学　位　論　文　要　旨

No. 1

コース名	生物環境科学	学籍番号	05D3010	氏　名	吉田　昼也

題目

Generation and Characterization of the Basic and Advanced Molecular Libraries in Evolutionary Molecular Engineering.
(進化分子工学における基本および高次分子ライブラリーの作成とその特性の研究)

要旨

Evolutionary molecular engineering usually adopts a strategy of repeating rounds of library construction and selection. Since there is a limitation in the number of molecules to be dealt with, we have to design the initial library carefully. Thus, how a library should be constructed is an essential problem in evolutionary molecular engineering. This is closely related to the landscape of the sequence space. Thus, I made two approaches to this problem: practical one and theoretical one.

In Chapter I entitled as "Molecular Design Guided by A Local Map of Sequence Space: DNA Aptamers that Inhibit Cathepsin E", a novel method to design functional molecules based on the local sequence map, which was drawn based on the acquired information (sequence-activity relationship of oligonucleotides), was introduced regarding cathepsin E (CE)-inhibitory DNA aptamers. This method was a kind of application of principal coordinate analysis (PCA) and the two coordinates selected here were found to be very effective to be used for drawing the local landscape. The current finding must be useful in various cases since *in vitro* selection experiments based on a library usually provide a set of molecules that have a common property for which the selection was aimed and enable us to draw a local sequence space map. The information on the landscape could be obtained from this method but not from the mere sequence alignment, which must be significant.

In Chapter II entitled as "GFPs of Insertion Mutation Generated by Molecular Size-altering Block Shuffling", the effect of insertion and deletion (*indel*) on the GFP activity (fluorescence) was made clear. To prepare such *indel* libraries, Y-ligation-based block shuffling (YLBS) technology was exploited. Some insertion mutations (GFP sequence between 164 and 181) found not to damage the fluorescence activity of GFP while no functional mutants were found in the deletion mutations. The clear-cut result that the GFP is insertion-tolerant but deletion-vulnerable (although this comes from a limited amount of data) has a profound meaning in both structural science and the sequence space problem of proteins. YLBS approach must be powerful in surveying a wide range of sequence space with a big or moderate stride, which is different from those of point mutation approaches.

コース名	生物環境科学	学籍番号	05D3010	氏　　名	吉田　昼也

In Chapter III entitled as " *In Vitro* Evolution High-powered: Demonstration by Acquisition of Cathepsin E-specific Inhibitors of Peptide", the development of a novel rapid molecular evolution system, eRAPANSY, is dealt. In this study, a library of octa-peptides were constructed using YLBS and was panned for cathepsin E-inhibitory peptides. As the first library, a random collection of octa-peptides were used (as the whole number of this library is 20^8 ($\approx 2.6\times10^{10}$), we can walk the entire sequence space exhaustively at least theoretically). Next, the block shuffling in the ASAC mode was performed using the peptides obtained from the primary library selection. As a result of the selection experiment using this library, peptides of higher activity were obtained successfully. The acquisition of smaller peptides by the selection of the secondary library (ASAC) is significant from the technological viewpoint. To obtain more compact molecules is constantly required in the field of drug discovery. In this study, the secondary library was constructed by expanding the sequence space [(variable 8 a.a. + constant 20 a.a.) → (variable 28 a.a.)]. As another type of the secondary library, “peptide pair” was proposed in our laboratory. This library can be constructed by combining peptides which had been selected in the preceding experiments thus proved to have abilities to bind the target molecule (in our experiment, cathepsin E) with an activity of either inhibition or activation or so. This method provides not only the fittest molecule for a particular function but also the information on he surface structure of the target molecule. We can make use of the activity information on the paired peptides to infer the surface structure based on the logic explained here. Exhaustive survey of the surface structure using peptide pairs enables us to take another approach to the study on proteins. Once we can get a 3-D structure of a particular set of protein and peptide inhibitor employing such technology, we may be able to determine the finding structures of the other peptides depending on the *epitope mapping* presented here and the molecular dynamic approach, thus saving a lot of energy required for the structural analysis of biomolecule interactions.

The block design problem is a symbolic one in this field, which had been mainly dealt in constructing the secondary or higher libraries in this study: ASAC (Chapter III) and pep-pair. However, this does not exclude the possibility that we will face with this block design problem also in constructing the primary library, which is usually built up without intimate information on desired molecules. In essence, blocks can be thought to be entities which compactly hold the information acquired through selections. In this study, this fact was discussed in relation to the landscape of the sequence space, where sequence made of blocks appear at a fixed distance (as lattice point) with a stride of the size of a block. To tackle this problem concretely, I employed a Molecular Dynamics approach (Chapter IV). That is, in order to implement a favorable nature to a starting library, I and my collaborators developed a computational method to predict the nature of peptide/protein folding in a realistic time scale and resources, finding that the nano-second MD simulation of peptides was sufficient to reveal the α-helix-forming propensity and others. Its usage for EME lies in the application of this knowledge and method to design initial blocks for the shuffling.

コース名	生物環境科学	学籍番号	05D3010	氏　　名	吉田　昼也

All of the knowledge obtained in *in vitro* evolution can be stored in the form of the sequence space landscape just as done in my study in Chapter I. Therefore, the concept of the sequence space is doubly beneficial: i) providing design principles and ii) serving as a platform to accumulate a huge amount of selection data (by forming a more and more precise landscape). Throughout my study, sequence space and block-shuffling were the key concepts technically or theoretically.

Recently, how proteins have evolved through evolution can be dealt with a voluminous amount of genome sequence data. It has already established that immune-related proteins such as Ig, TCR, MHC, and various cytokines have been constructed through domain shuffling. In the evolution of proteins, even smaller units (blocks) seem to have played important and active roles in boosting up the rate of evolution, also this still leaves a lot to be established. In this thesis, I discussed both of the basic and advanced libraries regarding to how to construct and how to use the information obtained from the selection of these libraries, which is important to create novel functional molecules.