

## NEW HOT PAPERS - 2009

March 2009



**Kuo-Chen Chou talks with *ScienceWatch.com* and answers a few questions about this month's New Hot Paper in the field of Biology & Biochemistry. The author has also sent along images of their work.**



### Article Title: Recent progress in protein subcellular location prediction

Authors: Chou, KC;Shen, HB

Journal: ANAL BIOCHEM

Volume: 370

Issue: 1

Page: 1-16

Year: NOV 1 2007

\* Gordon Life Sci Inst, San Diego, CA 92130 USA.

\* Gordon Life Sci Inst, San Diego, CA 92130 USA.

(addresses have been truncated)

### SW: Why do you think your paper is highly cited?

Information on the subcellular locations of proteins is important because it can provide useful insights about their functions, as well as how and in what kind of cellular environments they interact with each other and with other molecules. It is also fundamental and indispensable to systems biology because knowledge of the localization of proteins within cellular compartments can help us to understand the intricate pathways that regulate biological processes at the cellular level.

Although the subcellular locations of proteins can be determined by conducting various biochemical experiments, the approach by purely doing experiments is both time-consuming and costly. In particular, the number of newly-found protein sequences has been explosively increased. For instance, in 1986, the **Swiss-Prot databank** contained merely 3,939 protein sequence entries, but the number has since jumped to 408,099, according to version 56.7 of January 20, 2009, meaning that the number of protein sequence entries now is more than 103 times the number from about 23 years ago.

With the avalanche of gene products generated in the postgenomic age, the gap between newly-found protein sequences and the knowledge of their subcellular localization is becoming increasingly wide. Therefore, it has been highly desirable to develop computational methods by which one can quickly predict the subcellular locations of proteins, based on their sequence information alone.

During the past 15 years or so, many predictors have been developed in this regard. Our paper was focused on recent advancements, particularly on those predictors that distinguish themselves by having some remarkable features, such as the ability to be able to deal with proteins

[ScienceWatch Home](#)[Inside This Month...](#)[Interviews](#)[Featured Interviews](#)[Author Commentaries](#)[Institutional Interviews](#)[Journal Interviews](#)[Podcasts](#)[Analyses](#)[Featured Analyses](#)[What's Hot In...](#)[Special Topics](#)[Data & Rankings](#)[Sci-Bytes](#)[Fast Breaking Papers](#)[New Hot Papers](#)[Emerging Research Fronts](#)[Fast Moving Fronts](#)[Corporate Research Fronts](#)[Research Front Maps](#)[Current Classics](#)[Top Topics](#)[Rising Stars](#)[New Entrants](#)[Country Profiles](#)[About Science Watch](#)[Methodology](#)[Archives](#)[Contact Us](#)[RSS Feeds](#)

having multiple subcellular locations, the state-of-the-art prediction engines, the rigorous procedures for constructing high-quality organism-specific benchmark datasets, and the user-friendly web-servers accessible to the public. These features might be the reason for the paper being highly cited.

**SW: Does it describe a new discovery, methodology, or synthesis of knowledge?**

The paper did describe several new concepts and methodologies, such as pseudo amino acid composition, or PseAA composition, hybridization of the "higher level" approach with the *ab initio* approach, ensemble classifier, and how to deal with multiplex proteins which may simultaneously exist at, or move between, two or more different subcellular locations. Proteins with multiple locations or dynamic features of this kind are particularly interesting because they may have some special biological functions intriguing to investigators in both basic research and areas of drug discovery.

**SW: Would you summarize the significance of your paper in layman's terms?**

A cell is the most elementary unit of life. Its survival and replication will depend on the proper functions of many proteins therein; while the latter will depend on whether they are correctly located in their compartments or organelles (see Fig.1 [.PDF file]), called subcellular locations. It may cause various strange diseases should they occur at the wrong location sites. Therefore, one of the fundamental problems in cell biology and proteomics is to identify the subcellular locations and functions of these proteins, the cell's primary machinery.

To address this problem, a user-very-friendly web-server package called **Cell-PLoc** is described. It is freely accessible to the public via the web site. Cell-PLoc contains six predictors: Euk-mPLoc, Hum-mPLoc, Plant-PLoc, Gpos-PLoc, Gneg-PLoc, and Virus-PLoc, specialized for eukaryotic, human, plant, Gram-positive bacterial, Gram-negative bacterial, and virus proteins, respectively (see Fig.2 [.PDF file]).

For example, by using Euk-mPLoc, one can easily predict the subcellular location site(s) of a specific eukaryotic protein among the 22 possible location sites, as illustrated in Fig.1, with a high expected accuracy, in only about five seconds. Readers can also find a step-to-step guide for how to use the Cell-PLoc package in a recent protocol article (Chou KC, *et al.*, "Cell-PLoc: a package of Web servers for predicting subcellular localization of proteins in various organisms," *Nature Protocols*: 3: 153-62, 2008).

**SW: How did you become involved in this research, and were there any problems along the way?**

I was engaged with this research when I worked with the Pharmacia & Upjohn Pharmaceutical Company, because it could provide very useful information for drug discovery, and many project teams often came to my office asking about essentially the same question. This led me to design a web-server for predicting protein subcellular localization. Of course, it was available only within the company and its power was quite limited. After Pharmacia was merged with Pfizer in 2003 and I left the company, in collaboration with Dr. Hong-Bin Shen, my postdoc from Shanghai Jiaotong University, we eventually established the current Cell-PLoc which is much more powerful and freely accessible to the public.

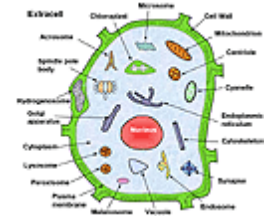
**SW: Where do you see your research leading in the future?**

We are prepared to establish a much more complete web-server package which can more efficiently predict protein subcellular localizations by covering a wider scope and increasing the accuracy expectancy, while also predicting all the other protein attributes.

**SW: Do you foresee any social or political implications for your research?**

The social implications I have foreseen is that publicly accessible web-servers will become more and more popular and these will have important impacts not only on science but also on the economy and other aspects of our lives.

**Kuo-Chen Chou, Ph.D.D.Sc.**  
**Chief Scientist**  
**Gordon Life Science Institute**  
**San Diego, CA, USA**  
**And**



22 different components or organelles in a eukaryotic cell.

View/download two accompanying slides and descriptions.



**Advisory Professor**  
**Shanghai Jiaotong University**  
**Shanghai, P. R. China**

**Web**

KEYWORDS: AMINO-ACID-COMPOSITION; SUPPORT VECTOR MACHINES; FUNCTIONAL DOMAIN COMPOSITION; STRUCTURAL CLASS PREDICTION; NEAREST-NEIGHBOR ALGORITHM; GRAM-NEGATIVE BACTERIA; LOCALIZATION PREDICTION; ENSEMBLE CLASSIFIER; CONOTOXIN SUPERFAMILY; FUSION CLASSIFIER.

 PDF

[back to top](#) 

2009 : March 2009 - New Hot Papers : Kuo-Chen Chou

[Scientific Home](#) | [About Scientific](#) | [Site Search](#) | [Site Map](#)

[Copyright Notices](#) | [Terms of Use](#) | [Privacy Statement](#)