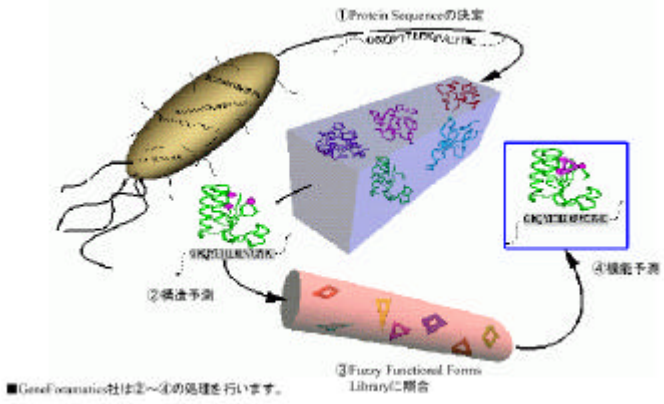
GeneFormatics 社　3次構造をベースにした

蛋白質機能予測サービスのご紹介

株式会社　理経
バイオインフォマティクス販売推進室
小澤　幸弘

１．緒　　言

近年では製薬会社や解析センターなどで Sequence データが大量に産出されています。その中からいかに早く有用な Gene を探し、創薬および特許に結び付けるかが重要になってきています。今回ご紹介する GeneFormatics 社の「蛋白質機能予測サービス」は構造･機能予測の分野で世界的に権威ある Dr. Jeffrey Skolnick 等により開発された Fuzzy Functional Form という技術を用いて Sequence データの機能を予測し創薬開発、特許戦略の支援をする革新的なサービスです。最近のレポートによると Sequence のホモロジーが高くても機能は異なる、また機能は同じでもホモロジーは低いと行った報告が出ています。今まで 1 次構造をベースに Sequence のホモロジーから機能を予測していたのに対し、同社は 3 次構造をベースに機能予測するため少なく見積もって 10～30%の精度の向上を産出します。また１０００Sequence を一日に処理できるため労力、コストそして期間を大幅に削減できます。



２．１　Novel Technology

GeneFormatics' Scientific Founders have developed leading-edge technology that is based on the idea that the structures of protein functional sites are three-dimensional, while genetic sequence information is inherently limited to one dimension. Currently, techniques are available for making function predictions based solely on the one-dimensional sequence information, but these methods produce inaccurate and incomplete results. Recognizing the limitations of these sequence-only techniques, the Scientific Founders at GeneFormatics have developed the Prediction Factory, which utilizes both the sequence and three-dimensional structural information to predict function. The Prediction Factory consists of two complementary processes: the Structure Prediction Factory and the Function Prediction Factory.

２．２　Structure Prediction Factory

The Structure Prediction Factory uses novel techniques for determining a protein's three-dimensional structure when only its sequence is known. This predicted structure is inexact, but can be created quickly without expending significant computational resources. However, it is sufficiently accurate for use in the Function Prediction Factory for determining the protein's functions and active site residues.

2.3 Function Prediction Factory

The Function Prediction Factory uses GeneFormatics' proprietary Fuzzy Functional Form? technology to rapidly identify genomic sites of interest, allowing comparative analysis of sequences that would otherwise not be recognized as being related.

Fuzzy Functional Forms are structural descriptors of the protein's functional sites. In the Function Prediction Factory, Fuzzy Functional Forms are applied to the low resolution protein structures created in the Structure Prediction Factory to find proteins with similar structural characteristics.

Protein function predictions made this way have been found to be much more accurate than those resulting directly from protein sequence data. An additional advantage is that the specific functional site is automatically identified in the protein's structure.

2.4 Remarkable Results

In peer-reviewed publications, GeneFormatics' Prediction Factory Technology has been shown to achieve a conservatively estimated 10-30% higher recognition rate of newly sequenced proteins over current sequence-based prediction methods. The Scientific Founders have validated this approach by correctly identifying all proteins that were already known to exhibit a given function for several genomes. More importantly, this technology has made several correct predictions for novel proteins of previously unknown function, that could not be predicted using current methods.

3.1 Function Annotation Services

GeneFormatics delivers comprehensive bioinformatics-based services for the next level of proteomics research ? studying and understanding the structure and function of proteins. The Prediction Factory, developed in the laboratories of the company's Scientific Founders, can produce detailed functional information for specific proprietary protein sequences. These multi-level, biologically relevant function annotations can aid in target recognition, significantly decreasing the time required to develop a new product starting from the gene protein sequence.

Function Annotation Services provide our customers with information about the likely functions of each sequence. Each sequence is annotated and the amino acid residues involved in activity and binding are explicitly identified. Models of the three-dimensional protein structures for each protein sequence may also be produced.

3.2 Genome Analysis Services

By far the most valuable method for sifting out proteins from genomic databases is the efficient prediction of protein function. At GeneFormatics, the Scientific Founders have created the technology to predict the biological function of a protein from the raw sequence data. Unlike other methods, GeneFormatics' Prediction Factory is not restricted to making predictions based on evolutionary relationships between genes, and (unlike current publicly available tools) it is able to identify the specific amino acid residues which are involved in the protein's functional activity.

GeneFormatics' solution is not genome specific, but rather is applicable to any genome, whether human, plant, animal or microbial. Through Genome Analysis Services, Geneformatics' technology can be applied quickly and easily to massive proprietary or public genome sequence databases. Correct identification of useful leads from this huge amount of data allows our customers to concentrate their research dollars on targets that have a high potential for success.

3.3 Model Building Services

GeneFormatics provides state-of-the-art analysis of proprietary sequences at significantly less cost than

that required for setting up an in-house bioinformatics effort, and with significantly more valuable information than that which is generated by other methods. Our Scientists and Staff have over 25 years of combined experience in protein structure prediction and in development, application, and analysis of genomics and proteomics bioinformatics tools.

We offer Model Building Services, which use our Structure Prediction Factory technology to provide our customers with approximate three-dimensional structural models based on proprietary sequences. These inexact models are accurate enough to be used to identify functional sites, but can be quickly generated in large numbers to screen whole genomes. More detailed, exact models then need only be produced for those proteins determined to be of functional interest. These detailed models, with specific identification of active-site residues and geometrics, can then be used in the next step of product development.

4.1 Scientific Founders

Jeffrey Skolnick, Ph.D.

Director of Computational and Structural Biology,
The Donald Danforth Plant Science Center

Dr. Skolnick received his Ph.D. in chemistry from Yale University in 1978. He was a postdoctoral fellow at Bell Labs in 1979, and then joined the faculty at Louisiana State University as an Assistant Professor. He moved to Washington University in St. Louis in 1982 and was promoted to Full Professor there in 1988. Dr. Skolnick was hired as a Full Professor in Molecular Biology at The Scripps Research Institute in 1989. In 1999, Dr. Skolnick was hired as the Director of Computational and Structural Biology at The Donald Danforth Plant Science Center in St. Louis. His research focuses on protein structure analysis, structure prediction, and function prediction. He is a member of two professional societies and has served on numerous NIH study sections. He is an editorial board member of three different journals and is a referee for 18 different peer-reviewed journals. Dr. Skolnick has published over 200 research articles during his scientific career and has received grants from the NIH, NSF, the American Chemical Society, and the Sloan Foundation.

Jacquelyn Fetrow, Ph.D.

Chief Scientific Officer

Dr. Fetrow received her Ph.D. in biochemistry from the Pennsylvania State University in 1986. She received an NIH postdoctoral fellowship which she served at the University of Rochester Medical School and at MIT's Whitehead Institute for Biomedical Research. In 1990, she joined the faculty at the University at Albany in Albany, NY and was promoted to Associate Professor with tenure in 1995. In 1998, she joined the faculty at The Scripps Research Institute as an Associate Professor where her research focused on protein structure and function prediction. In 1999, she moved to GeneFormatics as the Chief Scientific Officer. Dr. Fetrow is a member of eight professional societies, is an editorial board member for Proteins: Structure, Function & Genetics, serves as a referee for eight peer-reviewed journals, and has served on two NIH, one NSF, and one NASA study panels. She is the author of over 30 publications and has received research grants from the NIH, NSF, and American Chemical Society.

Andrzej Kolinski, Ph.D.

Professor of Computational Genomics,The Donald Danforth Plant Science Center;
Professor of Chemistry, University of Warsaw

Dr. Kolinski received his Ph.D. in chemistry from the University of Warsaw in 1979. He has been on the faculty of the University of Warsaw since he earned his doctorate, and is currently a Full Professor there. In 1989, Dr. Kolinski joined the

faculty at The Scripps Research Institute in La Jolla, CA. In 1999, he was hired as a Full Professor at The Donald Danforth Plant Science Center in St. Louis. Dr. Kolinski is also an International Scholar of the Howard Hughes Medical Institute. He has won several awards, including the Switoslawski Award for the Best Science Done in 1984-1989 at the University of Warsaw. His research focuses on protein folding and computer simulations of biopolymeric systems. Dr. Kolinski is an author of over 115 research publications.

Adam Godzik, Ph.D.

Associate Professor,Program Director, Bioinformatics and Biological Complexity, The Burnham Institute

Dr. Godzik received his Ph.D. in physics from the University of Warsaw in 1990 and became a visiting scientist at the European Molecular Biology Laboratories in 1990. Dr. Godzik then moved to The Scripps Research Institute where he was a postdoctoral fellow from 1990 to 1992 and an Assistant Professor from 1992 until 1998. He then moved to the Burnham Institute as an Associate Professor. In 1999 he became Program Director for the Bioinformatics and Biological Complexity program at the Burnham Institute. Dr. Godzik's research focuses on analysis of protein structure and folding and on prediction of protein structure and function. He is a member of two professional societies and is referee for eight peer-reviewed journals. He is an author on over 40 research publications. Dr. Godzik has received research grants from the NIH and the NSF.

4. 1 Reference

1.J. Skolnick, J.S. Fetrow: TIBTECH 18:34-39 (2000), From genes to protein structure and function: Novel applications of computational approaches in the genomic era.

2. A. Kolinski, P. Rotkiewicz, B. Ilkowski, J. Skolnick: Proteins 37:593-610 (1999), A method for the improvement of threading-based protein models.

3. J. Fetrow, N. Siew, J. Skolnick: FASEB Journal (1999), Structure-based functional motif identifies a potential disulfide oxidoreductase active site in the serine/threonine protein phosphatase-1 subfamily.

4. Zhang, L., Godzik, A., Skolnick, J., and Fetrow, J.S.: Folding and Design 1998;3(6):535-48. Functional analysis of E. coli proteins for members of the alpha/beta hydrolase family.

5.Fetrow, J.S., Godzik, A., and Skolnick, J.: J Mol Biol 1998 Oct 2;282(4):703-711, Functional analysis of the Escherichia coli genome using the sequence-to-structure-to-function paradigm: identification of proteins exhibiting the glutaredoxin/thioredoxin disulfide oxidoreductase activity.

6.Fetrow, J.S., and Skolnick, J.: J Mol Biol 1998 Sep 4;281(5):949-68, Method for prediction of protein function from sequence using the sequence-to-structure-to-function paradigm with application to glutaredoxins/thioredoxins and T1 ribonucleases.

7.L. Rychlewski, B. Zhang and A. Godzik: Protein Science 8:614-24 (1999), Functional insights from structural predictions: analysis of the Escherichia coli genome.

8.B. Zhang, L. Rychlewski, P. Pawlowski, J. Fetrow, J. Skolnick and A. Godzik: Protein Science 8:1104-1115 (1999), From fold predictions to function predictions: Automation of functional site conservation analysis for functional genome predictions.