Studies on Immune Clonal Selection Algorithm and Application of Bioinformatics

Hongbing Zhu\(^1\)*, Jun Wu\(^2\), Jinguang Gu\(^2\)

\(^1\) Hiroshima Kokusai Gakuin University, Japan
\(^2\) Wuhan University of Science and Technology, China

* Corresponding author’s Email: yekon7312@gmail.com

Abstract: Immune algorithms (IAs) are microscopic view of evolutionary algorithms (EAs) and applied in combinatorial optimization problems. This paper addresses to a clonal selection algorithm (CSA) that is one of the most representative IA and was applied into the protein structure prediction (PSP) on AB off-lattice model, in which the memory B cells of the CSA was innovated by employing different strategies: local search and global search in the phase of the mutation. And the CSA was further improved by adding aging operator to combat the premature convergence. However the pure aging operator didn’t achieve effective results and sometimes the optimum solution was eliminated. To resolve this problem, the current best solution was reserved by an antibody and it was not eliminated when its age reached its life span. In our experiments the improved algorithm was compared with the standard CSA and the pure aging CSA, which of the results demonstrated that the improved strategy with the memory B cells and long life aging was very effective to overcome premature convergence and to avoid trapped in the local-best solution, and it was also effective in keeping the diversity of the small size population. On the other hand, one novel hybrid algorithm Quantum Immune(QI), which combines Quantum Algorithm (QA) and Immune Clonal Selection(ICS) Algorithm, has been presented for dealing with multi-extremum and multi-parameter problem based on AB off-lattice model in the predicting 2D protein folding structure. Clonal Selection Algorithm was introduced into the hypermutation operators in the Quantum Algorithm to improve the local search ability, and double chains quantum coded was designed to enlarge the probability of the global optimization solution. It showed that the solution mostly trap into the local optimum, to escape the local best solution the aging operator is introduced to improve the performance of the algorithm. Experimental results showed that the lowest energies and computing-time of the improved Quantum Clonal Selection(QCS) algorithm were better than that of the previous methods, and the QCS was further improved by adding aging operator to combat the premature convergence. Compared with previous approaches, the improved QCS algorithm remarkably enhanced the convergence performance and the search efficiency of the immune optimization algorithm.

Keywords: Immune algorithms (IAs), Colnal selection algorithm, Protein structure prediction, AB Off-Lattice model, Aging operators, Memory B cells, Quantum Algorithm

1. Introduction

Prediction of protein folding structure is a key issue in bioinformatics [1], which helps to explain a variety of biological phenomena (e.g., the causes of disease, etc.), and can be used to predict or control of biological phenomena (e.g., to prevent and control disease). Forming protein folding is a complex evolutionary process [1]. In order to simulate the process of protein folding many simplified models had been proposed. One of the widely used models is the HP lattice-model that was proposed by Dill, et al. [2]. But the HP model
could not show the amino acid sequence of a protein and its native structure exactly. Stillinger, et al. proposed one AB off-lattice protein model [3] that uses only two types of residues, hydrophobic (A) and hydrophilic (B).

The AB model is still nontrivial to predict the native state for the protein folding problem, which of the prime cause is that the protein is a large flexible molecule system and its energy surface is a complex multi-dimension polyhedron that has many local minimum. Therefore, the global optimization problem is a sticking point of the protein structure prediction. In particular, for any N-residue of a given amino acid, the conformation is found once the energy value of the amino acid is equal to the minimum energy that is computed based on the minimum free-energy hypothesis.

Immune algorithms (IAs) are inspired by artificial immune systems, which protect the host organism against attacks from antigens and eliminate infected or old cells. Clonal selection algorithm (CSA) is one of the representative IA based on clonal selection principle and gives us one approach to solve the complex combinatorial problems.

In this paper we employed the CSA into the PSP and improved the algorithm by introducing memory B cells and aging operators in the computation. The memory B cells allowed two special mutation operators that made effective searching with local search and global search. The aging mechanism was designed to avoid premature convergence and to enforce diversity in the population during the process of the evolution. The result of our experiments showed that the improved strategies were effective for protein structure prediction on AB off-lattice model. And we proposed a novel hybrid algorithm that combines quantum algorithm and clonal selection algorithm to accurately search for the ground-state conformations of the protein. This algorithm was applied in the protein structure prediction on AB off-lattice model. In the experimental it was found that the Quantum Clonal Selection(QCS) algorithm performed better than the Clonal Selection Algorithm(CSA).

2. AB Off-Lattice Model

The AB off-lattice model utilizes two “amino acids”: hydrophobic residues and hydrophilic residues represented by the letters A and B respectively, to replace the real amino acids of the natural world. The model is composed of some successive residues by some fixed backbond and shown at Figure 1. The configuration of any n-mer is specified by the n-2 angles of bend \( \theta_2, ..., \theta_{n-1} \) at each of the nonterminal residues. Here \(-\pi \leq \theta_i < \pi, \theta_i = 0\) corresponds to linearity of successive bonds and the positive angles indicate counterclockwise rotation.

Residue species along the backbone are encoded by a set of binary variables \( \xi_i \), \( i = 1, ..., N \). When the i-th residue is A, \( \xi_i = 1 \), and when i-th residue is B, \( \xi_i = -1 \). The intramolecular potential-energy function \( \Phi \) is expressed as follows:

\[
\Phi = \sum_{i=2}^{n-1} V_1(\theta_i) + \sum_{j=i+2}^{n} \sum_{i=j}^{n} V_2(r_{ij}, \xi_i, \xi_j) \tag{1}
\]

where

\[
V_1(\theta_i) = 1/4(1 - \cos \theta_i) \tag{2}
\]

\[
V_2(r_{ij}, \xi_i, \xi_j) = 4(r_{ij}^{-12} - C(\xi_i, \xi_j)r_{ij}^{-6}) \tag{3}
\]

\( r_{ij} \) denotes distance between monomer \( i \) and \( j \) of the chain, which is written as one function of the intervening angles:

\[
r_{ij} = \left\{ \left[ \sum_{k=i+1}^{j-1} \cos \left( \sum_{l=i+1}^{k} \theta_l \right) \right]^2 + \left[ \sum_{k=i+1}^{j-1} \sin \left( \sum_{l=i+1}^{k} \theta_l \right) \right]^2 \right\}^{1/2} \tag{4}
\]

The coefficient \( C(\xi_i, \xi_j) \) is interaction between the pair residues:

\[
C(\xi_i, \xi_j) = \frac{1}{8} (1 + \xi_i + \xi_j + 5\xi_i\xi_j) \tag{5}
\]

It is apparent that the coefficient \( C(\xi_i, \xi_j) \) is +1 for an AA pair, +1/2 for an BB pair, and -1/2 for an AB pair. Consequently the first of these pairs may be regarded as strongly attracting, the second as weakly attracting, and the third as weakly repelling.
3. Immunological Clonal Selection

Clonal selection theory is an important theory of biological immune system. In 1958 the Australian scholar Burnet put forward the famous "clonal selection theory" [4]. This theory stated that immune cells select particular cells for clonal expansion. In this process the clones undergo hypermutation to generate the diversity of cells and to create higher affinity cells.

Figure 2 [5] shows the process of clonal selection that antigen selects particular lymphocytes for clonal expansion, which a hematopoietic stem cell undergoes differentiation and genetic rearrangement to produce immature lymphocytes with many different antigen receptors. Those that bind to antigens from the body’s own tissues are destroyed, while the rest mature into inactive lymphocytes. Most of these never encounter a matching foreign antigen, but those activated cells produce many clones of themselves.

In Figure 2 the activated cell proliferates, and then its cloning offspring occur somatic hypermutation (because the mutation rates are nine orders of magnitude higher than ordinary cell mutation rates). These high mutation rates increase the chance that the clones have different receptor structures from the parent, which brings about its different epitope affinities. The new B-cell clones have the opportunity to bind to pathogenic epitopes captured within the lymph nodes. If they do not bind they will die after a short time.

4. Algorithm

4.1 Immune Clonal Selection Algorithm

The clonal selection algorithm is comprised of clone, immune genetic and selection main operators [6, 7].

Table 1 depicts the pseudo-code of CSA.

The formula (6) shows the process of the algorithm.

\[ A(k) \xrightarrow{T^C} Y(k) \xrightarrow{T^C} Z(k) \bigcup A(k) \xrightarrow{T^C} A(k+1) \]  

where \( T^C \) is clone operation, which utilizes we apply dynamic clone that clones the antibodies according to their affinities and distances between the antibodies.

\( T^C \) is immune genetic operation, where the clone antibodies sets \( Y(k) \) to take place hypermutation to turn into \( Z(k) \). This operation will enable the clone antibodies to search the near neighborhood solution of \( A(k) \).

\( T^S \) is immune selection operation, which selects the best affinity antibody from \( Z(k) \) and \( A(k) \) to enter the next generation.

4.2 Improved Strategies

4.2.1 Memory B Cells

In the experiments it was found that the standard immune clonal selection algorithm was very hard to attain the exact result in contrast with the others EAs. So we employed the memory cells to divide the antibody sets into two parts and adopted different mutation methods for the two parts and the local adjustment strategy for the memory B cells to elevate the accuracy, which of the expression is as following:

\[ A(k) = [A_1(k), A_2(k), \ldots, A_n(k)] = [M_1(k), \ldots, M_r(k), N_{r+1}(k), \ldots, N_n(k)] \]  

\[ Y_i(k) = \begin{cases} M_i(k) & \text{if } i \leq r \\ N_i(k) & \text{if } r < i \leq n \end{cases} \]

where \( M_i \) and \( N_i \) denote different mutation methods respectively.

Let \( A_i(k) = \bar{\theta} = \{\theta_2, \theta_3, \ldots, \theta_{n-1}\} \)

\[ \bar{\theta} = \begin{cases} \bar{\theta} + \text{rand}(0,1)\Delta\theta & \text{if } i \leq r \\ \text{rand}(0,1)\pi & \text{if } r < i \leq n \end{cases} \]

where central moment \( \bar{\theta} = \frac{\sum_{i=1}^{r-1} \theta_i}{r} \), \( \Delta\theta = \bar{\theta}_{best} - \bar{\theta} \)
4.2.2 Aging Operator

For resolving the problem that the standard CSA was mostly trapped in the local-best solution, an aging operator was introduced into the standard CSA [8, 9]. The aging operator eliminates old B cells from the population of A to avoid premature convergence, which was implemented by extending the B cells data structure with a counter age. We defined the parameter \( T_B \) (and \( T_{BM} \) for the memory B cells \( T_B < T_{BM} \)) to set the maximum number of generations that is the time to remain in the population for the B cells.

If there is a cloning offspring better than its parents after hypermutation, it replaces its parents and sets the age to be 0. Otherwise, \( age + 1 \). When the age of a B cell is greater than \( T_B \) or the age of memory B cell is greater than \( T_{BM} \), it is eliminated from the current population, no matter what its fitness value is. Meanwhile the age value is set to be 0 to generate a random B cell replaces itself.

The above algorithm is pure aging operators and sometimes eliminated the global optimization. For this reason we improved the algorithm by retaining the antibody with best affinity. That is to say, the set of the antibody reserve one antibody that has the current best affinity as a long life span cell.

5. Experimental Results I

To evaluate the performance of the improved CSA for the protein structure prediction problem of 2-D AB off-lattice model, we employed the subclass of Fibonacci sequences [10] to test the effects by different IAs. Table 2 lists the Fibonacci sequences \( 13 \leq N \leq 55 \).

Table 2 Fibonacci sequences

<table>
<thead>
<tr>
<th>A.S.</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>( S_{13} )</td>
<td>ABBABBABABBAB</td>
</tr>
<tr>
<td>( S_{21} )</td>
<td>BABABBABABBBABABABBAB</td>
</tr>
<tr>
<td>( S_{34} )</td>
<td>ABBABBABABBABABABBABABBABABABBAB</td>
</tr>
<tr>
<td>( S_{55} )</td>
<td>BABABBABABABBABABBABABBABABBABAB</td>
</tr>
</tbody>
</table>

For the time reason, the experiments of the sequences \( 13 \) and \( 21 \) were finished with the population size \( n = 40 \), clone size \( n_c = 100 \) and the terminated parameter maximum generation \( T_{MAX} = 3000 \). All the experimental results reported in this section are average over 100 independent runs. The CSA with aging operator and memory B cells added two parameters \( T_B \) and \( T_{BM} \). In the aging clonal selection algorithm (ACSA), the aging parameters were set \( T_M = 5, T_{BM} = 10 \).

Table 3 shows the experimental results of the CSA with and without memory B cells. \( E_{MEAN} \) is average value of energy about 100 independent runs. \( E_{MIN} \) is the lowest energy obtained about the 100 independent experimental results. In the table NMB denotes the CSA without memory B cells and MB denotes with memory B cells.

<table>
<thead>
<tr>
<th>Sequence Length</th>
<th>NMB</th>
<th>MB</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( E_{MEAN} )</td>
<td>( E_{MIN} )</td>
</tr>
<tr>
<td>21</td>
<td>-4.3290</td>
<td>-5.2088</td>
</tr>
</tbody>
</table>

Table 4 shows the experimental results of the CSA with memory B cells and aging operator. In the table PA denotes pure aging operator, LA denotes long life aging operator.

Table 4 Results of the two algorithms with different aging strategy

<table>
<thead>
<tr>
<th>Sequence Length</th>
<th>PA</th>
<th>LA</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( E_{MEAN} )</td>
<td>( E_{MIN} )</td>
</tr>
<tr>
<td>21</td>
<td>-5.2560</td>
<td>-6.0177</td>
</tr>
</tbody>
</table>

It is seen from the two tables that the results of the improved algorithms are better than the original algorithms. Specially the results of the lowest energy of the clonal selection algorithm with memory B cells and long life aging operator strategy are obviously better than others.

6. Quantum Theory

In twenty century quantum mechanics is one of the most important scientific achievements. Today it still a powerful scientific field, which have being applied in many areas, such as computer and optics etc.

6.1 Single Qubit operation

The bit is the fundamental concept of classical computation and classical information. Quantum computation and quantum information are built upon an analogous concept, the quantum bit, or qubit for short. All of the Quantum computational development from the operation on the single qubit [11].

A single qubit is a vector \( |\varphi\rangle = \alpha|0\rangle + \beta|1\rangle \) parameterize by two complex numbers satisfying \( |\alpha|^2 + |\beta|^2 = 1 \). For this feature the expression for qubit can be rewritten in terms of trigonometric sines and
cosines as $|\varphi\rangle = \cos(\theta)|0\rangle + \sin(\theta)|1\rangle$. Quantum rotation gate is a important operation on a qubit. The quantum rotation gate must preserve its norm is 1, and thus is described by unitary matrices. So the quantum rotation gate can be described as following expression:

$$
U = \begin{bmatrix}
\cos(\triangle \theta) & -\sin(\triangle \theta) \\
\sin(\triangle \theta) & \cos(\triangle \theta)
\end{bmatrix}
$$

(10)

The qubit change its phase by the quantum rotation gate. The operation is follow: $|\varphi\rangle = U|\varphi\rangle = \cos(\theta + \triangle \theta)|0\rangle + \sin(\theta + \triangle \theta)|1\rangle$, U is the quantum rotation gate.

6.2 Double Chains Quantum scheme

In this paper we adopted double chain quantum scheme [12]. Here we coded the chain as the expression following:

$$
P_i = \begin{bmatrix}
\cos(\theta_1^i) & \cos(\theta_2^i) & \cdots & \cos(\theta_n^i) \\
\sin(\theta_1^i) & \sin(\theta_2^i) & \cdots & \sin(\theta_n^i)
\end{bmatrix}
$$

(11)

So one quantum solution has two solutions. One is $P_{ci} = (\cos(\theta_1^i) \sin(\theta_2^i) \cdots \sin(\theta_n^i))$ called sine solution, the other is $P_{si} = (\cos(\theta_1^i) \cos(\theta_2^i) \cdots \cos(\theta_n^i))$ called cosine solution. This scheme enlarge the probability to find the best solution. The $x_i$ is the ith qubit of the global optimum solution. As the Figure 3 shows there is four qubits satisfying with the $x_i$. Here we denoted:

Figure 3 Position of qubit i in unit circle for the optimum solution

$$
r_{ci} = \cos(\arccos(x_i)), \quad r_{si} = \cos(-\arccos(x_i))
$$

$$
r'_{ci} = \sin(\frac{\pi}{2} - \arccos(x_i)), \quad r'_{si} = \cos(\frac{\pi}{2} + \arccos(x_i))
$$

If the global optimum solution is $P = (x_1,x_2,\cdots,x_n)$, there are $2^{n+1}$ solutions exist by the quantum coded. On account of sine solution $P_{si}$ correspond with $2^n$ sines solutions, and cosine solution$P_{ci}$ correspond with $2^n$ cosines solutions.

The scheme makes the global solutions achieve exponential growth, and remarkable increase the probability to find optimum solution.

7. Quantum Clonal Selection Algorithm

The quantum algorithm has better global search ability but not good at local search, and the clonal selection algorithm has great advantage at search local optimum but it convergence speed is very slow [6]. Combining the two algorithm advantage we got the QCS algorithm, which had powerful local searching ability and quick convergence speed.

In the algorithm the problem is defined as antigen, the solutions are antibodies. The aim is to find the best antibody by its affinity, the evaluation function of the affinity is potential-energy, and the population is the antibodies set. The individual of the antibodies set is coded by expression (6).

7.1 Improved Strategy

It was found that in experiment the algorithm could trap in the local optimum. So the aging operation was introduced in this algorithm to escape the local-best solution [8, 9].

The aging operator eliminates old $B$ cells from the population of $A$ to avoid premature convergence, which was implemented by extending the $B$ cells data structure with a counter age. We defined the parameter $T_B$ (and $T_{BM}$ for the memory $B$ cells $T_B < T_{BM}$) to set the maximum number of generations that is the time to remain in the population for the $B$ cells.

If there is a cloning offspring better than its parents after hypermutation, it replaces its parents and sets the age to be 0. Otherwise, $age + 1$. When the age of $B$ cell is greater than $T_B$ or the age of memory $B$ cell is greater than $T_{BM}$, it is eliminated from the current population, no matter what its fitness value is. Meanwhile the age value is set to be 0 to generate a random $B$ cell replaces itself.

The above algorithm is pure aging operators and sometimes eliminated the global optimization. For this reason we improved the algorithm by retaining the antibody with best affinity. That is to say, the set of the antibody reserve one antibody that has the current best affinity as a long life span cell.

7.2 The Algorithm

The steps in detail of the Quantum Clonal Selection Algorithm based on AB off-lattice model is follow:

Step 1 Initialize the start population and the age of the individual, the terminal condition.

Step 2 Calculate the affinity of the each individual in the population, and sort the population by the affinity.
Step 3 Clone the population, each individual clone itself and have several offspring. The offspring inherit its father’s age add.

Step 4 Base on the father affinity the offspring occur hypermutation, which the higher fit antibodies search, to generate new antibodies which differ with its father.

Step 5 Calculate the offspring affinity: If there is a cloning offspring better than its parents after hypermutation, it replaces its parents to update the population and sets the age to be 0, else save its father’s antibody but age add 1.

Step 6 If there any antibodies of the population reach its life span, generate a new antibodies to instead of the old one and the age set 0.

Step 7 Judging the requirement whether satisfies the terminal condition. If not jump to Step 2.

8. Experimental Results II

We restricted the same Fibonacci sequences as Table 5 for comparison. The Fibonacci sequence is defined recursively by

\[ S_0 = A \quad S_1 = B \quad S_{i+1} = S_{i-1} \ast S_i \]

Where * is the concatenation operator. The lengths of the sequences are given by Fibonacci numbers \( N_{i+1} = N_{i-1} + N_i \). The first few sequences are \( S_2 = AB, S_3 = BAB, S_4 = ABBAB \), etc. Hydrophobic residues A occur isolated along the chain, while hydrophilic residues B occur either isolated or in pairs.

<table>
<thead>
<tr>
<th>Sequence</th>
<th>Length</th>
<th>( E_{\text{mean}} )</th>
<th>( E_{\text{min}} )</th>
<th>( E_{\text{mean}} )</th>
<th>( E_{\text{min}} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>21</td>
<td>-5.2855</td>
<td>-6.1447</td>
<td>-6.0051</td>
<td>-6.1879</td>
<td></td>
</tr>
</tbody>
</table>

In the experiments, the main parameters were set as following: the population size was 40, clone size was 100, the memory cells’ age set 10, normal cells’ age set 5.

To evaluate the effect of the algorithm, it compared with the CSA [12]. Here two sets of data were given from experiment. The average and minimum energy value was listed in Table 6, along with the Clonal S-election Algorithm(CSA) and Quantum Conal Algorithm(QCS). It was saw that both the average and minimum energy value obtained by the QCS were better than the CSA form Table 6.

The Figure 4 showed the convergence rate of the CSA and the QCS. The dashed line denoted the CSA, and dotted line was the QCS. In the picture the curve indicated that the convergence rate of the QCS was faster than the CSA.

9. Conclusion

In this paper the immune clonal selection algorithm was applied for PSP on 2-D AB off-lattice model. The standard clonal selection algorithms and improved clonal selection algorithms were compared in the PSP. From the experiment, the IA with memory B cells and aging operator obviously outperforms standard IAs. And it is very effective to overcome local-best solution or premature. The results proved that the IA with memory B cells and aging operator is advantage over the standard IA in AB off-lattice model for PSP.

In the experiment it was found that the parameters, like \( T_B, T_{BM}, n, n_c, T_{MAX} \), influenced accuracy of the results. So, it’s worthwhile to analyze the impact of these parameters in the future. Moreover, introducing other EAs to improve the IAs’ efficacy is also considered.

And the improved quantum clonal selection algorithm was applied for PSP on 2-D AB off-lattice model in this paper. The proposed algorithm could deal
with multi-extremum and multi-parameter problems. In the proposed algorithm, different strategies were adopted to make the proposed algorithm had different advantages. For examples, the aging operation strategy could keep the diversity of the population, and the aging operation strategy made it possible to accept poor solution as the current solution and thus made the algorithm had better hill-climbing capability and stronger local searching capability than many other mutation operators. In addition, double chains strategy could enlarge probability of the optimum, and thus could also search more solution space in once calculation. Compared with the previous algorithms, QCS had stronger capability of global searching. In the future work, we will improve the algorithm and make it more effective for long protein sequence prediction using multi-core computing platforms.

Acknowledgment

This work was partially supported by a grant from the NSF (Natural Science Foundation) of educational agency of Hubei Prov. under grant number D20091106 and the NSF of China under grant number 60803160.

References


