

Instytut Podstawowych Problemów Techniki  
Zakład Mechaniki i Fizyki Płynów  
Polska Akademia Nauk

# Wpływ ilości kopii genu na dynamikę sieci regulatorowych w komórce

Beata Hat-Plewińska

*Rozprawa doktorska  
wykonana pod kierunkiem  
doc. dr hab. Tomasza Lipniackiego*

Warszawa 2009

***Doc. dr hab. Tomaszowi Lipniackiemu***

*promotorowi, za opiekę naukową  
i pomoc w realizacji niniejszej rozprawy,  
za cierpliwość i mobilizację do pracy*

***Prof. dr hab. Tomaszowi Kowalewskiemu,***  
*Kierownikowi Zakładu Mechaniki i Fizyki Płynów,  
w którym była realizowana ta praca,  
za życzliwą pomoc*

*Wszystkim moim wspaniałym **Koleżankom i Kolegom** z ZMiFP  
za życzliwość i wsparcie*

*Serdecznie Dziękuję*

*Pracę tę dedykuję moim Rodzicom*

Główne rezultaty przedstawione w niniejszej Rozprawie zostały zawarte w następujących dwóch pracach:

B. Hat, P. Paszek, M. Kimmel, K. Piechor and T. Lipniacki. How the number of alleles influences gene autoregulation. *J. Stat. Phys.* 128:511-533 (2007)

B. Hat, K. Puszyński, T. Lipniacki. Exploring mechanisms of oscillations in *p53* and NF-kB systems. *IET Sys. Biol.*, praca przyjęta (2009)

W Rozprawie korzystam również (poza publikacjami innych autorów) z dwóch publikacji naszej grupy, których jestem współautorką:

K. Puszyński, B. Hat, T. Lipniacki. Oscillations and bistability in the stochastic model of *p53* regulation. *J. Theor. Biol.* 254: 452-465 (2008)

T. Lipniacki, B. Hat, J. R. Faeder, W. S. Hlavacek. Stochastic effects and bistability in *T* cell receptor signaling. *J. Theor. Biol.* 254:110-22 (2008)

Niniejsza Rozprawa Doktorska została sfinansowana w ramach grantu promotorskiego Ministerstwa Nauki i Szkolnictwa Wyższego nr 3345.

# Streszczenie

Komórka biologiczna może być rozpatrywana jako stochastyczny reaktor biochemiczny. Istotną rolę w opisie procesów biochemicznych w komórce odgrywają molekularne sieci regulatorowe. Poprzez sieć regulatorową rozumiemy układ wzajemnie reagujących ze sobą substratów (białek, mRNA, genów) odpowiedzialnych za regulację określonych biologicznych funkcji komórki (np. obrona przed infekcją wirusową, cykl komórkowy, apoptoza). Od strony matematycznej molekularna sieć regulatorowa może być postrzegana jako nieliniowy układ dynamiczny, w którym istotną rolę odgrywają efekty stochastyczne związane z małą ilością reagujących cząsteczek, na przykład kopii genów, mRNA lub białek. Układy takie mogą być matematycznie modelowane zarówno w przybliżeniu deterministycznym jak i poprzez równania stochastyczne. Przybliżenie deterministyczne, prowadzące do opisu sieci regulatorowej poprzez układ nieliniowych równań różniczkowych zwyczajnych (po zaniebaniu gradientów przestrzennych), umożliwia przeprowadzenie analizy bifurkacyjnej układu, znalezienie stanów stacjonarnych i cykli granicznych, które mogą być następnie powiązane z biologicznymi stanami komórki. Uwzględniając efekty stochastyczne, sieć regulatorową opisujemy poprzez ciągły w czasie proces Markowa prowadzący do równań Chapmana-Kołmogorowa.

Przedmiotem pracy jest analiza wpływu zmiany liczby kopii genu na dynamikę układu regulatorowego. Informacja o konsekwencjach zmiany liczby kopii genu nie była wykorzystywana podczas przeprowadzania i interpretacji eksperymentów oraz konstrukcji modeli matematycznych. Jak pokażemy, taka analiza dostarcza informacji o mechanizmach rządzących układami regulatorowymi oraz umożliwia interpretację i weryfikację wyników badań eksperymentalnych.

W pierwszej części pracy rozważamy stochastyczną ekspresję pojedynczego genu ze zmienną liczbą kopii. Najpierw analizujemy układ bez autoregulacji, dla którego bierzemy pod uwagę trzy główne źródła stochastyczności, a mianowicie zmianę aktywności genu, transkrypcję i degradację mRNA oraz translację i degradację białka. Wyznaczamy średnie oraz

wariancje rozkładu mRNA i białka dla dowolnej liczby kopii genu. W przybliżeniu ciągłym, w którym rozważamy efekty stochastyczne związane jedynie ze zmianą aktywności genu, wyznaczamy analitycznie średnie i wariancje funkcji gęstości prawdopodobieństwa rozkładu białka dla haploidalnego genu regulowanego przez swoje własne białko w pętli ujemnego i dodatniego sprzężenia zwrotnego. Następnie, dla układu  $N$ -ploidalnego genu, dla którego nie istnieje rozwiązanie analityczne, znajdujemy rozwiązanie numeryczne. W przypadku dodatniej autoregulacji, dla której zależność ekspresji od liczby kopii genu jest znacznie bardziej skomplikowana niż w przypadku ujemnej autoregulacji, znajdujemy dziewięć wzorców równoległej aktywacji haploidalnego, diploidalnego lub tetraploidalnego genu. Analiza tych wzorców pokazuje potencjalne konsekwencje straty kopii genu bądź jego duplikacji.

W drugiej części pracy rozważamy deterministyczny model ścieżki sygnałowej tymocytów sterujących nabytą odpowiedzią immunologiczną. Aktywacja ścieżki sygnałowej tymocytów mierzona jest poziomem kinazy *ERK*, która reguluje pętlę dodatniego sprzężenia zwrotnego zawierającego kaskadę sygnałową kinaz *ZAP*, *MEK*, *ERK*, oraz inhibuje pętlę ujemnego sprzężenia zwrotnego zawierającego fosfatazę *SHP*. Analizujemy wpływ liczby kopii genów kodujących kinazy *ZAP*, *MEK* oraz *ERK* na dynamikę tej kaskady oraz na aktywność tymocytu.

W trzeciej części pracy analizujemy dynamikę układów czynników transkrypcyjnych *p53* i *NF- $\kappa$ B*, związanych z odpowiedzią immunologiczną komórki, dla których dostępne są dane z eksperymentów na pojedynczych komórkach. Szczególną uwagę poświęcamy analizie sieci regulatorowej *p53*, jednej z najważniejszych ścieżek regulatorowych odpowiedzialnych za ochronę genomu komórki. Układy *p53* i *NF- $\kappa$ B*, w odpowiedzi na stymulację, generują trwałe oscylacje o zmiennej amplitudzie lecz dobrze zachowanym okresie, odporne na zmieniającą się liczbę poszczególnych kopii genu oraz na poziom stymulacji. Eksperymenty te nakładają ograniczenia na mechanizmy regulatorowe odpowiedzialne za oscylacje. Wprowadzamy nową metodę walidacji modeli ścieżek regulatorowych, która opiera się na analizie wrażliwości okresu oscylacji na liczbę kopii genów. Wykorzystując tę metodę, analizujemy istniejące

modele układu regulatorowego  $p53$  i pokazujemy, że typy bifurkacji oraz sprzężeń zwrotnych występujących w układzie są kluczowymi wyznacznikami wrażliwości okresu oscylacji na parametry, przy czym typ bifurkacji jest bardziej znaczący. W szczególności pokazujemy, że układy zawierające ujemne sprzężenie zwrotne i opóźnienie, w których oscylacje powstają na skutek przejścia przez nadkrytyczną bifurkację Hopfa, wykazują oscylacje o najbardziej odpornym okresie na zmianę liczby kopii genów i poziom stymulacji. Dla układu  $NF-\kappa B$ , w którym wszystkie deterministyczne modele generują tylko gasnące oscylacje, pokazujemy, że nasycenie w członie transkrypcji jest kluczowe, aby wyprodukować oscylacje o odpornym okresie.

Wyniki tej pracy umożliwiają powiązanie struktury modułów regulatorowych z ich wrażliwością na zmianę liczby kopii genów; można się spodziewać, że moduły regulatorowe, w których brak kopii genu powoduje chorobę haploinsufficiency, zawierają pętlę dodatniego sprzężenia zwrotnego, oraz odwrotnie, występowanie pętli dodatniego sprzężenia powoduje, że gen jest wrażliwy na zmianę liczby kopii. Spostrzeżenie to może być pomocne w analizie zarówno chorób genetycznych jak i mutacji nowotworowych, w których dochodzi do usunięcia bądź powielenia wybranych genów. Zaproponowana w Rozprawie metoda walidacji modeli ułatwi ich poprawną konstrukcję oraz właściwą interpretację i weryfikację wyników badań eksperymentalnych.

# Spis treści

<b>Streszczenie</b>	<b>4</b>
<b>1 Wprowadzenie</b>	<b>10</b>
1.1 Podstawy biologiczne ekspresji genów i sieci regulatorowych	10
1.2 Eksperymenty na pojedynczych komórkach	13
1.3 Liczba kopii genu w komórce	16
<b>2 Modelowanie ekspresji genów i sieci regulatorowych</b>	<b>19</b>
2.1 Proces stochastyczny	20
2.2 Proces Markowa	22
2.3 Równanie Chapmana-Kołmogorowa	23
2.4 Równanie Mistrzów	23
<b>3 Wpływ liczby kopii genu na jego regulację z uwzględnieniem efektów stochastycznych</b>	<b>26</b>
3.1 Stochastyczność w procesie ekspresji genów	26
3.2 Opis modelu ekspresji genu haploidalnego	26
3.2.1 Model dyskretny: równania Mistrzów	28
3.2.2 Średnie i wariancje poziomów mRNA i białka dla modelu dyskretnego bez autoregulacji	29
3.2.3 Przybliżenie ciągłe	30
3.2.4 Równania ciągłości dla funkcji gęstości prawdopodobieństwa (FGP) poziomów mRNA i białek	31
3.2.5 Średnie i wariancje poziomów mRNA i białka dla modelu ciągłego bez autoregulacji	33
3.2.6 Przybliżenie Keplera-Elstona	34
3.2.7 Średnia i wariancja poziomu białka dla modelu K-E bez autoregulacji	35



3.3	Analiza ekspresji genu diploidalnego	36
3.4	Analiza ekspresji genu $N$ -ploidalnego	38
<b>4</b>	<b>Wyniki analityczne</b>	<b>42</b>
4.1	Rozwiązanie modelu K-E z autoregulacją dla genu haploidalnego	42
4.2	Rozwiązanie modelu K-E bez autoregulacji dla genu diploidalnego	44
4.3	Rozwiązanie modelu K-E bez autoregulacji dla $N$ -ploidalnego genu	45
<b>5</b>	<b>Wyniki numeryczne</b>	<b>46</b>
5.1	Charakterystyczne rozwiązania dla układu bez autoregulacji	46
5.2	Charakterystyczne rozwiązania dla układu z autoregulacją	49
5.3	Brak kopii genu nie może być zrekompensowany przez podwojoną ekspresję pozostałej kopii genu	59
<b>6</b>	<b>Wpływ liczby kopii genów na dynamikę ścieżki sygnałowej tymocytu</b>	<b>62</b>
6.1	Tymocyty	62
6.2	Model ścieżki sygnałowej $TCR$	63
6.3	Analiza modelu	67
<b>7</b>	<b>Wpływ liczby kopii genów na dynamikę sieci regulatorowych czynników transkrypcyjnych <math>p53</math> oraz <math>NF-\kappa B</math></b>	<b>72</b>
7.1	Układ regulatorowy $p53 Mdm2$	72
7.2	Układ regulatorowy $NF-\kappa B$	74
7.3	Trwałe oscylacje w układach $p53$ i $NF-\kappa B$	76
7.4	Typy bifurkacji prowadzące do oscylacji	78
7.5	Modele układu regulatorowego $p53 Mdm2$	80
7.5.1	Kryteria poprawnego modelu	81
7.5.2	Analiza modeli opartych na dodatnim i ujemnym	

sprzężeniu zwrotnym	82
7.5.3 Model "testowy" z dodatnim i ujemnym sprzężeniem zwrotnym oparty na nadkrytycznej bifurkacji Hopfa	86
7.5.4 Analiza modeli opartych na ujemnym sprzężeniu zwrotnym i opóźnieniu	93
7.6 Analiza modeli układu regulatorowego $NF-\kappa B$	97
<b>8 Konkluzje</b>	<b>100</b>
<b>Dodatek A Momenty rozkładów dla mRNA i białka</b>	<b>106</b>
A.1 Model dyskretny, $N=1$	106
A.2 Model ciągły, $N=1$	114
A.3 Przybliżenie K-E, $N=1$ , $N=2$	119
<b>Dodatek B Uzasadnienie przybliżenia K-E</b>	<b>131</b>
<b>Dodatek C Niezmienniczość zbioru trajektorii <math>D_2</math></b>	<b>132</b>
<b>Dodatek D Metoda dyskretyzacji</b>	<b>135</b>
<b>Bibliografia</b>	<b>137</b>

# Rozdział 1

## Wprowadzenie

### 1.1 Podstawy biologiczne ekspresji genów i sieci regulatorowych

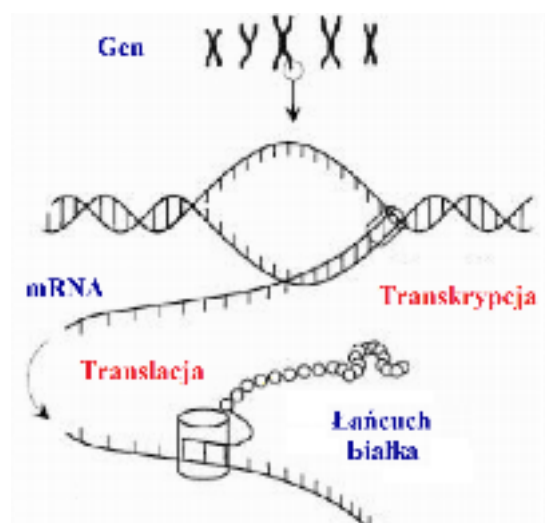
Informacja genetyczna zawarta w DNA zakodowana jest w sekwencji zasad azotowych: adeniny, guaniny, cytozyny lub tyminy, oznaczanych skrótami A, G, C i T. W cząsteczce RNA natomiast, zamiast tyminy występuje uracyl (U). Białka zbudowane są z dwudziestu różnych rodzajów aminokwasów. *Gen* jest odcinkiem łańcucha DNA, który koduje sekwencję aminokwasów w cząsteczce białka. Geny zlokalizowane są w długich fragmentach DNA zwanych *chromosomami*. Chromosomy dzielą się na *autosomy* - zawiadujące dziedziczeniem cech nie związanych z płcią, oraz chromosomy płci, czyli *allosomy*.

Wyższe organizmy (eukarionty) są w większości *diploidalne*, tzn. ich komórki somatyczne (komórki budujące organizm z wyłączeniem komórek płciowych) zawierają dwa zestawy (oznaczane przez  $2n$ ) chromosomów homologicznych (genomy), z których jeden pochodzi od organizmu matki, a drugi od organizmu ojca. Komórki *haploidalne* zawierają pojedynczy zestaw chromosomów ( $1n$ ). U gatunków rozmnażających się płciowo występują zarówno komórki diploidalne jak i haploidalne komórki płciowe (gamety). Liczba chromosomów danego gatunku jest stała i wynosi od 10 do 40. U człowieka kompletny zestaw chromosomów komórki somatycznej tworzą 22 pary autosomów i 1 para allosomów (nazywanych  $X$  i  $Y$ ), która odpowiada za determinację płci. Osobnik posiadający chromosomy  $XX$  jest płci żeńskiej, zaś osobnik z jednym chromosomem  $X$  i jednym chromosomem  $Y$  jest płci męskiej.

Informacja genetyczna zawarta w genach zostaje odczytana i przepisana na ich produkty (białka lub różne formy RNA) w procesie określanym jako *ekspresja genów*, i przekazywana

jest tylko w jednym kierunku (tzw. główny dogmat biologii): od DNA do RNA w procesie *transkrypcji*, a następnie do białka w procesie *translacji*. W komórkach eukariotycznych synteza RNA zachodzi w jądrze komórkowym, zaś synteza białka w cytoplazmie (Rys. 1.1).

Transkrypcja katalizowana jest przez enzym zwany *polimerazą RNA*. Zapoczątkowanie transkrypcji polega na związaniu się polimerazy RNA z *promotorem* - odcinkiem DNA, położonym w pobliżu sekwencji kodującej genu, który zawiera sekwencje rozpoznawane przez polimerazę RNA. Promotory eukariotyczne zawierają także sekwencje rozpoznawane przez *czynniki transkrypcyjne*, które wiążąc się z DNA umożliwiają związanie się polimerazy RNA z nicią DNA i rozpoczęcie transkrypcji. Polimeraza przesuwając się wzdłuż nici DNA służącej jako matryca i dobudowuje komplementarną nić mRNA zgodnie z regułą, że jednej zasadzie w transkrybowanej nici DNA odpowiada tylko jedna zasada w powstającej cząsteczce mRNA. Przetransportowany do cytoplazmy mRNA łączy się z *rybosomami*, które obejmują podjednostkami przesuwającą się nić mRNA, odczytują sekwencję zasad mRNA i tłumaczą ją w procesie translacji na sekwencję aminokwasową białka. W wyniku translacji dochodzi do ostatecznego "przetłumaczenia" informacji genetycznej zawartej w genie na konkretną sekwencję białka.



Rysunek 1.1 Przeływ informacji genetycznej

Po translacji nowo zsyntetyzowane białko może ulegać różnym modyfikacjom, w wyniku których może pełnić określone funkcje. Jedną z tych modyfikacji jest *fosforylacja*, która polega na przyłączeniu do białka reszty fosforanowej przez enzymy zwane *kinazami*. *Defosforylacja* polega na usunięciu reszty fosforanowej przez enzymy zwane *fosfatazami*. Przyłączenie lub odłączenie reszty fosforanowej reguluje aktywność enzymatyczną wielu białek, ich lokalizację w obrębie komórki czy też zdolność do tworzenia kompleksów. Inną potranslacyjną modyfikacją białka jest *ubikwitynacja*, która polega na dołączeniu do białka jednej bądź kilku pojedynczych cząsteczek innego białka - *ubikwityn*. Po dołączeniu ubikwityny białko może ulec rozkładowi w wieloenzymatycznym kompleksie białkowym, *proteasomie*.

Ekspresja genów jest ściśle regulowana. W większości przypadków regulacja zachodzi na poziomie transkrypcji. Głównym transkrypcyjnym mechanizmem regulacji jest inicjacja transkrypcji kontrolowana przez oddziaływanie czynników transkrypcyjnych z DNA, które może zarówno blokować transkrypcję, jak i ją umożliwić. Odpowiednio do pełnionej funkcji, czynniki transkrypcyjne nazywane są *reprezami* albo *aktywatorami*. Jeżeli białkowy produkt genu jest regulatorem własnej transkrypcji, to tworzy się układ *sprzężenia zwrotnego*. Jeżeli białko to działa na gen jak represor, tworzy się pętla *ujemnego* sprzężenia zwrotnego, jeżeli zaś białko działa jak aktywator, tworzy się pętla *dodatniego* sprzężenia zwrotnego.

Tempa zmiany stanu genu, transkrypcji, translacji oraz degradacji mRNA i białek mogą zmieniać się o kilka rzędów wielkości pomiędzy poszczególnymi genami. Górna granica tempa dla transkrypcji wynosi około 0.1 cząsteczek na sekundę, dla translacji - około 0.5 cząsteczek na sekundę i osiągana przez geny, których produkty są potrzebne w komórce w dużych ilościach w przeciągu krótkiego czasu; np. około 100 000 cząsteczek białka inhibitorowego nazywanego *I $\kappa$ B $\alpha$*  musi zostać zdegradowanych a następnie zresyntetyzowanych w przeciągu około 30 minut, aby właściwie kontrolować wczesną fazę odpowiedzi immunologicznej, [2]. Jednak istnieje duża liczba genów, których produkty są potrzebne w małych ilościach, rzędu kilku cząsteczek na komórkę, geny te ulegają ekspresji ze znacznie mniejszymi tempami lub ich ekspresja jest bardzo nieregularna. Tempa zmiany aktywacji genu są większe dla prokar-

iontów (rzędu  $1/s$ ) niż dla eukariontów. Białka są zazwyczaj bardziej stabilne (mają dłuższy czas degradacji) niż mRNA, jednak w obydwu przypadkach połówkowy czas degradacji może być rzędu minut bądź dni.

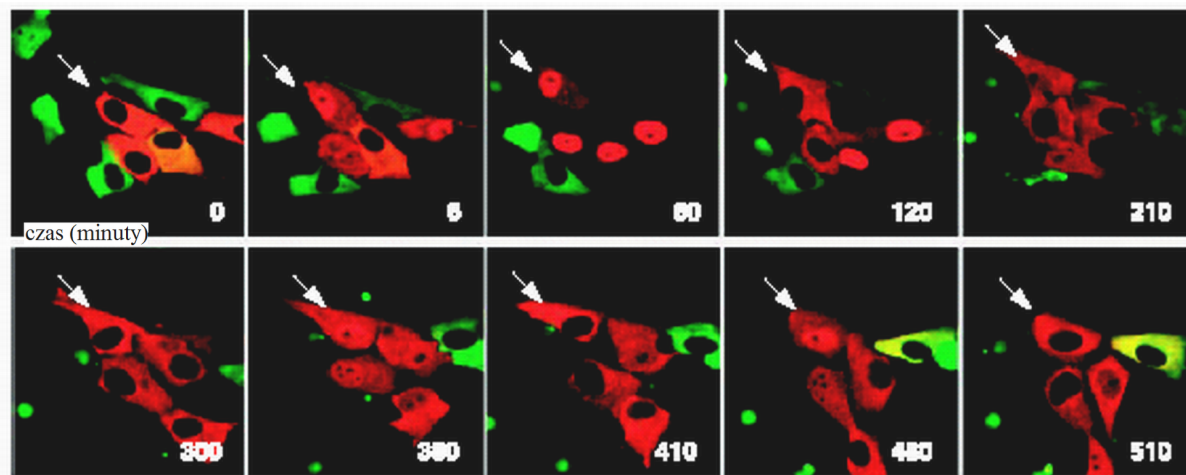
## 1.2 Eksperymenty na pojedynczych komórkach

Matematyczne modelowanie układów regulatorowych białek w pojedynczych komórkach jest obecnie gwałtownie rozwijającą się dziedziną na skutek powstania nowych technik eksperymentalnych, które umożliwiają śledzenie zmian poziomu konkretnego białka z podziałem na jądro i cytoplazmę w pojedynczej komórce w ciągu godzin a nawet dni, których nie można obserwować w eksperymentach populacyjnych. Rozwijające się eksperymenty na pojedynczych komórkach są bardzo trudne do przeprowadzenia, dlatego większość eksperymentów wykonywana jest na dużych populacjach komórek, [45], w których obserwuje się średnie poziomy mRNA lub białek. W przypadku analizy sieci regulatorowej czynnika transkrypcyjnego *p53*, gdy populacja komórek składa się z podpopulacji komórek proliferujących (rozmnażających się), komórek z zablokowanym cyklem komórkowym oraz komórek apoptycznych, taka uśredniona po populacji trajektoria nie odpowiada żadnemu biologicznemu procesowi.

Obecne techniki eksperymentalne umożliwiają analizę jedynie uniesmiertelnionych komórek rakowych, ze względu na łatwość ich hodowli oraz możliwość uzyskania powtarzalnych eksperymentów. Śledzenie w czasie ewolucji zdrowych niezmutowanych komórek (tzw. wild-type) jest niemożliwe ze względu na ich szybkie obumieranie.

Ponadto, aby przeprowadzić eksperyment dokonuje się *transfekcji*, czyli procesu wprowadzania do genomu komórek eukariotycznych dodatkowych kopii obcego genu kodującego białko docelowe wraz z dołączonym fluorescencyjnym znacznikiem. W wyniku transfekcji fluorescencyjnie oznaczone białko świeci pod wpływem światła lasera, co umożliwia jego obserwację pod mikroskopem oddzielnie w każdej komórce. Wyróżnia się dwa rodzaje transfekcji: *stabilną* oraz *przeźściową*. Transfekcja stabilna występuje wtedy, gdy wprowadzone kopie genu

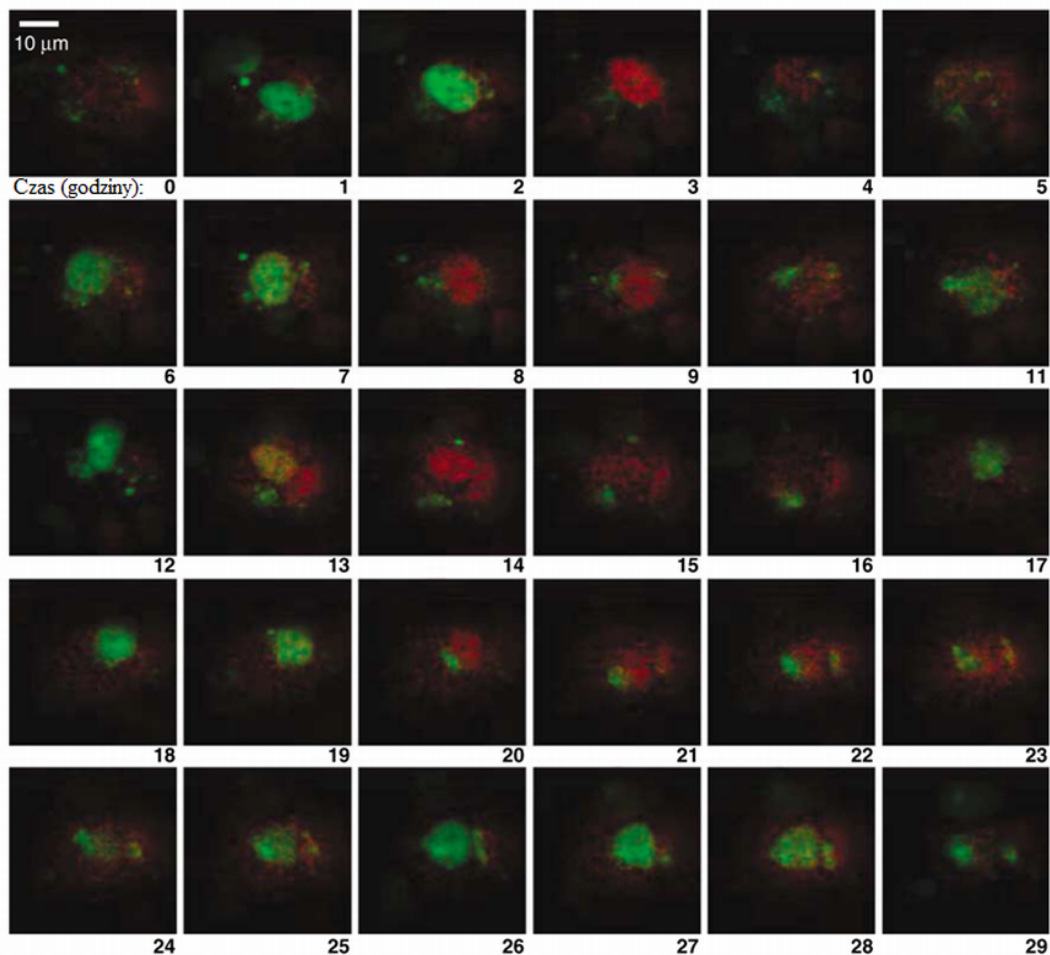
ulegają integracji z chromosomem; liczba tych sztucznie wprowadzonych kopii genu jest trudna do kontrolowania lecz pozostaje stała w tym samym klonie. Transfekcja przejściowa polega na wprowadzeniu do komórki *plazmidu* - cząsteczki DNA zawierającej sekwencje kodujące dane białko, występującej w komórce poza chromosomem, nie ulegającej integracji z chromosomem i zdolnej do autonomicznej replikacji; gen w transfekowanej komórce ulega ekspresji przejściowo, tzn. w krótkim okresie czasu (jego ekspresja zostaje zgubiona w miarę podziałów komórkowych) oraz jedynie w komórce, do której został wprowadzony.



Rysunek 1.2 (Nelson i in. 2004, grupa M.R.H. White) Oscylacyjne translokacje pomiędzy jądrem a cytoplazmą białka RelA-dsRed (białko tworzące kompleks  $NF-\kappa B$  z doczepionym fragmentem świecącym na czerwono) wywołane stymulacją  $TNF\alpha$  (czynnik martwicy nowotworu), jednocześnie degradacja i odbudowa inhibitora,  $I\kappa B\alpha$ -EGFP (zielone). Wykorzystywana linia komórkowa to SK-N-AS (komórki nowotworowe układu nerwowego), rodzaj przeprowadzonej transfekcji - przejściowa.

W naszej pracy korzystamy z wyników eksperymentów na pojedynczych komórkach dwóch grup: U. Alona, która analizuje białka tworzące sieć regulatorową czynnika transkrypcyjnego  $p53$  (Rys. 1.2), oraz M.R.H. White'a, która analizuje białka tworzące sieć regulatorową czynnika transkrypcyjnego  $NF-\kappa B$  (Rys. 1.3). Eksperymenty te pozwoliły

pokazać po raz pierwszy w grupie U. Alona istnienie niegasnących (w ciągu 72 godzin trwania eksperymentu) oscylacji w układzie regulatorowym  $p53|Mdm2$  o okresie ok. 6 godzin, nie możliwych do zaobserwowania w eksperymentach populacyjnych, [37],[57] oraz w grupie M.R.H. White'a niegasnących oscylacji w układzie  $NF-\kappa B$  o okresie ok. 100 minut, [72],[36].



Rysunek 1.3 (Geva-Zatorsky i in. 2006, grupa U. Alona) Oscylacje pomiędzy jądrem a cytoplazmą poziomów białek  $p53$ -CFP (zielone) i  $Mdm2$ -YFP (czerwone) w odpowiedzi na uszkodzenie DNA wywołane promieniowaniem gamma. Wykorzystywana linia komórkowa to MCF7 (komórki rakowe gruczołu piersiowego), rodzaj przeprowadzonej transfekcji - stabilna.



### 1.3 Liczba kopii genu w komórce

Liczba kopii genu w komórce nie jest stała. Istnieje wiele procesów, które zmieniają tę liczbę w czasie ewolucji komórki.

Podczas cyklu komórkowego, który u eukariontów trwa około 24 godzin, w fazie S w diploidalnych komórkach somatycznych następuje replikacja DNA i większość genów jest przejściowo tetraploidalna; następnie, w czasie mitozy, w komórce macierzystej, która po replikacji DNA ma podwójny ( $4n$ ) zestaw chromosomów, chromosomy potomne rozdzielają się, wokół nich odtwarzana jest otoczka jądrowa i tworzą się dwa jądra komórkowe. W wyniku mitozy powstają dwie komórki potomne z identyczną jak komórka macierzysta liczbą chromosomów i tym samym zestawem genów.

Mejoza (proces podziałów związanych z gametogenezą) prowadzi do zrekombinowania informacji genetycznej zapewniając stałą liczbę chromosomów w kolejnych pokoleniach organizmów, które rozmnażają się płciowo. W wyniku mejozy obie komórki potomne, gamety, mają haploidalną liczbę chromosomów. Z połączenia dwóch gamet powstaje zygota, która jest organizmem diploidalnym. Liczba chromosomów może ulec zmianie na skutek błędów podczas mejozy, które prowadzą do powstawania gamet o zwiększonej ilości materiału genetycznego.

U organizmów diploidalnych brak jednej kopii genu może prowadzić do *haploinsufficiency*, choroby wynikającej z niedoboru białka danego typu lub zbyt dużych fluktuacji jego poziomów. Rozwój nowotworu jest w wielu przypadkach związany z częściową multiplikacją genomu. Mitoza w komórkach nowotworowych prowadzi do zmienności liczby chromosomów w komórkach tego samego guza. W rezultacie w komórkach nowotworowych może być znacznie więcej kopii poszczególnych genów niż w komórce diploidalnej. W normalnej ewolucji komórki niektóre geny mogą być multiplikowane w odpowiedzi na stresogenne czynniki środowiskowe, gdy dane białko jest potrzebne w komórce w większej ilości.

W eksperymentach na pojedynczych komórkach analizuje się transfekowane komórki rakowe, w których geny występują w większej liczbie kopii niż w normalnych komórkach

diploidalnych. Na przykład komórki HeLa (komórki raka szyjki macicy), powszechnie używane jako standardowa linia komórkowa w wielu eksperymentach, zawierają trzy razy więcej DNA niż komórki wild-type, co powoduje że, przeciętnie, geny są heksaploidalne. Eksperymenty na pojedynczych komórkach są niezbędne do badania układów regulatorowych białek, ponieważ pokazują dynamikę poszczególnych komórek, zamaskowaną w eksperymentach populacyjnych. Jednakże, przeprowadzenie transfekcji, która zwiększa liczbę kopii genów, oraz wykorzystanie komórek rakowych, które ze swojej natury są silnie zmutowane zaś mechanizmy regulatorowe białek są w nich uszkodzone powoduje, że w eksperymentach tych analizuje się dynamikę silnie zaburzonego układu, co bardzo utrudnia interpretację wyników eksperymentalnych, i w efekcie zrozumienie mechanizmów regulatorowych układu nie zaburzonego, którego bezpośrednio nie możemy zbadać.

Pełne zrozumienie działania sieci regulatorowej zarówno w komórkach niezmutowanych jak i nowotworowych wymaga konstrukcji i analizy odpowiedniego modelu matematycznego. Budowa modelu sieci regulatorowej polega na odgadnięciu struktury układu dynamicznego, którego dynamika jest zgodna z eksperymentem. Naturalną metodą jest analiza wpływu wyłączenia (bądź wzmocnienia) poszczególnych elementów układu na funkcjonowanie całości. W eksperymencie polega to na usunięciu pewnych białek bądź zwiększeniu ich ilości poprzez wprowadzenie dodatkowych kopii genów.

Informacja o konsekwencjach zmiany liczby kopii genów nie była w praktyce uwzględniana podczas przeprowadzania i interpretacji eksperymentów. Stworzenie poprawnego modelu matematycznego opisującego dynamikę zarówno komórek normalnych jak i komórek ze zwiększoną liczbą kopii genów jest niezbędne z punktu widzenia przeprowadzanych eksperymentów. Poprawny model matematyczny oraz symulacje numeryczne pozwoliłyby ocenić, w jaki sposób zachowują się białka w rozważanej sieci w zależności od wymuszenia, jakim może być uszkodzenie DNA, oraz umożliwiłyby interpretację i weryfikację wyników badań eksperymentalnych. Analiza matematyczna może zatem ukierunkować badania eksperymentalne oraz przyczynić się do znacznego obniżenia kosztów badań oraz skrócenia ich czasu.

## Rozdział 2

# Modelowanie ekspresji genów i sieci regulatorowych

Niektóre z kluczowych biochemicznych interakcji w komórce zachodzą pomiędzy bardzo małą liczbą cząsteczek. Przykładem jest tu proces transkrypcji mRNA określonego genu, który zachodzi, jeżeli w rejonie promotora genu przyłączony jest odpowiedni czynnik transkrypcyjny lub kilka różnych czynników transkrypcyjnych. W rezultacie przyłączenie (z natury losowe) pojedynczej cząsteczki czynnika transkrypcyjnego prowadzi do transkrypcji dziesiątek cząsteczek mRNA a następnie syntezy tysięcy cząsteczek odpowiadającego im białka. Powoduje to, że trajektorie poszczególnych komórek w wielowymiarowej przestrzeni, której współrzędnymi są ilości poszczególnych mRNA, białek i ich kompleksów, odpowiednio w jądrze i cytoplazmie, są istotnie różne. Mechanizmy regulatorowe, w szczególności sprzężenia zwrotne sterujące transkrypcją, translacją, degradacją białek i mRNA oraz formowaniem się kompleksów białkowych, wykorzystują stochastyczność do różnicowania populacji komórek, bądź minimalizują ją, aby zapewnić poprawne (bliskie deterministycznemu) funkcjonowanie komórki.

Modele łączące deterministyczny opis szybkich kanałów reakcji w komórce opisywanych układami nieliniowych równań różniczkowych, ze stochastycznym modelowaniem wolnych kanałów reakcji prowadzą do losowych równań różniczkowych, w których „cząstki” dokonują stochastycznych przeskoków pomiędzy rodzinami deterministycznych trajektorii. Prowadzi to do opisu sieci regulatorowej poprzez ciągły w czasie, kawałkami deterministyczny proces Markowa. Matematyczny opis takiego procesu stanowią równania Chapmana-Kołmogorowa.

## 2.1 Proces stochastyczny

Zmienną losową (rzeczywistą) nazywamy dowolną rzeczywistą funkcję mierzalną  $Y$  (tzn. spełniającą warunek  $Y^{-1}(B) \in \mathcal{F}$  dla każdego zbioru borelowskiego  $B \subseteq \mathbb{R}$ ) określoną na przestrzeni probabilistycznej  $(\Omega, \mathcal{F}, P)$  o wartościach w zbiorze liczb rzeczywistych:  $Y : \Omega \rightarrow \mathbb{R}$ .

*Proces stochastyczny* jest rodziną  $\{Y_t\}_{t \in T}$  zmiennych losowych, które przyjmują wartości w przestrzeni mierzalnej  $(\mathbb{R}, \mathcal{F})$  i są indeksowane przez zbiór  $T$ , gdzie  $t$  interpretujemy jako czas. Dla ustalonego  $y \in \Omega$  funkcję  $Y_y(t)$  nazywamy *realizacją* (albo *trajektorią*) *procesu*. Realizacja procesu wyraża ewolucję w czasie wybranego zdarzenia losowego. Wartości procesu nazywamy *stanami*.

Jeśli zmienna losowa  $Y$  ma rozkład dyskretny, tzn.  $Y$  przyjmuje wartości  $y_i$  z prawdopodobieństwami  $p_i$ ,  $\{(y_i, p_i)_{i \in I}\}$ , gdzie  $I$  jest zbiorem przeliczalnym oraz  $\sum_{i \in I} p_i = 1$ , wówczas wartość oczekiwana zmiennej losowej  $Y$  zdefiniowanej na tym samym zbiorze stanów określona jest wzorem

$$E[Y] = \sum_{i \in I} y_i p_i,$$

jeśli zaś zmienna  $Y$  ma rozkład ciągły o gęstości  $p(y)$ , wówczas

$$E[Y] = \int_{-\infty}^{\infty} y p_Y(y) dy.$$

Następnie, mając zdefiniowaną wartość oczekiwaną można zdefiniować wariancję w następujący sposób: dla zmiennych losowych o rozkładzie dyskretnym:

$$Var[Y] = \sum_{i=1}^{\infty} (y_i - E[Y])^2 p_i$$

oraz dla zmiennych losowych o rozkładzie ciągłym z gęstością  $p(y)$ :

$$Var[Y] = \int_{-\infty}^{\infty} (y - E[Y])^2 p(y) dy.$$

Wartość oczekiwana i wariancja są w istocie szczególnymi przypadkami tak zwanych *momentów*. Liczbę

$$M^m[Y] := E[Y^m] = \begin{cases} \sum_i y_i^m p_i, & \text{dla } Y \text{ o rozkładzie dyskretnym} \\ \int_{-\infty}^{\infty} y^m p(y) dy, & \text{dla } Y \text{ o rozkładzie ciągłym} \end{cases}, \quad m = 1, 2, \dots$$

nazywamy *momentem rzędu  $m$*  zmiennej losowej  $Y$ .

*Prawdopodobieństwo łączne* przyjęcia przez proces stochastyczny  $Y$  wartości  $y_1$  w chwili  $t_1$ , wartości  $y_2$  w chwili  $t_2$ , itd., oznaczamy:

$$p(y_1, t_1; y_2, t_2; \dots; y_n, t_n), \quad n = 1, 2, \dots \quad (1)$$

Ustalając wartości zmiennej losowej  $Y$  w  $k$  zadanych chwilach  $t_1, \dots, t_k$ , można zdefiniować *prawdopodobieństwo warunkowe* przyjęcia określonych wartości w  $l$  innych chwilach  $t_{k+1}, \dots, t_{k+l}$ , które wynosi

$$p(y_{k+l}, t_{k+l}; \dots; y_{k+1}, t_{k+1} | y_k, t_k; \dots; y_1, t_1) = \frac{p(y_{k+l}, t_{k+l}; \dots; y_{k+1}, t_{k+1})}{p(y_k, t_k; \dots; y_1, t_1)}. \quad (2)$$

Prawdopodobieństwo warunkowe jest gęstością prawdopodobieństwa przyjęcia przez  $Y$  wartości  $y_{k+1}$  w momencie  $t_{k+1}$ , wartości  $y_{k+2}$  w momencie  $t_{k+2}$ , itd. jeżeli jej wartość w chwili  $t_1$  wynosi  $y_1$ , wartość w chwili  $t_2$  wynosi  $y_2$ , itd.

Prawdopodobieństwo warunkowe jest nieujemne oraz unormowane:

$$\int p(y_{k+l}, t_{k+l}; \dots; y_{k+1}, t_{k+1} | y_k, t_k; \dots; y_1, t_1) dy_{k+l} \dots dy_{k+1} = 1. \quad (3)$$

Proces stochastyczny jest *stacjonarny* jeśli, dla każdego  $n$ , prawdopodobieństwa (1) zależą jedynie od różnic czasów, tzn.

$$p(y_1, t_1 + \tau; y_2, t_2 + \tau; \dots; y_n, t_n + \tau) = p(y_1, t_1; y_2, t_2; \dots; y_n, t_n), \quad n = 1, 2, \dots \quad (4)$$

## 2.2 Proces Markowa

Jednym z najważniejszych przykładów procesów stochastycznych są *procesy Markowa*. Proces stochastyczny  $Y_n$  o takiej własności, że dla każdego  $n \in \mathbb{N}$  i każdego ciągu chwil  $t_1 < t_2, \dots < t_n \in T$  gęstość prawdopodobieństwa warunkowego  $y_n$  w chwili  $t_n$  zależy tylko od stanu układu  $y_{n-1}$  w chwili  $t_{n-1}$ , a nie zależy od przeszłości, czyli od  $y_1, \dots, y_{n-2}$ , tzn.

$$p(y_n, t_n | y_{n-1}, t_{n-1}; \dots; y_1, t_1) = p(y_n, t_n | y_{n-1}, t_{n-1}) \quad (5)$$

nazywamy *procesem Markowa*. Własność (5) nazywamy *własnością Markowa*.

Prawdopodobieństwo  $p(y_n, t_n | y_{n-1}, t_{n-1})$  nazywane jest *prawdopodobieństwem przejścia*. Proces Markowa wyznaczony jest przez rozkład początkowy oraz prawdopodobieństwo przejścia:  $p(y_1, t_1)$  i  $p(y_2, t_2 | y_1, t_1)$ . Na ich podstawie można odtworzyć całą hierarchię prawdopodobieństw łącznych, które w pełni określają proces:

$$p(y_1, t_1; y_2, t_2; \dots; y_n, t_n) = p(y_n, t_n | y_{n-1}, t_{n-1}) \dots p(y_3, t_3 | y_2, t_2) p(y_2, t_2 | y_1, t_1). \quad (6)$$

Na przykład dla rozkładu z  $t_1 < t_2$  otrzymujemy, z definicji prawdopodobieństwa warunkowego:

$$p(y_1, t_1; y_2, t_2) = p(y_2, t_2 | y_1, t_1) p(y_1, t_1). \quad (7)$$

Podobnie, dla rozkładu z  $t_1 < t_2 < t_3$ , korzystając dodatkowo z definicji procesów Markowa, otrzymujemy:

$$\begin{aligned} p(y_1, t_1; y_2, t_2; y_3, t_3) &= p(y_3, t_3 | y_1, t_1; y_2, t_2) p(y_1, t_1; y_2, t_2) = \\ &= p(y_3, t_3 | y_2, t_2) p(y_2, t_2 | y_1, t_1) p(y_1, t_1). \end{aligned} \quad (8)$$

## 2.3 Równanie Chapmana-Kołmogorowa

Całkując obie strony tożsamości (8) po  $y_2$  otrzymujemy dla  $t_1 < t_2 < t_3$ , z definicji prawdopodobieństwa łącznego (6) dla  $n = 3$ , następujące wyrażenie:

$$p(y_1, t_1; y_3, t_3) = p(y_1, t_1) \int p(y_2, t_2 | y_1, t_1) p(y_3, t_3 | y_2, t_2) dy_2. \quad (9)$$

Dzieląc obie strony przez  $p(y_1, t_1)$ , a następnie korzystając z definicji prawdopodobieństwa warunkowego (2) dla  $k = l = 1$  dostajemy następujący związek dla prawdopodobieństw przejścia, który zachodzi dla każdego procesu stochastycznego:

$$p(y_3, t_3 | y_1, t_1) = \int p(y_3, t_3 | y_2, t_2) p(y_2, t_2 | y_1, t_1) dy_2, \quad t_1 < t_2 < t_3. \quad (10)$$

Równanie to nazywa się *równaniem Chapmana-Kołmogorowa*. Jest to tożsamość, którą musi spełniać prawdopodobieństwo przejścia każdego procesu Markowa. Równanie Chapmana-Kołmogorowa orzeka, że proces rozpoczynający się w chwili  $t_1$  od wartości  $y_1$  osiąga wartość  $y_3$  w chwili  $t_3$ , przechodząc przez jedną z dopuszczalnych wartości  $y_2$  w pośredniej chwili  $t_2$ .

## 2.4 Równanie Mistrzów

*Równanie Mistrzów* jest pewną równoważną, różniczkową postacią równania Chapmana-Kołmogorowa.

Zapisując tożsamość (7) dla  $t_1 = t_2$  otrzymamy dodatkowy warunek:

$$p(y_2, t_1 | y_1, t_1) = \delta(y_2 - y_1). \quad (11)$$

Zapiszmy prawdopodobieństwo przejścia w procesie Markowa dla małych wartości różnicy czasu  $\Delta t$ . Uwzględniając warunek (11) mamy, z dokładnością do członów liniowych w  $\Delta t$ :

$$p(y, t + \Delta t | y', t) \approx \delta(y - y') (1 - a_0(y', t) \Delta t) + W_t(y | y') \Delta t. \quad (12)$$

Funkcja  $W_t(y | y')$  jest *prawdopodobieństwem przejścia w jednostce czasu* w chwili  $t$  od stanu  $y$  do  $y'$ . Dla stacjonarnych procesów Markowa  $W_t(y | y')$  nie zależy od czasu. Całkując obie strony równania (12) po  $y$ , z warunku unormowania (3), otrzymujemy następujące równanie

$$1 = 1 - a_0(y', t) \Delta t + \Delta t \int W_t(y | y') dy. \quad (13)$$

Stąd

$$a_0(y', t) = \int W_t(y | y') dy, \quad (14)$$

$a_0(y', t)$  jest całkowitym prawdopodobieństwem przejścia na jednostkę czasu ze stanu  $y'$  do jakiegokolwiek innego stanu.

Napiszmy równanie Chapmana-Kołmogorowa dla uporządkowanych chwil  $t_0 < t < t + \Delta t$ :

$$p(y, t + \Delta t | y_0, t_0) = \int p(y, t + \Delta t | y', t) p(y', t | y_0, t_0) dy'. \quad (15)$$

Stąd, po podstawieniu równania (13), otrzymujemy

$$p(y, t + \Delta t | y_0, t_0) \approx \left(1 - a_0(y', t) \Delta t\right) p(y, t | y_0, t_0) + \int \left\{W_t(y|y') \Delta t\right\} p(y', t | y_0, t_0) dy'. \quad (16)$$

Przepiszmy równanie (16) w postaci

$$p(y, t + \Delta t | y_0, t_0) = \int \left\{W_t(y|y') \Delta t\right\} p(y', t | y_0, t_0) dy' + p(y, t | y_0, t_0) \left(1 - \int W_t(y|y') dy' \Delta t\right). \quad (17)$$

Pierwszy wyraz po prawej stronie opisuje prawdopodobieństwo przejścia do stanu  $y$  w czasie  $\Delta t$ , natomiast drugi opisuje prawdopodobieństwo, że w tym czasie układ pozostanie w stanie  $y$ . W granicy  $\Delta t \rightarrow 0$  otrzymujemy różniczkową postać równania Chapmana-Kołmogorowa zwaną *równaniem Mistrzów*. (Równanie Mistrzów jest równaniem na prawdopodobieństwo przejścia w procesie  $Y(t)$ , a nie na gęstość prawdopodobieństwa  $p(y, t)$ ):

$$\frac{\partial p(y, t | y_0, t_0)}{\partial t} = \int \left\{W_t(y|y') p(y', t | y_0, t_0) - W_t(y'|y) p(y, t | y_0, t_0)\right\} dy'. \quad (18)$$

Pierwszy wyraz po prawej stronie opisuje zmianę  $p(y, t | y_0, t_0)$  wynikającą z przejść do stanu  $y$  w czasie  $\Delta t$ , natomiast drugi wyraz z ujemnym znakiem opisuje analogiczną zmianę wynikającą z opuszczenia stanu  $y$ . Równanie Mistrzów jest więc równaniem bilansu przyjść odniesionych do ustalonego stanu.

Jeśli zbiór wartości procesu  $Y$  jest dyskretnym zbiorem stanów ponumerowanych wskaźnikiem  $n$ , wówczas równanie Mistrzów przyjmuje postać:

$$\frac{dp_n(t)}{dt} = \sum_n \left\{W_{nn'}(t) p_{n'}(t) - W_{n'n}(t) p_n(t)\right\}'. \quad (19)$$



W tej postaci równanie *Mistrzów* jest równaniem "zysku i straty" opisującym "wzrost i spadek" prawdopodobieństwa każdego stanu  $n$ ; pierwszym wyrazem jest wzrost prawdopodobieństwa związany z przejściami z innych stanów  $n'$ , a drugim spadek związany z przejściami do innych stanów.

# Rozdział 3

## Wpływ liczby kopii genu na jego regulację z uwzględnieniem efektów stochastycznych

### 3.1 Stochastyczność w procesie ekspresji genów

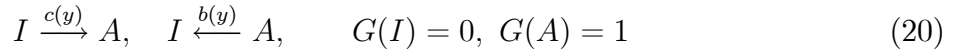
Stochastyczność w procesie ekspresji genów powstaje w wyniku przerw w aktywności genu (przełączania stanu genu) [12], transkrypcji i degradacji mRNA oraz translacji i degradacji białka ([16], [11], [28], [29], [9] oraz [17]). Stochastyczność ta powoduje, że populacja komórek wykazuje duże zróżnicowanie pomiędzy komórkami, jak to zaobserwowano na przykład u Takasuka i in. [27] oraz u Stirland i in. [24] dla komórek ssaków, u Raser i O’Shea [22] dla komórek drożdży (*Saccharomyces cerevisiae*) lub u Elowitz i in. [7] dla bakterii (*Escherichia coli*). U bakterii stochastyczność wynika z małej liczby cząsteczek mRNA oraz białka danego typu na komórkę. Komórki eukariontów natomiast są znacznie większe i zawierają znacznie więcej mRNA więc źródła stochastyczności związane są głównie ze zmianą aktywności genów.

### 3.2 Opis modelu ekspresji genu haploidalnego

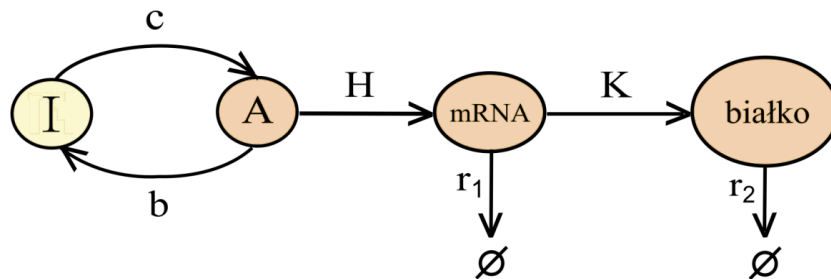
W tej części pracy rozważamy pojedynczą komórkę oraz koncentrujemy się na regulacji pojedynczego genu. Kontrola aktywności genu odbywa się poprzez przyłączanie bądź dysocjację białek regulatorowych. Zakładamy, że współczynniki aktywacji  $c$  i dezaktywacji  $b$  genu zależą od liczby cząsteczek syntetyzowanego białka [10], [14], [15],  $c = c(y)$ ,  $b = b(y)$ , co tworzy

pętlę sprzężenia zwrotnego (autoregulacji) (Rys. 3.1). W przypadku regulacji bez sprzężenia zwrotnego (włączenie lub wyłączenie genu nie zależy od jego białkowych produktów) zakładamy, że gen może przełączyć się do aktywnego stanu (oznaczonego przez  $A$ ) ze stałym współczynnikiem  $c = c_0$ , a wraca do nieaktywnego stanu (oznaczonego przez  $I$ ) ze stałym współczynnikiem  $b = b_0$  w wyniku przyłączenia bądź dysocjacji innych (pochodzących od innego genu) czynników transkrypcyjnych, o których zakłada się, że są obecne w stałej ilości w jądrze komórkowym.

Zakładamy, że tylko aktywny gen ulega transkrypcji. Oznaczmy przez  $H$  tempo transkrypcji mRNA (zsyntetyzowana liczba cząsteczek mRNA na sekundę), oraz przez  $K$  tempo translacji białka. Zakładamy, że tempa  $H$  i  $K$  są stałe. Zatem cząsteczki mRNA syntetyzowane są z tempem  $G(t)H$ , gdzie  $G$  jest binarną zmienną losową opisującą stan genu:  $G(A) = 1$  i  $G(I) = 0$  [12], [10], [20], [15]. Translacja białka zachodzi z tempem  $Kx(t)$ , gdzie  $x(t)$  jest liczbą cząsteczek mRNA; liczbę cząsteczek białka oznaczmy przez  $y(t)$ . Ponadto zakładamy, że cząsteczki mRNA i białka ulegają degradacji ze współczynnikami odpowiednio  $r_1$ ,  $r_2$ . Opisane reakcje można przedstawić następująco:



gdzie symbol  $\phi$  oznacza degradację.



Rysunek 3.1: Schematyczne przedstawienie ekspresji pojedynczego genu.

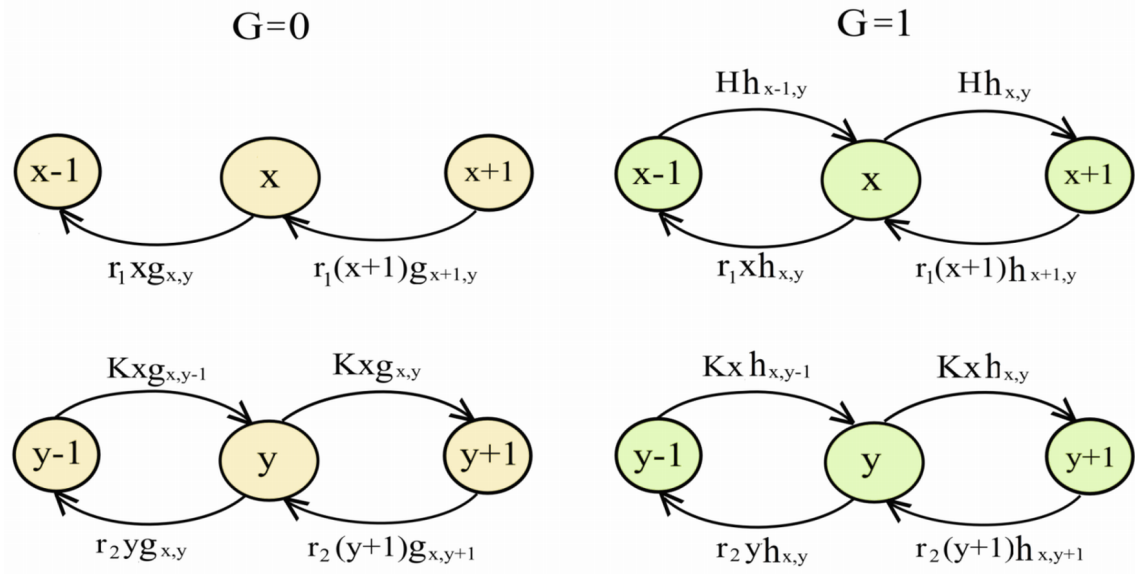
### 3.2.1 Model dyskretny: równania Mistrzów

Stan układu w pojedynczej komórce w dowolnej chwili czasu opisany jest przez trójkę zmiennych  $x(t)$ ,  $y(t)$ ,  $G(t)$ , gdzie  $x(t)$  jest liczbą cząsteczek mRNA,  $y(t)$  jest liczbą cząsteczek białka, a  $G$  jest zmienną binarną opisującą stan genu. Zdefiniujmy prawdopodobieństwa

$$g_{xy}(t) : = P[\# \text{ mRNA} = x; \# \text{ białka} = y; G = 0], \quad (23a)$$

$$h_{xy}(t) : = P[\# \text{ mRNA} = x; \# \text{ białka} = y; G = 1], \quad (23b)$$

$g_{xy}$  i  $h_{xy}$  są prawdopodobieństwami łącznymi, że liczba cząsteczek mRNA wynosi  $x$ , liczba cząsteczek białka wynosi  $y$ , gdzie  $x \in \mathbb{N}$ ,  $y \in \mathbb{N}$ , a stan genu jest, odpowiednio, nieaktywny ( $G = 0$ ) oraz aktywny ( $G = 1$ ). Rozkład całkowity  $\rho_{xy} = g_{xy} + h_{xy}$  opisuje łączny poziom mRNA oraz białka w komórce bez względu na stan genu,  $\sum_{x \in \mathbb{N}, y \in \mathbb{N}} \rho_{xy} = 1$ .



Rysunek 3.2: Diagramy ewolucji gęstości prawdopodobieństw  $g_{xy}$  i  $h_{xy}$  zadanych przez proces Markowa pierwszego rzędu.

Ewolucja w czasie prawdopodobieństw (23a)-(23b) [30], [10] dana jest przez następujący nieskończony układ równań Mistrzów (Rys. 3.2)

$$\begin{aligned} \frac{dg_{xy}}{dt} = & bh_{xy} - cg_{xy} + G(I)Hg_{(x-1,y)} + r_1(x+1)g_{(x+1,y)} - (G(I)H + r_1x)g_{xy} + \\ & + Kxg_{(x,y-1)} + r_2(y+1)g_{(x,y+1)} - (Kx + r_2y)g_{xy}, \end{aligned} \quad (24)$$

$$\begin{aligned} \frac{dh_{xy}}{dt} = & -bh_{xy} + cg_{xy} + G(A)Hh_{(x-1,y)} + r_1(x+1)h_{(x+1,y)} - (G(A)H + r_1x)h_{xy} \\ & + Kxh_{(x,y-1)} + r_2(y+1)h_{(x,y+1)} - (Kx + r_2y)h_{xy}. \end{aligned} \quad (25)$$

Pierwsze dwa prawostronne człony w równaniach (24)-(25) odpowiadają przepływowi prawdopodobieństwa związanego ze zmianą aktywności genu (przejściem pomiędzy  $g_{xy}$  i  $h_{xy}$ ), następne trzy człony opisują syntezę i degradację cząsteczek mRNA, zaś pozostałe trzy człony opisują syntezę i degradację cząsteczek białka. Zauważmy, że w równaniu (24) człony syntezy mRNA zerują się. Równania (24)-(25) dają zależny od czasu rozkład procesu stochastycznego.

### 3.2.2 Średnie i wariancje poziomów mRNA i białka dla modelu dyskretnego bez autoregulacji

Pomimo, że równań (24)-(25) nie potrafimy rozwiązać analitycznie nawet dla przypadku stacjonarnego,  $\frac{dg_{xy}}{dt} = \frac{dh_{xy}}{dt} = 0$ , to można wykorzystać je do obliczenia momentów rozkładów  $g_{xy}$ ,  $h_{xy}$  wykorzystując funkcje tworzące, które zdefiniowane są następująco [19]:

$$F(z, s) = \sum_{x,y} z^x s^y g_{xy}, \quad (26)$$

$$T(z, s) = \sum_{x,y} z^x s^y h_{xy}. \quad (27)$$

Jak pokażemy w Dodatku A1, dla haploidalnego genu bez sprzężenia zwrotnego (tzn. ze stałymi współczynnikami zmiany stanu genu  $c(y) = c_0$  i  $b(y) = b_0$ ) wartość oczekiwana liczby

cząsteczek mRNA i białka dla stacjonarnego układu (24)-(25) wynosi:

$$E_D[x] = \frac{c_0 H}{r_1(c_0 + b_0)} \quad (28)$$

$$E_D[y] = \frac{c_0 K H}{r_1 r_2 (c_0 + b_0)} = \frac{K}{r_2} E_D[x] \quad (29)$$

zaś wariancja liczby cząsteczek mRNA i białka wynosi:

$$Var_D[x] = \frac{b_0 r_1}{c_0(c_0 + b_0 + r_1)} E_D^2[x] + E_D[x] \quad (30)$$

$$Var_D[y] = \frac{b_0 r_1 r_2 (r_1 + r_2 + c_0 + b_0)}{c_0 (r_1 + r_2) (r_1 + c_0 + b_0) (r_2 + c_0 + b_0)} E_D^2[y] + \frac{K}{r_1 + r_2} E_D[y] + E_D[y] \quad (31)$$

### 3.2.3 Przybliżenie ciągłe

Aby uprościć zagadnienie wprowadzamy przybliżenie, spełnione u eukariontów, w którym stochastyczność związana z syntezą/degradacją mRNA oraz białka jest zaniedbana (patrz [14],[15],[19]). Przybliżenie to opiera się na fakcie, że typowe liczby cząsteczek mRNA i białek są znacznie większe niż liczba kopii genu. W rezultacie, fluktuacje związane ze stanem genu dają znacznie większy wkład do wariancji białka niż synteza mRNA i białka oraz procesy degradacji (patrz [19]). Zaniedbując stochastyczność syntezy oraz degradacji mRNA i białka traktujemy  $x$  i  $y$  jako ciągłe zmienne,  $x \in \mathbb{R}$  i  $y \in \mathbb{R}$ , oraz opisujemy proces (21)-(22) równaniami różniczkowymi. Dla haploidalnego genu otrzymujemy zatem następujący układ:

$$I \xrightarrow{c(y)} A, \quad I \xleftarrow{b(y)} A, \quad G(I) = 0, \quad G(A) = 1 \quad (32)$$

$$\frac{dx}{dt} = HG(t) - r_1 x, \quad (33)$$

$$\frac{dy}{dt} = Kx - r_2 y, \quad (34)$$

Równanie (33) opisuje proces transkrypcji oraz degradacji mRNA i jest stochastycznym (losowym) równaniem różniczkowym z losową zmienną binarną  $G(t)$ . Równanie (34), sprzężone

z (33), jest równaniem stochastycznym opisującym proces translacji i degradacji białka. Zarówno liczba cząsteczek mRNA jak i białka są ciągłymi zmiennymi losowymi, ograniczonymi przez rozwiązania stacjonarne równań (33)-(34): ograniczenie dolne dla nieaktywnego stanu genu ( $G = 0$ ):  $x = 0$ ,  $y = 0$ , oraz ograniczenie górne dla aktywnego stanu genu ( $G = 1$ ):  $x \in \left[0, \frac{H}{r_1}\right]$ ,  $y \in \left[0, \frac{KH}{r_1 r_2}\right]$ . Równania (33)-(34) generują zatem kawałkami deterministyczny, ciągły w czasie proces Markowa.

### 3.2.4 Równania ciągłości dla funkcji gęstości prawdopodobieństwa (FGP) poziomów mRNA i białka

W każdej chwili czasu  $t$  realizacje poziomów mRNA i białka,  $x(t)$  i  $y(t)$ , oraz stan przełącznika transkrypcyjnego  $G(t)$  tworzą trójkę zmiennych losowych, z których pierwsze dwie są ciągłe a trzecia jest binarna. Zatem ich wspólny rozkład łączny może być reprezentowany przez parę funkcji gęstości prawdopodobieństwa (FGP)  $g(x, y, t)$  i  $h(x, y, t)$  zdefiniowanych następująco:

$$\begin{aligned} g(x, y, t)\Delta x\Delta y: &= \Pr[x(t) \in (x, x + \Delta x); y(t) \in (y, y + \Delta y); G(t) = 0], \\ h(x, y, t)\Delta x\Delta y: &= \Pr[x(t) \in (x, x + \Delta x); y(t) \in (y, y + \Delta y); G(t) = 1]. \end{aligned}$$

Wyprowadzimy teraz równania ciągłości dla funkcji  $g(x, y, t)$  i  $h(x, y, t)$ .

Napiszmy równania bilansu dla prawdopodobieństw  $g$  i  $h$  z członami źródłowymi wynikającymi ze zmiany stanu genu:

$$\begin{aligned} \int_{A_{t+\Delta t}} g(x, y, t + \Delta t) dx dy - \int_{A_t} g(x, y, t) dx dy &= \Delta t \int_{A_t} b(y) h(x, y, t) dx dy + \quad (35) \\ &\quad - \Delta t \int_{A_t} c(y) g(x, y, t) dx dy \end{aligned}$$

$$\begin{aligned} \int_{A_{t+\Delta t}} h(x, y, t + \Delta t) dx dy - \int_{A_t} h(x, y, t) dx dy &= -\Delta t \int_{A_t} b(y) h(x, y, t) dx dy + \quad (36) \\ &\quad + \Delta t \int_{A_t} c(y) g(x, y, t) dx dy. \end{aligned}$$

Dzieląc (35)-(36) stronami przez  $\Delta t$  otrzymujemy iloraz różnicowy:

$$\frac{\int_{A_{t+\Delta t}} g(x, y, t + \Delta t) dx dy - \int_{A_t} g(x, y, t) dx dy}{\Delta t} = \int_{A_t} b(y) h(x, y, t) dx dy - \int_{A_t} c(y) g(x, y, t) dx dy$$

$$\frac{\int_{A_{t+\Delta t}} h(x, y, t + \Delta t) dx dy - \int_{A_t} h(x, y, t) dx dy}{\Delta t} = - \int_{A_t} b(y) h(x, y, t) dx dy + \int_{A_t} c(y) g(x, y, t) dx dy,$$

następnie, przechodząc do granicy  $\Delta t \rightarrow 0$ , otrzymujemy:

$$\frac{d}{dt} \int_{A_t} g(x, y, t) dx dy = \int_{A_t} b(y) h(x, y, t) dx dy - \int_{A_t} c(y) g(x, y, t) dx dy$$

$$\frac{d}{dt} \int_{A_t} h(x, y, t) dx dy = - \int_{A_t} b(y) h(x, y, t) dx dy + \int_{A_t} c(y) g(x, y, t) dx dy.$$

Obszar całkowania  $A_t$  zależy od czasu  $t$ , stosujemy więc twierdzenie Gaussa-Ostrogradzkiego, zgodnie z którym wchodzimy z pochodną pod znak całki:

$$\int_{A_t} \left\{ \frac{\partial g(x, y, t)}{\partial t} + \operatorname{div} \left[ \left( \frac{dx}{dt} /_{G=0}, \frac{dy}{dt} \right) g \right] \right\} dx dy = \int_{A_t} [b(y) h(x, y, t) - c(y) g(x, y, t)] dx dy$$

$$\int_{A_t} \left\{ \frac{\partial h(x, y, t)}{\partial t} + \operatorname{div} \left[ \left( \frac{dx}{dt} /_{G=1}, \frac{dy}{dt} \right) h \right] \right\} dx dy = \int_{A_t} [-b(y) h(x, y, t) + c(y) g(x, y, t)] dx dy,$$

stąd

$$\int_{A_t} \left\{ \frac{\partial g(x, y, t)}{\partial t} + \operatorname{div} \left[ \left( \frac{dx}{dt} /_{G=0}, \frac{dy}{dt} \right) g \right] - [b(y) h(x, y, t) - c(y) g(x, y, t)] \right\} dx dy = 0$$

$$\int_{A_t} \left\{ \frac{\partial h(x, y, t)}{\partial t} + \operatorname{div} \left[ \left( \frac{dx}{dt} /_{G=1}, \frac{dy}{dt} \right) h \right] - [-b(y) h(x, y, t) + c(y) g(x, y, t)] \right\} dx dy = 0.$$

Ponieważ obszar całkowania  $A_t$  jest dowolny, więc funkcja podcałkowa musi być równa zero, otrzymujemy zatem następujące równania ciągłości dla  $g(x, y, t)$  i  $h(x, y, t)$ :

$$\frac{\partial g}{\partial t} + \operatorname{div} \left[ \left( \frac{dx}{dt} /_{G=0}, \frac{dy}{dt} \right) g \right] = bh - cg \quad (37)$$

$$\frac{\partial h}{\partial t} + \operatorname{div} \left[ \left( \frac{dx}{dt} /_{G=1}, \frac{dy}{dt} \right) h \right] = -bh + cg. \quad (38)$$



Uwzględniając równania na pola prędkości  $\frac{dx}{dt}|_{G=0}$  i  $\frac{dx}{dt}|_{G=1}$  dane przez związki (33)-(34), otrzymujemy równania ciągłości w postaci:

$$\frac{\partial g}{\partial t} - \frac{\partial}{\partial x}(r_1 x g) + \frac{\partial}{\partial y}[(Kx - r_2 y) g] = bh - cg \quad (39)$$

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial x}[(H - r_1 x) h] + \frac{\partial}{\partial y}[(Kx - r_2 y) h] = -bh + cg. \quad (40)$$

Alternatywnie, powyższe dwa równania mogą być otrzymane z równań Mistrzów (24)-(25) poprzez rozwinięcie Kramersa-Moyala (pierwszego rzędu) [30]. Równań (39)-(40) nie potrafimy rozwiązać analitycznie, numerycznie równania te zostały rozwiązane w pracy Lipniacki i in. [15].

Funkcje  $g(x, y, t)$ ,  $h(x, y, t)$  można badać w obszarze  $D = [0, \infty) \times [0, \infty)$ . Zawężymy jednak nasze rozważania do obszaru  $D_1 \subset D$  ograniczonego przez następujące rozwiązania układu (33)-(34):  $(x_1(t), y_1(t))$  dla  $G = 0$  z warunkiem początkowym  $(x_1(0), y_1(0)) = (1, 1)$  oraz  $(x_2(t), y_2(t))$  dla  $G = 1$  z warunkiem początkowym  $(x_2(0), y_2(0)) = (0, 0)$ . Rozwiązania te zadają dwie krzywe w płaszczyźnie  $(x, y)$  łączące się w punkcie  $(0, 0)$  i  $(1, 1)$  (patrz Rys. 3.4). Obszar  $D_1$ , jak pokazano w [3], jest niezmienniczy.

### 3.2.5 Średnie i wariancje poziomów mRNA i białka dla modelu ciągłego bez autoregulacji

Policzmy średnie oraz wariancje mRNA i białka dla układu (33)-(34) w przypadku bez sprzężenia zwrotnego,  $c(y) = c_0$ ,  $b(y) = b_0$ . Równania (39)-(40) wykorzystamy do obliczenia momentów rozkładów  $g_{xy}$ ,  $h_{xy}$ . Podobnie jak w przypadku dyskretnym, definiujemy funkcje tworzące:

$$F(z, s) = \int_0^\infty \int_0^\infty e^{zx+sy} g(x, y) dx dy \quad (41)$$

$$T(z, s) = \int_0^\infty \int_0^\infty e^{zx+sy} h(x, y) dx dy. \quad (42)$$

Pierwsze i drugie pochodne  $F(z, s)$  i  $T(z, s)$  w punkcie  $z = s = 0$  dają odpowiednio momenty pierwszego i drugiego rzędu dla  $g(x, y, t)$  i  $h(x, y, t)$ . Pozwala to policzyć średnią i wariancję rozkładu brzegowego białka  $\rho_y = \int (g(x, y) + h(x, y)) dx$  dla modelu ciągłego [19] (patrz Dodatek A2):

$$E_C[x] = \frac{c_0 H}{r_1(c_0 + b_0)} \quad (43)$$

$$E_C[y] = \frac{c_0 H K}{r_1 r_2 (c_0 + b_0)} = \frac{K}{r_2} E_C[x] \quad (44)$$

$$Var_C[x] = \frac{c_0 b_0 H^2}{r_1^2 (r_1 + c_0 + b_0) (c_0 + b_0)^2} = \frac{b_0 H}{r_1 (r_1 + c_0 + b_0) (c_0 + b_0)} E_C[x] \quad (45)$$

$$Var_C[y] = \frac{b_0 r_1 r_2 (r_1 + r_2 + c_0 + b_0)}{c_0 (r_1 + r_2) (r_1 + c_0 + b_0) (r_2 + c_0 + b_0)} (E_C[y])^2. \quad (46)$$

Zauważmy, że otrzymaliśmy taką samą średnią dla białka jak w przypadku dyskretnym (29), zaś wariancja  $Var_C[y]$  różni się od wariancji w przypadku dyskretnym tylko członami liniowymi: pierwszy brakujący człon  $\frac{K}{r_1 + r_2} E[y]$  odpowiada transkrypcyjnemu szumowi, podczas gdy drugi człon  $E[y]$  odpowiada szumowi translacyjnemu.

### 3.2.6 Przybliżenie Keplera-Elstona (K-E) modelu ciągłego

Rozważając ciągle przybliżenie, układ (33)- (34) może być dalej uproszczony jeżeli założymy, że tempo degradacji mRNA jest duże w porównaniu z tempem przełączania genu, wówczas poziom mRNA osiąga stan równowagi znacznie szybciej niż poziom białka. W takim przypadku równanie (33) może zostać zastąpione przez swój stan stacjonarny  $x = \frac{HG}{r_1}$  (patrz Dodatek B).

Otrzymujemy wówczas

$$\frac{dy}{dt} = \frac{HKG}{r_1} - r_2 y. \quad (47)$$

Powyższe przybliżenie jest równoważne założeniu wprowadzonemu przez Keplera i Elstona (2001) [10], że białko jest syntetyzowane bezpośrednio z genu (przybliżenie K-E). Zróbmy

następującą zamianę zmiennych w równaniu (47):  $y^* = \frac{r_1 r_2}{HK} y$ ,  $t^* = r_2 t$ ; po opuszczeniu gwiazdek otrzymujemy równanie w bezwymiarowej postaci:

$$\frac{dy}{dt} = G - y. \quad (48)$$

Rozkład łączny dla losowych zmiennych  $(y, G)$ , opisujących stan genu w dowolnej chwili czasu, dany jest przez następującą parę FGP:

$$\Pr[y(t) \in (y, y + \Delta y), G(t) = 0] = g(y, t) \Delta y$$

$$\Pr[y(t) \in (y, y + \Delta y), G(t) = 1] = h(y, t) \Delta y.$$

Wówczas równania ciągłości dla FGP  $g(y, t)$  i  $h(y, t)$ , dla pola porędkości  $\frac{dy}{dt}$  zadanego przez (47), można zapisać:

$$\frac{\partial g}{\partial t} - \frac{\partial}{\partial y}(yg) = bh - cg \quad (49)$$

$$\frac{\partial h}{\partial t} + \frac{\partial}{\partial y}((1-y)h) = -bh + cg. \quad (50)$$

### 3.2.7 Średnia i wariancja poziomego białka dla modelu K-E bez autoregulacji

Średnia  $E_{KE}[y]$  oraz wariancja  $Var_{KE}[y]$  stacjonarnego rozkładu brzegowego białka wynoszą

$$E_{KE}[y] = \frac{c_0}{c_0 + b_0}, \quad (51)$$

$$Var_{KE}[y] = \frac{b_0}{c_0(1 + c_0 + b_0)} (E_{KE}[y])^2. \quad (52)$$

Zauważmy, że średnia  $E_{KE}[y]$  (przy  $\frac{r_1 r_2}{HK} = 1$ ) jest taka sama jak w poprzednim przypadku.

## 3.3 Analiza ekspresji dla genu diploidalnego

Rozważmy teraz autoregulowany gen z dwiema kopiami biorącymi udział w procesie transkrypcji. Oznacza to, że stan genu przyjmuje teraz trzy wartości,  $G \in \{0, 1, 2\}$ , tzn.  $G = 0$

oznacza, że obie kopie genu są wyłączone,  $G = 1$  oznacza, że jedna kopia genu jest aktywna a druga wyłączona, oraz  $G = 2$  oznacza, że obie kopie genu są aktywne.

### Równania ciągłości dla FGP poziomów mRNA i białek

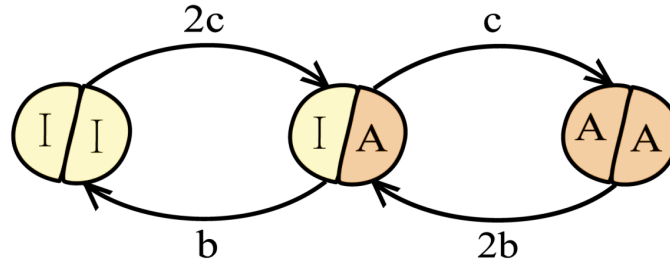
Równania (33)-(34) generują stochastyczne trajektorie opisane przez ciągły w czasie proces Markowa. W każdej chwili czasu  $t$  realizacje poziomów mRNA i białka  $x(t)$  i  $y(t)$  oraz przełącznika transkrypcji  $G(t)$  są trójką losowych zmiennych, z których pierwsze dwie są ciągłe a trzecia przyjmuje trzy wartości. Stąd ich rozkład łączny może być opisany przez trzy funkcje gęstości prawdopodobieństwa  $f_0(x, y, t)$ ,  $f_1(x, y, t)$ ,  $f_2(x, y, t)$ :

$$f_0(x, y, t)\Delta x\Delta y: = \Pr[x(t) \in (x, x + \Delta x); y(t) \in (y, y + \Delta y); G(t) = 0]$$

$$f_1(x, y, t)\Delta x\Delta y: = \Pr[x(t) \in (x, x + \Delta x); y(t) \in (y, y + \Delta y); G(t) = 1]$$

$$f_2(x, y, t)\Delta x\Delta y: = \Pr[x(t) \in (x, x + \Delta x); y(t) \in (y, y + \Delta y); G(t) = 2].$$

Przepływ prawdopodobieństw  $f_0$ ,  $f_1$  i  $f_2$  związanych ze zmianą aktywności kopii genów obrazuje Rys. 3.3.



Rysunek 3.3: Schemat przepływu prawdopodobieństw dla układu z dwiema kopiami genu.

Równania ciągłości dla funkcji  $f_0$ ,  $f_1$ ,  $f_2$  mają zatem postać:

$$\frac{\partial f_0}{\partial t} + \text{div}(f_0(\frac{dx}{dt}/_{G=0}, \frac{dy}{dt})) = bf_1 - 2cf_0 \quad (53)$$

$$\frac{\partial f_1}{\partial t} + \text{div}(f_1(\frac{dx}{dt}/_{G=1}, \frac{dy}{dt})) = -(b+c)f_1 + 2cf_0 + 2bf_2 \quad (54)$$

$$\frac{\partial f_2}{\partial t} + \text{div}(f_2(\frac{dx}{dt}/_{G=2}, \frac{dy}{dt})) = cf_1 - 2bf_2. \quad (55)$$

Podstawiając równania na pola prędkości  $\frac{dx}{dt}$  i  $\frac{dy}{dt}$  zadane przez (33)-(34) otrzymujemy:

$$\frac{\partial f_0}{\partial t} - r_1 x \frac{\partial f_0}{\partial x} + (Kx - r_2 y) \frac{\partial f_0}{\partial y} = b f_1 - (2c - r_1 - r_2) f_0 \quad (56)$$

$$\frac{\partial f_1}{\partial t} + (H - r_1 x) \frac{\partial f_1}{\partial x} + (Kx - r_2 y) \frac{\partial f_1}{\partial y} = -(b + c - r_1 - r_2) f_1 + 2c f_0 + 2b f_2 \quad (57)$$

$$\frac{\partial f_2}{\partial t} + (2H - r_1 x) \frac{\partial f_2}{\partial x} + (Kx - r_2 y) \frac{\partial f_2}{\partial y} = c f_1 - (2b - r_1 - r_2) f_2. \quad (58)$$

Funkcje  $f_0(x, y, t)$ ,  $f_1(x, y, t)$ ,  $f_2(x, y, t)$  można badać w zwartym obszarze  $D_2 \subset D = [0, \infty) \times [0, \infty)$  zdefiniowanym poniżej, który jak pokażemy w Dodatku C, jest zbiorem niezmienniczym. Rozwiązania układu (33)-(34):  $(x_1(t), y_1(t))$  dla  $G = 0$  z warunkiem początkowym  $(x_1(0), y_1(0)) = (2, 2)$  oraz  $(x_2(t), y_2(t))$  dla  $G = 2$  z warunkiem początkowym  $(x_2(0), y_2(0)) = (0, 0)$  zadają dwie krzywe w płaszczyźnie  $(x, y)$  łączące się w punkcie  $(0, 0)$  i  $(2, 2)$  (patrz Rys. 3.4).

### Średnia oraz wariancja białka dla układu bez autoregulacji

W układzie bez autoregulacji kopie genu "działają" niezależnie od siebie. Zatem wartość oczekiwana  $E_{II}[y]$  oraz wariancja  $Var_{II}[y]$  stacjonarnego rozkładu brzegowego białka dla układu z dwiema kopiami genu bez autoregulacji dana jest przez relacje:

$$E_{II}[y] = 2E_I[y] \quad (59)$$

$$Var_{II}[y] = 2Var_I[y], \quad (60)$$

gdzie  $E_I[y]$  oraz  $Var_I[y]$  są odpowiednio średnią oraz wariancją rozkładu białka dla układu z jedną kopią genu.

Niezależne obliczenia dla sprawdzenia poprawności związków (59)-(60) dla układów z jedną i dwiema kopiami genu w przybliżeniu K-E znajdują się w Dodatku A3.

### 3.4 Analiza ekspresji genu $N$ -ploidalnego

$N$ -ploidalny gen może istnieć w jednym z  $N + 1$  stanów. Dyskretna funkcja  $G_N(t) \in \{0, 1, \dots, N\}$ , gdzie  $G_N(t) = i$ , odpowiada stanowi, w którym  $i$  z  $N$  kopii jest aktywnych. Niech  $f_i(x, y, t)$ ,  $i \in \mathbb{Z}$ , oznacza FGP taką, że

$$P[x(t) \in (x, x + \Delta x), y(t) \in (y, y + \Delta y), G_N(t) = i] = f_i(x, y, t)\Delta x\Delta y.$$

Oczywiście  $f_i(x, y, t) = 0$  dla  $i \notin \{0, 1, \dots, N\}$ . Macierz przejścia  $T_N$  pomiędzy stanami  $f_i$  odpowiada współczynnikom przejścia dla pojedynczej kopii genu:

$$I \xrightarrow{c(y)} A, \quad I \xleftarrow{b(y)} A,$$

i ma postać:

$$T_N = \begin{array}{|c|c|c|c|} \hline & f_{i-1} & f_i & f_{i+1} \\ \hline f_{i-1} & & (N + 1 - i)c(y) & 0 \\ \hline f_i & ib(y) & & (N - i)c(y) \\ \hline f_{i+1} & 0 & (i + 1)b(y) & \\ \hline \end{array}, \quad i \in \{0, 1, \dots, N\}.$$

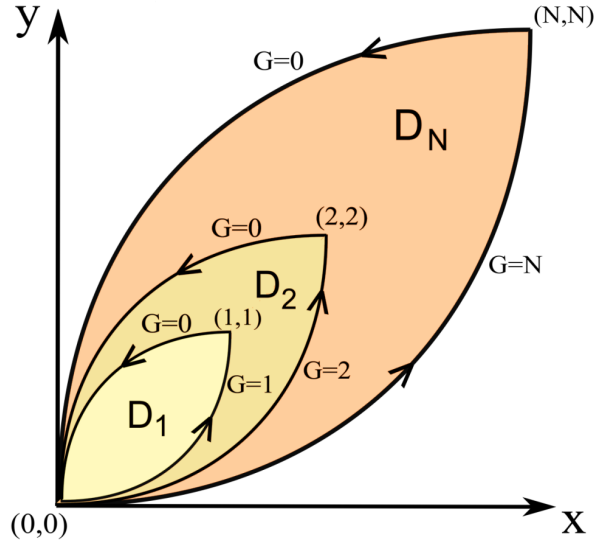
Ewolucja w czasie  $x$  i  $y$  dana jest przez równania:

$$\frac{dx}{dt} = HG_N(t) - r_1x \quad (61)$$

$$\frac{dy}{dt} = Kx - r_2y, \quad (62)$$

zaś ewolucja w czasie funkcji gęstości prawdopodobieństwa  $f_i$  dana jest przez następujący układ  $N + 1$  równań różniczkowych cząstkowych:

$$\begin{aligned} \frac{\partial f_i}{\partial t} + \frac{\partial}{\partial x}((Hi - r_1x)f_i) + \frac{\partial}{\partial y}((Kx - r_2y)f_i) &= -ib(y)f_i - (N - i)c(y)f_i + \\ &+ (N + 1 - i)c(y)f_{i-1} + (i + 1)b(y)f_{i+1}. \end{aligned}$$



Rysunek 3.4: Obszary  $D_1, D_2, \dots, D_N$  ograniczone przez trajektorie (61)-(62) dla  $G_N(t) \in \{0, 1, \dots, N\}$ .

Funkcje  $f_i(x, y, t)$  można badać w ograniczonym obszarze  $D_N \subset D = [0, \infty) \times [0, \infty)$ , zadanym w płaszczyźnie  $(x, y)$  przez rozwiązania układu (61)-(62):  $(x_1(t), y_1(t))$  dla  $G = 0$  z warunkiem początkowym  $(x_1(0), y_1(0)) = (N, N)$ , oraz  $(x_2(t), y_2(t))$  dla  $G_N = N$  z warunkiem początkowym  $(x_2(0), y_2(0)) = (0, 0)$  (Rys. 3.4). W podobny sposób, jak dla  $D_2$ , można pokazać, że zbiór  $D_N$  również jest zbiorem niezmienniczym.

### Średnia i wariancja białka dla układu bez autoregulacji

W układzie bez autoregulacji kopie genu "działają" niezależnie od siebie. Zatem, dla genu mającego  $N$  kopii, wartość oczekiwana  $E_N[y]$  oraz wariancja  $Var_{IN}[y]$  stacjonarnego rozkładu brzegowego białka dla układu bez autoregulacji dana jest przez relacje:

$$E_N[y] = NE_I[y] \quad (63)$$

$$Var_N[y] = NVar_I[y]. \quad (64)$$

Średnia i wariancja liczone na kopię genu,  $E_0 = \frac{E_N}{N}$  i  $Var_0 = \frac{Var_N}{N}$ , nie zależą od liczby kopii genu.

W przypadku z autoregulacją białko wyprodukowane z mRNA pochodzącego z pierwszej kopii genu może dodatnio bądź ujemnie regulować drugą kopię genu, w związku z czym z drugiej kopii genu może być wyprodukowana inna ilość białka niż z pierwszej, dlatego nie ma generalnej zasady wiążącej wariancję oraz średnią z liczbą kopii genu.

### Przybliżenie Keplera-Elstona

Rozważmy model regulacji  $N$ -ploidalnego genu w przybliżeniu K-E, który wykorzystamy do obliczeń numerycznych, a mianowicie:

$$\frac{dy}{dt} = G_N - y, \quad G_N \in \{0, 1, \dots, N\} \quad (65)$$

Łączny rozkład dla losowych zmiennych  $(y, G_N)$  opisany jest przez  $N + 1$  FGP:

$$f_i(y, t)\Delta y := \Pr[y(t) \in (y, y + \Delta y), G_N(t) = i].$$

Stacjonarne rozkłady  $\frac{\partial f_i}{\partial t} = 0$  dane są przez układ liniowych równań różniczkowych zwyczajnych

$$\frac{d}{dy}((i - y)f_i) = -ib(y)f_i - (N - i)c(y)f_i + (N + 1 - i)c(y)f_{i-1} + (i + 1)b(y)f_{i+1}. \quad (66)$$



# Rozdział 4

## Wyniki analityczne

Dla układu z autoregulacją rozwiązanie analityczne potrafimy znaleźć jedynie w przypadku przybliżenia K-E dla układu z jedną kopią genu. W pozostałych przypadkach układy z autoregulacją rozwiązujemy numerycznie. Dla układów bez autoregulacji, znając rozwiązanie układu z jedną kopią genu, potrafimy znaleźć metodą splotów rozwiązania analityczne układów z dowolną liczbą kopii genów.

### 4.1 Rozwiązanie stacjonarne układu K-E dla $N=1$ z autoregulacją

Układ (49)-(50) posiada analityczne rozwiązanie stacjonarne dla  $y \in (0, 1)$ . Dla  $\frac{\partial g}{\partial t} = \frac{\partial h}{\partial t} = 0$  układ ten ma postać:

$$-\frac{d}{dy}(yg) = bh - cg \quad (67)$$

$$\frac{d}{dy}((1-y)h) = -bh + cg. \quad (68)$$

Dodając stronami (67)-(68) otrzymujemy całkę pierwszą

$$\frac{d}{dy}[-yg + (1-y)h] = 0, \quad (69)$$

która implikuje warunek

$$-yg + (1-y)h = -g(1) = h(0). \quad (70)$$

Ponieważ funkcje  $g(y)$  i  $h(y)$  są z definicji nieujemne, warunek  $-g(1) = h(0)$  implikuje, że  $g(1) = h(0) = 0$ . Zatem, z równania (70), wyznaczamy  $h = \frac{yg}{(1-y)}$ . Po podstawieniu  $h$  do

równania (67) otrzymujemy:

$$\frac{d}{dy}(yg) = \frac{-by}{1-y}g + cg,$$

lub, zapisując w równoważnej postaci:

$$\frac{dg}{g} = \left( \frac{-b}{1-y} + \frac{c-1}{y} \right) dy.$$

Zatem układ (67)-(68) dla  $y \in (0, 1)$  ma analityczne rozwiązanie stacjonarne postaci

$$g(y) = \text{Exp} \left[ \int_0^y \left( \frac{-b}{(1-s)} + \frac{c-1}{s} \right) ds \right], \quad h(y) = \frac{yg(y)}{(1-y)}. \quad (71)$$

Dla współczynników zmiany stanu genu  $c(y) = c_0 + c_1y + c_2y^2$  i  $b(y) = b_0 + b_1y + b_2y^2$  rozwiązanie ma postać:

$$g(y) = A \cdot \exp^{[y(b_1+b_2+c_1) + \frac{1}{2}y^2(b_2+c_2)]} \cdot y^{(c_0-1)} \cdot (1-y)^{(b_0+b_1+b_2)} \quad (72)$$

$$h(y) = A \cdot \exp^{[y(b_1+b_2+c_1) + \frac{1}{2}y^2(b_2+c_2)]} \cdot y^{c_0} \cdot (1-y)^{(b_0+b_1+b_2-1)}, \quad (73)$$

oraz

$$\rho^{(1)}(y) = g(y) + h(y) = A \cdot \exp^{[y(b_1+b_2+c_1) + \frac{1}{2}y^2(b_2+c_2)]} \cdot y^{(c_0-1)} \cdot (1-y)^{(b_0+b_1+b_2-1)},$$

gdzie  $A$  jest stałą wybraną tak, aby  $\int_0^1 \rho(y) = 1$ .

Zauważmy, że rozwiązanie (72)-(73) układu (67)-(68) istnieje, jeżeli spełnione są następujące warunki zapewniające całkowalność  $g(y)$  i  $h(y)$ :

$$\begin{cases} c_0 > 0 \\ b_0 + b_1 + b_2 > 0. \end{cases}$$

Z biologicznego punktu widzenia warunek  $c_0 = 0$  oznaczałby, że gen wyłączy się i może nigdy więcej się nie włączyć. Zauważmy również, że

$$\begin{cases} c_0 < 1 \Rightarrow \lim_{y \rightarrow 0^+} g(y) = \infty \\ b_0 + b_1 + b_2 < 1 \Rightarrow \lim_{y \rightarrow 1^-} h(y) = \infty. \end{cases}$$

## 4.2 Rozwiązanie stacjonarne układu K-E dla $N=2$ bez autoregulacji

W przypadku bez sprzężenia zwrotnego,  $c(y) = c_0$ ,  $b(y) = b_0$ , potrafimy znaleźć analityczne rozwiązanie układu (49)-(50) dla  $\frac{\partial g}{\partial t} = \frac{\partial h}{\partial t} = 0$ . Oznaczmy różnymi kolorami mRNA i białka pochodzące z różnych kopii genu. Jeżeli współczynniki  $c$  i  $b$  nie zależą od  $y$ , wówczas rozkłady dla różnie "pokolorowanych" białek są niezależne. W układzie bez autoregulacji rozkład brzegowy białka jest splotem tych "kolorowych" rozkładów. Rozwiązanie  $f_0(y)$ ,  $f_1(y)$ ,  $f_2(y)$  układu (56)-(58) wyraża się w postaci splotu funkcji  $g(y)$  i  $h(y)$ :

$$f_0(y) = \int_0^y g^*(z)g^*(y-z)dz \quad (74)$$

$$f_1(y) = 2 \int_0^y g^*(z)h^*(y-z)dz \quad (75)$$

$$f_2(y) = \int_0^y h^*(z)h^*(y-z)dz \quad (76)$$

oraz

$$\rho^{(2)}(y) = \int_0^y \rho^{(1)}(z)\rho^{(1)}(y-z)dz,$$

gdzie  $g^* = \begin{cases} g & \text{dla } z \in (0, 1) \\ 0 & \text{w p.p.} \end{cases}$ ,

$\rho^{(2)}$  jest rozkładem brzegowym dla układu z dwiema kopiami genu, oraz  $g(y)$  i  $h(y)$  jest rozwiązaniem układu dla jednej kopii genu (49)-(50) ze stałymi współczynnikami zmiany stanu genu.

Dla rozwiązań  $g(y)$  i  $h(y)$

$$g(y) = y^{c_0-1}(1-y)^{b_0} \quad (77)$$

$$h(y) = y^{c_0}(1-y)^{b_0-1} \quad (78)$$

modelu w przybliżeniu K-E dla jednej kopii genu (49)-(50) ze stałymi współczynnikami zmienny stanu genu, rozwiązanie  $f_0(y)$ ,  $f_1(y)$ ,  $f_2(y)$  układu (56)-(58) przedstawia się następująco:

$$f_0(y) = \int_0^y z^{c_0-1} (1-z)^{b_0} (y-z)^{c_0-1} (1+z-y)^{b_0} dz = \quad (79)$$

$$= \int_0^y (z(y-z))^{c_0-1} ((1-z)(1+z-y))^{b_0} dz$$

$$f_1(y) = 2 \int_0^y z^{c_0} (1-z)^{b_0-1} (y-z)^{c_0-1} (1+z-y)^{b_0} dz \quad (80)$$

$$f_2(y) = \int_0^y z^{c_0} (1-z)^{b_0-1} (y-z)^{c_0} (1+z-y)^{b_0-1} dz = \quad (81)$$

$$+ \int_0^y (z(y-z))^{c_0} ((1-z)(1+z-y))^{b_0-1} dz.$$

## Rozwiązanie stacjonarne układu K-E dla $N$ -ploidalnego genu bez autoregulacji

Tak jak poprzednio, w układzie bez autoregulacji rozkład brzegowy białka jest splotem niezależnych rozkładów "kolorowych" białek. Wykorzystując odpowiednie sploty tych rozkładów możemy znaleźć analityczne rozwiązanie układu (66) dla dowolnego  $N$ . Rozkład brzegowy dla  $N$ -ploidalnego genu bez autoregulacji ma postać

$$\rho^{(N)}(y) = \int_0^y \rho^{(k)}(z) \rho^{(N-k)}(y-z) dz.$$

# Rozdział 5

## Wyniki numeryczne

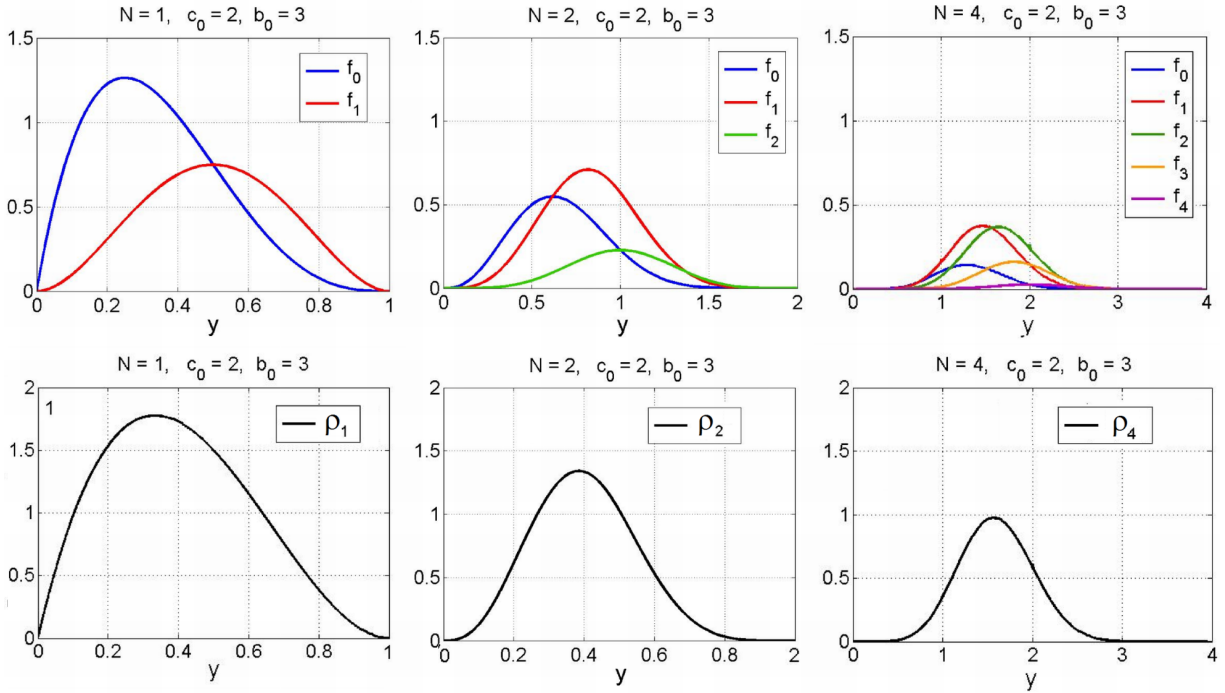
Przeanalizujemy charakterystyczne numeryczne rozwiązania układu (66) dla genu haplo- ( $N = 1$ ), diplo- ( $N = 2$ ) oraz tetraploidalnego ( $N = 4$ ), najpierw dla przypadku bez autoregulacji, a następnie z autoregulacją. Do obliczenia tych rozwiązań wykorzystamy metodę dyskretyzacji (patrz Dodatek D) wprowadzoną w pracy [15].

### 5.1 Charakterystyczne rozwiązania układu bez autoregulacji

Rozważmy rozwiązania stacjonarne  $f_k^{(N)}$  ( $k$  jest liczbą aktywnych kopii w  $N$ -ploidalnym genie) układu (66) bez autoregulacji,  $c(y) = c_0$ ,  $b(y) = b_0$ , dla  $N = 1, 2, 4$  oraz trzech par współczynników zmiany stanu genu  $c_0$  i  $b_0$ :  $c_0 = 2$  i  $b_0 = 3$ ;  $c_0 = 4$  i  $b_0 = 1$ ;  $c_0 = 0.8$  i  $b_0 = 0.8$ .

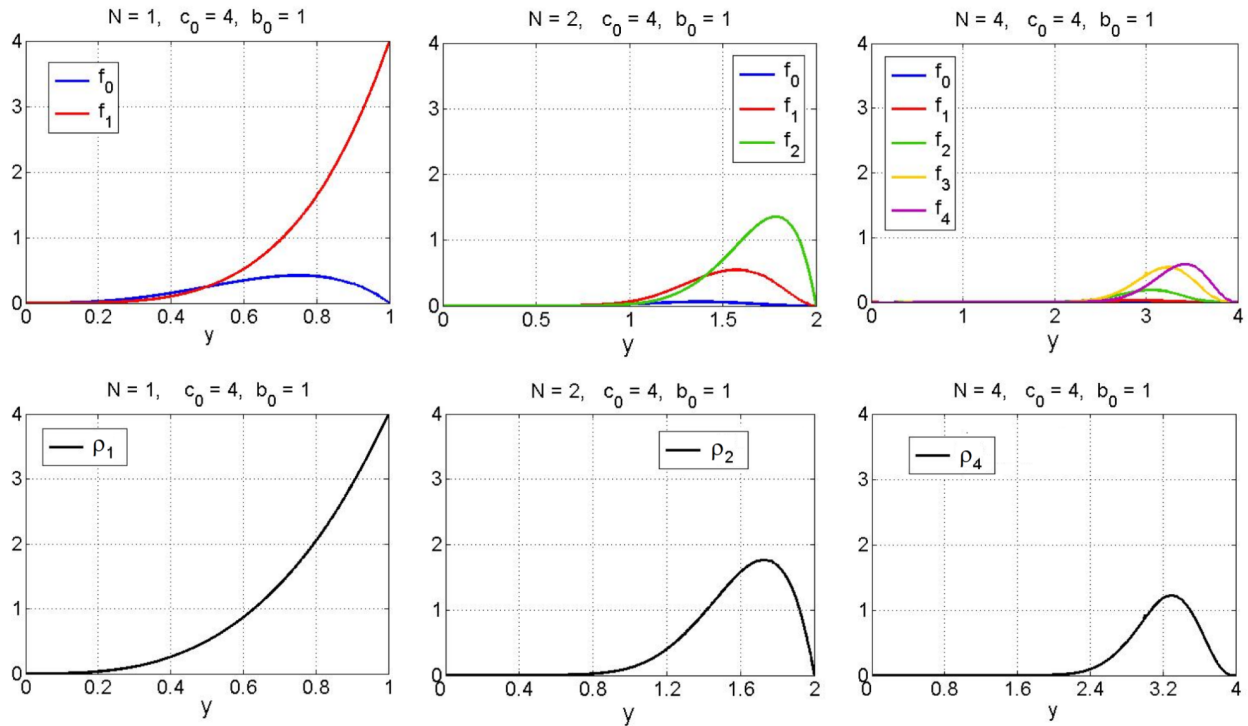
Dla współczynników  $c_0 = 2$ ,  $b_0 = 3$  (Rys. 5.5) funkcja rozkładu białka dla układu z jedną kopią genu odpowiadająca genowi w stanie nieaktywnym,  $f_0$ , osiąga maksimum dla mniejszej wartości  $y$  niż rozkład  $f_1$ , odpowiadający genowi aktywnemu (gdyż białko produkowane jest gdy gen jest włączony). Podobnie, dla układu z dwiema i czterema kopiami genu, pierwsze maksimum osiąga rozkład  $f_0$ , w którym wszystkie kopie genu są nieaktywne, drugie maksimum, dla większego  $y$ , osiąga rozkład  $f_1$ , w którym jedna z kopii jest aktywna, itd. Dla współczynników  $c_0 > 1$  i  $b_0 > 1$  profile rozkładów dla  $N = 1, 2, 4$  mają maksima wewnątrz obszaru oraz znikają na jego brzegach. Maksimum  $\rho(y)$  rozkładu brzegowego przesuwa się w kierunku rosnącego  $y$  ze wzrostem liczby kopii genu w układzie. Współczynnik dezaktywacji  $b_0$  jest większy niż współczynnik aktywacji  $c_0$  dlatego gen w układzie z jedną, dwiema i

czterema kopiami znajduje się głównie w stanie nieaktywnym, tzn. średnia ilość białka na kopię  $E_0[y] < \frac{1}{2}$ .



Rysunek 5.5: Profile rozkładów białka  $f_k$  oraz  $\rho_N(y)$  dla  $c_0 = 2$ ,  $b_0 = 3$ . Pierwsza kolumna: układ z jedną kopią genu,  $N = 1$ ; druga kolumna: układ z dwiema kopiami genu,  $N = 2$ ; trzecia kolumna: układ z czterema kopiami genu,  $N = 4$ .

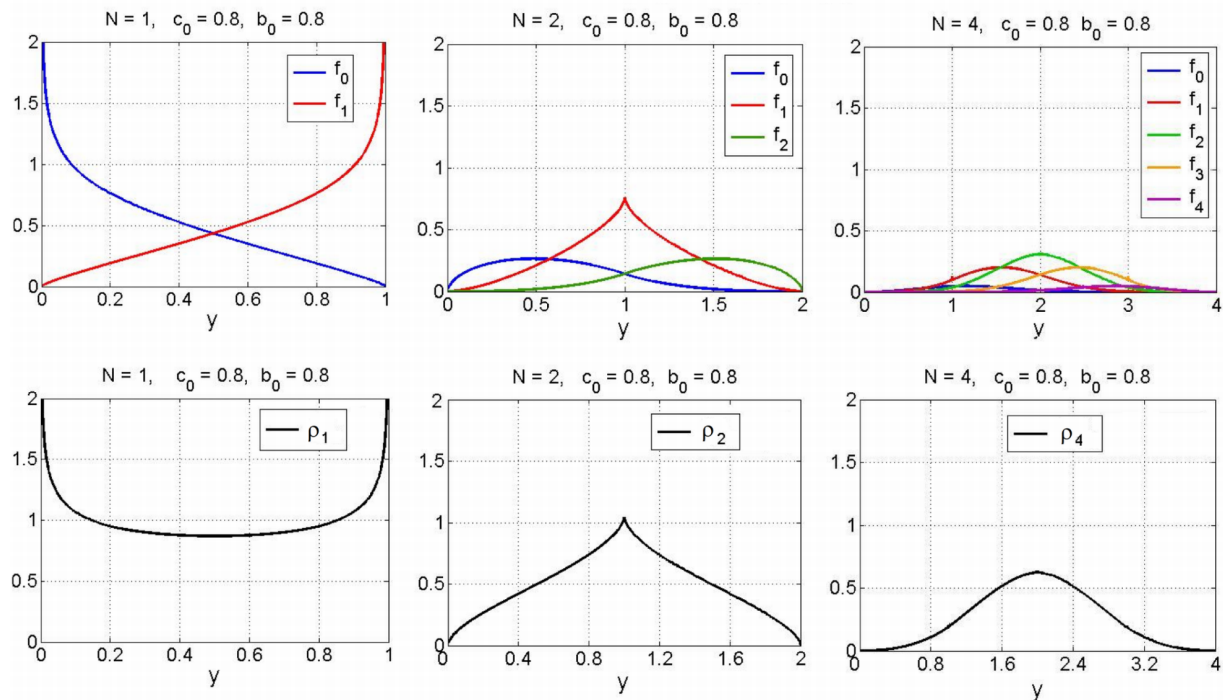
Jak widać na Rys. 5.6 (poniżej), ponieważ współczynnik aktywacji  $c_0$  jest znacznie większy niż współczynnik deaktywacji  $b_0$ , więc gen z  $N = 1$ ,  $N = 2$  i  $N = 4$  znajduje się głównie w stanie aktywnym (średnia ilość białka na kopię  $E_0[y] \geq \frac{1}{2}$ ), oraz profile rozkładów brzegowych osiągają maksima dla większej wartości  $y$  niż w poprzednim przypadku. Profile rozkładu brzegowego dla układu z jedną i dwiema kopiami bardzo różnią się od siebie: rozkład brzegowy dla układu z jedną kopią genu koncentruje się blisko  $y = 1$ .



Rysunek 5.6: Profile rozkładów białka  $f_k$  oraz  $\rho_N(y)$  dla  $c_0 = 4$ ,  $b_0 = 1$ . Pierwsza kolumna:  $N = 1$ ; druga kolumna:  $N = 2$ ; trzecia kolumna:  $N = 4$ .

Dla mniejszych wartości  $c_0$  i  $b_0$  (Rys. 5.7) można zaobserwować większą różnicę pomiędzy profilami rozkładów brzegowych w układach  $N = 1$ ,  $N = 2$  i  $N = 4$ , a mianowicie, rozkład dla  $N = 1$  ma dwa maksima, dla  $y = 0$  oraz  $y = 1$ , co sugeruje, że układ oscyluje pomiędzy dwoma stanami: z małą oraz dużą ilością białka.

Dla układu z jedną kopią genu  $\lim_{y \rightarrow 0} \rho^{(1)}(y) = \infty$  oraz  $\lim_{y \rightarrow 1} \rho^{(1)}(y) = \infty$ , podczas gdy dla układów z dwiema  $\rho_2(y)$  i czterema  $\rho_4(y)$  kopiami genu  $\lim_{y \rightarrow 0} \rho^{(1)}(y) = 0$  oraz  $\lim_{y \rightarrow 1} \rho^{(1)}(y) = 0$ .



Rysunek 5.7: Profile rozkładów poziomów białka  $f_k$  oraz  $\rho_N(y)$  dla  $c_0 = 0.8$ ,  $b_0 = 0.8$ .

Pierwsza kolumna:  $N = 1$ ; druga kolumna:  $N = 2$ ; trzecia kolumna:  $N = 4$ .

## 5.2 Charakterystyczne rozwiązania układu z autoregulacją

Rozważmy charakterystyczne rozwiązania układu (66) dla  $N = 1, 2, 4$  w przypadku z ujemną oraz dodatnią autoregulacją.

### Ujemne sprzężenie zwrotne

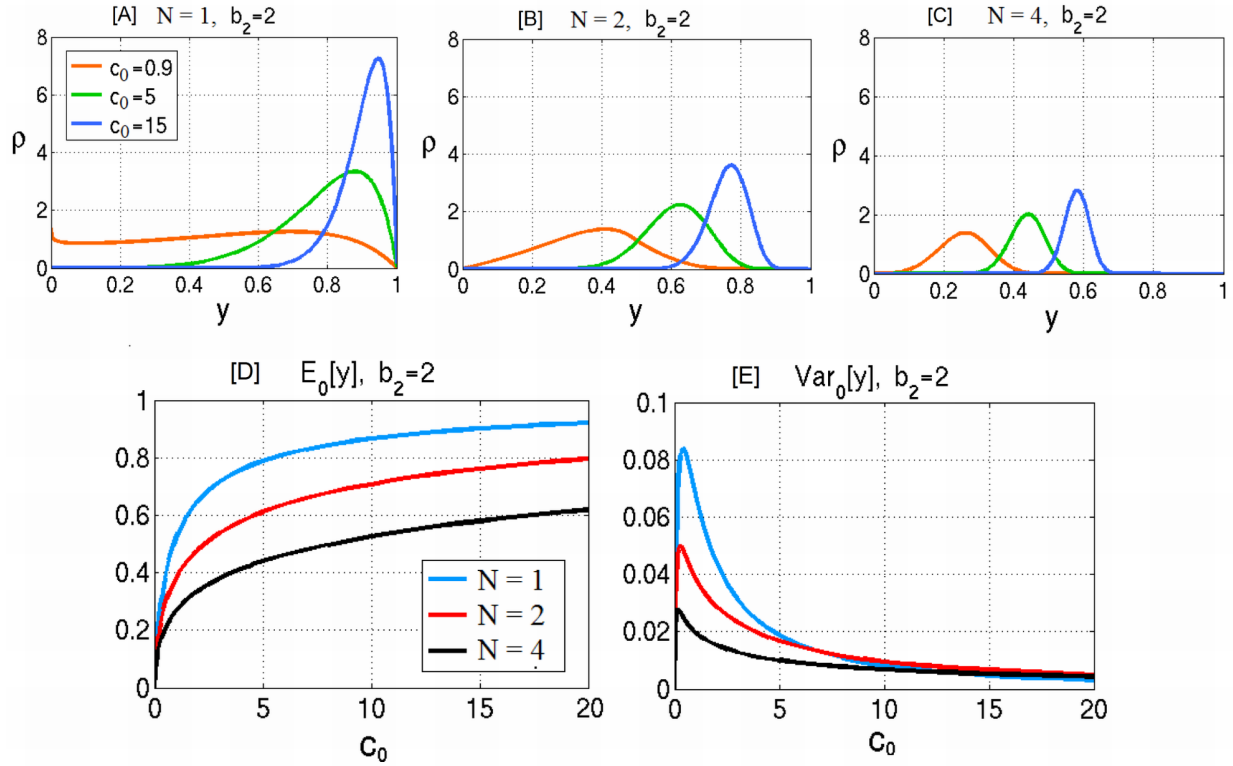
Rozważmy zewnętrzną indukcję samodezaktywującego się genu. Zakładamy, że współczynniki zmiany stanu genu  $c(y)$  i  $b(y)$  są postaci

$$c(y) = c_0, \quad b(y) = b_2 y^2 \quad (82)$$



Współczynnik  $c_0$  kontroluje tempo zewnętrznej lub spontanicznej aktywacji w wyniku przyłączenia czynników transkrypcyjnych niezależnych od analizowanego genu. Współczynnik  $b_2$  mierzy tempo dezaktywacji poprzez dysocjację białkowych homodimerów pochodzących od analizowanego genu, dlatego przyjmujemy, że współczynnik aktywacji genu jest proporcjonalny do  $y^2$  (kwadratu liczby monomerów). Jest to naturalne założenie, ponieważ białko powstaje jako monomer i bardzo często dwa homologiczne monomery łączą się ze sobą tworząc homodimery, które w wielu przypadkach są znacznie bardziej aktywne i stabilne niż monomery. Homodimery bardzo często pełnią funkcję białek regulatorowych. Produkcja homodimerów jest kwadratową funkcją koncentracji białka [10].

Na trzech Panelach (Rys. 5.8) mamy rozkład brzegowy białka dla trzech wartości spontanicznej aktywacji  $c_0$ , dla których, odpowiednio, układ z jedną, dwiema i czterema kopiami aktywuje się. Oś pozioma dla  $N = 2$  i  $N = 4$  została przeskalowana do odcinka  $[0, 1]$  dla lepszego porównania rozkładów. Profil rozkładu dla układu z dwiema kopiami genu jest węższy niż dla układu z jedną kopią, a szerszy od rozkładu dla układu z czterema kopiami, tzn. wraz ze wzrostem współczynnika spontanicznej aktywacji  $c_0$  oraz ze wzrostem liczby aktywnych kopii genu średnia oraz wariancja na kopię maleją.



Rysunek 5.8: Panele A, B, C: profile rozkładu brzegowego  $\rho_N(y)$ . Panel D: średnia  $E_0[y]$  białka na kopię genu. Panel E: wariancja  $Var_0[y]$  białka na kopię genu:  $b_2 = 2$ ,  $c_0 \in \{0.9, 5, 15\}$  dla  $N = 1, 2, 4$ .

Różnica pomiędzy rozkładami dla układów z jedną, dwiema i czterema kopiami jest znacznie bardziej widoczna w przypadku dodatniej autoregulacji.

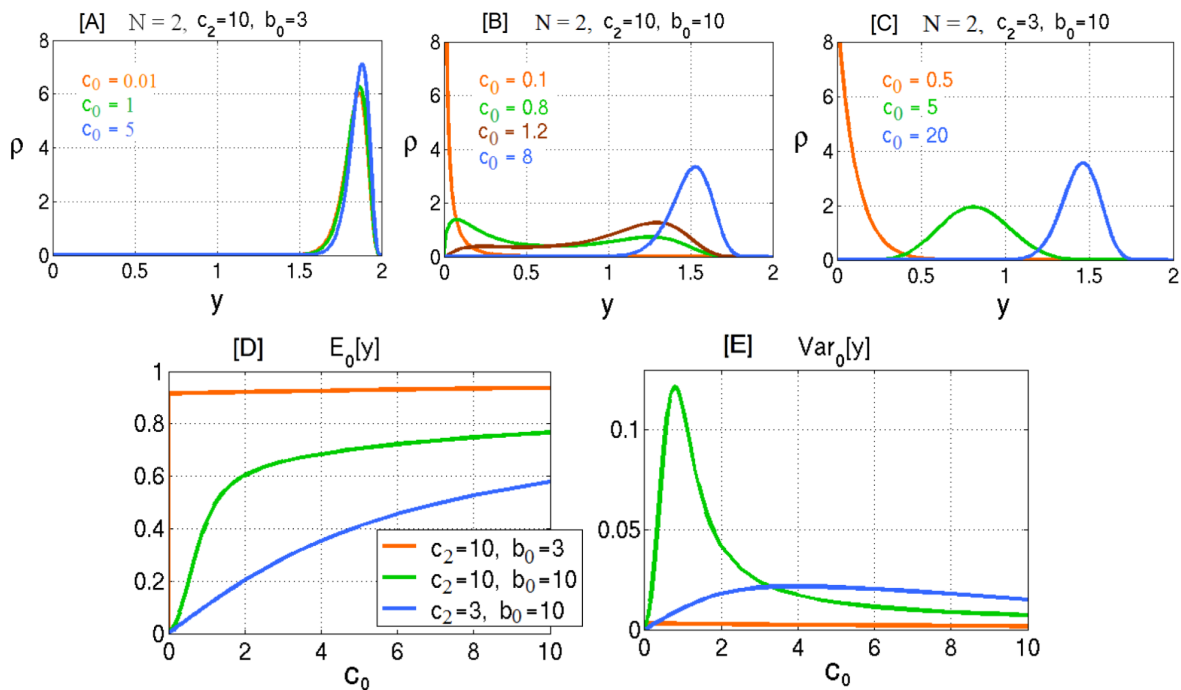
### Dodatnie sprzężenie zwrotne - wzorce aktywacji

Skoncentrujemy się teraz na regulacji samoaktywującego się genu. Zakładamy, że współczynniki zmiany stanu genu  $c(y)$  i  $b(y)$  są postaci

$$c(y) = c_0 + c_2 y^2, \quad b(y) = b_0.$$

Współczynnik  $c_2$  mierzy siłę samoaktywacji spowodowaną przyłączaniem się białkowych homodimerów. Współczynnik  $b_0$  oznacza dezaktywację w wyniku dysocjacji homodimerów z

promotora genu. Tempo aktywacji w wyniku przyłączenia pewnych czynników transkrypcyjnych jest kontrolowane przez współczynnik  $c_0$ . Rozważymy współczynniki  $c_2$  i  $b_0$  jako parametry charakteryzujące gen oraz dla danej pary  $c_2$  i  $b_0$  przeanalizujemy odpowiedź układu haploidalnego, diploidalnego i tetraploidalnego na zmieniającą się zewnętrzną stymulację  $c_0$ .



Rysunek 5.9: Trzy wzorce aktywacji genu diploidalnego (dodatnie sprzężenie zwrotne). Panele A, B, C: rozkłady brzegowe białka  $\rho(y)$  dla wzorców aktywacji {A}, {B} i {U}. Panele D, E: średnia na kopię  $E_0[y](c_0)$ , oraz E: wariancja na kopię  $Var_0[y](c_0)$ ; kolory: czerwony, zielony i niebieski odpowiadają wzorcom {A}, {B} i {U} aktywacji liczonym dla punktów (10,3), (10,10) i (3,10) w płaszczyźnie  $(c_2, b_0)$  (patrz Rys. 5.10 B).

$c_0 = 0$  prowadzi do singularności rozkładu  $\rho(y)$ . Numerycznie, bardzo małe  $c_0$  jest także wyzwaniem, ponieważ rozkład koncentruje się na  $y = 0$ . Aby uniknąć tych numerycznych trudności ograniczymy naszą numeryczną analizę do przypadku  $c_0 \geq 0.01$ . Z biologicznego punktu widzenia nie jest to znaczące ograniczenie gdyż każdy gen ulega spontanicznej aktywacji.

Zauważmy, że istnieją trzy wyróżniające się wzorce  $N$ -ploidalnej ( $N = 1, 2, 4$ ) aktywacji genu odpowiadające różnym parom  $(c_2, b_0)$ :

wzorzec  $\{A\}$  - jest trywialnym modelem, w którym gen pozostaje Aktywny (tzn.  $E_0[y] > 1/2$ ) dla  $c_0 \in [0.01, \infty)$ . Z biologicznego punktu widzenia oznacza to, że gen może być aktywowywany spontanicznie i pozostaje aktywny (Rys. 5.9 A),

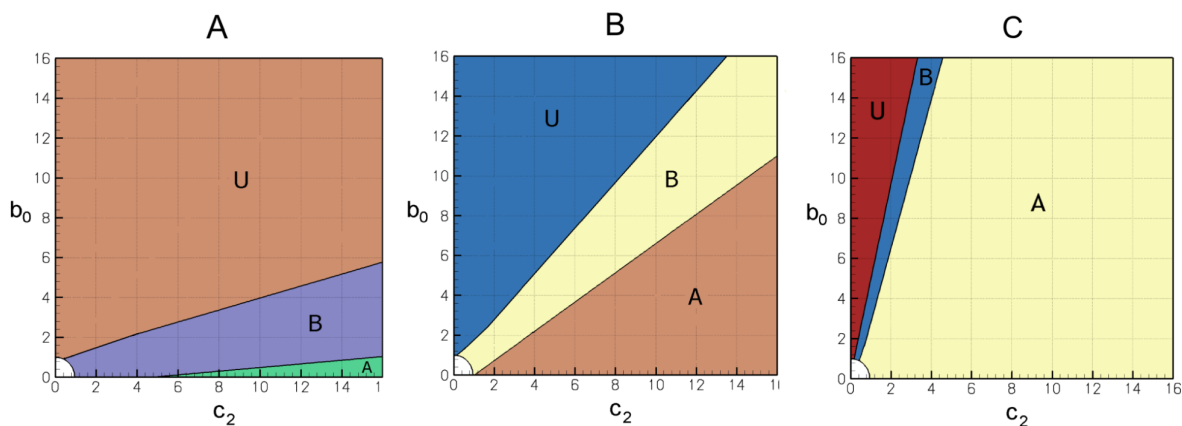
wzorzec  $\{B\}$  - w którym gen aktywuje się dla pewnego  $c_0 \in [0.01, \infty)$  oraz rozkład  $\rho(y)$  jest przejściowo (tzn. dla pewnego  $c_0$ ) Bimodalny (Rys. 5.9 B),

wzorzec  $\{U\}$  - w którym gen aktywuje się dla pewnego  $c_0 \in [0.01, \infty)$  lecz jego rozkład  $\rho(y)$  pozostaje Unimodalny dla wszystkich  $c_0 \in [0.01, \infty)$  (Rys. 5.9 C).

Gen znajdujący się we wzorcu  $\{A\}$  jest ciągle aktywny, średnia na kopię  $E_0[y](c_0)$  jest bliska 1 oraz układ jest niewrażliwy na siłę zewnętrznej stymulacji  $c_0$  (Rys. 5.9 D). Wariancja na kopię  $Var_0[y](c_0)$  pozostaje niska (Rys. 5.9 E). Wzorzec  $\{A\}$  jest "idealnym wyborem" dla genów, których produkt jest potrzebny w stałej ilości. We wzorcu  $\{B\}$  gen aktywuje się dla względnie niskiego  $c_0$  a profil aktywacji jest interesujący: dla pewnego  $c_0$  pojawia się drugie maksimum (dla większego  $y$ ) rozkładu  $\rho(y)$  (Rys. 5.9 B, zielona krzywa), które następnie staje się wyższe od pierwszego (Rys. 5.9 B, brązowa krzywa), ostatecznie dla większego  $c_0$  rozkład staje się unimodalny. Ten wzorzec aktywacji oznacza, że dla pewnego zakresu parametru aktywacji  $c_0$  populacja komórek dzieli się na dwie podpopulacje mające mniejszą i większą ilość białka. Ten wzorzec w pewnych sytuacjach mógłby być bezpieczniejszy dla komórki niż aktywacja unimodalna  $\{U\}$  pokazana na Rys. 5.9 C. Jeżeli sygnał nie jest wystarczająco silny tylko część komórek aktywuje się, podczas gdy reszta czeka na silniejszy sygnał do aktywacji.

Wzorzec  $\{B\}$  pojawia się dla większego stosunku  $\frac{c_2}{b_0}$  i w ten sposób układ aktywuje się (tzn.  $E_0[y] > 1/2$ ) dla mniejszej zewnętrznej stymulacji  $c_0$  z bardziej schodkowym profilem aktywacji (Rys. 5.9 D). Bimodalna aktywacja obecna jest także w profilu wariancji  $Var_0[y](c_0)$ , który ma wysokie ostre maksimum dla wartości  $c_0$ , dla której  $E_0[y](c_0) \approx 1/2$  (Rys. 5.9 E).

Na Rys. 5.10 B pokazujemy trzy sektory w płaszczyźnie  $(c_2, b_0)$  odpowiadające trzem wzorcom aktywacji  $\{A\}, \{B\}, \{U\}$  dla genu diploidalnego. Sektor dla haploidalnego genu jest strukturalnie podobny do sektora dla genu diploidalnego (Rys. 5.10 A), sektor  $\{A\}$  jest znacznie węższy, zaś sektor  $\{U\}$  jest znacznie szerszy. W przypadku genu tetraploidalnego sektor  $\{A\}$  jest szerszy, podczas gdy  $\{B\}$  i  $\{U\}$  są względnie wąskie w porównaniu z sektorami  $\{B\}$  i  $\{U\}$  dla genu diploidalnego (Rys. 5.10 C).

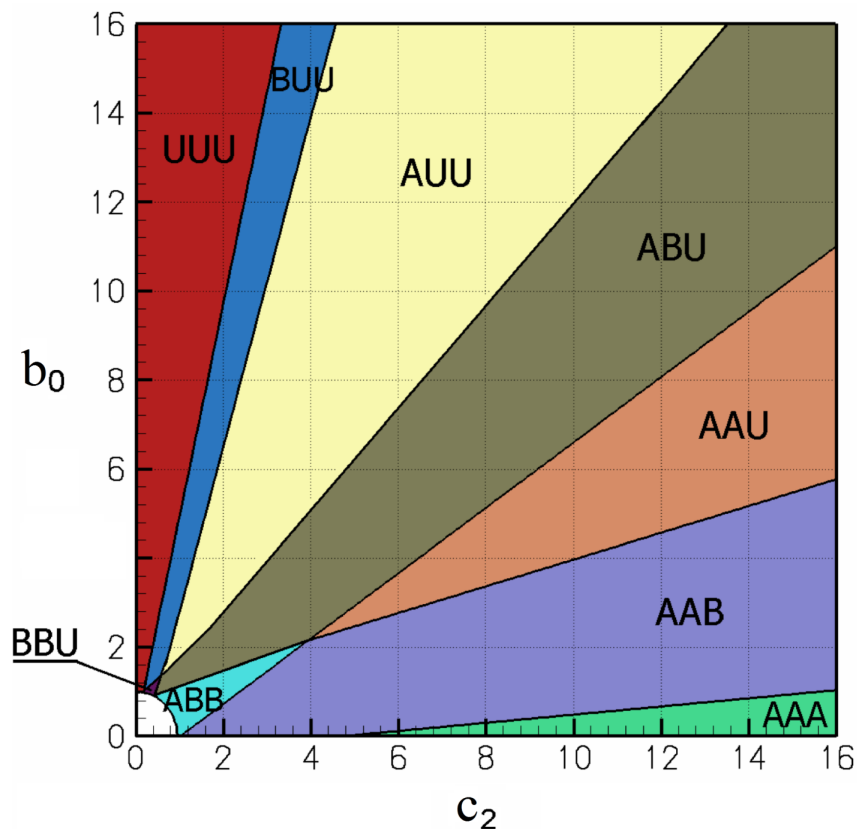


Rysunek 5.10: Trzy obszary w płaszczyźnie  $(c_2, b_0)$  odpowiadające trzem wzorcom aktywacji  $\{A\}, \{B\}, \{U\}$ . Panel A: gen haploidalny, Panel B: gen diploidalny, Panel C: gen tetraploidalny.

Na Rys. 5.10 A, 5.10 B i 5.10 C możemy wyróżnić mały obszar dla  $c_2^2 + b_0^2 \leq 1$ . Dla bardzo małego  $c_2, b_0$  i  $c_0$  rozkład  $\rho^{(4)}(y)$ , odpowiadający genowi tetraploidalnemu może mieć do pięciu maksimumów dla  $y = i, 0 \leq i \leq 4$ , a każde z nich odpowiada stanowi, w którym  $i$  z 4 kopii jest aktywne. Multimodalność, spowodowana przez bardzo rzadkie przeskakiwanie pomiędzy stanami genu, ma inny charakter niż bimodalność, wyłączamy więc obszar  $c_2^2 + b_0^2 \leq 1$  z dalszych rozważań.

Interesujące jest, że wzorec aktywacji diploidalnego genu (Rys. 5.10 B) jest najbardziej regularny, tzn. sektory  $\{A\}, \{B\}$  i  $\{U\}$  są podobnej wielkości. Geny haploidalne mogą znajdować się w obszarze  $\{A\}$  tylko dla bardzo małego stosunku  $\frac{c_2}{b_0}$ , podczas gdy tetraploidalne

geny znajdują się w większości w obszarze  $\{A\}$ , co powoduje, że są niewrażliwe na siłę zewnętrznego sygnału  $c_0$ .



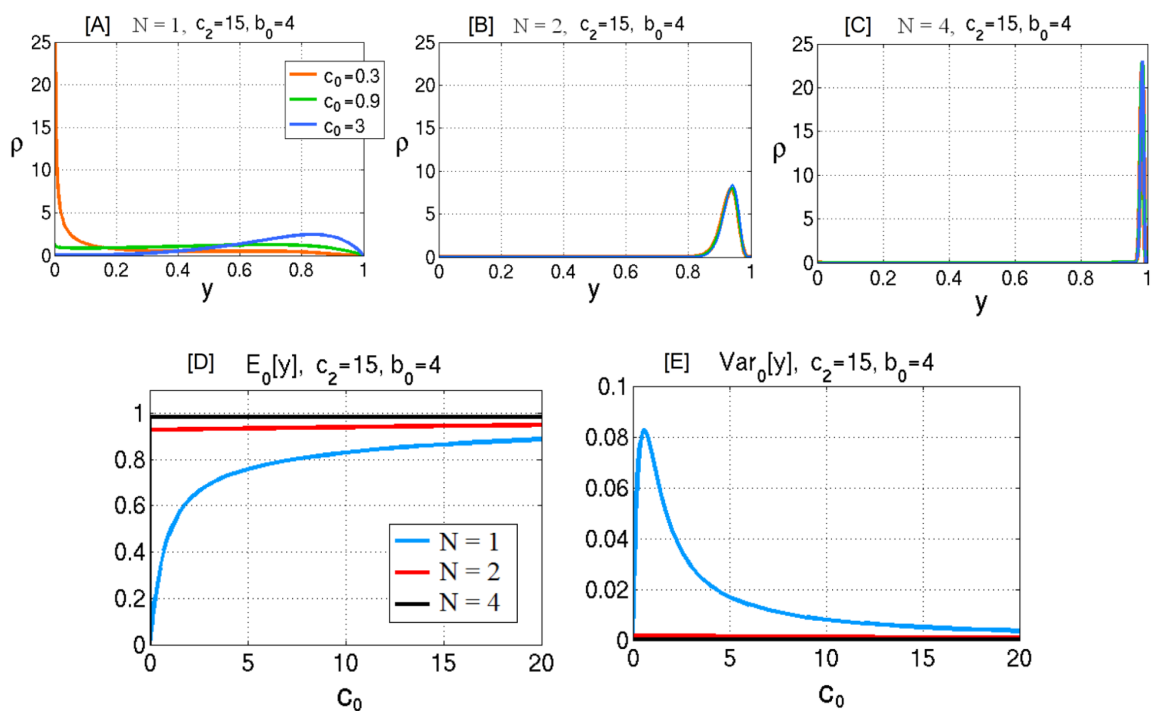
Rysunek 5.11: Dziewięć wzorców równoczesnej aktywacji genu haplo-, diplo- i tetraploidalnego. Profile aktywacji dla trzech wybranych punktów w płaszczyźnie  $(c_2, b_0)$  analizowane są na Rys. 5.12, 5.13 i 5.14.

Gdy połączymy te trzy obrazki (z Rys. 5.10) w jeden (Rys. 5.11), otrzymamy dziewięć możliwych wzorców równoczesnej aktywacji haploidalnego, diploidalnego i tetraploidalnego genu. To pozwala na analizę dla genów diploidalnych konsekwencji braku jednej kopii genu bądź duplikacji genu.

Zgodnie z naszym oznaczeniem sektor  $\{ABU\}$  wyznacza zakres parametrów  $(c_2, b_0)$ , dla których tetraploidalny gen aktywuje się we wzorcu  $\{A\}$ , diploidalny w  $\{B\}$  i haploidalny w

$\{U\}$ . Stąd, dla dowolnej pary  $(c_2, b_0)$  z sektora  $\{ABU\}$  strata kopii genu implikuje, że układ diploidalny aktywowałby się we wzorcu  $\{U\}$  zamiast w  $\{B\}$ , podczas gdy duplikacja genu przerzuci go do wzorca  $\{A\}$ .

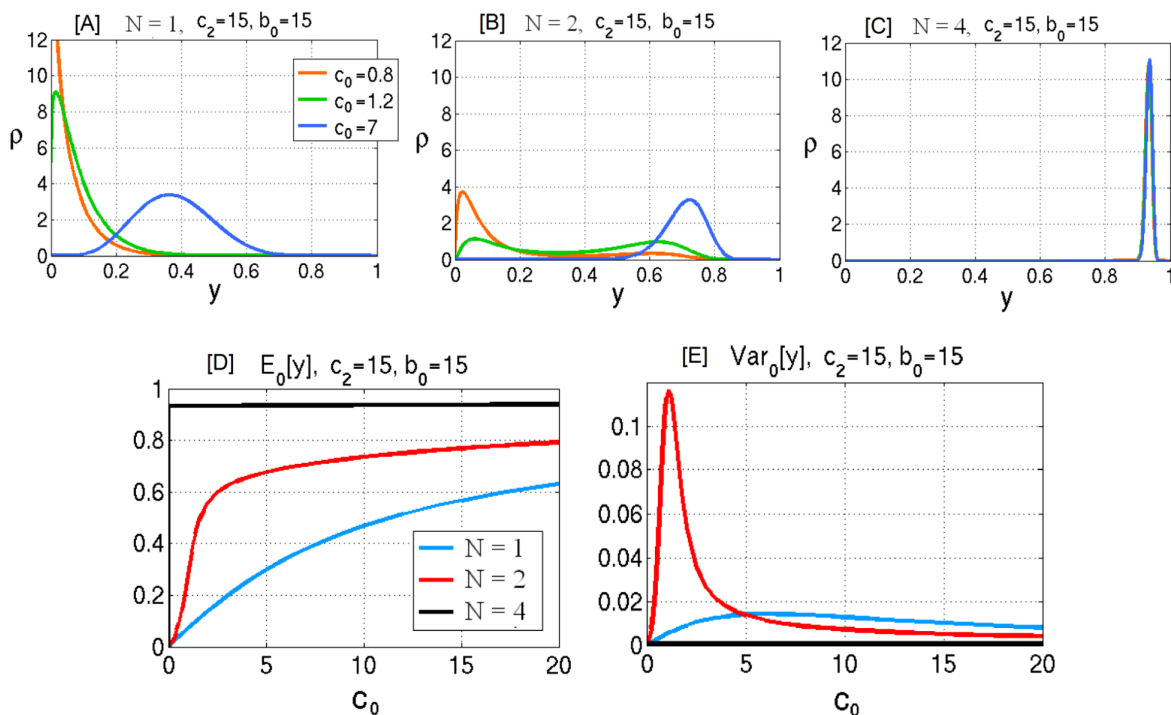
Gdy układ diploidalny "działa" w obszarze  $\{AAB\}$  lub  $\{AAU\}$ , wtedy duplikacja genu nie zmieni wzorca aktywacji, ale strata kopii genu zatrzyma ciągłą aktywność genu. Jeśli układ diploidalny znajduje się w sektorze  $\{AUU\}$ , wówczas strata kopii genu nie spowoduje zmiany wzorca aktywacji, lecz duplikacja genu przerzuci go do wzorca  $\{A\}$ .



Rysunek 5.12: Wzorec aktywacji  $\{AAB\}$ :  $c_2 = 15, b_0 = 4$ . Panele A, B i C: rozkłady brzegowe białka  $\rho(y)$  dla, odpowiednio, układu haplo-, diplo- i tetraploidalnego dla  $c_0 = 0.3$ ,  $c_0 = 0.6$  (w którym układ haploidalny ma  $\max[Var_0[y](c_0)]$ ), oraz  $c_0 = 3$ . Panel D: średnia na kopię  $E_0[y](c_0)$ . Panel E: wariancja na kopię  $Var_0[y](c_0)$ .

Na Rys. 5.12, 5.13 i 5.14 analizujemy trzy z dziewięciu możliwych przypadków:  $\{AAB\}$ ,  $\{ABU\}$ ,  $\{BUU\}$ . Przypadek  $\{AAB\}$  (Rys. 5.12) przypomina klasyczny obraz haploinsuf-

ficiency, patrz Cook i in. [5]. Diploidalny i tetraploidalny gen pozostaje aktywny, podczas gdy haploidalny gen aktywuje się poprzez rozkład bimodalny tak, że profil jego wariacji  $Var_0[y](c_0)$  ma ostre wysokie maksimum (Rys. 5.12 B, 5.12 E). Strata kopii genu zmniejsza średnią, ale co ważniejsze, zwiększa zmienność ekspresji genu. W tym przypadku, jeżeli produkt genu jest potrzebny na stałym poziomie strata kopii genu może prowadzić do choroby.

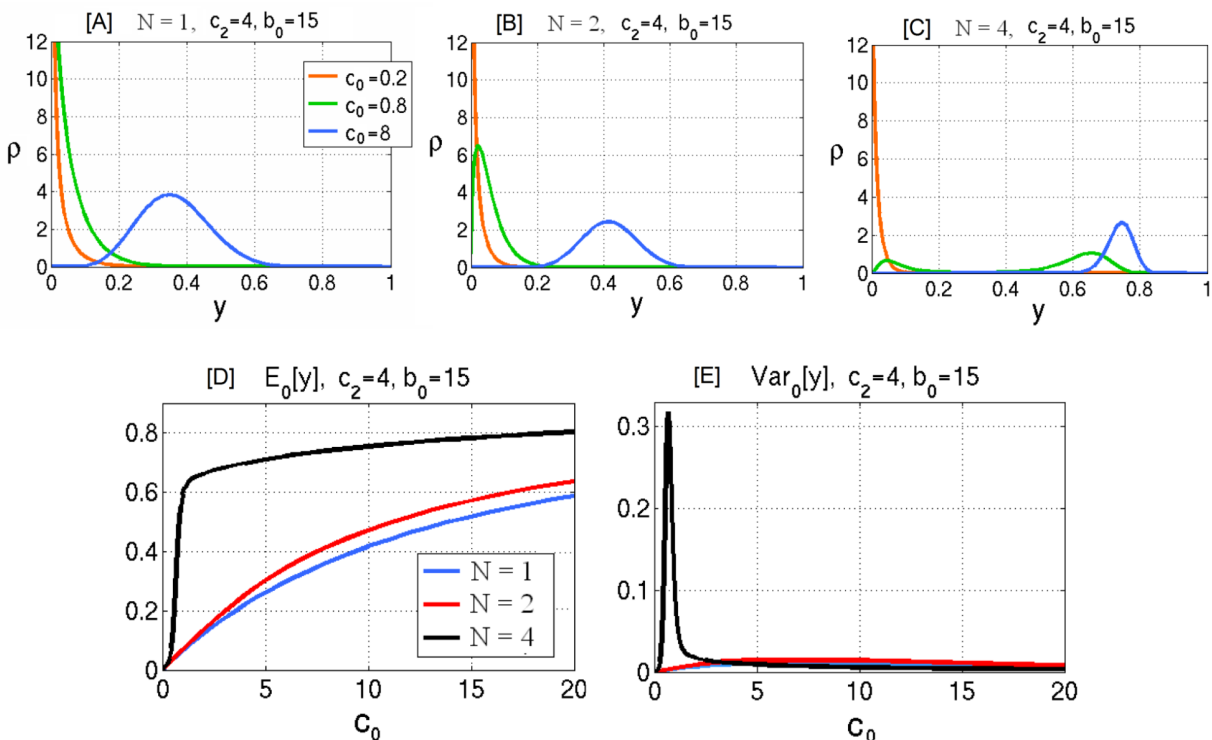


Rysunek 5.13: Wzorzec aktywacji {ABU}:  $c_2 = 15$ ,  $b_0 = 15$ . Panele A, B i C: rozkłady brzegowe białka  $\rho(y)$  dla, odpowiednio, układu haplo-, diplo- i tetraploidalnego dla  $c_0 = 0.8$ ,  $c_0 = 1.2$  (w którym układ diploidalny ma  $\max[Var_0[y](c_0)]$ ), oraz  $c_0 = 7$ . Panel D: średnia na kopię  $E_0[y](c_0)$ . Panel E: wariancja na kopię  $Var_0[y](c_0)$ .

Przypadek {ABU} (Rys. 5.13) jest zupełnie inny; układ diploidalny wykazuje przejściowo najszerszy rozkład białka (Rys. 5.13 B). Wariancja na kopię  $Var_0[y](c_0)$  jest w tym przypadku większa dla genu diploidalnego niż dla haploidalnego (Rys. 5.13 E). Aktywacja hap-



loidального genu jest znacznie bardziej stopniowa ze względnie małą wariancją. Tetraploidalny gen jest stale aktywny z wąskim rozkładem białka.

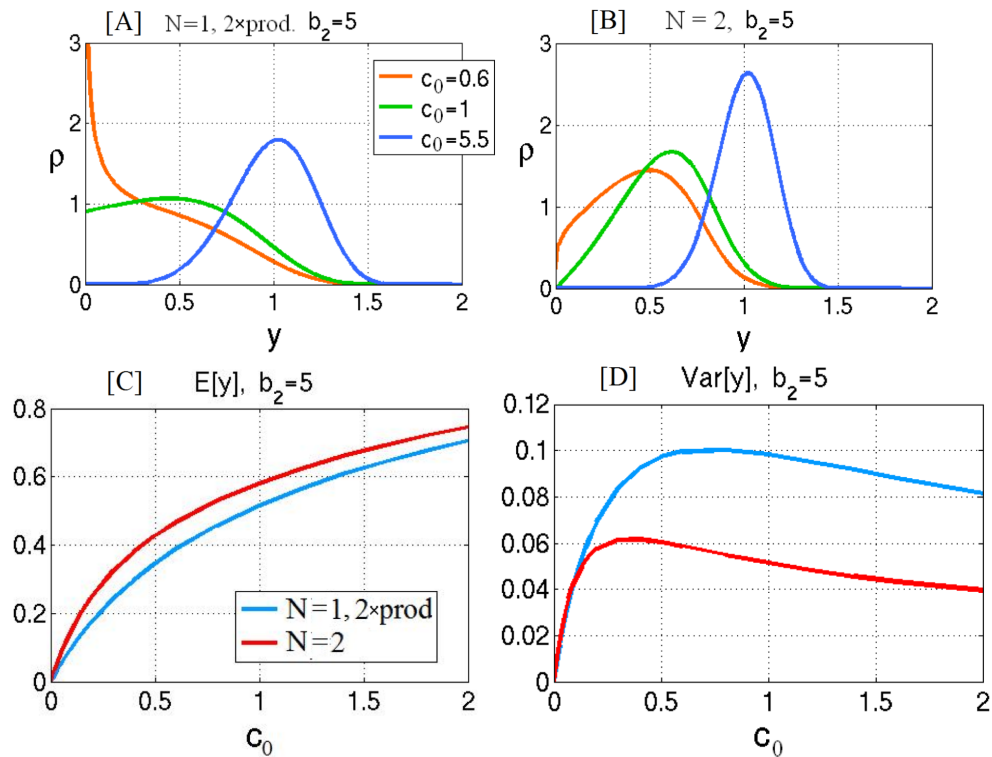


Rysunek 5.14: Wzorec aktywacji {BUU}:  $c_2 = 4$ ,  $b_0 = 15$ . Panele A, B i C: rozkłady brzegowe białka  $\rho(y)$  dla, odpowiednio, układu haplo-, diplo- i tetraploidalnego dla  $c_0 = 0.2$ ,  $c_0 = 0.8$  (w którym układ tetraploidalny ma  $\max[Var_0[y](c_0)]$ ), oraz  $c_0 = 8$ . Panel D: średnia na kopię  $E_0(y)(c_0)$ . Panel E: wariancja na kopię  $Var_0[y](c_0)$ .

W przypadku {BUU} (Rys. 5.14) geny haplo-, diplo- i tetraploidalne aktywują się wraz ze wzrostem stymulacji. Profil aktywacji (Rys. 5.14 D) jest najbardziej stromy dla genu tetraploidalnego oraz najbardziej łagodny dla genu haploidalnego. Wbrew intuicji, wariancja na kopię  $Var_0[y](c_0)$  jest największa dla tetraploidalnego genu ze względu na jego przejściowo bimodalny rozkład (Rys. 5.14 E).

### 5.3 Brak kopii genu nie może być zrekompensowany przez podwojoną ekspresję pozostałej kopii genu

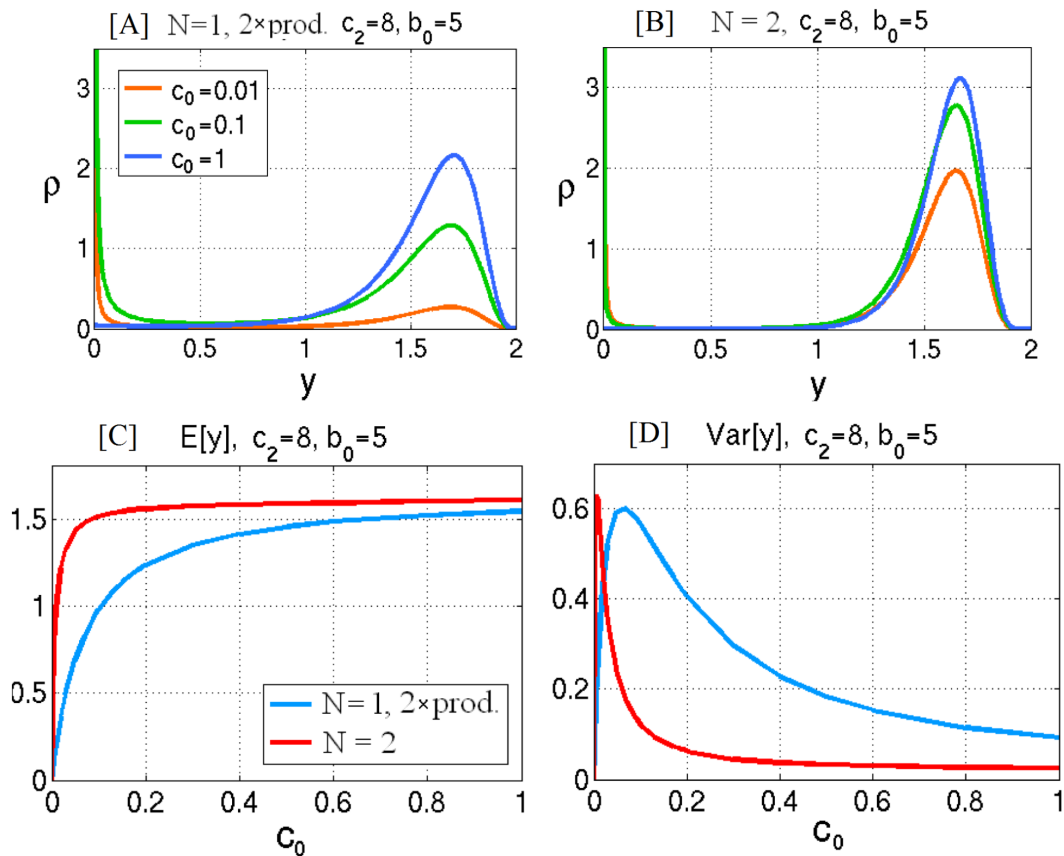
Chcielibyśmy sprawdzić, czy w przypadku braku jednej z dwóch kopii genu dwukrotne zwiększenie współczynnika syntezy mRNA dla pozostałej kopii może zrekompensować stratę jednej kopii genu. Okazuje się, że w przypadku dodatniego sprzężenia zwrotnego strata kopii może nie być w pełni zrekompensowana przez wyższą ekspresję pozostałej kopii genu.



Rysunek 5.15: Regulacja genu haploidalnego z podwojoną produkcją oraz genu diploidalnego (ujemne sprzężenie zwrotne):  $b_2 = 5$ . Panele A i B: rozkład brzegowy białka  $\rho(y)$  dla układu haplo- i diploidalnego dla  $c_0 = 0.6$ ,  $c_0 = 1$  oraz  $c_0 = 5.5$ . Panele D i E: średnia  $E[y](c_0)$  oraz wariancja  $\text{Var}[y](c_0)$  białka.

W przypadku bez autoregulacji, bazując na analizie z Podrozdziału 3.4, można wywnioskować, że haploidalny gen ma taką samą średnią  $E[y]$  lecz dwa razy większą wariancję  $\text{Var}[y]$  niż

gen diploidalny (Rys. 5.15 C, 5.15 D). W przypadku ujemnego sprzężenia zwrotnego średnia  $E[y](c_0)$  dla haploidalnego genu jest nieznacznie wyższa od średniej dla genu diploidalnego, zaś wariancja  $Var[y](c_0)$  jest około dwa razy większa, profile rozkładu białka  $\rho(y)$  dla układu haploidalnego i diploidalnego znacznie się różnią.



Rysunek 5.16: Regulacja genu haploidalnego z podwojoną produkcją oraz genu diploidalnego (dodatnie sprzężenie zwrotne):  $c_2 = 8$ ,  $b_0 = 5$ . Panele A i B: rozkład brzegowy białka  $\rho(y)$  dla układu haplo- i diploidalnego dla  $c_0 = 0.01$  (w którym gen diploidalny aktywuje się),  $c_0 = 0.1$  oraz  $c_0 = 1$ . Panele C i D: średnia  $E[y](c_0)$  oraz wariancja  $Var[y](c_0)$  białka.

Różnica pomiędzy haploidalnym i diploidalnym układem jest znacznie bardziej wyraźna w przypadku dodatniej autoregulacji (Rys. 5.16). Jak pokazano na Rys. 5.16 C, średnia  $E[y](c_0)$  dla genu diploidalnego (który aktywuje się dla współczynnika  $c_0 = 0.01$ ) jest

większa niż dla genu haploidalnego. Spowodowane jest to tym, że w przypadku dodatniej autoregulacji aktywacja jednej kopii genu powoduje produkcję białek, które aktywują drugą kopię genu. W przypadku pokazanym na Rys. 5.16 A, 5.16 B (czerwona krzywa) rozkłady białek odpowiadające haploidalnym i diploidalnym genom są bimodalne gdy oba te geny aktywują się. Jednakże, ponieważ gen diploidalny aktywuje się dla 10 razy mniejszej wartości  $c_0$  niż haploidalny, wariancja rozkładu białka jest przejściowo większa dla układu diploidalnego (Rys. 5.16 D). Dla większej wartości  $c_0$  haploidalny układ ma, jak można się było spodziewać, znacznie większą wariancję niż układ diploidalny.

# Rozdział 6

## Wpływ liczby kopii genów na dynamikę ścieżki sygnałowej tymocyту

### 6.1 Tymocyty

Limfocyty  $T$  (tymocyty, komórki  $T$ ) sterują nabytą odpowiedzią immunologiczną. Wyróżnia się dwa podstawowe rodzaje tymocytów: *pomocnicze* ( $Th$ ) oraz *cytotoksyczne* ( $Tc$ ). Limfocyty pomocnicze pobudzają i podtrzymują odpowiedź immunologiczną, zaś limfocyty cytotoksyczne mają zdolność do zabijania komórek oznaczonych przeciwciałami skierowanymi przeciwko ich antygenom powierzchniowym.

Każda komórka organizmu zawiera własne antygeny, które pozostają pod kontrolą układu immunologicznego i w normalnych warunkach rozpoznawane jako "swoje" nie powodują reakcji immunologicznej, oraz antygeny, które po ich "odczytaniu" traktowane są jako "obce" i powodują reakcję immunologiczną. Aby tymocyt mógł rozpoznać antygen musi on zostać przetworzony przez tzw. *komórki prezentujące antygen* ( $APC$ ) należące do układu komórek makrofagów i układu komórek dendrytycznych. Za tzw. prezentację antygenów limfocytom  $T$  odpowiedzialny jest kompleks peptydów antygenowych zwany *głównym układem zgodności tkankowej MHC*, który występuje na powierzchni komórek organizmu, a zwłaszcza na powierzchni makrofagów. Komórki  $APC$  mają zdolność rozkładania złożonego antygenu białkowego na fragmenty peptydowe, które po związaniu z cząsteczkami  $MHC$  są następnie transportowane na powierzchnię komórki. W ten sposób związany z błoną komórkową kompleks jest przygotowany do interakcji z obecną na tymocycie cząsteczką rozpoznającą nazywaną *receptorem antygenowym* ( $TCR$ ). Utworzony kompleks, rozpoznawalny przez ty-

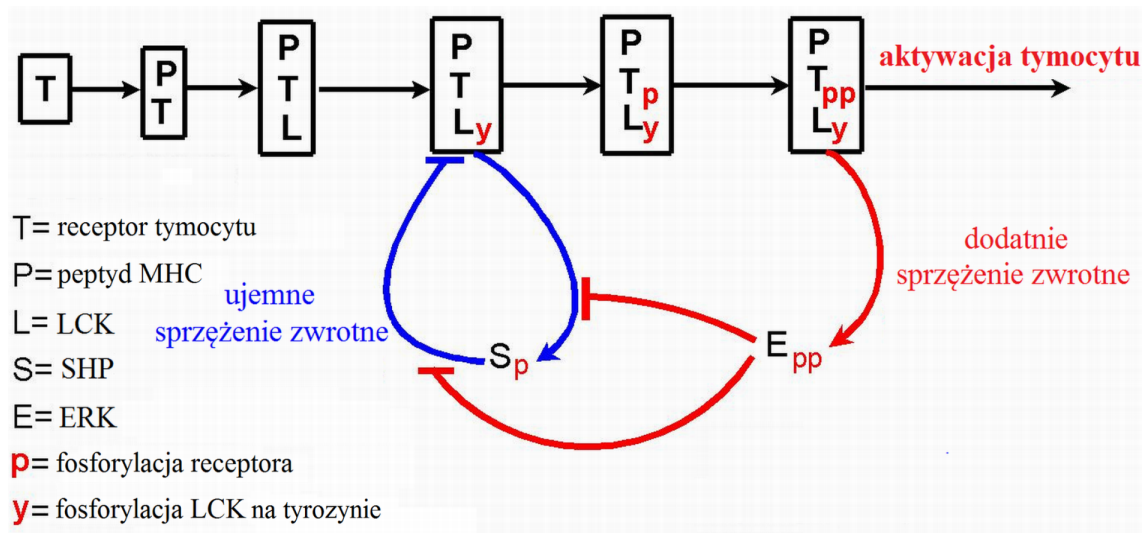
mocyt, umożliwia prezentację i rozpoznanie antygeny.

Istnieją dwa typy antygenów *MHC*: klasy I i klasy II. To czy poszczególne komórka będzie reagowała z antygenem prezentowanym przez cząsteczkę *MHC* klasy I czy *MHC* II uzależnione jest od dodatkowych cząsteczek występujących na powierzchni komórki, określanych jako markery układu zróżnicowania (*CD*). Markery *CD4*, występujące na limfocytach *Th*, są odpowiedzialne za wiązanie z cząsteczkami *MHC* II obecnymi głównie na komórkach *APC*, markery *CD8* są natomiast obecne na limfocytach *Tc* i odpowiadają za wiązanie z cząsteczkami *MHC* I. Kompleksy peptydów antygenowych połączonych z *MHC* II pobudzają i podtrzymują odpowiedź immunologiczną, zaś kompleksy peptydów antygenowych połączonych z *MHC* I biorą udział w obronie przeciwko patogenom wewnątrzkomórkowym, np. wirusom; jeżeli tego rodzaju antygen zostanie rozpoznany jako obcy, komórka prezentująca ulegnie eliminacji.

## 6.2 Model ścieżki sygnałowej *TCR*

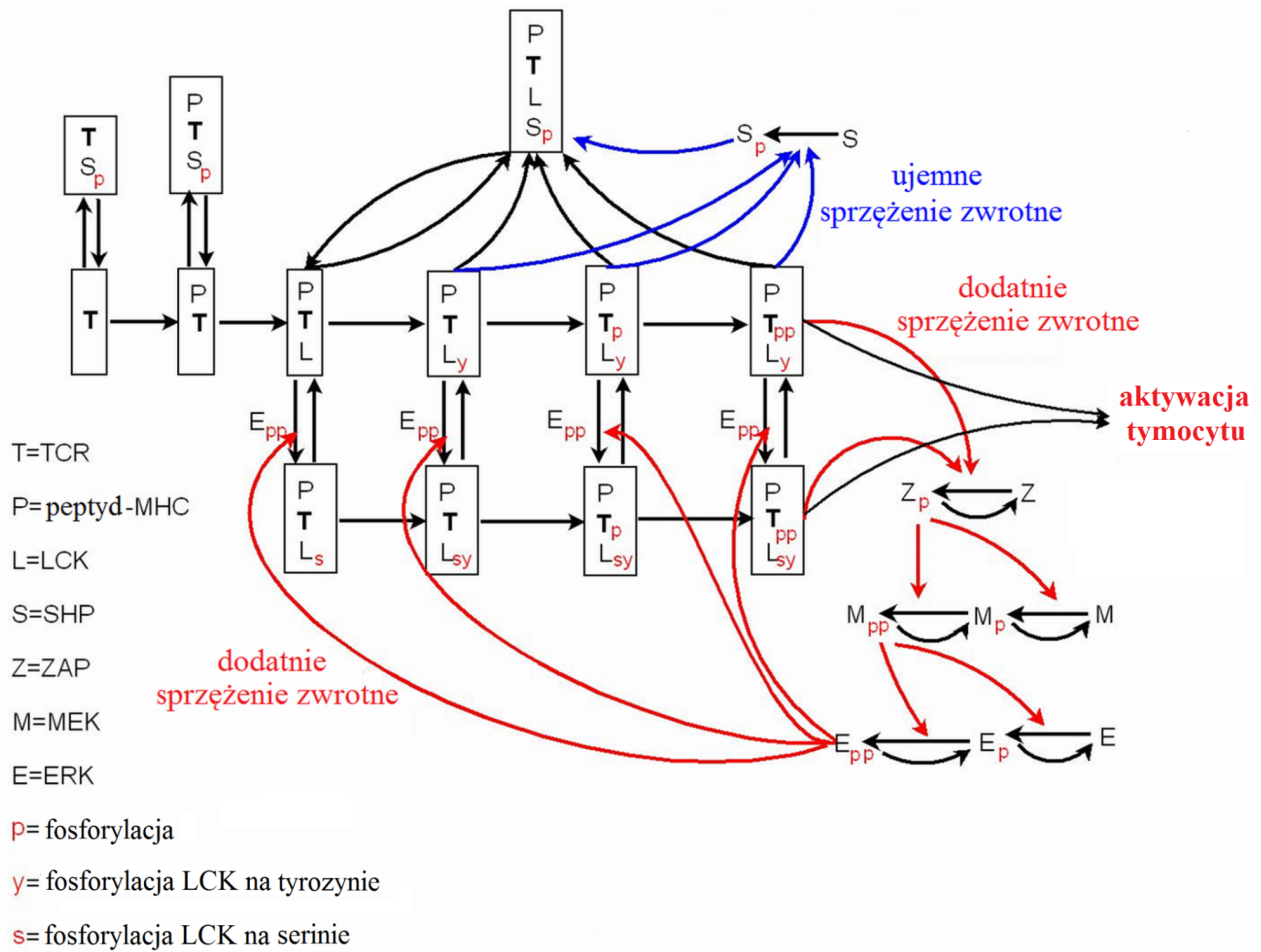
Receptor antygenowy *TCR* jest podstawowym elementem w komunikacji z komórkami układu immunologicznego, a jego stymulacja stanowi pierwszy krok we wzbudzaniu aktywności tymocytu.

Aktywacja receptora *TCR* rozpoczyna się od przyłączenia do *TCR* peptydów antygenowych *MHC* (Rys. 6.1). Następnie, do cytoplazmatycznej strony receptora przyłącza się kinaza *LCK*, która zostaje sfosforylowana. *LCK* ma dwa główne miejsca fosforylacji, na tyrozynie oraz serinie. Po dołączeniu do *TCR*, *LCK* ulega fosforylacji na tyrozynie. Sfosforylowane na tyrozynie *LCK* oznaczamy przez  $LCK_y$ . Zainicjowana zostaje pętla ujemnego sprzężenia zwrotnego:  $LCK_y$  fosforyluje i aktywuje fosfatazę *SHP* (sfosforylowane *SHP* oznaczamy przez  $SHP_p$ ),  $SHP_p$  przyłącza się do *TCR* i defosforyluje  $LCK_y$ .



Rysunek 6.1: Uproszczony schemat ścieżki sygnałowej *TCR*. Regulacja ujemnego sprzężenia zwrotnego (niebieskie linie) zachodzi poprzez działanie fosfatazy *SHP*, zaś regulacja dodatniego sprzężenia zwrotnego (czerwone linie) poprzez działanie kinazy *ERK*.

*LCK<sub>y</sub>* sekwencyjnie fosforyluje *TCR*. *TCR<sub>pp</sub>* (podwójnie sfosforylowane *TCR*) umożliwia autofosforylację i aktywację kinazy *ZAP* (Rys. 6.2). Procesy te zapoczątkowują kaskadę sygnałową kinaz *MEK* i *ERK*, w wyniku której *ERK* zostaje podwójnie sfosforylowany. Poprzez fosforylację *LCK* na serynie (oznaczane przez *LCK<sub>s</sub>*), *ERK<sub>pp</sub>* inicjuje pętlę dodatniego sprzężenia zwrotnego i tym samym inhibuje pętlę ujemnego sprzężenia zwrotnego, ponieważ fosforylacja *LCK<sub>y</sub>* na serynie uniemożliwia zarówno kinazie *LCK* sfosforylowanie *SHP*, oraz *SHP* sfosforylowanie *LCK*.

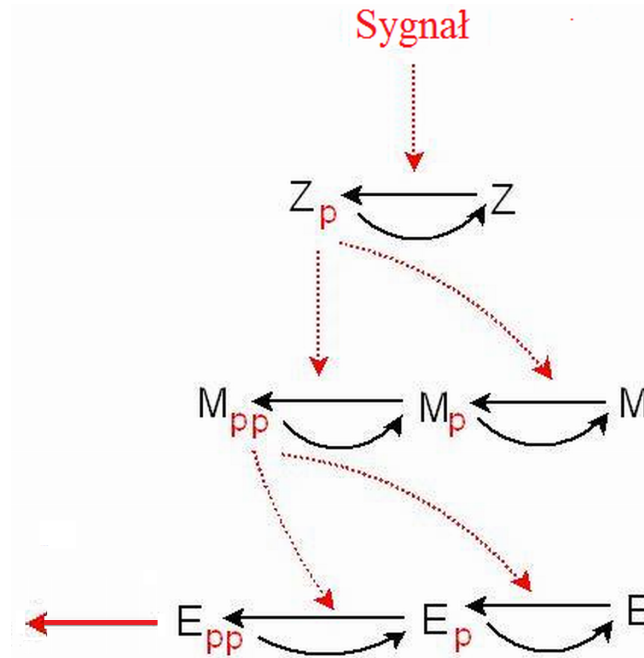


Rysunek 6.2: Pełny diagram ścieżki sygnałowej *TCR*.

Stochastyczny model Lipniacki i in. ścieżki sygnałowej *TCR* analizowany poniżej [74] zawiera pętlę ujemnego i dodatniego sprzężenia zwrotnego i reprezentowany jest przez 34 równania różniczkowe zwyczajne opisujące ewolucję poziomów białek danego typu przypadających na pojedynczy tymocyt. W Rozprawie rozważamy wersję deterministyczną modelu.

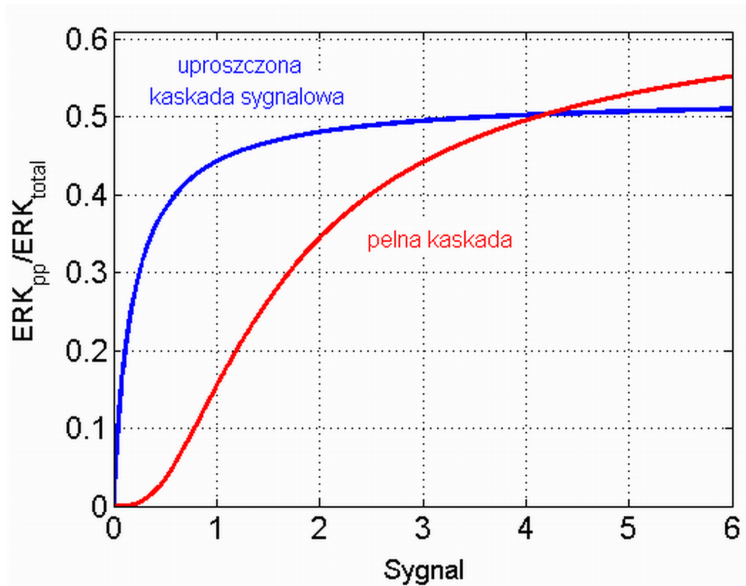
Za miarę poziomu aktywacji komórki przyjmujemy poziom  $ERK_{pp}$ : jeśli w danym tymocycie  $\frac{ERK_{pp \max}}{ERK_{total}} > \frac{1}{2}$  przyjmujemy, że komórka jest przejściowo aktywna, jeśli zaś w danym tymocycie  $\frac{ERK_{pp(\infty)}}{ERK_{total}} > \frac{1}{2}$  przyjmujemy, że komórka jest stale aktywna. Komórki, dla których  $\frac{ERK_{pp \max}}{ERK_{total}} < \frac{1}{10}$  przyjmujemy za nieaktywne.





Rysunek 6.3: Kaskada sygnałowa kinaz *ZAP* (*Z*), *MEK* (*M*) oraz *ERK* (*E*). Czerwone kropkowane linie reprezentują fosforylację.

Analizowana w naszym modelu kaskada sygnałowa zawiera pojedynczo sfosforylowane kinazy *ZAP*, *MEK* i *ERK*, oraz podwójnie sfosforylowane kinazy *MEK* i *ERK* (Rys. 6.3). Rozważmy również uproszczoną postać kaskady sygnałowej, w której pomija się pojedynczo sfosforylowane formy *MEK* i *ERK*. Porównaliśmy kinetykę obu typów kaskad sygnałowych, a mianowicie dla pełnej kaskady wyznaczyliśmy wartość *Sygnału*, dla którego  $\frac{ERK_{pp}(\infty)}{ERK_{total}} = \frac{1}{2}$ , a następnie w kaskadzie uproszczonej dobraliśmy współczynniki fosforylacji oraz defosforylacji poszczególnych kinaz tak, aby  $\frac{ERK_{pp}(\infty)}{ERK_{total}} = \frac{1}{2}$  dla tej samej wartości *Sygnału* co w pełnej kaskadzie.



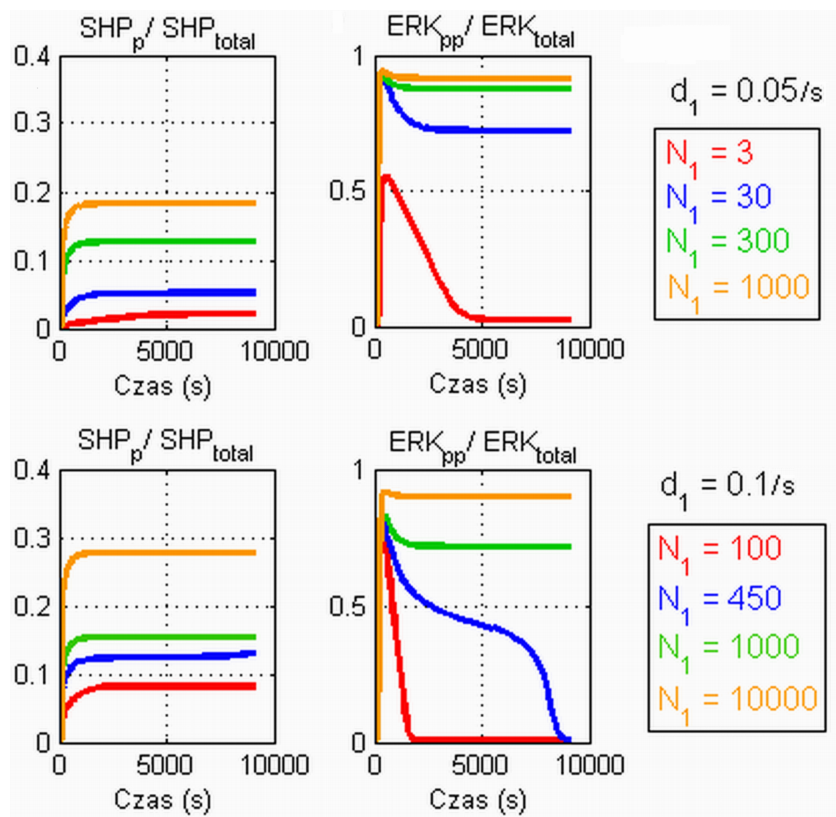
Rysunek 6.4: Profil  $ERK_{pp}(t)/ERK_{total}$  jako funkcji *Sygnalu* dla pełnej kaskady (czerwona krzywa) oraz kaskady uproszczonej (niebieska krzywa).

Wówczas profil  $ERK_{pp}(t)/ERK_{total}$  jako funkcja *Sygnalu*, dla pełnej kaskady jest zupełnie inny niż dla kaskady uproszczonej (Rys. 6.4), tzn. dla małej wartości *Sygnalu*, dla pełnej kaskady poziom  $MEK_{pp}$  zależy od drugiej potęgi *Sygnalu* oraz poziom  $ERK_{pp}$  zależy od czwartej potęgi *Sygnalu*, podczas gdy dla uproszczonej kaskady poziom  $MEK_{pp}$  oraz poziom  $ERK_{pp}$  zależą liniowo od *Sygnalu*.

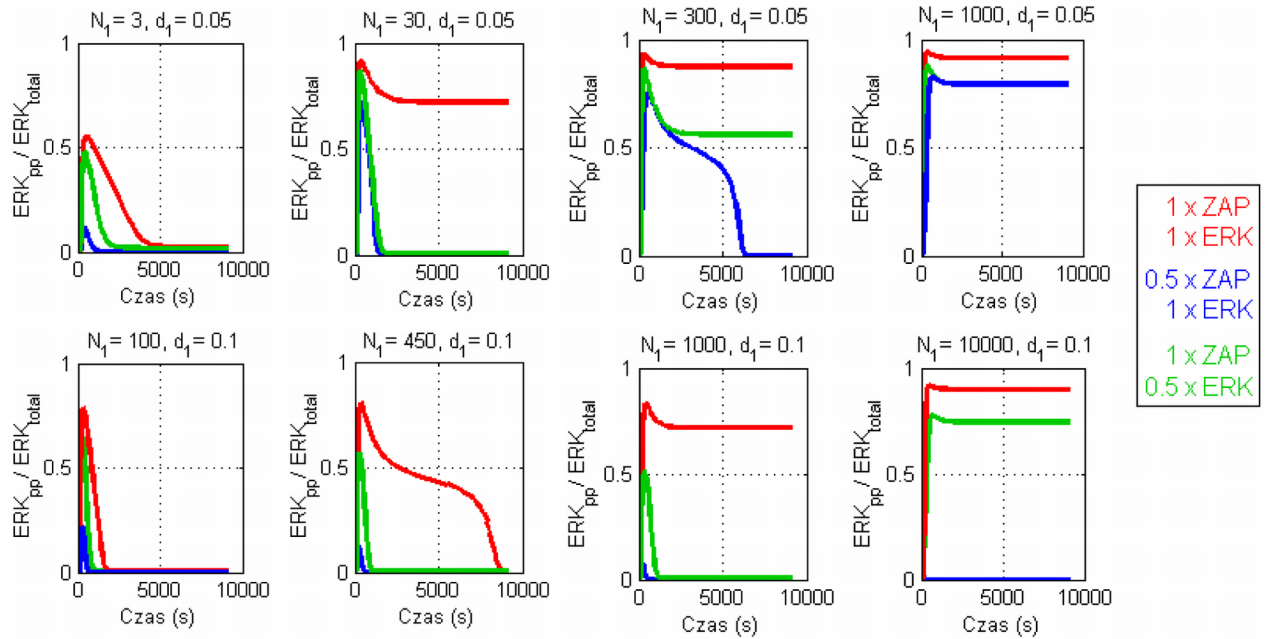
### 6.3 Analiza modelu

Chcemy zbadać jak liczby kopii genów kodujących poszczególne kinazy tworzących kaskadę sygnałową wpływają na dynamikę tej kaskady. W modelu Lipniacki i in. [74] rozważanych jest 100 000 cząsteczek *ZAP*, 100 000 cząsteczek *MEK* oraz 300 000 cząsteczek *ERK*. Zakładamy, że brak jednej kopii genu kodującej *ZAP*, *MEK* oraz *ERK* powoduje dwukrotne zmniejszenie całkowitej ilości odpowiednio *ZAP*, *MEK* oraz *ERK*, zaś zwiększenie liczby kopii genu trzech pociąga za sobą proporcjonalną zmianę ilości związanego z nim białka.

Jak pokazano na Rys. 6.5, poziomy  $SHP_p$  oraz  $ERK_{pp}$  zwiększają się wraz ze wzrostem liczby peptydów na tymocyt,  $N_1$ , i dążą asymptotycznie do stabilnego stanu stałego z wysokim poziomem  $SHP_p$  i  $ERK_{pp}$ . Dla współczynnika dysocjacji  $d_1 = 0.05$  komórka staje się przejściowo aktywna, tzn.  $\frac{ERK_{pp}^{max}}{ERK_{total}} > \frac{1}{2}$ , gdy do receptora  $TCR$  przyłączą się 3 peptydy, oraz stale aktywna gdy do receptora  $TCR$  przyłączy się 10 peptydów. Dla dwa razy większego współczynnika dysocjacji,  $d_1 = 0.1$ , komórka staje się przejściowo aktywna gdy do receptora  $TCR$  przyłączy się 15 peptydów, zaś do trwałej aktywacji komórki (tzn. dla  $\frac{ERK_{pp}(\infty)}{ERK_{total}} > \frac{1}{2}$ ) potrzeba około 480 peptydów.

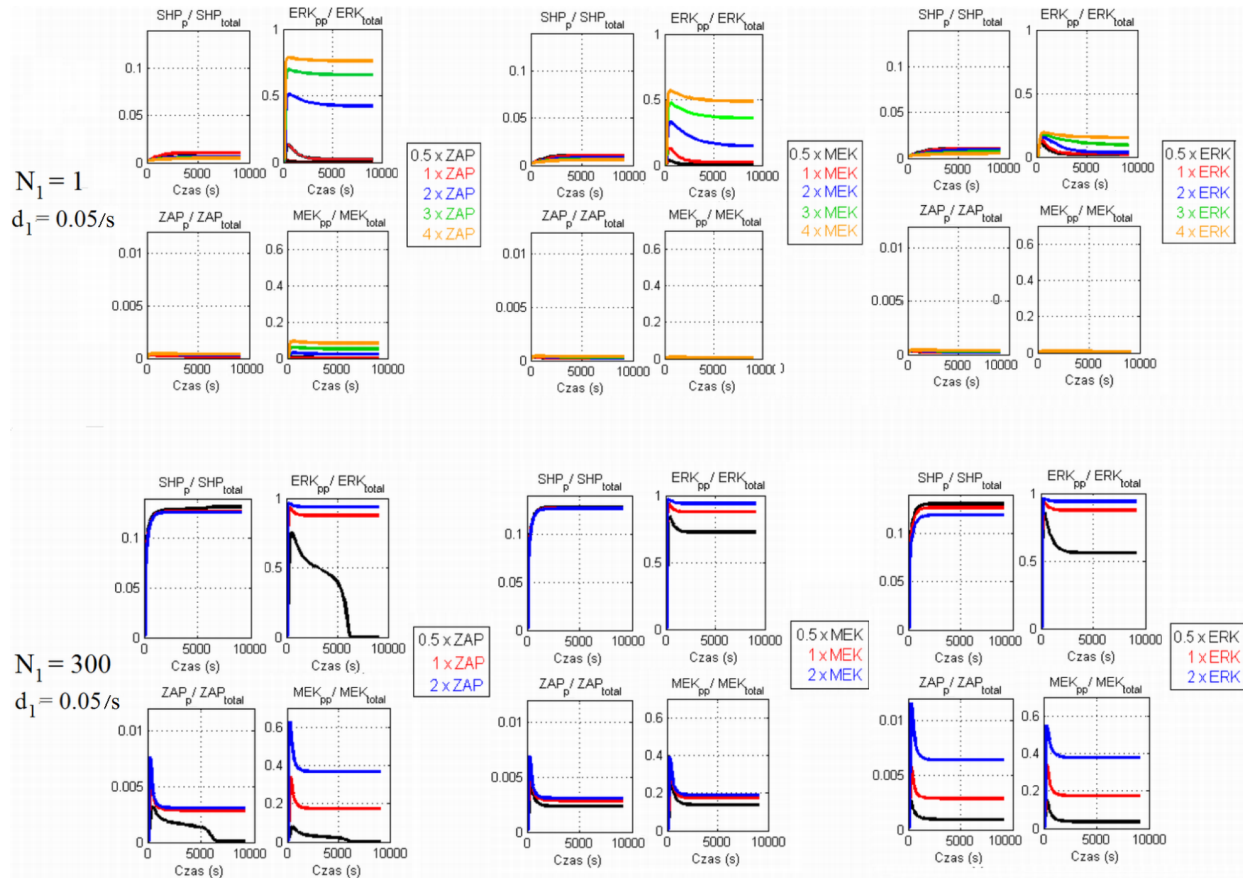


Rysunek 6.5: Poziom  $SHP_p(t)/SHP_{total}$  oraz  $ERK_{pp}(t)/ERK_{total}$  dla  $d_1 = 0.05$ ,  $d_1 = 0.1$  i odpowiednio czterech wartości  $N_1$ , gdzie  $d_1$  jest stałą dysocjacji peptydu  $MHC$  z kompleksu  $TCR$ , zaś  $N_1$  jest liczbą peptydów  $MHC$  na tymocyt. Warunek początkowy odpowiada komórce aktywnej.



Rysunek 6.6: Poziom  $ERK_{pp}(t)/ERK_{total}$  przy braku jednej kopii  $ZAP$  lub braku jednej kopii  $ERK$ , dla  $d_1 = 0.05$  (pierwszy wiersz) oraz  $d_1 = 0.1$  (drugi wiersz), i odpowiednio czterech wartości  $N_1$ . Kolor czerwony odpowiada tymocytom z normalną liczbą kopii genu  $ZAP$  i  $ERK$ , kolor niebieski - tymocytom z brakiem jednej kopii genu  $ZAP$  i normalną liczbą kopii  $ERK$ , oraz zielony - tymocytom z normalną liczbą kopii genu  $ZAP$  i brakiem jednej kopii  $ERK$ . Warunek początkowy odpowiada komórce aktywnej.

Jak pokazano na Rys. 6.6, brak jednej kopii genu  $ZAP$  powoduje większe obniżenie względnego poziomu  $ERK_{pp}(t)/ERK_{total}$  (Rys. 6.6, niebieska krzywa) niż brak jednej kopii genu  $ERK$  (zielona krzywa). Dla  $d_1 = 0.05$  przy braku jednej kopii genu  $ZAP$ , poziom  $ERK_{pp}(t)/ERK_{total}$  rośnie wraz ze wzrostem liczby przyłączonych peptydów i dąży asymptotycznie do stabilnego stanu stałego z dużą ilością  $ERK_{pp}(t)/ERK_{total}$  odpowiadającego komórce stale aktywnej (pierwszy wiersz), podczas gdy dla dwa razy większego współczynnika dysocjacji, przy braku jednej kopii  $ZAP$ , poziom  $ERK_{pp}(t)/ERK_{total}$  maleje wraz ze wzrostem liczby peptydów do bardzo niskiego poziomu odpowiadającemu komórce nieaktywnej (drugi wiersz).



Rysunek 6.7: Poziom  $ERK_{pp}(t)/ERK_{total}$  przy braku jednej kopii genu oraz zwielokrotnieniu liczby kopii genu odpowiednio  $ZAP$ ,  $MEK$  i  $ERK$ , dla  $N_1 = 1$  (pierwszy wiersz), oraz  $N_1 = 300$  (drugi wiersz);  $d_1 = 0.05/s$ . Kolory: czarny, czerwony, niebieski, zielony i pomarańczowy odpowiadają tymocytom z: brakiem jednej kopii genu, normalną liczbą kopii, dwa, trzy oraz cztery razy większą liczbą kopii genu odpowiednio  $ZAP$  (pierwsza kolumna),  $MEK$  (druga kolumna) oraz  $ERK$  (trzecia kolumna). Warunek początkowy odpowiada komórce aktywnej.

Jak pokazano na Rys. 6.7, wraz ze wzrostem liczby kopii genów  $ZAP$ ,  $MEK$  i  $ERK$  poziom  $SHP_p(t)/SHP_{total}$  maleje, zaś poziomy  $ERK_{pp}(t)/ERK_{total}$ ,  $ZAP_p(t)/ZAP_{total}$  oraz  $MEK_{pp}(t)/MEK_{total}$  rosną. Wzrost liczby kopii genu  $ZAP$  powoduje większą zmianę poziomu  $ERK_{pp}(t)/ERK_{total}$  niż wzrost liczby kopii genu  $ERK$  oraz  $MEK$ . Dla małej liczby peptydów

poziom  $SHP_p(t)/SHP_{total}$  jest niski (pierwszy wiersz), dla dużej liczby peptydów poziom  $SHP_p(t)/SHP_{total}$  jest wysoki.

Asymptotyczne zachowanie układu zależy od warunków początkowych. W komórce z wysokim poziomem  $SHP_p$  (50 000 cząsteczek), odpowiadającej komórce nieaktywnej, poziom  $ERK_{pp}$  przy braku jednej kopii  $ZAP$  i współczynnika dysocjacji  $d_1 = 0.1$  rośnie do bardzo wysokiego poziomu wraz ze wzrostem liczby peptydów (nie pokazano na Rysunku).

# Rozdział 7

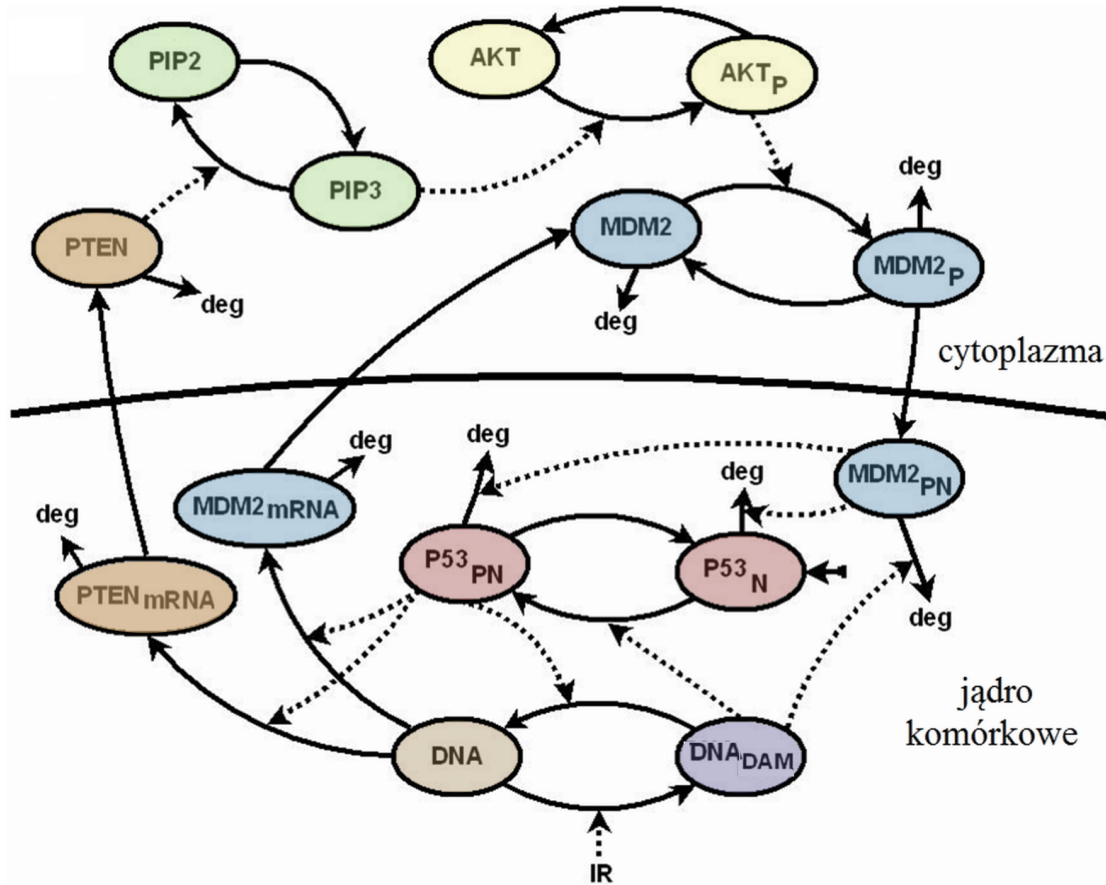
## Wpływ liczby kopii genów na dynamikę sieci regulatorowych czynników transkrypcyjnych $p53$ oraz $NF-\kappa B$

### 7.1 Układ regulatorowy $p53$

Jednym z najważniejszych białek supresorowych odpowiedzialnych za utrzymanie integralności genomu jest czynnik transkrypcyjny  $p53$ . Aktywne  $p53$  indukuje transkrypcję ponad stu genów kodujących białka odpowiedzialne za wstrzymanie cyklu komórkowego, naprawę DNA bądź indukcję procesu apoptozy. Gen kodujący białko  $p53$  jest najczęściej uszkodzonym genem w komórkach nowotworowych a jego mutacja występuje w ponad 50% nowotworów ([64]) Brak funkcjonalnego  $p53$  uniemożliwia zatrzymanie cyklu komórkowego, zachodzi wówczas nieograniczona replikacja uszkodzonego DNA, co zwiększa częstość mutacji i chroni zmutowane komórki przed apoptozą prowadząc do rozwoju nowotworu ([65]).

Aktywność  $p53$  podlega precyzyjnej kontroli (Rys. 7.1). W zdrowych komórkach  $p53$  jest nieaktywne i utrzymywane na niskim poziomie dzięki działaniu jego głównego inhibitora, białka  $Mdm2$ , które wiążąc się z  $p53$  powoduje jego ubikwitynację i szybką degradację ([66], [67], [68], [69], [70]). Jednocześnie  $p53$  jest czynnikiem transkrypcyjnym dla  $Mdm2$ . W odpowiedzi na uszkodzenie DNA,  $Mdm2$  ulega gwałtownej degradacji,  $p53$  aktywuje się a jego koncentracja w jądrze komórkowym rośnie. Cząsteczki  $p53$  przyłączają się do miejsc

regulatorowych w genomie indukując transkrypcję genów zaangażowanych w zatrzymanie cyklu komórkowego oraz naprawę DNA. Jednocześnie *p53* indukuje transkrypcję *Mdm2*. Nowo zsyntetyzowane *Mdm2* wchodzi do jądra komórkowego i łączy się z *p53* powodując jego degradację. Dopóki uszkodzenie DNA nie zostanie naprawione cały cykl powtarza się i trwa 5-7 godzin. Jeśli naprawa DNA nie jest możliwa lub trwa zbyt długo (powyżej ok. 15 godzin), *p53* indukuje transkrypcję czynników apoptycznych, które inicjują zaprogramowaną śmierć komórki.



Rysunek 7.1: Sieć regulatorowa *p53*.

Pomimo bardzo intensywnych badań zarówno eksperymentalnych jak i teoretycznych (ponad 60 tysięcy artykułów), mechanizmy regulujące *p53* nie są dostatecznie dobrze poz-



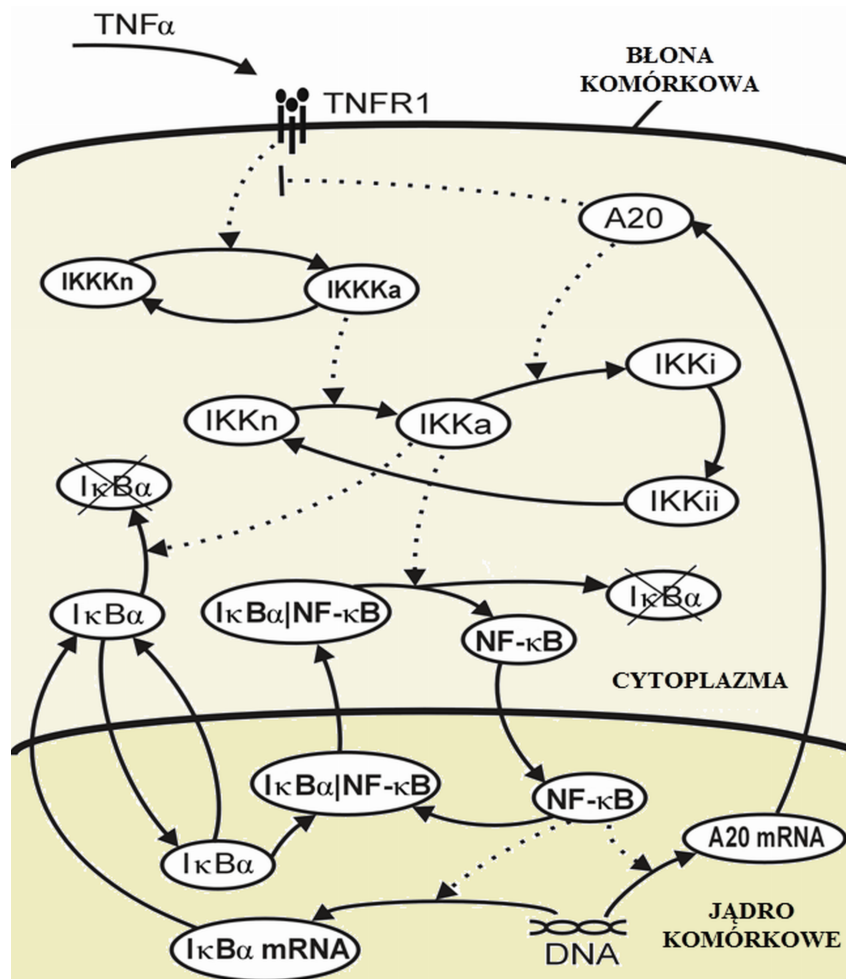
nane, ponieważ sieć regulatorowa  $p53|Mdm2$  jest bardzo złożona [71], występuje w niej wiele sprzężeń zwrotnych, w pracy [33] omawianych jest 7 sprzężeń ujemnych i 3 dodatnie. W zdrowych komórkach  $p53$  utrzymywane na niskim poziomie głównie poprzez pętlę ujemnego sprzężenia zwrotnego z  $Mdm2$ . Jeśli naprawa DNA nie jest możliwa, włącza się pętla dodatniego sprzężenia zwrotnego, która blokuje pętlę ujemnego sprzężenia, co powoduje wzrost  $p53$  do poziomu, przy którym może być zainicjowana apoptoza. Pętla dodatniego sprzężenia zwrotnego zawiera fosfatazę  $PTEN$  oraz kinazy  $PIP3$  i  $Akt$  i działa na zasadzie podwójnej negacji:  $p53 \rightarrow PTEN \dashv PIP3 \rightarrow Akt \rightarrow Mdm2 \dashv p53$ : najpierw  $p53$  indukuje syntezę  $PTEN$ , który działa jako inhibitor szlaku kinazy  $Akt$  poprzez katalizowanie reakcji odłączenia grupy fosforanowej od  $PIP3$  powodując jego dezaktywację. Aktywny  $PIP3$  jest niezbędny aby sfosforylować  $Akt$ , który z kolei inicjuje degradację  $p53$  poprzez fosforylację  $Mdm2$ , umożliwiając jego wejście do jądra i przyłączenie się do  $p53$ .

## 7.2 Układ regulatorowy $NF-\kappa B$

Czynnik transkrypcyjny  $NF-\kappa B$  jest głównym mediatorem wrodzonej odpowiedzi immunologicznej.  $NF-\kappa B$  reguluje ekspresję ponad stu genów kodujących białka odpowiedzialne za podział komórki, blokadę ścieżki apoptotycznej kaspaz (enzymów odpowiedzialnych za proces apoptozy) [38], [42] oraz zahamowanie rozprzestrzeniania się infekcji zanim organizm nie wypracuje adaptacyjnej, specyficznej do danej infekcji, odpowiedzi immunologicznej.

W zdrowych komórkach nieaktywne  $NF-\kappa B$  zlokalizowane jest w cytoplazmie poprzez wiązanie z jego głównym białkiem inhibitorowym  $I\kappa B\alpha$  (Rys. 7.2). Wiązanie  $I\kappa B\alpha$  z  $NF-\kappa B$  blokuje transport  $NF-\kappa B$  do jądra komórkowego. Stymulacja cytokinami takimi jak  $TNF\alpha$  (czynnik martwicy nowotworu) i IL-1 (Interleukin 1) prowadzi do aktywacji kompleksu kinaz białkowych IKK, które katalizują fosforylację  $I\kappa B\alpha$ , co prowadzi do jego ubiquitinizacji i szybkiej degradacji. Uwolnione  $NF-\kappa B$  jest bardzo szybko transportowane do jądra, gdzie przyłącza się do miejsc regulatorowych w promotorach wielu genów, m.in. swoich inhibitorów

*IκBα* i A20, indukując ich ekspresję oraz tworząc dwie pętle ujemnego sprzężenia zwrotnego: *NF-κB*–*IκBα* i *NF-κB*–A20–IKK.



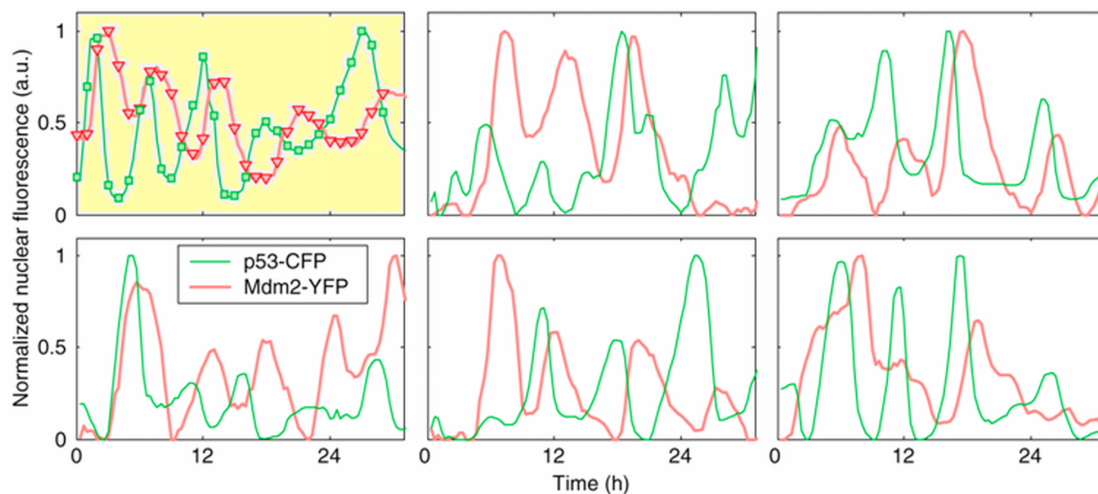
Rysunek 7.2: Sieć regulatorowa *NF-κB*.

W pętli sprzężenia *NF-κB*–*IκBα*, białko *NF-κB* po degradacji *IκBα* wchodzi do jądra komórkowego i indukuje transkrypcję *IκBα*; nowo zsyntetyzowane *IκBα* wchodzi do jądra, łączy się z *NF-κB* tworząc nieaktywny kompleks i transportuje go z powrotem do cytoplazmy. W pętli sprzężenia *NF-κB*–A20–IKK, białko A20, które nie jest syntetyzowane przed stymulacją, dezaktywuje kinazy IKK chroniąc nowo zsyntetyzowane *IκBα* przed degradacją. A20

jest inhibitorem kinazy IKK, która pośredniczy w fosforylacji  $I\kappa B\alpha$ ; gdy nie ma ekspresji A20, IKK pozostaje aktywne co prowadzi do szybkiej degradacji nowo zsyntetyzowanego  $I\kappa B\alpha$  [43], [44]. Dopóki uszkodzenie komórki nie zostanie naprawione cały cykl powtarza się i trwa 60-75 minut.

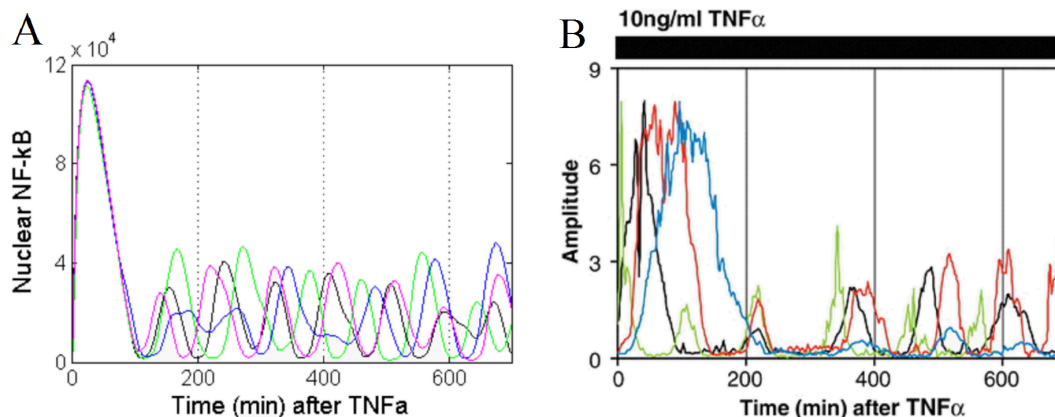
### 7.3 Trwałe oscylacje w układach $p53$ i $NF-\kappa B$

Wiele molekularnych ścieżek regulatorowych ma potencjał do generowania trwałych oscylacji o zmiennej amplitudzie lecz dobrze zachowanym okresie, które powstają w odpowiedzi na zewnętrzny stres lub stymulacje cytokinami. Eksperymenty na pojedynczych komórkach pokazują, że w układach regulatorowych  $p53$  i  $NF-\kappa B$  okres oscylacji w poszczególnych komórkach jest jednorodny w populacji, niewrażliwy na siłę stymulacji oraz nie jest podatny na nadekspresję czynników transkrypcyjnych  $p53$  i  $NF-\kappa B$ .



Rysunek 7.3: (Geva-Zatorsky i in. 2006, grupa U. Alona) Niegasnące oscylacje jądrowych poziomów oznaczonych fluorescencyjnie białek  $p53$ -CFP oraz  $Mdm2$ -YFP po uszkodzeniu DNA promieniowaniem gamma, obserwowane przez 72 godziny trwania eksperymentu. Lewy

górny obrazek: oscylacje w komórce pokazanej na Rys. 1.2; pozostałe obrazki: oscylacje w pięciu innych komórkach.



Rysunek 7.4: (Nelson i in. 2004, grupa M.R.H. White) Panel A: poziom jądrowego  $NF-\kappa B$  w czterech różnych komórkach. Panel B: analiza w czasie poziomów  $\frac{NF-\kappa B_{jadr}}{NF-\kappa B_{cyt}}$  dla czterech różnych komórek SK-N-AS; czarny pasek powyżej obrazka reprezentuje ciągłą stymulację komórek czynnikiem TNF.

Pomimo, że pojedyncza pętla ujemnego sprzężenia zwrotnego wraz z opóźnieniem może generować trwałe oscylacje, układy regulatorowe takie jak  $p53$  lub  $NF-\kappa B$  zawierają zarówno ujemne jak i dodatnie sprzężenia zwrotne (omawiane w [32] dla  $NF-\kappa B$ , oraz w [33] dla  $p53$ ). Własność ta, wraz z obserwacją, że układy oscylatorowe opierające się jedynie na ujemnym sprzężeniu zwrotnym i opóźnieniu zazwyczaj oscylują powyżej ograniczonego zakresu wartości parametrów [34] i [35], prowadzi do konkluzji, że dodatnie sprzężenie zwrotne jest konieczne, aby wygenerować trwałe oscylacje.

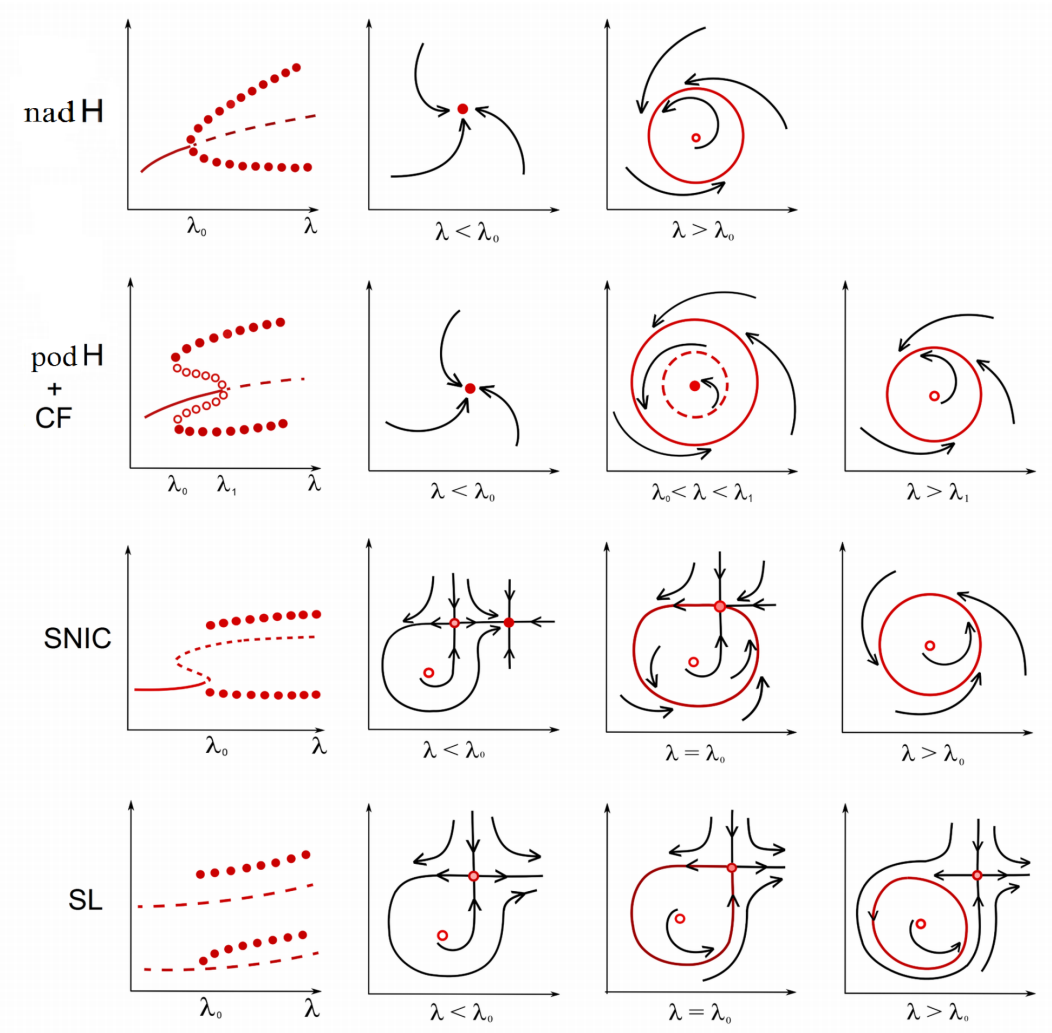
Proponujemy nową metodę walidacji modeli ścieżek molekularnych, która opiera się na analizie wrażliwości okresu oscylacji na zmieniającą się liczbę kopii poszczególnych genów oraz na poziom stymulacji. Za pomocą tej metody pokazujemy, że istniejące modele układu regulatorowego  $p53$ , w których oscylacje rodzą się w punkcie *bifurkacji siodło-węzeł* lub

podkrytyczną bifurkacją Hopfa (charakterystycznych dla układów ze współdziałającym dodatnim i ujemnym sprzężeniem zwrotnym) generują oscylacje o okresie bardzo wrażliwym na zmianę liczby kopii genów, co sugeruje, że mogą niewłaściwie tłumaczyć wyniki istniejących eksperymentów. Skoncentrujemy się na dwóch układach czynników transkrypcyjnych,  $p53$  oraz  $NF-\kappa B$ , o których wiadomo, że generują trwałe oscylacje o zmiennej amplitudzie lecz dobrze zachowanym okresie, i dla których dostępne są dane z eksperymentów na pojedynczych komórkach [36], [37]. Zaletą tych eksperymentów jest to, że pokazują dynamikę poszczególnych komórek, która maskowana jest w badaniach populacyjnych.

W eksperymentach przeprowadzonych przez Geva-Zatorsky'ego i in. [36] (Rys. 7.3) oraz Nelsona i in. [37] (Rys. 7.4), oscylacje zostały zwizualizowane przez fluorescencyjnie oznaczone białka w pojedynczych transfekowanych komórkach. Ponieważ układy regulatorowe  $p53$  i  $NF-\kappa B$  są nieliniowe, transfekowane komórki, które mają podwyższoną liczbę kopii genu odpowiednio  $p53$  lub RelA (kluczowy komponent  $NF-\kappa B$ ), mogą mieć dynamikę radykalnie różną od dynamiki normalnych komórek diploidalnych, zwłaszcza jeżeli, jak pokazaliśmy w Rozdziale 3, układ regulatorowy zawiera dodatnie sprzężenie zwrotne [39]. Jednakże eksperymenty pokazują, że okres oscylacji jest prawie taki sam w komórkach transfekowanych jak w normalnych diploidalnych komórkach analizowanych w eksperymentach populacyjnych. Ponadto, również długość okresu oscylacji nie jest wrażliwa na poziom stymulacji (dawkę promieniowania gamma w przypadku  $p53$ , oraz dawkę  $TNF\alpha$  w przypadku  $NF-\kappa B$ ). Te eksperymentalne obserwacje nakładają ograniczenia na mechanizmy regulatorowe odpowiedzialne za oscylacje - ograniczenia, które nie były brane pod uwagę w poprzednich badaniach teoretycznych. Typ bifurkacji jest kluczowym wyznacznikiem wrażliwości okresu oscylacji na parametry.

## 7.4 Typy bifurkacji prowadzące do oscylacji

Istnieją cztery główne typy bifurkacji prowadzące do oscylacji [40]: *nadH* (nadkrytyczna bifurkacja Hopfa), *podH* (podkrytyczna bifurkacja Hopfa), *SNIC* (siodło-węzeł) oraz *SL* (bifurkacja siodło-pętla) (Rys. 7.5).



Rysunek 7.5: Typy bifurkacji prowadzące do oscylacji: *nadH*, *podH*, *SNIC* oraz *SL*. Pierwsza kolumna: diagramy bifurkacji; kolejne kolumny: portrety fazowe dla różnych wartości parametru bifurkacyjnego  $\lambda$ .

1) W bifurkacji *nadH* oscylacje rodzą się w punkcie zerową amplitudą i skończonym (niezerowym) okresem.

2) W bifurkacji *podH* oscylacje rodzą się w punkcie ze skończonym okresem i skończoną amplitudą. Ten typ bifurkacji zazwyczaj związany jest z bifurkacją *CF* (*cyclic fold*), w której stabilne i niestabilne orbity okresowe zanikają (lub rodzą się). W takim układzie występuje histereza; stabilne oscylacje rodzą się w punkcie  $\lambda_1$  gdy parametr bifurkacji rośnie oraz znikają w punkcie  $\lambda_0 < \lambda_1$  gdy parametr bifurkacji maleje. Dla  $\lambda \in (\lambda_0, \lambda_1)$  współistnieją stabilny punkt stały oraz stabilny cykl graniczny.

3) W bifurkacji *SNIC*, punkty siodło i węzeł anihilują i powstaje cykl graniczny o dużym okresie (nieskończonym w punkcie bifurkacji). W punkcie bifurkacji układ ma orbitę homokliniczną.

4) Bifurkacja *SL* (lub *homokliniczna*) jest w pewnym sensie podobna do *SNIC*: w punkcie bifurkacji układ ma orbitę homokliniczną do siodła, która następnie daje początek orbicie okresowej o dużym okresie (nieskończonym w punkcie bifurkacji). Zauważmy, że "minimalny układ" generujący bifurkację *SL* dla  $\lambda < \lambda_0$  nie ma ani stabilnego stanu stałego ani cyklu granicznego, dlatego układ taki nie jest dobrym kandydatem na model biologiczny.

Bifurkacja *nadH* może powstawać w układach biologicznych, które mają pojedynczą pętlę ujemnego sprzężenia zwrotnego. Układy ze współdziałającym ujemnym i dodatnim sprzężeniem zwrotnym mogą generować bifurkację *podH* lub *SNIC*, ale także bifurkację *nadH*. Ogólnie, układy oparte na bifurkacji *nadH* generują oscylacje o względnie dobrze zachowanym okresie (który zmienia się powoli wraz z parametrem bifurkacyjnym), lecz o zmiennej amplitudzie. W przeciwieństwie do *nadH*, w bifurkacjach *SNIC* i *SL* amplituda jest raczej stała, ale okres oscylacji może być dowolnie duży (patrz także [41]).

W dalszej części Rozdziału 7 skoncentrujemy się na układach regulatorowych dwóch czynników transkrypcyjnych: *p53* oraz *NF- $\kappa$ B*.

## 7.5 Modele układu regulatorowego $p53|Mdm2$

Istnieją dwie główne klasy modeli zaproponowane, aby wyjaśnić oscylacje układu regulatorowego  $p53|Mdm2$ .

1) Grupa J. Tyson'a zaproponowała cztery modele oznaczane: Zhang I, II, III oraz Zhang IV, w których oscylacje powstają w wyniku współdziałania dodatniego i ujemnego sprzężenia zwrotnego [34, 47]. W modelach tych oscylacje rodzą się w punkcie bifurkacji *podH* (modele Zhang I i Zhang III [34]) lub w punkcie bifurkacji *SNIC* (modele Zhang II i Zhang IV).

2) W innych trzech modelach, Ma i in., Puszynski i in. oraz Batchelor i in., oscylacje generowane są w wyniku współdziałania ujemnego sprzężenia i opóźnienia w czasie, które spowodowane jest obecnością form pośrednich [48, 49] lub wprowadzone jest na poziomie transkrypcji i translacji mRNA *Mdm2* [50, 51]. W modelach tych oscylacje rodzą się w punkcie bifurkacji *nadH*.

Naszym celem jest weryfikacja, które modele potrafią wyjaśnić obserwowane oscylacyjne zachowanie układu  $p53|Mdm2$ , a w szczególności, odporność długości okresu oscylacji na zmianę liczby kopii genów *p53* i *Mdm2* oraz poziomu stymulacji. Najpierw analizujemy cztery modele zaproponowane przez grupę Tyson'a; następnie analizujemy nasz model "testowy", który opiera się na dwóch sprzężeniach zwrotnych, dodatnim i ujemnym, występujących w modelu Ciliberto [47], ale w naszym modelu oscylacje generowane są w punkcie bifurkacji *nadH* zamiast *SNIC*. Następnie analizujemy modele Ma i in. [50], Puszynski i in. [49] oraz Batchelor i in. [73], w których oscylacje generowane są w wyniku współdziałania ujemnego sprzężenia zwrotnego i opóźnienia oraz rodzą się w punkcie bifurkacją *nadH*.

Parametr opisujący liczbę kopii genu nie występuje jawnie w żadnym z analizowanych modeli. Wprowadzamy go poprzez założenie, że współczynnik produkcji mRNA lub białka (jeżeli nie występuje etap transkrypcji mRNA) jest proporcjonalny do liczby kopii genu *p53* lub *Mdm2*. Ścisłej, zakładamy, że pierwotne współczynniki syntezy odpowiadają komórkom diploidalnym i proporcjonalnie modyfikujemy te współczynniki dla komórek mających dowolną liczbę kopii genów.



### 7.5.1 Kryteria poprawnego modelu

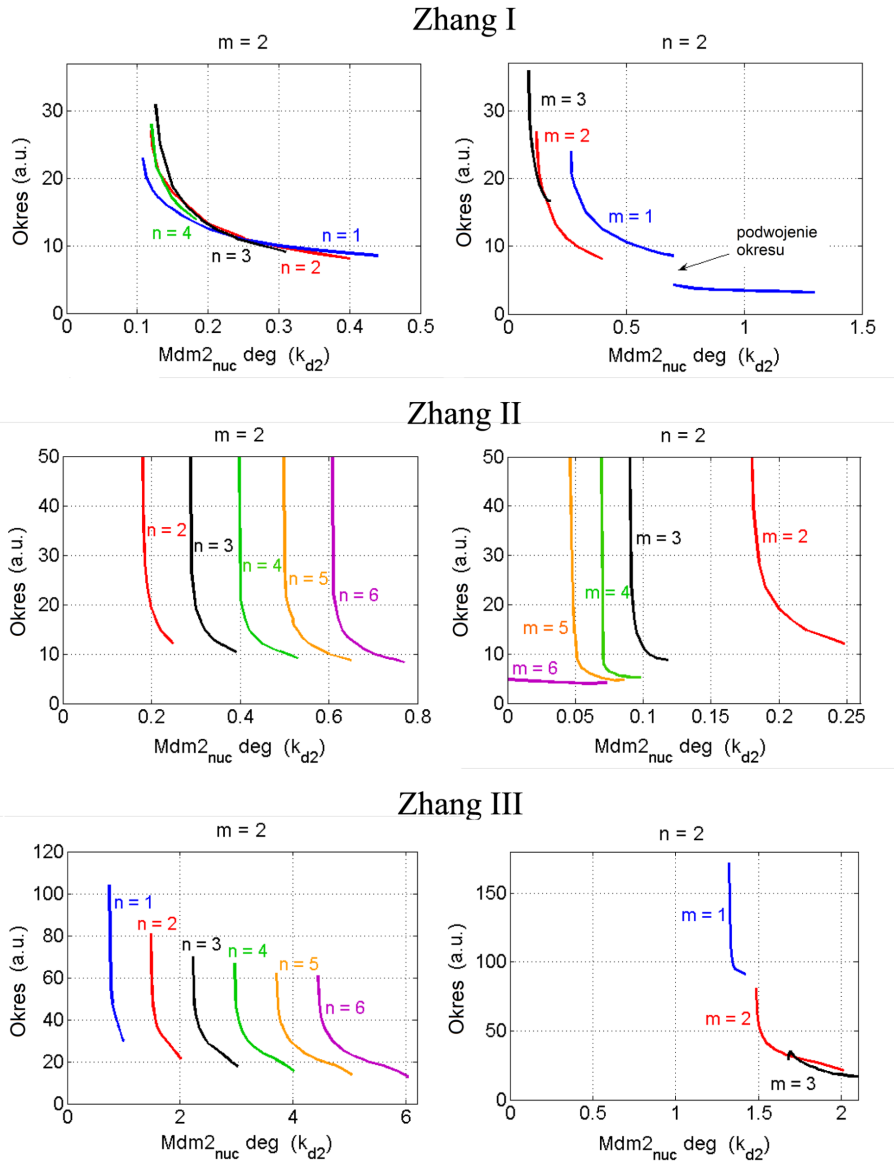
Uszkodzenie DNA powoduje gwałtowną zmianę współczynników degradacji  $p53$  i  $Mdm2$  ( $p53$  jest stabilizowane przez fosforylację, podczas gdy  $Mdm2$  destabilizuje się), co prowadzi do przejścia układu ze stanu stałego z niskim poziomem  $p53$  do trwałych oscylacji. Istnienie tych niegasnących oscylacji, o charakterystycznym okresie 5-6 godzin, lecz o bardzo zmiennej amplitudzie, zostało zademonstrowane przez Geva-Zatorsky i in. w eksperymentach na pojedynczych komórkach wykorzystujących fluorescencyjnie oznaczone  $p53$  i  $Mdm2$  [36].

Poprawny model ścieżki regulatorowej  $p53|Mdm2$  powinien spełniać (przynajmniej) następujące ograniczenia zadane przez eksperyment:

1. Oscylacje poziomów  $p53$  i  $Mdm2$  powinny być indukowane poprzez zmniejszanie się współczynnika degradacji  $p53$  lub zwiększanie się współczynnika degradacji  $Mdm2$ , następujące w wyniku uszkodzenia DNA.
2. Okres oscylacji poziomów  $p53$  i  $Mdm2$  powinien być odporny na zmianę liczby kopii genów  $p53$  i  $Mdm2$  oraz na poziom uszkodzenia DNA.
3. Zakres parametru degradacji  $Mdm2$ , dla którego istnieją oscylacje, nie powinien być zbyt wrażliwy na liczbę kopii genów  $p53$  i  $Mdm2$ .

### 7.5.2 Analiza modeli opartych na dodatnim i ujemnym sprzężeniu zwrotnym

W modelach Tyson'a [34] zakłada się, że oscylacje indukowane są w wyniku uszkodzenia DNA poprzez destabilizację  $Mdm2$  (zwiększenie współczynnika degradacji  $Mdm2$ ). Wyznaczyliśmy zatem okres oscylacji ze względu na współczynnik degradacji  $Mdm2$  dla komórek z różną liczbą kopii genów  $p53$  i  $Mdm2$  (Rys. 7.6 i 7.7). A mianowicie, rozważyliśmy komórki z  $n = 2$  kopiami  $Mdm2$  i  $m \in \{1, 2, 3, 4, 5, 6\}$  kopiami  $p53$ , oraz  $m = 2$  kopiami  $p53$  i  $n \in \{1, 2, 3, 4, 5, 6\}$  kopiami  $Mdm2$ .



Rysunek 7.6: Analiza modeli grupy Tyson'a: Zhang I, Zhang II oraz Zhang III. Okres oscylacji dla stabilnego cyklu granicznego jako funkcja parametru degradacji  $Mdm2$  ( $k_{d2}$ ). Kolory: niebieski, czerwony, czarny, zielony, pomarańczowy i fioletowy odpowiadają komórkom z  $n = 1, 2, 3, 4, 5, 6$  (liczba kopii genu  $Mdm2$ ) oraz ustalonej liczbie kopii genu  $p53$ :  $m = 2$  (lewa kolumna) lub komórkom z  $m = 1, 2, 3, 4, 5, 6$  (liczba kopii genu  $p53$ ) oraz ustalonej liczbie kopii genu  $Mdm2$ :  $n = 2$  (prawa kolumna).

W pierwszych trzech modelach rozważanych przez Zhang i in. [34] zakres współczynnika degradacji  $Mdm2$ , dla którego układ posiada stabilny cykl graniczny, jest bardzo wrażliwy na liczbę kopii genów  $p53$  i  $Mdm2$ , Rys. 7.6. Ponadto, dla danej liczby kopii genów  $p53$  i  $Mdm2$  okres oscylacji silnie zmienia się wraz ze zmianą współczynnikiem degradacji  $Mdm2$ . To znaczy:

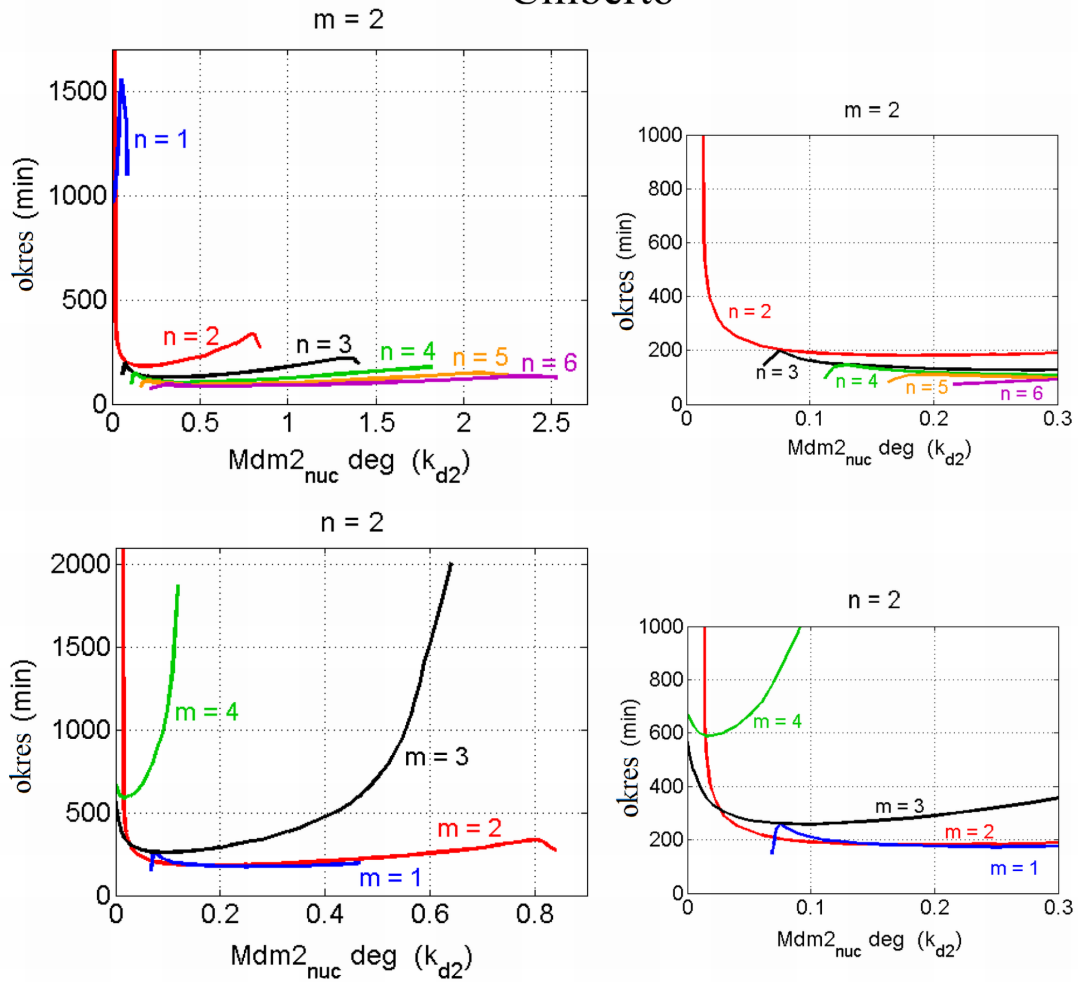
- w modelu Zhang I oscylacje rodzą się w punkcie bifurkacji  $podH$ , gdy współczynnik degradacji  $Mdm2$  rośnie, oraz znikają z bifurkacją  $CF$  z dużym lecz skończonym okresem, gdy współczynnik degradacji  $Mdm2$  maleje. Dla  $m = 2, n > 4$  lub  $n = 2, m > 3$  układ nie oscyluje.

- w modelu Zhang II, oscylacje rodzą się w punkcie bifurkacji  $SNIC$  z nieskończonym okresem. Dla  $n = 2$  i  $m = 6$  układ oscyluje dla dowolnie małego  $k_{d2}$  (w modelu tym  $k_{d2}$  kontroluje jedynie degradację  $Mdm2$  w jądrze komórkowym, podczas gdy degradację  $Mdm2$  w cytoplaźmie kontroluje niezależny parametr). Dla  $m = 2, n = 1$  układ nie oscyluje. Zakres oscylacji dla każdej pary  $(m, n)$  jest bardzo wąski i istnieje tylko małe zazębienie pomiędzy zakresami oscylacji dpowiadającymi różnym parom  $(m, n)$ .

- w modelu Zhang III, podobnie jak w modelu Zhang I, oscylacje rodzą się w punkcie bifurkacji  $podH$  z niezerową amplitudą i znikają, gdy współczynnik degradacji  $Mdm2$  maleje w punkcie bifurkacji  $CF$  z dużym, ale skończonym okresem). Dla  $n = 2, m > 3$  układ nie wykazuje oscylacji. Dla  $n = 2, m = 3$  oscylacje rodzą się w punkcie bifurkacją  $nadH$ . Podobnie jak w modelu Zhang II, zakres oscylacji w przestrzeni parametru  $k_{d2}$  dla każdej pary  $(m, n)$  jest bardzo wąski oraz istnieje tylko małe zazębienie pomiędzy zakresami odpowiadającymi różnym parom  $(m, n)$ .

Te trzy modele implikują bardzo szeroki rozkład okresu oscylacji w populacji komórek, który nie jest obserwowany eksperymentalnie. Ponadto, ponieważ dla każdej pary  $(m, n)$  układ oscyluje dla różnych zakresów parametru degradacji  $Mdm2$ , to na danym poziomie stymulacji oscylowałyby jedynie komórki dla szczególnych par  $(m, n)$ .

## Ciliberto



Rysunek 7.7: Analiza modelu Ciliberto grupy Tyson'a. Okres oscylacji dla stabilnego cyklu granicznego jako funkcja parametru degradacji  $Mdm2$  ( $k_{d2}$ ). Kolory: niebieski, czerwony, czarny, zielony, pomarańczowy i fioletowy odpowiadają komórkom z  $n = 1, 2, 3, 4, 5, 6$  (liczba kopii genu  $Mdm2$ ) oraz ustalonej liczbie kopii genu  $p53$ :  $m = 2$ ; lub komórkom z  $m = 1, 2, 3, 4, 5, 6$  (liczba kopii genu  $p53$ ) oraz ustalonej liczbie kopii genu  $Mdm2$ :  $n = 2$ . W prawej kolumnie pokazujemy powiększone fragmenty paneli z lewej kolumny.

Jak obserwujemy na Rys. 7.7, bardziej odporny jest model Ciliberto i in. (Zhang IV) [47]. W modelu tym ujemne sprzężenie powstaje gdy  $p53$  indukuje syntezę  $Mdm2$ , następnie

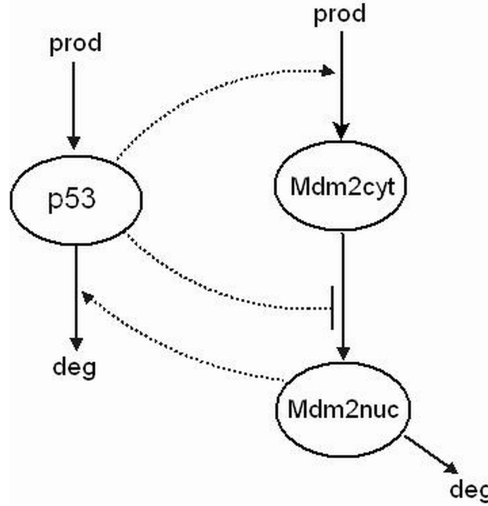
*Mdm2* jest transportowane do jądra, gdzie katalizuje degradację *p53*. Dodatkowo sprzężenie powstaje, gdy fosforylacja *Mdm2*, która jest konieczna do jego wejścia do jądra, jest inhibowana przez *p53* (poprzez długą ścieżkę sygnałową wymagającą obecności PTEN, PIP3 i Akt – nie zawartych jawnie w modelu). W ten sposób *p53* inhibuje swoją degradację. Dla komórek diploidalnych ( $m = n = 2$ ) model generuje oscylacje *p53* poprzez bifurkację *SNIC*. Dla komórek z jedną kopią genu *Mdm2* ( $m = 2, n = 1$ ) układ oscyluje dla dowolnie małego  $k_{d2}$ , współczynnika degradacji jądrowego *Mdm2*, który kontroluje degradację *Mdm2* w jądrze, podczas gdy dla  $m = 2$  i  $n \in \{3, 4, 5, 6\}$ , oraz  $m = 1$  i  $n = 2$  oscylacje rodzą się w punkcie bifurkacją *nadH* i tylko dla tych par  $(m, n)$  okres jest wolno zmieniającą się funkcją współczynnika degradacji *Mdm2* (dla innych parametrów okres oscylacji zmienia się silnie wraz z współczynnikiem degradacji). Dla  $n = 2, m \in \{3, 4\}$  układ oscyluje dla dowolnie małego  $k_{d2}$ , podczas gdy dla  $n = 2$  i  $m \in \{5, 6\}$  układ nie wykazuje stabilnych oscylacji.

Podsumowując, model Ciliberto i in. [47] generuje oscylacje o odpornym okresie tylko w przypadku transfekcji *Mdm2* ( $m = 2$  i  $n \in \{3, 4, 5, 6\}$ ), gdy oscylacje rodzą się w punkcie *nadkrytycznej* bifurkacji Hopfa. To sugeruje, że właściwy model powinien opierać się na *nadkrytycznej* bifurkacji Hopfa a nie na bifurkacji *SNIC* lub *podkrytycznej* bifurkacji Hopfa.

### 7.5.3 Model "testowy" z dodatnim i ujemnym sprzężeniem zwrotnym oparty na nadkrytycznej bifurkacji Hopfa

Skonstruowaliśmy model "testowy" układu regulatorowego  $p53|Mdm2$ , który zawiera dodatkowo i ujemne pętle sprzężenia zwrotnego, takie jak w oryginalnym modelu Ciliberto i in. [47], jednak nasz model ma zasadniczo inną dynamikę. W modelu Ciliberto oscylacje generowane są przez dwa typy bifurkacji (a nawet są obecne dla zerowego parametru degradacji *Mdm2*), w naszym modelu "testowym" oscylacje powstają w wyniku nadkrytycznej bifurkacji Hopfa dla  $m, n \in \{1, 2, 3, 4, 5, 6\}$ .

Model składa się tylko z trzech komponent: całkowitego  $p53$ , cytoplazmatycznego  $Mdm2$  ( $Mdm2_{cyt}$ ) oraz jądrowego  $Mdm2$  ( $Mdm2_{nuc}$ ). Ujemne sprzężenie zwrotne powstaje, gdy  $p53$  indukuje produkcję  $Mdm2$ , podczas gdy jądrowe  $Mdm2$  podwyższa degradację  $p53$ . Dodatnie sprzężenie zwrotne jest indukowane gdy  $p53$  blokuje import  $Mdm2$  do jądra (Rys. 7.8).



Rysunek 7.8: Diagram układu  $p53|Mdm2$ : model testowy. Linie kropkowane ze strzałkami reprezentują indukcję; linia kropkowana z kreską reprezentuje inhibicję.

Układ regulatorowy opisany jest przez trzy równania różniczkowe zwyczajne opisujące ewolucję poziomów cząsteczek całkowitego  $p53$ , cytoplazmatycznego  $Mdm2$  i jądrowego  $Mdm2$ :

$$\frac{d}{dt}p53 = ms_1 - k_{d1}p53(Mdm2_{nuc})^2, \quad (83)$$

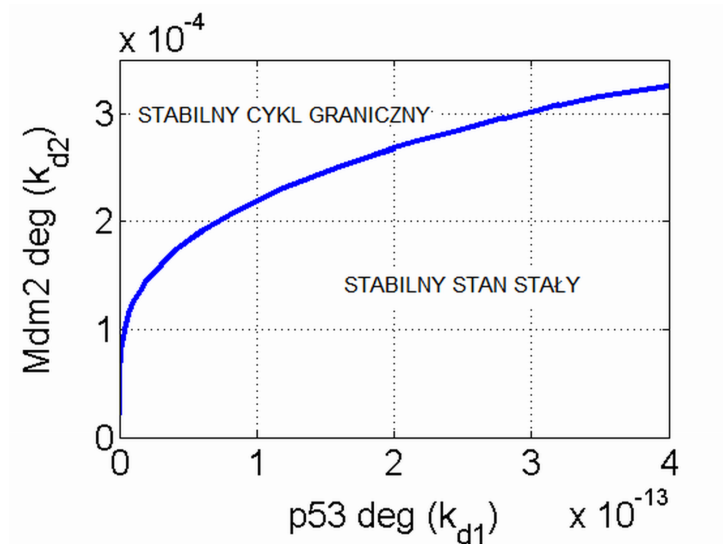
$$\frac{d}{dt}Mdm2_{cyt} = n \left( s_2 + s_3 \frac{(p53)^3}{s_4 + (p53)^3} \right) - k_1 \frac{k_2}{k_2 + p53} Mdm2_{cyt}, \quad (84)$$

$$\frac{d}{dt}Mdm2_{nuc} = k_1 \frac{k_2}{k_2 + p53} Mdm2_{cyt} - k_{d2}Mdm2_{nuc}, \quad (85)$$

gdzie  $m$  i  $n$  są liczbami kopii genów  $p53$  i  $Mdm2$ . Współczynniki  $s_1$ ,  $s_2$  i  $s_3$  są współczynnikami produkcji na kopię genu (patrz Tabela 1 dla wartości parametrów). Transkrypcja  $Mdm2$  regulowana jest głównie przez tetramery  $p53$ , dlatego założyliśmy, że Ciliberto i in.

[47], że  $p53$  indukuje transkrypcję  $Mdm2$  zgodnie z funkcją Hilla z wykładnikiem 3. Nieliniowość w członie degradacji  $p53$  wynika z faktu, że  $Mdm2$  musi przyłączyć kilka ubikwityn do  $p53$ , aby zainicjować jego degradację. Podobnie jak w [47], transport  $Mdm2$  z cytoplazmy do jądra jest inhibowany przez działanie  $p53$ .

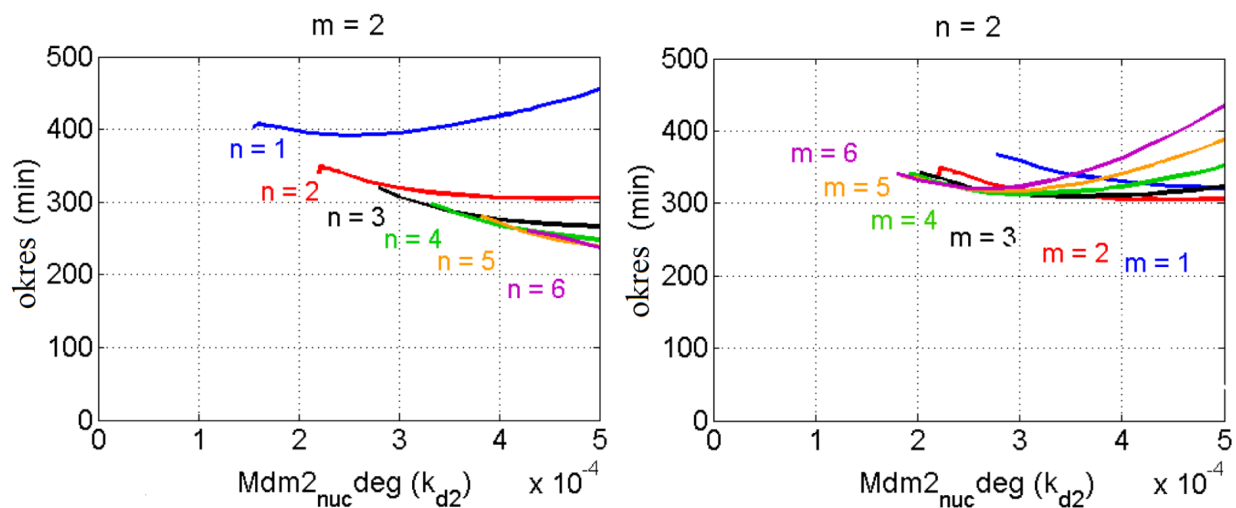
Uszkodzenie DNA nie występuje jawnie w naszym modelu, ale może być wprowadzone poprzez zwiększenie współczynnika degradacji  $Mdm2$  lub zmniejszenie współczynnika degradacji  $p53$ . Model ma wymaganą własność: przejście ze stabilnego punktu do stabilnego cyklu granicznego jest konsekwencją zwiększania degradacji  $Mdm2$  lub zmniejszania degradacji  $p53$  (Rys. 7.9). Pomimo, że nadkrytyczna bifurkacja Hopfa jest bardziej charakterystyczna dla układów tylko z ujemnym sprzężeniem zwrotnym, w tym modelu obecność dodatniego sprzężenia jest potrzebna do powstania oscylacji, tzn. oscylacje znikają po założeniu, że import  $Mdm2$  do jądra jest niezależny od  $p53$ .



Rysunek 7.9: Obszar stabilnych cykli granicznych oraz stabilnych punktów stałych w płaszczyźnie  $(p53 \text{ deg}, Mdm2 \text{ deg})$  dla komórek diploidalnych ( $m = n = 2$ ), dla modelu testowego.

Tak jak dla modeli Zhang i in. [34], zbadaliśmy zależność okresu oscylacji od współczynnika degradacji  $Mdm2$  i  $p53$  oraz liczby kopii genu  $Mdm2$ . Jak pokazujemy na Rys. 7.10,

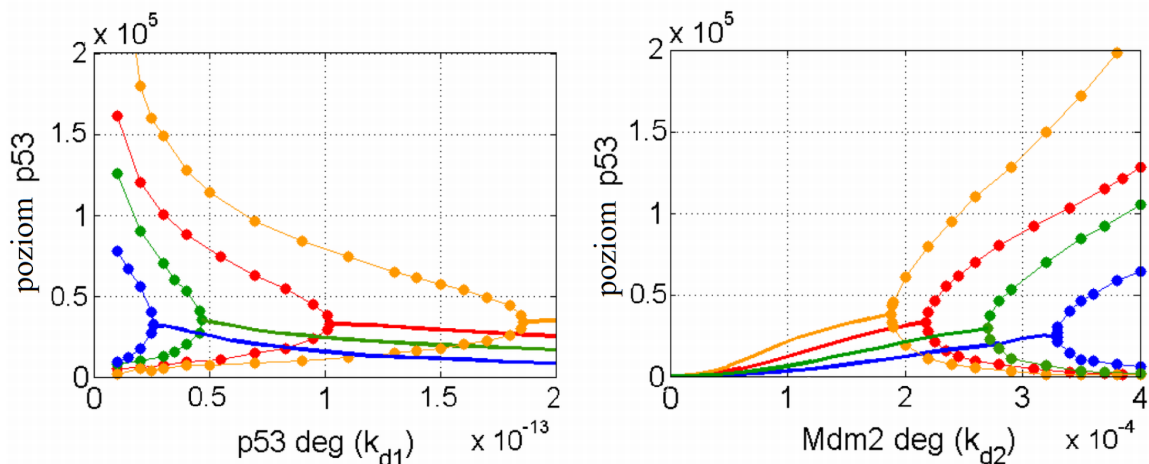
okres oscylacji słabo zależy od parametru degradacji *Mdm2* oraz liczby kopii genu, jednakże zakres współczynnika degradacji *Mdm2*, dla którego układ oscyluje, istotnie zależy od liczby kopii genów *p53* i *Mdm2*, Rys. 7.10 i 7.11.



Rysunek 7.10: Analiza modelu testowego. Okres oscylacji dla stabilnego cyklu granicznego jako funkcja parametru degradacji *Mdm2* ( $k_{d2}$ ). Kolory: niebieski, czerwony, czarny, zielony, pomarańczowy i fioletowy odpowiadają komórkom z  $N = 1, 2, 3, 4, 5, 6$  (liczba kopii genu *Mdm2*) oraz ustalonej liczbie kopii genu *p53*:  $m = 2$ ; lub komórkom z  $m = 1, 2, 3, 4, 5, 6$  (liczba kopii genu *p53*) oraz ustalonej liczbie kopii genu *Mdm2*:  $n = 2$ .

Jak pokazujemy na Rys. 7.11, transfekcja *Mdm2* ( $m = 2, n = 4$ ) powoduje zmniejszenie zakresu parametru, dla którego układ oscyluje, podczas gdy transfekcja *p53* ( $m = 4, n = 2$ ) zwiększa ten zakres. W kotransfekowanych komórkach mających zduplikowaną liczbę kopii *p53* i *Mdm2* ( $m = n = 4$ ) obszar oscylacji jest mniejszy niż w normalnych diploidalnych komórkach, ale większy niż w komórkach tylko z duplikowaną liczbą kopii *Mdm2*.



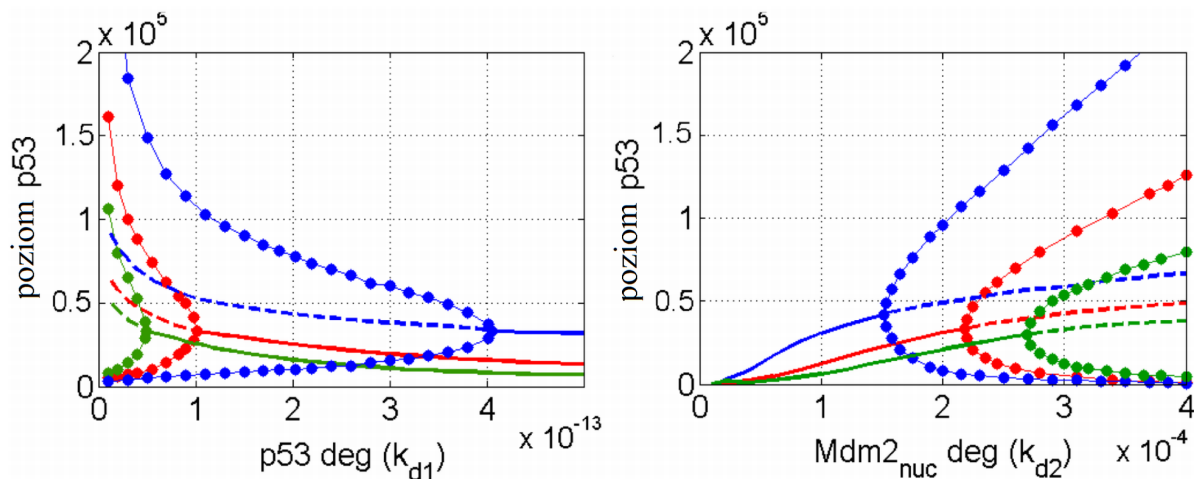


Rysunek 7.11: Diagramy bifurkacji dla normalnych oraz transfekowanych komórek dla modelu testowego. Punkty stałe oraz cykle graniczne wyznaczone są jako funkcje współczynnika degradacji  $p53$  (lewa kolumna) oraz  $Mdm2$  (prawa kolumna); czerwony: normalne komórki ( $m = n = 2$ ), niebieski: komórki z transfekcją  $Mdm2$  ( $m = 2, n = 4$ ) pomarańczowy: komórki z transfekcją  $p53$  ( $m = 4, n = 2$ ) oraz zielony: komórki z kotransfekcją  $Mdm2$  i  $p53$  ( $m = 4, n = 4$ ). Linie kropkowane reprezentują dolne i górne granice stabilnych cykli granicznych, linie ciągłe reprezentują stabilne stany stałe.

Jak pokazano na Rys. 7.12, brak jednej z dwóch kopii genu  $p53$  powoduje, że oscylacyjny obszar jest mniejszy, podczas gdy brak jednej kopii  $Mdm2$  powoduje, że obszar oscylacyjny jest większy.

Nasz model implikuje, że oscylacje mogą być generowane w wyniku transfekcji  $p53$ , lub tłumione (w komórkach z uszkodzonym DNA) w odpowiedzi na transfekcję  $Mdm2$ . Podobnie, brak jednej kopii genu  $Mdm2$  może generować oscylacje nawet gdy DNA nie jest uszkodzone, oraz brak kopii genu  $p53$  może uniemożliwić oscylacje pomimo, że DNA jest uszkodzone. Te przewidywania modelu są zgodne z wynikami eksperymentalnymi. Ponieważ oscylacje  $p53$  i  $Mdm2$  są konieczne, aby zainicjować transkrypcję genów zależnych od  $p53$  zaangażowanych w zatrzymanie cyklu komórkowego, naprawę DNA lub apoptozę, model może tłumaczyć  $p53$ -haploinsufficiency prowadzące do rozwoju nowotworu. Natomiast brak jednej kopii

genu *Mdm2* powoduje, że komórki są bardziej wrażliwe na uszkodzenie DNA i zatrzymuje nowotwór, jak zaobserwowano przez Alt i in. [52].



Rysunek 7.12: Diagramy bifurkacji dla normalnych komórek i komórek ze zmniejszoną ilością kopii genu dla modelu testowego. Punkty stałe oraz cykle graniczne rysowane są jako funkcje współczynnika degradacji *p53* (Panel lewy) oraz *Mdm2* (Panel prawy); czerwony: normalne komórki ( $m = n = 2$ ), niebieski: komórki z jedną kopią *Mdm2* ( $m = 2, n = 1$ ) oraz zielony: komórki z jedną kopią *p53* ( $m = 1, n = 2$ ). Linie kropkowane reprezentują dolne i górne granice stabilnych cykli granicznych, linie ciągłe reprezentują stabilne stały stałe.

W porównaniu z modelem Ciliberto i in. [47] nasz model testowy generuje oscylacje o okresie mniej wrażliwym na zmianę liczby kopii genu *p53* i *Mdm2* oraz na parametr degradacji *Mdm2*. Pokazuje to, że typ bifurkacji jest bardziej istotnym wyznacznikiem dynamiki układu niż istnienie w układzie dodatniego sprzężenia zwrotnego.

Zarówno model Ciliberto i in. [47] jak i nasz model testowy powinien być traktowany ostrożnie; obydwa modele opierają się na dodatnim sprzężeniu zwrotnym, w którym pośredniczy *PTEN*, który jest mało aktywny w linii komórkowej MCF-7 badanej przez Geva-Zatorsky i in. [36]. Ponadto, nawet w liniach komórkowych mających *PTEN* aktywowany

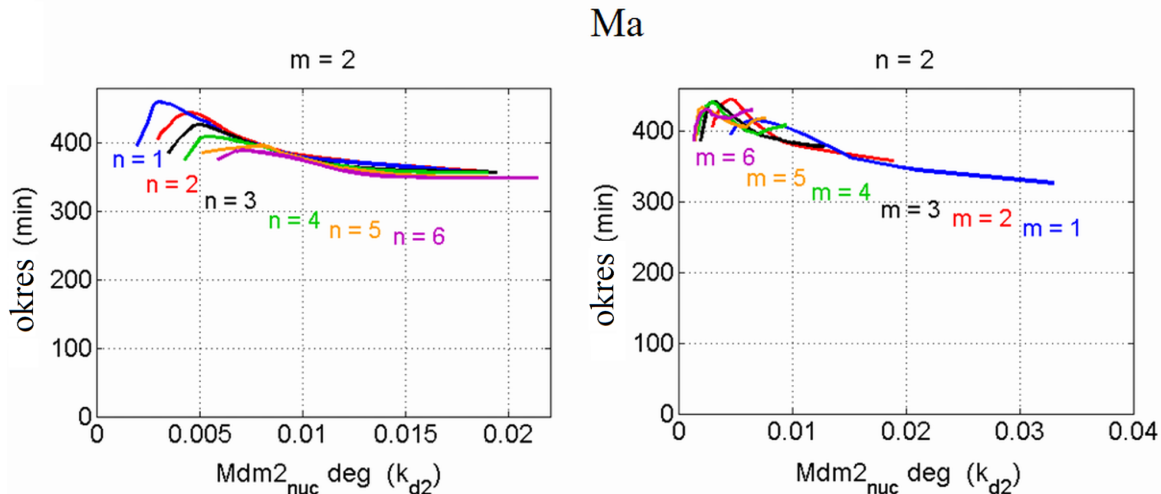
przez *p53* responsive *PTEN*, obecność form pośrednich w pętli dodatniego sprzężenia zwrotnego (tzn. *PTEN*, *PIP3* i *Akt*) wprowadza opóźnienie (pominięte w obydwu modelach), które może zasadniczo zmienić dynamikę układu, jak pokazaliśmy w pracy [49].

### 7.5.4 Analiza modeli opartych na ujemnym sprzężeniu zwrotnym i opóźnieniu

Przeanalizujemy trzy modele, w których oscylacje generowane są przez współdziałanie ujemnego sprzężenia zwrotnego łączącego *p53* z jego inhibitorem *Mdm2* oraz opóźnienia.

#### Model Ma i in.

W modelu Ma i in. [50] opóźnienie wprowadzone jest na poziomie transkrypcji i translacji *Mdm2*. W modelu tym, oscylacje rodzą się w punkcie nadkrytyczną bifurkacją Hopfa.

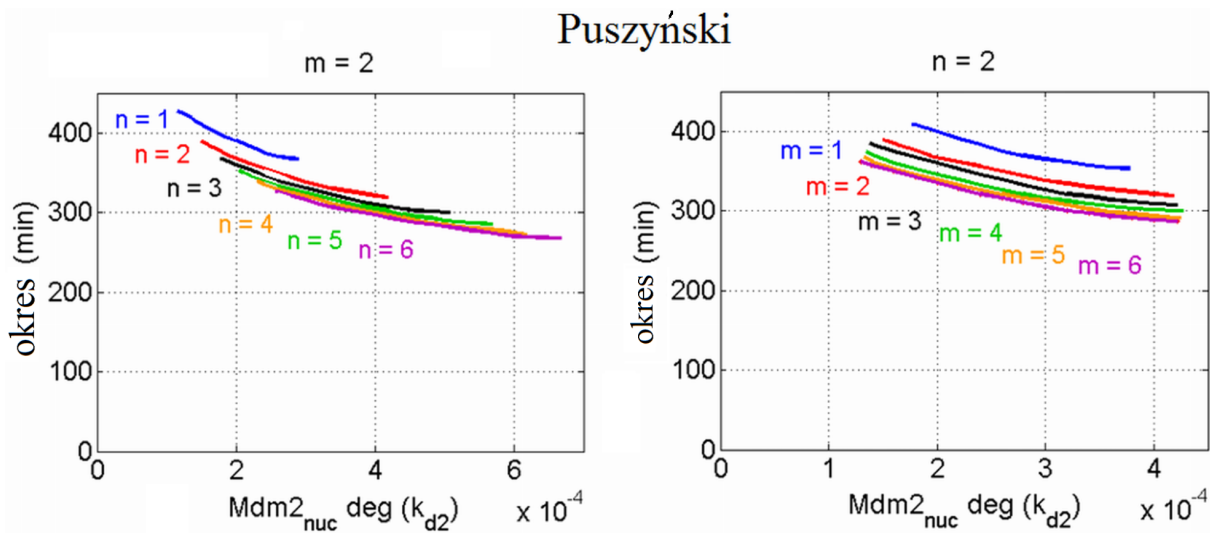


Rysunek 7.13: Analiza modelu Ma i in. [50]. Okres oscylacji dla stabilnego cyklu granicznego jako funkcja parametru degradacji *Mdm2* ( $k_{d2}$ ). Kolory: niebieski, czerwony, czarny, zielony, pomarańczowy i fioletowy odpowiadają komórkom ze zmienną liczbą kopii genu *Mdm2*:  $n = 1, 2, 3, 4, 5, 6$  i  $m = 2$  (lewa kolumna), oraz ze zmienną liczbą kopii *p53*:  $m = 1, 2, 3, 4, 5, 6$  i  $n = 2$  (prawa kolumna).

Jak pokazano na Rys. 7.13, okres oscylacji słabo zależy od współczynnika degradacji *Mdm2* (który kontrolowany jest przez uszkodzenie DNA) oraz od liczby kopii genów *p53* i *Mdm2*.

### Model Puszyńskiego i in.

Model Puszyńskiego i in. [49] opiera się na dwóch sprzężeniach zwrotnych: ujemnym i dodatnim. Jednakże, oscylacje generowane są wyłącznie przez ujemne sprzężenie oraz obecność form pośrednich (które wprowadzają opóźnienie) w pętli ujemnego sprzężenia zwrotnego, które zawiera *p53*, transkrypt *Mdm2*, cytoplazmatyczne *Mdm2*, cytoplazmatyczne sfosforylowane *Mdm2* i jądrowe sfosforylowane *Mdm2*, które kontroluje degradację *p53*.



Rysunek 7.14: Analiza modelu Puszyńskiego i in. [49]. Okres oscylacji dla stabilnego cyklu granicznego jako funkcja parametru degradacji *Mdm2* ( $k_{d2}$ ). Kolory: niebieski, czerwony, czarny, zielony, pomarańczowy i fioletowy odpowiadają komórkom ze zmienną liczbą kopii genu *Mdm2*:  $n = 1, 2, 3, 4, 5, 6$  i  $m = 2$  (lewa kolumna), oraz ze zmienną liczbą kopii *p53*:  $m = 1, 2, 3, 4, 5, 6$  i  $n = 2$  (prawa kolumna).

Dodatknie sprzężenie zwrotne zawiera jeszcze dłuższą od ujemnego sprzężenia pętlę oddziaływań zawierającą fosfatazę PTEN, PIP3 i Akt, i działa poprzez “podwójną negację”:

$p53 \rightarrow PTEN \dashv PIP3 \rightarrow Akt \rightarrow Mdm2 \dashv p53$ , blokując ujemne działanie  $p53$  i  $Mdm2$ . Uszkodzenie DNA prowadzi do destabilizacji  $Mdm2$  i fosforylacji  $p53$ , generując oscylacje poziomów  $p53$  i  $Mdm2$ , podczas których DNA może zostać naprawione. Jeżeli po około 15 godzinach naprawa DNA nie jest kompletna, sygnał przechodzący przez dodatnie sprzężenie zwrotne blokuje pętlę ujemnego sprzężenia i powoduje wzrost  $p53$  do poziomu, przy którym może być zainicjowana apoptoza.

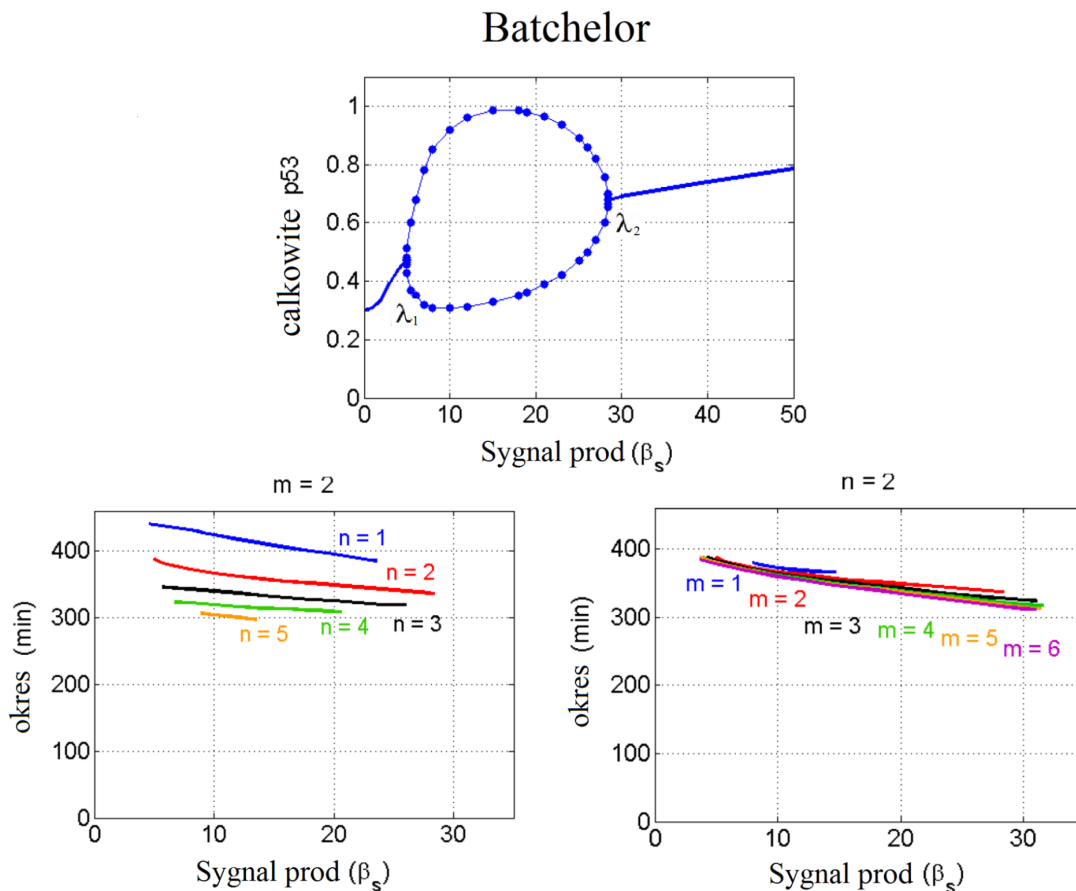
Z zablokowaną pętlą dodatniego sprzężenia zwrotnego oraz z wyłączoną naprawą DNA układ, w odpowiedzi na uszkodzenie DNA, generuje długotrwałe oscylacje. Aby przeanalizować model Puszyńskiego i in., wprowadzamy uszkodzenie DNA poprzez zmianę współczynnika degradacji  $Mdm2$  (pozostawiając niezmienny współczynnik fosforylacji  $p53$ ). Jak pokazano na Rys. 7.14, podobnie jak w modelu Ma i in. [50], oscylacje rodzą się w punkcie nadkrytyczną bifurkacją Hopfa oraz ich okres nie jest wrażliwy ani na degradację  $Mdm2$  ani na liczbę kopii genów  $p53$  i  $Mdm2$ .

### Model Batchelora i in.

Jako ostatni model układu  $p53$ , przeanalizujemy model zaproponowany przez grupę Galit Lahav, który opiera się na dwóch pętlach ujemnego sprzężenia zwrotnego [73]. Tak jak w poprzednich modelach, jedna pętla łączy  $p53$  oraz  $Mdm2$ , zaś druga pętla zawiera kinazy  $ATM$  i  $Chk2$ , które aktywują  $p53$ , oraz inhibitor  $Wip1$ . Pulsy poziomów  $p53$  oraz jego aktywność są ściśle związane z pulsami aktywnego  $ATM$ . Istnienie tej nowej pętli sprzężenia zwrotnego zostało udokumentowane przez serię eksperymentów wykonaną na linii komórek raka piersi MCF-7.

Model zawiera dwa opóźnienia bezpośrednio związane z kontrolowaną przez  $p53$  syntezą  $Mdm2$  oraz  $Wip1$ . Ponieważ w rozważanym modelu degradacja  $Mdm2$  kontrolowana jest przez  $Signal$  (reprezentujący aktywny  $ATM$  oraz  $Chk2$ ), który z kolei reguluje  $Mdm2$  poprzez  $p53$ . W naszej analizie jako parametr bifurkacyjny, zamiast współczynnika degradacji  $Mdm2$ , wybraliśmy współczynnik produkcji sygnału  $\beta_s$ . Parametr  $\beta_s$  kontroluje siłę sygnału w

komórkach z uszkodzonym DNA.



Rysunek 7.15: Analiza modelu Batchelora. Górny Panel: stabilne punkty stałe i cykle graniczne przedstawione są w funkcji współczynnika produkcji *sygnału*  $\beta_s$ ; układ zawiera dwa punkty bifurkacji: wraz ze wzrostem  $\beta_s$ , powstaje cykl graniczny w punkcie bifurkacji *nadH* ( $\lambda_1$ ), oraz znika w punkcie bifurkacji *nadH* ( $\lambda_2$ ). Okres oscylacji dla stabilnego cyklu granicznego przedstawiony jest jako funkcja  $\beta_s$ . Kolory: niebieski, czerwony, czarny, zielony, pomarańczowy i fioletowy odpowiadają komórkom z  $n = 1, 2, 3, 4, 5, 6$  (liczba kopii *Mdm2*) oraz ustaloną liczbą kopii genu *p53*  $m = 2$  (Panel prawy), lub komórkom z  $m = 1, 2, 3, 4, 5, 6$  (liczba kopii *p53*) oraz ustaloną liczbą kopii genu *Mdm2*  $n = 2$  (Panel prawy).

W modelu tym, oscylacje cyklu granicznego istnieją dla skończonego zakresu współczynnika produkcji *sygnału*  $\beta_s$ , oraz rodzą się i znikają w dwóch punktach bifurkacji *nadH*  $\lambda_1$  i  $\lambda_2$

(Rys. 7.15). Okres oscylacji słabo zależy od współczynnika produkcji sygnatu  $\beta_s$ , ponadto dla  $m, n \in \{2, 3, 4\}$  zakres  $\beta_s$  prawie nie zależy od  $m$  i  $n$ . Jedynie dla  $m = 1$  i  $n = 2$  zakres oscylacji jest znacznie mniejszy niż zakres oscylacji dla normalnych komórek ( $m = n = 2$ ). Oscylacje nie istnieją dla  $m = 2$  oraz  $n > 5$ , co implikuje, że trzykrotna nadekspresja *Mdm2* może zatrzymać oscylacje. Jednakże, efekt ten może zostać zrekompensowany przez transfekcję *p53*; komórki z  $m = n \in \{2, 3, 4, 5, 6\}$  oscyluje dla prawie tego samego zakresu parametru co komórki normalne.

## 7.6 Analiza modeli układu regulatorowego $NF-\kappa B$

W ostatnich latach rozwinięto kilka modeli sygnałowego układu regulatorowego  $NF-\kappa B$ , omawianych w [54] i [55]. Pierwszym był model z pętlą jednego sprzężenia zwrotnego, który zkoncentrował się na oddziaływaniu pomiędzy trzema izoformami  $I\kappa B$  [45]. Model ten został rozwinięty przez grupy Hoffmann'a i Levchenki włączając model Werner'a i in. [56]. Następnym był nasz model z dwiema pętlami sprzężenia zwrotnego [44], zawierający działanie zarówno inhibitorów  $NF-\kappa B$ ,  $I\kappa B\alpha$  jak i A20, które były dalej rozwijane, aby zawierały stochastyczność w członie aktywacji/deaktywacji genów  $I\kappa B\alpha$  i A20 [57]. Trwałe oscylacje jądrowo-cytoplazmatycznej lokalizacji  $NF-\kappa B$  w odpowiedzi na stymulację  $TNF\alpha$  zostały zaobserwowane przez Nelson'a i in. [37] w eksperymentach na pojedynczych komórkach SK-N-AS transfekowanych plazmidami powodującymi ekspresję RelA połączonego z czerwonym fluorescencyjnym białkiem (RelA-DsRed). Zauważono, że 91% tych komórek wykazywało długotrwałe oscylacje jądro-cytoplazma białka RelA i że oscylacje te są całkiem dobrze zsynchronizowane pomiędzy komórkami w przeciągu kilku pierwszych cykli. Potem jednak oscylacje te wypadają z fazy i stają się gasnące gdy obserwuje się je na poziomie populacji we wcześniejszych eksperymentach wykonanych przez Hoffmann i in. [45]. Sun i in. [46] zademonstrowali, że ponad 4000 cząsteczek mRNA oscylowało w odpowiedzi na stymulację  $TNF\alpha$ . We wszystkich trzech deterministycznych modelach [34, 44, 56], oscylacje, które pow-

stają w odpowiedzi na stymulację  $\text{TNF}\alpha$  są gasnące (nie ma tam żadnych cykli granicznych) dla zakładanych (dopasowanych) zbiorów parametrów. Jedynie w stochastycznym modelu [57], oscylacje są ciągłe z powodu stochastycznego rezonansu.

Matematyczne struktury tych modeli mają kilka ważnych różnic, jednakże ze względu na naszą analizę główna różnica znajduje się w formie transkrypcyjnej regulacji  $I\kappa B\alpha$ ; Oryginalny model Hoffmann-Levchenko [45] zakładał, że współczynnik transkrypcji jest proporcjonalny do kwadratu koncentracji jądrowego  $NF-\kappa B$ , podczas gdy w modelu Werner'a i in. [56], i w następnych modelach, zakładany jest kubiczny współczynnik transkrypcji. Przeciwnie, w modelu [44] współczynnik transkrypcji jest liniowy. W późniejszym stochastycznym modelu [57], aktywność genów  $I\kappa B\alpha$  i A20 regulowana jest przez przyłączenie i dysocjację dimerów  $NF-\kappa B$ , i zakłada się, że aktywny gen transkrybowany jest ze stałym współczynnikiem równym 4.5 mRNA/sekundę/kopię genu, bliskim fizjologicznego maksimum.

Jak pokazano w Barken'a i in. [58], model Hoffmann'a-Levchenko [45] generuje oscylacje o okresie bardzo wrażliwym na ilość całkowitego  $NF-\kappa B$ . Prowadzi to do konkluzji, że oscylacje zarejestrowane w układzie z nadekspresją przez Nelson i in. [37] nie pozwalają wywnioskować, czy oscylacje o tej samej trwałości, amplitudzie i okresie pojawiają się u morfogenetycznie niezmiennych komórek [58]. W rzeczywistości, zgodnie z modelem Hoffmann'a-Levchenko [45] czterokrotna nadekspresja  $NF-\kappa B$  powoduje, że drugi pik w poziomie jądrowego  $NF-\kappa B$  pojawia się nie w 2 h (jak normalnie), ale w 5.5 h ciągłej stymulacji  $\text{TNF}\alpha$  (w pracy [58] Rys. 1A). Jednakże, jak pokazano przez Nelson'a i in. [59], okres oscylacji jest prawie niewrażliwy na ekspresję  $NF-\kappa B$  i wystarczy zastąpić kwadratowy współczynnik transkrypcji przez liniowy aby zasadniczo zminimalizować wrażliwość długości okresu na całkowity poziom  $NF-\kappa B$  (w pracy [59] Rys. 2). Następnie, pokazaliśmy, że założenie, że współczynnik transkrypcji jest przyłączony, prowadzi do oscylacji o okresie prawie niewrażliwym na całkowity poziom  $NF-\kappa B$  ([57], Rys. 4.6).

Modele, w których współczynnik transkrypcji jest proporcjonalny do koncentracji czynnika transkrypcyjnego (zwłaszcza w wyższej potędze), bez jakiegokolwiek nasycenia, prowadzą



do transkrypcyjnych wydajności przekraczających fizjologiczne współczynniki i w ten sposób może niewłaściwie naśladować dynamikę komórki. W tym szczególnym przypadku, niewłaściwy wybór kwadratowego współczynnika transkrypcji prowadzi do wysokiej wrażliwości przewidywanego okresu całkowitego poziomu  $NF-\kappa B$ . Na wysokim poziomie czynnika transkrypcyjnego, współczynnik transkrypcji  $I\kappa B\alpha$ , inhibitora  $NF-\kappa B$ , przekracza wiele razy maksimum fizjologicznego współczynnika (patrz [60]), i powoduje, że  $I\kappa B\alpha$  kumuluje się na bardzo wysokim poziomie inhibując  $NF-\kappa B$  na wiele godzin, co wydłuża okres oscylacji. Dodatkowo, dłuższy czas degradacji akumulowanego  $I\kappa B\alpha$  wprowadza opóźnienie, które powoduje, że pojawia się stabilny cykl graniczny.

# Rozdział 8

## Konkluzje

W pierwszej części pracy (Rozdziały 3, 4, 5) przeanalizowaliśmy, poprzez stochastyczne modelowanie ekspresji genów, potencjalne konsekwencje zmiany liczby kopii genów na ich ekspresję. Analiza tego problemu została zapoczątkowana w pracy Cook'a i in. [5] na temat wpływu stochastycznej ekspresji genów na haploinsufficiency. Rozwinięcie modelowania stochastycznej ekspresji genów [10], [30], [25], [26], [15] pozwoliło nam zastosować bardziej rygorystyczne traktowanie tego problemu poprzez średnie funkcji gęstości prawdopodobieństwa białka. Rozważyliśmy ekspresję haploidalnych, diploidalnych i tetraploidalnych genów, najpierw bez autoregulacji, a następnie z dodatnią autoregulacją. W przypadku bez autoregulacji aktywność każdej kopii genu w  $N$ -ploidalnym genie jest niezależna i w ten sposób średnia i wariancja rozkładu białka jest proporcjonalna do liczby kopii genów pod warunkiem, że ekspresja pojedynczej kopii jest taka sama. Kiedy ekspresja białka na kopię jest zmodyfikowana, tak że średnia ekspresja  $N$ -ploidalnego genu jest równa ekspresji haploidalnego genu, wtedy (w ciągłym przybliżeniu, w którym transkrypcyjny i translacyjny szum jest zaniedbany) wariancja jest odwrotnie proporcjonalna do  $N$ . To wyjaśnia numeryczny wynik Cook'a i in. [5] gdzie, przeprowadzając symulacje Monte Carlo, otrzymano, że wariancja rozkładu białka maleje ze wzrostem liczby kopii (pod warunkiem, że średnia ekspresja każdej kopii jest stała).

W tej części pracy ograniczamy się do układów autoregulowanych, w których zależność ekspresji od liczby kopii genu jest znacznie bardziej skomplikowana. W tym przypadku ani poziom białka ani wariancja rozkładu białka nie są proporcjonalne do liczby kopii genu. Układy te mają dwie cechy, które powodują, że są wrażliwe na zmianę liczby kopii genów. Po pierwsze, wyższy poziom białka implikuje silniejsze sprzężenie, po drugie wyższa liczba

kopii genów zwiększa prawdopodobieństwo aktywacji genu.

Rozważając zewnętrzną indukcję haploidalnych, diploidalnych i tetraploidalnych genów z dodatnim sprzężeniem zwrotnym znaleźliśmy, że w każdym przypadku istnieją trzy charakterystyczne wzorce aktywacji:  $\{A\}$  - gdy aktywność genu jest podtrzymywana bez zewnętrznej stymulacji,  $\{U\}$  - gdy gen aktywuje się z unimodalnym profilem rozkładu białka, oraz  $\{B\}$  - gdy gen aktywuje się poprzez rozkład bimodalny. We wzorcach aktywacji  $\{U\}$  i  $\{B\}$  gen odgrywa rolę przełącznika, a ponadto we wzorcu  $\{B\}$  stochastyczność przejściowo dzieli populację komórek na dwie podpopulacje: jedną z wysoką a drugą z niską ekspresją białka. W pewnych przypadkach może być to najbezpieczniejsza strategia dla tkanki bądź kolonii bakterii. Nasza analiza pokazuje, że diploidalny gen ma najbardziej regularny wzorec aktywacji, tzn., sektory  $\{A\}, \{B\}, \{U\}$  w płaszczyźnie parametrów  $(c_2, b_0)$  są podobnej wielkości.

Główny wynik prezentowany jest na Rys. 5.11: rozważając równoległą aktywację haploidalnego, diploidalnego lub tetraploidalnego genu znaleźliśmy, że istnieje dziewięć wzorców aktywacji. Analiza tych wzorców pokazuje potencjalne konsekwencje straty oraz duplikacji kopii genu. Gdy układ znajduje się w jednym z dwóch wzorców  $\{AAU\}$  lub  $\{AAB\}$  wtedy strata kopii genu może zatrzymać ciągłą aktywność genu i prowadzić do choroby jeżeli wymagany jest stały poziom białka syntetyzowanego przez ten gen. Jak pokazano na Rys. 5.12 w przypadku  $\{AAB\}$  strata kopii genu prowadzi nie tylko do niższego poziomu białka, ale także do znacznie wyższej wariacji co implikuje przejściowy spadek poziomu białka do bardzo niskiej wartości. W przypadku gdy autoregulowany gen jest czynnikiem transkrypcyjnym regulującym transkrypcję innych genów, spadek jego poziomu propagowałby z kolei spadek poziomu innych białek co mogłoby prowadzić do choroby. Wśród ponad jego 500 czynników transkrypcyjnych kodowanych w ludzkim genomie, istnieje ponad 30, które mogą powodować chorobę poprzez haploinsufficiency [23].

Nasza analiza pokazuje, że duplikacja genu także może wpłynąć na jego funkcjonowanie. Na przykład, gdy układ znajduje się w modzie  $\{AUU\}$  lub  $\{ABU\}$  duplikacja genu z diploidalnego na tetraploidalny powodowałaby ciągłą aktywność. Podobnie, gdy układ znajduje się w

modzie  $\{AAB\}$  lub  $\{AAU\}$  duplikacja haploidalnego genu także powodowałaby jego ciągłą aktywność. W przypadku, gdy haploidalny lub diploidalny gen działa jako "przełącznik" jego duplikacja może prowadzić do choroby. W rzeczywistości, regulacja poprzez dodatnie sprzężenie zwrotne jest częstym elementem przełączania. Bazując na Rys. 5.11 możemy oczekiwać, że czasowa duplikacja genomu w fazie S cyklu komórkowego może powodować przeniesienie genów regulowanych przez dodatnie sprzężenie zwrotne do aktywnego stanu. To implikuje różną ekspresję genu w fazie G1 (poprzedzającej S) i G2 (następnej po S) cyklu komórkowego.

Na koniec pokazaliśmy, że strata jednej kopii genu może nie być zrekompensowana przez podwyższenie ekspresji pozostałej kopii. Nawet jeżeli haploidalny gen ma dwa razy większy współczynnik syntezy na kopię niż gen diploidalny, to nadal wymaga on silniejszego sygnału do aktywacji. Ma to ważne konsekwencje dla modelowania: aby właściwie opisać regulację genu diploidalnego należy wziąć pod uwagę fakt, że obie kopie genu mogą być aktywowane niezależnie. W przypadku przedstawionym na Rys. 5.16 do aktywacji genu haploidalnego potrzebny jest 10 razy silniejszy sygnał niż do aktywacji genu diploidalnego. Co więcej, aktywacja genu diploidalnego przebiega poprzez rozkład bimodalny, dlatego ma on przejściowo wyższą wariancję.

W drugiej części pracy (Rozdział 6) przeanalizowaliśmy wpływ liczby kopii genów kodujących poszczególne kinazy tworzące kaskadę sygnałową w ścieżce sygnałowej tymocytów, na dynamikę tej kaskady oraz aktywność tymocytu. Aktywacja ścieżki sygnałowej tymocytów generowana jest przez przyłączenie do kompleksu  $TCR$  peptydów antygeny i mierzona jest poziomem kinazy  $ERK_{pp}(t)/ERK_{total}$  w pojedynczej komórce tymocytu.  $ERK_{pp}$  reguluje pętlę dodatniego sprzężenia zwrotnego, zawierającego kaskadę sygnałową kinaz  $ZAP$ ,  $MEK$ ,  $ERK$ , oraz inhibuje pętlę ujemnego sprzężenia zwrotnego zawierającego fosfatę  $SHP_p$ . Naszym celem była analiza wpływu liczby kopii genów kodujących poszczególne kinazy tworzące kaskadę sygnałową na dynamikę tej kaskady oraz aktywność tymocytu. Współdziałanie ujemnego i dodatniego sprzężenia zwrotnego powoduje współistnienie dwóch

stanów stabilnych. Jak pokazano na Rys. 6.5, dla małej liczby peptydów układ osiąga stan stabilny odpowiadający nieaktywnej komórce, dla dużej liczby peptydów układ zbiega asymptotycznie do stanu stabilnego odpowiadającego komórce w stanie aktywnym. Asymptotyczne zachowanie trajektorii układu wyznaczone jest przez warunki początkowe. Rysunki 6.5, 6.6 i 6.7 przedstawiają komórkę, która w chwili początkowej znajduje się w stanie z niskim poziomem  $SHP_p$  (500 cząsteczek) oraz niskim poziomem  $ERK_{pp}$  (670 cząsteczek). Na przykład po przyłączeniu do tymocytu 30 peptydów, ze stałą przyłączenia  $d_1 = 0.05/s$ , układ zbiega do aktywnego stanu z wysokim poziomem  $ERK_{pp}(t)/ERK_{total}$ . Jeżeli natomiast początkowo komórka znajduje się w stanie z wysokim poziomem  $SHP_p(t)$ , wówczas pętla ujemnego sprzężenia jest silniejsza, komórka pozostaje nieaktywna i zbiega asymptotycznie do stanu z niskim poziomem  $ERK_{pp}(t)/ERK_{total}$ .

Aktywacja  $TCR$  jest wrażliwa na ilość przyłączonych peptydów antygeny oraz wielkość współczynnika dysocjacji antygeny. Jak pokazaliśmy na Rys. 6.5, aktywacja  $TCR$  jest bardziej wrażliwa na przyłączenie małej liczby peptydów antygeny, niż na przyłączenie dużej liczby peptydów, przy której poziom  $ERK_{pp}$  osiąga stan nasycenia. Ponadto, im jest większy współczynnik dysocjacji antygeny, tym większa liczba peptydów wymagana jest do aktywacji tymocytu. Ponieważ  $ERK_{pp}$  inhibuje pętlę ujemnego sprzężenia zwrotnego zawierającego  $SHP_p$ , zwiększenie poziomu  $ERK_{pp}$  powoduje znaczne obniżenie poziomu fosfatazy  $SHP_p$  (Rys. 6.7).

Jak pokazano na Rys. 6.7, poziom  $SHP_p$  słabo zależy od liczby kopii genów  $ZAP$ ,  $MEK$ ,  $ERK$ . Zwiększenie liczby kopii genu  $ERK$ ,  $ZAP$  i  $MEK$  powoduje zwiększenie poziomu  $ERK_{pp}$ . Dla małego współczynnika dysocjacji brak jednej kopii genu  $ZAP$  znacznie bardziej obniża poziom  $ERK_{pp}$  niż brak jednej kopii genu  $ERK$ , zaś zwiększenie liczby kopii genu  $ZAP$  znacznie bardziej podwyższa poziom  $ERK_{pp}(t)/ERK_{total}$  niż zwiększenie liczby kopii genu  $ERK$ . Analizowana w modelu kaskada sygnałowa zawiera pojedynczo sfosforylowaną kinazę  $ZAP$ ,  $MEK$  i  $ERK$ , oraz podwójnie sfosforylowane kinazy  $MEK$  i  $ERK$ . Taka postać kaskady sygnałowej powoduje, że dla małej wartości *Sygnалу* funkcja  $ERK_{pp}$  jest silnie nieliniowa,

podczas gdy dla prostszej postaci kaskady, z nie uwzględnioną formą pojedynczej fosforylacji, funkcja  $ERK_{pp}$  liniowo zależy od *Sygnatu*.

W trzeciej części pracy (Rozdział 7) zaproponowaliśmy nową metodę walidacji modeli ścieżek regulatorowych, w której dynamika układu dynamicznego badana jest numerycznie ze względu na liczbę kopii genu(ów). Opierając się na tym przybliżeniu zademonstrowaliśmy, że wszystkie modele układu *p53*, które opierają się na nadkrytycznej bifurkacji Hopfa, generują oscylacje o okresie odpornym na zmianę parametru degradacji *Mdm2* oraz zmianę liczby kopii genu, w przeciwieństwie do modeli, które opierają się na bifurkacjach *SNIC* lub podkrytycznej Hopfa.

Rozważając cztery modele, w których oscylacje rodzą się w punkcie nadkrytyczną bifurkacją Hopfa pokazaliśmy, że zakres oscylacji w obszarze parametru jest rosnącą funkcją liczby kopii genu *p53* oraz malejącą funkcją liczby kopii genów *Mdm2*. Zależność zakresu oscylacji od liczby kopii genu jest silniejsza w naszym modelu testowym, który opiera się na współdziałających pętlach dodatniego i ujemnego sprzężenia zwrotnego, niż w modelach Ma i in. [50] oraz Puszynski i in. [49], w których oscylacje generowane są wyłącznie przez ujemne sprzężenie i opóźnienie. Zgodnie z modelem testowym transfekcja *Mdm2* może zatrzymać oscylacje nawet w komórkach z uszkodzonym DNA, podczas gdy transfekcja *p53* może wzmacniać oscylacje nawet gdy DNA nie jest uszkodzone. Ponieważ oscylacje w układzie *p53|Mdm2* mogą być potencjalnie zwizualizowane w *p53*- lub *Mdm2*-transfekowanych komórkach, przewidywanie to może być wykorzystane do eksperymentalnej weryfikacji modelu testowego. Jeżeli model okazałby się poprawny, sugerowałoby to, że oscylacje obserwowane w komórkach z transfekcją *p53* są częściowo artefaktem układu eksperymentalnego. Jak zbadano przez Geva-Zatorsky i in. [36], 19 z 48 *p53*-transfekowanych komórek wykazywało oscylacje bez jakiegokolwiek wcześniejszego promieniowania gamma, podczas gdy silne promieniowanie dawką 10 Gy powodowało, że 37 z 54 komórek oscylowało. To może sugerować, że zarówno transfekcja *p53* jak i promieniowanie gamma indukuje oscylacje.

Inną możliwością jest, że dodatnie sprzężenie nie odgrywa żadnej roli w utrzymywaniu oscylacji  $p53$ , które powstają z współdziałania ujemnego sprzężenia zwrotnego i opóźnienia. Takie oscylacje są prawie niewrażliwe na liczbę kopii genu  $Mdm2$  i  $p53$ . Jak zaproponowano przez Puszyńskiego i in. [49], dodatnie sprzężenie związane z długim opóźnieniem, może być odpowiedzialne za przerwanie oscylacji poprzez syntezę  $p53$  do nawet wyższego poziomu, który umożliwia apoptotyczną odpowiedź w poważnie uszkodzonych komórkach.

Ze względu na układ  $NF-\kappa B$ , w którym wszystkie deterministyczne modele generują tylko gasnące oscylacje, pokazaliśmy, że nasycenie we współczynniku transkrypcji jest kluczowe, aby wyprodukować oscylacje o odpornym okresie. Modele, w których współczynnik transkrypcji jest proporcjonalny do koncentracji czynnika transkrypcyjnego w drugiej lub trzeciej potęgze, bez nasycenia, wydajność transkrypcji przekracza fizjologiczne współczynniki i w ten sposób może niewłaściwie naśladować dynamikę komórki. Modele te przewidują, że okres oscylacji gwałtownie rośnie wraz z poziomem transfekcji, co nie jest obserwowane w eksperymencie.

Eksperymenty na pojedynczych komórkach stają się wiodącą techniką wizualizującą zachowanie poszczególnych komórek, które jest słabo widoczne poprzez uśrednianie po populacji [61]. Eksperymenty te są wykonywane na komórkach z podwyższoną liczbą kopii genu i dlatego muszą być analizowane ze szczególną uwagą. Jak zademonstrowali Nelson i in. [37], [59] w układzie regulatorowym  $NF-\kappa B$ , oraz Geva-Zatorsky i in. [36] w układzie regulatorowym  $p53$ , oscylacje generowane w odpowiedzi na stres mają okres odporny zarówno na liczbę kopii genu czynnika transkrypcyjnego jak i na poziom stresu. Obserwacja ta stawia ograniczenie na strukturę układów regulatorowych, która jest odpowiedzialna za zwizualizowane oscylacje. Analizując modele układu  $p53$  pokazaliśmy, że oscylacje o odpornym okresie produkowane są w wyniku nadkrytycznej bifurkacji Hopfa, charakterystycznej dla układów opierających się na ujemnym sprzężeniu zwrotnym i opóźnieniu, a nie wyniku bifurkacji  $SNIC$  lub podkrytycznej bifurkacji Hopfa, które są charakterystyczne dla układów opierających się na współdziałającym ujemnym i dodatnim sprzężeniu zwrotnym. Jednakże

typ bifurkacji jest bardziej znaczący niż istnienie dodatnich sprzężeń zwrotnych; model "testowy" zawierający pętlę dodatniego i ujemnego sprzężenia zwrotnego, lecz oparty na nadkrytycznej bifurkacji Hopfa ma zadowalającą odporność.



# Dodatek A

## Momenty rozkładów dla mRNA i białka

Momenty rzędu pierwszego i drugiego rozkładów mRNA i białka wyznaczymy metodą funkcji tworzących momenty (FTM). Momenty liczymy dla układów bez autoregulacji.

### Model dyskretny, $N=1$

Niech dane będą funkcje FTM dla mRNA ( $x$ ) i białka ( $y$ ) o rozkładzie dyskretnym  $g_{xy}$ ,  $h_{xy}$ :

$$F(z, s) = \sum_{x,y} z^x s^y g_{xy} \quad (86)$$

$$T(z, s) = \sum_{x,y} z^x s^y h_{xy}. \quad (87)$$

Zauważmy, że  $F(z, s)|_{z=s=1}$  oraz  $T(z, s)|_{z=s=1}$  są prawdopodobieństwami, że gen jest, odpowiednio, aktywny oraz nieaktywny:

$$F(z, s)|_{z=s=1} = \sum_{x,y} g_{xy} = \frac{b_0}{c_0 + b_0} \quad (88)$$

$$T(z, s)|_{z=s=1} = \sum_{x,y} h_{xy} = \frac{c_0}{c_0 + b_0}. \quad (89)$$

Zróżniczkujemy FTM (86)-(87) po  $z$  i  $s$  :

$$\begin{aligned}\frac{\partial F(z, s)}{\partial z} &= \sum_{x,y} xz^{x-1}s^y g_{xy} \\ \frac{\partial F(z, s)}{\partial s} &= \sum_{x,y} yz^x s^{y-1} g_{xy} \\ \frac{\partial T(z, s)}{\partial z} &= \sum_{x,y} xz^{x-1}s^y h_{xy} \\ \frac{\partial T(z, s)}{\partial s} &= \sum_{x,y} yz^x s^{y-1} h_{xy}.\end{aligned}\tag{90}$$

Układ równań Mistrzów (24)-(25) pomnóżmy stronami przez  $z^x s^y$ , a następnie zsumujmy stronami po  $x$  i  $y$ , otrzymamy wówczas

$$\begin{aligned}\frac{d}{dt} \left( \sum_{x,y} z^x s^y g_{xy} \right) &= b_0 \left( \sum_{x,y} z^x s^y h_{xy} \right) - c_0 \left( \sum_{x,y} z^x s^y g_{xy} \right) + r_1 \left( \sum_{x,y} (x+1) z^x s^y g_{(x+1,y)} \right) + \\ &- r_1 \left( \sum_{x,y} x z^x s^y g_{xy} \right) + K \left( \sum_{x,y} x z^x s^y g_{(x,y-1)} \right) + r_2 \left( \sum_{x,y} (y+1) z^x s^y g_{(x,y+1)} \right) \\ &- K \left( \sum_{x,y} x z^x s^y g_{xy} \right) - r_2 \left( \sum_{x,y} y z^x s^y g_{xy} \right),\end{aligned}\tag{91}$$

$$\begin{aligned}\frac{d}{dt} \left( \sum_{x,y} z^x s^y h_{xy} \right) &= -b_0 \left( \sum_{x,y} z^x s^y h_{xy} \right) + c_0 \left( \sum_{x,y} z^x s^y g_{xy} \right) + H \left( \sum_{x,y} z^x s^y h_{(x-1,y)} \right) + \\ &+ r_1 \left( \sum_{x,y} (x+1) z^x s^y h_{(x+1,y)} \right) - H \left( \sum_{x,y} z^x s^y h_{xy} \right) - r_1 \left( \sum_{x,y} x z^x s^y h_{xy} \right) + \\ &+ K \left( \sum_{x,y} x z^x s^y h_{(x,y-1)} \right) + r_2 \left( \sum_{x,y} (y+1) z^x s^y h_{(x,y+1)} \right) + \\ &- K \left( \sum_{x,y} x z^x s^y h_{xy} \right) - r_2 \left( \sum_{x,y} y z^x s^y h_{xy} \right).\end{aligned}\tag{92}$$

Zapiszmy odpowiednie człony w równaniach (91)-(92) za pomocą związków (90):

$$\begin{aligned}
\sum_{x,y=0}^{\infty} (x+1)z^x s^y g_{(x+1,y)} &= \sum_{x,y=0}^{\infty} xz^{x-1} s^y g_{xy} = \frac{\partial F(z, s)}{\partial z} \\
\sum_{x,y=0}^{\infty} (x+1)z^x s^y h_{(x+1,y)} &= \frac{\partial T(z, s)}{\partial z} \\
\sum_{x,y=0}^{\infty} (y+1)z^x s^y g_{(x,y+1)} &= \frac{\partial F(z, s)}{\partial s} \\
\sum_{x,y=0}^{\infty} (y+1)z^x s^y h_{(x,y+1)} &= \frac{\partial T(z, s)}{\partial s} \\
\sum_{x,y=0}^{\infty} xz^x s^y g_{xy} &= z \sum_{x,y=0}^{\infty} xz^{x-1} s^y g_{xy} = z \frac{\partial F(z, s)}{\partial z} \\
\sum_{x,y=0}^{\infty} xz^x s^y h_{xy} &= z \frac{\partial T(z, s)}{\partial z} \\
\sum_{x,y=0}^{\infty} yz^x s^y g_{xy} &= s \sum_{x,y=0}^{\infty} yz^x s^{y-1} g_{xy} = s \frac{\partial F(z, s)}{\partial s} \\
\sum_{x,y=0}^{\infty} yz^x s^y h_{xy} &= s \sum_{x,y=0}^{\infty} yz^x s^{y-1} h_{xy} = s \frac{\partial T(z, s)}{\partial s} \\
\sum_{x,y=0}^{\infty} xz^x s^y g_{(x,y-1)} &= \sum_{x,y=0}^{\infty} xz^x s^{y+1} g_{xy} = sz \sum_{x,y=0}^{\infty} xz^{x-1} s^y g_{xy} = sz \frac{\partial F(z, s)}{\partial z} \\
\sum_{x,y=0}^{\infty} xz^x s^y h_{(x,y-1)} &= sz \frac{\partial T(z, s)}{\partial z} \\
\sum_{x,y=0}^{\infty} z^x s^y h_{(x-1,y)} &= \sum_{x,y=0}^{\infty} z^{x+1} s^y h_{xy} = z \sum_{x,y=0}^{\infty} z^x s^y h_{xy} = zT(z, s)
\end{aligned} \tag{93}$$

Po podstawieniu wyrażeń (93) do (91) i (92) otrzymamy następujące równania różniczkowe

cząstkowe na FTM:

$$\begin{aligned}
\frac{\partial F(z, s)}{\partial t} &= b_0 T(s, z) - c_0 F(z, s) + r_1 \frac{\partial F(z, s)}{\partial z} - r_1 z \frac{\partial F(z, s)}{\partial z} + \\
&\quad + K z s \frac{\partial F(z, s)}{\partial z} + r_2 \frac{\partial F(z, s)}{\partial s} - K z \frac{\partial F(z, s)}{\partial z} - r_2 s \frac{\partial F(z, s)}{\partial s}, \\
\frac{\partial T(z, s)}{\partial t} &= -b_0 T(s, z) + c_0 F(z, s) + H z T(z, s) + r_1 \frac{\partial T(z, s)}{\partial z} + \\
&\quad - H T(z, s) - r_1 z \frac{\partial T(z, s)}{\partial z} + K s z \frac{\partial F(z, s)}{\partial z} + r_2 \frac{\partial T(z, s)}{\partial s} + \\
&\quad - K z \frac{\partial T(z, s)}{\partial z} - r_2 s \frac{\partial T(z, s)}{\partial s}.
\end{aligned}$$

bądź, po uporządkowaniu:

$$\begin{aligned}
\frac{\partial F(z, s)}{\partial t} &= b_0 T(s, z) - c_0 F(z, s) - r_1 (z - 1) \frac{\partial F(z, s)}{\partial z} + \\
&\quad + K z (s - 1) \frac{\partial F(z, s)}{\partial z} - r_2 (s - 1) \frac{\partial F(z, s)}{\partial s}
\end{aligned} \tag{94}$$

$$\begin{aligned}
\frac{\partial T(z, s)}{\partial t} &= -b_0 T(s, z) + c_0 F(z, s) + H (z - 1) T(z, s) - r_1 (z - 1) \frac{\partial T(z, s)}{\partial z} \\
&\quad + K z (s - 1) \frac{\partial T(z, s)}{\partial z} - r_2 (s - 1) \frac{\partial T(z, s)}{\partial s}.
\end{aligned} \tag{95}$$

Równania (94)-(95) można wykorzystać do wyznaczenia momentów dowolnego rzędu dla funkcji  $g$  i  $h$ . W pracy tej wyznaczymy momenty pierwszego i drugiego rzędu. Z własności FTM, momenty pierwszego rzędu mRNA i białka dla  $G \in \{0, 1\}$  dane są przez związki:

$$E[x; G = 0] = \frac{\partial F(z, s)}{\partial z} \Big|_{z=s=1} \tag{96}$$

$$E[x; G = 1] = \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=1} \tag{97}$$

$$E[y; G = 0] = \frac{\partial F(z, s)}{\partial s} \Big|_{z=s=1} \tag{98}$$

$$E[y; G = 1] = \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=1}. \tag{99}$$

Położmy  $\frac{\partial F(z,s)}{\partial t} = \frac{\partial T(z,s)}{\partial t} = 0$ :

$$b_0 T(s, z) - c_0 F(z, s) - r_1 (z - 1) \frac{\partial F(z, s)}{\partial z} + \quad (100)$$

$$+ Kz (s - 1) \frac{\partial F(z, s)}{\partial z} - r_2 (s - 1) \frac{\partial F(z, s)}{\partial s} = 0 \quad (101)$$

$$-b_0 T(s, z) + c_0 F(z, s) + H (z - 1) T(z, s) - r_1 (z - 1) \frac{\partial T(z, s)}{\partial z} + \quad (102)$$

$$+ Kz (s - 1) \frac{\partial T(z, s)}{\partial z} - r_2 (s - 1) \frac{\partial T(z, s)}{\partial s} = 0, \quad (103)$$

po zróżniczkowaniu równń (101)-(103) osobno po  $z$  i  $s$ , a następnie po podstawieniu  $z = s = 1$ , otrzymamy rozwiązania stacjonarne układu równań algebraicznych: momenty całkowite pierwszego rzędu dla, odpowiednio, mRNA i białka:

$$E[x] = E[x; G = 0] + E[x; G = 1] \quad (104)$$

$$E[y] = E[y; G = 0] + E[y; G = 1]. \quad (105)$$

Zróżniczkujemy układ (101)-(103) po  $z$ :

$$b_0 \frac{\partial T}{\partial z} - c_0 \frac{\partial F}{\partial z} - r_1 \frac{\partial F}{\partial z} - r_1 (z - 1) \frac{\partial^2 F}{\partial z^2} + K (s - 1) \frac{\partial F}{\partial z} + \quad (106)$$

$$+ Kz (s - 1) \frac{\partial^2 F}{\partial z^2} + -r_2 (s - 1) \frac{\partial^2 F}{\partial s \partial z} = 0,$$

$$-b_0 \frac{\partial T}{\partial z} + c_0 \frac{\partial F}{\partial z} + HT + H (z - 1) \frac{\partial T}{\partial z} + \quad (107)$$

$$-r_1 \frac{\partial T}{\partial z} - r_1 (z - 1) \frac{\partial^2 T}{\partial z^2} \partial^2 T \partial z^2 + -r_2 (s - 1) \frac{\partial^2 T}{\partial s \partial z} = 0,$$

położmy  $z = s = 1$ :

$$b_0 \frac{\partial T}{\partial z} \Big|_{z=s=1} - c_0 \frac{\partial F}{\partial z} \Big|_{z=s=1} - r_1 \frac{\partial F}{\partial z} \Big|_{z=s=1} = 0$$

$$-b_0 \frac{\partial T}{\partial z} \Big|_{z=s=1} + c_0 \frac{\partial F}{\partial z} \Big|_{z=s=1} + HT \Big|_{z=s=1} - r_1 \frac{\partial T}{\partial z} \Big|_{z=s=1} = 0,$$

po podstawieniu (89) oraz uwzględnieniu (96) i (97) otrzymujemy momenty pierwszego rzędu dla mRNA:

$$E[x; G = 0] = \frac{\partial F(z, s)}{\partial z} \Big|_{z=s=1} = \frac{c_0 b_0 H}{r_1 (c_0 + b_0) (c_0 + b_0 + r_1)} \quad (108)$$

$$E[x; G = 1] = \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=1} = \frac{c_0 H (c_0 + r_1)}{r_1 (c_0 + b_0) (c_0 + b_0 + r_1)} \quad (109)$$

zatem, zgodnie z (104), moment całkowity pierwszego rzędu dla mRNA wynosi:

$$E_D[x] = \frac{c_0 H}{r_1(c_0 + b_0)}. \quad (110)$$

Aby otrzymać momenty pierwszego rzędu dla białka zróżniczkujemy równania (101)-(103)

po  $s$ :

$$\begin{aligned} b_0 \frac{\partial T}{\partial s} - c_0 \frac{\partial F}{\partial s} - r_1(z-1) \frac{\partial^2 F}{\partial z \partial s} + Kz \frac{\partial F}{\partial z} + Kz(s-1) \frac{\partial^2 F}{\partial z \partial s} + \\ -r_2 \frac{\partial F}{\partial s} - r_2(s-1) \frac{\partial^2 F}{\partial s^2} = 0 \end{aligned} \quad (111)$$

$$\begin{aligned} -b_0 \frac{\partial T}{\partial s} + c_0 \frac{\partial F}{\partial s} + H(z-1) \frac{\partial T}{\partial s} - r_1(z-1) \frac{\partial^2 T}{\partial z \partial s} + Kz \frac{\partial T}{\partial z} + \\ + Kz(s-1) \frac{\partial^2 T}{\partial z \partial s} - r_2 \frac{\partial T}{\partial s} - r_2(s-1) \frac{\partial^2 T}{\partial s^2} = 0, \end{aligned} \quad (112)$$

położmy  $z = s = 1$ :

$$\begin{aligned} b_0 \frac{\partial T}{\partial s} \Big|_{z=s=1} - c_0 \frac{\partial F}{\partial s} \Big|_{z=s=1} + K \frac{\partial F}{\partial z} \Big|_{z=s=1} - r_2 \frac{\partial F}{\partial s} \Big|_{z=s=1} = 0 \\ -b_0 \frac{\partial T}{\partial z} \Big|_{z=s=1} + c_0 \frac{\partial F}{\partial z} \Big|_{z=s=1} + K \frac{\partial T}{\partial z} \Big|_{z=s=1} - r_2 \frac{\partial T}{\partial s} \Big|_{z=s=1} = 0 \end{aligned}$$

po podstawieniu (108)-(109) oraz (98)-(99) otrzymujemy momenty pierwszego rzędu dla białka:

$$E_D[y; G = 0] = \frac{\partial F(z, s)}{\partial s} \Big|_{z=s=1} = \frac{c_0 b_0 K H (c_0 + b_0 + r_1 + r_2)}{r_1 r_2 (c_0 + b_0) (c_0 + b_0 + r_1) (c_0 + b_0 + r_2)} \quad (113)$$

$$E_D[y; G = 1] = \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=1} = \frac{c_0 K H [(c_0 + r_1)(c_0 + r_2) + b_0 c_0]}{r_1 r_2 (c_0 + b_0) (c_0 + b_0 + r_1) (c_0 + b_0 + r_2)} \quad (114)$$

zatem, zgodnie z (105) moment całkowity pierwszego rzędu dla białka wynosi:

$$E_D[y] = \frac{c_0 H K}{r_1 r_2 (c_0 + b_0)} = \frac{K}{r_2} E[x] \quad (115)$$

Aby wyznaczyć wariancję mRNA i białka należy policzyć następujące momenty drugiego

rzędu:

$$E[x(x-1); G = 0] = E[x^2; G = 0] - E[x; G = 0] = \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=1} \quad (116)$$

$$E[x(x-1); G = 1] = \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=1} \quad (117)$$

$$E[y(y-1); G = 0] = \frac{\partial^2 F(z, s)}{\partial s^2} \Big|_{z=s=1} \quad (118)$$

$$E[y(y-1); G = 1] = \frac{\partial^2 T(z, s)}{\partial s^2} \Big|_{z=s=1} \quad (119)$$

$$E[xy; G = 0] = \frac{\partial^2 F(z, s)}{\partial z \partial s} \Big|_{z=s=1} \quad (120)$$

$$E[xy; G = 1] = \frac{\partial^2 T(z, s)}{\partial z \partial s} \Big|_{z=s=1}, \quad (121)$$

przy czym

$$\frac{\partial^2 F(z, s)}{\partial z^2} = \sum_{x,y} x(x-1) z^{x-2} s^y g_{xy}$$

$$\frac{\partial^2 F(z, s)}{\partial s^2} = \sum_{x,y} y(y-1) z^x s^{y-2} g_{xy}$$

$$\frac{\partial^2 T(z, s)}{\partial z^2} = \sum_{x,y} x(x-1) z^{x-2} s^y h_{xy}$$

$$\frac{\partial^2 T(z, s)}{\partial s^2} = \sum_{x,y} y(y-1) z^x s^{y-2} h_{xy},$$

wtedy wariancję dla mRNA oraz białka można wyrazić następująco:

$$Var[x] = \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=1} + \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=1} + E[x] - (E[x])^2 \quad (122)$$

$$Var[y] = \frac{\partial^2 F(z, s)}{\partial s^2} \Big|_{z=s=1} + \frac{\partial^2 T(z, s)}{\partial s^2} \Big|_{z=s=1} + E[y] - (E[y])^2 \quad (123)$$

Różniczkując układ (106)-(107) po  $z$ , a następnie podstawiając  $z = s = 1$  uzyskamy równania na momenty drugiego rzędu dla mRNA:

$$\begin{aligned} -2r_1 \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=1} + b_0 \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=1} - c_0 \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=1} &= 0 \\ 2H \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=1} - 2r_1 \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=1} - b_0 \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=1} + c_0 \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=1} &= 0, \end{aligned}$$

stąd

$$\frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=1} = \frac{c_0 b_0 H^2 (c_0 + r_1)}{r_1^2 (c_0 + b_0) (c_0 + b_0 + r_1) (c_0 + b_0 + 2r_1)} \quad (124)$$

$$\begin{aligned} \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=1} &= \frac{c_0 H^2 (c_0 + r_1) (c_0 + 2r_1)}{r_1^2 (c_0 + b_0) (c_0 + b_0 + r_1) (c_0 + b_0 + 2r_1)} = \\ &= b_0 (c_0 + 2r_1) \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=1}, \end{aligned} \quad (125)$$

zatem, zgodnie z (122), wariancja dla mRNA wynosi:

$$Var_D[x] = \frac{c_0 H [(c_0 + b_0) (c_0 + b_0 + r_1) + b_0 H]}{r_1 (c_0 + b_0)^2 (c_0 + b_0 + r_1)} = \frac{b_0 r_1}{c_0 (c_0 + b_0 + r_1)} E^2[x] + E[x]. \quad (126)$$

Aby otrzymać momenty drugiego rzędu dla białka, różniczkujemy układ (111)-(112) po  $s$ , a następnie podstawiamy  $z = s = 1$ :

$$2K \frac{\partial^2 F}{\partial z \partial s} \Big|_{z=s=1} - 2r_2 \frac{\partial^2 F}{\partial s^2} \Big|_{z=s=1} + b_0 \frac{\partial^2 T}{\partial s^2} \Big|_{z=s=1} - c_0 \frac{\partial^2 F}{\partial s^2} \Big|_{z=s=1} = 0 \quad (127)$$

$$2K \frac{\partial^2 T}{\partial z \partial s} \Big|_{z=s=1} - 2r_2 \frac{\partial^2 T}{\partial s^2} \Big|_{z=s=1} - b_0 \frac{\partial^2 T}{\partial s^2} \Big|_{z=s=1} + c_0 \frac{\partial^2 F}{\partial s^2} \Big|_{z=s=1} = 0. \quad (128)$$

Aby wyznaczyć momenty mieszane  $\frac{\partial^2 F}{\partial z \partial s} \Big|_{z=s=1}$  i  $\frac{\partial^2 T}{\partial z \partial s} \Big|_{z=s=1}$ , różniczkujemy układ (111)-(112) po  $z$ , a następnie kładziemy  $z = s = 1$ :

$$b_0 \frac{\partial^2 T}{\partial z \partial s} \Big|_{z=s=1} - (c_0 + r_1 + r_2) \frac{\partial^2 F}{\partial z \partial s} \Big|_{z=s=1} + K \left( \frac{\partial F}{\partial z} \Big|_{z=s=1} + \frac{\partial^2 F}{\partial z^2} \Big|_{z=s=1} \right) = 0 \quad (129)$$

$$\begin{aligned} c_0 \frac{\partial^2 F}{\partial z \partial s} \Big|_{z=s=1} + (b_0 - r_1) \frac{\partial^2 T}{\partial z \partial s} \Big|_{z=s=1} + H \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=1} - r_2 \frac{\partial^2 T}{\partial s^2} \Big|_{z=s=1} + \\ + K \left( \frac{\partial T}{\partial z} \Big|_{z=s=1} + \frac{\partial^2 T}{\partial z^2} \Big|_{z=s=1} \right) = 0, \end{aligned} \quad (130)$$

po podstawieniu (108)-(109), (113)-(114) oraz (124)-(125), rozwiązujemy układ czterech równań (127)-(128) i (129)-(130) z czterema niewiadomymi; zgodnie z (123) otrzymujemy moment całkowity drugiego rzędu dla białka postaci:

$$Var_D[y] = \frac{b_0 r_1 r_2 (r_1 + r_2 + c_0 + b_0)}{c_0 (r_1 + r_2) (r_1 + c_0 + b_0) (r_2 + c_0 + b_0)} (E_D[y])^2 + K (r_1 + r_2) E_D[y] + E_D[y].$$



## Model ciągły, $N=1$

Niech dane będą funkcje tworzące momentów (FTM) dla mRNA ( $x$ ) i białka ( $y$ ) o rozkładzie ciągłym  $g_{xy}$ ,  $h_{xy}$ :

$$F(z, s) = \int_x \int_y e^{zx+sy} g(x, y) dx dy \quad (131)$$

$$T(z, s) = \int_x \int_y e^{zx+sy} h(x, y) dx dy. \quad (132)$$

Zauważmy, że  $g_{xy}$  oraz  $h_{xy}$  są prawdopodobieństwami, że gen jest, odpowiednio, aktywny oraz nieaktywny:

$$g_{xy} = F(z, s)|_{z=s=0} = \int_x \int_y g(x, y) dx dy = \frac{b_0}{c_0 + b_0}$$

$$h_{xy} = T(z, s)|_{z=s=0} = \int_x \int_y h(x, y) dx dy = \frac{c_0}{c_0 + b_0}.$$

Równania ciągłości (39)-(40) mnożymy stronami przez  $e^{zx+sy}$ , a następnie całkujemy po  $x$  i  $y$ :

$$\begin{aligned} & \int_x \int_y e^{zx+sy} \frac{\partial g(x, y)}{\partial t} dx dy - \int_x \int_y e^{zx+sy} \frac{\partial}{\partial x} (r_1 x g(x, y)) dx dy + \\ & + \int_x \int_y e^{zx+sy} \frac{\partial}{\partial y} [(Kx - r_2 y) g(x, y)] dx dy = \\ = & b_0 \int_x \int_y e^{zx+sy} h(x, y) dx dy - c_0 \int_x \int_y e^{zx+sy} g(x, y) dx dy, \\ & \int_x \int_y e^{zx+sy} \frac{\partial h(x, y)}{\partial t} dx dy + \int_x \int_y e^{zx+sy} \frac{\partial}{\partial x} [(H - r_1 x) h(x, y)] dx dy + \\ & + \int_x \int_y e^{zx+sy} \frac{\partial}{\partial y} [(Kx - r_2 y) h(x, y)] dx dy = \\ = & -b_0 \int_x \int_y e^{zx+sy} h(x, y) dx dy + c_0 \int_x \int_y e^{zx+sy} g(x, y) dx dy. \end{aligned} \quad (133)$$

$$\quad (134)$$

Zauważmy, że:

$$\begin{aligned}
& \int_y \int_x e^{zx+sy} \frac{\partial}{\partial x} (r_1 x g) dx dy = r_1 \int_y \left[ e^{zx+sy} x g - \int_x (z e^{zx+sy} x g) dx \right] dy = r_1 (1-z) \frac{\partial F}{\partial z} \\
& \int_x \int_y e^{zx+sy} \frac{\partial}{\partial y} [(Kx - r_2 y) g] dx dy = \tag{135} \\
= & \int_x \left[ e^{zx+sy} (Kx - r_2 y) g - \int_y (s e^{zx+sy} (Kx - r_2 y) g) dy \right] dx = K(1-s) \frac{\partial F}{\partial z} - r_2 (1-s) \frac{\partial F}{\partial s} \\
& \int_x \int_y e^{zx+sy} \frac{\partial}{\partial y} [(Kx - r_2 y) h(x, y)] dx dy = K(1-s) \frac{\partial T}{\partial z} - r_2 (1-s) \frac{\partial T}{\partial s} \\
& \int_x \int_y e^{zx+sy} \frac{\partial}{\partial x} [(H - r_1 x) h] dx dy = \\
= & \int_y \left[ e^{zx+sy} (H - r_1 x) h - \int_x (z e^{zx+sy} (H - r_1 x) h) dx \right] dy = (1-z) HT - r_1 (1-z) \frac{\partial T}{\partial z}
\end{aligned}$$

Po podstawieniu związków (135) do (133)-(134) otrzymujemy następujące równania różniczkowe cząstkowe na FTM:

$$\frac{\partial F(z, s)}{\partial t} + (r_1 z - Ks) \frac{\partial F(z, s)}{\partial z} + r_2 s \frac{\partial F(z, s)}{\partial s} = b_0 T - c_0 F \tag{136}$$

$$\frac{\partial T(z, s)}{\partial t} - zHT + (r_1 z - Ks) \frac{\partial T(z, s)}{\partial z} + r_2 s \frac{\partial T(z, s)}{\partial s} = -b_0 T + c_0 F. \tag{137}$$

Położmy  $\frac{\partial F(z, s)}{\partial t} = \frac{\partial T(z, s)}{\partial t} = 0$ :

$$b_0 T - c_0 F - (r_1 z - Ks) \frac{\partial F(z, s)}{\partial z} - r_2 s \frac{\partial F(z, s)}{\partial s} = 0 \tag{138}$$

$$-b_0 T + c_0 F + zHT - (r_1 z - Ks) \frac{\partial T(z, s)}{\partial z} - r_2 s \frac{\partial T(z, s)}{\partial s} = 0. \tag{139}$$

Różniczkując równania (138)-(139) po  $z$ , a następnie podstawiając  $z = s = 0$  otrzymujemy następujące równania na momenty pierwszego rzędu dla mRNA:

$$b_0 \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=0} - c_0 \frac{\partial F(z, s)}{\partial z} \Big|_{z=s=0} - r_1 \frac{\partial F(z, s)}{\partial z} \Big|_{z=s=0} = 0 \tag{140}$$

$$-b_0 \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=0} + c_0 \frac{\partial F(z, s)}{\partial z} \Big|_{z=s=0} + HT \Big|_{z=s=0} - r_1 \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=0} = 0, \tag{141}$$

stąd, momenty pierwszego rzędu dla mRNA wynoszą:

$$E_C[x; G = 0] = \frac{\partial F(z, s)}{\partial z} \Big|_{z=s=0} = \frac{Hc_0b_0}{r_1(c_0 + b_0 + r_1)(c_0 + b_0)} \quad (142)$$

$$E_C[x; G = 1] = \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=0} = \frac{Hc_0(c_0 + r_1)}{r_1(c_0 + b_0 + r_1)(c_0 + b_0)}, \quad (143)$$

zaś, zgodnie z (104), moment całkowity pierwszego rzędu dla mRNA wynosi:

$$E_C[x] = \frac{Hc_0}{r_1(c_0 + b_0)}. \quad (144)$$

Różniczkując równania (138)-(139) po  $s$  oraz podstawiając  $z = s = 0$  otrzymujemy następujące równania na momenty pierwszego rzędu dla białka:

$$\begin{aligned} b_0 \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=0} - c_0 \frac{\partial F(z, s)}{\partial s} \Big|_{z=s=0} + K \frac{\partial F(z, s)}{\partial z} \Big|_{z=s=0} - r_2 \frac{\partial F(z, s)}{\partial s} \Big|_{z=s=0} &= 0 \\ -b_0 \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=0} + c_0 \frac{\partial F(z, s)}{\partial s} \Big|_{z=s=0} + K \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=0} - r_2 \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=0} &= 0, \end{aligned}$$

stąd

$$\begin{aligned} E_C[y; G = 0] &= \frac{\partial F(z, s)}{\partial s} \Big|_{z=s=0} = \frac{KHc_0b_0(c_0 + b_0 + r_1 + r_2)}{r_1r_2(c_0 + b_0 + r_1)(c_0 + b_0 + r_2)(c_0 + b_0)} = \quad (145) \\ &= \frac{K(b_0 + r_2)}{r_2(c_0 + b_0 + r_2)} E[x; G = 0] + \frac{Kb_0}{r_2(c_0 + b_0 + r_2)} E[x; G = 1] \end{aligned}$$

$$\begin{aligned} E_C[y; G = 1] &= \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=0} = \frac{KHc_0[c_0b_0 + (c_0 + r_1)(c_0 + r_2)]}{r_1r_2(c_0 + b_0 + r_1)(c_0 + b_0 + r_2)(c_0 + b_0)} = \quad (146) \\ &= \frac{Kc_0}{r_2(c_0 + b_0 + r_2)} E[x; G = 0] + \frac{K(c_0 + r_2)}{r_2(c_0 + b_0 + r_2)} E[x; G = 1]. \end{aligned}$$

Zatem, zgodnie z (122), moment całkowity pierwszego rzędu dla białka wynosi:

$$E_C[y] = \frac{KHc_0}{r_1r_2(c_0 + b_0)} = \frac{K}{r_2} E_C[x]. \quad (147)$$

Po zróżniczkowaniu układu (106)-(107) dwukrotnie po  $z$  oraz po podstawieniu  $z = s = 0$

uzyskamy równania na momenty drugiego rzędu dla mRNA:

$$b_0 \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=0} - c_0 \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=0} - 2r_1 \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=0} = 0 \quad (148)$$

$$-b_0 \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=0} + c_0 \frac{\partial^2 F(z, s)}{\partial z^2} \Big|_{z=s=0} + 2H \frac{\partial T(z, s)}{\partial z} \Big|_{z=s=0} + \quad (149)$$

$$-2r_1 \frac{\partial^2 T(z, s)}{\partial z^2} \Big|_{z=s=0} = 0. \quad (150)$$

Po rozwiązaniu równań (148)-(150) otrzymujemy moment całkowity drugiego rzędu dla mRNA:

$$Var_C[x] = \frac{c_0 b_0 H^2}{r_1^2 (r_1 + c_0 + b_0) (c_0 + b_0)^2}. \quad (151)$$

Aby otrzymać momenty drugiego rzędu dla białka, najpierw różniczkujemy układ (111)-(112) dwukrotnie po  $s$  oraz podstawiamy  $z = s = 0$ :

$$b_0 \frac{\partial^2 T(z, s)}{\partial s^2} \Big|_{z=s=0} - c_0 \frac{\partial^2 F(z, s)}{\partial s^2} \Big|_{z=s=0} + 2K \frac{\partial^2 F(z, s)}{\partial z \partial s} \Big|_{z=s=0} + \quad (152)$$

$$-2r_2 \frac{\partial^2 F(z, s)}{\partial s^2} \Big|_{z=s=0} = 0$$

$$-b_0 \frac{\partial^2 T(z, s)}{\partial s^2} \Big|_{z=s=0} + c_0 \frac{\partial^2 F(z, s)}{\partial s^2} \Big|_{z=s=0} + 2K \frac{\partial^2 T(z, s)}{\partial z \partial s} \Big|_{z=s=0} + \quad (153)$$

$$-2r_2 \frac{\partial^2 T(z, s)}{\partial s^2} \Big|_{z=s=0} = 0,$$

aby obliczyć momenty mieszane  $\frac{\partial^2 F}{\partial z \partial s} \Big|_{z=s=1}$  i  $\frac{\partial^2 T}{\partial z \partial s} \Big|_{z=s=1}$ , różniczkujemy układ (140)-(141) po  $s$ , a następnie kładziemy  $z = s = 0$ :

$$b_0 \frac{\partial^2 T(z, s)}{\partial z \partial s} \Big|_{z=s=0} - c_0 \frac{\partial^2 F(z, s)}{\partial z \partial s} \Big|_{z=s=0} + \quad (154)$$

$$-r_1 \frac{\partial^2 F(z, s)}{\partial z \partial s} \Big|_{z=s=0} = 0$$

$$-b_0 \frac{\partial^2 T(z, s)}{\partial z \partial s} \Big|_{z=s=0} + c_0 \frac{\partial^2 F(z, s)}{\partial z \partial s} \Big|_{z=s=0} + H \frac{\partial T(z, s)}{\partial s} \Big|_{z=s=0} + \quad (155)$$

$$-r_1 \frac{\partial^2 T(z, s)}{\partial z \partial s} \Big|_{z=s=0} = 0,$$

rozwiązujemy układ czterech równań z czterema niewiadomymi i otrzymujemy wariancję rozkładu białka:

$$Var_C[y] = \frac{b_0 r_1 r_2 (r_1 + r_2 + c_0 + b_0)}{c_0 (r_1 + r_2) (r_1 + c_0 + b_0) (r_2 + c_0 + b_0)} (E_C[y])^2.$$

## Przybliżenie K-E, $N=1$ , $N=2$

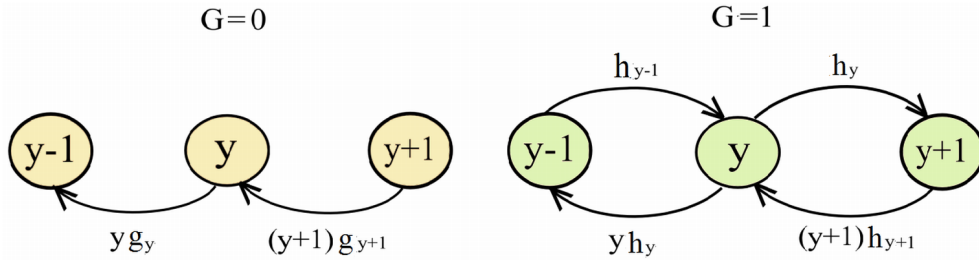
Aby sprawdzić poprawność związków (59)-(60) obliczymy wartość oczekiwaną oraz wariancję białka najpierw dla układu z jedną kopią genu, a następnie dla układu z dwiema kopiami genu. Dla uproszczenia rachunków obliczenia wykonamy dla modeli dyskretnych w przybliżeniu Keplera-Elstona.

Wykorzystując następujące równania Mistrzów opisujące ewolucję w czasie prawdopodobieństw  $g_y$ ,  $h_y$  (patrz Rys. A1):

$$\frac{dg_y}{dt} = b_0 h_y - c_0 g_y - y g_y + (y+1) g_{y+1} \quad (154)$$

$$\frac{dh_y}{dt} = -b_0 h_y + c_0 g_y - y h_y + (y+1) h_{y+1} + h_{y-1} - h_y \quad (155)$$

wyznamy wartość oczekiwaną oraz wariancję białka w przybliżeniu K-E dla układu z jedną kopią genu.



Rysunek A1: Diagram ewolucji gęstości prawdopodobieństw zadanych przez proces Markowa pierwszego rzędu dla układu z jedną kopią genu.

Niech dane będą funkcje FTM dla białka ( $y$ ) o rozkładzie  $g_y$ ,  $h_y$ :

$$F_0(z) : = \sum_{y=0}^{\infty} z^y g_y \quad (156)$$

$$F_1(z) : = \sum_{y=0}^{\infty} z^y h_y. \quad (157)$$

Wprowadźmy oznaczenie

$$g : = F_0(z) |_{z=1} = \sum_{y=0}^{\infty} g_y$$

$$h : = F_1(z) |_{z=1} = \sum_{y=0}^{\infty} h_y,$$

$g$  oraz  $h$  są prawdopodobieństwami, że gen jest, odpowiednio, aktywny oraz nieaktywny:

$$g = \frac{b_0}{c_0 + b_0} \quad (158)$$

$$h = \frac{c_0}{c_0 + b_0}.$$

Zapiszmy równania Mistrzów (154)-(155) dla  $\frac{\partial g_y}{\partial t} = \frac{\partial h_y}{\partial t} = 0$ :

$$b_0 h_y - c_0 g_y - y g_y + (y + 1) g_{y+1} = 0 \quad (159)$$

$$-b_0 h_y + c_0 g_y - y h_y + (y + 1) h_{y+1} + h_{y-1} - h_y = 0, \quad (160)$$

pomnóżmy (159)-(160) stronami przez  $y$ , oraz zsumujmy stronami po  $y$ ; otrzymamy wówczas następujący układ równań:

$$b_0 \sum_{y=0}^{\infty} y h_y - c_0 \sum_{y=0}^{\infty} y g_y - \sum_{y=0}^{\infty} y^2 g_y + \sum_{y=0}^{\infty} y (y + 1) g_{y+1} = 0 \quad (161)$$

$$-b_0 \sum_{y=0}^{\infty} y h_y + c_0 \sum_{y=0}^{\infty} y g_y - \sum_{y=0}^{\infty} y^2 h_y + \sum_{y=0}^{\infty} y (y + 1) h_{y+1} +$$

$$+ \sum_{y=0}^{\infty} y h_{y-1} - \sum_{y=0}^{\infty} y h_y = 0. \quad (163)$$

Wprowadźmy oznaczenie:

$$T_0 : = E_I[y; G = 0] = \frac{\partial F_0(z)}{\partial z} |_{z=1} = \sum_{y=0}^{\infty} y g_y$$

$$T_1 : = E_I[y; G = 1] = \frac{\partial F_1(z)}{\partial z} |_{z=1} = \sum_{y=0}^{\infty} y h_y.$$

Zauważmy, że

$$\begin{aligned} \sum_{y=0}^{\infty} y(y+1)g_{y+1} &= \sum_{y=0}^{\infty} (y+1)^2 g_{y+1} - \sum_{y=0}^{\infty} yg_{y+1} - \sum_{y=0}^{\infty} g_{y+1} = \sum_{y=0}^{\infty} y^2 g_y - T_0 \quad (164) \\ \sum_{y=0}^{\infty} yh_{y-1} &= \sum_{y=0}^{\infty} (y-1)h_{y-1} + \sum_{y=0}^{\infty} h_{y-1} = T_1 + h. \end{aligned}$$

Po podstawieniu wyrażeń (164) do (161)-(163) otrzymujemy następujące równania różniczkowe cząstkowe na FTM, z których wyznaczmy momenty pierwszego rzędu dla białka:

$$\begin{aligned} b_0 T_1 - c_0 T_0 - \sum_{y=0}^{\infty} y^2 g_y + \sum_{y=0}^{\infty} y^2 g_y - T_0 &= 0 \\ -b_0 T_1 + c_0 T_0 - \sum_{y=0}^{\infty} y^2 h_y + \sum_{y=0}^{\infty} y^2 h_y - T_1 + T_1 + h - T_1 &= 0, \end{aligned}$$

po uproszczeniu otrzymujemy następujący układ:

$$b_0 T_1 - c_0 T_0 - T_0 = 0 \quad (165)$$

$$-b_0 T_1 + c_0 T_0 + h - T_1 = 0. \quad (166)$$

Rozwiązując układ (165)-(166) otrzymamy momenty pierwszego rzędu dla białka:

$$T_0 = \frac{c_0 b_0}{(c_0 + b_0 + 1)(c_0 + b_0)} \quad (167)$$

$$T_1 = \frac{c_0(c_0 + 1)}{(c_0 + b_0 + 1)(c_0 + b_0)}, \quad (168)$$

zatem moment całkowity pierwszego rzędu białka dla układu z jedną kopią genu w przybliżeniu K-E wynosi:

$$E_I[y] = T_0 + T_1 = \frac{c_0}{c_0 + b_0}.$$

Aby otrzymać momenty drugiego rzędu mnożymy układ (159)-(160) przez  $y^2$  i sumujemy

jak poprzednio:

$$b_0 \sum_{y=0}^{\infty} y^2 h_y - c_0 \sum_{y=0}^{\infty} y^2 g_y - \sum_{y=0}^{\infty} y^3 g_y + \sum_{y=0}^{\infty} y^2 (y+1) g_{y+1} = 0 \quad (169)$$

$$-b_0 \sum_{y=0}^{\infty} y^2 h_y + c_0 \sum_{y=0}^{\infty} y^2 g_y - \sum_{y=0}^{\infty} y^3 h_y + \sum_{y=0}^{\infty} y^2 (y+1) h_{y+1} + \quad (170)$$

$$+ \sum_{y=0}^{\infty} y^2 h_{y-1} - \sum_{y=0}^{\infty} y^2 h_y = 0. \quad (171)$$

Wprowadźmy oznaczenie

$$R_0 : = \sum_{y=0}^{\infty} y^2 g_y$$

$$R_1 : = \sum_{y=0}^{\infty} y^2 h_y,$$

wtedy moment drugiego rzędu dla białka wyraża się następująco:

$$Var_I [y] = R_0 + R_1 + E_I[y] - (E_I[y])^2. \quad (172)$$

Zauważmy, że

$$\begin{aligned} \sum_{y=0}^{\infty} y g_{y-1} &= T_0 + g & (173) \\ \sum_{y=0}^{\infty} y^2 h_{y-1} &= \sum_{y=0}^{\infty} (y-1)^2 h_{y-1} + 2 \sum_{y=0}^{\infty} y h_{y-1} - \sum_{y=0}^{\infty} h_{y-1} = R_1 + 2T_1 + h \\ \sum_{y=0}^{\infty} y^2 g_{y-1} &= \sum_{y=0}^{\infty} (y+1)^2 g_{y-1} - 2 \sum_{y=0}^{\infty} y g_{y-1} - \sum_{y=0}^{\infty} g_{y-1} = R_0 - 2T_0 + g \\ \sum_{y=0}^{\infty} y^2 (y+1) g_{y+1} &= \sum_{y=0}^{\infty} (y+1)^3 g_{y+1} - 2 \sum_{y=0}^{\infty} y^2 g_{y+1} - 3 \sum_{y=0}^{\infty} y g_{y+1} - \sum_{y=0}^{\infty} g_{y+1} = \\ &= \sum_{y=0}^{\infty} (y+1)^3 g_{y+1} - 2(R_0 - 2T_0 + g) - 3(T_0 - g) - g = \\ &= \sum_{y=0}^{\infty} y^3 g_y - 2R_0 + T_0. \end{aligned}$$



Po podstawieniu związków (173) do (169)-(171) otrzymujemy:

$$b_0 R_1 - c_0 R_0 - \sum_{y=0}^{\infty} y^3 g_y + \sum_{y=0}^{\infty} y^3 g_y - 2R_0 + T_0 = 0$$

$$-b_0 R_1 + c_0 R_0 - \sum_{y=0}^{\infty} y^3 h_y + \sum_{y=0}^{\infty} y^3 h_y - 2R_1 + T_1 + R_1 + 2T_1 + h - R_1 = 0,$$

po uproszczeniu

$$b_0 R_1 - R_0(2 + c_0) + T_0 = 0 \quad (174)$$

$$-R_1(2 + b_0) + c_0 R_0 + 3T_1 + h = 0. \quad (175)$$

Podstawiamy (167)-(168), rozwiązujemy układ (174)-(175) i, zgodnie z (172), otrzymujemy wariancję rozkładu białka dla układu z jedną kopią genu w przybliżeniu K-E:

$$Var_I [y] = \frac{c_0 b_0}{(c_0 + b_0)^2 (1 + c_0 + b_0)}.$$

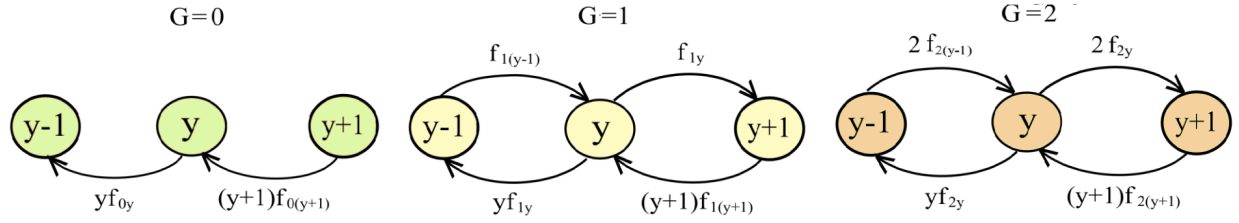
Wykorzystując następujące równania Mistrzów dla układu z dwiema kopiami genu bez autoregulacji, opisujące ewolucję w czasie prawdopodobieństw  $f_0(y)$ ,  $f_1(y)$ ,  $f_2(y)$  (patrz Rys. A2)

$$\frac{df_{0y}(y)}{dt} = b_0 f_{1y} - 2c_0 f_{0y} - y f_{0y} + (y + 1) f_{0,y+1} \quad (176)$$

$$\frac{df_{1y}(y)}{dt} = 2c_0 f_{0y} - (b_0 + c_0) f_{1y} + 2b_0 f_{2y} + f_{1,y-1} + (y + 1) f_{1,y+1} - f_{1y} - y f_{1y} \quad (177)$$

$$\frac{df_{2y}(y)}{dt} = c_0 f_{1y} - 2b_0 f_{2y} + 2f_{2,y-1} + (y + 1) f_{2,y+1} - 2f_{2y} - y f_{2y}, \quad (178)$$

wyznamy wartość oczekiwaną oraz wariancję białka w przybliżeniu K-E dla układu z dwiema kopiami genu.



Rysunek A2: Diagram ewolucji gęstości prawdopodobieństw zadanych przez proces Markowa pierwszego rzędu dla układu z dwiema kopiami genu.

Niech dane będą funkcje FTM dla białka ( $y$ ) o rozkładzie  $f_{0y}$ ,  $f_{1y}$ ,  $f_{2y}$ :

$$F_0(z) : = \sum_{y=0}^{\infty} z^y f_{0y} \quad (179)$$

$$F_1(z) : = \sum_{y=0}^{\infty} z^y f_{1y} \quad (180)$$

$$F_2(z) : = \sum_{y=0}^{\infty} z^y f_{2y}. \quad (181)$$

Wprowadźmy oznaczenie

$$f_0 : = F_0(z) |_{z=1} = \sum_{y=0}^{\infty} f_{0y}$$

$$f_1 : = F_1(z) |_{z=1} = \sum_{y=0}^{\infty} f_{1y}$$

$$f_2 : = F_2(z) |_{z=1} = \sum_{y=0}^{\infty} f_{2y},$$

gdzie  $f_i$ ,  $i \in \{0, 1, 2\}$  jest prawdopodobieństwem, że  $i$ -kopii genów jest aktywnych.

Z następujących równań dla członów odpowiadających przepływowi prawdopodobieństw

związanych ze zmianą aktywności kopii genu

$$-2c_0f_0 + b_0f_1 = 0 \quad (182)$$

$$2c_0f_0 - (b_0 + c_0)f_1 + 2b_0f_2 = 0 \quad (183)$$

$$c_0f_1 - 2b_0f_2 = 0 \quad (184)$$

$$f_0 + f_1 + f_2 = 1 \quad (185)$$

otrzymujemy:

$$f_0 = \frac{b_0^2}{(b_0 + c_0)^2} \quad (186)$$

$$f_1 = \frac{2b_0}{(b_0 + c_0)^2}$$

$$f_2 = \frac{c_0^2}{(b_0 + c_0)^2}.$$

Zróżniczkujemy FTM (179)-(181) po  $z$ :

$$\frac{\partial F_0(z)}{\partial z} = \sum_y yz^{y-1} f_{0y} \quad (187)$$

$$\frac{\partial F_1(z)}{\partial z} = \sum_y yz^{y-1} f_{1y} \quad (188)$$

$$\frac{\partial F_2(z)}{\partial z} = \sum_y yz^{y-1} f_{2y}. \quad (189)$$

Zapiszmy równania Mistrzów (176)-(178) dla  $\frac{\partial f_0(z)}{\partial t} = \frac{\partial f_1(z)}{\partial t} = \frac{\partial f_2(z)}{\partial t} = 0$ :

$$b_0f_{1y} - 2c_0f_{0y} - yf_{0y} + (y+1)f_{0,y+1} = 0 \quad (190)$$

$$2c_0f_{0y} - (b_0 + c_0)f_{1y} + 2b_0f_{2y} + f_{1,y-1} + (y+1)f_{1,y+1} - f_{1y} - yf_{1y} = 0 \quad (191)$$

$$c_0f_{1y} - 2b_0f_{2y} + 2f_{2,y-1} + (y+1)f_{2,y+1} - 2f_{2y} - yf_{2y} = 0 \quad (192)$$

a następnie pomnóżmy (190)-(192) stronami przez  $y$ , oraz zsumujmy stronami po  $y$ ,

otrzymamy wówczas następujący układ równań:

$$b_0 \sum_{y=0}^{\infty} y f_{1y} - 2c_0 \sum_{y=0}^{\infty} y f_{0y} - \sum_{y=0}^{\infty} y^2 f_{0y} + \sum_{y=0}^{\infty} y(y+1) f_{0,y+1} = 0 \quad (193)$$

$$2c_0 \sum_{y=0}^{\infty} y f_{0y} - (b_0 + c_0) \sum_{y=0}^{\infty} y f_{1y} + 2b_0 \sum_{y=0}^{\infty} y f_{2y} + \sum_{y=0}^{\infty} y f_{1,y-1} + \quad (194)$$

$$+ \sum_{y=0}^{\infty} y(y+1) f_{1,y+1} - \sum_{y=0}^{\infty} y f_{1y} - \sum_{y=0}^{\infty} y^2 f_{1y} = 0$$

$$c_0 \sum_{y=0}^{\infty} y f_{1y} - 2b_0 \sum_{y=0}^{\infty} y f_{2y} + 2 \sum_{y=0}^{\infty} y f_{2,y-1} + \sum_{y=0}^{\infty} y(y+1) f_{2,y+1} + \quad (195)$$

$$-2 \sum_{y=0}^{\infty} y f_{2y} - \sum_{y=0}^{\infty} y^2 f_{2y} = 0.$$

Wprowadźmy oznaczenia:

$$T_i := E[y; G = i] = \frac{\partial F_i(z)}{\partial z} \Big|_{z=1} = \sum_{y=0}^{\infty} y f_{iy}, \quad i \in \{0, 1, 2\}$$

$$E_{II}[y] = \sum_i T_i. \quad (196)$$

Zauważmy, że

$$\sum_{y=0}^{\infty} y f_{i,y-1} = \sum_{y=0}^{\infty} (y-1) f_{i,y-1} + \sum_{y=0}^{\infty} f_{i,y-1} = T_i + f_i, \quad i \in \{0, 1, 2\} \quad (197)$$

$$\sum_{y=0}^{\infty} y f_{i,y+1} = \sum_{y=0}^{\infty} (y+1) f_{i,y+1} - \sum_{y=0}^{\infty} f_{i,y+1} = T_i - f_i$$

$$\sum_{y=0}^{\infty} (y+1)^2 f_{i,y+1} = \sum_{y=1}^{\infty} y^2 f_{i,y} = \sum_{y=0}^{\infty} y^2 f_{iy}$$

$$\sum_{y=0}^{\infty} y(y+1) f_{i,y+1} = \sum_{y=0}^{\infty} (y+1)^2 f_{i,y+1} - \sum_{y=0}^{\infty} y f_{i,y+1} - \sum_{y=0}^{\infty} f_{i,y+1} = \sum_{y=0}^{\infty} y^2 f_{iy} - T_i$$

Po podstawieniu wyrażeń (197) do (193)-(195) otrzymujemy następujące równania różniczkowe

cząstkowe na FTM, z których wyznaczmy momenty pierwszego rzędu dla białka:

$$\begin{aligned}
b_0 T_1 - 2c_0 T_0 - \sum_{y=0}^{\infty} y^2 f_{0y} + \sum_{y=0}^{\infty} y^2 f_{0y} - T_0 &= 0 \\
2c_0 T_0 - (b_0 + c_0) T_1 + 2b_0 T_2 + T_1 + f_1 + \sum_{y=0}^{\infty} y^2 f_{1y} - T_1 - T_1 - \sum_{y=0}^{\infty} y^2 f_{1y} &= 0 \\
c_0 T_1 - 2b_0 T_2 + 2(T_2 + f_2) + \sum_{y=0}^{\infty} y^2 f_{2y} - T_2 - 2T_2 - \sum_{y=0}^{\infty} y^2 f_{2y} &= 0,
\end{aligned}$$

po uproszczeniu otrzymujemy następujący układ:

$$b_0 T_1 - (2c_0 + 1) T_0 = 0 \quad (198)$$

$$2c_0 T_0 - (b_0 + c_0 + 1) T_1 + 2b_0 T_2 + f_1 = 0 \quad (199)$$

$$c_0 T_1 - (2b_0 + 1) T_2 + 2f_2 = 0. \quad (200)$$

Rozwiązując układ (198)-(200) otrzymamy momenty pierwszego rzędu dla białka:

$$T_0 = E[y; G = 0] = \frac{2c_0 b_0^2}{(c_0 + b_0 + 1)(c_0 + b_0)^2} \quad (201)$$

$$T_1 = E[y; G = 1] = \frac{2c_0 b_0 (2c_0 + 1)}{(c_0 + b_0 + 1)(c_0 + b_0)^2} \quad (202)$$

$$T_2 = E[y; G = 2] = \frac{2c_0^2 (c_0 + 1)}{(c_0 + b_0 + 1)(c_0 + b_0)^2} \quad (203)$$

zatem, zgodnie z (196), moment całkowity pierwszego rzędu dla białka dla układu z dwiema kopiami genu w przybliżeniu K-E wynosi:

$$E_{II} [y] = \frac{2c_0}{c_0 + b_0} = 2E_I [y], \quad (204)$$

gdzie  $E_I [y]$  jest średnią rozkładu białka dla układu z jedną kopią genu w przybliżeniu K-E (51).

Aby otrzymać momenty drugiego rzędu mnożymy układ (190)-(192) przez  $y^2$  i sumujemy

jak poprzednio:

$$b_0 \sum_{y=0}^{\infty} y^2 f_{1y} - 2c_0 \sum_{y=0}^{\infty} y^2 f_{0y} - \sum_{y=0}^{\infty} y^3 f_{0y} + \sum_{y=0}^{\infty} y^2 (y+1) f_{0,y+1} = 0 \quad (205)$$

$$2c_0 \sum_{y=0}^{\infty} y^2 f_{0y} - (b_0 + c_0) \sum_{y=0}^{\infty} y^2 f_{1y} + 2b_0 \sum_{y=0}^{\infty} y^2 f_{2y} + \sum_{y=0}^{\infty} y^2 f_{1,y-1} + \quad (206)$$

$$+ \sum_{y=0}^{\infty} y^2 (y+1) f_{1,y+1} - \sum_{y=0}^{\infty} y^2 f_{1y} - \sum_{y=0}^{\infty} y^3 f_{1y} = 0$$

$$c_0 \sum_{y=0}^{\infty} y^2 f_{1y} - 2b_0 \sum_{y=0}^{\infty} y^2 f_{2y} + 2 \sum_{y=0}^{\infty} y^2 f_{2,y-1} + \sum_{y=0}^{\infty} y^2 (y+1) f_{2,y+1} + \quad (207)$$

$$- 2 \sum_{y=0}^{\infty} y^2 f_{2y} - \sum_{y=0}^{\infty} y^3 f_{2y} = 0.$$

Wprowadźmy oznaczenie

$$R_i := E_{II}[y^2; G = i] = \frac{\partial^2 F_i(z)}{\partial z^2} \Big|_{z=1} = \sum_{y=0}^{\infty} y^2 f_{iy}, \quad i \in \{0, 1, 2\}$$

wtedy

$$Var_{II}[y] = \sum_{i=0}^2 R_i + E_{II}[y] - (E_{II}[y])^2. \quad (208)$$

Zauważmy, że

$$\sum_{y=0}^{\infty} y f_{i,y-1} = \sum_{y=0}^{\infty} (y-1) f_{i,y-1} + \sum_{y=0}^{\infty} f_{i,y-1} = \sum_{y=0}^{\infty} y f_{i,y} + \sum_{y=0}^{\infty} f_{i,y} = T_i + f_i \quad (209)$$

$$\sum_{y=0}^{\infty} y^2 f_{i,y-1} = \sum_{y=0}^{\infty} (y-1)^2 f_{i,y-1} + 2 \sum_{y=0}^{\infty} y f_{i,y-1} - \sum_{y=0}^{\infty} f_{i,y-1} = R_i + 2T_i + f_i$$

$$\sum_{y=0}^{\infty} y^2 f_{i,y+1} = \sum_{y=0}^{\infty} (y+1)^2 f_{i,y+1} - 2 \sum_{y=0}^{\infty} y f_{i,y+1} - \sum_{y=0}^{\infty} f_{i,y+1} = R_i - 2T_i + f_i$$

$$\begin{aligned}
\sum_{y=0}^{\infty} y^2 (y+1) f_{i,y+1} &= \sum_{y=0}^{\infty} (y+1)^3 f_{i,y+1} - 2 \sum_{y=0}^{\infty} y^2 f_{i,y+1} - 3 \sum_{y=0}^{\infty} y f_{i,y+1} - \sum_{y=0}^{\infty} f_{i,y+1} = \\
&= \sum_{y=0}^{\infty} (y+1)^3 f_{i,y+1} - 2(R_i - 2T_i + f_i) - 3(T_i - f_i) - f_i = \\
&= \sum_{y=0}^{\infty} y^3 f_{iy} - 2R_i + T_i, \quad i \in \{0, 1, 2\}
\end{aligned}$$

Po podstawieniu związków (209) do (205)-(207) otrzymujemy:

$$\begin{aligned}
b_0 R_1 - 2c_0 R_0 - \sum_{y=0}^{\infty} y^3 f_{0y} + \sum_{y=0}^{\infty} y^3 f_{0y} - 2R_0 + T_0 &= 0 \\
2c_0 R_0 - (b_0 + c_0) R_1 + 2b_0 R_2 + R_1 + 2T_1 + f_1 + \sum_{y=0}^{\infty} y^3 f_{1y} - 2R_1 + T_1 - R_1 - \sum_{y=0}^{\infty} y^3 f_{1y} &= 0 \\
c_0 R_1 - 2b_0 R_2 + 2(R_2 + 2T_2 + f_2) + \sum_{y=0}^{\infty} y^3 f_{2y} - 2R_2 + T_2 - 2R_2 - \sum_{y=0}^{\infty} y^3 f_{2y} &= 0,
\end{aligned}$$

po uproszczeniu

$$b_0 R_1 - 2R_0(c_0 + 1) + T_0 = 0 \quad (210)$$

$$2c_0 R_0 - R_1(2 + c_0 + b_0) + 2b_0 R_2 + 3T_1 + f_1 = 0 \quad (211)$$

$$c_0 R_1 - 2R_2(b_0 + 1) + 5T_2 + 2f_2 = 0. \quad (212)$$

Podstawiamy (201)-(203), rozwiązujemy układ (210)-(212) i otrzymujemy, zgodnie z (208), wariancję rozkładu białka dla układu z dwiema kopiami genu w przybliżeniu K-E:

$$Var_{II}[y] = \frac{2c_0 b_0}{(c_0 + b_0)^2 (1 + c_0 + b_0)} = 2Var_I[y],$$

gdzie  $Var_I[y]$  jest wariancją rozkładu białka dla układu z jedną kopią genu w przybliżeniu K-E.

# Dodatek B

## Uzasadnienie przybliżenia K-E

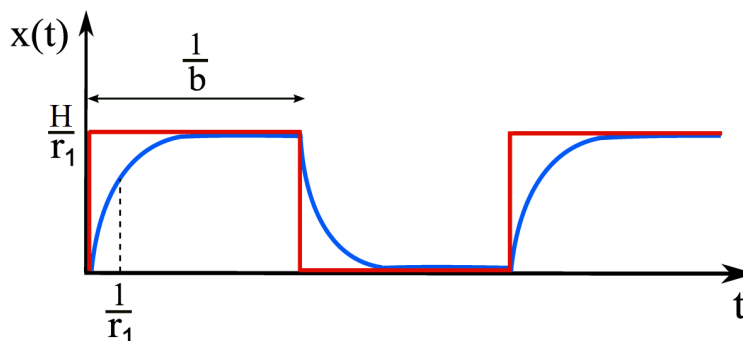
Układ (33)-(34) może zostać uproszczony poprzez założenie, że średni czas życia mRNA jest bardzo mały w porównaniu ze średnim czasem aktywności genu i średnim czasem nieaktywności genu, tzn.

$$\frac{1}{r_1} \ll \frac{1}{b}, \quad \frac{1}{r_1} \ll \frac{1}{c} \quad (213)$$

Rozwiązanie równania (33) przy ustalonym  $G(t) \in \{0, 1\}$  ma postać

$$x(t) = \frac{HG}{r_1} + (x_0 - \frac{HG}{r_1})e^{-r_1 t}, \quad (214)$$

jeśli założymy (213), wówczas trajektorię  $x(t)$  daną przez (214) (Rys. B1, niebieska krzywa) można zastąpić jego stanem stacjonarnym  $x(t) = \frac{HG}{r_1}$  (czerwona krzywa).



Rys. B1: Trajektorja  $x(t)$  zadana przez (214) (niebieska krzywa) oraz funkcja prostokątna (czerwona krzywa).



# Dodatek C

## Niezmienniczość zbioru trajektorii $D_2$

Niech  $D_2 \subset D = [0, \infty) \times [0, \infty)$ . Pokażemy, że jeżeli kinetyka komórki wyznaczona jest przez układ (215)-(216), wtedy  $D_2$  jest zbiorem niezmienniczym.

Dla uproszczenia obliczeń zrobmy zamianę zmiennych

$$x^* = \frac{r_1}{H}x, \quad y^* = \frac{r_1 r_2}{HK}y, \quad t^* = tr_1$$

wówczas, wprowadzając nowe zmienne do układu (33)-(34) i opuszczając gwiazdki otrzymujemy następujący układ:

$$\frac{dx}{dt} = -x + G, \tag{215}$$

$$\frac{dy}{dt} = r(x - y), \quad \text{gdzie } r = \frac{r_2}{r_1} \tag{216}$$

Aby udowodnić niezmienniczość zbioru  $D_2$ , należy pokazać, że trajektorie układu (215)-(216) wyznaczające dla  $G = 0$  oraz  $G = 2$  odpowiednio górną i dolną granicę obszaru  $D_2$  nie mogą opuścić tego obszaru.

Zapiszmy rozwiązania układu (215)-(216) w postaci ogólnej:

dla  $G = 0$ :

$$x_1(t) = x_0 e^{-t},$$

$$y_1(t) = -x_0 \frac{r}{1-r} e^{-t} + \left( x_0 \frac{r}{1-r} + y_0 \right) e^{-rt}, \quad r \neq 1$$

dla  $G = 1$ :

$$x_2(t) = 1 - (1 - x_0) e^{-t},$$

$$y_2(t) = 1 + (1 - x_0) \frac{r}{1-r} e^{-t} - \left( \frac{1}{1-r} - x_0 \frac{r}{1-r} - y_0 \right) e^{-rt}, \quad r \neq 1$$

dla  $G = 2$ :

$$\begin{aligned}x_3(t) &= 2 - (2 - x_0) e^{-t}, \\y_3(t) &= 2 + (2 - x_0) \frac{r}{1-r} e^{-t} - \left( \frac{2}{1-r} - x_0 \frac{r}{1-r} - y_0 \right) e^{-rt}, \quad r \neq 1\end{aligned}$$

Zauważmy, że przy  $t \rightarrow \infty$  trajektorie te zbiegają do punktów  $(0, 0)$ ,  $(1, 1)$ ,  $(2, 2) \in D_2$  dla odpowiednio  $G = 0, 1, 2$  niezależnie od warunków początkowych.

Skoncentrujmy się na na górnej granicy obszaru wyznaczonej przez trajektorię układu (215)-(216) dla  $G = 0$ ; dla dolnej obszaru granicy dowód przebiega analogicznie.

Aby pokazać, że trajektorie startujące z obszaru  $D_2$  nie opuszczają  $D_2$  wystarczy rozważyć następujące iloczyny wektorowe:

$$\vec{v}_{G=2} \times \vec{v}_{G=0} \cdot \vec{z} \text{ oraz } \vec{v}_{G=2} \times \vec{v}_{G=1} \cdot \vec{z}$$

gdzie  $\vec{v}_{G=2} = [\frac{dx}{dt}, \frac{dy}{dt}]|_{G=2}$ ,  $\vec{v}_{G=1} = [\frac{dx}{dt}, \frac{dy}{dt}]|_{G=1}$  oraz  $\vec{v}_{G=0} = [\frac{dx}{dt}, \frac{dy}{dt}]|_{G=0}$  są wektorami stycznymi do trajektorii granicznej wyznaczonej przez układ (215)-(216) dla, odpowiednio,  $G = 2, G = 1$  oraz  $G = 0$ ;  $\vec{z}$  jest wersorem.

Jeżeli iloczyn wektorowy będzie miał znak nieujemny, będzie to oznaczać, że lokalnie takie rozwiązanie również zostanie w  $D_2$  ze względu na ciągłość (włączając pochodne) wektora prędkości definiującego układ (215)-(216) oraz fakt, że każda trajektoria startująca z wnętrza  $D_2$ , zanim opuści  $D_2$ , powinna osiągnąć trajektorię graniczną oraz "wrócić" do  $D_2$  (z jednoznaczności rozwiązań).

Dla ustalenia uwagi przyjmijmy, że  $x < y$  oraz  $0 < x < 2$ ,  $0 < y < 2$ ,  $r > 0$ .

Wyznamy zatem iloczyny wektorowe następujących wektorów:

$$1. \vec{v}_{G=0} = [-x, r(x-y)] \text{ oraz } \vec{v}_{G=2} = [-x+2, r(x-y)] :$$

$$\vec{v}_{G=2} \times \vec{v}_{G=0} \cdot \vec{z} = (-x)r(x-y) - (-x+2)r(x-y) = -2r(x-y) > 0$$

2.  $\vec{v}_{G=0} = [-x, r(x-y)]$  oraz  $\vec{v}_{G=1} = [-x+1, r(x-y)]$  :

$$\vec{v}_{G=2} \times \vec{v}_{G=1G=1} \cdot \vec{z} = (-x)r(x-y) - (-x+1)r(x-y) = -r(x-y) > 0.$$

Zatem obszar  $D_2$  jest zbiorem niezmienniczym, co kończy dowód.

# Dodatek D

## Metoda dyskretyzacji

Układ (66) jest liniowy, jednak dla  $N > 1$  trudno jest go rozwiązać analitycznie lub nawet numerycznie. Nośnikiem stacjonarnych rozkładów  $f_i^{(N)}(y)$  jest przedział  $[0, N]$ . Z wyjątkiem bardzo szczególnych postaci  $b(y)$  i  $c(y)$  wszystkie funkcje  $f_i^{(N)}(y)$  dążą albo do zera albo do nieskończoności na krańcach przedziału  $[0, N]$  dlatego nie ma żadnego punktu startowego. Metoda obliczania stacjonarnych rozkładów  $f_i^{(N)}(y)$  układu (62) zastosowana w tej pracy została wprowadzona w [15]. Polega ona na zastąpieniu układu  $N + 1$  równań różniczkowych cząstkowych (62) przez układ  $N + 1$  liniowych równań różniczkowych zwyczajnych, bądź zastąpieniu układu (66) dla rozkładów stacjonarnych przez układ  $M(N + 1)$  liniowych równań algebraicznych, gdzie  $M$  jest rozmiarem siatki tak wybranym, aby zapewnić wymaganą dokładność. Ciągła zmienna  $y$  zastąpiona jest przez  $\frac{j}{N}$ ,  $j \in \{0, 1, \dots, N\}$ . Niech  $f_{ij}^{(N)}$  oznacza rozkład  $f_i^{(N)}$  w punkcie  $j$  siatki. Zdyskretyzowany układ (62) ma wówczas postać:

$$\begin{aligned} \frac{df_{ij}^{(N)}}{dt} = & \left[ -ibf_{ij}^{(N)} - (N - i)cf_{ij}^{(N)} + (N + 1 - i)cf_{i-1,j}^{(N)} + (i + 1)bf_{i+1,j}^{(N)} \right] + \\ & + if_{i,j-1}^{(N)} + \frac{j + 1}{N}f_{i,j+1}^{(N)} - if_{ij}^{(N)} - \frac{j}{N}f_{ij}^{(N)} \end{aligned}$$

W każdym z powyższych równań pierwsze cztery człony po prawej stronie odpowiadają przepływowi prawdopodobieństw  $f_i^{(N)}$ ,  $i \in \{0, 1, \dots, N\}$  związanych ze zmianą aktywności kopii genu. Pozostałe człony odpowiadają transportowi  $f_i^{(N)}$  odpowiednio z oraz do punktu siatki  $j$ . Kładąc  $\frac{df_{ij}^{(N)}}{dt} = 0$  liczymy rozkłady stacjonarne  $f_i^{(N)}$ . Otrzymujemy wówczas układ  $N \times (N + 1)$  równań algebraicznych. Funkcje  $f_i^{(N)}(y)$  muszą być następnie znormalizowane, aby rozkład brzegowy  $\rho(y) = \sum_i f_i^{(N)}(y)$  spełniał warunek  $\int_0^N \rho(y)dy = 1$ . Algorytm był testowany w przypadku  $N = 1$ , dla którego mamy rozwiązanie analityczne (66) i okazał się być bardzo dokładny dla  $M = 1000$ .

**Tabela 1: Parametry i definicje dla modelu testowego**

Symbol	Wartości	Przedziały	Opis	Komentarze
$s_1$	16	$s^{-1}$	współczynnik syntezy $p53$	
$s_2$	8	$s^{-1}$	konstytucyjny współczynnik syntezy $Mdm2$	
$s_3$	80	$s^{-1}$	współczynnik syntezy $Mdm2$ ( $p53$ -zależny)	
$s_4$	$10^5$	$s^{-1}$	stała nasycenia w syntezie $Mdm2$	
$k_{d1}$	$10^{-13}$	$s^{-1}$	degradacja $p53$ (indukowana przez $Mdm2$ )	wartość w punkcie bifurkacji dla komórek diploidalnych
$k_{d2}$	$2.2 \times 10^{-4}$	$s^{-1}$	współczynnik degradacji $Mdm2$	wartość w punkcie bifurkacji dla komórek diploidalnych
$k_1$	$3.5 \times 10^{-3}$	$s^{-1}$	jądrowy import $Mdm2$ (kontrolowany przez $p53$ )	
$k_2$	2300		stała nasycenia w jądrowym imporcie $Mdm2$	
$m$			liczba kopii genu $p53$	rozważane są różne $m$
$n$			liczba kopii genu $Mdm2$	rozważane są różne $n$

# Bibliografia

- [1] G.K. Basak, A. Bisi, and M.K. Ghosh. Stability of degenerate diffusions with state-dependent switching. *J. Math. Anal. Appl.* **240**:219-248 (1999)
- [2] W.J. Blake, M. Kaern, C.R. Cantor, and J.J. Collins. Noise in eukaryotic gene expression. *Nature* **422**:633-637 (2003)
- [3] A. Bobrowski. Degenerate convergence of semigroups related to a model of stochastic gene expression. *Semigroup forum* **73**:345-366 (2006)
- [4] R. Cheong, A. Bergmann, Sh.L. Werner, J. Regal, A. Hoffmann and A. Levchenko. Transient I $\kappa$ B kinase activity mediates temporal *NF- $\kappa$ B* dynamics in response to wide range of tumor necrosis factor- doses. *J. Biol. Chem.* **281**:2945-2950 (2006)
- [5] D.A. Cook, A.N. Gerber and S.J. Tapscott. Modeling stochastic gene expression: Implications for haploinsufficiency. *Proc. Natl. Acad. Sci. (USA)* **95**:15641-15646 (1998)
- [6] W.D. Cook and B.J. McCaw. Accommodating haploinsufficient tumor suppressor genes in Knudson's model. *Oncogene* **19**:3434-3438 (2000)
- [7] M.B. Elowitz, A.J. Levine, E.D. Siggia and P.S. Swain. Stochastic gene expression in a single cell. *Science* **297**:1183-1186 (2002)
- [8] D.T. Gillespie. Exact stochastic simulations of coupled chemical reactions. *J. Phys. Chem.* **81**:2340-2361 (1977)
- [9] M. Kaern, C.T. Elston, W.J. Blake and J.J. Collins. Stochasticity in gene expression: From theories to phenotypes. *Nature Review Genetics* **6**:451-464 (2005)

- [10] T.B. Kepler and T.C. Elston. Stochasticity in transcriptional regulation: origins, consequences, and mathematical representations. *Biophys. J.* **81**:3116-3136 (2001)
- [11] A.M. Kierzek, J. Zaim and P. Zielenkiewicz. The effect of transcription and translation initiation frequencies on the stochastic fluctuations in prokaryotic gene expression. *J. Biol. Chem.* **276**:8165-8172 (2001)
- [12] M.S.H. Ko. Stochastic model for gene induction. *J. Theor. Biol.* **153**:181-194 (1991)
- [13] T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon and M. Kimmel. Mathematical model of NF-kB regulatory module. *J. Theor. Biol.* **228**:195-215 (2004)
- [14] T. Lipniacki, P. Paszek, A.R. Brasier, B. Tian, H.-Q. Wang, B. Luxon and M. Kimmel. Stochastic regulation in early immune response. *Biophys. J.* **90**:725-742 (2006)
- [15] T. Lipniacki, P. Paszek, A. Marciniak-Czochra, A.R. Brasier and M. Kimmel. Transcriptional stochasticity in gene expression. *J. Theor. Biol.* **238**:348-367 (2006)
- [16] H.H. McAdams and A. Arkin. Stochastic mechanisms in gene expression. *Proc. Natl. Acad. Sci. (USA)* **94**:814-819 (1997)
- [17] J. Paulsson. Models of stochastic gene expression. *Physics of Life Reviews* **2**:157-175 (2005)
- [18] P. Paszek, T. Lipniacki, A.R. Brasier, B. Tian, D.E. Nowak and M. Kimmel. Stochastic effects of multiple regulators on expression profiles in Eukaryotes. *J. Theor. Biol.* **233**:423-433 (2005)
- [19] P. Paszek. Modeling stochasticity in gene regulation. Characterization in the terms of the underlying distribution function. *Bull. Math. Biol.* **69**:1567-601 (2007)
- [20] J.R. Pirone and T.C. Elston. Fluctuations in transcription factor binding can explain the graded and binary responses observed in inducible gene expression. *J. Theor. Biol.* **226**:111-121 (2004)

- [21] Ch.V. Rao, D.M. Wolf and A.P. Arkin. Control exploitation and tolerance of intracellular noise. *Nature* **420**:231-237 (2002)
- [22] J.M. Raser and E.K. O'Shea. Control of stochasticity in eukaryotic gene expression. *Science* **304**:1811-1814 (2004)
- [23] J.D. Seidman and Ch. Seidman. Transcription factor haploinsufficiency: when half a loaf is not enough. *J. Clinical Investigations* **109**:451-455 (2002)
- [24] J.A. Stirland, Z.C. Seymour, S. Windeatt, A.J. Norris, P. Stanley, M.G. Castro, A.S.I. Loudon, M.R.H. White and J.R.E. Davis. Real-time imaging of gene promoter activity using an adenoviral reporter construct demonstrates transcriptional dynamics in normal anterior pituitary cells. *J. Endocrinology* **178**:61-69 (2003)
- [25] Y. Tao. Intrinsic and external noise in an auto-regulatory genetic network. *J. Theor. Biol.* **229**:147-156 (2004)
- [26] Y. Tao. Intrinsic noise, gene regulation and steady-state statistics in a two-gene network. *J. Theor. Biol.* **231**:63-568 (2004)
- [27] N. Takasuka, M.R.H. White, C.D. Wood, W.R. Robertson and J.R.E. Davis. Dynamic changes in prolactin promoter activation in individual living lactotrophic cells. *Endocrinology* **139**:1361-1368 (1998)
- [28] M. Thattai, A. Oudenaarden. Intrinsic noise in gene regulatory networks. *Proc. Natl. Acad. Sci. (USA)* **98**:8614-8619 (2001)
- [29] R. Tomioka, H. Kimura, T.J. Kobayashi and K. Aihara. Multivariate analysis of noise in genetic regulatory networks. *J. Theor. Biol.* **229**:501-521 (2004)
- [30] N.G. van Kampen. Stochastic Processes in Physics and Chemistry (North-Holland, Amsterdam) (1992)



- [31] M.C. Walters, S. Fiering, J. Eidemiller, W. Magis, M. Groudine and D.I.K. Martin. Enhancers increase the probability but not the level of gene expression. *Proc. Natl. Acad. Sci. (USA)* **92**:7125-7129 (1995)
- [32] H. Li, and X. Lin. Positive and negative signaling components involved in TNF $\alpha$ -induced *NF- $\kappa$ B* activation. *Cytokine* **41**:1-8 (2008)
- [33] S.L. Harris and A.J. Levine. The *p53* pathway: positive and negative feedback loops. *Oncogene* **24**:2899-2908 (2005)
- [34] T. Zhang, P. Brazhnik and J.J Tyson. Exploring mechanisms of the DNA-damage response. *Cell Cycle* **6**:85-93 (2007)
- [35] T.Y.-C. Tsai, Y.S. Choi, W. Ma, J.R. Pomerening, C. Tang and J.E. Ferrel Jr. Robust, tunable biological oscillations from interlinked positive and negative feedback loops. *Science* **321**:126-129 (2008)
- [36] N. Geva-Zatorsky, N. Rosenfeld, S. Itzkovitz, R. Milo, A. Sigal, E. Dekel, T. Yarnitzky, Y. Liron, Y., P. Polak, G. Lahav and U. Alon. Oscillations and variability in the *p53* system. *Mol. Sys. Biol.* 2 2006.0033 (2006)
- [37] D.E. Nelson, A.E.C. Ihekweba, M. Elliot, J.R. Johnson, C.A. Gibney, B.E. Foreman, G. Nelson, V. See, C.A. Horton, D.G. Spiller, S.W. Edwards, H.P. McDowell, J.F. Unitt, E. Sullivan, R. Grimley, N. Benson, D. Broomhead, D.B. Kell and M.R.H. White. Oscillations in *NF- $\kappa$ B* signaling control the dynamics of gene expression. *Science* **306**:704-708 (2004)
- [38] A. Hoffmann and D. Baltimore. Circuitry of nuclear  $\kappa$ B factor signaling. *Immunol. Rev.* **210**:171-186 (2006)
- [39] B. Hat, P. Paszek, M. Kimmel, K. Piechor and T. Lipniacki. How the Number of Alleles Influences Gene Autoregulation. *J. Stat. Phys.* **128**:511-533 (2007)

- [40] S. Wiggins. Introduction to Applied Nonlinear Dynamical Systems and Chaos. Springer-Verlag (1990)
- [41] J.J. Tyson. Another turn for *p53*. *Mol. Sys. Biol.* 2006.003 (2006)
- [42] A.R. Brasier. The *NF- $\kappa$ B* regulatory networks. *Cardiovasc. Toxicol.* **6**:111-130 (2006)
- [43] E.G. Lee, D.L. Boone, S. Chai, S.L. Libby, M. Chien, J.P. Lodolce and A. Ma. Failure to regulate TNF-induced *NF- $\kappa$ B* and cell death responses in A20-deficient mice. *Science* **28**: 2350-2354 (2000)
- [44] T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon and M.Kimmel. Mathematical model of *NF- $\kappa$ B* regulatory module. *J. Theor. Biol.* **228**:195-215 (2004)
- [45] A. Hoffmann, A. Levchenko, M.L. Scott and D. Baltimore. The  $\text{I}\kappa\text{B}$  - *NF- $\kappa$ B* signaling module: Temporal control and selective gene activation. *Science* **298**:1241-1245 (2002)
- [46] L. Sun, G. Yang, M. Zaidi and J. Iqbal. TNF-induced gene expression oscillates in time. *Biochem. Biophys. Res. Commun.* doi:10.1016/j.bbrc.2008.03.114 (2008)
- [47] A. Ciliberto, B. Novak and J.J. Tyson. Steady states and oscillations in the *p53/Mdm2* network. *Cell Cycle* **4**:488-493 (2005)
- [48] R.L. Bar-Or, R. Maya, A. Lee, L.A. Segel, U. Alon, A.J. Levine and M. Oren. Generation of oscillations by the *p53-Mdm2* feedback loop: A theoretical and experimental study. *Proc. Nat. Acad. Sci. USA* **97**:11250-11255 (2000)
- [49] K. Puszynski, B. Hat and T. Lipniacki. Oscillations and bistability in the stochastic model of *p53* regulation. *J. Theor. Biol.* **254**:452-465 (2008)
- [50] L. Ma, J. Wagner, J.J. Rice, W. Hu, A.J. Levine and G.A. Stolovitzky. A plausible model for the digital response of *p53* to DNA damage. *Proc. Natl. Acad. Sci. USA* **102**: 14266-14271 (2005)

- [51] J. Wagner, L. Ma, J.J. Rice, W. Hu, A.J. Levine, G.A. Stolovitzky. *p53-Mdm2* loop controlled by a balance of its feedback strength and effective dampening using *ATM* and delayed feedback. *IEE Proc. Syst. Biol.* **152**:109-118 (2005)
- [52] J. Alt, T.C. Greiner, J.L. Cleveland and C.M. Eischen. *Mdm2* haploinsufficiency profoundly inhibits Myc-induced lymphomagenesis. *EMBO J.* **22**:1442-145 (2003)
- [53] S.M. Mendrysa, M.K. McElwee, J. Michalowski, K.A. O’Leary, K.M. Young and M.E. Perry. *Mdm2* is critical for inhibition of *p53* during Lymphopoiesis and the response to ionizing irradiation. *Mol. Cell. Biol.* **23**:462-473 (2003)
- [54] T. Lipniacki and M. Kimmel. Deterministic and stochastic models of *NF- $\kappa$ B* pathway. *Cardiovasc. Toxicol.* **7**:215-234 (2007)
- [55] R. Cheong, A. Hoffmann and A. Levchenko. Understanding *NF- $\kappa$ B* signaling via mathematical modeling. *Mol. Sys. Biol.* **4**:192 (2008)
- [56] S.L. Werner, D. Barken and A. Hoffmann. Stimulus specificity of gene expression programs determined by temporal control of IKK activity. *Science* **309**:1857-1861 (2005)
- [57] T. Lipniacki, P. Paszek, A.R. Brasier, B. Luxon and M. Kimmel, M. Stochastic regulation in early immune response. *Biophys. J.* **90**:725-742 (2006)
- [58] D. Barken, C.J.Wang, J. Kearns, R. Cheong, A. Hoffmann and A. Levchenko. Comment on ‘Oscillations in *NF- $\kappa$ B* signaling control the dynamics of gene expression’. *Science* **308**:52b (2005)
- [59] D.E. Nelson, C.A. Horton, V. See, J.R. Johnson, G. Nelson, D.G. Spiller, D.G.B. Kell and M.R.H. White. Response to comment on ‘Oscillations in *NF- $\kappa$ B* signaling control the dynamics of gene expression. *Science* **308**:52b (2005)

- [60] T. Lipniacki, K. Puszynski, P. Paszek, A.R. Brasier and M. Kimmel. Single TNF $\alpha$  trimers mediating *NF- $\kappa$ B* activation: Stochastic robustness of *NF- $\kappa$ B* signaling. *BMC Bioinformatics* **8**:376 (2007)
- [61] D. Longo and J. Hasty. Dynamics of single-cell gene expression. *Mol. Sys. Biol.* **2**:64 (2006)
- [62] K. Polyak, Y. Xia, J.L. Zweier, K.W. Kinzler and B. Vogelstein. A model for *p53*-induced apoptosis. *Nature* **389**:300-305 (1997)
- [63] S. Jin and A.J. Levine. The *p53* functional circuit. *J. Cell Sci.* **114**:4139-4140 (2001)
- [64] B. Vogelstein, D. Lane and A.J. Levine. Surfing the *p53* network. *Nature* **408**:307-310 (2000)
- [65] C.J. Lynch and J. Milner. Loss of one *p53* allele results in four-fold reduction of *p53* mRNA and protein: a basis for *p53* haplo-insufficiency. *Oncogene* **25**:3463-3470 (2006)
- [66] D.A. Freedman, L. Wu and A.J. Levine. Functions of the *Mdm2* oncoprotein. *Cell. Mol. Life Sci.* **55**:96-107 (1999)
- [67] Y. Barak, T. Juven, R. Haffner and M. Oren. *Mdm2* expression is induced by wild type *p53* activity. *EMBO J.* **12**: 461-468 (1993)
- [68] Y. Haupt, R. Maya, A. Kazaz and M. Oren. *Mdm2* promotes the rapid degradation of *p53*. *Nature* **387**:296-299 (1997)
- [69] M.H.G. Kubbutat, S.N. Jones and K.H. Vousden. Regulation of *p53* stability by *Mdm2*. *Nature* **387**:299-303 (1997)
- [70] X. Wu, J.H. Bayle, D. Olson, and A.J. Levine. The *p53-Mdm2* autoregulatory feedback loop. *Genes Dev.* **7**:1126-1132 (1993)

- [71] K.W. Kohn and Y. Pommier. Molecular interaction map of the *p53* and *Mdm2* logic elements, which control the Off–On switch of *p53* in response to DNA damage. *Biochem. Biophys. Res. Commun.* **331**:816-827 (2005)
- [72] G. Lahav, N. Rosenfeld, A. Sigal, N. Geva-Zatorsky, A.J. Levine, M.B. Elowitz and U. Alon. Dynamics of the *p53-Mdm2* feedback loop in individual cells. *Nat. Genet.* **36**:147-50 (2004)
- [73] Batchelor, E., Mock, C.S., Bhan, I., Loewer A. and Lahav, G. Recurrent Initiation: A mechanisms for triggering *p53* pulses in response to DNA damage. *Mol. Cell* **30**:277-289 (2008)
- [74] T. Lipniacki, B. Hat, J. R. Faeder, W. S. Hlavacek. Stochastic effects and bistability in *T* cell receptor signaling. *J. Theor. Biol.* **254**:110-22 (2008)