

Wsparcie aktywności naukowej doktorantów, studentów i  
Studenckich Kół Naukowych- Wydział Biotechnologii,  
Uniwersytet Wrocławski  
01.12.2015-15.12.2015  
Sprawozdanie

Michał Burdukiewicz

Dzięki środkom finansowym przyznanych przez Konsorcjum Wrocławskie Centrum Biotechnologii Doktoranckie Koło Naukowe Bioinformatyki zrealizowało projekt “Predykcja białek amyloidogennych”.

W trakcie badań opracowano probabilistyczny model miejsc amyloidogennych za pomocą metod analizy języka naturalnego. Stworzony model nie tylko w intuicyjny sposób opisuje strukturę miejsc inicjacji agregacji, ale również pozwala przewidywać białka amyloidowe. Dokładność predykcji naszego programu jest dużo większa niż innych istniejących programów tego typu. Przeprowadzenie obliczeń, niezbędnych do realizacji tego zadania badawczego, nie byłoby możliwe, gdyby nie modernizacja istniejącej infrastruktury dzięki środkom finansowym przyznanych przez Konsorcjum.

W ramach projektu przygotowano następujące komunikaty konferencyjne w postaci posteru:

1. Burdukiewicz M., Sobczyk P., Mackiewicz P., Kotulska M., *Amylo-Gram: n-gram analysis and prediction of amyloids*, **Bioinformatics in Toruń**, 16-18 czerwiec 2016, Toruń (Polska) (zał. 1)
2. Burdukiewicz M., Sobczyk P., Rödiger S., Mackiewicz P., Kotulska M., *AmyloGram: a novel predictor of amyloidogenicity*, **European Conference on Computational Biology**, 3-7 września 2016, Haga (Holandia) (zał. 2)

Wyniki badań zostaną również przedstawione w postaci wystąpienia podczas kongresu **German Conference on Bioinformatics** (12-15 września)

w Berlinie (Niemcy) (zał. 3) i opublikowane w **PeerJ Preprints**. Niezależnie trwa przygotowanie manuskryptu opisującego otrzymane rezultaty.

Prezentacja wyników na tak szeroką skalę nie byłaby możliwa, gdyby nie dofinansowanie wyjazdów zagranicznych przez Konsorcjum. Dzięki wsparciu finansowemu mamy możliwość przedyskutowania naszych rezultatów z wieloma naukowcami, co pozwoliło nam przeprowadzić znacznie ulepszyć stosowane metody badawcze.

## ZAŁĄCZNIK 1



Michał Burdukiewicz <michalburdukiewicz@gmail.com>

---

### poster bit 16 ok

1 wiadomość

wiesiek@fizyka.umk.pl <wiesiek@fizyka.umk.pl>  
Do: michalburdukiewicz@gmail.com

1 czerwca 2016 22:17

Szanowny Panie,

miło mi poinformować, iż komunikat pt.

AmyloGram: n-gram analysis and prediction of amyloids  
by

Michał Burdukiewicz<sup>1\*</sup>, Piotr Sobczyk<sup>2</sup>, Paweł Mackiewicz<sup>1</sup>, Małgorzata Kotulska

został zakwalifikowany na konferencję Bioinformatics in Torun do prezentacji  
w formie posteru. Numer posteru i sesji zostanie podany w czasie rejestracji.

Do zobaczenia w Toruniu.

Z poważaniem

Wiesław Nowak

Przewodniczący Komitetu Programowego

## ZAŁĄCZNIK 2



Michał Burdukiewicz <michalburdukiewicz@gmail.com>

---

### ECCB 2016 Notification of Poster presentation - 452

---

**ECCB 2016** <eccb2016@easychair.org>  
Do: Michał Burdukiewicz <michalburdukiewicz@gmail.com>

1 lipca 2016 16:55

Dear Michał,

It is our pleasure to inform you that your abstract AmyloGram: a novel predictor of amyloidogenicity with submission number 452 has been accepted for a poster presentation at ECCB 2016.

You will receive additional information by email if you submitted your poster for one of the sub tracks (ELIXIR, Training or Applications).

You can find more information about the poster presentations on the poster presenter page: <http://www.eccb2016.org/programme/posters/>

You can register for the conference via: <https://www.eccb2016.org/registration/>. Please notice that the early bird registration deadline is 29 July 2016!

Please note that we will open a F1000 channel for the ECCB 2016! The channel will be online on July 22nd and open for your poster submissions. More information about this will be available on our website soon.

If you need an invitation letter to be able to obtain a visa or funding, please follow the following procedure:

- \* register for the conference: you can create your own visa or presentation invitation letter via the registration procedure, but you will have to register first;
- \* to get access to your invitation letter, please click the 'Modify Registration' link stated in the confirmation email that will be sent to you after finishing your registration;
- \* you will find a 'Letter of Invitation' button at the bottom of the page. Follow this and you will find a form that you should fill out, after which you will be able to print your invitation letter.

If for any reason you decide to withdraw your poster presentation, please let us know via [posters@eccb2016.org](mailto:posters@eccb2016.org).

We are looking forward to seeing you in The Hague!

Cheers,

Matti Nykter  
ECCB 2016 Poster chair

For further questions, please contact: [posters@eccb2016.org](mailto:posters@eccb2016.org)

## ZAŁĄCZNIK 3



Michał Burdukiewicz <michalburdukiewicz@gmail.com>

### GCB 2016 notification for paper 29

1 wiadomość

GCB 2016 <gcb2016@easychair.org>  
Do: Michał Burdukiewicz <michalburdukiewicz@gmail.com>

6 lipca 2016 13:36

Dear Michał Burdukiewicz Burdukiewicz

thank you submitting your work

Prediction of amyloidogenicity based on the n-gram analysis  
to GCB 2016.

We are pleased to let you know that your paper has been accepted for presentation at GCB 2016 in Berlin. Attached you will find the reviewers comments. Please address them and submit the final version before August 7th.

Yours,

Bernhard Renard, Knut Reinert, Joachim Selbig

#### ----- REVIEW 1 -----

PAPER: 29  
TITLE: Prediction of amyloidogenicity based on the n-gram analysis  
AUTHORS: Michał Burdukiewicz, Piotr Sobczyk, Stefan Rödiger, Anna Duda-Madej, Paweł Mackiewicz and Małgorzata Kotulska

OVERALL EVALUATION: 2 (accept)  
REVIEWER'S CONFIDENCE: 4 (high)

#### ----- Review -----

The authors use a machine learning method to detect amyloidogenic peptides. Specifically, they select n-grams of encoded peptides (by physicochemical properties) and subsequently train a predictor using random forests. Prediction accuracy is then assessed and compared with other methods.

#### General:

The work fits well into the scope of the GCB'16 and should be accepted for the proceedings once the following revisions have been made.

- 1) There are many grammatical mistakes and essential descriptions of variables and formulas are missing. Normally this would be a minor issue, but the list is quite long (see at the end of the document).
- 2) The threshold for the classification as amyloidogenic is currently 0.5 as far as I understood. First, this is nowhere mentioned. Secondly, this may lead to many false positives/false negatives. You may think to use more conservative thresholds, i.e. akin to common statistics, a peptide would be predicted to be amylogenic (with >95% confidence) if the probability of not being amyloidogenic is < 0.05. Vice versa, a peptide may not be amyloidogenic with 95% confidence if the probability of being amyloidogenic is < 0.05. This leaves you with a range, where you cannot make conclusive statements, but this is ok. Please comment, or change. Also, multiple test correction (e.g. Benjamini-Hochberg Method) could be useful in the aforementioned context.
- 3) Quick permutation test (page 4): Please explain which hypothesis the p-value actually assesses/how it is computed.
- 4) The authors state that amylogram creates a highly interpretable outcome, however they fail to interpret it sufficiently. This part should be extended. The introduction and sections 3.1-3.3 can be shortened to accommodate for this additional discussion.

#### Minor:

- a) The tool doesn't work with data provided on github (e.g. /data/amyloid\_neg\_full.fasta). error: "n-gram too long". I guess the removal of the short sequences is not implemented.
- b) Page 5 under the contingency table: Shouldn't it be the product of two binomials, if x & y are independent, not multinomial?
- c) The classification and the training data is currently binary (amyloidogenic vs. non amyloidogenic). Which always creates a bias with regard to the actual processes being modelled. May it be useful, if not more accurate, to use a