出國報告（出國類別：其他：國際會議）

出席印尼峇里島 2016 ICCBS 國際研討會報告書

服務機關：國立高雄應用科技大學

姓名職稱：林煜勛/碩士生

派赴國家：印尼

出國期間：2016.06.23-2016.07.04

回國報告日期：2016.07.08

摘要

我們研究論文 Rooted Tree Optimization Algorithm for Protein Folding Prediction 投稿至 2016 5th International Conference on Bioinformatics and Biomedical Science (ICBBS 2016)國際研討會後被接受並且在 (6 月 25 日至 27 日)至印尼峇里島口頭報告論文，並參加他們所安排的當地參觀行程。首先早上有 4 位教授演講，而本論文報告被安排在 26 日下午 session 4，隔天有峇里島的一些景點、學校參觀行程，過程中交流了不少學術交流。最後感謝擁有補助可以讓我們更無負擔的到國外參加國際研討會增廣見聞，並且交流更多沒想過或未接觸的知識，以利日後有更好的研究。

關鍵詞：Hydrophobic-polar model、rooted tree optimization、protein structure prediction、ICBBS 2016 國際研討會、印尼峇里島。

目次

一、目的

我們研究論文 Rooted Tree Optimization Algorithm for Protein Folding Prediction 投稿至 2016 5th International Conference on Bioinformatics and Biomedical Science (ICBBS 2016)國際研討會後被接受並且在 (6 月 25 日至 27 日)至印尼峇里島口頭報告論文，並參加他們所安排的當地參觀行程。

二、過程

會議地點

Patra Jasa Bali Resort & Villas, Bali, Indonesia.

研討會簡述

ICBBS 2016 國際研討會由 association of the scientists and engineers in Chemical, Biological, & Environmental Engineering (CBEES) 協會主辦，本屆研討會主要研究領域包含三大主題：Bioinformatics and Computational Biology (包含 Protein structure, function and sequence analysis、Computational proteomics 及 Algorithms, models, software, and tools in Bioinformatics 等多項主題)、Biomedical Engineering (Biomedical imaging, image processing & visualization、Bioelectrical and neural engineering 及 Biomechanics and bio-transport)及 Other Related Topics (Biostatics、Biometric 及 Biomeasurement)。發表的論文將會分別收錄在 Journal of Life Sciences and Technologies (JOLST, ISSN: 2301-3672)與 International Journal of Pharma Medicine and Biological Sciences (IJPMBS, ISSN: 2278-5221)兩個期刊之中。

會議議程

Brief Schedule for Conferences

<table>
<tr><td>Day 1</td><td colspan="2">Afternoon, June 25, 2016 (Saturday) Venue: Lobby
Arrival Registration 13:30~17:00
(Committee Meeting 14:00~16:00)</td></tr>
<tr><td rowspan="11">Day 2</td><td colspan="2">June 26, 2016 (Sunday) 8:50~17:30
Venue: Gianyar Room & Klungkung Room
Arrival Registration, Keynote Speech, and Conference Presentation</td></tr>
<tr><td colspan="2">Morning Conference</td></tr>
<tr><td colspan="2">Venue: Gianyar Room
Opening Remarks 8:50~8:55
(Prof. Tjokorda Gde Tirta Nindhia, Engineering Faculty, Udayana University, Bali, Indonesia)
Keynote Speech I 8:55~9:30
Topic: “Sustainable Use and Zero Waste for Water Resources”
(Prof. Orawan Siriratpiriya, Environmental Research Institute of Chulaongkorn University, Thailand)
Keynote Speech II 9:30~10:05
Topic: “Indonesian Wild Silkworm Cocoon as Biomaterial”
(Prof. Tjokorda Gde Tirta Nindhia, Engineering Faculty, Udayana University, Bali, Indonesia)
Coffee Break & Photo Taking 10:05~10:40
Keynote Speech III 10:40~11:15
Topic: “Dietary Methylselenocysteine Prevents Mammary Carcinogenesis by Recoupling the Expression DNA Damage and Response Genes to the Circadian Clock”
(Prof. Helmut Zarbl, Rutgers, The State University of New Jersey, USA)
Keynote Speech IV 11:15~11:50
Topic: “In Situ Arsenic Removal in Groundwater for Rural Communities by Iron Sorption and Arsenic Immobilization”
(Prof. Solomon W. Leung, Environmental Engineering Civil and Environmental Engineering Department, Idaho State University)</td></tr>
<tr><td>Lunch 12:00~13:00</td><td>Venue: The Coffee Shop</td></tr>
<tr><td colspan="2">Afternoon Conferences</td></tr>
<tr><td>Session 1: 13:00~15:00
Venue: Gianyar Room
8 presentations-Topic: “Food Science & Biochemistry”</td><td>Session 2: 13:00~15:00
Venue: Klungkung Room
8 presentations-Topic: “Biomedicine”</td></tr>
<tr><td colspan="2">Coffee Break 15:00~15:30</td></tr>
<tr><td>Session 3: 15:30~17:30
Venue: Gianyar Room
8 presentations-Topic: “Environment”</td><td>Session 4: 15:30~17:30
Venue: Klungkung Room
8 presentations-Topic: “Bioinformatics & Medical”</td></tr>
<tr><td>Dinner 17:40</td><td>Venue: The Coffee Shop</td></tr>
<tr><td>Day 3</td><td colspan="2">June 27, 2016 (Monday) 9:00~17:00
One Day Visit & Tour</td></tr>
</table>

Tips: Please arrive at the conference room 10 minutes before the session begins to upload PPT into the laptop.

會議過程

Day 1 (2016.06.25)

第一天早上我們到 Bali State Polytechnic 參觀校區剛好遇上他們一年一度的慶典，雖然語言上沒辦法理解他們的意思，但是還是感覺得出一些一個莊重的民俗慶典氣氛，但他們跳完舞後，還有一些類似相聲的片段，逗得大家一直笑。慶典結束後享用他們學校的午餐，並且下午帶我們去看巴里島的一個新開發的海灘。

Day 2 (2016.06.26)

早上的行程由四位教授演講，他們分別來自泰國、印尼、美國、美國，從他們的身上學到了很多，它們很詳細的說明他們的研究，下面的這張圖是最後承辦人員提議大家來一張大合照，在場的氣氛感覺很棒，大家都為了吸收新的知識而認真的聽，而在休閒聊天的時候大家也都放得開盡情的交流，這是在台灣比較少見的情況，而在發問的時候大家也都踴躍提出自己的意見跟疑問。

下圖是中午的午餐，口味跟台灣部會差到很多，很美味在吃飯的途中也會聊聊剛剛的演講內容。

接著下午有 4 個 session 其中 session1 與 3 在同一個場地，2 跟 4 則是另外一個場地，我的報告是在 session4 在報告前先聽了前面不少人的報告，每個都很專業，讓我覺得我的程度還遠遠不夠，

Day 3 (2016.06.27)

第三天跟著研討會的參訪行程走，參觀了峇厘島的一些他們較為先進的醫療設備，並且參觀了巴里島的一些特色景點，其中有很多機會與其他研究員交流的機會，也交流了很多學術上的知識，也享受了印尼當地的一些特色，晚餐也在海邊的 一間餐廳用餐，最後去了一個大賣場購物買了些當地的名產。

三、心得及建議事項

我們去峇厘島時光是搭飛機加上轉機加上誤點，就花了 10 小時甚是煎熬，加上我在搭機時又會耳鳴並且疼痛，實在痛苦。但是一到了峇里島這些東西都拋到腦後了，在上一次的國際研討會去過日本，但那裏的生活機能等等都是與台灣類似，但這次第二次出國的地點峇厘島卻是跟我們台灣不太一樣，光是建築風格就跟我們大大不同，還好我們有兩個當地的朋友跟我們一起，在各方面上都省了很多麻煩。

我們報到後發現報告地點是一個很棒的飯店，讓我不禁想馬上進去裡面逛逛，之後承辦人員又剛好是跟我們語言共通的大陸四川人，真的很巧，這讓我們在日後形成等等在溝通上都省了不少麻煩，真的是太巧了。

第二天早上就開始聽 4 位教授的報告，它們在講述的時候使用英文完全不緊張，又快又流利，即便不是他們的母語一樣如此，這讓我一個光站上台就會非常緊張的人覺得非常的厲害。下午時在一個一個投稿者分享完他們的研究中，越接近我上台的時刻我越緊張，且台下的都是年紀比我們長不少的長輩，大家都很認真聽讓我壓力非常大，加上我英文又不好，當他們每個在盯著我認真想聽出我想表達什麼的時候，整個非常緊張。但經過了這次的經驗也讓我有很多的收穫，不管是在研究上還是個人上台經驗等等，都給我很多的幫助。

最後由衷的感謝學校能在出國旅遊費上補助我們，讓我們減輕很多經濟上的負擔，可以讓我們更專心在學習。

附錄

研討會行程

One Day Visit & Tour

June 27, 2016 (Monday) 9:00-17:00

(Tip: We will depart on time, please arrive at the Lobby before 9 a.m.)

1. Visit Turtle conservation at Serangan Island 09:00 - 11:00

The Turtle Conservation and Education Center (TCEC) opened by the governor of Bali, Mr Dewa Barata (20 January 2006) on Serangan island of Bali. TCEC is developed as part of the comprehensive strategy to eradicate illegal turtle trading on the island. Established on a land of 2.4 ha, the TCEC is trying to support the community of Serangan to find the alternatives beside illegal turtle business. The centre harnesses the potential of education, tourism, conservation and research, with a liberal sprinkling of business, to give endangered turtles one more chance on Serangan.

The four fundamental aspects to the centre include putting a definitive end to turtle trade, by encouraging the public not to consume turtle products (religious use or otherwise), and to generally support turtle conservation; providing turtles for rituals - without their killing - and monitoring turtle size and numbers, so that their use can be strictly controlled and regulated; offering employment opportunities for locals from Serangan; and finally, acting as a watchdog for turtle trade - in Serangan in particular and Bali in general.

2. Visit Udayana University (University hospital, Institute of peace and Democracy

(Photo session in front of Rectorat Building) 11:00-12:00

In the beginning of the 1960s, the people of Bali aspired to have a Tertiary Institution on the island. In order to realize this aspiration, on May 12th 1961, several figures from the educational sector, government, and community leaders conducted a conference led by Prof.Dr. Purbatjaraka,and assisted by Prof. Dr. Ida Bagus Mantra as secretary.

The conference discussed the steps required for the preparation of the establishment of a tertiary institution in Bali. An agreement was also reached for the formation of a committee led by dr. Anak Agung Made Djelantik, Head of the Board of Health in Bali, with a team of eight members.

Subsequently, the committee formed an institution named the Tertiary Education Institution of Bali, chaired by Ir. Ida Bagus Oka (Coordinator of Public Works Boards in the Southeast Islands Region); vice chaired by Dr. I Gusti Ngurah Gede Ngurah, assisted by two secretaries, Prof. Dr. Ida Bagus Mantra, and Drh. G.D. Teken Temadja. This institution succeeded in forming the Preparatory Committee for the establishment of Udayana University Bali on January 15th, 1962.

By a decision of the Directorate General of Higher Education, Ministry of Education and Culture of Indonesia, Udayana University (UNUD) was officially founded in August 17, 1962. Initially Unud consisted of four

faculties: Letters, Medicine, Veterinary Sciences and Animal Husbandry and Education and Teacher Training. The Faculty of Letters was actually established on 29th September 1958, however, the time it was a subsidiary of the Faculty of Letters of Airlangga University in Surabaya (East Java). This Faculty was thenintegrated into Udayana University in 1962. Although it was founded on August 17, the anniversary date of Udayana University is not August 17, but was choosen to be on September 29 to commemorate the date of establishment of the Faculty of Letters in 1958. Unud has develop rapidly, in 2015 the university has 13 faculties, 25 master programs and 10 doctoral programs.

Udayana University today's is listed as one of the 50"Promising Universities of Indonesia" published by the Ministry of Education of Republic Indonesia, out of nearly 2.500 higher education institutions around the country. The university has a strong position as one of the leading university particularly in the Eastern Indonesian Territory.

3. Lunch at Garuda Wisnu Kencana

Mandala Garuda Wisnu Kencana, or **Garuda Wisnu Kencana** (GWK), is a cultural park covering approximation 60 ha area located in Ungasan, Badung Regency, or about 10–15 minutes driving from Bali Ngurah Rai International Airport. It is devoted to the Hindu God Vishnu, and his mount, Garuda, the mythical bird who become his companion.

Currently, the statue of Vishnu is 23 metres (75.5 ft) high, although the original plan was for a 120-metre (390 ft) gold-plated Vishnu riding Garuda on top of an 11-storey entertainment complex. Garuda wing span will be 64 metres (210.0 ft) across. The idea was not without controversy, and religious authorities on the island complained that its massive size might disrupt the spiritual balance of the island, and that its commercial nature was inappropriate, but some groups agree with the project, because it will make new tourist attraction over barren land.

In 2013 Nyoman Nuarta and PT Alam Sutera Realty Tbk (IDX:ASRI) joined to build villas and apartments in the GWK area in exchange for Rp150 billion ($14.4 million). Nuarta plans to spend Rp20 billion to make another bust and to move the existing bust to another site 300 meters from the original site. It plans to spend additional Rp29 billion to make the new statue of stainless steel instead of galvanized steel as proposed previous design.

4. Tour to Uluwatu Temple

Uluwatu Temple (Indonesian: ***Pura (Luhur) Uluwatu***) is a Balinese sea temple (*pura segara*) in Uluwatu (Kuta South, Badung). The temple is regarded as one of the *sad kahyangan* and is dedicated to Sang Hyang Widhi Wasa in his manifestation as Rudra.

The temple (*pura* in Balinese) is built at the edge (*luh*) of a 70 meter high cliff or rock (*watu*) projecting into the sea. In folklore, this rock is said to be part of Dewi Danu's petrified barque.

Though a small temple was claimed to have existed earlier, the structure was significantly expanded by a Javanese sage, Empu Kuturan in the 11th Century. Another sage from East Java, Dang Hyang Nirartha is credited for constructing the padmasana shrines and it is said that he attained moksha here, an event called *ngeluhur* ("to go up") locally. This has resulted in the temple's epithet *Luhur*.

5. Dinner (farewell party) at Muaya Beach Jimbaran

研討會論文簡報

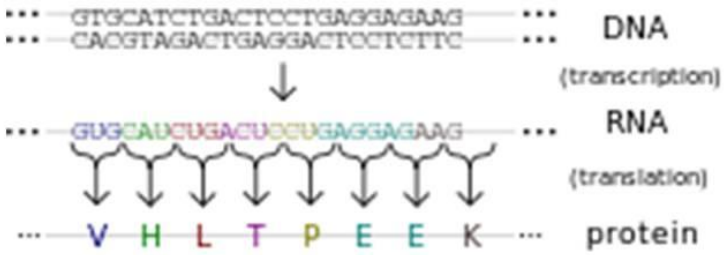
ROOTED TREE OPTIMIZATION ALGORITHM FOR PROTEIN FOLDING PREDICTION

lecturer ： Tu

DIRECTORY

PROTEIN INTRODUCTION

- DNA has four nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T).
- Each amino acid is encoded by three nucleotides.

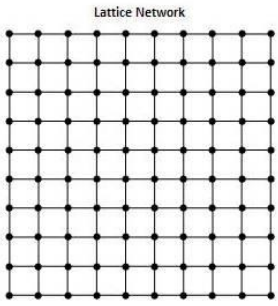


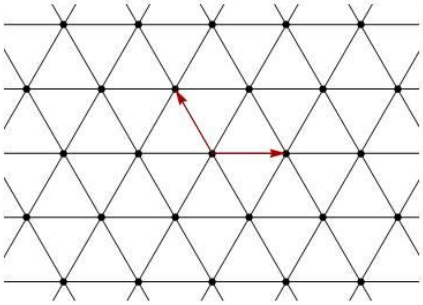
PROTEIN STRUCTURE PREDICTION

- The primary structure (amino acid sequence) correct fold to the tertiary structure can determine protein function.
- Protein tertiary structure has hydrophobic core to stabilize the structure usually.
- must explore the space of possible protein structures which is astronomically large.

HP MODEL

- This model cluster all amino acid to two groups.
- The folding space simplify to lattice





RTO ALGORITHM

- The algorithm simulate the root of tree finding the wet soil

- one part root grow from the best root (wettest place). (Rn)
- one part root affected grow according to previous best root. (Rc)
- one part will find the water randomly. (Rr)

SEARCH METHOD

- RTO having three ways to search

- The update formulas are not suitable about to solving the current problem

THE RATE TO NEAREST ROOT OF POPULATION (RN)

- Rn is a proportion parameter setting for all roots
- update from best root

$$X_{new(i,j)} = X_{best(i,j)}$$

$$X_{new(i,r_1)} = r_2 \in [\min, \max]$$

THE RATE TO CONTINUOUS ORIENTATION FOR ROOT OF POPULATION (RC)

- Rc is a proportion parameter setting for all roots
- The roots find water to grow to previous best root

$$X_{new(i,j)} = X_{old(i,j)} + c_2 \times (X_{best(i,j)} - X_{old(i,j)})$$

THE RATE TO RANDOM ROOT OF POPULATION (RR)

- Rr is a proportion parameter setting for all roots
- The update root random selection from previous root

$$X_{temporary(change)} = X_{temporary(original)} = X_{old(r)}$$

$$X_{temporary(change)(i,r_1)} = r_2 \in [\min, \max]$$

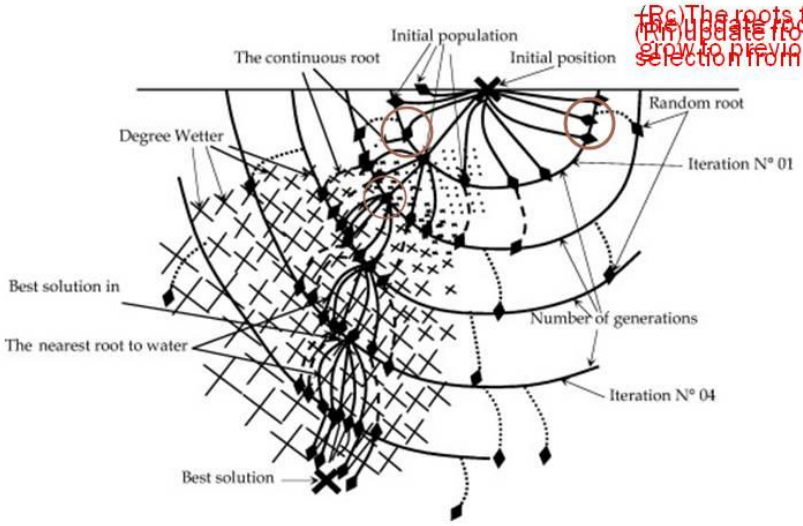
$$X_{new} = \begin{cases} Xtemporary(change), & \text{if this root better} \\ Xtemporary(original), & \text{else} \end{cases}$$

UPDATE TO ROOT

If the parameter setting Rn=0.3, Rc=0.2, Rr=0.5

Wetness (good fitness)	$root_1$	Use the chapter B.1 formula (1-2) to update
	$root_2$	Use the chapter B.1 formula (1-2) to update
	$root_3$	Use the chapter B.1 formula (1-2) to update
	$root_4$	Use the chapter B.2 formula (3) to update
	$root_5$	Use the chapter B.2 formula (3) to update
dry (bad fitness)	$root_6$	Use the chapter B.3 formula (4-6) to update
	$root_7$	Use the chapter B.3 formula (4-6) to update
	$root_8$	Use the chapter B.3 formula (4-6) to update
	$root_9$	Use the chapter B.3 formula (4-6) to update
	$root_{10}$	Use the chapter B.3 formula (4-6) to update

ROOTS OF PLANT BEHAVIOR WHEN THEY LOOK FOR WATER (SOLUTION)



※ This picture is get from RTO original paper.

PSEUDO CODE FOR RTO ALGORITHM

```
RTO pseudo code
Start
   // initialization;
   set a number to max for iteration and set current iteration equal to 0.
   set  number of root (population)
   set parameters Rn Rc Rr
   each root get random from search range.

while(current iteration<max iteration)
   Calculate to each root (fitness or witness)
   Reorder the root according to fitness
   determine the best root

            For(i to population size){
if(i/ population size < Rn) use the Rn formula to update root content (B.1)
if(i/ population size < Rn+Rc) use the Rc formula to update root content (B.2)
if(i/ population size > Rn+Rc) use the Rr formula to update root content (B.3)
                                    }
            End for
Current iteration+1
End while

END
```

DIFFERENCE PARAMETER COMPARISON

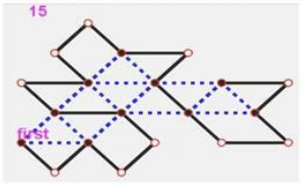
Parameter setting			Sequence				
Rn	Rc	Rr	1*1	2*1	3*1	4*1	5*1
0.2	0.3	0.5	15/14.9	17/15.6	12/11.9	24/23.2	40/39.0
0.5	0.3	0.2	15/14.93	16//14.6	12/12	24/22.8	38/36.6
0.3	0.5	0.2	15/14.93	17/14.5	12/11.9	24/22.7	39/36.6
1	0	0	15/14.7	16/14.1	12/11.7	2321.93	37/34.86
0	1	0	9/8	8/6	7/4.7	13/10.9	24/19.1
0	0	1	15/14.1	16/13.5	12/11.1	23/21.83	39/36.0

COMPARE BEST PREDICTION WITH OTHER ALGORITHM

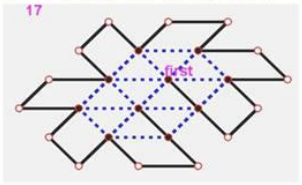
Sequence*1	length	Best	algorithm					
			SGA	HGA	TS	ERS-GA	HHGA	RTO
2(HP)PH2(HP)2(PH)HP2(PH)	20	**-15**	-11	**-15**	**-15**	**-15**	**-15**	**-15**
2H2P6(H2P)2H	24	**-17**	-10	-13	**-17**	-13	**-17**	**-17**
2PH2P3(2H4P)2H	25	**-12**	-10	-10	**-12**	**-12**	**-12**	**-12**
P2(2P2H)5P5H2(2H2P)2PH2(H2P)	36	**-24**	-16	-19	**-24**	-20	-23	**-24**
2PH2(P3H)5P10H6P2(2H2P)H2P5H	48	**-43**	-26	-32	-40	-32	**-41**	-40
2H3(PH)P4HPH2(3PH)4P2(H3P)HP4H3(PH)P2H	50	**-41**	-21	-23	nd	-30	-38	**-40**
P2(P3H)5H3P10HPH3P12H4P6HP2HPHP	60	**-70**	-40	-46	**-70**	-55	-66	-69
12H2(PH)3(2(2P2H)2PH)2(PH)11H	64	**-75**	-33	-46	-50	-47	-63	**-66**

*1 sequence content 2H=HH, 2(HP)=HPHP

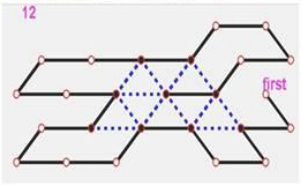
VISUALIZATION OF THE PREDICTION RESULT



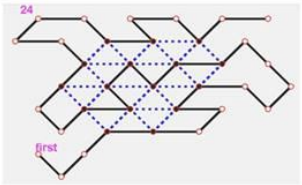
Fitness=-15



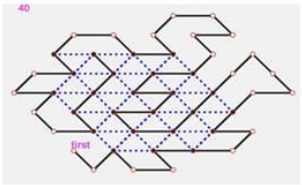
Fitness=-17



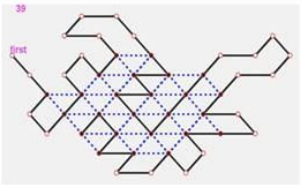
Fitness=-12



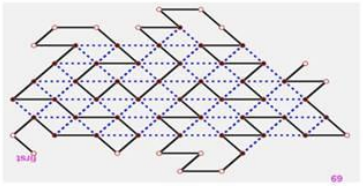
Fitness=-24



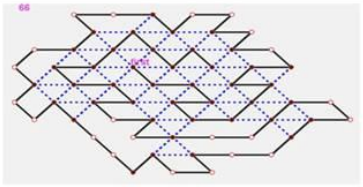
Fitness=-40



Fitness=-39



Fitness=-69



Fitness=-66

Thanks for listening

論文接受原稿

Rooted Tree Optimization Algorithm for Protein Folding Prediction

Cheng-Hong Yang, Yu-Shiun Lin
Department of Electronic Engineering, National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan
chyang@cc.kuas.edu.tw, joe29681195@yahoo.com

Li-Yeh Chuang
Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering, I-Shou University, Kaohsiung, Taiwan
chuang@isu.edu.tw

Hsueh-Wei Chang
Department of Biomedical Science and Environmental Biology, Kaohsiung Medical University, Kaohsiung, Taiwan
changhw@kmu.edu.tw

***Abstract*—Protein function depends on structural folding from the amino acid sequence. Correctly predicting the amino acid sequence is thus helpful for evaluating the protein structure and function. In 1995, a hydrophobic-polar model (HP model) was proposed to simplify the folding process. This model drastically simplifies the real folding space into a lattice and combines an optimization algorithm to predict the protein structure. Many optimization algorithms have been implemented with the HP model for protein structure prediction, but accuracy and speed still need to be improved. This study proposes a fairly new algorithm, namely the rooted tree optimization (RTO), to improve on current algorithm performance. RTO provide three ways to find optimal solution and with the HP model for protein structure prediction. The local search is designed to add to each iteration of RTO to improve performance of the triangular lattice model.**

***Index Terms*—Hydrophobic-polar model, rooted tree optimization, protein structure prediction**

I. INTRODUCTION

The hereditary information of all living organisms is stored in their DNA, which is made up of four nucleotides: adenine (A), cytosine (C), guanine (G), and thymine (T). Each amino acid is encoded by three nucleotides, which constitute one codon. A total of 64 codon combinations encode the amino acids and 20 amino acids have currently been identified [1, 2]. The structure of a protein can be divided into four levels. The primary structure is the amino acid sequence, while the secondary structure is the folded amino acid sequence. The tertiary structure is an individual protein formed using several folded secondary structures, which have independent activities and functions. An individual protein may be polymerized with other proteins, resulting in a polymer that forms a quaternary structure with more complex functions.

The primary structure (amino acid sequence) correctly folds to a tertiary structure that can determine protein function and control basic functions. Predicting protein structures has emerged as an important research issue, but such predictions require the investigation of many possible protein folding structures, making it very time and cost-intensive [3]. This paper proposes rooted tree optimization (RTO) algorithm with HP model to improve forecasting protein structure efficiency.

In 1985, Dill proposed the hydrophobic-polar protein folding model (HP model) [4], providing significant improvements for the prediction accuracy of protein structures, thus reducing study time and cost. Proteins usually have a hydrophobic core to stabilize the structure. As shown in Fig. 1, the HP model uses this to evaluate fitness, i.e., the interaction is taken -1 because the minimum free energy approaches its native state [5]. However, the whole simulation is processed using 2D or 3D lattices, and does not reinsert different amino acids at the same lattice point [6, 7], which is a self-avoiding walk.

The HP model for protein structure prediction is defined as nondeterministic polynomial-time-hard (NP-hard) [8, 9] problem which poses problems for machine-based solutions, because the model must account for all amino acid folding possibilities, especially in long sequences. Many algorithms perform well in the HP model, but are not suitable for use with long amino acid sequences.

The rooted tree optimization algorithm (RTO) for economic dispatch with valve-point effect performs well [10], so RTO is applied to protein structure prediction (PSP) problems because this study requires an optimal solution search function. However, the update formula to root cannot be directly applied in PSP.

Current→ Neighbor ↓	Hydrophobic	Polar
Hydrophobic	-1	0
Polar	0	0

Figure. 1. Update to root explanation diagram.

In the current study, we improve the RTO algorithm formula to update R_c, R_n and R_r for protein structure prediction problem. The new approach with the HP model performs well, but determining optimal parameter settings is still an issue. Therefore, the difference parameterset to R_c, R_n and R_r showy to result and discussion.

An algorithm using the HP model for a PSP problem is a folding direction simulation, e.g., encode number for 1 that fold the next amino acid to the left of the current amino acid so that the last amino acid does not required encoding, thus the entire amino acid sequence is folded to calculate the fitness. However, except the last amino acid are having encoding, i.e., folded structure has been defined while can see the Figure 2 to understanding.

II. METHOD

A. HP model

This model clusters all amino acids into two groups (hydrophobic and polar), and the folding space is simplified to a lattice. Different amino acids cannot be included on a single point, thus the amino acid sequence cannot be interrupted in prediction. The HP model generally uses square lattices and triangular lattices for prediction. To test the searching ability of the RTO algorithm, we use the triangular lattice because it must explore more possible folding configurations than the square lattice.

B. RTO algorithm

The rooted tree optimization (RTO) algorithm [10] is a fairly new algorithm proposed by Labbi in 2016. It simulates the root of tree searching for wet soil, in which one part of total root grows from the best root (i.e., the wettest place), one part of total root affects growth according to previous best root, and the surplus

part will randomly select one root to search for water randomly. The algorithm uses three functions to update the root, how to select the updated way to be determined by R_n, R_c and R_r before selection way needed to reorder all root. The original updated formulas are not suitable to solve the current problem (PSP), and were thus modified as follows.

```
RTO pseudo code
Start
   // initialization;
   set a number to max for iteration and set current iteration equal to 0.
   set  number of root (population)
   set parameters Rn Rc Rr
   each root get random from search range.

while(current iteration<max iteration)
   Calculate to each root (fitness or witness)
   Reorder the root according to fitness
   determine the best root

            For(i to population size){
if(i/ population size < Rn) use the Rn formula to update root content (B.1)
if(i/ population size < Rn+Rc) use the Rc formula to update root content (B.2)
if(i/ population size > Rn+Rc) use the Rr formula to update root content (B.3)
                                    }
            End for
Current iteration+1
End while

END
```

Figure. 2. RTO algorithm pseudo code

The RTO algorithm pseudo code shows explains the prediction process. First, set some parameter value and initialize each root. After, the wetness (fitness) calculation, sort the roots, best root determine and update each root to form a loop which updates the root according to parameter setting and formula from sections B.1, B.2 and B.3. This loop is executed continually until it achieves the maximum number of iterations.

B.1 The rate to nearest root of population (Rn)

R_n is a proportion parameter setting for all roots; 20 percent of all roots are updated from the best root if the R_n setting by 0.2. The best folding simulation is chosen to update the root to change the folding direction for one random amino acid. The change function follows as:

$$X_{new(i)} = X_{best} \tag{1}$$

$$X_{new(i,r_1)} = r_2 \in [\min, \max] \tag{2}$$

where X_{new} is next iteration's root, X_{best} is the previous iteration's best solution, *i* is current root of the population, r_1 is random for dimension to root of the population, and r_2 is a random number between min and max (i.e., minimum and maximum searching). Equation (1) copies the previous X_{best} to X_{new} while choosing a random dimension, then (2) changes the information between minimum and maximum search value.

B.2 Rate to continuous orientation for root of population (R_c)

R_c is a proportional parameter setting for all roots, i.e., 30 percent of all roots update if R_c is 0.3. The roots find water to grow to previous best root that function according the formula for HP model.

$$X_{new(i,j)} = X_{old(i,j)} + c_2 \times (X_{best(j)} - X_{old(i,j)}) \tag{3}$$

where *j* is the dimension of the root, X_{old} is the previous root, and c_2 is a parameter set between 0 to 1. Equation (3) like updated formula of particle swarm optimization algorithm [11].

B.3 Rate of random root of population (R_r)

R_r is a proportional parameter setting for all roots, i.e., 50 percent of all roots update if R_c is 0.5. The update root randomly selects multiple points (amino acids) from the previous root for mutating to a random direction. If the new root is wetter, it will replace the old root. The update formula is as follows:

$$X_{temporary(change)} = X_{temporary(original)} = X_{old(r)} \tag{4}$$

$$X_{temporary(change)(i,j)} = r_2 \in [\min, \max] \tag{5}$$

$$X_{new} = \begin{cases} Xtemporary(change), & \text{if this root better} \\ Xtemporary(original), & \text{else} \end{cases} \tag{6}$$

where $X_{temporary}$ is a temporary root, change is mutation from original, r_1 is a random value for the dimension to the root of the population, and r_2 is a random number between the minimum and maximum search. Equation (4) copies a random root from the previous population to $X_{temporary(change)}$ and $X_{temporary(original)}$. Equation (5) gives every dimension of $X_{temporary(original)}$ a mutation probability. Equation (6) selects the best temporary root for X_{new}.

C. Update to root

Figure 3 shows the root updating procedure. Every root is reordered from highest to lowest according to its wetness (fitness), and then updating methods are chosen proportionally. For example, if R_n, R_c, R_r are respectively 0.3, 0.2 and 0.5, then 30%, 20% and 50% of all roots will be respectively updated using R_n, R_c, and R_r. ,the three way introduction and formula reference the part B of methods section.

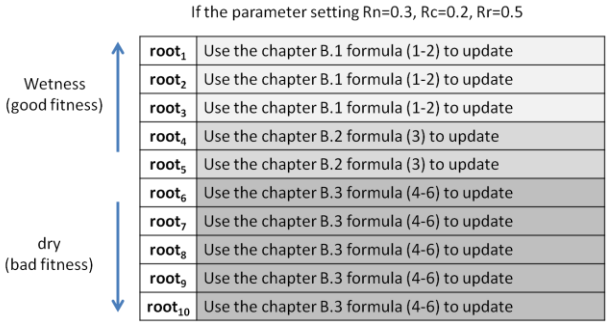


Figure 3. Root updating procedure

D. *Calculate fitness (wetness)*

In the HP model, fitness is the sum of hydrophobic-hydrophobic interaction that more assumes protein structure to native state, the number excludes adjacent hydrophobic protein sequences. In Fig. 1, the dashed lines indicate interaction.

III. RESULT

A. *Parameter setting*

TABLE I. PARAMETER SETTING TABLE

Parameter	value
C_2	0.2
Population	100
max iteration	100
search minimum	0
search maximum	5
R_n, R_c, R_r	see the next section

C_2 likes learning factor of PSO, so it is set at 0.2. Population and max iteration are setting at 100 while often set in a new test. The triangular lattice has six folding possibilities, so the search range is set between 0 and 5, respectively represent folding to the left, upper left, upper right, right, lower right and lower left. R_n, R_c, R_r are complex to setting, so a experiment shows difference result as follows.

B. *Different parameter comparison*

This section discusses different settings of Rn, Rc, and Rr because using different proportions to update the root will significantly affect the search ability. Overusing an update model for prediction loses the combined search effect. The sequence 1 to 5 references to Table III which as our test object because predicted poor on short sequence that less likely to have good results in the long sequence. Therefore, we set split points at lengths of 50. Lengths equal to or greater than 50 is defined as a long sequence, while those smaller than 50 are defined as short sequences. To change difference parameter to prediction can select the best setting to predict long sequences.

TABLE II. COMPARISON OF DIFFERENT R_N, R_C, R_R SETTINGS

Experiment	R_n*2	R_c*2	R_r*2	S-1*1	S-2*1	S-3*1	S-4*1	S-5*1
1(select)	0.2	0.3	0.5	**-15**	**-17**	**-12**	**-24**	**-40**
2	0.5	0.3	0.2	**-15**	-16	**-12**	**-24**	-38
3	0.3	0.5	0.2	**-15**	**-17**	**-12**	**-24**	-39
4	1	0	0	**-15**	-16	**-12**	-23	-37
5	0	1	0	-9	-8	-7	-13	-24
6	0	0	1	**-15**	-16	**-12**	-23	-38
7	0.5	0	0.5	**-15**	-16	**-12**	-23	-39

*1 Sequence refers to Table III.
*2 R_n, Rr, R_r are taken from chapter method.
Bold indicates best results.
(Select) is used in the next chapter parameter.

Experiments 4-6 show only a way to update which is not a good decision, especially when only using R_c gets the poor prediction results, but it is not a good idea to completely discards this way. Though discarded the R_c way in the experiment 7 have nice but compared to opposite not better. After determining the three formulas has its own meaning of existence, we adjust the ratio to approximately show as experiments 1 to 3, it was observed that excessive allocate setting to R_n is not a good choice, then half of population use R_c way can prediction well.

Sequences 1 and 3 predict a simpler folding structure. With the exception of using R_n only (experiment 5), all other settings found the best structure. Sequence 5 only a setting (experiment 1) predicts the best structure, and next section uses this setting as the basis for comparisons with other algorithms.

C. Prediction performance comparisons with other algorithms

We validate the performance of the proposed algorithm for protein structure prediction through experiments. RTO predict base on eight benchmark of Table III for experiments. Using the square HP model, several algorithms are able to predict the optimal structure, but performance is lower using the triangular HP model. Therefore, we apply the algorithms to eight sequences to simulate folding. Prediction results are compared with those of other algorithms based on the triangular lattice.

Table III compares the prediction results of SAG, HGA and TS. The first is shortest amino acid sequence among the eight tests, "-" indicates the algorithm produces no prediction for the current sequence, and the HGA algorithm predicts the best structure. In other sequences, the prediction performance of HGA and SGA is slightly worse, and RTO produces the same prediction solution as TS. TS outperforms RTO on sequence 7, but RTO performs better on the longest sequence (-66). Overall, RTO has good predictive ability regardless for both long and short sequences. TS generally produces good prediction results, but its prediction for sequence 8 includes some small defects.

Short sequences are easier to predict than long sequences because short sequence significantly reduces predict the structure likelihood relatively long sequence. Therefore, R_n produces a more detailed search of similar structures for the best solution of short sequences with good results. For long sequences, random searches using R_r produces good structure predictions for avoid falling into the area the best solution. R_c can close bad structure to better structure to prevent idle at bad prediction.

While the overall results are as we expected forecast, but still some sequence structure prediction are not the best solution. However, how to escape from the approximate optimum solution to finding a real best solution focuses on considerations in optimization algorithms always. The R_r only slightly improve the situation. After the study R_r will be focused to improve which maybe can reference modified hill climbing. Then the algorithm stability will be considered.

TABLE III. COMPARE THE PREDICTION SOLUTION WITH OTHER ALGORITHM

S (length) *3	Sequence*1 [12]	Best*2 [12]	SGA [13]	HGA [13]	TS [14]	RTO
1 (20)	2(HP)PH2(HP)2(PH)HP2(PH)	-15	-11	**-15**	**-15**	**-15**
2 (24)	2H2P6(H2P)2H	-17	-10	-13	**-17**	**-17**
3 (25)	2PH2P3(2H4P)2H	-12	-10	-10	**-12**	**-12**
4 (36)	P2(2P2H)5P5H2(2H2P)2PH2(H2P)	-24	-16	-19	**-24**	**-24**
5 (48)	2PH2(P3H)5P10H6P2(2H2P)H2P5H	-43	-26	-32	**-40**	**-40**
6 (50)	2H3(PH)P4HPH2(3PH)4P2(H3P)HP4H3(PH)P2H	-41	-21	-23	-	**-39**
7 (60)	P2(P3H)5H3P10HPH3P12H4P6HP2HPHP	-	-40	-46	**-70**	-69
8 (64)	12H2(PH)3(2(2P2H)2PH)2(PH)11H	-	-33	-46	-50	**-66**

*1 sequence content 2H=HH, 2(HP)=HPHP
*2 Best is the optimal prediction solution of sequence.
*3 sequence (length)
The bold to tag is best result in all experiment.
H is hydrophobic; P is polar.
Number is the number of H-H interactions.
- is no providing the optimal solution.

D. Prediction results visualization

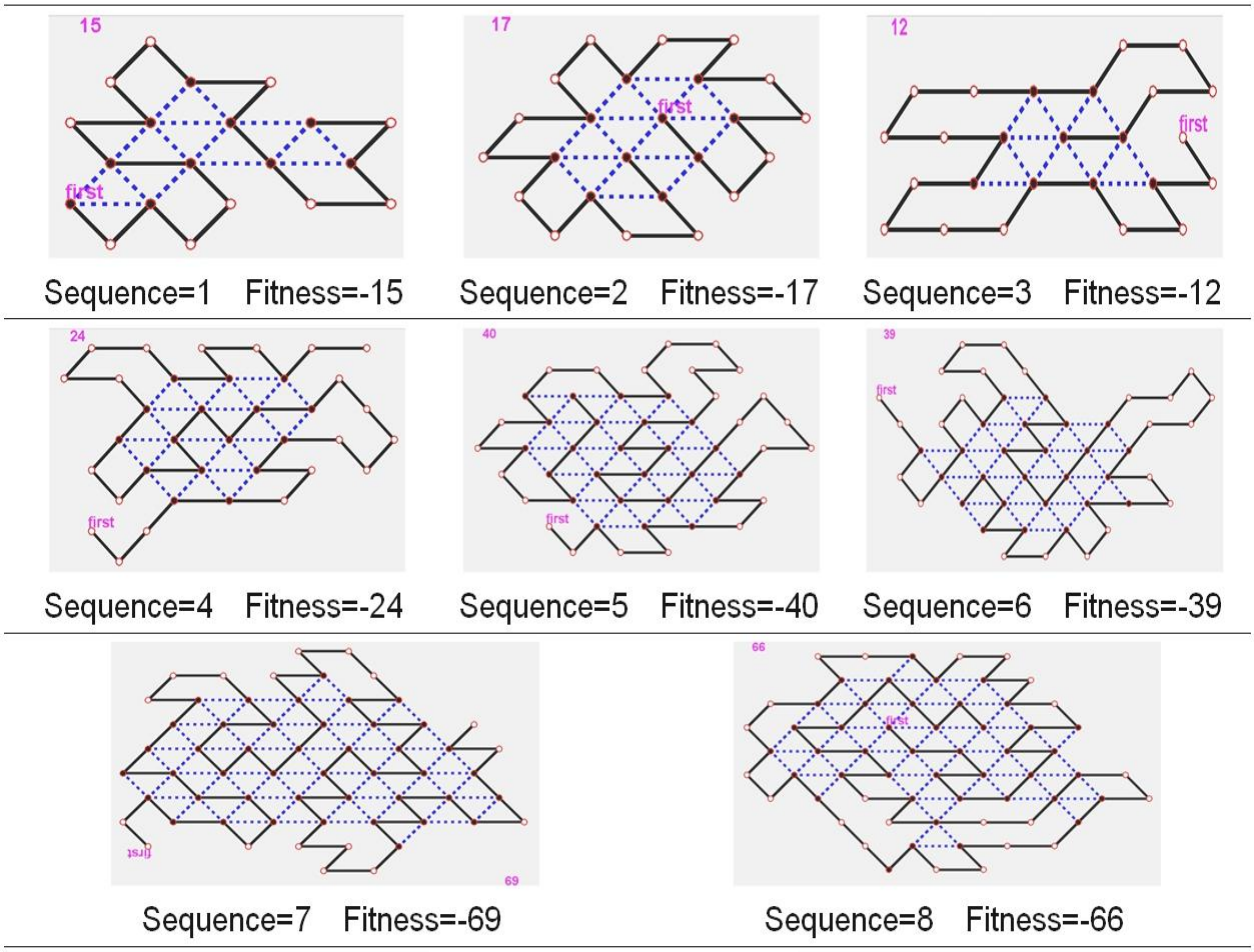


Figure. 4. Prediction visualization.

Figure 4 is RTO prediction result on the best visualization. In the figure, black point is hydrophobic, white point is polar, left top number is fitness in current prediction structure, dashed representative hydrophobic-hydrophobic interaction increase fitness each one. The solid line indicates amino acid connection of sequence, the first mark the sequence beginning (first amino acid).

Eight sequence predicted results visualization of figure as can be seen to the folding correctly or be close have many hydrophobic in center that like the natural protein folding.

IV. Discussion

In this study, we try a new idea to protein structure prediction with HP model, it achievement is well to long amino acid sequence in particular. In short sequence (1-4), TS and RTO can predict best structure that proves the prediction not just pure better than other algorithms, it is really can predict the correct structure on HP model.

The formula 3 of updated root like PSO updating velocity which closer general predictions to better result, the formula 1 and 2 detailed search similar structures of optimal solution, so a good prediction results obtained in the short sequence. Formula 4 to 6 random search to other structure which prevent Into the area optimal solution and repeated searching for similar structures.

Formula 1 to 6 are own creating according to experience and constantly test, these formula only can on the protein structure prediction with HP model, but this also shows that the current formula have flawed, we will consider more factors, other algorithm and formulas to improve these because the concept of three way to update population like mixing a local search in optimization algorithm.

As mentioned above, not only formula need to improve, the parameter setting is also a big problem and that may change follow to formula changing. The setting considerations also appear the RTO original paper because that setting direct impact on this algorithm searching ability.

V. Conclusion

The RTO algorithm exhibited high performance to the triangular lattice model. The algorithm effectively determined the folding mechanism for complicated and long amino acid sequences. Most of the algorithms in predicting long amino acid sequence structure are often not very good. We try to use a new algorithm and improve for ways to this problem. However, this attempt results can be seen good result from Table III, but if the

parameter and the formula have further improvement maybe perhaps more predictive power optimization.

ACKNOWLEDGMENT

This study was partly supported by the National Science Council of Taiwan for Grant NSC 103-2221-E-151-029-MY3.

References

[1] *F. Sanger and H. Tuppy, "The amino-acid sequence in the phenylalanyl chain of insulin. 1. The identification of lower peptides from partial hydrolysates," Biochemical Journal, vol. 49, p. 463, 1951.*

[2] *D. L. Nelson, A. L. Lehninger, and M. M. Cox, Lehninger principles of biochemistry: Macmillan, 2008.*

[3] *B. Qian, S. Raman, R. Das, P. Bradley, A. J. McCoy, R. J. Read, et al., "High-resolution structure prediction and the crystallographic phase problem," Nature, vol. 450, pp. 259-264, 2007.*

[4] *K. A. Dill, "Theory for the folding and stability of globular proteins," Biochemistry, vol. 24, pp. 1501-1509, 1985.*

[5] *C. B. Anfinsen, "Studies on the principles that govern the folding of protein chains," ed, 1972.*

[6] *B. P. Blackburne and J. D. Hirst, "Evolution of functional model proteins," The Journal of Chemical Physics, vol. 115, pp. 1935-1942, 2001.*

[7] *J. D. Hirst, "The evolutionary landscape of functional model proteins," Protein Engineering, vol. 12, pp. 721-726, 1999.*

[8] *P. Crescenzi, D. Goldman, C. Papadimitriou, A. Piccolboni, and M. Yannakakis, "On the complexity of protein folding," Journal of computational biology, vol. 5, pp. 423-465, 1998.*

[9] *B. Berger and T. Leighton, "Protein folding in the hydrophobic-hydrophilic (HP) model is NP-complete," Journal of Computational Biology, vol. 5, pp. 27-40, 1998.*

[10] *Y. Labbi, D. B. Attous, H. A. Gabbar, B. Mahdad, and A. Zidan, "A new rooted tree optimization algorithm for economic dispatch with valve-point effect," International Journal of Electrical Power & Energy Systems, vol. 79, pp. 298-311, 2016.*

[11] *J. Kennedy, "Particle swarm optimization," in Encyclopedia of machine learning, ed: Springer, 2011, pp. 760-766.*

[12] *M. K. Islam, M. Chetty, and M. Murshed, "Novel local improvement techniques in clustered memetic algorithm for protein structure prediction," in Evolutionary Computation (CEC), 2011 IEEE Congress on, 2011, pp. 1003-1011.*

[13] *M. T. Hoque, M. Chetty, and L. S. Dooley, "A hybrid genetic algorithm for 2D FCC hydrophobic-hydrophilic lattice model to predict protein folding," in AI 2006: Advances in Artificial Intelligence, ed: Springer, 2006, pp. 867-876.*

[14] *H.-J. Böckenhauer, A. Z. M. D. Ullah, L. Kapsokalivas, and K. Steinhöfel, "A local move set for protein folding in triangular lattice models," in Algorithms in Bioinformatics, ed: Springer, 2008, pp. 369-381.*

Cheng-Hong Yang

is a professor of the Department of Electronic Engineering at National Kaohsiung University of Applied Sciences, Tai-wan. He received his M.S. and Ph.D. degrees in computer engineering from North Dakota State University in 1988and 1992, respectively. His main areas of research are evolutionary computation, bioinformatics, and assistive tool implementation.

Yu-Shiun Lin

is a undergraduate student of the Department of Computer Science and Information Engineering at National Kaohsiung University of Applied Sciences, Kaohsiung, Taiwan. His research interests include Image recognition, algorithm design, bioinformatics, and data analysis.

Li-Yeh Chuang

is a professor and director of the Department of Chemical Engineering & Institute of Biotechnology and Chemical Engineering at IShou University, Kaohsiung, Taiwan. She received her M.S. degree from the Department of Chemistry at the University of North Carolina in 1989 and her Ph.D. degree from the Department of Biochemistry at North Dakota State University in 1994. Her main areas of research are Bioinformatics, Biochemistry and Genetic Engineering.

Hsueh-Wei Chang

He is a Professor and director in the Department of Biomedical Science and Environmental Biology at Kaohsiung Medical University. He received the MS degree from the Institute of Radiation Biology and the PhD degree from the Department of Life Sciences at Tsing Hua University, Hsinchu, Taiwan, in 1991 and 2000, respectively. His main areas of research are bioinformatics, genomics, cancer biomarker, and anticancer drug screening from natural products.