BIOTECHNO 2012

The Fourth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies

ISBN: 978-1-61208-190-8

March 25-20, 2012

St. Maarten, Netherlands Antilles

BIOTECHNO 2012 Editors

Petre Dini, Concordia University, Canada / China Space Agency Center-Beijing, China

Pascal Lorenz, University of Haute Alsace, France

BIOTECHNO 2012

Foreword

The Fourth International Conference on Bioinformatics, Biocomputational Systems and Biotechnologies (BIOTECHNO 2012), held between March 25-30, 2012 - St. Maarten, Netherlands Antilles, covered these three main areas: bioinformatics, biocomputational systems, and biotechnologies.

Bioinformatics deals with the system-level study of complex interactions in biosystems providing a quantitative systemic approach to understand them and appropriate tool support and concepts to model them. Understanding and modeling biosystems requires simulation of biological behaviors and functions. Bioinformatics itself constitutes a vast area of research and specialization, as many classical domains such as databases, modeling, and regular expressions are used to represent, store, retrieve and process a huge volume of knowledge. There are challenging aspects concerning biocomputation technologies, bioinformatics mechanisms dealing with chemoinformatics, bioimaging, and neuroinformatics.

Brain-computing, biocomputing, and computation biology and microbiology represent advanced methodologies and mechanisms in approaching and understanding the challenging behavior of life mechanisms. Using bio-ontologies, biosemantics and special processing concepts, progress was achieved in dealing with genomics, biopharmaceutical and molecular intelligence, in the biology and microbiology domains. The area brings a rich spectrum of informatics paradigms, such as epidemic models, pattern classification, graph theory, or stochastic models, to support special biocomputing applications in biomedical, genetics, molecular and cellular biology and microbiology. While progress is achieved with a high speed, challenges must be overcome for large-scale bio-subsystems, special genomics cases, bio-nanotechnologies, drugs, or microbial propagation and immunity.

Biotechnology is defined as the industrial use of living organisms or biological techniques developed through basic research. Bio-oriented technologies became very popular in various research topics and industrial market segments. Current human mechanisms seem to offer significant ways for improving theories, algorithms, technologies, products and systems. The focus is driven by fundamentals in approaching and applying biotechnologies in terms of engineering methods, special electronics, and special materials and systems. Borrowing simplicity and performance from the real life, biodevices cover a large spectrum of areas, from sensors, chips, and biometry to computing. One of the chief domains is represented by the biomedical biotechnologies, from instrumentation to monitoring, from simple sensors to integrated systems, including image processing and visualization systems. As the state-of-the-art in all the domains enumerated in the conference topics evolve with high velocity, new biotechnologes and biosystems become available. Their rapid integration in the real life becomes a challenge.

We welcomed technical papers presenting research and practical results, position papers addressing the pros and cons of specific proposals, such as those being discussed in the standard forums or in industry consortia, survey papers addressing the key problems and solutions on any of the above topics short papers on work in progress, and panel proposals.

We take here the opportunity to warmly thank all the members of the BIOTECHNO 2012 technical program committee as well as the numerous reviewers. The creation of such a broad and high quality conference program would not have been possible without their involvement. We also kindly thank all the authors that dedicated much of their time and efforts to contribute to BIOTECHNO 2012. We truly believe that, thanks to all these efforts, the final conference program consisted of top quality contributions.

We hope that BIOTECHNO 2012 was a successful international forum for the exchange of ideas and results between academia and industry and to promote further progress in biotechnology.

We are certain that the participants found the event useful and communications very open. The beautiful places of St. Maarten surely provided a pleasant environment during the conference and we hope you had a chance to visit the surroundings.

BIOTECHNO 2012 Chairs
Stephen Anthony, The University of New South Wales, Australia
Petre Dini, Concordia University, Canada / China Space Agency Center-Beijing, China

BIOTECHNO 2012

Committee

BIOTECHNO Advisory Chairs

Stephen Anthony, The University of New South Wales, Australia
Petre Dini, Concordia University, Canada / China Space Agency Center-Beijing, China

BIOTECHNO 2012 Technical Program Committee

Basim Alhadidi, Albalqa' Applied University - Salt, Jordan
Stephen Anthony, The University of New South Wales, Australia
Ganesharam Balagopal, Ontario Ministry of the Environment - Toronto, Canada
Siegfried Benkner, University of Vienna, Austria
Gilles Bernot, University of Nice Sophia Antipolis, France
Tom Bersano, University of Michigan, USA
Christian Blum, Universitat Politècnica de Catalunya - Barcelona, Spain
Razvan Bocu, University of Brasov, Romania
Magnus Bordewich, Durham University, UK
Sabin-Corneliu Buraga, "A. I. Cuza" University - Iasi, Romania
Yang Cao, Virginia Tech – Blacksburg, USA
Yili Chen, Monsanto Company - St. Louis, USA
Eugen Czeizler, Aalto University,Finland
Rolf Drechsler, University of Bremen, Germany
Lingke Fan, University Hospitals of Leicester NHS Trust, UK
Victor Felea, "Al.I. Cuza" University - Iasi, Romania
Jerome Feret, INRIA, France
Xin Gao, KAUST (King Abdullah University of Science and Technology), Saudi Arabia
Alejandro Giorgetti, University of Verona, Italy
Paul Gordon, University of Calgary, Canada
Radu Grosu, Stony Brook University – New York, USA
Jun-Tao Guo, The University of North Carolina at Charlotte, USA
Mahmoudi Hacene, University Hassiba Ben Bouali – Chlef, Algeria
Saman Kumara Halgamuge, University of Melbourne, Australia
Steffen Heber, North Carolina State University-Raleigh, USA
Attila Kertesz-Farkas, International Centre for Genetic Engineering and Biotechnology, Trieste, Italy
Daisuke Kihara, Purdue University - West Lafayette, USA
DaeEun Kim, Yonsei University - Seoul, South Korea
Marco Lübbecke, RWTH Aachen University, Germany
José Luis Oliveira, University of Aveiro, Portugal
Roger Mailler, The University of Tulsa, USA
Igor V. Maslov, EvoCo Inc. - Tokyo, Japan
Bud Mishra, NYU, USA
José Manuel Molina López, Universidad Carlos III de Madrid, Spain
Julián Molina, University of Malaga, Spain
Chris J. Myers, University of Utah, USA

Copyright Information

For your reference, this is the text governing the copyright release for material published by IARIA.

The copyright release is a transfer of publication rights, which allows IARIA and its partners to drive the dissemination of the published material. This allows IARIA to give articles increased visibility via distribution, inclusion in libraries, and arrangements for submission to indexes.

I, the undersigned, declare that the article is original, and that I represent the authors of this article in the copyright release matters. If this work has been done as work-for-hire, I have obtained all necessary clearances to execute a copyright release. I hereby irrevocably transfer exclusive copyright for this material to IARIA. I give IARIA permission or reproduce the work in any media format such as, but not limited to, print, digital, or electronic. I give IARIA permission to distribute the materials without restriction to any institutions or individuals. I give IARIA permission to submit the work for inclusion in article repositories as IARIA sees fit.

I, the undersigned, declare that to the best of my knowledge, the article is does not contain libelous or otherwise unlawful contents or invading the right of privacy or infringing on a proprietary right.

Following the copyright release, any circulated version of the article must bear the copyright notice and any header and footer information that IARIA applies to the published article.

IARIA grants royalty-free permission to the authors to disseminate the work, under the above provisions, for any academic, commercial, or industrial use. IARIA grants royalty-free permission to any individuals or institutions to make the article available electronically, online, or in print.

IARIA acknowledges that rights to any algorithm, process, procedure, apparatus, or articles of manufacture remain with the authors and their employers.

I, the undersigned, understand that IARIA will not be liable, in contract, tort (including, without limitation, negligence), pre-contract or other representations (other than fraudulent misrepresentations) or otherwise in connection with the publication of my work.

Exception to the above is made for work-for-hire performed while employed by the government. In that case, copyright to the material remains with the said government. The rightful owners (authors and government entity) grant unlimited and unrestricted permission to IARIA, IARIA's contractors, and IARIA's partners to further distribute the work.

Table of Contents

A Coupled Three-Step Network-Based Approach to Identify Genes Associated with Breast Cancer

Michael Netzer*, Xiaocong Fang†, Michael Handler*, Christian Baumgartner*
*Research Group for Clinical Bioinformatics,
Institute of Electrical, Electronic and Bioengineering, UMIT, 6060 Hall in Tirol, Austria
Email: michael.netzer|michael.handler|christian.baumgartner@umit.at
†Zhongshan Hospital, Fudan University
200032 Xuhui District, Shanghai,China
fangxiaocong@gmail.com

***Abstract*—New biomarker candidates for breast cancer recurrence are urgently needed to provide patients an optimal treatment and avoid "overtreatment" where patients needlessly suffer from the toxic side effects of chemotherapy. In this work, we present a new network-based approach to identify biomarker candidates for breast cancer recurrence. Our coupled three-step strategy first selects relevant genes by using a filter approach. In the second step, we infer a new type of network where the number of edges of each vertex (i.e., degree) represents the predictive value of the underlying gene. In the third step, we conduct a database search for biomedical interpretation of our findings. Using a breast cancer microarray dataset we could verify top ranked genes associated with breast cancer and other pathophysiological processes.**

***Index Terms*—biomarker discovery; microarray; feature selection; networks; breast cancer**

I. Introduction

Recent biotechnological advances in the "omics" sciences such as microarrays have led to high amounts of data. In particular, this data is characterized by a high number of features (e.g., genes) and a small number of samples (large p, small n problem) [1]. To identify highly predictive biomarker candidates, sophisticated feature selection approaches, including filters and wrappers, are required [2]. Wrapper approaches [3] use learning algorithms (i.e., a classifier if the dependent variable is categorical) and a search strategy to identify sets of highly discriminating features. Filter approaches that calculate a measure for every feature representing its predictive ability are independent of a classifier and generally have less extensive computational costs [4].

Advances of computer systems have also led to extensive biological network analysis [5]–[7], increasing understanding of interactions between genes, proteins and metabolites.

In this work, we introduce a new three-step network-based approach to identify genes related to breast cancer recurrence. New biomarker candidates demonstrating recurrence are urgently needed for diagnosis and to provide the patients with an optimal treatment plan. According to Weigelt et al. [8], currently, more than 80% of female patients with breast cancer receive adjuvant chemotherapy, though only approximately 40% relapse and ultimately die of metastatic breast cancer. Consequently, many patients are over-treated needlessly, suffering from toxic side effects of chemotherapy [8].

Section "Materials and Methods" describes our new three-step network-based approach. In the "Results and Discussion" section, we show our results for an example dataset. Finally, we conclude and discuss our method and findings.

II. Materials and Methods

A. Data

We used 59 preprocessed microarray spectra (platform GPL1390) from the study of Hoadley et al. [9] available at the GEO database [10] ($n_s = 59$). The data includes 29 controls and 30 cases (breast cancer recurrence within 36 months). In our experiments we only use genes where the ensembl gene id is available (n_f = 11638).

B. Computational approach

Given is a set of tuples (dataset) $T = \{x_i, y_i\}_{i=1}^{n_s}$ with $x_i \in \mathbb{R}^{n_f}$ and $y_i \in \{case, control\}$, where n_f is the number of features, n_s is the number of samples and y is the set of classlabels. Recently, we developed a new supervised approach to infer networks based on the ratios between metabolites [11]. Therefore, we first calculated all ratios R between features F, where $r_{ij} = |\log_2\left(\frac{f_i}{f_j}\right)|$ with $i > j$, and $f \in F, r \in R$.

Note that the logarithm induces symmetry of the ratios and their reciprocals, respectively. We then created a graph G with:

$$G_{ij} = \begin{cases} 1 & \text{if } |s_{ij}| > \tau \\ 0 & \text{else,} \end{cases} \quad (1)$$

for $i, j \in 1, \ldots, n_f$, where τ is a defined threshold. The score s representing the discriminatory ability was calculated using the BI filtering method [12]. Briefly, BI combines a discriminance measure (DA^*), the coefficient of variation CV and Δ_{change} representing the strength of the change. For the unpaired BI, we define $\Delta = \frac{\overline{x}}{\overline{x}_{ref}}$, where $\overline{x}$ is the mean of the comparison group (e.g., case) and $\overline{x}_{ref}$ is the mean of the reference group (e.g., control).

In our recent work, we could show that this new type of network, based on the ratios, outperforms other networks as

 ISBN: 978-1-61208-190-8

correlation networks in terms of accuracy (see also Netzer et al., 2011). However, this approach has several limitations: i) Creating the network for n features results in $\frac{n \cdot (n-1)}{2}$ comparisons and BI calculations ($\mathcal{O}(n^2)$); ii) by definition the values for f must be positive to calculate the logarithms and $\overline{x}_{ref}$ must be $\neq 0$ to calculate the BI scores.

In particular, when dealing with standard normal distributed microarray data ($N(0,1)$) we have a high number of features ($n > 10,000$) including negative values for f and $\overline{x}_{ref}$ is close to zero.

Therefore, in this work, we propose a new generic three-step strategy to overcome the afore mentioned restrictions for standard normal distributed datasets with a high number of features (e.g., preprocessed microarray data):

Step 1: In order to reduce the number of features we first perform a feature selection. Therefore, we use a filter method and calculate the score s representing the discriminatory ability of each feature. We remove all features f with a score less than a defined threshold ($s_f < \tau$). Finally, we obtain our reduced dataset T_r.

Step 2: Given a standard normal distributed dataset we calculate all differences D between features F in T_r, where $d_{ij} = |f_i - f_j|$ with $i > j$, and $f \in F, d \in D$. Similar to equation 1 we finally construct the graph with

$$G_{ij} = \begin{cases} 1 & \text{if } s_{ij} > \tau \\ 0 & \text{else,} \end{cases} \tag{2}$$

In our experiments, we used the information gain [13] to calculate the score s on the distances d for step 1 and 2. The information gain IG of a feature f is given by [14]

$$IG = H(Y) - H(Y|X), \text{ where} \tag{3}$$

$$H(Y) = -\sum_{y \in Y} p(y) \log_2(p(y)), \text{ and} \tag{4}$$

$$H(Y|X) = -\sum_{x \in X} p(x) \sum_{y \in Y} p(y|x) \log_2(p(y|x)), \tag{5}$$

where $H(Y)$ denotes the entropy for Y (class variable) and $H(Y|X)$ is the entropy of Y after observing X.

The information gain easily allows to identify features with no or less discriminatory ability. Therefore, we set $\tau = 0$ (filtering threshold) to remove features with no information regarding the class attribute. In addition, we also used the information gain because it can also deal with values ≤ 0.

Step 3: After the network is inferred the genes are ranked according to the topological descriptor of the vertices. In this study we ranked the genes according to their degree (i.e., number of edges).

To verify and interpret our findings a database search to multiple repositories such as the Database for Annotation, Visualization and Integrated Discovery (DAVID) [15] and the Kyoto Encyclopedia of Genes and Genomes (KEGG) database was conducted [16].

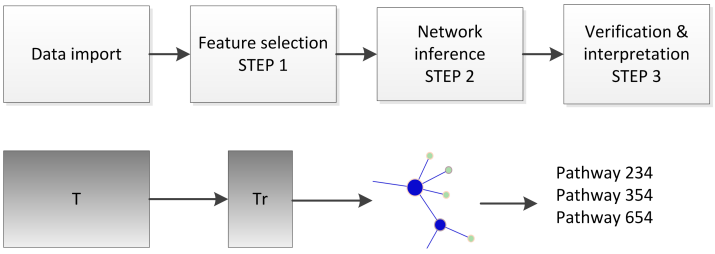


Fig. 1. The proposed workflow. The second row outlines the result of each process from row one. T denotes the entire dataset and T_r denotes the reduced dataset after feature selection.

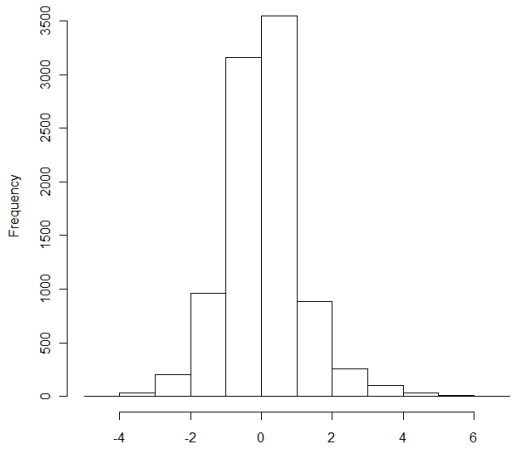


Fig. 2. Standard normal distribution of the dataset.

The feature selection was performed using Weka [17]. We used R [18] to implement the network approach methods. There exist several R packages for handling and analyzing graphs such as `graph` [19], `igraph` [20], `QuACN` [21] and `BioNet` [22]. The overall workflow including data import, feature selection, network inference and pathway analysis is depicted in Fig. 1.

The used hardware platform was an Intel Centrino 2x-1.83 GHz PC with 2048 MB RAM.

III. Results and Discussion

This works aims at identifying new genes related to breast cancer recurrence. The distribution of the used dataset is shown in Fig. 2. After feature selection of step 1 a total of 156 genes yielded an information gain > 0. The resulting network after applying step 2 to the reduced dataset is shown in Fig. 3. The ten top ranked genes using the degree (i.e., number of edges) are listed in Table I.

The related pathophysiological processes of the identified gene set are shown in Table II.

Kallilreins gene family are a group of serine proteases and known to be involved in several endocrine related malignancies

 ISBN: 978-1-61208-190-8

[23]. Kallikrein gene 7 (KLK7) was first reported in 1991 in the desquamation of stratum corneum [24] and later reported to be highly upregulated in ovarian carinomas [25]. Recent studies have reported that KLK7 can be up regulated primarily by estrogens and glucocorticoids [26], [27]. The expression of the KLK7 gene was supposed to be the most independent prognostic marker for the survival of patients with breast cancer [27].

Salivary amylase alpha 1 (AMY1) is involved in the starch and sucrose metabolism. It was reported to be highly expressed in individuals with high starch diets [28] and highly activated under psychosocial stress [29]. There is no existing evidence which supports the association between AMY1 and breast cancer, however, recent study reported AMY1 is an important modulator of cAMP-dependent protein kinase (PKA) which has versatile functions in cells [30].

The PH domain of PHLDA2 can compete with the PH domain of some other proteins, thereby interfering with their binding to phosphatidylinositol 4,5-bisphosphate (PIP2) and phosphatidylinositol 3,4,5-triphosphate (PIP3) in membrane lipids thus involved in various biological processes [15].

SnrpC belongs to the U1 small nuclear ribonucleoprotein C family and was reported to be involved in the splicing of mRNA [31]. The peptidylarginine deiminases 2 (PAD2) is a member of PADs which catalyze citrullination by converting arginine residues of proteins [32], [33]. It was also reported to play a role in inflammatory response, cell apoptosis [34], [35] as well as the gene regulation in the mammary glands [36].

Wee1 is another gene which was previously reported to be involved in tumourigenesis [37], [38] as well as the signaling regulation in breast cancer stem cells [39]. It was suggested to act as a tumor suppressor via regulating Cyclin and cyclin-dependent kinase complexes [37].

SOCS family proteins are part of a classical negative feedback system that regulates cytokine signal transduction of which SOCS2 appears to regulate the growth hormone/IGF1 signaling pathway [15], [40]. It is also suggested to play a role in the oncogenesis of ovary and mammary gland [41].

SOX5 is one of the high-mobility group (HMG) which has been recognized as a key player in the regulation of embryonic development [42] and in the determination of cell fate [43] reported to be involved in the progression of glioma and prostate cancer [44], [45].

KIAA0494 was supposed to play a role in the splicing of eukaryotic pre-mRNAs [46] but there are no more studies provided on its effect on tumorgenesis. CTP synthase (CTPS) plays a predominant role in CTP synthesis by converting UTP to CTP thus controling cell proliferation, differentiation and apoptosis [47]. Previous studies have demonstrated that CTPS depletion resulted in stabilization of wild-type p53 and showed antitumor effects in breast cancer cells [48].

Among the top 10 most related genes, we found that half of them had been previously reported to play roles in tumorogenesis, such as KLK7, Wee1, SOCS, SOX5 and CTPS. Furthermore, some had been studied in breast cancer (KL7, Wee1,SOCS, CTPS). However, the others were first suspected to be involved in the development or progression of breast cancer. Our results revealed that it is possible that these genes provide new targets for the control of breast cancer. However, further studies are warranted and essential to verify and validate these promising findings.

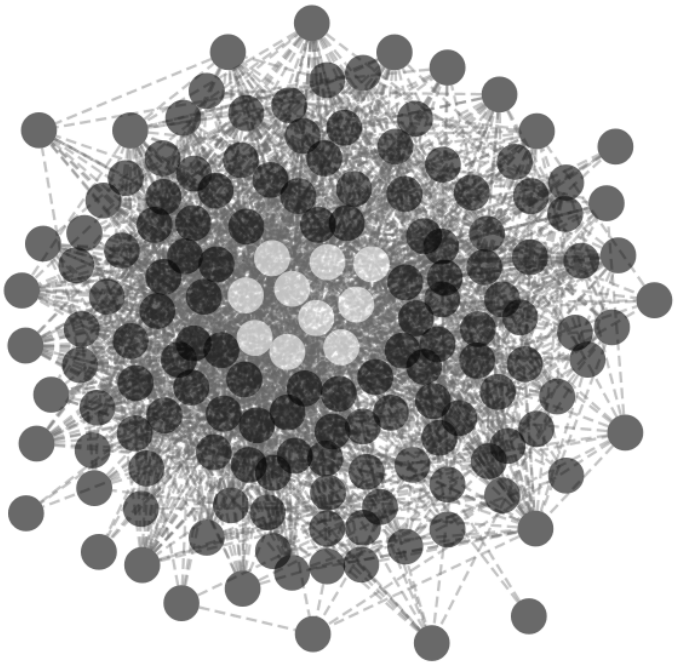


Fig. 3. The resulting network in which the ten white vertices represent the genes with highest degree. The network is plotted using Cytoscape [49].

TABLE I
THE GENE ID, DEGREE AND NAME OF THE TOP 10 RANKED FEATURES.

Rank	Gene ID	Degree	Gene name
1	ENSG00000169035	86	kallikrein-related peptidase 7 (KLK7)
2	ENSG00000051415	68	amylase, alpha 1A, 1B, 1C
3	ENSG00000181649	66	pleckstrin homology-like domain family A, member 2 (phlda2)
4	ENSG00000124562	54	small nuclear ribonucleoprotein polypeptide C
5	ENSG00000117115	52	peptidyl arginine deiminase, type II
6	ENSG00000166483	51	WEE1 homolog (S. pombe)
7	ENSG00000120833	49	suppressor of cytokine signaling 2
8	ENSG00000134532	47	SRY (sex determining region Y)-box 5
9	ENSG00000159658	45	KIAA0494
10	ENSG00000171793	43	CTP synthase

IV. CONCLUSION

In this work, we introduced a new network-based approach to identify biomarker candidates for breast cancer recurrence. Our main contribution is to propose a new workflow that

 ISBN: 978-1-61208-190-8

TABLE II
THE RELATED PATHOPHYSIOLOGICAL PROCESSES OF THE TOP 10 RANKED GENES. THE ASSOCIATED GENE NAMES CAN BE FOUND IN TABLE I.

Rank	Related pathophysiological processes
1	Play a role in the desquamation of the skin, the stratum corneum. Up-regulated in by estrogens and glucocorticoids.
2	involved in starch and sucrose metabolism, dalivary secretion, carbohydrate digestion and absorption.
3	compete to bind phosphoinositides in the membrane lipids with a broad specificity in various biological processes
4	associated with snRNP U1. and involved in the splicesome pathway
5	catalyzes the deimination of arginine residues of proteins
6	participate in the cell cycle with an increased synthesis during S and G2 phases
7	a negative regulator in the growth hormone/IGF1 signaling pathway
8	binds specifically to the DNA sequence 5'-AACAAT-3', overexpressed in glioma and prostate tumor
9	involved in the splicing of eukaryotic pre-mRNAs
10	catalyzes the ATP-dependent amination of UTP to CTP, play a role in pyrimidine metabolism

includes the filtering of relevant genes and inferring the network based on an information entropy measure. Our three-step strategy selected in the first step relevant features using the information gain. In the second step we inferred a new type of a network where the number of edges of each vertex (i.e., degree) represents the predictive ability of the underlying feature. Finally we used the DAVID and KEGG databases to verify and interpret top ranked genes.

Using our breast cancer microarray dataset from the GEO database we could identify a set of known and unexpected genes associated with breast cancer and other pathophysiological processes. The proposed generic method can also be applied to other biomedical questions (e.g., other diseases) or types of data such as metabolic datasets.

ACKNOWLEDGMENT

This work was supported by the Austrian Genome Research Program GEN-AU (Bioinformatics Integration Network, BIN III) and partly by the Tiroler Wissenschaftsfond.

REFERENCES

[1] A. A. Amini and M. J. Wainwright, "High-dimensional analysis of semidefinite relaxations for sparse," *The Annals of Statistics*, vol. 37, p. 28772921, 2009.

[2] M. Netzer, G. Millonig, M. Osl, B. Pfeifer, S. Praun, J. Villinger, W. Vogel, and C. Baumgartner, "A new ensemble-based algorithm for identifying breath gas marker candidates in liver disease using ion molecule reaction mass spectrometry." *Bioinformatics*, vol. 25, no. 7, pp. 941–947, Apr 2009.

[3] I. Inza, P. Larraaga, R. Blanco, and A. J. Cerrolaza, "Filter versus wrapper gene selection approaches in dna microarray domains." *Artif Intell Med*, vol. 31, pp. 91–103, Jun 2004.

[4] C. Baumgartner and A. Graber, *Successes and new directions in data mining*. Idea Group Inc., 2007, ch. 7, Data mining and knowledge discovery in metabolomics, pp. 141–166.

[5] F. Emmert-Streib and M. Dehmer, "Networks for systems biology: conceptual connection of data and function." *IET Syst Biol*, vol. 5, no. 3, pp. 185–207, May 2011.

[6] M. Dehmer, N. Barbarini, K. Varmuza, and A. Graber, "Novel topological descriptors for analyzing biological networks." *BMC Struct Biol*, vol. 10, p. 18, 2010.

[7] S. Bergmann, J. Ihmels, and N. Barkai, "Similarities and differences in genome-wide expression data of six organisms." *PLoS Biol*, vol. 2, no. 1, p. E9, Jan 2004.

[8] B. Weigelt, J. L. Peterse, and L. J. van 't Veer, "Breast cancer metastasis: markers and models." *Nat Rev Cancer*, vol. 5, no. 8, pp. 591–602, Aug 2005.

[9] K. A. Hoadley, V. J. Weigman, C. Fan, L. R. Sawyer, X. He, M. A. Troester, C. I. Sartor, T. Rieger-House, P. S. Bernard, L. A. Carey, and C. M. Perou, "Egfr associated expression profiles vary with breast tumor subtype." *BMC Genomics*, vol. 8, p. 258, 2007.

[10] T. Barrett, D. B. Troup, S. E. Wilhite, P. Ledoux, C. Evangelista, I. F. Kim, M. Tomashevsky, K. A. Marshall, K. H. Phillippy, P. M. Sherman, R. N. Muertter, M. Holko, O. Ayanbule, A. Yefanov, and A. Soboleva, "Ncbi geo: archive for functional genomics data sets–10 years on." *Nucleic Acids Res*, vol. 39, no. Database issue, pp. D1005–D1010, Jan 2011.

[11] M. Netzer, K. M. Weinberger, M. Handler, M. Seger, X. Fang, K. G. Kugler, A. Graber, and C. Baumgartner, "A computational strategy for the identification and kinetic analysis of metabolic biomarkers: A pilot study for profiling human response to physical exercise," *J Clin Bioinforma*, vol. 1, no. 34, 2011.

[12] C. Baumgartner, G. D. Lewis, M. Netzer, B. Pfeifer, and R. E. Gerszten, "A new data mining approach for profiling and categorizing kinetic patterns of metabolic biomarkers after myocardial injury." *Bioinformatics*, vol. 26, no. 14, pp. 1745–1751, Jul 2010.

[13] J. R. Quinlan, "Induction of decision trees," *Mach Learn*, vol. 1, no. 1, pp. 81–106, 1986.

[14] M. A. Hall, "Correlation-based feature selection for machine learning," Ph.D. dissertation, University of Waikato, 1999.

[15] G. Dennis, B. T. Sherman, D. A. Hosack, J. Yang, W. Gao, H. C. Lane, and R. A. Lempicki, "David: Database for annotation, visualization, and integrated discovery." *Genome Biol*, vol. 4, no. 5, p. P3, 2003.

[16] M. Kanehisa and S. Goto, "Kegg: kyoto encyclopedia of genes and genomes." *Nucleic Acids Res*, vol. 28, no. 1, pp. 27–30, Jan 2000.

[17] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," *SIGKDD Explorations*, vol. 11, 2009.

[18] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2011, ISBN 3-900051-07-0. [Online]. Available: http://www.R-project.org

[19] R. Gentleman, E. Whalen, W. Huber, and S. Falcon, *graph: graph: A package to handle graph data structures*, 2009, r package version 1.26.0.

[20] G. Csardi and T. Nepusz, "The igraph software package for complex network research," *InterJournal*, vol. Complex Systems, p. 1695, 2006. [Online]. Available: http://igraph.sf.net

[21] L. A. J. Mueller, K. G. Kugler, A. Dander, A. Graber, and M. Dehmer, "QuACN: an R package for analyzing complex biological networks quantitatively," *Bioinformatics*, vol. 27, no. 1, pp. 140–141, Jan 2011.

[22] D. Beisser, G. Klau, T. Dandekar, T. Mueller, and M. Dittrich, "Bionet an r-package for the functional analysis of biological networks," *Bioinformatics*, vol. 26, pp. 1129–1130, 2009.

[23] F. Fritzsche, T. Gansukh, C. A. Borgoo, M. Burkhardt, S. Pahl, E. Mayordomo, K.-J. Winzer, W. Weichert, C. Denkert, K. Jung, C. Stephan, M. Dietel, E. P. Diamandis, E. Dahl, and G. Kristiansen, "Expression of human kallikrein 14 (klk14) in breast cancer is associated with higher tumour grades and positive nodal status." *Br J Cancer*, vol. 94, no. 4, pp. 540–547, Feb 2006.

[24] A. Lundstrm and T. Egelrud, "Stratum corneum chymotryptic enzyme: a proteinase which may be generally present in the stratum corneum and with a possible involvement in desquamation." *Acta Derm Venereol*, vol. 71, no. 6, pp. 471–474, 1991.

[25] H. Tanimoto, L. J. Underwood, K. Shigemasa, M. S. Y. Yan, J. Clarke, T. H. Parmley, and T. J. O'Brien, "The stratum corneum chymotryptic enzyme that mediates shedding and desquamation of skin cells is highly overexpressed in ovarian tumor cells." *Cancer*, vol. 86, no. 10, pp. 2074–2082, Nov 1999.

 ISBN: 978-1-61208-190-8

[26] G. M. Yousef, A. Scorilas, A. Magklara, A. Soosaipillai, and E. P. Diamandis, "The klk7 (prss6) gene, encoding for the stratum corneum chymotryptic enzyme is a new member of the human kallikrein gene family - genomic characterization, mapping, tissue expression and hormonal regulation." *Gene*, vol. 254, no. 1-2, pp. 119–128, Aug 2000.

[27] M. Talieri, E. P. Diamandis, D. Gourgiotis, K. Mathioudaki, and A. Scorilas, "Expression analysis of the human kallikrein 7 (klk7) in breast tumors: a new potential biomarker for prognosis of breast carcinoma." *Thromb Haemost*, vol. 91, no. 1, pp. 180–186, Jan 2004.

[28] G. H. Perry, N. J. Dominy, K. G. Claw, A. S. Lee, H. Fiegler, R. Redon, J. Werner, F. A. Villanea, J. L. Mountain, R. Misra, N. P. Carter, C. Lee, and A. C. Stone, "Diet and the evolution of human amylase gene copy number variation." *Nat Genet*, vol. 39, no. 10, pp. 1256–1260, Oct 2007.

[29] U. M. Nater, N. Rohleder, J. Gaab, S. Berger, A. Jud, C. Kirschbaum, and U. Ehlert, "Human salivary alpha-amylase reactivity in a psychosocial stress paradigm." *Int J Psychophysiol*, vol. 55, no. 3, pp. 333–342, Mar 2005.

[30] M. Furusawa, T. Taira, S. M. M. Iguchi-Ariga, and H. Ariga, "Amy-1 interacts with s-akap84 and akap95 in the cytoplasm and the nucleus, respectively, and inhibits camp-dependent protein kinase activity by preventing binding of its catalytic subunit to a-kinase-anchoring protein (akap) complex." *J Biol Chem*, vol. 277, no. 52, pp. 50 885–50 892, Dec 2002.

[31] Y. Muto, D. P. Krummel, C. Oubridge, H. Hernandez, C. V. Robinson, D. Neuhaus, and K. Nagai, "The structure and biochemical properties of the human spliceosomal protein u1c." *J Mol Biol*, vol. 341, no. 1, pp. 185–198, Jul 2004.

[32] R. Raijmakers, A. J. W. Zendman, W. V. Egberts, E. R. Vossenaar, J. Raats, C. Soede-Huijbregts, F. P. J. T. Rutjes, P. A. van Veelen, J. W. Drijfhout, and G. J. M. Pruijn, "Methylation of arginine residues interferes with citrullination by peptidylarginine deiminases in vitro." *J Mol Biol*, vol. 367, no. 4, pp. 1118–1129, Apr 2007. [Online]. Available: http://dx.doi.org/10.1016/j.jmb.2007.01.054

[33] E. R. Vossenaar, A. J. W. Zendman, W. J. van Venrooij, and G. J. M. Pruijn, "Pad, a growing family of citrullinating enzymes: genes, features and involvement in disease." *Bioessays*, vol. 25, no. 11, pp. 1106–1118, Nov 2003.

[34] E. R. Vossenaar, T. R. D. Radstake, A. van der Heijden, M. A. M. van Mansum, C. Dieteren, D.-J. de Rooij, P. Barrera, A. J. W. Zendman, and W. J. van Venrooij, "Expression and activity of citrullinating peptidylarginine deiminase enzymes in monocytes and macrophages." *Ann Rheum Dis*, vol. 63, no. 4, pp. 373–381, Apr 2004.

[35] H. J. Lee, M. Joo, R. Abdolrasulnia, D. G. Young, I. Choi, L. B. Ware, T. S. Blackwell, and B. W. Christman, "Peptidylarginine deiminase 2 suppresses inhibitory kappab kinase activity in lipopolysaccharide-stimulated raw 264.7 macrophages." *J Biol Chem*, vol. 285, no. 51, pp. 39 655–39 662, Dec 2010.

[36] B. D. Cherrington, E. Morency, A. M. Struble, S. A. Coonrod, and J. J. Wakshlag, "Potential role for peptidylarginine deiminase 2 (pad2) in citrullination of canine mammary epithelial cell histones." *PLoS One*, vol. 5, no. 7, p. e11768, 2010.

[37] T. Yoshida, S. Tanaka, A. Mogi, Y. Shitara, and H. Kuwano, "The clinical significance of cyclin b1 and wee1 expression in non-small-cell lung cancer." *Ann Oncol*, vol. 15, no. 2, pp. 252–256, Feb 2004.

[38] S. Backert, M. Gelos, U. Kobalz, M. L. Hanski, C. Bhm, B. Mann, N. Lvin, A. Gratchev, U. Mansmann, M. P. Moyer, E. O. Riecken, and C. Hanski, "Differential gene expression in colon carcinoma cells and tissues detected with a cdna array." *Int J Cancer*, vol. 82, no. 6, pp. 868–874, Sep 1999.

[39] H. Wang, M. Huang, D. Y. Zhang, and F. Zhang, "Global profiling of signaling networks: study of breast cancer stem cells and potential regulation." *Oncologist*, vol. 16, no. 7, pp. 966–979, 2011.

[40] E. Rico-Bautista, A. Flores-Morales, and L. Fernndez-Prez, "Suppressor of cytokine signaling (socs) 2, a protein with multiple functions." *Cytokine Growth Factor Rev*, vol. 17, no. 6, pp. 431–439, Dec 2006.

[41] K. D. Sutherland, G. J. Lindeman, D. Y. H. Choong, S. Wittlin, L. Brentzell, W. Phillips, I. G. Campbell, and J. E. Visvader, "Differential hypermethylation of socs genes in ovarian and breast carcinomas." *Oncogene*, vol. 23, no. 46, pp. 7726–7733, Oct 2004.

[42] Y. Kamachi, M. Uchikawa, and H. Kondoh, "Pairing sox off: with partners in the regulation of embryonic development." *Trends Genet*, vol. 16, no. 4, pp. 182–187, Apr 2000.

[43] L. H. Pevny and R. Lovell-Badge, "Sox genes find their feet." *Curr Opin Genet Dev*, vol. 7, no. 3, pp. 338–344, Jun 1997.

[44] S. Ma, Y. P. Chan, B. Woolcock, L. Hu, K. Y. Wong, M. T. Ling, T. Bainbridge, D. Webber, T. H. M. Chan, X.-Y. Guan, W. Lam, J. Vielkind, and K. W. Chan, "Dna fingerprinting tags novel altered chromosomal regions and identifies the involvement of sox5 in the progression of prostate cancer." *Int J Cancer*, vol. 124, no. 10, pp. 2323–2332, May 2009.

[45] E. Tchougounova, Y. Jiang, D. Brster, N. Lindberg, M. Kastemar, A. Asplund, B. Westermark, and L. Uhrbom, "Sox5 can suppress platelet-derived growth factor b-induced glioma development in ink4a-deficient mice through induction of acute cellular senescence." *Oncogene*, vol. 28, no. 12, pp. 1537–1548, Mar 2009.

[46] K. Szafranski, S. Schindler, S. Taudien, M. Hiller, K. Huse, N. Jahn, S. Schreiber, R. Backofen, and M. Platzer, "Violating the splicing rules: Tg dinucleotides function as alternative 3' splice sites in u2-dependent introns." *Genome Biol*, vol. 8, no. 8, p. R154, 2007.

[47] K. J. M. Schimmel, H. Gelderblom, and H. J. Guchelaar, "Cyclopentenyl cytosine (cpec): an overview of its in vitro and in vivo activity." *Curr Cancer Drug Targets*, vol. 7, no. 5, pp. 504–509, Aug 2007.

[48] M. Huang, P. Whang, P. Lewicki, and B. S. Mitchell, "Cyclopentenyl cytosine induces senescence in breast cancer cells through the nucleolar stress response and activation of p53." *Mol Pharmacol*, vol. 80, no. 1, pp. 40–48, Jul 2011.

[49] M. S. Cline, M. Smoot, E. Cerami, A. Kuchinsky, N. Landys, C. Workman, R. Christmas, I. Avila-Campilo, M. Creech, B. Gross, K. Hanspers, R. Isserlin, R. Kelley, S. Killcoyne, S. Lotia, S. Maere, J. Morris, K. Ono, V. Pavlovic, A. R. Pico, A. Vailaya, P.-L. Wang, A. Adler, B. R. Conklin, L. Hood, M. Kuiper, C. Sander, I. Schmulevich, B. Schwikowski, G. J. Warner, T. Ideker, and G. D. Bader, "Integration of biological networks and gene expression data using cytoscape." *Nat Protoc*, vol. 2, no. 10, pp. 2366–2382, 2007.

 ISBN: 978-1-61208-190-8

Enhancing the Potency and Longevity of Highly Valuable Peptides Using Gene Fusion

Fuad Fares, Naiel Azzam, Rinat Bar-Shalom
Faculty of Natural Sciences
University of Haifa
Haifa, Israel,
ffares@sci.haifa.ac.il;
naiel_azzam@walla.com; rinalamir@gmail.com

Avri Havron Eyal Fima
PROLOR Biotech
Weizmann Science Park
Nes-Ziona, Israel
eyal@prolor-biotech.com;
avri@prolor-biotech.com

***Abstract* - One major issue regarding the clinical use of many peptides is their short half-life in the body, due to the rapid clearance from the circulation. To overcome this issue, the carboxyl-terminal peptide of human chorionic gonadotropin β subunit was ligated to the coding sequence of follitropin, thyrotropin, erythropoietin and growth hormone. This peptide contains 28 amino acids and four *O*-linked oligosaccharide recognition sites. It was postulated that the *O*-linked oligosaccharides add flexibility, hydrophilicity and stability to the protein. Ligation of this peptide to the coding sequence of these hormones has no effect on receptor binding and *in vitro* bioactivity. However it is dramatically increased half-life and bioactivity *in vivo*. Interestingly, the new analogs of follitropin and growth hormone were found not immunogenic in human. Follitropin already passed successfully clinical trials phase III and approved by The European Commission for human use. In addition, our results indicated that long acting growth hormone is not toxic in monkeys and it passed successfully clinical trials phase II in adults. Other strategy to stabilize the peptide hormones was to convert the heterodimeric structure to a single chain by fusing the subunits to a single gene with or without the carboxyl-terminal peptide of human chorionic gonadotropin β subunit. The single peptide chains of human thyroptropin were active and have longer half-life in the circulation.**

Keywords - recombinant proteins; long acting; follitropin; thyrotropin; human chorionic gonadotropin; growth hormone; erythropoietin.

I. INTRODUCTION

The family of glycoprotein hormones consists of Thyrotropin (TSH), Lutropin (LH), Follitropin (FSH) and Chorionic Gonadotropin (CG). These hormones are heterodimers, consisting of the noncovalent association of a common α subunit with a unique β subunit that confers biological specificity to the hormone [1] [2]. The individual subunits have no known biological activity. Thus, formation of heterodimer is essential for activity [1] [2] [3]. The α subunit within a species has an identical amino acid sequence in all four members of this hormone family. The β subunit shares considerable amino acid homology with one another, indicating that most likely they evolved from a common precursor [4] [5] [6]. The subunits achieve their tertiary structures by the formation of internal disulfide bonds; 5 in the α subunit and 2-6 in the β subunits. The location of cystein residues, which determines the 3-dimensional structure of the subunits by predicating their folding, in the β subunits is highly conserved among the various hormones as well as different species [4].

The glycoprotein hormones activate the target cells via adenylate cyclase-linked receptors through binding to the membrane receptors. FSH stimulates follicular development in the ovary and gametogenesis in the testes. LH, acts primarily in promoting luteinization of the ovary and in stimulating Leydig cell function of the testes. CG maintains the corpus luteum in the ovary during pregnancy. TSH is a major regulator of thyroid hormone synthesis and secretion from the thyroid gland. The thyroid hormones triiodothyronine (T_3) and thyroxine (T_4) regulate the synthesis and secretion of TSH from the pituitary [1] [3].

The subunits contain one (TSHβ and LHβ) or two (α, FSHβ and hCGβ) asparagine N-linked oligosaccharide chains [1] [2]. One unique structural difference among the subunits is the sequence of the carboxy terminus. TSH and LH subunits contain short hydrophobic stretches at their termini, deduced from gene sequencing [5]. However, the mature subunit does not contain the sequence due to a final deletion process. hCGβ subunit is distinguish from the other human β subunits in that it contains a unique 29 amino acid carboxyl-terminal peptide (CTP) bearing four *O*-linked oligosaccharide chains (Fig.1). It has been suggested that the *O*-linked oligosaccharide chains play an important role in the secretion of intact hCG from the cell, enhaced bioactivity and prolonged its circulating half-life *in vivo* [6]. Deletion of the *O*-linked oligosaccharide chains from hCG, didn't affect assemly of the subunits or secretion of the dimer from the cell and *in vitro* bioactivity. On the other hand, it was shown that truncated hCG without the CTP is 3 times less potent than intact hCG *in vivo* [7]. On the other hand, the *O*-linked oligosaccharide chains play a minor role in receptor binding and signal transduction. These findings indicate that the CTP of hCGβ and the associated *O*-linked oligosaccharides are not important

 ISBN: 978-1-61208-190-8

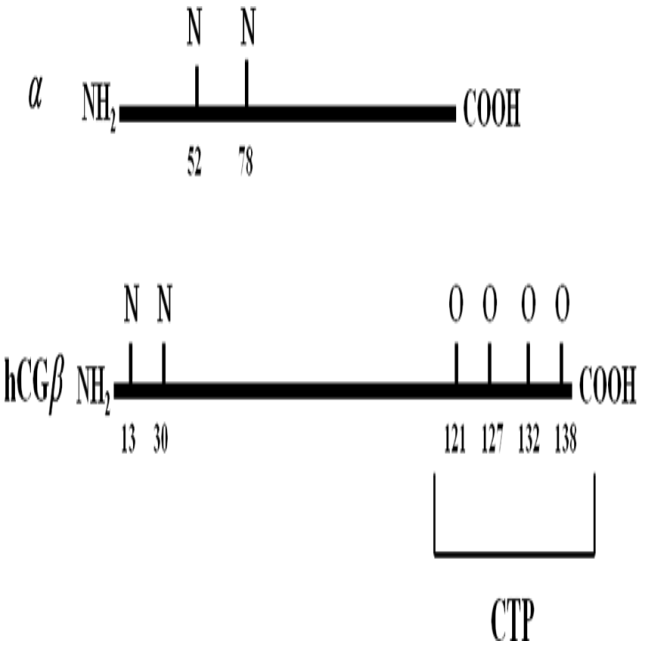


Figure 1. Human chorionic gonadotropin subunits. Localization of asparagine-linked carbohydrates (N) in both subunits are indicated. The CGβ subunit contain four sites of O-linked ologosaccharide chaines (O) in the CTP region.

for receptor binding or *in vitro* signal transduction, but are critical for *in vivo* bioactivity and half-life [6]

It was reported that the kidney is the main site of clearance for glycoprotein hormones [8]. On the other hand, much less hCG, which contains the CTP associatioted with the four *O*-linked oligosaccharide chains, is cleared by the kidney [9]. Other studies indicated that sialic acid plays an important role in the survival of glycoproteins in the circulation [1] [2]. It has been suggested that more negatively charged forms of glycoprotein hormones have longer half-lives, which may be related to a decrease glomerular filtration [10]. Thus, the presence of the CTP with its sialylated *O*-linked oligosaccharides may prolonged the circulating half-life of the hormone secondary to a decrease in renal clearance.

Erythropoietin (EPO) is a glycoprotein hormone produced primarily by cells of the peritubular capillary endothelium of the kidney [11] [13]. EPO is a member of an extensive cytokine family which also includes growth hormone, prolactin, interleukins 2 through 7, granulocyte colony-stimulating factor (G-CSF), granulocyte-macrophage colony-stimulating factor (GM-CSF), macrophage colony-stimulating factor (M-CSF), oncostatin-M, leukemia inhibitory factor and ciliaryneurotrophic factor. EPO production is stimulated by reduced oxygen content in the renal arterial circulation. Circulating EPO binds to EPO receptors on the surface of erythroid progenitors resulting in replication and maturation to functional erythrocytes by an incompletely understood mechanism.

The gene encoding human erythropoietin was cloned in 1985 leading to the production of recombinant human EPO (rhuEPO) [13]. The rhuEPO has been used successfully in a variety of clinical situations to increase production of red blood cells. Currently, this agent is licensed for use in the treatment of the anemia of renal failure, the anemia associated with HIV infection in zidovudine (AZT) treated patients, and anemia associated with cancer chemotherapy. Administration of rhuEPO has become routine in the treatment of anemia secondary to renal insufficiency where doses of 50-75 U/Kg given three times per week are used to gradually restore heamatocrit and eliminate transfusion dependency.

Human Growth Hormone (GH) is a member of family of closely related hormones that include prolactin and placental lactogen. GH regulates wide variety of physiological processes including growth and differentiation of muscle, bone and cartilage cells. GH is secreted by the somatotrophs of the anterior pituitary gland and acts on various tissues to promote growth and influence metabolism. GH antagonizes insulin effects by increasing hepatic gluconeogenesis and glycogenolysis by decreasing peripheral glucose consumption, effect that attributed to lipolytic effect. Disturbances of the carbohydrate metabolism can modulate the secretion of GH. The use of GH for the treatment of children with impaired linear growth has been accepted as an important therapeutic modality for many years. In addition, beneficial effects of GH replacement therapy in hypopituitary adults are well established. GH replacement therapy reduced body fat and increased body mass.

II. RESEARCH TOPICS

One major issue regarding the clinical use of glycoprotein hormones is their relatively short half-life *in vivo* due to their rapid clearance from the circulation when it's injected intravenously. Thus, the therapeutic protocol used in the treatment of glycoproteins, required frequent injections. The recommended therapy with rhuEPO is 2-3 times per week by subcutaneous or intravenous injections. Similarly, FSH or GH is injected daily. Therefore, we anticipated that ligation the CTP to the coding sequence of glycoprotein hormones will increase the half-life and bioactivity *in vivo*. This may reduce the number of injections per week. Therefore, it was hypothesized that FSH dimmer, EPO or GH containing the CTP would have a prolonged half-life and higher bioactivity *in vivo*. On the other hand, assembly of the hTSHβ and α-subunit is the rate limiting step in the production of functional heterodimer [3]. Thus, we hypothesized that converting hTSH to a single chain form could increase the half-life and expand the range of TSH structure – function studies. Here, we describe the construction of biologically active hTSH single chains with or without the CTP.

III. DESIGNINING GLYCOPROTEIN ANALOGS

To address the issue of the relatively short half-life of FSH, EPO and GH, the CTP of hCGβ was fused to the carboxyl-terminal of hFSHβ [14] [15], hEPO [16] and to the *N*- and *C*-terminal of human GH [17] coding sequences using overlapping PCR (Fig.2). Regarding the hTSH, the CTP was ligated to the C-terminal of β subunit. In addition the heterodimeric structure of hTSH was converted to a single chain by fusing the carboxyl end of hTSHβ subunit to the *N*-terminus of α-subunit in the absence or presence of the CTP sequence as a linker between the subunits (Fig. 3). It was noted that the *N*-and *C*- terminals of the hormones have no role in receptor binding or bioactivity. Therefore, it was hypothesized that ligation of these ends will not affect receptor binding or bioactivity. The designed chimeric genes were sequenced and ligateed into eukaryotic expression vectores. The plasmids were transfected into Chinese hamster ovary

 ISBN: 978-1-61208-190-8

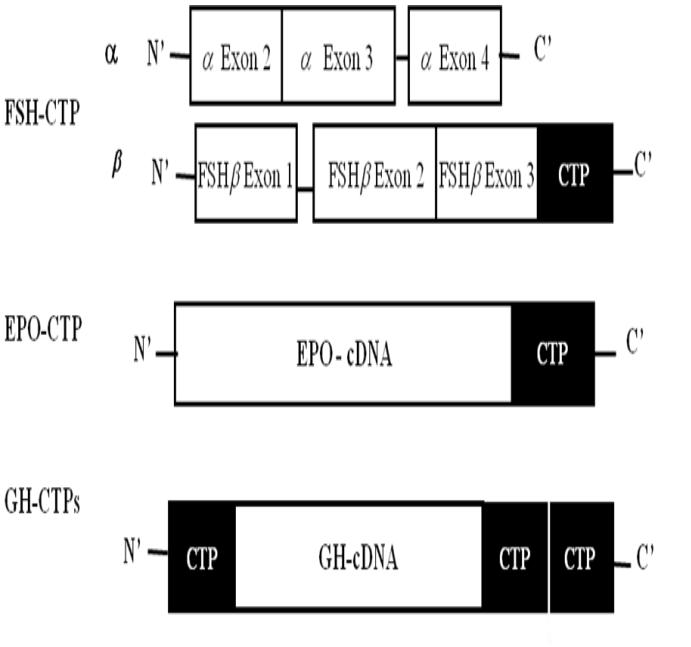


Figure 2. Construction of FSH-CTP, EPO-CTP and GH-CTP chimeric genes. The coding sequence of CTP was ligated to the 3rd end of hFSH*β gene or EPO* cDNA. One or two CTPs to be ligated to the N-terminal or to the C-terminal of GH cDNA, respectively.

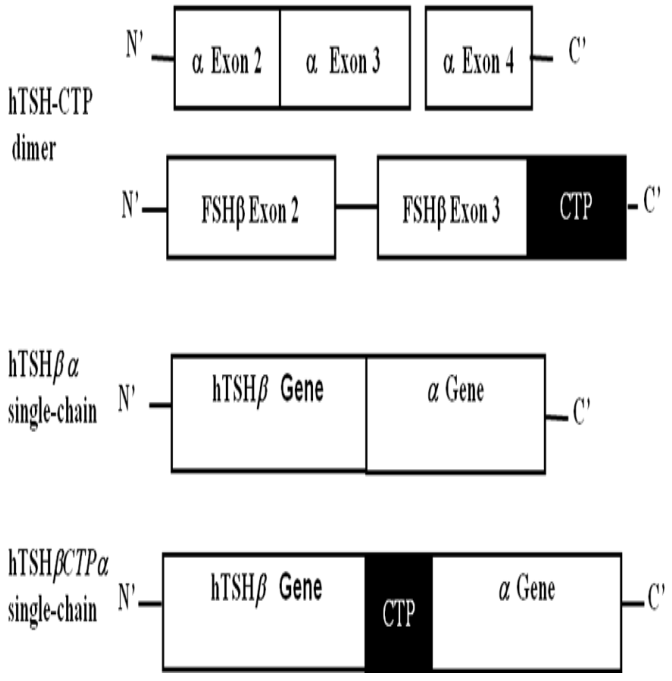


Figure 3. Construction of hTSH-CTP chimeric gene and single peptide chains with or without CTP as a linker between the subunits.

cells (CHO) and stable clones secreting the hormones were selected using selective antibiotics. Selected colonies resistant to antibiotics were harvested and screened for expression of the hormone. Collected media were concentrated and the hormones were detected immunoradiometric assay and a double antibody. Receptor binding was detected by radioreceptor assays. Bioactivity of the hormones was detected using *in vitro* and *in vivo* models.

IV. BIOACTIVITY OF THE DESIGNED ANALOGS

A. FSH

FSH induces aromatase enzyme in granulosa cells, therefore, signal transduction of the modified FSH dimer was assessed *in vitro* in the granulosa cell aromatase bioassay by measuring hormone-stimulated estrogen production. The detected steroidogenic activity of FSH-CTP chimeras was comparable to that of wild-type FSH. It was observed that ligation of CTP to the coding sequence of FSH*β* subunit, didn't affect receptor binding affinity or *in vitro* bioactivity. The *in vivo* bioactivity of wild-type FSH and the chimera was examined by determining ovarian weight augmentation and granulosa cell aromatase induction. It was found that ovarian weight increased significantly between animals treated with wild-type FSH and the FSH chimera. Treatment with FSH wild-type increased ovarian weight by 2 folds, where treatment with FSH-CTP increased ovarian weight by 3 folds. In addition, estrogen production by granulosa cells from chimera-treated rats increased 3- to 5-fold over that seen in rats treated with wild-type FSH. FSH-CTP was expressed in LDLD cells, a CHO cells mutated in the enzymes involved in addition of *O*-linked oligosaccharide chains. The results indicated that FSH-CTP lacking *O*-linked oligosaccharide have similar bioactivity of hFSH-WT (Fig.4).

Because the increased bioactivity of the chimeras may reflect a change in their *in vivo* longevity, the circulatory half-life of the hormones was detected. It seems that the clearance of the chimera is much slower than that of wild-type FSH; presumably. RIA determinations show that a high level of the chimera is still detectable in serum after 24 h and yet injected wild-type hFSH reaches basal level between 8 and 24 h.

The safety, pharmacokinetics and pharmacodynamics of FSH-CTP were studies in hypogonadotrophic hypogonadal male subjects as a phase I in multi-center study. The results indicated that FSH-CTP use is safe and does not lead to detectable formation of antibodies. Furthermore, pharmacokinetic and dynamic profile of FSH-CTP seemed to be promising. Compared with recombinant FSH – WT (Puregon), the half-life of FSH-CTP was increases 2-3 times [18] [19].

Further studies in *in vitro* fertilization (IVF) patients, indicated that a single dose of FSH-CTP is able to induce multifollicular growth comparing to daily injection of FSH-WT for 7 days [20]. According to the promising results described above in clinical trials, on January 28, 2010 the European Commission (EC) gave Merck & Co. marketing approval with unified labeling valid in all European Union Member States for FSH-CTP.

B. TSH

Because of TSH therapeutic potential, a longer acting analog of TSH was constructed by fusing the carboxyl-terminal extension peptide (CTP) of hCG*β* onto the coding sequence of TSH*β* subunit. When co-expressed either with alpha-subunit complementary DNA or alpha minigene in African green monkey (COS-7) or in human embryonic kidney (293) cells, the chimera was fully bioactive *in vitro* and

 ISBN: 978-1-61208-190-8

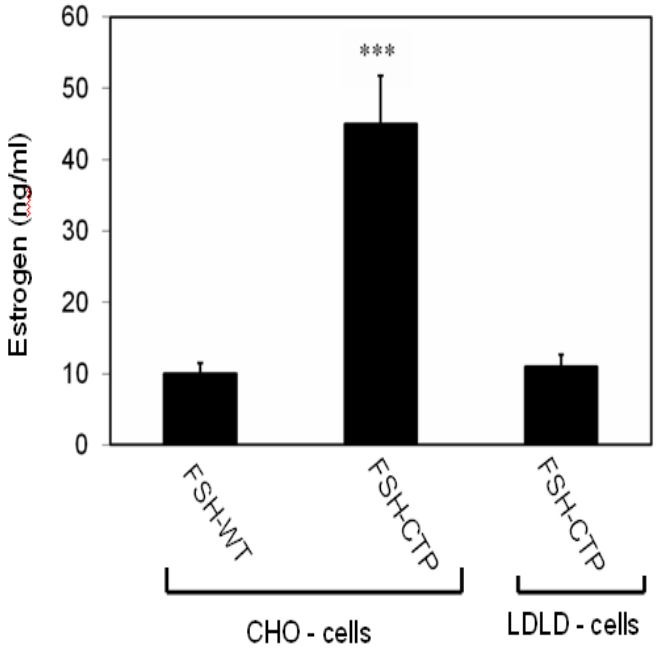


Figure 4. The effect of FSH on the level of estrogen in circulation of hypophysectomized rats. Recombinant FSH analogs were produced in CHO or LDLD cells'. ***P<0/001

exhibited enhanced *in vivo* potency associated with a prolonged plasma half-life. Ligation of the CTP to the C-terminal of TSHβ subunit did not affect the assembly and secretion of chimeric TSH [21]. Assembly of the glycoprotein hormone subunits is the rate-limiting step in the production of functional heterodimers [21]. To bypass this problem of dimerization of subunits, hTSHβ and α subunits were genetically fused in a single chain hormone with or without the CTP as a linker between the subunits. Single chains of hTSH retained a biologically active conformation similar to that of the wild-type heterodimer [22]. Moreover, it was found that the single peptide chains bind to the receptor with higher affinity and have higher bioactivity *in vivo* than the wild-type.

C. Erythropoietin

The CTP was ligated to the coding sequence at the C-terminal end of EPO. The *in vitro* biological activity of EPO-CTP was demonstrated by measuring their ability to stimulate proliferation of erythroid burst forming colonies (BFU-E) from human peripheral blood. BFU-E colonies were grown from blood of healthy donors using a microwell modification of the methylcellulose technique. The optimal formation of BFU-E colonies *in vitro* achieved by EPO-CTP was similar to that achieved by EPO-WT and rhuEPO using 1 U/ml of the protein. Receptor binding assay indicated that ligation of the CTP to the carboxyl-terminal of EPO has no significant effect on the affinity of the hormone to the receptor.

Further pharmacological evaluation of EPO-CTP, comparative pharmacodynamic studies of EPO-WT and commercial rhuEPO were performed in male C57BL mice (n=7/group) using different frequencies and wide dose range. The *in vivo* efficacy was obtained by measuring the mean values of haematocrit percentage in the blood. The results indicated that EPO-CTP is significantly (P<0.05) more efficient than EPO-WT when administered *IV* once a week with a dose of 200 mg/kg. EPO-CTP can successfully increase the haematocrit when administered once a week with a dose of 660 IU/kg. Once weekly dosing with the same concentration of commercial rhuEPO or EPO-WT was significantly (P<0.001) less efficient than once weekly dosing of EPO-CTP. An interesting observation is the ability of a single injection once a week of EPO-CTP (660 IU/kg) to increase the levels of haematocrit, whereas the same effect was achieved by administration of the same total dose of rhuEPO administered three times a week as 220 IU/kg per injection (Fig. 5).

These results indicated the importance of sustained blood levels, rather than total dose of EPO. These findings are consistent with the hypothesis that the ability of a single injection of EPO-CTP to increase haematocrit, results from its increased stability in the circulation. The increased biopotency of the chimera may reflect a change in their metabolic clearance *in vivo*. Detecting the half-lives of EPO analogs in mice indicated that a higher level of the chimera is still detectable in serum after 24 h. The half-life of EPO CTP is increased 2-3 folds comparing to EPO-WT. These data suggest that the mechanism of EPO metabolic clearance is affected by the presence of CTP [16].

D. Growth Hormone

Crystallographic studies indicated that *N*-terminal and *C*-terminal of GH are not important for binding of the hormone to its receptor. Therefore, CTP was fused to the *N*-terminal and *C*-terminal of hGH. The results indicate that ligation of CTP to the coding sequence of GH did not affect secretion of the chimeric protein into the medium. *In vivo* studies in hypophysectomized rats indicated that, bioactivity and pharmacokinetic parameters, MRT, AUC, Tmax, Cmax and half-life, of GH bearing the CTPs were dramatically enhanced. The estimated half-life of CTP-GH-CTP- CTP is increased by 4-5 folds comparing to Biotropin. These data suggest that the mechanism of GH metabolic clearance is affected by the presence of CTP [17]. Clinical trials phase I and phase II of GH-CTPs indicated that this peptide is safe for use and clinical trials of phase III is in the way.

V. CONCLUSIONS

Ligation of the CTP bearing four sites for *O*-linked oligosaccharide chains to different proteins indicated that the *O*-linked glycosylation recognition sites of the CTP are preserved. Moreover, this ligation is not involved in secretion, receptor binding and *in vitro* bioactivity. However, both the *in vivo* bioactivity and half-life in circulation of proteins bearing the CTP were significantly enhanced. Proteins containing the CTP could serve as long acting agonists for clinical use. This strategy may have wide applications for enhancing the *in vivo* bioactivity and half-life of diverse proteins. On the Other hand, converting the hetrerodimeric structure of the hormone to a single-chain polypeptide chain did not affect receptor binding affinity and bioactivity. However, this strategy is important for designing of new analogs of heterodimeric proteins.

 ISBN: 978-1-61208-190-8

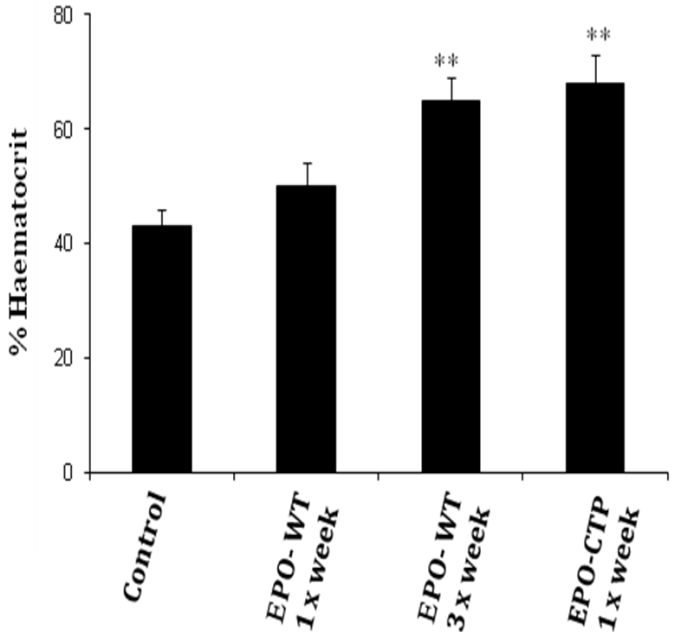


Figure 5. Ability of a single IV injection of EPO-CTP vs 3 IV injections of rHuEPO to increase haematocrit levels in mice. ICR mice (n=7/group) received a single IV injection/week for three weeks of EPO-CTP (660 IU/ kg) or wild-type rHuEPO (220 IU/kg) that were injected IV three times a week for three weeks. Control animals were injected IV with F-12 medium free of serum. Blood samples were collected three times a week and haematocrit levels were detected. Each point represents the group average of haematocrit (%) ± SE. **P<0/01.

VI. ACKNOWLEDGMENTS

This work was supported by grant from the Israel Science Foundation. Grant No. 617/01. We would like also to thank the Israel Ministry of Industry and Trade for supporting this research.

VII. REFERENCES

[1] I. Boime, and D. Ben-Menahem J. Glycoprotein hormone structure-function and analog design. Recent Prog Horm Res. vol. 54, pp. 271-288, 1999.

[2] M. Wallis, S.L. Howell, and K.W. Taylor. The biochemistry of the polypeptide hormones, New York: John Wiley and sons. pp. 147-183, 1985.

[3] M.W. Szkudlinski, V. Fremont, C. Ronin, and B.D. Weintraub. Thyroid-stimulating hormone and thyroid-stimulating hormone receptor structure-function relationship. Physiol. Rev. vol. 82, pp. 473-502, 2002.

[4] S.D. Gharib, M.E. Wierman, M.A. Shupnik, and W.W. Molecular Biology of the pituitary gonadotropins, Endocrine Rev. vol 11, pp. 177-199, 1990.

[5] Y. Hayashizaki, K. Miyai, K. Kato, and K. Matsubara. Molecular cloning of the human thyrotropin-β bubunit gene, FEBS Lett. vol. 188, pp. 394-400, 1985.

[6] M.M. Matzuk, A.J.W. Hsueh, P. Lapolt, A. Tsafriri, J.L. Keene, and I. Boime. The biological role of the carboxyl-terminal extension of human chorionic gonadotropin beta-subunit, Endocrinology. vol.126, pp. 376-383, 1990.

[7] M.M. Matzuk, M. Krieger, C.L. Corless, and I. Boime. Effect of preventing *O*-glycosylation on the secretion of human chorionic gonadotropin in Chinese hamster ovary cells. Proc. Natl. Acad. Sci. USA. vol. 84, pp. 6354-6358, 1987.

[8] H.K. Amin, and W.M. Hunter. Human pituitary follicle-stimulating hormone: distribution, plasma clearance, and urinary excretion as determined by radioimmunoassay, J. Endocrinol. vol. 48, pp. 307-317, 1977.

[9] J.R. Sowers, A.E. Pekary, J.M. Hershman, M. Kanter, and JJ. DiStefano. Metabolism of exogenous human chorionic gonadotropin in men, J. Endocrinol. vol. 80, pp. 83-89, 1979.

[10] L. Wide. The regulation of metabolic clearance rate of human FSH in mice by Variation of the molecular structure of the hormone, Acta Endocrinol. (Copenh) vol. 112, pp. 336-44, 1986.

[11] W. Jelkmann. Erythropoietin: Structure, control of production, and function. Physiol Rev. vol. 72, pp.449-487, 1992.

[12] S. Schuster, J.H. Wilson, A.J. Erslev, and J. Caro. Physiologic regulation and tissue localization of renal erythropoietin messenger RNA. Blood vol. 70, pp. 316-318, 1987.

[13] F.K. Lin, F.K.S. Suggs, C.H. Lin, J.K. Browne, R. Smalling, J.C. Egric, K.K. Chen, G.M. Fox, F. Martin, Z. Stabinsky, S.M. Badrawi, P.H. Lai, and E. Goldwasser. Cloning and expression of the human erythropoietin gene. Proc Natl Acad Sci USA vol. 82, pp. 7580-7584, 1985.

[14] F.A. Fares, N. Suganuma, K. Nishimori, P.S. LaPolt, A.J.W. Hsueh, and I. Boime. Design of a long-acting follitropin agonist by fusing the C-terminal sequence of the β chorionic gonadotropin subunit. Proc. Natl. Acad. Sci. USA. vol. 89, pp. 4304-4308, 1992.

[15] P.S. Lapolt, K. Nishimori, F.A. Fares, E. Perlas, I. Boime, and AJW. Hsueh. Enhanced stimulation of follicle maturation and ovulatory potential by long acting follicle-stimulating hormone agonist with extended carboxyl-terminal peptide, Endocrinology vol. 131, pp. 2514-2520, 1992.

[16] F. Fares, S. Ganem, T. Hajoj, and E. Agai. Development of a Long Acting Erythropoietin by Fusing the Carboxyl-Terminal Peptide of Human Chorionic Gonadotropin to the Coding Sequence of Human Erythropoietin. Endocrinology. vol. 148, pp. 5081- 5087, 2007.

[17] F. Fares, R. Guy, A. Bar-Ilan, Y. Felikman, and E. Fima. Designing a Long Acting Human Growth hormone by Fusing the Carboxyl-Terminal Peptide of Human Chorionic Gonadotropin β Subunit to the Coding Sequence of Human Growth hormone. Endocrinology. vol. 151, pp. 4410-4417, 2010.

[18] P.M.G. Bouloux, D.J. Handelsman, F. Jockenhovel, E. Nieschlag, J. Rabinovici, W.L. Frasa, J.J. deBie, G. Voortman, and J Itskovitz-Eldor, J. First human exposure to FSH-CTP in hypogonadotrophic hypogonadal males. Hum. Reprod. vol. 16, pp. 1592-1597, 2001.

[19] P. Devroey, B.C. Fauser, P. Platteau, N.G. Beckers, M. Dhont, and M. Mannaerts. Induction of multiple follicular development by single dose of long-acting recombinant follicle-stimulating hormone (FSH-CTP, Corifollitropin Alfa) for controlled ovarian stimulation before in vitro fertilization, J. Clin. Endocrinol. Metabol. vol. 89, pp. 2062-2070, 2004.

[20] N.G.M. Beckers, N.S. Mackolas, P. Devroey, P. Platteau, P.J. Boerrigter, and B.C.J.M. Fauser. First live birth after ovarian stimulation using a chimeric long-acting human recombinant follicle-stimulating hormone (FSH) agonist (recFSH-CTP) for *in vitro* fertilization, Fertil. Steril. vol. 79, pp. 621-623, 2003.

[21] L. Joshi, Y. Murata, F.E. Wondisford, M.W. Szkudlinski, R. Desai, and B.D Weintraub. Recombinant thyrotropin containing a β-subunit chimera with the human chorionic gonadotropin-β carboxy terminal is biologically active with a prolonged plasma half-life: role of carbohydrate in bioactivity and metabolic clearance. Endocrinolgy. vol. 136, pp. 3839-3848, 1995.

[22] F.A. Fares, S. Yamabe, D. Ben-Menahem, M. Pixley, A.J.W. Hsueh, and I. Boime. Conversion of thyrotropin heterodimer to a biologically active single-chain, Endocrinology vol. 139, pp. 2459-2464, 1998.

 ISBN: 978-1-61208-190-8

Predicting Gene Knockout Effects by Minimal Pathway Enumeration

Takehide Soh
Transdisciplinary Research Integration Center
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, Japan
soh@nii.ac.jp

Katsumi Inoue
National Institute of Informatics
2-1-2, Hitotsubashi, Chiyoda-ku, Tokyo, Japan
ki@nii.ac.jp

Tomoya Baba
Transdisciplinary Research Integration Center
1111 Yata, Mishima, 411-8540, Japan
tobaba@lab.nig.ac.jp

Toyoyuki Takada, Toshihiko Shiroishi
National Institute of Genetics
1111 Yata, Mishima, 411-8540, Japan
{ttakada,tshirois}@lab.nig.ac.jp

***Abstract*—In this paper, we propose a method to predict gene knockout effects for the cell growth by utilizing biological databases such as KEGG and EcoCyc, in which biological knowledge and experimental results have been collected. We construct biological networks from such databases and configure experimental conditions by giving source metabolites, target metabolites, and knockout genes. We then enumerate all minimal active pathways, which are minimal subsets of a given network using source metabolites to produce target metabolites. We simulate the effects of gene knockouts by measuring the difference of minimal active pathways between original networks and knockout ones. In the experiments, we applied it to predict the gene knockout effects on the glycolysis pathway of *Escherichia coli*. In the results, our method predicted three out of four essential genes, which are confirmed by the Keio collection containing comprehensive cell growth data obtained from biological experiments.**

Keywords-metabolic pathways; gene knockout; prediction method; minimal pathway; Keio collection.

I. INTRODUCTION

Living organisms, such as bacteria, fishes, animals, and humans, are kept alive by a huge number of intracellular chemical reactions. In *systems biology*, interactions of such chemical reactions are represented in a network called a *pathway*. Pathways have been actively researched in the last decade [1]–[3]. In addition, it is a biologically important subject to reveal the function of genes, which affect the phenotype of organisms. For model organisms such as *Escherichia coli* (*E. coli*), it has been approached by various methods. Constructing gene knockout organisms is an example of such methods [4]–[6]. However, it generally involves high costs and is limited by target genes and organisms.

In this paper, we propose a computation method to predict gene knockout effects by identifying *active pathways*, which are sub-pathways that produce target metabolites from source metabolites. We particularly focus on *minimal active pathways*, which are proposed by Soh and Inoue [7] and do not contain any other active pathways. In other words, all elements of each minimal active pathway are qualitatively essential to produce target metabolites. To predict gene knockout effects by the enumeration of minimal active pathways, we first introduce *extended pathways* that include relations between enzymatic reactions and genes. Then, we formalize the problem of finding minimal active pathways on the extended pathway with gene knockouts. After computing the solution of the problem, our method predicts gene knockout effects by collecting minimal active pathways that are still active under given gene knockouts.

To evaluate our method, we choose *E. coli* as our target organism, since it has been studied and much information about it is available on public resources. We apply our method to predict gene knockout effects on *E. coli* utilizing biological databases KEGG and EcoCyc, in which biological knowledge and experimental results have been collected. In the experiments, we compared our prediction and the cell growth of every single gene knockout *E. coli* strain, which was obtained from the Keio collection [4].

This paper is organized as follows. At first, we explain databases used in this paper and our research framework in Section II. We define the extended pathway in Section III. We formalize the problem of finding minimal active pathways on the extended pathway in Section IV and the effect of gene knockouts in Section V. Following that, we show our computational method in Section VI. In Section VII, we compare computational prediction and results of biological experiments. Following discussions in Section VIII, we conclude this paper in Section IX.

II. USED DATABASES AND RESEARCH FRAMEWORK

This section explains used databases and our research framework shown in Figure 1. In this paper, we particularly focus on *E. coli*. The metabolic pathway has been revealed by biochemical, molecular, and genetic studies, and *E. coli* is the organism in most detail. A large number of *E. coli* studies has contributed to several kinds of biological databases. In particular, we use the following two databases to construct our input network, called an extended pathway. One is *EcoCyc* [8]. It is a bioinformatics database that describes the genome and the biochemical machinery of *E. coli* K-12 MG1655. The EcoCyc project performs literature-based curation of the entire genome, metabolic pathways, etc. Specifically, it has been doing a literature-based curation

 ISBN: 978-1-61208-190-8

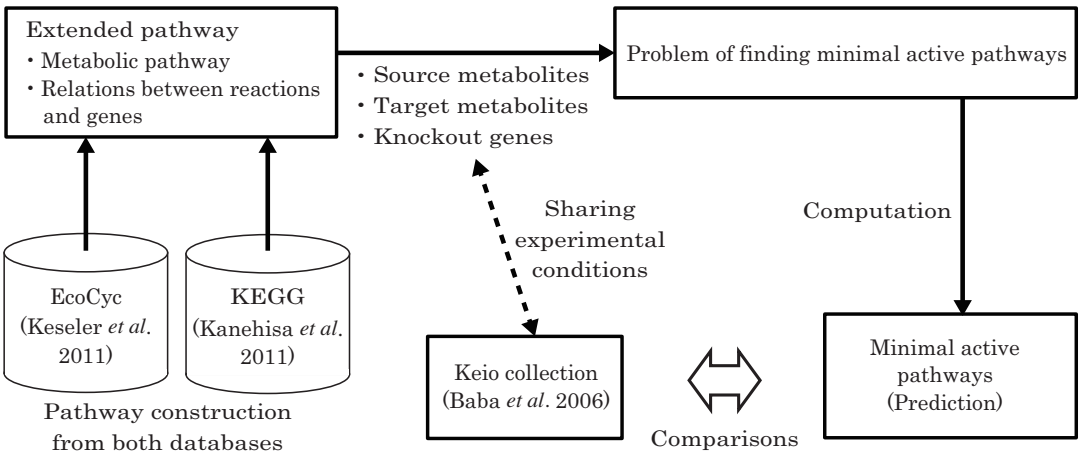


Figure 1. Used Databases and our Research Framework

from more than 19,000 publications. We construct metabolic pathways with EcoCyc. The other one is *Kyoto encyclopedia of genes and genomes* (KEGG) [9], which is a database resource that integrates genomic, chemical, and systemic functional information. In particular, gene catalogs in the completely sequenced genomes, from bacteria to humans, are linked to higher-level systemic functions of the cell, the organism, and the ecosystem. A distinguished feature of KEGG is that it provides useful application program interfaces (API). We connect enzymatic reactions of metabolic pathways to genes with this API.

Figure 1 shows our research framework using the two databases. At first, we construct our input network called *extended pathway* from them. We then construct the problem of finding minimal active pathways by giving source and target metabolites to the extended pathway. In addition, the condition of knockout genes is also added to the problem. Then, we compute minimal active pathways using source metabolites to produce target metabolites. In the case of wild cells, we usually obtain multiple minimal active pathways including bypass pathways. However, in the case of knockout cells, we lose some (or all) of them. In brief, we predict the effects of gene knockouts from how many pathways are lost from the case of wild cells.

To evaluate our prediction method, we usually need additional biological experiments. However, Baba *et al.* comprehensively experimented on the cell growth of every single gene knockout strain [4]. Thanks to this research, we can evaluate our method with comparative ease. We briefly explain this research as follows. The *E. coli* K-12 single gene knockout mutant set, named *Keio collection*, is constructed as a resource for systems biological analyses. Excluding repetitive genes, e.g., insertion sequences related genes, 4288 protein coding genes are targeted for the systematic single gene knockout experiments. Of those, 3985 genes are successfully disrupted, and those of single-gene knockout mutants are constructed as the Keio collection. On the other hand, 303 genes are not disrupted and they are thought to be essential gene candidates. Those single gene knockout mutants have the same genome background, which results in an advantage for distinct functional analysis of the targeted gene. The genome-wide relationship between the genome structure, i.e., genotype, and the phenomena, i.e., phenotype, which are analyzed by using the Keio collection has become available.

Although Figure 1 shows specific databases for *E. coli*, the research framework itself can be applied for other organisms whose pathway information is available, e.g., mice.

III. Extended Pathways

In this section, we explain how to represent metabolic pathways and their relations to genes. We then define the extended pathway.

To represent metabolic pathways, we commonly use bipartite directed graph representation as follows. Let M be a set of metabolites and R be a set of reactions. For M and R, $M \cap R = \emptyset$ holds. Let $A_M \subseteq (R \times M) \cup (M \times R)$ be a set of arcs. A *metabolic pathway* is represented in a directed bipartite graph $\mathcal{G}_{\mathcal{M}} = (M \cup R, A_M)$, where M and R are two sets of nodes, and A_M is a set of arcs. In addition to the metabolic pathway, we consider relations between enzymatic reactions and genes. Let G be a set of genes and A_G be a set of arcs such that $A_G \subseteq (G \times R)$. That is, A_G represents relations between enzymatic reactions and genes. Let N be a set of nodes such that $N = M \cup R \cup G$ and A be a set of arcs such that $A = A_M \cup A_G$. Then, the *extended pathway* is represented in a directed graph $\mathcal{G} = (N, A)$.

Figure 2 shows an example of the extended pathway. As the figure shows, it consists of two layers: the metabolic layer and the genetic layer. The genetic layer is the difference between the metabolic pathway and the extended pathway. In this example, the pathway consists of nodes of $M = \{m_1, m_2, \ldots, m_6\}$, $R = \{r_1, r_2, \ldots, r_7\}$, and $G = \{g_1, g_2, \ldots, g_8\}$. Each arc represents relations between

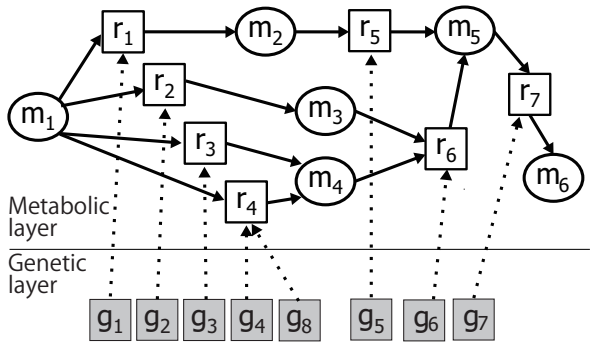


Figure 2. Example of an Extended Pathway

elements. For instance, the activation of the reaction r_6 needs the production of metabolites m_3 and m_4 and the expression of either g_4 or g_8. We will explain the interpretation of the extended pathway in detail in the next section.

IV. Minimal Active Pathways with Gene Knockouts

In the literature [7], the minimal active pathway is defined only on the metabolic pathway. On the other hand, in this section, we define the *minimal active pathway* on the extended pathway.

We here define $M_S \subset M$ as a set of source metabolites and $M_T \subset M$ as a set of target metabolites such that $M_S \cap M_T = \emptyset$. An *extended pathway instance* is represented in a four tuple $\pi = (N, A, M_S, M_T)$, where $N = M \cup R \cup G$, $A = A_M \cup A_G$. Let K be a set of genes such that $K \subseteq G$. We use K as a set of knockout genes in a given pathway. A *knockout instance* is represented in a five tuple $\pi_K = (N, A, M_S, M_T, K)$. If $K = \emptyset$ then π_K corresponds to π.

Let m, r be a metabolite and a reaction such that $m \in M$ and $r \in R$, respectively. A metabolite $m \in M$ is called a *reactant* of a reaction $r \in R$ when there is an arc $(m, r) \in A$. On the other hand, a metabolite $m \in M$ is called a *product* of a reaction $r \in R$ when there is an arc $(r, m) \in A$. Furthermore, a gene $g \in G$ is called a *corresponding gene* of a reaction $r \in R$ when there is an arc $(g, r) \in A$.

A reaction is called a *reversible reaction* if it can occur in both directions between reactants and products. In this paper, we distinguish a reversible reaction as two reactions. Suppose that there is a reversible reaction r_1 that has m_1 and m_2 as reactants and m_3 and m_4 as products. In this case, we split the reaction r_1 into two reactions r_{1a} and r_{1b} such that one of them has m_1 and m_2 as products and m_3 and m_4 as reactants.

Let $s : R \to 2^M$ be a mapping from a set of reactions to a power set of metabolites such that $s(r) = \{m \in M \mid (m, r) \in A\}$ represents the set of metabolites that are needed to turn the reaction r activatable. Let $p : R \to 2^M$ be a mapping from a set of reactions to a power set of metabolites such that $p(r) = \{m \in M \mid (r, m) \in A\}$ represents the set of metabolites that are produced by the reaction r. Let $c : R \to 2^G$ be a mapping from a set of reactions to a power set of genes such that $c(r) = \{g \in G \mid (g, r) \in A\}$ represents the set of genes that are corresponding genes of the reaction r. Let $p' : M \to 2^R$ be a mapping from a set of metabolites to a power set of reactions such that $p'(m) = \{r \in R \mid (r, m) \in A\}$. Let $c' : G \to 2^R$ be a mapping from a set of genes to a power set of reactions such that $c'(g) = \{r \in R \mid (g, r) \in A\}$.

Let t be an integer variable representing time. In this paper, the time is used to represent order relation between reactions to produce target metabolites from source metabolites. In the following, we explain important notions related to production of metabolites, activation of reactions, and expression of genes. Since we focus on gene knockouts, we suppose that almost all genes exist in the cell of a given organism. We also suppose that if genes exist, then they are expressed and available to construct enzymes needed for enzymatic reactions. The reason for this condition is that we want to simulate how the lack of corresponding genes affects metabolic pathway rather than how the existence of genes affects other elements. Although our pathway modeling is simple, it allows us to analyze a whole cell scale pathway. Let $\pi_K = (N, A, M_S, M_T, K)$ be a knockout instance, where $N = M \cup R \cup G$, $A = A_M \cup A_G$. Let $\mathcal{G} = (N, A)$ be an extended pathway. Let $M' \subset M$ be a subset of metabolites. A metabolite $m \in M$ is obviously *producible* at time $t = 0$ from M' on $\mathcal{G}$ if $m \in M'$ holds. A reaction $r \in R$ is *activatable* at time $t > 0$ from M' on $\mathcal{G}$ if the following two conditions are satisfied: (i) for every $m \in s(r)$, m is producible at time $t - 1$ from M', (ii) at least one corresponding gene $g \in c(r)$ is not included in K. A metabolite $m \in M$ is *producible* at time $t > 0$ from M' on $\mathcal{G}$ if there is at least one activatable reaction r at time t such that $m \in p(r)$. If r is activatable at time t, then r is activatable at time $t + 1$. If m is producible at time t, then m is producible at time $t + 1$.

Let $\mathcal{G}' = (N', A')$ be a sub-graph of $\mathcal{G}$, where $N' = M' \cup R' \cup G'$ and $A' = A'_M \cup A'_G$. Then, an active pathway of $\pi_K = (N, A, M_S, M_T, K)$ is defined as follows.

Definition 1: Active Pathway of Knockout Instance
A bipartite directed graph $\mathcal{G}'$ is an *active pathway* of π_K if it satisfies the following conditions:

- $M_T \subset M'$
- $M' = M_S \cup \{m \in M \mid (m, r) \subseteq A, r \in R'\} \cup \{m \in M \mid (r, m) \subseteq A, r \in R'\}$
- $A' = \{(m, r) \in A \mid r \in R'\} \cup \{(r, m) \in A \mid r \in R'\} \cup \{(g, r) \in A \mid g \notin K, r \in R'\}$
- $G' = \{g \in G \mid (g, r) \in A', r \in R'\}$
- For every $m \in M'$, m is producible from M_S on $\mathcal{G}'$

From Definition 1, active pathways include a set of metabolites, reactions, and genes, which are producible and activatable from M_S on $\mathcal{G}'$ such that all target metabolites M_T become producible. The number of active pathways

 ISBN: 978-1-61208-190-8

depends on the combination of M_S and M_T but an extended pathway generally has a large number of active pathways. We thus particularly focus on minimal ones rather than active pathways. We give the definition of minimal active pathways of π_K as follows. Let $\mathcal{G}$ and $\mathcal{G}'$ be extended pathways. We say that $\mathcal{G}$ is *smaller* than $\mathcal{G}'$ and represented in $\mathcal{G} \subset \mathcal{G}'$ if $R \subset R'$. An active pathway $\mathcal{G}$ is *minimal active pathway* of π_K *iff* there is no active pathway of π_K, which is smaller than $\mathcal{G}$. As this definition shows, we only need to see sets of reactions to compare two pathways. Thus, in the rest of this paper, we sometimes represent a minimal active pathway as a set of reactions.

Any reactions included in a minimal active pathway cannot be deleted to produce target metabolites. Intuitively, this indicates that each of the elements of a minimal active pathway is essential. In practice, minimal active pathways including a large number of reactions are considered to be biologically inefficient. We thus introduce a time limitation z and pathways that can make all target metabolites producible by $t = z$. In the following, we consider the problem of finding minimal active pathways with respect to π_K and z.

V. Knockout Effects

This section provides how to predict knockout effects. In the following, we give some definitions for the prediction. Let $\pi = (N, A, M_S, M_T)$ and $\pi_K = (N, A, M_S, M_T, K)$ be an extended pathway instance and a knockout instance, respectively. In addition, we denote the number of minimal active pathways of π as $|\pi|$ and the number of minimal active pathways of π_K as $|\pi_K|$. Obviously, $|\pi_K| \leq |\pi|$ holds. Then, the gene knockout effect, i.e., the prediction by the proposed method, is given by $E_K = |\pi| - |\pi_K|$. Let K_a and K_b be sets of knockout genes. If $E_{K_a} > E_{K_b}$ holds, then we say that the gene knockout effect of K_a is stronger than that of K_b. If $|\pi_K| = 0$, i.e., $E_K = |\pi|$, then we say that the knockout effect of K is *critical* to produce target metabolites. Various metabolites are known as vital metabolites, which means organisms cannot survive without them. That is, if some gene knockouts are critical to produce such metabolites, then a given organism cannot grow any more or dies. If $|K| = 1$ and its effect is critical to produce vital metabolites, then we say that the gene $g \in K$ is *essential*.

In the following, we explain the above definition with a specific example. Suppose that we are given a pathway instance $\pi = (N, A, M_S, M_T)$, where N and A are from the extended pathway shown in Figure 2, and the source metabolite is $M_S = \{m_1\}$ and the target metabolite is $M_T = \{m_6\}$. Obviously, $|\pi| = 3$ and the minimal active pathways of π are specifically as follows: $\{r_1, r_5, r_7\}, \{r_2, r_3, r_6, r_7\}, \{r_2, r_4, r_6, r_7\}$. Then, we consider the following knockout instances π_{K_1} and π_{K_2}, where $K_1 = \{g_1\}$ and $K_2 = \{g_6\}$. For π_{K_1}, minimal active pathways including r_1 can no longer be solutions, i.e., $|\pi_{K_1}| = 2$. For π_{K_2}, minimal active pathways including r_6 can no longer be solutions either. Thus, $\{r_2, r_3, r_6, r_7\}$ and $\{r_2, r_4, r_6, r_7\}$ are deleted from the solutions of π, i.e., $|\pi_{K_2}| = 1$. Consequently, we can say that the knockout effect of K_2 is stronger than that of K_1. Moreover, suppose that $K = \{g_7\}$. Then, there is no minimal active pathway of π_K and we say that the knockout effect of K is critical to produce m_6. If m_6 is a vital metabolite, we can simultaneously say that g_7 is an essential gene.

In addition to the number of remaining minimal active pathways after knockouts, an important factor in the prediction is the gain of ATPs. This is because pathways that are inefficient with respect to energy consumption will not be used in organisms. Let $|\pi^{a+}|$, $|\pi_K^{a+}|$ be the number of minimal active pathways of π and π_K, which gain ATPs, respectively. Then, the gene knockout effect with respect to ATP production is given by $E_K^{a+} = |\pi^{a+}| - |\pi_K^{a+}|$. In particular, it is important when we consider the glycolysis pathway since one of its main functions is to gain ATPs. However, we cannot find any pathways producing ATPs on some other pathways, i.e., minimal active pathways on them must consume ATPs. In this case, the number of minimal active pathways, which consume fewer ATPs, should be considered instead of $|\pi^{a+}|$ and $|\pi_K^{a+}|$.

VI. Computational Method

This section explains how to compute $|\pi_K|$. In this paper, we use the method of computing all minimal active pathways of π proposed by Soh and Inoue [7]. This method computes pathways through propositional encoding and minimal model generation. An advantage is that this method is flexible for adding biological constraints. Moreover, we can utilize SAT technologies, which have been developed actively in recent years.

In the following, we briefly explain the propositional encoding to compute minimal active pathways of π. Let i, j be integers denoting indices for metabolites and reactions. Let t be an integer variable representing time. Let $\pi = (N, A, M_S, M_T)$ be an extended pathway instance, where $N = M \cup R \cup G$, $A = A_M \cup A_G$. We introduce two kinds of propositional variables. Let $m_{i,t}^*$ be a propositional variable, which is $true$ if a metabolite $m_i \in M$ is producible at time t. Let $r_{j,t}^*$ be a propositional variable, which is $true$ if a reaction $r_j \in R$ is activatable at time t.

The encoding of the problem of finding minimal active pathways with respect to π_K and z is as follows.

$$\psi_1 = \bigwedge_{0 \leq t < z} \bigwedge_{m_i \in M} \left(m_{i,t}^* \rightarrow m_{i,t+1}^*\right)$$

$$\psi_2 = \bigwedge_{0 \leq t < z} \bigwedge_{r_j \in R} \left(r_{j,t}^* \rightarrow r_{j,t+1}^*\right)$$

$$\psi_3 = \bigwedge_{1 \leq t \leq z} \bigwedge_{r_j \in R} \left(r_{j,t}^* \rightarrow \bigwedge_{m_i \in s(r_j)} m_{i,t-1}^* \right)$$

$$\psi_4 = \bigwedge_{1\leq t\leq z} \bigwedge_{r_j\in R} \left(r^*_{j,t} \rightarrow \bigwedge_{m_i\in p(r_j)} m^*_{i,t} \right)$$

$$\psi_5 = \bigwedge_{m_i\in(M\setminus M_S)} \bigwedge_{1\leq t\leq z} \left(m^*_{i,t} \rightarrow m^*_{i,t-1} \vee \bigvee_{r_j\in p'(m_i)} r^*_{j,t} \right)$$

$$\psi_6 = \bigwedge_{m_i\in M_S} m^*_{i,0} \wedge \bigwedge_{m_{i'}\in M\setminus M_S} \neg m^*_{i',0}$$

$$\psi_7 = \bigwedge_{m_i\in M_T} m^*_{i,z}$$

The formulas ψ_1 and ψ_2 represent that once a metabolite (or a reaction) is made to producible (or activatable), then it remains in the producible (or activatable) state. The formula ψ_3 represents that if a reaction r_j is activatable at time t then its reactants must be producible at time $t-1$. The formula ψ_4 represents that if a reaction r_j is activatable at time t then its products must be producible at time t. The formula ψ_5 represents that if a reaction m_i is producible then either two states hold: the metabolite m_i is producible at $t-1$ or at least one reaction r_j is activatable. The formulas ψ_6 and ψ_7 represent source metabolites and target metabolites. We denote the conjunction of $\psi_1, \ldots, \psi_7$ as Ψ_z. Then, we can enumerate minimal active pathways with respect to π_K and z by computing minimal models of Ψ_z with respect to $V^z = \{r^*_{i,z} | r_i \in R\}$.

The computation for π is always needed to compare a wild cell and its mutant. We thus explain a method to compute all minimal active pathways of π_K for a set of knockout genes K. Actually, when the minimal active pathways of π are obtained, we do not need much additional computation. All minimal active pathways of π_K are obtained by selecting pathways that do not contain some $r \in R_K$, where $R_K = \{r \in c'(g) \mid g \in K\}$. The procedure is given as follows: (i) enumerate all minimal active pathways with respect to π and z, (ii) delete minimal active pathways including some $r \in R_K$, where $R_K = \{r \in c'(g) \mid g \in K\}$. As well as the above procedure, there is another way to compute all minimal active pathways with respect to π_K and z. The same is achieved by adding constraints, which inhibit the activation of each reaction in R_K, to the formula Ψ_z.

VII. Experimental Results

This section provides experimental results. At first, we describe experimental conditions. Then, we show the results of our prediction of knockout effects for glycolysis and amino acids biosynthesis.

A. Experimental conditions

We constructed extended pathways from EcoCyc [8] and KEGG [9]. Specifically, we use EcoCyc to construct metabolic pathways, which consists of 1222 metabolites and 1920 reactions. Moreover, we use KEGG to construct relations between enzymatic reactions and genes. In the following experiments, the entire extended pathway is constructed from these two databases. Each experiment has been done using a PC (3.2GHz CPU) running on OS X 10.6. For computation, we use a SAT solver Minisat2 [10]. Koshimura *et al.* proposed a procedure computing minimal models with SAT solvers [11]. We follow their procedure to generate minimal models by using a SAT solver.

To evaluate our method, we use the Keio collection as is described in Section II. In particular, we use their results on the MOPS medium whose main nutrient is glucose. Since these comparative data are obtained from every single gene knockout, in the following, we basically consider that the set of knockout genes K consists of one gene. Moreover, in the Keio collection, if a cell growth is less than 0.1 or not applicable (N.A.) then we say that the cell is strongly affected by a gene knockout.

B. Results for Glycolysis Analysis

First, we analyze the glycolysis pathway of *E. coli*. In accordance with the MOPS medium of the Keio collection [4], we choose source metabolites as follows: β-D-glucose-6-phosphate, H^+, H_2O, ATP, ADP, phosphate, and NAD^+. In addition, pyruvate is given as the target metabolite to analyze glycolysis.

We then compute all minimal active pathways from the entire metabolic pathway of *E. coli*. As we can see in biological literature such as the work of Ferguson *et al.* [12], glycolysis is known to a pathway constructed by eight steps. However, if some reactions are disabled, then *E. coli* is expected to use other bypass pathways by additional reactions. In this experiment, we thus give $z = 12$. Moreover, the number of reactions included in each pathway is limited to less than or equal to 12.

At first, we computed all minimal active pathways using the above conditions and obtained 75 minimal active pathways. We then connected 61 genes to reactions included in them by API on KEGG. Next, we computed minimal active pathways of for each gene knockout. This experiment was done within four seconds. Figure 3 shows the results of 61 gene knockouts. The x-axis denotes each gene knockout and the y-axis denotes the number of minimal active pathways. As is shown in the figure, we compute minimal active pathways of $\pi_{K_1}, \ldots, \pi_{K_{61}}$ such that $K_1 = \{\text{b4025}\}, K_2 = \{\text{b0963}\}, \ldots, K_{61} = \{\text{b2464}\}$. However, since some of the 61 genes construct isozymes, such single gene knockout K_i does not affect the number of minimal active pathways $|\pi_{K_i}|$. However, for reference, we compute the effect of the gene knockouts that disables all of them. For instance, b2133 and b1380 construct isozymes. In this case, the number of minimal active pathways in the figure shows the case of the gene knockout of both b2133 and b1380. For each gene knockout, we computed the gain of ATP in each minimal active pathway, which is calculated by counting the number of both reactions with the coefficient of ATP: ones

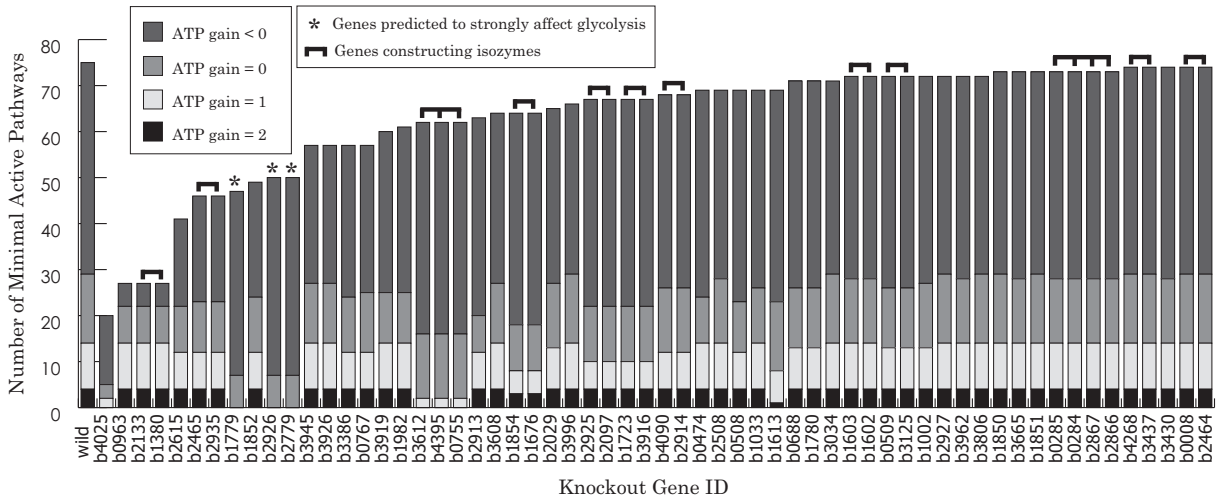


Figure 3. The Number of Minimal Active Pathways for each Gene Knockout on Glycolysis

consuming ATP and the other ones producing ATP. Minimal active pathways that produce the positive number of ATPs are more important than the others because producing ATP is a main function of glycolysis.

From the figure, we can see that *E. coli* keeps almost all minimal active pathways even by more than half of single gene knockouts. This is considered to indicate the robustness of *E. coli*. However, some gene knockouts dramatically reduce the number of minimal active pathways. In particular, the single gene knockouts of b1779, b2926, and b2779 destroy all minimal active pathways producing ATPs. Thus, they are predicted to strongly affect the glycolysis of *E. coli*.

To evaluate the above predictions, we compare them with the Keio collection. Table I compares the first 11 gene knockouts regarding the number of lost minimal active pathways. Column 1, Gene ID, shows identifiers of genes other than *wild*, which denotes an empty set of knockout genes, e.g., $K = \emptyset$. Other rows denote the result of single gene knockout. Column 2, Total, shows the total number of minimal active pathways, i.e., $|\pi_{K_i}|$. Columns 3 to 6 show the number of minimal active pathways, each of which denotes the gain of ATPs. Column 7, MOPS24hr, and Column 8, MOPS48hr, show the cell growth of *E. coli* after 24 hours and 48 hours, respectively. Note that N.A. (not applicable) refers to essential genes [4].

As the first row of Table I shows, we found 14 minimal active pathways that produce the positive number of ATPs on the wild cell of *E. coli* while there are 75 in total.

Distinguished single gene knockouts are K_8 = b1779, K_{10} = b2926, and K_{11} = b2779. Each gene knockout effect with respect to ATP production is $E^{a+}_{K_8} = E^{a+}_{K_{10}} = E^{a+}_{K_{11}} = 14$ and it is the strongest gene knockout effect with respect to ATP production, which is the important function of glycolysis. For this prediction, the Keio collection shows "N.A" for each gene knockout. Thus, in glycolysis, our predictions successfully agree with the results of the Keio collection.

Table I
11 SINGLE GENE KNOCKOUTS FOR GLYCOLYSIS

Gene ID	# of Minimal Active Pathways					Keio Collection [4]	
	Total	ATP Gain				MOPS24hr	MOPS48hr
		2	1	0	<0		
wild	75	4	10	15	46	0.219-0.392	0.216-0.480
b4025	20	0	2	3	15	0.137	0.542
b0963	27	4	10	8	5	0.293	0.371
b2133[a]	27	4	10	8	5	0.303	0.366
b1380[a]	27	4	10	8	5	0.357	0.393
b2615	41	4	8	10	19	N.A.	N.A.
b2465[b]	46	4	8	11	23	0.311	0.315
b2935[b]	46	4	8	11	23	0.317	0.327
b1779*	47	0	0	7	40	N.A.	N.A.
b1852	49	4	8	12	25	0.231	0.223
b2926*	50	0	0	7	43	N.A.	N.A.
b2779*	50	0	0	7	43	N.A.	N.A.

Table II
CRITICAL GENE KNOCKOUTS FOR AMINO ACIDS BIOSYNTHESIS

Gene ID	Unsynthesized Target	Keio Collection [4]	
		MOPS24hr	MOPS48hr
wild	-	0.219-0.392	0.216-0.480
b2153	MET	N.A.	N.A.
b2615	VAL, LEU, THR, ILE, LYS, MET	N.A.	N.A.
b0004	THR	0.000	0.000
b0003	THR	0.004	0.010
b3870	TRP, MET	0.005	0.015
b2329	TRP, PHE, TRP	0.009	0.020
b2838	LYS	0.012	0.021
b3389	PHE, TRP, MET	0.010	0.032
b0074	LEU	0.026	0.034
b3177	MET	0.283	0.293
b4019	MET	0.357	0.509

On the other hand, our method predicted that there are still minimal active pathways that produce the positive number of ATPs after the single gene knockouts of b4025, b0963, and b1852. Those remaining pathways are supposed to be used as bypass pathways. For instance, b4025 encoding glucosephosphate isomerase gene of glycolysis pathway that transfer D-glucose 6-phosphate to D-fluctose 6-phosphate.

 ISBN: 978-1-61208-190-8

However, pentose phosphate pathway is available as a bypass pathway from D-glucose 6-phosphate, resulting in the gene knockout slow-growth at starting MOPS24hr and same level of wild cell final growth at MOPS48hr. Moreover, the knockouts of b2133, b1380, b2465, and b2935 do not affect to the cell growth since they construct isozymes.

The single gene knockout of b2615 is different to the above gene knockouts. Our method predicts that this knockout does not affect the cell growth in terms of glycolysis. However, the Keio collection shows that this is an essential gene for *E. coli*. One assumption is that it affects other functions in the cell. In relation to this, we have additional experiments for amino acid generation in Section VII-C.

C. Results for Amino Acids Generation

We also applied our prediction method to predict gene knockout effects of the cell growth in terms of amino acid biosynthesis. Since we want to involve more genes for our prediction, we particularly focus on essential amino acids for humans, whose synthesis needs more reactions than others. In the experiments, we separately constructed pathway instances, each of which consists of the following eight amino acids as a target metabolites: L-valine (VAL), L-leucine (LEU), L-phenylalanine (PHE), L-isoleucine (ILE), L-threonine (THR), L-lysine (LYS), L-tryptophan (TRP) and L-methionine (MET). In addition, to produce the above amino acids, we added the following metabolites to the source metabolites used in the glycolysis analysis: coenzyme-A and sulfite. For each of the eight amino acids, the computation time is on average 255 seconds and the longest computation time is 877 seconds.

In contrast to the result of glycolysis, we found there are 11 single gene knockouts that destroy all minimal active pathways without the limitation of z. That is, no pathway can synthesize each target on the entire metabolic pathway of *E. coli* with those single gene knockouts. Obviously, they are predicted to be critical to produce each amino acid. Table II shows the cell growth of Keio collection. Column 1, gene ID, shows knockout genes predicted as critical by our prediction. Column 2, unsynthesized target, shows target amino acids, which cannot be synthesized with the knockout of the gene in Column 1. Columns 3 and 4 show the cell growth of *E. coli* after 24 hours and 48 hours, respectively. At first, the gene knockout of b2615 is predicted as critical for the cell growth in terms of six amino acids biosynthesis. This result is also supported by the Keio collection. We thus consider the essentiality of b2615 to be caused by its knockout effect in amino acids biosynthesis rather than glycolysis. Table II also shows that our method predicts that no way to produce target metabolites with each single gene knockout: b0004, b0003, b3870, b2329, b2838, b3389, and b0074. However, the Keio collection shows that *E. coli* survives with very low cell growth. One explanation for the results is that they are suspected to keep living by consuming unsynthesized amino acids from other individual cells. In this case, since the amino acids cannot be sustainably produced, those genes are recognized to be approximately essential for *E. coli*.

Furthermore, the result of the Keio collection shows that the knockouts of b3177 and b4019 are not critical, although our method predicts them to be critical. We have detailed discussions on these gene knockouts in the following section.

VIII. Discussion and Related Work

This section provides detailed discussion about the difference of our prediction and the cell growth of the Keio collection. Figure 4 shows the glycolysis pathway obtained from KEGG [9]. Each enzyme label is replaced to its corresponding gene identifier. The figure also shows four essential genes in terms of the glycolysis pathway confirmed by the Keio collection. Our method predicted three out of four essential genes. However, b2925 is not expected to be critical for the cell growth since the gene knockout cell keeps almost all minimal active pathways that gain ATPs even if we delete both b2097 and b2925. Specifically, the knockout lost only four minimal active pathways that gain one ATP (see Figure 3). Thus, two hypotheses come up. One is that the four lost minimal active pathways are the most important pathways in glycolysis. The other is that the essentiality is caused by the breakdown of other cell functions, similar to the case of b2615. Exploring this issue is a future topic.

The difference between b3177 and b4019 in terms of amino acid biosynthesis also introduces interesting issues. At first, we consider b4019, which constructs an enzymatic reaction methionine synthase. Its conversion is as follows: 5-methyltetrahydrofolate + L-homocysteine = tetrahydrofolate + L-methionine. In both KEGG and EcoCyc databases, there are two alternative reactions and their corresponding genes to the above reaction and b4019. A reaction S-adenosyl-L-methionine uses S-methyl-L-methionine instead of 5-methyltetrahydrofolate. On the other hand, a reaction 5-methyltetrahydropteroyltriglutamate uses 5-methyltetrahydropteroyltri-L-glutamate. However, both metabolites cannot be synthesized from the source metabolites. Specifically, S-methyl-L-methionine can be synthesized only from methionine, which is the target amino acid, and there is no reaction in the metabolic pathway of EcoCyc that can synthesize 5-methyltetrahydropteroyltri-L-glutamate. The gene b3177 is on folate biosynthesis and there is no alternative in the databases. Two hypotheses are as follows: there are unknown complementary genes, or there are unknown bypasses. For the above issues, we need to do more research on more databases and literature.

There are several researches on metabolic pathway analyses. Schuster *et al.* proposed a method called elementary mode analyses [13]. They focused on metabolic flux distributions corresponding to sets of reactions in metabolic pathways. A different point from our method is that their approach needs to define source metabolites strictly with a

 ISBN: 978-1-61208-190-8

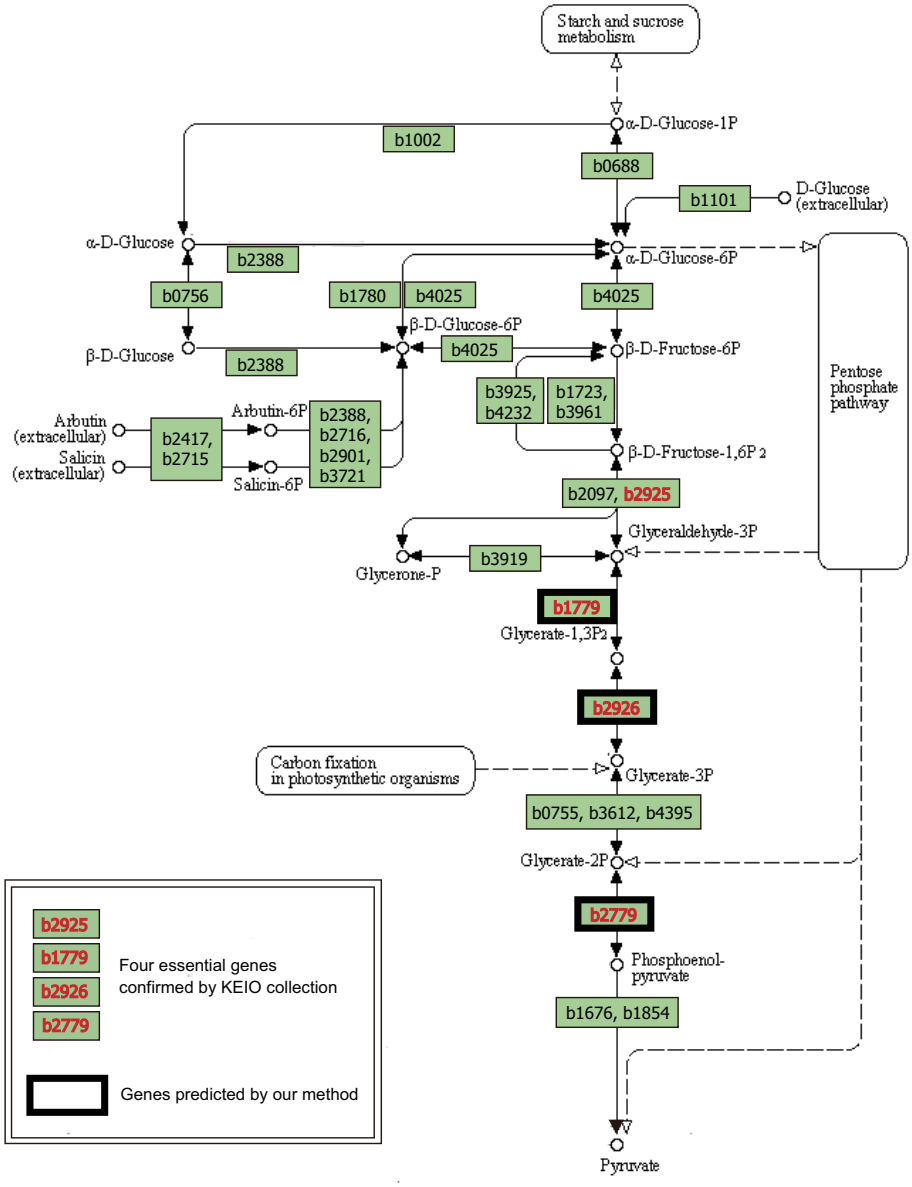


Figure 4. Glycolysis Pathway of *E. coli* K-12 MG1655 from KEGG [9]

fixed amount that must be consumed in flux. In contrast, our method treats them as candidates that will be utilized; thus, we can flexibly give source metabolites. Handorf *et al.* proposed the *inverse scope problem* [14]. This is the problem of finding necessary source metabolites from target metabolites. The two differences between their problem and our proposal are as follows. One is that they only computed the cardinality minimal solution. Unlike their approach, we can generate subset minimal solution by minimal model generation. Another one is that each of their solutions includes all reactions, which are activatable from source metabolites needed to generate target metabolites. For instance, if there are two ways to produce a metabolite from source metabolites then both are mixed in one solution, that is, we cannot distinguish between them. On the other hand, our method can distinguish between the two ways, and we think that it is important to identify functionally minimal active pathways. Schaub and Thiele applied answer set programming (ASP) to solve the inverse scope problem [15], while we use propositional encoding and minimal model generation to compute minimal active pathways.

IX. Conclusion and Future Work

In this paper, we propose a method to predict the knockout effect by enumerating minimal active pathways. We formalize the extended pathway and show the definition of minimal

 ISBN: 978-1-61208-190-8

active pathways on it. In addition, we present a computation method for the prediction. An advantage of our method is that it allows us to trace the reason for the prediction results, e.g., we can suggest the reason for the essentiality of three genes in the glycolysis pathway. This is an important feature that other methods do not have.

In the experiments, we applied our method to extended pathways of *E. coli* and made comparisons using the Keio collection. For the prediction of the knockout of 61 genes in the glycolysis pathway, our method predicted three essential genes, which correspond to the results of the Keio collection. Moreover, we found two essential genes and nine approximately essential genes in amino acids biosynthesis. However, for the knockout of b2925, b3177, and b4019, our prediction indicated different results from the Keio collection. Revealing the reason for this difference is a future work. Moreover, we plan to evaluate the efficiency of the computation method and compare it with other methods. Although we treat relations between genes and enzymatic reactions that have one-to-one relations, we intend to extend them to relations that are more complex such as multiple relations and consider interactions among genes. Following that, we plan to apply our method to other organisms such as mice. In addition to *E. coli*, mice are well known model organisms for human study, and information available on them has been accumulated in the last decade. In particular, chromosome substitution strains are used to reveal the function of genes [16]. In addition to gene knockouts, we could adapt our method to such strains. Although there is a large difference between *E. coli* and mice, the basic metabolism is same. This fact tells us that our method can also be a potential prediction method for mice.

ACKNOWLEDGMENTS

This research is supported in part by the 2008-2011 JSPS Grant-in-Aid for Scientific Research (A) No. 20240016, the research project of Systems Biology of Genetic Function, and the 2nd Transdisciplinary Seeds Research in the Transdisciplinary Research Integration Center. The authors would like to thank anonymous reviewers for their helpful comments.

REFERENCES

[1] H. D. Jong, "Modeling and simulation of genetic regulatory systems: A literature review," *Journal of Computational Biology*, vol. 9, pp. 67–103, 2002.

[2] M. Terzer, N. D. Maynard, M. W. Covert, and J. Stelling, "Genome-scale metabolit networks," *Systems Biology and Medicine*, vol. 1, no. 3, pp. 285 – 297, 2009.

[3] C. J. Tomlin and J. D. Axelrod, "Biology by numbers: mathematical modelling in developmental biology," *Nature Reviews Genetics*, vol. 8, no. 5, pp. 331 – 340, 2007.

[4] T. Baba, T. Ara, M. Hasegawa, Y. Takai, Y. Okumura, M. Baba, K. A. Datsenko, M. Tomita, B. L. Wanner, and H. Mori, "Construction of *Escherichia coli* K-12 in-frame, single-gene knockout mutants: the Keio collection," *Molecular Systems Biology*, vol. 2, no. 2006.0008, 2006.

[5] H. Mizoguchi, H. Mori, and T. Fujio, "*Escherichia coli* minimum genome factory," *Biotechnology and Applied Biochemistry*, vol. 46, no. 3, pp. 157–167, 2007.

[6] N. Ishii, K. Nakahigashi, T. Baba, M. Robert, T. Soga, A. Kanai, T. Hirasawa, M. Naba, K. Hirai, A. Hoque, P. Y. Ho, Y. Kakazu, K. Sugawara, S. Igarashi, S. Harada, T. Masuda, N. Sugiyama, T. Togashi, M. Hasegawa, Y. Takai, K. Yugi, K. Arakawa, N. Iwata, Y. Toya, Y. Nakayama, T. Nishioka, K. Shimizu, H. Mori, and M. Tomita, "Multiple high-throughput analyses monitor the response of *E. coli* to perturbations," *Science*, vol. 316, no. 5824, pp. 593–597, 2007.

[7] T. Soh and K. Inoue, "Identifying necessary reactions in metabolic pathways by minimal model generation," in *PAIS 2010, Proc. ECAI 2010*, 2010, pp. 277–282.

[8] I. M. Keseler, J. Collado-Vides, A. Santos-Zavaleta, M. Peralta-Gil, S. Gama-Castro, L. Muiz-Rascado, C. Bonavides-Martinez, S. Paley, M. Krummenacker, T. Altman, P. Kaipa, A. Spaulding, J. Pacheco, M. Latendresse, C. Fulcher, M. Sarker, A. G. Shearer, A. Mackie, I. Paulsen, R. P. Gunsalus, and P. D. Karp, "EcoCyc: a comprehensive database of *Escherichia coli* biology," *Nucleic Acids Research*, vol. 39, no. suppl 1, pp. D583–D590, 2011.

[9] M. Kanehisa, S. Goto, Y. Sato, M. Furumichi, and M. Tanabe, "KEGG for integration and interpretation of large-scale molecular data sets," *Nucleic Acids Research*, 2011.

[10] N. Eén and N. Sörensson, "An extensible SAT-solver," in *Proc. the 6th International Conference on Theory and Applications of Satisfiability Testing*, 2003, pp. 502–518.

[11] M. Koshimura, H. Nabeshima, H. Fujita, and R. Hasegawa, "Minimal model generation with respect to an atom set," in *Proc. the the 7th International Workshop on First-Order Theorem Proving*, 2009, pp. 49–59.

[12] G. P. Ferguson, S. Totemeyer, M. J. MacLean, and I. R. Booth, "Methylglyoxal production in bacteria: suicide or survival?" *Archives of Microbiology*, vol. 170, no. 4, pp. 209–218, 1998.

[13] S. Schuster, D. A. Fell, and T. Dandekar, "A general definition of metabolic pathways useful for systematic organization and analysis of complex metabolic networks," *Nature Biotechnology*, vol. 18, pp. 326–332, 2000.

[14] T. Handorf, N. Christian, O. Ebenhöh, and D. Kahn, "An environmental perspective on metabolism," *Journal of Theoretical Biology*, vol. 252, no. 3, pp. 530 – 537, 2008.

[15] T. Schaub and S. Thiele, "Metabolic network expansion with answer set programming," in *Proc. the 25th International Conference on Logic Programming*, 2009, pp. 312–326.

[16] T. Takada, A. Mita, A. Maeno, T. Sakai, H. Shitara, Y. Kikkawa, K. Moriwaki, H. Yonekawa, and T. Shiroishi, "Mouse inter-subspecific consomic strains for genetic dissection of quantitative complex traits," *Genome Research*, vol. 18, no. 3, pp. 500–508, 2008.

 ISBN: 978-1-61208-190-8

Effect of Insulating Layers in Loop Applicators for Cardiac Cryoablation

Michael Handler*, Gerald Fischer*, Michael Seger*, Roland Kienast*, Adrian Schuette† and Christian Baumgartner*
*Institute of Electrical, Electronic and Bioengineering
University for Health Sciences, Medical Informatics and Technology
Eduard Wallnöfer-Zentrum 1, 6060 Hall in Tirol, Austria
Email: {michael.handler|gerald.fischer|michael.seger|roland.kienast|christian.baumgartner}@umit.at
†Fraunhofer-Institute for Production Technology IPT
Steinbachstraße 17, 52074 Aachen, Germany
Email: adrian.schuette@ipt.fraunhofer.de

***Abstract*—Cryoablation for the treatment of cardiac arrhythmias inactivates the electrical conduction of arrhythmogenic tissue by freezing. This method has major advantages but also disadvantages compared to radiofrequency ablation. Modeling and simulation of the cryoablation process can be used to reduce the intervention duration and optimize the cryoablation result. In this work, different variants of cryo loops are modeled and simulated using the Finite Element Method. As the wall of the applicator tube has the same thickness in all dimensions the ablation efficiency in tissue direction is reduced by the heating flux of the streaming blood surrounding the applicator. The effects of different insulation layers in blood direction are compared by using heat flux estimations on stationary temperature distributions. As the applicator tube and the insulation layers consist of different fiber materials, anisotropic thermal conductivity is considered in the models. The estimation of thermal longitudinal and transversal conductivity values by a material modeling approach is presented in detail. It is concluded that insulation layers can have positive and unexpected negative effects depending on the insulation position.**

Keywords-cryoablation; loop applicator; finite element method; heat flux; anisotropic thermal conductivity

I. INTRODUCTION

Ablation as treatment for cardiac arrhythmias has been established as a method of choice whenever medical treatment fails or cannot be applied [1]. In this minimally invasive procedure, an applicator is inserted via a sheath through either the femoral artery or vein, or through the subclavian artery into the right or left heart's chambers (atria or ventricles) with X-ray guidance. The aim of ablation techniques in the treatment of cardiac arrhythmias is the electrical inactivation of the arrhythmogenic tissue either by heating (radiofrequency ablation) or freezing (cryoablation).

Radiofrequency (RF) ablation has become the gold standard of the catheter-based approach to tachyarrhythmia ablation [2]. In the late 1990s, catheter-based cryoablation became available as alternative to RF ablation [3]. Cryoablation enables the creation of an elongated lesion in one single ablation (freeze) cycle by using a loop applicator. An additional advantage is the application of *ice mapping*. This technique enables the cardiologist to temporarly inactivate the electrical conduction of small areas of cardiac tissue to test ablation scenarios by cooling to mild temperatures. Another advantage of cryoablation is that it leads to less endothelial disruption than RF ablation [3].

Disadvantages of cryoablation are (i) a higher recurrence rate [4] due to the possible reactivation of the elecrical conductivity, and (ii) a longer intervention duration compared to RF ablation [1].

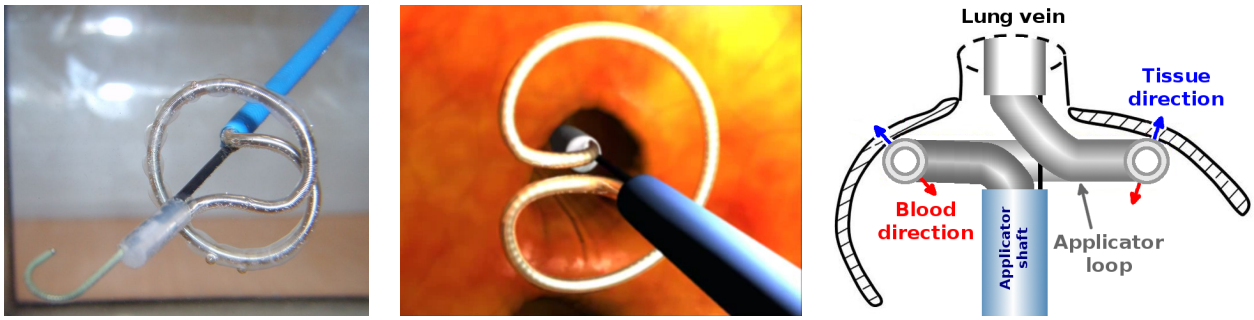
To improve the efficiency of cryoablation and thus reducing the previously mentioned disadvantages, modeling and simulation of the cryoablation process contributes a better understanding of the freezing effects. Seger et al. [1] published a model to simulate the effects of cryoablation mimicking the geometry of a loop shaped cryo-applicator as shown in Figure 1. Such a loop shape applicator enables the creation of a circumferential elongated lesion (approx. 7 cm length) around vessel structure. It can be applied in pulmonary vein ablation for the treatment of atrial fibrillation.

Only a part of the cooling power delivered by the refrigerant is withdrawn from the target tissue. A significant part of the cooling power will create a therapeutically unrelevant heat flow from the circulating blood stream to the cryo-applicator (Figure 1c). Thus, heat exchange between the refrigerant and the blood is considered an undesired loss of cooling power. The spatial limitations imposed by human anatomy onto the construction of catheters also limit the amount of refrigerant which can be used. Efficient use of the refrigerant is therefore crucial. The major scope of this simulation study is to investigate the effect of insulation layers assembled onto elongated cryo-applicators built from fiber reinforced polymers.

In this work, three applicator variants are presented (without insulation, inner insulation, outer insulation). The cooling flux profiles of these variants are compared with each other by creating two dimensional models of the applicators' cross sections. These cooling fluxes are estimated using stationary temperature distributions calculated by the *Finite Element Method (FEM)*.

For realizing mechanically stable loop applicators the material of the applicators consists of carbon fibers embedded

 ISBN: 978-1-61208-190-8



(a) (b) (c)

Figure 1. Loop applicator.
(a) Photo of active loop applicator in water (ablation loop partially surrounded by ice).
(b) Position of loop applicator inside the left atrium during cryoablation.
(c) Schematic illustration of applicator loop in left atrium with cross sectional cutthrough demonstrating tissue direction (blue) and blood direction (red) used in the models of Figure 5.

in a silicone matrix. These carbon fibers are spirally wound to form the applicators tube. The insulation layers consist of quartz fibers surrounded by a silicone matrix to withstand mechanical stress. The different thermal conductivities of these materials need to be considered in order to create a representative temperature distribution within the applicator. Therefore, the anisotropic conductivity parameters, longitudinal and transversal thermal conductivity, are estimated in this work for these fiber materials and considered with regard to the fiber orientation in the applicator variants.

In Section II, the mathematical background for the computation of the stationary temperature distribution and anisotropic conductivity parameters is explained and the estimation of the thermal flux to allow comparison of applicator variants is derived. Section III compares the thermal flux properties of the applicator variants, which are discussed in Section IV.

II. Methods

In this section, the heat diffusion equation used for the computation of the stationary temperature distributions is described. Subsequently the computation of the parameters of the anisotropic thermal conductivity is explained in detail.

A. Simulation of Stationary Temperature Distribution

For the models presented in this work, the general heat diffusion equation (1)

$$-\nabla \cdot k(\mathbf{r})\nabla[T(\mathbf{r})] = Q_r(\mathbf{r}), \quad \mathbf{r} \in \Omega \tag{1}$$

has been used to simulate the stationary temperature distributions [5].

The space dependent parameter k $[Wm^{-1}K^{-1}]$ represents the thermal conductivity; $\mathbf{r}$ holds the cartesian coordinates x and y $[m]$ in two dimensional models; T is the unknown stationary temperature $[K]$; Q_r defines the power contribution of distributed volumetric heat sources due to externally applied spatial heating/cooling $[Wm^{-3}]$; and Ω denotes the analyzed spatial domain. Q_r is set to 0 in the current models as the heating/cooling of tissue is integrated into the model by following two types of boundary conditions: *Dirichlet boundary conditions* define known temperatures at the related boundaries. *Cauchy boundary conditions* were used to model the thermal flux of blood/tissue (heating) and the refrigerant (cooling). The thermal flux induced by Cauchy boundary conditions is defined as [1]

$$\frac{\partial \dot{Q}(\mathbf{r})}{\partial \mathbf{n}_\Gamma(\mathbf{r})} = \alpha(\mathbf{r}) \cdot [T(\mathbf{r}) - T_c(\mathbf{r})]; \quad \mathbf{r} \in \Gamma \tag{2}$$

and depends on the difference between the boundary surface temperature T and the external Temperature T_c. $\dot{Q}$ is the heat power $[W]$, α the heat transition coefficient $[Wm^{-2}K^{-1}]$, Γ the boundary line and $\mathbf{n}_\Gamma$ the outward normal to Γ. Additionally, *homogenous Neumann boundary conditions* were used at the symmetry boundaries of the models.

For the numerical computation the analyzed spatial domain is discretized, boundary conditions are applied and the model is solved using the *Finite Element Method (FEM)*. Simulations for cryoablations using loop applicators have been performed similarly by [1]. In a first approximation, the creation of an elongated lesion can be modeled by an essentially cylindrical geometry. Thus, the cross section of the cryo-applicator tube is a two-dimensional plane, which is considered in this study.

B. Estimation of Thermal Longitudinal and Transversal Conductivity of Anisotropic Materials

At a macroscopic scale it is not possible to consider each single fiber embedded in the polymeric matrix of the tubing. Thus, we first estimate average material properties along and across the fibers assuming that the essentially parallel fibers are arranged in a regular pattern. The applicator tube modeled in this study consists of *carbon fibers* surrounded by a *silicone matrix* (Figure 2).

 ISBN: 978-1-61208-190-8

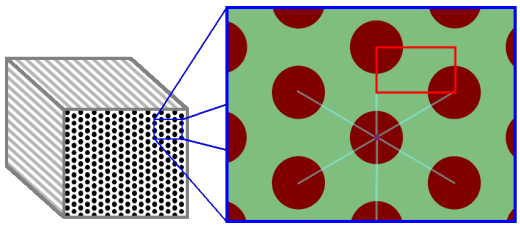


Figure 2. Sketch of the fiber material of the applicator tube. The material consists of equally distributed carbon fibers (red) surrounded by a silicone matrix (green). The red rectangle indicates analyzed domain for the evaluation of thermal transversal conductivity (see Figure 3).

For a proper evaluation of the stationary temperature distribution, the thermal longitudinal and transversal conductivity must be calculated. The average thermal longitudinal conductivity $\overline{k_l}$ is computed by a weighted average of both the fiber (k_f) and the matrix material thermal conductivities (k_m) (3):

$$\overline{k_l} = p_f \cdot k_f + (1 - p_f) \cdot k_m \tag{3}$$

p_f defines the volumetric density ($0 \leq p_f \leq 1$) of fiber in the material.

The thermal transversal conductivity was computed by generating a triangular mesh (Figure 3a) of the area within the red rectangle depicted in Figure 2. This area is sufficient for the computation of the thermal transversal conductivity because it represents the repetitive pattern for the cross section of the fiber material. The geometry of the simulated area was determined from the diameter of the carbon fibers (7 μm) and the volumetric density of carbon fibers in the material ($p_f = 0.6$).

To generate the model depicted in Figure 3a by using the fiber radius r and the fiber density in the material p_f the area covered by fiber A_f (area of a semicircle) can be calculated by (4)

$$A_f = \frac{r^2 \cdot \pi}{2}. \tag{4}$$

By using the known fiber density of the material, the simulation area A_s was computed as follows (5):

$$A_s = \frac{A_f}{p_f} \tag{5}$$

The height h is exactly the half of the diagonal of the model d. The parameter d describes the spatial difference between the centers of two neighboring fibers. As depicted in Figure 2 the height h (see left boundary of red rectangle indicating the simulated area) is exactly half the distance from the centers of the top fiber to the next fiber directly beneath of it. Using the Pythagorean theorem and defining that $d = 2h$, the relation between the width w and height h can be calculated by (6)

$$w = \sqrt{d^2 - h^2} = \sqrt{4h^2 - h^2} = \sqrt{3}h. \tag{6}$$

The product of width w and height h results in the size of the simulation area A_s. Therefore the width w can be determined by (7)

$$w = \frac{A_s}{h} = \frac{\sqrt{3}A_s}{w} = \sqrt{\sqrt{3}A_s}. \tag{7}$$

Summarized the width w and height h of the model can be calculated by (8)

$$w = \sqrt{\frac{\sqrt{3} \cdot r^2 \cdot \pi}{2 \cdot p_f}}, \quad h = \frac{w}{\sqrt{3}}. \tag{8}$$

Note that the mesh was created by using the software package Altair HyperMesh 10 (Altair Eng. Inc.) and for the visualization of the models AmiraDev 3.0 (TGS Template Graphics Software, Inc.) was used.

Using the result of the stationary temperature distribution of the model as shown in Figure 3b the average thermal transversal conductivity $\overline{k_t}$ can be estimated. The sum of heating power $\dot{Q}$ at the bottom of the model is defined as follows:

$$\dot{Q} = \overline{k_t} \cdot \triangle T \cdot \frac{w \cdot l}{h}, \text{ and} \tag{9}$$

$$\dot{Q} = -l \cdot \int_{\Gamma_b} k(\mathbf{r}) \cdot \nabla T(\mathbf{r}) \cdot \mathbf{n}_b d\mathbf{r}, \ \mathbf{r} \in \Gamma_b \tag{10}$$

where w and h are the width and height of the simulated area; l is the length of the fictive model volume in z-direction; ΔT is the temperature difference between the top and the bottom boundary (as Dirichlet boundary conditions were used $+1°C$ at the top and $-1°C$ at the bottom boundary of the model); Γ_b is the lower boundary of the model; $\mathbf{n}_b$ is the outward normal at Γ_b ($\mathbf{n}_b = \begin{pmatrix} 0 & -1 \end{pmatrix}^{\mathrm{T}}$). After assigning the thermal properties of Table I the stationary temperature distribution was computed (Figure 3b).

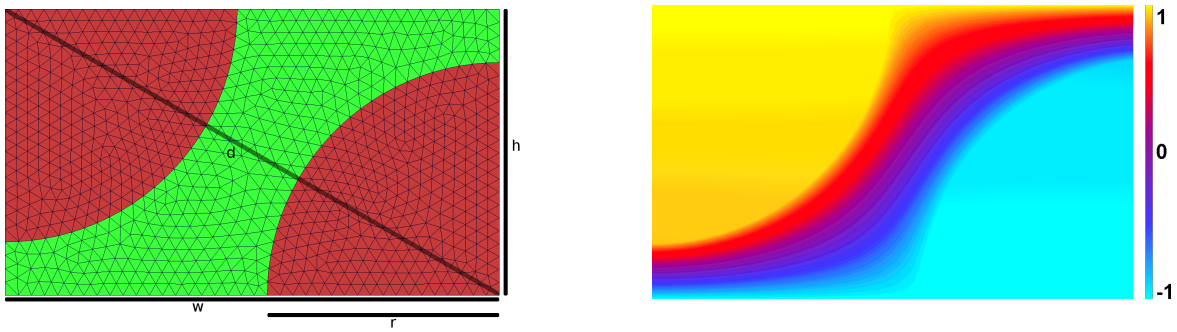
Equation (9) is used to calculate the power $\dot{Q}$ with known transversal conductivity $\overline{k_t}$. This equation was derived from the thermal resistance R defined by the ratio of temperature difference $\triangle T$ and the heating power $\dot{Q}$ at the bottom of the model depicted in Figure 3:

$$R = \frac{\triangle T}{\dot{Q}} \tag{11}$$

Furthermore, R represents the inverse of the average thermal transversal conductivity $\overline{k_t}$. After substituting R by $(\overline{k_t})^{-1}$ and rearranging (11) $\dot{Q}$ can be expressed by (9). Since the heat power $\dot{Q}$ is proportional to the width w and length l and inversely proportional to the height h of the model, the factor $w \cdot l/h$ was added.

As the transversal conductivity $\overline{k_t}$ is unknown, (10) was used to estimate $\dot{Q}$ with known stationary temperature distribution [5][6].

 ISBN: 978-1-61208-190-8



(a) (b)

Figure 3. FEM simulation of the selected red rectangular area in Figure 2 to compute the thermal transversal conductivity of the fiber material with carbon fiber density of 60 % and carbon fiber diameter of 7 μm.
(a) Meshed model (red: carbon fiber, green: silicone matrix)
(b) Stationary temperature distribution after applying constant temperatures of +1$^\circ C$ at the top boundary and -1$^\circ C$ at the bottom boundary of the model

From (9) and (10) we obtain

$$\overline{k_t} = -\frac{h}{w} \cdot \frac{1}{\Delta T} \cdot \int_{\Gamma_b} k(\mathbf{r}) \nabla T(\mathbf{r}) \cdot \mathbf{n}_b d\mathbf{r}; \quad \mathbf{r} \in \Gamma. \qquad (12)$$

For the insulating layers, a quartz fiber material was used. Computation of $\overline{k_l}$ and $\overline{k_t}$ for this material is done analogously to the carbon fiber material. The resulting values for $\overline{k_l}$ and $\overline{k_t}$ for both materials are listed in Table I.

Table I
MATERIAL PROPERTIES USED IN SIMULATIONS.
$\overline{k_l}$ AND $\overline{k_t}$ REFLECT THE COMPUTED LONGITUDINAL AND TRANSVERSAL CONDUCTIVITIES BASED ON (3) AND (12).

	Applicator	**Insulation**
Fiber material	Carbon	Quartz
p_f	0.6	0.5
k_f $[Wm^{-1}K^{-1}]$	17	1
Matrix material	Silicone	
k_m $[Wm^{-1}K^{-1}]$	0.17	
$\overline{k_l}$ $[Wm^{-1}K^{-1}]$	*10.27*	*0.59*
$\overline{k_t}$ $[Wm^{-1}K^{-1}]$	*0.66*	*0.36*

The parameters for the thermal conductivity of carbon and quartz fiber were taken out of the material description provided by the manufacturer of the material. The thermal conductivity of silicone was obtained from [7].

C. Thermal Flux Computation

We can now consider the cryo-applicator tubing at a macroscopic scale by using the average transversal and longitudinal conductivities computed above.

In this work, applicators with different insulation types in blood direction are compared with each other and with an applicator without insulation. To allow a comparison of ablation efficiency the *thermal flux* was computed over the applicator surface.

Therefore, models of the applicator variants were created (Figure 5). Due to the symmetry only the right half of two dimensional cuts of the applicator variants were modeled. The diameter of all applicator variants is 2.3 mm; the applicator tube without insulation has a thickness of 181 μm and the maximum insulation thickness for both insulated variants is 245 μm. The spatial domain of the models was discretized to allow the numerical computation of the stationary temperature distributions within the applicator variants during cryoablation.

For computation of the stationary temperature distribution the anisotropic thermal conductivities depicted in Table I were used. The carbon fibers of the applicators are aligned clockwise with a winding angle of $\beta = 80^\circ$, whereas the fibers of the insulating material are aligned perpendicular to the cross section (see Figure 4). Therefore, the longitudinal conductivity is not considered in the insulating material as only the transversal conductivity of the insulating material is needed.

Cauchy boundary conditions were assigned to the refrigerant boundary Γ_i ($\alpha = 3000\ Wm^{-2}K^{-1}, T_c = -90^\circ C$) and to the blood/tissue boundary Γ_o ($\alpha = 3000\ Wm^{-2}K^{-1}, T_c = 36.5^\circ C$) [7].

After computing the stationary temperature distribution of the applicators, the thermal flux into the applicator P $[Wm^{-2}]$ at the boundary Γ_o can be computed similarly to the Cauchy boundary condition in (2):

$$P(\mathbf{r}) = -\frac{\partial \dot{Q}(\mathbf{r})}{\partial \mathbf{n}_\Gamma(\mathbf{r})} = \alpha(\mathbf{r}) \cdot [T_c(\mathbf{r}) - T(\mathbf{r})], \quad \mathbf{r} \in \Gamma \quad (13)$$

P is obtained from the Cauchy boundary condition. Heat flow from the tissue to the applicator is defined as positive. Therefore the multiplicative factor -1 was added to (13).

 ISBN: 978-1-61208-190-8

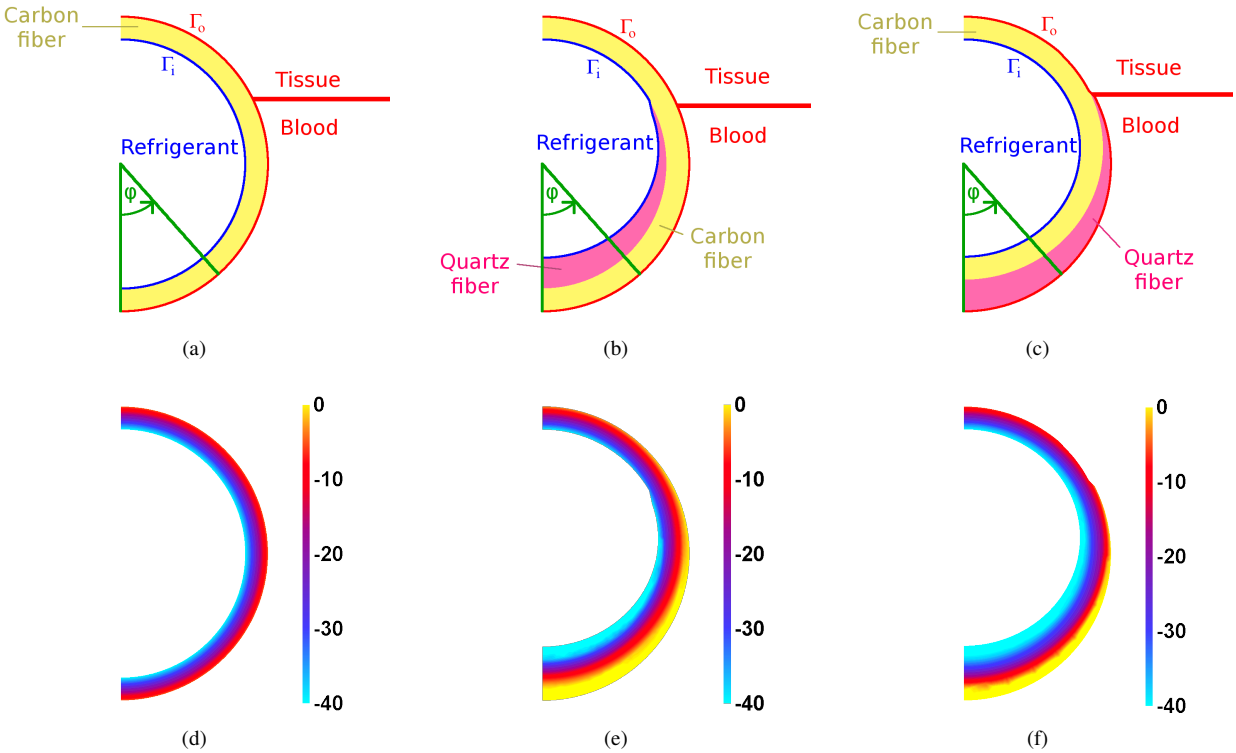


Figure 5. Model of applicator variants. Angle φ is used in Figure 6 for the comparison of the heat flux. (a) without insulation; (b) with inner insulation; (c) with outer insulation; (d) temperature distribution without insulation; (e) temperature distribution with inner insulation; (f) temperature distribution with outer insulation.

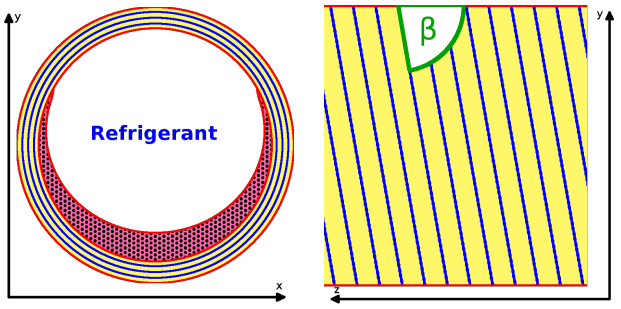


Figure 4. Fiber orientation in applicator models. The fiber orientation of the applicator tube is spirally wound with winding angle $\beta = 80°$ in z direction and the fiber orientation of the insulation layer is parallel to the z axis (winding angle 0°). (a) Cross section of the applicator variant with inner insulation (blue: carbon fibers of tube; yellow: silicone matrix of carbon fiber material; black: quartz fibers of insulation layer; pink: silicone matrix of insulation layer) (b) Perspective along the applicator surface depicting carbon fibers (blue) spirally wound with winding angle $\beta = 80°$ in z direction.

III. Results

Using the stationary temperature distributions of the models depicted in Figure 5 and the thermal flux equation (13) for the surface Γ_o the angle dependent thermal flux values were computed and compared with each other in Figure 6. The visualization of the thermal flux starts at the thickest insulation point in blood direction at $\varphi = 0°$ and proceeds counterclockwise with increasing φ in tissue direction until $\varphi = 180°$.

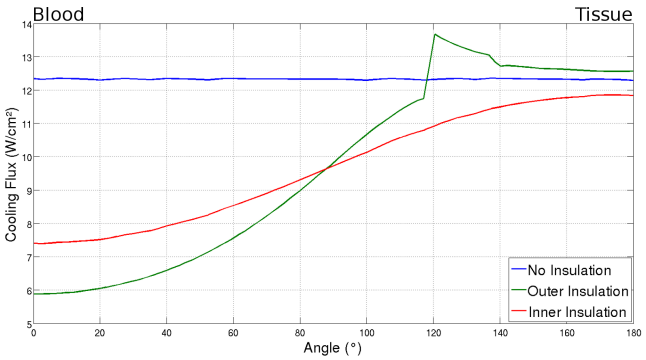


Figure 6. Comparison of angle dependent thermal flux values of the three applicator variants depicted in Figure 5 with $\varphi = 0°$ at the bottom (blood side) and $\varphi = 180°$ at the top (tissue side). Angle increases counterclockwise.

As depicted in Figure 6, the thermal flux over the surface of the applicator without insulation has a constant value of 12.3 W/cm^2.

The applicator with inner insulation shows a decrease of the cooling flux in blood direction. The minimal thermal flux

 ISBN: 978-1-61208-190-8

of 7.4 W/cm^2 is present at $\varphi = 0°$ (maximal thickness of insulation layer). It increases with rising φ and reaches its maximum of 11.8 W/cm^2 at $\varphi = 180°$.

The thermal flux profile of the applicator with outer insulation shows an even lower value of 5.9 W/cm^2 than both other variants at $\varphi = 0°$ (maximal thickness of insulation layer). Contrary to the inner insulated applicator the maximum cooling flux is located at $\varphi = 120°$ with the value of 13.7 W/cm^2 and reduces with increasing φ until it reaches the value of 12.6 W/cm^2 at $\varphi = 180°$.

IV. Conclusion and Future Work

The major scope of this study is the development of simulation tools, which should guide the development of efficient applicator tubings for the creation of elongated lesions. Three different applicator constructions using fiber reinforced polymers were compared.

As depicted in Figure 6 the best of the three applicator variants presented in this work (without/inner/outer insulation) is the outer insulated applicator. One of the main reasons for the efficiency of this applicator type is the anisotropic thermal conductivity property of the carbon fiber material that directs the cooling power from behind the insulation layer to the tissue front where it is required.

Obviously, the applicator variant without insulation (blue), showed no considerable variations of the thermal flux along the outer boundary. Note that the small variations are caused by spatial discretization and numerical effects.

The applicator variant with inner insulation (red) shows the desired effect of reducing the power loss in blood direction. As a negative side effect the thermal flux in tissue direction is also reduced. This may be caused by the anisotropy of the surrounding carbon fiber material. The inner insulation in blood direction allows for a stronger heating effect of the carbon fiber. Due to the higher thermal conductivity along the carbon fiber the temperature difference between the blood side (warm) and tissue side (cold) is reduced.

The outer insulation of the applicator showed the best results of the applicator variants presented in this investigation. In the direction of blood, a significant cooling flux reduction is seen. Additionally, there is an improved cooling flux in the tissue direction. The maximum at $\varphi = 120°$ (insulation end for both insulation variants) can be explained by the low temperatures of the carbon fiber material within the insulating layer and the low thermal conductivity of the insulation layer from $\varphi = 0°$ to $\varphi = 120°$. After the insulation layer vanishes at $\varphi = 120°$ the thermal conductivity rises immediately. At the end of the insulating layer the highest temperature gradient occurs that reduces progressively until $\varphi = 180°$.

The manufacturing of an applicator with an additional layer on the outside that does not surround the whole surface may cause mechanical instability of the applicator. Due to the missing linkage between the two layers it would be necessary to add additional material around the applicator to adhere the insulation to the applicator. Although this provides a possible solution to the problem, this method might reduce the cooling effect of the applicator in tissue direction.

Comparison was carried out by computing a stationary temperature field in a two-dimensional geometry. Such a model seems to be justified for making principle decisions (insulation inside or outside) when starting the development of a production process. Fiber reinforced polymers have been successfully applied in other fields of medical devices such as MR compatible materials [8] and have the potential to increase the efficiency of cryoablation in the future. The results obtained in this study are based on the assumption of an ideal production process, which delivers optimal contact of fibers and polymeric matrix material. In future studies the results should be validated by comparing the simulated predictions with manufactured tubings.

Acknowledgment

This work was supported by the K-Regio-Project of the Standortagentur Tirol, Innsbruck, Austria.

References

[1] M. Seger, G. Fischer, M. Handler, F. Hintringer, and C. Baumgartner, "Simulating the Spatio-Temporal Spread of Cooling Flux in Cardiac Tissue due to Catheterized Cryoablation," in *Proc. 8th IASTED Int. Conf. on Biomedical Engineering*. ACTA Press, 2011, pp. 83–88.

[2] D. Lustgarten, D. Keane, and J. Ruskin, "Cryothermal Ablation: Mechanism of Tissue Injury and Current Experience in the Treatment of Tachyarrhythmias," *Progress in Cardiovascular Diseases*, vol. 41, no. 6, pp. 481–498, 1999.

[3] A. Skanes, G. Klein, A. Krahn, and R. Yee, "Cryoablation: Potentials and Pitfalls," *J Cardiovasc Electrophysiol*, vol. 15, no. 10 Suppl, pp. S28–S34, 2004.

[4] B. Schwagten, Y. Van Belle, and L. Jordaens, "Cryoablation: how to improve results in atrioventricular nodal reentrant tachycardia ablation?" *Europace*, vol. 12, no. 11, p. 1522, 2010.

[5] J. Lienhard and J. Lienhard, *A Heat Transfer Textbook*, 4th ed. Phlogiston Press, 2011.

[6] M. Delfour, G. Payre, and J. Zolesio, "Optimal Design of a Minimum Weight Thermal Diffuser with Constraint on the Output Thermal Power Flux," *Applied Mathematics & Optimization*, vol. 9, no. 1, pp. 225–262, 1982.

[7] H. Kuchling, *Taschenbuch der Physik*. Hanser Verlag, 2007.

[8] A. Schuette, C. Brecher, S. Krüger, and G. Krombach, "New Materials for Magnetic Resonance Imaging (MRI) – Fiber-Reinforced Guide Wires and Catheters for Minimal Invasive Interventions," in *World Congress on Medical Physics and Biomedical Engineering, September 7-12, 2009, Munich, Germany*. Springer, 2009, pp. 163–165.

 ISBN: 978-1-61208-190-8