

混合型演算法應用於 DNA 序列編碼

楊正宏^{1,2}, 李奇翰², 莊麗月³

¹ 稻江科技暨管理學院網路系統學系

² 國立高雄應用科技大學電子工程系

³ 義守大學化學工程系

chyang@cc.kuas.edu.tw, 1094105349@cc.kuas.edu.tw, chuang@isu.edu.tw

摘要—DNA 計算依照 DNA 分子間化學反應進行計算，理想的 DNA 序列編碼不僅影響其計算效率，亦影響計算正確性，故設計穩定的 DNA 序列使 DNA 計算過程更具穩定且提高正確率，為目前 DNA 計算的研究重點。DNA 序列設計牽涉到一些相關聯的約束條件限制，由傳統方法設計出的 DNA 序列於計算時穩定性及正確性都略顯不足。故本研究提出以粒子族群最佳化結合侵入雜草演算法針對 DNA 序列做最佳化組合編碼，以便後續 DNA 計算更為穩定，進而提升實驗結果正確率。與其他文獻結果相比，本方法不僅編碼出更為穩定的 DNA 序列於 DNA 計算上，於尋找單極值類型之最佳化問題，亦較混合式演算法更具效率。

關鍵詞—DNA 計算、DNA 序列編碼、侵入雜草最佳化、粒子族群最佳化

一、前言

DNA 計算(DNA computing)是一種模擬生物 DNA 分子結構並借助現代分子生物學技術所開創的新計算概念，其利用 DNA、生物化學及分子生物學取代傳統以矽為基礎的電腦技術。DNA 計算藉由 Watson-Crick 互補鹼基對為特定生化反應提供有效幫助，如大規模並行計算及更大儲存容量[8][12]。Adleman 於 1994 年[1]提出利用 DNA 分子計算漢彌爾頓七個節點之有向路徑，並成功在 DNA 溶液試管進行實驗，此種計算方式為 NP-complete 問題帶來一種新的解決途徑。目前已有許多研究應用 DNA 計算來解決此類型問題，如 SAT 問題 (satisfaction problem) [3]、TSP 問題

(traveling salesman problem)[21]及最大集合問題 (maximal clique problem) [15]等。DNA 計算主要受到 DNA 序列間之雜交特異性反應影響，但亦可能因 DNA 分子間化學特性，而產生錯誤不良反應 [4]，錯誤的 DNA 計算可能會產生偽正類(false positive, FP)及偽負類(false negative, FN)兩種錯誤結果。良好的 DNA 編碼方式能降低此情形發生 [6]。故如何編碼良好 DNA 序列提升 DNA 計算的效率及結果正確性便成為目前 DNA 計算研究的重要課題。

DNA 計算基於 DNA 分子間生化反應來作為計算方式，因此當 DNA 序列產生雜交反應時，相當容易受生化反應各種因素影響(如反應物濃度、溫度及溶液 PH 值等)，且 DNA 計算之序列皆由兩類約束原則限制：組合約束及熱力學約束。故如何設計理想的 DNA 編碼成為 DNA 計算重要的研究課題之一。DNA 編碼研究目的在於當每一個 DNA 分子在實際生化反應過程中，皆對雜交反應有正向效果並降低錯誤雜交發生，以提高 DNA 計算效率及可靠性。實際 DNA 計算中，DNA 編碼問題主要降低相似距離之 DNA 序列出現機率，目前已有文獻針對此問題提出許多方法進行 DNA 編碼設計，如漢明距離(Hamming distance)及自由能(free energy, ΔG) 之動態規劃演算法[9][13]、模板映射[7]、“SCAN”程序[10]及模擬退火法(Simulated Annealing) [18]等。由於 DNA 編碼問題屬於多條件的最佳化問題，因此應用最佳化問題之啟發式演算法於尋找 DNA 最佳化編碼，亦有

相當不錯的表現，如基因演算法 (Genetic Algorithm, GA)[2]，粒子族群最佳化 (Particle Swarm Optimization, PSO)[5]。本文提出於相異評估條件下，以PSO演算法結合雜草侵入最佳化 (Invasive weed optimization, IWO)設計數條最佳化DNA序列，藉由改良PSO達到改善DNA編碼之目的，且與文獻GA[2]、GA/PSO[20]、傳統進化式演算法 (Conventional evolutionary algorithm, CEA)[17]之結果相比較，實驗證明本研究能有效改善DNA編碼最佳化。

本研究將在第二章節介紹DNA編碼需定義的約束條件，此外，除了介紹目前較為熱門的最佳化演算法PSO外，亦對Mehrabian及Lucas兩位學者於2006年提出之雜草侵入最佳化演算法作詳盡說明[14]。第三章節對本文使用方法之實驗結果進行分析及比較。第四章節將討論本方法之結論及未來應用的可能性。

二、研究背景

使用DNA進行計算時需編碼出一組不受DNA特定生物反應(如雜交、斷裂、二級結構的形成)所影響的DNA序列，避免實驗結果產生誤差，故可將DNA編碼視為DNA計算一個初始參數最佳化問題。目前DNA編碼常用的約束條件有Hamming、H-measure、Melting Temperature(Tm)、Similarity、Continuity及Hairpin等。其設計之DNA序列皆能滿足以上約束，且各約束間具有相互關聯。本研究在不同約束條件下進行最佳化DNA組合編碼，並將實驗結果與其他文獻比較。以下分別介紹：1.約束條件設計，2.粒子族群最佳化，3.侵入雜草最佳化及4.粒子族群最佳化結合雜草侵入。

1. 約束條件設計：

(1) Tm：一條DNA序列對是由兩條完全互補的DNA序列組成，當DNA到達一定溫度，即分離為兩條DNA序列，此時溫度稱為Melting

temperature(Tm)。在PCR實驗中，Tm預測是一個相當重要的預測，錯誤的Tm會導致預期外之生化反應進而影響結果。目前已有許多方法進行預測Tm，如GC% [19]或nearest neighbor mode [16]，本研究將採用nearest neighbor mode進行預測，其公式如下：

$$FTm(\Sigma) = \sum_{i=1}^m fTm(i) \quad (1)$$

$$fTm(i) = [Tm_{target}(i) - Tm(i)]^2 \quad (2)$$

$$Tm(i) = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln(Cs/\alpha)} - 273.15 \quad (3)$$

其中公式(3)中R為氣體常數 $1.987 \text{ cal} \cdot \text{K}^{-1} \cdot \text{mol}^{-1}$ ，Cs為核酸濃度， α 設定為4。

(2) Similarity：在一大群DNA序列中，某一條DNA和DNA群於同一個方向(5'→3')是否相似，稱之為DNA的相似性(Similarity)。相似性過高可能使多條相似DNA序列對一項反應引物同時產生不同化學反應，而造成DNA計算的結果紊亂。因此設計之DNA序列須具有較低相似性，DNA序列間的相似性以公式(4)(5)來評估：

$$Fsimilarity(\Sigma) = \sum_{i=1}^m \sum_{j=1}^m Similarity(i, j), \quad j \neq i \quad (4)$$

$$similarity(i, j) = \max_{0 \leq g \leq n} \max_{0 \leq k \leq n+g-1} S(i(-)^g i, shift^k(j)) \quad (5)$$

其中 $(-)^g$ 為g個間隙， $shift^k$ 為位移k個位置， $S(i(-)^g i, shift^k(j))$ 表示有多少相同的DNA分子在序列 $i(-)^g i$ 及序列 $shift^k(j)$ 中。

(3) H-measure：DNA序列間的H-measure類似於Similarity，與Similarity相異處在於H-measure是評估一群DNA序列中一條DNA於不同方向(3'

→5')是否會和其他DNA產生互補，而結合成雙股螺旋結構，以下列公式計算：

$$FH - measure(\Sigma) = \sum_{i=1}^m \sum_{j=1}^m Similarity(i, j) \quad (6)$$

$$H - measure(i, j) = \max_{0 \leq g \leq n} \max_{0 \leq k \leq n+g-1} C(i(-)^g i, \sigma^k(j^R)) \quad (7)$$

其中 j^R 為反轉該條DNA序列， $C(i(-)^g i, \sigma^k(j^R))$ 表示有多少相同的DNA分子在序列 $i(-)^g i$ 及序列 $\sigma^k(j^R)$ 中。

(4) Continuity：DNA序列中若分子連續性太高(如：AAAA)，可能出現不穩定的DNA序列結構而影響實驗結果，為避免設計出DNA序列連續性過高，本研究定義DNA連續性限制如下：

$$Fcon(\Sigma) = \sum_{i=1}^m con(i) \quad (8)$$

$$con(i) = \sum_{j=1}^{l-t+1} \sum_{\alpha \in bp} T(B_\alpha(x, j), t)^2 \quad (9)$$

$$B_\alpha(x, j) = \begin{cases} b, & x_j \neq \alpha, x_{j+k} = \alpha \\ & 1 \leq k \leq b, x_{j+k+1} \neq \alpha \\ 0, & otherwise \end{cases} \quad (10)$$

$$T(i, j) = \begin{cases} i, & i > j \\ 0, & otherwise \end{cases} \quad (11)$$

其中 $B_\alpha(x, j)$ 表示第 j 個鹼基連續性， t 為所設定的連續性門檻值。

(5) Hairpin：DNA序列中假使有太多互補特性分子，可能使DNA序列自行鏈結而成髮夾結構(如：AAACAGATACCGTTT)。若一般DNA序列出現此情況，則會導致結果錯誤，因此予以限制。

$$FHairpin(\Sigma) = \sum_{i=1}^m \sum_r^{(n-2*pl)(n-pl-\lfloor r/2 \rfloor)} \sum_{c=p\lfloor r/2 \rfloor} Hairpin(i, c) \quad (12)$$

其中 r 為最小的莖環長度， pl 為枝幹長度， c 為形成髮夾結構莖環中央的鹼基所在位置。

(6) GCcontent：DNA序列之分子G、C比例，會對 T_m 高低造成影響，因此GC在DNA序列中的比例為DNA序列相當重要的評估條件，一般來說，GC所佔比例約為40-60%。

$$FGC(\Sigma) = \sum_{i=1}^m GC(i) \quad (13)$$

2. 粒子族群最佳化(PSO)

PSO於1995年被 Kennedy 及 Eberhart 提出 [11]，其靈感來自於觀察自然界中鳥群飛行和魚群活動。其概念是依據個體和族群間經驗做移動參考，以找出最佳搜尋目標之啟發式學習演算法。PSO族群是由數個粒子所構成的，並且在 d 維度的搜尋空間中進行移動。首先，藉由隨機方式初始多個可能的解，其中第 i 個粒子的位置和速度可分別表示為 $x_i = (x_{i1}, x_{i2}, \dots, x_{id})$ 和 $v_i = (v_{i1}, v_{i2}, \dots, v_{id})$ ，且每個粒子的位置與速度皆限制在 $[X_{min}, X_{max}]_d$ 及 $[V_{min}, V_{max}]_d$ 的範圍。此外，所有粒子皆有各自的搜尋區域，並會將自身搜尋經驗記錄下來。對單一粒子而言，本身記錄最好適應值稱為 $pbest$ ，可以表示為 $p_i = (p_{i1}, p_{i2}, \dots, p_{id})$ ；然而在粒子群中，最好個體最佳適應值則稱為 $gbest$ ，可以表示為 $g = (g_1, g_2, \dots, g_d)$ 。PSO的位置和速度之更新公式如下：

$$V^{new} = \omega * V^{old} + c_1 * r_1 * (pbest - x^{old}) + c_2 * r_2 * (gbest - x^{old}) \quad (14)$$

$$x^{new} = x^{old} + V^{new} \quad (15)$$

其中 ω 為慣性權重值， c_1 、 c_2 分別為 $pbest$ 和 $gbest$ 的學習因子， r_1 、 r_2 為0~1的隨機亂數，在本研究中，我們將 c_1 、 c_2 設定為2[11]。

3. 侵入雜草最佳化(IWO)

侵入雜草最佳化是利用自然界雜草植物的韌性及快速繁殖能力所設計開發的一種最佳化演算法。雜草的適應性、穩健性及生長習性已被證明對環境有著強大適應能力。當數顆雜草種子到達一個新環境時，將會快速繁殖並將子代散播至此環境中最適合生長的地點，讓子代能快速到達最適合生存的地點以確保整個族群的競爭力，其方法描述如下：

- (1) 初始族群：隨機於 d 維空間中初始數株雜草
- (2) 繁殖：每株雜草都會依據適應函數值產生不同數量種子，種子的產生數量如圖1所示：

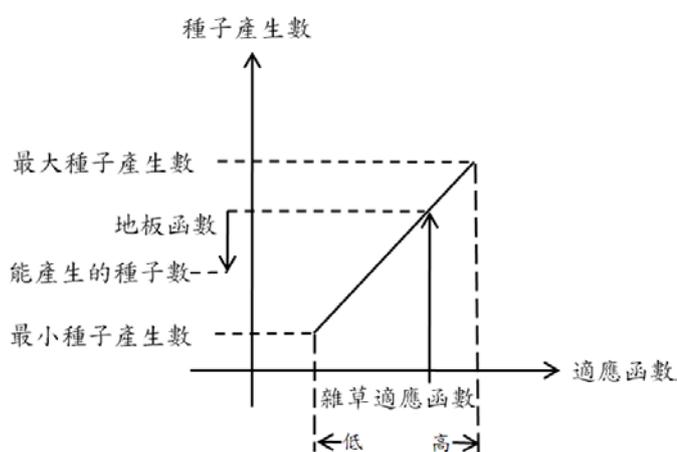


圖1 根據適應函數所能產生的種子數

- (3) 空間驅散：產生出的子代種子將會常態散佈於 d 維搜索空間內。在迭代過程中，使用以下公式確保產出子代種子會隨機分佈在自身母代附近。

$$\omega^{now} = \left[\frac{(\text{iterartive}_{\max} - \text{iterartive}_{\text{now}})}{\text{iterartive}_{\max}} \right]^n \quad (16)$$

$$* (\omega_{\text{inition}} - \omega_{\text{final}}) + \omega_{\text{final}}$$

其中 iterartive_{\max} 為最大迭代數， n 為非線性調變參數， ω_{inition} 及 ω_{final} 分別為3及0.001。

- (4) 生存競爭：當族群無法再產生有競爭力的後代時，此族群可能因無法適應環境變化及物種間競爭而滅絕。因此當雜草繁殖的子代散佈出去，使得族群達到最大數量時，需將各個母代及子代進行適應函數比較，適應函數較低之雜草所產出子代種子數量將會減少或無法產生子代種子而滅絕。

4. 粒子族群最佳化結合雜草侵入(PSO-IWO)

本研究使用方法將粒子族群最佳化及雜草侵入兩種方法相結合，利用粒子族群搜尋在更新過程中，以雜草快速繁殖特性將找到的區域快速搜尋並引導其他粒子往此方向移動。當其他粒子往此方向移動時，若找到適應性更佳的粒子，將改變其他粒子移動方向，往新的最佳粒子移動並搜尋出最佳解。本研究所用方法說明如下：

- (1) DNA粒子編碼：目前DNA序列都採用Watson-Crick模型[22]，將四種鹼基以較容易計算的方式呈現。本研究分別設定ACGT這四種鹼基為0、1、2、3，將這條DNA序列視為一個四進制數列。透過此四進制轉換為一般所使用的十進制，並使用最佳化演算法來找出最佳解，例如一條序列“AAAA”可以被視為 $0000_{(4)}=00_{(10)}$ ，“ACGA”可看做 $0121_{(4)}=25_{(10)}$ 。

- (2) 適應函數設計：DNA分子間的生化反應將影響DNA計算的結果，故編碼DNA序列後，需判斷是否合乎物理、化學及邏輯約束避免錯誤配對。一般來說，約束條件如H-measure、Similarity、Continuity、melting temperature及GC content等，都須考慮個別最佳化，故DNA序列編碼可視為多極值最佳化問題，其描述如下：

Optimize :

$$\left\{ FTm(\Sigma), Fsimilarity(\Sigma), FH - measure(\Sigma), \right. \\ \left. Fcon(\Sigma), FHairpin(\Sigma), FGC(\Sigma) \right\} \quad (17)$$

本研究制定此多目標函數為最小化問題，並且使用加權來處理各約束條件所產生的數值，藉權重加權的方法。轉換為單一目標最佳化問題，其適應函數轉換方程式如下：

$$Fitness = \sum_i wiFi(\Sigma) \quad (18)$$

$$i \in \left\{ Tm, similarity, H - measure, Hairpin \right\} \\ \left\{ GC, con \right\}$$

其中 wi 為權重加權值，本研究設定為1。

(3) 雜草繁殖：本研究利用PSO之快速搜尋能力，在解空間中搜尋全域最佳解，並使用IWO於 $gbest$ 附近快速繁殖產生子代種子，藉此加強區域搜尋的能力以找到最佳解。雜草繁殖的位置如下：

$$Seeds = Mod(gbest + \omega^{now} * Random(0,1), 4) \quad (19)$$

其中 $Mod(i, j)$ 為除以 j 後取餘數， $Random(0,1)$ 為隨機產生0~1之間的數， ω^{now} 為公式(16)計算之數值。

(4) 粒子更新：當在 $gbest$ 附近的種子繁殖到一定的數量時， $gbest$ 及子代適應函數將決定 $gbest$ 是否會被取代。當所繁殖出的子代適應函數值較 $gbest$ 更為優秀時，將取代 $gbest$ 成為新的 $gbest$ ，並引導剩下的粒子往此方向移動。

本方法的虛擬碼如下：

Pseudo-code for PSO-IWO procedure

```
01: begin
02: Randomly initialize particles swarm
03: while (the stopping criterion is not met)
```

```
04: Evaluate fitness of particle swarm
05: for n=1 to number of particles
06: Find  $pbest$ 
07: Find  $gbest$ 
08: Reproduction Seed by Eq. (19)
09: if fitness of Seed better than  $gbest$ 
10: change  $gbest$ 
11: end if
08: for  $d = 1$  to number of dimension of particle
09: update the position of particles by Eq. (14)-(15)
10: next  $d$ 
11: next  $n$ 
12: update the iterative value by Eq. (16)
13: next generation until stopping criterion
14: end
```

三、實驗結果及討論

本研究參數設定如下，迭代為 100 次，粒子數量為 20，每個粒子中有 7 條長度為 20 的 DNA 序列，繁殖種子為 10 顆種子， wi 為 1。在約束條件中，我們設定 Tm 核酸濃度 Cs 為 $1\mu M$ ，Continuity 為 3，Hairpin 形成至少 6 個鹼基對連結才會形成。

將編碼出的序列與其他方法編碼出的結果相比較。首先比較 Deaton *et al.*[6]使用 GA 模擬 Aldmen 實驗，其編碼出結果如表 1 所示。其 4 個約束條件：Similarity、H-measure、Continuity 及 Hairpin，比較結果如圖 2 所示。

在圖 2 中，可清楚看到 Similarity、Continuity、Hairpin 都比 GA 平均結果好，此實驗結果得知，編碼 DNA 序列將關係著之後的化學反應步驟是否能成功達到使用 DNA 計算來解決問題，因此不能只針對公式(18)一味尋找最小目標適應函數，而需考慮這些約束條件所編碼的 DNA 序列能夠穩定呈現各種化學反應的實驗結果，也就是需達到公式(17)中針對各個約束條件最佳化。由於 GA 利用交配及突變來跳脫區域最佳解，當迭代次數逐漸增大時，此機制跳脫機率將越來越低，因此實驗結果將只針對某些約束條件最佳化，對後續化學反應改善效果較為有限。本研究使用方法使用 PSO 尋找 $gbest$ 後，以 IWO 繁

殖方法於 *gbest* 附近繁殖,使 *gbest* 附近區域被快速搜尋,判斷此區域是否有更好的可能解,且 PSO 更新過程中其他粒子往 *gbest* 移動外亦尋找其路徑上是否有更佳解。故某個約束條件有較佳表現時,將以此為主要方向搜尋附近是否有更佳解。研究中發現,對各個約束條件做最佳化時,約束條件間將相互影響而使編碼的 DNA 序列有所不同。研究結果顯示本方法僅 H-measure 平均結果不如 GA,在其他約束條件表現上較為優秀。本研究方法 Similarity 為 50~55,而 H-measure 為 60~69,這些約束條件將會影響 DNA 序列是否會於化學反應時產生不良反應。DNA 序列的二級結構常用來預測 DNA 序列成為 RNA 時的功能形式,如何控制編碼 DNA 的二級結構亦是考量要素之一。

在 Continuity、Hairpin 這 2 個約束條件中,本研究所提出的實驗結果亦較優秀,平均數值均為 1.28,而在最容易影響化學反應成功與否的溫度上,PSO-IWO 所編碼的 DNA 序列在 T_m 及 GCcontent 皆較 GA 穩定,GA 最高解鏈溫度為 69.2009;但最低解鏈溫度為 48.4451。其表示在溫度上升過程中,有可能某幾條 DNA 序列失去活性,而使實驗結果產生誤差;反觀本研究所得之結果,最高解鏈溫度為 55.4303;最低解鏈溫度為 42.2240,相較之下更為穩定且降低誤差結果產生的可能性。由以上結果得知,本方法所編碼出的 DNA 序列將較 GA 所得結果在後續反應上較為穩定,且各種表現亦更優秀。

與 GA 之結果比較得知,本研究所提之方法能對多目標最佳化有顯著效果。但在多極值最佳化問題中,為使本研究方法能應用在不同問題上,我們和 Shin[17]應用於最短路徑問題(TSP)所編碼的結果相比較,其結果如表 2 所示,而比較結果如圖 3 所示。在尋找最短路徑本研究偏重利用公式(18)轉化為一個尋找最小化極值問題,相對於模擬 Aldmen 實驗編碼出來的 DNA 得考慮之後化學反應穩定使得各個約束條件需落在

理想範圍中,此實驗將尋找一個最小化最佳解而較不考慮各約束條件的理想範圍。在圖 3 中本方法結果雖然有一條 DNA 序列 T_m 不盡理想,但卻改善其他約束條件,像是 Similarity、H-measure、Hairpin、Continuity 則是有著相同的結果,顯示本方法在改善單極值問題最佳解有一定程度效果。

為表示本方法是將兩種演算法優點相結合,故與使用 GA/PSO[20]結果相比,其結果如表 3,比較結果如圖 4 所示,本研究結果在尋找最短路徑上較為優秀,在 H-measure 及 Continuity 皆有相當改善,於 Similarity 及 Hairpin 則小幅改善或具相同結果。在 GA/PSO 中,其方法分別經過 GA 及 PSO 更新才將其編碼出來,計算時間較為冗長。而本方法於 PSO 的更新過程前便先進入 IWO,而後將 PSO 粒子依照 *gbest* 方向開始做空間驅散,雖然亦提高 PSO 時間複雜度,卻可降低整體時間複雜度,且結果亦較為優秀。故由本方法得知,PSO 結合 IWO 將能有助於提升演算法的搜尋能力。

四、結論

由於 DNA 本身具不穩定性及複雜性,故採用 DNA 來計算前必需有穩定的 DNA 序列來進行計算,DNA 編碼屬於多極值目標最佳化問題,本研究提出一種新穎的演算法將 PSO 與 IWO 相結合,利用 PSO 搜尋全域最佳解結合 IWO 快速搜尋附近區域,針對數個約束條件進行 DNA 序列編碼,實驗結果顯示本方法和其他傳統演算法於此問題上有著更為優秀的表現,且將此問題轉換為單一極值最佳化問題上亦較混合 GA 與 PSO 演算法有效率。未來如何在加入更多及不同約束條件限制下使編碼 DNA 序列之計算結果正確性及穩定性提升將是本研究未來重點所在。

五、參考文獻

[1] L.M. Adelman, "Molecular computation of

- solutions to combinatorial problems”, *Science*, Vol. 266, pp. 1021-1024, 1994.
- [2] M. Arita, and A. Nishikawa, M. Hagiya, K. Komiya, H. Gouzu, K. Sakamoto, “Improving Sequence Design for DNA Computing”, in *Proceedings of Genetic and Evolutionary Computation Conference*, 2000.
- [3] R. S. Braich, N. Chelyapov, C. Johnson, P. W. K. Rothmund, and L. Adleman, “Solution of a 20-Variable 3-SAT Problem on a DNA Computer”, *Science*, Vol.296, pp.499-502, 2002.
- [4] A. Brenneman, and A. Condon, “Strand design for biomolecular computation”, *Theoretical Computer Science*, Vol.287, pp.39-58, 2002.
- [5] G. Z. Cui, Y. Y. Niu, Y. F. Wang, X. C. Zhang, and L. I. Pan, ” A new approach based on PSO algorithm to find good computational encoding sequences”, *Progress in Natural Science*, Vol.17, pp.712-716, 2007.
- [6] R. Deaton, R. C. Murphy, M. Garzon, and D. R. Franceschetti, S. E. Stevens, Jr., “Good Encodings for DNA-based Solutions to Combinatorial Problems” *Proceedings of 2nd DIMACS Workshop on DNA Based Computers*, pp.159-171, 1996.
- [7] A. G. Frutos, “Demonstration of a Word Design Strategy for DNA Computing on Surfaces”, *Nucleic Acids Research*, Vol.25, pp.4748-4757, 1997.
- [8] M. H. Garzon, and R. J. Deaton, “Biomolecular computing and programming”, *IEEE Transactions on Evolutionary Computation*, Vol. 3, pp. 236-250, 1999.
- [9] M. Garzon, P. Neathery, R. Deaton, R.C. Murphy, D.R. Franceschetti, and S.E. Stevens, Jr., “A New Metric for DNA Computing”, *Proceedings of the 2nd Annual Genetic Programming Conference*, USA: Stanford University, 1997.
- [10] A. J. Hartemink, D. K. Gifford, and J. Khodor, “Automated constraint-based nucleotide sequence selection for DNA computation” *Biosystem*, Vol.52, pp.227-235, 1999.
- [11] J. Kennedy, and R.C. Eberhart, “Particle swarm optimization, in: *Proceedings of IEEE International Conference on Neural Networks*, pp. 1942-1948, 1995.
- [12] C.C. Maley, “DNA computation: Theory, practice, and prospects”, *Evolutionary Computation*, Vol.6, pp. 201-229, 1998.
- [13] A. Marathe, A. E. Condon, and R. M. Corn, “On Combinatorial DNA Word Design”. *Proceedings of the 5th DIMACS Workshop on DNA-based Computers*. Cambridge, MA, USA : DIMACS Press, 1999.
- [14] A.R. Mehrabian, and C. Lucas, “A novel numerical optimization algorithm inspired from weed colonization”, *Ecological Informatics*, Vol.1, pp.355-366, 2006.
- [15] Q. Ouyang, P. D. Kaplan, S. Liu, and A. Libchaber, “DNA Solution of the Maximal Clique Problem”, *Science*, Vol.278, pp.446-449, 1997.
- [16] L. Santa, “A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics”, *Proc. Nat. Acad. Sci., USA*, pp.1460-1465, 1998.
- [17] S.Y. Shin, I.H. Lee, D. Kim, and B.T. Zhang, “Multiobjective evolutionary optimization of DNA sequences for reliable DNA computing”, *IEEE Transactions on Evolutionary Computation*, Vol.9, pp.143-158, 2005.
- [18] F. Tanaka, M. Nakatsugawa, M. Yamamoto,

T. Shiba, and A. Ohuchi, "Developing Support System for Sequence Design in DNA Computing", Proceedings of 7th International Workshop on DNA-Based Computers, 2001.

[19] J.G. Wetmur, "DNA probes: Applications of the principles of nucleic acid hybridization", Critical Reviews in Biochemistry and Molecular Biology, Vol.26, pp.227-259, 1991.

[20] C. Xu, Q. Zhang, B. Wang, and R. Zhang,

"Research on the DNA Sequence Design Based on GA/PSO algorithms", The 2nd International Conference on Bioinformatics and Biomedical Engineering, ICBBE 2008.

[21] K. Zimmermann, "Efficient DNA sticker algorithms for NP-complete graph problems", Computer Physics Communications, pp. 297-309, 2002.

表1 GA和PSO-IWO之DNA序列編碼

DNA序列(5'→3')	Similarity	H-measure	Hairpin	Continuity	GC%	Tm(°C)
PSO-IWO						
GCCAATACAATGAGGGCGGG	55	61	3	18	60	55.4303
AACTCTAATGAGTCGGTAGC	55	69	0	0	45	50.4841
CCGCAGCTTGCATATAACCT	53	66	0	0	50	48.8674
CATTACCATCCGCTTTAGAA	55	60	0	9	40	43.0013
CCAAGCAGCGATGAATACAG	51	65	0	0	50	50.4799
GCTCCCTTTCATACGAGCGG	50	64	0	18	40	42.2240
CCCGACTAGCGTCACGATCG	51	63	3	9	55	53.7547
Deaton's sequence(GA)						
ATAGAGTGGATAGTTCTGGG	64	55	3	9	45	52.6522
CATTGGCGGCGCGTAGGCTT	51	69	0	0	65	69.2009
GTTGTGACCGCTTCTGGGGA	63	60	0	16	60	60.8563
CAAAAACGACCAAAAGAGAG	45	58	0	41	40	52.7111
GATGGTGGTTAGAGAAGTGG	54	58	0	0	50	55.3056
TGTATCTCGTTTTAACATCC	50	61	4	16	35	48.4451
TTGTAAGCCTACTGCGTGAC	55	75	3	0	50	56.7055

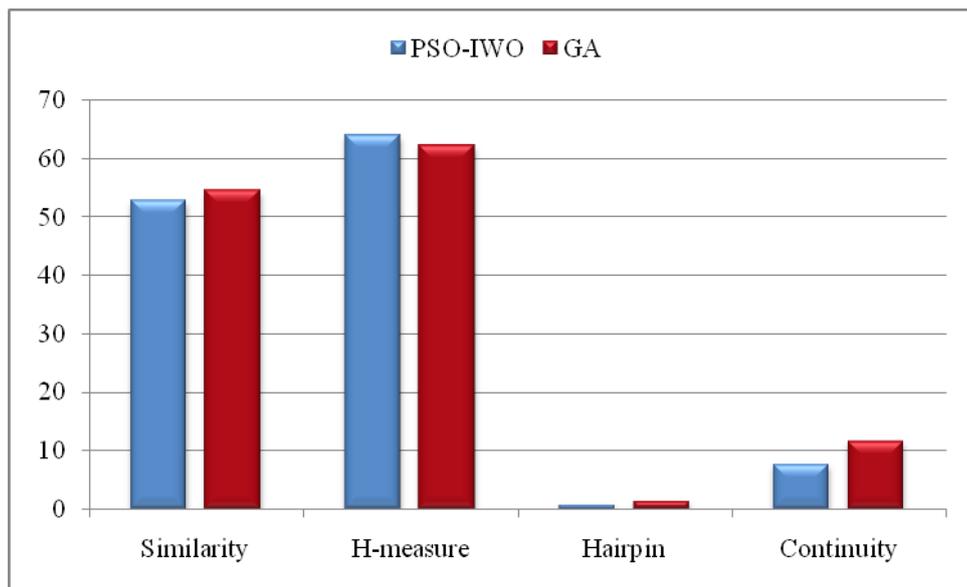


圖2 比較GA和PSO-IWO編碼之平均結果於4個約束條件

表2 CEA和PSO-IWO之DNA序列編碼

DNA序列(5'→3')	Similarity	H-measure	Hairpin	Continuity	GC%	Tm(°C)
PSO-IWO						
AGAGATAGCTGGGCCACGTA	51	62	0	9	55	53.4969
TTACTACTACCGGACGTAAGT	53	67	3	0	45	48.7496
ACTCAGAATCTCTGATCTTT	49	68	3	0	35	47.0155
TTACTGAAGACCAGAGGCAC	55	65	0	0	50	50.3165
ACCGTATGCTCAAGTTCCTA	55	63	0	0	45	46.0255
CGTGTAATGCCGGTGACCAC	49	62	3	0	60	52.6457
CCTCCTCAGTTGACCTATCC	52	61	0	0	55	39.6152
Shin's sequence(CEA)						
AGGCGAGTATGGGGTATATC	48	66	0	16	50	47.6070
CCTGTCAACATTGACGCTCA	57	66	3	0	50	50.6204
TTATGATTCCACTGGCGCTC	58	61	0	0	50	50.1205
ATCGTACTCATGGTCCCTAC	54	64	0	9	50	47.8464
CGCTCCATCCTTGATCGTTT	58	62	0	9	50	50.4628
CTTCGCTGCTGATAACCTCA	54	68	3	0	50	49.8103
GAGTTAGATGTCACGTCACG	51	67	3	0	50	48.3995

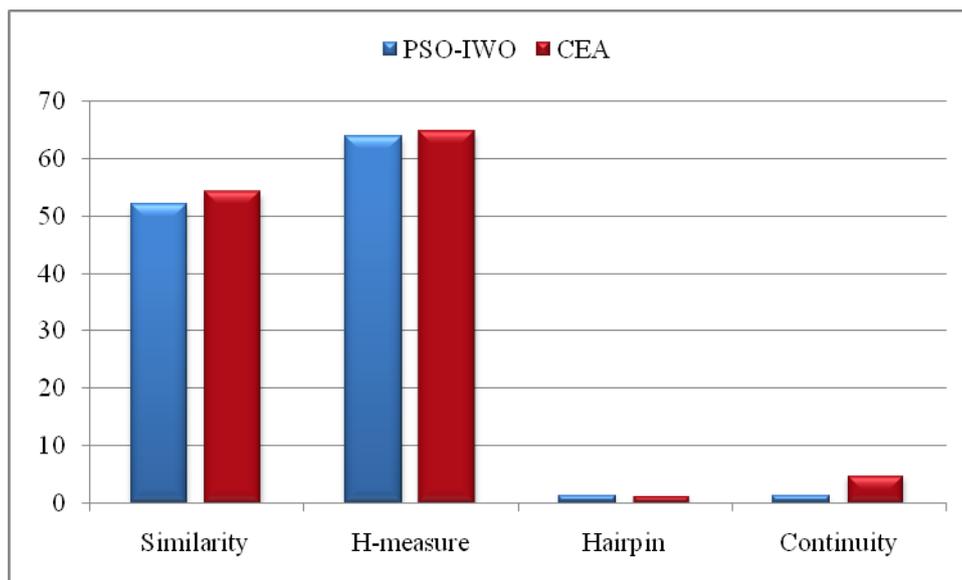


圖3 比較CEA和PSO-IWO編碼之平均結果於4個約束條件

表3 GA/PSO和PSO-IWO之DNA序列編碼

DNA序列(5'→3')	Similarity	H-measure	Hairpin	Continuity	GC%	Tm(°C)
PSO-IWO						
GCGCATATCTTCAATGTATC	50	68	0	0	40	48.5149
TCGCAATGCGGTCCTTACTT	55	62	0	0	50	53.2576
TTATTGTTATCCTATGTGCC	54	62	0	0	35	39.8866
GCTGTTAGATTCGTGTGTAG	53	58	0	0	45	50.0855
GAGCTACTTACATTGCTTAT	58	65	0	3	35	46.5020
TGCTATCTTGGAGTAGATCG	53	61	0	3	45	49.9901
AACTTGTGGTGGGTGGACTG	52	55	0	0	55	56.4718
Xu's sequence(GA/PSO)						
AGGAGGTCTCAGTTAGCATG	56	67	0	0	50	49.2149
TATCTAAGTCGCCGCAGATC	56	70	0	3	50	50.3076
TAGACTGGGTCTAGACGAG	54	65	9	0	50	51.4684
CTGCTACGCTTGTGATATCC	54	68	0	0	50	49.5876
TCCTAGATCCGTAGAGTCAC	55	70	0	0	50	45.0597
TTAGGAGTACACTCTCACGC	58	70	0	0	50	52.5508
ACGAACATCATCCCTACATCG	55	62	9	3	50	47.8391

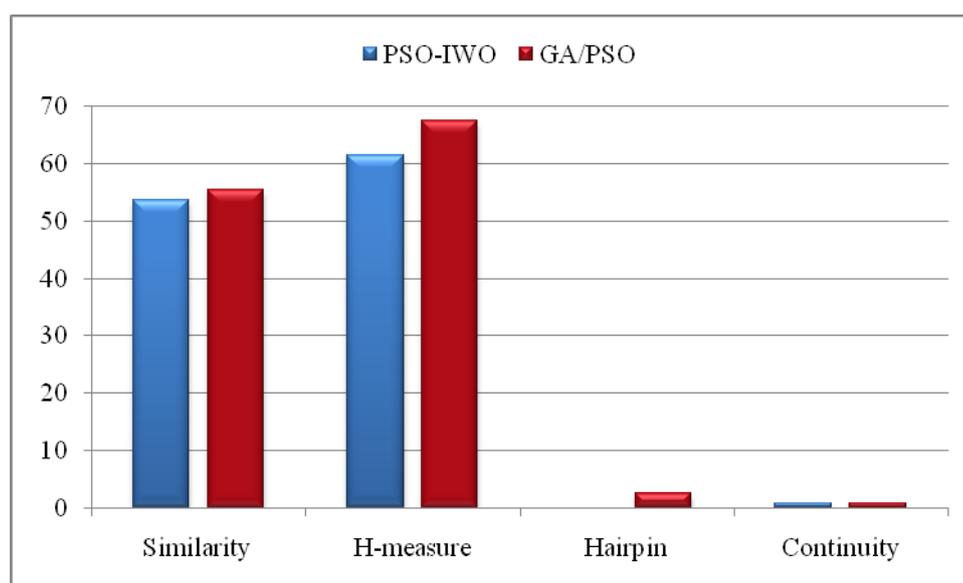


圖4 比較GA/PSO和PSO-IWO編碼之平均結果於4個約束條件