AUTOREFERAT

przedstawiający opis dorobku i osiągnięć naukowych,
w szczególności określonych w art. 16 ust. 2 ustawy

dr inż. Marta Szachniuk

Instytut Informatyki
Wydział Informatyki
Politechnika Poznańska
oraz
Zakład Bioinformatyki
Instytut Chemii Bioorganicznej
Polskiej Akademii Nauk w Poznaniu

Poznań, 10 marca 2015

[1] Imię i nazwisko

Marta Szachniuk

[2] Posiadane stopnie i tytuły naukowe

1. Tytuł magistra inżyniera informatyki uzyskany w 1998 roku na Wydziale Elektrycznym Politechniki Poznańskiej (tytuł pracy dyplomowej magisterskiej: "Automatyczne rozpoznawanie pisma – analiza metod"; promotor: dr inż. Jan Nikisch)
2. Tytuł magistra matematyki uzyskany w 1999 roku na Wydziale Budowy Maszyn i Zarządzania Politechniki Poznańskiej (tytuł pracy dyplomowej magisterskiej: "Pewne problemy topologii we współczesnym ujęciu"; promotor: dr Edward Ambrożko)
3. Stopień doktora nauk technicznych w zakresie informatyki uzyskany w 2005 roku na Wydziale Informatyki i Zarządzania Politechniki Poznańskiej (tytuł rozprawy doktorskiej: "Kombinatoryczna analiza widm 2D-NOESY w spektroskopii Magnetycznego Rezonansu Jądrowego cząsteczek RNA"; promotor: prof. dr hab. inż. Jacek Błażewicz)

[3] Dotychczasowe zatrudnienie w jednostkach naukowych

- 08.09.2003 – 31.08.2005: asystent, Pracownia Bioinformatyki, Instytut Chemii Bioorganicznej Polskiej Akademii Nauk w Poznaniu
- 01.09.2005 – 31.10.2013: adiunkt, Zakład Bioinformatyki, Instytut Chemii Bioorganicznej Polskiej Akademii Nauk w Poznaniu
- 01.10.2005 – 30.09.2008: asystent naukowy, Instytut Informatyki, Politechnika Poznańska
- 01.04.2010 – 30.04.2012: specjalista do spraw administracyjnych, Instytut Informatyki, Politechnika Poznańska
- od 01.10.2010: adiunkt, Instytut Informatyki, Politechnika Poznańska
- od 01.11.2013: starszy specjalista, Zakład Bioinformatyki, Instytut Chemii Bioorganicznej PAN w Poznaniu
- 11-17.10.2010, 22-27.10.2011, 26-30.11.2012: naukowiec wizytujący, Instituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti", Consiglio Nazionale delle Ricerche, Rzym, Włochy
- 04-10.07.2011, 22-28.07.2011, 24-29.11.2013: naukowiec wizytujący, CSIRO Mathematics, Informatics and Statistics Department, The Commonwealth Scientific and Industrial Research Organisation, Clayton, Australia

[4] Wskazanie osiągnięcia naukowego wynikającego z art. 16 ust. 2 ustawy z dnia 14 marca 2001 r. o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki (Dz. U. nr 65, poz. 595 ze zm.)

1. Tytuł osiągnięcia naukowego

   Informatyczne aspekty opartej na NMR analizie struktur RNA.

2. Lista prac wchodzących w skład osiągnięcia naukowego

   [P1] M. Szachniuk. Assigning NMR Spectra of Irregular RNAs by Heuristic Algorithms, *Bulletin of the Polish Academy of Sciences Technical Sciences* 63(1), 2015, *w druku*.

   [P2] M. Szachniuk, M.C. De Cola, G. Felici, D. de Werra, J. Blazewicz. Optimal pathway reconstruction on 3D NMR maps, *Discrete Applied Mathematics* 182, 2015, 134-149.

   [P3] M. Szachniuk, M.C. De Cola, G. Felici, J. Blazewicz. The Orderly Colored Longest Path Problem - a survey of applications and new algorithms, *RAIRO Operations Research* 48(1), 2014, 25-51.

   [P4] M. Szachniuk, M. Malaczynski, E. Pesch, E.K. Burke, J. Blazewicz. MLP accompanied beam search for the resonance assignment problem, *Journal of Heuristics* 19(3), 2013, 443-464.

   [P5] M. Popenda, M. Szachniuk, M. Blazewicz, S. Wasik, E.K. Burke, J. Blazewicz, R.W. Adamiak. RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures, *BMC Bioinformatics* 11, 2010, 231.

   [P6] M. Szachniuk, M. Popenda, R.W. Adamiak, J. Blazewicz. An assignment walk through 3D NMR spectrum, *Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2009, 215-219.

   [P7] M. Popenda, M. Blazewicz, M. Szachniuk, R.W. Adamiak. RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures, *Nucleic Acids Research* 36(1), 2008, D386-D391.

3. Omówienie celu osiągnięcia naukowego w/w prac oraz osiągniętych wyników wraz z omówieniem ich ewentualnego wykorzystania

   3.1. Wprowadzenie

   Od wielu lat obserwuje się rosnące znaczenie informatyki we wspomaganiu badań eksperymentalnych. We współczesnej biologii, chemii czy fizyce, praca z komputerowymi modelami problemów często zastępuje eksperyment tradycyjny lub wspomaga jego przebieg. Pozwala to na znaczne obniżanie kosztów eksperymentu i umożliwia testowanie najśmielszych nawet pomysłów. Przedstawione tutaj badania, prowadzone w dziedzinie informatyki, zostały zainspirowane problemami współczesnej bioinformatyki i biologii struktur RNA. Uzyskane wyniki znalazły zastosowanie między innymi w tych obszarach nauki.

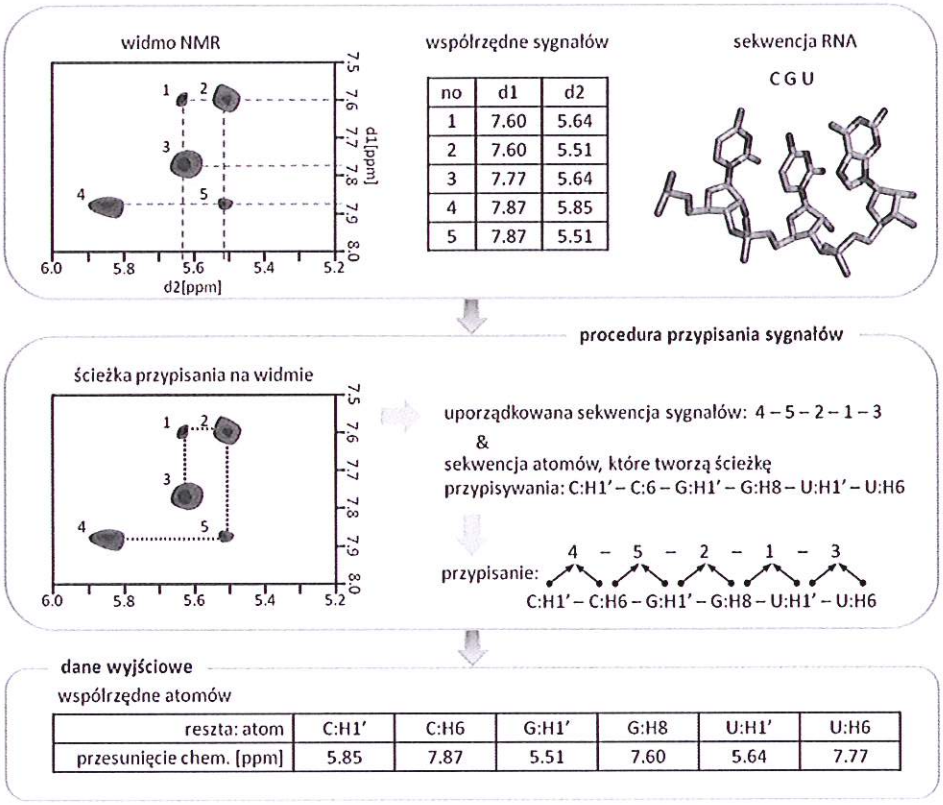
Jednym z najważniejszych zagadnień współczesnej biologii, biomedycyny, biochemii czy biotechnologii jest rozpoznawanie budowy organizmów żywych na poziomie molekularnym i atomowym. Poznanie struktur białek i kwasów nukleinowych (DNA, RNA) nie tylko przyczynia się do wyjaśnienia ich biologicznych funkcji, ale staje się także przyczynkiem do projektowania nowych leków, identyfikowania chorób, otrzymywania nowych odmian hodowlanych zwierząt i roślin, itp. Pełne trójwymiarowe struktury cząsteczek określane są za pomocą rentgenografii strukturalnej lub spektroskopii magnetycznego rezonansu jądrowego (ang. NMR - Nuclear Magnetic Resonance). Ta ostatnia jest jedyną metodą, która umożliwia analizę struktury przestrzennej kwasów nukleinowych i białek w roztworach wodnych w warunkach zbliżonych do fizjologicznych.

Metodyka analizy strukturalnej za pomocą spektroskopii magnetycznego rezonansu jądrowego ma wiele aspektów. Szereg eksperymentów NMR, zarówno homo- jak i heterojądrowych, których wyniki obrazowane są na jedno- lub wielowymiarowych widmach pozwala na pomiar parametrów, które są następnie podstawą do zrekonstruowania struktury badanego związku. Podstawowym elementem analizy jest identyfikacja sygnałów rezonansowych, którą wykonuje się w początkowym etapie wyznaczania struktury przestrzennej badanego związku za pomocą NMR.

W ramach prezentowanych badań: (i) zaprojektowano nową metodę identyfikacji sygnałów rezonansowych i zastosowano ją w analizie dwuwymiarowych widm NMR regularnych struktur RNA [P4], (ii) przeanalizowano problem identyfikacji sygnałów na dwuwymiarowych widmach struktur nieregularnych, sformułowano inspirowany tym zagadnieniem problem poszukiwania rozłącznych ścieżek NOE w grafie NOESY i zastosowano dwa algorytmy heurystyczne z modyfikacjami do analizy przykładowych grafów reprezentujących widma struktur nieregularnych [P1], (iii) opracowano model grafu spektralnego reprezentujący problem przypisywania sygnałów na widmach trójwymiarowych i zdefiniowano problem poszukiwania ścieżki transferowej w tym grafie [P2], (iv) przeprowadzono dowody NP-trudności problemu konstruowania ścieżek przypisywania na homo- i heterojądrowym trójwymiarowym widmie NMR [P2], (v) zdefiniowano problem poszukiwania ścieżki o uporządkowanej sekwencji kolorów w grafie z k-pokolorowanymi krawędziami [P3], (vi) przeanalizowano nowo opracowany model grafowy i możliwości jego zastosowania do różnych problemów obliczeniowych [P3], (vii) zaproponowano modele oparte na sieciach przepływowych i algorytmy do analizy trójwymiarowych widm NMR [P2, P6], (viii) opracowano model bazy danych do przechowywania danych o rozwiązanych eksperymentalnie strukturach RNA oraz wyszukiwarkę fragmentów strukturalnych o zadanym wzorcu sekwencyjnym i drugorzędowym [P5, P7]. W/w badania były kontynuacją studiów prowadzonych przed uzyskaniem stopnia doktora, w ramach których zaproponowano model grafu NOESY reprezentujący problem konstruowania ścieżki przypisywania na dwuwymiarowym widmie struktury regularnej, przeanalizowano złożoność obliczeniową tego problemu oraz opracowano pierwsze algorytmy do jego rozwiązania i zastosowano je dla przykładowych danych spektralnych.

3.2. Identyfikacja sygnałów rezonansowych na widmach NMR

Celem identyfikacji sygnałów rezonansowych jest określenie które atomy badanej cząsteczki wygenerowały dany sygnał zarejestrowany na widmie podczas eksperymentu NMR. Zatem identyfikacja sygnałów polega na przypisaniu tych sygnałów do odpowiednich atomów badanej cząsteczki. Jeden sygnał może być przypisany do pary, trójki, czwórki, itd. atomów, zależnie od tego ile wymiarów ma analizowane widmo NMR. Proces przypisywania rozpoczyna się od wyznaczenia na widmach NMR ścieżek, które łączą odpowiednie sygnały i obrazują drogi przenoszenia magnetyzacji w łańcuchach RNA. Na przykład na widmach dwuwymiarowych tzw. ścieżka NOE łączy się sygnały pomiędzy atomami H1' a atomami H6/H8 reszt pirymidynowych i purynowych wzdłuż łańcucha RNA. Rysunek 1 pokazuje ogólną ideę przypisywania sygnałów poprzez konstrukcję ścieżki NOE.



no	d1	d2
1	7.60	5.64
2	7.60	5.51
3	7.77	5.64
4	7.87	5.85
5	7.87	5.51

reszta: atom	C:H1'	C:H6	G:H1'	G:H8	U:H1'	U:H6
przesunięcie chem. [ppm]	5.85	7.87	5.51	7.60	5.64	7.77

Rys. 1. Ogólna koncepcja przypisywania sygnałów [P1].

Po identyfikacji sygnałów, ich właściwości (tj. współrzędne w przestrzeni NMR podane w wartościach przesunięcia chemicznego) są przypisywane do odpowiednich atomów. W ten sposób określa się położenia (przesunięcia chemiczne) wybranych atomów cząsteczki w przestrzeni NMR, które następnie wykorzystuje się do obliczenia kształtu całej cząsteczki, tj. współrzędnych atomowych cząsteczki w trójwymiarowej przestrzeni kartezjańskiej.

3.3. Modele i algorytmy w analizie widm dwuwymiarowych

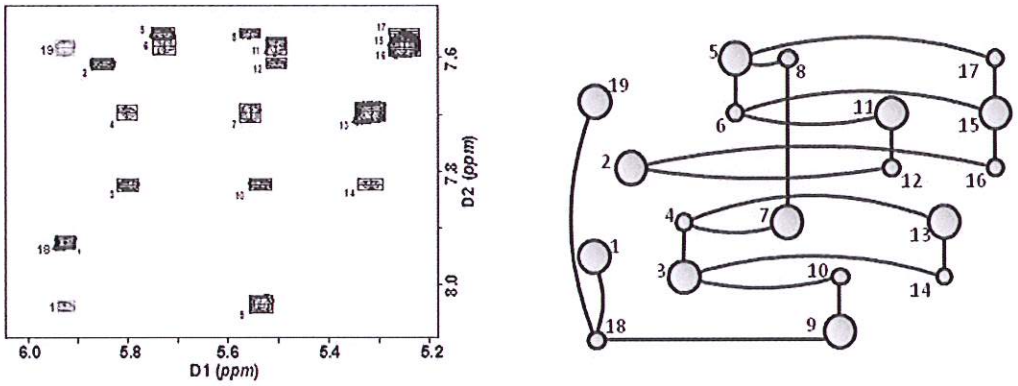
Grafowa reprezentacja przestrzeni rozwiązań

Podczas badań prowadzonych przez uzyskaniem stopnia doktora zdefiniowano nowy rodzaj grafu (nazwany grafem NOESY), który reprezentuje dwuwymiarowe widma NOESY oraz relacje, jakie powinna spełniać poszukiwana na widmie ścieżka przekazywania magnetyzacji. W analizie widm dwuwymiarowych prowadzonej po doktoracie opierano się na tej definicji [P1, P4]:

Definicja 1. Graf NOESY.

Niech $G=(V,E)$ będzie nieskierowanym grafem położonym na płaszczyźnie XxY, gdzie V jest zbiorem wierzchołków a E oznacza zbiór krawędzi. G jest grafem NOESY reprezentującym widmo S, jeśli spełnione są następujące warunki:

a. $|V|=N$, gdzie N jest liczbą sygnałów zarejestrowanych na widmie S
b. dla każdego $i=1..N$: wierzchołek $v_i(x_i,y_i)\in V$ reprezentuje sygnał $s_i(x_i,y_i)\in S$
c. dla każdego $i=1..N$: $v_i\in V$ ma wagę $w_i\in\{0,1\}$, gdzie $w_i=0$ jeśli $s_i\in S$ reprezentuje sygnał internukleotydowy, $w_i=1$ jeśli $s_i\in S$ reprezentuje sygnał intranukleotydowy
d. dla każdego $i,j=1..N$: $e(v_i,v_j)\in E$ jeśli $w_i\neq w_j$ oraz $((x_i=x_j,\ y_i\neq y_j)$ lub $(x_i\neq x_j,\ y_i=y_j))$.



Rys. 2. Przykładowe widmo dwuwymiarowe i odpowiadający mu graf NOESY.

Problem konstruowania ścieżki przypisywania na widmie NMR został sformułowany w postaci problemu poszukiwania ścieżki NOE w grafie NOESY. Stwierdzono, że dowolna ścieżka NOE znaleziona w grafie NOESY odpowiada dokładnie jednej ścieżce NOE znalezionej na widmie, któremu odpowiada analizowany graf. W zależności od rodzaju struktury RNA, regularnej (mającej postać dwóch komplementarnych nici) lub nieregularnej (struktura zawierająca fragmenty jedno- i dwuniciowe), na widmie poszukuje się odpowiednio jednej ścieżki lub dwóch rozłącznych ścieżek przypisywania [P1].

Poszukiwanie pojedynczej ścieżki w grafie NOESY

Definicja pojedynczej ścieżki przypisywania w grafie NOESY w przypadku idealnym:

Definicja 2. Ścieżka NOE.

Niech $G=(V,E)$ będzie grafem NOESY ze zbiorem krawędzi E i zbiorem wierzchołków $V=V_0 \cup V_1$, gdzie $v_i \in V_0$ jeśli $w_i=0$, $v_i \in V_1$ jeśli $w_i=1$ dla każdego $i=1..|V|$. Sekwencję wierzchołków $P=v_1,...,v_l$, gdzie $v_i \in V$, $i \in \{1,...,|V|\}$, nazywamy ścieżką NOE w grafie G jeśli spełnione są następujące warunki:

a. $v_1 \in V_1$
b. P nie zawiera cykli, tj. każdy wierzchołek występuje w ścieżce co najwyżej raz
c. dwa wierzchołki v_k, v_l mogą sąsiadować w ścieżce P jeśli $w_k \neq w_l$
d. każde dwie krawędzie incydentne ze wspólnym wierzchołkiem w P są do siebie prostopadłe
e. żadne dwie krawędzie w P nie leżą na tej samej linii poziomej ani pionowej
f. $|P|=2|V_1|-1$.

Proponując algorytmy optymalizacyjne dla problemu konstruowania ścieżki NOE zdefiniowano również ścieżkę dopuszczalną oraz funkcję oceny ścieżki, której wartość podlega optymalizacji w procesie poszukiwania najlepszego rozwiązania. Definicja ścieżki dopuszczalnej uwzględnia fakt, iż ścieżki na rzeczywistych widmach NMR rzadko spełniają wszystkie warunki podane w definicji 2 [P4].

Definicja 3. Dopuszczalna ścieżka NOE.

Niech $G=(V,E)$ będzie grafem NOESY ze zbiorem krawędzi E i zbiorem wierzchołków V. Sekwencję wierzchołków $P=v_n,...,v_l$, gdzie $v_i \in V$, $i \in \{1,...,|V|\}$, nazywamy dopuszczalną ścieżką NOE w grafie G jeśli P nie zawiera cykli (każdy wierzchołek występuje w ścieżce co najwyżej raz) oraz każde dwie krawędzie incydentne w P są nierównoległe.

W pracy [P4] zaprojektowano algorytm poszukiwania optymalnej ścieżki NOE w grafach reprezentujących dwuwymiarowe widmo NMR regularnych cząsteczek RNA. Algorytm wykorzystuje następującą funkcję oceny rozwiązań:

Definicja 4. Funkcja oceny ścieżki NOE.

Niech $P=v_1,v_2,..v_n$ będzie dopuszczalną ścieżką NOE w grafie $G=(V,E)$ oraz $n \leq |V|$. Funkcję $F(P)$ nazywamy funkcją oceny ścieżki NOE P jeśli $F(P)=\frac{1}{2}\sum_{i=1}^{6} \varphi_i \overline{y_i}$, gdzie ϕ_i są współczynnikami wagi, a $\overline{y_i}$, dla $i=1..6$, przyjmują następujące wartości:

$$\overline{y_1} = \begin{cases} 1 \text{ jeśli } v_1 \text{ jest początkowym wierzchołkiem ścieżki} \\ 0 \text{ w przeciwnym przypadku} \end{cases}$$

$$\overline{y_2} = \sum_{i=1}^{n-1} a_{i,i+1}, \text{ gdzie } a_{i,i+1} = \begin{cases} 1 \text{ gdy } w_i, w_{i+1} \text{ są w tym samym przedziale wartości} \\ 0 \text{ w przeciwnym przypadku} \end{cases}$$

$$\overline{y_3} = \sum_{i=2}^{n-1} b_{i-1,i,i+1},$$

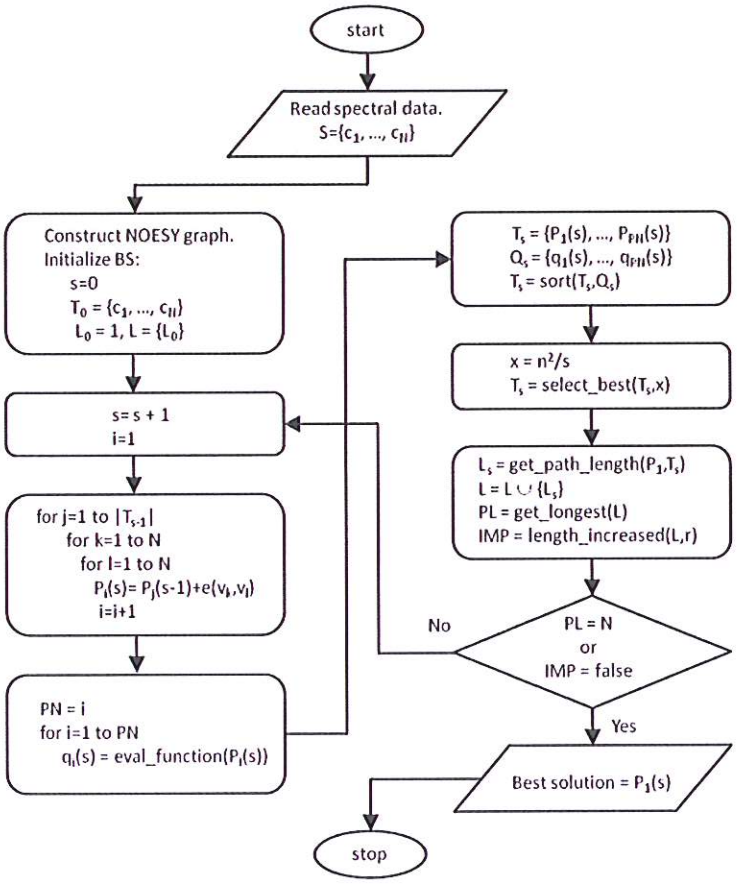
$$\text{gdzie } b_{i-1,i,i+1} = \begin{cases} 1 & \text{jeśli obie krawędzie } (v_{i-1},v_i),(v_i,v_{i+1}) \text{ są pionowe/poziome} \\ 0{,}8 & \text{jeśli jedna krawędź jest pionowa/ pozioma a druga nieznana} \\ 0 & \text{jeśli co najmniej jedna z krawędzi jest diagonalna} \end{cases}$$

$$\overline{y_4} = \sum_{i=1}^{n-1} c_{i-1,i,i+1}, \text{ gdzie } c_{i-1,i,i+1} = \begin{cases} 1 \text{ gdy } (v_{i-1},v_i),(v_i,v_{i+1}) \text{ są współliniowe} \\ 0 \text{ w przeciwnym przypadku} \end{cases}$$

$$\overline{y_5} = \sum_{i=1}^{n} d_i, \text{ gdzie } d_i = \begin{cases} 1 \text{ gdy } v_i \text{ nie odpowiada żadnemu sygnałowi H5-H6} \\ 0 \text{ w przeciwnym przypadku} \end{cases}$$

$$\overline{y_6} = \sum_{i=1}^{n-1} e_{i,i+1}, \text{ gdzie } e_{i,i+1} \text{ jest odchyleniem krawędzi od pionu/poziomu}$$

Algorytm BS-MLP zaproponowany w [P4] stosuje metodę przeszukiwania z wiązką (ang. beam search) oraz sieć neuronową (perceptron wielowarstwowy) do oszacowania wartości wag φ_i w funkcji oceny ścieżki $F(P)$. Algorytm działa iterując trzy główne fazy: faza pierwsza – przedłużenie bieżącej ścieżki, faza druga – ocena jakości bieżącego rozwiązania z wykorzystaniem perceptronu, faza trzecia – wybór najlepszych rozwiązań. Schemat działania algorytmu przedstawia Rysunek 3.



Rys. 3. Schemat działania algorytmu BS-MLP [P4].

Algorytm został przetestowany dla zbioru danych spektralnych. Jego działanie porównano z działaniem innych algorytmów (przeszukiwania tabu, ewolucyjny) [P4].

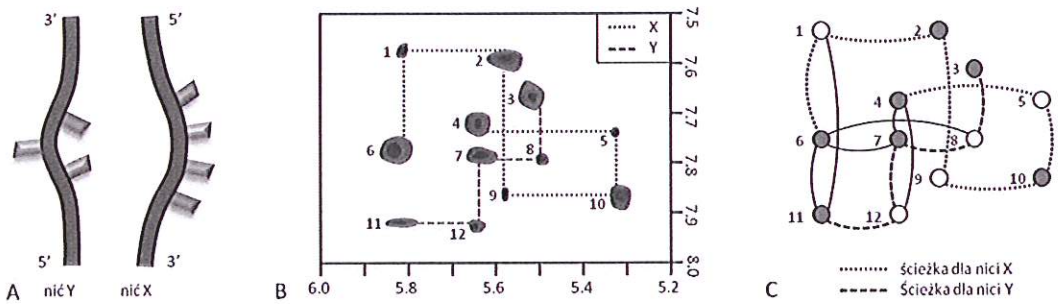
Poszukiwanie rozłącznych ścieżek w grafie NOESY

W grafie, który reprezentuje idealne widmo cząsteczki nieregularnej, złożonej z *K*-nici poszukuje się *K* rozłącznych ścieżek NOE [P1], które spełniają następujące warunki:

$$\bigcup_{i=1}^{K} P_i = V \quad \text{oraz} \quad \bigcap_{i=1}^{K} P_i = \emptyset,$$

gdzie P_i oznacza *i*-tą ścieżkę w grafie *G*.

W grafie reprezentującym widmo z błędami pozytywnymi lub negatywnymi ścieżki muszą spełniać warunek: $\bigcup_{i=1}^{K} P_i \leq V$.



Rys. 4. Przykładowy fragment struktury (A), widma NMR (B) i rozłącznych ścieżek w grafie (C) [P1].

Obserwacja rzeczywistych danych pozyskiwanych z eksperymentów dla cząsteczek nieregularnych przyczyniła się do zdefiniowania dopuszczalnego grafu NOESY [P1]:

Definicja 5. Dopuszczalny graf NOESY.

Niech *G*=(*V*,*E*) będzie nieskierowanym grafem położonym na płaszczyźnie Euklidesowej *X*x*Y*, gdzie *V* jest zbiorem wierzchołków a *E* oznacza zbiór krawędzi. *G* jest dopuszczalnym grafem NOESY reprezentującym widmo *S*, jeśli spełnione są następujące warunki:

a. $|V|=N$, gdzie *N* jest liczbą sygnałów zarejestrowanych na widmie *S*
b. dla każdego i=1..*N*: wierzchołek $v_i(x_i,y_i) \in V$ reprezentuje sygnał $s_i(x_i,y_i) \in S$
c. dla każdego i=1..*N*: $v_i \in V$ ma wagę w_i równą objętości sygnału $s_i \in S$.

W pracy [P1] zaproponowano modyfikację dwóch algorytmów heurystycznych, algorytmu przeszukiwania tabu oraz algorytmu przeszukiwania z wiązką. Celem modyfikacji była możliwość analizy widm struktur nieregularnych i wyznaczanie na tych widmach rozłącznych ścieżek przypisywania. Działanie algorytmów przetestowano na wybranych rzeczywistych widmach NMR.

3.4. Modele i algorytmy w analizie widm trójwymiarowych

Podstawowy grafowy model problemu

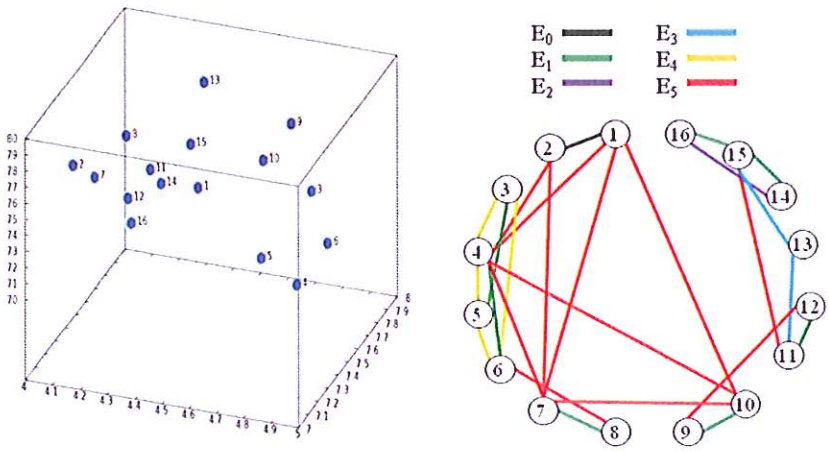
W pracy [P6] po raz pierwszy przeanalizowano problem przetwarzania trójwymiarowych widm NMR cząsteczek RNA. Zdefiniowano tam reguły identyfikacji sygnałów rezonansowych na widmach trójwymiarowych, z uwzględnieniem dwóch rodzajów widm: homo- i heterojądrowych, naszkicowano model grafu spektralnego oraz ścieżkę przypisywania w tym grafie. Dalsze szczegółowe badania problemu w wersji trójwymiarowej przedstawiono w pracy [P2]. Pokazano, że przestrzeń poszukiwań w problemie przypisywania sygnałów pochodzących z trójwymiarowego eksperymentu NMR może być reprezentowana w postaci grafu z pokolorowanymi krawędziami, gdzie liczba kolorów *k* wynosi 6.

Definicja 6. Graf spektralny.

Niech $G=(V,E)$ będzie grafem nieskierowanym, gdzie V jest zbiorem wierzchołków a E oznacza zbiór krawędzi. G jest grafem spektralnym reprezentującym trójwymiarowe widmo S, jeśli spełnione są następujące warunki:

a. $|V|=N$, gdzie N jest liczbą sygnałów zarejestrowanych na widmie S
b. dla każdego $i=1..N$: wierzchołek $v_i \in V$ reprezentuje sygnał $s_i \in S$
c. dla każdego $i,j=1..N$: $e(v_i,v_j) \in E$ jeśli $s_i,s_j \in S$ mają jedną lub dwie wspólne współrzędne, a różnią się pozostałymi
d. dla każdego $i,j=1..N$: krawędź $e(v_i,v_j) \in E$ ma przypisany kolor (etykietę) $c(v_i,v_j)$ ze zbioru C={0,1,2,3,4,5}:

$$c(v_i,v_j) = \begin{cases} 0 \text{ jeśli } x_i \neq x_j,\ y_i = y_j,\ z_i = z_j \\ 1 \text{ jeśli } x_i = x_j,\ y_i \neq y_j,\ z_i = z_j \\ 2 \text{ jeśli } x_i = x_j,\ y_i = y_j,\ z_i \neq z_j \\ 3 \text{ jeśli } x_i = x_j,\ y_i \neq y_j,\ z_i \neq z_j \\ 4 \text{ jeśli } x_i \neq x_j,\ y_i \neq y_j,\ z_i = z_j \\ 5 \text{ jeśli } x_i \neq x_j,\ y_i = y_j,\ z_i \neq z_j \end{cases}$$



Rys. 5. Przykładowe widmo trójwymiarowe i odpowiadający mu graf spektralny [P2].

Zatem graf spektralny $G=(V,E)$ jest grafem z k-pokolorowanymi krawędziami ($k=|C|=6$), co oznacza, że zbiór krawędzi E składa się z $k=6$ podzbiorów, $E=E_0 \cup E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$:

$E_0=\{(v_i,v_j)\in E : x_i \neq x_j,\ y_i=y_j,\ z_i=z_j\}$

$E_1=\{(v_i,v_j)\in E : x_i=x_j,\ y_i \neq y_j,\ z_i=z_j\}$

$E_2=\{(v_i,v_j)\in E : x_i=x_j,\ y_i=y_j,\ z_i \neq z_j\}$

$E_3=\{(v_i,v_j)\in E : x_i=x_j,\ y_i \neq y_j,\ z_i \neq z_j\}$

$E_4=\{(v_i,v_j)\in E : x_i \neq x_j,\ y_i \neq y_j,\ z_i=z_j\}$

$E_5=\{(v_i,v_j)\in E : x_i \neq x_j,\ y_i=y_j,\ z_i \neq z_j\}$

Problem konstruowania ścieżki przypisywania na trójwymiarowym widmie NMR został sformułowany w postaci problemu poszukiwania ścieżki transferowej w grafie spektralnym. Ścieżka transferowa zachowuje predefiniowany porządek kolorów krawędzi. Schemat porządku kolorów zależy od rodzaju widma trójwymiarowego, homo- lub heteronuklearnego, reprezentowanego przez analizowany graf spektralny z k–pokolorowanymi krawędziami [P2]:

Definicja 7. Ścieżka transferowa (wersja heteronuklearna).

Niech $G=(V,E)$ będzie grafem spektralnym ze zbiorem wierzchołków V i zbiorem krawędzi $E=E_0 \cup E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$. Sekwencję wierzchołków $P=v_n,..,.v_l$, gdzie $v_i \in V$, $i \in \{1,...,|V|\}$, nazywamy heteronuklearną ścieżką transferową w grafie G jeśli spełnione są następujące warunki:

a. dla każdej krawędzi $e_i \in P$, $i \geq 1$: $e_i(v_i,v_{i+1}) \in E$
b. dla każdej krawędzi $e_i \in P$, $i \geq 1$: jeśli $e_i \in E_w$ to $e_{i+1} \notin E_w$
c. ścieżka P jest zbudowana z krawędzi należących do podzbioru $E_0 \cup E_3$ lub $E_1 \cup E_5$ lub $E_2 \cup E_4$
d. w ścieżce P krawędzie ułożone są naprzemiennie względem kolorów
e. $|P| \geq m$, gdzie $m \in \{1,...,|E|\}$.

Definicja 8. Ścieżka transferowa (wersja homonuklearna).

Niech $G=(V,E)$ będzie grafem spektralnym ze zbiorem wierzchołków V i zbiorem krawędzi $E=E_0 \cup E_1 \cup E_2 \cup E_3 \cup E_4 \cup E_5$. Sekwencję wierzchołków $P=v_n,..,.v_l$, gdzie $v_i \in V$, $i \in \{1,...,|V|\}$, nazywamy homonuklearną ścieżką transferową w grafie G jeśli spełnione są następujące warunki:

a. dla każdej krawędzi $e_i \in P$, $i \geq 1$: $e_i(v_i,v_{i+1}) \in E$
b. dla każdej krawędzi $e_i \in P$, $i \geq 1$: jeśli $e_i \in E_w$ to e_{i+1}, $e_{i+2} \notin E_w$
c. ścieżka P jest zbudowana z krawędzi należących do podzbioru $E_0 \cup E_1 \cup E_2$
d. w ścieżce P krawędzie ułożone są naprzemiennie względem kolorów
e. $|P| \geq m$, gdzie $m \in \{1,...,|E|\}$.

Ścieżka transferowa jest najdłuższą ścieżką o predefiniowanym porządku kolorów, jaką można skonstruować w grafie spektralnym. Idealna ścieżka transferowa jest ścieżką Hamiltona w grafie spektralnym.

Złożoność obliczeniowa problemu

W pracy [P2] przeprowadzono analizę złożoności obliczeniowej problemu konstruowania ścieżki przypisywania sygnałów na trójwymiarowej mapie NMR. Przedstawiono dowód NP-zupełności wersji decyzyjnej problemu w wariancie homonuklearnym oraz w wariancie heteronuklearnym. W obu przypadkach zademonstrowano, że algorytm niedeterministyczny rozwiązuje badany problem w czasie wielomianowym. Następnie wybrano znany NP-zupełny problem poszukiwania ścieżki Hamiltona w grafie i pokazano proces transformacji wielomianowej do naszego problemu. Analizę złożoności obliczeniowej problemu dwuwymiarowego przeprowadzono w pracy doktorskiej i nie będziemy się tą kwestią tu zajmować.

Problem poszukiwania najdłuższej ścieżki o predefiniowanym porządku kolorów

Analiza problemu konstruowania ścieżek transferowych w grafie reprezentującym trójwymiarowe widmo NMR zainicjowała sformułowanie ogólnego problemu grafowego poszukiwania najdłuższej uporządkowanej kolorystycznie ścieżki w grafie z k-pokolorowanymi krawędziami [P3].

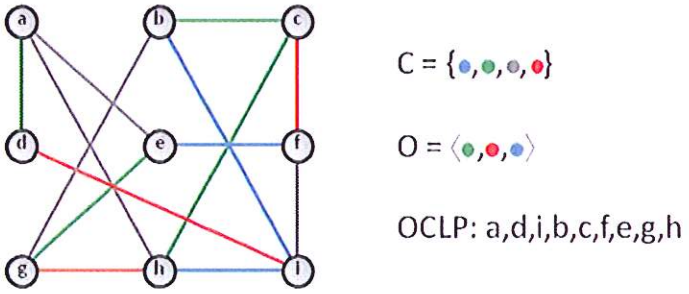
Definicja 9. Graf z k-pokolorowanymi krawędziami.

Niech $G=(V,E)$ będzie grafem nieskierowanym, gdzie V jest zbiorem wierzchołków, $|V|=n$, E oznacza zbiór krawędzi, $|E|=m$. Niech C będzie zbiorem kolorów (etykiet). G jest grafem z k-pokolorowanymi krawędziami jeśli jego krawędzie są pokolorowane za pomocą k kolorów, $k\in[2,|C|]$, zgodnie z funkcją $f:E\rightarrow C$.

Definicja 10. Uporządkowana kolorystycznie ścieżka w grafie (OCP).

Niech $G=(V,E)$ będzie grafem nieskierowanym z k-pokolorowanymi krawędziami zgodnie z funkcją $f:E\rightarrow C$, gdzie C jest zbiorem kolorów, a $k\in[2,|C|]$. Niech $O=\langle c_i,c_{i+1},\ldots,c_l\rangle$ będzie sekwencją kolorów ze zbioru C, która definiuje porządek kolorów. Ścieżkę P w grafie G nazywamy ścieżką uporządkowaną kolorystycznie jeśli $|P|\geq 3$ i kolory kolejnych krawędzi w ścieżce są zgodne z sekwencją O.

W pracy [P3] przeanalizowano zastosowania zaproponowanego modelu grafowego dla różnych problemów rzeczywistych oraz zebrano informacje o problemach poszukiwania ścieżek w grafach z kolorowymi krawędziami i ich złożoności obliczeniowej.



Rys. 6. 4-pokolorowany graf G, porządek kolorów O oraz ścieżka Hamiltona zachowująca ten porządek.

Modele wykorzystujące sieci przepływowe

W pracy [P3] zaproponowano 3 modele sieci przepływowych dla problemu poszukiwania najdłuższej ścieżki uporządkowanej kolorystycznie. Dwa spośród nich (wybrane na podstawie testów wykonanych w pracy [P3]) zastosowano dla problemu konstruowania ścieżek transferowych, a wyniki tej pracy zaprezentowano w [P2].

Niech $G=(V,E)$ będzie oryginalnym grafem z c-pokolorowanymi krawędziami, gdzie $|V|=n$, $|E|=m$. Niech $O=<c_i,c_{i+1},\ldots,c_l>$ będzie sekwencją kolorów ze zbioru C, $|C|=c$, która definiuje porządek kolorów, jaki ma być zachowany przez ścieżkę w grafie G. W modelach LPnPP, LPcPP, LPcCP zaproponowano transformacje grafu G do odpowiednich grafów przepływowych.

- Model 1. LPnPP: Problem najdłuższej ścieżki w n-dzielnym grafie acyklicznym

  Transformacja $G=(V,E)$ do n-dzielnego grafu skierowanego $G'=(V',A')$:

  i. Utwórz zbiór wierzchołków V' z n kopii zbioru V. Dodatkowo dodaj do V' dwa wierzchołki: źródło (s) oraz ujście (t). i-ta kopia zbioru V w G' tworzy partycję L_i. Możemy też powiedzieć, że zbiór V' składa się z n poziomów V^i, gdzie każdy poziom zawiera n kopii wierzchołka v_i.

  ii. Do zbioru A' dodaj łuki o zerowym koszcie: łączące źródło z wszystkimi wierzchołkami podzbioru L_1 oraz łuki wychodzące łączące każdy wierzchołek (z wyjątkiem źródła i ujścia) z ujściem. Dla każdej krawędzi $(v_i,v_j)\in E$ z kolorem c_y, dodaj do A' odpowiadające jej łuki (v_i^r,v_j^{r+1}) i (v_j^r,v_i^{r+1}), gdzie $r=1,\ldots,n-1$: r mod $c = y$ lub (r mod $c = 0$ oraz $y=c$). Łuki te mają koszt równy 1.

  Najdłuższa prosta ścieżka P ze źródła do ujścia grafu G' taka, że każdy poziom zbioru V' jest odwiedzany dokładnie raz, jest uporządkowaną kolorystycznie ścieżką w grafie G.

  Sformułowanie LPnPP w postaci problemu optymalizacyjnego [P3]:

  Maksymalizuj $\sum_{(v_i^r,v_j^{r+1})\in A'} p_r^{ij} x_r^{ij}$

  przy spełnieniu warunków:

  $$\sum_{(v_i^r,v_j^{r+1})\in A'} x_r^{ij} - \sum_{(v_j^{r-1},v_{ij}^r)\in A'} x_r^{ji} = 0 \quad \forall v_i^r \in V' - \{s,t\}$$

  $$\sum_{(s,v_j^1)\in A'} x_1^{sj} = 1$$

  $$\sum_{(v_i^r,t)\in A'} x_r^{it} = 1$$

  $$\sum_{(v_j^{r-1},v_i^r)\in A',v_i^r\in V^l} x_r^{ji} \le 1 \qquad l = 1,2,\ldots,n$$

  $$x_r^{ij} \in \{0,1\} \qquad \forall (v_i^r, v_j^{r+1}) \in A'$$

  gdzie x_r^{ij} jest zmienną decyzyjną skojarzoną z łukiem $(v_i^r,v_j^{r+1})\in A'$, natomiast p_r^{ij} jest kosztem tego łuku.

- Model 2. LPcPP: Problem najdłuższej ścieżki w *c*-dzielnym grafie cyklicznym

  Transformacja $G=(V,E)$ do *c*-dzielnego grafu skierowanego $G^*=(V^*,A^*)$:

  i. Utwórz zbiór wierzchołków V^* z *c* kopii zbioru *V* (*c* jest liczbą kolorów w grafie *G*). Dodatkowo dodaj do V^* dwa wierzchołki: źródło (*s*) oraz ujście (*t*). *i*-ta kopia zbioru *V* w G^* tworzy partycję L_i. Możemy też powiedzieć, że zbiór V^* składa się z *n* poziomów V^i, gdzie każdy poziom zawiera *c* kopii wierzchołka v_i.

  ii. Do zbioru A^* dodaj łuki o zerowym koszcie: łączące źródło z wszystkimi wierzchołkami podzbioru L_1 oraz łuki wychodzące łączące każdy wierzchołek (z wyjątkiem źródła i ujścia) z ujściem. Dla każdej krawędzi $(v_i,v_j)\in E$ z kolorem c_y, dodaj do A^* odpowiadające jej dwa łuki o koszcie równym 1:
     - (v_i^r,v_j^{r+1}) i (v_j^r,v_i^{r+1}) jeśli $r=y$, $r\in\{1,\dots,c-1\}$
     - (v_i^c,v_j^1) i (v_j^c,v_i^1) jeśli $y=c$.

  Sformułowanie LPcPP w postaci problemu optymalizacyjnego [P3]:

  Maksymalizuj $\sum_{(v_i,v_j)\in A^*} p^{ij}x^{ij}$

  przy spełnieniu warunków:

  $$\sum_{(v_i,v_j)\in A^*} x^{ij} - \sum_{(v_j,v_i)\in A^*} x^{ji} = 0 \quad \forall\, v_i \in V^* - \{s,t\}$$

  $$\sum_{(s,v_j^1)\in A^*} x^{sj} = 1$$

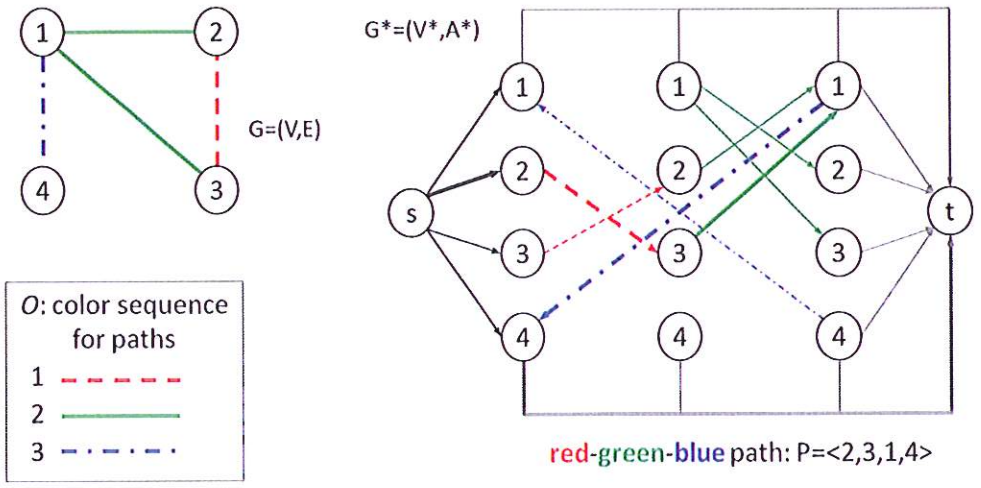
  $$\sum_{(v_i,t)\in A^*} x^{it} = 1$$

  $$\sum_{(v_i,v_j)\in A^*,\, v_i\in V^l} x^{ij} \le 1 \qquad l = 1,2,\dots,n$$

  $$\sum_{(v_i,v_j)\in\gamma} x^{ij} \le |\gamma| - 1 \qquad \gamma \in \hat{\gamma}$$

  $$x^{ij} \in \{0,1\} \qquad \forall (v_i,v_j) \in A^*$$

  gdzie x^{ij} jest zmienną decyzyjną skojarzoną z łukiem $(v_i,v_j)\in A^*$, natomiast p^{ij} jest kosztem tego łuku.



Rys. 7. Przykładowy graf *G* i odpowiadająca mu sieć przepływowa G^* [P3].

- Model 3. LPcCP: Problem najdłuższej ścieżki w *c*-spójnym grafie cyklicznym

  Transformacja $G=(V,E)$ do *c*-spójnego grafu skierowanego $G^{\#}=(V^{\#},A^{\#})$:

  i. Utwórz zbiór wierzchołków $V^{\#}$ z *c* kopii zbioru *V*. Dodatkowo dodaj do $V^{\#}$ dwa wierzchołki: źródło (*s*) oraz ujście (*t*). *i*-ta kopia zbioru *V* w $G^{\#}$ tworzy partycję L_i. Możemy też powiedzieć, że zbiór $V^{\#}$ składa się z *n* podzbiorów V^i, gdzie każdy podzbiór zawiera *c* kopii wierzchołka v_i. Każda kopia wierzchołka v_i ma przypisany kolor c_j.

  ii. Zbiór $A^{\#}$ jest tworzony w następujący sposób: (a) Dla każdego $V^i \subset V^{\#}$ wszystkie wierzchołki są połączone łukami o koszcie zerowym w cykl tak, że sekwencja kolorów wierzchołków jest zgodna z wzorcem *O*. W ten sposób otrzymujemy *c*-spójne podgrafy $G^i=(V^i,A^i)$, gdzie $|V^i|=|A^i|=c$. Oznaczmy przez A' sumę wszystkich podzbiorów A^i. (b) Dla każdej krawędzi $(v_i,v_j)\in E$ z kolorem c_y, dodaj do $A^{\#}$ odpowiadające jej łuki (v_i^y,v_j^y) i (v_j^y,v_i^y). Łuki te mają koszt równy 1. Oznaczmy przez A'' zbiór wszystkich łuków odpowiadających krawędziom z grafu *G*. (c) Do zbioru $A^{\#}$ dodaj łuki o zerowym koszcie: łączące źródło z wszystkimi wierzchołkami $v_i^1 \in V^{\#}$ oraz łuki wychodzące łączące każdy wierzchołek (z wyjątkiem źródła i ujścia) z ujściem. Te łuki należą do A''.

  Sformułowanie LPcCP w postaci problemu optymalizacyjnego [P3]:

  Maksymalizuj $\sum_{(v_i^l,v_j^l)\in A''} p_l^{ij} x_l^{ij}$

  przy spełnieniu warunków:

  $$\sum_{(v_j^l,v_i^l)\in A''} x_l^{ji} - \sum_{(v_i^l,v_i^r)\in A'} x_{lr}^{i} = 0 \quad \forall v_i^l \in V^{\#} - \{s,t\}$$

  $$\sum_{(v_i^r,v_i^l)\in A'} x_{rl}^{i} - \sum_{(v_i^l,v_i^l)\in A''} x_l^{ji} = 0 \quad \forall v_i^l \in V^{\#} - \{s,t\}$$

  $$\sum_{(s,v_j^1)\in A''} x_1^{sj} = 1$$

  $$\sum_{(v_i^l,t)\in A''} x_l^{it} = 1$$

  $$\sum_{(v_i^l,v_i^r)\in A'} x_{lr}^{i} \le 1 \qquad l = 1,2,\dots,n$$

  $$\sum_{(v_i^l,v_i^l)\in A''} x_l^{ij} \le 1 \qquad l = 1,2,\dots,n$$

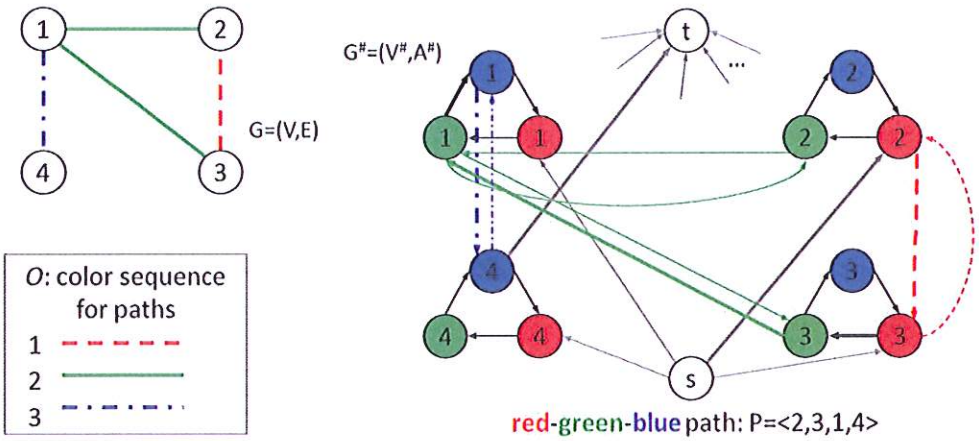
  $$\sum_{(v_j^l,v_i^l)\in A''} x_l^{ji} \le 1 \qquad l = 1,2,\dots,n$$

  $$\sum_{(v_i^l,v_j^l)\in\gamma} x_l^{ij} + \sum_{(v_i^l,v_j^r)\in\gamma} x_{lr}^{i} \le |\gamma| - 1 \qquad \gamma \in \hat{\gamma}$$

  $$x_l^{ij} \in \{0,1\} \qquad \forall (v_i^l, v_j^l) \in A''$$

  $$x_{lr}^{i} \in \{0,1\} \qquad \forall (v_i^l, v_i^r) \in A'$$

  gdzie x_l^{ij} jest zmienną decyzyjną skojarzoną z łukiem $(v_i^l,v_j^l)\in A''$, x_{lr}^{i} jest zmienną decyzyjną skojarzoną z łukiem $(v_i^l,v_i^r)\in A^i$, p_l^{ij} jest kosztem łuku $(v_i^l,v_j^l)\in A''$.



Rys. 8. Przykładowy graf *G* i odpowiadająca mu sieć przepływowa $G^{\#}$ [P3].

W publikacji [P3] pokazano wyniki eksperymentu, w którym porównano zaprezentowane wyżej modele LPnPP, LPcPP, LPcCP rozwiązane metodami liniowego programowania całkowitoliczbowego. Eksperyment pokazał przewagę modeli LPcPP i LPcCP nad pierwszą zaproponowaną reprezentacją LPnPP. W konsekwencji, w pracy [P2] dwa lepsze modele wykorzystano do rozwiązania problemu konstruowania ścieżek transferowych w grafach reprezentujących trójwymiarowe widma NMR. Badania wykonane w ramach [P2] pokazały, że LPcPP jest bardziej efektywny w przetwarzaniu mniejszych instancji (ok. 50-ciu wierzchołków w grafie), natomiast LPcCP nadaje się bardziej dla instancji większych (75 i 100 wierzchołków). Ponadto zastosowanie LPcCP pozwala na znaczącą (30-60% w zależności od zastosowanej procedury separowania cykli) oszczędność czasu obliczeń w stosunku do drugiego modelu, zwłaszcza dla instancji o rozmiarze większym od 50. Szczegółowa analiza wyników eksperymentu obliczeniowego podana jest w pracy [P2].

Przedstawione modele mogą znaleźć zastosowanie w rozwiązywaniu innych problemów, pojawiających się nie tylko w obszarze bioinformatyki i biologii obliczeniowej, ale np. w transporcie, logistyce, grach, zagadnieniach teorii krat, itp. Modele te pozwalają reprezentować zwłaszcza problemy wielowymiarowe i wieloparametryczne (gdzie liczba wymiarów / parametrów przekracza 2), w których każdy wymiar lub parametr może odpowiadać pojedynczemu kolorowi w grafie z pokolorowanymi krawędziami. Kilka takich problemów przedstawiono w pracy [P3].

3.5. Przetwarzanie informacji o rozwiązanych strukturach RNA

Struktury cząsteczek biologicznych, które zostały określone eksperymentalnie są przechowywane w strukturalnych bazach danych. Podstawową bazą, w której umieszczane są informacje o strukturach trzeciorzędowych białek, DNA i RNA jest Protein Data Bank. Baza ta jednak nie wspomaga zaawansowanego wyszukiwania interesujących użytkownika motywów strukturalnych i gromadzi tylko podstawowe informacje o strukturze. W ramach prowadzonych badań zaprojektowane zostało repozytorium RNA FRABASE, które przechowuje dodatkowe informacje o strukturach RNA (np. wartości kątów torsyjnych, których badanie jest istotne podczas analizy

struktur w eksperymencie NMR) [P7]. Informacje te są obliczane przez skrypty współpracujące z modułem pozyskiwania danych do bazy. Ponadto zaprojektowana została wyszukiwarka, która pozwala w bazie RNA FRABASE wyszukiwać fragmenty o interesującym dla użytkownika wzorcu sekwencyjnym lub drugorzędowym. Pozwoliło to wykorzystać zgromadzone w bazie informacje w analizie struktur nieregularnych i wyszukaniu odpowiednich cząsteczek RNA do analizy w pracy [P1]. Praca [P5] była kontynuacją projektu przedstawionego w pracy [P7]. Rozszerzono w niej schemat bazy danych, tak, aby możliwe było gromadzenie danych o wszystkich modelach NMR uzyskanych podczas eksperymentalnego określania struktury. Umożliwiono również bardziej efektywne wyszukiwanie małych fragmentów i motywów strukturalnych. Dane gromadzone w bazie zostały wykorzystane m.in. w eksperymentach obliczeniowych w pracach [P2], [P4] i [P6]. W pracach tych wykorzystano metodę wstecznego obliczania danych spektralnych na podstawie informacji o modelu NMR-owskim cząsteczki, a uzyskane tak widma zostały włączone do zbioru testowego i zbioru uczącego.

Baza danych RNA FRABASE jest obecnie jednym z najpopularniejszych narzędzi bioinformatycznych, szeroko wykorzystywanym w środowisku naukowym. Świadczy o tym m.in. liczba użytkowników (400 tys. użytkowników ze 150 krajów świata), liczba wizyt na stronie połączona z aktywnym wykorzystaniem funkcji systemu (2 mln sesji) oraz dzienne obciążenie systemu (860 sesji dziennie, co daje 1 nową sesję na każde dwie minuty doby). Łączna liczba cytowań obu publikacji o bazie RNA FRABASE wynosi obecnie 50 (Web of Science). Dane te zebrano w dniu 10.03.2015.

3.6. Kontynuacja badań oraz wykorzystanie wyników

Badania zaprezentowane w punktach 3.2-3.5 stały się zaczątkiem kolejnych projektów w tematyce modelowania i rozwiązywania problemów w bioinformatyce strukturalnej cząsteczek RNA. W ramach informatycznej analizy związanej z eksperymentami NMR rozpoczęto badania parametrów kątowych struktury RNA. Opracowano kątową (trygonometryczną) reprezentację struktury oraz transformację modelu klasycznego (opartego na współrzędnych atomowych) do tej reprezentacji [A9]. Reprezentację tę wykorzystano do klastrowania zbioru wszystkich określonych eksperymentalnie struktur nukleotydów RNA oraz do zbudowania biblioteki reprezentatywnych konformacji nukleotydowych [A18]. Zaproponowano metodę MCQ4Structures [A9] do porównywania trygonometrycznych modeli struktur RNA i wykorzystano ją m.in. do oceny jakości struktur przewidzianych komputerowo, zgłoszonych w międzynarodowym konkursie RNA Puzzles [A17]. Rozszerzając spektrum możliwości w ocenie jakości struktur i weryfikacji ich poprawności zaprojektowano nowe narzędzie (RNApdbee) do ekstrakcji struktury drugorzędowej RNA ze współrzędnych atomowych [A11]. RNApdbee jest obecnie jedynym istniejącym narzędziem, które potrafi wydobyć informacje o strukturach drugorzędowych zawierających pseudowęzły wysokiego rzędu i poprawnie je sklasyfikować. Zaproponowano również metodę oceny jakości struktur RNAlyzer opartą na deskryptorach [A7]. Wykorzystując zasoby oraz wyszukiwarkę bazy RNA FRABASE zaprojektowano nową metodę przewidywania struktury trójwymiarowej

RNA i zaimplementowano ją w systemie RNAComposer [A6]. System wykorzystano m.in. do przewidywania struktur ryboprzełączników, które - ze względu na duży rozmiar - trudno jest ustalić za pomocą innych metod komputerowych [A13]. Modele przewidziane przy wykorzystaniu metody RNAComposer brały również udział w kolejnych edycjach konkursu RNA Puzzles, uzyskując wysokie oceny asesorów, zwłaszcza na tle innych modeli obliczanych metodami automatycznymi. Jedną z ostatnich koncepcji badawczych było połączenie funkcjonalności systemów RNAComposer i RNApdbee dla celu wyznaczania rozszerzonej struktury drugorzędowej cząsteczek RNA zawierającej pełny opis wiązań kanonicznych i niekanonicznych [A19]. Obecnie RNAComposer jest najszybszą w pełni zautomatyzowaną metodą modelowania struktur RNA o wysokiej rozdzielczości. O jego popularności w środowisku naukowym świadczą statystyki użycia (10 tys. użytkowników, 56 sesji dziennie) oraz liczba cytowań (31). Ponadto, zarówno RNAComposer jak i RNA FRABASE są wykorzystywane dla celów edukacyjnych [A16], a zajęcia, w ramach których studenci poznają i uczą się praktycznego wykorzystania tych systemów znajdują się w programach nauczania kilku uniwersytetów i szkół wyższych w Polsce i na świecie.

Wymienione w tym rozdziale metody i narzędzia, będące kontynuacją lub rozszerzeniem przedstawianego osiągnięcia naukowego, zostały opublikowane w pracach, których pełna lista przytoczona jest w sekcji 5 autoreferatu.

3.7. Podsumowanie

Badania zaprezentowane w niniejszej sekcji, wykorzystując metody informatyczne, wnoszą istotny wkład w dziedzinę bioinformatyki i biologii obliczeniowej zwłaszcza w zakresie poznawania struktur cząsteczek RNA. Pozwalają one na wykonywanie eksperymentów, które w warunkach laboratoryjnych byłyby kosztowne i miałyby bardzo ograniczony zasięg (RNApdbee [A11], RNAComposer [A6]). Umożliwiają przetwarzanie danych eksperymentalnych, których ręczna analiza byłaby niemożliwa dla dużych zbiorów danych jakimi coraz częściej dysponuje eksperymentator (analiza widm NMR [P1,P2,P4,P6], RNA FRABASE [P5,P7], MCQ4Structures [A9]). Ułatwiają weryfikację uzyskiwanych danych strukturalnych oraz ocenę działania innych metod bioinformatycznych (MCQ4Structures [A9], RNApdbee [A11], RNAlyzer [A7]). Umożliwiają gromadzenie informacji o strukturze oraz jej efektywne przeszukiwanie (RNA FRABASE [P5,P7]). Wreszcie – wspomagają edukację w zakresie bioinformatyki strukturalnej i biologii [A16]. Z kolei reprezentacje grafowe zainspirowane problemami biologicznymi wnoszą nowy pierwiastek w dziedzinę modelowania za pomocą grafów i pozwalają na rozwiązywanie zagadnień mających źródło w innych dziedzinach nauki i życia (np. problemy transportu i logistyki) [P3]. Przenikanie się problematyki z tych dwóch różnych i dalekich od siebie dziedzin nauki istotnie przyczynia się do rozwoju ich obu.

[5] Osiągnięcia naukowo-badawcze ogółem

Po uzyskaniu stopnia doktora (w 2005 roku) podjęłam badania w dziedzinie informatyki, w obszarze bioinformatyki, opracowując metody wspomagające analizę strukturalną RNA. Wyniki moich badań zostały opisane w publikacjach naukowych, których pełne zestawienie uporządkowane chronologicznie przedstawiam poniżej. W zestawieniu A podano wartości współczynnika Impact Factor (IF) według listy Journal Citation Reports (JCR) oraz punkty MNiSW, w obu przypadkach zgodnie z rokiem publikacji. Zestawienie B zawiera listę prac w czasopismach, które w roku publikacji nie znajdowały się w bazie Journal Citation Reports.

A. Publikacje w czasopismach znajdujących się w bazie Journal Citation Reports

[1] R.W. Adamiak, J. Blazewicz, P. Formanowicz, Z. Gdaniec, M. Kasprzak, M. Popenda, M. Szachniuk. An algorithm for an automatic NOE pathways analysis in 2D NMR spectra of RNA duplexes, *Journal of Computational Biology* 11(1), 2004, 163-180 (doi: 10.1089/106652704773416948).
Impact Factor: **3,241**
Punkty KBN: 21

[2] J. Blazewicz, M. Szachniuk, A. Wojtowicz. RNA tertiary structure determination: NOE pathway construction by tabu search, *Bioinformatics* 21(10), 2005, 2356-2361 (doi: 10.1093/bioinformatics/bti351).
Impact Factor: **6,019**
Punkty KBN: 21

przed uzyskaniem stopnia doktora ↑

po uzyskaniu stopnia doktora ↓

[3] M. Popenda, M. Blazewicz, M. Szachniuk, R.W. Adamiak. RNA FRABASE version 1.0: an engine with a database to search for the three-dimensional fragments within RNA structures, *Nucleic Acids Research* 36(1), 2008, D386-D391 (doi:10.1093/nar/gkm786).
Impact Factor: **6,878**
Punkty MNiSW: 40

[4] M. Popenda, M. Szachniuk, M. Blazewicz, S. Wasik, E.K. Burke, J. Blazewicz, R.W. Adamiak. RNA FRABASE 2.0: an advanced web-accessible database with the capacity to search the three-dimensional fragments within RNA structures, *BMC Bioinformatics* 11, 2010, 231 (doi:10.1186/1471-2105-11-231).
Impact Factor: **3,028**
Punkty MNiSW: 35

[5] M. Radom, A. Rybarczyk, R. Kottmann, P. Formanowicz, M. Szachniuk, F.O. Gloeckner, D. Rebholz-Schuhmann, J. Blazewicz. Poseidon: an information retrieval and extraction system for metagenomic marine science, *Ecological Informatics* 12, 2012, 10-15 (doi:10.1016/j.ecoinf.2012.07.003).
Impact Factor: **1,961**
Punkty MNiSW: 20

[6] M. Popenda, M. Szachniuk, M. Antczak, K.J. Purzycka, P. Lukasiak, N. Bartol, J. Blazewicz, R.W. Adamiak. Automated 3D structure composition for large RNAs, *Nucleic Acids Research* 40(14), 2012, e112 (doi:10.1093/nar/gks339).
Impact Factor: **8,278**
Punkty MNiSW: 40

[7] P. Lukasiak, M. Antczak, T. Ratajczak, J.M. Bujnicki, M. Szachniuk, M. Popenda, R.W. Adamiak, J. Blazewicz. RNAlyzer – novel approach for quality analysis of RNA structural models, *Nucleic Acids Research* 41(12), 2013, 5978-90 (doi:10.1093/nar/gkt318).
Impact Factor: **8,808**
Punkty MNiSW: 40

[8] M. Szachniuk, M. Malaczynski, E. Pesch, E.K. Burke, J. Blazewicz. MLP accompanied beam search for the resonance assignment problem, *Journal of Heuristics* 19(3), 2013, 443-464 (doi:10.1007/s10732-013-9220-3).
Impact Factor: **1,359**
Punkty MNiSW: 35

[9] T. Zok, M. Popenda, M. Szachniuk. MCQ4Structures to compute similarity of molecule structures, *Central European Journal of Operations Research* 22(3), 2014, 457-474 (doi:10.1007/s10100-013-0296-5).
Impact Factor: **0,787**
Punkty MNiSW: 20

[10] M. Szachniuk, M.C. De Cola, G. Felici, J. Blazewicz. The Orderly Colored Longest Path Problem - a survey of applications and new algorithms, *RAIRO Operations Research* 48(1), 2014, 25-51 (doi:10.1051/ro/2013046).
Impact Factor: **0,333**
Punkty MNiSW: 15

[11] M. Antczak, T. Zok, M. Popenda, P. Lukasiak, R.W. Adamiak, J. Blazewicz, M. Szachniuk. RNApdbee - a webserver to derive secondary structures from pdb files of knotted and unknotted RNAs, *Nucleic Acids Research* 42(W1), 2014, W368-W372 (doi:10.1093/nar/gku330).
Impact Factor: **8,808**
Punkty MNiSW: 40

[12] N. Szostak, F. Royo, A. Rybarczyk, M. Szachniuk, J. Blazewicz, A. del Sol, J.M. Falcon-Perez. Sorting signal targeting mRNA into hepatic extracellular vesicles, *RNA Biology* 11(7), 2014, 836-844 (doi:10.4161/rna.29305).
Impact Factor: **4,841**
Punkty MNiSW: 35

[13] K.J. Purzycka, M. Popenda, M. Szachniuk, M. Antczak, P. Lukasiak, J. Blazewicz, R.W. Adamiak. Automated 3D RNA structure prediction using the RNAComposer method for riboswitches, in Shi-Jie Chen, Donald H. Burke-Aguero (eds.) *Methods in Enzymology: Computational Methods for Understanding Riboswitches* 553, Elsevier, 2014, 3-34 (doi: 10.1016/bs.mie.2014.10.050). Rozdział na zaproszenie.
Impact Factor: **2,194**
Punkty MNiSW: 20

[14] M. Szachniuk, M.C. De Cola, G. Felici, D. de Werra, J. Blazewicz. Optimal pathway reconstruction on 3D NMR maps, *Discrete Applied Mathematics* 182, 2015, 134-149 (doi:10.1016/j.dam.2014.04.010).
Impact Factor: **0,677**
Punkty MNiSW: 25

[15] M. Szachniuk. Assigning NMR Spectra of Irregular RNAs by Heuristic Algorithms, *Bulletin of the Polish Academy of Sciences Technical Sciences* 63(1), 2015, *w druku*.
Impact Factor: **1,000**
Punkty MNiSW: 25

[16] M. Biesiada, K.J. Purzycka, M. Szachniuk, J. Blazewicz, R.W. Adamiak. Automated RNA 3D structure prediction with RNAComposer: guidance for molecular biologists, in Doug Turner, David Mathews (eds.) *Methods in Molecular Biology: RNA Structure Determination*, Springer, Humana Press, 2015, *w druku*. Rozdział na zaproszenie.
Impact Factor: **1,290**
Punkty MNiSW: -

[17] Z. Miao, R.W. Adamiak, M.-F. Blanchet, M. Boniecki, J.M. Bujnicki, S.-J. Chen, C. Cheng, G. Chojnowski4, F.-C. Chou, P. Cordero, J.A. Cruz, A, Ferre-D'Amare, R. Das, F. Ding, N.V. Dokholyan, S. Dunin-Horkawicz, W. Kladwang, A. Krokhotin, G. Lach, M. Magnus, F. Major, T.H. Mann, B. Masquida, D. Matelska, M. Meyer, A. Peselis, M. Popenda, K.J. Purzycka, A. Serganov, J. Stasiewicz, M. Szachniuk, A. Tandon, S. Tian, J. Wang, Y. Xiao, X. Xu, J. Zhang, P. Zhao, T. Zok, E. Westhof. RNA-Puzzles Round II: Assessment of RNA structure prediction programs applied to three large RNA structures, *RNA*, 2015, *w druku*.
Impact Factor: **5,088**
Punkty MNiSW: 35

manuskrypty zgłoszone do druku ↓

[18] T. Zok, M. Antczak, M. Riedel, D. Nebel, T. Villmann, P. Lukasiak, J. Blazewicz, M. Szachniuk. Building the library of RNA 3D nucleotide conformations using clustering approach, zgłoszony do druku w *International Journal of Applied Mathematics and Computer Science*.

[19] A. Rybarczyk, N. Szostak, M. Antczak, T. Zok, M. Popenda, R.W. Adamiak, J. Blazewicz, M. Szachniuk. New in silico approach to assess RNA secondary structures with non-canonical base pairs, zgłoszony do druku w *BMC Bioinformatics*.

B. Pozostałe publikacje (niewymienione w punkcie A)

[1] M. Szachniuk, R.W. Adamiak, P. Formanowicz, Z. Gdaniec, M. Kasprzak, M. Popenda, J. Blazewicz. A combinatorial analysis of 2D NMR spectra of RNA duplexes, *Currents in Computational Molecular Biology*, 2003, 345-346.

[2] J. Blazewicz, M. Szachniuk, A. Wojtowicz. Evolutionary algorithm for a reconstruction of NOE paths in NMR spectra of RNA chains, *Bulletin of the Polish Academy of Sciences. Technical Sciences* 52(3), 2004, 221-230.

[3] J. Blazewicz, M. Szachniuk, A. Wojtowicz. Evolutionary approach to NOE paths assignment in RNA structure elucidation, *Proceedings of the 2004 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2004, 206-213.

[4] M. Szachniuk, L. Popenda, Z. Gdaniec, R.W. Adamiak, J. Blazewicz. NMR analysis of RNA bulged structures: tabu search application in NOE signal assignment, *Proceedings of the 2005 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2005, 175-178.

przed uzyskaniem stopnia doktora ↑

po uzyskaniu stopnia doktora ↓

[5] W. Paczkowski, M. Milawski, L. Popenda, M. Szachniuk. A new tool for fundamental analysis of NMR spectra, *TU Clausthal Technical Report Series* IfI-07-03, 2007, 8-11.

[6] M. Szachniuk, M. Popenda, L. Popenda. Strategies of signal assignment in NMR spectra of RNAs with different structural motifs, *TU Clausthal Technical Report Series* IfI-07-03, 2007, 12-14.

[7] M. Blazewicz, M. Popenda, R.W. Adamiak, M. Szachniuk. Project of RNA fragments database, *TU Clausthal Technical Report Series* IfI-07-03, 2007, 25-27.

[8] M. Szachniuk, M. Popenda, S. Klemczak, J. Blazewicz. An analysis of three-dimensional NMR spectra in RNA structure determination process, *Research report* RA-001/2007, 2007, Poznan Supercomputing and Networking Center.

[9] M. Szachniuk, M. Blazewicz, M. Popenda, R.W. Adamiak. On the design of RNA fragments structural database, *RECOMB'08 poster book*, 2008, 18-19.

[10] M. Szachniuk, M. Popenda, L. Popenda, J. Blazewicz. Constructing transfer pathways in multidimensional NMR spectra of RNAs, *RECOMB'08 poster book*, 2008, 62-63.

[11] M. Szachniuk, M. Radom, A. Rybarczyk, P. Formanowicz, J. Blazewicz. From document processing to an identification of marine microorganisms' habitat specificity, *TU Clausthal Technical Report Series* IfI-08-03, 2008, 61-63.

[12] J. Blazewicz, R.W. Adamiak, P. Lukasiak, M. Popenda, M. Szachniuk, M. Antczak, G. Palik. 3D-RNA-Pred: an automatic building of the three-dimensional RNA structures, *TU Clausthal Technical Report Series* IfI-08-03, 2008, 45-48.

[13] T. Zok, M. Popenda, M. Szachniuk. Comparison of RNA structures - concepts and measures, *TU Clausthal Technical Report Series* IfI-08-03, 2008, 42-44.

[14] M. Szachniuk, M. Popenda, L. Popenda. Graphs in NMR analysis of RNAs, *TU Clausthal Technical Report Series* IfI-09-12, 2009, 52-59.

[15] M. Szachniuk, M. Popenda, R.W. Adamiak, J. Blazewicz. An assignment walk through 3D NMR spectrum, *Proceedings of the 2009 IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology*, 2009, 215-219.

[16] M. Radom, R. Kottmann, A. Rybarczyk, P. Formanowicz, M. Szachniuk, F.O. Gloeckner, J. Blazewicz. Cerberus: a new information retrieval tool for marine metagenomics, *Foundations of Computing and Decision Sciences* 35(2), 2010, 107-126.

[17] T. Zok, M. Szachniuk, M. Antczak. Comparison of RNA structures in torsional angle space, *Machine Learning Reports* 01/2011, 2011, 14-18.

[18] M. Szachniuk, M. Popenda, M. Antczak, K.J. Purzycka, P. Lukasiak, J. Blazewicz, R.W. Adamiak. RNAComposer: the art of composing RNA structures, *Machine Learning Reports* 01/2011, 2011, 26-29.

[19] M. De Cola, G. Felici, M. Szachniuk. The Orderly Colored Longest Path Problem, *CNR-IASI Technical Report* 29/2012, 2012.

[20] M. Szachniuk, J. Sarzynska, J. Blazewicz. Turning data into folds using RNAComposer, *AIP Conference Proceedings* 1559(1), 2013, 353-354.

W ramach prowadzonych badań uczestniczyłam w tworzeniu narzędzi bioinformatycznych, które obecnie są udostępnione społeczności naukowej i szeroko przez nią wykorzystywane. W badaniach, w których efekcie powstały w/w narzędzia pełniłam rolę projektanta, konsultanta naukowego, analityka lub kierownika projektu. Część publikacji z wykazów A, B przedstawia te narzędzia oraz/lub pokazuje wyniki badań uzyskane z ich wykorzystaniem. Zestawienie C przedstawia listę współtworzonych przeze mnie narzędzi bioinformatycznych wraz z informacją o ich dostępności, statystykach użycia i publikacji.

C. Narzędzia bioinformatyczne

[1] RNA FRABASE – baza danych i wyszukiwarka fragmentów strukturalnych cząsteczek RNA
URL: http://rnafrabase.cs.put.poznan.pl
Publikacje: NAR 2008, BMC Bioinformatics 2010.
Liczba odwiedzin (na dzień 10.03.2015): **2 706 385**

[2] RNAComposer – automatyczny system do przewidywania struktury trzeciorzędowej RNA
URL(1): http://rnacomposer.cs.put.poznan.pl
URL(2): http://rnacomposer.ibch.poznan.pl
Publikacje: NAR 2012, MIE 2014, MIMB 2015.
Liczba odwiedzin (sumarycznie na dzień 10.03.2015): **97 042**

[3] RNApdbee – system do obliczania struktury drugorzędowej RNA na podstawie współrzędnych atomowych, wraz klasyfikacją wiązań kanonicznych i niekanonicznych oraz identyfikacją i klasyfikacją pseudowęzłów
URL: http://rnapdbee.cs.put.poznan.pl
Publikacje: NAR 2014, BMC Bioinformatics (zgłoszona).
Liczba odwiedzin (na dzień 10.03.2015): **1 540**

[4] RNAlyzer – system oceny jakości modeli strukturalnych RNA
URL(1): http://rnalyzer.cs.put.poznan.pl
URL(2): http://rnassess.cs.put.poznan.pl
Publikacje: NAR 2013, NAR (zgłoszona).
Liczba odwiedzin: brak danych

[5] MCQ4Structures – program do porównywania trzeciorzędowych struktur RNA w przestrzeni kątów torsyjnych (aplikacja autonomiczna)
URL: http://bio.cs.put.poznan.pl/programs/519227ba9dfb896927000001
Publikacje: CEJOR 2014, RNA 2015 (w druku).
Liczba odwiedzin: brak danych

[6] BS-MLP – program do wyznaczania ścieżek NOE na dwuwymiarowych widmach NMR cząsteczek RNA (aplikacja autonomiczna)
URL: http://www.cs.put.poznan.pl/mszachniuk/bsmlp.zip
Publikacje: JHEU 2013, Bull Pol Acad Sci 2015 (w druku).
Liczba odwiedzin: brak danych

[6] Wskaźniki naukometryczne

- Liczba wszystkich publikacji w czasopismach z bazy JCR: **19**, w tym **17** wydanych po uzyskaniu stopnia doktora, włączając w to publikacje w czasopismach z wysokim IF (*Nucleic Acids Research*, *BMC Bioinformatics*, *RNA Biology*) oraz **2** rozdziały na zaproszenie (*Methods in Enzymology*, *Methods in Molecular Biology*).
- Liczba publikacji według Web of Science: 17, w tym 14 wydanych po uzyskaniu stopnia doktora (nie zarejestrowano publikacji najnowszych; zarejestrowano 4 publikacje w materiałach konferencyjnych, które nie widnieją w bazie JCR – są to publikacje nr [3], [4], [15] i [20] z zestawienia B).
- Sumaryczny IF wszystkich publikacji: **64,59**
- Sumaryczna punktacja MNiSW i KBN wszystkich publikacji: **467**
- Sumaryczny IF publikacji wydanych po uzyskaniu stopnia doktora: **55,33**
- Sumaryczna punktacja MNiSW publikacji wydanych po uzyskaniu stopnia doktora: **425**
- Sumaryczna liczba cytowań (wg Web of Science): **117**
- Liczba cytowań z pominięciem cytowań własnych (wg Web of Science): **81**
- Indeks Hirscha (wg Web of Science): **6**

Powyższe dane zebrano w dniu 10.03.2015.

[7] Dorobek dydaktyczny

Począwszy od roku akademickiego 1999/2000 prowadzę zajęcia na Politechnice Poznańskiej. W latach 1999-2005 były to zajęcia laboratoryjne z przedmiotu "Wprowadzenie do programowania" dla studentów dziennych I roku, na I stopniu kierunku Informatyka (Wydział Elektryczny oraz Wydział Informatyki i Zarządzania Politechniki Poznańskiej). Liczba studentów / semestr: 60. Liczba godzin lekcyjnych: 90, w tym 60 godzin stanowiły zajęcia prowadzone w języku polskim, a 30 godzin – zajęcia w języku angielskim.

Od roku 2005 prowadzę zajęcia praktyczne z modelowania i programowania algorytmów, w ramach przedmiotu "Algorytmy i struktury danych". Przedmiot prowadzony jest dla studentów dziennych I roku, na I stopniu kierunku Informatyka (Wydział Informatyki Politechniki Poznańskiej) oraz na I stopniu makrokierunku Bioinformatyka (Wydział Informatyki Politechniki Poznańskiej i Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu). Liczba studentów / semestr: 60. Liczba godzin lekcyjnych: 120.

W roku akademickim 2006/2007 zaczęłam prowadzić zajęcia praktyczne w ramach przedmiotu "Projekt zespołowy", dla studentów dziennych III roku, na I stopniu kierunku Informatyka (Wydział Informatyki Politechniki Poznańskiej). Liczba studentów: 12-16. Liczba godzin lekcyjnych: 15.

W latach 2010-2012 prowadziłam cykl autorskich wykładów w ramach przedmiotu "Podstawy bioinformatyki" dla studentów dziennych I roku Bioinformatyki, na I stopniu makrokierunku Bioinformatyka (Wydział Informatyki Politechniki Poznańskiej i Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu). Liczba studentów: 30. Liczba godzin lekcyjnych: 15.

Począwszy od roku akademickiego 2011/2012 wykładam przedmiot "Bioinformatyka strukturalna" dla studentów dziennych I roku, na II stopniu makrokierunku Bioinformatyka (Wydział Informatyki Politechniki Poznańskiej i Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu). Jest to przedmiot autorski, podczas którego zapoznaję studentów z aktualnymi problemami w dziedzinie badania struktur cząsteczek biologicznych oraz przedstawiam możliwe rozwiązania tych problemów metodami bioinformatycznymi. Liczba studentów: 20. Liczba godzin lekcyjnych: 15.

Od 2002 roku sprawuję opiekę nad pracami magisterskimi i inżynierskimi wykonywanymi przez studentów kierunku Informatyka na Wydziale Informatyki (wcześniej: Wydziale Informatyki i Zarządzania) Politechniki Poznańskiej, natomiast od roku 2012 – również nad pracami licencjackimi i magisterskimi wykonywanymi przez studentów makrokierunku Bioinformatyka (Wydział Informatyki Politechniki Poznańskiej i Wydział Biologii Uniwersytetu im. Adama Mickiewicza w Poznaniu). W tym czasie zaproponowałam 29 tematów prac dyplomowych, w tym 20 tematów prac magisterskich, 6 tematów prac inżynierskich oraz 3 tematy prac licencjackich. W ok. połowie tych prac pełniłam rolę promotora, a w pozostałych – rolę opiekuna pracy i promotora pomocniczego. 24 spośród w/w prac dyplomowych zostało przedstawionych i obronionych na Wydziale Informatyki i Zarządzania lub na Wydziale Informatyki Politechniki Poznańskiej. 5 prac dyplomowych jest w toku, a ich obrona planowana jest na sesję letnią bieżącego roku akademickiego. Poniżej przedstawiam listę tematów wraz z autorami i rokiem obrony (dla prac ukończonych):

A. Prace dyplomowe magisterskie

[1] A. Wójtowicz (2003) Kombinatoryczna analiza wyników eksperymentu NMR.
[2] S. Klemczak (2005) Trzeciorzędowa struktura RNA - problematyka informatyczna.
[3] Ł. Kościński (2006) Analiza algorytmiczna NMR-owskich widm łańcuchów proteinowych.
[4] M. Błażewicz (2008) Wyszukiwarka fragmentów RNA - projekt i implementacja.
[5] W. Paczkowski (2008) Baza drugorzędowych struktur cząsteczek RNA.
[6] M. Miławski (2008) Algorytmy wspomagające rozwiązywanie i wyszukiwanie struktur RNA.
[7] M. Małaczyński (2009) Konstruowanie ścieżek transferowych na widmach NMR.
[8] N. Bartol (2010) RNAComposer: budowanie przestrzennej struktury RNA.
[9] A. Starikiewicz (2010) RNACompare: narzędzie do porównywania struktur RNA.
[10] T. Żok (2011) Nowa metoda porównywania struktur molekularnych.
[11] S. Kaczmarek (2011) Algorytmy modelowania struktury RNA na GPU.
[12] S. Bosy (2011) Repozytorium elementów strukturalnych kwasów nukleinowych - projekt i implementacja.
[13] P. Mulczyński (2011) Statystyka elementów strukturalnych RNA.
[14] J. Porzuczek (2012) Bakteryjne sRNA w mikrobiomie człowieka - bioinformatyczna analiza danych.
[15] K. Janas (2012) Wizualizacja struktury drugorzędowej RNA na ekranie dotykowym.
[16] M. Szewczyk (2012) Wysoko-przepustowa statystyczna analiza parametrów struktur białkowych.
[17] T. Tokarek (-) Bioserwer: administracja systemem i wizualizacja danych.
[18] M. Magowski (-) Animowany tutorial: narzędzia analizy strukturalnej RNA.
[19] J. Kowalska (-) Identyfikacja wzorców strukturalnych w roślinnym mikro-RNA .
[20] K. Rybicka (-) Bioinformatyczna analiza sekwencji i struktur mikro-RNA.

B. Prace dyplomowe inżynierskie

[21] M. Miławski, W. Paczkowski (2007) Automatyczna analiza widm NMR cząsteczek RNA.
[22] Ł. Chojnacki, P. Kiec, G. Skibiński, D. Zieliński (2008) NMR-corner: Internetowa platforma do obliczeń NMR-owskich.
[23] N. Bartol, G. Kaczmarek, A. Starikiewicz, J. Staszak (2009) Baza danych fragmentów RNA.
[24] T. Żok (2010) Porównywanie i łączenie fragmentów RNA.
[25] S. Bosy, K. Klukowski, P. Mulczyński, R. Szulc (2010) RNAComposer: baza danych i system obsługi użytkowników hurtowych.
[26] M. Jackowski, S. Szałata, M. Szewczyk, T. Tokarek (2011) Serwer narzędzi bioinformatycznych.

C. Prace dyplomowe licencjackie

[27] M. Przybyła (2013) Łamigłówki strukturalne w świecie RNA.
[28] J. Kowalska (2013) RNA w szklanej kuli, czyli o przewidywaniu struktur.
[29] W. Tomyślak (-) Tydzień z życia bazy struktur RNA – projekt edukacyjny.

Wiele spośród wymienionych wyżej prac dyplomowych zostało wyróżnionych na Politechnice Poznańskiej (prace nr [1], [4], [5], [7], [8], [10], [16], [24] - obrony z wyróżnieniem). Dwie prace magisterskie, których byłam promotorem otrzymały **wyróżnienia w Konkursie na Najlepszą Pracę Magisterską z Bioinformatyki** organizowanym przez Polskie Towarzystwo Bioinformatyczne. Były to: praca magisterska [4] autorstwa Marka Błażewicza (wyróżnienie w roku 2009) oraz praca magisterska [10] autorstwa Tomasza Żoka (wyróżnienie w roku 2012). W roku 2013, w kolejnej edycji tego samego konkursu przyznano wyróżnienie pracy magisterskiej [16] autorstwa Marka Szewczyka, której byłam opiekunem. Dwie wydane przeze mnie prace magisterskie (nr [14], [16]) zostały zrealizowane w ramach współpracy naukowej z Uniwersytetem w Luksemburgu.

Badania prowadzone w ramach prac dyplomowych nr [1], [2], [4], [7], [10], [15] i [28] były prezentowane na seminariach oraz krajowych i międzynarodowych konferencjach naukowych, m.in. na CIBCB: IEEE Symposium on Computational Intelligence in Bioinformatics and Computational Biology (USA 2004, 2005, 2009), ICOLE'07: Bioinformatics, Operations Research and Machine Learning - Recent Trends and Developments (Austria 2007, 2011), EURO: European Congress on Operational Research (Czechy 2007), RECOMB: Annual International Conference on Research in Computational Molecular Biology (Singapur 2008), Multi-Pole Approach to Structural Biology Conference (Warszawa 2011), Sympozjum Polskiego Towarzystwa Bioinformatycznego (Gdańsk 2012), Bioinformatics'13 (Toruń 2013).

Wyniki badań przedstawionych w pracach magisterskich nr [1], [4], [7], [8] i [10] stały się elementami **publikacji** w Bioinformatics, Nucleic Acids Research, Journal of Heuristics oraz Central European Journal of Operations Research, które są wiodącymi czasopismami z dziedziny bioinformatyki, biologii i badań operacyjnych.

[8] Popularyzacja nauki

Wyniki swoich badań wielokrotnie (**96** razy) prezentowałam na konferencjach i seminariach naukowych w kraju i zagranicą. Po uzyskaniu stopnia doktora (od 2006 roku) przygotowałam i przedstawiłam **72** referaty ustne i **24** prezentacje plakatowe, w tym (stan na 10.03.2015):

- **5** wykładów na zaproszenie, na międzynarodowych konferencjach i seminariach naukowych (*EurOPT, BIT, MultiPole, seminarium na UW, seminarium na UMK*),
- **41** wykładów na międzynarodowych konferencjach zagranicznych (m.in. *EURO, INFORMS, MultiPole, CBBM, ICOLE, EurOPT, IEEE CIBCB, ECCO, GO, AIRO, META, ORP3*),
- **11** wykładów na międzynarodowych konferencjach organizowanych w Polsce (m.in. *BIT, CBBM, COMBI, EURASNET, PTBI*),
- **8** prezentacji plakatowych na międzynarodowych konferencjach zagranicznych (m.in. *RECOMB, ECCB, Bioinformatics, BIOTECHNO, CMLS*),
- **16** prezentacji plakatowych na międzynarodowych konferencjach organizowanych w Polsce (m.in. *BIOCONNECT, BIT*),
- **2** wykłady na seminariach w zagranicznych ośrodkach naukowych,
- **13** wykładów na seminariach w krajowych ośrodkach naukowych.

[9] Współpraca naukowa, udział w projektach badawczych, działalność organizacyjna

W ramach prowadzonych przeze mnie badań współpracuję z naukowcami z innych jednostek i ośrodków badawczych w kraju i zagranicą. Część badań zakończyła się wspólnymi publikacjami, część była przedmiotem zdań realizowanych w ramach międzynarodowych projektów badawczych. Wynikiem współpracy jest również udział w organizacji konferencji, komitetach redakcyjnych, recenzowanie manuskryptów i projektów badawczych, itp. W poniższych zestawieniach przedstawiam wyniki tej współpracy, skupiając się na okresie po uzyskaniu stopnia doktora.

A. Udział w finansowanych projektach badawczych

[1] *Informatyczne aspekty analizy strukturalnej biomolekuł*, grant 3T11F00227, 2004-2007, wykonawca.

[2] *Środowisko i metagenomy – bioinformatyczny system do odkrywania i przypisywania funkcji do wzorców genomowych uwarunkowanych środowiskowo*, Projekt europejski FP6 METAFUNCTIONS, 2005-2008, wykonawca i kierownik zadania.

[3] *Racjonalne projektowanie oraz standaryzowana ewaluacja nowych szczepionek genetycznych*, Projekt europejski FP6 COMPUVAC, 2005-2009, wykonawca.

[4] *Optymalizacja w bioinformatyce*, Projekt europejski FP6 BIOPTRAIN, 2005-2009, kierownik zadania.

[5] *Pierwszorzędowa i przestrzenna struktura kwasów nukleinowych i białek - analiza i algorytmy*, grant NN519314635, 2008-2011, wykonawca.

[6] *Zaawansowane algorytmy bioinformatyczne w biologii molekularnej*, PAN-CNR Joint Research Project, 2010-2012, wykonawca.

[7] *Rozwój Infrastruktury Europejskiego Centrum Bioinformatyki i Genomiki*, grant POIG.02.02-30-009/09, 2010-2012, wykonawca.

[8] *Informatyczne modele i metody w biologii komórkowej*, grant 2012/05/B/ST6/03026, 2013-2015, główny wykonawca.

[9] *Automatyczne, wysokoprzepustowe modelowanie struktur przestrzennych RNA. Integrowanie informatyki z eksperymentem biochemicznym*, grant Maestro 2012/06/A/ST6/00384, 2013-2015, wykonawca.

B. Współpraca w ramach badań naukowych niefinansowanych

[1] Współpraca w zakresie analizy złożoności obliczeniowej problemów biologicznych. Prof. Dominique de Werra, Ecole Polytechnique Federale de Lausanne, Lausanne, Szwajcaria.

[2] Zastosowania algorytmów klastrowania w bioinformatyce strukturalnej. Prof. Thomas Villmann, University of Applied Science, Mittweida, Niemcy.

[3] Analiza właściwości algorytmów dokładnych i przybliżonych. Prof. Erwin Pesch, University of Siegen, Siegen, Niemcy.

[4] Optymalizacja procesów obliczeniowych w sieciach przepływowych. Dr Giovanni Felici, Institute of Systems Analysis and Computer Science "A. Ruberti", National Research Council, Rzym, Włochy.

[5] Eksperymenty laboratoryjne w badaniach egzosomu. Dr Juan M. Falcon-Perez, Bizkaia Technology Park, Bizkaia, Hiszpania.

C. Udział w komitetach organizacyjnych i programowych konferencji, organizacja sesji

[1] Przewodnicząca Komitetu Organizacyjnego konferencji IV EURO WG Conference on Operational Research in Computational Biology, Bioinformatics and Medicine, Poznań-Biedrusko, 26-28.06.2014.

[2] Członek Komitetu Organizacyjnego **5** konferencji krajowych i międzynarodowych (VI Sympozjum PTBI 2013, CBBM 2012, I Polish-Italian Workshop on Bioinformatics 2010, II Sympozjum PTBI 2009, COMBI 2007).

[3] Członek Komitetu Programowego **8** konferencji międzynarodowych (CBBM 2014, PRIB 2013 i 2014, CIBB 2013, IEEE CIBCB 2007, 2008 i 2013, CEC 2009).

[4] Organizator strumieni i sesji tematycznych na **9** konferencjach międzynarodowych (EURO 2009, 2010, 2012, 2013 i 2015, ECCO 2011 i 2012, IFORS 2011, EurOPT 2008).

[5] Przewodnicząca sesji na **16** konferencjach międzynarodowych (GO 2014, V Sympozjum PTBI 2012, IFORS 2011, CBBM 2008 i 2014, EURO 2009 i 2010, EurOPT 2008, COMBI 2007, ICOLE 2007, 2008, 2010, 2011, 2012, 2013 i 2014).

D. Udział w komitetach redakcyjnych czasopism

Edytor specjalnego wydania RAIRO-Operations Research: Operations Research in Computational Biology, Bioinformatics and Medicine (2015).

E. Recenzowanie publikacji, prac konkursowych i projektów badawczych

[1] Recenzent dla **5** międzynarodowych czasopism z dziedziny bioinformatyki, informatyki oraz biologii obliczeniowej: Nucleic Acids Research (7 recenzji), Central European Journal of Operations Research (2 recenzje), IJOCTA - An International Journal of Optimization and Control: Theories and Applications (1 recenzja), Source Code for Biology and Medicine (2 recenzje), TOP – Journal of the Spanish Society of Statistics and Operations Research (1 recenzja).

[2] Recenzent dla **11** konferencji międzynarodowych: PRIB 2014, CIBB 2013, PRIB 2013, CEC 2011, CEC 2010, IEEE CIBCB 2010, IEEE CIBCB 2009, EurOPT 2008, PRIB 2007, CEC 2007, CEC 2006 (sumarycznie 23 recenzje).

[3] Recenzent **2** wniosków NSF (National Science Foundation, USA): "Pattern Look-Up: RNA Search Engine Case Study" (A. Arslan, 2011), "BIGDATA: Collaborative Research: An RNA Comprehensive Database and Analysis Infrastructure" (R.R. Gutell, A. Arslan, 2012).

[4] Recenzent w Konkursie Polskiego Towarzystwa Bioinformatycznego na Najlepszą Pracę Magisterską z Bioinformatyki (1 recenzja).

F. Staże i wizyty naukowe

[1] 2011, 2013. The Commonwealth Scientific and Industrial Research Organisation (CSIRO), Clayton, Australia. 2 wizyty naukowe (łącznie na okres 3 tygodni).

[2] 2010-2012. Institute of Systems Analysis and Computer Science "A. Ruberti", National Research Council, Rzym, Włochy. 3 wizyty naukowe (łącznie na okres 3 tygodni).

G. Członkostwo w organizacjach i towarzystwach naukowych

[1] Polskie Towarzystwo Bioinformatyczne. Członek założyciel (od 2008r.).

[2] Polskie Towarzystwo Bioinformatyczne. Wiceprezes (od 2011r., obecnie II kadencja).

[3] EURO Working Group on Operational Research in Computational Biology, Bioinformatics and Medicine. Członek założyciel (od 2006r.).

[4] EURO Working Group on Operational Research in Computational Biology, Bioinformatics and Medicine. Członek Zarządu (od 2009r.).

[5] IEEE Computational Intelligence Society. Członek (2005-2007).

[6] IEEE Women in Engineering. Członek (2005-2007).

[10] Nagrody i wyróżnienia

[1] Nagroda Asocjacji Europejskich Towarzystw Badań Operacyjnych (EURO) za Najlepszą Rozprawę Doktorską w Europie w dziedzinie badań operacyjnych (2006).

[2] Nagroda JM Rektora Politechniki Poznańskiej za wyróżniającą się rozprawę doktorską (2006).

[3] Nagroda Rady Naukowej Instytutu Chemii Bioorganicznej PAN w Poznaniu za najlepszą pracę eksperymentalną roku – RNA FRABASE (2009).

[4] Nagroda Rady Naukowej Instytutu Chemii Bioorganicznej PAN w Poznaniu za najlepszą pracę eksperymentalną roku – RNAComposer (2013).

[5] Wyróżnienie w Konkursie Polskiego Towarzystwa Bioinformatycznego na Najlepszą Pracę Magisterską przyznane za pisaną pod moim kierunkiem pracę Marka Błażewicza pt. "Wyszukiwarka fragmentów RNA – projekt i implementacja" (2009).

[6] Wyróżnienie w Konkursie Polskiego Towarzystwa Bioinformatycznego na Najlepszą Pracę Magisterską przyznane za pisaną pod moim kierunkiem pracę Tomasz Żoka pt. "Nowa metoda porównywania struktur molekularnych" (2012).

[7] Wyróżnienie w Konkursie Polskiego Towarzystwa Bioinformatycznego na Najlepszą Pracę Magisterską przyznane za pisaną pod moją opieką (współpromotor) pracę Marka Szewczyka pt. "Wysoko-przepustowa statystyczna analiza parametrów struktur białkowych" (2013).

Marta Szachniuk
.....
(podpis)