論文題目　　　　Peptides Designed with Genetic Algorithms
　　　　　　　　（遺伝的アルゴリズムを用いて設計したペプチド）

氏　　名　　　　張　　武　明

The goal of this research is to develop and implement a methodology for the design of peptides with predefined structure. This thesis research established an evolutionary *de novo* peptide design system based on the genetic algorithms (GAs) and chemical synthesis method, and furthermore, investigated the impact of GAs on the design of peptide with predefined structure. While the irrational design provides a flexible formulation for the design of peptides with predefined structure, the GAs implement the optimization search of the combinatorial space thereof. However, the design of peptide with predefined structure is a challenging problem to the GA-based method. How to effectively convert the design goal into GA representation is the first question. How to efficiently obtain the target peptides is the second. In order to solve these problems, an experimental system was established on helical conformations due to their simple structure, diverse oligomers, and easy determination and interpretation using circular dichroism (CD) spectrum.

In Chapter 1, the motivation, goal, and significance of this research was proposed. They are to develop and implement a methodology for the design of peptides with predefined structure. This thesis research thus investigated the impact of GAs on the design of peptide with predefined structure, and established an evolutionary *de novo* peptide design system based on the GAs and chemical synthesis method.

The most important and basic concepts related to protein folding and design, the evolutionary algorithms, especially the GAs, and , furthermore, evolutionary design were then introduced. After brief review of work done by previous researchers, the irrational design problem was formalized as a combination of the evolutionary algorithm with experimental measurement, in addition to discussing the evaluation of protein properties and basic operations of the GAs for irrational protein design. In this thesis research, the challenging problems were how to effectively convert the design goal into GA representation and how to efficiently obtain the target peptides.

An experimental system based on helical model structure was created to solve these problems because the helical conformations possess the characteristics, such as, simple structure, diverse oligomers, easy determination and interpretation using CD spectrum, etc. The solutions to

the challenging problems were thus induced on the helical peptide design system: conversion of the design goal into a feasible sequence space for GA search; establishment of a new GA enabling to efficiently search the sequence space; construction of a more efficient and robust peptide design system comprising the above two constituents. Following that, it provided a discussion of the objective, scope, and contributions for this research effort. This chapter ended by outlining the contents of the remaining chapters.

In Chapter 2, a method converting irrational design problem satisfying design goal into GA representation was proposed. Different from all other GA applications, the objective in this research was a structure, not a parameter. Some description of the structure has to be used as the fitness function, but none can tell one structure from all other structures. It means that using a description of peptides GAs can evolve the fitness evaluation of the peptides in full search space, but cannot identify the detailed structures of the peptides in the search process. Pre-selection of the areas in the full search space provides a way to solve this problem. The search space was first divided into feasible areas, where the structures of peptides satisfy the design goal, and infeasible areas, where the structures of the peptides possibly violate the design goal. The peptides with predefined structure can thus be searched by GA in the constrained sequence space or the feasible areas in full sequence space. A method based on model structure of peptides was proposed to create constrained sequence space satisfying both the design goal and GA requirements.

The monomeric helix was selected as the model structure of the design problem. After a literature review of *de novo* design of monomeric helical peptides, a corresponding search space comprising only three amino acids Glu, Lys, and Ala was established by the developed method. The helicity determined by CD spectrum was used as the fitness evaluation of peptides. Furthermore, a simple GA was selected to the evolutionary search of the helix. Consequently, a practical peptide design system was established by combining the GA representation of monomeric helical peptides with a simple GA search method.

Sixteen helical peptides with the helicity better than the model peptide were obtained through a four-generation evolution with this system. This method was successfully evaluated by the attempt to the monomeric helical peptide design using the simple GA. The results indicated that the GA showed its efficacy and efficiency in searching for the target peptides, and furthermore, that the helicity combining with the feasible sequence space for monomeric helix provided an effective evaluation of the status of peptide pools in the search process. The statistical analysis revealed that the primary determinants of the monomeric helix were the electrostatic attractions between the oppositely charged residues Glu and Lys spaced three or four residues. Although this monomeric helical peptide design was used as a case study, the design principles might be applied to other kinds of problems as well. Indeed, the chapter could be seen as a guide for GA usage in peptide design.

In Chapter 3, a method of improving the efficiency of peptide design was proposed. The major limitation of the evolutionary method developed in this thesis research was the experimental effort required to the determination of the structure of the peptides. Whereas an ill setting often results in either a low efficiency in design or poor quality solutions at the end of the GA run, the GAs set at suitable parameter setting can fulfill their work with less number of generations and smaller size of populations. Thus, the minimization of the experimental effort in peptide design could be reached by exploring for a specific GA for the peptide design.

A numerical structure characterization based on physico-chemical properties of amino acids was established to evaluate the structure of *de novo* sequences. The performance of GAs could thus be evaluated by measuring their evolutionary speed in the numerical fitness landscape of this peptide design problem. With this benchmark, the task of setting GA parameters, such as, crossover rates, mutation rates, step numbers of multi-step variation, and population sizes, could be handled by the simulations. New operations: multi-step crossover-mutation or multi-step variation (msv) and elitist strategy called SeedLibrary, were created by modifying the genetic operations in simple GA to perform a more efficient and more robust exploration of target peptides than that of simple GA, even with a small population size. The GAs containing these new operations was called msvGAs in this research.

As results of simulations, a suitable multi-step variation operation and corresponding parameter setting were identified. Closer inspection of the developed algorithms and their performance revealed that more robust and more powerful search was realized with the proposed msvGA and the suitable parameter setting. Therefore, an improved GA, the msvGA set at the identified configuration and parameter setting, was achieved. This GA characterized less sensitive to parameter setting and more efficient in the search of target peptides than the simple GA, thus providing a way to search the sequence space with both the minimized size of population and minimized number of generations. In addition, different from almost all previous studies, where GAs were set by the simulations on function optimization, here, the msvGAs was evaluated by the simulations on the virtual protein design problem. The suitable parameter setting identified in this study provided a suitable parameter setting to an analogous protein design problem, the real peptide design system, according to the principle of "parameter setting by analogy".

In Chapter 4, the observation of the feasible sequence space in monomeric helical peptide design in Chapter 2 and the examination of the improved GA designed by an analogy problem in Chapter 3 were further extended to the experiment of parallel homodimeric coiled-coil peptide design. This extension allows simultaneous evolution of complexity structures of peptides even with a small population size.

The feasible sequence space for this design was established by using the general information about the structure, which had been studied by previous researchers. As the results of synthesis experiment, twenty homodimeric coiled-coil peptides having the helicity more than that

of the referent peptide GCN4p, the model structure of this study, were obtained through a five-generation evolution. Therefore, these results verified the effect of the improved GA in this real design of peptides with parallel homodimeric helical topology. Results also indicated that the improved GA did indeed make the design system more efficient and robust than that using a simple GA, and demonstrated that the target peptides were efficiently achieved even with only a half of the population size used by a simple GA in the monomeric helical peptide design.

The size-exclusion chromatography HPLC experimental results furthermore showed that the established homodimeric helical peptide search space effectively excluded the infeasible sequences, which could not be determined by only a description of helix, the CD signal. Thus, a combination of the fitness evaluation of peptide structure, which was used in GA evolution, with the constrained search space, which provides the search range for the GA, can perform an effective evaluation of peptide structure for the GA.

Furthermore, with the statistical analysis of the sequence patterns in the evolutionary process the dynamics of the GAs were explained. The GAs converge the determinants of structure at very different speeds; thus, the significant determents are converged in a few generations and the target peptides could be obtained in these generations. These statistics of the homodimeric helical peptides showed that the important intra-helical and inter-helical interactions were converged in only five generations. Among them, the most important interactions were those between the oppositely charged amino acids Glu or Gln and Lys or Arg. The statistical results furthermore emphasized the importance of the orientation of the interactions and spacing between the interacting residues.

In Chapter 5, the thesis was concluded with the summary of the research effort, including a discussion of the benefits provided by the proposed methods. The effort converting design goal into constrained sequence space overcome the potential infeasibility of the sequences in full combinatorial sequence space, thus providing a platform for structure design using GAs. Furthermore, the efficiency of the peptide design was improved by a GA containing multi-step crossover-mutation operation, new elitist strategy, and suitable parameter setting. The evolutionary *de novo* design system comprising these identified constituents could be used to design the peptide having specific structure more effectively and more efficiently. The chapter finally listed the suggestions for future work in the same direction.