Molecular Immunocomputing with Application to Alphabetical Pattern Recognition Mimics the Characterization of ABO Blood Type

Su Dong Kim
Biointelligence Lab
Interdisciplinary Program in Cognitive Science
Seoul National University
Seoul 151-742, Korea
sdkim@bi.snu.ac.kr

Ki-Roo Shin
Biointelligence Lab
School of Computer Science & Engineering
Seoul National University
Seoul 151-742, Korea
krshin@bi.snu.ac.kr

Byoung-Tak Zhang
Biointelligence Lab
School of Computer Science & Engineering
Seoul National University
Seoul 151-742, Korea
btzhang@bi.snu.ac.kr

Abstract - We propose the concept of *molecular immunocomputing* which is a kind of peptide computing. Molecular immunocomputing is basically implemented by direct antigen-antibody biomolecular recognition on the basis of Fab fragments diversity. In this paper, we consider how molecular immunocomputing can tell the alphabet "O" from the characters "A" and "B" similar to the characterization of ABO blood type. To implement molecular immunocomputing, the two-dimensional figures are coded on the one-dimensional DNA strings. The informations coded on the virtual DNA strings are transcribed to virtual RNA sequences, and then translated to the polypeptide sequences. The resulting peptide sequences are artificially synthesized and coupled to the carrier proteins. The resulting conjugate proteins are injected as input antigens to immunize the experimental animals. After immunization, we could purify the corresponding antibodies. These antibodies can be arrayed onto the protein microarray chips. The recent developments in the field of protein microarrays show the feasibility of molecular immunocomputing.

1 Introduction

Generally, the term *memory* implies the capacity to encode, store, and retrieve information. The variety of approaches that have been taken to the memory are – cognitive, neuropsychological, neurobiological, and computational (Schacter & Tulving, 1994).

In the field of immunology, the term (immunological) *memory* is the attribute of the immune system mediated by memory cells whereby a second encounter with an antigen induces a heightened state of immune reactivity (Goldsby *et al*., 2000).

The immune system processes peptide patterns using mechanisms that in some cases correspond closely to existing algorithms for processing information (e.g., genetic algorithm), and it is capable of exquisitely selective and well-coordinated responses, in some cases responding to fewer than ten molecules. Some of the techniques used by the immune system include learning (affinity maturation of B cells, negative selection of B and T cells, and evolved biases in the germ-line), memory (cross-reactivity and the secondary response), massively parallel and distributed computations with highly dynamic components (on the order of 10^8 different varieties of receptors and 10^7 new lymphocytes produced each day), and the use of combinatorics to address the problem of scarce genetic resources (V-region libraries) (Forrest and Hofmeyr, 2000).

From an information-processing perspective, the immune system is a highly parallel intelligent system (Dasgupta, 1999). Therefore, immunocomputing has been considered as a new computing approach replicating the principles of information processing by proteins and immune networks (Tarakanov *et al.*, 2003).

In the immune system, pattern recognition is implemented as binding. When an immune system detector (including B cells, T cells, or antibodies) binds to a peptide, we say that the immune system has recognized the pattern encoded by the peptide (Forrest & Hofmeyr, 2000).

It was natural that someone could solve pattern recognition problems by immunocomputing, because natural immune system had solved peptide pattern recognition problems. Tarakanov and Skormin had already considered this problem (Tarakanov & Skormin, 2002). In this work they took a formal immune system, on the notion of *formal protein* and *formal immune network*.

Although there is no generally accepted the precise concept of biomolecular computing (Rambidi, 1997), it can be defined as computations performed by biomolecules (McCaskill, 2000). In 1994, Adleman had used to solve computational problems by DNA hybridization (Adleman, 1994). From this ground-breaking experiment sprang the rapidly evolving field of DNA computation (Reif, 2002). In a similar way antibodies which specifically recognize peptide sequences can be used for calculation (Hug & Schuler, 2001).

In Watson-Crick DNA base-pairing, the purine bases (A or G) can form hydrogen bonds with pyrimidine bases (C or T). Despite the relative weakness of the hydrogen bonds holding the base pairs together, each DNA molecule contains so many base pairs that the complementary chains never spontaneously separate under physiological conditions (Watson & Crick, 1953).

If DNA is exposed to near-boiling temperatures, so many base pairs fall apart that the double helix separates into its two complementary chains (Watson *et al.*, 1992). The traditional DNA computing methods have depended on the relatively fixed interrelationships between 4 different DNA bases via hydrogen bonds.

Living things contain not only nucleic acids (DNA and RNA) but also proteins, polysaccharides, and lipids. In this paper, our discussion would be focused on proteins, especially on antibodies, with application to computation.

Amino Acid	Three Letter Symbol	One Letter Symbol
Alanine	Ala	A
Asparagine *or* aspartic acid	Asx	B
Cysteine	Cys	C
Aspartic acid	Asp	D
Glutamic acid	Glu	E
Phenylalanine	Phe	F
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Lysine	Lys	K
Leucine	Leu	L
Methionine	Met	M
Asparagine	Asn	N
Proline	Pro	P
Glutamine	Gln	Q
Arginine	Arg	R
Serine	Ser	S
Threonine	Thr	T
Valine	Val	V
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Glutamine *or* glutamic acid	Glx	Z

Table 1: The three and one letter symbols for the "standard" amino acid residues [Adapted from Voet & Voet, 1990].

All proteins are composed of the 20 "standard" amino acids (Table 1). The immunoglobulins, the major proteins of the immune system, form an essential biological defense system in higher animals. Antibody is a protein (immunoglobulin), consisting of two identical heavy chains and two identical light chains, that recognizes a particular epitope on an antigen and facilitates clearance of that antigen. Antigen is any substance that binds specifically to an antibody or a T-cell receptor. The interaction between an antibody and an antigen depends on four types of noncovalent forces: (1) hydrogen bonds, in which hydrogen atom is shared between two electronegative atoms; (2) ionic bonds between oppositely charged residues; (3) hydrophobic interactions, in which water forces hydrophobic groups together; and (4) van der Waals interactions between the outer electron clouds of two or more atoms (Fig. 1). The exquisite specificity of antigen-antibody interactions has led to the development of a variety of immunologic assays. These assays can be used to detect the presence of either antibody or antigen and have played vital roles in diagnosing diseases, monitoring the level of the humoral immune response, and identifying molecules of biological or medical interest (Goldsby *et al.*, 2000).

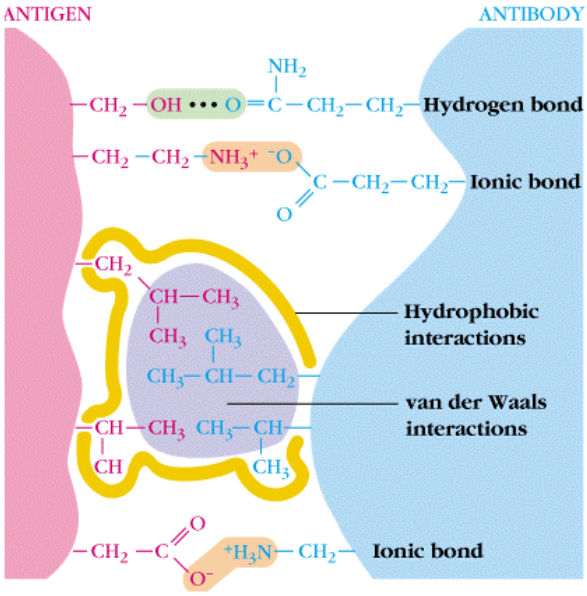

Figure 1: The noncovalent interactions that form the basis of antigen-antibody (Ag-Ab) binding. [Adapted from Goldby *et al.* (2000)].

Tarakanov and his colleagues applied immuno-computing method to immunoassay-based diagnostic arrays (Tarakanov *et al.*, 2002). In this work, they detected the interactions between protein G (pG) with immunoglobulin G (IgG). Immunoglobulins bind the bacterial pG by means of their Fc-fragment. Therefore, the pG-IgG interacting system cannot utilize the potential applicability of the specific antigen-antibody recognition through Fab fragments (Fig. 2).

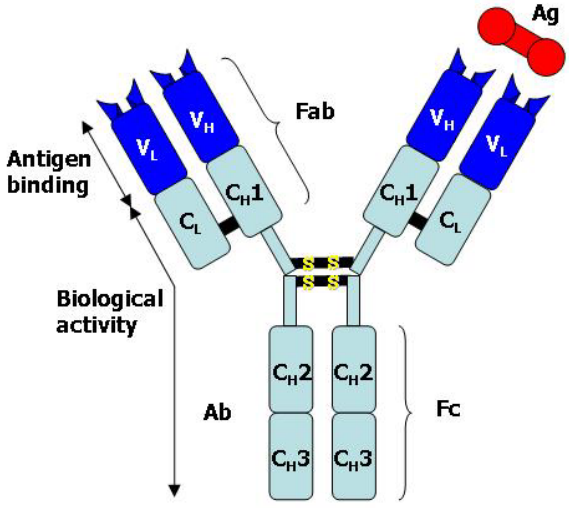

Figure 2: Structure of an antibody with its antigen.

In this paper, we propose the concept of *molecular immunocomputing*. Molecular immunocomputing is a kind of peptide computing which was introduced by Hug and Schuler (Hug & Schuler, 2001). We show that numeral pattern recognition problem can be applied with molecular immunocomputing method, directly based on real immune systems. This method is a kind of computation implemented by bio-molecules, specifically antigens and antibodies.

2 Molecular Immunocomputing Mimics the Characterization of ABO Blood Type

Molecular immunocomputing is inspired by both DNA computing and immunology. Molecular immunocomputing is basically implemented by molecular antigen-antibody recognition. In principle, it attempts to mimic the characterization of ABO blood type (Fig. 3).

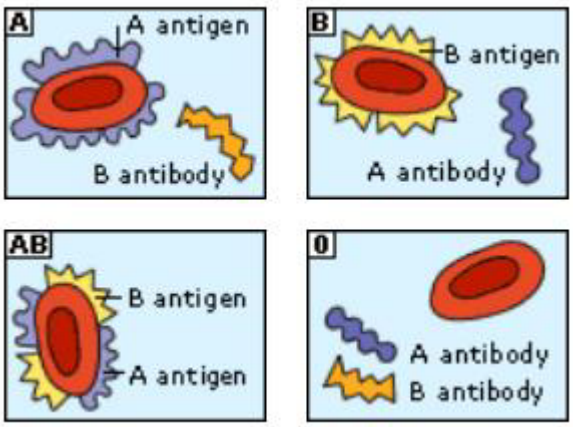

Figure 3: The ABO blood grouping system.
[Adapted from http://www.nobel.se/medicine/educational/landsteiner/readmore.html]

Pattern recognition is the basic function performed by immune system based on the features of recognizing antigen by antibodies (Tarakanov & Skormin, 2002). The molecular recognition between an antibody and an antigen involves various noncovalent interactions between the antigenic determinant, or epitope, of antigen and the variable-region (V_H / V_L) domain of the antibody molecule (Fig. 2), particularly the hypervariable regions, or complementarity-determining regions (CDRs). Although antigen-antibody reactions are highly specific, in some cases antibody elicited by one antigen can cross-react with an unrelated antigen. Such cross-reactivity occurs if two different antigens share an identical epitope or if antibodies specific for one epitope also bind to an unrelated epitope possessing similar chemical properties.

Cross-reactivity is often observed among poly-saccharide antigens that contain similar oligosaccharide residues. The ABO blood group antigens, for example, are glycoproteins expressed on red blood cells, or erythrocytes. Subtle differences in the terminal sugar residues distinguish the A and B blood group antigens (Fig. 4). An individual lacking one or both of these antigens will have serum antibodies to the missing antigen(s). A type O individual thus has anti-A and anti-B antibodies; a type A individual has anti-B; and a type B individual has anti-A (Fig. 3).

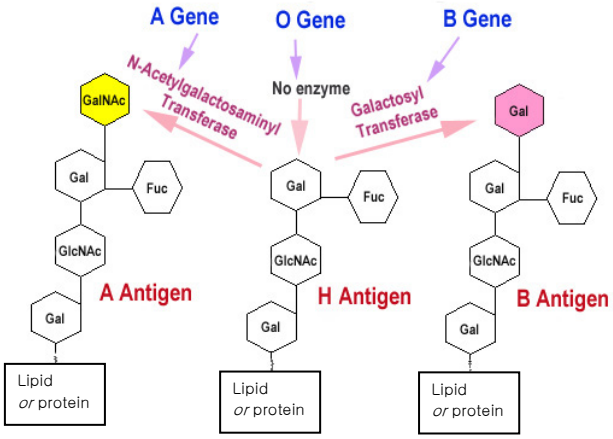

Figure 4: The structure of terminal sugars in the A, B, and O blood antigens. (Gal; Galactose, GlcNAc; *N*-acetylglucosamine, Fuc; Fucose, and GalNAc; *N*-acetylgalactosamine)

Antibodies to the A, B, and O antigens on erythrocytes are called isohemagglutinins. If a type A individual is accidentally transfused with blood containing type B cells, the anti-B isohemagglutinins will bind to the B blood cells and mediate their destruction by means of complement-mediated lysis. The ABO system provides a simplified model of the concept that immune recognition of genetically foreign components plays a role in graft rejection (Goldsby *et al*., 2000).

The ABO blood group is clinically the most important blood group system (Yip, 2000). The ABO blood group system was discovered by Karl Landsteiner at the beginning of the 20th century. He received the 1930 Nobel Prize for Physiology or Medicine for his discovery of the major blood groups and his development of the ABO system of blood typing, which made blood transfusion a routine medical practice (Owen, 2000). However, it was only in 1990 that the nucleotide sequence of the ABO gene was determined (Yamamoto *et al.*, 1990), and in 1995 that its genomic organization was elucidated (Yamamoto *et al.*, 1995).

Since the ABO gene was cloned and the molecular basis of the three major alleles delineated about 10 years ago, the gene has increasingly been examined by a variety of DNA-based genotyping methods and analyzed in detail by DNA sequencing. There is extensive sequence heterogeneity underlying the major ABO alleles that produce normal blood groups A, B, AB, and O when in correct combination with other alleles. There are over 70 ABO alleles reported to date and these alleles highlight the extensive sequence variation in the coding region of the gene. Excluding the common alleles, about half of the remaining alleles are due to new mutations and the other half can better be explained by intragenic recombination

(both crossover and gene conversion) between common alleles (Yip, 2002). Although polymerase chain reaction-restriction enzyme fragment length polymorphism (PCR-RFLP) (Lee & Chang, 1992) and allele-specific PCR (Ugozzoli & Wallace, 1992) methods are useful for detecting variations in the recognition sequences of certain restriction enzymes or specific primers, they cannot necessarily be used to detect alleles with unknown substitutions in the amplified DNA fragments. However, polymerase chain reaction-single-strand conformation polymorphism (PCR-SSCP) method can detect point mutations at various positions in the amplified fragments, and it can be used to determine the ABO genotypes as well as to detect undescribed ABO alleles (Ogasawara *et al.*, 1996).

Based on DNA sequence diversities, 6 billion people can be classified with the ABO blood group. Similarly, various two-dimensional patterns featured on pixel images coded by DNA sequences can be assigned to the groups with restricted numbers such as the alphabetical character sets or the numeral pattern sets. In molecular immunocomputing, pattern recognition processes mimic the characterization of ABO blood type.

3 Methods

For example, we can consider the alphabetical character recognition by molecular immunocomputing. The implementation of molecular immunocomputing proceeds with following two stages.

3.1 Designing the peptide sequences

We consider two-dimensional (2-D) 7×7 pixel images of the alphabetical character sets. Each pixel is coded by one base of the DNA nucleotide set: $B = \{A, C, G, T\}$. There are four types of DNA nucleotides, distinguished by, and denoted by, their bases, written A, C, G, and T (Sella & Ardell, 2002).

We consider the standard operating template. The template is composed of 7×7 pixel grids. Each pixel has four kinds of DNA code operators (Table 2).

Operator	Operation
W	*Transition*
X	*Matched-transversion*
Y	*Mismatched-transversion*
Z	*Back-mutation*

Table 2: Four kinds of DNA code operators.

There are two subsets of DNA nucleotide set B: $Y = \{C, T\}$, called *pYrimidines*, and $R = \{A, G\}$, called *puRines*. The mutations within these sets are called *transitions*, and the mutations between them are called *transversions*. Therefore, four different transitions and eight different transversions are possible (Gardner *et al.*, 1991). The "*matched-transversions*" could be defined as the transversion between the common Watson-Crick base-pairs. Then the "*mismatched-transversions*" could be defined as the transversion between the rare mismatched AC and GT base-pairs caused by tautomeric shifts (Harris *et al.*, 2003). The *back-mutation* occurs at the same site in the mutated sequence as the original mutation, restoring the original nucleotide sequence (Gardner *et al.*, 1991).

After overlapping the 2-D pixel image coded by DNA bases with the standard operating template, each pixel undergoes a kind of DNA code mutation (Table 2). We select one-dimensional (1-D) DNA string from this mutation result.

The DNA string can be virtually transcribed to the corresponding RNA transcript. To the 3'-end position of RNA transcript, guanine (G) and uracil (U) are added. As a result, the 3'-end codon of RNA transcript is assigned to cysteine.

The RNA transcript can be virtually translated to peptide sequence according to the universal genetic code (Fig. 5). A noncanonical genetic code might be considered as an alternative choice (Keeling & Doolittle, 1997).

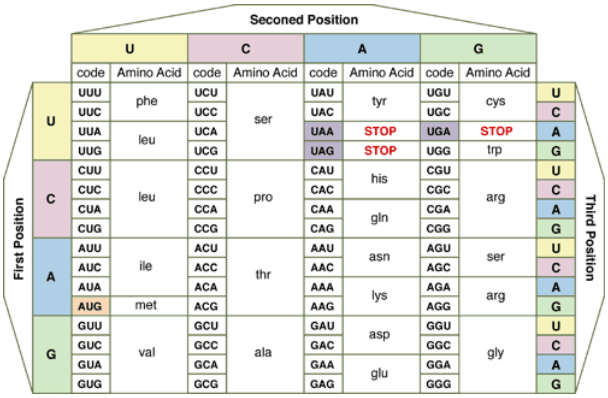

Figure 5: The universal genetic code.

The resulting peptide should be used for the immunogen. The peptides containing hydrophilic amino acids and proline residues would be preferable (Harlow & Lane, 1988).

3.2 Purifying the corresponding antibodies

We can artificially synthesize the peptide in the previous stage from a number of commercial sources. We can couple the synthetic peptides as haptens with the carrier proteins. By itself, a hapten cannot function as an immunogenic epitope. But when a multiple molecules of a single hapten are coupled to a carrier protein, or nonimmunogenic homopolymer, the hapten becomes accessible to the immune system and often functions as the immunodominant determinant (Goldsby *et al.*, 2000).

The two most commonly used carrier proteins are keyhole limpet hemacyanin (KLH) and bovine serum albumin (BSA). Sulfhydryl groups found on cysteine side chains are used for coupling to the carrier proteins.

We can immunize some experimental animals. The five most commonly used laboratory animals are rabbits, mice, rats, hamsters, and guinea pigs. Rabbits are chosen for the routine production of antisera. They are easy to keep and handle, can be safely and repeatedly bled, and the antibodies they produce are well characterized and

easily purified (Fig. 6).

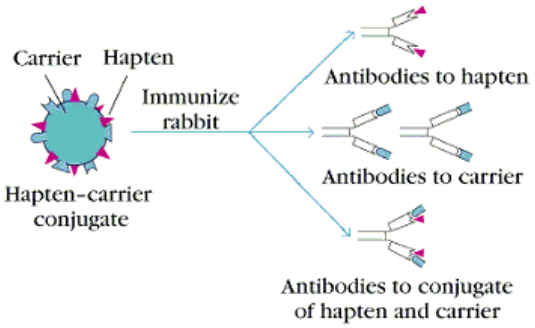

Figure 6: Hapten-carrier conjugation and immunization. [Adapted from Goldby *et al.* (2000)].

Once a good titer has developed against the antigen of interest, regular boosts and bleeds are performed to collect the maximum amount of serum. Antibodies from serum can be purified using conventional methods involving precipitation and column chromatography (Harlow & Lane, 1988).

3.3 Preparation of protein microarray chips

Microarray-based assays using nucleic acid-nucleic acid interactions (DNA chips) are well established and protein microarrays are just becoming popular (Templin *et al.*, 2002). There are two general types of protein microarrays. Firstly, analytical microarrays in which antibodies, antibody mimics or other proteins are arrayed and used to measure the presence and concentrations of proteins in complex mixtures. Secondly, functional protein micro-arrays, in which sets of proteins or even an entire proteome are prepared and arrayed for a wide range of biochemical activities.

The most common form of analytical arrays are antibodies/antibody mimic arrays in which antibodies (or similar reagents) that bind specific antigens are arrayed on a glass slide at high density. A lysate is passed over the array and the bound antigen is detected after washing. Detection is usually carried out by using labeled lysates or using second antibody that recognizes the antigen of interest (Zhu & Snyder, 2003).

The antibody arrays can be prepared as previously reported (Haab *et al.*, 2001a). Antibodies and antigens that are provided in glycerol solutions were transferred to a glycerol-free, phosphate-buffered saline (PBS) solution (137 mM NaCl, 2.7 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.4). A robotic device can be used to print hundreds of specific antibody or antigen solutions in an array on the surface of poly-$_L$-lysine coated microscopic slides. Two complex protein samples, one serving as a standard for comparative quantitation, the other representing an experimental sample in which the protein quantities are to be measured, could be labeled by covalent attachment of spectrally resolvable fluorescent dyes. The conjugate protein solutions and dye (such as Cy3 and Cy5) solutions are prepared. Each experimental protein mixture is labeled with the dye Cy5 (red fluorescence) and then mixed with the Cy3-labelled (green fluorescence) 'reference' mixture. The reference mixture contains each of the same sample proteins at a constant concentration. The variation across the microarrays in the red-to-green (R/G) ratio measured for each antibody or antigen spot should reflect the variation in the concentration of the corresponding binding partner in the set of mixes (Fig. 7; Haab *et al.*, 2001a).

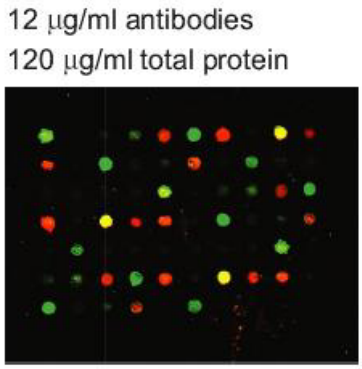

Figure 7: An example of antibody microarray data. [Adapted from Haab *et al.* (2001a)]

4 Discussion

To design a peptide sequence coded for a computational problem, we adopt the DNA computing concept implemented by nature. We choose the four kinds of mutation mechanisms as the transforming functions of DNA base-shifting operations.

Various structural features of some kinds of peptide-antibody complexes are compared as follows (Fig. 8).

Cholera toxin peptide 3 (CTP3) is a 15-residue peptide (VEVPGSQHIDSQKKA) corresponding in sequence to an immunogenic loop on the surface of the B-subunits of both cholera toxin and the heat-labile toxin from *Escherichia coli*. (Fig. 8a: Shoham, 1993). The Fab structure of an antibody PC283 complexed with its corresponding 15-residue peptide antigen, PS1 (HQLDPAFGANSTNPD), derived from the hepatitis B virus surface antigen was also determined (Fig. 8b: Nair *et al.*, 2000). The anti-Syrian hamster prion protein (SHaPrP) monoclonal Fab 3F4 alone, as well as the complex with its cognate peptide epitope (SHaPrP 104-113; KPKTNMKHMA), have been determined to atomic resolution (Fig. 8c: Kanyo *et al.*, 1999). The molecular surface of the antigen binding sites is highly specific to the corresponding peptide (Fig. 8d: Lescar *et al.*, 1997).

We could assign, for example, the peptide epitope SHaPrP 104-113 as an alphabet character "O," the CTP3 as "A," and PS1 as "B." Then we are able to implement a pattern classification by molecular immunocomputing. The given peptides could be modified (*e. g.*, *N*-linked glycosylation in A, B, and O antigens), and regarded as the characters representing alternative type of font (*e. g.*, "$\mathcal{A}$," "$\mathcal{B}$," and "$\mathcal{O}$," and so forth). Each corresponding antibody can bind to both its modified and unmodified peptide partners with different affinities. This method

could be applicable to the pattern classification problem with handwritten characters.

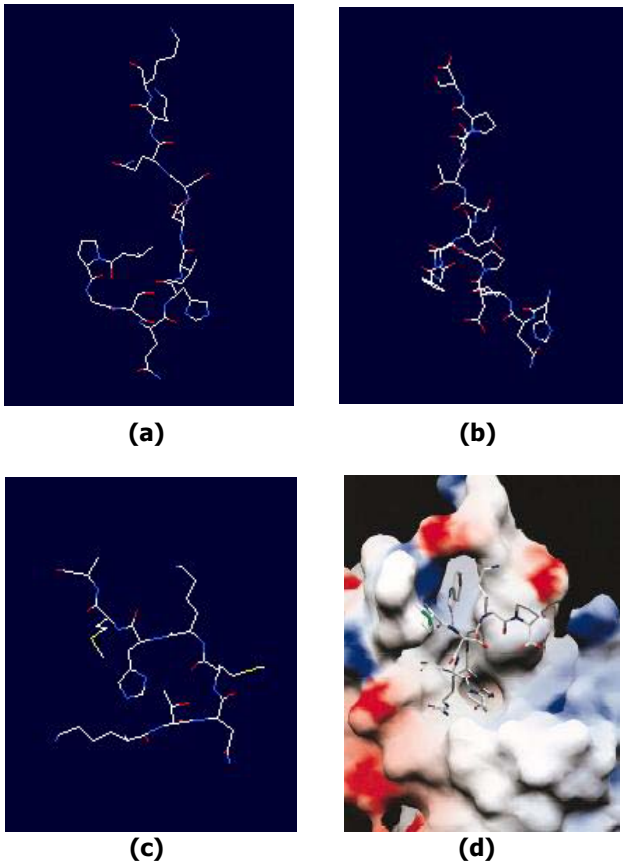

Figure 8: Structures of the peptide CTP3 (a), PS1 (b), SHaPrP (c), and peptide with antigen binding site (d). [(a), (b), and (c): drawed by the Swiss-PdbViewer (Guex & Peitsch, 1997); (d): adapted from Nicholls *et al.*, 1991.]

Antibodies are the traditional reagent of choice for detecting proteins in complex mixtures. However, polyclonal antibodies are often not specific and are expensive to produce, and the conventional hybridoma method of producing highly specific monoclonal (derived from a single cell) antibodies is also time-consuming, laborious and costly. Recently, alternative methods, such as phage antibody-display, ribosome display, systematic evolution of ligands by exponential enrichment (SELEX), mRNA display, and affibody display, have been developed to expedite the production of antibodies and/or antibody mimics (Haab *et al.*, 2001b). All of these approaches involve the construction of large repertoires of viable regions with potential binding activity, which are then selected by multiple rounds of affinity purifications. The binding affinity of the resulting candidate clones can be further improved using maturation strategies. However, the ideal selection system, which is not only fast, robust, sensitive, and of low cost, but automated and minimized, is yet to be fully developed (Templin *et al.*, 2002).

Although the number of different antibody combining sites the immune system generate may be difficult to calculate with any degree of precision, the potential antibody diversity would be well above 10^{11} (Goldsby *et al.*, 2000). In one of the largest studies to date, Sreekumar *et al.* spotted 146 distinct antibodies on glass (Sreekumar *et al.* 2001). Proteins are often present in a very large dynamic range (10^6); thus, reagents that might have high affinity for one protein, but are low affinity for another will still exhibit detection of the lower affinity protein if it is much more prevalent. Haab *et al.* have investigated the ability of 115 well-characterized antibody-antigen pairs to react in high-density microarrays on modified glass slides (Haab *et al.*, 2001a). 30 % of the pairs showed the expected linear relationships, indicating that a fraction of the antibodies were suitable for quantitative analysis.

To avoid this problem, many groups have turned to using sandwich assays, in which the first antibody is spotted on the array and then the antigen is detected with a second antibody that recognizes a different part of the proteins. This approach dramatically increases the specificity of the antigen detection, but required that a least two high-quality antibodies exist for each antigen to be detected (Zhu & Snyder, 2003). Sandwich immunoassays were miniaturized and parallelized and performed in a micorarray format (Templin *et al.*, 2002). This was recently demonstrated by the parallel determination of different cytokine levels in biological samples with high specificity and sensitivity (Huang, 2001).

Microarray immunoassays are of general interest for all diagnostic applications where several parameters of one sample have to be analyzed in parallel (Mendoza *et al.*, 1999). Accurate quantification with protein microarrays can be achieved by including positive and negative control spots and/or internal calibration spots. This will finally lead to robust and reliable diagnostic assays (Templin *et al.*, 2002).

5 Conclusion

The advantages of peptide computing in comparison to a DNA computing are that at each position 20 different building blocks instead of 4 are possible. More than 20 amino acids that are used in the living world can be used for the peptide computing, *e. g.* $_D$-amino acids and amino acids which are not encoded by the genetic code. Therefore, the number of possible peptide sequences increases dramatically (Hug & Schuler, 2001).

In contrast to a DNA strand which has only one possible reverse complement of the same length, in peptide-antibody interaction:

1. different antibodies can recognize different regions or structures of the same epitopes,
2. the binding affinity of different antibodies recognizing the same epitope can be different (Balan *et al.*, 2002), and
3. same antibody can recognize different antigens with different affinity (Foote, 2003).

These factors add additional flexibility for doing computation with peptides.

Meanwhile, immunocomputing approaches based on mathematical model recently implemented via biochip to achieve biomolecular computers (Tarakanov *et al.*, 2002).

There might be some contact points between the peptide computing and the immunocomputing with bioarrays. We hope that our molecular immunocomputing concept could mediate to connect these two types of computing.

Acknowledgments

This research was supported in part by the Ministry of Education & Human Resources Development under the BK21-IT Program, the Ministry of Commerce, Industry and Energy through the MEC project, and the NRL Program from Korean Ministry of Science & Technology. The ICT at Seoul National University provided research facilities for this study.

Bibliography

L. M. Adleman. (1994) Molecular computation of solutions to combinatorial problems. *Science* 266. 1021-1024.

M. S. Balan et al. (2002) Peptide computing – universality and complexity. in N. Jonoska and N. Seeman (ed.) *DNA Computing: revised papers / 7th International Workshop on DNA Based Computers, DNA 7, Tampa, FL, USA, June 10 – 13, 2001*. 290-299. Berlin: Springer-Verlag.

D. Dasgupta. (1998) (ed.) *Artificial Immune Systems and Their Applications*. Berlin: Springer-Verlag.

J. Foote. (2003) Isomeric antibodies. *Science* 299. 1327-1328.

S. Forrest and S. A. Hofmeyr. (2000) Immunology as Information Processing. In L.A. Segel and I. R. Cohen (ed.), *Design Principles for Immune System & Other Distributed Autonomous Systems*. 361-387. Oxford: Oxford Univ. Press.

E. J. Gardner *et al.* (1991) *The Principles of Genetics*. New York: John Wileys & Sons.

R. A. Goldsby, *et al*. (2000) *Kuby immunology*. New York: W. H. Freeman & Co.

N. Guex & M. C. Peitsch. (1997) SWISS-MODEL and Swiss-PdbViewer: an environment for comparative protein modeling. *Electrophoresis* 18 (15). 2714-2723.

B. B. Haab *et al.* (2001a) Protein microarrays for highly parallel detection and quantitation of specific proteins and antibodies in complex solutions. *Genome Biol.* 2 (2). Research0004.

B. B. Haab *et al.* (2001b) Advances in protein microarray technology for protein expression and interaction profiling. *Curr. Opin. Drug Discov. Devel.* 4 (1). 16-23.

E. Harlow and D. Lane. (1988) *Antibodies: A Laboratory Manual*. New York: Cold Spring Harbor Laboratory.

V. H. Harris *et al.* (2003) The effect of tautomeric constant on the specificity of nucleotide incorporation during DNA replication: support for the rare tautomer hypothesis of substitution mutagenesis. *J. Mol. Biol.* 326 (5). 1389-1401.

R. P. Huang. (2001) Simultaneous detection of multiple proteins with an array-based enzyme-linked immunosorbent assay (ELISA) and enhanced chemiluminescence (ECL). *Clin. Chem. Lab. Med.* 39. 209-214.

H. Hug and R. Schuler. (2001) Strategies for the development of a peptide computer. *Bioinformatics* 17 (4). 364-368.

Z. F. Kanyo *et al.* (1999) Antibody Binding Defines a Structure for an Epitope that Participates in the $PrP^{C} \rightarrow PrP^{Sc}$ Conformational Change. *J. Mol. Biol.* 293 (4).855-63.

P. J. Keeling and W. F. Doolittle (1997) Widespread and ancient distribution of a noncanonical genetic code in diplomonads. *Mol. Biol. Evol.* 14 (9). 895-901.

J. C. Lee and J. G. Chang. (1992) ABO genotyping by polymerase chain reaction. *J. Forensic Sci.* 37. 1269-1275.

J. Lescar *et al.* (1997) Three-dimensional structure of an Fab-peptide complex: structural basis of HIV-1 protease inhibition by a monoclonal antibody. *J. Mol. Biol.* 267. 1207-1222.

J. McCaskill. (2000) Biomolecular computing. *ERCIM News No.* 43. [available at http://www.ercim.org/publication/Ercim_News/enw43/mc_caskill1.html]

L. G. Mendoza *et al.* (1999) High-throughput microarray-based enzyme-linked immunosorbent assay (ELISA). *Biotechniques* 27. 778-788.

D. T. Nair *et al.* (2000) Crystal structure of an antibody bound to an immunodominant peptide epitope: novel features in peptide-antibody recognition. *J. Immunol.* 165. 6949-6955.

A. Nicholls *et al.* (1991) Protein folding and association: insights from the interfacial and thermodynamic properties of hydrocarbons. *Proteins* 11 (4). 281-296.

Nobel e-Musium. (2003) Blood Groups, Blood Typing and Blood Transfusions. [available at http://www.nobel.se/medicine/educational/landsteiner/readmore.html]

K. Ogasawara *et al.* (1996) Extensive polymorphism of ABO blood group gene: three major lineages of the alleles for the common ABO phenotypes. *Hum. Genet.* 97. 777-783.

R. Owen. (2000) Karl Landsteiner and the first human marker locus. *Genetics* 155. 995-998.

N. G. Rambidi. (1997) Biomolecular computer: roots and promises. *Biosystems* 44. 1-15.

J. Reif. (2002) Computing. Successes and Challenges. *Science* 296. 478-479.

D. L. Schacter and E. Tulving. (ed.) (1994) *Memory Systems 1994*. Cambridge: The MIT Press.

G. Sella and D. H. Ardell. (2002) The impact of message mutation on the fitness of a genetic code. in L. F. Landweber and E. Winfree (ed.) *Evolution as Computation: DIMCS Workshop, Princeton, January 1999*. 140-159. Berlin: Springer-Verlag.

M. Shoham. (1993) Crystal structure of an anticholera toxin peptide complex at 2.3 Å. *J. Mol. Biol.* 232 (4). 1169-1175.

A. Sreekumar, *et al.* (2001) Profiling of cancer cells using protein microarrays: discovery of novel radiation-

regulated proteins. *Cancer Research* 61. 7585-7593.
A. O. Tarakanov and V. A. Skormin. (2002) Pattern Recognition by Immunocomputing. *Proceedings of the 2002 Congress on Evolutionary computation*. 938-943.
A. O. Tarakanov, *et al.* (2002) Immunocomputing for Bioarrays. *Proceedings of the 1st International Conference on Artificial Immune Systems 2002*. 32-40.
A. O. Tarakanov, *et al.* (2003) *Immunocomputing: Principles and Applications.* Berlin: Springer Verlag.
M. F. Templin, *et al.* (2002) Protein microarray technology. *TRENDS in Biotechnology* 20 (4). 160-166.
L. Ugozzoli and R. B. Wallace. (1992) Application of an allele-specific polymerase chain reaction to the direct determination of ABO blood group genotypes. *Genomics* 12. 670-674.
D. Voet and J. G. Voet. (1990) *Biochemistry.* New York: John Wiley & Sons.
J. D. Watson and F. H. C. Crick. (1953) General implications of the structure of deoxyribonucleic acid. *Nature* 171. 964-967.
J. D. Watson *et al.* (1992) *Recombinant DNA.* New York: W. H. Freeman & Co.
F. Yamamoto *et al.* (1990) Molecular genetic basis of the histoblood group ABO system. *Nature* 345. 229-233.
F. Yamamoto *et al.* (1995) Genomic organization of human histo-blood group ABO genes. *Glycobiology* 5 (1). 51-58.
S. P. Yip. (2000) Single-tube multiplex PCR-SSCP analysis distinguishes 7 common ABO alleles and readily identifies new alleles, *Blood* 95 (4). 1487-1492.
S. P. Yip. (2002) Sequence variation at the human ABO locus, *Ann. Hum. Genet.* 66. 1-27.
H. Zhu and M. Snyder. (2003) Protein chip technology. *Current Opinion in Chemical Biology* 7. 55-63.