

دمج طريقة مقدر الترجيح الأعظم مع طريقة قيد الجيل باستخدام الخوارزمية الجينية لتقدير معلمة توزيع بولتزمان المتمثل بمعلمة الـ (RNA)

م.م. انتصار إبراهيم إلياس
كلية الإدارة والاقتصاد
جامعة نورو

Intisar_sulayman@yahoo.com

م.م. فرح سعد نشاط حميد
كلية الهندسة
جامعة دهوك

stat.farah@gmail.com

المستخلص:

مع التقدم العلمي الهائل الذي تشهده التكنولوجيا في الوقت الراهن، ظهرت أنماط جديدة من الأنظمة سميت بالأنظمة الذكية التي سرعان ما طُوِّرت واستخدمت في العديد من التطبيقات الحالية ومن هنا جاءت الخوارزميات الجينية (Genetic Algorithms) حيث استخدمت هذه الخوارزمية في هذا البحث لدراسة توزيع بولتزمان الذي يخضع له تركيب الحامض النووي الرايبوسومي RNA وتضمن اقتراح خوارزمية جينية تدمج بين طريقة مقدر الترجيح الأعظم مع طريقة قيد الجيل لتقدير معلمة توزيع بولتزمان المتمثل بمعلمة الـ RNA. وقد أظهرت النتائج ان الخوارزمية الجينية المدمجة أفضل لتقدير معلمة سلسلة الـ RNA من الطرق السابقة.

ولقد تم استخدام البرنامج Matlab في كتابة خوارزميات البحث وإيجاد النتائج.
الكلمات المفتاحية: الخوارزمية الجينية، توزيع بولتزمان، الترجيح الأعظم، قيد الجيل.

Combining Maximum Likelihood Estimate with Constraint Generation (ML-CG) Method by Using the Genetic Algorithm to Estimate the Parameter of Boltzmann's Distribution Represented by RNA

Assist. Lecturer Farah Saad Nashat Hameed
College of Engineering
University of Duhok

Assist. Lecturer Intisar Ibrahim Allyas
College of Administration and Economics
Nawroz University

Abstract:

With the enormous scientific progress that technology is witnessing at the moment, new types of systems have emerged called smart systems. That have been developed and used in many current applications Genetic algorithms. Were used in this research to study the distribution of Boltzmann, which is subject to the composition of ribosomal RNA, and included the suggestion of a genetic algorithm that combines the method of the maximum likelihood with the method of generation constraint to estimate the Boltzmann distribution parameter represented by the RNA parameter.

The results showed that the embedded genetic algorithm is better for estimating the RNA string parameter than previous methods.

Matlab has been used in writing research algorithms and finding results.

Keywords: genetic algorithm, Boltzmann distribution, Maximum Likelihood, generation constraint.

١. المقدمة:

الخوارزمية الجينية (Genetic Algorithms) هي طريقة من طرائق البحث العلمي لإيجاد الحل الأمثل وتندرج تحت مظلة الذكاء الاصطناعي وتصنف كأحدى طرائق الخوارزميات التطورية التي تعتمد على تقليد عمل الطبيعة، وقد وصفت بالجينية نظرا لاعتمادها الشديد على محاكاة عمل الجينات الوراثية في الكائنات الحية واستفادتها من مفهوم المعالجة المتوازية للتوصل للحل الأمثل. تعتبر الخوارزمية الجينية أحد أهم أدوات الذكاء الاصطناعي إذ إن برنامج الخوارزمية الجينية يتصف بخصائص البرنامج الذكي ألا وهي (التفكير والاستنتاج والتعلم) وهذا ما يتميز به البرنامج الذكي عن البرامج التقليدية الأخرى.

تعتبر الخوارزميات الجينية من التقنيات الهامة في البحث عن الخيار الأمثل من بين مجموعة حلول متوفرة لتصميم معين، إذ تقوم هذه المعالجة الوراثية بتمرير المزايا المثلى من خلال عمليات التوالد المتعاقبة، وتدعيم هذه الصفات، وتكون لهذه الصفات القدرة الأكبر على دخول عملية التوالد، وإنتاج جيل أمثل، وبتكرار الدورة الوراثية تتحسن نوعية الذرية تدريجياً.

توجد الخوارزميات الجينية في التطبيقات المعلوماتية الإحيائية (bioinformatics) وعلوم الحاسوب والهندسة والاقتصاد والكيمياء والصناعات التحويلية (manufacturing) والرياضيات والفيزياء وغيرها من الميادين. (عبد الغني، ٢٠١٦: ٤٨) حيث ظهرت هذه الخوارزمية في أوائل السبعينات عندما قام العالم John Holland في عام ١٩٧٠ بتقديمها على أنها خوارزمية بحث احتمالية تقوم على أساس آلية الانتقاء الطبيعي و الوراثة الطبيعية ووضح ان آلية عمل الخوارزمية الجينية تبدأ باختيار مجموعة من البيانات او الحلول وهي الكورموسومات (Population) (Tsoukalas, L., & Uhrig, R., 1997: 4) وغالبا ما يتم تمثيل البيانات بالأرقام الثنائية، ثم بعد ذلك تقوم بحساب دالة الأمثلة (fitness function) لكل كروموسوم وهي دالة تقييم النتائج الوسيطة و النهائية.

اما المراحل الأساسية للخوارزمية فتكون كلاتي:

- أ. البداية Start: توليد مجتمع عشوائي من الكروموسومات، أي بمعنى إيجاد حلول عشوائية مناسبة للمسألة.
- ب. اللياقة Fitness: هي دالة الهدف (Objective Function) وهي دالة مناسبة للحل في الخوارزمية الجينية حسب التطبيق المستخدم.
- ج. مجتمع جديد New Population: توليد جيل جديد بتكرار الخطوات الآتية إلى أن يكتمل الجيل، وتتضمن:

- الاختيار Selection: يتم اختيار مجموعة من الكروموسومات (Chromosomes Parents) من المجتمع الابتدائي استناداً إلى دالة اللياقة (أفضل القيم التي لها فرص أكبر للاختيار).
- التداخل الابدالي Crossover: إجراء إحدى عمليات التعابر للحصول على الذرية (Offspring) ويكون بين كروموسومين.
- الطفرة Mutation: مع احتمالية وجود الطفرة يتم عمل الطفرة للكروموسومات الجديدة بموقع معين في الكروموسوم، وتجري بين الجينات في الكروموسوم الواحد.
- الاستبدال (Replacement): عملية وضع الكروموسومات الجديدة المتكونة في الجيل الجديد للحلول محل المجتمع الابتدائي.
- الاختبار Test: عند تحقق شرط التوقف، فإن الخوارزمية الجينية تتوقف وتعيد الحل الجيد من آخر جيل متكون.
- الدورة Loop: يتم الرجوع إلى الخطوة ٢.

إن كل تكرار لهذه العملية يسمى بالجيل (Generation)، وبعد نهاية التنفيذ يقوم الباحث بتقديم تقرير عن الحقائق التي تم التوصل إليها. (ثابت، ٢٠٠٥: ١٤) و (Zhang, 2003: 145) وتقوم بتقييم الجيل الجديد بالاعتماد على دالة الأمثلية واتخاذ قرار لتكرار العمليات الأساسية اعلاه او قبول النتائج المرحلية والاكتفاء بها وبحسب متطلبات الحل.

تكون الخوارزميات الجينية مفيدة وفعالة عندما يكون فضاء البحث كبير أو معقد أو عندما يكون من الصعب استخدام طرائق البحث التقليدية لحل المسألة ولهذا اثبتت هذه الخوارزمية نجاح كبير في كثير من المجالات وانتشر استخدامها.

ومن هنا ظهرت الحاجة لدراسة أحد التراكيب الحيوية لجسم الانسان وهو الحامض النووي الرايبوسومي RNA الذي يخضع لتوزيع بولتزمان.

إن جزيئة الـ RNA هي سلسلة لقواعد النيوكليوتيد، وهناك أربعة أنواع من القواعد وهي: أدنين (A)، سايتوسين (C)، كوانين (G)، واليوراسيل (U).

إن الأزواج A=U و C≡G تدعى بقواعد واتسن-كريك (Watson-Crick) المكملة التي تكون ذات شكل مستقر وترتبط بواسطة أوامر هيدروجينية وعندما تشكل هذه القواعد الاتصال تدعى عندئذٍ هذه الأزواج بـ (الأزواج الأساسية)، هناك زوج آخر وهو G=U يدعى بـ (زوج التذبذب: Wobble Pair) وهو يشكل زوجاً أساسياً قوياً آخر. (Ding, 2006: 213)

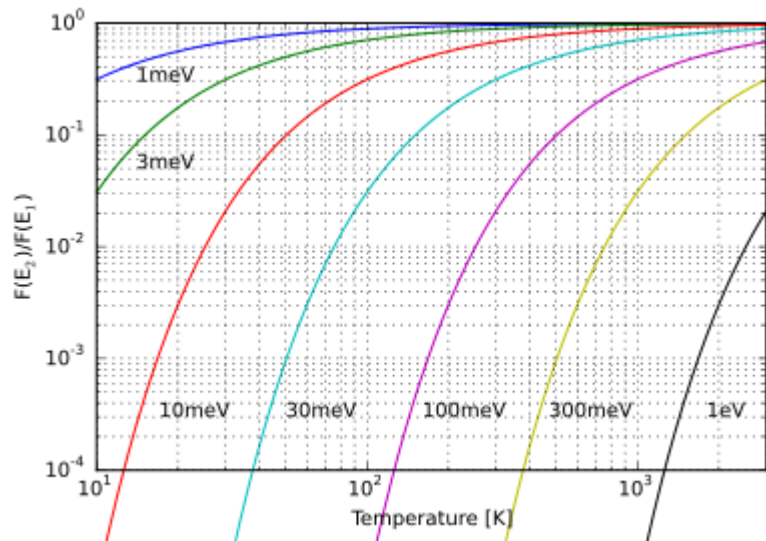
٢. توزيع بولتزمان Boltzmann's Distribution:

إن توزيع بولتزمان (والذي يسمى أيضاً توزيع Gibbs) (Landau and Lifshitz, 1980: 57) هو توزيع الاحتمال، أو قياس الاحتمال، وهو توزيع الترددات للجسيمات في نظام ما على عدد الحالات المحتملة. وهناك العديد من البحوث والدراسات عن هذا التوزيع ففي عام ١٩٨٩ قدم العالمان Tyagi و Bhattacharya توزيع بولتزمان لأول مرة كنموذج لعمر الايون وناقشا نظرية بيز وتقدير الحد الأدنى للتباين غير المتحيز لمعلماته (Tyagi, R.K., 1989: 73-79)، وفي عام 1998 قام العالمان Rani و Chaturvedi بتقدير كلاسيكي لمعلمات توزيع بولتزمان بعد تعميمها وذلك عن طريق ادخال

معلمة او اكثر للتوزيع (Chaturvedi,A. and Rani, U., 1998: 113-120)، اما في عام ٢٠٠٥ قام العالمان Bekker و Roux بدراسة توزيع بولتزمان وتقدير معلماته باستخدام نظرية (Empirical Baye's estimation) (Bekker,A.J.J.J and Roux, J.J., 2005: 2169-2178)، ان تقدير العينات الكبيرة يمكن ان تكون مكلفة او حتى مستحيلة لهذا طور كل من العالم Kazmi و اخرون في عام ٢٠١٢ طريقة الترجيح الاعظم لتوزيع بولتزمان وذلك بإضافة اوزان للبيانات وتمكنوا من تقدير المعلمات لهذه العينات (Kazmi,S.,Aslam,M. and Ali, S., 2012: 197-218).
يعبر عن توزيع بولتزمان في المعادلة (1.1):

$$F(\text{state}) \propto e^{-\frac{E}{kT}}$$

حيث ان E هي طاقة الحالة (والتي تختلف من حالة إلى أخرى) و kT (ثابت التوزيع) وهو نتاج درجة الحرارة الثابتة والثرموديناميكية لبولتزمان، والشكل (١-١) يوضح التوزيع:



الشكل (١-١) يوضح احتمالية توزيع بولتزمان والتي تعتمد على التغير في الطاقة ودرجة الحرارة المصدر: (Mcquarrie, 2000: 22).

اما في الميكانيكا الإحصائية والرياضيات فإن توزيع بولتزمان هو توزيع احتمالي يعطي احتمالية أن يكون النظام في حالة معينة كدالة لطاقة تلك الحالة ودرجة حرارة النظام الذي يتم تطبيق التوزيع عليه.

وكما موضح بالمعادلة (1.2): (Mcquarrie, 2000)

$$p_i = \frac{e^{-\varepsilon_i/kT}}{\sum_{j=1}^M e^{-\varepsilon_j/kT}}$$

حيث p_i هو احتمالية الحالة i ، ε_i طاقة الحالة i ، k ثابت بولتزمان، T درجة حرارة النظام و M هو عدد كافة الحالات التي يمكن الوصول إليها من خلال النظام، ويشار إليها عند بعض المؤلفين بالرمز Q او الرمز Z وكما في المعادلة (1.3):

$$Q = \sum_{i=1}^M e^{-\varepsilon_i/kT}$$

لذلك، يمكن أيضا أن تكتب صيغة توزيع بولتزمان كما في المعادلة (1.4):

$$p_i = \frac{1}{Q} e^{-\varepsilon_i/kT}$$

وهذا هو التوزيع الذي يزيد من النظام الديناميكي الحراري (1.5):

$$H(p_1, p_2, \dots, p_M) = - \sum_{i=1}^M p_i \log_2 p_i$$

ان القيد $\sum_{i=1}^M p_i$ يخضع الى قيمة طاقة متوسطة خاصة (والتي يمكن إثباتها باستخدام مضاعفات لاجرانج).

والدالة الجزئية يمكن ان تحسب إذا علمنا طاقات المستويات التي يمكن الوصول إليها من خلال النظام المدروس.

وبين التوزيع أن الحالات ذات الطاقة المنخفضة ستحتل دائمًا باحتمالية أكبر للانفعال من الحالات ذات الطاقة الأعلى والتي يمكن أن تعطينا أيضا العلاقة الكمية بين احتمالات الحالتين التي تشغلها. يتم إعطاء نسبة الاحتمالات للحالتين i و j كما المعادلة (1.6):

$$\frac{p_i}{p_j} = e^{(\varepsilon_j - \varepsilon_i)/kT}$$

حيث p_i هو احتمالية الحالة i ، p_j : احتمالية الحالة j ، و ε_i و ε_j هي الطاقة للحالات i و j ، على التوالي. وكثيرا ما يستخدم توزيع بولتزمان لوصف توزيع الجسيمات، مثل الذرات أو الجزيئات، على حالات الطاقة التي يمكن الوصول إليها. إذا كان لدينا نظام يتكون من العديد من الجسيمات، فإن احتمالية وجود جسيم في الحالة i هي احتمالية عملية، إذا اخترنا جسيماً عشوائياً من هذا النظام وتحقق من الحالة الموجودة فيه، فسنجد أنه في الحالة i . هذا الاحتمال يساوي عدد الجسيمات في الحالة i مقسوماً على إجمالي عدد الجسيمات في النظام، أي جزء من الجزيئات التي تشغل الحالة i .

$$p_i = \frac{N_i}{N}$$

حيث N_i هو عدد الجسيمات في الحالة i و N هو العدد الإجمالي للجسيمات في النظام. قد نستخدم توزيع بولتزمان لإيجاد هذا الاحتمال الذي، كما رأينا، يساوي جزء من الجسيمات الموجودة في الحالة i . لذا فإن المعادلة التي تعطي جزء من الجسيمات في الحالة i كدالة لطاقة تلك الحالة هي (1.8) (Atkins, 2010: 33)

$$\frac{N_i}{N} = \frac{e^{-\varepsilon_i/kT}}{\sum_{j=1}^M e^{-\varepsilon_j/kT}}$$

هذه المعادلة لها أهمية كبيرة في التحليل الطيفي، حيث انه في التحليل الطيفي نلاحظ خطأ طيفيًا من الذرات أو الجزيئات التي تهتم بالانتقال من حالة إلى أخرى. (Atkins, 2010: 128) (Atkins and De Paula, 2009: 55) لكي يكون هذا ممكناً ، يجب أن يكون هناك بعض الجسيمات في الحالة الأولى قابلة للخضوع لعملية الانتقال. قد نجد أن هذا الشرط يتحقق من خلال إيجاد جزء من الجسيمات في الحالة الأولى. من المرجح جداً أنه لا يتم ملاحظة التحول في درجة الحرارة التي تم حسابها، وبشكل عام، فإن الجزء الأكبر من الجزيئات في الحالة الأولى يعني عدداً أكبر من التحولات إلى الحالة الثانية (Skoog, Holler and Crouch, 2006: 89) وهذا يعطي خط طيفي أقوى. ومع ذلك، هناك عوامل أخرى تؤثر على شدة الخط الطيفي، مثل ما إذا كان ذلك ناتجاً عن انتقال مسموح به أو ممنوع.

وتعرف نسبة توزيع بولتزمان المحسوبة للحالتين بـ (عامل بولتزمان) وتمثل فقط بالاعتماد على التغير في الطاقة للحالات وتمثل بالمعادلة (1.9):

$$\frac{F(\text{state2})}{F(\text{state1})} = e^{\frac{E_1 - E_2}{kT}}$$

٣. تعظيم دالة الترجيح لتوزيع بولتزمان باستخدام الخوارزمية الجينية:

Maximizing the Likelihood Function of Boltzmann Distribution by Using Genetic Algorithm

تعرف دالة كثافة الاحتمال لتوزيع بولتزمان بالشكل الآتي:

$$P(x; \theta) = \begin{cases} \exp \frac{-\Delta G(x; \theta)}{kT} & \text{if } \Delta G(x; \theta) > 0 \\ 1 & \text{if } \Delta G(x; \theta) \leq 0 \end{cases} \quad (1.10)$$

إذ أن:

ΔG : التغير في الطاقة، وتعرف بالصيغة الآتية:

$$\Delta G = \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta \quad (1.11)$$

k: ثابت بولتزمان.

T: درجة الحرارة المطلقة.

سوف يتم إيجاد تقدير المعلمة θ لتوزيع بولتزمان الشرطي باستخدام طريقة الترجيح الأعظم، وهنا سنعرف سلسلة الـ (RNA) بـ (x)، ونعرف الاحتمالية لتركيب سلسلة الـ (RNA) بـ (y)، ونستخدم النموذج الخطي-اللوجاريتمي الشرطي لتوزيع بولتزمان (Boltzmann Distribution) ثم نقوم بتعظيم دالة الترجيح لتوزيع بولتزمان باستخدام الخوارزمية الجينية. إن دالة كثافة الاحتمال لتوزيع بولتزمان الشرطي (Conditional Boltzmann Distribution) تعرف كالتالي: (Andronescu, 2006)

$$p(y/x, \theta) = \frac{1}{\delta(x, \theta)} \exp\left(\frac{-1}{kT} \Delta G(x, y, \theta)\right) \quad (1.12)$$

حيث ان ΔG معرفة في المعادلة (1.11) وأن الدالة الجزئية $\delta(x, \theta)$ تمثل بـ:

$$\delta(x, \theta) = \sum_y \exp\left(\frac{-1}{kT} \Delta G(x, y, \theta)\right) \quad (1.1.12)$$

وبالتعويض عن δ في المعادلة (1.12) نحصل على المعادلة الآتية:

$$p(y/x, \theta) = \frac{\exp\left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta\right)}{\sum_y \exp\left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta\right)} \quad (2.1.12)$$

ولإيجاد مقدر الترجيح الأعظم لتوزيع بولتزمان الشرطي نتبع الخطوات الآتية:
أولاً. نجد دالة الإمكان $L(\theta)$ كالآتي:

$$L(\theta) = \prod_{r=1}^n p(y/x, \theta) \quad (1.13)$$

$$\prod_{r=1}^n = \sum_{m=1}^{16} \sum_{m=1}^{16} \quad \text{إذ أن:}$$

$$= \prod_{r=1}^n \frac{\exp\left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta\right)}{\sum_y \exp\left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta\right)} \quad (1.1.13)$$

$$= \left(\frac{\exp \sum_{r=1}^n \left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta \right)}{\prod_{r=1}^n \sum_y \exp\left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta\right)} \right) \quad (2.1.13)$$

ثانياً. نأخذ اللوغاريتم لدالة $L(\theta)$ لتصبح المعادلة كالآتي:

$$\ln L(\theta) = \log\left(\frac{\exp \sum_{r=1}^n \left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta \right)}{\prod_{r=1}^n \sum_y \exp\left(\frac{-1}{kT} \sum_x \frac{1}{y!} \sum_{i=0}^{y-1} (-1)^i \frac{y!}{i!(y-i)!} (y-i)^x \theta\right)} \right) \quad (3.1.13)$$

وبتبسيط المعادلة الأخيرة نحصل على ما يلي:

$$\log L(\theta) = - \sum_{x=1}^4 \sum_{y=1}^4 \left(\ln \sum_{m=1}^{16} \sum_{m=1}^{16} \exp\left(\frac{-1}{kT} \sum_x \frac{1}{y(m)!} \sum_{i=0}^{y-1} (-1)^i \frac{y(m)!}{i!(y(m)-i)!} (y(m)-i)^{x(m)} \theta\right) - \right. \quad (4.1.13)$$

$$\left. \sum_{m=1}^{16} \sum_{m=1}^{16} \left(\frac{-1}{kT} \sum_x \frac{1}{y(m)!} \sum_{i=0}^{y-1} (-1)^i \frac{y(m)!}{i!(y(m)-i)!} (y(m)-i)^{x(m)} \theta\right) \right)$$

٤. تقليل قيد الجيل لتوزيع بولتزمان باستخدام الخوارزمية الجينية:

Minimizing Constraint Generation (CG) of Boltzmann Distribution Using Genetic Algorithm

الطريقة البديلة لطريقة الترجيح في تقدير المعلمة θ هي طريقة قيد الجيل والتي اقترحت من قبل (Andronescu) وآخرون في عام (٢٠٠٦) والطريقة تتلخص بتقدير قيمة θ لتوزيع بولتزمان بحل نظام القيود الموضح في المعادلة الآتية: (Andronescu, 2006)

$$\Delta G(x, y_x, \theta) \leq \Delta G(x, y, \theta) \quad (1.14)$$

$$Y_x \setminus \{ y_x \} \in y$$

إذ أن:

y_x : تمثل مجموعة كل التراكيب الثانوية للمتسلسلة x .

هذه القيود تضمن أن لكل سلسلة x جميع تراكيبها الثانوية y غير المثلى والتي يرمز لها بـ (y_x) لها أعلى طاقة حرة دنيا. (من خلال ذلك نفرض أنه لا يوجد تركيب آخر له نفس الطاقة الحرة الدنيا (Minimum Free Energy: MFE) كالتركيب المعروف لدينا.

قد يحدث أن هذا النظام من القيود يكون غير قابل للتطبيق، أي أنه لا يوجد حل للمعلمة (θ) الموجودة في توزيع بولتزمان والتي تعرف كل القيود بشكل آني. وللتعامل مع التطبيقات غير الممكنة، سنضيف المتغيرات البطيئة $\delta_{x,y} \geq 0$ إلى القيود، والتي سوف يتم تقليل قيمتها، وهذا يؤدي إلى استرخاء (Relaxation) القيود (أي تحسين قيمتها) كما موضح في المعادلة الآتية:

$$\Delta G(x, y_x, \theta) \leq \Delta G(x, y, \theta) + \delta_{x,y} \quad (1.15)$$

وعلى اعتبار أن دالة الطاقة ΔG معرفة سابقاً في المعادلة (1.11)، ويمكن تحويل المتراحة (1.15) إلى معادلة معرفة كالآتي:

$$\Delta G(x, y_x, \theta) - \Delta G(x, y, \theta) = \delta_{x,y} \quad (1.16)$$

وبالتالي يجب أن تتحقق دالة قيد الجيل والممثلة بالمعادلة (1.17) التالية:

$$\text{minimize } \|\delta\|^2 \quad (1.17)$$

مع تحقق القيد:

$$\delta \geq 0 \quad (1.18)$$

٥. دمج طريقة مقدر الترجيح الأعظم مع طريقة قيد الجيل باستخدام الخوارزمية الجينية:

Combining Maximum Likelihood Estimate with Constraint Generation (ML-CG) Method by Using the Genetic Algorithm

تم هنا دمج كل من طريقة الترجيح الأعظم (M.L.E.) مع طريقة قيد الجيل (CG) واعتبرنا دالة الهدف لعملية الترجيح كالاتي:

$$\text{Target Function} = \begin{cases} \max \log L(\theta) & \text{if } \delta < 0 \\ \text{minimize } \|\delta\|^2 & \text{if } \delta \geq 0 \end{cases} \quad (1.19)$$

وخطوات الخوارزمية المقترحة لعملية الدمج تكون بالشكل الآتي:

١-٥. الخوارزمية الجينية المقترحة لإيجاد المعلمة θ باستخدام دمج طريقة مقدر الترجيح الأعظم مع طريقة قيد الجيل

Proposed Genetic Algorithm for Finding the Value of the Parameter θ by Using Combined Maximum Likelihood Estimate and Constraint Generation (ML-CG) Methods

١. البيانات الأولية (Initial Data): وهي قراءة لمجموعة من المتغيرات التي استخدمت في الخوارزمية:

- العدد m : وهو يمثل الطول لكل من سلسلتي x و y .
- $x(m)$: يمثل متجه لأحد النيوكليوتيدات الداخلة في تركيب الـ RNA ويمثل القواعد النتروجينية: (أدينين A، يوراسيل U، سايتوسين C، كوانين G) وكل عنصر مكرر أربع مرات، إذ تم تشفير هذه التراكيب ورمزنا للأدينين بالرمز (٠٠٠)، ولليوراسيل بالرمز (٠٠١)، وللسايتوسين بالرمز (١١١)، والكوانين بالرمز (٠١٠).
- $y(m)$: يمثل متجه التراكيب الناتجة من تداخل هذه القواعد النتروجينية وللتكرارات أعلاه (فقط التراكيب المقبولة التي فيها يرتبط الأدينين A مع اليوراسيل U بأصرتين هيدروجينيتين $A=U$ ويرتبط السايتوسين C مع الكوانين G بثلاث أوأصر هيدروجينية $C \equiv G$)، وبهذا سوف يصبح تركيب الـ AU هو (٠٠٠٠٠١)، وتركيب الـ CG هو (١١١٠١٠)، وتركيب الـ UA هو (٠٠١٠٠٠)، وتركيب الـ GC هو (٠١٠١١١).
- K : يمثل ثابت الغاز.
- T : تمثل درجة الحرارة المطلقة.

٢. إنشاء الجيل الابتدائي (Initial Generation): الجيل الابتدائي في هذه الخوارزمية يحدث من تزاوج الافراد الافضل وغالبا ما تتم هذه العملية بتبادل انصاف التمثيل الثنائي بين الابوين حيث ان أحدهما يمثل القواعد النتروجينية والآخر يمثل تراكيب هذه القواعد النتروجينية، وقد تم وضع القيم الحقيقية في جينات الكروموسومين أي أن التشفير كان تشفيراً حقيقياً للكروموسومات.

٣. قيمة الجودة (Fitness Value): إن قيمة الجودة في هذه الخوارزمية تمثل بالمعادلة (١٩، ١) إذ أنه:

- عندما δ أقل من الصفر سوف يتم حساب تقدير الترجيح الأعظم M.L.E. والتي سوف تعظم قيمتها عن طريق الخوارزمية الفرعية الاتية:

الخطوة الأولى والثانية كما في بداية الخوارزمية.

الخطوة الثالثة: قيمة الجودة (Fitness Value): إن قيمة الجودة في هذه الخوارزمية تمثل مُقدر الترجيح الأعظم لتوزيع بولتزمان والممثلة بالمعادلة (٤,١,١٣)، إذ تم استخدام اللوغاريتم لأنه الأساس في استخراج المقدرات التي تجعل المقدر أعظم ما يمكن.

الخطوة الرابعة: تم استخدام الانتقاء من النوعين الآتين (Uniform, Roulette).

الخطوة الخامسة: تم استخدام التعابر من الأنواع الآتية: (Scattered, Intermediate, Heuristic, Single Point)

الخطوة السادسة: تم استخدام الطفرة من الأنواع الآتية (Gaussian, Uniform).

وتم التحديد المسبق لعدد الأجيال في توقف الخوارزمية الجينية.

- اما عندما δ أكبر أو تساوي الصفر سوف يتم حساب معيار الطول (Norm) وبعدها سيتم إيجاد أقل قيمة لها، عن طريق الخوارزمية الفرعية الآتية:

الخطوة الأولى والخطوة الثانية كما في الخوارزمية الاساسية.

الخطوة الثالثة: قيمة الجودة (Fitness Value): إن قيمة الجودة في هذه الخوارزمية ممثلة بالمعادلة (1.17) إذ استخدم الرمز δ ليمثل ناتج طرح دوال الطاقة في الحالتين: عند تكرار عدد التراكيب $\Delta G(x, y_x, \theta)$ ، وعند ثبوت عدد التراكيب التي رُمز لها بـ $\Delta G(x, y, \theta)$ ، أي أن:

$$\delta = \Delta G(x, y_x, \theta) - \Delta G(x, y, \theta)$$

وفي هذه الحالة يجب تحقق الشرط التالي وهو أن δ أكبر أو تساوي الصفر، عندها سوف يتم حساب معيار الطول (Norm) لـ δ بعدها سوف يتم تقليل قيمتها.

الخطوة الرابعة: تم استخدام الانتقاء من النوعين الآتين (Uniform, Roulette).

الخطوة الخامسة: تم استخدام التعابر من الأنواع الآتية: (Scattered, Intermediate, Heuristic, Two Point, Single Point)

الخطوة السادسة: تم استخدام الطفرات الآتية (Gaussian, Uniform).

وتم التحديد المسبق لعدد الأجيال في توقف الخوارزمية الجينية.

بعد الانتهاء من حساب قيمة دالة الجودة تستكمل الخوارزمية الاساسية عملها بالخطوة الآتية:

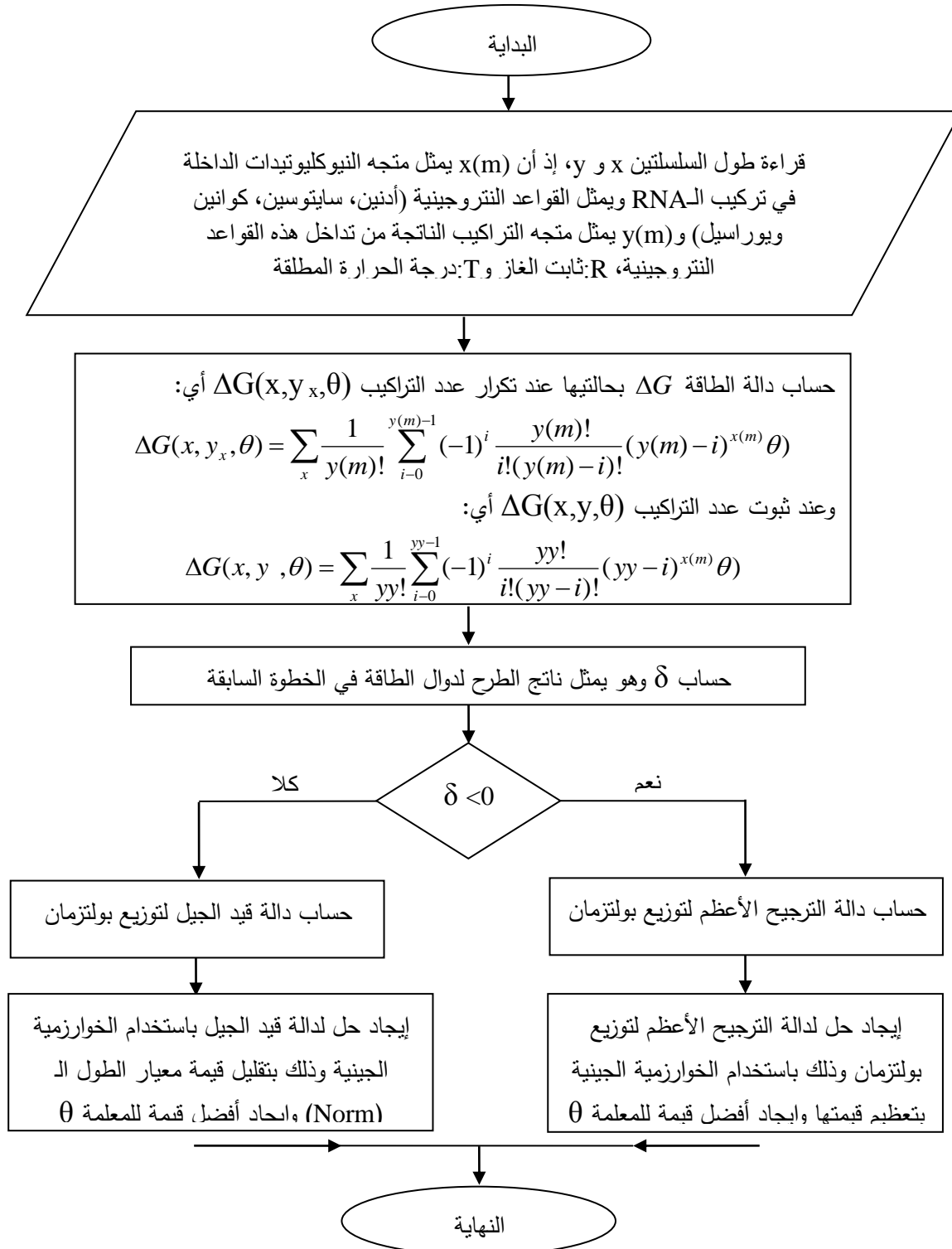
٤. تم استخدام انتقاء عجلة الروليت والانتقاء المنتظم.

٥. تم استخدام التعابر من الأنواع الآتية: (Scattered, Intermediate, Heuristic, Single Point)

٦. تم استخدام الطفرة من الأنواع الآتية (Gaussian, Uniform).

وتم التحديد المسبق لعدد الأجيال في توقف الخوارزمية الجينية.

والشكل التالي يمثل المخطط الانسيابي للخوارزمية الجينية المقترحة:



الشكل (٢-١): يمثل المخطط الانسيابي لدمج طريقة الترجيح الأعظم مع طريقة قيد الجيل لتوزيع بولتزمان وإيجاد قيمة المعلمة θ

الخوارزمية من اعداد الباحثان.

تم تصميم برنامج بلغة MATLAB ثم تم ايجاد قيمة دالة الجودة ومن خلاله تم تحديد قيمتي المعلمة التي تحقق أعلى قيمة لدالة ترجيح وأقل قيمة لدالة قيد جيل (أي ستحقق القيمة العظمى لدالة الترجيح في حالة δ أقل من الصفر وتحقق القيمة الدنيا لدالة قيد الجيل في حالة δ أكبر أو تساوي الصفر) وتم اختيار عدة أنواع لكل من (الانتقاء، التداخل الابدالي والطفرة) وكذلك عدد مرات توليد الأجيال تراوحت ما بين (٢٥-١٥٠٠) جيل وملاحظة النتائج بعد كل جيل واعتمد على التحديد المسبق لعدد الأجيال لبيان مدى التقرب من الحل الأمثل. والجدول (١-١) يوضح النتائج التي تم الحصول عليها وضحت هذه النتائج في الشكلين (٣-١) و (٤-١) على التوالي.

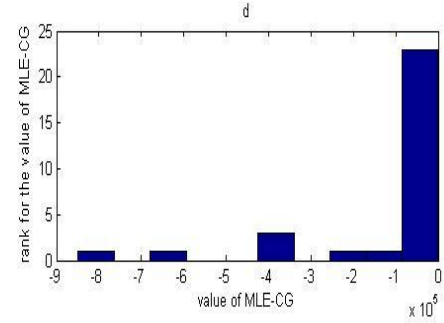
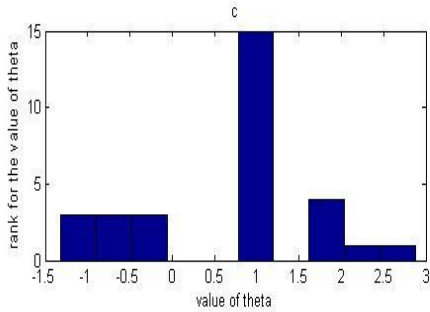
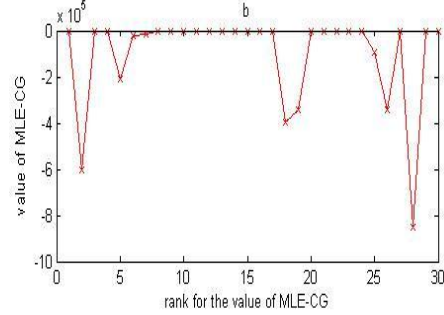
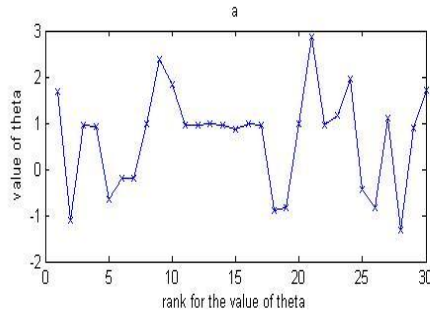
من خلال الخوارزميات أعلاه استطعنا الحصول على أكبر قيمة لدالة الترجيح لتوزيع بولتزمان وأقل قيمة لدالة قيد الجيل للتوزيع نفسه باستخدام الخوارزمية الجينية المقترحة وحصلنا على أمثل قيمة للمعلمة θ وهي (-0.1867) في حالة مقدر الترجيح الاعظم و(٢,٨٧٧٢) في حالة تقدير قيد الجيل.

الجدول (١-١) يمثل دمج طريقة مقدر الترجيح الأعظم M.L.E. مع طريقة قيد الجيل CG لتوزيع بولتزمان وإيجاد أمثل قيمة للمعلمة θ

	عدد الأجيال	نوع الانتقاء	نوع التداخل الابدالي (التعابر)	نوع الطفرة	قيمة θ	*MLE **CG
1.	100	Uniform	Scattered	Gaussian	1.6802	-0.60821*10 ¹
2.	100	Uniform	Single Point	Gaussian	-1.1047	-0.60071422*10 ⁶
3.	49	Uniform	Single Point	Uniform	0.9473	-0.19333*10 ¹
4.	25	Uniform	Single Point	Uniform	0.9271	-0.18519*10 ¹
5.	150	Uniform	Scattered	Gaussian	-0.6536	-0.20924291*10 ⁶
6.	29	Uniform	Scattered	Gaussian	-0.2012	-0.19819812*10 ⁵
7.	100	Uniform	Single Point	Gaussian	-0.1867	-0.16704016*10 ^{5*}
8.	52	Uniform	Intermediate	Uniform	0.9713	-0.20324*10 ¹
9.	100	Stochastic Uniform	Scattered	Gaussian	2.3818	-0.12222*10 ²
10.	100	Roulette	Scattered	Gaussian	1.8310	-0.72234*10 ¹
11.	100	Roulette	Heuristic	Uniform	0.9554	-0.19665*10 ¹
12.	100	Roulette	Intermediate	Uniform	0.9689	-0.20226*10 ¹
13.	73	Roulette	Intermediate	Uniform	0.9891	-0.20773*10 ¹
14.	84	Roulette	Intermediate	Uniform	0.9544	-0.19623*10 ¹
15.	99	Roulette	Scattered	Uniform	0.8698	-0.16299*10 ¹
16.	77	Roulette	Heuristic	Uniform	0.9901	-0.21122*10 ¹
17.	64	Roulette	Single Point	Uniform	0.9461	-0.19286*10 ¹

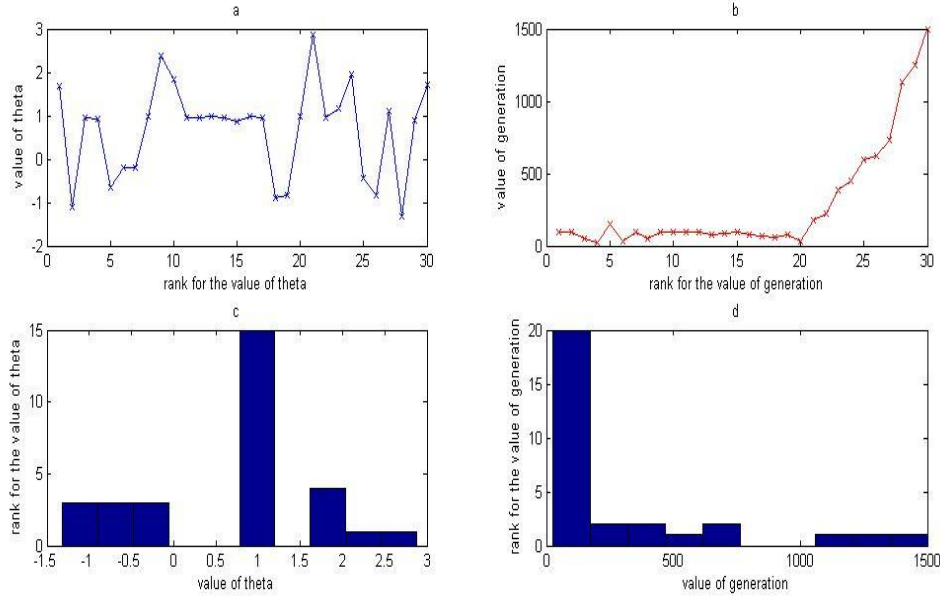
	عدد الأجيال	نوع الانتقاء	نوع التداخل الابدالي (التعابر)	نوع الطفرة	قيمة θ	*MLE **CG
18.	59	Uniform	Heuristic	Gaussian	-0.8992	-0.39596007×10^6
19.	80	Uniform	Intermediate	Gaussian	-0.8334	-0.34018509×10^6
20.	35	Uniform	Scattered	Gaussian	0.9943	-0.21300×10^1
21.	174	Roulette	Scattered	Gaussian	2.8772	$-0.17959 \times 10^{2**}$
22.	221	Uniform	Intermediate	Uniform	0.9423	-0.19132×10^1
23.	387	Uniform	Heuristic	Uniform	1.1791	-0.29955×10^1
24.	454	Roulette	Single Point	Gaussian	1.9556	-0.82397×10^1
25.	597	Stochastic Uniform	Scattered	Gaussian	-0.4373	-0.93656995×10^5
26.	619	Roulette	Single Point	Gaussian	-0.8236	-0.33950690×10^6
27.	735	Uniform	Heuristic	Uniform	1.1191	-0.26981×10^1
28.	1129	Uniform	Intermediate	Gaussian	-1.3164	-0.84873490×10^6
29.	1250	Roulette	Single Point	Uniform	0.8995	-0.17431×10^1
30.		Stochastic Uniform	Scattered	Gaussian	1.7154	-0.63398×10^1

المصدر: من اعداد الباحثان بالاعتماد على مخرجات برنامج (Matlab).
والشكل الآتي يمثل نتائج دمج دالة الترجيح ودالة قيد الجيل لتوزيع بولتزمان مع المعلمة المقدرة θ .



إذ أن:

- (a): يمثل الرسم البياني بين قيمة المعلمة θ ورتبتها، ونلاحظ أن أكثر القيم للمعلمة θ تذبذبت حول الـ ١.
- (b): يمثل الرسم البياني بين قيمة الـ MLE-CG ورتبتها، ونلاحظ أن أكثر القيم للـ MLE-CG تذبذبت حول الـ 0.
- (c): يمثل المدرج التكراري لـ (a).
- (d): يمثل المدرج التكراري لـ (b).
- المصدر من اعداد الباحثتان بالاعتماد على مخرجات برنامج Matlab.
- والشكل الآتي يمثل نتائج دمج دالة الترجيح ودالة قيد الجيل لتوزيع بولتزمان مع عدد الأجيال للخوارزمية الجينية.



إذ

a)

c)

c)

d)

المصدر من اعداد الباحثتان بالاعتماد على مخرجات برنامج Matlab.

الاستنتاجات والتوصيات

أولاً. الاستنتاجات

توصل البحث الى النتائج الاتية وكما وضحت في الجدول (١-١) وكل من الشكلين (٣-١) (٤-١):
- أعظم قيمة لمقدر الترجيح لتوزيع بولتزمان عند القيمة $(-0.16704016 \times 10^5)$ وقيمة θ هي (-0.1867) والتي ظهرت عند الجيل (100) إذ كان:

١. الانتقاء من نوع (Uniform).

٢. التداخل الابدالي من نوع (Single Point).

٣. الطفرة من نوع (Gaussian).

- أقل مقدر لقيد الجيل لتوزيع بولتزمان عند القيمة (-0.17959×10^2) وقيمة θ هي (2.8772) والتي ظهرت عند الجيل (174) إذ كان:

١. الانتقاء من نوع (Roulette).

٢. التداخل الابدالي من نوع (Scattered).

٣. الطفرة من نوع (Gaussian).

ثانياً. التوصيات

١. اقتراح استخدام طريقة مقدر الانكماش (Shrinkage Estimator Method: SEM) في المستقبل لتقدير معلمة توزيع بولتزمان ومقارنته مع الخوارزمية الجينية.

٢. اقتراح استخدام خوارزمية العلاقات المستقبلية (Future Relationships: FR) المدمجة مع طريقة الامكان الاعظم لتوزيع بولتزمان والتي يرمز لها بـ BL-FR واستخدامها في المستقبل لتقدير معلمات الطاقة للحامض النووي الرايبوسومي RNA وذلك بإدخال سلاسل جديدة من تراكيب الحامض.

٣. اقتراح استخدام الخوارزمية الجينية في تقدير معلمة الحامض النووي منقوص الاوكسجين DNA من خلال معرفة التوزيع الاحصائي الذي يتبعه.

المصادر العربية:

١. ثابت، همسة معن، (٢٠٠٥)، "بعض تطبيقات الخوارزمية الجينية في حل مسائل الأمثلية"، رسالة ماجستير، جامعة الوصل، كلية علوم الحاسبات والرياضيات، قسم الإحصاء.

٢. عبد الغني، عادل، (٢٠١٦)، "الخوارزمية الجينية"، جامعة ديالى، كلية التربية للعلوم الصرفة.

المصادر الاجنبية:

1. Andronescu, M.; Condon, A.; Hoos, H. H.; Mathews, D. H. and Murphy, K., (2006), Efficient Parameter Estimation for RNA Secondary Structure Prediction, Oxford University Press.
2. Atkins, P.W., (2010), Quanta, W.H. Freeman and Company, New York.
3. Atkins, P.W. and De Paula J., (2009), Physical Chemistry, Oxford University Press, Oxford UK.
4. Bekker, A.J.J.J and Roux, J.J.J, (2005), Communications in Statistics Theory and Methods, Boston.
5. Chaturvedi, A. and Rani, U., (1998), Journal of Statistical Research, Oxford UK.

6. Ding, Y.; Chan, C. Y. and Lawrence, C. E., (2006), **Clustering of RNA Secondary Structures with Application to Messenger RNAs**, Elsevier Ltd, Oxford University.
7. Kazmi, S., Aslam, M. and Ali, S., (2012), **SOURCE International Journal of Applied Science & Technology**, New York.
8. Landau, L. D. and Lifshitz, E. M., (1980), **Statistical Physics**, Oxford University.
9. Mcquarrie, A., (2000), **Statistical Mechanics**, University Science Books, California.
10. Skoog, D. A., Holler, F. J. and Crouch, S. R., (2006), **Principles of Instrumental Analysis**, Boston.
11. Tsoukalas, L., and Uhringm R., (1997), **Fuzzy and Neural Approaches in Engineering**, Wiley.
12. Tyagi, R.K and Bhattacharya, S.K, (1989), **Statistic**, California.
13. Zhang, H., (2003), **Genetic Algorithms**:
<http://www.cs.toronto.edu/pub/cuty/Book/AI/genetical.pdf>