

Małgorzata KARDYŃSKA, Jarosław ŚMIEJA  
Politechnika Śląska

## **ANALIZA WRAŻLIWOŚCI MODELI SZLAKÓW SYGNAŁOWYCH ZBUDOWANYCH NA PODSTAWIE POMIARÓW W DYSKRETNYM CZASIE**

**Streszczenie.** Modele szlaków sygnałowych cechuje wysoki stopień skomplikowania, a występujące w nich parametry często są trudne do wyznaczenia w sposób eksperymentalny. Dlatego podczas ich tworzenia pomocne jest przeprowadzenie analizy wrażliwościowej, która pozwala zbadać odporność układu na zmiany parametrów. Ponadto parametry te powiązane są z konkretnymi procesami biochemicznymi oraz molekułami, dlatego analiza wrażliwości modeli biologicznych może wskazać potencjalne cele molekularne dla nowych leków. Należy jednak pamiętać, że o ile same modele mają zazwyczaj charakter ciągły, to zbudowane są one na podstawie pomiarów w dyskretnym czasie. W niniejszej pracy prezentujemy nową metodę tworzenia rankingów parametrów w oparciu o wyznaczanie Hessianu funkcji błędów definiującej dopasowanie modelu do dyskretnych danych pomiarowych.

## **SENSITIVITY ANALYSIS OF SIGNALING PATHWAY MODELS BASED ON DISCRETE-TIME MEASUREMENTS**

**Summary.** Models of signaling pathways are usually very complex and include parameters that are often difficult to determine in an experimental way. Therefore, when we develop these models sensitivity analysis is a useful tool providing information about their robustness. Moreover, since each parameter is associated with particular biochemical process and molecules, sensitivity analysis may indicate molecular drug targets affecting a given pathway. It should be stressed, however, that even though the models are usually continuous-time, they are based on discrete-time measurements. In this paper we present a new method of creating parameter rankings, based on an Hessian of a cost function which describes the fit of the model to a discrete experimental data.

### **1. Wprowadzenie**

Szlaki sygnałowe (nazywane także ścieżkami sygnałowymi) są kaskadami procesów biochemicznych takich jak produkcja / degradacja białek, aktywacja / dezaktywacja enzymów, transkrypcja genów bądź jej hamowanie. Procesy te mogą być zainicjowane na skutek wymuszenia zewnętrznego (np. związanie ligandu z receptorem zlokalizowanym na powierzchni błony komórkowej) bądź wewnętrznego (np. uszkodzenie DNA), a regulowane są poprzez dodatnie i ujemne pętle sprzężeń zwrotnych. Szlaki sygnałowe

można opisać językiem matematyki na różne sposoby, jednak w tej pracy skupimy się wyłącznie na modelach opisanych za pomocą równań różniczkowych zwyczajnych.

Wraz z rozwojem nowych metod eksperymentalnych nasza wiedza na temat szlaków sygnałowych nieustannie się poszerza, i co za tym idzie, w literaturze pojawiają się nowe modele matematyczne opisujące w coraz dokładniejszy sposób procesy zachodzące w szlakach sygnałowych [12]. Modele te cechuje zwykle wysoki stopień skomplikowania, a występujące w nich parametry często są trudne do wyznaczenia w sposób eksperymentalny i obciążone dużą niepewnością pomiarową [7]. Ponadto należy zdawać sobie sprawę ze specyfiki danych pozyskiwanych w eksperymentach biologicznych. Są to pomiary dyskretne, wykonywane w kilkunasto- kilkudziesięciominutowych odstępach czasu (np. metodą Western blot). Nawet obserwacje przeżyciowe komórek, przedstawione w postaci filmów, w istocie są sekwencją zdjęć wykonywanych w kilkunastominutowych odstępach czasu. Dodatkowo dane tego typu bardzo często są odniesione do pomiarów dla próbki kontrolnej, a nie wyrażone w jednostkach bezwzględnych.

Z tych wszystkich powodów jednym z etapów przy tworzeniu modeli szlaków sygnałowych powinno być przeprowadzenie analizy wrażliwościowej, która pozwala zbadać odporność układu na zmiany parametrów oraz wpływ poszczególnych parametrów na dynamikę układu [10]. Analiza wrażliwości modeli matematycznych nie jest pojęciem nowym, jednak nie zawsze metody stosowane w automatyce można z powodzeniem wykorzystać do analizy modeli biologicznych [11]. Dlatego też konieczne jest opracowanie nowych metod analizy wrażliwości układów biologicznych, które uwzględniać będą specyficzny charakter danych eksperymentalnych, na których oparty został model.

W tej pracy prezentujemy nową metodę tworzenia rankingów parametrów w oparciu o wyznaczanie Hesjanu funkcji błędu opisującej dopasowanie modelu do dyskretnych danych pomiarowych [2]. Rankingi te pozwalają zidentyfikować najważniejsze parametry modelu, które powinny zostać wyznaczone z największą dokładnością, ale mogą też stanowić np. podpowiedź przy poszukiwaniu celów molekularnych dla nowych leków [9]. W celu lepszego zrozumienia metoda została przetestowana w pierwszej kolejności na prostym modelu układu dynamicznego a następnie na wybranym modelu szlaku sygnałowego.

## 2. Metodyka

Analizę wrażliwości modelu na zmiany parametrów można najogólniej podzielić na analizę lokalną i globalną. Metody lokalnej analizy wrażliwości badają wpływ niewielkich odchyłek wartości parametrów na odpowiedź modelu, podczas gdy w metodach globalnych wartości parametrów zmieniane są w znacznie szerszym zakresie [6], [5]. Biorąc pod uwagę ten podział metodę zaprezentowaną w tym tekście można zaklasyfikować jako metodę lokalną.

Niech analizowany model będzie opisany równaniem różniczkowym zwyczajnym:

$$\frac{d\mathbf{X}}{dt} = F(\mathbf{X}, \mathbf{U}, \theta), \quad \mathbf{X}(t_0) = \mathbf{X}_0, \quad (1)$$

gdzie  $\mathbf{X}$  jest wektorem zmiennych stanu, reprezentującym najczęściej stężenia bądź liczby molekuł białek, enzymów czy transkryptów uczestniczących w danym szlaku sygnałowym,  $\mathbf{U}$  jest wektorem sygnałów wejściowych, który zazwyczaj w układach

biologicznych jest skalar,  $\theta$  jest wektorem parametrów modelu, natomiast  $\mathbf{X}_0$  jest wektorem warunków początkowych. Rozwiązanie tak zdefiniowanego modelu będzie określone przez:

$$\mathbf{X}(\theta, t). \quad (2)$$

Dopasowanie zmiennej  $x_s(\theta, t) \in \mathbf{X}(\theta, t)$  do danych eksperymentalnych można opisać nieliniową funkcją kosztów [2]:

$$C_s(\theta) = \sum_n \frac{1}{2} \frac{(x_s(\theta, t) - x_{s,n})^2}{\sigma_{s,n}^2} = \sum_n \frac{1}{2} r_{s,n}^2, \quad (3)$$

gdzie  $x_{s,n}$  jest stężeniem lub liczbą molekuł  $s$  w  $n$ -tej próbce zmierzonej z niepewnością pomiarową równą  $\sigma_{s,n}$ ,  $x_s(\theta, t)$  jest rozwiązaniem modelu w chwili  $t$  równej chwili pobrania  $n$ -tej próbki, podczas gdy  $r_{s,i}$  jest błędem modelu w stosunku do pomiarów molekuł  $s$ . W przypadku gdy model idealnie odzwierciedla dane pomiarowe funkcja błędu  $C_s$  będzie przyjmować wartość równą 0, a zestaw parametrów  $\theta$ , dla których otrzymujemy takie rozwiązanie, oznaczać będziemy  $\theta^*$ .

Aby określić wrażliwość modelu na zmiany parametrów wyznaczamy Hesjan funkcji kosztów  $C_s$  w punkcie  $\theta^*$ . Ponieważ parametry biochemiczne często mają różne jednostki, a ich wartości mogą się różnić między sobą nawet o kilka rzędów wielkości, Hesjan wyznaczamy dla  $\log \theta$ , co pozwala zniwelować wpływ wartości parametrów [3]:

$$H_{j,k}^{C_s} = \frac{\partial^2 C_s}{\partial \log \theta_j \partial \log \theta_k}, \quad (4)$$

gdzie  $j$  i  $k$  oznaczają odpowiednio  $j$ -ty i  $k$ -ty parametr.

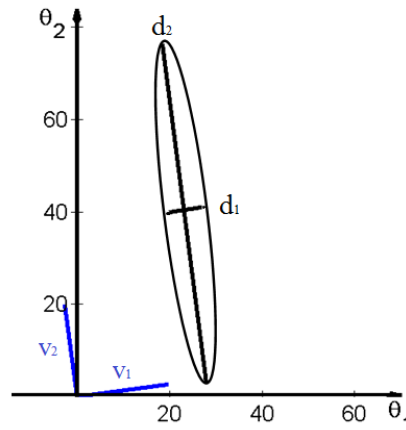
Alternatywnie można posłużyć się przybliżeniem Hesjanu  $H^{C_s}$  - macierzą informacyjną Fishera  $J^T J$  [13], gdzie:

$$J = \frac{\partial r_{s,n}}{\partial \log \theta}. \quad (5)$$

Hesjan funkcji kosztów  $H^{C_s}$  jest macierzą dodatnio określoną i symetryczną, a więc posiada rzeczywiste wartości własne  $\lambda$  oraz wektory własne  $v$  [8]. Hesjan ten opisuje pewną wielowymiarową płaszczyznę obrazującą odchyłki modelu od danych pomiarowych. Jeżeli w modelu występuje  $N_p$  parametrów, płaszczyzna ta będzie  $N_p$ -wymiarową elipsoidą umieszczoną w przestrzeni parametrów, przy czym kąt nachylenia osi tej elipsoidy względem osi parametrów jest zdefiniowany przez wektory własne Hesjanu  $H^{C_s}$ , natomiast długość tych osi (oznaczona jako  $d_i$ ) jest proporcjonalna do:

$$d_i = \frac{1}{\sqrt{\lambda_i}}. \quad (6)$$

Przyjmuje się, że najkrótsze osie elipsoidy nazywane są "stiff" i wyznaczają one kierunki zmian parametrów, które skutkować będą dużymi zmianami w odpowiedzi modelu, podczas gdy najdłuższe osie, nazywane "sloppy", reprezentują kierunki wzdłuż których zmiana parametrów, nawet w szerokim zakresie, nie będzie powodować gorszego dopasowania modelu do danych eksperymentalnych [1]. Z tego też powodu metodę nazywa się często analizą wrażliwości sloppy/stiff. Znaczenie wartości własnych i wektorów własnych Hesjanu  $H^{C_s}$  zobrazowano na przykładzie hipotetycznego modelu o dwóch



Rys. 1. Przykład elipsy obrazującej błąd odpowiedzi modelu względem danych eksperymentalnych dla protego modelu o dwóch parametrach:  $\theta_1$  i  $\theta_2$ . Przez  $d_1$  i  $d_2$  oznaczono długości dwóch osi elipsy, które są proporcjonalne do wartości własnych Heszjanu  $H^{Cs}$  - odpowiednio  $\lambda_1$  i  $\lambda_2$ , podczas gdy  $v_1$  i  $v_2$  reprezentują wektory własne Heszjanu  $H^{Cs}$ .

parametrach:  $\theta_1$  i  $\theta_2$  (rys. 1). W tym przypadku Heszjan obrazuje elipsę w przestrzeni parametrów  $\theta_1/\theta_2$ .

Należy jednak pamiętać, że modele szlaków sygnałowych są znacznie bardziej złożone i nierzadko występuje w nich po kilkadziesiąt parametrów. Dlatego graficzna prezentacja elipsoidy obrazującej błąd odpowiedzi modelu względem danych eksperymentalnych nie jest możliwa. W takim przypadku znacznie bardziej czytelnym sposobem prezentacji wyników analizy wrażliwościowej są rankingi parametrów. W niniejszej pracy prezentujemy własny sposób tworzenia rankingów parametrów w oparciu o wartości własne i wektory własne Heszjanu  $H^{Cs}$ , w którym znaczenie  $j$ -tego parametru można zdefiniować jako:

$$R_j = \sum_i \left| \frac{v_{j,i}}{d_i} \right|, \quad (7)$$

gdzie  $d_i$  jest długością elipsoidy wzdłuż  $i$ -tej osi, a  $v_{j,i}$  jest elementem  $i$ -tego wektora własnego powiązanego z  $j$ -tym parametrem.

### 3. Wyniki

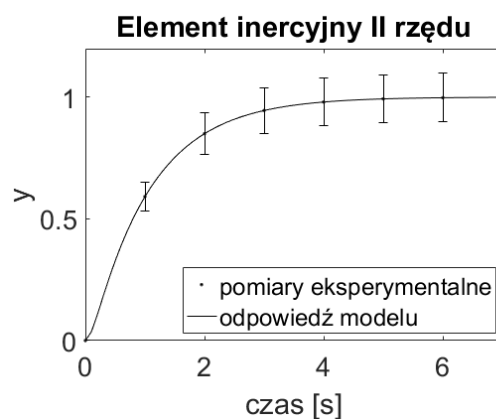
W tej sekcji przedstawione zostaną wyniki analizy wrażliwości dwóch przykładowych modeli: w pierwszej kolejności dla prostego układu dynamicznego (inercji II rzędu), w celu lepszego zrozumienia opisanego metody, a następnie dla modelu szlaku sygnałowego p53. Rankingi parametrów porównane zostaną ze standardowymi rankingami stworzonymi na podstawie pól powierzchni pod funkcjami wrażliwości.

#### 3.1. Element inercyjny II rzędu

Rozważamy element inercyjny II rzędu opisany transmitancją:

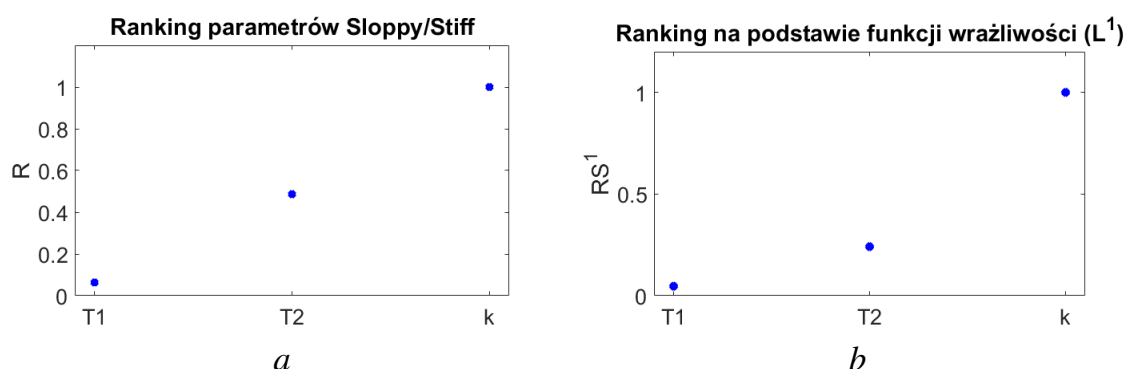
$$K(s) = \frac{X(s)}{U(s)} = \frac{k}{(1 + sT_1)(1 + sT_2)}, \quad (8)$$

o parametrach  $T_1 = 0, 1$ ,  $T_2 = 1$  oraz  $k = 1$ . Model ten oparto na danych pomiarowych wygenerowanych na potrzeby tego przykładu, przy czym założono, że dane te charakteryzują się niepewnością  $\sigma$  na poziomie 10% wartości pomiaru. Zarówno pomiary jak i odpowiedź modelu o zadanych parametrach przedstawiono na rysunku 2.



Rys. 2. Odpowiedź modelu dla elementu inercyjnego II rzędu o parametrach  $T_1 = 0, 1$ ,  $T_2 = 1$  oraz  $k = 1$  na tle danych pomiarowych.

Postępując zgodnie z procedurą opisaną w poprzednim rozdziale przeprowadzono analizę wrażliwości tego modelu, a wyniki przedstawiono na rysunku 3.



Rys. 3. Rankingi parametrów dla elementu inercyjnego II rzędu otrzymane z wykorzystaniem: metody sloppy/stiff (a) oraz funkcji wrażliwości (b).

Przedstawione rankingi parametrów są do siebie podobne, jednak można zaobserwować pewne różnice dla parametru  $T_2$  (parametr ten decyduje o szybkości z jaką układ dochodzi do stanu ustalonego). Różnice te wynikają z uwzględnienia niepewności pomiarowej w funkcji kosztów  $C$  (równanie 3). O ile sama niepewność pomiarowa jest stała, to analizując rysunek 2 można zauważyć, że wariancja pomiarów wykonanych w czasach 1s i 2s (przed osiągnięciem stanu ustalonego) jest mniejsza niż pozostałych pomiarów. W efekcie odchyłki odpowiedzi modelu w tych punktach będą traktowane jako bardziej istotne niż w punktach o dużej wariancji - stąd wyższa pozycja parametru  $T_2$  w rankingu parametrów otrzymanym naszą metodą.

### 3.2. Model szlaku sygnałowego p53

Analizowany model szlaku sygnałowego p53 opisany jest za pomocą trzech równań różniczkowych zwyczajnych [4]:

$$\frac{d(p53)}{dt} = m s_1 - k_{d1} \cdot (p53) \cdot (Mdm2_{nuc})^2, \quad (9)$$

$$\frac{d(Mdm2_{cyt})}{dt} = n \cdot \left( s_2 + \frac{s_3 \cdot (p53)^3}{s_4^3 + (p53)^3} \right) - k_1 k_2 \cdot \frac{Mdm2_{cyt}}{k_2 + (p53)}, \quad (10)$$

$$\frac{d(Mdm2_{nuc})}{dt} = \frac{k_1 k_2 \cdot (Mdm2_{cyt})}{k_2 + (p53)} - k_{d2} \cdot (Mdm2_{nuc}), \quad (11)$$

gdzie kolejne zmienne opisują liczby cząsteczek białek p53 (całkowitego), Mdm2 w cytoplazmie i Mdm2 w jądrze komórkowym. Jest to minimalny model oddający oscylacyjny charakter odpowiedzi białka p53 na wymuszenia takie jak uszkodzenia DNA w komórce. Ze względu na bardzo duże uproszczenia w literaturze modele takie określane są często mianem "toy models". W modelu tym parametry  $m$  i  $n$  oznaczają liczbę kopii genów p53 i Mdm2,  $s_1$ ,  $s_2$ ,  $s_3$  i  $s_4$  są parametrami określającymi szybkość produkcji poszczególnych białek w przeliczeniu na kopię genu,  $k_{d1}$  i  $k_{d2}$  są szybkościami degradacji odpowiednio dla p53 i Mdm2, a  $k_1$  i  $k_2$  określają szybkość importu białka Mdm2 do jądra komórkowego. Wartości poszczególnych parametrów zostały zamieszczone w tabeli 1.

Tabela 1

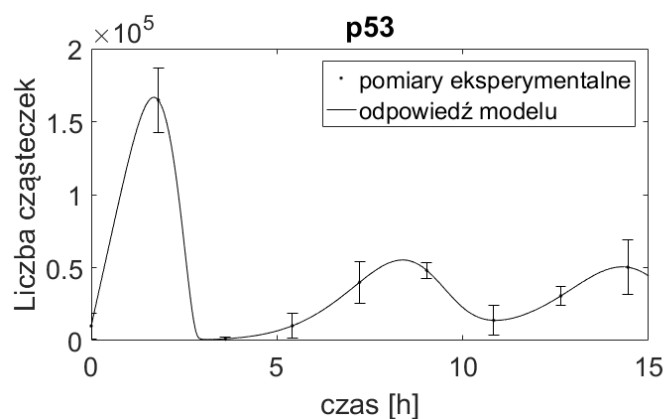
Parametry modelu p53/Mdm2.

|           |                     |                     |                  |                |                    |
|-----------|---------------------|---------------------|------------------|----------------|--------------------|
| Parametr  | $s_1$               | $s_2$               | $s_3$            | $s_4$          | $k_{d1}$           |
| Wartość   | 16                  | 8                   | 80               | $1 \cdot 10^5$ | $1 \cdot 10^{-13}$ |
| Jednostka | $s^{-1}$            | $s^{-1}$            | $s^{-1}$         | —              | $s^{-1}$           |
| Parametr  | $k_{d2}$            | $k_1$               | $k_2$            | $m$            | $n$                |
| Wartość   | $2.2 \cdot 10^{-4}$ | $3.5 \cdot 10^{-3}$ | $2.3 \cdot 10^3$ | 2              | 2                  |
| Jednostka | $s^{-1}$            | $s^{-1}$            | —                | —              | —                  |

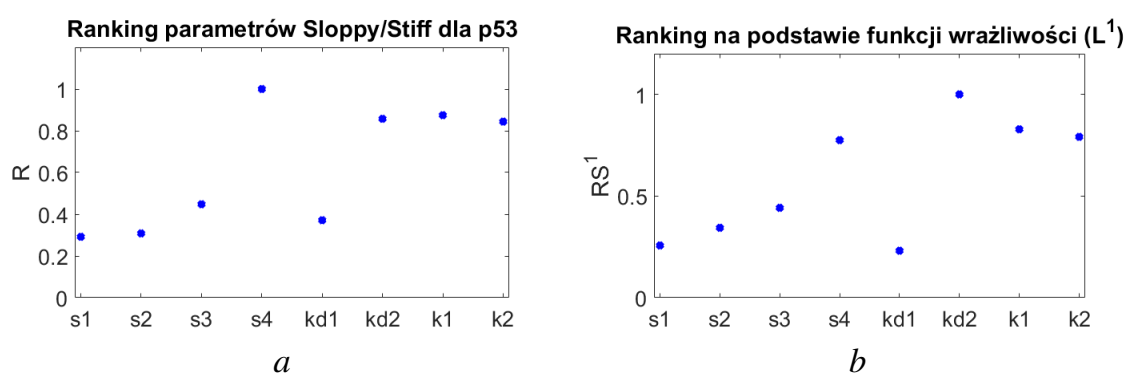
W tej pracy skupimy się wyłącznie na zmiennej p53 i jej dopasowaniu do danych pomiarowych. Cechą charakterystyczną pomiarów biologicznych jest ich duża niepewność pomiarowa, przy czym na niepewność tą wpływa szereg różnych czynników. W efekcie wariancja pomiarów w kolejnych punktach pomiarowych może się znacznie różnić, co pokazano na rysunku 4 (zaznaczone punkty pomiarowe służą jedynie ilustracji - wartości wyjścia zostały wybrane na podstawie symulacji, a odchylenie standardowe jest arbitralnie przyjęte).

Przeprowadzono analizę wrażliwości modelu p53/Mdm2 korzystając z metody funkcji wrażliwości oraz metody opisanej w poprzednim rozdziale. Wyniki przedstawiono na rysunku 5.

Podobnie jak miało to miejsce w przypadku elementu inercyjnego, obserwujemy pewne różnice w otrzymanych rankingach. Ranking parametrów otrzymany na podstawie funkcji wrażliwości wskazuje na duży wpływ parametru  $k_{d2}$  na dynamikę modelu. Jest to parametr wpływający przede wszystkim na amplitudę oscylacji p53. Analizując



Rys. 4. Odpowiedź modelu dla zmiennej  $p53$  na tle danych pomiarowych.



Rys. 5. Rankingi parametrów dla zmiennej  $p53$  otrzymane z wykorzystaniem: metody sloppy/stiff (a) oraz funkcji wrażliwości (b).

jednak rysunek 4 stwierdzamy, że ze względu na dużą niepewność pomiarową amplituda ta nie jest dokładnie poznana. W związku z tym w rankingu uwzględniającym wpływ niepewności pomiarowej pozycja tego parametru jest niższa. Podobnie można tłumaczyć różnice dla innych parametrów.

#### 4. Wnioski

W pracy pokazano nową metodę tworzenia rankingów parametrów, opracowaną w oparciu o znaną metodę analizy wrażliwości parametrów sloppy/stiff. W metodzie tej uwzględniono wpływ niepewności pomiarowej, która stanowi istotny problem podczas analizy wszelkiego rodzaju biologicznych danych eksperymentalnych.

Na przykładzie dwóch prostych modeli pokazano, że zaprezentowana metoda umożliwia stworzenie wiarygodnych rankingów parametrów, które pozwalają zidentyfikować parametry najistotniej wpływające na jakość dopasowania modelu do danych eksperymentalnych. Pozwala to wytypować parametry, które powinny być wyznaczone z największą dokładnością w badaniach eksperymentalnych. Dzięki temu metoda ta może zostać wykorzystana do planowania kolejnych eksperymentów biologicznych i w konsekwencji umożliwia poprawę efektywności wykorzystania środków przeznaczonych na badania eksperymentalne.

## 5. Podziękowania

Praca została sfinansowana z BKM/506/RAU1/2016/5. Do obliczeń wykorzystano klaster Ziemowit (<http://www.ziemowit.hpc.polsl.pl>), stworzony w ramach projektu POIG.02.01.00-00-166/08 (BIO-FARMA) i rozbudowany w projekcie POIG.02.03.01-00-040/13 (Syscancer).

## LITERATURA

1. Brown K.S., Sethna J.P.: Statistical mechanical approaches to models with many poorly known parameters, *Phys Rev E* 68, 2003, p. 021904.
2. Daniels B.C., Chen Y.J., Sethna J.P., Gutenkunst R.N., Myers C.R.: Sloppiness, robustness, and evolvability in systems biology, *Current Opinion in Biotechnology* 19, 2008, p. 389–395.
3. Gutenkunst R.N., Waterfall J.J., Casey F.P., Brown K.S., Myers C.R., Sethna J.P.: Universally sloppy parameter sensitivities in systems biology models, *PLoS Comp. Biol.* 3(10), 2007, p. 1871–1878.
4. Hat B., Puszyński K., Lipniacki T.: Exploring mechanisms of oscillations in p53 and nuclear factor- $\kappa$ B systems, *IET Systems Biology* 3, 2009, p. 342–355.
5. Iglesias P., Ingalls B. (editors): *Control theory and systems biology*, MIT Press, 2010.
6. Leis J., Kramer M.: Sensitivity analysis of systems of differential and algebraic equations, *Computers & Chemical Engineering* 9, 1985, p. 93–96.
7. Maerkl S.J., Quake S.R.: A systems approach to measuring the binding energy landscapes of transcription factors, *Science* 315, 2007, p. 233–237.
8. Mannakee B.K., Ragsdale A.P., Transtrum M.K., Gutenkunst R.N.: Sloppiness and the Geometry of Parameter Space, In: Geris L, Gomez-Cabrero D, editors. *Uncertainty in Biology, Studies in Mechanobiology, Tissue Engineering and Biomaterials* 17. Switzerland: Springer; 2016. p. 271–298.
9. Marin-Sanguino A., Gupta S.K., Voit E.O., Vera J.: Biochemical pathway modeling tools for drug target detection in cancer and other complex diseases, *Methods in Enzymology* 487, 2011, p. 319–369.
10. Rand D.A.: Mapping the global sensitivity of cellular network dynamics, *J. Royal Society Interface* 5, 2008, p. 59–69.
11. Śmieja J., Kardynańska M., Jamroz A.: The meaning of sensitivity functions in signaling pathways analysis, *DCDS-series B* 10, 2014, p. 2697–2707.
12. Williams R, Timmis J, Qvarnstrom EE.: Computational models of the NF- $\kappa$ B signalling pathway, *Computation* 2(4), 2014, p. 131–158.
13. Waterfall J.J., Casey F.P., Gutenkunst R.N., Brown K.S., Myers C.R., Brouwer P.W., et al.: Sloppy-Model Universality Class and the Vandermonde Matrix, *Phys. Rev. Lett.* 97(15), 2006, p. 150601.