

Modelowanie układów reakcji biochemicznych

Autoreferat rozprawy doktorskiej

Mikołaj Rybiński

1 WSTĘP

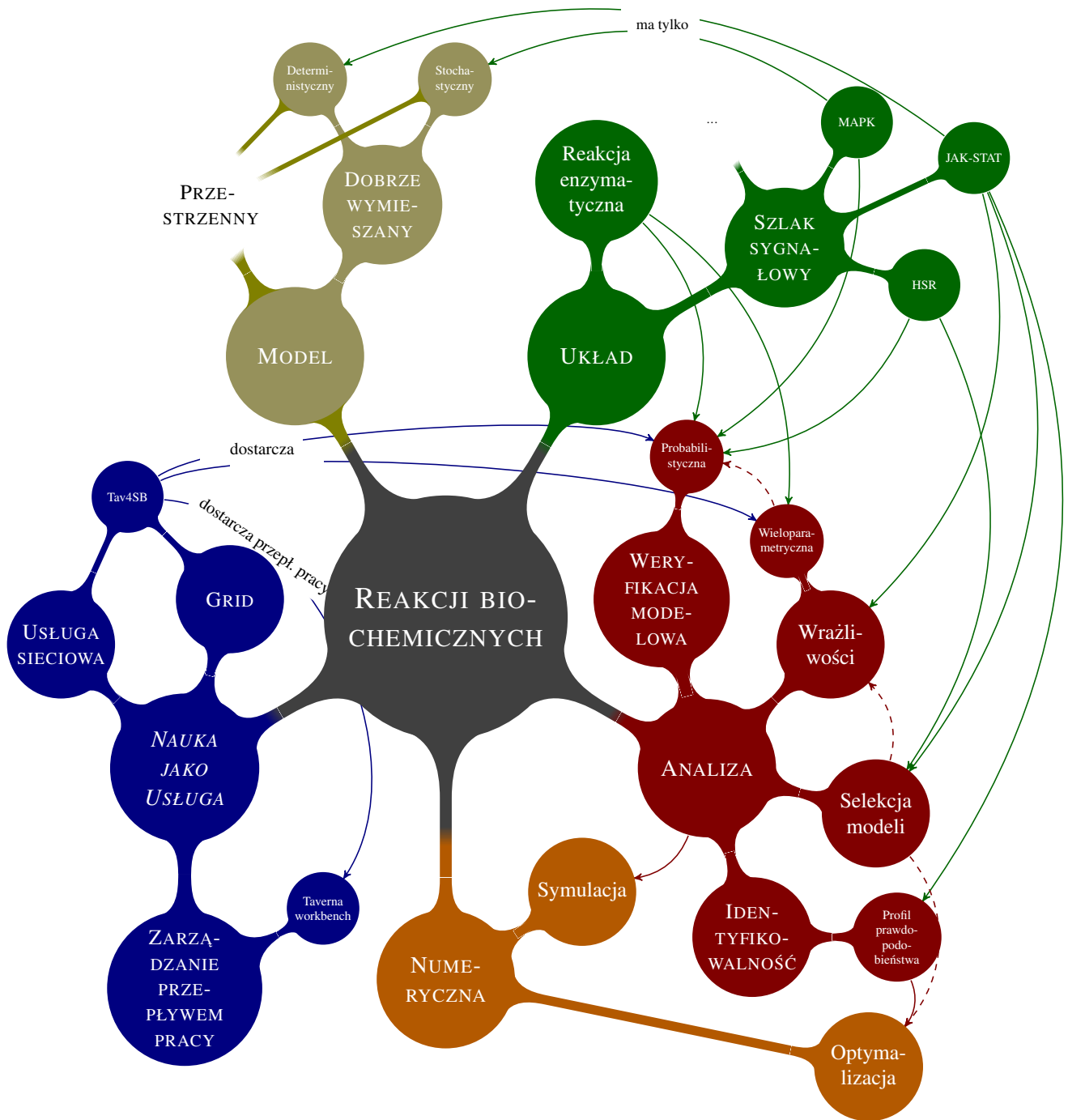
Modelowanie matematyczne dynamiki molekularnych systemów biologicznych uzupełnia technologie eksperymentalne, które są używane do identyfikacji i wyjaśnienia roli elementów badanego systemu. Proces tworzenia i udoskonalania modelu pomaga formułować hipotezy i tym samym sugeruje projekt eksperymentu, który sprawdzi zarówno te hipotezy jak i sam model. Potwierdzony eksperymentalnie model pozwala wnioskować na temat mechanizmów leżących u podstaw interesującej nas funkcji komórki.

Konstrukcja i analiza mechanistycznych modeli molekularnych systemów biologicznych jest częścią niedawno ugruntowanych, wysoce interdyscyplinarnych dziedzin biologii systemów i biologii obliczeniowej. Modelowanie obejmuje sieci regulacji genów (Alon, 2007), przekaźnictwo sygnałów oraz szlaki metaboliczne (Wolkenhauer et al., 2008; Klipp et al., 2009). W tej rozprawie opiszemy i zastosujemy metody poprawne dla jednorodnych w przestrzeni modeli układów reakcji biochemicznych. Jednak skupimy się przede wszystkim na praktycznych zastosowaniach dla modeli wewnątrzkomórkowych szlaków sygnałowych.

Wyniki dotyczą wielu aspektów analizy układów reakcji biochemicznych. Podsumowanie podjętych tematów oraz ich wzajemne relacje przedstawione są na Rysunku 1.1.

Rysunek 1.1: Mapa koncepcyjna rozprawy, przedstawiająca rozważane systemy biologiczne i powiązane metody. Przedstawiony jest również niewykorzystany koncept modeli uwzględniających niejednorodność w przestrzeni. Strzałki reprezentują dodatkowe relacje pomiędzy różnymi konceptami, zgodnie z ich ujęciem w rozprawie. Etykiety strzałek opisują charakter relacji. Dla przykładu, w rozprawie, szlak sygnałowy JAK-STAT ma tylko deterministyczną wersję modelu natomiast projekt TAV4SB dostarcza metody wieloparametrycznej analizy wrażliwości. Wszystkie nieetykietowane strzałki prowadzące od systemów biologicznych do metod numerycznych i metod analizy oznaczają, które z tych metod były użyte w studium przypadku stosownym do wskazującego systemu biologicznego. Podobnie, wszystkie nieetykietowane strzałki prowadzące od metod analizy oznaczają relację wykorzystania danej metody, lub, w przypadku przerywanych strzałek, możliwości jej wykorzystania. Na przykład, selekcja modeli wykorzystuje numeryczną optymalizację, jeżeli rozważymy kryterium dopasowania do danych, natomiast bayesowska selekcja modeli już nie wymaga numerycznej optymalizacji.

Rysunek na następnej stronie.



2 PRZEKAŹNICTWO SYGNAŁU WEWNĄTRZ KOMÓRKI

Podstawowe procesy komórkowe, tj. wzrost, podział czy ruchliwość, są sterowane komunikacją zarówno pomiędzy komórkami jak i wewnątrz komórek (Alberts et al., 2002). Pierwsza, odbywa się przez bezpośredni kontakt pomiędzy komórkami albo przez rozpuszczalne czynniki typu hormony czy cytokiny. Sygnał ze środowiska odbierany jest przez znajdujące się na membranie receptory. Wywołują one kaskadę reakcji biochemicznych w cytoplazmie i w jądrze komórki. W efekcie działania tej kaskady, jedno lub więcej białek wpływa na zachowanie komórki zazwyczaj przez zmianę poziomu ekspresji docelowych genów. Przekaznictwo sygnału może również bezpośrednio wpłynąć na kształt lub ruch komórki albo na jej metabolizm. Szlaki sygnałowe, odpowiedzialne za przetwarzanie konkretnych sygnałów, są częścią całej, wewnątrzkomórkowej sieci reakcji. W większości przypadków, aktywność szlaku sygnałowego jest samoregulowana przez jego elementy, pozwalając komórce kontrolować chwilową odpowiedź na zewnętrzny bodziec (np. wewnątrzkomórkowy sygnał jest tłumiony po udanej zmianie profilu ekspresji docelowych genów).

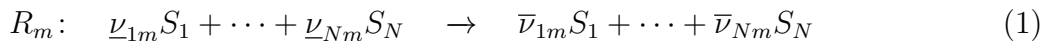
Większość reakcji odpowiedzialnych za przekaznictwo sygnału wewnątrz komórki, to reakcje katalizowane przez enzymy, które aktywują lub dezaktywują kolejne białka na szlaku. Może to się odbywać, przykładowo, przez dodanie lub, odpowiednio, usunięcie grupy fosforanowej, która jest dostarczana przez cząsteczki typu ATP. Modyfikacja fosforanowa białka powoduje zmianę konformacji jego struktury oraz zmianę jego energii swobodnej. Te procesy to, odpowiednio, *fosforylacja* i *defosforylacja* (enzymy, które katalizują takie reakcje są nazywane *kinazami* i *fosfatazami*). Reakcje enzymatyczne są podstawowym elementem skomplikowanych układów przekaznictwa sygnału wewnątrzkomórkowego.

Aktywność elementów szlaku sygnałowego to koncepcja biologiczna, która może być utożsamiana z wydajnością procesu technologicznego, generującego określone liczby molekuł. Dlatego nie jest błędem definiowanie aktywności w terminach ilości danego gatunku. Nota bene, liczba molekuł pojedynczego typu białka wewnątrz komórki jest zazwyczaj rzędu 10^2 – 10^3 .

3 MODELOWANIE

Reakcje biochemiczne odbywają się w wodzie — fizjologicznym środowisku komórki. Dla uproszczenia, zakładamy jednorodność w przestrzeni, tzn. zakładamy, że biomolekuły są dobrze wymieszane oraz, że efekty związane z dyfuzją nie są brane pod uwagę. Co więcej, zakładamy, że zbiór kanałów przez które odbywa-

ją się reakcje jest ustalony. Taki zbiór kanałów reakcji oraz reagujących w nich substancji (gatunków) nazywamy *układem reakcji biochemicznych*, tzn.



gdzie $\underline{\nu}_{nm}$ i $\bar{\nu}_{nm}$ oznaczają ilości molekuł n -tej substancji, które są, odpowiednio, konsumowane i produkowane m -tej w reakcji.

Zależny od czasu stan takiego biochemicznego, kinetycznego układu można opisać w terminach przestrzeni stanów, będących liczbą molekuł $\#\vec{S}$ wektora gatunków \vec{S} . Jednostką ilościowego opisu molekuł jest *mol*, który w przybliżeniu zawiera $N_A \approx 6.022 \cdot 10^{23} \text{ mol}^{-1}$ molekuł (N_A nazywa się *stałą Avogadro*). *Stężenie molowe* definiuje się jako liczbę moli substancji w ustalonej objętości roztworu V , tzn.

$$[S] = \frac{\#S}{N_A \cdot |V|} \quad \text{M}, \quad (2)$$

gdzie $|V|$ to objętość podana w litrach a M, wynoszące $\text{mol} \cdot \text{dm}^{-3}$, oznacza jednostkę stężenia molowego.

Złożone układy biologiczne można modelować w sposób deterministyczny, stochastyczny lub hybrydowy. Pierwszy sposób wykorzystuje klasyczną teorię równań różniczkowych, drugi natomiast procesy stochastyczne lub teorię stochastycznych równań różniczkowych. Jak już wcześniej wspominaliśmy, oba typy modeli wynikają z upraszczających rzeczywistość założeń. W szczególności, zakładamy, że proces dyfuzji zachodzi natychmiastowo, co zapewnia równomierny rozkład substancji na ograniczonej przestrzeni. Zakładamy również, że parametry środowiska biochemicznego, takie jak temperatura, są mniej więcej stałe. Pozawala to uznać zbiór reakcji i , co ważniejsze, ich parametrów kinetycznych, za ustalone w ciągu całej rozważanej ewolucji systemu.

Opis kinetyki większości modeli wynika z klasycznego *prawa działania mas* (zob. np. [Lund, 1965](#)). Modele deterministyczne opisują zmianę stężeń reagentów w czasie i te modele nie uwzględniają efektów fluktuacji (które występują w rzeczywistości). Oznacza to, że dla zadanych warunków początkowych model deterministyczny zawsze zwróci ten sam wynik. Jest to tzw. zagadnienie początkowe dla równań różniczkowych pierwszego rzędu, które są matematyczną reprezentacją modelu deterministycznego. W najpowszechniej rozważanym przypadku ciągłego czasu i jednorodności w przestrzeni, bez efektów opóźnień, deterministyczny model biochemiczny reprezentują *równania szybkości reakcji* (ang. reaction rate equations; RRE):

$$\frac{d[\vec{S}](t)}{dt} = \vec{f}([\vec{S}](t)) = \mathbf{C} \cdot \vec{v}([\vec{S}](t)). \quad (3)$$

Zmienne $[\vec{S}] = ([S_1], \dots, [S_N])$ odpowiadają reagującym gatunkom, a nieróżniczkowe prawe strony równań $\vec{f}([\vec{S}](t))$ odpowiadają szybkościom reakcji, dając zazwyczaj zbiór wysoce nieliniowych równań. Dokładniej, w równaniu (3), $[\vec{S}](t) \in \mathbb{R}_{\geq 0}^N$ i $\vec{v}([\vec{S}](t)) \in \mathbb{R}_{\geq 0}^M$ są nieujemnymi wektorami odpowiednio stanu stężeń i szybkości reakcji w chwili t , a C to macierz stechiometryczna.

Modele stochastyczne opisują stan systemu w danym punkcie czasu poprzez liczbę cząstek każdego z gatunków. Takie modele są najczęściej reprezentowane przez *chemiczne równanie główne* (ang. chemical master equation; CME), które jest alternatywą do równań Chapmana-Kołomogorowa. CME opisuje ewolucję rozkładu prawdopodobieństwa *procesu Markowa z czasem ciągłym* (ang. continuous-time Markov process; CTMC), nad zbiorem wszystkich osiągalnych stanów \vec{s} (Kampen, 2007):

$$\frac{d\mathbb{P}(t; \vec{s}, \vec{s}_0)}{dt} = \sum_{m=1}^M \left(a_m(\vec{s} - \vec{v}_m) \mathbb{P}(t; \vec{s} - \vec{v}_m, \vec{s}_0) - a_m(\vec{s}) \mathbb{P}(t; \vec{s}, \vec{s}_0) \right), \quad (4)$$

gdzie \vec{s}_0 to rozkład początkowy, a_m to funkcje intensywności reakcji, \vec{v}_m to m -ta kolumna macierzy stechiometrycznej C , a

$$\mathbb{P}(t; \vec{s}, \vec{s}_0) = \mathbb{P} \left(\#\vec{S}(t) = \vec{s} \mid \#\vec{S}(0) = \vec{s}_0 \right), \quad (5)$$

gdzie $\#\vec{S}(t) \in \mathbb{N}^N$ jest nieujemną, wielowymiarową, dyskretną zmienną losową wyrażającą liczbę molekuł wszystkich reagujących gatunków.

4 METODY NUMERYCZNE

Zarówno RRE jak i CME to *równania różniczkowe zwyczajne* (ang. ordinary differential equations; ODE) pierwszego rzędu i zarówno RRE i przede wszystkim CME są analitycznie nierozwiązywalne we wszystkich poza najprostszymi przypadkami (por. Laurenzi, 2000; Jahnke and Huisinga, 2007). Jest tak ze względu na wysoki poziom nieliniowości oraz dużą liczbę zmiennych, gęsto powiązanych ze sobą reakcjami zmiennych lub, odpowiednio, ze względu na rozmiar przestrzeni stanów, która obejmuje wszystkie możliwe układy molekuł w systemie, osiągalne przez zadany zbiór kanałów reakcji. Najprostszym rozwiązaniem jest wykorzystanie metod numerycznych.

Zarówno dla modeli deterministycznych, jak i stochastycznych, dostępne jest wiele metod numerycznych o wysokiej, jak i niskiej precyzji. Podstawowym typem metod są symulacje numeryczne albo równań różniczkowych (zob. np. Butcher, 2008) albo procesu Markowa (tzw. symulacje stochastyczne; zob. np. Gillespie,

2007; Pahle, 2009). Pozwalają one symulować ewolucję w czasie reagentów, stając tym samym numeryczną podstawę do dalszej analizy modelu.

Symulowane trajektorie modelu, po przekształceniu do formy, którą w rzeczywistości obserwuje się w danych, używane są do estymacji parametrów modelu. Powszechnie używaną funkcją w zadaniu dopasowywania modelu do danych jest *błąd kwadratowy* (ang. sum of squared errors; SSE), który prowadzi do optymalizacyjnego problemu najmniejszych kwadratów. Istnieje wiele numerycznych algorytmów do rozwiązywania tego problemu (zob. np. Nocedal and Wright, 1999), jednak ich wadą jest znajdowanie lokalnych minimów, których jest dużo już w przypadku stosunkowo prostych modeli szlaków sygnałowych (Moles et al., 2003; Rodriguez-Fernandez et al., 2006). Z tego względu stosuje się globalne metody Monte Carlo takie jak symulowane wyżarzanie (Kirkpatrick et al., 1983; Černý, 1985).

5 ANALIZA

Ujęcie matematyczne (deterministyczne lub stochastyczne) określa strukturę kinetycznego opisu dla zadanego modelu biochemicznego opisującego system biologiczny. Jak już wcześniej było wspomniane, standardowymi reprezentacjami są RRE dla ujęcia deterministycznego i CME/CTMC dla ujęcia stochastycznego (Wolkenhauer et al., 2004; Aldridge et al., 2006; Goutsias, 2007). Niektóre z powszechnie używanych metod analizy są niezależne od matematycznego modelu. To stwierdzenie obejmuje, dla przykładu, niektóre z metod analizy wrażliwości lub metody selekcji modelu.

Bezpośrednią techniką porównywania modeli jest kryterium *dobroci dopasowania* (ang. goodness of fit; GOF) takie jak wspomniany SSE, tzn.:

$$F(\vec{p}) := \sum_{i=1}^N (f_i(\vec{p}))^2, \quad (6)$$

gdzie ważone residuum to:

$$f_i(\vec{p}) := \frac{Y_i^{\mathcal{M}}(\vec{p}) - Y_i^{\mathcal{D}}}{\sigma_i}, \quad (7)$$

z wartością zmiennej obserwowalnej $Y_i^{\mathcal{M}}(\vec{p})$, pozyskanej z modelu \mathcal{M} dla zadanego zestawu parametrów \vec{p} , i z odpowiadającą jej próbką $Y_i^{\mathcal{D}}$ z danych \mathcal{D} . W przypadku gdy zakłada się, że residua są niezależnymi zmiennymi losowymi o standardowym rozkładzie normalnym, tzn. $Y_i^{\mathcal{M}} \sim \mathcal{N}(Y_i^{\mathcal{D}}, \sigma_i^2)$ niezależnie dla $i = 1, \dots, N$, SSE jako miara GOF daje tzw. *test chi-kwadrat* dla problemu selekcji modeli (χ^2).

Kryterium GOF nie jest dobre do oceny modeli biologicznych ze względu na towarzyszącą im losowość i, co ważniejsze, ze względu na szum w danych (Myung

and Pitt, 2004). Problem polega na tym, że model może uzyskać dobry wynik według kryterium GOF dopasowując się do szumu, zamiast do regularności leżącej u podstaw badanego zjawiska (tzw., ang., overfitting). Dlatego metody selekcji modelu realizują regułę “brzytwy Occama” i karzą za złożoność modelu, promując uogólnienie.

Dobrze znanymi metodami implementującymi zasadę uogólnienia są *kryteria informacyjne Akaike* (ang. Akaike information criterion; AIC; Akaike, 1974) oraz *bayesowskie* (ang. Bayesian information criterion; BIC; Schwarz, 1978), a także najogólniejsza z nich metoda *bayesowskiej selekcji modeli* (ang. Bayesian model selection; BMS) i bezpośrednio z nią powiązany *współczynnik Bayesa* (ang. Bayes factor; BF) — złoty standard porównywania modeli w statystyce Bayesowskiej (Kass and Raftery, 1995; Myung et al., 2009):

$$B_{12} = \frac{\mathbb{P}(\mathcal{D}|\mathcal{M}_1)}{\mathbb{P}(\mathcal{D}|\mathcal{M}_2)}, \quad (8)$$

gdzie

$$\mathbb{P}(\mathcal{D}|\mathcal{M}_i) = \int \mathbb{P}(\mathcal{D}|\mathcal{M}_i, \vec{p}) \cdot \mathbb{P}(\vec{p}|\mathcal{M}_i) d\vec{p}. \quad (9)$$

to brzegowe prawdopodobieństwo pochodzenia danych \mathcal{D} z modelu \mathcal{M}_i , z funkcją gęstości prawdopodobieństwa rozkładu *a priori* na wartościach parametrów $\mathbb{P}(\vec{p}|\mathcal{M}_i)$, oraz

$$\mathbb{P}(\mathcal{D}|\mathcal{M}_i, \vec{p}) = \prod_{i=1}^N \mathbb{P}_i^{\mathcal{M}_i, \vec{p}}(Y_i^{\mathcal{D}}), \quad (10)$$

gdzie $\mathbb{P}_i^{\mathcal{M}_i, \vec{p}}$ to funkcja gęstości prawdopodobieństwa rozkładu normalnego $\mathcal{N}(Y_i^{\mathcal{M}}(\vec{p}), \sigma_i^2)$.

Wadą BF, w przeciwieństwie do AIC lub BIC, jest złożoność obliczeniowa. Bardziej uniwersalnym ograniczeniem wspomnianych miar uogólnienia jest fakt, że wyrażają relację pomiędzy modelem a danymi w postaci pojedynczej liczby (Myung et al., 2009), co powoduje, że nieistotne wyniki selekcji modelu stają się bezużyteczne. W naszym podejściu rozwiązujemy ten problem zastosowaniem analizy wrażliwości do zadania selekcji modelu.

Analiza wrażliwości bada relacje pomiędzy niepewnymi wartościami parametrów modelu a własnością obserwowalnego wyniku jego działania. Matematyczne modele układów reakcji biochemicznych są z natury nieliniowe. Dla takich modeli najodpowiedniejsze są metody *globalnej analizy wrażliwości* (ang. global sensitivity analysis; GSA Saltelli et al., 2005). GSA bada zakres wartości każdego z parametrów jednocześnie, w przeciwieństwie do metod badających wpływ każdego parametru oddzielnie, takich jak te wykorzystujących pochodne względem

parametrów. Przykładową implementacją GSA jest *wieloparametryczna analiza wrażliwości* (ang. multi-parameter sensitivity analysis; MPSA; Young et al., 1978; Hornberger and Spear, 1981). Jest to tzw. filtrująca metoda Monte Carlo, która odwzorowuje przestrzeń parametrów na zadane akceptowalne i nieakceptowalne regiony wartości obserwowalnego wyniku.

W naszym podejściu posługujemy się koncepcją krzepkości (ang. robustness) biologicznych systemów, opartą na analizie wrażliwości. Niska wrażliwość może oznaczać krzepkość, ale również, w zbyt skomplikowanym modelu, nieidentyfikowalność parametrów. W tym celu, uzupełniamy badania o *analizę identyfikowalności* (ang. identifiability analysis; IA) opartą na profilu prawdopodobieństwa (Raue et al., 2009). IA parametrów pozwala wyjaśnić wykonywalność pomiarów eksperymentalnych, które, w efekcie, mogą pozwolić precyzyjnie zidentyfikować poprawną strukturę układu reakcji.

Jako oryginalne rozwiązanie proponujemy rozszerzenie metody MPSA, łącząc ją z techniką *probabilistycznej weryfikacji modelowej* (ang. probabilistic model checking; PMC; Rutten et al., 2004; Kwiatkowska et al., 2007). Jest to technika formalnej weryfikacji systemu, który wykazuje się stochastycznym zachowaniem. W zastosowaniach biologicznych jako model matematyczny wybiera się CTMC (zob. np. Heath et al., 2008; Kwiatkowska et al., 2008), a badaną własność wyraża się w *ciągłej logice stochastycznej* (ang. continuous stochastic logic; CSL; Aziz et al., 1996). Wykorzystujemy również samą probabilistyczną weryfikację modelową (PMC) jako metodę analizy.

W końcu, staramy się realizować ideę “Nauki jako Usługi” (ang. Science as a Service; SaaS; Foster, 2005, 2011). Jest to idea, zgodnie z którą każdy badacz może wygodnie wykonywać skomplikowane analizy na dużych zbiorach danych lub wymagające złożonych obliczeń. Pisząc wygodnie mamy na myśli to, że analiza może być wykonana na przeciętnym komputerze osobistym, w preferowanym miejscu, bez wkładania istotnego wysiłku w aspekty techniczne. Z technicznego punktu widzenia SaaS wpisuje się w koncepcję “chmury obliczeniowej”.

Bardziej precyzyjne hasła charakteryzujące SaaS to: systemy zarządzania przepływami pracy, takie jak TAVERNA WORKBENCH (Hull et al., 2006), które pozwalają na powtarzalne eksperymenty *in silico*; publicznie dostępne *usługi sieciowe* (ang. Web services; WS; The World Wide Web Consortium, 2002), które dostarczają wymaganej funkcjonalności; oraz obliczeniowe środowiska gridowe, które umożliwiają wykorzystanie wyspecjalizowanego, fizycznie rozproszonego sprzętu.

6 PROBLEMY BADAWCZE I WYNIKI

Pierwszym badanym przez nas zagadnieniem jest wybór najbardziej wiarygodnego wariantu aktywacji szlaku sygnałowego JAK-STAT. Rozwiązanie tego zadania selekcji modelu jest oparte wyłącznie na opisie deterministycznym (ODE). Wykorzystujemy tu koncepcję krzepkości opartą na analizie wrażliwości. Następnie poszerzamy zakres matematycznych modeli układów reakcji biochemicznych do jednorodnego w przestrzeni wariantu stochastycznego. W tym celu wykorzystujemy technikę PMC do analizy prostego modelu reakcji enzymatycznej, obrazując ideę analizy wrażliwości specyficznej dla badanej własności. Przy tej okazji zbudowaliśmy wspierające narzędzia programistyczne w ramach projektu TAV4SB. W końcu, wykorzystujemy zarówno deterministyczne jak i stochastyczne metody w studium przypadku modelu odpowiedzi na szok termiczny. Badamy zjawisko termotolerancji i efekt łączonej terapii hipertermią oraz lekami przeciwnowotworowymi. Koncepcyjne związki pomiędzy używanymi modelami oraz metodami podjętymi w rozprawie przedstawia Rysunek 1.1.

Nasze wyniki ilustrują możliwości praktycznego zastosowania oraz potencjał technik takich jak analiza wrażliwości czy probabilistyczna weryfikacja modelowa, oraz zestandaryzowanego, łatwo dostępnego oprogramowania, do analizy kinetycznych modeli molekularnych systemów biologicznych.

6.1 SELEKCJA MODELU JAK-STAT

Rodzina szlaków JAK-STAT jest jednym z głównych mechanizmów sygnalizacji w komórkach eukariotycznych (Aaronson and Horvath, 2002). Ewolucyjnie zachowane role tego mechanizmu obejmują kontrolę nad podstawowymi procesami, takimi jak wzrost komórki czy apoptoza. Deregulacja sygnalizacji poprzez szlak JAK-STAT często jest powiązana z kancerogenezą. Szlaki z rodziny JAK-STAT stają się hiperaktywne w wielu ludzkich nowotworach (Yu and Jove, 2004). Z tego powodu, elementy tych szlaków są atrakcyjnym celem dla leków, których projektowanie wymaga jak najbardziej adekwatnych modeli. Pomimo, że sygnalizacja JAK-STAT jest relatywnie prosta, niejednoznaczności w procesie aktywacji receptorów przeszkadzają w klarownym wyjaśnieniu molekularnego mechanizmu leżącego u podstaw tego procesu.

W rozprawie porównujemy cztery warianty obliczeniowego modelu szlaku sygnałowego JAK1/2-STAT1. Rozważane warianty obejmują znane, podstawowe rozbieżności w mechanizmie aktywacji receptora cytokinowego w kontekście wszystkich kluczowych elementów szlaku. Wykonujemy analizę porównawczą z wykorzystaniem prawa kinetyki mas. Badane różnice są tak marginalne, że wszystkie mode-

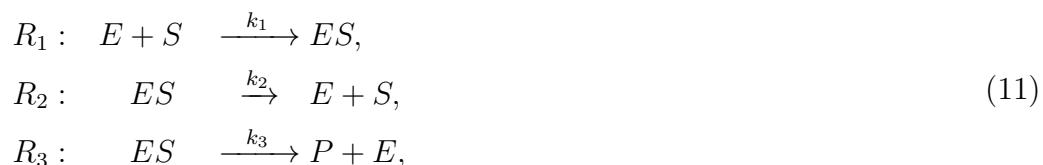
le spełniają kryterium GOF do tego stopnia, że również metoda BMS nie pozwala istotnie wyróżnić jednego modelu. Z tego powodu badamy porównawczo zmiany w krzepkości wariantów szlaku JAK1/2-STAT1 za pomocą GSA, uzupełnionej o IA. Zarówno BMS jak i GSA są użyte do analizy modelu dla różnych wartości parametrów. Odkrywamy, że BMS i GSA, ograniczone do komponentu aktywacji receptora, nieznacznie promują najmniej skomplikowany model. Dalsza, wnikliwa GSA wszystkich parametrów modelu, umotywowana koncepcją krzepkości, pozwala nam wykazać, że dokładna kolejność reakcji wiązania się ligandów do receptorów i reakcji dimeryzacji receptorów, nie jest tak ważna jak uprzednie łączenie się receptorów na membranie przy absencji ligandów.

Główne wyniki studium przypadku szlaku JAK-STAT to:

- zaproponowanie rygorystycznych modeli dla różnych wariantów mechanizmu aktywacji szlaku JAK1/2-STAT1,
- ewaluacja metod selekcji modelu w przypadku analizy *de novo* wielkoskalowego modelu szlaku sygnałowego,
- identyfikacja najbardziej krzepkiej konstrukcji mechanizmu aktywacji receptorów szlaku sygnałowego JAK1/2-STAT1.

6.2 ANALIZA WRAŻLIWOŚCI SPECYFICZNA DLA BADANEJ WŁASNOŚCI

Rozważamy prosty model reakcji enzymatycznej z trzema reakcjami:



gdzie nazwy gatunków S , E , ES oraz P oznaczają odpowiednio: substrat, enzym, kompleks substrat-enzym oraz produkt. Zadanie polega na ocenie tego, która z dwóch reakcji “do przodu”, reakcja formacji kompleksu R_1 czy reakcja konwersji substratu R_3 , ma większy wpływ na zachowanie się systemu. Pokażemy, że w pewnym stopniu odpowiedź zależy od własności, którą opiszemy zachowanie się systemu.

MPSA była stosowana do deterministycznych modeli szlaków sygnałowych z funkcją błędu opartą na trajektoriach zmiennych RRE. W naszym zadaniu, bezpośrednie podejście to ocena ważności parametrów w/w reakcji przy użyciu SSE trajektorii produktu P , dla obecnie analizowanego zestawu wartości parametrów, względem referencyjnego zestawu. W takim ujęciu, parametr k_3 istotnie dominuje nad parametrem k_1 . Jest to oczekiwany wynik.

Teraz, rozważmy następujące pytanie: ile razy średnio musi zajść reakcja asocjacji kompleksu enzym-substrat zanim ilość produktu P osiągnie połowę swojej granicznej wartości? To pytanie można wyrazić w terminach formuły CSL, rozszerzonej w narzędziu PRISM o mechanizm nagród (zob. np. Kwiatkowska et al., 2006):

$$R_{\{\#R_1=?\}} \left(\diamond \left(P > 0.5 \cdot \lim_{t \rightarrow \infty} P(t) \right) \right). \quad (12)$$

Z tak zdefiniowanym zachowaniem się systemu, skupionym na wydajności enzymów, używamy PMC aby obliczyć błąd w MPSA, względem stochastycznego wariantu referencyjnego modelu. W tym ujęciu, można zaobserwować, że parametr k_1 ma większy wpływ na zachowanie systemu niż parametr k_3 , chociaż dominacja nie jest tak silna jak w poprzednim przypadku.

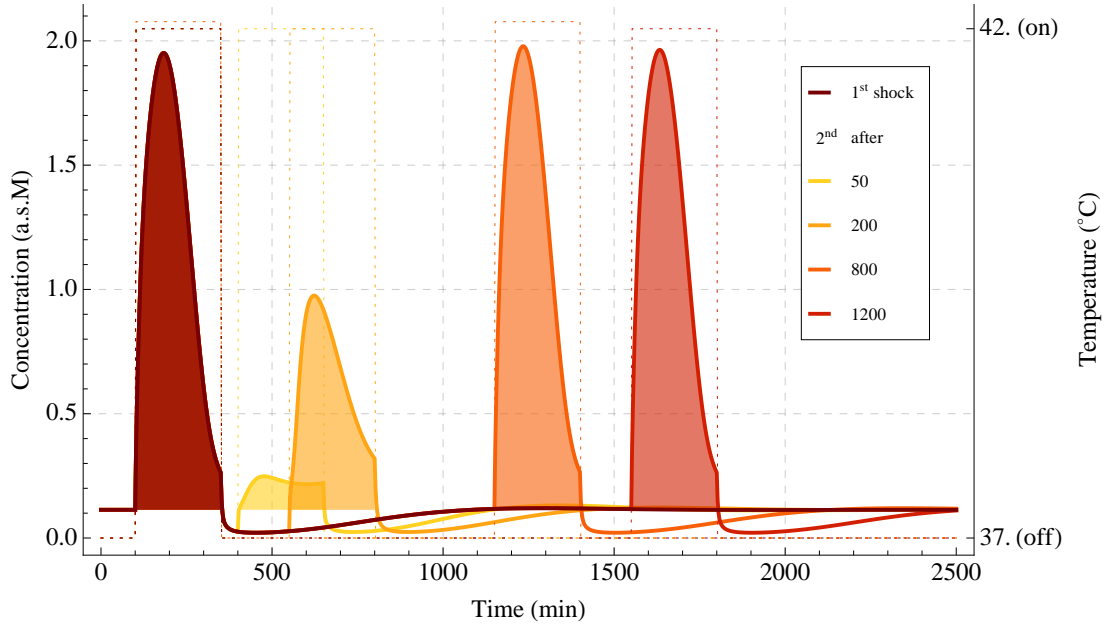
6.3 SKUTECZNOŚĆ LECZENIA HIPERTERMIA

Wielomodalne strategie onkologiczne, które łączą chemioterapię lub radioterapię z hipertermią (tzn. podniesieniem temperatury obszaru ciała dotkniętego nowotworem) mają potencjał poprawienia skuteczności niechirurgicznych metod leczenia nowotworu (Hildebrandt et al., 2002; Neznanov et al., 2011; Wust et al., 2002). Hipertermia angażuje *mechanizm odpowiedzi na szok termiczny* (ang. heat-shock response; HSR), którego główny element (*białka szoku termicznego*; ang. heat-shock proteins; HSP) bezpośrednio zapobiega zamierzonej apoptozie komórek nowotworowych. Co więcej, komórki nowotworowe posiadają już częściowo aktywowany HSR, przez co hipertermia może być dla nich bardziej toksyczna niż dla zwykłych komórek (Neznanov et al., 2011). Jednakże, HSR wywołuje termotolerancję, tzn. komórki potraktowane hipertermią wykazują brak podatności na kolejny szok termiczny (por. Rys. 6.1 Wust et al., 2002). Z tego powodu uważamy, że stosowanie łączonej terapii, opierającej się na hipertermii, wymaga dokładnej analizy.

Dostosowujemy model z pracy Szymańska and Żylicz (2009) i proponujemy jego stochastyczne rozszerzenie, które następnie analizujemy używając *aproksymacyjnych technik PMC* (APMC; Nimal, 2010). Estymujemy globalną funkcję poziomu denaturacji białek:

$$V_{\text{den}}(T) = 1 - \frac{1}{1 + \left(\frac{T - T_0}{T_{0.5} - T_0} \right)^{n_T}}, \quad (13)$$

gdzie T_0 i $T_{0.5}$ to temperatury, dla których $V_{\text{den}}(T)$ wynosi odpowiednio 0 i 0.5 oraz n_T to parametr stromości. Dalej, wyprowadzamy poprawną szybkość denaturacji



Rysunek 6.1: Zjawisko termotolerancji w odpowiedzi na szok termiczny: aktywność zdenaturowanych białek podczas dwóch, kolejnych szoków termicznych (przerywana linia) o temperaturze 5°C ponad poziom homeostazy przy 37°C . Siła zatrucia zdenaturowanymi białkami (pokolorowany obszar) zależy od różnicy w czasie pomiędzy szokami termicznymi. Co ciekawe, aktywność zdenaturowanych białek podczas drugiego szoku może być nawet nieznacznie wyższa niż podczas pierwszego szoku, jak zostało to zilustrowane dla różnicy czasu 800 min.

natywnych białek (cf. [Peper et al., 1998](#)):

$$Q_{\text{den}}^T(t) = m_{11} \times \frac{V_{\text{den}}(T)}{1 - V_{\text{den}}(T)} \times [P_{\text{native}}](t) \quad \frac{\mu\text{M}}{\text{min}}, \quad (14)$$

gdzie m_{11} jest dobrze przybliżane przez stałą kinetyczną reakcji ponownego związania zdenaturowanych białek S . Następnie, uściślamy pojęcie termotolerancji podczas n -tego szoku termicznego ($n > 1$):

$$\mathcal{D}_n = 1 - \frac{\mathcal{R}_n}{\mathcal{R}_1}, \quad (15)$$

gdzie n -ta odpowiedź jest zdefiniowana jako:

$$\mathcal{R}_n = \max_{t_n < t < t_n + \Delta t_n} \{ \#S(t) - \#S^* \}, \quad (16)$$

przy czym $\#S^* = \mathbb{E}_\pi(\#S)$ jest wartością średnią w stanie stacjonarnym π , $[t_n, t_n + \Delta t_n]$ jest przedziałem czasu trwania szoku termicznego, a pierwsza odpowiedź $\mathcal{R}_1 > 0$. Z pomocą tych definicji, obliczamy rozmiar i czas trwania indukowanej przez HSR termotolerancji (por. Rys. 6.1).

W końcu, charakteryzujemy ilościowo efekt łączonej terapii hipertermii i cytotoksycznej inhibicji ponownego zwijania zdenaturowanych białek. Poprzez mechanistyczne modelowanie HSR uzasadniamy powszechne przekonanie, że łączenie różnych niechirurgicznych strategii kuracji nowotworu zwiększa skuteczność terapii.

6.4 PROJEKT TAV4SB

Postęp w modelowaniu systemów biologicznych jest mocno uzależniony od dostępności wyspecjalizowanych, wspieranych komputerowo narzędzi. W tym celu TAVERNA WORKBENCH ułatwia integrację narzędzi programistycznych przeznaczonych dla nauk przyrodniczych oraz dostarcza ujednoliconego, opartego na przepływach pracy, środowiska dla eksperymentów obliczeniowych w biologii.

Projekt *Taverna dla biologii systemów* (ang. Taverna services for Systems Biology; TAV4SB) dostarcza zbioru nowych operacji usługi sieciowej, które poszerzają funkcjonalność TAVERNA WORKBENCH w dziedzinie biologii systemów. Operacje TAV4SB pozwalają wykonywać symulacje numeryczne lub weryfikację modelową, dla odpowiednio deterministycznych lub stochastycznych semantyk modeli biologicznych. Oprócz tej funkcjonalności, TAV4SB umożliwia konstrukcję wysoko-poziomowych eksperymentów. Jako ilustrację możliwości oferowanej przez nasz projekt prezentujemy zastosowanie MPSA. Do wizualizacji wyników analizy modelu dostarczona jest również elastyczna operacja służąca do sporządzania wykresów.

Operacje TAV4SB wykonywane są w prostym środowisku gridowym, które integruje heterogeniczne oprogramowanie, takie jak MATHEMATICA (Wolfram Research, Inc., 2008), PRISM (Hinton et al., 2006) i SBML ODE SOLVER (Machné et al., 2006). Kod źródłowy, instrukcja użytkownika, kontakt, pełna dokumentacja dostępnych operacji usługi sieciowej, przepływy pracy i inne dodatkowe zasoby można znaleźć na stronie internetowej projektu TAV4SB:

<http://bioputer.mimuw.edu.pl/tav4sb/>

7 ARTYKUŁY I WSPÓŁAUTORZY

Większość wyników przedstawionych w rozprawie pochodzi z recenzowanych artykułów. Opis matematycznych modeli i metod analizy wrażliwości można znaleźć w pracy [Charzyńska et al. \(2012\)](#). Studium przypadku dotyczące szlaku JAK-STAT jest obszernie opisane w pracy [Rybiński and Gambin \(2012\)](#), skąd zaczerpnięta jest zdecydowana większość tekstu na ten temat (zob. również [Rybiński and Gambin, 2007, 2009](#)). Opis projektu TAV4SB został pierwotnie opublikowany w pracy [Rybiński et al. \(2011\)](#). Rozszerzona wersja, ze studium przypadku analizy wrażliwości specyficznej dla badanej własności, została przedstawiona w pracy [Rybiński et al. \(2012\)](#), napisanej wraz ze Sławomirem Lasotą i Anną Gambin. Oprogramowanie powstało w ramach wspólnej pracy Mikołaja Rybińskiego z Pawłem Banasikiem ([Banasik, 2008](#)) oraz Michałem Lulą ([Lula, 2009](#)). Nieopublikowane wyniki przedstawione są w oparciu o szkice dwóch publikacji, współtworzonych przez Mikołaja Rybińskiego, Zuzannę Szymańską, Sławomira Lasotę i Annę Gambin. Wyniki te dotyczą mechanizmu odpowiedzi na szok termiczny i jego powiązania z hipertermią oraz termotolerancją.

LITERATURA

- Aaronson, D. S., Horvath, C. M., 2002. A road map for those who don't know JAK-STAT, *Science*, 296, p. 1653.
- Akaike, H., 1974. A new look at the statistical model identification, *IEEE Transactions on Automatic Control*, 19(6), p. 716.
- Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P., 2002. *Molecular Biology of the Cell*, Garland, 4th ed.
- Aldridge, B. B., Burke, J. M., Lauffenburger, D. A., Sorger, P. K., 2006. Physicochemical modelling of cell signalling pathways, *Nature Cell Biology*, 8(11), p. 1195.
- Alon, U., 2007. *An Introduction to Systems Biology: Design Principles of Biological Circuits*, Chapman & Hall/CRC.
- Aziz, A., Sanwal, K., Singhal, V., Brayton, R., 1996. Verifying continuous time markov chains, in: *Proceedings of the 8th International Conference on Computer Aided Verification (CAV'96)*, LNCS, vol. 1102, Springer, p. 269.
- Banasik, P., 2008. Rozszerzenie funkcjonalności programu Taverna dla biologii systemów, Master's thesis, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw.
- Butcher, J. C., 2008. *Numerical Methods for Ordinary Differential Equations*, John Wiley & Sons.
- Charzyńska, A., Nałęcz, A., Rybiński, M., Gambin, A., 2012. Sensitivity analysis of mathematical models of signaling pathways. In press (*BioTechnologia*).
- Foster, I., 2005. Service-oriented science, *Science (New York, N.Y.)*, 308(5723), p. 814.
- Foster, I., 2011. Globus online: Accelerating and democratizing science through Cloud-Based services, *Internet Computing, IEEE*, 15(3), p. 70.
- Gillespie, D. T., 2007. Stochastic simulation of chemical kinetics, *Annual Review of Physical Chemistry*, 58, p. 35.
- Goutsias, J., 2007. Classical versus stochastic kinetics modeling of biochemical reaction systems, *Biophysical Journal*, 92(7), p. 2350.

- Heath, J., Kwiatkowska, M., Norman, G., Parker, D., Tymchyshyn, O., 2008. Probabilistic model checking of complex biological pathways, *Theoretical Computer Science*, 391(3), p. 239.
- Hildebrandt, B., Wust, P., Ahlers, O., Dieing, A., Sreenivasa, G., Kerner, T., Felix, R., Riess, H., 2002. The cellular and molecular basis of hyperthermia, *Critical Reviews in Oncology/Hematology*, 43(1), p. 33.
- Hinton, A., Kwiatkowska, M., Norman, G., Parker, D., 2006. PRISM: A tool for automatic verification of probabilistic systems, in: *Proceedings of the 12th International Conference on Tools and Algorithms for the Construction and Analysis of Systems (TACAS'06)*, *LNCS*, vol. 3920, Springer, p. 441.
- Hornberger, G. M., Spear, R. C., 1981. An approach to the preliminary analysis of environmental systems, *Journal of Environmental Management*, 12(1), p. 7.
- Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P., Oinn, T., 2006. Taverna: a tool for building and running workflows of services, *Nucleic Acids Research*, 34(Web Server issue), p. W729.
- Jahnke, T., Huisinga, W., 2007. Solving the chemical master equation for monomolecular reaction systems analytically, *Journal of Mathematical Biology*, 54(1), p. 1.
- Kampen, N. G. V., 2007. *Stochastic Processes in Physics and Chemistry*, Elsevier.
- Kass, R. E., Raftery, A. E., 1995. Bayes factors, *Journal of the American Statistical Association*, 90(430), p. 773.
- Kirkpatrick, S., Gelatt, J., C D, Vecchi, M. P., 1983. Optimization by simulated annealing, *Science (New York, N.Y.)*, 220(4598), p. 671.
- Klipp, E., Liebermeister, W., Wierling, C., Kowald, A., Lehrach, H., 2009. *Systems Biology: A Textbook*, Wiley-VCH.
- Kwiatkowska, M., Norman, G., Pacheco, A., 2006. Model checking expected time and expected reward formulae with random time bounds, *Computers & Mathematics with Applications*, 51(2), p. 305–.
- Kwiatkowska, M., Norman, G., Parker, D., 2007. Stochastic model checking, in: Bernardo, M., Hillston, J. (Eds.), *Formal Methods for Performance Evaluation*, *LNCS*, vol. 4486, Springer Berlin / Heidelberg, p. 220.

- Kwiatkowska, M., Norman, G., Parker, D., 2008. Using probabilistic model checking in systems biology, *ACM SIGMETRICS Performance Evaluation Review*, 35(4), p. 14.
- Laurenzi, I. J., 2000. An analytical solution of the stochastic master equation for reversible bimolecular reaction kinetics, *The Journal of Chemical Physics*, 113(8), p. 3315.
- Lula, M., 2009. Rozproszone środowisko dla biologii systemów, Master's thesis, Faculty of Mathematics, Informatics and Mechanics, University of Warsaw.
- Lund, E. W., 1965. Guldberg and waage and the law of mass action, *Journal of Chemical Education*, 42(10), p. 548.
- Machné, R., Finney, A., Müller, S., Lu, J., Widder, S., Flamm, C., 2006. The SBML ODE solver library: a native API. for symbolic and fast numerical analysis of reaction networks, *Bioinformatics*, 22, p. 1406.
- Moles, C. G., Mendes, P., Banga, J. R., 2003. Parameter estimation in biochemical pathways: a comparison of global optimization methods, *Genome Research*, 13(11), p. 2467.
- Myung, J. I., Pitt, M. A., 2004. Model comparison methods, *Methods in Enzymology*, 383, p. 351.
- Myung, J. I., Tang, Y., Pitt, M. A., 2009. Evaluation and comparison of computational models, *Methods in Enzymology*, 454, p. 287.
- Neznanov, N., Komarov, A. P., Neznanova, L., Stanhope-Baker, P., Gudkov, A. V., 2011. Proteotoxic stress targeted therapy (PSTT): induction of protein misfolding enhances the antitumor effect of the proteasome inhibitor bortezomib, *Oncotarget*, 2(3), p. 209.
- Nimal, V., 2010. Statistical Approaches for Probabilistic Model Checking, Master's thesis, Oxford University Computing Laboratory.
- Nocedal, J., Wright, S. J., 1999. *Numerical Optimization*, Springer.
- Pahle, J., 2009. Biochemical simulations: stochastic, approximate stochastic and hybrid approaches, *Briefings in Bioinformatics*, 10(1), p. 53.
- Peper, A., Grimbergen, C. A., Spaan, J. A., Souren, J. E., van Wijk, R., 1998. A mathematical model of the hsp70 regulation in the cell, *International Journal of Hyperthermia: The Official Journal of European Society for Hyperthermic Oncology, North American Hyperthermia Group*, 14(1), p. 97.

- Raue, A., Kreutz, C., Maiwald, T., Bachmann, J., Schilling, M., Klingmüller, U., Timmer, J., 2009. Structural and practical identifiability analysis of partially observed dynamical models by exploiting the profile likelihood, *Bioinformatics* (Oxford, England), 25(15), p. 1923.
- Rodriguez-Fernandez, M., Mendes, P., Banga, J. R., 2006. A hybrid approach for efficient and robust parameter estimation in biochemical pathways, *Biosystems*, 83(2-3), p. 248.
- Rutten, J. J. M. M., Kwiatkowska, M., Norman, G., Parker, D., 2004. *Mathematical Techniques for Analyzing Concurrent and Probabilistic Systems*, American Mathematical Society, Centre de Recherches Mathematiques.
- Rybiński, M., Gambin, A., 2007. Details of the JAK-STAT signalling pathway model, *Acta Biochimica Polonica*, supplement, 54, p. 63.
- Rybiński, M., Gambin, A., 2009. Model selection of the JAK-STAT pathway activation mechanism, *Acta Biochimica Polonica*, supplement, 56, p. 58.
- Rybiński, M., Gambin, A., 2012. Model-based selection of the robust JAK-STAT activation mechanism. In press (*Journal of Theoretical Biology*).
- Rybiński, M., Lula, M., Banasik, P., Lasota, S., Gambin, A., 2012. Tav4SB: integrating tools for analysis of kinetic models of biological systems, *BMC Systems Biology*, 6(1), p. 25.
- Rybiński, M., Lula, M., Lasota, S., Gambin, A., 2011. Tav4SB: grid environment for analysis of kinetic models of biological systems, in: *Short Abstracts of the 7th International Symposium on Bioinformatics Research and Applications (IS-BRA'11)*, p. 92.
- Saltelli, A., Ratto, M., Tarantola, S., Campolongo, F., 2005. Sensitivity analysis for chemical models, *Chemical Reviews*, 105, p. 2811.
- Schwarz, G., 1978. Estimating the dimension of a model, *The Annals of Statistics*, 6(2), p. 461.
- Szymańska, Z., Żylicz, M., 2009. Mathematical modeling of heat shock protein synthesis in response to temperature change, *Journal of Theoretical Biology*, 259(3), p. 562.
- The World Wide Web Consortium, 2002. Web services activity, <http://www.w3.org/2002/ws/> (Last accessed: 17 May 2012).

- Černý, V., 1985. Thermodynamical approach to the traveling salesman problem: An efficient simulation algorithm, *Journal of Optimization Theory and Applications*, 45(1), p. 41.
- Wolfram Research, Inc., 2008. *Mathematica Edition: Version 7.0*, Wolfram Research, Inc., Champaign, Illinois.
- Wolkenhauer, O., Ullah, M., Kolch, W., Cho, K., 2004. Modeling and simulation of intracellular dynamics: choosing an appropriate framework, *IEEE Transactions on Nanobioscience*, 3(3), p. 200.
- Wolkenhauer, O., Wellstead, P. E., Cho, K., 2008. *Systems Biology*, Portland Press.
- Wust, P., Hildebrandt, B., Sreenivasa, G., Rau, B., Gellermann, J., Riess, H., Felix, R., Schlag, P. M., 2002. Hyperthermia in combined treatment of cancer, *The Lancet Oncology*, 3(8), p. 487.
- Young, P. C., Spear, R. C., Hornberger, G. M., 1978. Modeling badly defined systems: some further thoughts, in: *Proceedings of the Simulation Special Interest Group Conference (SIMSIG-78)*, Australian National University, Canberra, p. 24.
- Yu, H., Jove, R., 2004. The STATs of cancer—new molecular targets come of age, *Nature reviews. Cancer*, 4, p. 97.