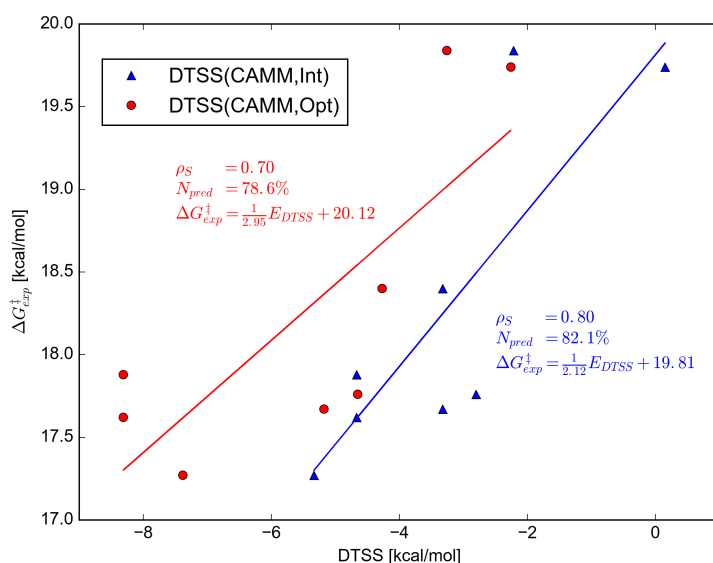
Aby to zbadać, powtórzono procedurę z uwzględnieniem wzajemnego oddziaływania pomiędzy resztami naładowanymi, pomijając pozostałe aminokwasy; dla każdego białka przeanalizowano w ten sposób wszystkie kombinacje rotamerów uwzględnionych aminokwasów (eliminując te, dla których najmniejsza odległość od łańcucha głównego, reagentów lub innej z uwzględnionych reszt była mniejsza od 2Å), wybierając tę, dla której energia elektrostatycznego układu (w obecności substratu) była najniższa<sup>11</sup>. Aby dokonać tego w rozsądnym czasie (dla P<sub>6</sub> bowiem wiąże się to z koniecznością przeanalizowania ok. 21 milionów kombinacji), konieczne było zastąpienie multipoli CAMM ładunkami atomowymi: pola siłowego Amber94 w przypadku aminokwasów i Merza-Kollmana [137] w przypadku

<sup>11</sup> Poczyniono tutaj założenie, że związany z zajściem reakcji proces przeniesienia protonu jest szybszy od obrotu reszt aminokwasowych, kiedy już następuje.

Tabela 3.4: TSS- najważniejsze rotamery ( $|E_{TSS}| > 1.0$  kcal/mol). W tabeli podano aminokwas, pozycję, skrajne wartości TSS oraz odpowiadające im rotamery (zapis zgodny z definicją w bazie Dynameonics [131]).

		MIN	MAX	Zakres
Arg	202	-46.01 g-, g+, t, t	-23.02 t, g-, t, t	22.99
Lys	146	-30.43 g-, g+, g+, t	-19.36 t, g-, t, t	11.08
Arg	123	-22.60 t, g-, g+, g+	-15.36 g+, g-, t, t	7.24
Lys	19	-22.07 g+, g-, t, g-	-14.96 t, g+, t, t	7.11
Asn	224	-4.62 g+, Og+	3.60 g-, Og+	8.22
Gln	7	-3.54 t, t, Og-	5.85 t, t, Ng-	9.39
Met	12	-2.00 g+, g+, t	-0.32 g-, g+, t	1.68
His	84	-1.51 t, Cg-	1.20 g-, Nt	2.71
Leu	12	-1.43 g+, g+	-1.25 t, g+	0.18
Thr	146	-1.39 g-	-0.82 g+	0.57
Val	12	-1.31 g+	-1.28 g-	0.02
Tyr	84	-1.31 g-, t	-0.31 g-, g	1.00
Gln	123	0.08 t, g-, Ng+	2.18 t, g+, Ng+	2.10
Phe	77	2.07 t, t	2.57 t, g	0.50
Ile	77	2.23 g-, g-	2.44 t, g+	0.21
Glu	19	14.88 t, g+, t	19.86 g+, g-, g-	4.97
Glu	146	18.72 t, g-, g+	27.96 g+, g-, g-	9.24
Asp	224	28.66 t, g+	42.42 g-, g-	13.76
Asp	7	31.05 g-, g-	42.36 t, g+	11.31

reagentów. Wiąże się to z pewną utratą dokładności, pomimo niezerowego ładunku układu reagującego (*vide* Rysunek S.10 w Suplemencie), jednak nie aż tak duża, by spodziewać się innej wynikowej kombinacji rotamerów. Dla znalezionych w ten sposób rotamerów wykorzystano odpowiednie energie CAMM uzyskane w poprzednim skanie. Porównanie wyników obu analiz z eksperymentalnymi aktywnościami enzymów przedstawia Rysunek 3.13. Zauważyć na nim można, że uwzględnienie wzajemnego odpychania się reszt naładowanych poprawiło opis najaktywniejszych mutantów (w których takich reszt było najwięcej). Prześledzenie wkładów DTSS aminokwasów w poszczególnych pozycjach (Tabela 3.5) pozwala określić jakościowe znaczenie reszty Asp224 i mutacji Lys19Thr; pierwsza reszta ułatwia optymalną orientację Arg202, mającej największy wkład katalityczny- jej usunięcie powoduje zmianę wkładów Lys146 i Arg202 wskutek ich wzajemnego odpychania (*vide* Rysunek 3.14); reszta Lys19, pomimo stabilizującej wartości DTSS, wymuszała suboptymalne położenie Lys146 (Tabela 3.5). Można to jeszcze bardziej uwidocznic, rozważając hipotetycznego mutantu Asp224Thr (P<sub>8</sub>); usunięcie reszty kwasu asparaginowego z tej pozycji (*vide* Tabela 3.5) powoduje zmianę konformacji Arg202 na korzyść Lys146; jednakże w modelu *in vacuo* korzystniejsza orientacja Arg202 nie równoważy destabilizującego wpływu Asp224.



Rysunek 3.13: Porównanie DTSS(CAMM) dla zbioru optymalnych rotamerów (Opt) i konfiguracji uwzględniającej oddziaływanie pomiędzy aminokwasami (Int).

Efekt ten można wyjaśnić, wprowadzając model ciągłego rozpuszczalnika. Zaletą tego zabiegu jest to, że pozwala on na oszacowanie wpływu solwatacji na elektrostatyczną część DTSS za pomocą efektywnej stałej dielektrycznej (*vide* sekcja S.7 w Suplemencie<sup>12</sup>). Wprowadzenie tego elementu do procedury przesiewania zbioru rotamerów modyfikuje oddziaływania pomiędzy bardziej odległymi resztami, jak np. Lys19 i Lys146: brak tego pierwszego aminokwasu już nie wpływa bezpośrednio na Lys146, lecz Arg202 (Tabela 3.6), która leży bliżej. Oddziaływanie blisko leżących reszt Asp224 i Arg202 dalej jednak pozostaje, z tą różnicą, że dzięki uwzględnieniu rozpuszczalnika w opisany sposób brak reszty kwasu asparaginowego rzeczywiście wiąże się z gorszą sumaryczną stabilizacją.

Podsumowując, wykorzystanie biblioteki rotamerów i elektrostatycznego przybliżenia DTSS pozwoliło wykazać znaczenie naładowanych reszt zoptymalizowanych drogą eksperymentów LDE w enzymie P<sub>7</sub>. Spośród sześciu pozycji, w których wprowadzano reszty tego typu, trzy (Glu19, Arg123 i Glu146) miały negatywny wkład DTSS i zostały zastąpione innymi aminokwasami; ich znaczenie dla danego mutantu musiało wiązać się z innym czynnikiem lub musiały być uwikłane w bardziej złożone korelacje, niż ujęte w zaprezentowanej analizie. Równocześnie jednak nie wszystkie reszty prezentujące stabilizu-

<sup>12</sup> W zaprezentowanych tutaj rozważaniach wykorzystano efektywne promienie Borna obliczone za pomocą programu *bornRadius* opracowanego w zespole Alexieja Onufrieva [153,154]



Tabela 3.5: Wkłady DTSS (CAMM) reszt aminokwasowych z kombinacji rotamerów o najniższej energii dla analizowanych białek (uszeregowanych według rosnącej aktywności katalitycznej). Dla jasności nazwy aminokwasów podano w kodzie jednoliterowym, oznaczając je kolorem zgodnym z ich ładunkiem formalnym (czerwony-dodatni, niebieski-ujemny).

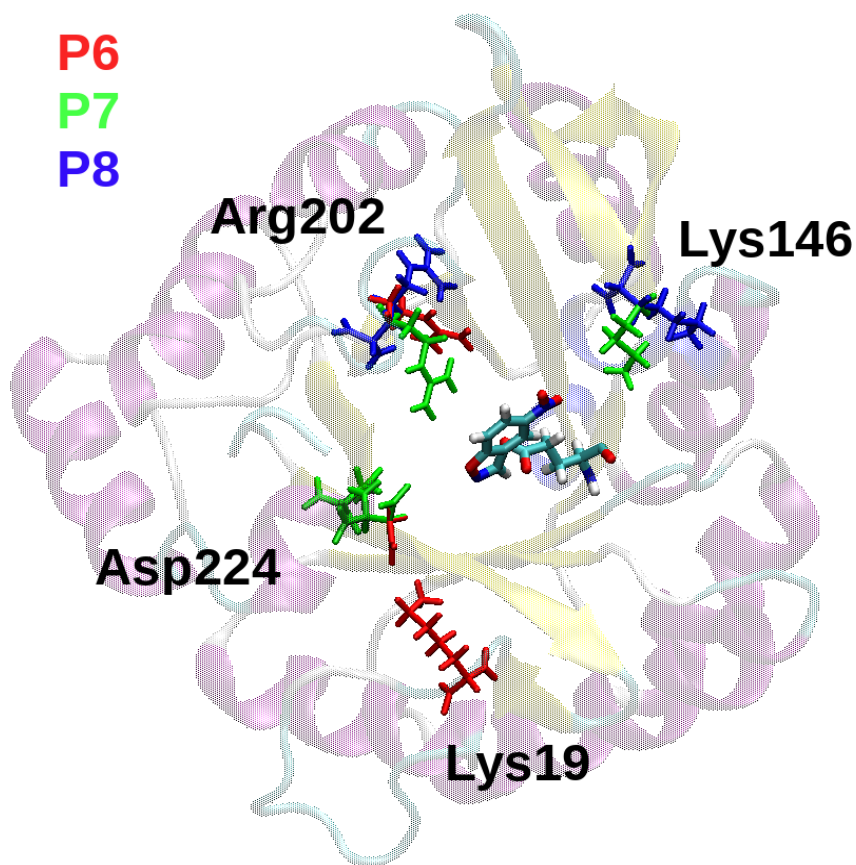
Enzym	Reszta										Suma		
	7		19		123		146		202			224	
P <sub>0</sub>	I	0.00 <sup>1</sup>	<b>K</b>	-0.85	Q	0.00	<b>K</b>	-1.36	G	0.00	N	0.00	-2.21
P <sub>1</sub>	I	0.00	<b>E</b>	0.40	<b>R</b>	1.75	T	0.00	<b>R</b>	-4.66	<b>D</b>	2.66	0.15
P <sub>2</sub>	Q	0.00	<b>K</b>	-0.93	Q	0.00	T	0.00	<b>R</b>	-4.66	<b>D</b>	2.26	-3.32
P <sub>4</sub>	<b>D</b>	0.02	<b>K</b>	-0.93	Q	0.00	<b>K</b>	-1.36	<b>R</b>	-4.66	<b>D</b>	2.26	-4.66
P <sub>5</sub>	<b>D</b>	-1.12	<b>E</b>	0.40	Q	0.00	T	0.00	<b>R</b>	-4.66	<b>D</b>	2.66	-2.80
P <sub>3</sub>	<b>D</b>	-1.12	<b>K</b>	-0.93	Q	0.00	<b>E</b>	1.19	<b>R</b>	-4.66	<b>D</b>	2.26	-3.32
P <sub>6</sub>	<b>D</b>	0.02	<b>K</b>	-0.93	Q	0.00	<b>K</b>	-1.36	<b>R</b>	-4.66	<b>D</b>	2.26	-4.66
P <sub>7</sub>	<b>D</b>	-1.12	T	0.00	Q	0.00	<b>K</b>	-2.13	<b>R</b>	-4.66	<b>D</b>	2.66	-5.33
P <sub>8</sub>	<b>D</b>	-1.19	T	0.00	Q	0.00	<b>K</b>	-3.18	<b>R</b>	-2.78	T	0.00	-7.14

1) Dla reszt nienaładowanych przyjęto wartość DTSS równą zero

Tabela 3.6: Wkłady DTSS (ładunki atomowe i efektywne stałe dielektryczne) reszt aminokwasowych z kombinacji rotamerów o najniższej energii dla analizowanych białek (uszeregowanych według rosnącej aktywności katalitycznej). Dla jasności nazwy aminokwasów podano w kodzie jednoliterowym, oznaczając je kolorem zgodnym z ich ładunkiem formalnym (czerwony-dodatni, niebieski-ujemny).

Enzym	Reszta										Suma		
	7		19		123		146		202			224	
P <sub>0</sub>	I	0.00 <sup>1</sup>	<b>K</b>	-0.12	Q	0.00	<b>K</b>	-0.60	G	0.00	N	0.00	-0.72
P <sub>1</sub>	I	0.00	<b>E</b>	0.04	<b>R</b>	0.40	T	0.00	<b>R</b>	-1.95	<b>D</b>	0.73	-0.78
P <sub>2</sub>	Q	0.00	<b>K</b>	-0.15	Q	0.00	T	0.00	<b>R</b>	-0.98	<b>D</b>	0.21	-0.92
P <sub>4</sub>	<b>D</b>	-0.24	<b>K</b>	-0.15	Q	0.00	<b>K</b>	-0.89	<b>R</b>	-0.70	<b>D</b>	0.21	-1.75
P <sub>5</sub>	<b>D</b>	-0.47	<b>E</b>	0.04	Q	0.00	T	0.00	<b>R</b>	-1.95	<b>D</b>	0.73	-1.65
P <sub>3</sub>	<b>D</b>	-0.24	<b>K</b>	-0.15	Q	0.00	<b>E</b>	0.33	<b>R</b>	-1.95	<b>D</b>	0.21	-1.79
P <sub>6</sub>	<b>D</b>	-0.24	<b>K</b>	-0.15	Q	0.00	<b>K</b>	-0.89	<b>R</b>	-0.70	<b>D</b>	0.21	-1.75
P <sub>7</sub>	<b>D</b>	-0.47	T	0.00	Q	0.00	<b>K</b>	-0.89	<b>R</b>	-1.95	<b>D</b>	0.73	-2.53
P <sub>8</sub>	<b>D</b>	-0.24	T	0.00	Q	0.00	<b>K</b>	-0.89	<b>R</b>	-0.70	T	0.00	-1.82

1) Dla reszt nienaładowanych przyjęto wartość DTSS równą zero



Rysunek 3.14: Porównanie rotamerów w P<sub>6</sub> (kolor czerwony), P<sub>7</sub> (zielony) i hipotetycznym P<sub>8</sub> (niebieski), uzyskanych na drodze skanu rotamerów z uwzględnieniem ich wzajemnego odpychania (model GB). Widać, że obecność reszty Lys19 powoduje przesunięcie w jej kierunku Asp224 i oddalenie się Arg202 od reagenta; brak Asp224 (P<sub>8</sub>) również skutkuje suboptymalnym położeniem tej reszty.

jący wkład DTSS zostały zachowane w P<sub>7</sub>, jak również nie wszystkie występujące w tym białku reszty naładowane wykazują efekt stabilizujący; z pośród czołowej czwórki aminokwasów (Asp7, Lys19, Lys146 i Arg202), Lys19 jest nieobecna (co wiąże się z dwukrotnym zwiększeniem stałej szybkości  $k_{cat}$ ), gdyż jej odpychanie z pozostałymi resztami uniemożliwia przyjęcie optymalnych dla nich konformacji. Reszta Asp224, pomimo destabilizującego wkładu DTSS, okazuje się konieczna dla stabilizacji korzystnej konformacji Arg202, mającej ze wszystkich wymienionych reszt najsilniejszy wpływ.

### Symulacje dynamiki molekularnej

Rozważania poprzedniej sekcji porównano z wynikami symulacji Dynamiki Molekularnej (MD) wszystkich 8 wariantów białka (w kompleksie z substratem); dla każdego z nich uzyskano 30 ns trajektorii fazy produkcji w zespole NPT. Symulacje przeprowadzono za pomocą programu Gromacs 4.6.4 [155], wykorzystując pole siłowe Amber94 i model wody TIP3P oraz ładunki TS i substratu wspomniane we wcześniejszym teście. Wybór pola siłowego był podyktowany wcześniejszymi obliczeniami ONIOM. Następnie oszacowano energie DTSS (wraz z dekompozycją na wkłady od różnych reszt aminokwasowych) za pomocą metody MMPBSA<sup>13</sup> za pomocą programu *g\_mmpbsa* [156,157], uwzględniając tylko wkład *polarny* (oddziaływania elektrostatyczne wraz elektrostatyczną częścią energii solwatacji).

Dla każdej trajektorii przeanalizowano rotamery, jakie przyjmują analizowane aminokwasy. Okazuje się, że nie tylko dystrybucje są różne dla tego samego aminokwasu w różnych trajektoriach, ale zdecydowanie inne od tych z bazy Dymameonics [131]. Przedstawiono to na przykładzie Arg202, która w Tabeli 3.3 odznaczała się zarówno najbardziej stabilizującą energią DTSS, jak i największym zakresem jej zmian. Zestawienie umieszczono na Rysunku 3.15.

Na pierwszy rzut oka wydaje się, że w enzymach P<sub>1</sub>, P<sub>7</sub> i P<sub>5</sub> Arg202 przyjmuje dramatycznie różne orientacje niż w pozostałych białkach. Jest to jednak tylko pozorna niezgodność; główne rotamery we wszystkich trajektoriach odpowiadają położeniu łańcucha bocznego w kierunku rozpuszczalnika, nad kanałem wiążącym substrat (podobnie do zielonej orientacji Arg202 na Rysunku 3.14). Różnica wynika z obrotu łańcucha głównego- Arg202 leży bowiem w obrębie pętli łączącej beta-harmonijkę z alfa helisą.

Prawdą jest natomiast, że każda z 7 dystrybucji jest zdominowana przez pojedynczy rotamer, podczas gdy w bazie przodują cztery rotamery. Może to być z jednej strony zwiastunem konieczności użycia wersji bazy, w której populacje rotamerów są zależne od konformacji łańcucha głównego (ang. *backbone-dependent*, niewykorzystana tutaj), z drugiej zaś oznaka tego, że nawet 30 ns symulacja może nie wystarczać, by należycie uchwycić swobodę konformacyjną łańcuchów bocznych aminokwasów takich jak Arg202.

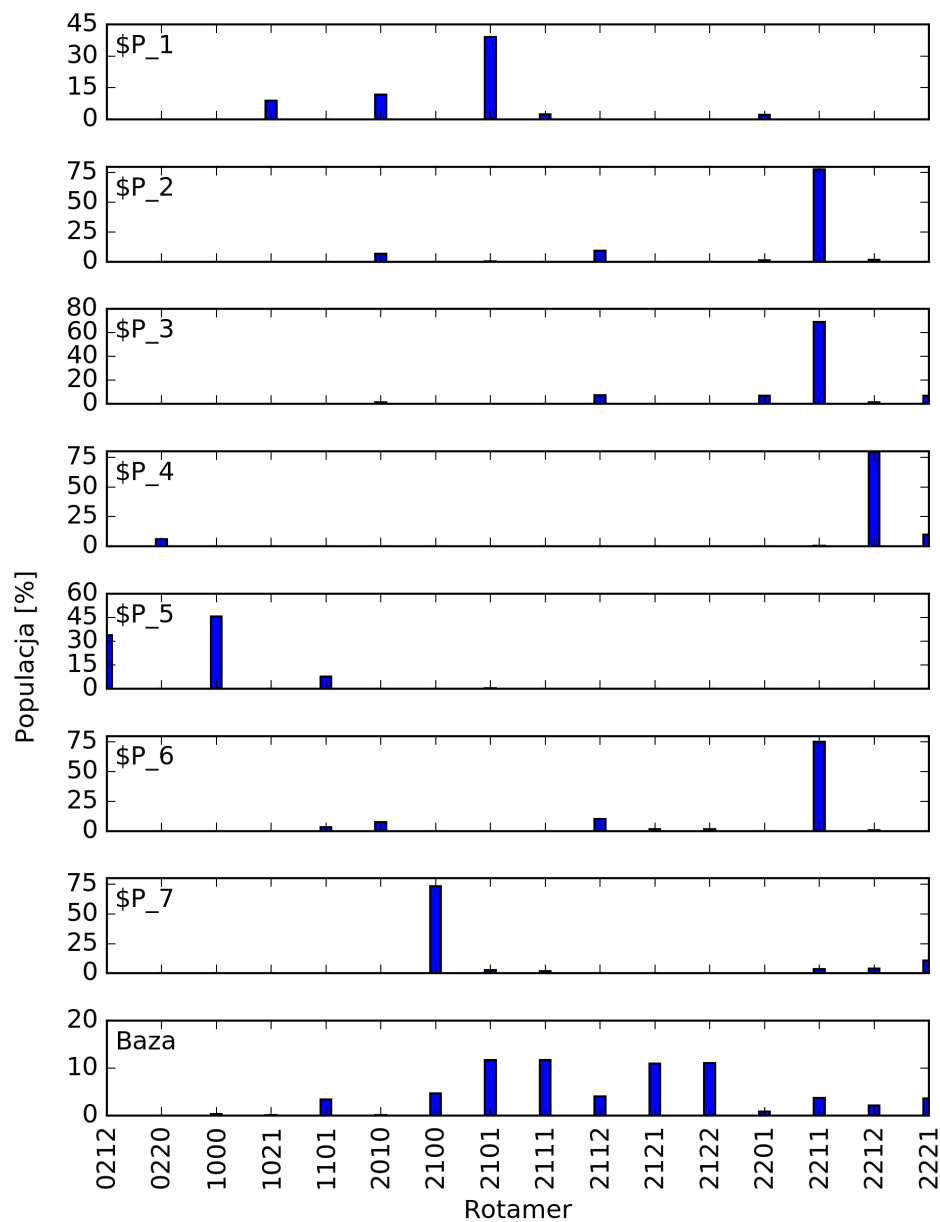
Skupiając się na najaktywniejszym enzymie (P<sub>7</sub>), można zauważyć, że do najbardziej stabilizujących aminokwasów należą te same reszty, co w Tabeli 3.3; choć wartości liczbowe różnią się zarówno od liczonych *in vacuo* energii CAMM jak i wartości uwzględniających rozpuszczalnik w bardziej przybliżony sposób (model Borna), można je uznać za jakościowo zgodne.

Tabela 3.7: Wkłady DTSS(MMPBSA) mutowanych aminokwasów P<sub>7</sub>

Aminokwas		DTSS(MMPB) kcal/mol
Asp	7	-5.302
Arg	202	-3.650
Lys	146	-2.258
Asp	224	3.819

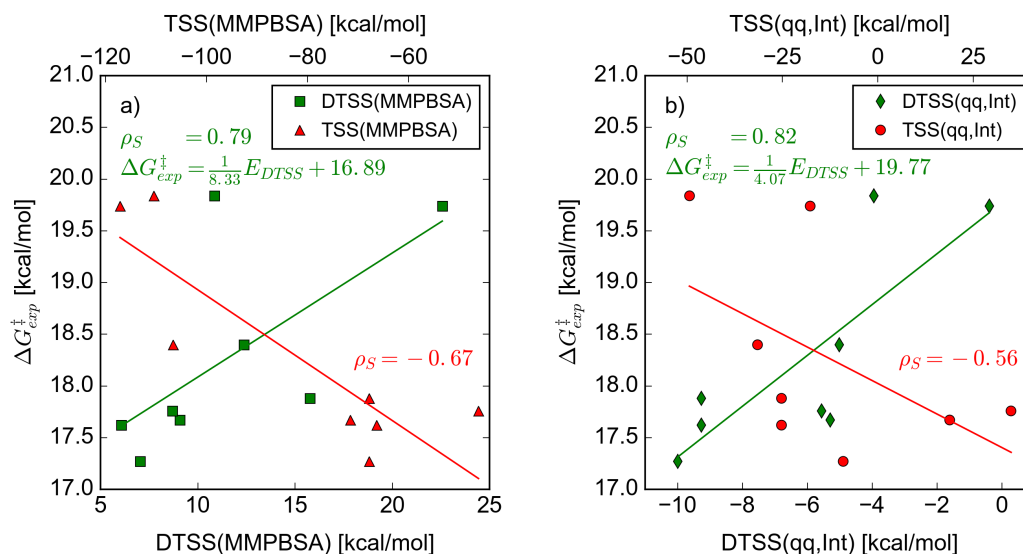
Pod względem zgodności z eksperymentalnymi barierami, pominięcie oddziaływania między aminokwasami przy wyborze rotamerów prowadzi do wyników znacznie gorzej odpowiadających wartościom eksperymentalnym (Rysunki 3.16 i 3.17). Jeżeli za miarę zgodności przyjąć współczynniki korelacji

<sup>13</sup> Popularny sposób wyznaczania energii wiązania ligandów przez białka; rozszerza on opis za pomocą mechaniki molekularnej (MM) o model ciągłego rozpuszczalnika: równanie Poissona Boltzmanna (PB) lub uogólniony model Borna (ang. *Generalized Born*, GB) oraz empiryczny człon zależny od powierzchni dostępnej rozpuszczalnikowi (ang. *Solvent Accessible Area*, SA).



Rysunek 3.15: Populacje rotamerów Arg202 wzdłuż trajektorii mutantów oraz w bazie Dynamics. Dla czytelności symbole rotamerów zapisano za pomocą cyfr:  $g+ \rightarrow 0, t \rightarrow 1, g- \rightarrow 2$ , np.  $g-, g+, g-, t \rightarrow 2211$ .

Pearsona ( $\rho_P$ ) i Spearmana ( $\rho_S$ )<sup>14</sup>, dla zestawu optymalnych rotamerów (w którym pominięto oddziaływania pomiędzy aminokwasami)  $\rho_P = 0.79$  i  $\rho_S = 0.70$ , natomiast uwzględnienie oddziaływań pomiędzy aminokwasami prowadzi do wartości, odpowiednio, 0.83 i 0.80 (w przypadku modelu GB nawet do 0.89 i 0.93)<sup>15</sup>. Jak widać na Rysunku 3.17, dotyczy to szczególnie najaktywniejszych z analizowanych enzymów (w których naładowanych aminokwasów było najwięcej). Dodać także można, że bariera „zerowego DTSS” (punkt przecięcia z osią  $OY$  na Rysunkach 3.13 i 3.17) zgadza się w obu przypadkach z „barierą własną” ok. 20.6 kcal/mol<sup>16</sup>. Warto też porównać wartości DTSS z energiami stabilizacji stanu przejściowego (TSS), na których są oparte konwencjonalne modele analizy aktywności katalitycznej (Rysunek 3.16); TSS wykazuje *antykorrelację* z eksperymentem: *bardziej* aktywny mutant *w mniejszym stopniu* stabilizuje stan przejściowy. Jest to w zgodzie z ostatnią obserwacją zespołu Warshela [100], według której ewolucyjna ścieżka w tej rodzinie eliminaz Kempa wiedzie drogą rosnącej destabilizacji substratu (GSD), w przeciwieństwie do rodziny HG3, w której kolejne mutanty wykazują większą stabilizację TS. Konieczność uwzględnienia korelacji pomiędzy resztami zgadza się także z ostatnią pracą zespołu Teresy Head-Gordon [98], aczkolwiek wydaje się, że przedstawiona analiza konfiguracji aminokwasowych jest bardziej przejrzysta; trzeba także zaznaczyć, że obecne rozważania mają charakter „entalpowy” w tym sensie, że entropia związana z dystrybucją poszczególnych struktur w zespole termodynamicznym nie jest analizowana<sup>17</sup>.



Rysunek 3.16: Porównanie DTSS i TSS. a) MMPBSA, b) Konfiguracje rotamerów znalezione w procedurze skanowania bazy rotamerów z uwzględnieniem oddziaływania pomiędzy resztami (model ładunkowy).

### Przewidywanie wpływu mutacji na aktywność katalityczną

Bazując na prezentowanych w tej pracy wynikach, zarys algorytmu znajdowania nowych mutacji, mających poprawić aktywność katalityczną, można przedstawić w sposób następujący.

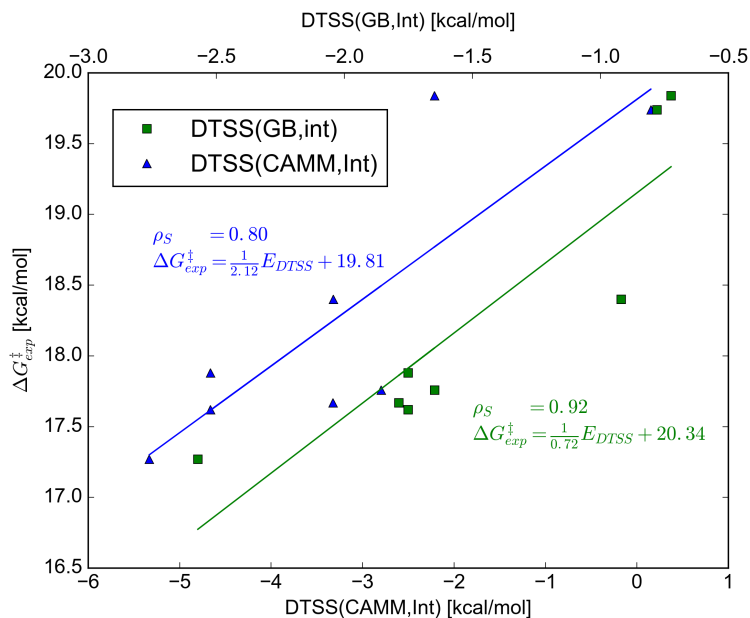
1. Określ miejsca w sekwencji, w których potencjalnie można wprowadzić naładowany aminokwas

<sup>14</sup> Tzw. współczynnik korelacji rangowej; można go uznać za miarę monotonicznej zależności pomiędzy zmiennymi (współczynnik Pearsona określa zależność liniową).

<sup>15</sup> W sensie statystycznym odpowiada to w przybliżeniu przejściu z poziomu istotności  $2\sigma$  na  $3\sigma$ .

<sup>16</sup> Jest to różnica energii elektronowych TS i substratu; energię punktu zerowego pominięto.

<sup>17</sup> Podjęto próbę rozszerzenia DTSS w kierunku opisu energii swobodnej (*vide* Supplement, sekcja S.4, lecz nie uzyskano wyników mogących zostać zastosowanymi do prezentowanej analizy mutantów).



Rysunek 3.17: Porównanie DTSS(CAMM) i DTSS(GB) z eksperymentalnymi barierami aktywacji. Zestawy rotamerów wyznaczone były z uwzględnieniem oddziaływania pomiędzy aminokwasami (Int).

o korzystnym DTSS. W tym celu np. wybierz te reszty, dla których wartość pola katalitycznego  $\Delta_S$  na atomie węgla  $C_\beta$  ma moduł większy od 1 kcal/(mol·e).

2. Pomiń te pozycje, które mogą zawierać reszty istotne z innego powodu (np. uczestniczące w jakimś ważnym elemencie strukturalnym białka).
3. Prześledź wkłady DTSS(CAMM) rotamerów aminokwasów o ładunku zgodnym z  $\Delta_S$ ; eliminuj te rotamery, które wchodzą w konflikty steryczne z otoczeniem.
4. Mając zestaw potencjalnych kandydatów, prześledź ich wzajemne oddziaływania.

Jeśli chodzi o ostatni punkt, pozostaje kwestią do rozważenia w późniejszych badaniach, jak uwzględnić potencjalne mutacje niezgodne z  $\Delta_S$ , ale korzystne ze względu na oddziaływania z innymi resztami, na podobieństwo omówionego w poprzedniej sekcji przykładu Asp224. Być może należałoby dodatkowo rozważyć miejsca w sekwencji w otoczeniu aminokwasu o najsilniejszym wkładzie DTSS. Inną kwestią, którą można podjąć w dalszych dociekaniach, jest wybór efektywnych stałych dielektrycznych  $\epsilon_{ij}$  pomiędzy różnymi miejscami białka. Na przykładzie KE07, wytypowano w ten sposób pozycje o numerach 171 (znak pola sugerujący aminokwas naładowany dodatnio) i 99 (znak pola sugerujący aminokwas naładowany ujemnie), w których jak dotąd nikt nie umieścił reszt zgodnych z  $\Delta_S$ ; na tym etapie pominięto pozycje, w których występowały reszty hydrofobowe (mogące być istotne dla poprawnego fałdowania białka) jak również te, które leżały w bezpośrednim sąsiedztwie już zoptymalizowanych reszt. Rotamery o najmniejszych i największych wkładach, przedstawia Tabela 3.2.2.

Ze względu na położenie, reszty Arg lub Lys w pozycji 171 mają podobny charakter oddziaływania do Arg202 (w rzeczy samej leżą „po drugiej stronie” centrum aktywnego). Przewiduje się jednak, że mutant zawierający dodatni aminokwas w obu pozycjach 202 i 171 będzie miał mniejszą aktywność niż suma efektów każdej z tych mutacji z osobna, gdyż wzajemnie odpychanie reszt nie pozwoli im równocześnie przyjąć optymalnej konformacji. Znacznie bardziej obiecująca jest proponowana mutacja Lys99Glu; ponieważ jednak reszta ta uczestniczy w jednym z dwóch sąsiednich mostków solnych, należałoby „obrócić” cały ten układ (poprzez mutacje Lys99Glu/Arg5Glu/Glu46Lys/Glu167Arg).

Przedstawiony schemat podobny jest do proponowanego przez zespół Warshela w pracy [93]; istnieje jednak kilka zasadniczych różnic. Po pierwsze, w prezentowanej pracy zastosowane jest rozwinięcie multipolowe, natomiast ładunki atomowe traktowane są jedynie jako jego przybliżenie; po drugie, prezentowane podejście *oddolne* daje szczegółowy wgląd w przestrzeń fazową układu bez konieczności

Tabela 3.8: Mutacje mogące prawdopodobnie zwiększyć aktywność P<sub>7</sub> wedle modelu DTSS(CAMM).

Aminokwas	najniższy DTSS kcal/mol	najwyższy DTSS kcal/mol
Arg 171	-9.62 t,t,t,t	-1.20 g+,g+,t,t
Lys 171	-6.53 t,t,t,t	-1.40 g+,g+,t,t
Asp 99	-4.20 t,g+	-2.54 g-,g-
Glu 99	-5.41 t,g-,g-	-2.30 g-,g+,g+

przewodzenia długich symulacji, włączając wzajemne oddziaływanie reszt, podczas gdy autorzy wspomnianej pracy kandydatów wyselekcjonowanych na zasadzie podobnej do  $\Delta_S$  poddają symulacjom EVB. W tym drugim przypadku zaznaczyć należy, że analiza możliwych konfiguracji rotamerów zaprezentowana w sekcji 3.2.2 dostarczyła wyjaśnienia także mutacji niezgodnych z samym tylko polem katalitycznym. Warto także zwrócić uwagę, że, zgodnie z [93], przeprowadzane na jednym rdzeniu CPU symulacje EVB dla jednego enzymu zajmują ok. 17 godzin; dla kontrastu, analiza konfiguracji rotamerów przedstawiona w poprzednich akapitach zajęła w najbardziej wymagającym przypadku (P<sub>6</sub>, w którym trzeba było przeanalizować ok. 21 milionów kombinacji) ok. 1 godziny przy prowadzeniu obliczeń jednowątkowych<sup>18</sup>.

### Wniosek 5

*Śledzenie wkładu DTSS aminokwasów wraz z analizą dostępnych dla nich konfiguracji (rotamerów) pozwala na uzyskanie szczegółowego wglądu w ich znaczenie z punktu widzenia katalizy. Wyjaśnienie znaczenia niektórych mutacji wymaga jednakże uwzględnienia wzajemnych oddziaływań pomiędzy nimi, co również jest możliwe w obrębie elektrostatycznego opisu oddziaływań. Tego rodzaju „skan” stanowi atrakcyjną metodę pośrednią między wyborem reszt zgodnych z  $\Delta_S$ , a symulacjami (EVB, obliczenia QM/MM połączone z MD) powszechnie proponowanymi jako dalszy etap weryfikacji, umożliwiając jeszcze większe zawężenie obszaru poszukiwań w relatywnie krótkim czasie.*

<sup>18</sup> Wykorzystano tutaj model ładunkowy; praktycznie tym samym kosztem obliczeniowym dla aminokwasów naładowanych można zastosować multipolową reprezentację  $\Delta_S$  w połączeniu z opisem aminokwasu za pomocą jego ładunku formalnego, *vide* sekcja 3.1.2.





# Rozdział 4

## Podsumowanie

Przedmiotem niniejszej pracy są trzy zagadnienia związane z zastosowaniem koncepcji różnicowej stabilizacji stanu przejściowego (DTSS) oraz pól katalitycznych do projektowania biokatalizatorów. Przede wszystkim zbadano granice stosowalności różnych przybliżeń oraz przeanalizowano za pomocą tych technik szereg zmutowanych wersji jednego enzymu, co pozwoliło sprawdzić zdolność metody do określania wpływu mutacji na aktywność katalityczną. Istotną cechą metodologii jest jej zakorzenienie w pierwszych zasadach (*ab initio*), do których zgodnie ze strategią *oddolną* (ang. *bottom-up*) sukcesywnie dobudowuje się kolejne elementy, przyjmując możliwie najmniejszą liczbę założeń. Wyniki prezentowanych dociekań pokazują, że sukcesywna, racjonalna analiza kolejnych czynników stanowi konkurencyjne podejście wobec powszechnie stosowanych metod *odgórnych* (ang. *top-down*), być może nawet stanowi sposób, by przeciąć gordyjski węzeł przybliżeń stosowanych m.in. w metodach QM/MM (np. ONIOM, EVB), wykorzystujących pola siłowe.

- **Atomowe rozwinięcie multipolowe stanowi uniwersalniejszy model opisu oddziaływań elektrostatycznych.** Niemniej jednak, w układach naładowanych elektrycznie potencjał elektrostatyczny jest zdominowany przez człon  $R^{-1}$ , przez co można opisać  $\Delta_S$  z dobrą dokładnością za pomocą ładunków Merza-Kollmana lub podobnych. Wówczas także anizotropowość ładunku otoczenia (katalizatora) wydaje się być drugorzędym czynnikiem (sekcja 3.1.1). Natomiast może być ona istotna w przypadku opisu stanów przejściowych, w których atomy uczestniczące w tworzeniu lub rozrywaniu nowych wiązań chemicznych wykazują duże odstępstwa od izotropowości. Wydaje się, że w układach naładowanych można z wystarczającą dokładnością opisać pole katalityczne  $\Delta_S$  za pomocą ładunków atomowych, tym samym zmniejszając koszt obliczeniowy; jednakże nawet wówczas może się to wiązać z pewnym błędem (sekcje 3.2.2 i Rysunek S.10).
- Odpowiedzią na opisany wyżej problem może być obserwacja z sekcji 3.1.2 (*vide* Rysunek S.9 w Suplemencie). **Reprezentacja aminokwasów naładowanych (dających najsilniejszy wkład do elektrostatycznej części DTSS) za pomocą ich ładunku formalnego umieszczonego w miejscu grupy funkcyjnej, w połączeniu z  $\Delta_S$  uzyskanym za pomocą multipoli atomowych, daje wynik bliski wartości  $\Delta E_{mtp}^{(10)}$ .** Potrzebny czas obliczeniowy (wyznaczony na przykładzie dwóch reszt liczących ok. 20 atomów i rozwinięcia CAMM do członu  $R^{-5}$  włącznie) jest wówczas zbliżony do modelu opartego na ładunkach atomowych.

Zastosowanie DTSS do projektowania enzymów wiąże się z takimi problemami jak wpływ struktury na wynik przewidywań czy kwestia traktowania reakcji wieloetapowych. Równocześnie ważna jest kwestia, czy elektrostatyczny wkład do energii DTSS rzeczywiście pozwala określić zmianę bariery aktywacji wskutek mutacji.

- W przypadku izomerazy ketosteroidowej (KSI), katalizującej reakcję polegającą na dwóch kolejnych transferach protonu, wprowadzono model różnicowej stabilizacji intermediatu (DISS; sekcja 3.2.1). **Pokazano, że daje on wyniki równoważne DTSS oraz poprawnie opisuje wpływ analizowanych mutacji na aktywność katalityczną.** Poprzez dekompozycję energii oddziaływań w ramach obliczeń SAPT zweryfikowano elektrostatyczny charakter DISS w całym szeregu mutantów. Analiza składowej elektrostatycznej pozwala na ilościowe

określenie wpływu mutacji na aktywność katalityczną, także mutantów nieujętych we wstępnych obliczeniach (Y55F i Y30F).

- Na tym przykładzie porównano ze sobą dwa zestawy struktur mutantów: wygenerowanych metodą *najmniejszego zaburzenia* i uzyskanych poprzez klasteryzację trajektorii z symulacji dynamiki molekularnej. **Reprezentacja członu elektrostatycznego za pomocą multipoli CAMM okazała się mniej podatna na zaburzenia strukturalne w stosunku do innych metod;** w porównaniu z ich kosztem obliczeniowym i dokładnością uzyskiwanych wyników jest konkurencyjna. Wyniki DISS(CAMM) wykazały zgodność bliską ilościowej z eksperymentem, były także lepsze od modelu opartego na ładunkach atomowych.

Najważniejszym celem prowadzonych badań jest opracowanie metody pozwalającej określić znaczenie mutacji wprowadzonych do teozymów drogą eksperymentów kierunkowanej ewolucji (LDE); równocześnie ważne jest, by związany z nią koszt obliczeniowy był relatywnie niski, by można było ją efektywnie zastosować do komputerowego projektowania enzymów.

- Na przykładzie mutantów teozymu KE07 (sekcja 3.2.2), przetestowano bibliotekę CAMM dla rotamerów aminokwasów białkowych. Jest ona niezbędnym elementem, jeżeli zastosowanie atomowych momentów multipolowych ma być opłacalne. **Prześledzenie możliwych konformacji reszt wprowadzanych w eksperymentach LDE wskazało cztery aminokwasy charakteryzujące się dużym, stabilizującym wkładem DTSS: Arg202, Asp7, Lys146 i Lys19. Trzy z nich znajdują się w najbardziej aktywnym enzymie z tej rodziny.**
- Wyjaśnienie stwierdzonego eksperymentalnie destabilizującego wpływu Lys19 wymaga uwzględnienia oddziaływań pomiędzy resztami aminokwasowymi. Prześledzenie możliwych kombinacji rotamerów aminokwasów z różnych pozycji z uwzględnieniem ich wzajemnego oddziaływania pokazało, że **obecność tej reszty utrudnia optymalną orientację Arg202, mającej najsilniejszy wpływ katalityczny.** W podobny sposób stwierdzono, że **rola reszty Asp224, mającej destabilizujący wkład DTSS, polega na ułatwieniu optymalnej orientacji Arg202.**
- Prezentowany skan możliwych konfiguracji rotamerów aminokwasów wraz z analizą ich oddziaływań pomiędzy sobą i z reagentami daje szczegółowy wgląd w charakterystykę przestrzeni fazowej układu, równocześnie **wykazując znacznie mniejszy koszt obliczeniowy w porównaniu z konwencjonalnymi symulacjami** (EVB, obliczenia QM/MM połączone z MD) proponowanymi jako narzędzie służące weryfikacji proponowanych mutantów. Dla przykładu, zgodnie z [93], **symulacje EVB jednego enzymu na pojedynczym rdzeniu trwają ok. 17 godzin,** podczas gdy w tych samych warunkach **proponowany skan konfiguracji rotamerów,** w najgorszym analizowanym przypadku (ok. 21 milionów kombinacji), trwał w podobnych warunkach (podobny procesor, obliczenia jednowątkowe) **ok. 1 godziny.**

Zaprezentowane wyniki pozwalają sądzić, że prezentowana metoda może zostać zastosowana przy interpretacji mutacji oraz ich przewidywaniu. Zaproponowana analiza konfiguracji rotamerów stanowi pomost pomiędzy prostym przewidywaniem w oparciu o pole katalityczne a przeprowadzaniem długich symulacji, co z może pozwolić na szybsze racjonalne przeszukiwanie zbioru potencjalnych mutacji. Podsumowując, w skróconej formie można najważniejsze wnioski z prezentowanej pracy sformułować następująco.

1. Technika pól katalitycznych reprezentująca alternatywne, całkowicie nieempiryczne podejście *od-dolne* (ang. *bottom-up*) może stanowić narzędzie do teoretycznego projektowania biokatalizatorów, analizy przestrzeni konformacyjnej oraz nieaddytywnych efektów związanych z wielokrotnymi mutacjami.
2. Metoda różnicowej stabilizacji stanu przejściowego (DTSS) stanowiaca uogólnienie metody stabilizacji stanu przejściowego (TSS) oraz destabilizacji substratów (GSD) umożliwiła jakościową interpretację roli mutacji w drugiej strefie koordynacyjnej eliminazy Kempa KE07.
3. Zastosowanie biblioteki atomowych momentów multipolowych dla rotamerów aminokwasów pozwala na modelowanie niskim kosztem obliczeniowym efektów wymienionych w punkcie 1.

# Suplement

## S.1 Przegląd pojęć związanych z rozwinięciem multipolowym

Obliczenie potencjału elektrostatycznego czy energii oddziaływania za pomocą rozwinięcia multipolowego sprowadza się do sumowania w ogólnym przypadku nieskończonego szeregu. Ponieważ w praktyce stosuje się sumy skończone, zdefiniować należy sposób sumowania oraz *rzęd* rozwinięcia<sup>1</sup>. Niech  $l_A$  ( $l_B$ ) oznacza rząd momentu multipolowego w  $A$  ( $B$ ), tj. sumę wykładników przy  $x$ ,  $y$  i  $z$  w równaniu 2.35 (strona 31); liczba ta jest związana z liczbą ładunków o takim samym module, za pomocą których można przedstawić dany multipol oraz z zależnością potencjału danego multipola od odległości  $R$  (Tabela S.1). Energia oddziaływania dwóch multipoli o rzędach  $l_A$  i  $l_B$  zmienia się wraz z odległością jak  $R^{-(l_A+l_B+1)}$ . Wówczas rozważaną sumę można przedstawić na dwa sposoby: jako sumę po *mo-*

Tabela S.1: Rzędy kilku najniższych multipoli.

Rząd $l$	Nazwa	Liczba równorzędnych ładunków punktowych $2^l$	Zależność potencjału od odległości $\sim R^{-l-1}$
0	monopol	1	$R^{-1}$
1	dipol	2	$R^{-2}$
2	kwadrupol	4	$R^{-3}$
3	oktupol	8	$R^{-4}$
4	heksadekapol	16	$R^{-5}$

*mentach* lub po *eksponentach* (wykładnikach). W pierwszym *rzęd*  $L$  oznacza najwyższy uwzględniony multipol, zaś w sumie zatrzymuje się wszystkie człony zawierające multipole mniejsze lub równe od niego ( $l \leq L$ ); w drugim przypadku rząd dotyczy największej sumy rzędów multipoli  $L = l_A + l_B$  występujących w jednym członie, związanych z największym uwzględnionym wykładnikiem  $R^{-(l_A+l_B+1)}$  (Rysunek S.1).

Drugi sposób sumowania, w przeciwieństwie do pierwszego, daje wynik niezależny od wyboru układu współrzędnych. Jest to, zgodnie z równaniem S.1, sposób sumowania rozwinięcia CAMM.

$$E_{CAMM}^{k \leq L} = \sum_{a \in A} \sum_{b \in B} \sum_{\substack{k_a \\ k_a + k_b \leq L}} \sum_{k_b} \mathbf{M}_a^{(k_a)} [k_a] \mathbf{T}^{(k_a + k_b)} [k_b] \mathbf{M}_b^{(k_b)} \quad (\text{S.1})$$

## S.2 SAPT– komentarze szczegółowe

### S.2.1 Delty Murrella i Landshoffa

W praktycznych zastosowaniach, gdy nie mamy do dyspozycji *idealnej* funkcji falowej cząsteczek, energia Heitlera-Londona  $E_{IE}^{HL}$  *nie jest równa* pierwszej poprawce SAPT  $E^{(1)}$ . Oznaczmy tyldą przybliżone funkcje falowe  $\tilde{\psi}_{A,0}$  i  $\tilde{\psi}_{B,0}$  oraz energie  $\tilde{E}_{int}^{HL}$  i  $\tilde{E}^{(1)}$ . Wówczas zachodzi relacja (S.2), gdzie  $\Delta_M$  i  $\Delta_L$  to odpowiednio delty Murrella i Landshoffa [158, 159].

$$\tilde{E}_{int}^{HL} = \tilde{E}^{(1)} + \Delta_M + \Delta_L \quad (\text{S.2})$$

<sup>1</sup> W przypadku *atomowego* rozwinięcia multipolowego, centra oznaczają oddziałujące ze sobą atomy; jeżeli rozwinięcie jest molekularne, są to centra (np. środki masy) molekul.

		$l_A$					
		0	1	2	3	...	
$L = 0$	—	<b>0</b>	$R^{-1}$	$R^{-2}$	$R^{-3}$	$R^{-4}$	...
$L = 1$	—	<b>1</b>	$R^{-2}$	$R^{-3}$	$R^{-4}$	$R^{-5}$	...
$L = 2$	—	<b>2</b>	$R^{-3}$	$R^{-4}$	$R^{-5}$	$R^{-6}$	...
$L = 3$	—	<b>3</b>	$R^{-4}$	$R^{-5}$	$R^{-6}$	$R^{-7}$	...
$L = 4$	—	...	...	...	...	...	...
$L = 5$	—	...	...	...	...	...	...
$L = 6$	—	...	...	...	...	...	...
...		...	...	...	...	...	...

Rysunek S.1: Schemat przedstawiający sumowanie po wykładniku. Kolumny (wiersze) odpowiadają momentom centrum  $A$  ( $B$ ), w komórkach wpisano potęgę odległości  $R$  odpowiadającą odpowiedniemu członowi; dla rozwinięcia rzędu  $L$  do sumy wchodzą człony powyżej linii oznaczonej odpowiednim kolorem.

Wielkości te można zdefiniować poprzez operatory Focka  $\hat{F}_A$  i operatory potencjału fluktuacyjnego  $\hat{W}_A$  monomerów  $A$  i  $B$ , przyjmując za  $\tilde{\psi}$  przybliżone rozwiązania równania Hartree-Focka. Przyjmując oznaczenia (S.3)-(S.7), wkłady pochodzące od monomerów  $\Delta_M^A$  i  $\Delta_L^A$  (i analogiczne wielkości dla drugiej cząsteczki) można wyrazić poprzez (S.8) i (S.9).

$$\Delta_M = \Delta_M^A + \Delta_M^B \quad (\text{S.3})$$

$$\Delta_L = \Delta_L^A + \Delta_L^B \quad (\text{S.4})$$

$$\tilde{\phi}_a^{(0)} = \tilde{\psi}_{A,0} \tilde{\psi}_{B,0} \quad (\text{S.5})$$

$$\langle \hat{F}_A \rangle = \langle \tilde{\psi}_{A,0} | \hat{F}_A | \tilde{\psi}_{A,0} \rangle \quad (\text{S.6})$$

$$\langle \hat{W}_A \rangle = \langle \tilde{\psi}_{A,0} | \hat{W}_A | \tilde{\psi}_{A,0} \rangle \quad (\text{S.7})$$

$$\Delta_L^A = \frac{\langle \tilde{\phi}^{(0)} | \hat{\mathcal{A}} (\hat{F}_A - \langle \hat{F}_A \rangle) | \tilde{\phi}^{(0)} \rangle}{\langle \tilde{\phi}^{(0)} | \hat{\mathcal{A}} | \tilde{\phi}^{(0)} \rangle} \quad (\text{S.8})$$

$$\Delta_M^A = \frac{\langle \tilde{\phi}^{(0)} | \hat{\mathcal{A}} (\hat{W}_A - \langle \hat{W}_A \rangle) | \tilde{\phi}^{(0)} \rangle}{\langle \tilde{\phi}^{(0)} | \hat{\mathcal{A}} | \tilde{\phi}^{(0)} \rangle} \quad (\text{S.9})$$

Wykazano [160, 161], że delta Landshoffa jest równa zero dla dokładnych rozwiązań HF (w tzw. granicy Hartree-Focka lub w bazie zupełnej) oraz w bazie dimeru, natomiast delta Murrella jest rzędu  $O(S^4)$  [161]. Tak więc  $E_{ex}^{(1)}$  w rachunku HVPT, zdefiniowana jako różnica  $E_{int}^{HL} - E_{el}^{(1)}$  liczona w bazie dimeru, różni się od  $E_{ex}^{(1)}$  o  $\Delta_M$  (S.10). Jeżeli obliczenia  $E_{ex}^{(10)}$  (SAPT) byłyby przeprowadzane w bazie monomerów, skutkowałoby to różnicą względem  $E_{ex}^{(10)}$  (HVPT) o  $\Delta_L$  i jeszcze jedną poprawkę wyrażającą różnicę energii elektrostatycznych w bazie dimeru i monomerów.

$$E_{ex}^{(10)}(HVPT) = E_{ex}^{(10)}(SAPT) + \Delta_M \quad (\text{S.10})$$

## S.2.2 Pierwsza poprawka SAPT w bazach MO i AO

Dla celów dydaktycznych warto rozpisać wyrażenia na  $E_{el}^{(1)}$  i  $E_{ex}^{(1)}$  (równania S.11 i S.12, wspomniane w sekcji 2.2.2) na orbitale molekularne (ang. *Molecular Orbitals*, MO) cząsteczek A i B.

$$E^{(1)} = \frac{\langle \psi_{A,0} \psi_{B,0} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle + \langle \psi_{A,0} \psi_{B,0} | \hat{V} \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle + O(S^4)}{1 + \langle \psi_{A,0} \psi_{B,0} | \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle + O(S^4)} \quad (\text{S.11})$$

$$E_{ex}^{(1)} = \langle \psi_{A,0} \psi_{B,0} | \hat{V} \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle - \langle \psi_{A,0} \psi_{B,0} | \hat{V} | \psi_{A,0} \psi_{B,0} \rangle \langle \psi_{A,0} \psi_{B,0} | \hat{P}^{AB} | \psi_{A,0} \psi_{B,0} \rangle + O(S^4) \quad (\text{S.12})$$

Niech funkcja stanu podstawowego  $\phi_0^{(0)}$  dana będzie iloczynem wyznaczników Slatera molekuł  $A$  i  $B$  (S.13); oznaczmy przez  $i, \bar{i}$  spinorbitale molekularne  $A$ , oraz przez  $j, \bar{j}$  spinorbitale molekularne  $B$  (równocześnie litery bez nadkreśleń  $i, j, k, l$  oznaczają odpowiednie *orbitale* molekularne).

$$\phi_0^{(0)} = |\cdots i \bar{i} \cdots | \cdot | \cdots j \bar{j} \cdots | \quad (\text{S.13})$$

Wprowadźmy następujące oznaczenia dla całek nakładania (S.14), całek dwuelektronowych (S.15) i jednoelektronowych całek oddziaływania z potencjałem jąder atomowych (S.16).

$$\langle i | j \rangle \equiv \langle i(1) | j(1) \rangle \quad (\text{S.14})$$

$$\langle ij | kl \rangle \equiv \langle i(1)j(1) | r_{12}^{-1} | k(2)l(2) \rangle \quad (\text{S.15})$$

$$V_{a,ij} \equiv \langle i(1) | -Z_a r_{a1}^{-1} | j(1) \rangle \quad (\text{S.16})$$

Wówczas  $E_{el}^{(10)}$  i  $E_{ex}^{(10)}$  dane są, odpowiednio, równaniami (S.17) i (S.18).

$$E_{el}^{(1)} = 4 \sum_{i \in A} \sum_{j \in B}^{occ} \langle ii | jj \rangle + 2 \sum_{i \in A} \sum_{bin B}^{occ} V_{b,ii} + 2 \sum_{a \in A} \sum_{j \in B}^{occ} V_{a,jj} + \sum_{a \in A} \sum_{b \in B} Z_a Z_b R_{ab}^{-1} \quad (\text{S.17})$$

$$E_{ex}^{(1)} = \left\{ -2 \sum_{i \in A} \sum_{j \in B}^{occ} \left[ \langle ij | ij \rangle + \left( 2 \sum_{i' \in A} \langle i' i' | ij \rangle - \langle ii | ij \rangle \right) + 2 \sum_{j' \in B} \langle ij | j' j' \rangle - \langle ij | jj \rangle + \sum_{a \in A} V_{a,ij} + \sum_{b \in B} V_{b,ij} \right] \langle i | j \rangle + \left( \langle ii | jj \rangle - 2 \sum_{i' \in A} \langle i' i' | jj \rangle - 2 \sum_{j' \in B} \langle ii | j' j' \rangle - \sum_{a \in A} V_{a,jj} - \sum_{b \in B} V_{b,ii} \right) \langle i | j \rangle^2 \right\} \left( 1 - 2 \sum_{i \in A} \sum_{j \in B}^{occ} \langle i | j \rangle^2 \right)^{-1} \quad (\text{S.18})$$

Ponieważ wiodące w obu równaniach są wkłady dwuelektronowe, koszt obliczeniowy wyznaczenia obu członów skaluje się z rozmiarem bazy funkcyjnej  $N$  jak  $O(N^4)$ . Aby to lepiej unaocznić, przydatne będzie rozpisanie członu elektrostatycznego w bazie orbitali atomowych (ang. *Atomic Orbitals*, AO).

### S.2.3 Całka elektrostatyczna w bazie AO

Niech  $\{\phi_k\}$  stanowi bazę cząsteczki  $A$ , a  $\{\phi_l\}$  bazę cząsteczki  $B$ ; wówczas  $|i\rangle = \sum_k c_{ik} |\phi_k\rangle \equiv \sum_k c_{ik} |k\rangle$  i  $|j\rangle = \sum_l c_{jl} |\phi_l\rangle \equiv \sum_l c_{jl} |l\rangle$ , gdzie współczynnik  $c_{ik}$  oznacza udział orbitalu atomowego  $k$  w rozwinięciu orbitalu molekularnego  $i$ . Skorzystajmy także z definicji *macierzy gęstości* (równania S.19 i S.20), będącej obok funkcji falowej jednym z wyników metody SCF.

$$P_{kk'}^A = 2 \sum_{i \in A}^{occ} c_{ik}^* c_{ik'} \quad (\text{S.19})$$

$$P_{ll'}^B = 2 \sum_{j \in B}^{occ} c_{jl}^* c_{jl'} \quad (\text{S.20})$$

Wówczas całki jedno- i dwuelektronowe w równaniu (S.17) zapisać za pomocą (S.21)-(S.24).

$$\begin{aligned} V_{a,jj} &= \langle j(1) | -Z_a r_{a1}^{-1} | j(1) \rangle \\ &= \sum_l \sum_{l' \in B} c_{jl}^* c_{jl'} \langle l(1) | -Z_a r_{a1}^{-1} | l'(1) \rangle \end{aligned} \quad (\text{S.21})$$

$$\begin{aligned} 2 \sum_{a \in A} \sum_{j \in B}^{occ} V_{a,jj} &= 2 \sum_{a \in A} \sum_{j \in B}^{occ} \sum_l \sum_{l' \in B} c_{jl}^* c_{jl'} \langle l(1) | -Z_a r_{a1}^{-1} | l'(1) \rangle \\ &= \sum_l \sum_{l' \in B} c_{jl}^* c_{jl'} \left( 2 \sum_{j \in B}^{occ} c_{jl}^* c_{jl'} \right) \left\langle l(1) \left| - \sum_{a \in A} Z_a r_{a1}^{-1} \right| l'(1) \right\rangle \\ &\equiv \sum_l \sum_{l' \in B} c_{jl}^* c_{jl'} P_{ll'}^B V_{A,ll'}^{nuc} \end{aligned} \quad (\text{S.22})$$

$$\begin{aligned} \langle ii | jj \rangle &= \sum_l \sum_{l' \in B} c_{jl}^* c_{jl'} \langle ii | ll' \rangle \\ &= \sum_{k, k' \in A} \sum_l \sum_{l' \in B} c_{ik}^* c_{ik'} c_{jl}^* c_{jl'} \langle kk' | ll' \rangle \end{aligned} \quad (\text{S.23})$$

$$\begin{aligned} 4 \sum_{i \in A} \sum_{j \in B}^{occ} \langle ii | jj \rangle &= 4 \sum_{i \in A} \sum_{j \in B}^{occ} \sum_{k, k' \in A} \sum_l \sum_{l' \in B} c_{ik}^* c_{ik'} c_{jl}^* c_{jl'} \langle kk' | ll' \rangle \\ &= \sum_{k, k' \in A} \sum_l \sum_{l' \in B} \left( 2 \sum_{i \in A}^{occ} c_{ik}^* c_{ik'} \right) \left( 2 \sum_{j \in B}^{occ} c_{jl}^* c_{jl'} \right) \langle kk' | ll' \rangle \\ &= \sum_{k, k' \in A} \sum_l \sum_{l' \in B} P_{kk'}^A P_{ll'}^B \langle kk' | ll' \rangle \end{aligned} \quad (\text{S.24})$$

Pozwala to ostatecznie wyrazić  $E_{el}^{(1)}$  za pomocą (S.25). Jak można zauważyć, baza rozmiaru  $N$  na każdej z molekuł skutkuje koniecznością wyznaczenia  $N^4$  całek dwuelektronowych i  $2N^2$  całek jednoelektronowych, co jest powodem skalowalności rzędu  $O(N^4)$ .

$$\begin{aligned} E_{el}^{(1)} &= \sum_{k, k' \in A} \sum_l \sum_{l' \in B} P_{kk'}^A P_{ll'}^B \langle kk' | ll' \rangle + \sum_{k, k' \in A} P_{kk'}^A V_{B,kk'}^{nuc} \\ &\quad + \sum_l \sum_{l' \in B} P_{ll'}^B V_{A,ll'}^{nuc} + \sum_{a \in A} \sum_{b \in B} Z_a Z_b R_{ab}^{-1} \end{aligned} \quad (\text{S.25})$$

### S.3 Dynamiczne pole katalityczne

Oprócz wiodącej interpretacji [122],  $\vec{\Delta}_d$  można zinterpretować na dwa inne sposoby.

**Optymalizacja geometrii.** Jako gradient  $-\Delta_S$ , można je wykorzystać w optymalizacji geometrii katalizatora pod kątem jego jak najlepszego dostosowania do pola katalitycznego, względnie optymalnej reorientacji reagentów względem nieruchomego katalizatora (tzw. wewnętrzne pole katalityczne) [162].

**Katalityczne dipole.** Podczas, gdy  $-\Delta_S$  zawiera informację o optymalnym rozkładzie ładunku idealnego katalizatora,  $\vec{\Delta}_d$  oznacza to samo dla położenia katalitycznych dipoli.

Dowód drugiej interpretacji jest dość prosty. Przyjmując zbiór dipoli elektrycznych otoczenia  $\{\vec{\mu}_i\}_{env}$  w miejsce ładunków  $\{q_i\}_{env}$ , energia ich oddziaływania z substratem (stanem przejściowym) dana jest równaniem (S.26), co daje wyrażenie na energię DTSS (S.27).

$$E_{IE}(TS) \approx - \sum_{i \in env} \vec{\mu}_i \cdot \vec{F}_{TS}(\vec{r}_i) \quad (S.26)$$

$$E_{IE}(TS) - E_{IE}(S) \approx - \sum_{i \in env} \vec{\mu}_i \cdot (\vec{F}_{TS} - \vec{F}_S)(\vec{r}_i) = - \sum_{i \in env} \vec{\mu}_i \cdot \vec{\Delta}_d(\vec{r}_i) \quad (S.27)$$

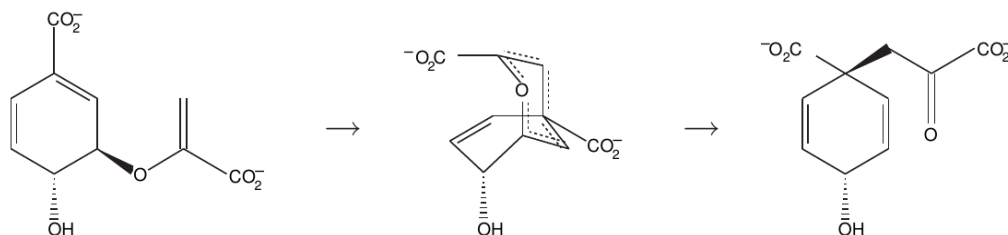
Energia (S.27) ma wartość najmniejszą (optymalną), kiedy dipole  $\vec{\mu}_i$  znajdują się w miejscach o możliwie dużym module dynamicznego pola katalitycznego  $|\vec{\Delta}_d(\vec{r}_i)|$ , a ich kierunki są z tym polem zgodne (oczywiście, możliwie duża wartość modułu  $|\vec{\mu}_i|$  też jest z tej perspektywy pożądana).

Takie zastosowanie  $\vec{\Delta}_d$  może być przydatne, kiedy rozważane są reszty aminokwasowe obojętne elektrycznie; wówczas  $\vec{\Delta}_d$  może wskazać optymalne rozmieszczenie wiązań wodorowych (*vide* sekcja S.3.1), które występują np. w tzw. dziurach oksyanionowych stabilizujących produkty przejściowe reakcji (jak we wspomnianej we Wstępie KSI). Biorąc pod uwagę, że pole katalityczne ma charakter co najmniej dipolowy (ze względu na zasadę zachowania ładunku), energia DTSS pochodząca z takich oddziaływań zależy od odległości  $R$  jak  $R^{-3}$  (dla porównania, wkład katalityczny reszt obdarzonych ładunkiem zmienia się jak  $R^{-2}$ ); w obliczu szczególnego zainteresowania „drugą sferą koordynacyjną” (*vide* sekcja 1.3.3), a także faktu, że dowolny multipol można wyreprezentować odpowiednim układem ładunków, interpretacja ta traci na znaczeniu, gdyż wkłady pochodzące od odległych dipoli przyćmiewane są przez te od reszt obdarzonych ładunkiem.

### S.3.1 Testy

#### Próba wskazania optymalnej katalitycznie konformacji za pomocą $\vec{\Delta}_d$ .

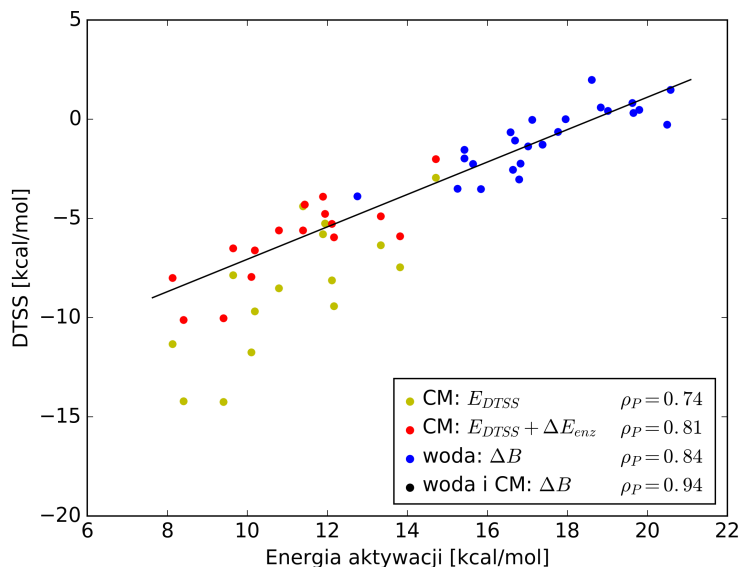
Jak wspomniano w Rozdziale 1, przy opisie zjawisk związanych z katalizą enzymatyczną istotna jest także informacja o lokalnym wycinku przestrzeni fazowej układu (dostępnych termicznie konformacjach), co pozwala na określenie energii swobodnej aktywacji. Aby zbadać z tej perspektywy koncepcję DTSS i pól katalitycznych, skorzystano ze struktur mutazy choryzmianowej udostępnionych w ramach współpracy z grupą prof. Mulhollanda z Uniwersytetu w Bristolu [37]. Schemat katalizowanej reakcji przedstawia Rysunek S.2. Analizie poddano 16 ścieżek reakcji wyznaczonych za pomocą obliczeń QM/MM dla 16 klatek pochodzących z symulacji dynamiki molekularnej [163]. Dodatkowo przeanalizowano także 24 ścieżki uzyskane analogicznie dla reakcji w roztworze wodnym.



Rysunek S.2: Schemat reakcji przegrupowania choryzmyanu do preferianu, katalizowanej przez mutazę choryzmianową.

W pierwszej kolejności poszukiwano odpowiedzi na następujące pytanie: czy konformacja o optymalnym DTSS odpowiada konformacji o najniższej energii aktywacji reakcji? Czy  $\vec{\Delta}_d$  może wskazać taki ruch reszt aminokwasowych, jaki ku niej wiedzie? Niestety, nie udało się znaleźć takiej zgodności. Przyczynę takiego stanu rzeczy najłatwiej zilustrować, analizując korelację pomiędzy DTSS a barierą aktywacji. Trzeba jednakże zaznaczyć, że w pracy [163] energie DTSS wyliczono dokładnie jako różnicę barier aktywacji w próżni i enzymie  $B_G - B_E$ , tym samym wielkość ta uwzględniała energię wewnętrzną samego białka; aby uniknąć nieporozumień, w dalszym tekście wielkość ta (różnica barier) będzie

określana jako  $\Delta B$ , natomiast termin DTSS zarezerwowany zostanie dla różnicy energii oddziaływań, będącej przybliżeniem  $\Delta B$  na podstawie równania 2.1 z Rozdziału 2. Dodatkowo, porównując tak zdefiniowane energie DTSS z barierami aktywacji, łącznie potraktowano reakcje zachodzącą w roztworze wodnym i w enzymie. Uwzględnienie obu tych czynników — zawarcia w zbiorze porównywanych danych obu reakcji oraz dodania do energii oddziaływania energii wewnętrznej enzymu w różnych konformacjach — pozwoliło autorom uzyskać wysoką korelację pomiędzy  $\Delta B$  a barierą aktywacji (współczynnik korelacji Pearsona równy 0.96 [163]). Ograniczenie zestawu danych do reakcji przebiegającej w enzymie skutkuje obniżeniem współczynnika korelacji do 0.84, natomiast pominięcie zmian energetycznych wynikających z deformacji enzymu prowadzi do współczynnika korelacji 0.74. Obrazowo przedstawione jest to na Rysunku S.3. Wynika stąd, że decydujący wkład do  $\Delta B$  pochodzi od oddziaływania



Rysunek S.3: Zależność pomiędzy DTSS a bariera aktywacji w mutacji choryzmianowej dla różnych zestawów danych.

CM – DTSS(CAMM) w mutacji choryzmianowej, CM (z MM) – DTSS(CAMM) plus energia wewnętrzna enzymu  $\Delta E_{enz}$  (z Mechaniki Molekularnej), woda – DTSS(CAMM) dla ścieżek w roztworze wodnym, woda + CM – prosta najlepszego dopasowania dla zestawu zawierającego łącznie woda i CM (z MM).

ze środowiskiem białkowym. Drugorzędym, chociaż istotnym czynnikiem jest zmiana energii samego enzymu  $\Delta E_{enz}$  (równanie 2.1 w Rozdziale 2). Zauważyć tutaj można, że konformacja o bardziej ujemnym DTSS ma z reguły wyższy wkład  $\Delta E_{enz}$ ; wielkości te są silnie skorelowane: ich współczynnik korelacji wynosi  $-0.89^2$ . Wyniki te sugerują, że — pomimo obiecujących rezultatów poprzednich analiz tego rodzaju [106] — wykorzystanie samego pola katalitycznego do określenia optymalnych deformacji termicznych enzymu może być niewystarczające.

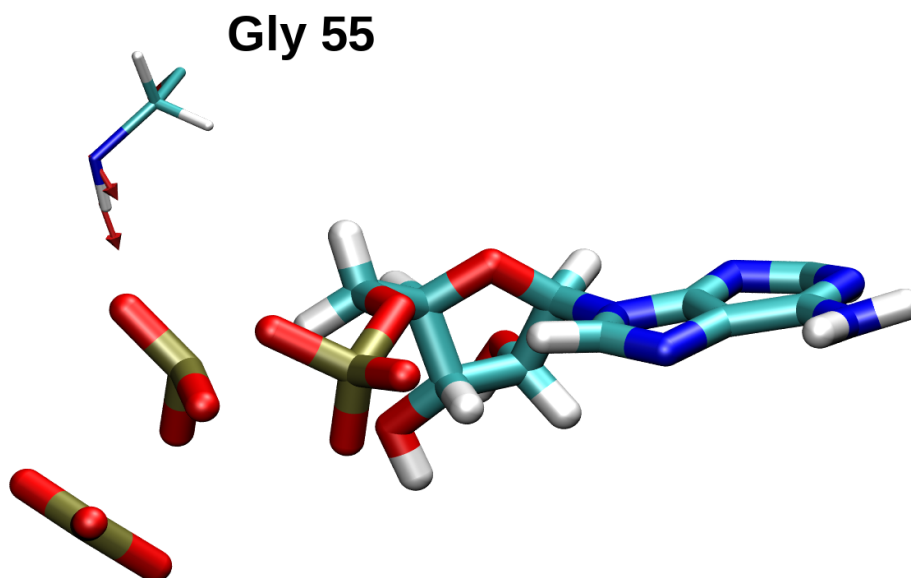
### Katalityczne dipole

Próby zastosowania drugiej interpretacji ( $\vec{\Delta}_d$  jako predyktor położenia momentów dipolowych o znaczeniu katalitycznym, np. wiązań wodorowych) przyniosły nieco lepsze rezultaty. Pierwszym jest zgodność dynamicznego pola katalitycznego w PKA z orientacją wiązania wodorowego Gly55, aminokwasu mającego najbardziej stabilizujący wkład DTSS spośród reszt obojętnych elektrycznie (*vide* Rys. S.4) – wektor  $\vec{\Delta}_d$  w miejscu atomu wodoru jej grupy aminowej miał stosunkowo dużą wartość; niestety, w przypadku pozostałych, nawet wykazujących stabilizujący DTSS, nie udało się znaleźć podobnej korelacji, o klarownej interpretacji. Wynikać to może zarówno z nieoptymalnej orientacji donora wiązania wodorowego, jak i z arbitralności definicji momentu dipolowego grupy funkcyjnej (szczególnie w przypadku reszt takich jak np. asparagina).

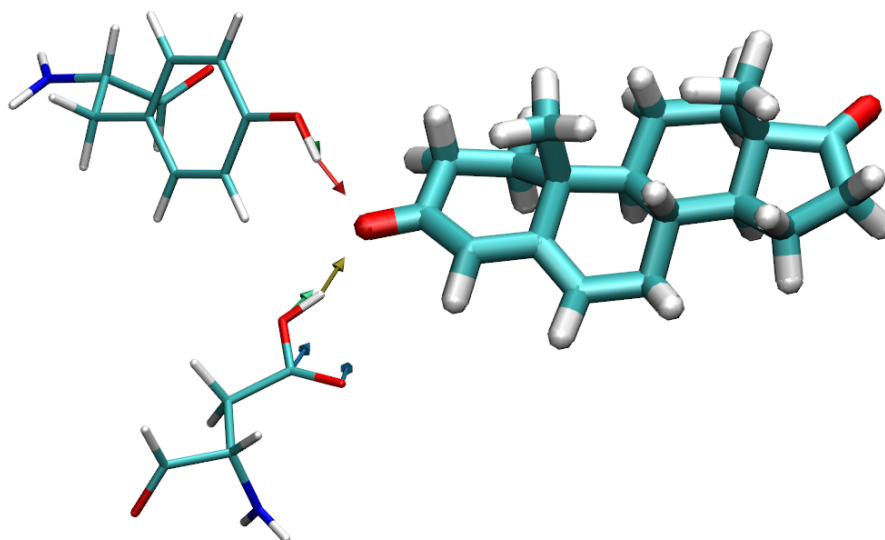
<sup>2</sup> Podobne powiązania pomiędzy DTSS a  $B_0$  i  $\Delta E_{enz}$  a  $B_0$  są rzędu 0.40–0.5.



Drugim przykładem jest centrum aktywne KSI (sekcja 3.2.1), gdzie wyjątkowo duże wektory  $\vec{\Delta}_d$  pozostają w zgodności z orientacją wiązań wodorowych dziury oksyanionowej (reszty Tyr14 i Asp99), które mają duże znaczenie katalityczne w tym enzymie (Rysunek S.5).



Rysunek S.4: Wektor dynamicznego pola katalitycznego w PKA oraz orientacja Gly55.



Rysunek S.5: Wektor dynamicznego pola katalitycznego w KSI oraz orientacje Tyr14 i Asp99.

## S.4 Powiązanie z metodą zaburzenia energii swobodnej (FEP)

Wyrażenie na różnicową stabilizację stanu przejściowego (2.1) można także wyprowadzić z metody zaburzenia energii swobodnej (ang. *Free Energy Perturbation*, FEP). Niech  $\beta = (k_B T)^{-1}$ , zaś różnica energii układu pomiędzy substratem a stanem przejściowym dana będzie przez  $\Delta E = E(TS) - E(S)$ ; wówczas energię swobodną reakcji wyrazić można przez formułę Zwanziga (S.28) [164], w którym  $\langle \dots \rangle_S$  oznacza średnią po zespole substratów. To wyrażenie można dalej rozwinąć w szereg kumulant (S.29) [165].

$$\Delta G^\ddagger = -\beta^{-1} \ln \langle \exp(-\beta \Delta E) \rangle_S \quad (\text{S.28})$$

$$\Delta G^\ddagger = -\beta^{-1} \sum_{n=1}^{\infty} \frac{\kappa_n(\Delta E)}{n!} (-\beta)^n = \langle \Delta E \rangle_S - \frac{\beta}{2} \text{Var}_S(\Delta E) + \dots \quad (\text{S.29})$$

Jeżeli energię układu (hamiltonian) poddać dekompozycji na wkłady: reagentów  $E_r$ , enzymu  $E_{enz}$  i oddziaływania enzym-reagenty  $U$ , różnicę  $\Delta E$  można wyrazić w myśl (2.1) za pomocą (S.30); zaniedbując  $\Delta E_{enz}$  jako wielkość niewielką lub co najmniej stałą w szeregu porównywanych mutantów, (S.29) można wyrazić przez (S.31).

$$\begin{aligned} \Delta H &= E_r(TS) - E_r(S) + U(TS) - U(S) + E_{enz}(TS) - E_{enz}(S) \\ &= B_0 + E_{DTSS} + \Delta E_{enz} \end{aligned} \quad (\text{S.30})$$

$$\begin{aligned} -\beta^{-1} \ln \langle \exp(-\beta(B_0 + E_{DTSS})) \rangle_S &= \langle \Delta B_0 \rangle_S + \langle E_{DTSS} \rangle_S - \\ &- \frac{\beta}{2} (\text{Var}_S(B_0) + \text{Var}_S(E_{DTSS}) + 2\text{Cov}_S(B_0, E_{DTSS})) + \dots \end{aligned} \quad (\text{S.31})$$

W wyrażeniach (S.29) i (S.31)  $\text{Var}_S$  i  $\text{Cov}_S$  oznaczają wariancję i kowariancję wyznaczoną po zespole substratów. Jeżeli  $B_0$  i  $E_{DTSS}$  są (choćby w przybliżeniu) nieskorelowane i odpowiednie człony krzyżowe w (S.31) znikają<sup>3</sup>, równanie to można przegrupować tak, by odpowiadało sumie wielkości zależnych energii samych reagentów i energii oddziaływania (S.32).

$$\Delta G^\ddagger \approx -\beta^{-1} [\ln \langle \exp(-\beta \Delta B_0) \rangle_S + \ln \langle \exp(-\beta E_{DTSS}) \rangle_S] \quad (\text{S.32})$$

Z kolei biorąc z (S.31) tylko człony zerowego rzędu względem  $\beta$  uzyskujemy podział energii aktywacji pomiędzy barierę własną  $B_0$  a energię DTSS, z których każdy przyczynek jest reprezentowany przez średnią po zespole (S.33).

$$\Delta G^\ddagger \approx \langle B_0 \rangle_S + \langle E_{DTSS} \rangle_S \quad (\text{S.33})$$

Wraz z obserwacją Cooper i Kästnera, że średnia eksponencjalna (wyrażona przez równanie S.28) jest lepszym estymatorem energii swobodnej aktywacji [166] niż np. średnia arytmetyczna, przedstawione wyprowadzenie daje podstawy, by w analizie stosować uśrednione energie DTSS, zarówno w postaci średniej arytmetycznej (popularny estymator wielkości w (S.33)), jak i eksponencjalnej.

### S.4.1 Korelacja pomiędzy DTSS a barierą własną

Przedstawione powyżej rozważania postanowiono zweryfikować w oparciu o wspomniane wyniki dla mutazy choryzmianowej (sekcja S.3.1). Niech  $\langle \dots \rangle$  oznacza średnią arytmetyczną, zaś  $\langle \dots \rangle_{exp}$  średnią eksponencjalną, która jest dobrym estymatorem energii swobodnej [166]. Wówczas  $\langle B_{enz} \rangle$  i  $\langle B_{enz} \rangle_{exp}$  (Tabela S.2) są w dobrej zgodności z eksperymentalnymi wartościami entalpii  $\Delta H^\ddagger$  i entalpii swobodnej aktywacji  $\Delta G^\ddagger$ : odpowiednio 12.7 i 9.1 kcal/mol [167]. Ponieważ estymatorem entalpii aktywacji jest średnia arytmetyczna, poszczególne przyczynki ( $B_0$ ,  $\Delta E_{enz}$  i  $E_{DTSS}$ ) są w oczywisty sposób addytywne; w przypadku entalpii swobodnej, jak pokazuje Tabela S.3, założenie addytywności wiąże się z błędem rzędu 5 kcal/mol. Dopiero uwzględnienie wzajemnej zależności<sup>4</sup>  $\Delta E_{enz}$  i  $E_{DTSS}$  pozwala uzyskać błąd rzędu 1.4 kcal/mol. Sugeruje to, że nie tylko wkład od  $\Delta E_{enz}$  do entalpii swobodnej

<sup>3</sup> Dla niezależnych zmiennych  $X$  i  $Y$  zachodzi addytywność kumulant  $\kappa_n(X + Y) = \kappa_n(X) + \kappa_n(Y)$ .

<sup>4</sup> Nie został tutaj użyty termin „korelacja”, gdyż, ze względu na wolną zbieżność szeregu kumulant w tym przypadku, wzajemne powiązanie pomiędzy zmiennymi wykracza poza wariancję i kowariancję (kumulanta drugiego rzędu) występującą w wyrażeniu na współczynnik korelacji.

Tabela S.2: Średnie i średnie eksponencjalne składowych bariery aktywacji.

$X$	$\langle X \rangle$	$\langle X \rangle_{exp}$
$B_{enz}$	11.2 ± 0.9	9.4 ± 0.3
$E_{DTSS}$	-8.2 ± 0.9	-13.01 ± 0.31
$\Delta E_{enz}$	2.1 ± 0.4	0.29 ± 0.35
$B_0$	17.31 ± 0.31	16.55 ± 0.17
$B_0 + \Delta E_{enz}$	—	16.9 ± 0.3
$B_0 + E_{DTSS}$	—	5.6 ± 0.3
$E_{DTSS} + \Delta E_{enz}$	—	-8.87 ± 0.31

Tabela S.3: Błąd oszacowania energii swobodnej związany z nieaddytywnością składowych bariery aktywacji.

$X$	$X - \langle B_{enz} \rangle_{exp}$
$\langle B_0 \rangle_{exp} + \langle E_{DTSS} \rangle_{exp} + \langle \Delta E_{enz} \rangle_{exp}$	-5.57
$\langle B_0 + \Delta E_{enz} \rangle_{exp} + \langle E_{DTSS} \rangle_{exp}$	-5.51
$\langle B_0 \rangle_{exp} + \langle E_{DTSS} + \Delta E_{enz} \rangle_{exp}$	-1.37

aktywacji jest istotny, ale także jego powiązanie z energią DTSS; zależność ta, jak można się przekonać analizując Tabelę S.2, wydaje się być nawet istotniejsza od samego  $\Delta E_{enz}$ : różnica pomiędzy  $\langle E_{DTSS} + \Delta E_{enz} \rangle_{exp}$  a  $\langle E_{DTSS} \rangle_{exp}$ , równa 4.14 kcal/mol, jest prawie dwukrotnie większa od  $\langle \Delta E_{enz} \rangle$ , nie wspominając już o  $\langle \Delta E_{enz} \rangle_{exp}$ .

Wyniki te prowadzą do wniosku, że także przy uśrednianiu energii aktywacji po zespole istotne jest uwzględnienie wkładu własnego enzymu  $\Delta E_{enz}$ ; w przypadku energii swobodnej, estymowanej średnią eksponencjalną, jeszcze istotniejsze jest wzajemne powiązanie tej wielkości z energią DTSS. Bariera własna  $B_0$  jest względnie addytywna z pozostałymi członami (w analizowanym przykładzie wiąże się to z niedoszacowaniem  $\Delta G^\ddagger$  o ok. 1.4 kcal/mol), lecz pominięcie zależności DTSS z  $\Delta E_{enz}$  wiąże się z dalszym niedoszacowaniem rzędu 4.2 kcal/mol, co na obecnym etapie raczej przekreśla możliwość oszacowania zmian wkładu entropowego do energii swobodnej za pomocą DTSS. Istnieje możliwość, że część tych błędów w szeregu mutantów ma systematyczny (częściowo stały) charakter, jednakże rozsądzenie tej kwestii bez przeanalizowania w powyższy sposób co najmniej kilku mutantów jest niemożliwe.

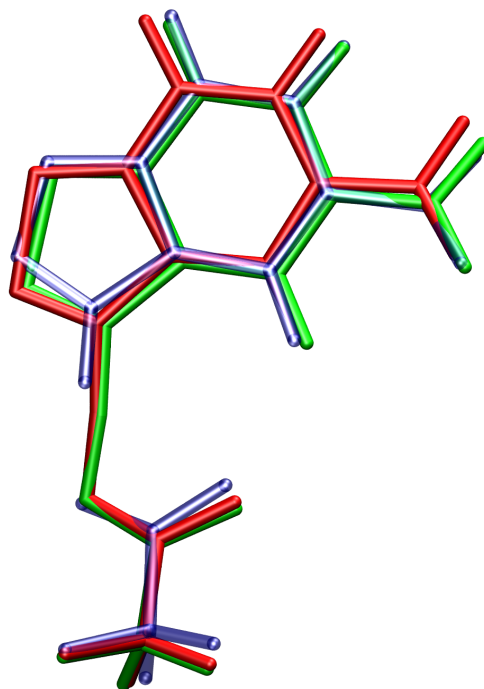
## S.5 Obliczenia ONIOM

Stan przejściowy, substrat i ścieżkę reakcji uzyskano metodą ONIOM [90] z wykorzystaniem funkcjonału B3LYP/6-311+G(d,p) w części QM (benzoizoksazol wraz z resztą octanową reprezentującą Glu101; wiązanie  $C_\gamma$ - $C_\delta$  w obliczeniach QM zastąpiono łącznikowym atomem wodoru) oraz pola siłowego Amber94, stosując osadzenie mechaniczne (ang. *mechanical embedding*) do opisu interakcji pomiędzy regionami. Podejście to w sposób jawny pomija polaryzację reagentów wskutek oddziaływania otoczenia białkowego. Uwzględnienie tego efektu jest szczególnie zalecane w przypadku stanów przejściowych reakcji chemicznych; ponieważ napotkano problemy techniczne w optymalizacji z wykorzystaniem osadzenia elektrostatycznego oraz trójwarstwowej wersji ONIOM z osadzeniem mechanicznym, posłużono się procedurą refitowania ładunku regionu QM [135].

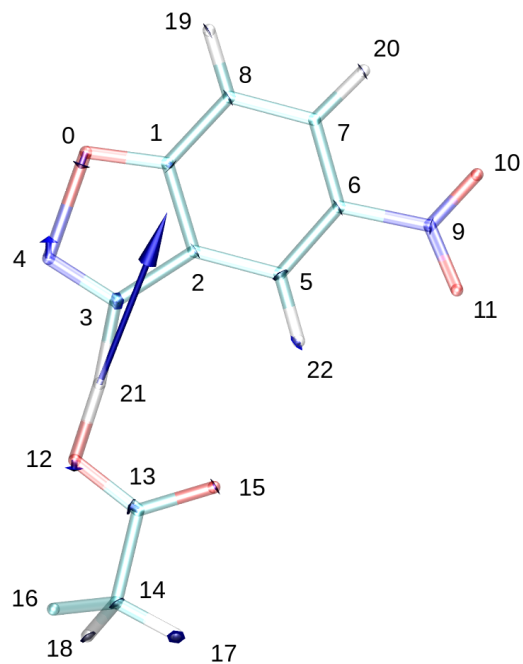
Już po pierwszej rundzie uzyskano strukturę, która tylko nieznacznie różniła się od poprzedniej (Rys. S.6). W drugiej rundzie refitowania ładunków i optymalizacji geometrii, maksymalne przesunięcie atomów było mniejsze od 0.01 Å. Tym samym uznano, że efekt polaryzacji na strukturę nie jest w tym wypadku znaczący. Znaczenie ma natomiast dla energii aktywacji – ekstrapolowana metodą ONIOM bez refitowania ładunków bariera (pomijając poprawkę na energię drgań zerowych) wynosiła 10.20 kcal/mol, natomiast po dostosowaniu ładunków regionu QM 7.46 kcal/mol.

Znalezienie punktu siodłowego potwierdzono przez obliczenie drgań normalnych – znajdowała się w nich tylko jedna częstość urojona (moduł częstości 914.59  $\text{cm}^{-1}$ ) odpowiadająca przenoszeniu protonu (Rysunek S.7). Drugim istotnym elementem rzeczonoego drgania jest rozciąganie wiązania  $O_0$ - $N_4$  wraz z

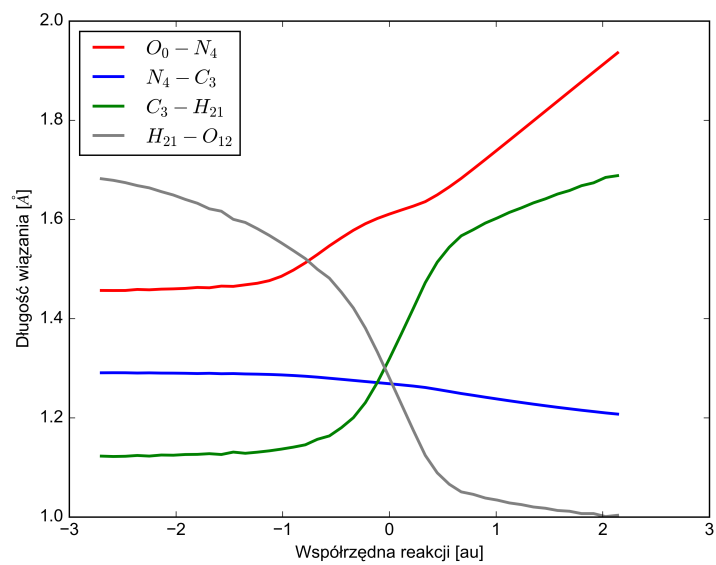
odrywaniem atomu wodoru  $H_{21}$  przez tlen  $O_{12}$  Glu101. Wiązanie to ulega osłabieniu już na początku reakcji (Rys. S.8).



Rysunek S.6: Porównanie struktur: substratu (przezroczysty niebieski), pierwotnego stanu przejściowego (czerwony) oraz TS po reflowaniu ładunków (zielony).

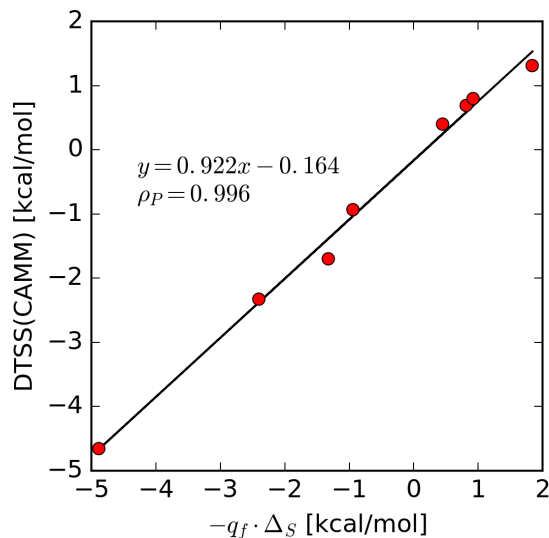


Rysunek S.7: Drganie odpowiadające częstości urojonej oraz przyjęta numeracja atomów.

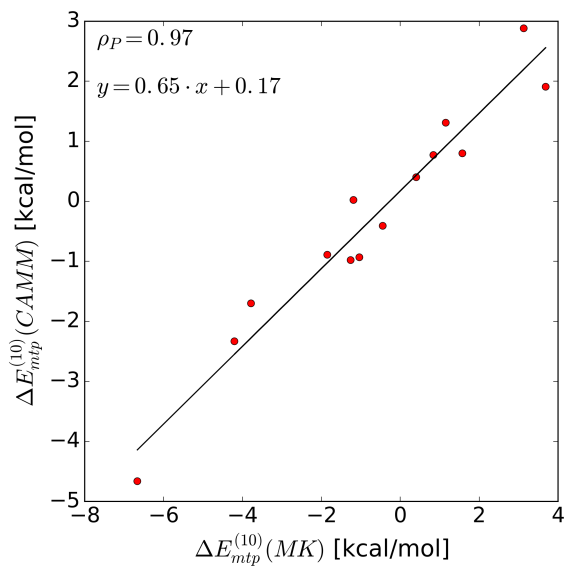


Rysunek S.8: Wykres zmian długości wybranych wiązań na ścieżce reakcji.

## S.6 Reprezentacja $\Delta_S$



Rysunek S.9: Iloczyn ładunku formalnego i pola katalitycznego (obliczonego za pomocą CAMM) dla naładowanych aminokwasów w szeregu mutantów eliminazy Kempa KE07 w porównaniu z  $E_{mtp}^{(10)}$  (vide sekcja 3.1.2).



Rysunek S.10: Porównanie  $\Delta E_{mtp}^{(10)}$  wyznaczonego z ładunków Merza-Kollmana [137] z wartością obliczoną za pomocą CAMM.

## S.7 Model ciągłego rozpuszczalnika

Za punkt wyjścia do uwzględnienia efektów rozpuszczalnikowych w obliczeniach DTSS(CAMM) posłużył uogólniony model Borna (S.34), który można rozważać jako analityczne przybliżenie rozwiązania równania Poissona-Boltzmanna. Onufriev wraz ze współpracownikami wprowadził szereg poprawek do

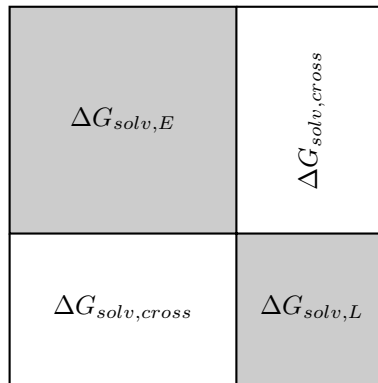
tego modelu, m.in. uwzględniając zależność nie tylko od lokalnego otoczenia (zawartą w efektywnych promieniach Borna atomów  $\{R_i\}$  oraz funkcji  $f_{ij}$  w równaniu (S.34)), ale także od globalnego kształtu cząsteczki (*Analytic Linearized Poisson-Boltzmann*, ALPB [168]) oraz znaku ładunku atomu [154]. Jednakże testy na analizowanych białkach wykazały, że poprawka globalna z [168] nie jest istotna. Warto zaznaczyć, że w prowadzonych testach wykorzystano ulepszoną procedurę wyznaczania efektywnych promieni Borna [153, 169].

Obliczenia efektywnych promieni Borna i energii solwatacji w modelu ALPB wykonano za pomocą programu *bornRadius* udostępnionym przez zespół Onufrieva.

$$\Delta G_{solv} = -\frac{1}{2} \left( \frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}} \right) \sum_{i,j} \frac{q_i q_j}{f_{ij}} \quad (\text{S.34})$$

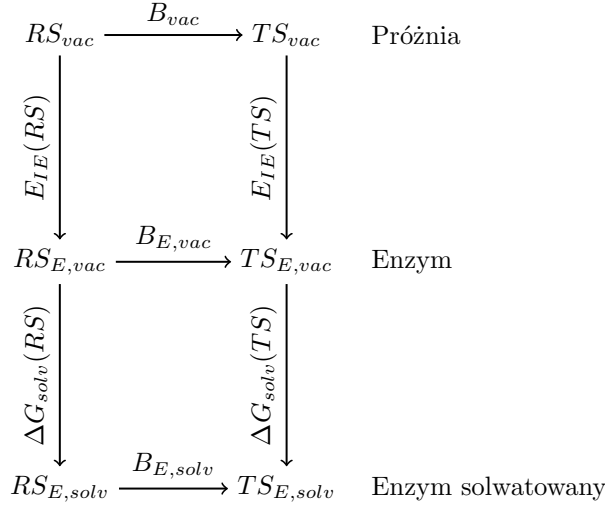
$$f_{ij} = \sqrt{r_{ij}^2 + R_i R_j \exp \left( -\frac{r_{ij}^2}{4R_i R_j} \right)}$$

Zanim dostosujemy równanie (S.34) do obliczeń energii DTSS za pomocą rozwinięcia multipolowego CAMM dla układu białko-ligand, należy zauważyć, że sumę po wszystkich parach indeksów  $i$  i  $j$  możemy pojmować jako sumowanie wszystkich elementów macierzy  $\mathbf{G}_{solv}$  o elementach przez te indeksach zdefiniowanych. Ponieważ kolejność sumowania nie gra roli, niech pierwsze  $N_P$  indeksów numeruje atomy białkowe, a kolejnych  $N_L$  atomy reagenta (liganda). Wówczas w macierzy  $\mathbf{G}_{solv}$  możemy wyróżnić dwa rodzaje bloków (Rysunek S.11): dwa diagonalne bloki (zaznaczone na Rysunku na szaro), których elementy zawierają tylko atomy enzymu ( $i, j \leq N_P$ ) lub liganda ( $N_P < i, j \leq N_P + N_L$ ), oraz dwa identyczne bloki pozadiagonalne zawierające człony mieszane. Suma członów mieszanych w (S.34) jest bardzo podobna do wzoru Coulomba na energię oddziaływania pomiędzy dwoma zestawami ładunków, z czego zaraz skorzystamy.



Rysunek S.11: Bloki macierzy  $\mathbf{G}_{solv}$

Rozważmy cykl termodynamiczny przedstawiony na Rys. S.12. Pierwsze dwa piętra (licząc od góry) prowadzą do omawianej wcześniej energii DTSS. Uwzględnienie solwatacji (ostatnie piętro) prowadzi do wzoru (S.35). Różnicową energię solwatacji  $\Delta \Delta G_{solv}$  można uprościć, spoglądając na nią jako na sumę elementów macierzy  $\mathbf{G}_{solv}(TS) - \mathbf{G}_{solv}(RS)$ , przyjmując taką samą kolejność numeracji atomów w obu składnikach sumy oraz patrząc na ich strukturę blokową (Rys. S.11 i równanie (S.36)). Jeżeli zmiana położenia atomów białkowych pomiędzy substratem a TS jest znikoma (co jest jak najbardziej uzasadnione w przypadku rozważanych w projekcie reakcji przeniesienia protonu), wówczas diagonalny blok odpowiadający atomom enzymu w przybliżeniu się zeruje. Jeśli idzie o blok liganda, to biorąc pod uwagę, że mutacje analizowane w rozważanym enzymie znajdowały się w drugiej sferze koordynacyjnej, można go przyjąć za w przybliżeniu taki sam w obrębie rozważanych mutantów. Tym samym naprawdę istotnym czynnikiem pozostaje człon mieszany.



Rysunek S.12: Cykl termodynamiczny

$$B_{E,solv} = B_{vac} + E_{IE}(TS) - E_{IE}(RS) + \Delta G_{solv}(TS) - \Delta G_{solv}(RS) = B_{vac} + E_{DTSS} + \Delta\Delta G_{solv} \quad (S.35)$$

$$\Delta\Delta G_{solv} = \Delta\Delta G_{solv,L} + \Delta\Delta G_{solv,E} + 2 * \Delta\Delta G_{solv,cross} \approx const + 0 + 2 \cdot \Delta\Delta G_{solv,cross} \quad (S.36)$$

$$E_{IE,q-q} + 2 * \Delta G_{solv,cross} = \sum_{i<j} \frac{q_i q_j}{r_{ij}} - \left( \frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}} \right) \sum_{i<j} \frac{q_i q_j}{f_{ij}} = \sum_{i<j} \frac{q_i q_j}{r_{ij}} \left[ 1 - \left( \frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}} \right) \frac{r_{ij}}{f_{ij}} \right] \\ = \sum_{i<j} \frac{q_i q_j}{\varepsilon_{ij} r_{ij}} \quad (S.37)$$

$$\text{gdzie } \varepsilon_{ij}^{-1} = 1 - \left( \frac{1}{\varepsilon_{in}} - \frac{1}{\varepsilon_{out}} \right) \frac{r_{ij}}{f_{ij}}$$

Dzięki temu zabiegowi można wprowadzić efektywne międzyatomowe stałe dielektryczne poprzez połączenie członu oddziaływania elektrostatycznego z członem mieszanym energii solwatacji (S.37). Tak więc przy założeniu niezmienności energii solwatacji białka pomiędzy RS a TS oraz jego pomijalnej zmianie geometrii udało się sprowadzić wyrażenie (S.35) do (S.38).

$$B_{E,solv} = B_{vac} + E_{DTSS,diel} + const \sim E_{DTSS,diel} \quad (S.38)$$

Przeprowadzenie podobnego rozważania dla multipoli CAMM formalnie wymagałoby wyprowadzenia wzoru (S.34) dla szeregu multipolowego. Ponieważ jednak wykracza to poza założenia projektu, postanowiono potraktować efektywne stałe dielektryczne  $\varepsilon_{ij}$  (S.37) jako czynniki skalujące przyczynki od atomów  $i$  i  $j$  do energii oddziaływania, tym samym przekształcając równanie (S.39) do (S.40).

$$\Delta E_{CAMM}^{k \leq L} = \sum_{a \in A} \sum_{\substack{b \in B \\ k_a + k_b \leq L}} \sum_{k_a} \sum_{k_b} \mathbf{M}_a^{(k_a)}[k_a] \mathbf{T}^{(k_a + k_b)}[k_b] \mathbf{M}_b^{(k_b)} \quad (S.39)$$

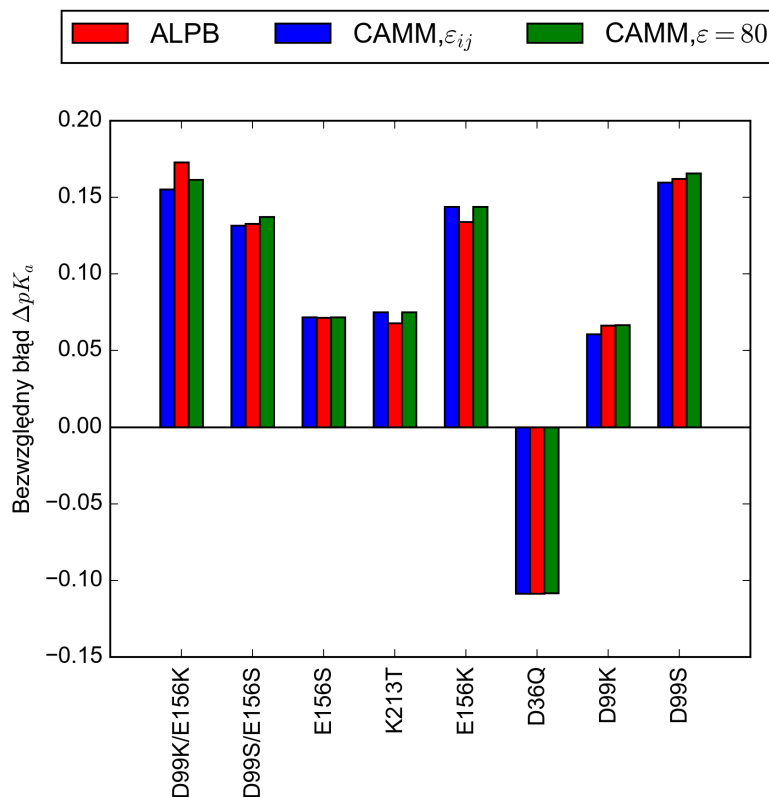
$$\Delta E_{CAMM, \varepsilon_{ab}}^{k \leq L} = \sum_{a \in A} \sum_{b \in B} \varepsilon_{ab}^{-1} \sum_{\substack{k_a \\ k_b \\ k_a + k_b \leq L}} \sum_{k_a} \sum_{k_b} \mathbf{M}_a^{(k_a)}[k_a] \mathbf{T}^{(k_a + k_b)}[k_b] \mathbf{M}_b^{(k_b)} \quad (S.40)$$

### S.7.1 Walidacja: subtylizyna

Równania (S.37) i (S.40) mogą znaleźć zastosowanie, kiedy w grę wchodzi dowolna energia różnicowa. Przykładem może być zmiana  $pK_a$  jednego z aminokwasów w białku pod wpływem mutacji.



Jako system testowy wybrano pracę opisującą zmiany  $pK_a$  His64 subtylizyny pod wpływem mutacji w innych pozycjach białka [170]. W tym wypadku wyniki uzyskane dla próżni wykazują pozytywną korelację z eksperymentem ( $\rho_P = 0.98, \rho_S = 1.0$ ), a model z międzyatomowymi stałymi dielektrycznymi daje minimalnie lepsze rezultaty. Praktycznie wszystkie analizowane w tym teście modele dają ten sam wynik (Rysunek S.13). Jedną z przyczyn obserwowanych rozbieżności może być reprezentowanie każdego wariantu enzymu pojedynczą strukturą. Jednakże celem testu nie było lepsze odtworzenie wartości eksperymentalnych w porównaniu z innymi modelami, lecz zweryfikowanie poprawności modelu CAMM z międzyatomowymi stałymi dielektrycznymi; wybrano w tym celu przykładowy zestaw naładowanych elektrycznie reszt, gdyż wówczas można się spodziewać zbieżności wyników pochodzących z rozwinięcia multipolowego z modelem ładunkowym, na którym opierają się dostępne metody obliczenia energii solwatacji.



Rysunek S.13: Błąd wyznaczenia zmiany  $pK_a$  pod wpływem mutacji. *ALPB*- analityczne zlinearyzowane równanie Poissona-Boltzmanna [168], *CAMM, $\epsilon = 80$* - CAMM przeskalowane przez stałą dielektryczną wody(80), *CAMM, $\epsilon_{ij}$* - CAMM z międzyatomowymi stałymi dielektrycznymi.

## S.8 Biblioteka multipoli atomowych dla aminokwasów

### S.8.1 Motywacja utworzenia biblioteki

Wykonane dotąd analizy aktywności katalitycznej przy zastosowaniu teorii oddziaływań międzycząsteczkowych w ramach modelu DTSS wskazują na dominującą predykcijną rolę multipolowej składowej oddziaływań elektrostatycznych [37–39].

Zastosowanie energii oddziaływania elektrostatycznego jako predyktora energii wiązania lub aktywności katalitycznej ma dwie podstawowe zalety: nieempiryczny charakter oraz bardzo niskie skalowanie czasu obliczeniowego wraz z rozmiarem układu. Pierwsza cecha wiąże się z wyprowadzeniem modelu z pierwszych zasad (*ab initio*), w przeciwieństwie do szeregu metod stosowanych w modelowaniu makromolekuł, zależnych od parametrów dopasowywanych tak, by model jak najlepiej odtwarzał dane

eksperymentalne. Ponadto metodologia *ab initio* pozwala na systematyczną poprawę modelu np. w ramach rachunku zaburzeń [111]. Druga cecha jest szczególnie widoczna, gdy porównamy czas potrzebny na obliczenie energii oddziaływania za pomocą atomowego rozwinięcia multipolowego z innymi metodami *ab initio*. Przyjmując następujące oznaczenia:  $A$  – liczba atomów,  $N$  – liczba funkcji bazy,  $O(f)$  – górne ograniczenie skalowania czasu obliczeniowego, zestawienie wybranych metod obliczeniowych wygląda następująco:

- CAMM i inne atomowe rozwinięcia multipolowe –  $O(A^2)$ ,
- Kompletna energia oddziaływań elektrostatycznych wraz ze składową penetracyjną –  $O(N^4)$ ,
- SCF (HF, DFT) –  $O(N^4)$ ,
- MP2 –  $O(N^5)$ ,
- metody sprzężonych klasterów (CC) –  $O(N^6)$  i więcej.

Należy dodać, że w powyższym zestawieniu nie uwzględniono czasu obliczeń niezbędnego do wyznaczenia atomowych multipoli rozważanych podukładów. Ta część procesu obliczeniowego stanowi główne ograniczenie zastosowania rozważanego modelu, o ile bowiem czas wyliczenia energii oddziaływania pomiędzy dwoma indywiduami mieści się w granicach kilku milisekund (dla rozwinięcia rzędu  $R^{-5}$ ), o tyle czas potrzebny na obliczenie funkcji falowej każdego z nich liczy się najczęściej w minutach. Jednakowoż, ponieważ w białkach będących głównym obiektem zainteresowania w projekcie występuje ograniczona ilość aminokwasów, całość obliczeń można by znacznie usprawnić, dysponując przeliczonymi dla nich wcześniej tablicami multipoli atomowych.

### S.8.2 Zależność konformacyjna CAMM

Problem przenaszalności konformacyjnej atomowych multipoli różnie jest traktowany w literaturze. Popelier [171, 172] używa sieci neuronowych oraz metod uczenia maszynowego do modelowania wartości poszczególnych momentów w przestrzeni konformacyjnej. Podejście to wymaga jednak dość gęstej siatki punktów referencyjnych, by uzyskać wynik z rozsądną dokładnością. Z kolei Jakobsen i Jensen [138] wyznaczają najbardziej istotne atomowe momenty multipolowe pod kątem ich zdolności do odtwarzania potencjału elektrostatycznego w różnych konformacjach. Jak dotąd zastosowali oni swoją analizę do kilku zaledwie aminokwasów o niedużej przestrzeni konformacyjnej. Istotną trudnością w ich metodzie jest szybko rosnąca złożoność obliczeniowa związana z koniecznością analizy różnych kombinacji atomowych multipoli. Momenty multipolowe CAMM, które są podstawowym rozwinięciem multipolowym wykorzystywanym w projekcie, mogą poprawnie odtwarzać zależność energii oddziaływania od zmian kąta torsyjnego [173, 174], jednak trudno uznać je za w pełni przenaszalne pomiędzy konformerami [175]. Postanowiono więc skorzystać z biblioteki rotamerów Dynameonics [131], w pierwszej kolejności skupiając się na rotamerach niezależnych od konformacji głównego łańcucha polipeptydowego. Dla każdego rotameru z bazy obliczono multipole CAMM z gęstości elektronowej Hartree-Focka w bazach 6-311G(3d,2p) oraz 6-31G(d).

### S.8.3 Rodzaje interpolacji – testy 1D i 2D

Ponieważ w konkretnej strukturze kąty dwuściennie definiujące dany rotamer mogą różnić się od wartości średniej podanej w bazie [131], rozważono kilka mechanizmów transformacji rotamerów bazy.

1. *Najbliższy rotamer.* Wybranie rotameru najbliższego danemu, następnie wykonanie kilku obrotów celem dopasowania kątów dwuściennych definiujących rotamer.
2. *Interpolacja liniowa.* Wybranie rotamerów 'oskrzydlających' dany względem każdego kąta dwuściennego (w ilości  $2^{N_{rot}}$ , gdzie  $N_{rot}$  oznacza ilość kątów dwuściennych), a następnie liniowa interpolacja pomiędzy nimi.
3. *Interpolacja trygonometryczna.* Podobnie do punktu 2, lecz z wykorzystaniem szeregów Fouriera.

Niezależnie od sposobu interpolacji, wstępnie należy dla każdego rotameru wybranego z bazy dokonać  $N_{rot}$  obrotów względem kolejnych kątów dwuściennych w łańcuchu bocznym, by dopasować jego

konformację do wartości docelowej. Jest to istotne, gdyż obrotom podlegają nie tylko współrzędne atomów, ale także osadzone na nich tensory momentów multipolowych. Interpolacja samych tablic opisujących multipole bez uprzedniej zmiany orientacji wiązałyby się z wprowadzeniem dodatkowego źródła błędu.

We wszystkich schematach interpolacji, każdy kąt dwuścienny definiujący rotamer traktowany jest jako zmienna niezależna. Na obecnym etapie prac pominięto kwestię konformacji łańcucha głównego (kątów Ramachandrana  $\psi$  i  $\phi$ ), skupiając się jedynie na łańcuchu bocznym aminokwasów.

### Interpolacja liniowa.

Niech  $\{x_i\}$  oznaczają argumenty znanych punktów funkcji, zaś  $\{y_i\}$  ich wartości. Za punkt wyjściowy przyjmijmy formułę Lagrange'a (równanie S.41) [176].

$$p_{int}(x) = \sum_{i=0}^N y_i l_i(x) \quad (\text{S.41})$$

$$\text{gdzie } l_i(x) = \prod_{i \neq j} \frac{x - x_j}{x_i - x_j}$$

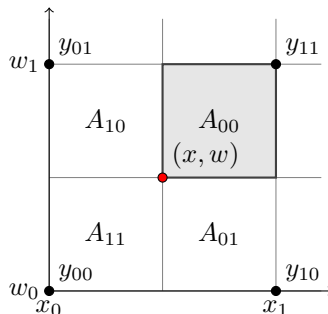
W przypadku jednowymiarowej interpolacji pomiędzy dwoma punktami, równanie (S.41) redukuje się do (S.42).

$$p_1(x) = y_0 \frac{x - x_1}{x_0 - x_1} + y_1 \frac{x - x_0}{x_1 - x_0} = y_0 \frac{x_1 - x}{x_1 - x_0} + y_1 \frac{x - x_0}{x_1 - x_0} \quad (\text{S.42})$$

Formułę (S.41) łatwo jest uogólnić na przypadek dwuwymiarowy (S.43) i analogicznie na wyższe ilości wymiarów.

$$p_{int}(x, w) = \sum_{i=0}^N \sum_{j=0}^M y_{ij} l_i(x) l_j(w) \quad (\text{S.43})$$

W zastosowaniu interpolacji multipoli warto posłużyć się interpretacją geometryczną (Rys. S.14). Iloczyn wielomianów  $l_i(x)$  i  $l_j(w)$  można interpretować jako stosunek powierzchni  $A_{ij}$  do całkowitej powierzchni prostokąta  $A_{00} + A_{01} + A_{10} + A_{11}$ . Należy zwrócić uwagę, że prostokąt determinujący udział wartości funkcji  $y_{ij}$  w sumie ważonej leży *naprzeciwko* tego wierzchołka względem punktu, w którym szukamy wartości interpolowanej. Powyższą analogię łatwo jest przenieść na wyższe wymiary, zastępując powierzchnie objętościami, hiperobjętościami itd. Należy zaznaczyć, że analogia ta dotyczy jednorodnej siatki, tzn. równomiernie rozłożonych rotamerów. W przypadku rzeczywistym, w którym rozkład ten jest zaburzony, można potraktować analogię geometryczną jako przybliżenie lub skorzystać z innego schematu interpolacji, np. Sheparda.



Rysunek S.14: Geometryczna interpretacja dwuwymiarowej interpolacji liniowej

### Interpolacja trygonometryczna w jednym wymiarze.

Dla nieparzystej ilości punktów referencyjnych  $N = 2K + 1$ , połączenie formuły Lagrange'a (S.41) z zespoloną postacią szeregu Fouriera prowadzi do wzoru na interpolację (S.44).

$$p(x) = \sum_{k=0}^{2K} y_k t_k(x) \quad (\text{S.44})$$

$$\text{gdzie } t_k(x) = \prod_{m=0, m \neq k}^{2K} \frac{\sin \frac{1}{2}(x - x_m)}{\sin \frac{1}{2}(x_k - x_m)}$$

Dla równomiernie rozmieszczonych punktów upraszcza się to do (S.45).

$$t_k = \frac{\text{sinc} \frac{1}{2} N(x - x_k)}{\text{sinc} \frac{1}{2}(x - x_k)} \quad (\text{S.45})$$

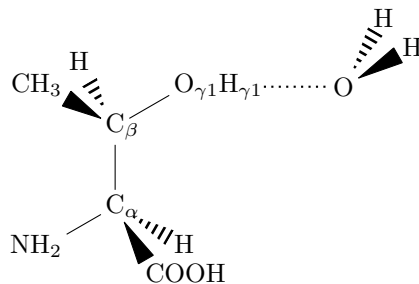
$$\text{gdzie } \text{sinc}(x) = \frac{\sin(x)}{x}$$

Dla parzystej ilości równomiernie rozmieszczonych punktów wzór (S.45) przechodzi w (S.46).

$$t_k = \frac{\text{sinc} \frac{1}{2} N(x - x_k)}{\text{sinc} \frac{1}{2}(x - x_k)} \cos \frac{1}{2}(x - x_k) \quad (\text{S.46})$$

### Porównanie algorytmów w jednym wymiarze.

Aby porównać ze sobą wymienione w poprzedniej podsekcji algorytmy, przeprowadzono test w układzie jednowymiarowym. Jako układ testowy wybrano kompleks treoniny z wodą wedle schematu na rys. S.15; wszystkie energie oddziaływania uzyskano z rozwinięcia multipolowego CAMM( $R^{-5}$ ) wyznaczonego z gęstości HF/6-311G\*\*. Za wartość kąta dwuściennego N-C $_{\alpha}$ -C $_{\beta}$ -O $_{\gamma 1}$  przyjęto  $-60.85^{\circ}$ ; energię oddziaływania treonina-woda analizowano jako funkcję kąta C $_{\alpha}$ -C $_{\beta}$ -O $_{\gamma 1}$ -H $_{\gamma 1}$  (nazwijmy go  $\theta$ ), utrzymując geometrię samego wiązania wodorowego. Kąt  $\theta$  zmieniano co  $10^{\circ}$ , dla każdej wartości wyliczając momenty multipolowe CAMM dla treoniny. Jako punkty referencyjne dla procedur interpolacji przyjęto zestawy rotamerów  $\theta \in \{0^{\circ}, 120^{\circ}, 240^{\circ}\}$  i  $\theta \in \{0^{\circ}, 60^{\circ}, 120^{\circ}, 180^{\circ}, 240^{\circ}, 300^{\circ}\}$ . Wynik przedstawia Tabela S.4.



Rysunek S.15: Kompleks treoniny z wodą używany do testów interpolacji multipoli.

Jak można się spodziewać, wykorzystanie większej ilości punktów referencyjnych daje wynik obarczony prawie dwukrotnie mniejszym błędem. Widać także, że różnice pomiędzy różnymi metodami interpolacji nie są duże, szczególnie interpolacja liniowa i trygonometryczna prowadzą do zbliżonych wyników. Chociaż interpolacja wykorzystująca zaledwie 3 struktury odniesienia jest wyraźnie gorsza od tej wykorzystującej 6, wydaje się ona dobrym kompromisem pomiędzy dokładnością obliczeń a wielkością bazy, zwłaszcza jeśli uwzględnić fakt, że odchylenia standardowe kątów dwuściennych w bazie Dynameonics mieszczą się w granicach kilkunastu stopni. Rysunek S.16 pokazuje, że dla siatki trójpunktowej błędy poszczególnych technik interpolacji w zakresie do  $20^{\circ}$  od poszczególnych struktur referencyjnych (jasno zielone prostokąty na wykresie) mieszczą się w granicach 0.2–0.3 kcal/mol (2–3% energii oddziaływania).

Niestety, zarówno interpolacja liniowa jak i trygonometryczna nie odtwarzają poprawnie całkowitego momentu dipolowego treoniny. W przeprowadzonym teście jako miary błędu interpolowanego momentu dipolowego względem wartości referencyjnej użyto dwóch wielkości: różnicy długości wektorów  $\Delta\mu = |\vec{\mu}| - |\vec{\mu}_{ref}|$  oraz kąta  $\alpha$  pomiędzy nimi. Wyniki prezentuje Tabela S.5.

Tabela S.4: Statystyki testu z treoniną

Tryb	$N_{grid}$	RMSD [kcal/mol]	MaxAE [kcal/mol]
Najbliższy	6	0.18	0.42
Int. liniowa	6	0.09	0.20
Int. trygonometryczna	6	0.08	0.27
Najbliższy	3	0.34	0.78
Int. liniowa	3	0.31	0.59
Int. trygonometryczna	3	0.32	0.66

$N_{grid}$ - ilość punktów siatki, RMSD- średni błąd kwadratowy, MaxAE- maksymalny błąd bezwzględny

Tabela S.5: Błędy interpolowanego momentu dipolowego

		Int. liniowa	Int. trygonometryczna
RMSD( $\Delta\mu$ )	[D]	0.47	0.92
MAPE( $\Delta\mu$ )	[%]	14.03	34.54
MaxAE( $\Delta\mu$ )	[D]	1.07	1.68
MaxAPE( $\Delta\mu$ )	[%]	48.00	103.52
$\bar{\alpha}$	[°]	9.56	27.90
$\sigma(\alpha)$	[°]	8.08	18.59

RMSD- średni błąd kwadratowy, MAPE- średni bezwzględny błąd procentowy, MaxAE- maksymalny błąd bezwzględny, MaxAPE- maksymalny bezwzględny błąd procentowy

Porównując wskaźniki dla obu metod interpolacji, widać, że pod tym względem odtwarzania całkowitego momentu dipolowego interpolacja trygonometryczna radzi sobie znacznie gorzej od liniowej. Aby odpowiedzieć na pytanie, dlaczego w tym teście interpolowane multipole CAMM wypadły zdecydowanie gorzej od testu energii oddziaływania, dla interpolacji liniowej dokonano dekompozycji powyższych wskaźników. Wykorzystano fakt, że całkowity moment dipolowy zestawu atomowych multipoli CAMM  $\{q_i, \vec{\mu}_i, \theta_i \dots\}$  (gdzie  $i$  indeksuje atomy o współrzędnych  $\vec{r}_i$ ) można przedstawić jako sumę (S.47), gdzie  $\vec{M}_q$  oznacza składową “skalarną”, a  $\vec{M}_\mu$  “wektorową”. Wskaźniki błędów każdej z tych składowych względem odpowiednich wartości odniesienia przedstawia Tabela S.6

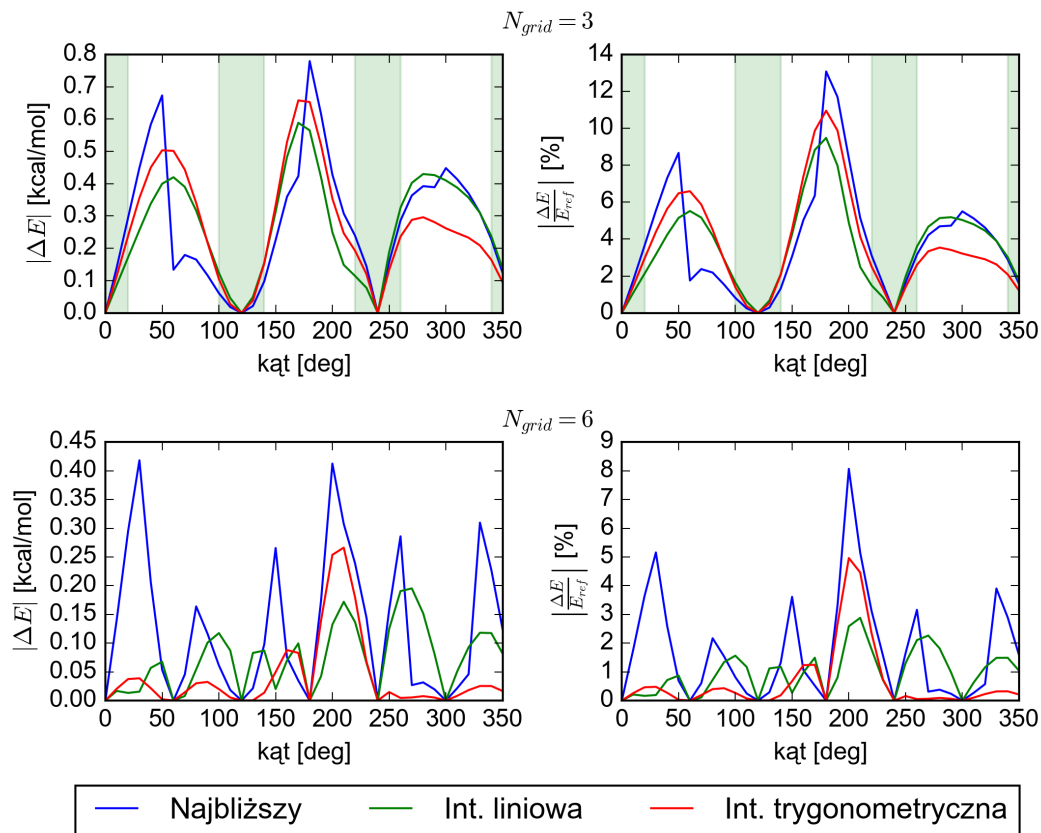
$$\vec{\mu}_{mol} = \sum_i^{N_{at}} q_i \vec{r}_i + \sum_i^{N_{at}} \vec{\mu}_i = \vec{M}_q + \vec{M}_\mu \quad (\text{S.47})$$

Tabela S.6: Błędy interpolowanego momentu dipolowego

		$\vec{M}_q$	$\vec{M}_\mu$
RMSD( $\Delta\mu$ )	[D]	0.48	0.04
MAPE( $\Delta\mu$ )	[%]	21.41	2.05
MaxAE( $\Delta\mu$ )	[D]	1.06	0.10
MaxAPE( $\Delta\mu$ )	[%]	88.39	7.08
$\bar{\alpha}$	[°]	14.55	2.40
$\sigma(\alpha)$	[°]	13.32	1.42

RMSD- średni błąd kwadratowy, MAPE- średni bezwzględny błąd procentowy, MaxAE- maksymalny błąd bezwzględny, MaxAPE- maksymalny bezwzględny błąd procentowy

Na podstawie tego porównania za przyczynę złego odtwarzania całkowitego momentu dipolowego przez interpolowane multipole CAMM można uznać silną zależność ładunków Mullikena od kąta dwu-



Rysunek S.16: Wykres błędów interpolacji w teście z treoniną.

ściennego. Wiąże się to z tym, że ładunek atomowy nie jest wartością spodziewaną (podobnie jak każdy inny multipol atomowy) [177]. Stąd należy z ostrożnością podchodzić do każdej analizy populacyjnej. W testach uwzględniających energię oddziaływania błąd ten najwidoczniej był kompensowany przez wyższe multipole (w prezentowanych zestawieniach do heksadekapoli włącznie). Możliwe, że zmiana analizy populacyjnej w definicji multipoli CAMM mogłaby poprawić ten wynik, wykracza to jednak poza zakres prezentowanych badań. Należy przy tym zaznaczyć, że pomimo arbitralności w podziale gęstości elektronowej pomiędzy atomy, niezależnie od przyjętego schematu granica szeregu odpowiada wartości spodziewanej energii oddziaływania elektrostatycznego.

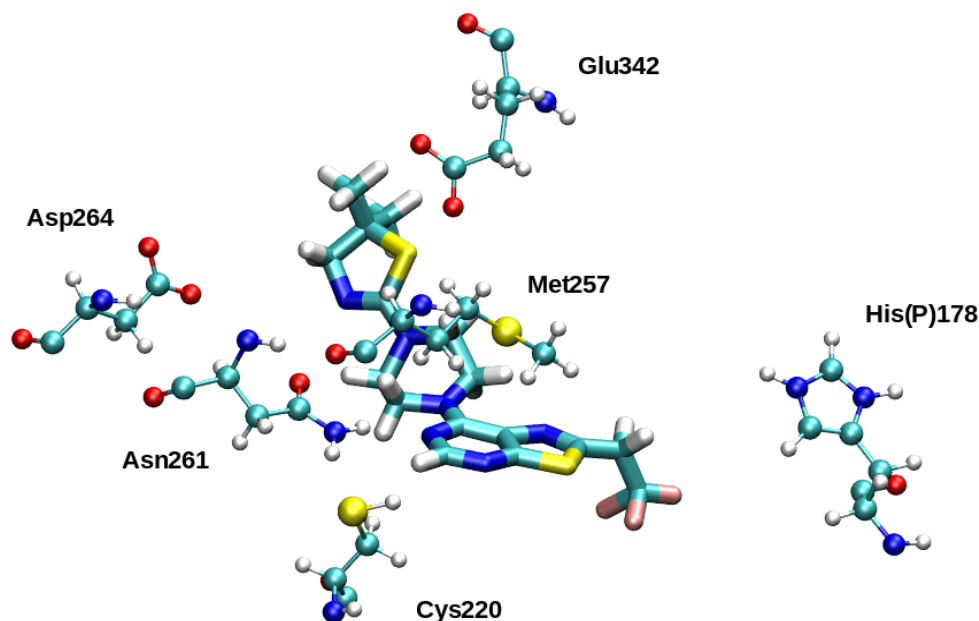
Kolejny test, jaki przeprowadzono na tym układzie, to test na próbce 200 losowo wygenerowanych rotamerów treoniny. Kąty dwuścienne losowano w dwóch etapach, zgodnie z rozkładem zdefiniowanym w bazie Dymaonics. W pierwszej kolejności rotamer zgodnie z prawdopodobieństwem jego wystąpienia; następnie kąty definiujące rotamer zostały wylosowane zgodnie z rozkładem normalnym zdefiniowanym przez ich wartości średnie i odchylenia standardowe ( w przypadku wylosowania wartości spoza przedziału definiującego rotamer, przyjęto wartość graniczną pomniejszoną o  $1^\circ$ ). Średnie błędy kwadratowe dla interpolacji liniowej (pomiędzy rotamerami w bazie, a więc trójpunktowej) i wyboru najbliższego rotameru na tej próbce wyniosły odpowiednio 0.10 i 0.08 kcal/mol, natomiast maksymalne błędy bezwzględne 0.21 i 0.17 kcal/mol.

#### S.8.4 Testy 2D i 3D

Aby sprawdzić działanie proponowanego algorytmu na bardziej złożonym przypadku, wybrano kilka aminokwasów z centrum aktywnego meniny wraz z jednym z jej inhibitorów (Rysunek S.17) [178]. Pominięto te reszty aminokwasowe, które oddziaływały z inhibitorem poprzez łańcuch główny, gdyż na tym etapie w centrum uwagi postawiono konformację łańcucha bocznego. Ze względu na rozmiar liganda, użyto funkcji falowych uzyskanych metodą HF/6-31G\*. Spośród wybranych aminokwasów, kwas

glutaminowy i metionina mają trzy stopnie swobody, pozostałe zaś reszty dwa. Jak pokazują Tabele S.8 i S.7, błąd energii oddziaływania w rozważanym przykładzie jest podobny dla wszystkich rozważanych aminokwasów, niezależnie od siły ich oddziaływania. Błąd interpolacji w znikomym stopniu zależy od wyboru techniki.

W obecnej implementacji, czas wczytania 'najbliższego' rotameru wraz z korektą kątów dwusiecznych jest rzędu kilku sekund, natomiast w przypadku interpolacji liniowej lub Sheparda jest on średnio  $2^{N_{rot}}$  razy dłuższy. Wciąż jednak jest to czas konkurencyjny w stosunku do obliczeń funkcji falowej dla danego aminokwasu.



Rysunek S.17: Wybrane aminokwasy (kule i cylindry) w centrum aktywnym meniny wraz z inhibitorem (tuby).

Tabela S.7: Statystyki testu z meniną [kcal/mol]

	MaxAPE	MaxAE	MAPE	RMSD
Najbliższy	55.416	0.370	20.648	0.234
Int. liniowa	44.081	0.369	18.722	0.232
Int. Sheparda	41.310	0.377	18.029	0.244

RMSD- średni błąd kwadratowy, MAPE- średni bezwzględny błąd procentowy, MaxAE- maksymalny błąd bezwzględny, MaxAPE- maksymalny bezwzględny błąd procentowy

Tabela S.8: Błędy energii oddziaływania z inhibitorem interpolowanych aminokwasów  $\Delta E_i = E_i - E_{ref}$  (wszystkie wartości w kcal/mol).

Aminokwas	$E_{ref}$	$\Delta E_{Najblizy}$	$\Delta E_{liniowa}$	$\Delta E_{Sheparda}$
H178	2.834	0.150	0.185	0.153
C220	-0.397	0.220	0.175	0.164
M257	-1.494	-0.223	-0.190	-0.143
N261	-4.181	-0.251	-0.280	-0.369
D264	1.845	-0.085	-0.088	-0.087
E342	-0.983	0.370	0.369	0.377

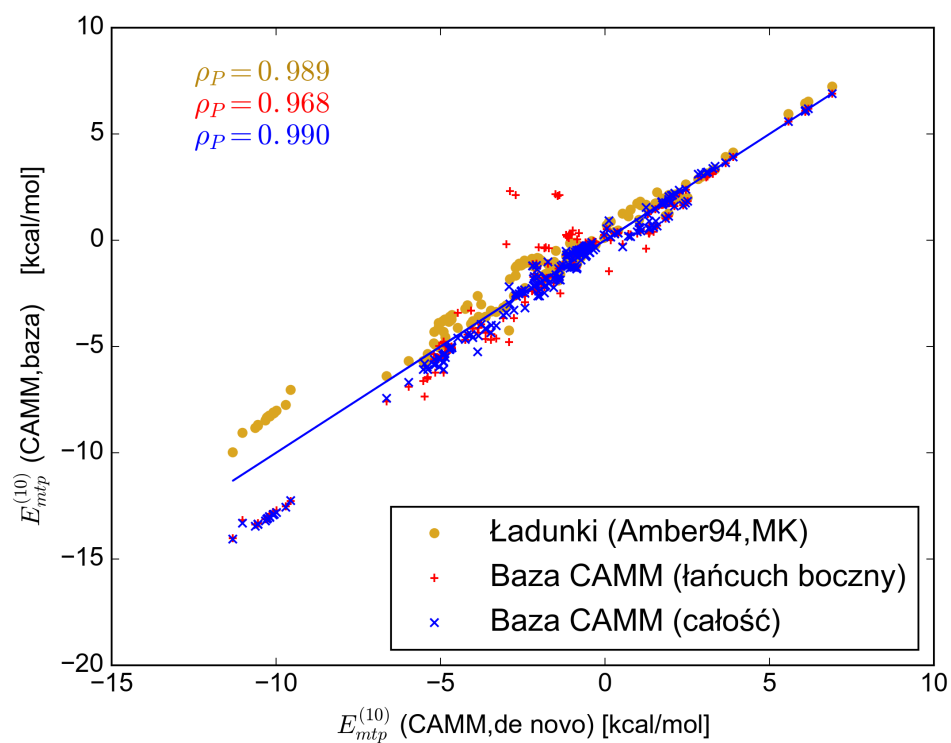
### S.8.5 Uwzględnienie konformacji łańcucha głównego

Przedstawione powyżej wyniki dotyczyły interpolacji łańcucha bocznego aminokwasów; w zestawie testowym pominięto z tego powodu reszty, które oddziaływały z inhibitorami poprzez łańcuch główny. Aby to poprawić, ogólnie rzecz biorąc, należałoby rozważyć wersję bazy zależną od konformacji łańcucha głównego i rozszerzyć schemat interpolacji o dwa kolejne stopnie swobody. Jako rozwiązanie tymczasowe przygotowano „łatkę” wykorzystującą reszty glicyny. Przygotowano zestaw konformacji tego aminokwasu zmieniając każdy z kątów dwuściennych co  $60^\circ$ . Następnie dla każdej reszty fituje się glicynę z bazy dwukrotnie: raz do docelowej konformacji łańcucha głównego (*target*), drugi raz do konformacji tegoż łańcucha, jaki ma dany aminokwas z bazy (*base*). Energię oddziaływania z taką resztą (*aa*) szacuje się w sposób przedstawiony w równaniu (S.48).

$$E_{int}(aa, target) \approx E_{int}(aa, base) + E_{int}(Gly, target) - E_{int}(Gly, base) \quad (S.48)$$

Porównanie obu metod — pierwotnej i z proponowaną poprawką — przedstawia Rysunek S.18. Można zaobserwować znaczną poprawę opisu oddziaływań aminokwasów słabiej oddziałujących z inhibitorami z rozważanego zestawu, jak również pewną poprawę w stosunku do ładunków pola siłowego (aminokwasy) i Merza-Kollmana (inhibitory); wykorzystane tutaj zostały wszystkie aminokwasy centrum aktywnego.





Rysunek S.18: Porównanie pierwotnej wersji bazy (w której dostosowywany był tylko łańcuch boczny), wersji z poprawką na łańcuch główny, ładunkami atomowymi oraz energią CMM liczoną *de novo*. Niebieska linia dana jest równaniem  $y = x$ . Dodatkowo zamieszczono także wartości współczynników korelacji Pearsona, w kolorach odpowiadających danym zestawom danych.



# Bibliografia

- [1] R. D. Astumian. How molecular motors work - insights from the molecular machinist's toolbox: the Nobel prize in Chemistry 2016. *Chemical Science*, 8(2):840–845, 2017.
- [2] T. Eriksson, S. Bjorkman, and P. Hoglund. Clinical pharmacology of thalidomide. *European Journal of Clinical Pharmacology*, 57(5):365–376, 2001.
- [3] F. R. Bisogno, M. G. Lopez-Vidal, and G. de Gonzalo. Organocatalysis and biocatalysis hand in hand: combining catalysts in one-pot procedures. *Advanced Synthesis & Catalysis*, 359(12, SI):2026–2049, 2017.
- [4] A. Radzicka and R. Wolfenden. A proficient enzyme. *Science*, 267(5194):90–93, 1995.
- [5] M. Polanyi. Über Adsorptionskatalyse. *Zeitschrift für Elektrochemie*, 27:142–150, 1921.
- [6] L. Pauling. Molecular architecture and biological reactions. *Chemical Engineering and News*, 161:707–709, 1946.
- [7] L. Pauling. Chemical achievement and hope for the future. *American Scientist*, 36(1):51–58, 1948.
- [8] L. Pauling. The nature of forces between large molecules of biological interest. *Nature*, 161:707–709, 1948.
- [9] H. Eyring. The activated complex in chemical reactions. *Journal of Chemical Physics*, 3:107, 1935.
- [10] R. Wolfenden. Transition state analogues for enzyme catalysis. *Nature*, 223:704–705, 1969.
- [11] R. Blomberg, Hajo Kries, D. M. Pinkas, P. R. E. Mittl, M. G. Grutter, H. K. Privett, S. L. Mayo, and D. Hilvert. Precision is essential for efficient catalysis in an evolved kemp eliminase. *Nature*, 503(7476):418–421, 2013.
- [12] V. L. Schramm. Enzymatic transition states and transition state analog design. *Annual Review of Biochemistry*, 67(1):693–720, 1998.
- [13] T. Humphry, S. Iyer, O. Iranzo, J. R. Morrow, J. P. Richard, P. Paneth, and A. C. Hengge. Altered transition state for the reaction of an RNA model catalyzed by a dinuclear zinc(II) catalyst. *Journal of The American Chemical Society*, 130(52):17858–17866, 2008.
- [14] J. Bigeleisen and M. G. Mayer. Calculation of equilibrium constants for isotopic exchange reactions. *Journal of Chemical Physics*, 15(5):261–267, 1947.
- [15] J. Bigeleisen and M. Wolfsberg. Theoretical and experimental aspects of isotope effects in chemical kinetics. *Advances in Chemical Physics*, 1:15–76, 1958.
- [16] J. F. Morrison and W. W. Cleland. Isotope exchange studies of mechanism of reaction catalyzed by adenosine triphosphate - creatine phosphotransferase. *Journal of Biological Chemistry*, 241(3):673–&, 1966.
- [17] S. N. HO, H. D. Hunt, R. M. Horton, J. K. Pullen, and L. R. Pease. Site-directed mutagenesis by overlap extension using the polymerase chain-reaction. *Gene*, 77(1):51–59, 1989.

- [18] A. Tramontano, K. D. Janda, and R. A. Lerner. Catalytic antibodies. *Science*, 234(4783):1566–1570, 1986.
- [19] S. J. Pollack, J. W. Jacobs, and P. G. Schultz. Selective chemical catalysis by an antibody. *Science*, 234(4783):1570–1573, 1986.
- [20] E. Z. Eisenmesser, O. Millet, W. Labeikovsky, D. M. Korzhnev, M. Wolf-Watz, D. A. Bosco, J. J. Skalicky, L. E. Kay, and D. Kern. Intrinsic dynamics of an enzyme underlies catalysis. *Nature*, 438:117–121, 2005.
- [21] D. D. Boehr, D. McElheny, H. J. Dyson, and P. E. Wright. The dynamic energy landscape of dihydrofolate reductase catalysis. *Science*, 313(5793):1638–1642, 2006.
- [22] T. H. Rod, J. L. Radkiewicz, and C. L. Brooks. Correlated motion and the effect of distal mutations in dihydrofolate reductase. *Proceedings of The National Academy of Sciences of The United States of America*, 100(12):6980–6985, 2003.
- [23] E. Z. Eisenmesser, D. A. Bosco, M. Akke, and D. Kern. Enzyme dynamics during catalysis. *Science*, 295(5559):1520–1523, 2002.
- [24] J. S. Fraser, M. W. Clarkson, S. C. Degnan, R. Erion, D. Kern, and T. Alber. Hidden alternative structures of proline isomerase essential for catalysis. *Nature*, 462(7273):669–U149, 2009.
- [25] S. Nunez, D. Antoniou, V. L. Schramm, and S. D. Schwartz. Promoting vibrations in human purine nucleoside phosphorylase. A molecular dynamics and hybrid quantum mechanical/molecular mechanical study. *Journal of The American Chemical Society*, 126(48):15720–15729, 2004.
- [26] D. Antoniou, S. Caratzoulas, C. Kalyanaraman, J. S. Mincer, and S. D. Schwartz. Barrier passage and protein dynamics in enzymatically catalyzed reactions. *European Journal of Biochemistry*, 269(13):3103–3112, 2002.
- [27] A. Warshel and R. P. Bora. Perspective: Defining and quantifying the role of dynamics in enzyme catalysis. *Journal of Chemical Physics*, 144(18), 2016.
- [28] S. C. L. Kamerlin and A. Warshel. At the dawn of the 21st century: Is dynamics the missing link for understanding enzyme catalysis? *Proteins-Structure Function and Bioinformatics*, 78(6):1339–1375, 2010.
- [29] G. Li and Q. Cui. What is so special about Arg 55 in the catalysis of cyclophilin A? Insights from hybrid QM/MM simulations. *Journal of the American Chemical Society*, 125(49):15028–15038, 2003.
- [30] D. T. Major, A. Heroux, A. M. Orville, M. P. Valley, P. F. Fitzpatrick, and J. Gao. Differential quantum tunneling contributions in nitroalkane oxidase catalyzed and the uncatalyzed proton transfer reaction. *Proceedings of The National Academy of Sciences of The United States of America*, 106(49):20734–20739, 2009.
- [31] S. C. L. Kamerlin and A. Warshel. An analysis of all the relevant facts and arguments indicates that enzyme catalysis does not involve large contributions from nuclear tunneling. *Journal of Physical Organic Chemistry*, 23(7, SI):677–684, 2010.
- [32] M. H. M. Olsson and A. Warshel. Solute solvent dynamics and energetics in enzyme catalysis: The SN2 reaction of dehalogenase as a general benchmark. *Journal of the American Chemical Society*, 126(46):15167–15179, 2004.
- [33] A. Warshel. Electrostatic basis of structure-function correlation in proteins. *Accounts of Chemical Research*, 14(9):284–290, 1981.
- [34] A. Warshel, P. K. Sharma, Z. T. Chu, and J. Aqvist. Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: Phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry*, 46(6):1466–1476, 2007.

- [35] A. Warshel, P. K. Sharma, M. Kato, Y. Xiang, H. Liu, and M. H. M. Olsson. Electrostatic basis for enzyme catalysis. *Chemical Reviews*, 106(8):3210–3235, 2006.
- [36] P. K. Sharma, Z. T. Chu, M. H. M. Olsson, and A. Warshel. A new paradigm for electrostatic catalysis of radical reactions in vitamin b12 enzymes. *Proceedings of the National Academy of Sciences*, 104(23):9661–9666, 2007.
- [37] B. Szeferczyk, A. J. Mulholland, K. E. Ranaghan, and W. A. Sokalski. Differential transition-state stabilization in enzyme catalysis: Quantum chemical analysis of interactions in the chorismate mutase reaction and prediction of the optimal catalytic field. *Journal of the American Chemical Society*, 126(49):16148–16159, 2004.
- [38] P. Szarek, E. Dyguda-Kazimierowicz, A. Tachibana, and W. A. Sokalski. Physical nature of intermolecular interactions within camp-dependent protein kinase active site: Differential transition state stabilization in phosphoryl transfer reaction. *Journal of Physical Chemistry B*, 112(37):11819–11826, 2008.
- [39] W. Beker, M. van der Kamp, A. J. Mulholland, and W. A. Sokalski. Rapid estimation of catalytic efficiency by cumulative atomic multipole moments: application to ketosteroid isomerase mutants. *Journal of Chemical Theory and Computation*, 13:945–955, 2017.
- [40] S. D. Fried, S. Bagchi, and S. G. Boxer. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science*, 346(6216):1510–1514, 2014.
- [41] A. Natarajan, F. Yabukarski, V. Lamba, J. P. Schwans, F. Sunden, and D. Herschlag. Comment on "Extreme electric fields power catalysis in the active site of ketosteroid isomerase". *Science*, 349(6251), 2015.
- [42] S. D. Fried and S. G. Boxer. Response to comments on "Extreme electric fields power catalysis in the active site of ketosteroid isomerase". *Science*, 349(6251), 2015.
- [43] D. A. Kraut, P. A. Sigala, B. Pybus, C. W. Liu, D. Ringe, G. A. Petsko, and D. Herschlag. Testing electrostatic complementarity in enzyme catalysis: Hydrogen bonding in the ketosteroid isomerase oxyanion hole. *PLoS Biology*, 4(4):501–519, 2006.
- [44] A. Warshel, P. K. Sharma, Z. T. Chu, and J. Aqvist. Electrostatic contributions to binding of transition state analogues can be very different from the corresponding contributions to catalysis: Phenolates binding to the oxyanion hole of ketosteroid isomerase. *Biochemistry*, 46(6):1466–1476, 2007.
- [45] T. C. Bruice and S. J. Benkovic. Chemical basis for enzyme catalysis. *Biochemistry*, 39(21):6267–6274, 2000.
- [46] E. Y. Lau, K. Kahn, P. A. Bash, and T. C. Bruice. The importance of reactant positioning in enzyme catalysis: A hybrid quantum mechanics/molecular mechanics study of a haloalkane dehalogenase. *Proceedings of The National Academy of Sciences of The United States of America*, 97(18):9937–9942, 2000.
- [47] A. Shurki, M. Štrajbl, J. Villà, and A. Warshel. How much do enzymes really gain by restraining their reacting fragments? *Journal of the American Chemical Society*, 124(15):4097–4107, 2002.
- [48] W. P. Jencks. Binding-energy, specificity, and enzymic catalysis - Circe effect. *Advances In Enzymology And Related Areas of Molecular Biology*, 43:219–410, 1975.
- [49] A. Warshel, J. Florian, M. Strajbl, and J. Villa. Circe effect versus enzyme preorganization: What can be learned from the structure of the most proficient enzyme? *ChemBioChem*, 2(2):109–111, 2001.
- [50] M. Kazemi, F. Himo, and J. Aqvist. Enzyme catalysis by entropy without Circe effect. *Proceedings of The National Academy of Sciences of The United States of America*, 113(9):2406–2411, 2016.

- [51] J. Aqvist, M. Kazemi, G. V. Isaksen, and Bjorn O. Brandsdal. Entropy and Enzyme Catalysis. *Accounts of Chemical Research*, 50(2):199–207, 2017.
- [52] C. Kramer, A. Spinn, and K. R. Liedl. Charge anisotropy: Where atomic multipoles matter most. *Journal of Chemical Theory and Computation*, 10(10):4488–4496, 2014.
- [53] B. Gutte and R. B. Merrifield. Total synthesis of an enzyme with ribonuclease A activity. *Journal of The American Chemical Society*, 91(2):501, 1969.
- [54] B. Gutte and R. B. Merrifield. Synthesis of ribonuclease-A. *Journal of Biological Chemistry*, 246(6):1922, 1971.
- [55] B. Gutte. Synthetic 70-amino acid residue analog of ribonuclease S-protein with enzymic activity. *Journal of Biological Chemistry*, 250(3):889–904, 1975.
- [56] B. Gutte, M. Daumigen, and E. Wittschieber. Design, synthesis and characterization of a 34-residue polypeptide that interacts with nucleic-acids. *Nature*, 281(5733):650–655, 1979.
- [57] R. Moser, R. M. Thomas, and B. Gutte. An artificial crystalline DDT-binding polypeptide. *FEBS Letters*, 157(2):247–251, 1983.
- [58] J. S. Richardson and D. C. Richardson. The *de novo* design of protein structures. *Trends in Biochemical Sciences*, 14(7):304–309, 1989.
- [59] W. F. Degrado, Z. R. Wasserman, and J. D. Lear. Protein design, a minimalist approach. *Science*, 243(4891):622–628, 1989.
- [60] V. Kohler. Protein Design Methods And Applications Second Edition Preface. In Kohler, V., editor, *Protein Design: Methods And Applications, Second Edition*, volume 1216 of *Methods in Molecular Biology*, pages V–VI. Humana Press Inc, 999 Riverview Dr, Ste 208, Totowa, NJ 07512-1165 USA, 2014.
- [61] P. Huang, S. E. Boyken, and D. Baker. The coming of age of *de novo* protein design. *Nature*, 537:320–327, 2016.
- [62] L. Vanderkam. Finding order: Jane Richardson. *Scientific American*, May 28 2008.
- [63] W. F. Degrado, S. P. Ho, P. C. Weber, F. R. Salemme, W. Wilcox, S. Eshita, P. Pryciak, and D. Eisenberg. Design of a Synthetic 4-Helical Bundle Protein. *Biophysical Journal*, 49(2, 2):A573, 1986.
- [64] A. J. Gamble and A. F. A. Peacock. De Novo Design of Peptide Scaffolds as Novel Preorganized Ligands for Metal-Ion Coordination. In Kohler, V., editor, *Protein Design: Methods And Applications, Second Edition*, volume 1216 of *Methods in Molecular Biology*, pages 211–231. Humana Press Inc, 999 Riverview DR, STE 208, Totowa, NJ 07512-1165 USA, 2014.
- [65] G. Winter and A. R. Fersht. Engineering Enzymes. *Trends in Biotechnology*, 2(5):115–118, 1984.
- [66] A. R. Fersht, J. P. Shi, J. Knilljones, D. M. Lowe, A. J. Wilkinson, D. M. Blow, P. Brick, P. Carter, M. M. Y. Waye, and G. Winter. Hydrogen-Bonding and Biological Specificity Analyzed By Protein Engineering. *Nature*, 314(6008):235–238, 1985.
- [67] H. W. Hellinga and F. M. Richards. Construction of New Ligand-Binding Sites In Proteins of Known Structure. 1. Computer-Aided Modeling of Sites With Predefined Geometry. *Journal of Molecular Biology*, 222(3):763–785, 1991.
- [68] E. Quemeneur, M. Moutiez, J. B. Charbonnier, and A. Menez. Engineering cyclophilin into a proline-specific endopeptidase. *Nature*, 391(6664):301–304, 1998.
- [69] D. N. Bolon and S. L. Mayo. Enzyme-like proteins by computational design. *Proceedings of The National Academy of Sciences of The United States of America*, 98(25):14274–14279, 2001.
- [70] E. C. Hayden. Chemistry - Designer debacle. *Nature*, 453(7193):275–278, 2008.

- [71] L. Jiang, E. A. Althoff, F. R. Clemente, L. Doyle, D. Roethlisberger, A. Zanghellini, J. L. Gallaher, J. L. Betker, F. Tanaka, C. F. Barbas, III, D. Hilvert, Kendall N. Houk, Barry L. Stoddard, and D. Baker. De novo computational design of retro-aldol enzymes. *Science*, 319(5868):1387–1391, 2008.
- [72] D. Roethlisberger, O. Khersonsky, A. M. Wollacott, L. Jiand, J. DeChancie, J. Betker, J. L. Gallaher, E. A. Althoff, A. Zanghellini, O. Dym, S. Albeck, K. N. Houk, D., D. S. Tawfik, and D. Baker. Kemp elimination catalysts by computational enzyme design. *Nature*, 453:190–195, 2008.
- [73] J. B. Siegel, A. Zanghellini, H. M. Lovick, G. Kiss, A. R. Lambert, J. L. St. Clair, J. L. Gallaher, D. Hilvert, M. H. Gelb, B. L. Stoddard, K. N. Houk, F. E. Michael, and D. Baker. Computational Design of an Enzyme Catalyst for a Stereoselective Bimolecular Diels-Alder Reaction. *Science*, 329(5989):309–313, 2010.
- [74] M. J. Byrne, N. R. Lees, L. Han, M. W. van der Kamp, A. J. Mulholland, J. E. M. Stach, C. L. Willis, and P. R. Race. The catalytic mechanism of a natural diels-alderase revealed in molecular detail. *Journal of the American Chemical Society*, 138(19):6095–6098, 2016.
- [75] S. D. Khare, Y. Kipnis, P. J. Greisen, R. Takeuchi, Y. Ashani, M. Goldsmith, Y. F. Song, J. L. Gallaher, I. Silman, H. Leader, J. L. Sussman, B. L. Stoddard, D. S. Tawfik, and D. Baker. Computational redesign of a mononuclear zinc metalloenzyme for organophosphate hydrolysis. *Nature Chemical Biology*, 8(3):294–300, 2012.
- [76] C. H. Arnaud. Proteins made to order. *Chemical Engineering and News*, 90:8, 2012.
- [77] C. H. Arnaud. Enzyme by design. *Chemical Engineering and News*, 91(33):26–27, 2013.
- [78] S. Borman. Protein designers roll out a barrel. *Chemical Engineering and News*, 93:6, 2015.
- [79] H. Kries, R. Blomberg, and D. Hilvert. De novo enzymes by computational design. *Current Opinion in Chemical Biology*, 17(2):221–228, 2013.
- [80] F. Richter, A. Leaver-Fay, S. D. Khare, S. Bjelic, and D. Baker. De Novo Enzyme Design Using Rosetta3. *PLoS ONE*, 6(5), 2011.
- [81] G. R. Nosrati and K. N. Houk. SABER: A computational method for identifying active sites for new reactions. *Protein Science*, 21(5):697–706, 2012.
- [82] B. I. Dahiyat and S. L. Mayo. Protein design automation. *Protein Science*, 5(5):895–903, 1996.
- [83] C. A. Denard, H. Ren, and H. Zhao. Improving and repurposing biocatalysts via directed evolution. *Current Opinion in Chemical Biology*, 25:55–64, 2015.
- [84] K. Swiderek, Tunon I., V. Moliner, and J. Bertran. Computational strategies for the design of new enzymatic functions. *Archives of Biochemistry and Biophysics*, 582:68–79, 2015.
- [85] O. Khersonsky, D. Roethlisberger, O. Dym, C. J. Jackson, D. Baker, and D. S. Tawfik. Evolutionary optimization of computationally designed enzymes: Kemp eliminases of the ke07 series. *Journal of Molecular Biology*, 396(4):1025–1042, 2010.
- [86] H. K. Privett, G. Kiss, T. M. Lee, R. Blomberg, R. A. Chica, L. M. Thomas, D. Hilvert, K. N. Houk, and S. L. Mayo. Iterative approach to computational enzyme design. *Proceedings of The National Academy of Sciences of The United States of America*, 109(10):3790–3795, 2012.
- [87] In D. M. York and T. Lee, editors, *Multi-scale Quantum Models for Biocatalysis*, volume 4 of *Challenges And Advances In Computational Chemistry And Physics*, chapter 1, pages 1–154. Springer, 2000. J. Leszczyński, series editor.
- [88] A. N. Alexandrova, Roethlisberger D., D. Baker, and W. L. Jorgensen. Catalytic mechanism and performance of computationally designed enzymes for kemp elimination. *Journal of American Chemical Society*, 130:15907–15915, 2008.

- [89] M. Sparta and A. N. Alexandrova. Computational design and characterisation of artificial enzymes for kemp elimination. *Molecular Simulation*, 37(7):557–571, 2011.
- [90] S. Dapprich, I. Komaromi, K. S. Byun, K. Morokuma, and M. J. Frisch. A new oniom implementation in gaussian98. part i. the calculation of energies, gradients, vibrational frequencies and electric field derivatives. *Journal of Molecular Structure: THEOCHEM*, 461–462:1–21, 1999.
- [91] G. Kiss, D. Röthlisberger, D. Baker, and K. N. Houk. Evaluation and ranking of enzyme designs. *Protein Science*, 19:1760–1773, 2010.
- [92] M. P. Frushicheva, J. Cao, Z. T. Chu, and A. Warshel. Exploring challenges in rational enzyme design by simulating the catalysis in artificial Kemp eliminase. *Proceedings of The National Academy of Sciences of The United States of America*, 107(39):16869–16874, 2010.
- [93] M. P. Frushicheva, J. Cao, and A. Warshel. Challenges and advances in validating enzyme design proposals: The case of kemp eliminase catalysis. *Biochemistry*, 50:3849–3858, 2011.
- [94] M. P. Frushicheva, M. J. L. Mills, P. Schopf, M. K. Singh, R. B. Prasad, and A. Warshel. Computer aided enzyme design and catalytic concepts. *Current Opinion in Chemical Biology*, 21:56–62, 2014.
- [95] A. Labas, E. Szabo, L. Mones, and M. Fuxreiter. Optimization of reorganization energy drives evolution of the designed Kemp eliminase KE07. *Biochimica et Biophysica Acta-Proteins and Proteomics*, 1834(5):908–917, 2013.
- [96] K. Świderek, Tunon I., V. Moliner, and J. Bertran. Protein flexibility and preorganization in the design of enzymes. the kemp elimination catalyzed by hg3.17. *ACS Catalysis*, 5:2587–2595, 2015.
- [97] S. Osuna, G. Jimenez-Oses, Elizabeth L. Noey, and K. N. Houk. Molecular Dynamics Explorations of Active Site Structure in Designed and Evolved Enzymes. *Accounts of Chemical Research*, 48(4):1080–1089, 2015.
- [98] A. Bhowmick, S. C. Sharma, H. Honma, and T. Head-Gordon. The role of side chain entropy and mutual information for improving the de novo design of kemp eliminases ke07 and ke70. *Physical Chemistry Chemical Physics*, 18:19386, 2016.
- [99] D. Baker. An exciting but challenging road ahead for computational enzyme design. *Protein Science*, 19(10):1817–1819, 2010.
- [100] G. Jindal, B. Ramachandran, R. P. Bora, and A. Warshel. Exploring the development of ground-state destabilization and transition-state stabilization in two directed evolution paths of kemp eliminases. *ACS Catalysis*, 7:3301–3305, 2017.
- [101] W. A. Sokalski. The physical nature of catalytic activity due to the molecular environment in terms of intermolecular interaction theory - derivation of simplified models. *Journal of Molecular Catalysis*, 30(3):395–410, 1985.
- [102] K. N. Houk and F. Liu. Holy Grails for Computational Organic Chemistry and Biochemistry. *Accounts of Chemical Research*, 50(3):539–543, 2017.
- [103] F. Himo. Recent trends in quantum chemical modeling of enzymatic reactions. *Journal of the American Chemical Society*, 139(20):6780–6786, 2017.
- [104] J. C. Kromann, A. S. Christensen, Q. Cui, and J. H. Jensen. Towards a barrier height benchmark set for biologically relevant systems. *PEERJ*, 4, 2016.
- [105] A. K. Das and M. Meuwly. Empirical Force Fields for Mechanistic Studies of Chemical Reactions in Proteins. In Voith, G. A., editor, *Computational Approaches For Studying Enzyme Mechanism, Pt A*, volume 577 of *Methods in Enzymology*, pages 31–55. 2016.



- [106] B. Szeftczyk, F. Claeysens, A. J. Mulholland, and W. A. Sokalski. Quantum chemical analysis of reaction paths in chorismate mutase: Conformational effects and electrostatic stabilization. *International Journal of Quantum Chemistry*, 107(12):2274–2285, 2007.
- [107] W. A. Sokalski and R. A. Poirier. Cumulative Atomic Multipole Representation of The Molecular Charge-distribution and Its Basis Set Dependence. *Chemical Physics Letters*, 98(1):86–92, 1983.
- [108] A. Warshel. Energetics of enzyme catalysis. *Proceedings of National Academy of Sciences of the United States*, 75:5250–5254, 1978.
- [109] W. J. Albery and J. R. Knowles. Evolution of Enzyme Function And Development of Catalytic Efficiency. *Biochemistry*, 15(25):5631–5640, 1976.
- [110] A. J. Mulholland and W. G. Richards. Modeling enzyme reaction intermediates and transition states: Citrate synthase. *Journal of Physical Chemistry B*, 102(34):6635–6646, 1998.
- [111] Lucjan Piela. *Idee chemii kwantowej*. PWN, 2011.
- [112] R. Moszyński. Theory of intermolecular forces: an introductory account. In *Molecular Materials with Specific Interactions — Modeling and Design*, volume 4 of *Challenges And Advances in Computational Chemistry and Physics*, chapter 1, pages 1–154. Springer, 2007. J. Leszczyński, series editor.
- [113] B. Jeziorski and W. Kolos. Symmetry Forcing In Perturbation-Theory of Weak Inter-Molecular Interactions. *International Journal of Quantum Chemistry*, 12(1):91–117, 1977.
- [114] C. Møller and M. S. Plesset. Note on an approximation treatment for many-electron systems. *Physical Reviews*, 46:618–622, 1934.
- [115] R. Bukowski and K. Szalewicz. Complete ab initio three-body nonadditive potential in Monte Carlo simulations of vapor-liquid equilibria and pure phases of argon. *Journal of Chemical Physics*, 114(21):9518–9531, 2001.
- [116] M. Jeziorska, B. Jeziorski, and J. Cizek. Direct Calculation of The Hartree-Fock Interaction Energy Via Exchange Perturbation Expansion - The He = He Interaction. *International Journal of Quantum Chemistry*, 32(2):149–164, 1987.
- [117] E. G. Hohenstein and C. D. Sherrill. Wavefunction methods for noncovalent interactions. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 2(2):304–326, 2012.
- [118] K. Kitaura and K. Morokuma. New Energy Decomposition Scheme For Molecular-Interactions Within Hartree-Fock Approximation. *International Journal of Quantum Chemistry*, 10(2):325–340, 1976.
- [119] W. A. Sokalski, S. Roszak, P. C. Hariharan, and J. J. Kaufman. Improved scf interaction energy decomposition scheme corrected for basis set superposition effect. *International Journal of Quantum Chemistry*, 23(3):847–854, 1983.
- [120] W. A. Sokalski, S. Roszak, and K. Pecul. An efficient procedure for decomposition of the scf interaction energy into components with reduced basis set dependence. *Chemical Physics Letters*, 153(2-3):153–159, 1988.
- [121] W. A. Sokalski and S. M. Roszak. Efficient techniques for the decomposition of intermolecular interaction energy at scf level and beyond. *Journal of Molecular Structure-Theochem*, 80:387–400, 1991.
- [122] W. A. Sokalski. Nonempirical modeling of the static and dynamic properties of the optimum environment for chemical reactions. *Journal of Molecular Structure-Theochem*, 138:77–87, 1986.
- [123] A. J. Stone. Distributed Multipole Analysis, Or How To Describe A Molecular Charge-Distribution. *Chemical Physics Letters*, 83(2):233–239, 1981.

- [124] A. J. Misquitta, A. J. Stone, and F. Fazeli. Distributed Multipoles from a Robust Basis-Space Implementation of the Iterated Stockholder Atoms Procedure. *Journal of Chemical Theory and Computation*, 10(12):5405–5418, 2014.
- [125] K. M. Langner. Nieempiryczne metody w analizie i elektrostatycznym modelowaniu oddziaływań w biocząsteczkach. Rozprawa doktorska, Politechnika Wroclawska 2010.
- [126] M.W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. Nguyen, S. J. Su, T. L. Windus, and J. A. Dupuis, M. Montgomery. General atomic and molecular electronic-structure system. *Journal of Computational Chemistry*, 14:11347–1363, 1993.
- [127] W. Beker, K. M. Langner, E. Dyguda-Kazimierowicz, M. Feliks, and W. A. Sokalski. Low-Cost Prediction of Relative Stabilities of Hydrogen Bonded Complexes from Atomic Multipole Moments for Overly Short Intermolecular Distances. *Journal of Computational Chemistry*, 34(21):1797–1799, 2013.
- [128] W. Giedroyc-Piasecka, E. Dyguda-Kazimierowicz, W. Beker, M. Mor, A. Lodola, and W. A. Sokalski. Physical Nature of Fatty Acid Amide Hydrolase Interactions with Its Inhibitors: Testing a Simple Nonempirical Scoring Model. *Journal of Physical Chemistry B*, 118(51):14727–14736, 2014.
- [129] K. Pernal, R. Podeszwa, K. Patkowski, and K. Szalewicz. Dispersionless Density Functional Theory. *Physical Review Letters*, 103(26), 2009.
- [130] K. M. Langner, W. Beker, E. Dyguda-Kazimierowicz, and Sokalski. W. A. Tracking molecular charge distribution along reaction paths with atomic multipole moments. *Structural Chemistry*, 26, 2016.
- [131] A. D. Scouras and V. Daggett. The dynamomics rotamer library: Amino acid side chain conformations and dynamics from comprehensive molecular dynamics simulations in water. *Protein Science*, 20(2):341–352, 2011.
- [132] W. A. Sokalski and A. Sawaryn. Correlated Molecular And Cumulative Atomic Multipole Moments. *Journal of Chemical Physics*, 87(1):526–534, 1987.
- [133] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, G. Cheeseman, J. R. and Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakaajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, T. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, and D. J. Fox. *Gaussian 09, Revision A.02*. Gaussian, Inc., Wallingford C. T., 2016.
- [134] P. Tao and H. B. Schlegel. A Toolkit to Assist ONIOM Calculations. *Journal of Computational Chemistry*, 31:2363–2369, 2010.
- [135] P. Tao, J. F. Fisher, Q. Shi, T. Vreven, S. Mobashery, and H. B. Schlegel. Matrix Metalloproteinase 2 Inhibition: Combined Quantum Mechanics and Molecular Mechanics Studies of the Inhibition Mechanism of (4-Phenoxyphenylsulfonyl)methylthiirane and its Oxirane Analogue. *Biochemistry*, 48(41):9839–9847, 2009.
- [136] F. A. Momany. Determination of Partial Atomic Charges From Abinitio Molecular Electrostatic Potentials - Application To Formamide, Methanol, and Formic-Acid. *Journal of Physical Chemistry*, 82(5):592–601, 1978.
- [137] B. H. Besler, K. M. Merz, and P. A. Kollman. Atomic Charges Derived From Semiempirical Methods. *Journal of Computational Chemistry*, 11(4):431–439, 1990.

- [138] S. Jakobsen and F. Jensen. Searching the Force Field Electrostatic Multipole Parameter Space. *Journal of Chemical Theory and Computation*, pages 1824–1832, 2016.
- [139] L. Chirlian and M. M. Francl. Atomic Charges Derived From Electrostatic Potentials - A Detailed Study. *Journal of Computational Chemistry*, 8(6):894–905, 1987.
- [140] C. M. Breneman and K. B. Wiberg. Determining Atom-Centered Monopoles From Molecular Electrostatic Potentials - The Need For High Sampling Density In Formamide Conformational-Analysis. *Journal of Computational Chemistry*, 11(3):361–373, 1990.
- [141] H. Oberg and T. Brinck. Fragment molecular orbital study of the cAMP-dependent protein kinase catalyzed phosphoryl transfer: a comparison with the differential transition state stabilization method. *Physical Chemistry Chemical Physics*, 18(22):15153–15161, 2016.
- [142] S. D. Fried, S. Bagchi, and S. G. Boxer. Extreme electric fields power catalysis in the active site of ketosteroid isomerase. *Science*, 346, 2014.
- [143] A. Natarajan, F. Yabukarski, V. Lamba, J. P. Schwans, F. Sunden, and D. Herschlag. Comment on "extreme electric fields power catalysis in the active site of ketosteroid isomerase". *Science*, 349(6251), 2015.
- [144] S. D. Fried and S. G. Boxer. Response to comments on "extreme electric fields power catalysis in the active site of ketosteroid isomerase". *Science*, 349(6251), 2015.
- [145] M. W. van der Kamp, R. Chaudret, and A. J. Mulholland. Qm/mm modelling of ketosteroid isomerase reactivity indicates that active site closure is integral to catalysis. *FEBS Journal*, 280(13):3120–3131, 2013.
- [146] W. J. Houck and R. M. Pollack. Temperature effects on the catalytic activity of the d38e mutant of 3-oxo- $\delta$ -5-steroid isomerase: Favorable enthalpies and entropies of activation relative to the nonenzymatic reaction catalyzed by acetate ion. *Journal of the American Chemical Society*, 126(50):16416–16425, 2004.
- [147] D. K. Chakravorty, A. V. Soudackov, and S. Hammes-Schiffer. Hybrid quantum/classical molecular dynamics simulations of the proton transfer reactions catalyzed by ketosteroid isomerase: Analysis of hydrogen bonding, conformational motions, and electrostatics. *Biochemistry*, 48(44):10608–10619, 2009.
- [148] M. Ito and T. Brinck. Novel approach for identifying key residues in enzymatic reactions: Proton abstraction in ketosteroid isomerase. *The Journal of Physical Chemistry B*, 118(46):13050–13058, 2014.
- [149] S. C. L. Kamerlin and A. Warshel. The empirical valence bond model: theory and applications. *Wiley Interdisciplinary Reviews-Computational Molecular Science*, 1(1):30–45, 2011.
- [150] D. Kim, D. S. Jang, G. H. Nam, G. Choi, J. Kim, N. Ha, M. Kim, B. Oh, and K. Y. Choi. Contribution of the hydrogen-bond network involving a tyrosine triad in the active site to the structure and function of a highly proficient ketosteroid isomerase from *Pseudomonas putida* biotype b. *Biochemistry*, 39(16):4581–4589, 2000.
- [151] G. Choi, N. Ha, M. Kim, B. Hong, B. Oh, and K. Y. Choi. Pseudoreversion of the catalytic activity of y14f by the additional substitution(s) of tyrosine with phenylalanine in the hydrogen bond network of  $\delta$ 5-3-ketosteroid isomerase from *Pseudomonas putida* biotype b. *Biochemistry*, 40(23):6828–6835, 2001.
- [152] E. Dyguda-Kazimierowicz, W. A. Sokalski, and J. Leszczynski. Gas-phase mechanisms of degradation of hazardous organophosphorus compounds: Do they follow a common pattern of alkaline hydrolysis reaction as in phosphotriesterase? *Journal of Physical Chemistry B*, 112(32):9982–9991, 2008.
- [153] B. Aguilar, R. Shadrach, and A. V. Onufriev. Reducing the secondary structure bias in the generalized born model via r6 effective radii. *Journal of Chemical Theory And Computation*, 6:3613–3630, 2010.

- [154] A. Mukhopadhyay, B. H. Aguilar, I. S. Tolokh, and A. V. Onufriev. Introducing charge hydration asymmetry into the generalized Born model. *Journal of Chemical Theory and Computation*, 10(4):1788–1794, 2014.
- [155] H. J. C. Berendsen, D. van der Spoel, and R. van Drunen. Gromacs: A message-passing parallel molecular dynamics implementation. *Computer Physics Communications*, 91:43–56, 1995.
- [156] R. Kumari, R. Kumar, Open Source Drug Discovery Consortium, and A. Lynn. g\_mmpbsa - a gromacs tool for high-throughput mm-pbsa calculations. *Journal of Chemical Information and Modeling*, 54:1951–1962, 2014.
- [157] N. A. Baker, D. Sept, S. Joseph, M. J. Holst, and J. A. McCammon. Electrostatics of nanosystems: Application to microtubules and the ribosome. *Proceedings of the National Academy of Sciences of the United States of America*, 98:10037–10041, 2001.
- [158] J. N. Murrell and A. J. C. Varandas. Perturbation Calculations of rare-gas potentials near van der Waals minimum. *Molecular Physics*, 30(1):223–236, 1975.
- [159] R. Landshoff. Quantenmechanische Berechnung des Verlaufes der Gitterenergie des Na-Cl-Gitters in Abhängigkeit vom Gitterabstand. *Zeitschrift für Physik*, 102(3):201–228, Mar 1936.
- [160] A. Froman and P. O. Lowdin. Virial theorem and cohesive energies of solids, particularly ionic crystals. *Journal of Physics And Chemistry of Solids*, 23(Jan-F):75–&, 1962.
- [161] B. Jeziorski, M. Bulski, and L. Piela. 1st-Order perturbation treatment of short-range repulsion in a system of many closed-shell atoms or molecules. *International Journal of Quantum Chemistry*, 10(2):281–297, 1976.
- [162] B. Szefczyk. *Teoretyczne metody badania właściwości katalitycznych enzymów oraz inhibicji na przykładzie mutazy choryzmianowej i amoniakolizy feniloalaninowej*. Rozprawa doktorska, Politechnika Wroclawska, 2005.
- [163] F. Claeysens, K. E. Ranaghan, N. Lawan, S. J. Macrae, F. R. Manby, J. N. Harvey, and A. J. Mulholland. Analysis of chorismate mutase catalysis by QM/MM modelling of enzyme-catalysed and uncatalysed reactions. *Organic & Biomolecular Chemistry*, 9(5):1578–1590, 2011.
- [164] R. W. Zwanzig. High-temperature equation of state by a perturbation method. i. nonpolar gases. *Journal of Chemical Physics*, 22:1420–1426, 1954.
- [165] G. Archontis and M. Karplus. Cumulant expansion of the free energy: Application to free energy derivatives and component analysis. *Journal of Chemical Physics*, 105(24):11246–11260, 1996.
- [166] A. M. Cooper and J. Kaestner. Averaging techniques for reaction barriers in QM/MM simulations. *ChemPhysChem*, 15(15, SI):3264–3269, 2014.
- [167] P. Kast, M. Asif-Ullah, and D. Hilvert. Is chorismate mutase a prototypic entropy trap? Activation parameters for the Bacillus subtilis enzyme. *Tetrahedron Letters*, 37(16):2691–2694, 1996.
- [168] G. Sigalov, A. Fenley, and A. V. Onufriev. Analytical linearized poisson-boltzman n approach: Beyond the generalized born approximation. *Journal of Chemical Physics*, 124, 2006.
- [169] B. Aguilar and A. V. Onufriev. Efficient computation of the total solvation energy of small molecules via the r6 generalized born model. *Journal of Chemical Theory And Computation*, 8:2404–2411, 2012.
- [170] M. J. E. Sternberg, F. R. F. Hayes, A. J. Russell, P. G. Thomas, and A. R. Fersht. Prediction of electrostatic effects of engineering of protein charges. *Nature*, 330:86–88, 1987.
- [171] M. J. L. Mills and P. L. A. Popelier. Intramolecular polarisable multipolar electrostatics from the machine learning method Kriging. *Computational and Theoretical Chemistry*, 975:42–51, 2011.

- [172] M. J. L. Mills and P. L. A Popelier. Polarisable multipolar electrostatics from the machine learning method Kriging: an application to alanine. *Theoretical Chemistry Accounts*, 131(1137), 2012.
- [173] W. A. Sokalski, J. Lai, N. Luo, S. Sun, M. Shibata, and R. Ornstein. Abinitio Study of The Electrostatic multipole nature of torsional potentials in  $\text{CH}_3\text{SSCH}_3$ ,  $\text{CH}_3\text{SSH}$ , and  $\text{HOOH}$ . *International Journal of Quantum Chemistry*, 40:61–71, 1991.
- [174] K. Strasburger and W. A. Sokalski. Intramolecular Electrostatic interactions studied by cumulative atomic multipole moment expansion with improved convergence. *Chemical Physics Letters*, 221(1-2):129–135, 1994.
- [175] P. Kędzierski and W. A. Sokalski. Analysis of the transferability of atomic multipoles for amino acids in modeling macromolecular charge distribution from fragments. *Journal of Computational Chemistry*, 22(10), 2001.
- [176] K. E. Atkinson. *An introduction to numerical analysis*. John Wiley & Sons, 1989.
- [177] C. W. Kern and M. Karplus. Analysis of charge distributions: hydrogen fluoride. *Journal of Chemical Physics*, 40:1374, 1964.
- [178] W. Jedwabny, T. Cierpicki, J. Grembecka, E. Dyguda-Kazimierowicz,. Validation of approximate nonempirical model for menin-mlt interaction inhibitors. *MedChemComm*, doi: 10.1039/C7MD00170C, 2017.



# Spis rysunków

1.1	Reakcja katalizowana przez dekarboksylazę orotydylanową. . . . .	10
1.2	Reakcja eliminacji Kempa oraz jej analog stanu przejściowego. . . . .	11
1.3	Analiza cytowań prac zawierających frazy <i>dynamic effect</i> i <i>enzyme catalysis</i> . . . . .	12
1.4	Schemat przedstawiający ideę elektrostatycznej preorganizacji . . . . .	13
1.5	Schemat reakcji katalizowanej przez KSI. . . . .	14
1.6	Schemat algorytmu projektowania nowych enzymów [79] (tłumaczenie i przedruk za zgodą wydawnictwa Elsevier). . . . .	17
2.1	Schematyczne przedstawienie różnicowej stabilizacji Stanu Przejściowego (DTSS). . . . .	22
2.2	Różnicowa stabilizacja intermediatu (DISS). . . . .	23
2.3	Schematyczne przedstawienie przykładowego pola katalitycznego . . . . .	30
3.1	Porównanie RMSD MEP różnych modeli rozkładu ładunku . . . . .	35
3.2	Model oddziaływania dipol-anion . . . . .	36
3.3	Błąd wyznaczenia wartości energii oddziaływania elektrostatycznego . . . . .	36
3.4	Schemat reakcji katalizowanej przez kinazę białkową A. . . . .	37
3.5	Porównanie biblioteki CAMM oraz metody FMO na przykładzie kinazy białkowej A. . . . .	37
3.6	Porównanie biblioteki CAMM oraz ładunków atomowych na przykładzie kinazy białkowej A. . . . .	38
3.7	Numeracja atomów ciężkich w substracie KSI. . . . .	39
3.8	Oddziaływania elektrostatyczne na ścieżce reakcji w KSI. . . . .	40
3.9	Zależność pomiędzy aktywnością katalityczną a DISS dla różnie wygenerowanych struktur. . . . .	41
3.10	Porównanie wartości DISS otrzymanych różnymi metodami. . . . .	42
3.11	Dekompozycja energii oddziaływań w mutantach KSI . . . . .	42
3.12	Skale czasowe różnych ruchów występujących w białkach . . . . .	44
3.13	Porównanie DTSS(CAMM,Opt) i DTSS(CAMM,Int). . . . .	48
3.14	Porównanie rotamerów DTSS(GB,Int). . . . .	50
3.15	Populacje rotamerów Arg202 wzdłuż trajektorii MD . . . . .	52
3.16	Porównanie DTSS i TSS w rodzinie KE07. . . . .	53
3.17	Porównanie DTSS(CAMM) i DTSS(GB) w rodzinie KE07. . . . .	54
S.1	Schemat przedstawiający sumowanie po wykładniku. . . . .	60
S.2	Schemat reakcji przegrupowania choryzmianu do preferianu . . . . .	63
S.3	Zależność pomiędzy DTSS a barierą aktywacji w mutacji choryzmianowej. . . . .	64
S.4	Wektor dynamicznego pola katalitycznego w PKA oraz orientacja Gly55. . . . .	65
S.5	Wektor dynamicznego pola katalitycznego w KSI oraz orientacje Tyr14 i Asp99. . . . .	65
S.6	Porównanie struktur uzyskanych metodą ONIOM. . . . .	68
S.7	Drganie odpowiadające częstości urojonej. . . . .	69
S.8	Zmiany długości wiązań na ścieżce reakcji. . . . .	69
S.9	Iloczyn ładunku formalnego i pola katalitycznego. . . . .	70
S.10	Porównanie $\Delta E_{mtp}^{(10)}$ wyznaczonego z ładunków Merza-Kollmana [137] z wartością obliczoną za pomocą CAMM. . . . .	70
S.11	Bloki macierzy $\mathbf{G}_{solv}$ . . . . .	71
S.12	Cykl termodynamiczny . . . . .	72
S.13	Błąd wyznaczenia zmiany $pK_a$ pod wpływem mutacji. . . . .	73

S.14 Geometryczna interpretacja dwuwymiarowej interpolacji liniowej . . . . .	75
S.15 Kompleks treoniny z wodą . . . . .	76
S.16 Wykres błędów interpolacji w teście z treoniną. . . . .	78
S.17 Model centrum aktywne meniny. . . . .	79
S.18 Porównanie różnych wersji bazy multipoli. . . . .	81



# Spis tabel

3.1	Porównanie reprezentacji aminokwasów naładowanych. . . . .	39
3.2	Mutacje wprowadzone do teozymu KE07 . . . . .	45
3.3	DTSS– najważniejsze rotamery . . . . .	46
3.4	TSS– najważniejsze rotamery. . . . .	47
3.5	Wkłady DTSS (CAMM) reszt aminokwasowych. . . . .	49
3.6	Wkłady DTSS (GB). . . . .	49
3.7	Wkłady DTSS(MMPBSA) mutowanych aminokwasów P <sub>7</sub> . . . . .	51
3.8	Teoretycznie korzystne mutacje. . . . .	55
S.1	Rzędy kilku najniższych multipoli. . . . .	59
S.2	Średnie i średnie eksponencjalne składowych bariery aktywacji. . . . .	67
S.3	Błąd oszacowania energii swobodnej. . . . .	67
S.4	Statystyki testu z treoniną . . . . .	77
S.5	Błędy interpolowanego momentu dipolowego . . . . .	77
S.6	Błędy interpolowanego momentu dipolowego . . . . .	77
S.7	Statystyki testu z meniną . . . . .	79
S.8	Błędy energii oddziaływania związane z interpolacją multipoli. . . . .	80



# Spis publikacji

- W. Beker, M.W. van der Kamp, A.J. Mulholland, W.A. Sokalski; „Rapid Estimation of Catalytic Efficiency by Cumulative Atomic Multipole Moments: Application to Ketosteroid Isomerase Mutants”; *Journal of Chemical Theory and Computation*, **13** (2), 945–955 (2017) ; **IF=5.310**
- K. M. Langner, W. Beker, E. Dyguda-Kazimierowicz, W. A. Sokalski; „Tracking molecular charge redistribution along reaction paths by atomic multipole expansion”; *Structural Chemistry*, **27** (2), 429-438 (2016); **IF=2.178**
- W. Beker, A. Stachowicz-Kuśnierz, J. Zaklika, A. Ziobro, J. Lipiński, P. Ordon, L. Komorowski; „Atomic polarization justified Fukui indices and the affinity indicators in aromatic heterocycles and nucleobases”; *Computational and Theoretical Chemistry*, **1065**, 42-49 (2015); **IF=1.549**
- W. Giedroyc-Piasecka, E. Dyguda-Kazimierowicz, W. Beker, M. Mor, A. Lodola, W. A. Sokalski; „Physical Nature of Fatty Acid Amide Hydrolase Interactions with Its Inhibitors: Testing a Simple Nonempirical Scoring Model”; *Journal of Physical Chemistry B*, **118** (51), 14727-14736 (2014); **IF=3.177**
- W. Beker, K. M. Langner, E. Dyguda-Kazimierowicz, M. Feliks, W. A. Sokalski; „Low cost prediction of relative stabilities of hydrogen bonded complexes from atomic multipole moments for overly short intermolecular distances”; *Journal of Computational Chemistry*, **34** (21), 1797-1799 (2013); **IF=3.229**
- W. Beker, P. Szarek, L. Komorowski, J. Lipiński, „Reactivity patterns of imidazole, oxazole, and thiazole as reflected by the polarization justified Fukui functions”; *Journal of Physical Chemistry A*, **117** (7), 1596-600 (2013); **IF=2.847**
- K. M. Langner, W. Beker, W. A. Sokalski; „Robust Predictive Power of the Electrostatic Term at Shortened Intermolecular Distances”; *Journal of Physical Chemistry Letters*, **3**, 2785-2789 (2012); **IF=9.353**

## W przygotowaniu do druku

- M. Chojnacka, M. Feliks, W. Beker, W. A. Sokalski; *Predicting substituent effects on activation energy changes by static catalytic fields.*
- W. Beker, W. A. Sokalski; *Bottom-up nonempirical approach reducing search space in enzyme design guided by catalytic fields*



# Spis wystąpień konferencyjnych

## Postery

- W. Beker, W. A. Sokalski; *Combined Molecular Dynamics and Differential Transition State Stabilisation approach to Kemp eliminase KE07 mutants*; Modelling and Design of Molecular Materials 2016; 26 – 30.06.2016; Trzebnica, Poland
- W. Beker, W. A. Sokalski; *Testing applicability of Catalytic Fields in enzyme design: artificial enzymes and their mutants*; "Novel Enzymes" Conference; 11 – 14.10.2016; Groningen (Netherlands)
- W. Beker, E. Dyguda-Kazimierowicz, W.A. Sokalski; *Analysis of catalytic activity of PT3 theozyme mutants through the lenses of intermolecular interaction theory*; TSRC Summer School "Intermolecular Interactions: New Challenges for ab initio Theory"; 10 – 14.08.2015; Telluride (CO), USA
- W. Beker, W.A. Sokalski; *Analysis of spectral shift among redesigned hCRBPII mutants with a simple electrostatic model*; Summer School "Supramolecular Nanosystems"; 27 – 31.07.2015; Würzburg, Germany
- W. Beker, E. Dyguda-Kazimierowicz, W. A. Sokalski; *Comparison of the enzyme-reactant stabilization along reaction path with reaction force and reaction force constant*; WATOC; 5 – 10.10.2014; Santiago, Chile
- W. Beker, J. Zaklika, A. Ziobro, J. Lipiński, L. Komorowski, P. Ordon; *Atomic polarization justified Fukui index as affinity indicator in aromatic heterocycles and nucleobases*; Modelling and Design of Molecular Materials 2014; 29.06 – 3.07.2014; Kudowa-Zdrój, Poland
- W. Beker, W. A. Sokalski; *Catalytic activity of Kemp eliminase theozyme (KE07) mutants interpreted by electrostatic multipole term of differential transition state stabilization*; Modelling and Design of Molecular Materials 2014; 29.06 – 3.07.2014; Kudowa-Zdrój, Poland
- W. Beker, E. Dyguda-Kazimierowicz, W.A. Sokalski; *Modeling the Catalytic Activity of Enzyme Mutants with Differential Transition State Stabilization Approach*; Modelling and Design of Molecular Materials 2014; 29.06 – 3.07.2014; Kudowa-Zdrój, Poland
- W. Beker, E. Dyguda-Kazimierowicz, W.A. Sokalski; *Understanding catalytic activity of PT3 mutants using Cumulative multipole Expansion*; Interface between experimental and theoretical approaches to energy-related enzyme catalysis; 4 – 6.06.2014; London, Great Britain
- W. Beker, P. Szarek, L. Komorowski, J. Lipiński; *Polarization justified Fukui functions for imidazole, oxazole and thiazole*; „Density functional Theory conference”; 28.08–5.09.2012; Athens, Greece

## Wystąpienia ustne

- Wiktor Beker, W. Andrzej Sokalski, *Understanding Kemp Eliminase KE07 mutants with Differential Transition State Theory*, 3<sup>rd</sup> Symposium on Weak Molecular Interactions, Opole 27 – 29.03.2017

- Wiktor Beker, *Projektowanie de novo enzymów: perspektywy*, 3-cia Studencka Konferencja Matematyczno-Biologiczno-Informatyczna „Liczby-Komputery-Życie”, Kraków 19 – 21.04.2013  
(**Nagroda za najlepszy wykład studencki**)