



Uniwersytet im. Adama Mickiewicza  
w Poznaniu  
Wydział Chemii

---

Marta Hajdziona

**Optymalizacja metod analizy Försterowskiego  
rezonansowego przeniesienia energii (FRET)  
pojedynczych cząsteczek**

ROZPRAWA DOKTORSKA

Promotor  
prof. dr hab. Andrzej Molski

Poznań 2011



Dziękuję

mojemu Promotorowi,  
prof. dr. hab. Andrzejowi Molskiemu  
za przekazaną wiedzę i cenne uwagi,  
oraz za pokazanie mi  
jak być dobrym naukowcem i szefem

Rodzicom  
za trud wychowania i przekazane wartości

Markowi  
za pomoc i wsparcie w chwilach zwątpienia  
oraz za to, że dziś używam Linuxa

Kasi  
za słowa otuchy  
i pomoc w załatwieniu wielu spraw  
oraz innym członkom Pracowni Dynamiki,  
za miłą atmosferę w pracy



mojej babci Helenie,  
bez której miłości i wsparcia  
nie zasłabym tak daleko...



# Spis treści

Spis rysunków	5
Spis tabel	9
Wykaz skrótów	11
<b>1 Wstęp</b>	<b>13</b>
<b>2 Część literaturowa</b>	<b>17</b>
2.1 Spektroskopia fluorescencyjna pojedynczych cząsteczek . . . . .	17
2.1.1 Fluorescencja i FRET . . . . .	17
2.1.2 Pomiar FRET na poziomie pojedynczych cząsteczek . . . . .	21
2.1.3 Wady i zalety spFRET . . . . .	22
2.1.4 Alternatywne techniki FRET . . . . .	24
2.1.5 Analiza i estymacja parametrów w eksperymentach spFRET .	24
2.2 Wybrane przykłady eksperymentalne . . . . .	26
2.2.1 Rybozomy . . . . .	26
2.2.2 Fałdowanie rybozymów . . . . .	31
2.2.2.1 Rybozym <i>Tetrahymena thermophila</i> . . . . .	32
2.2.2.2 Rybozym typu hairpin . . . . .	33
2.2.2.3 Inne układy . . . . .	34
2.2.3 Aktywność katalityczna rybozymów i jej sprzężenie z dynamiką konformacyjną . . . . .	35
2.2.3.1 Rybozym <i>Tetrahymena thermophila</i> . . . . .	35
2.2.3.2 Rybozym typu hairpin . . . . .	36

2.2.3.3	Inne układy . . . . .	36
2.2.4	Analiza danych w badaniach pojedynczych cząsteczek RNA . . . . .	37
2.3	Podsumowanie rozdziału i wnioski . . . . .	38
<b>3</b>	<b>Cel i zakres pracy</b>	<b>39</b>
<b>4</b>	<b>Materiały i metody</b>	<b>43</b>
4.1	Symulacja danych . . . . .	43
4.1.1	Teoria . . . . .	43
4.1.2	Przykład symulacji . . . . .	45
4.2	Estymacja parametrów kinetycznych . . . . .	48
4.2.1	Metoda największej wiarygodności (ML) . . . . .	48
4.2.1.1	Teoria . . . . .	48
4.2.1.2	Przykład . . . . .	51
4.2.2	Analiza on-off . . . . .	53
4.2.2.1	Teoria . . . . .	53
4.2.2.2	Przykład . . . . .	54
4.3	Statystyczna analiza danych . . . . .	55
4.4	Rozróżnianie modeli kinetycznych . . . . .	58
4.4.1	Teoria . . . . .	58
4.4.2	Przykład . . . . .	60
<b>5</b>	<b>Wyniki</b>	<b>63</b>
5.1	Model 2-stanowy . . . . .	63
5.1.1	Wpływ zarejestrowanej liczby fotonów . . . . .	63
5.1.1.1	Wyniki dla eksperymentów 2-kolorowych . . . . .	63
5.1.1.2	Wyniki dla eksperymentów 1-kolorowych . . . . .	68
5.1.1.3	Porównanie metody ML i analizy on-off . . . . .	71
5.1.1.4	Podsumowanie rozdziału . . . . .	73
5.1.2	Wpływ szybkości przejść pomiędzy stanami . . . . .	75
5.1.2.1	Zmianie ulega jedna ze stałych szybkości . . . . .	75
5.1.2.2	Zmieniają się obie stałe szybkości . . . . .	81

---

5.1.2.3	Podsumowanie rozdziału . . . . .	84
5.1.3	Wpływ rozdzielenia intensywności . . . . .	85
5.1.3.1	Wyniki dla eksperymentów 2-kolorowych . . . . .	85
5.1.3.2	Wyniki dla eksperymentów 1-kolorowych . . . . .	89
5.1.3.3	Podsumowanie rozdziału . . . . .	90
5.2	Modele 3- i 4- stanowe . . . . .	92
5.2.1	Wpływ liczby stopni swobody na jakość estymowanych parametrów . . . . .	93
5.2.2	Wpływ długości trajektorii na jakość estymowanych parametrów	95
5.2.3	Podsumowanie rozdziału . . . . .	97
5.3	Parametry startowe użyte do dopasowania . . . . .	98
5.3.1	Modele 4-stanowe . . . . .	98
5.3.2	Model 3-stanowy . . . . .	103
5.3.3	Model 2-stanowy . . . . .	105
5.3.4	Podsumowanie rozdziału . . . . .	106
5.4	Rozróżnianie modeli kinetycznych . . . . .	108
5.4.1	Przykładowy eksperyment rozróżniania modeli . . . . .	108
5.4.2	Wpływ długości trajektorii na rozróżnianie modeli kinetycznych	111
5.4.3	Wpływ parametrów startowych . . . . .	114
5.4.4	Podsumowanie rozdziału . . . . .	114
<b>6</b>	<b>Podsumowanie i wnioski</b>	<b>115</b>
	<b>Bibliografia</b>	<b>120</b>



# Spis rysunków

2.1	Diagram Jabłońskiego . . . . .	19
2.2	Zależność wydajności FRET od odległości pomiędzy donorem i akceptorem . . . . .	19
2.3	Zasada techniki spFRET . . . . .	20
2.4	Obszar wzbudzenia dla zespołu cząsteczek i pojedynczej cząsteczki oraz schemat unieruchomiania cząsteczek . . . . .	21
2.5	Przykładowe trajektorie fluorescencji dla FRET dwustanowego . . . . .	24
2.6	Przykłady mechanizmów aktywności katalitycznej rybozymów . . . . .	28
2.7	Budowa rybozymów typu hairpin . . . . .	29
2.8	Modyfikacje rybozemu typu hairpin . . . . .	30
2.9	Budowa rybozemu <i>Tetrahymena thermophila</i> . . . . .	31
2.10	Struktura formy sfałdowanej Diels-Alderazy . . . . .	32
2.11	Mechanizm katalizy enzymu <i>Tetrahymena thermophila</i> . . . . .	35
2.12	Mechanizm katalizy formy minimalnej rybozemu typu hairpin . . . . .	36
4.1	Macierz stałych kinetycznych $\mathbf{Q}$ . . . . .	44
4.2	Macierze intensywności $\mathbf{\Lambda}^D$ oraz $\mathbf{\Lambda}^A$ . . . . .	45
4.3	Schematy modeli 3- i 4-stanowych . . . . .	49
4.4	Analiza on-off: obliczanie trajektorii fluorescencji . . . . .	54
4.5	Schemat przedstawiający sposób prowadzenia eksperymentów komputerowych . . . . .	55
4.6	Przykład analizy statystycznej otrzymanych wyników . . . . .	56
4.7	Schemat postępowania w eksperymentach rozróżniania modeli . . . . .	59

4.8	Przykład wartości funkcji wiarygodności, AIC i BIC w dopasowaniu różnych modeli do danych z symulacji modelu 3L. . . . .	61
4.9	Przykładowe histogramy wartości BIC . . . . .	61
4.10	Przykładowy eksperyment rozróżniania modeli, gdy symulowano model 3L . . . . .	62
4.11	Wartości $\Delta$ BIC . . . . .	62
5.1	Odchylenia standardowe $\kappa_{12}$ dla różnych intensywności i długości trajektorii . . . . .	68
5.2	Odchylenia standardowe $\kappa_{12}$ dla różnych intensywności i długości trajektorii dla eksperymentu 1-kolorowego . . . . .	71
5.3	Analiza on-off: trajektorie o różnej szerokości okna czasowego . . . . .	72
5.4	Wpływ szybkości $k_{21}$ na wyniki dopasowań w eksperymencie 2-kolorowym . . . . .	77
5.5	Wykres korelacji estymat stałych szybkości, gdy zmieniano wartość $k_{21}$ . . . . .	78
5.6	Wpływ szybkości $k_{21}$ na wyniki dopasowań w eksperymencie 1-kolorowym . . . . .	80
5.7	Wpływ szybkości $k_{12} = k_{21} = k$ na wyniki dopasowań w eksperymencie 2-kolorowym . . . . .	82
5.8	Wykres korelacji estymat stałych szybkości, gdy zmieniano obie wartości stałych szybkości $k$ . . . . .	83
5.9	Rozkłady zliczeń fotonów dla różnych par intensywności . . . . .	86
5.10	Wyniki dopasowań dla różnych wartości $I_1/I_h$ . . . . .	87
5.11	Wykres pudełkowy dla różnych wartości $I_1/I_h$ . . . . .	88
5.12	Wpływ rozdzielania intensywności w eksperymencie 1-kolorowym . . . . .	90
5.13	Parametry startowe użyte do dopasowania modelu 4L . . . . .	99
5.14	Parametry startowe użyte do dopasowania modelu 4C . . . . .	99
5.15	Wyniki dopasowań dla różnych punktów startowych dla modelu 4L . . . . .	100
5.16	Wyniki dopasowań dla różnych punktów startowych dla modelu 4C . . . . .	101
5.17	Wyniki dopasowań modelu 3L dla różnych parametrów startowych . . . . .	103
5.18	Nieskalowane parametry startowe użyte do dopasowania modelu 2-stanowego (5.9) . . . . .	105
5.19	Wyniki dopasowań dla różnych parametrów startowych . . . . .	106

---

5.20 Modele wykorzystane w eksperymentach rozróżniania modeli kinetycznych . . . . .	109
5.21 Przykłady rozróżniania modeli 2, 3L, 4L, 5L, 3C, 4C . . . . .	110
5.22 Modele wybierane przez AIC i BIC dla różnych długości trajektorii .	113



# Spis tabel

4.1	Wartości funkcji wiarygodności oraz AIC i BIC: symulowano model 3L, dopasowywano modele 2-, 3-, 4-, 5-stanowe . . . . .	60
5.1	Porównanie liczby cykli dla eksperymentów dwukolorowych . . . . .	64
5.2	Wartości średnie i odchylenia standardowe estymat stałych przejść dla różnych intensywności i długości trajektorii dla modelu 2-stanowego .	65
5.3	Wartości średnie i odchylenia standardowe estymat intensywności dla różnych intensywności i długości trajektorii dla modelu 2-stanowego .	66
5.4	Wartości średnie i odchylenia standardowe oszacowań estymat stałych szybkości dla różnych intensywności i długości trajektorii dla modelu 2-stanowego, dla eksperymentu 1-kolorowego . . . . .	69
5.5	Wartości średnie i odchylenia standardowe estymat intensywności dla różnej liczby fotonów na stan i długości trajektorii dla modelu 2-stanowego, dla eksperymentu 1-kolorowego . . . . .	69
5.6	Analiza on-off modelu 2-stanowego dla różnych wielkości okna czasowego w trajektorii . . . . .	72
5.7	Analiza on-off: wyniki dla różnych długości trajektorii . . . . .	73
5.8	Wartości średnie i odchylenia standardowe estymat intensywności i stałych kinetycznych dla różnych wartości stałych przejść dla $n = 5 \times 10^4$ fotonów, dla eksperymentu 2- kolorowego . . . . .	76
5.9	Wartości średnie i odchylenia standardowe estymat intensywności i stałych kinetycznych dla różnych wartości stałych kinetycznych dla traj $n = 2,5 \times 10^4$ fotonów, 1-kolorowej . . . . .	80
5.10	Wartości średnie i odchylenia standardowe estymat intensywności i stałych kinetycznych dla dopasowań, gdzie zmieniano obie stałe kinetyczne . . . . .	81

---

5.11 Wyniki dopasowań stałych szybkości dla różnego rozdzielania intensywności w eksperymencie 2-kolorowym . . . . .	86
5.12 Wyniki dopasowań dla różnych wartości $I_1/I_h$ w eksperymencie 1-kolorowym . . . . .	89
5.13 Wartości średnie i odchylenia standardowe dopasowanych parametrów dla modeli 3- i 4- stanowych, dla $I_i = 120$ . . . . .	94
5.14 Dopasowanie modeli 3- i 4- stanowych, kiedy $I_i = 80$ , dla różnych długości trajektorii $n$ . . . . .	96
5.15 Wyniki dopasowań modelu 4L dla jednego zestawu danych i 50 różnych zestawów parametrów startowych . . . . .	100
5.16 Wyniki dopasowań modelu 4C dla jednego zestawu danych i 50 różnych zestawów parametrów startowych . . . . .	101
5.17 Wyniki dopasowań różnych parametrów startowych modelu 3L, w eksperymencie 2-kolorowym . . . . .	104
5.18 Statystyka rozróżniania pomiędzy modelami 2-, 3- i 4- stanowymi . . .	111

# Wykaz skrótów

- FRET - (ang. *Förster/Fluorescence Resonance Energy Transfer*), Försterowskie/fluorescencyjne rezonansowe przeniesienie energii
- spFRET - (ang. *Single-Pair Förster/Fluorescence Resonance Energy Transfer*), FRET na poziomie pojedynczych cząsteczek
- SMS - (ang. *Single-Molecule Spectroscopy*), spektroskopia pojedynczych cząsteczek
- MMPP - (ang. *Markov Modulated Poisson Processes*), Markowsko modulowane procesy Poissona
- MMMPP - (ang. *Marked Markov Modulated Poisson Processes*), znakowane Markowsko modulowane procesy Poissona
- ML - (ang. *Maximum Likelihood*), funkcja największej wiarygodności
- AIC - (ang. *Akaike information criterion*), Kryterium Akaikego
- BIC - (ang. *Bayesian information criterion*), Kryterium Bayesa
- APD - (ang. *avalanche photodiodes*), fotodioda lawinowa
- CCD - (ang. *Charge Coupled Device*), matryca CCD
- TIR - (ang. *Total Internal Reflection*), całkowite wewnętrzne odbicie
- ALEX - (ang. *Alternating-Laser Excitation*), wzbudzenie przemienne
- DAza - (ang. *Diels-Alderase ribozyme*), rybozym Diels-Alderaza
- HMM - (ang. *Hidden Markov Models*), ukryte modele Markowa
- 3C - model 3-stanowy, cykliczny
- 3L - model 3-stanowy, liniowy
- 4C - model 4-stanowy, cykliczny
- 4L - model 4-stanowy, liniowy



# Rozdział 1

## Wstęp

Spektroskopia fluorescencyjna znajduje zastosowanie w badaniu dynamiki i mechanizmu działania makrocząsteczek biologicznych [1, 2]. Wykorzystując technikę FRET (ang. *Förster Resonance Energy Transfer*) pojedynczych cząsteczek możliwe jest badanie zachowania poszczególnych cząstek takich jak DNA, RNA i białka [3].

Zjawisko FRET obserwuje się, gdy w odpowiednio małej odległości znajdują się obok siebie dwa barwniki fluorescencyjne, między którymi istnieje możliwość przeniesienia energii z jednego z nich (donora) na drugi (akceptor) [4, 5]. Zasięg oddziaływania Försterowskiego przeniesienia energii wynosi ok. 100 Å. Wydajność takiego przeniesienia zależy od zakresu nakładania się widm absorpcji akceptora i emisji donora oraz jest odwrotnie proporcjonalna do szóstej potęgi odległości między barwnikami.

Ze względu na małe odległości i wysoką czułość na zmianę położenia między fluoroforami technika ta może być wykorzystywana do mierzenia zmian odległości wewnątrz makrocząsteczek biologicznych [6, 7]. Znaczny rozwój technologii pozwolił na tak duże zwiększenie rozdzielczości, że możliwe jest obserwowanie fotonów pochodzących od pojedynczej cząsteczki barwnika [8]. Analizując zarejestrowane fotony fluorescencji uzyskać można informacje na temat dynamiki badanej cząsteczki [9]. Dzięki temu możliwe było odkrycie zjawisk niewidocznych w zespole cząstek, ze względu na efekt uśrednienia. Wykorzystując metody badania pojedynczych cząsteczek odkryto takie zjawiska jak dynamiczny i statyczny nieporządek w działaniu enzymów, a także efekt pamięci, między innymi w rybozymach [10, 11].

Eksperymentalisci często korzystają z analizy histogramów czasów on-off w celu estymacji parametrów kinetycznych procesu [8, 12]. Metoda on-off jest stosunkowo

prosta i pozwala badać dynamikę różnych procesów. Wariantem tej metody jest dopasowanie modelu do kumulatywnych histogramów czasów on i off [13, 14]. Rozdzielczość analizy on-off determinowana jest jednak wielkością okna czasowego, wybranego do obliczenia trajektorii FRET, a także doбором progów rozdzielającego stany o różnej wydajności FRET. W wyniku tego trudne jest oszacowanie szybkości procesów, których czas jest porównywalny z wielkością okna czasowego przyjętego w celu obliczenia trajektorii fluorescencji [15]. Trajektorie muszą także być długie, ponieważ konieczne jest zmierzenie odpowiedniej liczby czasów przebywania w konkretnych stanach. W każdym kanale trajektorii fluorescencji musi zostać zarejestrowana odpowiednia liczba fotonów w celu zminimalizowania wpływu tła i umożliwienia rozdzielenia stanów. Jeżeli intensywności w dwóch stanach nie są całkowicie rozdzielone, pomimo optymalnego doboru wartości progowej, nieuniknione jest popełnianie błędów w kategoryzowaniu poszczególnych kanałów jako należących do któregoś ze stanów. Pogarsza to znacznie jakość otrzymywanych wyników.

Niestety opisane powyżej warunki nie zawsze są możliwe do osiągnięcia ze względu na właściwości materiału biologicznego, jak i na własności fotofizyczne barwników fluorescencyjnych. W eksperymentach na poziomie pojedynczych cząsteczek często badany obiekt unieruchamia się, by uniemożliwić cząstce oddyfundowanie z objętości konfokalnej [8, 16, 17, 18, 19]. Konsekwencją tego jest fakt, że barwniki fluorescencyjne są stale naświetlane, co przyspiesza fotowysbielenie fluoroforu i znacznie skraca czas obserwacji. Konieczne jest więc poszukiwanie nowych metod analizy danych, które pozwalałyby odtwarzać parametry kinetyczne na podstawie krótkich trajektorii, charakteryzując się przy tym wysoką rozdzielczością. Ze względu na duże skomplikowanie procesów zachodzących na poziomie pojedynczych makrocząsteczek, do poszukiwania nowych metod analizy i ich optymalizacji wykorzystuje się symulacje komputerowe [20].

Metody symulacyjne pozwalają przewidzieć wynik eksperymentu, lub stwierdzić, czy dany układ jest odpowiedni do wykonywania pomiarów, znacznie redukując czas i koszty przeprowadzanych doświadczeń. Ważną częścią badań jest też opracowanie statystyczne otrzymanych wyników, pozwalające zrozumieć istotę procesów i oszacować, jaki błąd niosą ze sobą warunki pomiarowe [21]. W celu uzyskania danych do analizy przeprowadza się symulację, w wyniku której otrzymać można np. trajektorię czasów międzydetekcyjnych, a następnie analizować ją przy użyciu dowolnych metod. Zaletą takiego działania jest fakt, że przystępując do

---

analizy posiada się dokładne informacje na temat modelu kinetycznego, jaki został użyty, znana jest także wartość parametrów kinetycznych, którymi ten model się charakteryzował. W celu określenia właściwości statystycznych estymowanych parametrów z użyciem danej metody analizy stosuje się metodę Monte Carlo, polegającą na wielokrotnym powtórzeniu eksperymentu komputerowego, a następnie analizie statystycznej otrzymanych wyników.

W niniejszej pracy skupiono się na zbadaniu właściwości metody, opartej na analizie największej wiarygodności trajektorii czasów międzydetekcyjnych pojedynczych, unieruchomionych cząsteczek [21, 22, 23]. Trajektorie czasów międzydetekcyjnych symulowano wykorzystując MMMPP (ang. *Marked Markov Modulated Poisson Processes*) znakowane Markowsko modulowane procesy Poissona [24, 25]. Dobierając parametry procesu opierano się na eksperymentach FRET pojedynczych cząsteczek w badaniu dynamiki rybozymów [26]. Następnie przeprowadzano estymację parametrów kinetycznych dla zadanego modelu. Badano jaki wpływ na jakość dopasowania mają takie czynniki jak długość trajektorii, intensywność fluorescencji, wartości stałych kinetycznych, rozdzielanie intensywności świecenia barwników, a także wartości parametrów startowych, użytych do dopasowania. W wielu przypadkach modele kinetyczne działania cząsteczek biologicznych są bardziej skomplikowane, niż 2-stanowe [10, 27, 28]. Sprawdzano stosowalność metody z wykorzystaniem dopasowania funkcji największej wiarygodności (ang. *Maximum Likelihood*) ML w odzyskiwaniu parametrów kinetycznych dla modeli składających się z większej liczby stanów (3, 4 i więcej) oraz o różnej topologii (modele liniowe i cykliczne). Oprócz estymacji parametrów badano możliwość rozróżniania modeli stosując kryteria Akaike i Bayesa [29, 30].

Niniejsza praca składa się z następujących części: W rozdziale 2 przedstawiono podstawy teoretyczne FRET, omówiono techniki pomiarowe w spFRET oraz przykłady eksperymentalne wykorzystania FRET pojedynczych cząsteczek w badaniu dynamiki rybozymów. W rozdziale 3 zaprezentowano cel i zakres pracy. W rozdziale 4 przedstawiono materiały i metody badawcze, z których korzystano w niniejszej pracy. Rozdział 5 poświęcono wynikom jakie uzyskano z symulacji i dopasowań metodą największej wiarygodności i porównano je z wynikami uzyskanymi dla analizy on-off. Brano pod uwagę różne czynniki, takie jak intensywność, wartość stałych kinetycznych, rozdzielanie intensywności, które mogą mieć wpływ na jakość estymowanych parametrów. Dopasowywano też modele

3- i 4- stanowe. Przedstawiono także w jaki sposób dokonać można rozróżniania modeli kinetycznych wykorzystując AIC i BIC. W rozdziale 6 dokonano całościowego podsumowania pracy oraz zaproponowano możliwe kierunki dalszych badań.

# Rozdział 2

## Część literaturowa

### 2.1 Spektroskopia fluorescencyjna pojedynczych cząsteczek

---

W niniejszym rozdziale przedstawione zostaną podstawy teoretyczne spektroskopii fluorescencyjnej, a w szczególności FRET (ang. *Förster Resonance Energy Transfer*), Försterowskiego rezonansowego przeniesienia energii. Zaprezentowany będzie sposób, w jaki wykonuje się pomiary fluorescencji na poziomie pojedynczych cząsteczek i aparatura pomiarowa, niezbędna do przeprowadzenia eksperymentu. Pokazane będzie jakie nowe możliwości niesie ze sobą obserwowanie makrocząsteczek biologicznych na poziomie pojedynczych molekuł oraz problemy z jakimi należy się zmierzyć wykonując eksperymenty spFRET. Krótko scharakteryzowane także zostaną metody analizy, które wykorzystuje się do estymacji parametrów kinetycznych w spektroskopii pojedynczych cząsteczek.

---

#### 2.1.1 Fluorescencja i FRET

Istnieją cząsteczki (chromofory), które w wyniku naświetlania falą o odpowiedniej długości, potrafią zaabsorbować tą energię i ulec wzbudzeniu elektronowemu do stanu singletowego. Stan wzbudzony jest stanem o wysokiej energii i układ dąży

do oddania nadmiaru energii i powrotowi do stanu podstawowego. Istnieje kilka możliwości takiej dezaktywacji. Jedną z nich jest wypromieniowanie energii w postaci kwantu promieniowania świetlnego, zjawisko to nazywamy fluorescencją. Ze względu na efekt przesunięcia batochromowego energia emitowanych fotonów jest zawsze mniejsza od energii światła wzbudzającego. Widmo emisji fluorescencji jest przesunięte względem widma absorpcji w stronę fal dłuższych. W wyniku tego wzbudzając cząstkę na przykład falą o długości 400 nm (światło niebieskie), obserwuje się fotony fluorescencji o energii na przykład w zakresie fal czerwonych (ok. 500 nm).

Jeżeli w odpowiednio bliskiej odległości od siebie znajdują się barwniki fluorescencyjne takie, że widmo emisji jednego z nich nakłada się z widmem absorpcji drugiego, istnieje możliwość bezpromienistego przeniesienia energii wzbudzenia z jednego z nich (donora) na drugi (akceptor). Zjawisko takie nazywa się FRET (ang. *Förster Resonance Energy Transfer*), Försterowskie rezonansowe przeniesienie energii [4, 5]. W wyniku wzbudzenia światłem o długości fali w zakresie maksimum absorpcji donora obserwuje się emisję fluorescencji akceptora (rys. 2.1).

Barwniki wykorzystane w eksperymencie FRET muszą charakteryzować się dużą jasnością (współczynnik ekstynkcji  $\varepsilon > 50000 \text{ M}^{-1}\text{cm}^{-1}$ ), wydajność kwantowa fluorescencji powinna być większa od 0,1. Powinny być także fotostabilne, małe i rozpuszczalne w wodzie oraz posiadać grupy, umożliwiające przyłączenie ich do cząsteczek biologicznych. Idealna zaś para donor-akceptor powinna charakteryzować się tym, że oba barwniki mają podobną wydajność fluorescencji i detekcji, widma emisji i absorpcji powinny jak najbardziej się nakładać.

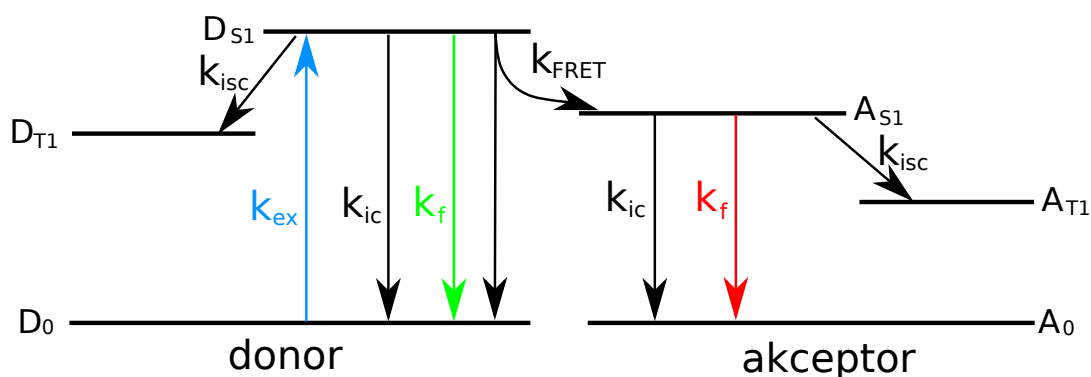
Wydajność przeniesienia energii zmniejsza się gwałtownie ze wzrostem odległości między fluoroforami [31] i można zapisać ją wzorem :

$$E = \frac{1}{1 + (R/R_0)^6} \quad (2.1)$$

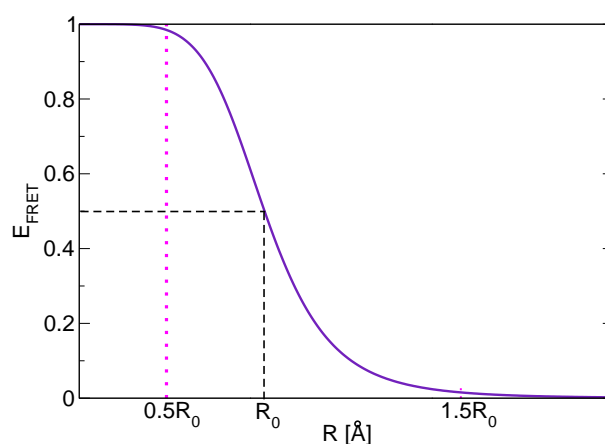
gdzie  $E$  – wydajność FRET,  $R$  – odległość między donorem i akceptorem, a  $R_0$  – odległość, dla której wydajność FRET wynosi 0,5.

Najwyższą czułość na zmianę odległości metoda FRET ma w przedziale od  $0,5R_0$  do  $1,5R_0$  (rys. 2.2). Poniżej  $0,5R_0$  i powyżej  $1,5R_0$  krzywa osiąga plateau i nie obserwuje się znacznej zmiany wydajności FRET przy zmianie odległości pomiędzy barwnikami, a tym samym FRET jest mniej czuły na zmiany odległości.

Wydajność zależy także od ułożenia dipoli donora i akceptora. Zasięg przeniesienia energii FRET to ok.  $100\text{Å}$ . Ze względu na wysoką czułość na



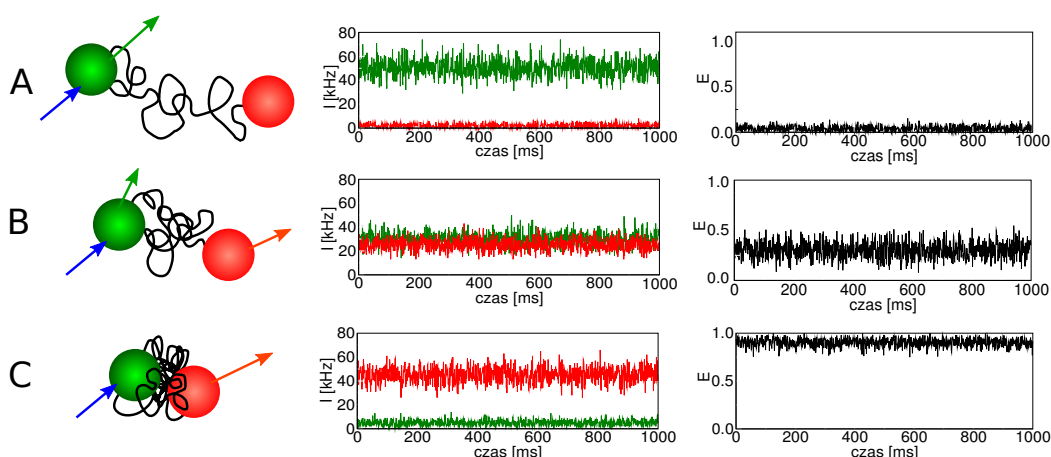
**Rysunek 2.1:** Diagram Jabłońskiego dla dwóch barwników, pomiędzy którymi następuje przeniesienie energii.  $D_0$ ,  $A_0$ : stan podstawowy,  $D_{T1}$ ,  $A_{T1}$ : stan trypletowy,  $D_{S1}$ ,  $A_{S1}$ : singletowy stan wzbudzony odpowiednio donora i akceptora. Wzbudzenie donora do stanu singletowego  $k_{ex}$ : niebieska strzałka.  $k_f$ : fluorescencja donora (zielona strzałka) i akceptora (czerwona strzałka),  $k_{ic}$ : bezpromienista dezaktywacja stanu singletowego,  $k_{isc}$ : przejście do stanu trypletowego.  $k_{FRET}$ : przeniesienie energii FRET.



**Rysunek 2.2:** Zależność wydajności FRET od odległości pomiędzy donorem i akceptorem. Różowe, kropkowane linie oznaczają wartości  $0,5R_0$  i  $1,5R_0$ , w tym zakresie technika FRET jest bardzo czuła na zmiany odległości pomiędzy barwnikami. Czarna kropkowana linia:  $R_0$ , czyli odległość, dla której wydajność FRET wynosi  $E_{FRET} = 0,5$ .

zmianę położenia między fluoroforami zjawisko FRET może być wykorzystywane do mierzenia zmian odległości wewnątrz makrocząsteczek biologicznych [1, 10].

Wraz z postępem technicznym stało się możliwe obserwowanie zjawiska FRET dla pojedynczych par donor-akceptor (spFRET - ang. *single-pair FRET*). W idealnych warunkach światło laserowe wzbudza jedynie cząsteczkę donora, a fluorescencja akceptora pochodzi jedynie od bezpromienistego przeniesienia energii



**Rysunek 2.3:** Zasada techniki spFRET. Po lewej: cząsteczka donora (zielona) i akceptora (czerwona) przyłączone do biopolimeru. Na środku: trajektorie fluorescencji donora (linia zielona) i akceptora (linia czerwona). Po prawej: trajektorie FRET. A) Duża odległość między fluoroforami powoduje niską wydajność przeniesienia energii i obserwuje się tylko fluorescencję donora. B) Odległość między fluoroforami jest porównywalna z  $R_0$  (wzór 2.1), tak że oba barwniki (donor i akceptor) emitują fotony fluorescencji. C) Mała odległość między fluoroforami powoduje wysoką wydajność przeniesienia energii i obserwuje się głównie fluorescencję akceptora.

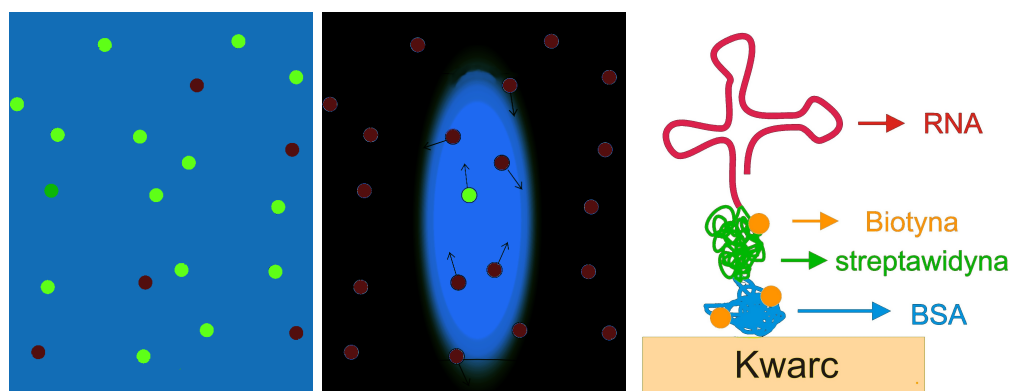
z donora na akceptor. Wydajność przeniesienia energii definiuje się wtedy jako:

$$E = \frac{I_A}{I_A + I_D} \quad (2.2)$$

gdzie:  $E$  – wydajność przeniesienia energii,  $I_A$  – intensywność fluorescencji akceptora,  $I_D$  – intensywność fluorescencji donora. Zasadę działania techniki spFRET przedstawiono na rys. 2.3.

Technika FRET może być stosowana do badania zespołu swobodnie dyfundujących cząsteczek [32] lub też do pary donor-akceptor przyłączonej do unieruchomionej cząsteczki 2.4. Spektroskopia pojedynczych cząsteczek wymaga dokładnego oczyszczania próbek i konieczności sprawdzenia, czy oba barwniki przyłączyły się. Konieczne jest także zastosowanie sprzętu pomiarowego, gdzie obszar wzbudzenia będzie odpowiednio mały oraz detektora zdolnego rejestrować fotony fluorescencji pochodzące od pojedynczej cząsteczki barwnika. Należy także przeprowadzić analizę statystyczną dla wielu pojedynczych cząsteczek. Niemniej jednak obserwowanie pojedynczych cząsteczek niesie ze sobą korzyści takie jak: możliwość zaobserwowania zmian zachodzących wewnątrz cząsteczki, możliwość

śledzenia zachowania pojedynczej molekuly, a nie tylko całego zespołu, możliwość zaobserwowania efektów, uśredniających się w zespole cząsteczek.



**Rysunek 2.4:** Obszar wzbudzenia dla zespołu cząsteczek (po lewej) i dla pojedynczej cząsteczki (pośrodku). Na niebiesko oznaczono objętość konfokalną, na zielono barwniki, które w danym momencie świecą, na brązowo - barwniki, które obecnie nie emitują fotonów. W eksperymencie na zespole cząsteczek objętość konfokalna jest na tyle duża, by wzbudzać i obserwować fluorescencję wielu barwników jednocześnie. W eksperymencie na poziomie pojedynczej cząsteczki objętość konfokalna jest tak mała, by z dużym prawdopodobieństwem można założyć, że w danym momencie obserwuje się fotony pochodzące od pojedynczego barwnika. Po prawej: przykład unieruchomienia pojedynczej cząsteczki RNA.

### 2.1.2 Pomiar FRET na poziomie pojedynczych cząsteczek

Ze względu na ograniczoną liczbę emitowanych przez cząsteczkę barwnika fotonów oraz ich niewielką ilość w stosunku do pomiarów na zespole cząstek, pomiary z rozdzielczością pojedynczej cząsteczki mają kilka wymogów. Należy do nich na przykład konieczność wydajnej detekcji z jak najniższym szumem. Niezbędne jest wykorzystanie wysoko wydajnych przyrządów optycznych (obiektywów, filtrów, soczewek). W celu zwiększenia wydajności detekcji fotonów stosuje się detektory charakteryzujące się wysoką wydajnością [16, 33, 34, 35]. Są to zazwyczaj detektory punktowe o wysokiej rozdzielczości (rzędu kilkudziesięciu nanosekund), np. fotodiody lawinowe APD (ang. *avalanche photodiodes*) lub matryca CCD (ang. *Charge Coupled Device*) detektor powierzchniowy, o nieco niższej rozdzielczości czasowej (maksymalnie 1ms) [36]. Matryca CCD jest układem wielu elementów światłoczułych. Padający kwant promieniowania powoduje efekt fotoelektryczny wewnętrzny, w wyniku czego powstaje ładunek elektryczny, który jest proporcjonalny do ilości światła padającego na matrycę.

Im większy jest obszar podlegający detekcji tym większy jest poziom szumu i dlatego też stosuje się techniki pozwalające ograniczyć obserwowaną powierzchnię. Główne dwie, to mikroskopia konfokalna i wykorzystanie całkowitego wewnętrznego odbicia TIR (ang. *Total Internal Reflection*).

W mikroskopii konfokalnej wykorzystuje się wzbudzenie ciągłe, lub pulsacyjne. Emitowane fotony fluorescencji dwóch barwników rozdzielane są przy użyciu lustro dichroicznego. W celu ograniczenia obserwowanej objętości wzdłuż osi optycznej mikroskopu stosuje się przesłonę z małą szczeliną (ang. *pinhole*), przez którą przechodzą fotony, które są później rejestrowane w detektorze. Wszystkie pozostałe błyski nie podlegają detekcji. Zastosowanie przesłony zmniejsza obserwowany obszar do objętości femtolitrów [37]. Stosuje się tu zazwyczaj detektory punktowe APD.

Natomiast w technice TIR wykorzystywany jest fakt, że światło przy przejściu z ośrodka o wyższym do ośrodka o niższym współczynniku załamania pod kątem większym od kąta granicznego ulega całkowitemu odbiciu wewnętrznemu. W wyniku tego, cząsteczki znajdujące się na styku faz (np. bezpośrednio na powierzchni płytki) mogą ulec wzbudzeniu pod wpływem fali zanikającej. Jako detektory wykorzystywane są zazwyczaj matryce CCD.

### 2.1.3 Wady i zalety spFRET

W celu wykonania pomiarów FRET konieczne jest przyłączenie do badanej cząsteczki obu barwników. Jeśli przyłączy się tylko akceptor, fluorescencja nie będzie obserwowana, natomiast w przypadku przyłączenia tylko donora, wydajność FRET wyniesie 0. Poprawne przyłączanie barwników może być kłopotliwe w przypadku dużych cząsteczek rybozymów [38, 39] i wielu białek [40]. FRET nie jest czuły na odległości spoza przedziału 2-8 nm, kiedy odległość  $R_0 = 5$  nm, natomiast można obserwować zmiany rzędu 0, 3nm, kiedy odległość pomiędzy barwnikami mieści się w przedziale  $0,6R_0$  do  $1,5R_0$ .

Ze względu na istnienie ruchów Browna, cząsteczki zawieszane w roztworze mogą opuścić objętość wzbudzania. Aby tego uniknąć, cząsteczki unieruchamia się [41, 42, 43]. Jeden ze sposobów polega na pokryciu powierzchni kwarcowej glikolem polietylenowym (PEG), a następnie nałożeniu białka będącego pochodną awidyny, np. neutrawidynę lub streprawidynę (rys. 2.4, po prawej). Białka te wykazują duże powinowactwo do biotyny, do której przyłącza się z kolei badaną cząsteczkę kwasu nukleinowego (DNA, lub RNA). Na przykład, dla kompleksu

biotyna i streptawidyna, bardzo silne oddziaływania niekowalencyjne prowadzą do stałej dysocjacji  $4 \times 10^{-14}$  M. Unieruchomienie może jednak wpływać na właściwości badanych molekuł. Istnieje niebezpieczeństwo, że pozbawiona pewnej liczby stopni swobody cząsteczka będzie zachowywać się inaczej niż w warunkach naturalnych. Można to (do pewnego stopnia) zweryfikować porównując wyniki otrzymane dla pojedynczych, unieruchomionych cząsteczek z wynikami otrzymanymi dla zespołu cząsteczek w roztworze.

Należy także pamiętać o *crosstalk* and *crossexcitation*. *Crosstalk* wynika z faktu nakładania się widm emisji donora i akceptora, a przez to fotony fluorescencji pochodzące od obu barwników nie są całkowicie rozdzielone przed detekcją. W konsekwencji część fotonów emitowanych przez donor podlega detekcji w kanale akceptora, a część fotonów akceptora może być rejestrowanych w kanale donora. Wielkość *crosstalk* można oszacować mierząc w obu kanałach fluorescencję samego donora i samego akceptora, a następnie brać przy pomiarach pod uwagę poprawkę. Natomiast *crossexcitation* może nastąpić, kiedy widma wzbudzenia donora i akceptora nie są całkowicie rozdzielone i wtedy następuje niewielkie bezpośrednie wzbudzenie akceptora w wyniku naświetlaniem fali o długości odpowiadającej maksimum absorpcji donora. Wielkość *crossexcitation* można oszacować przeprowadzając eksperyment wstępny: mierząc fluorescencję samego akceptora, kiedy poddawany jest wzbudzeniu falą o długości charakterystycznej maksimum absorpcji donora.

Innym problemem jest fakt, że barwniki mają ograniczoną trwałość i mogą wyemitować tylko określoną liczbę fotonów zanim utracą swoje własności fluorescencyjne, zjawisko to nazywamy fotowybielaniem. Czas życia barwników można wydłużyć przeprowadzając eksperyment w warunkach beztlenowych, niestety brak tlenu powoduje wydłużenie czasu życia stanu trypletowego [3], co nie jest zjawiskiem pożądanym. Istnieją jednak także inne czynniki zmniejszające fotowybielanie: Trolox [3], oksydaza glukozy [44]. W spFRET cząsteczka jest unieruchomiona, fluorofory znajdują się pod ciągłym działaniem promieniowania wzbudzającego, co może znacznie przyspieszyć ich fotowybielanie. Konieczne jest zatem zoptymalizowanie intensywności wzbudzenia, wybór odpowiednich barwników, a także zastosowanie efektywnych metod analizy danych.

Pomimo trudności pomiarowych, FRET jest narzędziem pozwalającym obserwować zachowanie makrocząsteczek biologicznych, w warunkach coraz

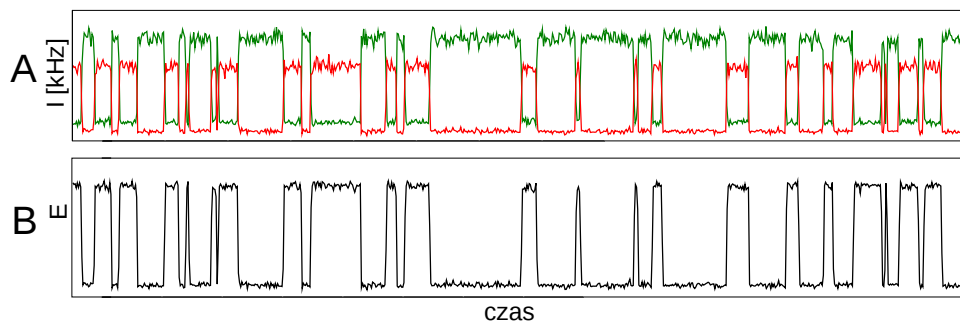
bliższych naturalnym. Umożliwił także odkrycie wielu cech białek, DNA i rybozymów, które były nieuchwytnie, kiedy mierzono układy innymi metodami.

### 2.1.4 Alternatywne techniki FRET

Istnieje możliwość przyłączenia do cząsteczki trzech barwników fluorescencyjnych i uzyskanie dzięki temu dodatkowego kanału informacji o stanie badanej molekuly. W większości przypadków do trójkolorowego FRET wykorzystuje się wzbudzenie przemienne ALEX (ang. *Alternating-Laser Excitation*) [45]. Badaną cząsteczkę z przyłączonymi barwnikami wzbudza się światłem o zmieniającej się skokowo długości fali, odpowiadającej maksimum wzbudzenia każdego z barwników [46, 47, 48]. Prowadzono także badania z wykorzystaniem czterech fluoroforów [49, 50]. Analiza danych dla FRET więcej niż 2-kolorowego jest skomplikowana, ponieważ należy uwzględnić możliwość przeniesienia energii pomiędzy wszystkimi barwnikami.

### 2.1.5 Analiza i estymacja parametrów w eksperymentach spFRET

W spektroskopii spFRET rejestruje się w dwóch kanałach fotony fluorescencji donora i akceptora. W badaniach cząsteczek biologicznych trajektorie wykazują często dwie lub więcej dyskretne wartości  $E$ . W najprostszym przypadku, kiedy badana cząsteczka może znajdować się w jednym z dwóch stanów, obserwuje się dwie różne wydajności FRET: np.  $E_{\text{on}} = 0,8$  i  $E_{\text{off}} = 0,2$  (rys. 2.5). Zaprezentowane na rys. 2.5 trajektorie przedstawiają najprostszy model 2-stanowy.



**Rysunek 2.5:** A) Przykładowe trajektorie fluorescencji pojedynczej cząsteczki donora ( $I_D$  – linia zielona) i akceptora ( $I_A$  – linia czerwona) oraz B) obliczona na ich podstawie trajektoria spFRET (2.2).

Mając do dyspozycji trajektorie fluorescencji donora i akceptora możliwe jest przeprowadzenie analizy i odpowiedź na pytanie: jak szybko układ zmienia się pomiędzy poszczególnymi stanami. W tym celu wykorzystać można trajektorie fluorescencji danego barwnika, lub też trajektorie FRET. Możliwa jest także analiza globalna fluorescencji obu barwników. Najczęściej stosowaną przez eksperymentatorów techniką jest analiza on-off, w której dopasowywana jest krzywa wykładnicza do histogramów czasów, które układ spędził w każdym ze stanów. Istnieje także szereg innych metod, takich jak analiza histogramów zliczeń fotonów, PCH (ang. *Photon Counting Histogram*) zastosowana do trajektorii fluorescencji [51] oraz dopasowanie do histogramów FRET. Na dzień dzisiejszy metody te pozwalają na estymację parametrów kinetycznych dla modelu 2-stanowego. Inną, coraz częściej wykorzystywaną metodą, są ukryte modele Markova, HMM (ang. *Hidden Markov Modelling*) dla trajektorii FRET [52, 53, 54], z wykorzystaniem których estymować można parametry dla modeli bardziej skomplikowanych, niż 2-stanowe. Istnieją także metody nie wymagające łączenia fotonów w grupy w celu obliczenia trajektorii. Jedną z nich jest dopasowanie funkcji największej wiarygodności ML (ang. *Maximum Likelihood*) do trajektorii czasów międzydetekcyjnych [21, 22].

Niejednokrotnie cząsteczki biologiczne działają według bardziej skomplikowanych schematów, na przykład 3-, 4-, lub więcej stanowych, o różnej organizacji. Zwiększa się wtedy liczba parametrów, które należy poddać estymacji. Analiza staje się wtedy nieco bardziej skomplikowana, niemniej jednak możliwa do przeprowadzania. To zagadnienie omawiane będzie szerzej w rozdziałach 5.2 oraz 5.3.

## 2.2 Wybrane przykłady eksperymentalne

---

W niniejszym rozdziale zaprezentowane będą przykłady zastosowania techniki FRET na poziomie pojedynczych cząsteczek do badania dynamiki i funkcji molekuł biologicznych. Ze względu na dużą ilość różnorodnych badań skupiono się tutaj na pracach eksperymentalnych dotyczących rybozymów. Przedstawione będą wyniki jakie zostały otrzymane dla różnych typów rybozymów i porównam je z tymi otrzymanymi dla zespołu cząsteczek, z uwzględnieniem korzyści jakie mogą przynieść badania na poziomie pojedynczych cząsteczek.

---

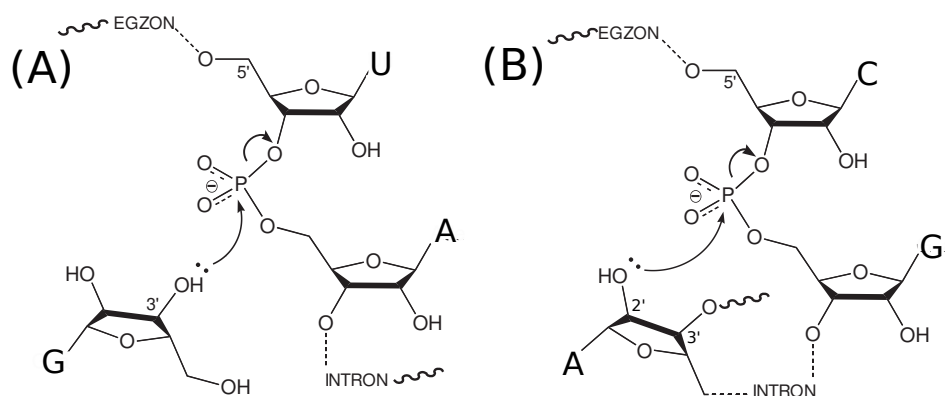
### 2.2.1 Rybozymy

Przez długi czas uważano, że jedynymi substancjami, które mogą katalizować reakcje w żywych organizmach są białka (enzymy). Dopiero w latach 80-tych XXw. Thomas Cech, w trakcie badań nad splicingiem RNA mających na celu znalezienie enzymu katalizującego wycinanie intronów RNA u *Tetrahymena thermophila*, stwierdził, że w badanych cząsteczkach splicing zachodzi nawet w nieobecności białek. Oznaczało to, że sama cząsteczka RNA posiada właściwości katalityczne [55]. W tym samym czasie do podobnego wniosku doszedł Sidney Altman badając dojrzewanie tRNA (ang. *transfer RNA*) w rybosomach [56]. Takie cząsteczki - RNA mające właściwości katalityczne - nazwano rybozymami, a ich odkrywcy w 1989 r. otrzymali za swoje badania Nagrodę Nobla w dziedzinie chemii. Aktywność enzymatyczna rybozymów jest interesująca z kilku powodów. Po pierwsze, uczestniczą one w ważnych reakcjach, których poznanie jest niezbędne, aby zrozumieć mechanizmy rządzące metabolizmem komórkowym. Po drugie, rybozymy mają zazwyczaj prostszą budowę od enzymów białkowych, składają się bowiem jedynie z czterech rodzajów monomerów, a dodatkowo można lepiej przewidzieć ich strukturę z uwagi na komplementarność pomiędzy poszczególnymi zasadami Watsona-Cricka (adeniną i uracylem oraz guaniną i cytozyną). Ponieważ rybozymy są mniej złożone niż enzymy białkowe, mechanizm ich działania powinien być prostszy do rozwikłania. Po trzecie, wykazują one działanie selektywne, co może być wykorzystane w terapii celowanej w RNA, choć technika ta jest na razie we wczesnym stadium rozwoju,

a jej zastosowanie będzie wymagało rozwiązania problemów związanych m.in. ze stabilnością chemiczną oligonukleotydów w środowisku komórkowym, transportem do odpowiedniego miejsca w komórce oraz fałdowaniem do aktywnej konformacji. Po czwarte, najnowsze osiągnięcia spektroskopii pojedynczych cząsteczek pozwalają na obserwowanie fałdowania i aktywności pojedynczych cząsteczek rybozymów, dzięki czemu otrzymać można informacje, które w eksperymencie na zespole cząsteczek byłyby niedostępne z powodu uśredniania. Do tej pory ukazał się szereg prac przeglądowych dotyczących rybozymów [6, 26, 57, 58, 59, 60, 61] i wykorzystujących w celu ich zbadania technikę spFRET.

Cząsteczki RNA, na poziomie struktury pierwszorzędowej, składają się z jednoniciowych łańcuchów. Poszczególne ich fragmenty mogą jednak łączyć się ze sobą wiązaniami wodorowymi, które wytwarzają się pomiędzy komplementarnymi zasadami (adenina z uracylem i cytozyna z guaniną). Efektem tego jest powstanie rozmaitych motywów strukturalnych, takich jak pętle, wybrzuszenia, czy spinki. Tworzą one strukturę drugorzędową cząsteczki RNA. Dalsze oddziaływanie tych elementów daje przestrzennie sfałdowaną strukturę trzeciorzędową. Ponieważ każdy rybonukleotyd niesie ze sobą jeden ładunek ujemny, do odpowiedniego sfałdowania się RNA, tj. powstania prawidłowej (natywnej) struktury trzeciorzędowej RNA konieczna jest obecność jonów dodatnich, które będą ekranować elektrostatyczne odpychanie między elementami składowymi łańcucha. W komórce najczęściej funkcję tę spełniają kationy magnezowe lub sodowe. Po opublikowaniu odkryć Cecha i Altmana pojawiło się wiele doniesień o aktywności katalitycznej RNA, a także o samo-tnących i samo-składających się cząsteczkach [62, 63]. W ramach ogólnego podziału rybozymów ze względu na budowę i rodzaj przeprowadzanej reakcji wyróżnić można introny grup I i II, RNA RNazy P, rybozymy typu hairpin, hammerhead, VS (ang. *Varkud Satellite*), HDV (ang. *Hepatitis Delta Virus*) oraz spliceosomy [58] i makrokompleksy rybonukleoproteinowe. Inny podział rybozymów opiera się na wielkości cząsteczek: istnieją rybozymy małe, o niezbyt skomplikowanej budowie oraz duże, złożone z licznych pętli i części helikalnych.

Złożona struktura trójwymiarowa rybozymów pozwala im na przeprowadzenie reakcji katalitycznej w ściśle określonym miejscu. Rybozymy potrafią nie tylko przeorganizować swoje miejsce aktywne, ale także katalizować reakcje poprzez zbliżenie do siebie dwóch substratów [64]. Reakcje przeprowadzane przez rybozymy obejmują m.in.: transestryfikację (np. rybozym typu hairpin), transfer nukleotydu (introny grupy I i II), hydrolizę (RNA RNazy P), a także syntezę wiązania



**Rysunek 2.6:** Dwa przykłady mechanizmów aktywności katalitycznej rybozymów. A) Intron grupy I: nukleofilowy atak tlenu 3' guanozyny, która nie należy do rozcinanego łańcucha RNA. B) Intron grupy II: nukleofilowy atak tlenu 2' nukleotydu, będącego częścią intronu wycinanego łańcucha. W obu przypadkach produktem jest egzon z grupą hydroksylową na końcu 3'. Dwa oddzielone wcześniej intronem egzony łączą się, przy czym powstaje intron, na którego końcu 5' znajduje się reszta guanozyny.

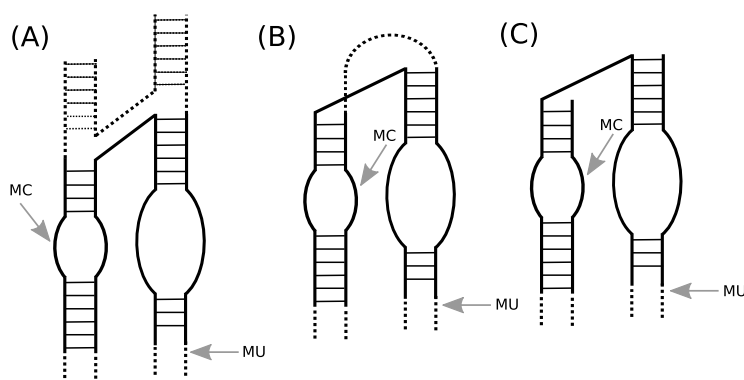
peptydowego (rybosom). Pomimo, że w niektórych przypadkach w żywych organizmach działanie rybozymów wspomagane jest przez białka, pojedyncze cząsteczki RNA również wykazują aktywność katalityczną. Przykłady mechanizmów reakcji katalizowanych przez rybozomy przedstawiono na rys. 2.6.

Większość znanych obecnie rybozymów katalizuje reakcje jednoobrotowe, czyli takie, w wyniku których następuje nieodwracalna zmiana rybozymu, a w konsekwencji utrata właściwości katalitycznych. Niektóre rybozomy mogą jednak przeprowadzać wiele obrotów reakcji, są zatem katalizatorami w tradycyjnym znaczeniu tego terminu i nie zmieniając się w wyniku reakcji, podobnie jak enzymy białkowe. Przykładem może tu być rybozym występujący w rybosomie [65], a także RNA RNazy P [66]. W badaniach *in vitro* często wykorzystuje się pochodne rybozymów “jednoobrotowych”, jednak zmodyfikowane w taki sposób, aby mogły katalizować wiele obrotów reakcji.

W dalszej części rozdziału przedstawiono dwa najlepiej zbadane dotąd rybozomy: rybozym typu hairpin przeprowadzającym reakcję transestryfikacji 2'O oraz rybozym *Tetrahymena thermophila* (intron grupy I), zaliczanych odpowiednio do grupy małych i dużych rybozymów. W obu przypadkach są to cząsteczki, które *in vivo* katalizują tylko jeden obrót reakcji. Dla porównania omówiono także inne układy, m.in. RNA RNazy P, który katalizuje reakcję hydrolizy oraz



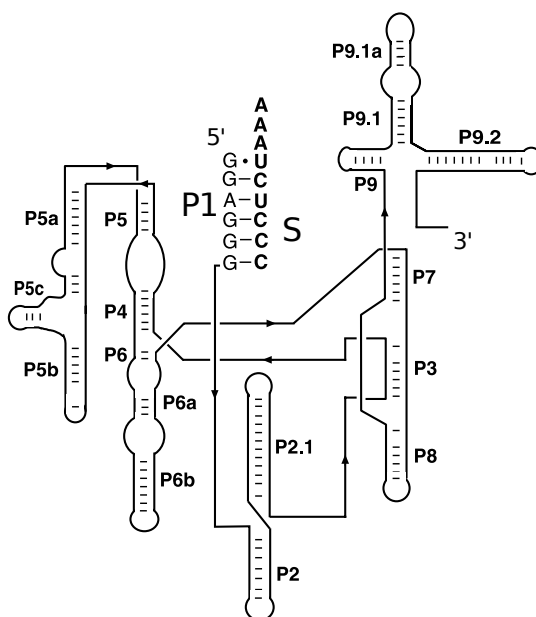
W eksperymentach *in vitro* badano zarówno naturalną formę rybozomu (2.8A), jak i jego różne modyfikacje. W formie naturalnej, po zajściu reakcji katalitycznej jedna z części substratu pozostaje połączona z rybozymem [15, 28, 69], co uniemożliwia przyłączenie kolejnego substratu. Podobnie zachowuje się uproszczona forma rybozomu, w której usunięte zostały dwa ramiona nie zawierające pętli wewnętrznych (2.8B) [70, 71]. Dalsza modyfikacja, dzięki której obie części rozciętego substratu odłączają się, pozwala na odtworzenie rybozomu po reakcji rozcinania i umożliwia przyłączanie kolejnego substratu [7, 11, 72, 73]. Taką formę nazywać będziemy formą minimalną (2.8C).



**Rysunek 2.8:** Modyfikacje rybozomu typu hairpin wykorzystywane w badaniach pojedynczych cząsteczek. Linią ciągłą oznaczono regiony aktywnie katalityczne, a liniją kropkowaną obszary, które mają wpływ na szybkość reakcji. Strzałkami zaznaczono miejsce unieruchamiania (MU) cząsteczki oraz miejsce cięcia (MC) łańcucha. A) Forma naturalna, B) forma uproszczona – pozbawiona ramion nie zawierających pętli wewnętrznych, C) forma minimalna, która może katalizować wielu obrotów reakcji.

Rybozym *Tetrahymena thermophila* katalizuje wycinanie intronów (niekodujących fragmentów) własnego łańcucha RNA. Należy on do grupy dużych rybozymów i składa się z wielu helis oraz pętli wewnętrznych (rys. 2.9).

RNaza P odgrywa kluczową rolę w dojrzewaniu końca 5' tRNA u bakterii [56], archeowców oraz u organizmów eukariotycznych [74]. RNaza P *in vivo* występuje w postaci holoenzymu, tzn. w akcie katalitycznym współuczestniczy kofaktor białkowy. U bakterii kofaktor ten składa się z jednej cząsteczki białka, u archeowców z czterech, natomiast u organizmów eukariotycznych z dziewięciu lub dziesięciu cząsteczek białka [66]. W warunkach *in vitro* RNaza P wykazuje (w odpowiednim pH i w obecności jonów dodatnich) właściwości katalityczne nawet w nieobecności części białkowych. Co ciekawe, szybkość reakcji prowadzonych przez rybozymy pochodzące z organizmów eukariotycznych potrafi być nawet do kilku rzędów wielkości niższa



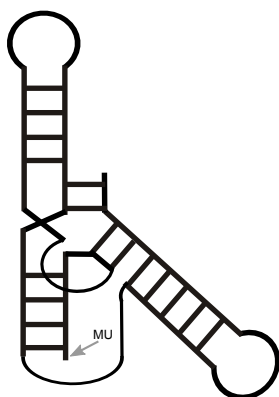
Rysunek 2.9: Budowa rybozomu *Tetrahymena thermophila*.

niż w przypadku rybozymów bakteryjnych [74]. Przyczyną tego mogą być różnice w budowie miejsca przyłączenia substratu u tych organizmów.

Innym ciekawym przykładem rybozimu jest DAza, niewielki rybozym katalizujący cykloaddycję Dielsa-Aldera [64], którego schemat przedstawiono na rys. 2.10. Reakcja ta polega na stereospecyficznym formowaniu się wiązania węgiel-węgiel między antracenenem (dien) i imidem kwasu maleinowego (dienofil). Rybozym ten katalizuje więc reakcję międzycząsteczkową między substratami nie będącymi cząsteczkami RNA. W tym przypadku rolę rybozimu jako katalizatora jest zbliżenie do siebie substratów. Działanie DAzy jest inhibitowane przez powstający produkt [68].

### 2.2.2 Fałdowanie rybozymów

W działaniu katalitycznym rybozymów bardzo istotna jest odpowiednia struktura trzeciorzędowa. Aby kontrolować proces uzyskiwania zdolności katalitycznych przez cząsteczkę biopolimeru, niezbędne jest zatem zrozumienie zależności drogi fałdowania od czynników zewnętrznych, takich jak stężenie jonów, czy temperatura. Dotyczy to zarówno enzymów RNA, jak i enzymów białkowych. Warunkiem poprawnego sfałdowania RNA jest obecność dodatnich jonów metali (np. sodowych, magnezowych), od których stężenia zależy położenie punktu początkowego



**Rysunek 2.10:** Struktura formy sfałdowanej Diels-Alderazy. Zaznaczono miejsce unieruchomienia (MU), które stosuje się w badaniach na poziomie pojedynczych cząsteczek.

fałdowania na powierzchni energii swobodnej. Punkt ten z kolei determinuje drogę, jaką cząsteczka musi przebyć, aby osiągnąć stan natywny. Sfałdowane cząsteczki rybozymów wykazują lokalną dynamikę konformacyjną. Tak na przykład w cząsteczce rybozymu typu hairpin położenie helikalnych ramion fluktuuje pomiędzy dwoma pozycjami. Konformację, w której ramiona zawierające pętle wewnętrzne znajdują się blisko siebie określa się mianem formy zwiniętej (ang. *docked*), zaś konformację, w której ramiona te oddalone są od siebie na pewną odległość nazywa się formą rozwiniętą (ang. *undocked*).

### 2.2.2.1 Rybozym *Tetrahymena thermophila*

Mechanizm fałdowania rybozymu *Tetrahymena thermophila* był badany zarówno w eksperymentach dla zespołów cząsteczek [75, 76] jak i dla pojedynczych cząsteczek [8, 77, 78]. Stwierdzono, że rybozym ten może się znajdować w 3 stanach: rozfałdowanym, niepoprawnie sfałdowanym i poprawnie sfałdowanym (natywnym). Forma rozfałdowana charakteryzuje się wydajnością FRET na poziomie  $E = 0,1$ ; sfałdowana niepoprawnie:  $E = 0,3$ ; a poprawnie sfałdowana:  $E = 0,9$ . W procesie fałdowania niezbędna jest obecność jonów magnezowych, jednak położenie punktu startowego, a tym samym drogi fałdowania zależy przede wszystkim od stężenia jonów sodowych w roztworze. Dla wysokiego stężenia ( $> 250 \text{ mmol/dm}^3$ ) fałdowanie przebiega bardzo szybko do stanu natywnego, natomiast przy średnim stężeniu ( $150 - 250 \text{ mmol/dm}^3$ ) obserwuje się stan przejściowy, z którego cząsteczka może przejść do stanu natywnego, lub niepoprawnie sfałdowanego. Z kolei dla

małego stężenia jonów sodowych ( $< 150 \text{ mmol/dm}^3$ ) obserwuje się na drodze fałdowania, oprócz stanu przejściowego, pułapkę energetyczną. Stwierdzono także, że do poprawnego sfałdowania rybozomu nie jest konieczne wcześniejsze wytworzenie struktury drugorzędowej.

Rybozym *Tetrahymena thermophila*, podobnie jak rybozym Hairpin, wykazuje lokalne fluktuacje struktury. W badaniach na zespole cząsteczek wyznaczono stałą równowagową zwijania ( $K_{\text{dock}} = 7 \pm 3$ ), stwierdzono, że tworzenie struktury trzeciorzędowej przebiegać może różnymi drogami, a także zaobserwowano długo żyjące stany rozfałdowane. Badania pojedynczych cząsteczek pozwoliły na wyznaczenie stałych szybkości zwijania i rozwijania [8] oraz kształtu powierzchni energii swobodnej fałdowania [77]. Stała rozwijania otrzymana na podstawie wyników eksperymentalnych wynosiła ( $k_{\text{undock}} = 1,62 \pm 0,08 \text{ s}^{-1}$ ) i zwijania ( $k_{\text{dock}} = 0,224 \pm 0,015 \text{ s}^{-1}$ ), a stała równowagi zwijania ( $K_{\text{dock}} = k_{\text{dock}}/k_{\text{undock}} = 7,2 \pm 0,8 \text{ s}^{-1}$ ) [8].

### 2.2.2.2 Rybozym typu hairpin

Formą naturalną rybozomu hairpin jest tzw. struktura Holliday'a, w której dwa ramiona wyposażone w pętle wewnętrzne znajdują się blisko siebie. Wykazano jednakże, że aktywność katalityczną mogą wykazywać również formy zmodyfikowane, w których ramiona nie zawierające pętli zostały usunięte (rys. 2.8). Sugeruje to, że na przebieg reakcji najważniejszy wpływ będzie wywierać pozycja ramion posiadających pętle. W celu zbadania jej kinetyki, przygotowano rybozym oznakowany fluorescencyjnie na dwóch ramionach z pętlami wewnętrznymi. W eksperymencie zaobserwowano trzy różne wydajności FRET. Wartości wysoka ( $E = 0,81$ ) i niska ( $E = 0,15$ ) odnosiły się odpowiednio do rybozomu w formie zwiniętej i rozwiniętej z przyłączonym substratem, natomiast wartość pośrednia ( $E = 0,38$ ) odpowiadała wolnemu rybozymowi. Stwierdzono też, że rybozym z przyłączonym substratem wykazuje inną dynamikę konformacyjną przed i po jego rozcięciu.

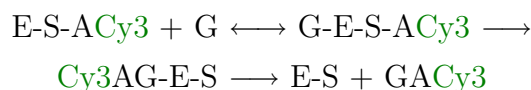
Badając formę minimalną (rys. 2.8C), udało się wyznaczyć stałą zwijania rybozomu z nierozciętym substratem, której wartość była taka sama jak ta otrzymana w badaniach na zespole cząsteczek ( $0,008 \text{ s}^{-1}$ ). Stwierdzono jednak, że poszczególne cząsteczki wykazują różne stałe rozwijania. Najwolniejsza stała rozwijania ( $0,005 \text{ s}^{-1}$ ) zgadzała się z wynikami otrzymanymi dla zespołu cząsteczek. W przypadku innych cząsteczek obserwowane stałe szybkości rozwijania odbiegały od tej wartości, zaobserwowano jednak efekt pamięci: każda cząsteczka wykazywała

tylko jedną, konkretną stałą szybkości rozwijania, która nie zmieniała się w trakcie eksperymentu. Dla porównania, rybozym z rozciętym substratem wykazuje stałą szybkości zwijania równą  $0,02 \text{ s}^{-1}$ , przy czym zaobserwowano różnorodność kinetyczną stałej rozwijania. Tak więc stała zwijania rybozymu z przyłączonym, rozciętym substratem jest 2,5 razy większa niż stała zwijania rybozymu, do którego przyłączony jest nie rozcięty substrat. Porównując wyniki eksperymentalne otrzymane dla formy minimalnej i naturalnej rybozymu stwierdzono, że forma minimalna wymaga dwa do trzy razy większego stężenia jonów magnezowych, a jej stała zwijania jest o trzy rzędy wielkości mniejsza niż stała zwijania dla formy naturalnej.

### 2.2.2.3 Inne układy

Innym dobrym modelem do badań kinetyki fałdowania jest domena katalityczna RNazy P RNA z *Bacillus Subtilis*, ponieważ obszar jej energii swobodnej nie zawiera pułapek energetycznych [79]. Dodatkowo znaczny rozmiar pozwala na stosowanie dla niej zasad fałdowania dużych cząsteczek RNA. W badaniach na poziomie pojedynczych cząsteczek potwierdzono wcześniejsze spostrzeżenia o fałdowaniu z udziałem stanów przejściowych. Stwierdzono też, że na drodze fałdowania występują trzy takie stany, a nie, jak przedtem uważano, jeden. Warunki początkowe (stężenie jonów dodatnich, pH roztworu) determinują punkt startowy fałdowania na powierzchni energii swobodnej, wywierając tym samym bezpośredni wpływ na drogę fałdowania (tzn. dobór stanów przejściowych które będzie osiągać cząsteczka w trakcie fałdowania).

Wykorzystując technikę FRET pojedynczych cząsteczek do badania DAzy stwierdzono, że jej fałdowanie do formy aktywnej przebiega poprzez stan przejściowy. Cząsteczki DAzy były znakowane fluorescencyjnie, Cy3 i Cy5, działającymi odpowiednio jako donator i jako akceptor. Stan rozfałdowany stanowił kłębek statystyczny, w którym nie wytworzyły się jeszcze wiązania wodorowe między komplementarnymi parami zasad. Po dodaniu jonów jednododatnich ( $\text{Na}^+$ ) wytwarzała się struktura drugorzędowa, zawierająca części helikalne i pętle wewnętrzne. W ten sposób cząsteczka osiągała stan pośredni. Do utworzenia prawidłowej struktury trzeciorzędowej (tzn. przejścia do stanu sfałdowanego) konieczne było dodanie do układu jonów dwudodatnich, dzięki którym fragmenty struktury drugorzędowej zaczynały oddziaływać ze sobą.



**Rysunek 2.11:** Mechanizm katalizy enzymu *Tetrahymena thermophila* dla oligonukleotydowego substratu (S-A = CCCUCUA), do którego dołączony jest barwnik fluorescencyjny Cy3. Guanozyna (G) przyłącza się do rybozemu E, który katalizuje rozcięcie substratu, a produkt (GACy3) oddziela się od rybozemu.

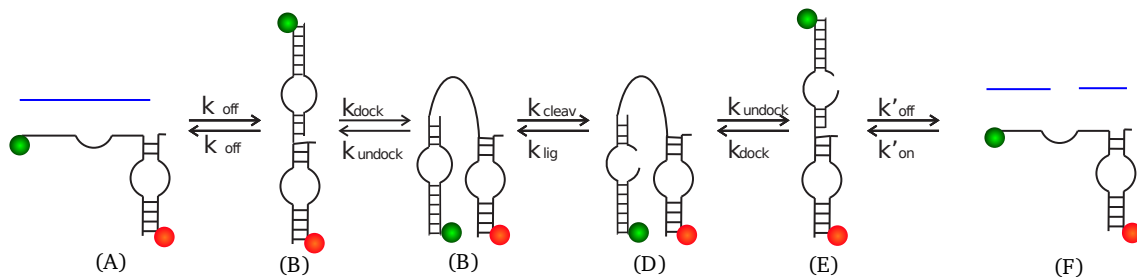
### 2.2.3 Aktywność katalityczna rybozymów i jej sprzężenie z dynamiką konformacyjną

Badania spFRET pokazują, że poszczególne cząsteczki rybozymów różnią się od siebie pod względem aktywności katalitycznej. Rozkłady aktywności katalitycznej rybozymów są zgodne z wartościami średnimi uzyskanymi w badaniach zespołów cząsteczek [7, 8]. Stwierdzono, że w przypadku rybozymów aktywność ta jest najprawdopodobniej ściśle związana z dynamiką konformacyjną, której rybozomy, podobnie jak inne makrocząsteczki biologiczne, ulegają. Na przykład dla rybozemu typu hairpin dynamika ta polega na zmianie częstości przejść pomiędzy stanem zwiniętym i rozwiniętym rybozemu, gdy przyłączony substrat ulega rozcięciu.

#### 2.2.3.1 Rybozym *Tetrahymena thermophila*

Zbadano aktywność katalityczną *Tetrahymena thermophila* dla pojedynczych, unieruchomionych cząsteczek [8]. Schemat reakcji jest przedstawiony na rysunku 2.11. Jako że po rozcięciu produkt oznakowany barwnikiem fluorescencyjnym oddziela się od rybozemu, możliwe było określenie momentu rozcięcia. Widoczne było ono jako nagłe zmniejszenie intensywności fluorescencji.

Etapem limitującym szybkość jest etap rozcinania substratu. Zależność szybkości rozcięcia  $v$  od stężenia guanozyny  $[G]$  była zgodna z równaniem Michaelisa-Menten [80],  $v = k[G]/(K_{1/2} + [G])$ , ze stałą szybkości rozcięcia  $k = 0,21 \pm 0,01 \text{ s}^{-1}$  i stałą Michaelisa-Menten  $K_{1/2} = 0,26 \pm 0,04 \text{ mmol/dm}^3$ . Wyniki te zgadzały się z tymi otrzymanymi dla zespołu cząsteczek, które wynosiły  $k_{\text{max}} = 0,21 \pm 0,01 \text{ s}^{-1}$ , a  $K_{1/2} = 0,30 \pm 0,05 \text{ mmol/dm}^3$ . Reakcja rozcinania zachodziła tylko wtedy, gdy rybozym znajdował się w konformacji poprawnie zwiniętej, dla której wydajność spFRET wynosiła  $E = 0,9$ .



**Rysunek 2.12:** Mechanizm katalizy formy minimalnej rybozomu typu hairpin. Do cząsteczki przyłączono dwa barwniki fluorescencyjne: Cy3 jako donor (zielony) i Cy5 jako akceptor (czerwony). (A) – wolny rybozom i przyłączający się do niego substrat, (B) rozwinięty rybozom z przyłączonym substratem, (C) zwinięty rybozom z przyłączonym substratem, (D) zwinięty rybozom z rozciętym substratem, (E) rozwinięty rybozom z rozciętym substratem, (F) wolny rybozom i uwolnione produkty.

### 2.2.3.2 Rybozom typu hairpin

Mechanizm aktywności katalitycznej formy minimalnej rybozomu typu hairpin przedstawiono na rys. 2.12. Ponieważ poszczególne etapy reakcji mają porównywalne szybkości, zależności kinetyczne są złożone i wymagają analizy numerycznej [7].

Wydatność FRET w stanie zwiniętym wynosi ok. 0,8, natomiast dla formy rozwiniętej ok. 0,15. Rybozom (zarówno w formie minimalnej jak i naturalnej) do którego przyłączony jest rozcięty substrat wykazuje szybkie przejścia między stanami zwiniętym i rozwiniętym. Dzięki temu na podstawie trajektorii FRET można określić dokładnie moment rozcięcia. Zjawisko to widoczne jest jako przejście z długotrwałego stanu wysokiego FRET ( $E = 0,8$ ) do szybkich fluktuacji pomiędzy wartościami  $E = 0,2$  i  $0,8$ . Obserwacja tego aktu katalitycznego możliwa jest na drodze pośredniej, dzięki znajomości dynamiki konformacyjnej form rybozomu z przyłączonym rozciętym i nierozciętym substratem. Dla formy minimalnej (rys. 2.8C) wyznaczono stałą równowagi rozcinania  $K \approx 0,5$ , oraz przedział,  $0,1$  do  $0,4 \text{ s}^{-1}$ , w którym mieści się stała szybkości rozcinania, a także analogiczny przedział,  $0,2$  do  $0,8 \text{ s}^{-1}$ , dla stałej ligacji [7]. Dla formy naturalnej stałe rozcinania i ligacji wahają się według różnych autorów [28, 69], między  $0,6$  a  $1 \text{ min}^{-1}$  dla stałej rozcinania i  $18-21 \text{ min}^{-1}$  dla stałej ligacji.

### 2.2.3.3 Inne układy

Porównano kinetykę działania RNA RNazy P dla organizmów prokariotycznych (bakterie, archeowce), jak i eukariotycznych [74, 81], w warunkach kiedy możliwy

jest tylko pojedynczy obrót reakcji. Stwierdzono, że bez udziału białek, stałe rozcinania przeprowadzanego przez rybozomy eukariotyczne (ludzki i *G. lamblia*) wynoszą odpowiednio  $2,6 \times 10^{-5}$  i  $3,5 \times 10^{-6} \text{ min}^{-1}$ , podczas gdy prokariotyczny RNA RNazy P przeprowadza proces rozcinania ze stałą szybkości  $8,4 \text{ min}^{-1}$ . Powodem tak dużych rozbieżności może być fakt nieobecności w rybozomach eukariotycznych fragmentów odpowiedzialnych za przyłączanie substratu, które występują w organizmach prokariotycznych. Po modyfikacji rybozomu bakteryjnego, w ramach której usunięto te fragmenty, stwierdzono, że stała rozcinania zmniejszyła się, jednak tylko do wartości  $2,5 \times 10^{-3} \text{ min}^{-1}$ . Pozostaje więc nadal różnica dwóch rzędów wielkości. Stwierdzono, że najprawdopodobniej w wyniku ewolucji RNA RNazy P coraz bardziej „uzależniał” się od pomocy kofaktorów białkowych, choć nawet u eukariotów nie utracił do końca swoich właściwości katalitycznych.

#### 2.2.4 Analiza danych w badaniach pojedynczych cząsteczek RNA

Przegląd metod statystycznej analizy danych w zastosowaniu do badań nad pojedynczymi cząsteczkami znaleźć można w pracach [82, 83]. W enzymologii, analiza danych pojedynczych cząsteczek RNA polega często na analizie trajektorii spFRET wykazujących skoki pomiędzy stanami o różnej wydajności przeniesienia energii fluorescencji  $E$ . Przy niskim poziomie szumu, liczbę stanów FRET można w prosty sposób określić wizualnie z wykresu trajektorii. Na tej samej zasadzie można ustalić progi pozwalające przypisać zmierzone wartości  $E$  do poszczególnych stanów, a następnie wyznaczyć długości interwałów czasu przebywania w tych stanach. Stałe szybkości przejść pomiędzy stanami wyznacza się analizując histogramy czasów przebywania w poszczególnych stanach FRET.

Wyznaczenie liczby stanów i rozdzielenie trajektorii na odcinki odpowiadające poszczególnym stanom mogą być trudne, jeśli szum eksperymentalny jest duży, jak ma to często miejsce w badaniach spFRET. W takim przypadku można zastosować metody eliminacji szumów, co skutkuje wygładzeniem trajektorii spFRET [13]. Pomimo tego, nie we wszystkich przypadkach udaje się przy pomocy analizy on-off określić szybkości przejść pomiędzy stanami bez przeprowadzania dodatkowych doświadczeń [15] i wciąż poszukuje się nowych metod statystycznej analizy danych [84]. Interesujące możliwości daje modelowanie trajektorii spFRET jako ukrytego procesu Markowa [52], lub modelowanie zliczeń w kanałach donora

jako modulowanych procesów Poissona [21, 22, 23], a następnie dopasowanie metodą największej wiarygodności (ML). Metoda ML nie wymaga łączenia fotonów w klasy w celu przeprowadzenia analizy. Dzięki temu unika się uśredniania otrzymanych wyników i możliwe jest zwiększenie rozdzielczości estymowania parametrów kinetycznych.

## 2.3 Podsumowanie rozdziału i wnioski

W niniejszym rozdziale przedstawiono przegląd najnowszej literatury dotyczącej badań nad fałdowaniem i właściwościami katalitycznymi rybozymów. Dzięki zastosowaniu FRET pojedynczych cząsteczek możliwe było odkrycie różnorodności kinetycznej dla formy minimalnej rybozemu hairpin. Pozwoliło także stwierdzić, że niektóre reakcje przeprowadzane przez rybozomy są znacznie bardziej skomplikowane, niż pokazywały to badania dla zespołu cząsteczek. W przedstawionych przykładach do analizy wykorzystywano dopasowania do histogramów czasów on-off, co nie pozwoliło określić wartości najszybszych przejść i konieczne było wprowadzanie modyfikacji w cząsteczkach RNA. Zastosowanie alternatywnych metod odzyskiwania parametrów kinetycznych pozwoliłoby uniknąć tego problemu i wydobyć wszystkie szybkości zmian dla omawianego modelu.

# Rozdział 3

## Cel i zakres pracy

Spektroskopia pojedynczych cząsteczek jest narzędziem dającym duże możliwości w badaniu zachowania pojedynczych makrocząsteczek. Jak pokazano w rozdziale 2.2, w przypadku rybozymów standardowa analiza on-off nie zawsze pozwala wydobyć wszystkie możliwe informacje z przeprowadzonego eksperymentu ze względu na ograniczoną rozdzielczość czasową i konieczność znalezienia optymalnej wartości progowej.

W najnowszej literaturze pojawiają się doniesienia o wykorzystaniu alternatywnych do on-off metod analizy danych, dzięki którym oszacowywać można parametry kinetyczne z większą rozdzielczością. Są to na przykład dopasowanie do histogramów FRET, ukryte modele Markova HMM (ang. *Hidden Markov Models*) [52]. Metody te wymagają jednak do przeprowadzenia analizy danych w postaci trajektorii fluorescencji, co oznacza, że zarejestrowane fotony muszą być przed analizą grupowane, wykorzystując wybrane okno czasowe, a tym samym często uniemożliwiają oszacowanie szybszych zmian w modelu. Istnieje jednak metoda, nie wymagająca obliczania trajektorii fluorescencji i opierając się na dopasowaniu metodą największej wiarygodności ML (ang. *Maximum Likelihood*) do trajektorii zliczeń fotonów [22].

Najlepsza metoda analizy to taka, która pozwala z jak najwyższą precyzją i dokładnością oszacować parametry kinetyczne badanego modelu, mając do dyspozycji jak najkrótszą trajektorię. Algorytm dopasowujący powinien zbiegać się do minimum globalnego z punktów jak najbardziej oddalonych od teoretycznych wartości parametrów. Dodatkowo w stosunkowo prosty sposób powinna istnieć możliwość odtwarzania parametrów kinetycznych dla modeli bardziej skomplikowanych, niż 2-stanowy. Ważnym aspektem jest też możliwość

określenia najbardziej prawdopodobnego modelu kinetycznego dla poszczególnych danych eksperymentalnych. Możliwość zastosowania metody ML do estymacji parametrów kinetycznych w eksperymentach spFRET nie była dotąd testowana dla szerokiego zestawu parametrów. W takich badaniach wykorzystuje się często symulacje komputerowe metodą Monte Carlo. Symulowane są dane o zadanych parametrach, a następnie wykonywane jest dopasowanie. Taki eksperyment powtarza się wielokrotnie w celu określenia właściwości statystycznych otrzymanych wyników.

**Celem głównym** niniejszej pracy jest określenie możliwości zastosowania metody ML do estymacji parametrów i rozróżniania modeli kinetycznych w eksperymentach FRET pojedynczych cząsteczek.

W trakcie pracy porównywano wyniki otrzymane metodą ML z najczęściej stosowaną analizą on-off. Przy użyciu przygotowanych programów wykonywano symulacje i dla wybranych modeli kinetycznych odtwarzano dane analogiczne do tych otrzymanych w rzeczywistym eksperymencie. Jako algorytm symulacji wybrano znakowane Markowsko modulowane procesy Poissona MMMPP (ang. *Marked Markov Modulated Poisson Processes*). Następnie wykorzystując te dane analizowano je przy użyciu metody największej wiarygodności (ML) w celu odtworzenia parametrów modelu. Sprawdzano także jak przedstawiałyby się wyniki, jeśli do analizy wykorzystano by fotony tylko jednego barwnika.

**Cele szczegółowe to:**

1. Stworzenie i walidacja programów do symulacji trajektorii zliczeń fotonów o dwóch różnych kolorach przy użyciu algorytmu MMMPP, a także programów do analizy danych metodą on-off i ML. Programy uwzględniają modele o różnej liczbie stanów i skomplikowości, w dopasowaniu metodą ML istnieje także możliwość uwzględnienia warunku równowagi szczegółowej.
2. Określenie optymalnych i granicznych (czyli takich, poza którymi niemożliwe jest odtworzenie modelu kinetycznego) parametrów eksperymentalnych, które pozwalają odzyskać parametry kinetyczne z eksperymentów FRET pojedynczych cząsteczek dla badanych metod analizy migotania fluorescencji. Badano zależność jakości dopasowania od takich czynników jak:
  - liczba zarejestrowanych fotonów (od  $n = 10^3$  do  $n = 5 \times 10^4$  fotonów)
  - długość trajektorii mierzona liczbą obserwowanych cykli

- intensywność fluorescencji, rozumiana tutaj dwojako: jako liczba fotonów danego barwnika emitowana w jednostce czasu w danym stanie  $I_s^c$  oraz jako sumaryczna liczba fotonów (suma fotonów pochodzących od obu barwników) obserwowana w danym stanie  $I_s$ , gdzie  $I$ , to intensywność,  $s$  stan, natomiast  $c$  oznacza kolor fotonów fluorescencji.  $I_s^c$  wahało się w przedziale od 0,2 do 80, a  $I_s$  od 1 do 100.
  - szybkość zmian konformacyjnych  $k_{ij}$  zmieniano w przedziale od wolnych przejść ( $k_{ij} = 1$ ) do szybkich ( $k_{ij} = 100$ ), jednostką była odwrotność czasu.
  - rozdzielenie intensywności, dla modelu 2-stanowego rozumiane jako iloraz intensywności niższej ( $I_l$ ) do wyższej ( $I_h$ ) dla danego koloru barwnika.  $I_l/I_h$  zmieniało się w przedziale od 0,2 (dobre rozdzielenie) do 0,8 (gdzie intensywności w obu stanach są do siebie bardzo zbliżone.)
  - różna liczba stanów (2, 3, 4, 5)
  - różna topologia modeli (liniowe i cykliczne)
3. Stwierdzenie, jakie konsekwencje niesie ze sobą wzięcie do analizy fotonów tylko jednego koloru
  4. Zbadanie wpływu parametrów startowych dopasowania na jakość estymat
  5. Określenie możliwości rozróżniania modeli kinetycznych w eksperymentach FRET
    - wykorzystanie AIC oraz BIC oraz odpowiedź na pytanie które z tych kryteriów jest bardziej wydajne
    - rozróżnianie pomiędzy modelami o różnej liczbie stanów
    - rozróżnianie modeli o takiej samej liczbie stanów, a różnej topologii, na przykład liniowe i cykliczne
  6. Określenie optymalnych warunków, jakie powinny być zachowane w eksperymencie, by zmaksymalizować prawdopodobieństwo poprawnego odtworzenia parametrów kinetycznych.

**Metodologia** skupiała się wokół trzech aspektów:

1. Praca analityczna, obejmująca opracowanie na podstawie dostępnych w literaturze informacji algorytmów symulacji danych oraz metod analizy.

Należało wybrać optymalne dla rozpatrywanego problemu algorytmy, które miały zostać zaimplementowane. Ta część pracy polegała także na zaplanowaniu parametrów przeprowadzanych symulacji tak, by były one jak najbliższe rzeczywistemu eksperymentowi. Jako przykład wybrano reakcje rybozymów, ponieważ katalityczne RNA może charakteryzować się stałymi kinetycznymi, które różnią się między sobą nawet o kilka rzędów wielkości. Poszukiwano także czynników, które są ważne dla eksperymentu i w jakikolwiek sposób mogą wpływać na jakość dopasowania np. intensywność fluorescencji, długość trajektorii.

2. Praca programistyczna, w zakres której wchodzi przygotowanie programów i przeprowadzenie symulacji. Programy do symulacji wyników eksperymentalnych napisane zostały w języku C++. Skrypty dopasowujące stworzone są przy użyciu pakietu SciLab [85, 86], niektóre napisane są w języku C++. Przed przystąpieniem do głównych symulacji zarówno programy symulacyjne, jak i dopasowujące poddane zostały optymalizacji i walidacji.
3. Analiza statystyczna otrzymanych wyników. W badaniach posługiwano się metodą Monte Carlo w celu zbadania zachowania i właściwości statystycznych, jakimi charakteryzują się parametry, odzyskane przy wykorzystaniu poszczególnych metod analizy danych. Powtarzano każdy eksperyment wielokrotnie i obliczano wartości średnie, odchylenia standardowe, mediany, kwartyle oraz rysowano histogramy i wykresy pudełkowe otrzymanych z dopasowań wartości. Analiza statystyczna ma odpowiedzieć na pytanie, czy otrzymywane wyniki dopasowań są obciążone i jaki jest ich rozkład statystyczny. Dla większej przejrzystości i czytelności w niniejszej pracy skupiono się głównie na wartościach średnich i odchyleniach standardowych otrzymanych wyników.

# Rozdział 4

## Materiały i metody

### 4.1 Symulacja danych

---

W niniejszym rozdziale przedstawiony zostanie algorytm MMMPP (ang. *Marked Markov Modulated Poisson Processes*) znakowane Markowsko modulowane procesy Poissona, który posłużył do symulacji trajektorii czasów międzydetekcyjnych. Trajektorie te były następnie wykorzystane w celu odtworzenia parametrów modelu.

---

#### 4.1.1 Teoria

Najprostszym modelem kinetycznym jest model 2-stanowy:



gdzie  $k_{12}$  i  $k_{21}$ , to stałe kinetyczne przejść między stanami 1 i 2. Intensywności fluorescencji w stanach 1 i 2 oznaczane będą odpowiednio jako  $I_1$  oraz  $I_2$ .

Symulacje danych przeprowadzano z wykorzystaniem MMMPP (ang. *Marked Markov Modulated Poisson Processes*), znakowanych Markowsko modulowanych procesów Poissona [24, 25, 87]. Modele Markova początkowo stosowano w automatycznym rozpoznawaniu mowy, w późniejszych latach wykorzystano je także w innych dziedzinach [88, 89]. Stosując teorię łańcuchów Markova do zjawiska zmian

konformacyjnych w cząsteczkach zakłada się, że makromolekuła biologiczna może znajdować się w jednym z  $m$  stanów, a liczba stanów jest skończona i policzalna.

Dynamika zmian między stanami opisana jest przez macierz  $\mathbf{Q}$  (nazywaną także generatorem infinitezymalnym), zawierającą elementy  $k_{ij}$ . Jeśli cząsteczka w czasie  $t$  znajduje się w stanie  $i$ , to może ona w przedziale czasu  $(t, t + dt)$  "przeskoczyć" do stanu  $j \neq i$  z prawdopodobieństwem równym  $k_{ij}dt$ , lub też może pozostać w stanie  $i$  z prawdopodobieństwem  $1 - \sum_{j \neq i} k_{ij}dt$ . Nediagonalne elementy macierzy  $\mathbf{Q}$  są dodatnie i przedstawiają stałe przejść między poszczególnymi stanami. Natomiast na diagonalnej  $k_{ii}$  znajduje się ujemna wartość sumy całego wiersza ( $k_{ii} = -\sum_{j \neq i} k_{ij}$ ).  $\sum_{j \neq i} k_{ij}$  wyraża stałą ucieczki ze stanu  $i$ , a  $1/\sum_{j \neq i} k_{ij}$  jest średnim czasem przebywania w danym stanie. Na rys. 4.1 przedstawiono schematycznie w jaki sposób tworzy się macierz kinetyczną na przykładzie modelu 2-stanowego.

$$\begin{array}{c|cc} \text{stan} & 1 & 2 \\ \hline 1 & - & k_{12} \\ 2 & k_{21} & - \end{array} \Rightarrow \mathbf{Q} = \begin{bmatrix} -k_{12} & k_{12} \\ k_{21} & -k_{21} \end{bmatrix}$$

**Rysunek 4.1:** Przykład tworzenia macierzy kinetycznej dla modelu 2-stanowego.

W eksperymentach spektroskopii fluorescencyjnej informacja o tym, w którym ze stanów układ się w danym momencie znajduje zawarta jest w postaci rejestrowanych fotonów. Stałe szybkości detekcji fotonów w każdym ze stanów przedstawiane są w postaci macierzy  $\mathbf{\Lambda}$ , w której na diagonalnej znajdują się intensywności fluorescencji  $I_i$  dla stanów  $i = 1, \dots, m$ ; a wszystkie nediagonalne elementy, to zera. Kiedy badana cząsteczka znajduje się w stanie  $i$  w momencie  $(t, t + dt)$ , kolejne fotony generowane są z prawdopodobieństwem równym  $I_i dt$ . Czasy detekcji kolejnych fotonów  $t_0, t_1, \dots, t_n$  tworzą modulowane Markowsko Procesy Poissona (MMPP), które są wykorzystywane do modelowania trajektorii zliczeń fotonów dla pojedynczych, unieruchomionych cząsteczek.

Wykorzystując MMPP możliwe jest także generowanie trajektorii czasów międzydetekcyjnych dla dwóch kolorów barwników. W takim przypadku mówi się o MMMPP (ang. *Marked Markov Modulated Poisson Processes*), znakowanych Markowsko modulowanych procesów Poissona. Zakłada się wtedy, że informacja o danym stanie zawarta jest w dwóch kanałach detekcji fluorescencji: kanale donora (D) i akceptora (A). Otrzymuje się wtedy nową informację o stanie układu, a jest nią kolor wyemitowanego fotonu. Zdefiniować więc można dwie macierze intensywności:  $\mathbf{\Lambda}^D$  dla donora i  $\mathbf{\Lambda}^A$  dla akceptora. Model kinetyczny można zapisać tak jak

przedstawiono na równaniu 4.2. Model 2-stanowy jest wtedy opisany 6 parametrami: dwoma stałymi szybkości i 4 intensywnościami fluorescencji.



gdzie  $I_1^D$  oraz  $I_1^A$  oznaczają intensywności fluorescencji donora i akceptora w stanie pierwszym, a  $I_2^D$  oraz  $I_2^A$  intensywności donora i akceptora w stanie drugim.

Przykład macierzy  $\Lambda^D$  oraz  $\Lambda^A$  dla modelu 2-stanowego przedstawiono na rys. 4.2.

$$\Lambda^D = \begin{bmatrix} I_1^D & 0 \\ 0 & I_2^D \end{bmatrix} \quad \Lambda^A = \begin{bmatrix} I_1^A & 0 \\ 0 & I_2^A \end{bmatrix}$$

**Rysunek 4.2:** Przykład macierzy intensywności  $\Lambda$  dla modelu 2-stanowego. Na diagonalnej znajdują się intensywności fluorescencji poszczególnych barwników w kolejnych stanach.

Dla eksperymentów dwukolorowych prawdopodobieństwo wyemitowania fotonu jest równe  $(I_i^D + I_i^A)dt$ . Prawdopodobieństwo zarejestrowania wyemitowanego fotonu w kanale donora wynosi  $P_i^D = I_i^D / (I_i^D + I_i^A)$ , natomiast prawdopodobieństwo zarejestrowania fotonu w kanale akceptora jest równe  $P_i^A = 1 - P_i^D$ . Każde pojawienie się fotonu opisane jest dwoma parametrami: czasem detekcji  $t_i$  oraz kolorem fotonu  $c_i = D$  albo  $A$ .

### 4.1.2 Przykład symulacji

Program do symulacji MMMPP napisany został w języku C++ i nosi nazwę `gen_BMAP`. Pobiera on dane modelu i symuluje trajektorię czasów międzydetekcyjnych. Przy wywołaniu należy podać nazwę pliku zawierającego parametry modelu. Wyniki symulacji zapisywane są w pliku *SimulationOutput.txt*. Domyślna długość trajektorii, to  $10^5$  fotonów. Istnieje także możliwość podania długości trajektorii w momencie wywołania programu. Poniżej przedstawiono w jaki sposób przebiega generacja  $10^4$  fotonów dla eksperymentu 2-kolorowego, 2-stanowego (4.3). Jako jednostkę czasu przyjęto ms.



W powyższym schemacie u góry podano intensywność fluorescencji donora, a na dole intensywność fluorescencji akceptora w danym stanie. I tak w stanie 1

intensywność donora  $I_1^D = 10 \text{ ms}^{-1}$ , akceptora  $I_1^A = 40 \text{ ms}^{-1}$ , natomiast w stanie 2 intensywności donora i akceptora wynoszą odpowiednio:  $I_1^D = 40 \text{ ms}^{-1}$ , akceptora  $I_1^A = 10 \text{ ms}^{-1}$ , a  $k_{12} = k_{21} = 1 \text{ ms}^{-1}$ . W tym przypadku macierz kinetyczna i macierze intensywności wyglądają następująco:

$$\mathbf{Q} = \begin{bmatrix} -1 & 1 \\ 1 & -1 \end{bmatrix}, \quad \mathbf{\Lambda}^D = \begin{bmatrix} 10 & 0 \\ 0 & 40 \end{bmatrix}, \quad \mathbf{\Lambda}^A = \begin{bmatrix} 40 & 0 \\ 0 & 10 \end{bmatrix}.$$

Aby wykonać symulację w pierwszej kolejności należy przygotować plik *SimulationParameters.txt*:

```

1 BMAP-EM Parameter Estimation Output M = 2, N = 2
2 -----
3 -----
4 D[0][i][j] |      j=0      |      j=1      |
5 -----+-----+-----
6 i=0         |    -51.000000   |     1.000000   |
7 i=1         |     1.000000   |    -51.000000   |
8 -----
9 -----
10 D[1][i][j] |      j=0      |      j=1      |
11 -----+-----+-----
12 i=0         |    10.000000   |     0.000000   |
13 i=1         |     0.000000   |    40.000000   |
14 -----
15 -----
16 D[2][i][j] |      j=0      |      j=1      |
17 -----+-----+-----
18 i=0         |    40.000000   |     0.000000   |
19 i=1         |     0.000000   |    10.000000   |
20 -----

```

zawierający informację o liczbie stanów ( $N$ ) oraz kolorów ( $M$ ), a także macierze  $\mathbf{\Lambda}^c$ , oznaczane w programie jako macierze  $\mathbf{D}[c]$ , gdzie  $c$  oznacza kolor fotonów. Dla modelu 2-stanowego  $c = [1, 2]$ . Plik *SimulationParameters.txt* zawiera także macierz:

$$\mathbf{D}[0] = \mathbf{Q} - \sum_{c=1}^N \mathbf{D}[c] \quad (4.4)$$

Następnie wywołuje się program symulacyjny `gen_BMAP`, podając jako parametry plik zawierający macierze oraz sumaryczną liczbę fotonów, jaka ma być symulowana (parametr  $s$ ). Wywołanie programu dla tego przypadku będzie wyglądało następująco:

```
./gen_BMAP SimulationParameters.txt -s10000
```

W wyniku symulacji otrzymuje się plik tekstowy o nazwie *SimulationOutput.txt*, którego część przedstawiono poniżej:

```
1 0.000252 1
2 0.000880 1
3 0.009040 2
4 0.001136 1
5 0.002654 2
6 0.003407 1
7 0.010152 1
8 0.024843 1
9 0.003877 1
10 0.012945 1
```

Pierwsza kolumna zawiera czasy międzydetekcyjne, czyli czasy, jakie upłynęły pomiędzy detekcją poszczególnych fotonów w eksperymencie komputerowym, druga natomiast zawiera informację o kolorze danego fotonu, czy jest to foton donora (1), czy akceptora (2).

## 4.2 Estymacja parametrów kinetycznych

---

W tym rozdziale przedstawiono, w jaki sposób przebiega estymacja parametrów kinetycznych z wykorzystaniem ML i analizy on-off.

---

### 4.2.1 Metoda największej wiarygodności (ML)

#### 4.2.1.1 Teoria

W analizie metodą największej wiarygodności (ML) korzystano bezpośrednio z trajektorii czasów międzydetekcyjnych i numerycznie maksymalizowano funkcję wiarygodności (4.5). Sposób numerycznej maksymalizacji funkcji wiarygodności prezentowany był w kilku publikacjach [24, 90, 91]. W niniejszej pracy maksymalizację przeprowadzano przy użyciu pakietu `ip2bmap`, który został napisany przez Lindemanna [90] do modelowania ruchu pakietów internetowych. Dla wybranych zestawów parametrów wyniki otrzymane przy wykorzystaniu oprogramowania opartego na `ip2bmap` porównywano z dopasowaniami wykonanymi przy użyciu pakietu `HiddenMarkov` w języku R, napisanym przez Harte i dostępnym na <http://cran.at.r-project.org/web/packages/HiddenMarkov>. Wyniki były porównywalne. Program został przystosowany do rozpatrywanego tutaj problemu i zoptymalizowany.

Oznaczmy  $X(t)$  jako stan układu, który determinuje częstość pojawiania się fotonów w czasie  $t$  i  $X_k$ ,  $k \geq 0$ , oznaczmy jako stan układu bezpośrednio po rejestracji  $k$ -tego fotonu  $X_k = X(t_k)$ . Wtedy  $X_k = i$ , jeżeli zliczenie nastąpiło, kiedy układ znajdował się w  $i$ -tym stanie. Niech  $Y_k$ ,  $k \geq 0$  będzie czasem pomiędzy pojawieniem się  $(k - 1)$ -szego i  $k$ -fotonu,  $Y_k = t_k - t_{k-1}$ ,  $Y_0 = 0$ . Kluczowym spostrzeżeniem, z którego korzysta się, aby wyprowadzić wzór na wiarygodność zaobserwowanej sekwencji czasów detekcji fotonów jest fakt, że funkcja dwóch zmiennych  $\{(X_k, Y_k); k \geq 0\}$  jest odnawialnym procesem Markova.

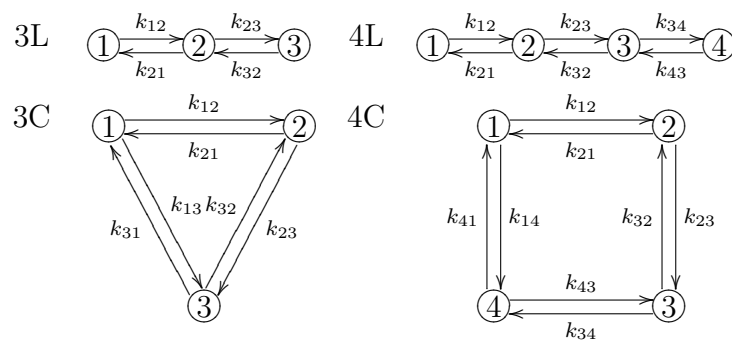
Warto tutaj zwrócić uwagę, że stany  $X_k$  w momencie  $t_k$  (momencie pojawienia się  $k$ -tego fotonu) nie są znane. Funkcja wiarygodności dla trajektorii zliczeń fotonów charakteryzujących się różnymi kolorami  $(t_0, c_0), (t_1, c_1), \dots, (t_n, c_n)$  zapisana jest wzorem:

$$L = \mathbf{P}_{\text{ini}} \prod_{k=1}^n \{\exp[(\mathbf{Q} - \mathbf{\Lambda})\Delta t_k] \mathbf{\Lambda}^{c_k}\} \mathbf{1} \quad (4.5)$$

gdzie  $\mathbf{P}_{\text{ini}} = [P_1^{\text{ini}}, P_2^{\text{ini}}, \dots, P_n^{\text{ini}}]$  jest wektorem liniowym opisującym początkowy rozkład łańcucha Markova, a  $\mathbf{1}$  jest wektorem jedyńkowym  $\mathbf{1} = [1, 1, \dots, 1]^T$ . Prawdopodobieństwa znalezienia się układu w danym stanie w czasie  $t_0$  zależą od wartości  $\mathbf{P}_{\text{ini}}$ , czynnik  $\exp[(\mathbf{Q} - \mathbf{\Lambda})\Delta t_k]$  determinuje prawdopodobieństwa przejścia pomiędzy stanami, bez rejestracji fotonów w czasie  $\Delta t_k$ . Prawdopodobieństwo detekcji fotonu w czasie  $t_k$  przedstawia macierz  $\mathbf{\Lambda}$ . Pomnożenie całości przez macierz jedyńkową  $\mathbf{1}$  jest sumą wszystkich opisanych wcześniej czynników na konkretne stany w czasie  $t_k$ . Analogiczne wyrażenie było stosowane przez Kou [92] dla trajektorii opóźnienia fotonów pojedynczej cząsteczki oraz Gopich i Szabo [22] dla dwukolorowych trajektorii fotonów.

Opisaną powyżej procedurę można z powodzeniem stosować w przypadku modelu 2-stanowego oraz modeli liniowych, ponieważ można w ich przypadku jednoznacznie określić prawdopodobieństwo przebywania w danym stanie bezpośrednio ze stałych szybkości zmian pomiędzy stanami.

W niniejszej pracy badano także możliwość wykorzystania metody ML do estymacji parametrów kinetycznych modeli 3- i 4- stanowych, zarówno liniowych (L), jak i cyklicznych (C). Na rys. 4.3 przedstawiono modele kinetyczne rozpatrywane w rozdziałach 5.2, 5.3 i 5.4.



**Rysunek 4.3:** Schematy 3- i 4- stanowych modeli kinetycznych, które są rozpatrywane w niniejszej pracy. Literą “L” oznaczane będą modele liniowe, natomiast litera “C” modele cykliczne.

Dla modelu 3L stosunek prawdopodobieństw przebywania w stanach, które mogą po sobie następować wygląda tak, jak przedstawiono w dwóch pierwszych równaniach we wzorze 4.6, dodatkowo założyć należy, że układ w danym momencie znajduje się w którymś z trzech stanów, stąd dodać należy równanie 3 we wzorze 4.6.

$$\left\{ \begin{array}{l} \frac{P_1}{P_2} = \frac{k_{21}}{k_{12}} \\ \frac{P_2}{P_3} = \frac{k_{32}}{k_{23}} \\ P_1 + P_2 + P_3 = 1 \end{array} \right. \quad (4.6)$$

Rozwiązując powyższe równanie otrzymuje się prawdopodobieństwa przebywania w danym stanie, wyrażone przy pomocy stałych szybkości:

$$\begin{aligned} P_1 &= \frac{k_{21}k_{32}}{k_{21}k_{32} + k_{23}k_{12} + k_{12}k_{32}} \\ P_2 &= \frac{k_{12}k_{32}}{k_{21}k_{32} + k_{23}k_{12} + k_{12}k_{32}} \\ P_3 &= \frac{k_{12}k_{23}}{k_{21}k_{32} + k_{23}k_{12} + k_{12}k_{32}} \end{aligned} \quad (4.7)$$

Natomiast dla modelu 3C należy wziąć pod uwagę jeszcze jedną zależność: przejścia pomiędzy stanami 1 i 3 (układ równań 4.8)

$$\left\{ \begin{array}{l} \frac{P_1}{P_2} = \frac{k_{21}}{k_{32}} \\ \frac{P_2}{P_3} = \frac{k_{23}}{k_{31}} \\ \frac{P_3}{P_1} = \frac{k_{13}}{k_{12}} \\ P_1 + P_2 + P_3 = 1 \end{array} \right. \quad (4.8)$$

W celu rozwiązania powyższego równania konieczne jest uwzględnienie warunku równowagi szczegółowej [93] w postaci:

$$k_{12}k_{23}k_{31} = k_{21}k_{13}k_{32}$$

Przyjmując, że

$$k_{32} = \frac{k_{12}k_{23}k_{31}}{k_{21}k_{13}}$$

otrzymano rozwiązanie:

$$\begin{aligned} P_1 &= \frac{k_{21}k_{31}}{k_{21}k_{31} + k_{12}k_{31} + k_{21}k_{13}} \\ P_2 &= \frac{k_{12}k_{31}}{k_{21}k_{31} + k_{12}k_{31} + k_{21}k_{13}} \\ P_3 &= \frac{k_{21}k_{13}}{k_{21}k_{31} + k_{12}k_{31} + k_{21}k_{13}} \end{aligned} \quad (4.9)$$

Stosując warunek równowagi szczegółowej zakłada się, że w stanie równowagi  $P_i^{\text{eq}}k_{ij} = P_j^{\text{eq}}k_{ji}$ , przy czym  $\mathbf{P}^{\text{eq}}$  jest rozkładem równowagi stanów takich że  $\mathbf{P}^{\text{eq}}\mathbf{Q} = 0$ . Warunek równowagi szczegółowej [94] może być przedstawiony wzorem:

$$\text{diag}(\mathbf{P}^{\text{eq}})\mathbf{Q} = [\text{diag}(\mathbf{P}^{\text{eq}})\mathbf{Q}]^T \quad (4.10)$$

gdzie  $\text{diag}(\mathbf{P}^{\text{eq}})$  jest diagonalną macierzy zawierającą elementy  $P_i^{\text{eq}}$ , a  $\mathbf{A}^T$  to transponowana macierz  $\mathbf{A}$ . W standardowym algorytmie EM warunek równowagi szczegółowej nie jest wzięty pod uwagę. W celu jego zaimplementowania wprowadzono następujące poprawki w każdym kroku iteracyjnym: po pierwsze, estymaty  $\mathbf{Q}$  i  $\mathbf{P}^{\text{eq}}$  otrzymywane są według algorytmu Rydena [24]. Następnie obliczana jest macierz

$$\mathbf{Q}_s = \{\text{diag}(\mathbf{P}^{\text{eq}})\mathbf{Q} + [\text{diag}(\mathbf{P}^{\text{eq}})\mathbf{Q}]^T\}/2 \quad (4.11)$$

Macierz stałych przejść, z uwzględnieniem warunku równowagi szczegółowej otrzymuje się w następujący sposób:

$$\mathbf{Q} = [\text{diag}(\mathbf{P}^{\text{eq}})]^{-1}\mathbf{Q}_s \quad (4.12)$$

gdzie  $\mathbf{A}^{-1}$ , to macierz odwrotna do macierzy  $\mathbf{A}$ .

Powyższy zabieg wymusza otrzymanie modelu zachowującego warunek równowagi szczegółowej po dopasowaniu i przybliża go do właściwego minimum, algorytm dopasowujący te właśnie parametry bierze jako parametry startowe do kolejnego kroku iteracyjnego.

#### 4.2.1.2 Przykład

W analizie ML jako dane do dopasowania wykorzystuje się bezpośrednio plik otrzymany z symulacji metodą MMMPP. Program nosi nazwę `em_BMAP` i napisany jest w języku C++. Program pobiera dane w postaci pliku *SimulationOutput.txt*, otrzymanego w wyniku symulacji oraz plik z parametrami startowymi dopasowania *InputParameters.txt*, należy także podać liczbę stanów w modelu oraz ile kolorów barwników było symulowanych. W przypadku przedstawionych tu badań, w większości przypadków było to 2 (donor i akceptor). Można także zdefiniować warunek zakończenia dopasowania oraz maksymalną liczbę iteracji, jaka ma być wykonana.

Przykładowo dla dwukolorowego eksperymentu, dla modelu 3-stanowego i maksymalnej liczby iteracji równej 5000 program należy wywołać poleceniem:

```
./em_BMAP -N3 -M2 -n5000 SimulationOutput.txt
```

Samo dopasowanie polega na wykorzystaniu rachunku macierzowego. Program korzysta z macierzy intensywności oraz stałych przejść:  $\mathbf{D}[m][i][j]$ , gdzie  $m =$

$1, \dots, M$  oznacza macierze zawierające intensywności dla konkretnych stanów,  $\mathbf{D}[0][i][j] = \mathbf{Q} - \sum_{m=1}^m \mathbf{\Lambda}^m$ . W każdej iteracji obliczana jest wartość funkcji wiarygodności, a zakończenie działania algorytmu dopasowującego następuje, gdy zmiany każdego elementu w konkretnych elementach macierzy  $\mathbf{D}$  pomiędzy iteracją  $u - 1$  i  $u$  są mniejsze od zadanego parametru  $\epsilon$  (dla przedstawianych tu wyników  $\epsilon = 10^{-4}$ ). W kodzie programu wygląda to następująco:

```

1 for(UINT m = 0; m <= _MaxBatchSize; m++)
2 {
3     for(UINT i~= 0; i~< _StateNo; i++)
4     {
5         for(UINT j = 0; j < _StateNo; j++)
6         {
7             if(abs(DOld[m][i][j]) > 0.0)
8             {
9                 RelDiff = abs((DOld[m][i][j] - DNew[m][i][j])) / abs(DOld[m][i][j]);
10                MaxRelChange = max(MaxRelChange, RelDiff);
11            }
12            AbsDiff = abs((DOld[m][i][j] - DNew[m][i][j]));
13            MaxAbsChange = max(MaxAbsChange, AbsDiff);
14            if(abs((DOld[m][i][j] - DNew[m][i][j])) > _Epsilon) {IsConv = false;}
15        }
16    }
17 }

```

`_MaxBatchSize`, to liczba kolorów fotonów fluorescencji, `DOld[m][i][j]` - stan macierzy  $\mathbf{D}$  po  $u - 1$  iteracji, `DNew[m][i][j]` - stan macierzy  $\mathbf{D}$  po  $u$ -tej iteracji. W linii 14 program sprawdza, czy został osiągnięty warunek zakończenia dopasowania (różnica pomiędzy parametrami macierzy  $\mathbf{D}$  jest mniejsza od wartości progowej  $\epsilon$ ). Jeżeli nie, dopasowanie trwa dalej. Jeśli natomiast warunek zostanie spełniony, dopasowanie kończy się i wyniki zapisywane są do plików. Program zapisuje otrzymane w wyniku dopasowania parametry: intensywności (diagonalne macierzy  $\mathbf{D}[m][i][j]$  dla  $m \neq 0$ ) oraz stałe przejść (niediagonalne elementy macierzy  $\mathbf{D}[0][i][j]$ ). Zapisywana jest także liczba iteracji, po której zakończyło się dopasowanie oraz wartość funkcji wiarygodności dla otrzymanych parametrów. Poniżej przedstawiono plik wynikowy dla czterech dopasowań intensywności akceptora dla modelu 2-stanowego (4.3)

1	38.803981	9.865109	7	23493.8662
2	40.118872	10.112122	10	23520.0702
3	41.293537	10.199432	10	23849.3602
4	39.495205	10.321648	9	23646.3362

W pierwszej i drugiej kolumnie znajdują się wyniki dopasowań intensywności akceptora dla stanu pierwszego i drugiego, w trzeciej kolumnie znajduje się informacja o liczbie iteracji, po których algorytm dopasowujący zbiegł się, czwarta kolumna przedstawia wartości funkcji wiarygodności dla poszczególnych dopasowań.

## 4.2.2 Analiza on-off

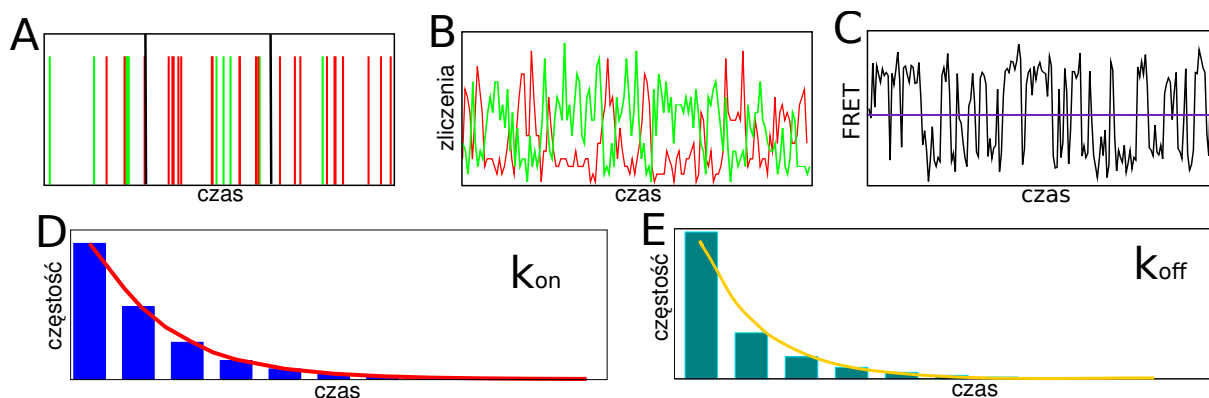
### 4.2.2.1 Teoria

W pierwszym etapie analizy on-off konieczne jest obliczenie trajektorii fluorescencji i trajektorii FRET. W tym celu trajektorię czasów detekcji, uzyskaną w wyniku symulacji dzielono na przedziały o wybranej długości. Przedział taki nazywać będziemy oknem czasowym ( $h$ ). Następnie zliczano ile fotonów danego koloru zarejestrowano w kolejnych przedziałach. Okno czasowe musi być na tyle małe, by intensywności fluorescencji w różnych stanach nie uśredniały się, ale jednocześnie na tyle szerokie, by obserwować kilka fotonów każdego z barwników w danym oknie. Po dokonaniu wyboru szerokości okna czasowego i obliczeniu trajektorii fluorescencji donora i akceptora obliczano trajektorię FRET. Następnie wybierano wartość progową  $E_p = (E_{\text{on}} + E_{\text{off}})/2$ , rozdzielająca trajektorię FRET na okresy o wysokiej (on) wydajności FRET ( $E \geq E_p$ ) i okresy o niskiej (off) wydajności FRET ( $E < E_p$ ). Kolejne czasy przebywania w stanach on i off są zmiennymi losowymi o rozkładach zależnych od mechanizmu zmian konformacyjnych. W celu wyznaczenia stałych szybkości przejść, do histogramu czasów on i off dopasowywano krzywe wynikające z modelu zmian konformacyjnych, w najprostszym przypadku jest to krzywa wykładnicza ( $Ae^{-t}$ ). Otrzymuje się w ten sposób stałe kinetyczne ucieczki ze stanów on i off ( $k_{\text{on}}$  i  $k_{\text{off}}$ ), będące odwrotnościami średnich czasów przebywania w każdym ze stanów ( $\tau_{\text{on}}$  i  $\tau_{\text{off}}$ ). Wariantem tej metody jest dopasowywanie modelu do kumulatywnych histogramów czasów on i off [13, 14]. W przypadku, gdy mierzone intensywności charakteryzują się dużymi odchyleniami i wartości intensywności w różnych stanach nakładają się na siebie, możliwe jest przeprowadzenie wygładzania trajektorii [13, 95].

Obliczano dwie trajektorie fluorescencji, a z nich trajektorię FRET, a następnie przeprowadzano analizę. Możliwe było także przeprowadzenia analizy globalnej wykorzystując trajektorię donora, akceptora i FRET. Wiadomo bowiem, że  $k_{\text{off}}^{\text{D}} = k_{\text{on}}^{\text{A}} = k_{\text{on}}^{\text{FRET}}$ , a  $k_{\text{on}}^{\text{D}} = k_{\text{off}}^{\text{A}} = k_{\text{off}}^{\text{FRET}}$ .

### 4.2.2.2 Przykład

Wykonano symulację trajektorii czasów międzydetekcyjnych dla modelu przedstawionego w równaniu 4.3, dla  $n = 5 \times 10^4$  fotonów. Dopasowania metodą on-off wykonywano przy pomocy skryptu pakietu SciLab [85, 86]. Skrypt o nazwie `on_off_analysis.sce` jako dane wejściowe pobiera plik *SimulationOutput.txt* z trajektorią czasów międzydetekcyjnych (rys. 4.4A) oraz szerokość okna czasowego  $h$ . Następnie oblicza na podstawie danych trajektorie fluorescencji donora i akceptora (rys. 4.4B). Dla omawianego przypadku  $h = 0,22$ . Następnie, przy użyciu podanej wartości progowej (w tym przypadku było to 0,5) trajektoria FRET była dzielona na stany on i off (rys. 4.4C). Skrypt posiada także możliwość automatycznego obliczenia optymalnej wartości progowej.



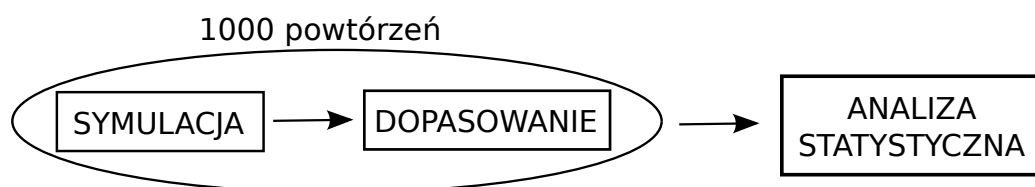
**Rysunek 4.4:** Przykład analizy on-off. A) Trajektoria czasów międzydetekcyjnych, otrzymana w wyniku symulacji, i podzielona na okna czasowe, według których oblicza się trajektorie fluorescencji B). C) trajektoria FRET obliczana na podstawie trajektorii fluorescencji, z zaznaczonym progiem rozdzielającym stany o różnej wartości FRET (linia fioletowa). D), E) histogramy czasów on i off z dopasowanymi krzywymi wykładniczymi.

Obliczano histogramy czasów przebywania w stanach on i off i dopasowywano krzywe wykładnicze (rys. 4.4D, E), dzięki czemu odzyskiwane są stałe kinetyczne ucieczki z konkretnych stanów. Wyniki dopasowań zapisywane były do plików.

## 4.3 Statystyczna analiza danych

W tym rozdziale przedstawiona będzie metodologia eksperymentów komputerowych, które zostały przeprowadzone w celu zbadania właściwości statystycznych parametrów estymowanych przy użyciu metody największej wiarygodności (ML) oraz porównaniu jej z analizą on-off. Przedstawione zostanie w jaki sposób w następnych rozdziałach badane będą właściwości statystyczne omawianych metod analizy danych. Opisany także będzie sposób prezentacji wyników dopasowań używanych w następnych rozdziałach.

Symulowano trajektorie czasów międzydetekcyjnych wykorzystując MMMPP dla modeli z różną liczbą stanów  $m$ . Uzyskane w wyniku symulacji trajektorie wykorzystywano jako dane do analizy w celu odzyskania parametrów modelu. W pierwszym etapie eksperymentów jako parametry startowe do dopasowań stosowano parametry użyte w symulacji. Każdy eksperyment komputerowy powtarzany był  $10^3$  razy, a dopasowane parametry ( $\hat{k}_{ij}$ ,  $\hat{I}_i$ ) były zapisywane. Schemat postępowania w trakcie eksperymentów przedstawiono na rys. 4.5

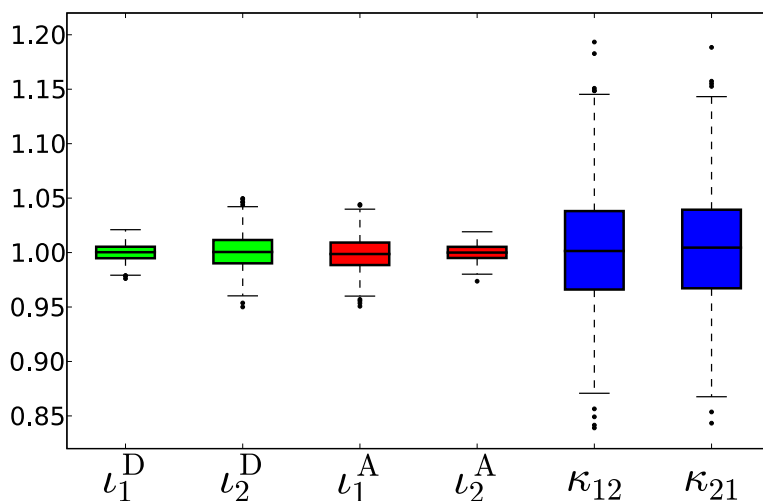
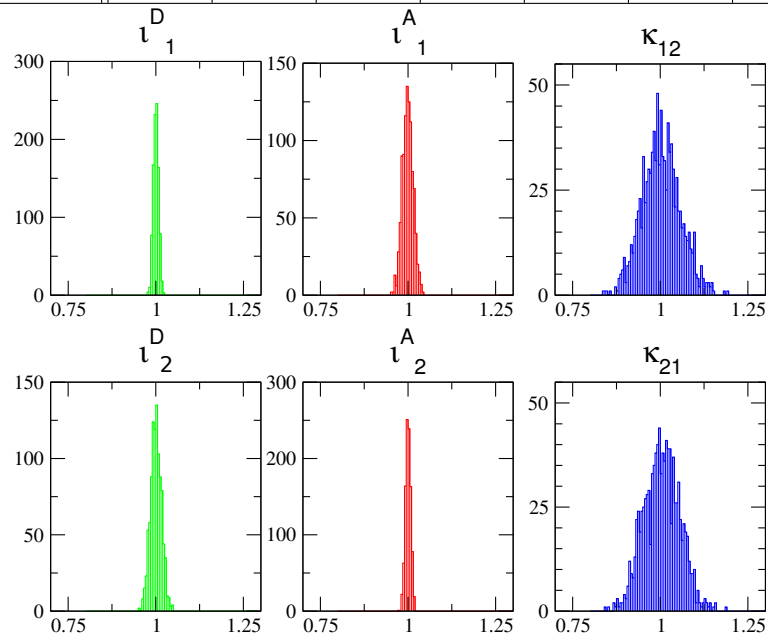


**Rysunek 4.5:** Schemat przedstawiający sposób prowadzenia eksperymentów komputerowych, których wyniki przedstawiono w niniejszej pracy.

W celu ujednoczenia sposobu prezentacji, estymowane parametry poddawane były normalizacji przez podzielenie ich przez wartości użyte w symulacji. Skalowane parametry oznaczane będą jako  $\kappa_{ij} = \hat{k}_{ij}/k_{ij}$ ,  $\iota_i = \hat{I}_i/I_i$ . Zaznaczyć także należy, iż zarówno stałe kinetyczne jak i intensywności mają jednostki odwrotności czasu. Dzięki skalowaniu przedstawiane w niniejszej pracy parametry są bezwymiarowe, co czyni wyniki dopasowań bardziej ogólnymi.

Analiza statystyczna dokonywana była poprzez obliczenie wartości średnich ( $\mu$ ), odchyłeń standardowych ( $\sigma$ ), minimalnych i maksymalnych wartości otrzymanych

parametr	$\mu$	$\sigma$	min	max	Qu1	med	Qu3
$l_1^D$	1,000	0,007	0,976	1,021	0,995	1,000	1,005
$l_1^A$	1,001	0,016	0,950	1,050	0,990	1,000	1,012
$l_2^D$	0,999	0,015	0,951	1,044	0,989	0,999	1,009
$l_2^A$	1,000	0,007	0,974	1,019	0,995	1,000	1,005
$\kappa_{12}$	1,003	0,055	0,839	1,193	0,966	1,001	1,038
$\kappa_{21}$	1,004	0,051	0,843	1,188	0,967	1,005	1,039



**Rysunek 4.6:** Przykład analizy statystycznej dla  $10^3$  dopasowań metodą ML dla modelu 4.3,  $n = 5 \times 10^4$  fotonów. U góry: tabela statystyk otrzymanych estymat, pośrodku: histogramy wartości dopasowanych, na dole: wykres pudełkowy otrzymanych dopasowań. We wszystkich przypadkach przedstawiono wartości znormalizowane.

estymat, a także median i kwartyli. W celu graficznego przedstawienia rozrzutu wyników wykonywano histogramy otrzymanych parametrów oraz wykresy pudełkowe. Przykładową statystykę dla dopasowania metodą ML do modelu 4.3, dla trajektorii o długości  $n = 5 \times 10^4$  fotonów przedstawiono na rys. 4.6.

W niniejszym opracowaniu, w celu uzyskania większej klarowności i przejrzystości prezentacji, oparto się głównie na wartościach średnich oraz odchyleniach standardowych.

Analogicznie przeprowadzano analizę danych dla modeli 3- i 4- stanowych. Ponieważ jako parametry startowe dopasowania podawano parametry użyte w symulacji, należy zakładać, że otrzymane w tej części eksperymentów wyniki są najlepszymi jakie jest możliwość otrzymać.

W następnym etapie badań sprawdzano, jak na jakość estymat wpływa wartość parametrów startowych podanych do algorytmu dopasowującego. Przeprowadzano symulację wybranego modelu, o określonych parametrach, następnie wielokrotnie uruchamiano dopasowanie do tych samych danych, ale z różnymi parametrami startowymi, coraz bardziej oddalonymi od tych stosowanych do symulacji trajektorii czasów międzydetekcyjnych.

## 4.4 Rozróżnianie modeli kinetycznych

---

W tym rozdziale omówiono w jaki sposób rozróżniano modele kinetyczne. Przedstawiona będzie metodologia eksperymentów komputerowych oraz kryteria AIC oraz BIC, które posłużyły do wyboru najbardziej prawdopodobnego spośród proponowanych modelu kinetycznego. Przedstawione także zostanie, w jaki sposób w rozdz. 5.4 prezentowane będą wyniki dotyczące rozróżniania modeli kinetycznych.

---

### 4.4.1 Teoria

Na tym etapie badań do danych dopasowywane były modele inne, niż te użyte w symulacji. Rozróżnianie modeli możliwe było tylko dla analizy metodą ML. Jako wynik dopasowania metodą ML otrzymywano pliki z informacją o dopasowanych intensywnościach i stałych kinetycznych, liczbę iteracji, po których algorytm zbiegł się, a także końcową wartością funkcji wiarygodności, będącą oszacowaniem jakości dopasowania parametrów do danych eksperymentalnych. Dzięki temu, biorąc pod uwagę wartość funkcji wiarygodności z dopasowań różnych modeli do konkretnych danych eksperymentalnych, można określić najbardziej prawdopodobny spośród proponowanych model układu. W miarę zwiększania się liczby stanów w dopasowywanym modelu, a co za tym idzie także stopni swobody układu dopasowywanego funkcja wiarygodności wykazuje tendencję, by przyjmować większą wartość dla modeli bardziej skomplikowanych, co może skutkować wyborem niepoprawnego, bardziej skomplikowanego modelu. Z tego powodu niemożliwe jest określenie, który model jest najbardziej prawdopodobny tylko na podstawie wartości funkcji wiarygodności. Rozwiązanie tego problemu można znaleźć biorąc pod uwagę oprócz funkcji wiarygodności także liczbę dopasowywanych parametrów. Akaike zaproponował następującą modyfikację wartości funkcji wiarygodności w celu wyeliminowania wpływu liczby parametrów na ocenę dobroci dopasowania [29]. Kryterium informacyjne Akaikego, AIC (ang. *Akaike Information Criterion*) można zapisać wzorem:

$$AIC = -2 \ln LL + 2p \quad (4.13)$$

gdzie:

$LL$  - wartość funkcji wiarygodności dla dopasowanego modelu,  $p$  - liczba dopasowywanych parametrów.

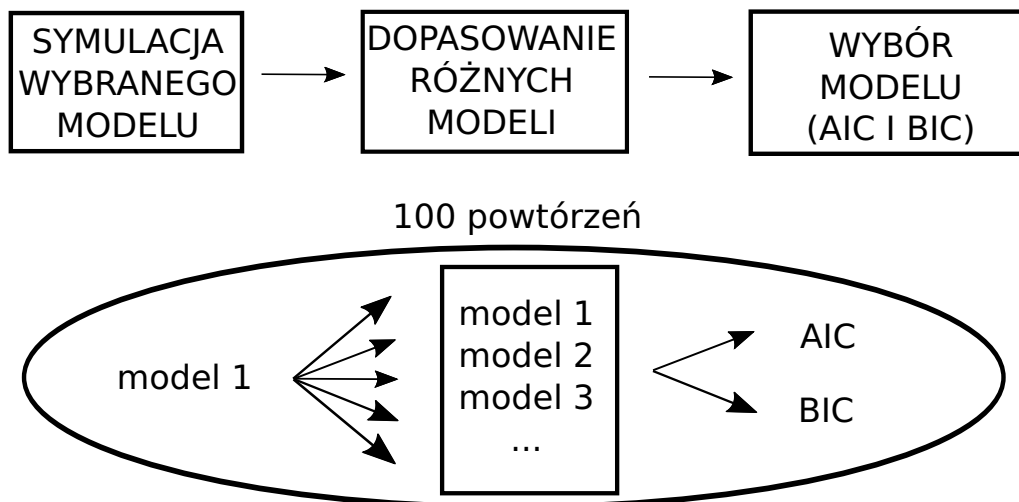
Dla danej trajektorii czasów międzydetekcyjnych lepszy jest ten model, dla którego wartość AIC, opisana wzorem 4.13 jest mniejsza. Podobne kryterium, to BIC (ang. *Bayesian Information Criterion*) [30] bierze ono pod uwagę nie tylko wartość funkcji wiarygodności oraz liczbę parametrów, ale także wielkość próby, czyli w tym wypadku liczbę fotonów w trajektorii czasów międzydetekcyjnych  $n$  (4.14).

$$\text{BIC} = -2 \ln LL + p \ln n \quad (4.14)$$

Kryteria AIC i BIC były wykorzystane w rozróżnianiu modeli kinetycznych w pracach [96] oraz [97].

W niniejszej pracy eksperyment rozróżniania modeli dla danego zestawu parametrów powtarzany był  $10^2$  razy, a otrzymane wyniki były zapisywane.

Schemat, według którego przeprowadzono symulacje przedstawiono na rys. 4.7.



**Rysunek 4.7:** Schemat przedstawiający sposób, w jaki postępowano w eksperymentach rozróżniania modeli kinetycznych

### 4.4.2 Przykład

Symulowano model 3L (4.15)

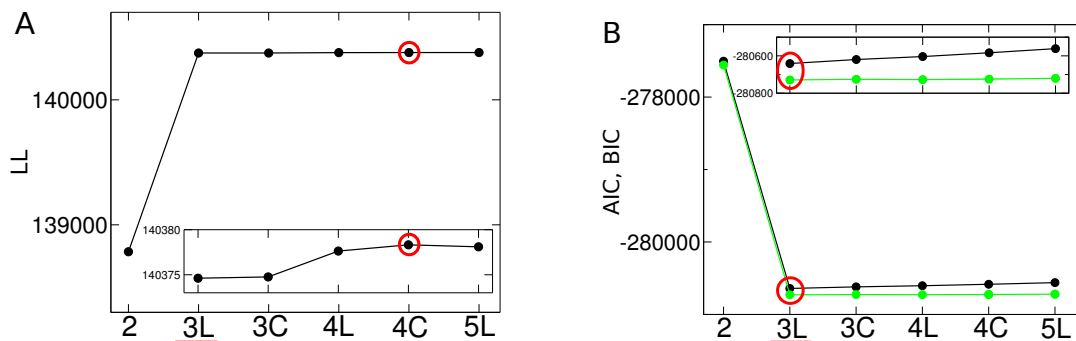


Do otrzymanej trajektorii czasów międzydetekcyjnych dopasowywano modele 2, 3L, 3C, 4L, 4C oraz 5L. Wyniki przedstawiono w tab. 4.1 i na rys. 4.8. Wartość funkcji wiarygodności przy dopasowaniu do modelu 2-stanowego jest zdecydowanie mniejsza od tych otrzymanych dla pozostałych modeli. Można więc wykluczyć model 2-stanowy jako najbardziej prawdopodobny. Natomiast pozostałe wartości LL są do siebie zbliżone, a największą z nich charakteryzuje się model 4C, niewiele niższą wartością LL charakteryzuje się model 5L (tab. 4.1 i rys. 4.8A). Wynika to z faktu, że w modelach tych dopasowywanych jest najwięcej parametrów (odpowiednio 16 i 18), a przez to łatwiej jest dopasować model do danych, ponieważ dostępnych jest więcej stopni swobody. Bazowanie więc w rozróżnianiu modeli kinetycznych na wartości funkcji wiarygodności powoduje, że uprzywilejowane będą modele bardziej skomplikowane. Aby wyeliminować efekt uprzywilejowania modeli z większą liczbą parametrów, można zastosować na przykład kryteria AIC lub BIC. Jak to przedstawiono w tab. 4.1 i na rys. 4.8B, zastosowanie zarówno AIC, jak i BIC pozwoliło na wybór poprawnego modelu jako najbardziej prawdopodobny.

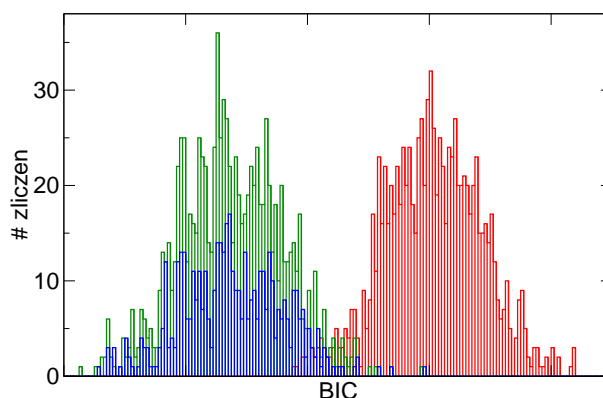
**Tabela 4.1:** Wartości funkcji wiarygodności oraz AIC i BIC, kiedy symulowano model 3L (zaznaczony na zielono) i dopasowywano modele 2, 3L, 3C, 4L, 4C, 5L. Modele wybrane jako najbardziej prawdopodobne pogrubiono.

model	l. par.	LL	AIC	BIC
2	6	138784,8	-277557,6	-277504,6
3L	10	140374,6	<b>-280729,2</b>	<b>-280641,0</b>
3C	12	140374,8	-280725,5	-280619,7
4L	14	140377,6	-280727,3	-280603,8
4C	16	<b>140378,3</b>	-280724,6	-280583,5
5L	18	140378,1	-280720,2	-280561,4

Powtórzono eksperyment rozróżniania modeli dla przedstawionego przykładu (4.15)  $10^3$  razy i na rys. 4.9 przedstawiono histogramy otrzymanych wartości BIC dla dopasowań modeli 2- 3L i 4L.



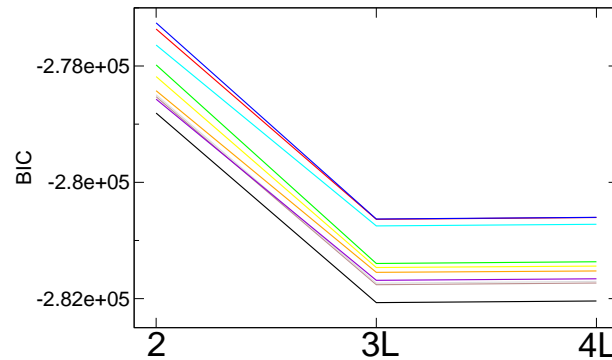
**Rysunek 4.8:** Przykład rozróżniania modeli (A) z wykorzystaniem funkcji wiarygodności oraz (B) AIC (zielony) i BIC (czarny). Symulowano model 3L i do otrzymanych danych dopasowywano modele 2, 3L, 3C, 4L, 4C oraz 5L. Podkreślono model symulowany, a kółkiem zaznaczono model wybrany jako najbardziej prawdopodobny.



**Rysunek 4.9:** Histogramy wartości BIC dla 10<sup>3</sup> eksperymentów rozróżniania modeli, gdy symulowany był model 3L, a dopasowywane były modele 2-stanowy (czerwony), 3L (zielony) oraz 4L (niebieski). Na rysunku przedstawiono wartości BIC tylko z tych eksperymentów, dla których algorytm dopasowujący zbiegł się. W przypadku dopasowania modeli 2 i 3L jest to 10<sup>3</sup> wartości, a w przypadku 4L: 475 wartości.

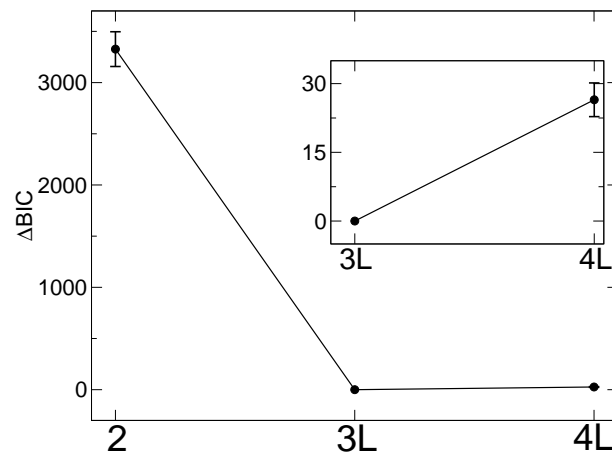
Dla dopasowania modelu 4-stanowego nie we wszystkich przypadkach algorytm dopasowujący zbiegł się po 10<sup>4</sup> iteracjach. Histogram sporządzono tylko z wartości BIC, dla których dopasowanie zakończyło się sukcesem, dlatego też histogram wartości BIC dla modelu 4-stanowego (niebieski) obliczony był dla mniejszej próby. Z histogramu na rys. 4.9 widać jednak wyraźnie, że biorąc pod uwagę rozrzut statystyczny wartości BIC z serii eksperymentów można z dość dużą dokładnością rozdzielić wartości BIC dla modelu 2-stanowego od tych dla modeli 3- i 4- stanowych. Nie ma natomiast możliwości rozróżnienia wartości BIC dla dwóch pozostałych modeli. Analogicznie wygląda to dla AIC. Niemniej jednak przyglądając się dopasowaniom różnych modeli do tych samych danych (rys. 4.10), w każdym przypadku różnice pomiędzy wartościami BIC dla poszczególnych modeli

dopasowywanych do danej trajektorii czasów międzydetekcyjnych utrzymują się na podobnym poziomie. Różnią się natomiast między sobą bezwzględnymi wartościami BIC.



**Rysunek 4.10:** Wartości BIC dla 10 eksperymentów, gdzie symulowany był model 3-stanowy liniowy i dopasowywane modele 2-, 3-, 4-stanowe liniowe.

Obliczono więc dla każdego eksperymentu wartość  $\Delta\text{BIC} = \text{BIC} - \text{BIC}_{\min}$ . W przypadku rozpatrywania  $\Delta\text{BIC}$  model wybrany ma wartość  $\Delta\text{BIC} = 0$ . Wartości  $\Delta\text{BIC}$  dla pozostałych modeli są większe. Na rys. 4.11 przedstawiono wartości średnie i odchylenia standardowe dla  $10^3$  eksperymentów rozróżniania modeli. Wartości  $\Delta\text{BIC}$  otrzymane dla dopasowania modeli 2- i 4- stanowych nie nakładają się na siebie i różnice w  $\Delta\text{BIC}$  są większe, niż rozrzut statystyczny dopasowań dla danego modelu. Można więc wykluczyć wpływ losowości na wynik rozróżniania modeli.



**Rysunek 4.11:** Wartości średnie i odchylenia standardowe  $\Delta\text{BIC}$  dla symulacji modelu 3-stanowego i dopasowania modeli 2- 3- i 4- stanowych, gdzie wszystkie dopasowania zbiegły się po mniej, niż  $10^4$  iteracji.

# Rozdział 5

## Wyniki

### 5.1 Model 2-stanowy

#### 5.1.1 Wpływ zarejestrowanej liczby fotonów

---

W tym rozdziale przedstawiono jak na jakość estymowanych parametrów dla modelu 2-stanowego wpływa liczba zarejestrowanych fotonów. Sprawdzono wpływ zarówno ogólnej liczby fotonów w trajektorii  $n$ , jak i liczby fotonów, przez które reprezentowany jest konkretny stan ( $I_1 = I_1^D + I_1^A$  i  $I_2 = I_2^D + I_2^A$ ). Stosunek intensywności niższej do wyższej w każdym kolorze pozostawał stały i wynosił  $I_1^D/I_1^A = I_2^D/I_2^A = 1 : 4$ .

---

##### 5.1.1.1 Wyniki dla eksperymentów 2-kolorowych

Symulowano trajektorie czasów międzydetekcyjnych dla modelu 2-stanowego:

$$\begin{array}{|c|} \hline I_1^D \\ \hline I_1^A \\ \hline \end{array} \begin{array}{c} \xrightarrow{1} \\ \xleftarrow{1} \end{array} \begin{array}{|c|} \hline I_2^D \\ \hline I_2^A \\ \hline \end{array} \quad (5.1)$$

Migotanie donora i akceptora było symetryczne ( $I_1^D = I_2^A$  oraz  $I_2^D = I_1^A$ ), różna była natomiast sumaryczna intensywność w każdym stanie i wynosiła od  $I_1 = I_2 = 1$  do  $I_1 = I_2 = 50$ . Długości trajektorii wahały się w zakresie  $n = 10^3$  do  $n = 5 \times 10^4$ . Należy pamiętać, że w tak zaprojektowanym eksperymencie dla danej liczby fotonów w trajektorii ulegają zmianie dwa parametry. Po pierwsze zmienia

się intensywność, która niesie ze sobą informację odnośnie tego, w którym ze stanów układ się w danym momencie znajduje. Przy niskich intensywnościach ( $I = 1$ ), kiedy  $k_{12} = k_{21} = 1$ , szybkość detekcji fotonów jest taka sama jak szybkość przeskoków układu pomiędzy stanami. Wtedy jeden stan jest reprezentowany średnio przez jeden foton. A zatem informacja na temat stanu jest niewielka. Natomiast dla tych samych stałych szybkości, kiedy  $I = 50$ , na jeden stan obserwuje się średnio 50 fotonów i w tych warunkach informacja o stanie jest większa, a więc jakość dopasowanych parametrów powinna się polepszać. Drugim ze zmieniających się parametrów, przy różnych intensywnościach i takiej samej sumarycznej liczbie fotonów jest liczba obserwowanych cykli. Przez cykl rozumie się tutaj przeskok ze stanu pierwszego do drugiego i z powrotem, a więc liczba cykli oznacza średnią liczbę pojawień się każdego ze stanów w trajektorii. W tabeli 5.1 przedstawiono średnią liczbę cykli dla zaprezentowanych w tab. 5.2 parametrów.

**Tabela 5.1:** Liczba cykli dla eksperymentów dwukolorowych, dla różnych wartości  $I_1^D + I_1^A = I_2^D + I_2^A = I$  oraz długości trajektorii  $n$ .

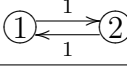
$I$	$n = 10^3$	$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
1	500	2500	5000	25000
5	100	500	1000	5000
10	50	250	500	2500
50	10	50	100	500
100	5	25	50	250

W omawianym modelu stałe kinetyczne są identyczne i obie przyjmują wartość  $k_{12} = k_{21} = 1$ . Średni czas przebywania w danym stanie jest równy odwrotności sumy wszystkich stałych szybkości ucieczki z tego stanu ( $\tau_i = 1/\sum_{j \neq i} k_{ij}$ ). Średni więc czas przebywania w każdym ze stanów wynosi  $\tau_i = 1$ . W takim przypadku średnia liczba fotonów przypadająca na dany stan jest równa intensywności w tym stanie. Dla takiego przypadku średnia liczba cykli (przeskoków ze stanu 1 do 2 i z powrotem) może być obliczona ze wzoru:  $n/2I$ , czyli dzieląc liczbę fotonów w trajektorii przez średnią liczbę fotonów przypadającą na cykl ( $I_1 + I_2 = 2I$ ).

Tabela 5.2 przedstawia wartości średnie i odchylenia standardowe dla stałych kinetycznych z  $10^3$  dopasowań. Kilka rzeczy wartych jest zauważenia.

1. We wszystkich przedstawionych przypadkach dla danej intensywności odchylenia standardowe dopasowanych parametrów zmniejszają się w miarę

**Tabela 5.2:** Wartości średnie i odchylenia standardowe z  $10^3$  skalowanych dopasowanych stałych szybkości dla modelu 2-stanowego o różnych długościach trajektorii i intensywnościach, z wykorzystaniem dwóch kolorów barwników.

	$I$		$n = 10^3$	$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
$I_1^D = 0,8 \quad I_2^D = 0,2$ $I_1^A = 0,2 \quad I_2^A = 0,8$	1	$\kappa_{12}$	1,15 (0,68)	1,00 (0,24)	1,01 (0,15)	1,001 (0,065)
		$\kappa_{21}$	1,09 (0,65)	1,02 (0,24)	1,00 (0,16)	1,003 (0,063)
$I_1^D = 4 \quad I_2^D = 1$ $I_1^A = 1 \quad I_2^A = 4$	5	$\kappa_{12}$	1,03 (0,29)	1,00 (0,12)	1,007 (0,087)	1,000 (0,040)
		$\kappa_{21}$	1,04 (0,29)	1,01 (0,12)	1,006 (0,088)	0,999 (0,038)
$I_1^D = 8 \quad I_2^D = 2$ $I_1^A = 2 \quad I_2^A = 8$	10	$\kappa_{12}$	1,03 (0,29)	1,01 (0,12)	1,01 (0,08)	0,999 (0,037)
		$\kappa_{21}$	1,02 (0,29)	1,00 (0,12)	1,00 (0,08)	1,000 (0,038)
$I_1^D = 40 \quad I_2^D = 10$ $I_1^A = 10 \quad I_2^A = 40$	50	$\kappa_{12}$	1,08 (0,41)	1,01 (0,17)	1,01 (0,12)	1,003 (0,055)
		$\kappa_{21}$	1,06 (0,43)	1,01 (0,17)	1,00 (0,12)	1,004 (0,051)
$I_1^D = 80 \quad I_2^D = 20$ $I_1^A = 20 \quad I_2^A = 80$	100	$\kappa_{12}$	1,16 (0,60)	1,03 (0,23)	1,01 (0,16)	1,000 (0,070)
		$\kappa_{21}$	1,16 (0,63)	1,02 (0,23)	1,01 (0,16)	0,999 (0,069)

wydłużania trajektorii, np. dla  $I = 5$  wahają się w przedziale od ok. 29% dla  $n = 10^3$  do ok. 4% dla  $n = 5 \times 10^4$ .

2. Dla niskich intensywności ( $I = 1$ , co przy stałych kinetycznych równych 1 oznacza, że średni czas przebywania w danym stanie jest porównywalny z czasem pomiędzy pojawieniem się poszczególnych fotonów) możliwe jest odtworzenie parametrów dla dłuższych trajektorii ( $n = 10^4$  oraz  $5 \times 10^4$  fotonów), informacja o stanie jest w tym przypadku mała, duże jest natomiast próbkowanie: 5000 i 25000 cykli. Odchylenia standardowe  $10^3$  dopasowań wynoszą ok. 15% dla  $n = 10^4$  i ok. 6,5% dla  $n = 5 \times 10^4$ .
3. Dla najkrótszej trajektorii ( $n = 10^3$ ) fotonów i najniższej intensywności ( $I = 1$ ) oraz najwyższej intensywności ( $I = 100$ ) obserwowano obciążenie estymowanych parametrów w górę o ok. 10%, a rozrzut statystyczny był znaczny i wynosił ok. 60%.
4. Porównując  $I = 1$  oraz  $I = 5$ , zwiększanie intensywności w każdym przypadku, przy niezmienniej długości trajektorii, powoduje polepszenie otrzymanych wyników (np. dla  $n = 5 \times 10^3$  odchylenia standardowe zmniejszają się z ok. 24% do ok. 12%), natomiast porównując  $I = 10$  i  $I = 50$  obserwuje się pogorszenie statystyki wyników dla wszystkich długości trajektorii (np. dla  $n = 5 \times 10^3$  odchylenie standardowe rośnie z ok. 12% do ok. 17%).

Na wyniki dopasowań wpływają dwa parametry: liczba fotonów reprezentujących dany stan oraz liczba cykli. Należy pamiętać, że ze wzrostem intensywności przy stałej długości trajektorii maleje liczba cykli.

**Tabela 5.3:** Wartości średnie i odchylenia standardowe z  $10^3$  skalowanych dopasowanych intensywności dla modelu 2-stanowego o różnych długościach trajektorii i intensywnościach, z wykorzystaniem dwóch kolorów barwników.

$\textcircled{1} \xleftrightarrow[1]{1} \textcircled{2}$	$I$		$n = 10^3$	$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
$I_1^D = 0,8$ $I_2^D = 0,2$ $I_1^A = 0,2$ $I_2^A = 0,8$	1	$\iota_1^D$	1,05 (0,20)	1,005 (0,077)	1,005 (0,052)	0,999 (0,021)
		$\iota_2^D$	0,96 (0,52)	0,97 (0,24)	1,00 (0,16)	0,997 (0,066)
		$\iota_1^A$	0,89 (0,52)	1,00 (0,23)	0,99 (0,16)	1,000 (0,067)
		$\iota_2^A$	1,03 (0,21)	1,010 (0,082)	1,000 (0,052)	1,001 (0,021)
$I_1^D = 4$ $I_2^D = 1$ $I_1^A = 1$ $I_2^A = 4$	5	$\iota_1^D$	1,004 (0,081)	1,000 (0,035)	1,002 (0,025)	1,001 (0,011)
		$\iota_2^D$	0,99 (0,19)	0,995 (0,087)	0,998 (0,062)	1,000 (0,028)
		$\iota_1^A$	0,99 (0,20)	1,005 (0,087)	1,001 (0,061)	0,999 (0,027)
		$\iota_2^A$	1,008 (0,081)	1,002 (0,034)	0,999 (0,025)	1,000 (0,011)
$I_1^D = 8$ $I_2^D = 2$ $I_1^A = 2$ $I_2^A = 8$	10	$\iota_1^D$	1,005 (0,067)	1,002 (0,029)	1,000 (0,021)	1,0010 (0,0089)
		$\iota_2^D$	1,00 (0,15)	1,000 (0,067)	1,000 (0,048)	0,999 (0,021)
		$\iota_1^A$	1,00 (0,16)	0,994 (0,067)	0,997 (0,046)	1,000 (0,022)
		$\iota_2^A$	1,000 (0,065)	1,000 (0,028)	1,000 (0,021)	1,0002 (0,0091)
$I_1^D = 40$ $I_2^D = 10$ $I_1^A = 10$ $I_2^A = 4$	50	$\iota_1^D$	1,001 (0,055)	1,001 (0,024)	1,000 (0,016)	1,0001 (0,0072)
		$\iota_2^D$	0,99 (0,12)	1,001 (0,048)	1,000 (0,035)	1,001 (0,016)
		$\iota_1^A$	1,00 (0,12)	1,000 (0,051)	0,999 (0,035)	0,999 (0,015)
		$\iota_2^A$	1,004 (0,057)	0,999 (0,024)	1,000 (0,016)	1,0001 (0,0071)
$I_1^D = 80$ $I_2^D = 20$ $I_1^A = 20$ $I_2^A = 80$	100	$\iota_1^D$	0,998 (0,057)	1,000 (0,023)	1,000 (0,016)	1,0000 (0,0079)
		$\iota_2^D$	1,00 (0,11)	1,001 (0,047)	1,001 (0,036)	1,000 (0,014)
		$\iota_1^A$	1,00 (0,12)	0,997 (0,048)	1,001 (0,034)	1,001 (0,015)
		$\iota_2^A$	1,000 (0,056)	1,000 (0,023)	1,000 (0,016)	1,0001 (0,0069)

Dla  $I = 1$  obserwowana liczba fotonów na jeden stan jest mała (średni czas pomiędzy fotonami jest równy średniemu czasowi przebywania w tym stanie). Uzyskuje się więc z pomiaru niewielką informację na temat tego, w którym stanie układ się w danym momencie znajduje. Z tego powodu dla  $n = 10^3$  (500 cykli) otrzymuje się wynik obciążony, a odchylenie standardowe wynosi ok. 68%. Wydłużenie trajektorii do  $n = 10^4$  (obserwuje się ok. 5000 cykli) pozwala odzyskiwać

wyniki nie wykazujące obciążenia i z akceptowalnym rozrzutem statystycznym, ok. 15%.

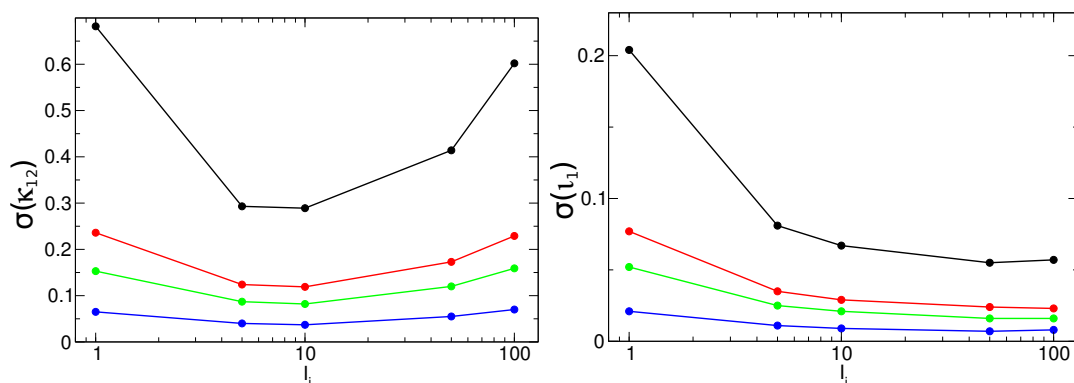
Zwiększając intensywność do  $I = 5$  uzyskuje się większą informację na temat stanu układu, ponieważ średni czas pomiędzy fotonami jest 5 razy mniejszy od średniego czasu przebywania w danym stanie i następuje rejestracja średnio 5 fotonów na stan. Dla  $n = 10^3$ , pomimo obserwacji ok. 100 cykli (mniej, niż dla  $I = 1$ ) poprawiają się wyniki w stosunku do  $I = 1$  dla tej samej liczby fotonów w trajektorii: wartość średnia nie jest obciążona, a odchylenie standardowe wynosi ok. 30%. Akceptowalne wyniki, tj odchylenie standardowe ok. 12% otrzymuje się dla  $I = 5$  już przy  $n = 5 \times 10^3$  fotonów (500 cykli). Średnia liczba fotonów na stan wynosząca 5 jest dostatecznie wysoka, by z dużą dokładnością odtworzyć parametry modelu.

Utrzymując długość trajektorii na tym samym poziomie i zwiększając intensywność do  $I = 50$  zwiększa się informację na temat danego stanu - średnio 50 fotonów na stan, zmniejsza się natomiast liczbę obserwowanych cykli (np. do 10 dla  $n = 10^3$  i  $I = 50$ ). Wtedy otrzymuje się wynik z 16% obciążeniem i 60% rozrzutem statystycznym. Tym razem wzrost liczby fotonów na stan nie rekompensuje zmniejszenia liczby cykli w trajektorii, co skutkuje większym rozrzutem wyników dopasowań.

Należy jednak pamiętać, że rozpatrywany był przypadek, kiedy zwiększano intensywność utrzymując sumaryczną liczbę fotonów na stałym poziomie. Natomiast dla takiej samej liczby cykli, zwiększając intensywność w danym stanie (jednocześnie zwiększając całkowitą liczbę fotonów), uzyskuje się wyniki dopasowań z mniejszymi odchyleniami standardowymi. Odchylenia standardowe dla  $\kappa_{12}$  z tab. 5.2 przedstawiono graficznie na rys. 5.1. Dla każdej długości trajektorii optymalna liczba obserwowanych fotonów na stan wynosi 5 do 10. Wynika z tego kolejny wniosek, a mianowicie, że dobierając intensywność fluorescencji, mając do dyspozycji małą liczbę fotonów należy wziąć po uwagę, że optymalna szybkość emisji fotonów w danym stanie powinna być 5 do 10 razy wyższa od średniej stałej ucieczki z tego stanu.

W tab. 5.3 przedstawiono wartości średnie oraz odchylenia standardowe intensywności dla powyższych eksperymentów. Dla każdego zestawu parametrów estymaty intensywności charakteryzują się mniejszymi odchyleniami standardowymi, niż estymaty stałych kinetycznych. W miarę podwyższania intensywności, dla danej długości trajektorii estymaty intensywności charakteryzują

się coraz mniejszymi odchyleniami standardowymi (rys. 5.1), np. dla  $n = 10^3$  oraz  $I = 1$  odchylenie standardowe wynosi ok. 20% dla wyższych ( $I^c = 0,8$ ) i 52% dla niższych ( $I^c = 0,2$ ) intensywności, natomiast dla  $I = 100$  wynosi 5,7% dla niższych ( $I^c = 20$ ) i ok. 11% dla wyższych ( $I^k = 80$ ) intensywności. Porównując  $I = 50$  oraz  $I = 100$  nie obserwuje się już poprawienia jakości estymat intensywności dla danej długości trajektorii.



**Rysunek 5.1:** Odchylenia standardowe  $\kappa_{12}$  oraz  $l_1^D$  dla różnych intensywności w danym stanie. Kolorami oznaczono różną długość trajektorii:  $10^3$  - czarny,  $5 \times 10^3$  - czerwony,  $10^4$  - zielony,  $5 \times 10^4$  - niebieski.

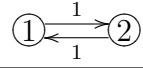
### 5.1.1.2 Wyniki dla eksperymentów 1-kolorowych

Wykonano symulacje i dopasowania dla takich samych zestawów parametrów, lecz dla eksperymentu jednokolorowego. Rejestrowano migotanie jednego tylko barwnika, liczba fotonów w trajektorii była równa sumarycznej liczbie fotonów donora i akceptora w eksperymentach dwukolorowych. Oznacza to, że w eksperymentach 1-kolorowych obserwowano dwukrotnie większą liczbę cykli. Wyniki dopasowań przedstawiono w tab. 5.4 oraz tab. 5.5.

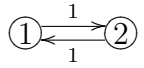
Porównując wyniki otrzymane w eksperymentach jednokolorowych oraz wyniki dopasowań dla eksperymentów dwukolorowych można wyciągnąć następujące wnioski.

1. W każdym przypadku intensywności dopasowują się z mniejszym rozrzutem statystycznym, niż stałe kinetyczne.
2. Dla niskich intensywności, w których średnia szybkość emisji fotonów danego koloru jest mniejsza od średniej szybkości zmian pomiędzy stanami, odchylenia standardowe dla eksperymentów 1-kolorowych są wyższe, niż te

**Tabela 5.4:** Wartości średnie i odchylenia standardowe stałych szybkości z  $10^3$  skalowanych dopasowanych stałych szybkości dla modelu 2-stanowego, eksperyment 1-kolorowy o różnych długościach trajektorii i intensywnościach.

		$n = 10^3$	$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
$I_1 = 0,2$	$\kappa_{12}$	1,04 (0,68)	1,00 (0,33)	1,011 (0,22)	0,999 (0,067)
$I_2 = 0,8$	$\kappa_{21}$	1,07 (0,83)	1,03 (0,35)	1,012 (0,22)	1,004 (0,081)
$I_1 = 1$	$\kappa_{12}$	1,05 (0,39)	1,01 (0,16)	1,004 (0,10)	1,001 (0,047)
$I_2 = 4$	$\kappa_{21}$	1,04 (0,40)	1,00 (0,16)	1,009 (0,11)	1,001 (0,050)
$I_1 = 2$	$\kappa_{12}$	1,03 (0,30)	1,00 (0,12)	1,007 (0,088)	1,001 (0,039)
$I_2 = 8$	$\kappa_{21}$	1,02 (0,32)	1,00 (0,13)	1,007 (0,092)	1,003 (0,041)
$I_1 = 10$	$\kappa_{12}$	1,06 (0,33)	1,01 (0,13)	1,004 (0,096)	1,001 (0,042)
$I_2 = 40$	$\kappa_{21}$	1,05 (0,33)	1,01 (0,14)	1,003 (0,097)	1,001 (0,045)
$I_1 = 20$	$\kappa_{12}$	1,13 (0,47)	1,03 (0,17)	1,02 (0,13)	1,000 (0,051)
$I_2 = 80$	$\kappa_{21}$	1,05 (0,39)	1,01 (0,16)	1,00 (0,12)	1,001 (0,054)

**Tabela 5.5:** Wartości średnie i odchylenia standardowe z  $10^3$  skalowanych dopasowanych intensywności dla modelu 2-stanowego, eksperyment 1-kolorowy o różnych długościach trajektorii i intensywnościach.

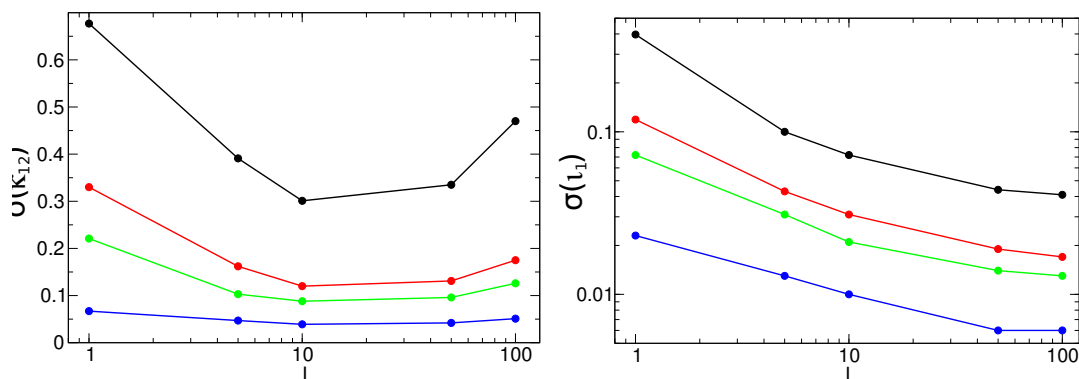
		$n = 10^3$	$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
$I_1 = 0,2$	$\iota_1$	0,852 (0,658)	0,976 (0,350)	0,981 (0,231)	1,001 (0,071)
$I_2 = 0,8$	$\iota_2$	1,096 (0,397)	1,021 (0,119)	1,006 (0,072)	1,002 (0,023)
$I_1 = 1$	$\iota_1$	0,969 (0,277)	0,989 (0,122)	1,000 (0,081)	1,000 (0,037)
$I_2 = 4$	$\iota_2$	1,013 (0,100)	1,000 (0,043)	1,001 (0,031)	1,000 (0,013)
$I_1 = 2$	$\iota_1$	0,994 (0,173)	0,995 (0,074)	1,000 (0,053)	1,000 (0,024)
$I_2 = 8$	$\iota_2$	1,003 (0,072)	0,998 (0,031)	1,001 (0,021)	1,001 (0,010)
$I_1 = 10$	$\iota_1$	0,996 (0,095)	1,001 (0,042)	0,999 (0,028)	1,000 (0,013)
$I_2 = 40$	$\iota_2$	1,005 (0,044)	1,001 (0,019)	1,000 (0,014)	1,000 (0,006)
$I_1 = 20$	$\iota_1$	1,002 (0,084)	1,000 (0,036)	1,001 (0,026)	1,000 (0,012)
$I_2 = 80$	$\iota_2$	1,002 (0,041)	1,001 (0,017)	1,000 (0,013)	1,000 (0,006)

dla eksperymentów 2-kolorowych (np. dla  $I_1 = 0,2$ ,  $I_2 = 0,8$  oraz  $I = 1$ , dla  $n = 10^4$  odchylenia standardowe stałych szybkości dla eksperymentu 1-kolorowego wynosi ok. 22%, natomiast dla eksperymentu 2-kolorowego ok. 15%). Wprowadzenie drugiego koloru zwiększa prawdopodobieństwo

pojawienia się fotonu, a tym samym powoduje, że ilość informacji o stanie układu jest większa.

3. Gdy szybkość emisji fotonów jest większa od szybkości zmian pomiędzy stanami ( $I_1^D + I_1^A = I_2^D + I_2^A = 50$ ), odchylenia standardowe stałych szybkości dla eksperymentu dwukolorowego rosną i stają się wyższe od tych otrzymanych dla jednego koloru (np. dla  $I_1 = 20$ ,  $I_2 = 80$  oraz  $I = 100$ , dla  $n = 10^4$  odchylenia standardowe dla eksperymentu 1-kolorowego wynoszą ok. 13%, natomiast dla eksperymentu 2-kolorowego ok. 16%). Ma to związek z faktem zmniejszenia obserwowanej liczby cykli. Kiedy obecny jest tylko jeden barwnik, na cykl (przeskok ze stanu pierwszego do drugiego i z powrotem) przypada średnio 50 fotonów, natomiast gdy pojawia się drugi barwnik, na cykl przypada 100 fotonów. Przy takiej samej sumarycznej liczbie fotonów zmniejsza się o połowę liczba obserwowanych cykli.
4. Porównując  $I = 10$  dla przypadku 1-kolorowego i  $I = 5$  dla przypadku 2-kolorowego rozpatruje się sytuację, gdzie w eksperymencie 1- i 2- kolorowym liczba fotonów na cykl jest równa i wynosi 10. Porównując takie same długości trajektorii obserwuje się porównywalne odchylenia standardowe odtworzonych wartości parametrów (np. odchylenia standardowe dla  $n = 5 \times 10^3$  wynoszą ok. 12% dla obu przypadków).

Należy więc tak dobierać parametry, by intensywność była na tyle wysoka, by średni czas między pojawianiem się fotonów był krótszy od średniego czasu przebywania układu w danym stanie, innymi słowy szybkość emisji fotonów powinna być większa od szybkości zmian między stanami. Należy pamiętać, że wydłużenie trajektorii może pozwolić na estymację poprawnych parametrów przy niższej intensywności, ponieważ wielkość obserwowanej próby (liczba cykli) zrekompensują zmniejszanie się intensywności, czyli informacji na temat danego stanu. Intensywność powinna być jednak na tyle niska, by obserwować jak największą liczbę cykli. Dzięki takiemu zabiegowi liczebność próby cykli jest możliwie największa, a próbkowanie poszczególnych stanów wystarczające do poprawnego odtworzenia parametrów kinetycznych. Także w eksperymencie jednokolorowym obserwuje się pogorszenie jakości dopasowań stałych kinetycznych dla zbyt wysokiej (100 fotonów na cykl) intensywności w stosunku do obserwowanej liczby cykli, a optymalna intensywność, to 2 i 8 w stanach 1 i 2 (rys. 5.2), natomiast w eksperymencie 2-kolorowym wynosi 5-10.

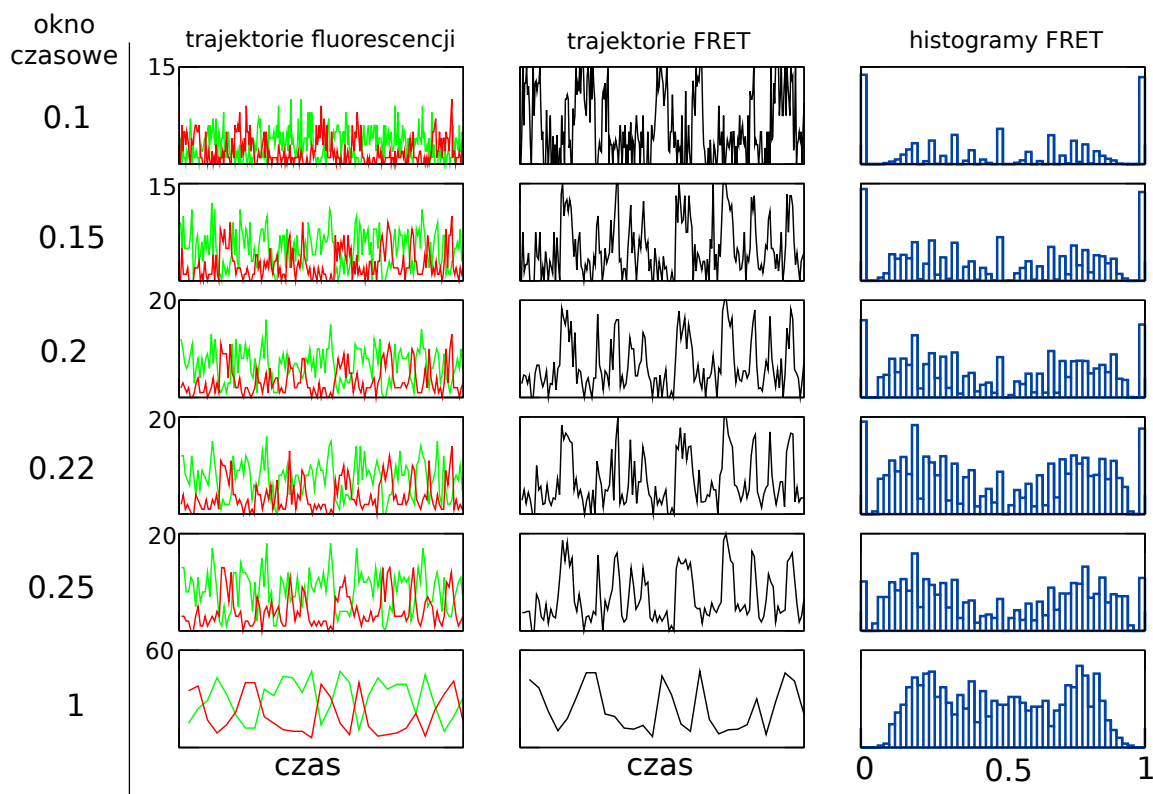


**Rysunek 5.2:** Odchylenia standardowe  $\kappa_{12}$  oraz  $t_1$  dla różnych intensywności w danym stanie w eksperymencie 1-kolorowym. Kolorami oznaczono różną długość trajektorii:  $10^3$  - czarny,  $5 \times 10^3$  - czerwony,  $10^4$  - zielony,  $5 \times 10^4$  - niebieski.

### 5.1.1.3 Porównanie metody ML i analizy on-off

Wykonano także dopasowania metodą on-off dla parametrów przedstawionych w tab. 5.2 oraz 5.3. Pierwszym etapem było dobranie optymalnej wielkości okna czasowego w celu sporządzenia trajektorii fluorescencji donora i akceptora. Wielkość okna czasowego wahała się od niskich wartości ( $h = 0,1$ ) do dużych, równych średniemu czasowi przebywania w danym stanie ( $h = 1$ ). Dla  $I = 1$ ,  $I = 5$  oraz  $I = 10$  niemożliwe było dobranie takiej szerokości okna czasowego i progu, by odtworzyć prawidłowe parametry. Na rys. 5.3 przedstawiono trajektorie fluorescencji donora i akceptora oraz trajektorie FRET i histogramy FRET dla różnych wielkości okna czasowego dla  $I = 50$ ,  $k_{12} = k_{21} = 1$ .

Okno czasowe  $h = 0,1$  jest zbyt małe w stosunku do intensywności i często zdarza się sytuacja, że nie zarejestrowano żadnego fotonu w danym przedziale dla donora lub akceptora, wtedy wydajność FRET wynosi odpowiednio 1 i 0. Pozostałe intensywności są na tyle małe, że stany niskiego i wysokiego FRET nie są dostatecznie rozdzielone i na histogramie FRET nie obserwuje się dwóch wyraźnych maksimów. Okno czasowe  $h = 1$  jest duże w stosunku do średniego czasu przebywania w danym stanie i także w tym przypadku nie rozróżnia się w histogramie FRET dwóch maksimów. Natomiast stosując okno  $h = 0,2$ ;  $0,22$ ;  $0,25$  można zaobserwować na histogramach FRET dwa maksima. Kolejnym etapem prowadzącym do otrzymania końcowego wyniku jest dobór progu rozdzielającego stany o różnej wydajności FRET. Wartość progu  $th$  przyjęto jako połowę wartości pomiędzy stanami niskiej ( $E_{\text{off}} = 0,2$ ) i wysokiej ( $E_{\text{on}} = 0,8$ ) wydajności FRET,  $th = 0,5$ . W tab. 5.6 przedstawiono wyniki dopasowań metodą on-off dla



**Rysunek 5.3:** Przykłady trajektorii fluorescencji (po lewej), trajektorie FRET (pośrodku), histogramy FRET (po prawej) dla modelu 2-stanowego, gdzie  $I = 50$ ,  $k_{12} = k_{21} = 1$  i różnych szerokości okna czasowego.

różnych okien czasowych i na tej podstawie stwierdzono, że estymaty najbliższe wartościom użytym w symulacji otrzymuje się, kiedy  $h = 0,22$ . Wtedy okno czasowe jest ok. 5 razy krótsze od średniego czasu przebywania w danym stanie, a intensywności fluorescencji są wystarczająco wysokie, by na histogramie zliczeń fotonów zaobserwować można dwa maksima.

**Tabela 5.6:** Wartości średnie i odchylenia standardowe otrzymane w wyniku  $10^3$  dopasowań metodą on-off modelu 2-stanowego, gdzie  $I_1^D = 40$ ,  $I_2^D = 10$ ,  $I_1^A = 10$ ,  $I_2^A = 40$ ,  $k_{12} = k_{21} = 1$  dla różnych wielkości okna czasowego.

$h$	$\kappa_{12}$	$\kappa_{12}$
0,15	1,409(0,094)	1,411(0,095)
0,2	1,082(0,066)	1,080(0,068)
0,22	1,013(0,060)	1,011(0,063)
0,25	0,941(0,057)	0,940(0,057)

Analogiczną analizę przeprowadzono dla  $I = 100$ . W tym przypadku optymalną wielkością okna czasowego okazało się  $h = 0,15$ .

**Tabela 5.7:** Analiza on-off. Wartości średnie i odchylenia standardowe  $10^3$  skalowanych parametrów dopasowanych, dla  $k_{12} = k_{21} = 1$ ;  $I_1^D = 40$ ,  $I_2^D = 10$ ,  $I_1^A = 10$ ,  $I_2^A = 40$  i  $I_1^D = 80$ ,  $I_2^D = 20$ ,  $I_1^A = 20$ ,  $I_2^A = 80$  oraz różnych długości trajektorii.

$I$	$h$		$n = 10^3$	$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
50	0,22	$\kappa_{12}$	1,02(0,58)	1,07(0,29)	1,05(0,16)	1,013(0,060)
		$\kappa_{21}$	1,00(0,56)	1,04(0,28)	1,04(0,16)	1,011(0,063)
100	0,15	$\kappa_{12}$	-	0,93 (0,42)	1,02 (0,29)	1,000 (0,086)
		$\kappa_{21}$	-	0,94 (0,41)	1,03 (0,27)	1,030 (0,084)

W tab. 5.7 przedstawiono wyniki  $10^3$  dopasowań otrzymanych przy dopasowaniu metodą on-off dla  $I = 50$  i  $I = 100$ , dla optymalnych, wybranych wcześniej wielkości okna czasowego. Dla  $I = 50$ , dla wszystkich długości trajektorii możliwa była estymacja parametrów kinetycznych. Odchylenia standardowe zmniejszają się w miarę wydłużania trajektorii. Dla  $I = 100$  estymacja parametrów kinetycznych była możliwa dopiero, gdy  $n = 5 \times 10^3$  fotonów, lecz dla tej długości trajektorii obserwowano obciążenie estymowanych parametrów o ok. 6% w dół. Dopiero dla  $n = 10^4$  wartości średnie były bliskie teoretycznym. W porównaniu z dopasowaniem metodą największej wiarygodności, wyniki otrzymane z wykorzystaniem analizy on-off charakteryzują się nieco wyższymi odchyleniami standardowymi. Dodatkowo stosując ML możliwa była estymacja parametrów dla  $I < 50$ .

#### 5.1.1.4 Podsumowanie rozdziału

Wykorzystując metodę ML można z powodzeniem odtworzyć parametry modelu 2-stanowego nawet dla trajektorii liczącej  $5 \times 10^3$  fotonów. Istnieją trzy czynniki, związane z liczbą fotonów i decydujące o jakości dopasowania i są to: średnia liczba fotonów emitowanych w danym stanie, sumaryczna liczba fotonów w trajektorii oraz wielkość wynikająca z dwóch pozostałych: liczba cykli informująca o tym, ile średnio razy każdy stan pojawił się w trajektorii. Przy stałej długości trajektorii  $n$  optymalna sumaryczna liczba fotonów donora i akceptora przypadająca na stan  $I$  w eksperymentach 2-kolorowych wynosi 5-10. Kiedy średnia liczba fotonów reprezentująca dany stan wynosi 50, to możliwa jest estymacja parametrów kinetycznych dla trajektorii zawierającej 50 cykli z odchyleniem standardowym w okolicy 17%. Dla większej liczby cykli ( $2,5 \times 10^4$ ) możliwe jest odtworzenia

parametrów kinetycznych z odchyleniem standardowym ok. 6,4% nawet gdy średnio na stan przypada 1 foton.

Porównując wyniki dopasowań dla eksperymentów jednokolorowych stwierdzono, że dla takich samych długości trajektorii i intensywności barwników, lepsze wyniki otrzymuje się, gdy analizowane są trajektorie zliczeń fotonów obu barwników, zwłaszcza gdy intensywności te są niskie (porównywalne, lub mniejsze od stałej szybkości ucieczki z tego stanu). Dla wyższych intensywności analiza 1 i 2 kolorów fluorescencji daje analogiczne wyniki, gdy liczba fotonów na cykl jest jednakowa dla obu eksperymentów.

Porównując ML z dopasowaniem metodą on-off stwierdzono, że wykorzystując ML możliwe jest odtworzenie parametrów kinetycznych modelu dla szerszego zakresu intensywności, niż dla analizy on-off. Korzystając z analizy on-off możliwe było poprawne odtworzenie parametrów kinetycznych dla przypadku, gdy  $I \geq 50$ . Dla niższych intensywności niemożliwe było obliczenie trajektorii FRET, którą udałooby się rozdzielić na 2 stany o różnej intensywności. Przyczyną była zbyt niska intensywność i niemożność doboru odpowiedniego okna czasowego ( $h$ ) do obliczenia trajektorii fluorescencji. Kiedy  $h$  było małe, zdarzało się, że ani w kanale donora, ani akceptora nie obserwowano fotonów, co uniemożliwiało obliczenie FRET. Natomiast poszerzając  $h$  tak, by obserwować fotony w kanale choćby jednego barwnika wartości intensywności w obu stanach uśredniały się (ponieważ okno czasowe dla tworzenia trajektorii było większe od średniego czasu przebywania w danym stanie) i niemożliwe było rozdzielenie intensywności na dwa różne stany. Jednak nawet dla przypadku, gdy intensywności były na tyle wysokie, by obliczyć trajektorię FRET z widocznymi dwoma stanami, w analizie on-off obserwowano wyższe odchylenia standardowe dopasowywanych parametrów, niż w dopasowaniach metodą ML. Wynika to z faktu, że w analizie ML nie ma konieczności łączenia fotonów w grupy w celu obliczenia trajektorii fluorescencji, a informację o parametrach modelu uzyskuje się bezpośrednio z trajektorii czasów międzydetekcyjnych.

### 5.1.2 Wpływ szybkości przejść pomiędzy stanami

W niniejszym rozdziale wykorzystano model 2-stanowy do sprawdzenia, jaki jest wpływ wartości stałych szybkości zmian kinetycznych na jakość odtwarzanych parametrów. Parametry dobrano tak, by zapewnić dobre rozdzielanie intensywności pomiędzy stanami, a także zapewnić dostateczną długość trajektorii. Przeprowadzono dwa eksperymenty. W jednym stała szybkości  $k_{12}$  przyjmowała wartość 1, zmieniano natomiast  $k_{21}$  w szerokim zakresie: od wolnych przejść w stosunku do emitowanej liczby fotonów ( $k_{21} = 1$ ) do szybkich ( $k_{21} = 50$ ), gdzie szybkość emisji fotonów była porównywalna z szybkością ucieczki ze stanu 2. Celem było doprowadzenie do sytuacji, w której główny wpływ na jakość dopasowań będzie miała szybkość przejść ze stanu 2 do stanu 1. W drugim eksperymencie zmieniano obie stałe szybkości  $k_{12} = k_{21} = k$  w zakresie od wolnych ( $k = 1$ ) do szybkich ( $k = 100$ ) i badano wpływ wartości  $k$  na jakość dopasowań.

W celu zbadania wpływu szybkości zmian pomiędzy stanami na jakość dopasowania wybrano następujący model:



Badania podzielono na dwie części: w pierwszej badano sytuację, kiedy jedna ze stałych kinetycznych nie zmienia się, a druga staje się coraz szybsza. W drugim zaś etapie zmieniano obie stałe szybkości.

#### 5.1.2.1 Zmianie ulega jedna ze stałych szybkości

W pierwszym etapie badań symulowano trajektorie czasów międzydetekcyjnych zmieniając parametr  $k_{21}$  w zakresie od wolnych do szybkich przejść. Schemat symulowanego modelu przedstawiono we wzorze 5.3.



Stosunkowo wysokie intensywności, ich dobre rozdzielanie i długie trajektorie powodowały, że główny wpływ na jakość dopasowania miała wartość  $k_{21}$ . W przypadku rejestracji fotonów w dwóch kanałach otrzymuje się informację o danym stanie w postaci fotonów fluorescencji o poszczególnych kolorach. Informacja o stanie pierwszym zawarta jest w  $I_1 = I_1^D + I_1^A$ , natomiast o stanie drugim:  $I_2 = I_2^D + I_2^A$ . Dla przyjętych w tym eksperymencie komputerowym parametrów będzie to  $I_2 = I_1 = 50$  fotonów na stan. Długość symulowanych trajektorii wynosiła  $n = 5 \times 10^4$  fotonów. Wartość stałej szybkości  $k_{21} = 1$  jest wolna w stosunku do szybkości emitowania fotonów w stanie 2 ( $I_2 = 50$ ), staje się natomiast szybka w stosunku do  $I_2$ , kiedy osiąga wartość  $k_{21} = 50$ .

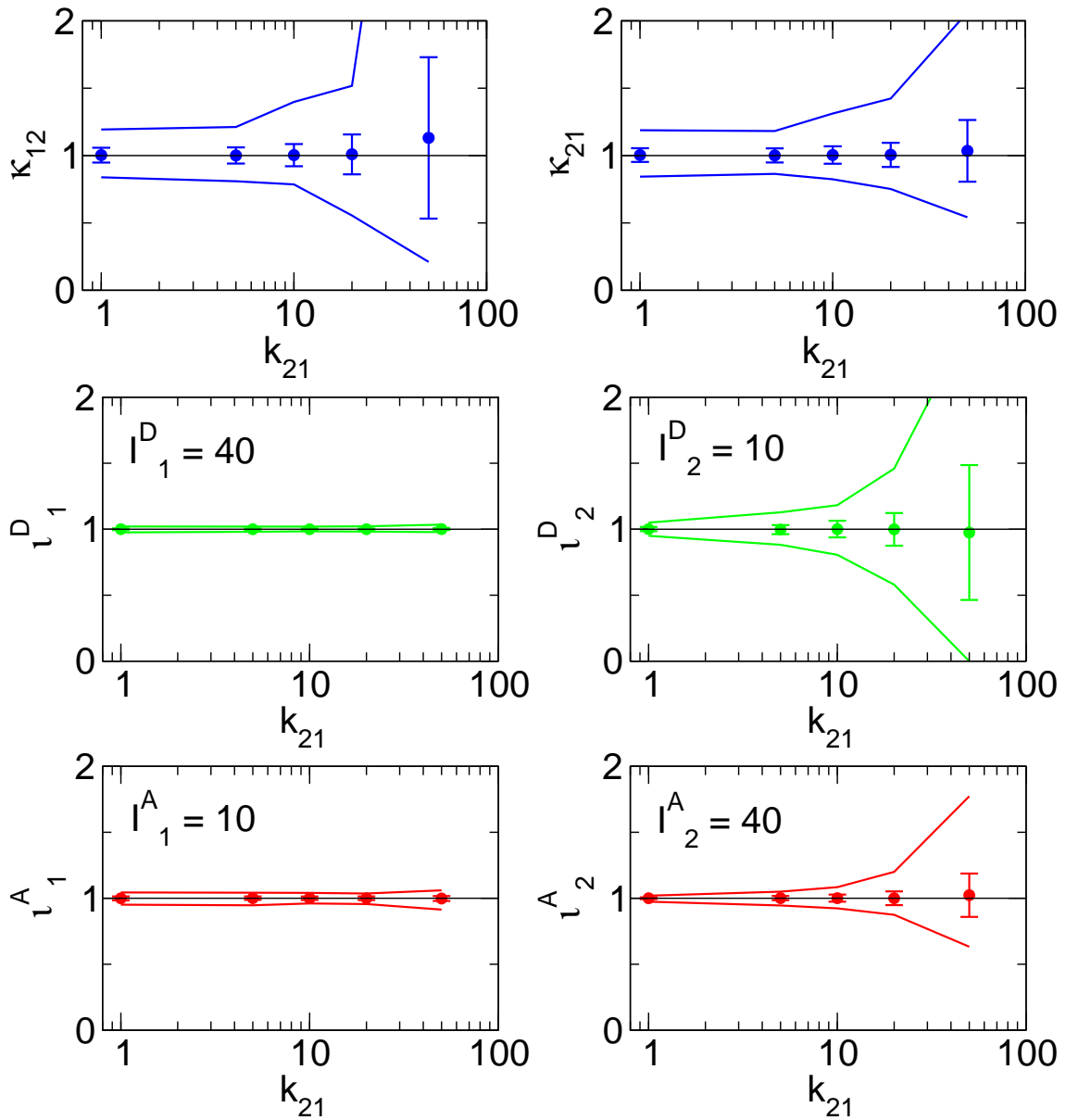
Tabela 5.8 przedstawia wartości średnie i odchylenia standardowe skalowanych parametrów dopasowanych dla modelu 5.3. Na rys. 5.4 przedstawiono wartości średnie i odchylenia standardowe oraz wartości minimalne i maksymalne estymat otrzymanych w  $10^3$  eksperymentach. Na rys. 5.5 wykreślono korelację pomiędzy estymatami stałych szybkości, kiedy zmieniała się wartość  $k_{21}$ .

**Tabela 5.8:** Wartości średnie i odchylenia standardowe dla  $10^3$  skalowanych dopasowanych parametrów stałych przejść oraz intensywności dla modelu 5.3 oraz różnych wartości  $k_{21}$ . Długość trajektorii wynosiła  $n = 5 \times 10^4$  fotonów. W ostatniej kolumnie, oznaczonej jako "zb" przedstawiono procentowy udział eksperymentów, dla których procedura dopasowująca zbiegła się po mniej, niż  $5 \times 10^3$  iteracji.

$k_{21}$	cykle	stałe kinetyczne		donor		akceptor		%zb
5	833	$\kappa_{12}$	1,001 (0,060)	$\iota_1^D$	1,0000 (0,0060)	$\iota_1^A$	1,001 (0,013)	100
		$\kappa_{21}$	1,001 (0,052)	$\iota_2^D$	0,997 (0,034)	$\iota_2^A$	1,001 (0,016)	
10	910	$\kappa_{12}$	1,003 (0,082)	$\iota_1^D$	1,000 (0,0060)	$\iota_1^A$	1,000 (0,013)	100
		$\kappa_{21}$	1,004 (0,065)	$\iota_2^D$	1,001 (0,063)	$\iota_2^A$	1,001 (0,027)	
20	952	$\kappa_{12}$	1,01 (0,15)	$\iota_1^D$	1,0000 (0,0061)	$\iota_1^A$	1,000 (0,014)	100
		$\kappa_{21}$	1,005 (0,090)	$\iota_2^D$	1,00 (0,12)	$\iota_2^A$	1,000 (0,052)	
50	980	$\kappa_{12}$	1,11 (0,57)	$\iota_1^D$	1,0010 (0,0069)	$\iota_1^A$	0,998 (0,019)	98,9
		$\kappa_{21}$	1,02 (0,24)	$\iota_2^D$	0,98 (0,52)	$\iota_2^A$	1,02 (0,17)	

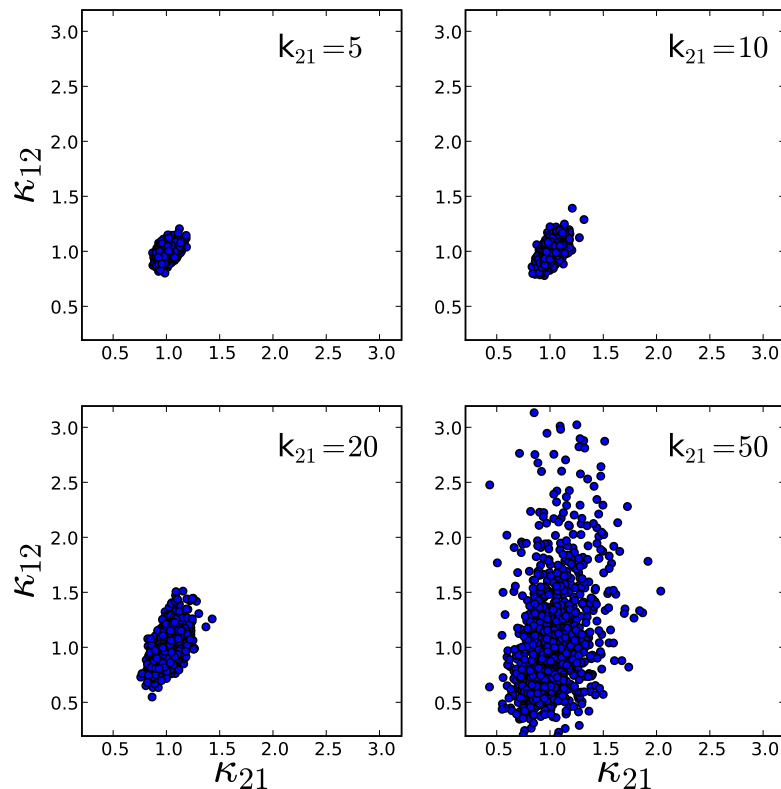
Z przedstawionych wyników można wyciągnąć następujące wnioski:

1. Dla stałych kinetycznych w miarę zwiększania wartości  $k_{21}$  do 20 (średnia szybkość ucieczki ze stanu 2 jest porównywalna ze średnią szybkością emitowania fotonów), odchylenia standardowe dopasowanych wartości rosną (dla  $k_{21} = 1$  odchylenia standardowe wynoszą  $\sigma(\kappa_{12}) = 0,055$ ,  $\sigma(\kappa_{21}) = 0,051$ ,



**Rysunek 5.4:** Wartości średnie i odchylenia standardowe dla  $10^3$  znormalizowanych parametrów dla modelu 5.3, gdzie  $k_{21}$  zmieniało się od 1 do 50. W każdym przypadku długość trajektorii  $n = 5 \times 10^4$ . Liniami ciągłymi oznaczono wartości minimalne i maksymalne.

a dla  $k_{21} = 20$  odchylenia standardowe wynoszą  $\sigma(\kappa_{12}) = 0,15$ ,  $\sigma(\kappa_{21}) = 0,090$ . Wartości średnie natomiast pozostają w okolicy teoretycznej. Możliwe jest więc poprawne odtworzenie wartości parametrów kinetycznych nawet jeśli średni czas przebywania w stanie drugim, w naszym przypadku  $k_{21}^{-1}$  jest tego samego rzędu wielkości co średni czas międzydetekcyjny w tym stanie ( $I_2^{-1}$ ).



**Rysunek 5.5:** Wykres korelacji estymat stałych szybkości dla  $10^3$  znormalizowanych parametrów dla modelu 5.3, gdy  $k_{12} = 1$ , a  $k_{21}$  zmieniało się od 1 do 50. W każdym przypadku długość trajektorii wynosiła  $n = 5 \times 10^4$ .

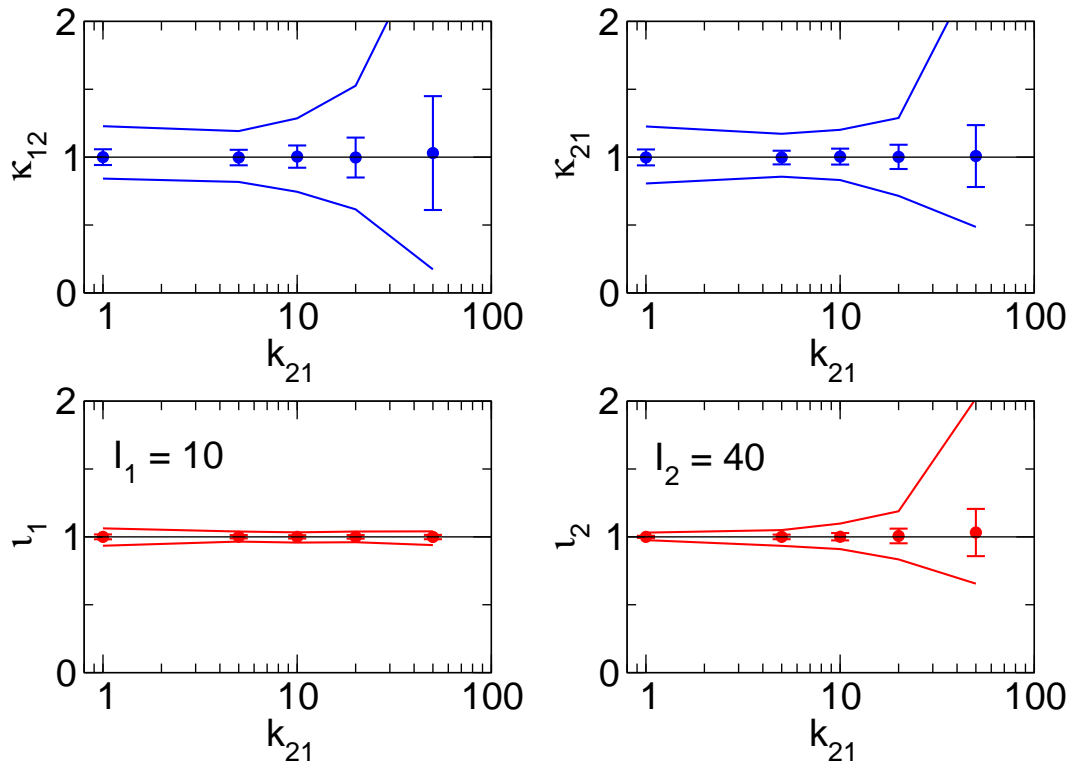
2. Kiedy  $k_{21}$  osiąga wartość 50, wartości średnie wykazują obciążenie ( $\kappa_{12} = 1,11 \pm 0,58$ ,  $\kappa_{21} = 1,02 \pm 0,24$ ).
3. Należy zwrócić uwagę na fakt, że otrzymano lepsze estymaty dla stałej kinetycznej  $k_{21}$ , niż dla  $k_{12}$ .
4. Dopasowania intensywności są lepsze dla stanu pierwszego, w którym w każdym eksperymencie komputerowym średni czas przebywania układu był stały i wynosił  $1/k_{12} = 1$ . Jednocześnie wyniki dopasowań dla donora w stanie pierwszym charakteryzują się mniejszymi wartościami odchyłeń standardowych. Wynika to z faktu wyższej intensywności donora w stanie 1, a co za tym idzie większą liczbę obserwowanych fotonów donora.
5. Dopasowania intensywności dla stanu 2 pogarszają się systematycznie ze zwiększaniem stałej  $k_{21}$ . Średni czas przebywania w stanie 2 jest bowiem

determinowany wartością stałej  $k_{21}$  ( $\tau_2 = 1/k_{21}$ ). W tych warunkach otrzymano gorsze estymaty dla donora, ponieważ intensywność donora jest tutaj niższa od intensywności akceptora.

6. Dla  $k_{21} = 5$  i  $10$  estymaty stałych szybkości wykazują niewielką korelację, ale porównywalne rozrzuty statystyczne. Dla  $k_{21} = 20$  wyniki nadal są skorelowane, a rozrzut  $\kappa_{21}$  jest mniejszy, niż dla  $\kappa_{12}$ . Dla  $k_{21} = 50$  rozrzut wyników jest bardzo duży, a korelacja nieco zanika. Nadal lepsze estymaty otrzymuje się dla  $k_{21}$ .

Sprawdzano także jak przedstawiają się wyniki, jeśli dopasowanie byłoby przeprowadzone dla jednej tylko trajektorii fluorescencji. Przeprowadzono eksperyment, gdzie stałe szybkości były analogiczne jak dla eksperymentu 2-kolorowego. Mierzono fluorescencję jednego tylko barwnika dla  $I_1 = 10$ ,  $I_2 = 40$ , (intensywności analogicznie jak dla akceptora). Długość trajektorii dobierano tak, by obserwować jednakową średnią liczbę cykli (przeskoków ze stanu 1 do 2 i z powrotem), co dla eksperymentów 2-kolorowych. Dla symulacji dwóch kolorów długość trajektorii  $n = 5 \times 10^4$  fotonów, intensywności dla obu stanów i kolorów były symetryczne ( $I_1^D = I_2^A$  oraz  $I_2^D = I_1^A$ ). Aby więc uzyskać taką samą liczbę cykli dla eksperymentu jednokolorowego należy przyjąć, że długość trajektorii wynosi  $n = 2,5 \times 10^4$  fotonów. W tab. 5.9 i na rys. 5.6 przedstawiono wyniki dla różnych wartości  $k_{21}$  z  $10^3$  dopasowań.

Porównując dopasowanie 1- i 2-kolorowe można wyciągnąć następujące wnioski: dla analogicznej liczby cykli w eksperymencie 1- i 2-kolorowym otrzymuje się podobne wyniki dopasowań. Przy czym odchylenia standardowe dla niskich wartości  $k_{21}$  (równych 1) są nieco niższe dla eksperymentu 2-kolorowego (np. dla eksperymentu 2-kolorowego  $\sigma(\kappa_{12}) = 0,055$ ,  $\sigma(\kappa_{21}) = 0,051$ , a dla jednokolorowego  $\sigma(\kappa_{12}) = 0,058$ ,  $\sigma(\kappa_{21}) = 0,059$ ). Natomiast kiedy  $k_{21}$  staje się szybkie, i przyjmuje wartość 50, lepsze wyniki otrzymano dla analizy trajektorii zliczeń jednego tylko barwnika (np. dla eksperymentu 2-kolorowego  $\sigma(\kappa_{12}) = 0,57$ , wartość średnia  $\kappa_{12}$  wykazuje też obciążenie w górę wynoszące 0,1,  $\sigma(\kappa_{21}) = 0,24$ , a dla jednokolorowego  $\sigma(\kappa_{12}) = 0,42$ ,  $\sigma(\kappa_{21}) = 0,23$ ). Ma to prawdopodobnie związek z faktem, że w analizie 1 barwnika w stanie krótko żyjącym ( $s = 2$ ) intensywność fluorescencji równa jest  $I_2 = 40$ , natomiast w eksperymencie 2-kolorowym  $I_2^A = 40$  oraz  $I_2^D = 10$ . Pomimo faktu, że sumaryczna liczba fotonów w stanie 2 dla eksperymentu 2-kolorowego wynosi  $50/k_{21}$ , to składa się ona z dwóch składników: liczby fotonów donora ( $10/k_{21}$ ) i akceptora ( $40/k_{21}$ ), kiedy informacja o stanie 2 w przypadku



**Rysunek 5.6:** Wartości średnie i odchylenia standardowe dla  $10^3$  znormalizowanych parametrów dla modelu 2-stanowego, gdzie  $I_1 = 10$ ,  $I_2 = 40$ ,  $k_{12} = 1$ , a  $k_{21}$  zmieniało się od 1 do 50. W każdym przypadku długość trajektorii wynosiła  $n = 2,5 \times 10^4$ . Liniami ciągłymi oznaczono wartości minimalne i maksymalne.

**Tabela 5.9:** Wartości średnie i odchylenia standardowe dla  $10^3$  skalowanych dopasowanych parametrów stałych przejść oraz intensywności dla modelu 2-stanowego, w eksperymencie 1-kolorowym, gdzie  $k_{12} = 1$ ,  $I_1 = 10$ ,  $I_2 = 40$  oraz różnych wartości  $k_{21}$ . Długość trajektorii wynosiła  $n = 2,5 \times 10^4$  fotonów. W ostatniej kolumnie, oznaczonej jako "zb" przedstawiono procentowy udział eksperymentów, dla których procedura dopasowująca zbiegła się po mniej, niż  $5 \times 10^3$  iteracji.

$k_{21}$		$\mu(\sigma)$		$\mu(\sigma)$	%zb
5,0	$\kappa_{12}$	0,997 (0,057)	$\iota_1$	1,001 (0,012)	100
	$\kappa_{21}$	0,997 (0,050)	$\iota_2$	1,000 (0,017)	
10,0	$\kappa_{12}$	1,004 (0,082)	$\iota_1$	1,000 (0,012)	100
	$\kappa_{21}$	1,004 (0,059)	$\iota_2$	1,001 (0,027)	
20,0	$\kappa_{12}$	1,00 (0,15)	$\iota_1$	1,000 (0,013)	100
	$\kappa_{21}$	1,0020 (0,089)	$\iota_2$	1,007 (0,054)	
50,0	$\kappa_{12}$	1,03 (0,42)	$\iota_1$	0,999 (0,016)	90,5
	$\kappa_{21}$	1,01 (0,23)	$\iota_2$	1,032 (0,17)	

analizy trajektorii jednego koloru wynosi  $(40/k_{21})$  fotonów w stanie 2. I to właśnie fakt, że obserwuje się średnio tylko  $10/k_{21}$  fotonów donora w stanie 2 może powodować te różnice.

### 5.1.2.2 Zmieniają się obie stałe szybkości

W drugim eksperymencie zmieniano obie stałe kinetyczne i przyjmowały one jednakowe, coraz wyższe wartości.

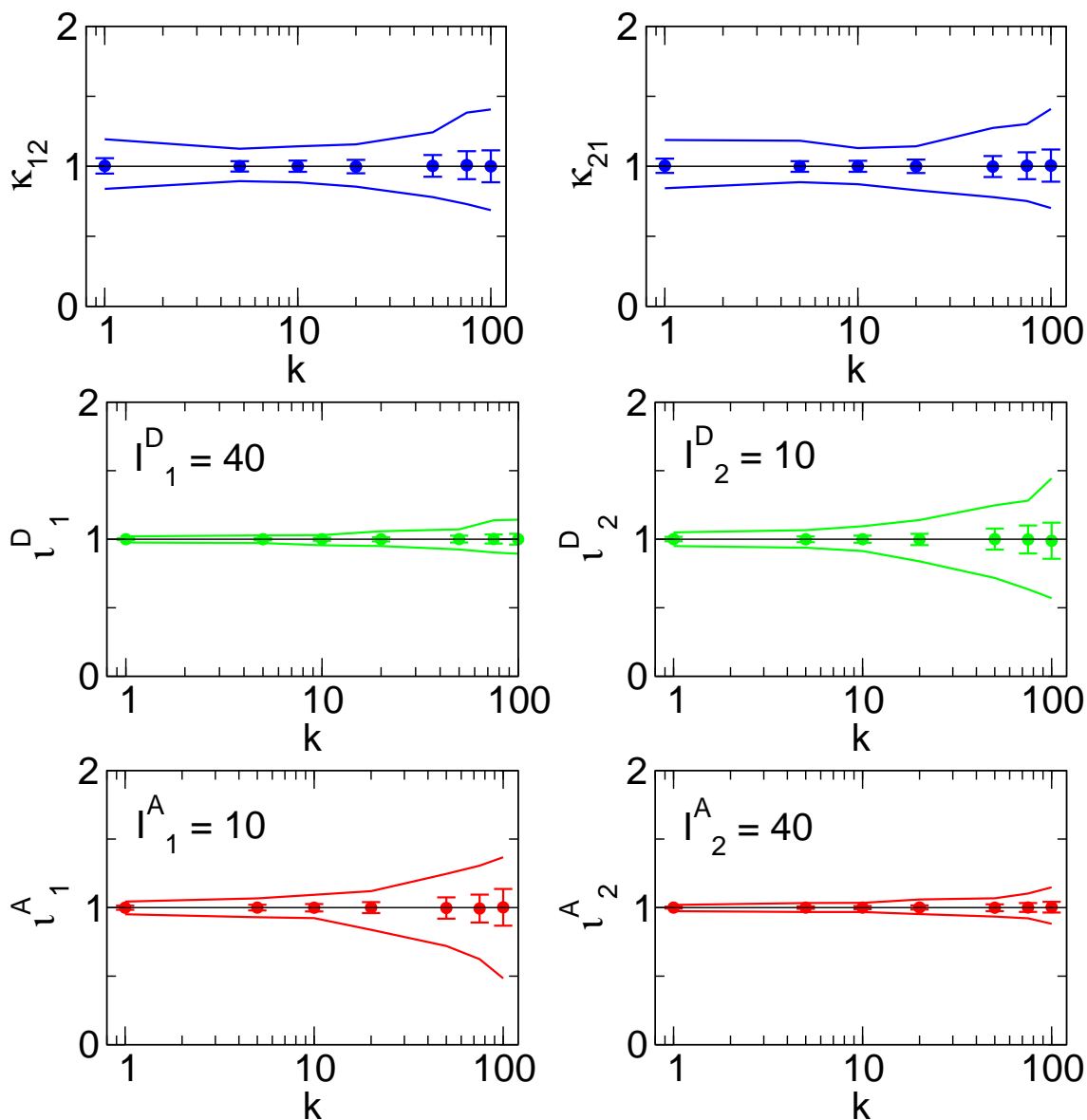


Długość trajektorii była taka sama jak w poprzednim eksperymencie i wynosiła  $n = 5 \times 10^4$  fotonów.

**Tabela 5.10:** Wartości średnie i odchylenia standardowe dla  $10^3$  skalowanych dopasowanych parametrów stałych przejść oraz intensywności dla modelu 5.4,  $k_{12} = k_{21} = k$ . Długość trajektorii wynosiła  $n = 5 \times 10^4$  fotonów. W drugiej kolumnie przedstawiono jaka była średnia liczba cykli w każdym z eksperymentów, a w ostatniej kolumnie, oznaczonej jako "zb" przedstawiono procentowy udział eksperymentów, dla których procedura dopasowująca zbiegła się po mniej, niż  $5 \times 10^3$  iteracji.

$k$	cykle	stałe kinetyczne		donor		akceptor		%zb
5	2500	$\kappa_{12}$	1,000 (0,037)	$\iota_1^D$	1,0000 (0,0090)	$\iota_1^A$	1,000 (0,021)	100
		$\kappa_{21}$	0,999 (0,038)	$\iota_2^D$	0,999 (0,021)	$\iota_2^A$	1,0000 (0,0090)	
10	5000	$\kappa_{12}$	1,001 (0,040)	$\iota_1^D$	1,000 (0,011)	$\iota_1^A$	0,999 (0,027)	100
		$\kappa_{21}$	1,000 (0,039)	$\iota_2^D$	1,000 (0,026)	$\iota_2^A$	1,000 (0,011)	
20	10000	$\kappa_{12}$	0,998 (0,048)	$\iota_1^D$	0,999 (0,015)	$\iota_1^A$	1,000 (0,040)	100
		$\kappa_{21}$	1,000 (0,048)	$\iota_2^D$	0,998 (0,041)	$\iota_2^A$	1,001 (0,015)	
50	25000	$\kappa_{12}$	1,003 (0,078)	$\iota_1^D$	1,001 (0,025)	$\iota_1^A$	0,997 (0,078)	100
		$\kappa_{21}$	0,998 (0,075)	$\iota_2^D$	1,001 (0,077)	$\iota_2^A$	0,999 (0,025)	
75	37500	$\kappa_{12}$	1,01 (0,10)	$\iota_1^D$	1,002 (0,033)	$\iota_1^A$	0,99 (0,10)	100
		$\kappa_{21}$	1,004 (0,096)	$\iota_2^D$	1,00 (0,10)	$\iota_2^A$	1,001 (0,032)	
100	50000	$\kappa_{12}$	1,00 ( 0,11)	$\iota_1^D$	1,000 ( 0,039)	$\iota_1^A$	1,00 ( 0,13))	80,9
		$\kappa_{21}$	1,00 ( 0,11)	$\iota_2^D$	0,99 ( 0,13)	$\iota_2^A$	1,003 ( 0,039)	

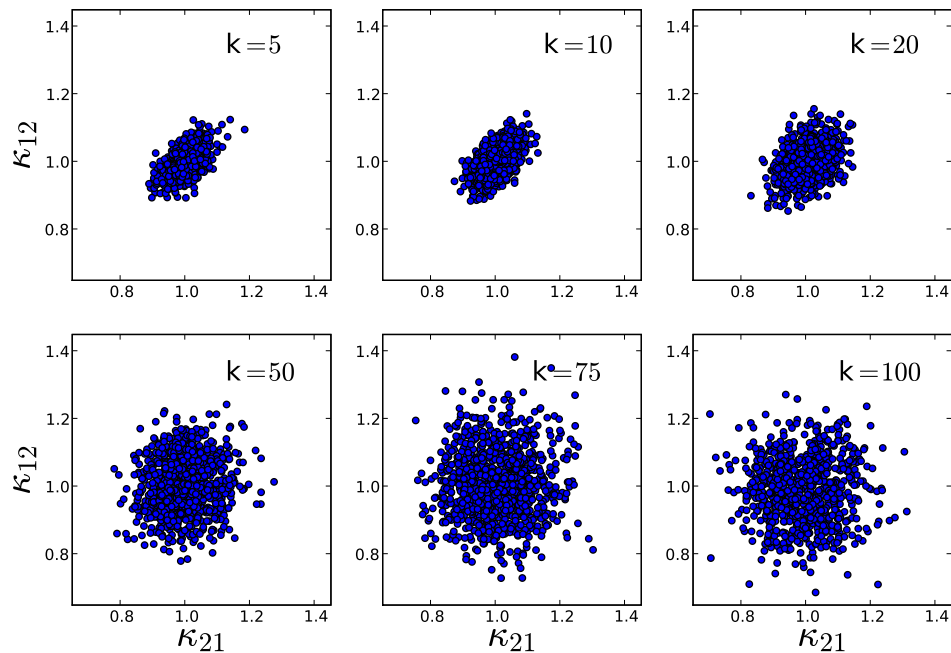
Wartości średnie i odchylenia standardowe z  $10^3$  dopasowań umieszczono w tab. 5.10, a na rys. 5.7 przedstawiono dodatkowo wartości minimalne i maksymalne



**Rysunek 5.7:** Wsrtości średnie i odchylenia standardowe dla  $10^3$  znormalizowanych parametrów dla modelu 2-stanowego, gdzie  $I_1^D = 40$ ,  $I_2^D = 10$ ,  $I_1^A = 10$ ,  $I_2^A = 40$ , a  $k$  zmieniało się od 1 do 100. W każdym przypadku długość trajektorii  $n = 5 \times 10^4$ . Liniami ciągłymi oznaczono wartości minimalne i maksymalne.

dla tych dopasowań. Wykresy na rys. 5.8 przedstawiają korelację pomiędzy estymatami stałych szybkości. Kilka rzeczy wartych jest zauważenia.

1. W miarę zwiększania wartości stałych kinetycznych, odchylenia standardowe estymowanych parametrów rosną (np. dla  $k = 5$  odchylenia standardowe stałych szybkości wynoszą ok. 0,04, dla  $k = 50$  przyjmują wartość ok. 0,076, natomiast dla  $k = 100$  rosną do ok. 0,11).



**Rysunek 5.8:** Wykres korelacji estymat stałych szybkości dla  $10^3$  znormalizowanych parametrów dla modelu 5.3, kiedy  $k_{12} = k_{21} = k$  zmieniało się od 1 do 100. W każdym przypadku długość trajektorii  $n = 5 \times 10^4$ .

2. Nawet kiedy stałe szybkości są dwukrotnie większe ( $k = 100$ ) od intensywności w danym stanie ( $I = 50$ ), co oznacza, że średnio rejestruje się jeden foton na dwa stany, możliwa jest estymacja poprawnych parametrów.
3. Kiedy  $k = 100$ , nie wszystkie dopasowania zbiegły się po mniej, niż  $5 \times 10^3$  iteracjach, te natomiast eksperymenty, w których dopasowanie zakończyło się dawały wyniki bliskie teoretycznym (odchylenie standardowe ok. 0,11). Kiedy wartość  $k$  podniesiono do 150 (danych nie umieszczono), w większości przypadków algorytm dopasowujący nie zbiegł się po  $5 \times 10^3$  iteracjach. Liczba zakończonych poprawnie dopasowań wzrosła, kiedy maksymalną liczbę iteracji zwiększono do  $10^4$ .
4. W miarę zwiększania wartości stałych szybkości zmniejszała się liczba fotonów na stan i obserwowano systematyczne pogarszanie się własności statystycznych estymowanych intensywności fluorescencji. Na przykład dla  $k = 5$  odchylenia standardowe wyższych intensywności ( $I_h$ ) wynosiły 0,009, natomiast niższych ( $I_l$ ) ok. 0,21, dla  $k = 50$  odchylenia intensywności wynosiły ok. 0,025 i ok. 0,077, a dla  $k = 100$  przyjmowały wartości ok. 0,04 i ok. 0,13.

5. Dla  $k = 5$  estymaty wykazywały korelację, ale ich rozrzut był porównywalny, w miarę zwiększania wartości  $k$  korelacja zanikała, a rozrzut wyników zwiększał się.

W eksperymentach 2-kolorowych, gdzie zmieniała się tylko jedna stała szybkości, dla  $k_{21} = 50$  odchylenia standardowe estymowanych stałych szybkości były wysokie (0,57 i 0,24). Natomiast kiedy zmieniano obie stałe szybkości, nawet dla  $k_{12} = k_{21} = 100$  odchylenia standardowe stałych szybkości wynosiły 0,12 i 0,13. Różnica w wynikach ma prawdopodobnie związek z liczbą obserwowanych cykli. Dla eksperymentu, gdzie zmieniała się tylko jedna stała szybkości, obserwowano dla  $k_{21} = 50$  jedynie ok. 980 cykli, a stan  $s = 2$  był reprezentowany przez średnio jeden foton. W trajektorii więc stan drugi pojawiał się średnio 980 razy i był on słabo reprezentowany. Natomiast gdy obie stałe szybkości ulegały zmianie, dla  $k = 50$  obserwowano średnio 37500 cykli. Pomimo faktu, że oba stany były słabo reprezentowane (średnio jeden foton na stan), liczebność próby była wysoka i to umożliwiło estymację poprawnych parametrów kinetycznych.

### 5.1.2.3 Podsumowanie rozdziału

Stosując analizę ML można odtworzyć parametry kinetyczne nawet w przypadku, gdy średnia szybkość detekcji fotonów w danym stanie jest porównywalna lub wyższa od szybkości ucieczki z tego stanu.

Dla krótszych trajektorii o liczbie cykli nie przekraczającej  $10^3$  czynnikiem decydującym o jakości dopasowania jest liczba fotonów reprezentujących dany stan. Kiedy jedna ze stałych szybkości ma stosunkowo niską wartość (np.  $k_{12} = 1$ ), a druga wysoką (np.  $k_{21} = 20$ ), dopasowania stałej szybkości, która ma wyższą wartość charakteryzują się nieco niższymi odchyleniami standardowymi.

Natomiast kiedy obserwowanych jest wiele cykli (ok.  $2 \times 10^4$ ) możliwa jest estymacja parametrów kinetycznych, które są szybsze od stałych szybkości emisji fotonów. Dla tych przypadków estymaty intensywności są tym gorsze, im gorzej reprezentowany jest dany stan. Niemniej jednak dla dużej liczby obserwowanych cykli możliwe jest estymowanie parametrów procesów, które są dwukrotnie szybsze od szybkości emisji fotonów.

Dopasowania intensywności w większości przypadków charakteryzują się niższymi odchyleniami standardowymi i zależą od średniej liczby fotonów,

reprezentującej dany stan. Im więcej fotonów obserwowanych jest w danym stanie, tym niższe odchylenie standardowe estymowanych wartości.

### 5.1.3 Wpływ rozdzielania intensywności

---

Ważnym czynnikiem wpływającym na jakość dopasowania jest informacja o tym, w którym ze stanów układ się w danym momencie znajduje. W przypadku eksperymentów FRET jest to intensywność fluorescencji poszczególnych barwników w każdym ze stanów. Im lepiej rozdzielone są intensywności fluorescencji w poszczególnych stanach tym łatwiej odtworzyć poprawne parametry modelu. W niniejszym rozdziale przedstawione będą wyniki symulacji i dopasowań dla modelu, gdzie  $k_{12} = k_{21} = 1$ , a sumaryczna intensywność obu barwników w każdym ze stanów  $I = 50$ . Wydajności FRET, a co za tym idzie wartości intensywności coraz mniej różnią się od siebie.

---

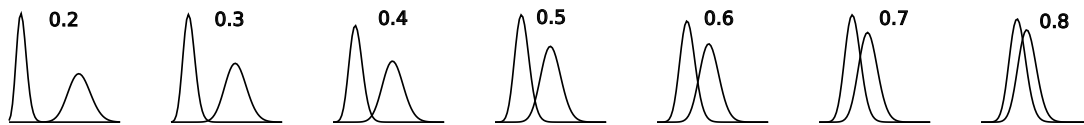
#### 5.1.3.1 Wyniki dla eksperymentów 2-kolorowych

Wykonywano symulacje dla modelu 2-stanowego, w którym obie stałe szybkości  $k_{12}$  oraz  $k_{21}$  były równe 1.



Suma intensywności w obu stanach była równa i wynosiła  $I_s^D + I_s^A = 50$ , różna była natomiast wydajność FRET w obu stanach. Stan charakteryzujący się wyższą wydajnością FRET oznaczano przez “on“, a stan o niższej wydajności FRET: “off“. Natomiast intensywności fluorescencji oznaczano:  $I_{\text{on}}^A = I_{\text{off}}^D = I_h$  (wysoka intensywność fluorescencji),  $I_{\text{off}}^A = I_{\text{on}}^D = I_l$  (niska intensywność fluorescencji). Długości trajektorii wynosiły  $n = 5 \times 10^3$ ,  $n = 10^4$ ,  $n = 5 \times 10^4$ . Dla tak dobranych parametrów liczba cykli była jednakowa dla każdej długości trajektorii i wynosiła odpowiednio 50, 100 i 500. Badano układy, dla których rozkłady zliczeń fotonów w obu stanach są dobrze rozdzielone (niskie wartości  $I_l/I_h \leq 0,4$ ), a co za tym idzie różniły się od siebie szybkości emisji fotonów w obu stanach oraz takie, dla których

intensywności w obu stanach były do siebie podobne  $I_1/I_h > 0,5$  (rys. 5.9), szybkości emisji fotonów w obu stanach były zbliżone.

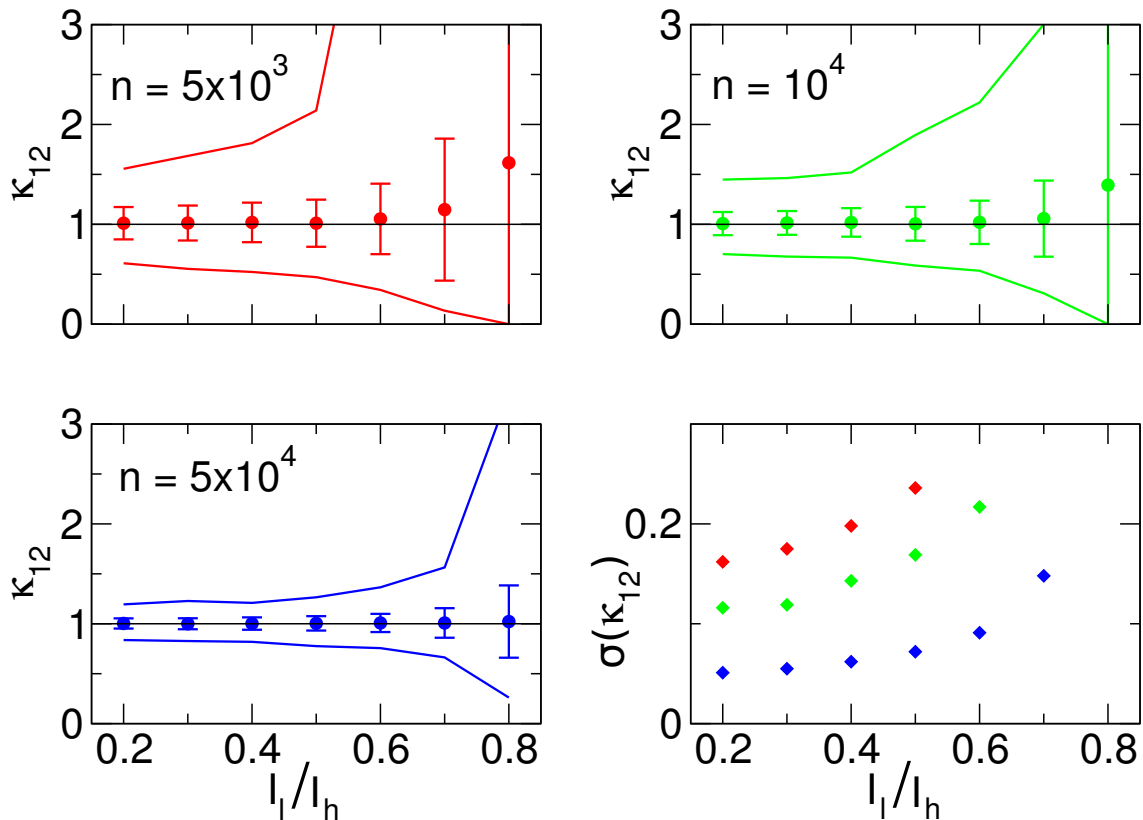


**Rysunek 5.9:** Rozkłady zliczeń fotonów dla różnych wartości  $I_1/I_h$ . Dla niskich wartości  $I_1/I_h \approx 0,2$  intensywności w obu stanach różnią się znacznie, ale gdy  $I_1/I_h \approx 0,5$ , rozkłady zliczeń fotonów zaczynają się nakładać, co powoduje pogorszenie się wyników dopasowania.

**Tabela 5.11:** Wartości średnie oraz odchylenia standardowe dla  $10^3$  skalowanych parametrów dopasowanych do modelu 5.5 oraz dla różnego rozdzielenia intensywności w obu stanach ( $I_1/I_h$ ). Długości trajektorii wynosiły  $n = 5 \times 10^3$ ,  $n = 10^4$ ,  $n = 5 \times 10^4$  fotonów.

$E$	$I_1/I_h$	$I$		$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
$E_{\text{on}}=0,83$ $E_{\text{off}}=0,17$	0,2	$I_h=41,67$	$\kappa_{12}$	1,01 (0,16)	1,01 (0,12)	1,003 (0,051)
		$I_1=8,33$	$\kappa_{21}$	1,01 (0,17)	1,00 (0,12)	1,000 (0,051)
$E_{\text{on}}=0,77$ $E_{\text{off}}=0,23$	0,3	$I_h=38,46$	$\kappa_{12}$	1,01 (0,17)	1,01 (0,12)	1,000 (0,055)
		$I_1=11,54$	$\kappa_{21}$	1,02 (0,19)	1,01 (0,12)	1,001 (0,054)
$E_{\text{on}}=0,71$ $E_{\text{off}}=0,29$	0,4	$I_h=35,71$	$\kappa_{12}$	1,02 (0,20)	1,02 (0,14)	1,002 (0,062)
		$I_1=14,29$	$\kappa_{21}$	1,02 (0,20)	1,02 (0,14)	1,001 (0,062)
$E_{\text{on}}=0,67$ $E_{\text{off}}=0,33$	0,5	$I_h=33,33$	$\kappa_{12}$	1,01 (0,24)	1,00 (0,17)	1,004 (0,072)
		$I_1=16,67$	$\kappa_{21}$	1,01 (0,24)	1,00 (0,16)	1,003 (0,072)
$E_{\text{on}}=0,62$ $E_{\text{off}}=0,38$	0,6	$I_h=31,25$	$\kappa_{12}$	1,05 (0,35)	1,02 (0,22)	1,008 (0,091)
		$I_1=18,75$	$\kappa_{21}$	1,03 (0,32)	1,01 (0,21)	1,011 (0,092)
$E_{\text{on}}=0,59$ $E_{\text{off}}=0,41$	0,7	$I_h=29,41$	$\kappa_{12}$	1,15 (0,71)	1,06 (0,38)	1,01 (0,15)
		$I_1=20,59$	$\kappa_{21}$	1,11 (0,67)	1,05 (0,38)	1,01 (0,14)
$E_{\text{on}}=0,56$ $E_{\text{off}}=0,44$	0,8	$I_h=27,78$	$\kappa_{12}$	1,6 (4,0)	1,4 (2,1)	1,02 (0,36)
		$I_1=22,22$	$\kappa_{21}$	1,81 (4,65)	1,4 (2,5)	1,04 (0,34)

Przez intensywność rozumie się szybkość emisji fotonów, a czas pomiędzy poszczególnymi fotonami jest odwrotnie proporcjonalny do wartości intensywności. Im wyższa intensywność tym częściej następuje detekcja fotonów. W miarę zbliżania się do siebie wartości intensywności, czasy międzydetekcyjne stawały się w obu stanach porównywalne i coraz trudniej było stwierdzić, w którym ze stanów układ się w danym momencie znajdował.

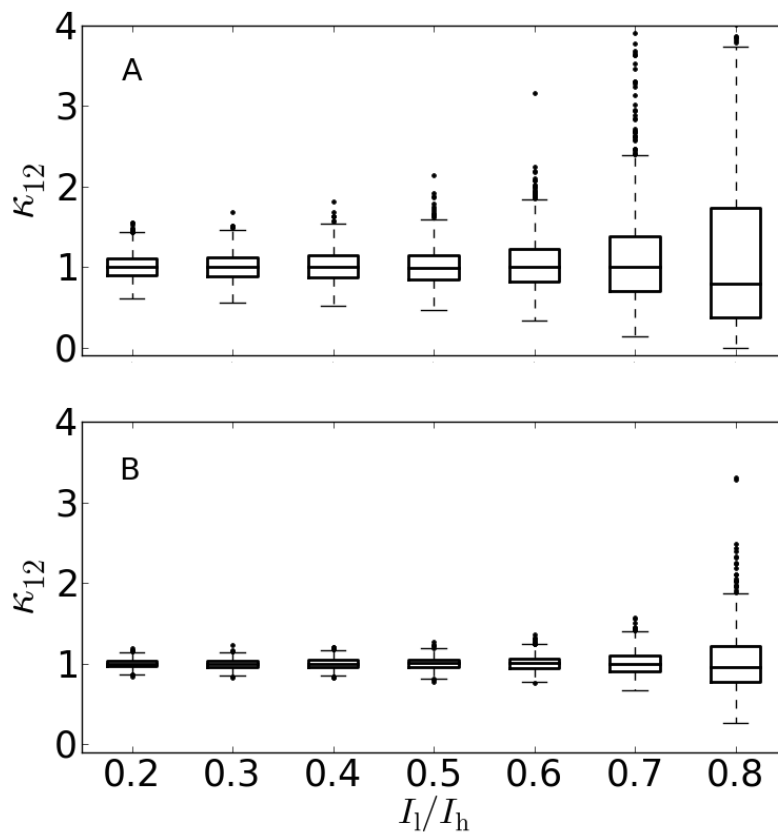


**Rysunek 5.10:** Wartości średnie i odchylenia standardowe dla  $10^3$  znormalizowanych parametrów dla modelu 5.5. Zmieniano wartość ilorazu intensywności ( $I_1/I_h$ ) w zakresie od 0,2 do 0,8. Przedstawiono wyniki dla różnych długości trajektorii:  $n = 5 \times 10^3$  (czerwony),  $n = 10^4$  (zielony) i  $n = 5 \times 10^4$  (niebieski) fotonów. Przedstawiono także odchylenia standardowe dla różnych wartości rozdzielania i różnych długości trajektorii. Liniami ciągłymi oznaczono wartości minimalne i maksymalne.

W tab. 5.11 przedstawiono wartości średnie oraz odchylenia standardowe dla  $10^3$  dopasowań do modeli 2-stanowych z różnym stopniem rozdzielania intensywności. Te same wyniki przedstawiono na rys. 5.10. Gdy intensywności w obu stanach są dobrze rozdzielone ( $I_1/I_h \approx 0,2$ , rys. 5.9, tab. 5.11) otrzymuje się niskie odchylenia standardowe z  $10^3$  dopasowań nawet, gdy długość trajektorii wynosi  $n = 5 \times 10^3$  fotonów (np. dla  $I_1/I_h = 0,2$   $\sigma(\kappa_{12}) = 0,16$ ), natomiast dla  $n = 5 \times 10^4$   $\sigma(\kappa_{12}) = 0,051$ .

Natomiast gdy wartość  $I_1/I_h$  rośnie, powodując coraz większe nakładanie się rozkładów zliczeń fotonów (rys. 5.9) wartości odchyłeń standardowych dopasowanych parametrów rosną (np. dla  $I_1/I_h = 0,5$ ,  $n = 5 \times 10^3$   $\sigma(\kappa_{12}) = 0,24$ , a dla  $n = 5 \times 10^4$   $\sigma(\kappa_{12}) = 0,072$ ).

Wreszcie gdy wartość  $I_1/I_h$  wzrosła do 0,7, dla krótszych trajektorii ( $n = 5 \times 10^3$ ) wartości średnie odbiegają od teoretycznych, dla dłuższych trajektorii wartości średnie pozostają w okolicy wartości teoretycznej, ale odchylenia standardowe są znaczne, np. dla  $I_1/I_h = 0,7$ ,  $n = 10^4$   $\sigma(\kappa_{12}) = 0,38$ . Jakość dopasowania poprawia wydłużenie trajektorii. Kiedy  $I_1/I_h$  osiąga wartość 0,8 nie udało się odtworzyć zadowalających wyników dla żadnej z długości trajektorii.



**Rysunek 5.11:** Wykres pudełkowy dla  $10^3$  znormalizowanych estymat parametrów dla modelu 5.5. Zmieniano wartość ilorazu intensywności ( $I_1/I_h$ ) w zakresie od 0,2 do 0,8, dla długości trajektorii wynosiły  $n = 5 \times 10^3$  (A) oraz  $n = 5 \times 10^4$  (B).

Na rys. 5.11 przedstawiono wyniki dopasowań dla różnych wartości  $I_1/I_h$  i różnych długości trajektorii w postaci wykresu pudełkowego. Dla trajektorii o długości  $n = 5 \times 10^3$  (A) obserwowano obciążenie wartości dopasowanych w górę. Dla  $n = 5 \times 10^4$  (B), nie obserwowano obciążenia dla  $I_1/I_h < 0,5$ . Rozrzut statystyczny wyników zmniejszył się znacznie dla dłuższych trajektorii.

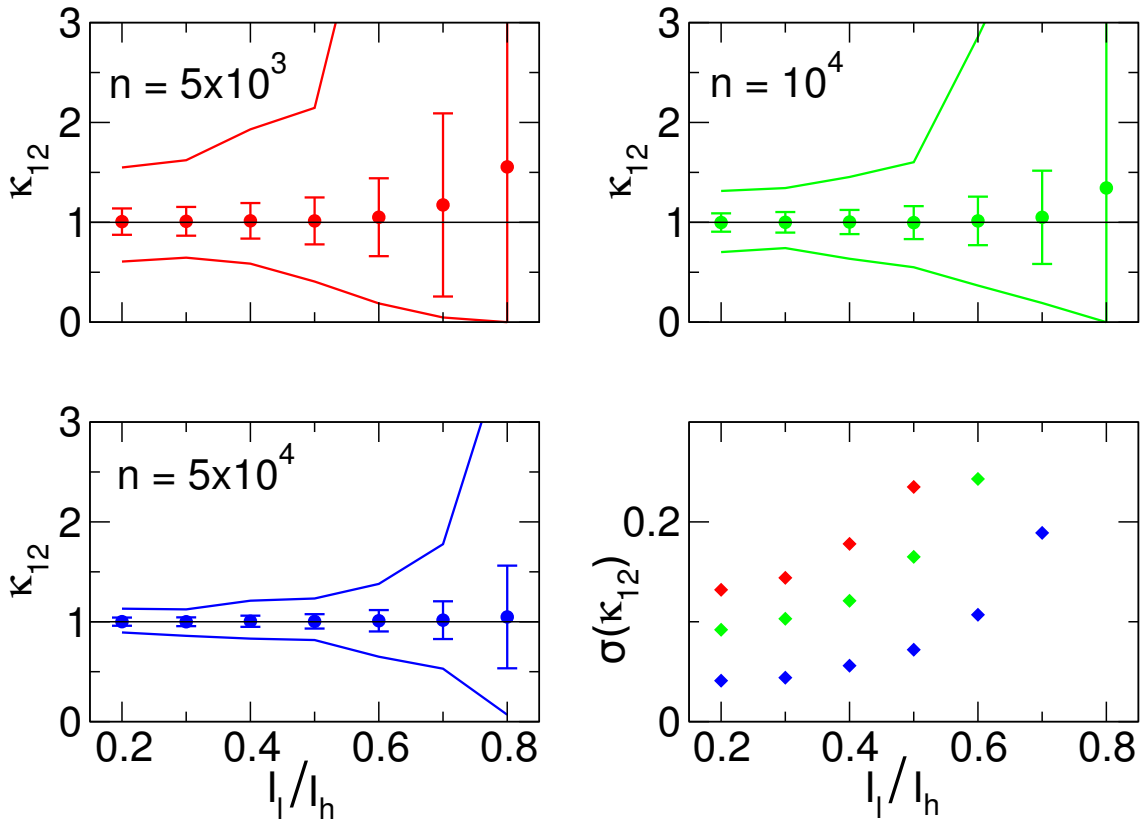
## 5.1.3.2 Wyniki dla eksperymentów 1-kolorowych

Analogiczne symulacje powtórzono dla eksperymentu 1-kolorowego, gdzie sumaryczna liczba fotonów była taka sama jak w eksperymencie 2-kolorowym (liczba cykli była zatem dwukrotnie większa, niż dla eksperymentów 2-kolorowych). Wyniki dopasowań przedstawiono w tab. 5.12 oraz na rys. 5.12.

**Tabela 5.12:** Wartości średnie oraz odchylenia standardowe dla  $10^3$  skalowanych parametrów dopasowanych do modelu 2-stanowego, dla eksperymentu 1-kolorowego, gdzie  $k_{12} = k_{21} = 1$ ,  $I_1 + I_2 = 50$  oraz dla różnego rozdzielenia intensywności w obu stanach ( $I_1/I_h$ ). Długości trajektorii wynosiły  $n = 5 \times 10^3$ ,  $n = 10^4$ ,  $n = 5 \times 10^4$  fotonów.

$I_1/I_h$	$I_h$	$I_1$		$n = 5 \times 10^3$	$n = 10^4$	$n = 5 \times 10^4$
0,2	41,67	8,33	$\kappa_{12}$	1,01 (0,13)	0,998 (0,092)	1,001 (0,041)
			$\kappa_{21}$	1,016 (0,132)	1,010 (0,091)	1,002 (0,039)
0,3	38,46	11,54	$\kappa_{12}$	1,01 (0,14)	1,00 (0,10)	1,000 (0,044)
			$\kappa_{21}$	1,015 (0,144)	1,01 (0,10)	1,003 (0,044)
0,4	35,71	14,29	$\kappa_{12}$	1,01 (0,18)	1,00 (0,12)	1,005 (0,056)
			$\kappa_{21}$	1,022 (0,182)	1,00 (0,12)	1,004 (0,054)
0,5	33,33	16,67	$\kappa_{12}$	1,01 (0,23)	1,00 (0,16)	1,004 (0,072)
			$\kappa_{21}$	1,017 (0,225)	1,01 (0,16)	1,000 (0,070)
0,6	31,25	18,75	$\kappa_{12}$	1,05 (0,39)	1,01 (0,24)	1,01 (0,11)
			$\kappa_{21}$	1,061 (0,362)	1,03 (0,25)	1,01 (0,10)
0,7	29,41	20,59	$\kappa_{12}$	1,17 (0,92)	1,05 (0,47)	1,02 (0,19)
			$\kappa_{21}$	1,161 (0,889)	1,05 (0,55)	1,01 (0,19)
0,8	27,78	22,22	$\kappa_{12}$	1,5 (2,6)	1,3 (6,6)	1,05 (0,51)
			$\kappa_{21}$	1,320 (2,192)	1,2 (1,6)	1,05 (0,53)

Kiedy intensywności są od siebie dobrze rozdzielone nieznacznie lepsze wyniki otrzymuje się dla dopasowania do trajektorii jednokolorowej przy tej samej sumarycznej liczbie fotonów w trajektorii (np. dla  $I_1/I_h = 0,2$ ,  $n = 10^4$  dla 2 kolorów  $\sigma(\kappa_{12}) = 0,12$ , a dla 1 koloru  $\sigma(\kappa_{12}) = 0,092$ ). Natomiast gdy ( $I_1/I_h$ ) przyjmuje wartość 0,6 odchylenia standardowe dla dopasowań 2-kolorowych stają się niższe od tych otrzymanych w wyniku dopasowań do jednej tylko trajektorii (np. dla  $n = 10^4$  dla 2 kolorów  $\sigma(\kappa_{12}) = 0,22$ , a dla 1 koloru  $\sigma(\kappa_{12}) = 0,24$ ). Pomimo o połowę mniejszej liczby obserwowanych cykli, dla trudniejszych warunków dopasowania w symulacjach 2-kolorowych otrzymuje się lepsze wyniki. Niemniej jednak porównując wyniki dla takiej samej liczny cykli ( $5 \times 10^3$  fotonów dla 1 koloru i  $10^4$  fotonów



**Rysunek 5.12:** Wartości średnie i odchylenia standardowe  $10^3$  znormalizowanych parametrów dla modelu 2-stanowego, w eksperymencie 1-kolorowym, gdzie  $I_1 + I_2 = 50$ ,  $k_{12} = k_{21} = 1$ . Zmieniała się wartość ilorazu intensywności niższej do wyższej ( $I_1/I_h$ ) w zakresie od 0,2 do 0,8. Przedstawiono wyniki dla różnych długości trajektorii:  $n = 5 \times 10^3$  (czerwony),  $n = 10^4$  (zielony) i  $n = 5 \times 10^4$  (niebieski) fotonów. Przedstawiono także odchylenia standardowe dla różnych wartości rozdzielnie i różnych długości trajektorii. Liniami ciągłymi oznaczono wartości minimalne i maksymalne.

dla 2 kolorów), w każdym przypadku dopasowania 2-kolorowe dają lepsze estymaty. Kiedy ( $I_1/I_h$ ) osiąga wartość 0,8, z dopasowania otrzymuje się wyniki obciążone (dla  $n = 5 \times 10^3$  oraz  $n = 10^4$ ), lub też ze znacznymi odchyleniami standardowymi (dla  $n = 5 \times 10^4$ ). Zarówno dla eksperymentów 1- jak i 2-kolorowych obserwowano znaczący wpływ długości trajektorii.

### 5.1.3.3 Podsumowanie rozdziału

Wykorzystując analizę ML możliwe jest odtworzenie parametrów kinetycznych, gdy rozkłady zliczeń fotonów w obu stanach nakładają się na siebie. Jakość estymowanych parametrów jest zależna od długości trajektorii, ale dla trajektorii  $n = 5 \times 10^4$  fotonów możliwe jest otrzymanie poprawnych estymat nawet gdy

stosunek intensywności w stanie niższym do intensywności w stanie wyższym ( $I_1/I_h = 0,7$ ). Dla takiej samej długości trajektorii, liczonej w liczbie cykli dopasowania ML do trajektorii 2-kolorowej charakteryzują się mniejszym rozrzutem statystycznym, niż dla 1-kolorowej. Natomiast biorąc pod uwagę długość trajektorii mierzoną jako liczba fotonów (w tym przypadku liczba cykli w trajektorii 2-kolorowej jest o połowę niższa, niż dla trajektorii 1-kolorowej), dla dobrego rozdzielania intensywności otrzymuje się nieznacznie lepsze wyniki dopasowując do trajektorii jednokolorowej. Natomiast gdy szybkości emisji fotonów w obu stanach stają się porównywalne ( $I_1/I_h = 0,6$ ), lepsze wyniki daje dopasowanie do trajektorii 2-kolorowej.

## 5.2 Modele 3- i 4- stanowe

---

W niniejszym rozdziale przedstawione zostaną wyniki symulacji i dopasowań modeli 3- i 4- stanowych, zarówno liniowych jak i cyklicznych. Zbadany zostanie wpływ liczby zarejestrowanych fotonów oraz średnia liczba pojawień się każdego ze stanów w trajektorii na jakość estymowanych parametrów. W modelach więcej niż 2-stanowe pojawia się zawsze pytanie jaki wpływ na jakość estymat ma liczba dopasowywanych parametrów. Także ten aspekt będzie badany w tym rozdziale.

---

Dopasowanie metodą ML wykonywane jest w wyniku przekształcania macierzy intensywności  $\mathbf{\Lambda}^D$ ,  $\mathbf{\Lambda}^A$  oraz generatora infinitezimalnego  $\mathbf{Q}$ . Dzięki temu możliwe jest rozpatrywanie modelu o dowolnej liczbie stanów i dowolnej topologii połączeń pomiędzy nimi. Jediną zmianą jaką należy wprowadzić w programie jest przygotowanie plików, w których rozmiary macierzy będą odpowiadały rozmiarom i wartościom parametrów modelu, który ma być symulowany i do którego mają być dopasowane parametry, w przypadku modeli cyklicznych w dopasowaniu należy uwzględnić warunek równowagi szczegółowej. W niniejszej pracy rozpatrywane będą modele 3- i 4- stanowe liniowe oraz cykliczne, przedstawione na rys. 4.3.

Przeprowadzono symulacje i dopasowania dla modeli 3- i 4- stanowych. W każdym stanie suma intensywności donora i akceptora była jednakowa ( $I_i^D + I_i^A = const$ ). Celem tej części badań było sprawdzenie, czy możliwe jest dopasowanie do modeli bardziej skomplikowanych, niż model 2-stanowy. Nie prowadzono tutaj badań dla szerokiego zakresu parametrów, lecz wybrano kilka przykładów w celu ilustracji jakości wyników. W planowaniu eksperymentów komputerowych wykorzystano wiedzę zdobytą na systematycznym badaniu modelu 2-stanowego.

Intensywności fluorescencji dobierano tak, by były one dobrze od siebie rozdzielone i tak, aby każdy stan był dostatecznie reprezentowany. Stałe kinetyczne przejść pomiędzy stanami były tak dobrane, by średni czas przebywania w danym stanie  $i$  był jednostkowy ( $1/\sum_{i \neq j} k_{ij} = 1$ ). Dzięki takiemu założeniu należy przyjąć, że udział każdego ze stanów w trajektorii jest jednakowy. Układ więc przebywał porównywalny czas w każdym ze stanów. Założenie takie sprawia, że możliwe

jest bezpośrednio porównywanie wyników otrzymanych dla dopasowań do modelu 2-stanowego z tymi, otrzymanymi dla bardziej skomplikowanych modeli.

Sprawdzono jak może wpływać liczba stopni swobody układu na jakość estymat. Należy bowiem pamiętać, że dla modelu 3L dopasowywane jest 10 parametrów (6 intensywności i 4 stałe szybkości), dla modelu 3C: 12 parametrów (6 intensywności i 6 stałych szybkości), dla modelu 4L: 14 parametrów (8 intensywności i 6 stałych szybkości) oraz dla modelu 4C: 16 parametrów (8 intensywności i 8 stałych szybkości). Następnie badano wpływ długości trajektorii na jakość estymowanych parametrów.

### 5.2.1 Wpływ liczby stopni swobody na jakość estymowanych parametrów

Pierwsza część eksperymentu miała na celu sprawdzenie jaki wpływ na jakość dopasowania ma liczba stopni swobody modelu. Parametry dobrano tak, by zminimalizować wpływ innych czynników na jakość dopasowania. W przypadku wszystkich symulacji sumaryczna liczba fotonów donora i akceptora na jeden stan wynosiła 120. Długość trajektorii dobierana była tak, by obserwować porównywalną liczbę pojawień się każdego ze stanów w trajektorii. Dla modelu 2-stanowego długość trajektorii wynosiła 24000 fotonów, dla modeli 3-stanowych wynosiła 36000 fotonów, a dla modeli 4-stanowych: 48000 fotonów. W tab. 5.13 przedstawiono wyniki dopasowań stałych kinetycznych dla modeli 3- i 4- stanowych oraz, dla porównania, modelu 2-stanowego.

We wszystkich przypadkach możliwe było odtworzenie parametrów modelu, a algorytm dopasowujący zbiegał się po mniej, niż  $5 \times 10^3$  iteracjach. Cztery rzeczy warte są zauważenia:

1. Intensywności dopasowywały się ze znacznie mniejszymi odchyleniami standardowymi (wszystkie wartości były mniejsze niż 0,05), przy czym wartość odchylenia zależała od wartości intensywności: im większa intensywność tym mniejsze odchylenie standardowe estymat.
2. W przypadku dopasowywania stałych kinetycznych z  $10^3$  eksperymentów, dla modelu 2-stanowego obserwowano znacznie mniejszy rozrzut statystyczny otrzymanych estymat (0,1), niż dla pozostałych modeli, odpowiednio:  $\sim 0,13$  dla 3L,  $\sim 0,17$  dla 3C,  $\sim 0,15$  dla 4L i  $\sim 0,17$  dla 4C. Ma to związek z

**Tabela 5.13:** Wartości średnie i odchylenia standardowe dla  $10^3$  znormalizowanych dopasowań, kiedy symulowane były modele 2, 3L, 3C, 4L, 4C. Długości trajektorii dobrano tak, by obserwować porównywalną liczbę pojawień się każdego ze stanów w trajektorii, a średnia liczba fotonów reprezentująca dany stan wynosiła 120, dla modeli 3L i 3C intensywności wynosiły  $I_1^D = 10$ ,  $I_2^D = 60$ ,  $I_3^D = 110$ ,  $I_1^A = 110$ ,  $I_2^A = 60$ ,  $I_3^A = 10$ , a dla 4L, 4C:  $I_1^D = 110$ ,  $I_2^D = 80$ ,  $I_3^D = 40$ ,  $I_4^D = 10$ ,  $I_1^A = 10$ ,  $I_2^A = 40$ ,  $I_3^A = 80$ ,  $I_4^A = 110$ .

model	stałe kinetyczne		intensywności				par. dop.
2	$\kappa_{12}$	1,01 (0,10)	$\iota_1^D$	1,000 (0,033)	$\iota_1^A$	1,000 (0,010)	6
	$\kappa_{21}$	1,01 (0,10)	$\iota_2^D$	1,0000 (0,0090)	$\iota_2^A$	1,000 (0,032)	
3L	$\kappa_{12}$	1,01(0,13)	$\iota_1^D$	0,999(0,038)	$\iota_1^A$	1,000(0,011)	10
	$\kappa_{21}$	1,00(0,13)	$\iota_2^D$	0,999(0,011)	$\iota_2^A$	1,000(0,011)	
	$\kappa_{23}$	1,01(0,13)	$\iota_3^D$	1,000(0,011)	$\iota_3^A$	0,999(0,040)	
	$\kappa_{32}$	1,01(0,13)					
3C	$\kappa_{12}$	1,00(0,17)	$\iota_1^D$	0,997(0,033)	$\iota_1^A$	1,0000(0,0090)	12
	$\kappa_{13}$	1,02(0,17)	$\iota_2^D$	1,000(0,014)	$\iota_2^A$	1,000(0,013)	
	$\kappa_{21}$	1,01(0,17)	$\iota_3^D$	1,000(0,010)	$\iota_3^A$	1,003(0,034)	
	$\kappa_{23}$	0,99(0,16)					
	$\kappa_{31}$	1,01(0,17)					
	$\kappa_{32}$	1,00(0,16)					
4L	$\kappa_{12}$	1,02(0,16)	$\iota_1^D$	1,000(0,048)	$\iota_1^A$	1,000(0,013)	14
	$\kappa_{21}$	1,01(0,15)	$\iota_2^D$	1,000(0,016)	$\iota_2^A$	1,000(0,011)	
	$\kappa_{23}$	1,02(0,14)	$\iota_3^D$	1,000(0,011)	$\iota_3^A$	1,000(0,016)	
	$\kappa_{32}$	1,01(0,14)	$\iota_4^D$	1,000(0,012)	$\iota_4^A$	0,998(0,046)	
	$\kappa_{34}$	1,02(0,16)					
	$\kappa_{43}$	1,02(0,16)					
4C	$\kappa_{12}$	1,01(0,17)	$\iota_1^D$	1,000(0,034)	$\iota_1^A$	1,000(0,010)	16
	$\kappa_{14}$	1,01(0,15)	$\iota_2^D$	1,000(0,018)	$\iota_2^A$	1,000(0,012)	
	$\kappa_{21}$	1,01(0,17)	$\iota_3^D$	1,001(0,012)	$\iota_3^A$	0,999(0,018)	
	$\kappa_{23}$	1,00(0,17)	$\iota_4^D$	1,000(0,010)	$\iota_4^A$	0,999(0,035)	
	$\kappa_{32}$	1,00(0,17)					
	$\kappa_{34}$	1,01(0,17)					
	$\kappa_{41}$	1,01(0,14)					
	$\kappa_{43}$	1,01(0,18)					

faktem, że w modelu 2-stanowym dopasowywanych było tylko 6 parametrów.

Natomiast dla modeli 3- i 4- stanowych liczba stopni swobody była większa, co spowodowało zwiększenie się rozrzutu statystycznego otrzymanych wyników.

3. Spośród modeli 3- i 4- stanowych najniższym rozrzutem estymat stałych szybkości charakteryzuje się model 3-stanowy liniowy (3L) i jest to ok. 0,13.
4. Wyniki dla dopasowania modeli 3C, 4L oraz 4C można uznać za porównywalne. Ich odchylenia standardowe wahały się od 0,14 do 0,18.

### 5.2.2 Wpływ długości trajektorii na jakość estymowanych parametrów

Przeprowadzono także symulacje dla modeli 3- i 4- stanowych dla różnych długości trajektorii liczonych jako sumaryczna liczba zarejestrowanych fotonów.

Tym razem przyjęto, że sumaryczna liczba fotonów na stan wynosi  $I_i = 80$ . Wartości dopasowanych intensywności były zdecydowanie lepsze od estymat stałych szybkości, dlatego też w tab. 5.14 przedstawiono tylko  $10^3$  znormalizowanych dopasowań dla stałych szybkości. Wszystkie dopasowania zbiegły się po mniej, niż  $10^4$  iteracjach. Kilka rzeczy wartych jest zauważenia:

1. Widoczny jest znaczny wpływ długości trajektorii na otrzymane wyniki. Dla przykładu rozpatrując model 3L odchylenia standardowe  $\kappa_{12}$  wynoszą odpowiednio 0,35; 0,22 oraz 0,10 dla  $n = 5 \times 10^3$ ,  $10^4$  i  $5 \times 10^4$ .
2. Dla modeli 3-stanowych, kiedy długość trajektorii wynosiła  $n = 10^4$  fotonów możliwa była estymacja parametrów z odchyleniem standardowym poniżej 0,3, przy czym wyniki dla modelu 3L były nieco lepsze. Dla  $n = 5 \times 10^4$  modele 3-stanowe estymowano z odchyleniem standardowym poniżej 0,15.
3. Dla modeli 4-stanowych estymaty dla  $n = 5 \times 10^3$  charakteryzowały się znacznym rozrzutem statystycznym, a niektóre wartości wykazywały obciążenie w górę. Dla  $n = 5 \times 10^4$  możliwa była estymacja parametrów kinetycznych z rozrzutem statystycznym poniżej 0,2.

**Tabela 5.14:** Wartości średnie i odchylenia standardowe dla  $10^3$  znormalizowanych dopasowań, kiedy symulowane były modele 3L, 3C, 4L, 4C dla różnych długości trajektorii  $n$ . Sumaryczna liczba fotonów na stan wynosiła  $I_i = 80$ . Intensywności dla 3L, 3C wynosiły:  $I_1^D = 70$ ,  $I_2^D = 40$ ,  $I_3^D = 10$ ,  $I_1^A = 10$ ,  $I_2^A = 40$ ,  $I_3^A = 70$ . Intensywności dla 4L, 4C wynosiły:  $I_1^D = 70$ ,  $I_2^D = 40$ ,  $I_3^D = 10$ ,  $I_4^D = 30$ ,  $I_1^A = 10$ ,  $I_2^A = 40$ ,  $I_3^A = 70$ ,  $I_4^A = 50$

model		$5 \times 10^3$	$10^4$	$5 \times 10^4$
3L	$\kappa_{12}$	1,08 (0,35)	1,04 (0,22)	1,01 (0,10)
	$\kappa_{21}$	1,04 (0,33)	1,02 (0,22)	1,00 (0,10)
	$\kappa_{23}$	1,06 (0,33)	1,04 (0,22)	1,002 (0,096)
	$\kappa_{32}$	1,09 (0,35)	1,05 (0,23)	1,005 (0,098)
3C	$\kappa_{12}$	1,04 (0,41)	1,01 (0,27)	1,00 (0,11)
	$\kappa_{13}$	1,04 (0,35)	1,01 (0,23)	1,00 (0,10)
	$\kappa_{21}$	1,05 (0,41)	1,02 (0,27)	1,01 (0,12)
	$\kappa_{23}$	1,06 (0,40)	1,04 (0,28)	1,00 (0,12)
	$\kappa_{31}$	1,02 (0,35)	1,02 (0,24)	1,00 (0,11)
	$\kappa_{32}$	1,04 (0,38)	1,03 (0,28)	1,00 (0,12)
4L	$\kappa_{12}$	1,15 (0,55)	1,07 (0,29)	1,01 (0,12)
	$\kappa_{21}$	1,08 (0,45)	1,05 (0,30)	1,01 (0,13)
	$\kappa_{23}$	1,07 (0,46)	1,04 (0,30)	1,01 (0,12)
	$\kappa_{32}$	1,09 (0,50)	1,06 (0,34)	1,00 (0,14)
	$\kappa_{34}$	1,11 (0,75)	1,07 (0,41)	1,01 (0,18)
	$\kappa_{43}$	1,22 (1,19)	1,08 (0,43)	1,02 (0,16)
4C	$\kappa_{12}$	1,11 (0,75)	1,05(0,45)	1,01 (0,20)
	$\kappa_{14}$	1,07 (0,65)	1,05 (0,44)	1,01 (0,18)
	$\kappa_{21}$	1,09 (0,66)	1,06 (0,47)	1,01 (0,17)
	$\kappa_{23}$	1,08 (0,70)	1,04 (0,41)	1,00 (0,16)
	$\kappa_{32}$	1,08 (0,71)	1,07 (0,49)	1,00 (0,20)
	$\kappa_{34}$	1,156 (1,024)	1,06 (0,52)	1,01 (0,21)
	$\kappa_{41}$	1,146 (2,080)	1,03 (0,39)	1,02 (0,15)
	$\kappa_{43}$	1,21 (1,97)	1,01 (0,45)	1,01 (0,18)

### 5.2.3 Podsumowanie rozdziału

Wykorzystując metodę ML możliwe jest odtworzenie parametrów kinetycznych modeli składających się z więcej niż dwóch stanów. Przy porównywalnej liczbie pojawień się każdego ze stanów w trajektoriach, takiej samej intensywności na stan i jednakowym średnim czasie przebywania w każdym ze stanów, obserwuje się nieznaczne zwiększenie rozrzutu statystycznego dla modeli z większą ilością stopni swobody. Ważnym czynnikiem wpływającym na jakość dopasowania jest próbkowanie stanów, czyli ile razy dany stan pojawił się w trajektorii. Znaczący wpływ na jakość dopasowań ma długość trajektorii. Dla  $n = 5 \times 10^4$  estymaty z  $10^3$  dopasowań charakteryzowały się odchyleniami standardowymi poniżej 0,2. Aby estymować poprawnie parametry modelu dla przedstawionych tutaj przykładów konieczne jest, by każdy ze stanów pojawił się w trajektorii ok. 50 razy, wtedy rozrzut statystyczny dopasowań nie powinien przekroczyć 0,3.

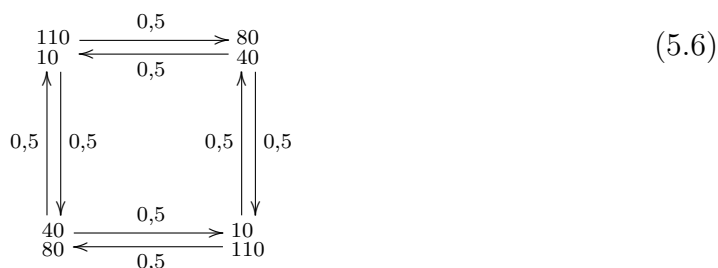
### 5.3 Parametry startowe użyte do dopasowania

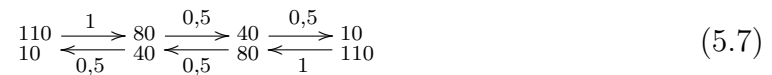
W niniejszym rozdziale badano wpływ parametrów startowych dopasowania na jakość otrzymywanych estymat. Wybrano modele "trudniejsze" do dopasowania, czyli z dużą liczbą stopni swobody, lub takie, dla których przypuszczano, że dopasowanie może zależeć od wyboru punktów startowych. Dopasowanie wykonywano używając parametrów startowych różniących się od wartości teoretycznych. Badano dwa aspekty. W pierwszym testowano przypadek, gdy wszystkie parametry są różne od teoretycznych. Natomiast w drugim zakładano, że znane są intensywności, a parametry startowe dopasowania stałych szybkości zmieniały się w szerokim zakresie.

W większości przypadków parametry startowe dla algorytmu dopasowującego były równe wartościom użytym w symulacji. Zakładać więc należy, że podane wyniki dopasowań są najlepszymi, jakie można otrzymać. Do wiedzy o wartościach parametrów modelu nie ma dostępu w przypadku rzeczywistego eksperymentu. Pierwszym krokiem w dopasowaniu jest wstępne oszacowanie rzędu wielkości estymowanych parametrów. Testowano jak dalekie od teoretycznych mogą być parametry startowe użyte do dopasowania, by w wyniku dopasowania metodą ML otrzymać wynik zgodny z rzeczywistością.

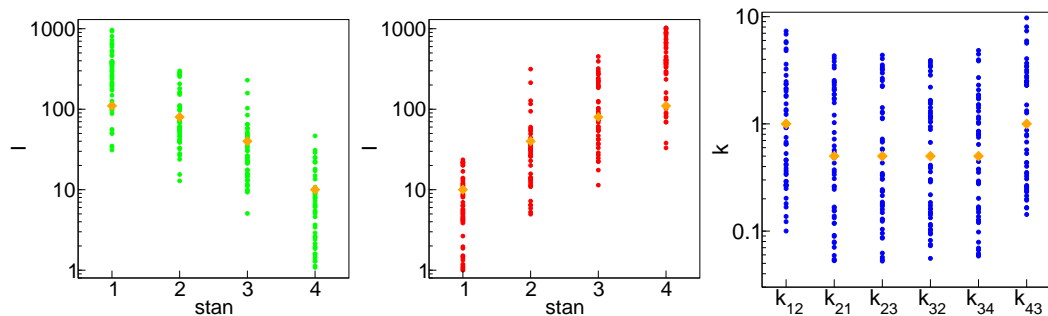
#### 5.3.1 Modele 4-stanowe

W pierwszym etapie jako model symulowany wybrano modele bardziej skomplikowane: 4L i 4C (5.6, 5.7), w których dopasowywano odpowiednio 14 i 16 parametrów:

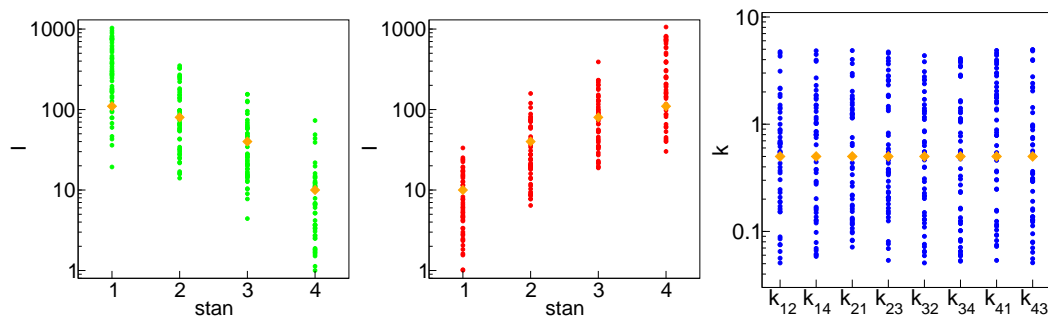




Wykonywano symulację, a następnie dopasowywano z 50 różnymi parametrami startowymi do tej samej trajektorii czasów międzydetekcyjnych. Parametry startowe dopasowania różniły się od tych użytych w symulacji maksymalnie o 1 rząd wielkości, a więc były maksymalnie 10 razy mniejsze lub większe od parametrów użytych do symulacji. Parametry startowe losowano z rozkładu jednostajnego. Po wylosowaniu startowych intensywności fluorescencji, sortowano je malejąco dla donora i rosnąco dla akceptora. Na rys. 5.13 przedstawiono parametry użyte do dopasowań modelu 4L (5.7), natomiast na rys. 5.14 parametry użyte do dopasowań modelu 4C (5.6).



**Rysunek 5.13:** Nieskalowane parametry startowe użyte do dopasowania modelu 4L (5.7). Na zielono oznaczono startowe intensywności donora, na czerwono: startowe parametry dla akceptora, na niebiesko: parametry startowe dopasowania stałych kinetycznych. Pomarańczowe punkty oznaczają wartości parametrów użyte do symulacji.

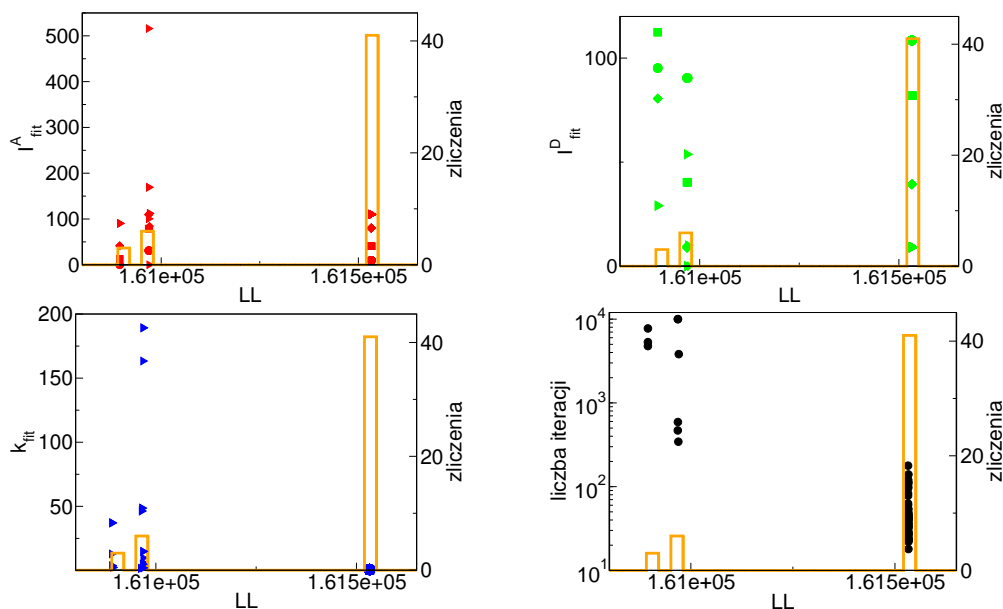


**Rysunek 5.14:** Nieskalowane parametry startowe użyte do dopasowania modelu 4C (5.6). Na zielono oznaczono startowe intensywności donora, na czerwono: startowe parametry dla akceptora, na niebiesko: parametry startowe dopasowania stałych kinetycznych. Pomarańczowe punkty oznaczają wartości parametrów użyte do symulacji.

Następnie wykonywano dopasowania. W tab. 5.15 i na rys. 5.15 przedstawiono wyniki dopasowań dla 50 różnych parametrów startowych przedstawionych na rys. 5.13 dla jednego zestawu danych symulowanego dla modelu 4L (5.7).

**Tabela 5.15:** Skalowane wyniki dopasowań stałych kinetycznych modelu 4L (5.7) dla jednego zestawu danych i 50 różnych zestawów parametrów startowych. “1” oznacza liczbę dopasowań, które zakończyły się daną wartością funkcji wiarygodności.

LL	1	$\kappa_{12}$	$\kappa_{21}$	$\kappa_{23}$	$\kappa_{32}$	$\kappa_{34}$	$\kappa_{43}$
161533	28	1,19	0,43	0,52	0,55	0,46	1,13
161532	13	1,19	0,43	0,52	0,55	0,46	1,11
160894	3	38,41	11,89	1,54	0,42	0,61	0,498
160969	1	0,49	0,64	0,45	1,13	0,00	49,41
160967	1	0,49	0,64	0,45	1,13	0,069	163,32
160968	1	0,49	0,64	0,45	1,18	0,40	9,79
160970	1	0,49	0,64	0,54	13,18	50,26	5,81
-	2	Algorytm nie zbiegł się po $10^4$ iteracji					



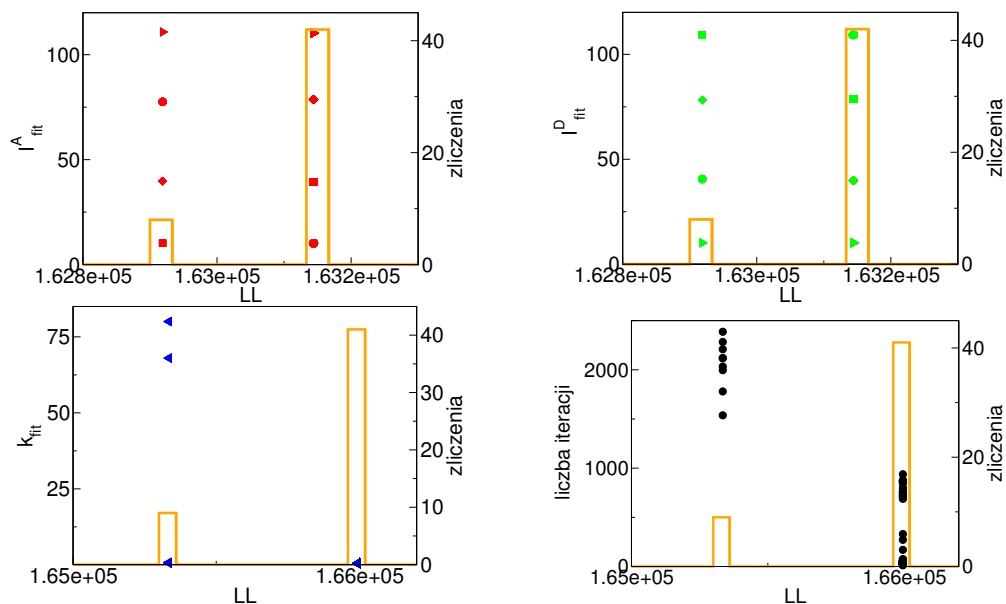
**Rysunek 5.15:** Wyniki dopasowań intensywności (donor: zielony, akceptor: czerwony) i stałych szybkości: niebieski dla różnych punktów startowych dla modelu 4L (5.7). Wartości parametrów oznaczono na lewej osi. Kółka: intensywności w stanie 1, kwadraty: w stanie 2, romby: w stanie 3 i trójkąty: w stanie 4. Pomarańczowy: histogram wartości funkcji wiarygodności dla 50 dopasowań z różnymi parametrami startowymi. Liczbę zliczeń przedstawiono na prawej osi.

Dla 2 spośród 50 zestawów parametrów algorytm nie zbiegł się po  $10^4$  iteracjach. Dla pozostałych 48 zestawów algorytm zbiegł się do 7 minimów o różnych wartościach funkcji wiarygodności. Dwie wartości funkcji wiarygodności, dla których otrzymano w sumie 41 dopasowań są bardzo bliskie sobie (LL =

161532 i 161533), a otrzymane parametry są bliskie teoretycznym. Dla pozostałych 7 dopasowań wartości funkcji wiarygodności były niższe (LL od 160970 do 160894), a wartości parametrów odbiegały od teoretycznych. W metodzie ML za najpoprawniejszy wynik uważa się ten, który charakteryzuje się najwyższą wartością funkcji wiarygodności. W przedstawionym tutaj przykładzie dla 41 zestawów punktów startowych algorytm zakończył się w tym samym miejscu, wartości funkcji wiarygodności były tutaj największe, a estymowane parametry bliskie teoretycznym.

**Tabela 5.16:** Skalowane wyniki dopasowań stałych kinetycznych modelu 4C (5.6) dla jednego zestawu danych i 50 różnych zestawów parametrów startowych. “l” oznacza liczbę dopasowań, które zakończyły się daną wartością funkcji wiarygodności.

LL	l	$\kappa_{12}$	$\kappa_{14}$	$\kappa_{21}$	$\kappa_{23}$	$\kappa_{32}$	$\kappa_{34}$	$\kappa_{41}$	$\kappa_{43}$
164227	38	0,68	0,67	0,49	0,50	0,64	0,42	0,69	0,46
164226	2	0,68	0,67	0,49	0,49	0,63	0,42	0,69	0,46
163979	8	0,47	1,00	0,51	1,17	0,80	0,54	1,059	0,77
163978	2	1,18	0,50	0,80	0,55	0,79	1,05	0,46	1,00



**Rysunek 5.16:** Wyniki dopasowań intensywności (donor: zielony, akceptor: czerwony) i stałych szybkości: niebieski dla różnych punktów startowych dla modelu 4C (5.6). Wartości parametrów oznaczono na lewej osi. Kółka: intensywności w stanie 1, kwadraty: w stanie 2, romby: w stanie 3 i trójkąty: w stanie 4. Pomarańczowy: histogram wartości funkcji wiarygodności dla 50 dopasowań z różnymi parametrami startowymi. Liczbę zliczeń przedstawiono na prawej osi.

W przypadku dopasowań do danych modelu 4C (5.6), kiedy algorytm dopasowujący rozpoczynano z 50 różnych, losowych parametrów startowych obserwowano 4 wartości funkcji wiarygodności. Wyniki dopasowań przedstawiono w tab. 5.16 oraz na rys. 5.16. Dwa maksima były sobie bliskie ( $LL = 164227$  i  $164226$ ), a estymaty miały wartości bliskie teoretycznym. Wyników takich otrzymano w sumie 40. W pozostałych 10 przypadkach algorytm dopasowujący kończył się w dwóch bliskich sobie maksimach, a estymowane parametry były dalekie od wartości teoretycznych. W tych przypadkach wartości funkcji wiarygodności były niższe od poprzednich ( $LL = 163979$  i  $163978$ ).

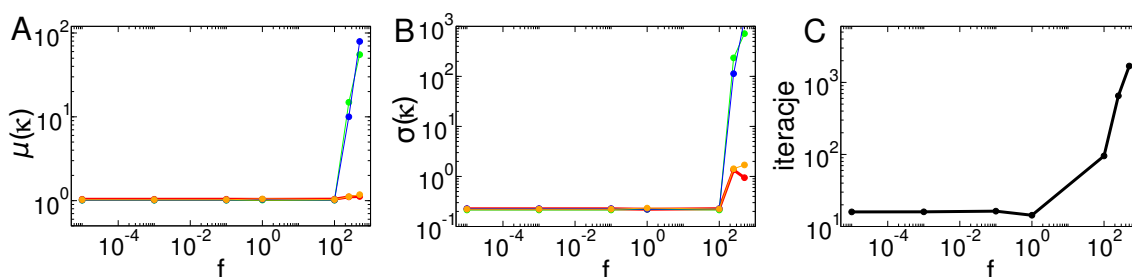
W przypadku dopasowania modeli 4L i 4C nie obserwowano istotnego wpływu wartości parametrów startowych na jakość estymowanych parametrów, gdy odbiegały one losowo od wartości teoretycznej o jeden rząd wielkości, a algorytm dopasowujący powtarzano 50 razy dla różnych parametrów. Na podstawie wartości funkcji wiarygodności możliwe było wybranie poprawnego wyniku dopasowania.

### 5.3.2 Model 3-stanowy

W rzeczywistym eksperymencie możliwe jest określenie z dużym prawdopodobieństwem wartości intensywności. Dlatego też w drugim etapie badań zakładano, że znane są intensywności fluorescencji w poszczególnych stanach. Tym razem wybrano model 3L (5.8). Przeprowadzono symulację, a następnie dopasowywano ten model dla szerokiego zakresu parametrów startowych.



Intensywności podawane algorytmowi dopasowującemu były równe użytym w symulacji, natomiast wszystkie stałe szybkości mnożono przez taki sam czynnik  $f$ , który zmieniał się w zakresie  $10^{-5}$  do  $5 \times 10^2$ . W tab. 5.17 oraz na rys. 5.17 przedstawiono wartości średnie i odchylenia standardowe oraz średnią liczbę eksperymentów, gdzie algorytm dopasowujący zbiegł się, a także średnią liczbę iteracji, po których zakończyło się dopasowanie.



**Rysunek 5.17:** Skalowane wyniki dopasowań modelu 3L dla różnych parametrów startowych, kiedy parametry startowe intensywności były równe parametrom symulacji, a startowe parametry stałych kinetycznych były równe parametrom symulacji mnożonym przez czynnik  $f$  ( $k_{\text{start}} = kf$ ).  $k_{12}$  - czerwony,  $k_{21}$  - zielony,  $k_{23}$  - niebieski,  $k_{32}$  - pomarańczowy.

Kiedy wartości intensywności podane do algorytmu dopasowującego były równe parametrom użytym w symulacji, a wartości startowe stałych kinetycznych były mniejsze od tych, użytych w symulacji ( $f < 1$ ), wartości średnie oraz odchylenia standardowe dopasowanych wartości pozostawały na tym samym poziomie. Nie zmieniała się także w znaczący sposób średnia liczba iteracji. Kiedy  $f > 1$ , czyli parametry startowe dopasowania były większe od tych, użytych w symulacji, średnia liczba iteracji zwiększała się, a parametry dopasowywane pozostawały stałe aż do  $f > 100$ . Kiedy wartości startowe dopasowania były więcej, niż  $10^2$  razy większe od teoretycznych obserwowano drastyczne pogorszenie się wartości parametrów

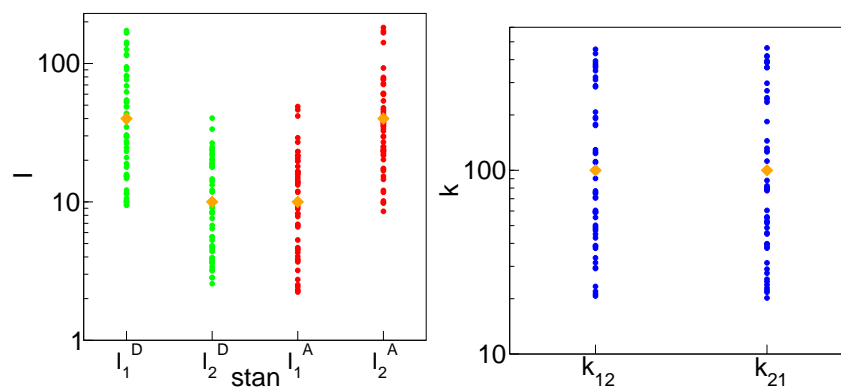
dopasowanych, a liczba iteracji zwiększyła się gwałtownie. Dla  $f > 100$  wartości startowe stałych kinetycznych były większe od średnich intensywności w danym stanie i być może dlatego algorytm dopasowujący zakańczył się w niewłaściwym maksimum.

**Tabela 5.17:** Wartości średnie i odchylenia standardowe z  $10^3$  skalowanych dopasowań różnych parametrów startowych modelu 3L (5.8), w eksperymencie 2-kolorowym. Każdorazowo jako parametry startowe dla intensywności były równe teoretycznym, natomiast wszystkie stałe szybkości mnożono przez czynnik  $f$ . W ostatniej kolumnie przedstawiono procentowy udział eksperymentów, w których algorytm zbiegł się po  $10^4$  iteracjach.

f	dopasowanie		średnia liczba iteracji	%zb
0,00001	$\kappa_{12}$	1,04 (0,23)	5,93 (4,83)	100
	$\kappa_{21}$	1,01 (0,21)		
	$\kappa_{23}$	1,02 (0,23)		
	$\kappa_{32}$	1,02 (0,22)		
0,001	$\kappa_{12}$	1,04 (0,23)	15,99 (7,31)	100
	$\kappa_{21}$	1,01 (0,21)		
	$\kappa_{23}$	1,02 (0,23)		
	$\kappa_{32}$	1,02 (0,22)		
0,1	$\kappa_{12}$	1,04 (0,23)	16,32 (13,41)	100
	$\kappa_{21}$	1,01 (0,21)		
	$\kappa_{23}$	1,02 (0,23)		
	$\kappa_{32}$	1,02 (0,22)		
100	$\kappa_{12}$	1,04 (0,23)	95,38 (44,10)	100
	$\kappa_{21}$	1,01 (0,21)		
	$\kappa_{23}$	1,02 (0,23)		
	$\kappa_{32}$	1,03 (0,22)		
250	$\kappa_{12}$	1,11 (1,35)	65,97 (1147,40)	96
	$\kappa_{21}$	14,89 (234,37)		
	$\kappa_{23}$	10,01 (113,20)		
	$\kappa_{32}$	1,12 (1,42)		
500	$\kappa_{12}$	1,12 (0,94)	1688,18 (1895,86)	81,4
	$\kappa_{21}$	55,20 (715,66)		
	$\kappa_{23}$	79,27 (1371,88)		
	$\kappa_{32}$	1,18 (1,70)		

### 5.3.3 Model 2-stanowy

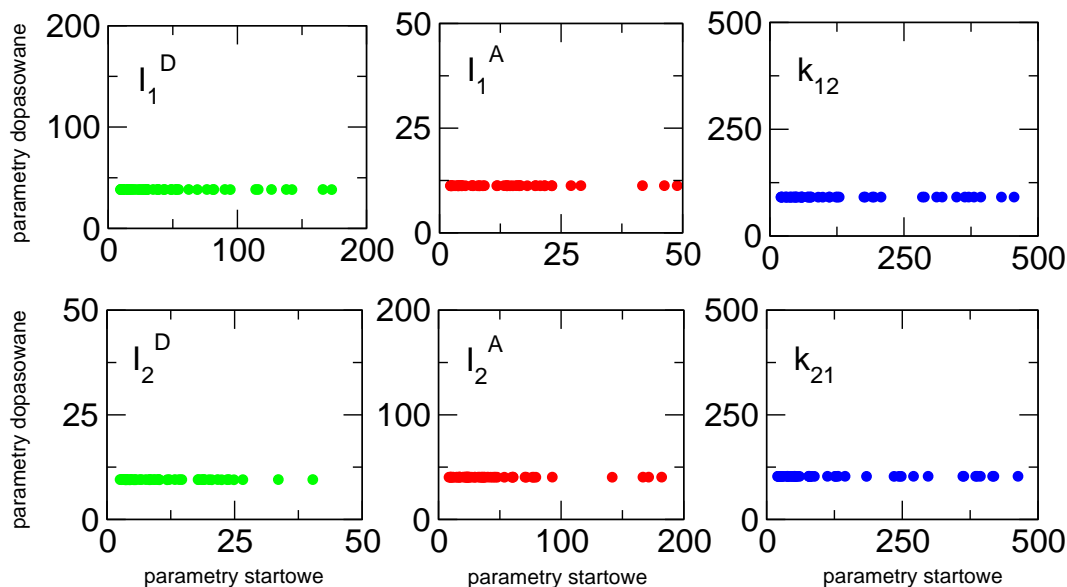
Wykonano także dopasowania z wykorzystaniem różnych parametrów startowych dla modelu 2-stanowego. Wybrano model (5.9), w którym stałe szybkości ucieczki z każdego ze stanów są większe od średnich intensywności w tych stanach. Dla tak dobranych parametrów obserwuje się średnio jeden foton na dwa stany. Długość trajektorii wynosiła  $5 \times 10^4$  fotonów, co oznacza, że obserwowano średnio  $5 \times 10^4$  cykli.



**Rysunek 5.18:** Nieskalowane parametry startowe użyte do dopasowania modelu 2-stanowego (5.9). Na zielono oznaczono startowe intensywności donora, na czerwono: startowe parametry dla akceptora, na niebiesko: parametry startowe dopasowania stałych kinetycznych. Pomarańczowe punkty oznaczają wartości parametrów użyte do symulacji.

Parametry startowe dopasowania losowano z rozkładu jednostajnego i różniły się maksymalnie pół rzędu wielkości od wartości teoretycznych. Były więc one maksymalnie pięć razy większe lub pięć razy mniejsze od wartości parametrów użytych do symulacji. Każdorazowo parametry startowe intensywności sortowano malejąco dla donora i rosnąco dla akceptora. Losowano 50 zestawów parametrów i dopasowywano je do tych samych danych. Na rys. 5.18 przedstawiono parametry, które użyto jako startowe do dopasowania.

Powtarzano eksperyment dla dziesięciu różnych zestawów danych i dla każdego zestawu uruchamiano algorytm dopasowujący z wylosowanymi parametrami startowymi. Wartości funkcji wiarygodności przy dopasowaniu do tych samych danych były jednakowe, a parametry dopasowane były takie same do drugiego miejsca znaczącego. Na rys. 5.19 przedstawiono zależność wartości otrzymanych estymat od parametrów użytych jako startowe dla algorytmu dopasowującego dla przykładowego zestawu danych.



Rysunek 5.19: Estymaty dla różnych parametrów startowych, dla modelu 5.9.

### 5.3.4 Podsumowanie rozdziału

Algorytm dopasowania metodą ML zbiega się do maksimum globalnego startując z parametrów różniących się znacznie od wartości teoretycznych. Nawet kiedy parametry startowe dopasowania były maksymalnie dziesięć razy większe, lub 10 razy mniejsze możliwa była poprawna estymacja parametrów modelu. Kiedy dopasowanie powtarzano 50 razy, dla różnych, losowo wybieranych parametrów startowych w większości przypadków algorytm osiągał maksimum globalne. Kiedy dopasowanie kończyło się w jednym z maksimum lokalnych, otrzymywano wartości parametrów różniące się znacznie od teoretycznych, ale jednocześnie dla tych przypadków wartość funkcji wiarygodności była niższa, co pozwala uznać te wyniki jako mało prawdopodobne.

Znając wartość intensywności w każdym ze stanów możliwe jest odtworzenie parametrów kinetycznych modelu nawet w przypadku, gdy wszystkie stałe szybkości podane do algorytmu dopasowującego były o 5 rzędów wielkości mniejsze lub o 2 rzędy wielkości większe od teoretycznych. Kiedy parametry startowe stałych szybkości były znacznie większe od intensywności, niemożliwe było odtworzenie poprawnych parametrów modelu.

Nie obserwowano także wpływu parametrów startowych na jakość estymowanych parametrów, kiedy w symulowanym modelu wartości stałych szybkości były dwa razy większe, niż intensywności fluorescencji w danym stanie, a parametry startowe dopasowania różniły się o pół rzędu wielkości.

Korzystając z estymacji metodą ML możliwe jest odtworzenie poprawnych parametrów modelu nawet gdy wszystkie parametry różnią się od użytych w symulacji losowo o 1 rząd wielkości. Kiedy znane są intensywności, stałe kinetyczne użyte jako punkty startowe dopasowania mogą różnić się od użytych w symulacji o kilka rzędów wielkości. Prawdopodobieństwo wyboru poprawnych parametrów zwiększa także fakt, że możliwa jest ocena jakości odtworzonych parametrów poprzez śledzenie wartości funkcji wiarygodności. Powtarzając wielokrotnie dopasowanie z różnymi parametrami startowymi należy wybrać te dopasowania, dla których wartość funkcji wiarygodności była największa.

## 5.4 Rozróżnianie modeli kinetycznych

---

W niniejszym rozdziale przedstawione będą przykłady rozróżniania modeli z wykorzystaniem metody opartej na maksymalizacji funkcji wiarygodności. Do wyboru najbardziej prawdopodobnego modelu wykorzystano kryteria AIC (ang. *Akaike information criterion*) oraz BIC (ang. *Bayesian information criterion*). Pokazane zostaną przykłady, a także wpływ długości trajektorii na efektywność rozróżniania modeli.

---

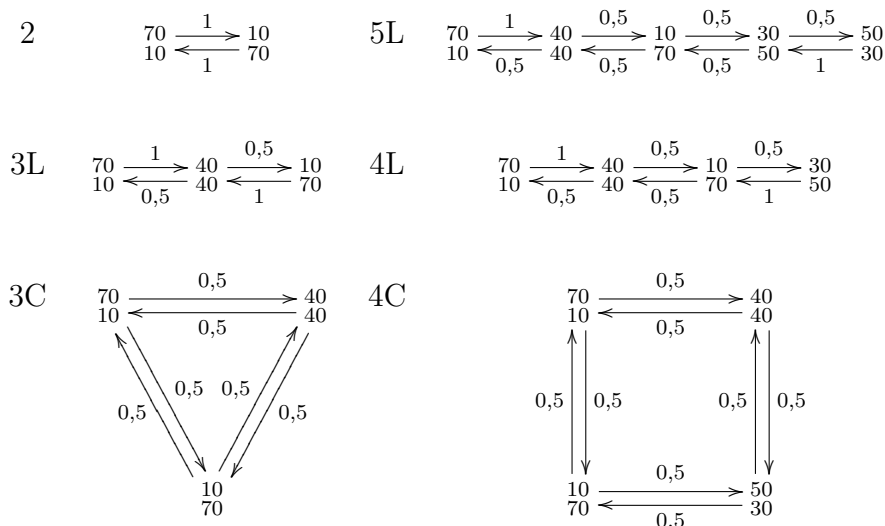
Procedura wyboru najbardziej prawdopodobnego modelu wyglądała w następujący sposób:

1. Symulacja trajektorii czasów międzydetekcyjnych dla wybranego modelu
2. Dopasowanie do otrzymanej trajektorii czasów międzydetekcyjnych modeli o różnej liczbie stanów i różnej topologii
3. Obliczenie AIC i BIC na podstawie otrzymanych wyników dopasowań
4. Wybór najbardziej prawdopodobnego modelu (posiadającego najniższą wartość AIC lub BIC) dla symulowanych danych

### 5.4.1 Przykładowy eksperyment rozróżniania modeli

Wykonano eksperyment, w którym symulowano modele 2, 3L, 3C, 4L, 4C oraz 5L, których schematy przedstawiono na rys. 5.20, dla długości trajektorii  $n = 5 \times 10^4$  fotonów, a następnie do otrzymanych trajektorii dopasowywano wszystkie wymienione wyżej modele. Parametry dobierano tak, by średnie czasy przebywania w każdym ze stanów były równe i wynosiły 1. Po zakończeniu algorytmu dopasowującego dla wszystkich modeli, z wartości funkcji wiarygodności otrzymanej na podstawie dopasowania obliczano wartości AIC i BIC, a następnie jako najbardziej prawdopodobny model wybierano ten z najniższą wartością IC. Przykładowe wyniki dla BIC przedstawiono na rys. 5.21. Dla ujednocnienia wyników na wykresach przedstawiono względne wartości BIC w postaci różnicy pomiędzy

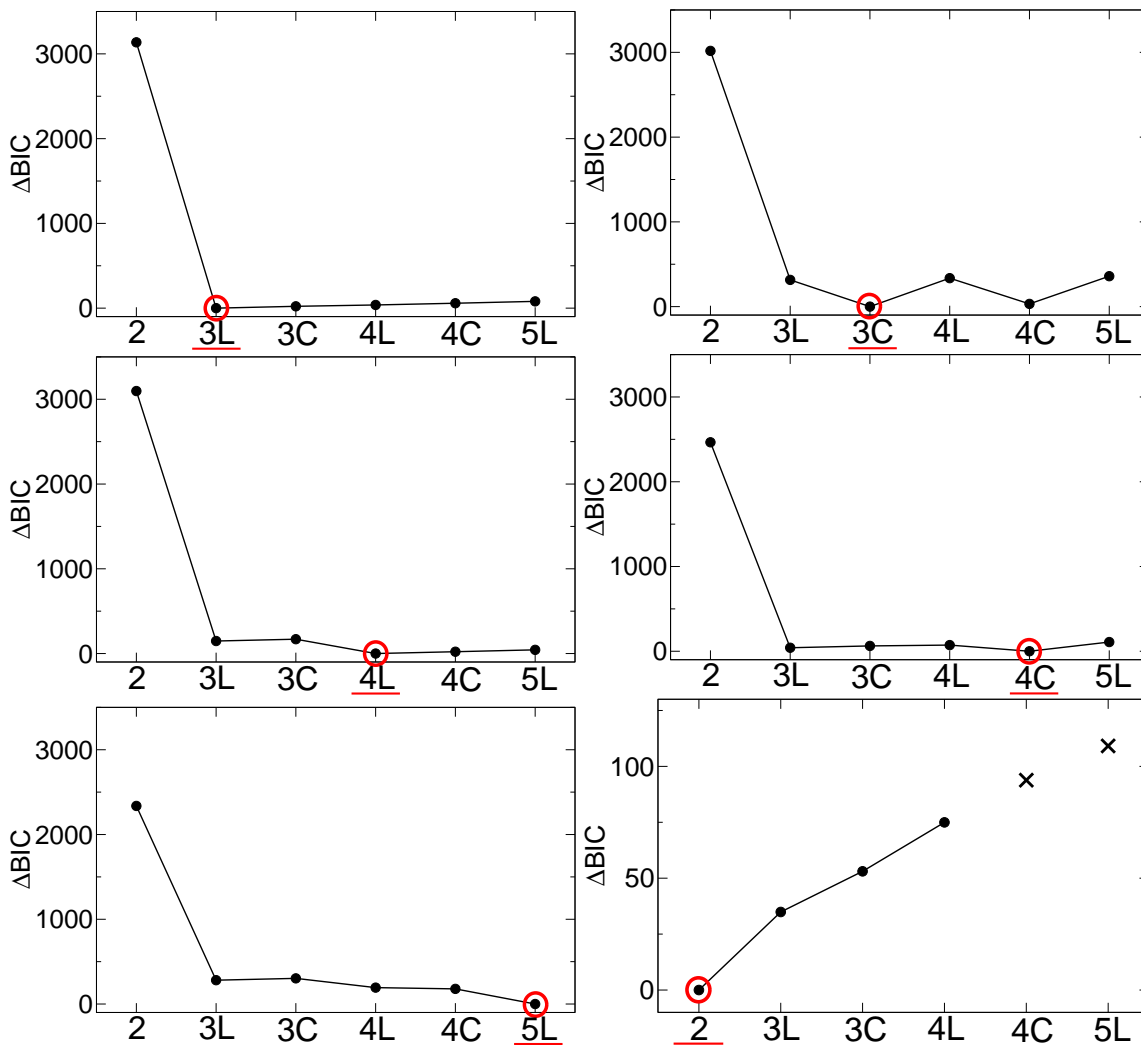
**Rysunek 5.20:** Modele kinetyczne, które wykorzystano w analizie rozróżniania modeli. W każdym przypadku sumaryczna liczba fotonów na stan wynosiła  $I_i = 80$ , a średnie czasy przebywania w każdym ze stanów były jednostkowe.



poszczególnymi BIC, a wartością najmniejszą otrzymaną dla dopasowań różnych modeli do danej trajektorii czasów międzydetekcyjnych.

We wszystkich przypadkach przy pomocy BIC wybrano model poprawny jako najbardziej prawdopodobny. Kiedy symulowany model był bardziej skomplikowany niż 2-stanowy, wartość BIC dla tego modelu była znacznie większa od pozostałych dopasowań. W przypadku modeli liniowych obserwowano spadek wartości BIC, a następnie niewielkie wypłaszczenie po osiągnięciu minim. Gdy symulowano model 3C, w każdym przypadku wartości BIC dla 3C i 4C miały najbardziej zbliżone wartości spośród wszystkich pozostałych. Rozróżniać można więc nie tylko pomiędzy modelami o różnej liczbie stanów, ale także pomiędzy modelami posiadającymi taką samą liczbę stanów, lecz różną topologię.

**Rysunek 5.21:** Przykłady rozróżniania modeli, kiedy symulowano modele 2, 3L, 3C, 4L, 4C oraz 5L i wszystkie te modele dopasowywano dla otrzymanej trajektorii czasów międzydetekcyjnych. W każdym przypadku model symulowany podkreślono, a wybrany przez BIC zaznaczono okręgiem. Kiedy symulowany był model 2 algorytm dopasowujący nie zbiegł się po  $10^4$  iteracjach, na rysunku zaznaczono to krzyżykami.



### 5.4.2 Wpływ długości trajektorii na rozróżnianie modeli kinetycznych

Wykonano także eksperyment, w którym symulowano modele 3L, 3C, 4L oraz 4C i dopasowywano modele 2, 3L, 3C, 4L, 4C oraz 5L dla różnych długości trajektorii:  $n = 5 \times 10^3$ ,  $10^4$  oraz  $5 \times 10^4$ . Każdorazowo dany model dopasowywano wykorzystując dwa różne zestawy parametrów startowych. Spośród dwóch wyników dopasowań tego samego modelu z różnymi parametrami startowymi wybierano ten wynik, który charakteryzował się wyższą wartością funkcji wiarygodności i to ta wartość używana była w etapie rozróżniania modeli. W tab. 5.18 przedstawiono, ile razy kryteria AIC oraz BIC pozwoliły wybrać poprawny model w 100 eksperymentach.

**Tabela 5.18:** Statystyka rozróżniania pomiędzy modelami 2-, 3- i 4-stanowymi. Liczba eksperymentów, w których kryteria AIC oraz BIC pozwoliły wybrać poprawny model w 100 eksperymentach. Symulowano modele 3L, 3C, 4L, 4C dla różnych długości trajektorii i dopasowywano te modele, a także model 2-stanowy i 5L.

kryterium model symulowany	AIC				BIC			
	3L	3C	4L	4C	3L	3C	4L	4C
$n = 5 \times 10^3$	83	75	81	42	100	87	1	0
$n = 10^4$	86	90	96	90	100	100	44	3
$n = 5 \times 10^4$	85	87	98	100	100	100	100	97

Kilka kwestii wartych jest zauważenia:

1. Dla trajektorii o długości  $n = 5 \times 10^4$  zarówno przy użyciu AIC jak i BIC otrzymywano zadowalające wyniki (ponad w 85 przypadkach został wybrany poprawny model). W większości eksperymentów dla tej długości trajektorii lepsze wyniki dawało BIC, wyjątkiem jest sytuacja, w której symulowano model 4C. Wtedy to BIC wybrał poprawny model w 9 przypadkach, natomiast AIC w 100 przypadkach.
2. Większą stabilnością wyboru poprawnego modelu wykazuje się AIC, niezależnie od liczby stanów i topologii modeli. Możliwe było wybranie poprawnego modelu jako najbardziej prawdopodobny w ponad 75% eksperymentów. Wyjątek stanowi przypadek, gdy symulowano model 4C dla  $n = 5 \times 10^3$ ; poprawny model został wówczas wybrany jedynie 42 razy. Kiedy symulowano model 4C dla  $n = 5 \times 10^4$ , poprawny model został wybrany we wszystkich przypadkach.

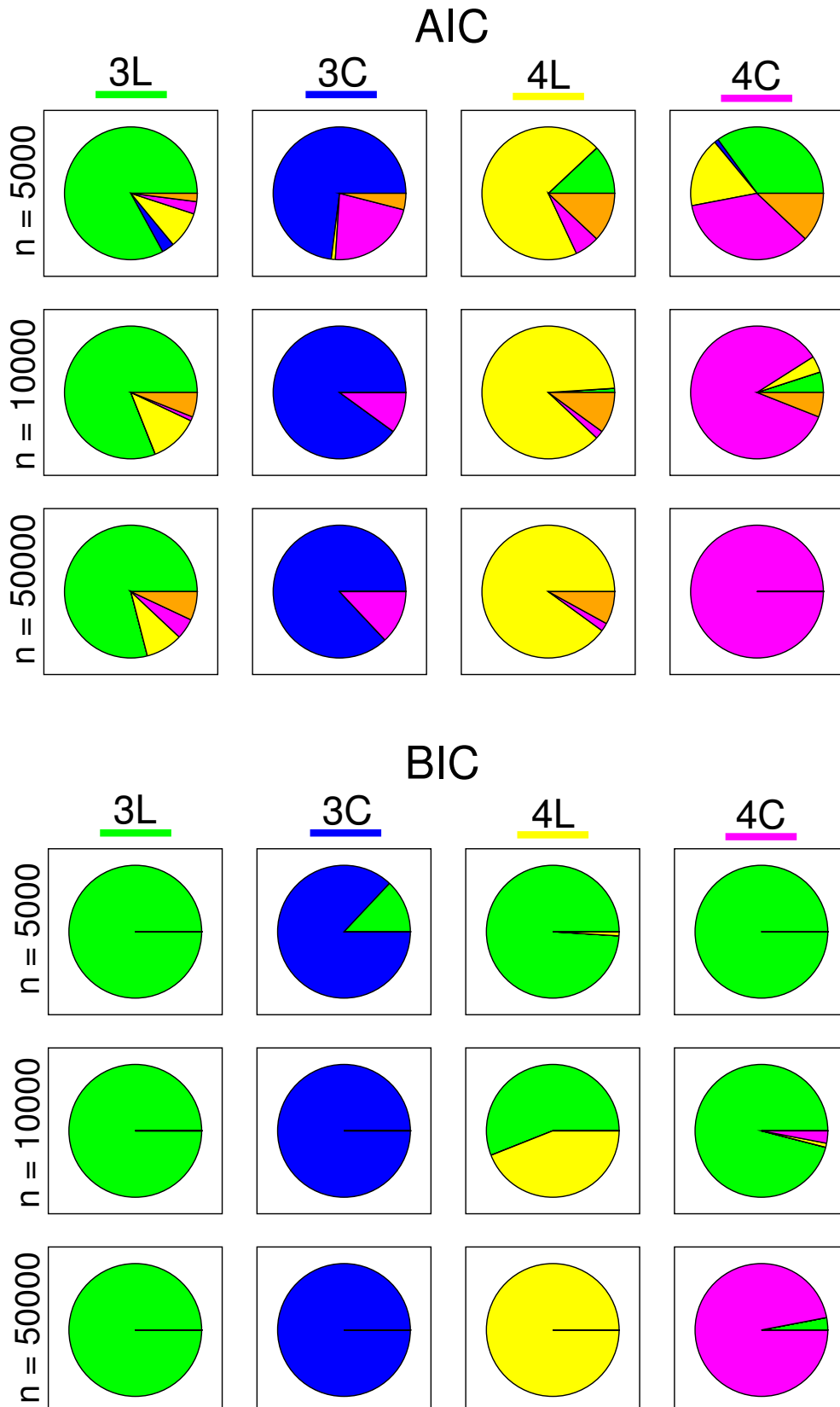
3. Kiedy do rozróżniania modeli wybrano kryterium BIC, trafność rozróżniania modeli była różna: od zera poprawnych wyborów dla 4C,  $n = 5 \times 10^3$  do 100 kiedy symulowano model 3L, dla każdej długości trajektorii. Kryterium BIC daje możliwość poprawnego wyboru modeli 3-stanowych nawet, gdy długość trajektorii wynosi  $n = 5 \times 10^3$  fotonów. Nie daje natomiast poprawnych wyników, kiedy symulowany model jest 4-stanowy, a trajektoria jest krótsza niż  $5 \times 10^4$  fotonów.

Na rys. 5.22 przedstawiono jakie modele były wybierane w eksperymentach, których wyniki przedstawiono w tab. 5.18. Na rysunku zaznaczono, który z modeli był symulowany. Kolorami oznaczano modele, które wybrane były jako najbardziej prawdopodobne przez poszczególne kryteria. Na zielono oznaczono odsetek wybranych modeli 3L, na niebiesko: odsetek wybranych modeli 3C, na żółto: modeli 4L, na różowo: modeli 4C i na pomarańczowo: 5L. W każdym przypadku dopasowywano także model 2-stanowy, ale nie został on ani razu wybrany jako najbardziej prawdopodobny.

W przypadku AIC kiedy symulowana trajektoria była krótka  $n = 5 \times 10^3$  fotonów, z dużym prawdopodobieństwem wybierany był model poprawny, ale także każdy z pozostałych mógł być uznany za najlepszy. Kiedy natomiast wydłużano trajektorię do  $n = 10^4$ , gdy symulowano model 3L, w większości przypadków AIC wybierało model poprawny, a w większości pozostałych dopasowań za najbardziej prawdopodobny uznawany był model 4L, lub 5L. Tak samo działo się w przypadku modelu 3C: dla większości eksperymentów wybrany był poprawny model, natomiast w pozostałych przypadkach za najbardziej prawdopodobny uznawany był model 4C. Tak więc AIC ma tendencję do wyboru za najbardziej prawdopodobny model o poprawnej topologii, czasami wybór pada na model o takiej samej topologii, lecz większej liczbie parametrów.

Kiedy do rozróżniania modeli wybrano kryterium BIC, dawało ono bardzo dobre wyniki w przypadku długiej trajektorii ( $n = 5 \times 10^4$  fotonów) oraz dla modeli 3-stanowych. W pozostałych przypadkach uprzywilejowany był model 3L i to on był najczęściej wybierany za najbardziej prawdopodobny, nawet jeśli symulowano model 4L lub 4C. W żadnym przypadku nie został wybrany model 5L.

**Rysunek 5.22:** Modele wybierane jako najbardziej prawdopodobne z wykorzystaniem kryteriów AIC i BIC dla różnych modeli i różnych długości trajektorii. Na zielono zaznaczano ile razy wybrany został model 3L, na niebiesko: 3C, na żółto: 4L i na różowo: 4C, na pomarańczowo: 5L.



### 5.4.3 Wpływ parametrów startowych

W celu zbadania wpływu parametrów startowych na trafność rozróżniania modeli kinetycznych przeprowadzono symulację wybranego modelu, a następnie dopasowanie różnych modeli do otrzymanych danych. Dopasowanie każdego z modeli powtarzano 20 razy, dla różnych parametrów startowych. Jako modele do symulacji w tym eksperymencie wybrano modele 3-stanowe, które dobrze rozróżniały się we wcześniejszych eksperymentach. Długość trajektorii wynosiła  $n = 10^4$ . Symulowano modele 3C i 3L oraz dokonywano dopasowania dla modeli 2, 3C oraz 3L. Spośród 20 dopasowań z użyciem różnych parametrów startowych wybierano to, które charakteryzowały się najwyższą wartością funkcji wiarygodności i tej wartości używano w późniejszym rozróżnianiu modeli. Otrzymane wyniki były porównywalne z tymi, otrzymanymi dla rozróżniania modeli z wykorzystaniem dwóch różnych zestawów parametrów.

### 5.4.4 Podsumowanie rozdziału

Dla krótszych trajektorii AIC pozwala rozróżniać szerszy zakres modeli kinetycznych, niż BIC. Jednak dla rozróżniania modeli 3-stanowych lepsze wyniki otrzymuje się dla BIC. AIC ma tendencję do wybierania modeli o poprawnej topologii, ale bardziej skomplikowanych, natomiast BIC do wyboru modelu 3L. Nie obserwowano znaczącego wpływu parametrów startowych dopasowania na trafność wyboru modeli kinetycznych. Dla trajektorii o długości  $n = 5 \times 10^4$ , korzystając z dowolnego kryterium rozróżniania modeli możliwy jest wybór modelu prawdziwego jako najbardziej prawdopodobnego.

# Rozdział 6

## Podsumowanie i wnioski

W niniejszej pracy sprawdzano możliwość zastosowania metody największej wiarygodności (ML) w estymacji parametrów kinetycznych i rozróżnianiu modeli w eksperymentach FRET pojedynczych, unieruchomionych cząsteczek. Parametry symulacji dobierano w oparciu o reakcje rybozymów. Wyniki otrzymane metodą ML porównywano z tymi, otrzymywanymi dla najczęściej stosowanej przez eksperymentalistów analizy on-off.

Stworzono i walidowano programy służące do symulacji trajektorii czasów międzydetekcyjnych metodą MMMPP, a następnie badano wpływ różnych czynników na jakość estymowanych parametrów oraz porównywano otrzymane wyniki z estymatami uzyskanymi przy użyciu analizy on-off. Sprawdzano także czy wykorzystując metodę ML możliwa jest estymacja parametrów kinetycznych dla modeli wielostanowych mających układ liniowy lub też cykliczny. W większości symulacji jako parametry startowe dopasowania wykorzystywano wartości teoretyczne zakładając, że wyniki tak uzyskane będą najlepszymi dla danego zestawu parametrów.

Badano także, jaki wpływ na jakość estymat mają parametry startowe dopasowania. Symulowano dane i dopasowywano do nich model rozpoczynając od punktów coraz bardziej odległych od wartości teoretycznych.

Przeprowadzono eksperymenty mające na celu wybór najbardziej prawdopodobnego modelu kinetycznego spośród kilku zaproponowanych. W rozróżnianiu modeli kinetycznych wykorzystano wartość funkcji wiarygodności uzyskanej w wyniku dopasowania i obliczano wartości AIC lub BIC, a następnie na podstawie kryteriów AIC i BIC wybierano model. Rozróżniano pomiędzy modelami

liniowymi o różnej liczbie stanów, a także pomiędzy modelami składającymi się z takiej samej liczby stanów, ale o różnej topologii.

Z przeprowadzonych doświadczeń komputerowych wyciągnięto następujące wnioski:

- możliwe jest odzyskiwanie parametrów modelu nawet dla krótkich ( $n = 5 \times 10^3$  fotonów) trajektorii
- na jakość dopasowania duży wpływ ma długość trajektorii
- na jakość estymowanych intensywności ma wpływ tylko fakt, ile fotonów danego koloru było zarejestrowanych w konkretnym stanie
- kiedy obserwowana jest mała liczba cykli (do ok.  $10^3$ ) i jedna ze stałych szybkości jest wolna (około rząd wielkości mniejsza niż szybkość emisji fotonów), możliwa jest estymacja parametrów kinetycznych, kiedy wartość drugiej stałej szybkości jest rzędu szybkości emisji fotonów. Kiedy obserwuje się dużą liczbę cykli (rzędu  $10^4$  i większą), obie stałe szybkości mogą być nawet dwukrotnie wyższe od szybkości detekcji fotonów, by otrzymać poprawne estymaty stałych kinetycznych
- najlepsze wyniki otrzymuje się, gdy szybkość emisji fotonów jest 5 do 10 razy większa od szybkości zmian pomiędzy stanami
- możliwe jest poprawne odtworzenie parametrów w przypadku, kiedy intensywności fluorescencji w poszczególnych stanach mają bliskie sobie wartości (czasy pomiędzy rejestracją kolejnych fotonów w poszczególnych stanach są porównywalne). Im dłuższa trajektoria tym intensywności mogą być bliższe sobie. Np. dla  $5 \times 10^4$  fotonów parametry kinetyczne odtwarzane są poprawnie nawet dla  $I_1/I_h = 0,6$
- możliwe jest także odzyskiwanie parametrów dla bardziej niż 2-stanowe skomplikowanych modeli. Rozrzut statystyczny otrzymanych dopasowań zależy między innymi od ilości dopasowywanych parametrów i długości trajektorii.
- wykorzystując kryteria AIC i BIC możliwe jest prawidłowe wybranie najbardziej prawdopodobnego modelu dla trajektorii liczących  $10^4$  fotonów.

- możliwe jest rozróżnianie zarówno pomiędzy modelami o różnej liczbie stanów, jak i pomiędzy modelami składającymi się z takiej samej liczby stanów, ale o różnej topologii
- nie obserwowano istotnego wpływu parametrów startowych na jakość estymowanych wartości. Parametry startowe mogą odbiegać o rząd wielkości od wartości teoretycznych, ale powtarzając dopasowanie dla jednego zestawu danych, rozpoczynając algorytm z 50 różnych parametrów startowych, w większości przypadków otrzymuje się wynik bliski teoretycznemu. Dla przypadków, gdy wartości estymat odbiegają od tych, użytych do generacji danych, wartość funkcji wiarygodności jest niższa, co pozwala odrzucić te wyniki jako najbardziej prawdopodobne

Metoda dopasowania do histogramów czasów on-off jest skuteczna, gdy intensywności fluorescencji są dobrze rozdzielone i obserwuje się dostatecznie dużą populację stanów on i off. Średnia długość czasu przebywania w każdym ze stanów musi być kilkakrotnie większa od szerokości okna czasowego, użytego do obliczenia trajektorii intensywności fluorescencji. Istotną rzeczą jest odpowiedni dobór położenia progu rozdzielającego stany, a rozdzielczość odtwarzanych parametrów kinetycznych jest zależna od szerokości okna czasowego w trajektorii intensywności fluorescencji. W przypadku niedostatecznego rozdzielania stanów, wygładzanie trajektorii może okazać się pomocne i zwiększyć zakres, w którym możliwa jest poprawna estymacja parametrów.

Metoda ML daje większe możliwości w porównaniu z analizą on-off. Przede wszystkim do analizy wykorzystywana jest bezpośrednio trajektoria zliczeń fotonów. Nie ma tu konieczności łączenia fotonów w klasy, co pozwala uniknąć poszukiwania optymalnej szerokości okna czasowego, a także nie powoduje zmniejszenia się rozdzielczości metody, tak jak to jest w przypadku metod korzystających z trajektorii fluorescencji. Unika się także problemu z doбором progu rozdzielającego stany o różnej intensywności, konieczności wyboru optymalnej wartości progowej i ewentualnego wygładzania trajektorii.

ML charakteryzuje się dużą rozdzielczością. Stosując tę metodę możliwe jest odtworzenie parametrów kinetycznych modelu, które nie są możliwe do zaobserwowania w analizie on-off, jak to było dla rybozemu typu hairpin w [15]. Dla dłuższych trajektorii (liczących ok.  $10^4$  cykli) możliwa jest estymacja parametrów kinetycznych nawet, gdy są one dwa razy wyższe od szybkości emisji fotonów fluorescencji. Jednocześnie dla krótszych trajektorii wystarczające jest, by

zaobserwować 5-10 fotonów fluorescencji w czasie przebywania układu w danym stanie, by z dużą dokładnością oszacować szybkość ucieczki z tego stanu, dla analizy on off jest to ok. 50. Metoda ML nie wymaga długich trajektorii, a intensywność fluorescencji może być niska w stosunku do tej, jaka wymagana jest dla analizy on-off.

Parametry startowe dopasowania mogą losowo nawet o rząd wielkości różnić się od użytych w symulacji, a mimo to możliwe jest poprawne odtworzenie parametrów. Jeśli jako parametry startowe użyje się intensywności bliskie stosowanym w symulacjach, to startowe wartości stałych szybkości mogą różnić się znacznie (nawet kilka rzędów wielkości) od wartości prawdziwych, a pomimo to algorytm dopasowujący z dużym prawdopodobieństwem zakończy dopasowanie w minimum globalnym i dzięki temu odtworzone zostaną poprawne parametry kinetyczne. Kolejną zaletą ML jest fakt, że możliwe jest dopasowanie do modeli wielostanowych, o dowolnej topologii (liniowych, cyklicznych i innych).

W połączeniu z AIC i BIC, ML daje możliwość rozróżniania modeli i pozwala z dużym prawdopodobieństwem wybrać model najbardziej prawdopodobny spośród proponowanych.

Przeprowadzone w niniejszej rozprawie badania wykazują, że zaproponowana przez Szabo [22] analiza danych dopasowująca metodą największej wiarygodności do trajektorii czasów międzydetekcyjnych może znaleźć zastosowanie w estymacji parametrów kinetycznych w spektroskopii FRET pojedynczych cząsteczek. Dzięki wielu zaletom, metoda ML zdaje się otwierać nowe ścieżki w analizie danych otrzymanych z eksperymentów FRET pojedynczych, unieruchomionych cząsteczek i pozwoli analizować dane ze znacznie większą, niż dotychczas rozdzielczością.

Analiza metodą ML wykonywana jest z wykorzystaniem rachunku macierzowego i można rozszerzyć ją na analizę dla trzech kolorów fotonów fluorescencji, lub bardziej złożonych modeli i w prosty sposób przeprowadzać analizę danych w eksperymentach FRET pojedynczych cząsteczek dla rybozymów i innych cząstek biologicznych. Dzięki symulacjom komputerowym możliwe jest lepsze planowanie eksperymentów przeprowadzanych na poziomie pojedynczych cząsteczek. Jeśli pomiary dokonane będą z wykorzystaniem mikroskopu konfokalnego i detektora punktowego, pozwalającego mierzyć trajektorię zliczeń fotonów, wykorzystując metodę ML można estymować szybkości przejść o wartościach większych od intensywności fluorescencji. Ponieważ w metodzie ML do przeprowadzenia poprawnej analizy wystarczające jest detekcja ok. 5-10 fotonów na czas życia stanu, możliwe jest

użycie niższej intensywności wzbudzenia, a dzięki temu wydłużenie czasu obserwacji badanej molekuly.



# Bibliografia

- [1] S. Weiss. *Science*, 283:1676, 1999.
- [2] A. Gershenson, L. M. Gierasch. *Curr. Op. Str. Biol.*, 21:32, 2010.
- [3] I. Rasnik, S. A. McKinney, T. Ha. *Nat. Methods*, 3:891, 2006.
- [4] T. Forster. *Z. aturforsch. A*, 4:321, 1949.
- [5] T. Forster. *Academic Press*, New York:93, 1967.
- [6] S. Myong, B. C. Stevens, T. Ha. *Structure*, 14:633, 2007.
- [7] X. Zhuang, H. Kim, M. J. B. Pereira, H. P. Babcock, N. G. Walter, S. Chu. *Science*, 296:1473, 2002.
- [8] X. Zhuang, L. E. Bartley, H P. Babcock, R. Russel, T. Ha, D. Herschlag, S. R. Chu. *Science*, 288:2048, 2000.
- [9] B. Schuler. *Biophys. J.*, 100:373a, 2011.
- [10] Z. Xie, N. Srividya, T. R. Sosnick, T. Pan, N. F. Scherer. *PNAS*, 101:534, 2004.
- [11] D. Rueda, G. Bokinsky, M. M. Rhodes, M. J. Rust, X. Zhuang, N. G. Walter. *PNAS*, 101:10066, 2004.
- [12] J. S. Smith, K. T. Lee, X. Qu, Z. Xie, J. Pesic, T. R. Sosnick, T. Pan, N. F. Scherer. *J. Mol. Biol.*, 378:943, 2008.
- [13] G. Haran. *Chem. Phys.*, 307:137, 2004.
- [14] S. Liu, G. Bokinsky, N. G. Walter, X. Zhuang. *PNAS*, 104:12634, 2007.
- [15] E. Tan, T. J. Wilson, M. K. Nahas, R. M. Clegg, D. M. J. Lilley, T. Ha. *PNAS*, 100:9308, 2003.

- [16] T. Ha. *Methods*, 25:78, 2001.
- [17] E. W. Merrill S. J. Sofia, V. V. Premnath. *Macromolecules*, 31:5059, 1998.
- [18] B. Okumus, T. J. Wilson, D. M. Lilley, T. Ha. *Biophys. J.*, 87:2798, 2004.
- [19] I. Cisse, B. Okumus, C. Joo, T. Ha. *PNAS*, 104:12646, 2007.
- [20] M. Hajdziona, A. Molski. *Chem. Phys. Lett.*, 470:363, 2009.
- [21] M. Hajdziona, A. Molski. *J. Chem. Phys.*, 134:054112, 2011.
- [22] I. Gopich, A. Szabo. *J. Phys. Chem. B*, 113:10965, 2009.
- [23] M. Hajdziona, A. Molski. *wysłana do druku*.
- [24] T. Ryden. *Comput. Stat. Data Anal.*, 21:431, 1996.
- [25] T. Ryden. *Commun. Statist. Stoch. Models*, 10:795, 1994.
- [26] M. A. Ditzler, E. A. Aleman, D. Rueda, N. G. Walter. *Biopolymers*, 87:302, 2007.
- [27] M. J. B. Pereira, E. N. Nikolova, S. L. Hiley, D. Jaikaran, R. A. Collins, N. G. Walter. *J. Mol. Biol.*, 382:496, 2008.
- [28] M. K. Nahas, T. J. Wilson, S. Hohng, K. Jarvie, D. M. J. Lilley. *T. Ha Nature Struct. Mol. Biol.*, 11:1107, 2004.
- [29] H. Akaike. *IEEE Trans. Autom. Control.*, 19:716, 1974.
- [30] G. Schwartz. *Ann. Stat.*, 6:461, 1978.
- [31] M. Dahan, A. A. Deniz, T. Ha, D. S. Chemla, P. G Schultz, S. Weiss. *Chem. Phys.*, 245:85, 1999.
- [32] A. A. Deniz. *PNAS*, 96:3670, 1999.
- [33] D. P. Fromm, W. E. Moerner. *Rev. Sci. Instrum*, 74:3595, 2003.
- [34] R. Roy, S. Hohng, T. Ha. *Nature Meth.*, 6:507, 2008.
- [35] A. A. Deniz, S. Mukhopadhyay, E. A. Lemke. *J. R. Soc. Interface*, 5:15, 2008.
- [36] S. Weiss, X. Michalet, O. H. W. Siegmund, J. V. Vallerga, P. Jelinsky, J. E. Millaud. *J. Mod. Opt.*, 54:239, 2007.

- [37] J. Pawley. *Handbook of biological confocal microscopy*, New York:Plenum Press, 1989.
- [38] T. Pan, G. J. Smith, T. R. Sosnick, N. F. Scherer. *RNA*, 11:234, 2005.
- [39] J. D. Puglisi, M. Dorywalska, S. C. Blanchard, R. L. Gonzalez, H. D. Kim, S. Chu. *Nucleic Acids Res.*, 33:182, 2005.
- [40] S. Weiss, M. Jager, E. Nir. *Protein Sci.*, 15:640, 2006.
- [41] E. Rhoades, E. Gussakovsky, G. Haran. *PNAS*, 100:3197, 2003.
- [42] G. U. Nienhaus, E. V. Kuzmankina, C. D. Heyes. *PNAS*, 100:3197, 2005.
- [43] T. Ha, I. Rasnik, S. A. McKinney. *Nat. Methods*, 38:542, 2005.
- [44] R. E. Benesch. *Science*, 118:1826, 1953.
- [45] A. N. Kapanidis, T. A. Laurence, N. K. Lee, E. Margeat, X. Kong, S. Weiss. *Acc Chem Res.*, 38:523, 2005.
- [46] M. D. Barkley, H. M. Watrob, C. P. Pan. *JACS*, 125:7336, 2003.
- [47] P. Tinnefeld, J. Ross, P. Buschkamp, D. Fetting, A. Donnermeyer, C. M. Roth. *J. Phys. Chem. B*, 111:321, 2007.
- [48] S. Weiss, N. K. Lee, A. N. Kapanidis, H. R. Koh, Y. Korlann, S. O. Ho, Y. Kim, S. K. Kim N. Gassman. *Biophys. J.*, 92:303, 2007.
- [49] S. Hohng, J. Lee, S. Lee, K. Ragunathan, C. Joo, T. Ha. *Angewandte Chemie International Edition*, 49:9922, 2010.
- [50] P. Tinnefeld, I. H. Stein, C. Steinhauer. *JACS*, DOI:10.1021/ja1105464, 2011.
- [51] A. Molski M. Hajdziona. *Chem. Phys. Lett.*, 470:363, 2009.
- [52] S. A. McKinney, Ch. Joo, T. Ha. *Biophys. J.*, 91:1941, 2006.
- [53] S. Jung, R. M. Dickson. *J. Phys. Chem. B*, 113:13886, 2009.
- [54] Y. Liu, J. Park, K. A. Dahmen, Y. R. Chemla, T. Ha. *J. Phys. Chem. B*, 114:5386, 2010.
- [55] T. R. Cech. *Biosci. Rep.*, 10:317, 1990.

- [56] S. Altman. *Biosci. Rep.*, 10:239, 1990.
- [57] M. Hajdziona, A. Molski. *Wiad. Chem.*, 64:173, 2010.
- [58] N. G. Walter, D. R. Engelke. *Biologist*, 49:199, 2002.
- [59] X. Zhuang. *Ann. Rev. Biophys. Struct.*, 34:399, 2005.
- [60] R. Zhao, D. Rueda. *Methods*, 49:112, 2009.
- [61] D. Rueda, K. S. Karunatilaka. *Chem. Phys. Lett.*, 476:1, 2009.
- [62] X. Chen, A. D. Ellington. *Chem. Biodiversity*, 4:633, 2007.
- [63] N. G. Walter, S. A. Woodson, R. T. Batey. *Non-Protein Coding RNAs, Springer Series in Biophysics*. Springer, 2009.
- [64] J. J. Agresti, B. T. Kelly, A. Jaschke, A. D. Griffiths. *PNAS*, 102:16170, 2005.
- [65] H. D. Kim, G. U. Nienhaus, T. Ha, J. W. Orr, J. R. Williamson, S. Chu. *PNAS*, 99:4284, 2002.
- [66] V. Gopalan. *PNAS*, 104:2031, 2007.
- [67] Y. Kobitski, A. Nierth, M. Helm, A. Jaschke, G. U. Nienhaus. *Nucl. Ac. Research*, 35:2047, 2007.
- [68] R. Wombacher, S. Keiper, S. Suhm, A. Serbanov, D. J. Patel, A. Jaschke. *Angew. Chem. Int. Ed*, 45:2469, 2006.
- [69] T. J. Wilson, M. Nahas, T. Ha, D. M. J. Lilley. *Biochem. Soc. Trans.*, 33:461, 2005.
- [70] S. Liu, G. Bokinsky, N. G. Walter, X. Zhuang. *PNAS*, 104:12639, 2007.
- [71] G. Bokinsky, D. Rueda, V. K. Misra, M. M. Rhodes, A. Gordus, H. P. Babcock, N. G. Walter, X. Zhuang. *PNAS*, 100:9302, 2003.
- [72] N. G. Walter, K. J. Hampel, K. M. Brown, J. M. Burke. *EMBO*, 17:2378, 1998.
- [73] N. G. Walter, J. M. Burke, D. P. Millar. *Nat. Str. Biol.*, 6:544, 1999.
- [74] E. Kikovska, S. G. Svard, L. A. Kirsebom. *PNAS*, 104:2062, 2007.
- [75] D. Herschlag, T. R. Cech. *Biochemistry*, 29:10159, 1990.

- [76] R. Russell, D. Herschlag. *J. Mol. Biol.*, 308:839, 2001.
- [77] R. Russel, X. Zhuang, H. P. Babcock, I. S. Millett, S. Doniach, S. Chu, D. Herschlag. *PNAS*, 99:155, 2002.
- [78] T. H. Lee, L. J. Lapidus, W. Zhao, K. J. Travers, D. Herschlag, S. Chu. *Biophys. J.*, 92:3275, 2007.
- [79] X. Fang, T. Pan, T. R. Sosnick. *Nature, Str. Biol.*, 6:1091, 1999.
- [80] P. Toti, L. Sbordone, C. Sbordone, C. Bauer. *J. Biol. Phys.*, 32:473, 2006.
- [81] T. Persson, S. Cuzic, R. K. Hartmann. *J. Biol. Chem.*, 278:43394, 2003.
- [82] E. Barkai, F. Brown, M. Orrit, H. Yang (Eds.). *Theory and Evaluations of Single-molecule signals*, World Scientific Publishing Co. Inc., 2008.
- [83] M. Lippitz, F. Kulzer, M. Orrit. *ChemPhysChem*, 6:770, 2005.
- [84] H. S. Chung, I. V. Gopich, K. McHale, T. Cellmer, J. M. Louis, W. A. Eaton. *J. Phys. Chem. A*, 10.1021/jp1009669.
- [85] <http://www.scilab.org/>.
- [86] C. Gomez, C. Bunks, J.-P. Chancelier, F. Delebecque, M. Goursat, R. Nikoukhah, S. Steer. *Engineering and Scientific Computing with Scilab*. Birkhauser, 1999.
- [87] W. Fischer, K. Meier-Hellstern. *Perform. Eval.*, 18:149, 1992.
- [88] H. Heffes, D. M. Lucantoni. *J. Sel. Areas Comm.*, 4:856, 1986.
- [89] L. R. Rabiner. *Proc. IEEE*, 77:257, 1989.
- [90] A. Klemm, C. Lindemann, M. Lohmann. *Perform. Eval.*, 54:149, 2003.
- [91] W. J. J. Roberts, Y. Ephraim. *IEEE Signal Process. Lett.*, 13:373, 2006.
- [92] S. C. Kou, X. S. Xie, J. S. Liu. *J. R. Stat. Soc. Ser. C*, 54:469, 2005.
- [93] W. J. Bruno, J. E. Pearson. *J. Comp. Biol.*, 13:1574, 2006.
- [94] I. Yang, W. J. Bruno, W. S. Hlavacek, J. E. Pearson, J. E. *Biophys. J.*, 91:1136, 2006.

- [95] S. H. Chung, R. A. Kennedy. *J. Neurosci. Meth.*, 40:71, 1991.
- [96] M. Andrec, R. M. Levy, D. S. Talaga. *J. Phys. Chem. A*, 107:745, 2003.
- [97] M. Jäger, A. Kiel, D.-P. Hertel, F. Hamprecht. *ChemPhysChem*, 10:2486, 2009.