

**dr hab. Anna Gambin**

## **Curriculum Vitae**

### **Personal data:**

- Born 7.06.1971.
- Married (1995),
- three children: Gabriela (2000), Richard (2003), Andrew (2007).

### **Education and scientific degrees**

- October 1990 - May 1995: computer science in the Faculty of Mathematics, Informatics and Mechanics (MIM in short), Warsaw University. MSc in Computer Science with honours (05.1995), thesis title: *Shared-Memory Simulation on a Faulty Distributed Memory Machine*.
- October 1995 – December 1999: PhD studies in Warsaw University. PhD in computer science with honours (01.2000), thesis title: *Combinatorial Methods in Approximation Algorithms for Markov Chains with Large State Space*.
- October 2008: habilitation in the Faculty of MIM, Warsaw University. Habilitation title: *Contextual alignment of molecular sequences*.

### **Academic work experience**

- October 1997 – September 2001: research assistant in Institute of Informatics, Warsaw University.
- October 2001 – September 2002: post-doctoral fellow in CNRS - Centre de Génétique Moléculaire, Gif-sur-Yvette, France.
- October 2002 onwards: assistant professor in Institute of Informatics, Warsaw University.

### **Research interest**

Computational Molecular Biology: gene networks (mathematical modelling of signalling pathways, inferring gene regulation networks); peptide mass spectrometry (MS sample classification in medical diagnosis, biomarker selection techniques, clustering methods); sequences analysis (contextual alignment, statistical significance of the alignment, comparative genomics of transposable elements, seed design for protein alignment); mathematical modelling of protein folding; phylogenetics (efficient tree comparison techniques, metrics on the space of phylogenies); detecting genomic aberrations (aCGH).

## Scientific leadership potential

**Interdisciplinary projects.** The cooperation with the Mass Spectrometry Lab. stimulated my leadership skills – I have supervised a large interdisciplinary project: 9 graduate students, in close cooperation with physicians from *Warsaw Cancer Center, Institute of Mother and Child* and molecular biologists from Institute of Biochemistry and Biophysics, Polish Acad. of Sci. We have design new algorithms and developed complete set of tools for processing and analysis of MS data: starting from raw spectra interpretation (detection of peptide signals) through various biomarker selection methods, concluding with classification task (which supports medical diagnosis). Our results are presented in several papers published in high-impact international journals: [2],[3], [5], [8], [10], [12], [16].

**International cooperation.** My interdisciplinary cooperation started during my post-docs stay at Center of Molecular Genetics (CNRS, Gif-sur-Yvette). The research on contextual alignment for protein sequences was conducted in close cooperation with the group of prof. P. Słonimski (CNRS) (see [13], [18], [19], [20]).

My further interest in sequence analysis resulted in scientific projects on comparative genomics of transposon sequences. These are both international (Polonium and Eco-Net projects with LIFL, Lille - the group of prof. G. Kucherov) and interdisciplinary (cooperation with Agricultural University of Cracow on plant genomes and with Institute of Oceanology, Polish Academy of Sciences, on fish genomes) (see [1], [11]).

Second project conducted with the french group aimed in designing seeds for protein homology search. We published our results in [4], [6].

Additionally, I initiated the international cooperation with:

- prof. P.Stankiewicz, (*Baylor College of Medicine, Houston*– on analysis of comparative genome hybridization data),
- prof. M. Vingron (Max Planc Institute for Molecular Genetics, Berlin – on modelling and analysis of transcriptional regulation),
- prof. A. Czumaj (Centre for Discrete Mathematics and its Applications, Warwick – on algorithms for huge biological networks),
- dr R. Haluk (Pacific Northwest National Laboratory, Richland – on modelling signalling pathways) and
- prof. M. Kwiatkowska (Oxford University Computing Laboratory – on applications of model checking to systems biology).

**PhD supervision.** Currently I supervise four PhD students at my faculty. My first student, Bogusław Kluge, is the author of three high-impact journal paper on algorithms and statistical modelling from MS data. Mikołaj Rybiński works on analysis of signalling pathways. He has presented the results concerning JAK-STAT

pathway on CMSB conference. Maciej Sykulski has started working on analysis of high-throughput aCGH data.

### **Selected scientific visits**

- Laboratoire d'Informatique Fondamentale de Lille (LIFL), Lille, France, 2006;
- LORIA, Nancy, France, 2005, 2004, 2001;
- Limburg University Center, Belgium, 2004;
- Center for Molecular Genetics (CNRS), Gif sur Yvette, France, 2001 – 2002;
- Max Planck Institute für Informatik, Sarbrücken, Germany, 2000;
- University Marne la Vallee, Paris, 1998;
- Heinz Nixdorf Institute, Paderborn, Germany, 1997;
- Dortmund University, Germany, 1995 – 1996.

### **Teaching (graduate courses)**

*Statistical analysis of molecular data ; Advanced methods of Computational Molecular Biology ; Bioinformatics: a practical approach ; Introduction to Computational Biology ; Phylogenetics trees .* (I am the author of the program of these lectures.)

### **Graduate students**

Co-supervisor of a seminar for graduate students: *Computational Molecular Biology*. Over 20 MSc dissertations were completed under my supervision.

### **Recent international projects participation**

- Polish-Flemish project *Biological Databases* (2004–2005)
- site leader in the Polish-French project Polonium *Détection et analyse phylogénétique des répétitions dans les séquences protéiques* (2005)
- site leader in the Egide Eco-Net project *Analyse comparative des sequences biologiques* (2005–2006)
- NEUPROCF European Union LSHG-CT-2005-512044 grant *Development of new technologies for low abundance proteomics: application to cystic fibrosis* (2006-2008)

## Other activities:

- Reviewer for: Bioinformatics, BMC Bioinformatics, J. of Mathematical Biology, Bulletin Polish Acad. of Sci., Fundamenta Informaticae, RECOMB, ECCB, WABI, SOFSEM, STACS, ICALP, WEA, ESA, SIRROCO, LATA, FCT, MFCS, Polish Scientific Publisher PWN
- Reviewer of projects supported by Polish Ministry of Science and Higher Education.
- Founding member of Polish Bioinformatics Society.
- Member of organising committee of ACM Regional Collegiate Programming Contest,
- Cooperation with Polish Children's Fund (lectures, reviews)

## Funding ID

I am currently a co-investigator in the following grants:

- Ministry of Science and Higher Education grant PBZ-MNiI-2/1/2005 *Application of modern functional genomics and bioinformatics to analysis and modelling biological processes with fundamental significance in medicine and agriculture* (2006-2009).
- Ministry of Science and Higher Education grant N N206 356036 *Mathematical modeling and algorithmic analysis of molecular signaling pathways.*
- Ministry of Science and Higher Education grant N N301 013436 *Comparative genomics of plant transposons.*

## Recent invited presentations

- talk at Dagstuhl Seminar *Formal Methods in Molecular Biology*, February 2009
- talk at Dagstuhl Seminar *Managing and Mining Genome Information: Frontiers in Bioinformatics*, November 2005
- invited lecture given at the Opening Ceremony of the Academic Year 2006-2007, Faculty of MIM, Warsaw University, October 2006
- invited talk on 190th Anniversary of Warsaw University, December 2004

## Program committees

- Berlin-Warsaw Computational Biology Workshop, Warsaw, Poland, 2009.
- COMBI '07: IV COMputational BIology Workshop, Uniejow, Poland, 2007.
- Berlin-Warsaw Computational Biology Workshop, Berlin, Germany, 2007.
- COMBI '06: IV COMputational BIology Workshop, Jadwisin, Poland, 2006.
- Annual Conference on Current Trends in Theory and Practice of Informatics SOFSEM'2000, Milovy, Czech Republic, 2000.
- member of the program committee of Interdisciplinary PhD Program, Warsaw University.
- founding member of Polish Bioinformatics Society.

## Scholarships and Special Achievements:

- scientific scholarship from the rector of the Warsaw University (2003-2008)
- the NATO Science Fellowship (2002)
- postdoctoral fellowship from the CNRS within the framework of the *Centre Franco-Polonais de Biotechnologie des Plants* (2001)
- the Annual Stipend for Young Scientists from the *Foundation for Polish Science* (2000/2001)
- distinction of the PhD thesis by the Scientific Board of the Faculty of MIM, Warsaw University (2000)
- one-year PhD scholarship from the *German Academic Exchange Service (DAAD)* (1995-1996)
- scholarship of the EU program ALTEC (1993-1995)

## Main scientific accomplishments

***Parallel computing and Markov modelling.*** My MSc thesis was devoted to studying different models of parallel computing, in particular those exhibiting memory faults. The thesis was based on results published at best european conferences (ICALP and ESA). These were pioneering papers, highly cited later on. Then I switched to approximation algorithms, analysis of very large Markov chains, and applications in communication networks. The most interesting results in my PhD thesis were new combinatorial methods in approximation algorithms for Markov Chains with large state space, especially the efficient combinatorial aggregation method. These results and accompanying algorithms went significantly beyond the current state-of-the-art and were published in good international conferences and journals. My paper on application of this method to Google matrix problem has been recently published in SIAM Journal of Scientific Computing.

***Molecular sequences analysis.*** My excellent knowledge of mathematical modelling gave me a very good starting point to work in computational biology. At the beginning my research interest focused on the molecular sequence analysis. In my works a new model of sequence alignment was proposed, which takes into account context-dependency. The efficient analogues of Smith-Watermann and BLAST algorithms was designed together with the rigorous analysis of the statistical significance (papers published in Bioinformatics, Journal of Computational Biology, and other high quality international journals).

***High-throughput data analysis.*** Recently I am working mostly on inferring gene regulation networks, mathematical modelling of signalling pathways and peptide mass spectrometry. Since 2002 I cooperate intensively with the Mass Spectrometry Laboratory, Polish Academy of Sciences, on projects aiming at medical diagnosis methods based on peptide mass spectrometry. My research contributions include: (i) efficient algorithm for mass spectra reduction and interpretation, (ii) mathematically sound model of data clustering, (iii) new method for consensus biomarker selection based on rank aggregation. To perform the classification, besides using few of best known classifiers (like SVM, random forests), I proposed an original boosting method based on the Hannan-consistent game playing strategies.

In my work I confront major challenges in the field of MS analysis (algorithms for multidimensional mass spectra interpretation and spectra alignments) and propose new and original ideas in statistical data analysis, e.g., biomarker consensus method based on Markov chain aggregation.

My experience with analysis of microarray data comprise both transcriptomic and comparative genomic platform. Recently, I have started close cooperation with *Institute of Mother and Child* and *Baylor College of Medicine, Houston* on detection of genomic aberration causing different diseases (e.g. autism, mental retardation, epilepsy).

## Selected recent publications (last 4 years)

- [1 ] Dariusz Grzebelus, Mirosława Gładysz, Alicja Macko-Podgórnica, Tomasz Gambin, Barbara Golis, Roksana Rakoczy, Anna Gambin, *Population dynamics of miniature inverted-repeat transposable elements (MITEs) in *Medicago truncatula**, **Gene**, 2009.
- [2 ] Bogusław Kluge, Wojciech Niemiro, Anna Gambin, *Modeling exopeptidase activity from LC-MS data*, **Journal of Computational Biology** 16 (2), pp. 395-406, 2009.
- [3 ] Marta Łuksza, Bogusław Kluge, Jerzy Ostrowski, Jakub Karczmarski, Anna Gambin, *Two-Stage Model-Based Clustering for Liquid Chromatography Mass Spectrometry Data Analysis*, **Statistical Applications in Genetics and Molecular Biology**, 2009 Vol. 8 : Iss. 1, Article 15.
- [4 ] Eugenia Furletova, Anna Gambin, Gregory Kucherov, Sławomir Lasota, Laurent Noé, Mikhail Roytberg, Ewa Szczurek, *On subset seeds for protein alignment*, **IEEE/ACM Transactions on Computational Biology and Bioinformatics**, 6(3), pp. 483-494, 2009.
- [5 ] Anna Gambin, Ewa Szczurek, Janusz Dutkowski, Magda Bakun, Michał Dadlez *Classification of peptide mass fingerprint data by novel no-regret boosting method*, **Computers in Biology and Medicine**, 39(5):460-73, 2009.
- [6 ] Mikhail A. Roytberg, Eugenia Furletova, Anna Gambin, Gregory Kucherov, Sławomir Lasota, Laurent Noé, Ewa Szczurek, *Efficient Seeding Techniques for Protein Similarity Search*. BIRD, pp. 466-478, 2008.
- [7 ] Anna Gambin, Piotr Krzyżanowski, Piotr Pokarowski, *Aggregation Algorithms for Perturbed Markov Chains with Applications to Networks Modeling*. **SIAM J. on Sci. Comput.** 31(1), pp. 45-73, 2008.
- [8 ] Magdalena Bakun, Michał Dadlez, Janusz Dutkowski, Anna Gambin, Jakub Karczmarski, Bogusław Kluge, Krzysztof Kowalczyk, Jerzy Ostrowski, Jarosław Poznański, Jerzy Tiuryn, *Automated reduction and interpretation of multidimensional mass spectra for analysis of complex peptide mixtures*. **Int. J. Mass Spectrometry** 260(1), pp. 20-30, 2007.
- [9 ] Łukasz Bolikowski, Anna Gambin, *New Metrics for Phylogenies*. **Fund. Inform.** 78(2), pp. 199-216, 2007.
- [10 ] Janusz Dutkowski, Anna Gambin, *On consensus biomarker selection*. **BMC Bioinformatics** 8(S5), 2007.
- [11 ] Anna Gambin, Tomasz Gambin, Dariusz Grzebelus, Gregory Kucherov, Sławomir Lasota, *Diversity and structure of PIF/Harbinger-like elements in the genome of *Medicago Truncatula**. **BMC Genomics** 8(409), 2007.

- [12] | Anna Gambin, Ewa Szczurek, *No-Regret Boosting*. ICANNGA, pp. 422-431, 2007.
- [13] | Anna Gambin, Piotr Wojtalewicz, *CTX-BLAST: context sensitive version of protein BLAST*. **Bioinformatics** 23(13), pp. 1686-1688, 2007.
- [14] | Katarzyna Bożek, Anna Gambin, Jerzy Tiuryn, Bartosz Wilczyński, *Automated modeling of genetic control in Arabidopsis thaliana*. **J. Fruit and Ornamental Plant Research** 14(S1), pp. 163-171, 2006.
- [15] | Norbert Dojer, Anna Gambin, Andrzej Mizera, Jerzy Tiuryn, Bartosz Wilczyński, *Applying dynamic Bayesian networks to perturbed gene expression data*. **BMC Bioinformatics** 7(249), 2006.
- [16] | Anna Gambin, Jakub Karczmarski, Bogusław Kluge, Marta Łuksza, Jerzy Ostrowski, *Efficient Model-Based Clustering for LC-MS Data*. WABI, pp. 32-43, 2006.
- [17] | Anna Gambin, Sławomir Lasota, M. Rutkowski, *Analyzing stationary states of gene regulatory network using Petri nets*. **In Silico Biology** 6(0010), 2006.
- [18] | Anna Gambin, Jerzy Tiuryn, Jerzy Tyszkiewicz, *Alignment with context dependent scoring function*. **J. Computational Biology** 13(1), pp. 81-101, 2006.
- [19] | Anna Gambin, Rafał Otto, *Contextual Multiple Alignment (Context helps in aligning orphan genes)*. **J. Biomedicine and Biotechnology** 2, pp. 124-131, 2005.
- [20] | Anna Gambin, Piotr Słonimski, *Hierarchical clustering based upon contextual alignment of proteins: a different way to approach phylogeny*. **C.R. Biologie Acad. Science Paris** 328, pp. 11-22, 2005.
- [21] | Anna Gambin, Damian Wójtowicz, *Almost FPRAS for Lattice Models of Protein Folding*. WEA, pp. 534-544, 2005.