Specific Amino Acid Substitutions Improve the Activity and Specificity of an Antimicrobial Peptide & Serodiagnosis by Immunosignature: a Multiplexing Tool for Monitoring the Humoral Immune Response to Dengue

by

Xiao Wang

A Thesis Presented in Partial Fulfillment
of the Requirements for the Degree
Master of Science

Approved April 2013 by the
Graduate Supervisory Committee:

Stephen Albert Johnston, Chair
Joseph Blattman
Charles Arntzen

ARIZONA STATE UNIVERSITY

August 2013

ABSTRACT

Random peptide microarrays are a powerful tool for both the treatment and diagnostics of infectious diseases. On the treatment side, selected random peptides on the microarray have either binding or lytic potency against certain pathogens cells, thus they can be synthesized into new antimicrobial agents, denoted as synbodies (synthetic antibodies). On the diagnostic side, serum containing specific infection-related antibodies create unique and distinct "pathogen-immunosignatures" on the random peptide microarray distinct from the healthy control serum, and this different mode of binding can be used as a more precise measurement than traditional ELISA tests. My thesis project is separated into these two parts: the first part falls into the treatment side and the second one focuses on the diagnostic side.

My first chapter shows that a substitution amino acid peptide library helps to improve the activity of a recently reported synthetic antimicrobial peptide selected by the random peptide microarray. By substituting one or two amino acids of the original lead peptide, the new substitutes show changed hemolytic effects against mouse red blood cells and changed potency against two pathogens: *Staphylococcus aureus* and *Pseudomonas aeruginosa*. Two new substitutes are then combined together to form the synbody, which shows a significantly antimicrobial potency against *Staphylococcus aureus* (<0.5uM).

In the second chapter, I explore the possibility of using the 10K Ver.2 random peptide microarray to monitor the humoral immune response of dengue. Over 2.5 billion people (40% of the world’s population) live in dengue transmitting areas. However, currently there is no efficient dengue treatment or vaccine. Here, with limited dengue

patient serum samples, we show that the immunosignature has the potential to not only distinguish the dengue infection from non-infected people, but also the primary dengue infection from the secondary dengue infections, dengue infection from West Nile Virus (WNV) infection, and even between different dengue serotypes. By further bioinformatic analysis, we demonstrate that the significant peptides selected to distinguish dengue infected and normal samples may indicate the epitopes responsible for the immune response.

ACKNOWLEDGMENTS

Here I would like to thank many people without whom I could not complete my thesis work successfully. First of all, I would like to particularly thank my advisor and committee chair Dr. Stephen Albert Johnston for his continuous help and support for my studies. Without him, I did not have the chance to do this project and got good training at the Center of Innovative Medicine (CIM, Biodesign Institute). I also want to thank my committee members, Dr. Charles Arntzen, Dr. Yixin Shi and Dr. Joseph Blattman, for their insightful suggestions and comments on my project.

In addition, I sincerely appreciate the help and guidance from the people at CIM. Thanks Dr. Chris Diehnelt, Dr. Andrey Loskutov and Dr. Valeriy Domenyuk, for their direct help in the study of antimicrobial peptides. Thanks Dr. Zbigniew Cichacaz, Dr. Bart Legutki and Dr. Phillip Stafford, for their tremendous patience in teaching me microarray-related techniques. Thanks Dr. Nidhi Gupta and Dr. Zhan-gong Zhao, for their help with all my chemical synthesis. Thanks our lab technician John Lainson for all his technical support.

Finally, I would like to thank my labmates and classmates, Kurt Whittemore, Hu Duan, Josh Richer, Dr. Rebecca Halperin, Krupa Navalkar, Dr. Luhui Shen, for their useful discussions and critical talks. They kindly shared with me lots of their experiences on doing research. Without their help, I could not get familiar with so many techniques so quickly.

TABLE OF CONTENTS

Page

LIST OF TABLES

LIST OF FIGURES

Chapter 1

AN AMINO ACID SUBSTITUTION LIBRARY HELPS TO IMPROVE A RECENTLY REPORTED SYNTHETIC ANTIMICROBIAL PEPTIDE

1.1 Introduction

Antibiotic resistance is a continuing problem in antimicrobial drug treatment. However, new antibiotic development is always lagging behind bacteria evolution. As a result, more and more antibiotic-resistant strains create the risk of uncontrollable infection spreading. Our lab recently developed a random peptide microarray based high-throughput technique for screening for new antimicrobial candidates. Our overall goal is to develop antimicrobial agents, particularly ones that would be bacterial specific. The idea is that bacterial specific agents would slow the spread of microbial resistance and spare "good" bacteria. Here we show that one of the peptide candidates we selected previously can be further improved by a mutant library approach.

In order to improve the de novo synthetic antimicrobial peptide, RWRRHKHFKRPHRKHKRGSC, we created a single substitution peptide library (containing 304 peptides). We have shown that by single amino acid replacement, selected peptide variants have a minimum inhibitory concentration range (MICrange) of 5 μ M-10 μ M against *S.aureus* (SA) and 10 μ M->40 μ M against *P.aeruginosa* (PA). The hemolytic activity (%hemolysis) of the peptide candidates exhibits a range from 4% to 100% cell lysis. We further showed that by adding the single mutants together into double mutants, we observed additive effects on their inhibitory potency against both SA and PA. Finally, the selected mutants were combined together into a class of affinity agents called synbodies, which had MIC less than 1 μ M against SA.

As a follow-up procedure of our microarray-based synthetic antimicrobial peptides discovery system, we have demonstrated the peptide mutant library as an effective method to improve our selected candidates with more antimicrobial potency, more specificity and less hemolytic activity.

In addition, we also tested the concept of using a random peptide as the antibody recruiting molecule (ARM) to increase the macrophage-bacteria interaction. We constructed the molecules called synbody effectors, which consisted of a bacteria binding arm and an antibody recruiting arm. The initial results indicated that synbody effectors were able to increase the macrophage absorption of pathogenic *S.aureus*.

1.2 Background and significance

1.2.1 Antibiotics and drug-resistance

"No action today, no cure tomorrow", claimed the WHO World Health Day 2011, accurately stressing the problem of antimicrobial resistance. On the one hand, more and more multi-drug resistant strains, such as multi-drug-resistant TB (MDR-TB)[1], extensively drug-resistant TB (XDR-TB) and Methicillin/Vancomycin-resistant *Staphylococcus aureus* (MRSA/VRSA)[2] are continually being identified these years. On the other hand, fewer and fewer new antibiotics have been introduced since the 1990s[3]. Multiple mechanisms exist in bacteria to confer drug-resistance. For example, the resistance against tetracycline can be caused by the active efflux system located on the bacteria membrane; the resistance against streptomycin can be explained as the chemical inactivation by the enzyme produced in the bacteria. The development of drug resistance is almost inevitable after the introduction of any new antibiotic, especially for those

targeting redundant metabolic pathways. Both vertical and horizontal gene transfer contribute to the wide-spread appearance of acquired antibiotic resistance. As a result, we are pushed to the edge of developing new antimicrobial treatments.

1.2.2 Antimicrobial peptides (AMPs)

Antimicrobial peptides (AMP), as a class of unconventional antibiotics, presently represent a promising solution due to their unique antimicrobial potency and the rare development of drug-resistance[4,5]. Natural antimicrobial peptides are host-defense molecules with diverse structures and functions produced by all living organisms, including bacteria, fungi, plants, invertebrates, vertebrates as well as human beings.

Most natural antimicrobial peptides contain 20~50 amino acids with a net positive charge and 40%~60% hydrophobicity. A typical α -helical antimicrobial peptide has a membranlytic mode of action: the positively charged AMPs are first attracted by the negatively charged bacteria surface (G+: the peptidoglycan layer; G-: outer membrane), then they assemble on the bacteria cytoplasmic membrane with different possible structures (classical pores; toroidal pores; carpet mode). The insertion of AMPs causes the direct rupture of the bacteria membrane, has allowing little chance for the bacteria to develop drug resistance.

According to the Antimicrobial Peptide Database (APD, http://aps.unmc.edu/AP/main.php)[6], 1664 antibacterial peptides have been reported (09/2012), among them, 799 are from fishes and amphibians, 180 are from insects, 161 are from bacteria, 90 are from plants, 49 are from human sources, and only 27 are synthetic.

1.2.3 Synthetic antimicrobial peptides selection by random peptide microarray

Natural antimicrobial peptides have limitations/disadvantages, such as high toxicity, low specificity and special structures, which make chemical synthesis difficult. Many efforts have been made to create synthetic antimicrobial peptides to overcome these shortcomes. Generally speaking, there are two kinds of approaches. One is a rational peptide sequence design based on the comparison of already demonstrated antimicrobial peptides, especially for the α -helical class. The new synthetic peptides are usually generated from one or two selected peptide scaffolds with limited amphipathicity and charge alterations[7,8]. Another approach is to select the random peptides with antimicrobial capacity from a phage display library. A successful example was shown by Alessandro Pini for his selection of anti-*E.coli* peptides[9]. Although a phage peptide library contains a large number of variants with diverse structures, the technique itself is relatively complex to perform.

Our group recently developed a random peptide microarray-based technique to screen for new bacteria-binding and lytic peptides in an efficient way. The basic concept is to apply dye-labeled bacteria cells directly onto the microarray. The peptides which can specifically bind with the bacteria cells will show high fluorescent signals and thus be selected. Since all the peptides on the microarray are well-characterized prior to the spotting and screening, it saves a lot of time for the identification step. We show that the whole selection procedure can be finished in less than two hours, which might be quite useful in some urgent circumstances, for example in the face of unknown pathogens.

1.2.4 Synthetic antibody and synbody effectors

Antibody therapy against microbes is an old idea, used even before the widespread usage of antibiotics. Passive antibody treatment has been shown useful in many infectious diseases, such as Anthrax (*Bacillus anthracis*), Hepatitis B (*Hepatitis* B virus), Pneumonia (*Streptococcus pneumoniae*), etc. However, among 15 FDA-approved mAbs (1986-2004), only one mAb, palivizumab, has been licensed for the infectious disease (RSV). One of the major disadvantages of antibody-based therapies is the high cost of drug development, production and administration[10].

In order to avoid the high cost of natural antibodies production, our group previously came up with the idea to chemically synthesize a class of agents called synthetic antibodies (synbodies), which can bind with selected targets with high affinity. A typical synbody molecule is produced by two relatively low affinity peptides (selected from the random peptide microarray) joined together with a scaffold to achieve the high affinity (based on the equation $Kd = Kd1 \times Kd2$)[11]. The targets of a synbody could be either protein (for example, TNF-α [12]) or the intact bacteria/ virus cells (for example, influenza virus, Norovirus and *Staphylococcus aureus*, unpublished results)

In addition, in order to mimic the immunomodulatory function of natural antibodies, such as the Antibody-Dependent Cell-mediated Cytotoxicity (ADCC) and Complement Dependent Cytotoxicity (CDC), we intend to further engineer the synbody into an Antibody-Recruiting Molecule (ARM). ARM can be separated into two parts, one part can bind with specific pathogen cells, and the other part can recruit self-antibodies to exert certain biological functions[13]. This idea has been proved by Parker et al in their design for targeting the HIV gp120[14]. Here, as a first demonstration, I showed the

possibility of adding a modulatory peptide arm to the synbody to make it a synbody effector.

1.3 Materials and methods

1.3.1 Peptides synthesis and substitution library creation

A

		1R	2W	3R	4R	5H	6K	7H	8F	9K	10R	11P	12H	13R	14K	15H	16K	17R
1	A	M1	M21	M41	M61	M81	M101	M121	M141	M161	M181	M201	M221	M241	M261	M281	M301	M321
2	R	M2	M22	M42	M62	M82	M102	M122	M142	M162	M182	M202	M222	M242	M262	M282	M302	M322
3	N	M3	M23	M43	M63	M83	M103	M123	M143	M163	M183	M203	M223	M243	M263	M283	M303	M323
4	D	M4	M24	M44	M64	M84	M104	M124	M144	M164	M184	M204	M224	M244	M264	M284	M304	M324
5	d-K	M5	M25	M45	M65	M85	M105	M125	M145	M165	M185	M205	M225	M245	M265	M285	M305	M325
6	E	M6	M26	M46	M66	M86	M106	M126	M146	M166	M186	M206	M226	M246	M266	M286	M306	M326
7	Q	M7	M27	M47	M67	M87	M107	M127	M147	M167	M187	M207	M227	M247	M267	M287	M307	M327
8	G	M8	M28	M48	M68	M88	M108	M128	M148	M168	M188	M208	M228	M248	M268	M288	M308	M328
9	H	M9	M29	M49	M69	M89	M109	M129	M149	M169	M189	M209	M229	M249	M269	M289	M309	M329
10	I	M10	M30	M50	M70	M90	M110	M130	M150	M170	M190	M210	M230	M250	M270	M290	M310	M330
11	L	M11	M31	M51	M71	M91	M111	M131	M151	M171	M191	M211	M231	M251	M271	M291	M311	M331
12	K	M12	M32	M52	M72	M92	M112	M132	M152	M172	M192	M212	M232	M252	M272	M292	M312	M332
13	M	M13	M33	M53	M73	M93	M113	M133	M153	M173	M193	M213	M233	M253	M273	M293	M313	M333
14	F	M14	M34	M54	M74	M94	M114	M134	M154	M174	M194	M214	M234	M254	M274	M294	M314	M334
15	P	M15	M35	M55	M75	M95	M115	M135	M155	M175	M195	M215	M235	M255	M275	M295	M315	M335
16	S	M16	M36	M56	M76	M96	M116	M136	M156	M176	M196	M216	M236	M256	M276	M296	M316	M336
17	T	M17	M37	M57	M77	M97	M117	M137	M157	M177	M197	M217	M237	M257	M277	M297	M317	M337
18	W	M18	M38	M58	M78	M98	M118	M138	M158	M178	M198	M218	M238	M258	M278	M298	M318	M338
19	Y	M19	M39	M59	M79	M99	M119	M139	M159	M179	M199	M219	M239	M259	M279	M299	M319	M339
20	V	M20	M40	M60	M80	M100	M120	M140	M160	M180	M200	M220	M240	M260	M280	M300	M320	M340

B

ORI(original)	[H]RWRRHKHFKRPHRKHKR[OH]
MUT155	[H]RWRRHKHPKRPHRKHKR[OH]
MUT83	[H]RWRRNKHFKRPHRKHKR[OH]
MUT165	[H]RWRRHKHFd-KRPHRKHKR[OH]
MUT114	[H]RWRRHFHFKRPHRKHKR[OH]
MUT157	[H]RWRRHKHTKRPHRKHKR[OH]
MUT110	[H]RWRRHIHFKRPHRKHKR[OH]
MUT153	[H]RWRRHKHMKRPHRKHKR[OH]
MUT120	[H]RWRRHVHFKRPHRKHKR[OH]
MUT22	[H]RRRRHKHFKRPHRKHKR[OH]
MUT222	[H]RWRRHKHFKRPRRKHKR[OH]
Comb(155&83)	[H]RWRRNKHPKRPHRKHKR[OH]
Comb(165&114)	[H]RWRRHFHFd-KRPHRKHKR[OH]
Comb(165&83)	[H]RWRRNKHFd-KRPHRKHKR[OH]
Comb(157&83)	[H]RWRRNKHTKRPHRKHKR[OH]
Comb(165&110)	[H]RWRRHIHFd-KRPHRKHKR[OH]
Comb(157&110)	[H]RWRRHIHTKRPHRKHKR[OH]
Comb(153&83)	[H]RWRRNKHMKRPHRKHKR[OH]
Comb(165&120)	[H]RWRRHVHFd-KRPHRKHKR[OH]
Comb(153&114)	[H]RWRRHFHMKRPHRKHKR[OH]
Comb(22&222)	[H]RRRRHKHFKRPRRKHKR[OH]
Comb(222&114)	[H]RWRRHFHFKRPRRKHKR[OH]
Comb(22&114)	[H]RRRRHFHFKRPHRKHKR[OH]
Comb(22&114&222)	[H]RRRRHFHFKRPRRKHKR[OH]

C

ORI(original)	[Ace]RWRRHKHFKRPHRKHKR[NH2]
MUT83	[Ace]RWRRNKHFKRPHRKHKR[NH2]
MUT165	[Ace]RWRRHKHFd-KRPHRKHKR[NH2]
MUT114	[Ace]RWRRHFHFKRPHRKHKR[NH2]
MUT157	[Ace]RWRRHKHTKRPHRKHKR[NH2]
MUT110	[Ace]RWRRHIHFKRPHRKHKR[NH2]
MUT120	[Ace]RWRRHVHFKRPHRKHKR[NH2]
Comb(165&114)	[Ace]RWRRHFHFd-KRPHRKHKR[NH2]
Comb(157&83)	[Ace]RWRRNKHTKRPHRKHKR[NH2]
Comb(165&110)	[Ace]RWRRHIHFd-KRPHRKHKR[NH2]
Comb(165&120)	[Ace]RWRRHVHFd-KRPHRKHKR[NH2]

Figure 1.1 The list of peptides synthesized and tested for the project. A. Single mutant peptide library. The first row is the original sequence RWRRHKHF-KRPHRKHKR, each of its position is replaced by 20 different amino acids in the second column, including one D amino acid (d-K). The orange color indicates the same amino acid replacement as control. B. Double mutant library containing ten single mutants and their double combinations with [H] as N terminal and [OH] as C terminal. C. Double mutant library containing six single mutants and their double combinations with [Ace] as N terminal and [NH2] as C terminal.

All peptides tested in this project were synthesized as three production runs by Sigma-Aldrich custom peptide library PEPscreen®. The first order (order number: 55435) contained 340 peptides, which were all single amino acid mutants of the original peptide [Ace]RWRRHKHFKRPHRKHKR[NH2] (Figure 1.1A). These 340 peptides were used for the initial screening for both antimicrobial and hemolytic activities. The second and third orders (order number: 3007196372 and 58302) contained 23 and 10 peptides, respectively (Figure 1.1B, 1.1C). They were picked and constructed as double/ triple mutants based on the initial screening of single mutants.

1.3.2 Peptide concentration determination by BCA method

All the ordered peptides were received as the lyophilized forms in 2ml tubes. Although these peptides were labeled as containing a known amount (mg), we did not know their solubility. So the concentrations needed to be determined. All peptides were dissolved in 100uL PBS and 100uL 50% acetonitrile. Since the Nanodrop cannot be used to measure the peptide concentrations with low extinction coefficient, in order to measure hundreds of peptides efficiently, we used the Pierce BCA Protein Assay Kit from Thermo Scientific. BSA standards of 2.0mg/ml Albumin Standard to 1.5mg/ml, 1.0mg/ml, 0.75mg/ml, 0.5mg/ml, 0.25mg/ml, 0.125mg/ml and 0.025mg/ml were used The working reagent was a mix of 50 parts of BCA Reagent A with 1 part of BCA Reagent B. Since our sample size were limited, 10 ul of each unknown sample and standard were used. Plates were incubated at 37 degree for 60 minutes. A SpectraMax 190 Absorbance Microplate Reader (Molecular Devices) was used for measurements at wavelength 562 nm. Dissolved peptides were stored in -20^{o}C. Each time these peptides were used, we

tested the concentrations again in case of peptide degradation, which was observed each time we tested.

1.3.3 Bacteria strains and culture

The two bacteria strains we tested, *Pseudomonas aeruginosa* (PA) PAO-1 and *Staphylococcus aureus* (SA) UAB637, were provided by the Center for Infectious Diseases and Vaccinology (CIDV), the Biodesign Institute at Arizona State University (ASU). MH (Mueller-Hinton) medium was used as the standard medium for both bacteria growth and antibiotic susceptibility test. Typically MH contains 30% beef infusion, 1.75% casein hydrolysate, 0.15% starch and 1.7% agar (PH adjusted to neutral at 25%). The working colony plates were prepared directly from the -80^{o}C degree original strain stock. Colony forming units at OD=0.5 were measured each time while making new working plates to make sure the bacteria number was the same for every experiment. For both SA and PA, we use the bacteria at the index growth period. A single colony was incubated with 5mL MH medium in a 50 mL tube (RPM250, 37^{o}C degree) for 6~7 hours. 100uL bacteria culture was measured by spectrometer at 600nm to make the OD=0.5. The final concentrations of the working culture was set as $2x10^{5}$CFU/mL.

1.3.4 Synbody and synbody effectors synthesis (HPLC,MALDI, Silver stain)

Synbody and synbody effectors are the conjugation of multiple peptides with linker molecules. The chemistry used for synbody synthesis is different based on different linkers applied. Synbody synthesis includes four steps, which are listed below.

Synbody synthesis: Peptide and scaffold are mixed together as a ratio of 2:1 and reacted in the solution phase (V[water]:V[ACN]=7:3; PH=6.5~7.5 adjusted by TEA); The mixture is incubated in room temp for at least 3 hours.

MALDI check: The raw products need to be checked by MALDI first. Samples are diluted in the MALDI matrix (100ul volume α-Cyano-4-hydroxycinnamic acid; 166ul water 2X; 166ul ACN; 40ul 10% TFA in water) and spotted on the MALDI MSP 96 target polished steel. MALDI is measured by MicroFlex and the results are analyzed by flexAnalysis.

HPLC primary check: Purification of the raw products is done by the HPLC Agilent Technology 1260 series. Briefly speaking, the raw products are mixed together with buffer A (0.1% TFA) and buffer B (90% CAN and 0.1% TFA) as the ratio 1:1:1, and the mixture is ran on the HPLC, the correct peaks are collected and combined together in a 50mL tube, finally the solution containing the purified products is lyophilized by the Labconco Freezone18 Dry system overnight.

Silver stain check: All the synbodies and synbody effectors were finally checked by silver stain. Pierce® Silver Stain Kit (Thermo #24612) was used for this purpose. The samples are firstly ran on Invitrogen NuPAGE® Novex® 4-12% Bis-Tris Gel. After running, the gel is fixed, sensitized, stained, and developed. The stain results were visualized by ChemiDoc™ XRS+ System

.

1.3.5 Broth microdilution susceptibility test

For the susceptibility assays, both PA and SA were grown in the Fluka-Mueller Hinton Broth/ Agar at 37℃. Selected peptides were measured twice by the BCA Protein

Assay Kit from Thermo Scientific Pierce to make sure the correct test concentrations. The MIC assay was slightly modified from the Clinical and Laboratory Standard[15]. ~$2x10^5$ bacteria were incubated with peptides in a total volume 100μL (90uL bacteria culture with 10uL peptide in PBS) aerobically at 35±2℃ for 18h. The endpoint reading was performed by a microplate ELISA reader (Molecular Devices SpectraMax M5) at 600nm absorbance. CORNING 3641 96-well flat bottom nonbinding surface polystyrene ELISA plates were used to reduce the peptide attachment. The standard definition of Minimum Inhibition Concentration (MIC) is the lowest concentration of an antibiotic which can inhibit the visible growth of certain bacteria. It is notable that there is not a quantitative criteria for measuring the "visible growth". In order to make our tests comparable with each other, we defined the MIC as the turning point below 20% Relative Growth (RG%: compared with the non-treated control group).

1.3.6 Hemolytic activity assay

The hemolytic activity assay on the mouse red blood cells used here was adapted from Shin et al[16]. Briefly, female BALB/C mice blood cells were used. Fresh erythrocytes were rinsed three times with PBS (35 nM phosphate buffer, 150 mM NaCl ,pH 7.0), centrifuged for 15min at 900g and then plated in 96-well microtiter plates (Nunc). The plates were incubated 1h at 37 ℃ with 100μL of the peptide solution and then centrifuged 1000g for 5 minutes. 100μL aliquots of the supernatant were transferred to a new 96-well plate for reading using a microplate ELISA reader (Molecular Devices SpectraMax M5) at 414nm absorbance. The % hemolysis was calculated by the formula: [(A414 in the peptide solution – A414 in PBS)/(A414 in 0.1% Triton-X 100 – A414 in

PBS)] x 100. PBS and 0.1%Triton-X100 were used as the negative and positive control, respectively. It's notable that the incubation time of the hemolytic assay is very sensitive, when measuring lots of samples simultaneously at least two people are needed to make precise time control.

1.3.7 Synbody effectors: mouse immunization and antibody titer determination

In order to test our synbody effectors in vitro, we produced the mouse anti-serum against the immunomodulatory peptide arms of the synbody effectors.

Conjugation of peptide to preactivated KLH: The preactivated KLH (Thermo SCIENTIFIC Imject Maleimide Activated mcKLH) is resuspended in PBS/EDTA to get 10.0mg/ml; 2mg/100ul peptide is mixed up with 200ul activated KLH gently and incubated at room temperature for 4 hours; After Pre-washing the slide-a-lyzers in PBS, the mixture is loaded into the slide-a-lyzers and dialyze for four PBS changes (the first one is overnight, the next three for one hour each); Remove from slide-a-lyzer and adjust buffer to 10% glycerol.

Immunization: the peptides conjugated to KLH were mixed 1:1 with pierce Imject Alum. 250ul of 1mg/ml peptide in PBS were mixed with 250ul Imject Alum for ~1 hour. 50ug per mouse was injected in a total volume of 100ul. Mouse: BALB/c mice from Charles River DOB: 19 JAN 2012 female. All mice were injected three times in total, and checked bleed periodically.

Mouse IgG endpoint ELISA: Antigen plate preparation: antigens are diluted at 1.0ug/ml in carbonate/bicarbonate buffer (0.2M Na_2CO_3 and 0.2M $NaHCO_3$) and dispensed 100ul into each well of Nunc Maxisorp flat bottom plate (100ng/well). Incubate at 4 degree

overnight. ELISA: The antigen plates are blocked by 5% nonfat milk in TBST for blocking 1 hour at 37 degree; serially diluted sera samples are added as primary across plate to make the dilution: 1:400, 1:800, 1:1600 etc; the plates are then incubated at 37 degree for one hour and washed three times with TBST; Finally the ELISA results are colored by the HRP-ABTS reaction (1.0% SDS as the terminator), and measured under 405nM absorbance.

1.3.8 Peptide sub-array binding test

We used the microarray platform to test the binding between selected peptides and pathogen cells. ***Peptide sub-array preparation:*** All the sub-arrays were prepared in our lab. Polymer slides were made from glass slides. Briefly the procedure was to. 1)clean the glass slides with piranha solution; 2). incubate in silanization solution (1% 3-Glycidoxypropyl-trimethoxysilane in anhydrous toluene) for 30min at 40 degree; 30 Wash with toluene three times and dry in oven for 40 min 120 degree; 4) Incubate in coating solution (6mg/ml polyethylenimine branched in 10% ethanol) for 1hr, wash and dry; 5) Apply 200ul crosslinker solution (10mM SMCC in 1*PBS with 1mM EDTA, PH7.2) on the surface and incubate for 1hr, wash and dry. The final spotting concentration for each peptide is 1 mg/ml. Spotting buffer: 40mM HEPES, 20nM TCEP and 10mM EDTA. The selected peptides were printed automatically by a Nanoprinter 60.

Pathogen cell binding assay: *Staphylococcus aureus* (SA) UAB637 was used at OD value 0.5. 10mg cells were washed by PBS two times and dispensed into three parts: 8mg for control; 1mg for labeling with AF647 (5ul in 500ul 1*?PBS); 1mg for labeling with CTO (10mM in DMSO, diluted to 10uM). Cells were incubated with dyes for 1hr at 37

degree, then washed and centrifuged at 5000g for 5mins. The final hybridization solution contained 1mg cells, 0.03% sodium azide, 3% BSA and 0.05% tween20. Previously prepared polymer slides were prewashed with TFA, DMF, ethanol and water sequentially. Then the slides were blocked with blocking buffer (3%BSA, 0.05%twee n20, 0.134mg/ml mercaptohexanol, 1*TBS) for 1hr at RT in humidity chamber. The slides incubated with the prepared cells using the Agilent hybridization chamber system for 1hr at 37^{o}C degree. The slides were washed and then scanned in an Aglient scanner at 555nm and 647nm.

1.3.9 Minimum Bactericidal Concentration (MBC) Test

The minimum bactericidal concentration test was modified from Nakatsuji's protocol[17]. The incubation plates were prepared as 100ul/well, containing 10ul testing peptides (400uM, 200uM, 100uM, 50uM, 25uM and 125uM) and 90ul 10^{7}CFU/ml *Staphylococcus aureus* (SA) UAB637 in PBS. It's notable that in order to prevent the bacteria growth during the incubation period, the incubation solution is PBS rather than medium. After incubation at 37^{o}C for 5 hours, the mixture of peptide and bacteria of each well was diluted into 15~30 CFU/15 ul with medium, and then spotted on CORNING low profile square BioAssay large dishes. PBS containing no bacteria was spotted as the negative control, bacteria incubated with kanamycin was spotted as the positive control.

1.3.10 Macrophage absorption assay

The concept of the macrophage absorption assay is to incubate macrophage with certain bacteria cells and then separate the bacteria cells remaining in the solution from

those absorbed by the macrophage. This assay was used to measure the in vitro viability of macrophage phagocytosis[18,19].

Macrophage: the macrophage strain we used here was J774A.1 ATCC. It is a mouse source macrophage which is active in antibody dependent phagocytosis. Cells were grown in ATCC-formulated Dulbecco's Modified Eagle's Medium with 10% FBS at 37 degree, 5% CO_2. The subcultivation ratio is 1:5. The best cell density for the macrophage absorption assay was determined using a Roche Cell Proliferation Kit II (XTT). Note: the macrophage growth time for XTT test should be the same as that to be used in the macrophage absorption assay, which is 12 hours here.

Macrophage absorption assay: Since this assay contains contents of both mammalian cells and bacteria cells, it's important to separate the procedures into two parts. PART A (tissue culture room): Macrophage cell culturing: Cells are seeded at $5x10^4$ per well (100ul) in a 96-well tissue culture plate, overnight incubation. Media is removed and the cellswashed with PBS three times. Plate control: no macrophage seeded wells. PART B (biosafety hood outside): Bacteria cells were cultured to OD=0.5 ($1.2x10^9$ CFU/ml), make dilutions in HBSS/HBSS+10%NMS (normal mouse serum, fresh)/ HBSS+synbody effectors. Bacteria were added at multiplicity of infection (MOI) of 10($5x10^5$) volume 100ul. Incubate in 37ºC, 5% CO_2, 30mins. Each well gently was mixed gentlyand the supernatants removed. The plate was washed with PBS 3 times. 0.1% cold sodium deoxycholate in PBS 100ul was added to each well for 5mins on ice. Each well was mixed thoroughly. 50ul was taken into 450ul PBS to make 10 times dilution. 50ul of the dilution was spread on MH agar plates and incubated at 37ºC overnight. The final count on the plates will be the number of bacteria cells absorbed by macrophage.

1.4 Results

1.4.1 Characterization of the original lead antimicrobial peptide RWRR

Previously, we have reported using a 10,000 random peptide microarray technique to screen for efficient bacteria-binding and lytic peptides. One of the peptides we selected showed a good array pattern as a bacteria lytic peptide (CTO-AF+; Bacteria membrane signal: Cell Tracker Orange and nucleus acid dye AF555). The sequence of the original peptide is [Ace]RWRRHKHFKRPHRKHKR-GSC[NH2] (Molecular weight=2688.169), the last three amino acids GSC is a general linker we used to connect with the microarray surface. This lead peptide contains 30% arginine, 20% lysine and 20% histidine with a high net charge of +14. The total hydrophobic ratio is 15% and the protein-binding potential (Boman index) is 6.31kcal/mol. It is notable that this is a linear peptide without any α -helix structures (predicted by using the AGADIR program[20], %helicity/residue=0.37% at pH 7, 278^{o} K and an ionic strength of 0.1. A). According to the classification by Yechiel Shai, our lead peptide belongs to the AMPs enriched with one or two amino acids (Pro-Arg, Trp or His) and without Cys. Other examples are PR-39, Indolicidin and Histatin 1[21]. As a starting point of our mutant library creation, we tested this lead peptide for its antimicrobial effects on both *S.aureus*(SA) and *P.aeruginosa*(PA).

Standard broth microdilution based antimicrobial susceptibility test (modified from the CLSI standards) was applied for validating the antimicrobial potency of our RWRR lead peptide (Figure 1.2). The lead peptide showed a distinct selectivity between SA(MIC=7.5μM) and PA(MIC=40μM). The dramatic shift observed from the minimum relative growth to the maximum relative growth indicated a membranlytic antimicrobial peptide mode of action[22] (Figure 1.2A). The time-based killing assay also showed a complete inhibition against SA over 22 hours above 5μM (Figure 1.2B).

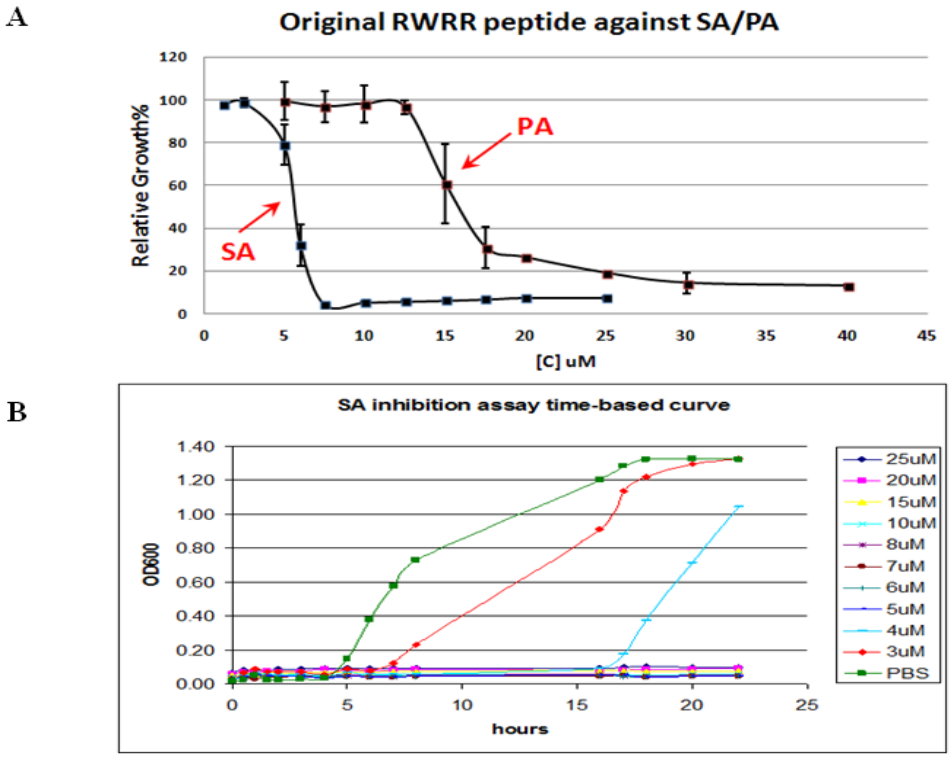


Figure 1.2. In vitro antimicrobial effects of the lead RWRR peptide against *S. aureus* (SA) and *P. aeruginosa* (PA) A. Broth microdilution based antimicrobial susceptibility test (modified from the CLSI standards) for the lead RWRR peptide. Both SA and PA were adjusted to ~$2x10^5$ CFU/mL at the starting point, the end point was read as OD600 value after 18 hours incubation with the peptide series dilutions. Results are shown as the relative growth % compared with the non-peptide treated control. B. *S. aureus* growth over time with the lead peptide treatments. The SA starting concentration was ~$2x10^5$ CFU/mL. This result corroborates with the broth microdilution in A.

1.4.2 Lead peptide mutant library creation and screening

We created a single mutant peptide library containing 304 RWRR single amino acid variants. The first 17 amino acids of the lead peptide were substituted by 19 different amino acids (A, R, N, D, d-K, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V), except for cysteine. The N terminal and C terminal of the peptide were acetylated and amidated, respectively. Based on the antimicrobial susceptibility test of the lead peptide, we chose 5μM and 15μM as the starting screening concentration for SA and PA, respectively and 18 hours as the incubation time. The hemolytic assay was performed at peptide concentration of 50μM (Figure 1.3).

For the bacteria inhibition assays, we defined the fold change value FC_{SA} and FC_{PA} as RG%(Lead)/RG%(Mut) (RG%: bacteria relative growth % compared with non-peptide treated control). Among the 301 single peptide mutants we tested (3 had insufficient sample amounts), 23 mutants showed improved inhibitory effects on SA (FC_{SA}>1.5) with the most significant FC_{SA}=17.45 (MUT122); while 84 mutants showed significant decreased potency against SA (FC_{SA}<0.08). In comparison, 9 mutants had FC_{PA}>1.5 with the highest FC_{PA}=10.95 (MUT114); while 19 mutants had FC_{PA}<0.08. It's notable that although the fold changes here were not the fold changes of minimum inhibitory concentration (MIC), they indicated the correct trends of improvement. We also defined

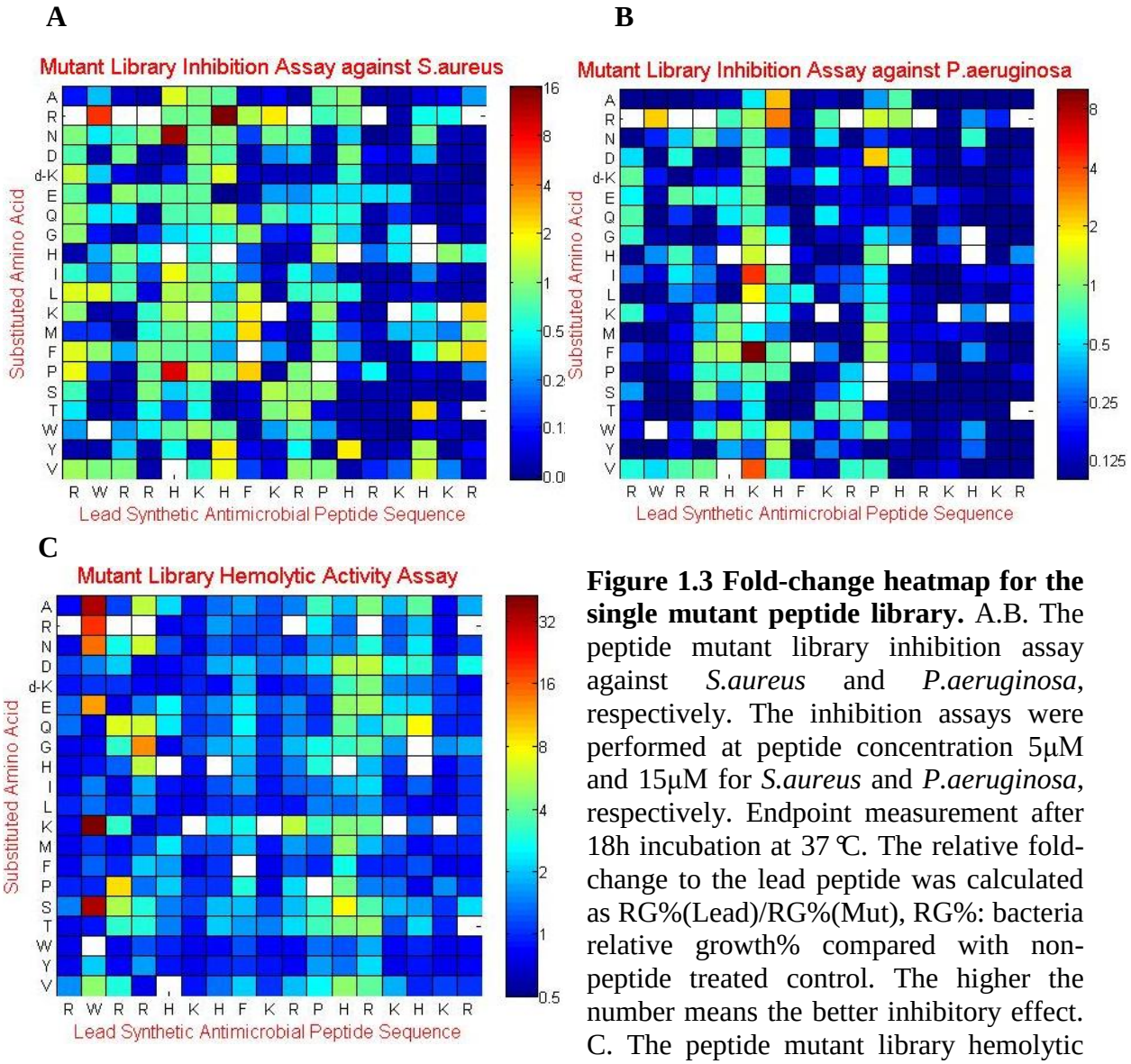


Figure 1.3 Fold-change heatmap for the single mutant peptide library. A.B. The peptide mutant library inhibition assay against *S.aureus* and *P.aeruginosa*, respectively. The inhibition assays were performed at peptide concentration 5μM and 15μM for *S.aureus* and *P.aeruginosa*, respectively. Endpoint measurement after 18h incubation at 37 ℃. The relative fold-change to the lead peptide was calculated as RG%(Lead)/RG%(Mut), RG%: bacteria relative growth% compared with non-peptide treated control. The higher the number means the better inhibitory effect. C. The peptide mutant library hemolytic activity assay. The fold-change was determined by the formula: Hemolysis%(Lead)/Hemolysis%(Mut), Zero and 100% hemolysis were determined in PBS and 0.1%Triton-X 100; assays were performed at peptide concentration 50uM. The higher the number means the better improvement on the hemolytic activity. Single Mutant nomenclature: "R3P" represents the arginine in the position 3 (position 1-16 from the left N-terminal to the right C-terminal) is substituted by a proline. Heatmap were created by MATLAB, the white features indicate the same amino acid substitutes or missing data.

the fold change in hemolytic assay FChemo as Hemolysis%(Lead)/Hemolysis%(Mut) (Zero and 100% hemolysis were determined in PBS and 0.1%Triton-X 100). 89 out of 301 mutants we tested showed at least twice less hemolysis than the lead peptide (FChemo>2).

A close scrutiny into these heat maps revealed some hot substitution positions/ amino acids (Figure 1.4). For example, 31.6% substitutions in position 6 (lysine) improved the peptide's anti-PA potency; 63.2% substitutions in position 13 (arginine) and 12 (histidine) decreased the hemolytic effects by at least two-folds. 27.3% and 45.5% arginine replacement improved the peptide's anti-SA and anti-PA potency, respectively. After all, it seems that the net charge impacts both the antimicrobial effects and the hemolytic activity.

A

FC(hemo)>2		1R	2W	3R	4R	5H	6K	7H	8F	9K	10R	11P	12H	13R	14K	15H	16K	17R	
1	A	0.85	32.19	1.14	5.75	2.17	0.89	1.36	1.70	1.18	1.43	3.35	1.92	4.45	1.95	3.77	0.86	1.82	35.3%
2	R		18.34			0.76	0.91	1.72	1.23	1.09		2.40	1.39		1.26	1.93	0.78		18.2%
3	N	0.77	14.21	2.78	6.33	1.23	0.78	1.28	1.04	1.19	0.95	1.56	1.58	2.84	1.25	2.39	0.77	1.09	29.4%
4	D	1.14	1.47	2.08	0.77	0.83	0.96	1.85	2.08	1.36	1.68	2.21	5.48	5.94	2.82	3.16	1.09	2.77	47.1%
5	d-K	0.92	1.05	1.03	0.86	0.77	0.98	0.84	1.88	0.82	0.94	1.18	3.39	4.58	1.55	1.48	1.09	0.87	11.8%
6	E	1.31	11.53	0.78	1.47	2.57	1.08	1.53	2.37	1.16	1.57	0.96	4.90	5.10	2.15	2.22	1.08	0.96	41.2%
7	Q	1.33	0.80	6.63	6.28	2.31	1.03	1.25	2.62	0.99	1.47	1.03	3.73	2.04	3.71	7.77	0.99	0.97	47.1%
8	G	0.81	0.85	3.07	12.86	0.79	1.22	1.76	2.12	1.05	1.94	2.79	3.17	3.83	1.71	1.31	1.55	1.75	35.3%
9	H	0.77	0.90	1.27	5.66		0.93		1.83	0.99	1.44	3.01		1.90	1.15		0.83	1.14	15.4%
10	I	0.82	1.47	0.79	1.85	0.88	0.84	1.20	1.63	0.89	1.23	1.38	1.92	2.21	0.91	1.30	0.80	1.06	5.9%
11	L	0.94	1.39	0.94	1.53	0.82	0.79	1.02	1.09	0.91	0.90	1.44	1.37	1.63	1.01	0.85	0.79	1.02	0.0%
12	K	0.83	42.84	3.13	0.75	0.97		2.25	2.69		6.00	3.06	4.48	4.06		1.35		1.18	61.5%
13	M	0.77	1.02	0.82	1.19	2.38	1.09	1.15	1.90	0.90	1.73	1.32	4.67	1.93	1.30	0.82	0.98	1.00	11.8%
14	F	0.77	1.28	0.94	2.02	1.68	0.77	1.01		0.80	1.20	0.94	2.65	0.98	0.98	1.20	0.81	0.82	12.5%
15	P	0.93	0.95	9.11	1.36	2.04	1.03	1.00	3.01	0.80	2.17		4.04	1.49	1.39	0.80	1.59	0.96	31.3%
16	S	1.44	33.06	5.40	3.00	1.46	1.09	1.96	1.90	1.13	1.87	3.14	7.58	3.57	1.95	1.65	1.01	2.16	41.2%
17	T	0.80	0.92	3.23	3.01	1.45	0.93	2.05	1.63	1.45	2.21	3.43	4.45	4.74	1.21	2.27	0.94	10.93	52.9%
18	W	0.78		0.81	1.20	0.93	0.77	0.86	1.20	0.82	0.83	1.36	1.18	1.84	0.83	0.96	0.79	0.84	0.0%
19	Y	0.77	2.07	0.85	1.72	0.92	0.77	0.98	1.02	0.90	0.83	1.74	1.38	2.22	0.81	0.91	0.79	0.83	11.8%
20	V	1.59	4.72	2.89	0.77	22.76	0.93	1.40	2.00	1.04	1.96	2.61	4.82	3.80	1.78	0.98	1.10	1.07	41.2%
		0.0%	42.1%	47.4%	42.1%	31.6%	0.0%	10.5%	36.8%	0.0%	15.8%	47.4%	63.2%	63.2%	15.8%	31.6%	0.0%	15.8%	

B

FC(PA)>1		1R	2W	3R	4R	5H	6K	7H	8F	9K	10R	11P	12H	13R	14K	15H	16K	17R	
1	A	0.08	0.08	0.09	0.12	0.13	0.51	2.30	0.11	0.13	0.08	0.36	0.77	0.08	0.09	0.10	0.08	0.09	5.9%
2	R		2.07			0.75	1.08	3.12	0.12	0.77		1.36	1.08		0.11	0.32	0.20		45.5%
3	N	0.07	0.19	0.43	0.88	0.31	0.80	0.21	0.11	0.48	0.08	0.22	0.14	0.12	0.09	0.67	0.10	0.08	0.0%
4	D	0.46	0.08	0.62	0.11	0.09	0.85	0.49	0.08	0.14	0.17	2.11	0.65	0.14	0.08	0.18	0.09	0.11	5.9%
5	d-K	0.86	0.19	0.11	0.15	0.18	0.92	0.14	0.20	0.55	0.19	0.16	0.73	0.12	0.08	0.09	0.11	0.13	0.0%
6	E	0.53	0.12	0.86	0.62	0.59	0.87	0.07	0.07	0.14	0.16	0.13	0.14	0.22	0.17	0.13	0.08	0.12	0.0%
7	Q	0.79	0.09	0.21	0.11	0.51	0.71	0.22	0.08	0.49	0.18	0.16	0.21	0.09	0.15	0.08	0.09	0.09	0.0%
8	G	0.69	0.12	0.11	0.09	0.18	1.19	0.54	0.09	0.14	0.13	0.46	0.32	0.08	0.28		0.09	0.07	6.3%
9	H	0.10	0.32	0.70	0.25		1.39		0.15	0.10	0.11	0.88		0.15	0.15		0.10	0.32	7.7%
10	I	0.29	0.15	0.51	0.31	0.13	4.39	0.83	0.11	0.22	0.24	0.52	0.13	0.09	0.08	0.17	0.17	0.16	5.9%
11	L	0.09	0.11	0.33	0.23	0.10	1.71	0.45	0.57	0.12	0.30	0.51	0.17	0.12	0.10	0.12	0.08	0.17	5.9%
12	K	0.67	0.17	0.13	0.44	0.88		0.54	0.13		0.12	0.66	0.17	0.13		0.34		0.20	0.0%
13	M	0.09	0.09	0.16	0.40	1.04	0.56	0.30	0.10	0.11	0.11	1.19	0.14	0.15	0.17	0.14	0.09	0.09	11.8%
14	F	0.21	0.14	0.15	1.02	1.21	10.95	0.84		0.31	0.08	1.13	0.17	0.13	0.09	0.33	0.12	0.09	25.0%
15	P	0.12	0.08	0.16	0.66	0.61	0.85	0.78	0.08	0.12	0.43		0.18	0.22	0.10	0.11	0.08	0.15	0.0%
16	S	0.36	0.10	0.08	0.88	0.32	0.60	0.11	0.10	0.22	0.55		0.09	0.09	0.10	0.09	0.09	0.08	0.0%
17	T	0.16	0.10	0.10	0.22	0.16	0.52	0.08	0.08	0.71	0.80	0.18	0.10	0.09	0.11	0.11	0.10		0.0%
18	W	0.17		0.18	0.62	1.18	0.64	1.32	0.39	0.16	0.41	0.82	0.20	0.11	0.13	0.36	0.09	0.07	12.5%
19	Y	0.11	0.10	0.17	0.08	0.38	0.27	1.16	0.08	0.32	0.15	0.10	0.12	0.08	0.12	0.26	0.08	0.08	5.9%
20	V	0.64	0.48	0.80	0.88		3.68	0.67	0.16	0.13	0.68	0.74	0.08	0.13	0.15	0.11	0.09	0.08	0.0%
		0.0%	5.3%	0.0%	5.3%	16.7%	31.6%	21.1%	0.0%	0.0%	0.0%	22.2%	5.3%	0.0%	0.0%	0.0%	0.0%	0.0%	

C

FC(SA)>1.5		1R	2W	3R	4R	5H	6K	7H	8F	9K	10R	11P	12H	13R	14K	15H	16K	17R	
1	A	0.13	0.34	0.09	0.07	1.52	0.90	0.79	0.09	0.12	0.07	0.79	1.03	0.08	0.07	0.10	0.11	0.28	5.9%
2	R		6.10			0.88	0.83	17.45	1.17	2.13		0.65	0.88		0.07	0.49	0.51		27.3%
3	N	0.91	0.43	0.65	0.95	13.15	0.86	0.97	0.16	0.85	0.70	0.08	0.35	0.05	0.08	0.81	0.08	0.08	5.9%
4	D	0.75	0.08	0.89	0.07	0.08	1.00	0.74	0.07	0.25	0.34	0.08	0.97	0.11	0.09	0.33	0.07	0.08	0.0%
5	d-K	1.31	0.36	0.10	0.07	0.14	0.78	1.59	0.09	0.08	0.08	0.08	0.62	0.08	0.07	0.08	0.07	0.05	5.9%
6	E	0.81	0.09	0.99	0.70	0.78	0.78	0.06	0.07	0.27	0.26	0.40	0.39	0.40	0.38	0.07	0.07	0.08	0.0%
7	Q	1.02	0.41	0.43	0.07	0.71	0.70	1.24	0.15	0.74	0.40	0.51	0.55	0.08	0.15	0.09	0.07	0.09	0.0%
8	G	1.03	0.13	0.07	0.15	0.40	0.48	0.56	1.02	0.14	0.12	0.70	0.35	0.07	0.44		0.09	0.08	0.0%
9	H	0.08	0.29	0.92	0.51		0.56		0.18	0.07	0.08	1.26		0.21	0.42		1.07	0.55	0.0%
10	I	0.59	0.22	0.67	0.18	1.75	0.78	0.56	0.21	0.09	0.51	0.23	0.08	0.07	0.07	0.25	0.09	0.07	0.0%
11	L	1.58	1.63	0.70	0.09	1.20	1.00	0.34	1.31	0.07	0.60	0.65	0.56	0.08	0.09	0.10	0.08	0.08	11.8%
12	K	0.99	0.08	0.08	0.62	1.27		0.86	2.05		0.09	1.02	0.10	0.08		0.51		2.44	15.4%
13	M	0.15	0.15	0.05	0.55	0.76	0.94	0.38	2.29	0.11	0.08	0.56	0.17	0.09	0.33	0.32	0.24	1.27	5.9%
14	F	1.65	1.00	0.30	0.95	0.80	0.80	0.31		0.27	0.08	0.93	0.31	0.08	0.07	0.52	1.50	2.37	18.8%
15	P	1.66	0.15	0.07	0.79	9.08	1.26	0.86	2.36	0.08	0.93		0.13	0.47	0.08	0.10	0.08	0.24	18.8%
16	S	0.64	0.08	0.08	0.94	0.37	0.62	0.07	0.10	1.19	1.02	1.05	0.08	0.08	0.07	0.07	0.07	0.08	0.0%
17	T	0.44	0.08	0.08	0.52	0.15	0.51	0.08	0.09	0.97	1.20	0.10	0.07	0.08	0.07	2.34	0.09		6.3%
18	W	0.28		0.27	0.44	0.80	1.08	0.81	0.07	0.29	0.76	0.27	0.08	0.08	0.06	0.10	0.09	0.08	0.0%
19	Y	0.08	0.08	0.35	0.07	0.53	0.09	2.14	0.11	0.09	0.46	0.08	2.12	0.07	0.08	1.23	0.07	0.11	11.8%
20	V	1.09	0.95	0.96	0.07		0.62	1.66	0.16	0.10	1.03	0.93	0.08	0.14	0.19	1.45	0.23	0.09	6.3%
		15.8%	10.5%	0.0%	0.0%	16.7%	0.0%	21.1%	15.8%	5.3%	0.0%	0.0%	5.3%	0.0%	0.0%	5.6%	5.3%	11.1%	

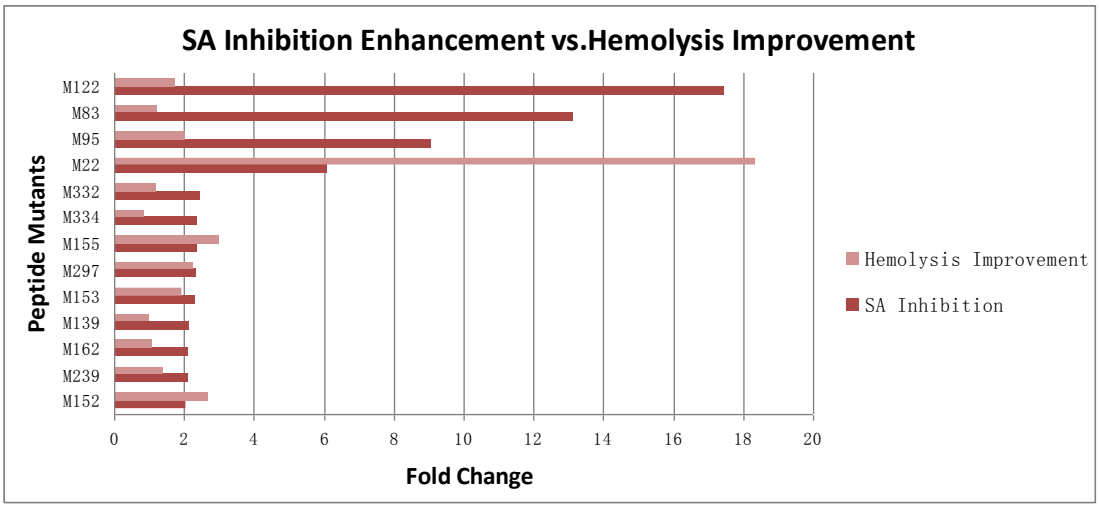
Figure 1.4. Statistics of fold-change heatmap for the single mutant peptide library. Green color represents the decreased hemolysis fold change (A), improved inhibitory fold change on SA (B) and PA (C). Empty positions are the same amino acid substitutes; Red colors are missing data.

1.4.3 Selected single substituted peptides showed changed specificity against *Pseudomonas aeruginosa* (PA) and *Staphylococcus aureus* (SA)

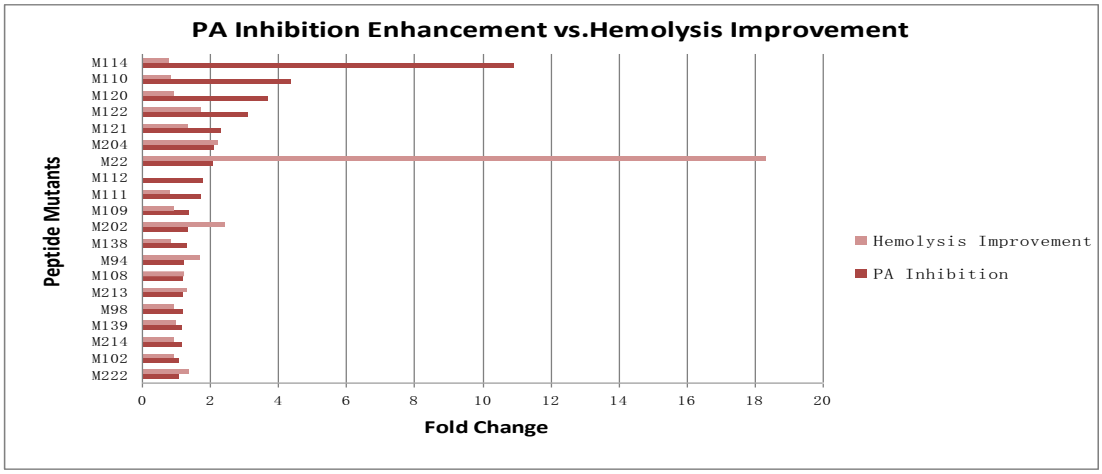
The next step we selected several promising single substituted peptides based on three different criteria (Figure 1.5). These peptides were tested for their minimal inhibitory concentrations (MICs). We noticed that most of the selected peptides showed

the consistent trend of MIC change compared with the initial library screening results, and the MIC change against PA is more significant than that against SA (Table 1.1, Figure 1.6). We also defined the peptide specificity against SA and PA as MIC(PA)/MIC(SA), and we showed that single amino acid substitutions can move the specificity value from the original 5.33 to a range of 1.5 (towards PA)~>6.67(towards SA) (Figure 1.3, Table 1.1). Interestingly, variants with the same position substituted by different amino acids had similar MIC and specificity shift, as we observed for MUT152(F8K); MUT153(F8M); MUT155(F8P); MUT157(F8T) and MUT110(K6I); MUT114(K6F); MUT120(K6V).

A



B



C

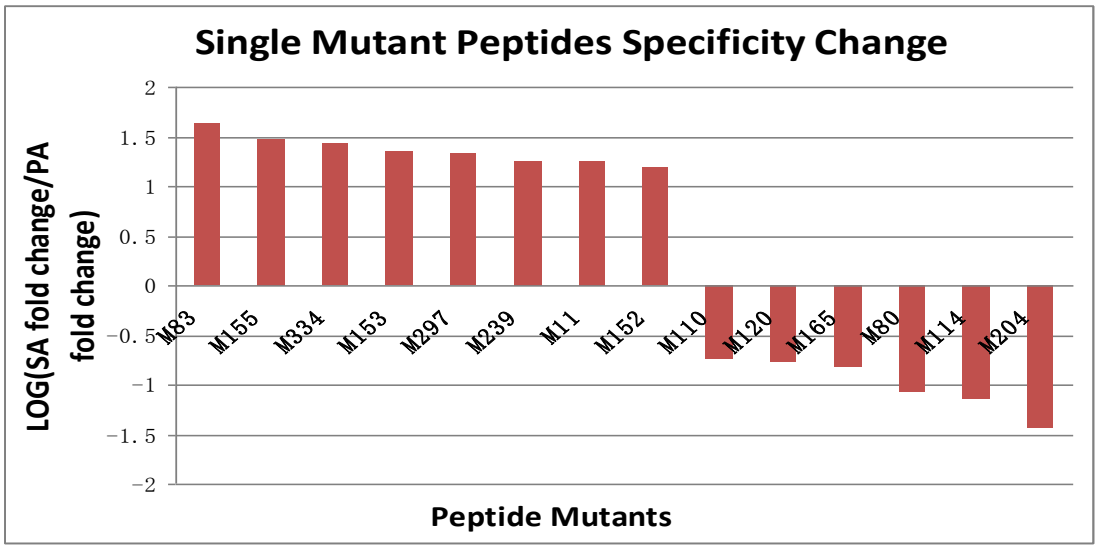


Figure 1.5. Different criteria used for selecting the single mutants. We chose the single mutants based on the improvement of inhibition against SA and hemolysis (A); the improvement of inhibition against PA and hemolysis (B); the change of inhibition selectivity between SA and PA (C).

Test Peptides	*S. aureus*		*P. aeruginosa*		Specificity
	MIC1/2(uM)[a]	MIC(uM)	MIC1/2(uM)	MIC(uM)	MIC(PA)/MIC(SA)
ORI	6	7.5	17.5	40	5.33
Single Mutants					
MUT83(H5N)	5.5	6	22.5	>40	>6.67
MUT152(F8K)	6.5	7.5	30	>40	>5.33
MUT153(F8M)	6.5	7.5	30	>40	>5.33
MUT155(F8P)	6.5	7.5	40	>40	>5.33
MUT157(F8T)	6.5	7.5	35	40	5.33
MUT110(K6I)	8.5	10	12.5	15	1.50
MUT114(K6F)	8.5	10	12.5	17.5	1.75
MUT120(K6V)	8.5	10	12.5	17.5	1.75
MUT165(K9d-K)	6.5	10	12.5	15	1.50
MUT22(W2R)	5	7.5	12.5	20	2.67
MUT222(H12R)	5.5	6	12.5	17.5	2.92
MUT290(H15I)	5.5	6	15	40	6.67
Double Mutants					
MUT157&83	2.75	5	>40	>40	>8.00
MUT110&165	5	6	8.5	10	1.67
Test Peptides	*S. aureus*		*P. aeruginosa*		Specificity
	MIC1/2(uM)	MIC(uM)	MIC1/2(uM)	MIC(uM)	MIC(PA)/MIC(SA)
ORI*[b]	2.75	5	<5	7.5	1.50
Single Mutants					
MUT83*(H5N)	5.5	7.5	11	12.5	1.67
MUT153*(F8M)	2.75	5	7.5	10	2.00
MUT155*(F8P)	2.5	5	7.5	10	2.00
MUT157*(F8T)	2.75	5	8.5	17.5	3.50
MUT110*(K6I)	6	10	6.5	7.5	0.75
MUT114*(K6F)	2.75	5	<5	<5	<1
MUT120*(K6V)	2.75	5	<5	7.5	1.50
MUT165*(K9d-K)	2.75	5	<5	7.5	1.50
MUT22*(W2R)	2.75	5	<5	<5	<1
MUT222*(H12R)	2.75	5	7.5	7.5	1.50
Double Mutants					
MUT22&222*	<2.5	2.5	<5	<5	<2
MUT83&153*	7.5	12.5	17.5	25	2.00
MUT165&120*	2.75	5	<5	<5	<1

a.MIC1/2 and MIC: inhibiting over 50% and 90% of the overnight bacteria growth, respectively. Determined by the concentration based inhibitory curves
b.the star symbol (*) represents the peptide N-[Ace] terminal and C-[NH2] terminal were deprotected to N-[H],C-[OH]

Table 1.1 Selected peptide mutants' minimum inhibitory concentration (MIC) comparison. The MIC endpoints and MIC1/2 points were determined by the broth microdilution susceptibility test with different peptide dilution series (*S.aureus*: 25μM, 20μM, 17.5μM, 15μM, 12.5μM, 10μM, 7.5μM, 6μM, 5μM, 2.5μM, 1.25μM; *P.aeruginosa*: 40μM, 30μM, 25μM, 20μM, 17.5μM, 15μM, 12.5μM, 10μM, 7.5μM, 5μM). In general, the MIC of PA and the specificity changed more significantly than the MIC of SA by mutation.

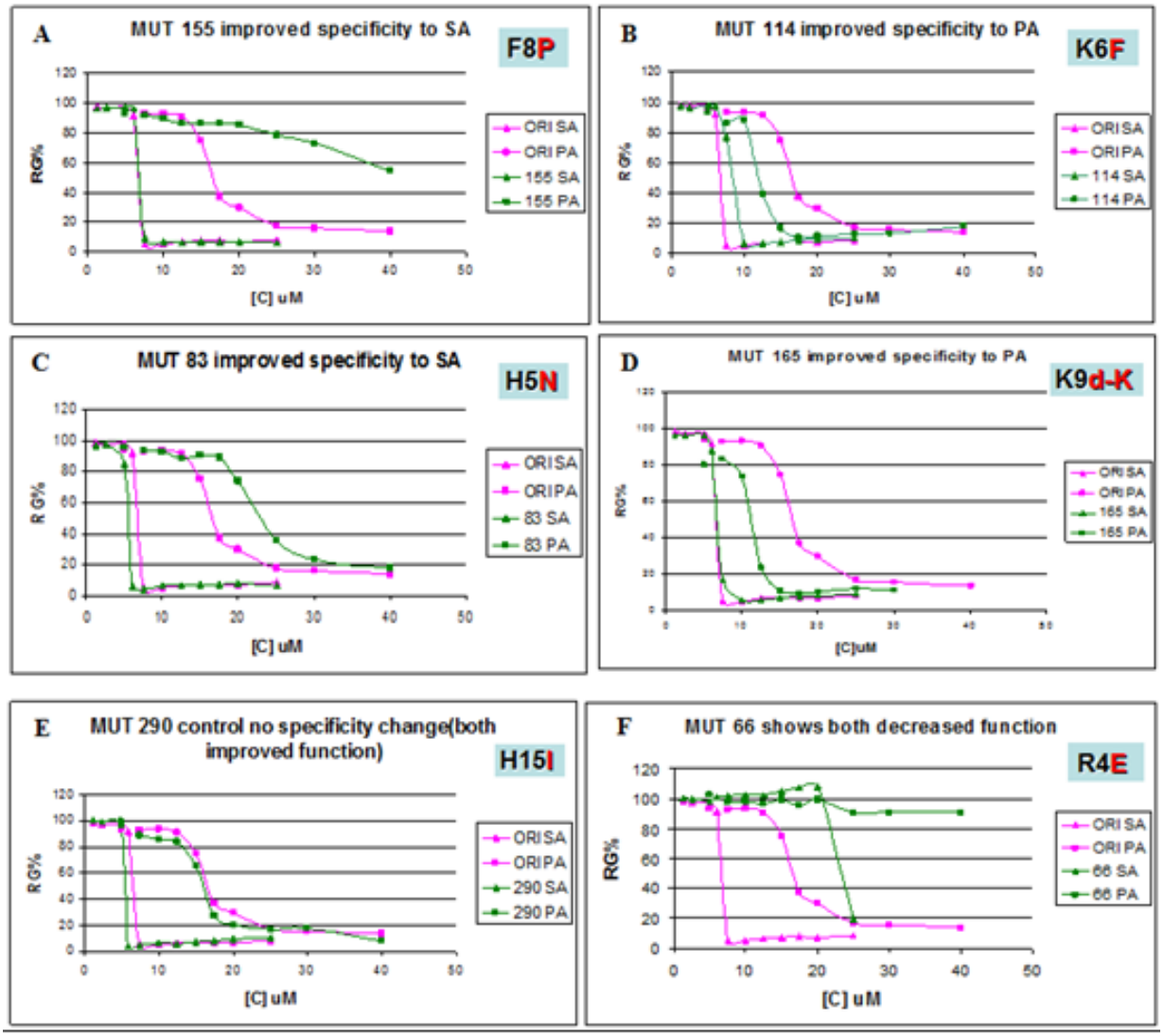
1.4.4 Selected single substituted peptides showed changed specificity against *Pseudomonas*

In one of our tests, we synthesized the selected peptides without N and C terminal protection (previous peptides had [ACE] group at the N terminal and [NH2] group at the C terminal). All the peptides showed undifferentiated MIC improvement against both SA and PA, especially PA (Table 1.1). The possible reason was the arginine at the N terminal plays an important role while it is deprotected in the solution.

1.4.5 There is an additive effect of double substituted peptide variants

To better understand the substitution-function relationship, we synthesized several double substituted peptide variants based on the single peptide variants we selected in 4.3. Previous work using this approach for protein targets had demonstrated additive improvements (REF). On the one side, we showed that by combining two defective single variants together, there was an even worse performance (Figure 1.7 A,B). This could be explained as the loss of functional structure. On the other side, two single variants with improved potency also had additive improvement (Figure 1.7C,D). For

example, the MUT22*(W2R) + MUT222*(H12R) exhibited the best MIC we observed (2.5μM) against SA, which might be caused by the increase of arginine %[23]. Importantly, all the screening and selection processes were performed without any knowledge of the peptide structures. This would be a benefit for us developing the MOVE THIS



UPsynthetic antimicrobial peptides for unknown bacteria targets.

Figure 1.6. Selected single mutant peptides show distinct specificities against *S.aureus*(SA) and *P.aeruginosa*(PA) by the broth microdilution susceptibility test. MUT: peptide mutant; ORI: lead peptide; RG%: bacteria relative growth% compared with non-peptide treated control; [C]uM: peptide concentration. A. MUT155(F8P) shows improved specificity to SA, MUT153(F8M) and MUT157(F8T) show the similar change.

B. MUT114(K6F) shows improved specificity to PA, MUT110(K6I) and MUT120(K6V) show the similar change. C. MUT83(H5N) shows improved specificity to SA. D. MUT165(K9d-K) shows increased specificity to PA. E. MUT290(H15I) shows slight improvement for both SA and PA but no specificity change. F. MUT66(R4E) shows significant decrease for both SA and PA, which indicates the importance of arginine in position 4.

* represents the peptide N-[Ace] terminal and C-[NH2] terminal deprotected to N-[H],C-[OH]

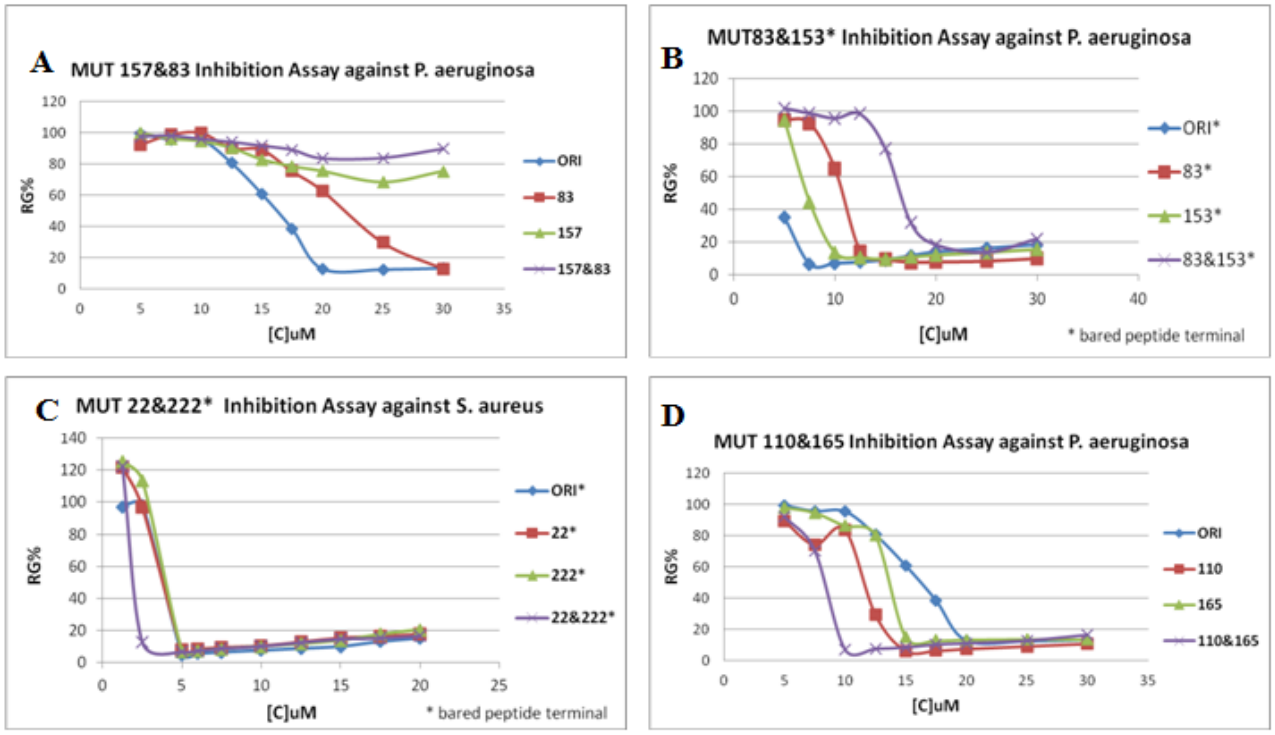


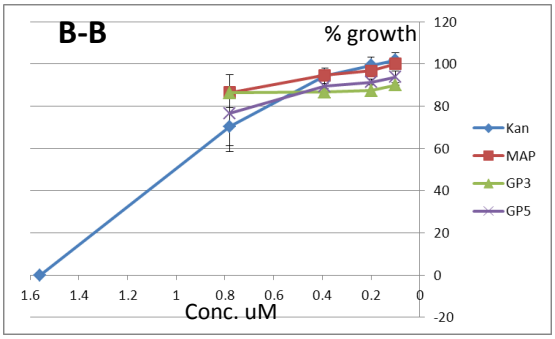
Figure 1.7. Selected double mutant peptides show additive effects on their antimicrobial potency. MUT: peptide mutant; ORI: lead peptide; RG%: bacteria relative growth% compared with non-peptide treated control; [C]uM: peptide concentration. A. MUT157*&83* shows additive decreased potency against PA, it also shows slight additive improved potency against SA (see Table1). B. MUT83&153* shows additive decreased potency against PA. C. MUT22*&222* shows additive improved potency against SA. D. MUT110&165 shows additive improved potency against PA.

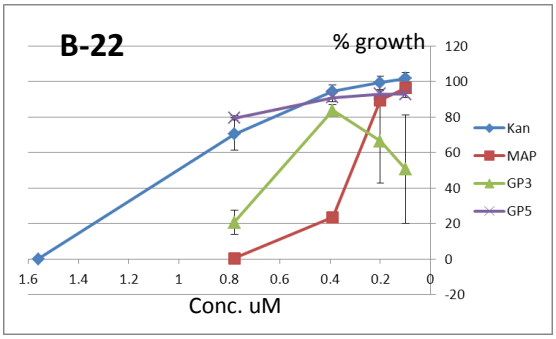
1.4.6 Synbodies made from the single variants exhibit great potency against SA

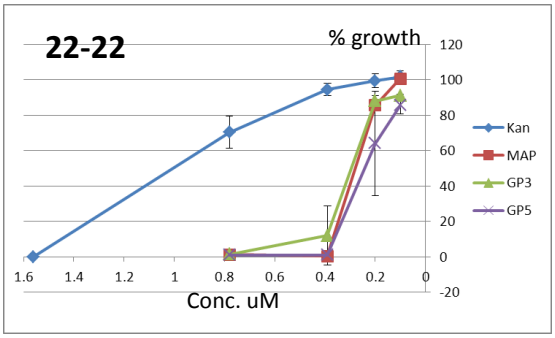
Previously, our group had shown that the antimicrobial effect of synbody (combination of one lytic peptide with one binding peptide) was much better than that of a single lytic peptide[24]. Here we constructed several synbodies to test this notion. As shown in Figure 1.8, the B-B synbody control (B represents SA binding peptide without any observed inhibitory effects, B: DRIFHKMQHKPYKIKKRGSC) has no inhibitory effects at uM level, however, both the combination of the binding peptide with one lytic peptide or the combination of two lytic peptides show strong inhibitory effects at the nM level. The best of them, 152-152, shows the MIC against SA at 200nM. It's noticeable that kanamycin can interact with the 30S subunit of prokaryotic ribosomes, which can indirectly inhibit the protein translation and kill the bacteria. This mechanism is different than the antimicrobial peptides, which can cause the membranlytic effects and directly kill the bacteria. This difference might explain why all the constructed synbodies work better than the kanamycin positive control. We also tried to test the influence of synbody scaffolds on its antimicrobial effect. We observed a slightly better performance by using the MAP scaffold than the two GP scaffolds.

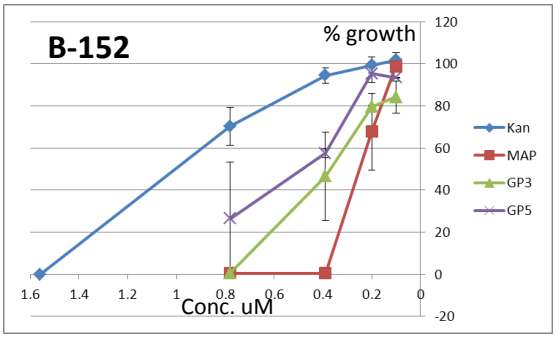
Homo-bivalent SB

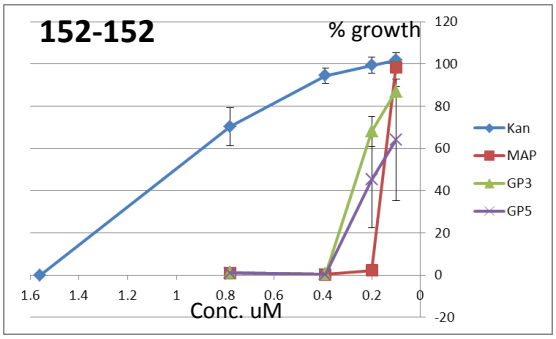
Hetero-bivalent SB

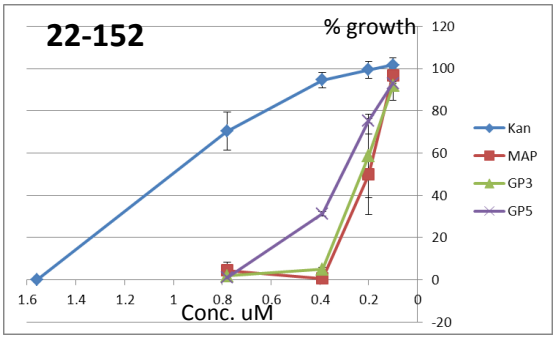


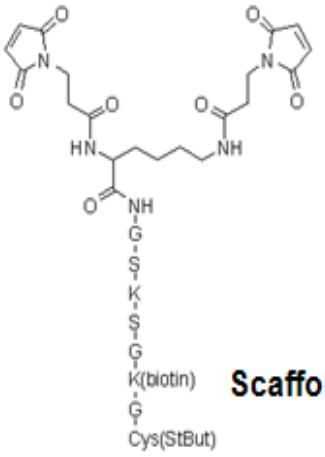




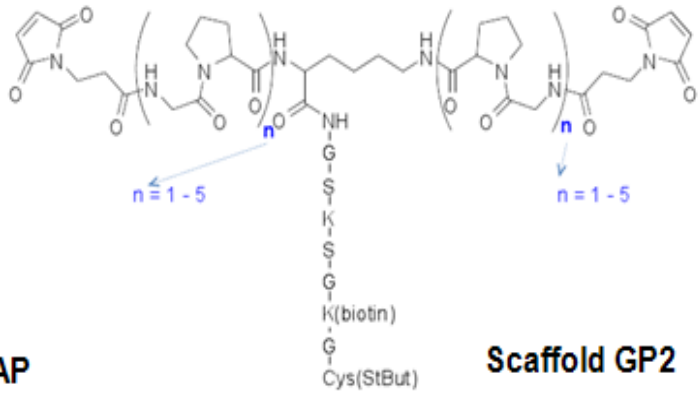








Scaffold MAP



Scaffold GP2

Figure 1.8. Constructed synbodies show significant improvements of antimicrobial effects against SA. Three peptides were used for the synthesis of synbodies: B: DRIFHKMQHKPYKIKKRGSC was selected for its high binding to SA on the peptide microarray, but no inhibitory effect on SA. 22(W2R) and 152(F8K) were two lytic peptides (shown in Table 1.1). The structures of two scaffolds MAP and GP were shown below.

1.4.7 The selection/verification of the synbody effectors binding arm and modulatory arm

As demonstrated in 4.6, combining individual peptides into a synbody is a promising strategy of creating new antimicrobial agent. The concept of a traditional synbody contains one binding peptide, which is used for increase the affinity to pathogen cells, and one lytic peptide, which is used for killing the bacteria. However, the existence of lytic peptide has many drawbacks, including the possible high toxicity. A new design of a synbody effector does not include the lytic peptide; instead, we add a modulatory arm to two binding peptides (Figure 1.9 A,B).

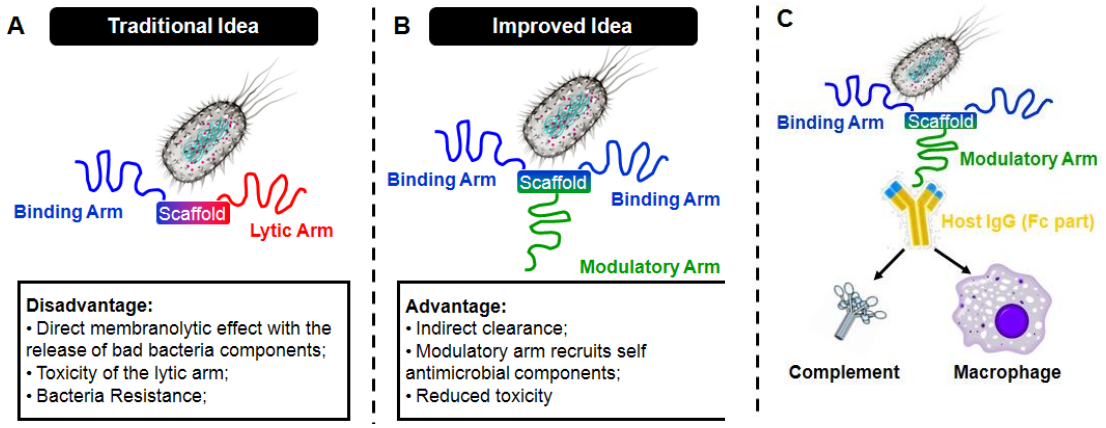


Figure 1.9. Synbody effector: a new concept of performing antimicrobial function. The traditional and new idea are shown as A and B. The deletion of lytic arm from the traditional design might have several advantages. The modulatory arm could be any molecule which can recruit the host self-immunity. One classic example is the Antibody Recruiting Molecules (ARMs), as shown in C. It can recruit self antibody to exert either Antibody-Dependent Cell-mediated Cytotoxicity (ADCC) or Complement Dependent Cytotoxicity (CDC).

One significant feature for cell-extrinsic recognition of pathogens by innate immune cells, for example macrophage, is the physical co-occurrence of the antigen and a PAMP (pathogen-associated molecular patterns) within the same particle. This kind of association ensures the efficient activation of adaptive immunity by the innate immune system. The introduction of a modulatory arm is to mark the pathogens for the innate immune cells and boost the recognition between them.

Different kinds of modulatory arms could be used for this purpose. 1. TLR (Toll-Like Receptor) peptide binders. Most of them are lipopeptides with diacylated or triacylated lipid modification of the N-terimal cysteine residue. TLR1/2 and TLR2/6 are their major receptors. An example is lipoprotein FSL-1-TLR6/2 agonist (Pam2CGDPKHPKSF). 2. N-formyl peptides are important chemotactic factors which are produced by pathogens. They can bind with N-formyl peptide receptor family on innate immune cells, namely FPR (N-formyl peptide receptor), FPRL1 (FPR like-1) and FPRL2 (FPR like-2). Host-derived antimicrobial peptide LL-37 can also bind with FPR. However, LL-37 is too complex to synthesize. There are already some reported N-formyl peptides (both natural and synthetic). 3. Antibody Recruiting Molecules (ARMs). ARMs are kinds of molecules that can act as the antibody Fc part. They will recruit host self antibodies and utilize these antibodies to exert both ADCC and CDC functions (Figure 1.9C). Here, we constructed the NO.3 as a demonstration of the idea.

In order to construct and test a synbody effector shown in Figure 9C, several components are needed: binding arm (can bind with SA, but not inhibit); modulatory arm (cannot bind with nor inhibit SA, high immunogenicity); antibody which can bind with

the modulatory arm; macrophages which are sensitive to antibody Fc part. According to our previous screening, peptide NO.27 (CSGPHPQYRHGPKHYIRTQM) was used as the bacteria binding arm; peptides NO.9 (CSGALTVHKQCHKLGGTVLP), NO.10 (CSGVGWHKSLQQQGPELPPQ) and NO.34 (CSGPHFMFEPSVVRPNYKQA) were chosen as the modulatory arm; peptide ORI (RWRRHKHFKRPHRKHKRGSC) as the positive control. Note: the peptide numbers used here are not from the substitution library creation.

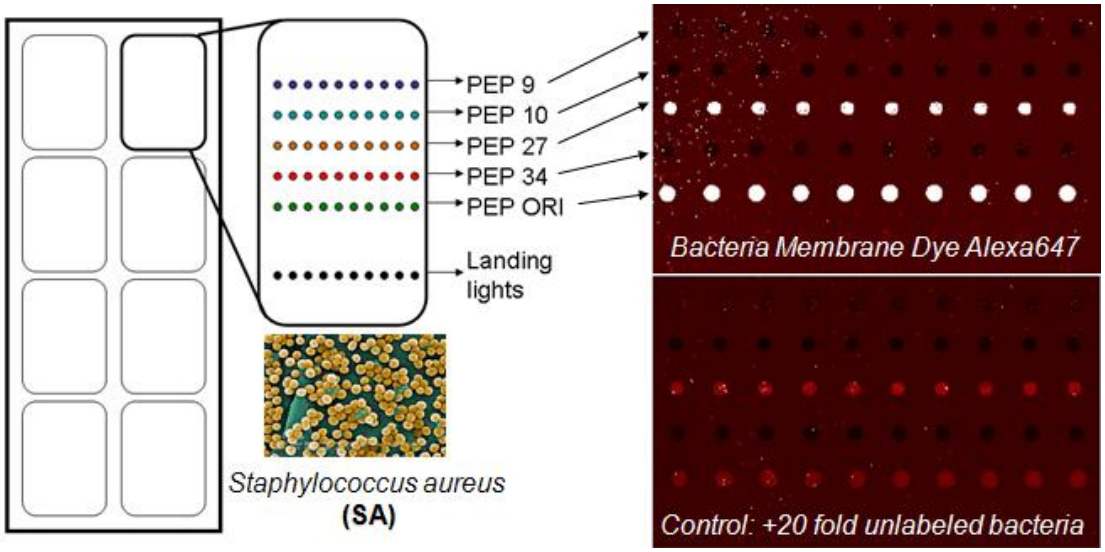


Figure 1.10. Synbody effector binding test. Peptide NO.9, NO.10, NO.34. NO.27. were printed on custom polymer microarray and tested for their bacteria binding capacity.

As shown in Figure 1.10, testing peptides were printed on custom slides as described in the method part. SA was screened on the slides as the pathogen target. The result showed significantly that the NO.27 (the binding arm) bind with the pathogen cells, as well as the positive control (ORI). However, the other three peptides (NO.9 NO.10 NO.34) were not able to bind with the bacteria at all. A 20 fold unlabeled bacteria was also given as a competitive control, which demonstrated the binding specificity.

In the next step, we showed that all the selected peptides had no significant inhibitory influence on SA (Figure 1.11), which was good, since it would not be able to interfere with the following macrophage absorption assay. It is notable that in our MBC test, the ORI showed stronger inhibitory effect against SA than the positive control kanamycin, which corroborated with the fact we observed previously in our synbody test (4.6). Immunized mice also showed significant IgG response against the selected modulatory arms NO9, NO10, NO34 (Figure 1.12).

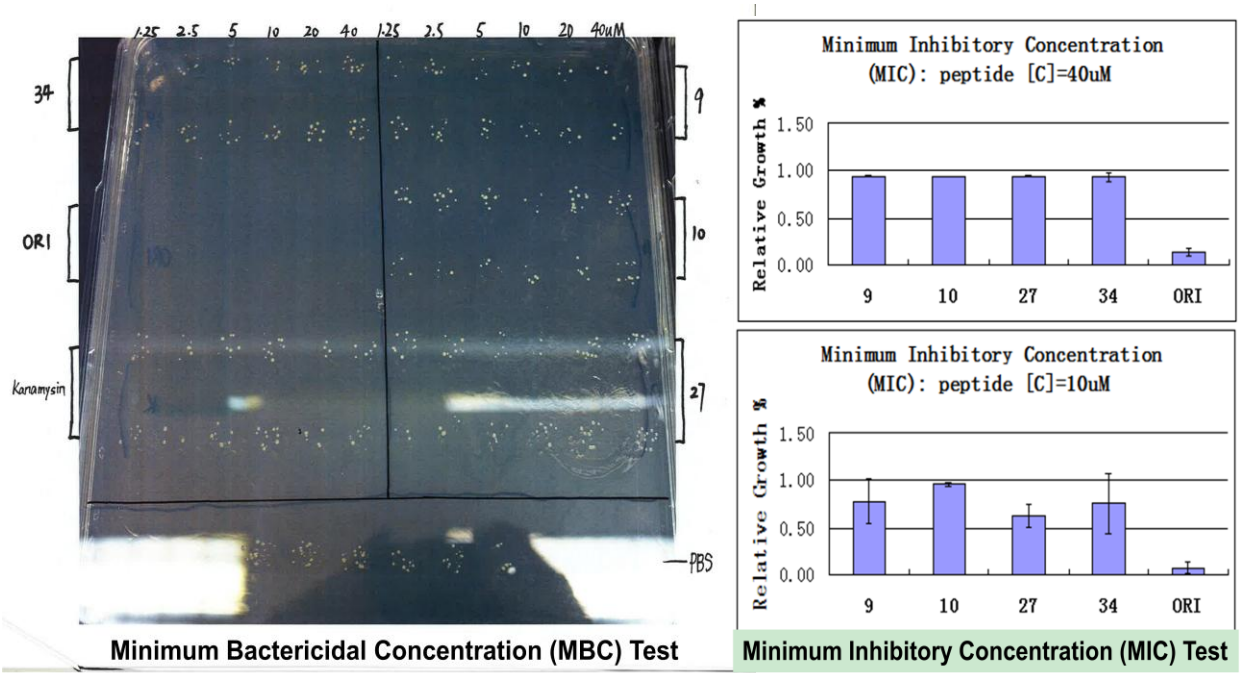
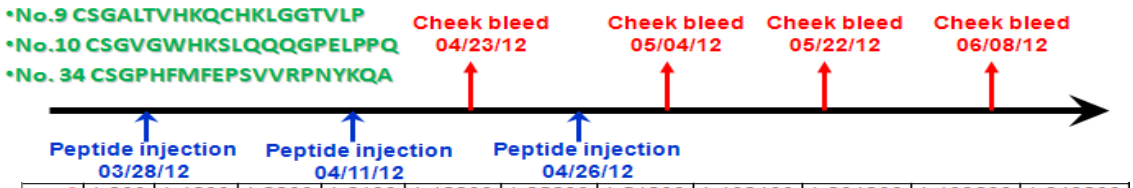


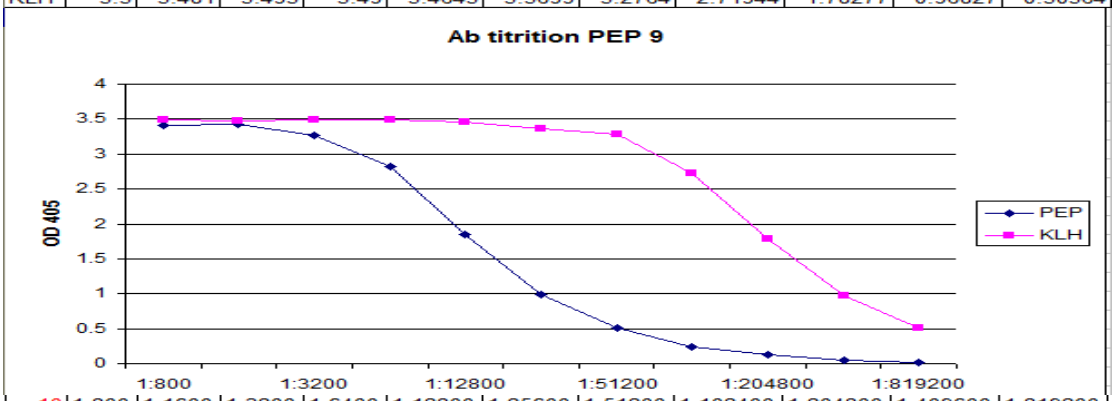
Figure 1.11. Synbody effector antimicrobial test. Peptide NO.9, NO.10, NO.34. NO.27. were tested for their anti-SA effects. Both the MBC test (left) and the MIC test (right) showed no significant anti-SA effects of selected peptides.

1.4.8 Chemical synthesis and purification of synbody effectors

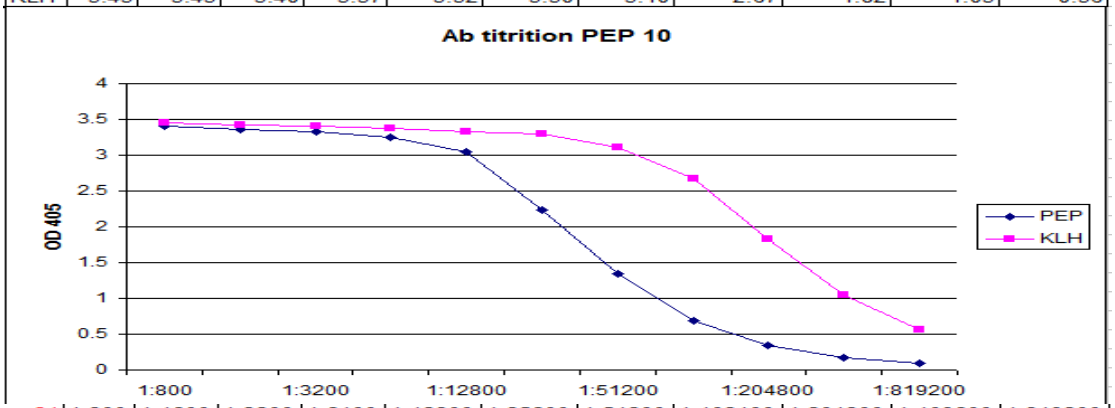
The protocol of synbody effector synthesis is listed in the method part. In general, we firstly constructed the binding part of the synbody effector, which consists of two peptide NO.27 and a GP2 scaffold. Since the GP2 scaffold is protected by a Cys-StBu group, IN the next step we deprotected its Cys to make it activated. In the third step, we connected this binding part with the selected modulatory peptide: NO10, NO34 with a maleimide crosslinker. Finally, we purified the product by HPLC and verified it by both mass spectrometer and silver stain. Figure 1.13 shows the structure of the synbody effector and final synthesis reports. As shown in the silver stain, our final product contained a significant amount of synbody effectors target, though there also existed non-reacted binding part (27-27). The final products were lyophilized and stored at minus 20 degree. The proper amount of product was weighted out and used for test each time.



9	1:800	1:1600	1:3200	1:6400	1:12800	1:25600	1:51200	1:102400	1:204800	1:409600	1:819200
PEP	3.42	3.424	3.272	2.827	1.8512	0.9908	0.5136	0.23705	0.1201	0.0485	0.0209
KLH	3.5	3.481	3.493	3.49	3.4645	3.3699	3.2764	2.71944	1.78277	0.96827	0.50364



10	1:800	1:1600	1:3200	1:6400	1:12800	1:25600	1:51200	1:102400	1:204800	1:409600	1:819200
PEP	3.40	3.36	3.33	3.25	3.04	2.23	1.35	0.68	0.34	0.17	0.09
KLH	3.45	3.43	3.40	3.37	3.32	3.30	3.10	2.67	1.82	1.05	0.56



34	1:800	1:1600	1:3200	1:6400	1:12800	1:25600	1:51200	1:102400	1:204800	1:409600	1:819200
PEP	3.49	3.48	3.40	3.35	3.04	2.30	1.33	0.65	0.31	0.14	0.06
KLH	3.47	3.47	3.43	3.42	3.39	3.31	3.26	2.91	2.25	1.32	0.72

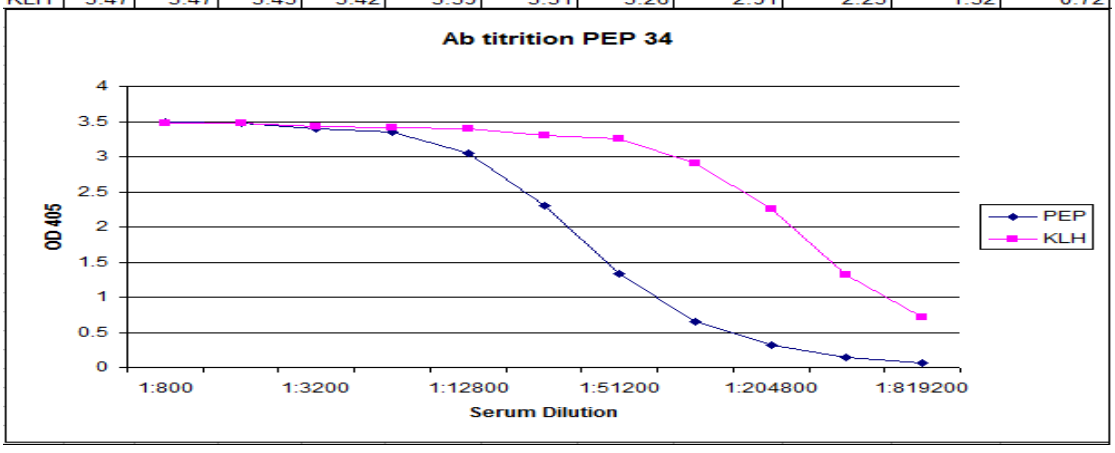


Figure 1.12. Synbody effector modulatory arms IgG end point titration.

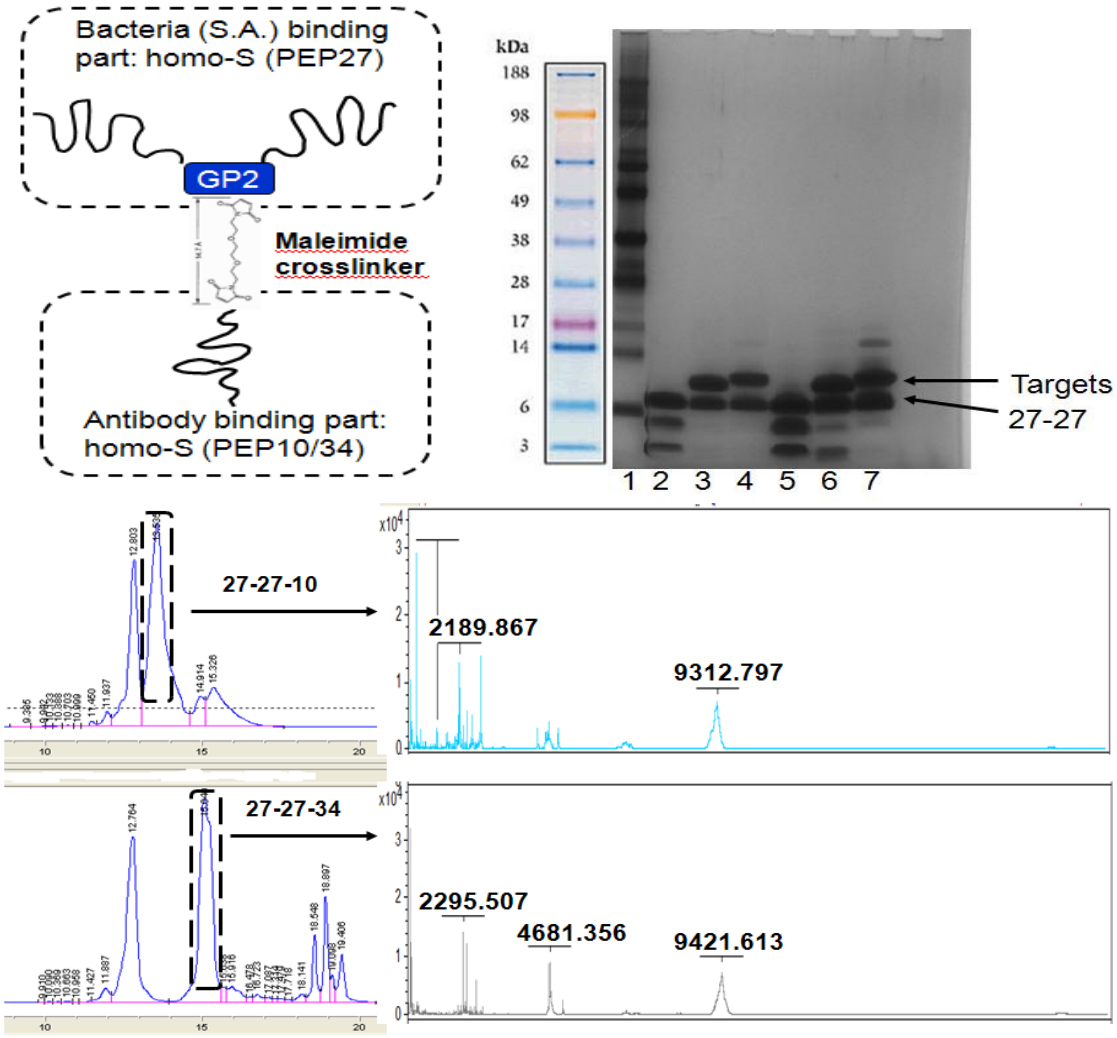
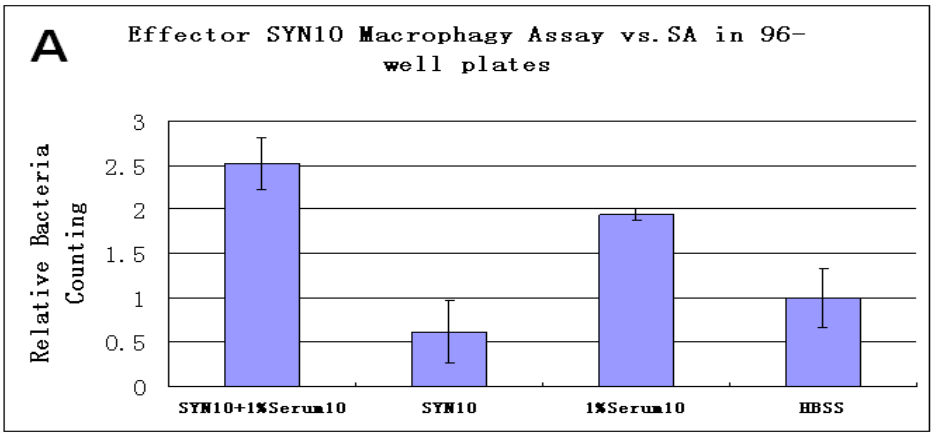


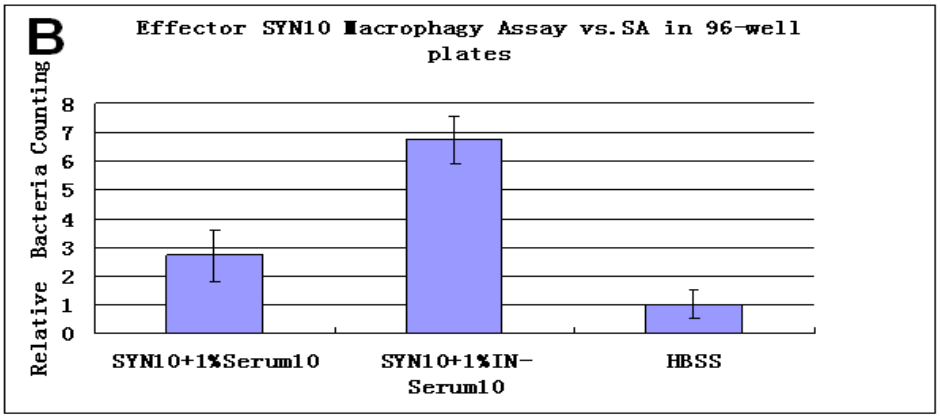
Figure 1.13. Synbody effector synthesis and purification. The structures of synbody effectors (top left) contains the bacteria binding part and the antibody recruiting part. The length of maleimide crosslinker is set up as 14.7 A^{o}. The final product was tested by both silver stain (top right) and HPLC (bot). For silver stain: Marker SeeBlue; 1.27-27(6ug); 2.27-27-34(6ug); 3.27-27-10(6ug); 4.27-27(15ug); 5.27-27-34(15ug); 6.27-27-10(15ug). For HPLC: the target synbodies were concentrated in the second peak with the molecule weight 9312 and 9421. The final product also contained the side products of single (MW 2189 and 2295) and double peptides (MW 4681), which could not be separated by HPLC.

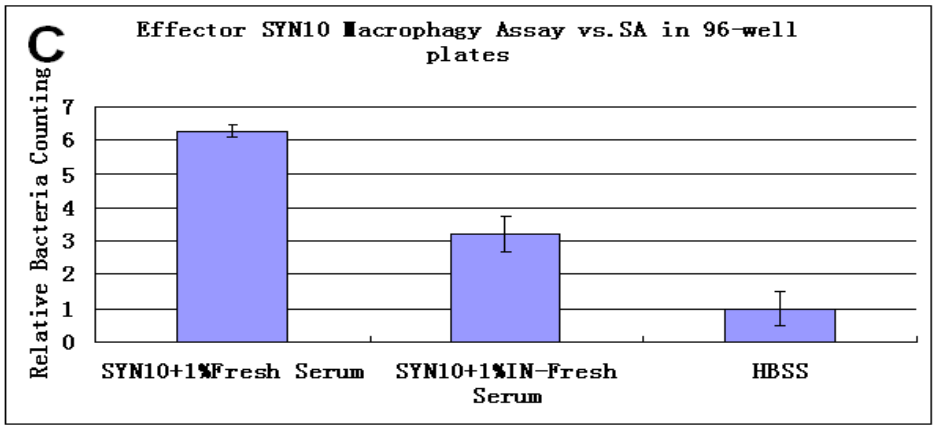
1.4.9 Synbody effectors can increase the macrophage absorption of bacteria

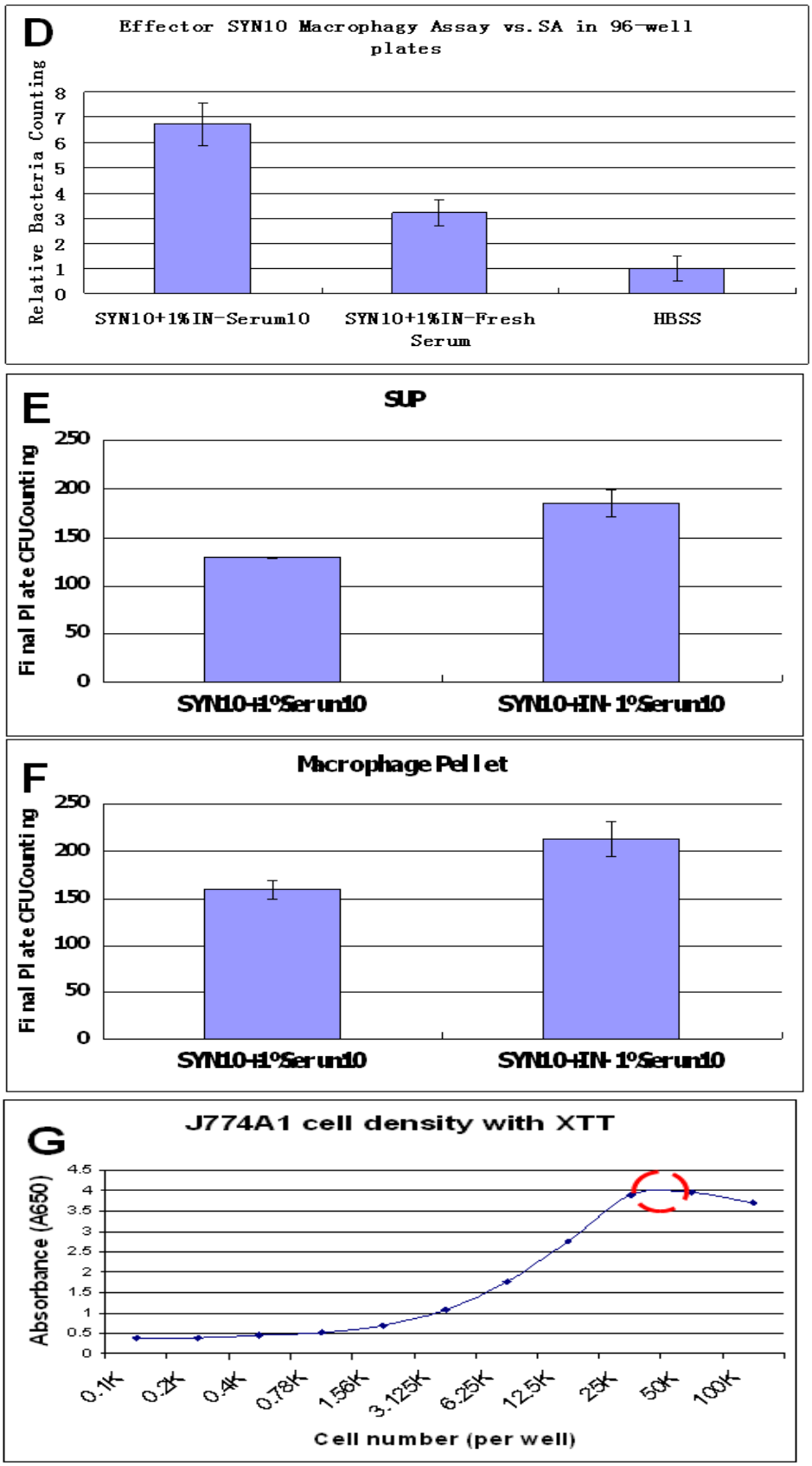
We establish a macrophage absorption array for testing the in vitro performance of the constructed synbody effectors, which are suppose to be able to recruit the antibody and increase the macrophage absorption (not distinguishing between the engulfment and

attachment). Our results showed that synbody effectors can increase the macrophage absorption of SA regardless of the interference of complements in serum. Our result also indicated the possible enhancement of CDC by synbody effectors.








D
Effector SYN10 Macrophagy Assay vs.SA in 96-well plates
Relative Bacteria Counting
0
1
2
3
4
5
6
7
8
SYN10+1%IN-Serum10
SYN10+1%IN-Fresh Serum
HBSS
E
SUP
Final Plate CFUCounting
0
50
100
150
200
250
SYN10+1%Serum10
SYN10+IN-1%Serum10
F
Macrophage Pellet
Final Plate CFUCounting
0
50
100
150
200
250
SYN10+1%Serum10
SYN10+IN-1%Serum10
G
J774A1 cell density with XTT
Absorbance (A650)
0
0.5
1
1.5
2
2.5
3
3.5
4
4.5
0.1K
0.2K
0.4K
0.78K
1.56K
3.125K
6.25K
12.5K
25K
50K
100K
Cell number (per well)

gureFigure 1.14. Synbody effector macrophage absorption assay. All assays: [C]SYN10=40uM. A. Correct combination "SYN10+Serum10" increases the macrophage adhesion/ phagocytosis; B,C,D. The possible interference by complements in the serum. E,F. The enhancement of CDC by synbody effecters. G. The starting macrophage density was determined as 50K/well (red circle) by the XTT test. SYN10: synbody effecter with modulatory arm peptide NO.10; Serum10: mouse serum immunized by peptide NO.10; IN-: heat (56 ℃) complement inactivated serum; Fresh serum: non-immunized mouse serum; Relative bacteria counting: the relative colony plate counting normalized by the HBSS control.

As shown in Figure 1.14A, the correct combination of synbody effector NO.10 and its corresponding serum NO.10 gave us the highest macrophage absorption. Complement in the serum played a complex role in this experiment: on the one hand, it could increase the macrophage assay result by its opsonization function; on the other hand, it could also decrease the assay result by killing the bacteria cells (Alternative Pathway and classic pathway). The results in Figure 1.14B and Figure 1.14C indicated that the synbody effector might trigger the complement classic pathway and decrease the assay result. In order to eliminate the complement influence, we used heat (56 ℃) to inactivate the complements in the serum. Figure 1.14D shows that the synbody effector treated group significantly increased the macrophage absorption. Figure 1.14E,F shows the total bacteria number of SYN10+1%Serum10 is lower than that of SYN10+IN-1%Serum10. This probably suggests that the complement classic pathway helped kill more bacteria cells.

1.4.10 Use 10K Ver.3 microarray to select the pre-existing Ab peptide binders

In our previous in vitro assay, we established an artificial test system including the modulatory peptide and the corresponding antibody produced against it. Here we show that without the pre-immunization, the antibodies in non-immunized mouse serum

are also able to bind with certain random peptide targets on microarray (Figure 1.15). In order to pick up the peptide binders caused mainly by affinity rather than avidity, that is the spot signal intensity is caused by the sum of individual peptide-antibody bonding, but not the combined synergistic strength of all the peptides in the same spot, we used the NSB27 slide, which was designed to have 9nm distance between each peptide in one spot. Two incubation times were tested, 1 hour and 16 hours. As we observed, 16 hours incubation displayed a much wider dynamic range than 1 hour incubation, which means the binding signal in 1 hour is mainly caused by avidity rather than affinity. The real affinity pairs stand out when the incubation time became longer (16h). Some top examples are shown here. It is notable that all peptides on 10K Ver 3 microarray contain D-amino acids. In the future, we could choose some of them as the candidates of synbody modulatory arm.

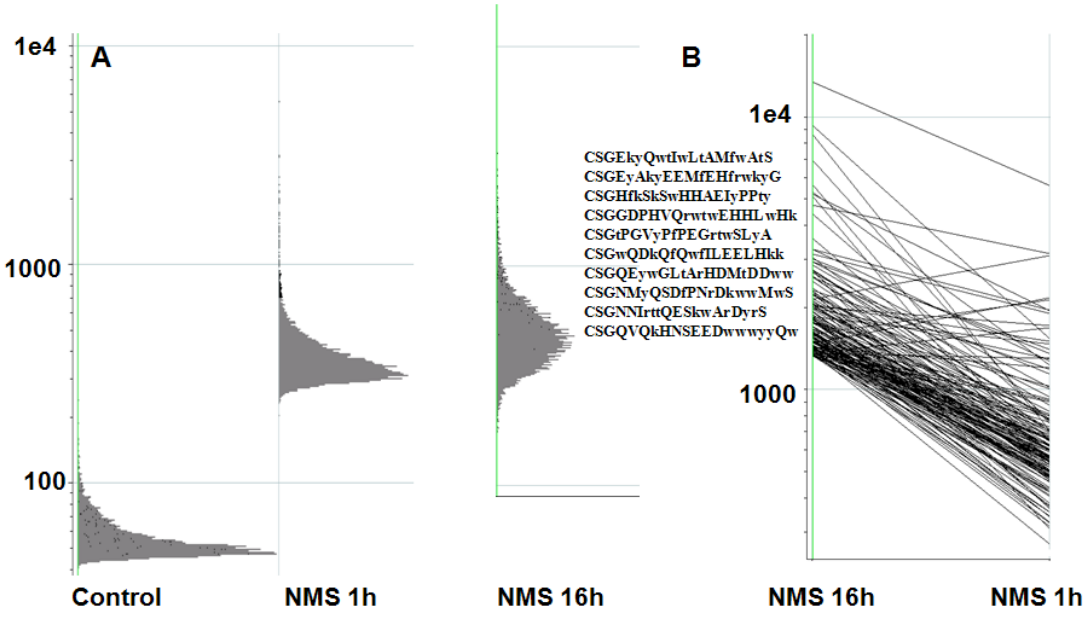


Figure 1.15. 10K NSB 27 Ver.3 microarray screening for a normal mouse serum pool. A pool of ten naïve BALB/c mouse serum samples was screened on the 10K NSB 27 Ver.3 microarray. A: Compared with the non-serum control, both 1 hour and 16 hour incubation showed significant antibody binding signal. NMS: naïve mouse serum; B: Left: the top 154 intensity spots with 16 hour incubation. Right: the corresponding spots intensity with 1 hour incubation.

1.5 Discussions

Antimicrobial peptide (AMP) represents an old, host native defense mechanism, which exists in virtually all kinds of multicellular organisms. Generally speaking, natural AMPs derived mostly from invertebrates and plants have the advantages: wide spectrum antimicrobial capacity, less bacteria-resistance occurrence chance and possible immunomodulatory effects. However, the disadvantages such as low specificity, relative high hemolytic activity and high discovery costs of synthesis and screening become the hurdles of transferring them into therapeutically valuable agents[25]. Numeral strategies have been used to make improvements, for example, rational design of the peptide α helicity /hydrophobicity /charges[26,27]; truncation of existing peptide sequences[25]; dendrimeric peptide constructions[28].

Here, we demonstrated that the creation of a peptide single variants library helped to improve the synthetic antimicrobial peptide RWRR we generated from our random peptide microarray. One of the selected single variants, MUT165(K9d-K) showed improved efficacy against PA but not SA. This finding supports the notion that the diastereomeric analogs have different impacts on different AMPs[21], and it indicates that the specific secondary structure-function relationship might be important for some AMPs but not the others. The selected single variants were further combined together to make either double variants or bivalent synbodies. For the double variants, we have shown the

best candidates MUT22&222* with the minimal inhibitory concentration (MIC) of 2.5μM (SA) and <5μM (PA), which are among the top of previously literature reported AMPs[29,30].

Traditionally, people in the field pay more attention to the AMPs ability of distinguishing mammalian cells and pathogen cells. Recently, more studies now focus on the selectivity within the pathogen cells (bacteria versus fungi[27]; Gram positive bacteria versus Gram negative bacteria[21]). This is an important issue since broad-spectrum antibiotics will more easily give rise to the drug resistance, so it will be best to narrow down the peptides' toxic effect solely on the specific pathogen cells but not the other gut flora. Besides the antimicrobial potency improvements, we also observed that the selectivity between SA and PA can be significantly shifted by either single or double mutation (the specificity index ranges from 0.75 to >8). It's noticeable that our lead peptide RWRR initially has better inhibitory potency against SA than PA (specificity index=5.33), this might explain why there are less changes for SA than PA. We can also create the substitution library based on a lead peptide having the same potency versus SA and PA (for example, peptide HRPRKKFHKFPRKWRRHGSC has 30μM MIC for both SA and PA, unpublished data), in this case, we expect to observe more diverse specificity changes.

On the other hand, we provided the evidence of constructing the synbody effector as an antibody recruiting molecule. As shown by our demonstration, the synbody acted as a bivalent linker between the pathogen cells and the antibodies, and then the antibody Fc part could potentially increase the macrophage phagocytosis. This reminds me of some other peptide-based immunomodulators, for example, tuftsin and soymetide-13, which

are able to stimulate macrophage and neutrophil phagocytosis, respectively. Conversely, a computationally designed peptide from CD47 was recently shown to inhibit phagocytic clearance and enhance delivery of nanoparticles[31]. In the future, it's valuable to add these peptide immunomodulators as part of synbody effector to check the synergetic effect.

Chapter 2

SERODIAGNOSIS BY IMMUNOSIGNATURE: A MULTIPLEXING TOOL FOR MONITORING THE HUMORAL IMMUNE RESPONSE TO DENGUE INFECTION

2.1 Introduction

Dengue virus (DENV) is a mosquito-transmitted virus belonging to the genus Flavivirus. It is endemic in more than one hundred tropical and subtropical countries around the world. Although it has been studied for more than fifty years, currently there is no specific treatment or vaccine available for dengue infection or significant strategies to control dengue transmission. As the control and eradication strategies of dengue mainly rely on rapid and effective laboratory diagnosis, in this chapter we try to use an antibody profiling assay, which is based on the 10K random peptide microarray technique in our lab, to monitor the humoral immune response of dengue. We showed that based on IgG, IgM or IgA immunosignaturing, dengue infected patients can be distinguished from healthy individuals, and there's a potential difference between the dengue primary and secondary infection. Furthermore, we can also detect the signature differences with some other related arthropod-borne viruses, such as West Nile Virus (WNV), and from malaria which confounds existing diagnostics. Finally, we checked the

possibility to use the 10K microarray, as well as a de novo made NS1 dengue protein subarray to differentiate the four dengue serotypes.

2.2 Background and significance

2.2.1 Dengue and dengue infection

Dengue is the disease caused by any one of four closely related dengue virus subtypes (DENV 1, DENV 2, DENV 3, or DENV 4). It's one of the most rapidly worldwide-spreading arthropod-borne/ mosquito-borne viral infections. During the last 50 years, the number of impacted districts and infected people increased dramatically. According to CDC, about one-third of the world population is currently living in areas at risk of dengue (including the South-East Asian Region; Western Pacific Region; Central/ South America and African Region). Most reported cases in United State are imported from the tropic endemic areas, however, outbreaks were also reported in Hawaii, Texas and most recently Florida. Dengue has been included in 2005 World Health Assembly resolution as an example of disease that might require a public health emergency of international concern beyond national borders[32].

Dengue viruses is a member of small enveloped single-strand RNA(+) flavivirus. It's mostly transmitted by mosquitoes of the genus Aedes. Dengue infections have a wide range of clinical outcomes from self-limiting no symptoms to dengue fever to life-threatening dengue hemorrhagic fever (DHF)/ dengue shock syndrome (DSS) with the mortality rate as high as 40%[33]. A typical severe infection includes the febrile phase (Day2-Day7), which is indistinguishable from the other influenza-like illness (ILI); critical phase (Day3-Day7) with increased hematocrit and plasma loss; recovery phase[32].

All four human dengue serotypes have the similar clinical manifestations, however, there's no long-term cross protective immunity between each other. What's even worse, the limited cross-reactive immune response to other serotypes increases the risk for developing DHF and DSS when the patient gets the secondary infection, a phenomenon commonly attributed to antibody dependent enhancement (ADE)[34]. It's noticeable that there's no approved vaccine or specific treatment for dengue infection, and little do we know about its pathological immune responses in human.

2.2.2 Current dengue diagnosis

The accurate and efficient laboratory diagnosis of dengue is of great importance for clinical concern (early detection), academic research, vaccine development and surveillance activities[32]. An ideal dengue test has the specifications such as "distinguish between dengue and other diseases with similar clinical presentations", "distinguish between dengue serotypes", "distinguish between the primary and secondary infection",etc[35].

Currently used diagnostic methods include traditional viral isolation and identification, viral RNA detection by PCR, antigen detection by NS1 assay, immunohistochemistry test for infected tissues and serological tests based on either hemagglutination-inhibition(HI) or ELISA(IgM/IgG/ IgA). The direct detection of either virus or viral antigens usually takes longer time and requires expertise and appropriate facilities. Since viral particles and antigens are usually cleared in historic samples, it's also not possible to use the direct methods for measurement. The indirect antibody detection techniques, especially the anti-dengue IgM based technique has become the

most widely used serologic test, for example the standard MAC (immunoglobulin M capture) ELISA assay. Different commercially available anti-DENV IgM tests have quite distinct sensitivity and specificity. Their general limitations include the inability to detect secondary infection samples with low IgM titers; inability to identify the dengue serotypes; potential false positive with other infections caused by antibody cross-reactivity (such as West Nile virus, St.Louis encephalitis and Malaria[36]) and potential delayed diagnosis (detectable IgM antibody six to ten days after onset[37]). In order to tell different dengue serotypes, RT-PCR is usually used. For example, the CDC DENV-1-4 Real Time RT-PCR Assay is approved by FDA to test the acute phase dengue infection within 5 days, but not those samples taken 5 or more days after the onset of symptoms. Another "gold standard" test, Dengue Plaque Reduction Neutralization Test (PRNT), is also used to characterize and quantify circulating levels of anti-DENV neutralizing antibody (NAb). WHO has already published a guideline for the PRNT test. However, PRNT is very labor intensive and therefore not amenable to be high throughput and large scale usage[38,39].

What is more important, none of the methods mentioned above are multiplexing, which means the negative result could not be interpreted into other diseases. In order to reduce the burden of diagnosis, a multiplexing diagnostic tool is needed especially in the areas with the other co-circulating arboviruses. Multiplexed microsphere-based flow cytometric assay has the potential to detect multiple viruses simultaneously, for example, different influenza viruses[40], West Nile virus and St.Louis encephalitis virus[41]. Microarray based techniques, including both DNA microarray, peptide microarray and protein microarray are also promising to detect different infectious diseases in parallel,

for example the peptide array created by Andresen et al to distinguish closely related herpes viruses and hepatitis c virus[42] and the proteome array designed to detect different infectious gram-negative bacteria[43].

2.2.3 10k random peptide microarray and immunosignaturing

Peptide microarray technology, as one kind of Small Molecule Microarrays (SMMs) has been widely used to expand our understanding of humoral immune responses to diseases in a high-throughput way[44]. These attempts include 1.Biomarker identification of M.tuberculosis (MTB)[45], SARS[46,47] and Alzheimer's disease[48]; 2.Epitope mapping for food allergens, such as those from lentil[49] and milk[50]; 3.Immunodiagnosis of Echinococcus Spp.[51], Toxoplasma gondii infection[52] and HIV[53]. However, none of the microarray platforms designed above is multiplexing, which means it's impossible to use one for testing another.

Our group has come up with the idea of creating a random peptide microarray as a universal diagnostic tool, which is not specifically designed for any pathogens, however, is able to generate distinct signal patterns-immunosignatures[54] to distinguish different kinds of humoral immune responses. The capacity of our array has been tested and reported on influenza (mouse)[55], valley fever[54], Alzheimer's disease and brain cancer[56]. One general benefit of using peptide-based microarray other than ELISA or a protein-based microarray is it is more distinguishable signal output[42]. The output of peptide microarray is based on each individual kind of antibody but not the sum of a heterogeneous family of antibodies, as in an ELISA. So we expect to tell even subtle changes in the whole antibodyome. However, since the peptides we used on the

microarray are all 17-mers, it's quite possible that certain peptides contain more than one epitope/ mimotope, thus can bind with different antibodies.

ELISA and epitope microarraies always contain the real antigens coated for detecting the specific antibodies, however, the peptides on the random peptide microarray have no intended homology to any specific pathogen antigens, which makes it difficult to track back to the real proteome. Currently, we do not have a good strategy to decipher the information on 10K microarray, due to its relatively low resolution.

2.3 Materials and methods

2.3.1 Serum sources: dengue, West Nile Virus and Malaria

Dengue infected patient sera samples tested here are from several different sources. One is the **PVD201** (Seracare, 20 dengue samples plus 1 normal control, it's notable that the real value of all these individuals is determined by ELISA, there's no clinical information for the patients. Since different ELISA methods give us different results, the individuals with only a few tests showing positive but not the others have the possibility to be false positive), Figure 2.1. By the definition of Panbio ELISA IgG/IgM Duo Assay, we are able to identify three of these samples (PVD201-10, 16, 20) as primary infections and the others as either secondary infections or non-infectious. Another source is Panel Dengue **DS689G** (Seracare, 5 samples, for these samples, we only know their IgG titers), **DS689M** (Seracare, 8 samples, for these samples, we only know their IgM titers), both PVD201 and DS689G/M samples were kindly provided by the Lawrence Livermore National Labs (LLNLs). The third one is the dengue **WHO reference panel** from NIBSC (four monovalent anti-sera collected from single donors

and confirmed by 50% PRNT test: 02/300 (anti-DEN 1), 02/296 (anti-DEN 2), 02/274 (anti-DEN 3), 02/298 (anti-DEN 4). One negative control plasma pool comprising of pre-bleeds of the Japanese Encephalitis vaccines 02/184 (Negative control human serum)). The other patient samples tested here include three malaria positive plasma (Seracare DS-774) and West Nile Virus (WNV, 10 samples from different Seracare panels). The normal samples were collected from laboratory volunteers, and were randomly chosen for testing.

2.3.2 10K random peptide microarray screening procedures

The general processing, alignment and data extraction is based on the Standard Operating Procedure for Binding Antibodies to the CIM10K. A brief description is listed below:

a. 10K Ver.2 slides preparation. In order to reduce the technical errors caused by batch-to-batch differences, we tried to use the slides from the same batch or the same spotting plate for each experiment. Slides are firstly pre-washed by the pre-wash buffer (7.33% acetonitrile, 33% isopropanol and 0.55% triflouroacetic acid in nanopure water), then manually washed by 1xTBST three times and then ddH2O three times. The were then dried by centrifuge (800g 5min).

b. Binding of Antibodies to CIM10K by Tecan Automated Slide Processing Station. **Blocking buffer:** 5 ml of 30% BSA, 6.9 ul of Mercaptohexanol, 25 ul of Tween 20, 5 ml 10x PBS, 40 ml ddH20. Slides are incubated with the blocking buffer for 1 hour at 27^{o}C; **Incubation buffer**: 5 ml of 30% BSA, 25 ul of Tween 20, 5 ml 10x PBS, 40 ml ddH20. Primary, secondary and tertiary are all diluted in the incubation buffer, and

Multiple IgG Tests

Sample	Panbio Dengue DuoCassette	BIOLINE Dengue Duo	Calbiotech ELISA	EUROIMMUN ELISA	Focus DxSelect	Panbio ELISA	SD ELISA
PVD201-01	NEG	POS	1.5	4.9	6.2	2.4	>8.9
PVD201-02	NEG	NEG	1.8	5.3	5.7	1.2	1.6
PVD201-03	NEG	NEG	2.2	4.8	7.1	1.3	1.1
PVD201-04	NEG	NEG	1.7	2	3.6	0.5	0.5
PVD201-05	NEG	NEG	2.1	5.1	7	4.1	1.9
PVD201-06	NEG	NEG	2.3	4.4	6.3	4.5	3.4
PVD201-07	POS	POS	1.3	4.7	7	7	8.8
PVD201-08	NEG	NEG	2.3	4.3	6.5	2.6	1.2
PVD201-09	NEG	NEG	2.1	4.4	5.7	0.7	2.1
PVD201-10	NEG	NEG	1.3	3.8	2.2	0.4	2.5
PVD201-11	NEG	NEG	2.1	4.2	4.8	0.5	1
PVD201-12	NEG	NEG	2.1	3	5.3	0.9	0.7
PVD201-13	NEG	NEG	2	4.7	5.8	1.3	2.1
PVD201-14	NEG	NEG	0.4	0.1	0	0	0.1
PVD201-15	NEG	NEG	2.4	3.7	5.2	1.3	0.8
PVD201-16	NEG	POS	1.4	4.1	2.5	0.7	2.4
PVD201-17	NEG	POS	1.5	5.2	5.9	2.3	>8.9
PVD201-18	NEG	NEG	1.9	3.1	5.4	2.3	0.7
PVD201-19	NEG	NEG	2.2	3.8	5.8	0.6	0.5
PVD201-20	NEG	POS	1.2	4.4	2.2	0.7	3
PVD201-21	NEG	POS	1.7	5.4	6.8	4.3	>8.9
Dengue IgG 9240594			2.308				
Dengue IgG 9240291			2.203				
Dengue IgG 9242886			2.49				
Dengue IgG 9242893			2.586				
Dengue IgG 9242911			2.494				

Multiple IgM Tests

Sample	Panbio Dengue DuoCassette	BIOLINE Dengue Duo	Calbiotech ELISA	EUROIMMUN ELISA	Focus DxSelect	InBIOS ELISA	Panbio ELISA	SD ELISA
PVD201-01	POS	POS	1.7	1	2.8	5.64	>5.8	6
PVD201-02	NEG	NEG	0.3	0.5	0.2	1.03	0.2	0.4
PVD201-03	NEG	NEG	0.4	0.3	0.8	1.09	1.1	0.3
PVD201-04	NEG	NEG	0.1	0.1	0.6	0.97	0.4	0.2
PVD201-05	NEG	NEG	0.2	0.3	0.5	1.38	0.8	0.5
PVD201-06	NEG	POS	0.8	1.1	0.9	1.7	0.7	2.6
PVD201-07	POS	POS	2	1.7	3.2	25.05	>5.8	>7.7
PVD201-08	NEG	NEG	0.6	0.7	0.5	1.24	0.9	1
PVD201-09	POS	NEG	0.9	1	0.5	1.71	0.3	1.2
PVD201-10	NEG	POS	1	1	2.8	6.85	>5.8	>7.7
PVD201-11	NEG	NEG	0.2	0.4	0.2	1.17	0.2	0.4
PVD201-12	NEG	NEG	0.2	0.2	0.3	1.08	0.2	0.4
PVD201-13	NEG	NEG	0.2	0.3	0.3	1.26	0.3	0.6
PVD201-14	NEG	NEG	0.2	0.1	0.2	1.06	0.3	0.3
PVD201-15	NEG	NEG	0	0.2	0.6	1.18	0.4	0.3
PVD201-16	POS	POS	0.7	0.2	0.3	4.51	>5.8	6.8
PVD201-17	POS	POS	1.6	1	2.6	5.23	>5.8	5.5
PVD201-18	NEG	NEG	0.4	0.4	0.4	1.39	0.3	0.8
PVD201-19	NEG	NEG	0.1	0.1	0.9	1.42	0.8	0.4
PVD201-20	POS	POS	0.9	1	2.5	6.08	>5.8	>7.7
PVD201-21	POS	POS	1.6	3.9	2.9	6.47	>5.8	3.9
Dengue IgM 9254165						29.23		
Dengue IgM 9254166						8.04		
Dengue IgM 9254167						7.29		
Dengue IgM 9254539						7.42		
Dengue IgM 9254540						6.96		
Dengue IgM 9254543						10.58		
Dengue IgM 9253421						6.93		
Dengue IgM 9253425						4.91		

Figure 2.1. Dengue ELISA results provided by Seracare. TOP: IgG tesets, the signal to cutoff ratios(s/co) >1.0 is considered to be positive and marked in green; BOT: IgM tests, for InBIOS ELISA, the ratios(s/co)>2.84 are considered to be positive, for the other tests, the ratios(s/co)>1.0 are considered to be positive; The country of origin: Colombia (PVD201-01,07,17) Honduras (PVD201-02,03,04,05,06,08,09,11,12,13,15,18,19) Ecuador (PVD201-10,16,20,21) Control United States (PVD201-14)

incubated for 1 hour at 37^0C. For the IgG detection, we use 1:500 diluted sera as the primary, 5nM IgGh andlnM biotin-labeled goat anti-mouse as the secondary, 5nM Alexaflour647 labeled streptavidin as the tertiary. For the IgM detection, we use 1:500 diluted sera as the primary, 5nM DyLight 649-conjugated affiniPure Rabbit Anti-Human IgM Fc5u Fragment as the secondary. For the IgA detection, we use 1:50 diluted sera as the primary, 1nM Alexa Fluor 647-conjugated AffiniPure Goat Anti-Human Serum IgA as the secondary.

c. The slides are scanned by Agilent scanner at 647nm. Alignment is done semi-automatically in GenePix Pro 6.0. The slides with good correlation between duplicates are chosen for analysis. Those with bad correlation will be repeated to achieve the satisfied correlation (>80%, calculated by a custom QCPC application)

2.3.3 Statistical analysis by GeneSpring

Genespring GX 7.3 from Agilent Technologies is the major tool we use to perform our statistical analysis. All the aligned microarray data were imported into the Genespring, every spot on the microarray is addressable to one random peptide sequence. **Data Normalization**: any measurements less than 0.01 (cutoff value) were set to the 0.01. We usually assumed that most of the features on the microarray were unaffected by the disease conditions compared with the normal control, so all of the measurements on each chip were divided by a specified percentile value (50.0% by default), denoted by Normalized Data. For those cases which most of the features on the microarray were affected by the disease conditions, this step was neglected, denoted as Raw Data.

Parameters: Experimental parameters were set up as individuals, such as Dengue PVD20101, Normal 158; groups, such as Dengue PVD201, Normals.

Principal Component Analysis (PCA) is a tool for reducing the complexity of multi-dimensions data by doing covariance analysis and discovering a number of principal components that define most of the data variability. PCA is not a clustering tool, but it can roughly show the relationships between different conditions. In this study, PCA (on conditions) is performed; each individual sample is a condition (a dimension) and is compressed and plotted on either a two-dimensional or three-dimensional scatter dot figure, the distance between each spot represents the difference between them.

Feature selection: several methods in Genespring were used to select the features to distinguish different groups, such as the dengue infection vs. normal. The purpose of **One-way ANOVA** is to test for a significant difference in the medium expression levels between two or more selected groups. In other words, the peptide features with the least p-value means they are expressed differently by at least one of the groups analyzed (Parametric test, assume variances equal, the false discovery rate is set up at least 0.05. No multiple testing correlation). **Filter on Volcano Plot**: In many cases, we just need to compare two different conditions, for example dengue infected serum with normal serum. Volcano plot is a convenient way to display both fold-change values and p-values interactively. We usually set up the fold change=2, and pick up the top 20/50/100 features by the p-value. **Draw expression profile**: for the cases in which most of the features are significant (p-value<0.05) between different conditions, or those just have one individual sample (for example, our WHO dengue panel contains just one individual sample for each serotype 1/2/3/4), we use the “draw expression profile function” to select the

features expressed higher in one or more conditions. Basically, the "Draw expression profile" creates an artificial expression line among all the other real features, and then the "Find similar genes" function let us find the features with similar expression pattern with this artificial line (for example, higher in dengue serotype 1 but not the others). The similarities between features are measured by Pearson correlation, Spearman correlation or Distance. The selected features can be shown as a line graph, bar graph, box plot, or scatter plot.

Clustering: Hierarchical clustering is done for both genes (selected peptide features) and conditions (dengue/ WNV/ Malaria/ Normal), which are two independent processes. The clustering is solely based on the feature expression, the features with similar expression patterns will be put close to each other. The similarities between features are measured by Pearson correlation, Spearman correlation or Distance with calculated confidence levels (bootstrapping). The color bar expression range (high/normal/low expression) is set up manually to achieve the best visual contrast. The clustering pattern is determined by both the peptide features used and the conditions (individual samples) selected. Note: We did not intentionally exclude any conditions in clustering, and the clustering result just shows the similarity between each condition, which means there's no cut-off line between any two conditions.

Classification: for the dengue IgG immunosignature, we separated it into training group and testing group, the significant features selected from the training group were used to classify the individuals from testing group. Weka, an open source software containing many machine learning algorithms for data mining, is used for this purpose. Naïve Bayes method is used to make 10-fold cross validation of selected samples.

2.3.4 Custom NS1 subarray creation

The custom NS1 subarray was created based on the dengue non-structural protein 1 (NS1). First, the amino acid sequences of 22 dengue NS1 proteins (3 serotype1; 9 serotype2; 5 serotype3; 5 serotype4. Obtained from NCBI taxonomy browser) are aligned by MEGA 5.10 ClustalW. After the alignment, all the sequences are chopped into non-overlapping 17-mers, and are exported into Excel. Note: every 17-mers has a tag indicated the dengue strain/serotype it was generated from. Second, all the 17-mers are chopped into continuous 5-mers, 6-mers, 7-mers….11mers, Third, all the 5-mers are compared with each other, and the 17-mers containing the 5-mers, which belong to more than one dengue serotypes are deleted. The remaining 17-mers are called the selection at "**5-mers level**". We can get the similar selections at "6/7/8/9/10/11-mers level". Note: 5-mers level is the strictest criteria for selection, while 11-mers is the loosest criteria for selection. Finally, the selections at 10-mer level (58 17-mers) combing with 7 real NS1 epitopes (reported by IEDB) were sent to Sigma for synthesis. A CSG- linker was added to each 17-mers. The NTerm and CTerm are [H] and [NH2], respectively.

The ordered peptides arrived as lyophilized power. They were firstly dissolved in 30% acetonitrile to make the concentration 2mg/ml. Microarray spotting buffer: 40mM HEPES, 20mM TCEP, 2mM EDTA, PH=7.5. The peptide stock and the spotting buffer were mixed together by half and half. The final peptide spotting concentration is 1mg/ml. The NSB 9 slides were used for spotting and activated by SMCC, an amine-to-sulfhydryl crosslinker. Printing was processed automatically by Nanoprint 60. All the peptide were printed three times, and randomly located on each array. The screening procedures of the NS1 subarray is similar with that of 10K array: 1:500 diluted sera as the primary

incubation for 16 hour; 5nM Alexa Fluor 647-conjugated AffiniPure Goat Anti-Human Serum IgG as the secondary for 1 hour. In order to decrease the cross-reactivity, all the incubation temperature here is 27 degree rather than 37 degree.

2.4 Results

2.4.1 CIM 10K Ver2 microarray can distinguish the dengue serum from the normal serum by the IgG/IgM/IgA immunosignature

The first fundamental question we would like to ask is whether by using immunosignaturing, we could separate the dengue infected serum samples with the non-infected normal serum samples. Here, we do not consider the potential differences in the dengue serotypes/ genotypes, primary/ secondary infection, sample collection time points, clinical outcomes etc. We just want to see whether there's a particular "dengue pattern", which is common among the dengue variants we tested.

Besides the IgG immunosignature, we also tested the IgM and IgA immunosignature of dengue infections in two independent experiments, each with its own sample set and secondary antibody. In fact, the IgM ELISA test is more widely used for dengue serologic tests than IgG, especially for the primary infection early detection. IgA ELISA tests for dengue are currently being developed by several groups[57] due to the fact that dengue specific IgA appears one day after IgM and disappeared within 45 days following detection. Since IgM and IgA might recognize different epitopes and mimotopes than IgG, we expect to obtain additional valuable information for the diagnosis.

The dengue samples we tested for the **IgG immunosignature** included the PVD201 (20 samples) and the Seracare DS689G (5 samples). Those tested for **IgM immunosignature** included only the PVD201 (14 samples, 7 of 14 have positive IgM ELISA titers, denoted by Dengue IgM+; the other 7 have negative IgM ELISA titers but positive IgG ELISA titers, denoted by Dengue IgM-/IgG+); those tested for **IgA immunosignature** included only the PVD201 (6 samples). Note that all these samples were from the Americas. The peptides selected for each isotype were different. As shown in Figure 2.2, by correctly selecting significant peptide features, our microarray can distinguish the dengue serum from the normal serum by either IgG, IgM or IgA immunosignature, at least in this training set.

We tested most of our dengue samples by the **IgG immunosignature**. 25 dengue samples and 8 normal samples are tested. The sample size is big enough to do the class prediction. Three quarters of the PVD201 (15 samples) and three quarters of the normal (6 samples) were randomly selected and used as the training group. The other one quarter of each group along with the DS689G (5 samples) are used as the test group. Based on the training set, 50 features are selected (1-Way T-Test of the training set;P=2.1239e-5; fold change>2; parametric test assuming variances the equal) as significant between the normal and dengue. As shown in Figure 2.3, the Naïve Bayes 10-fold cross validation for all samples (25dengue+8normal) shows 97% accuracy with only one dengue misclassified as normal. The weighted average false positive rate is 0.01(by Weka classification).

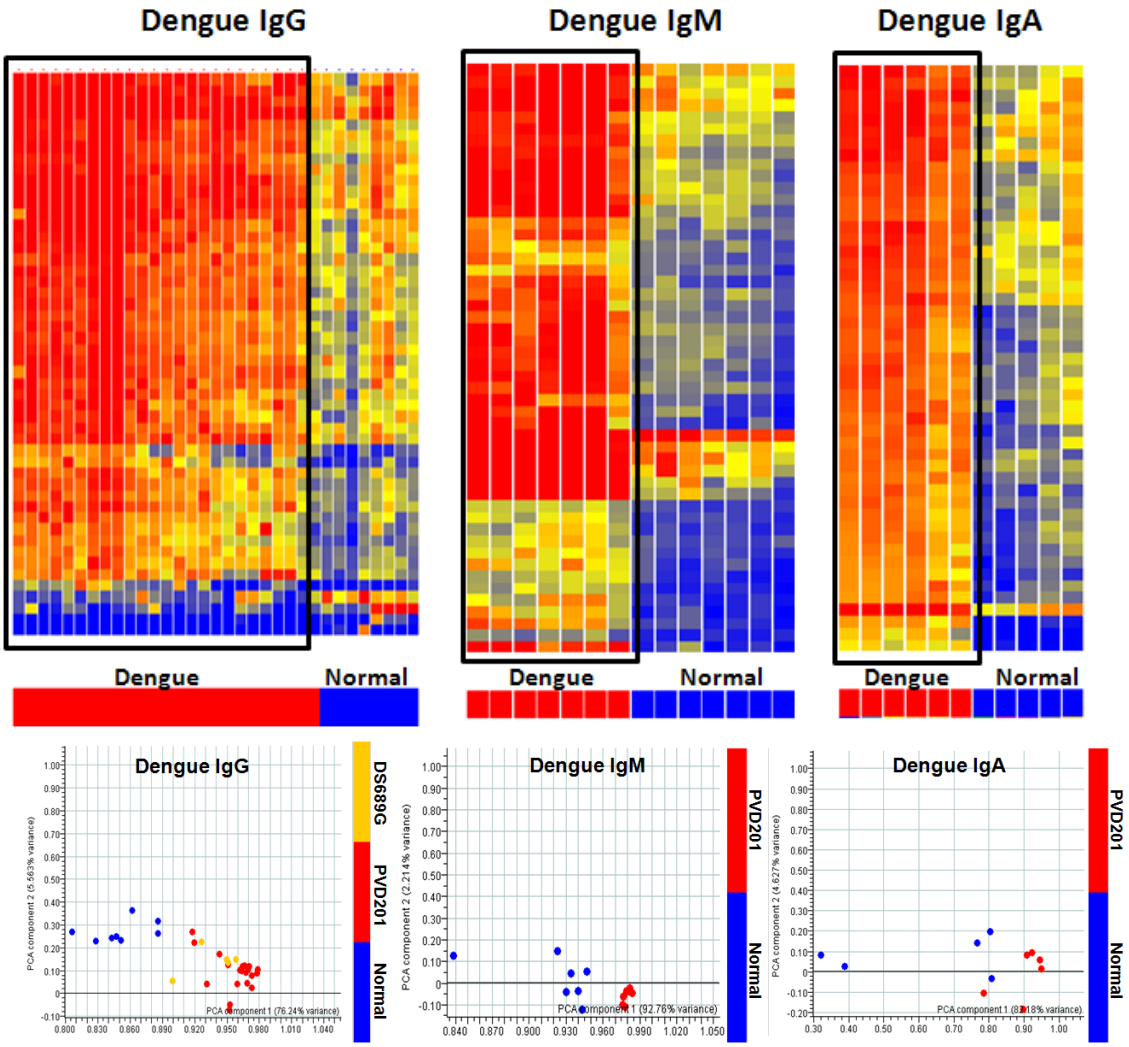


Figure 2.2. CIM 10K Ver2 microarray can distinguish the dengue serum with the normal serum by the IgG/IgM/IgA immunosignature. Distance clustering (top three), PCA by conditions (bottom three). **IgG immunosignature** (Left): 50 peptides selected by the T-Test of the training set (P=2.1239e-5; fold change>2), dengue PVD201(20), dengue DS689G(5), normal(8); **IgM immunosignature** (Middle): 50 peptides selected by the T-Test (P= 8.7729e-11; fold change>2) between the dengue PVD201 IgM+ (7) and normal(7); **IgA immunosignature** (Right): 49 peptides selected by the T-Test (P= 4.8604e-8; fold change>2), dengue PVD201(6), normal(5). Note: the feature selections of IgG and IgM immunosignature were based on the 50% normalized data, the selection of IgA immunosignature was based on the raw data.

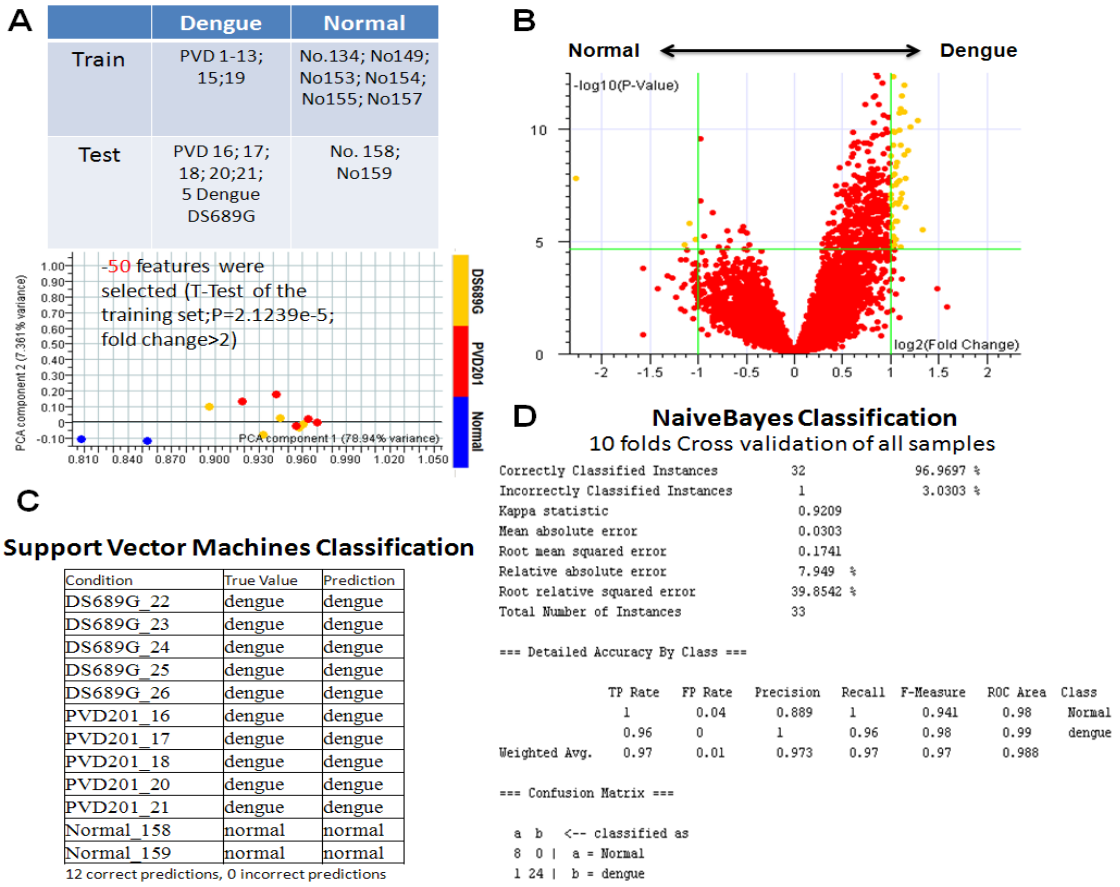
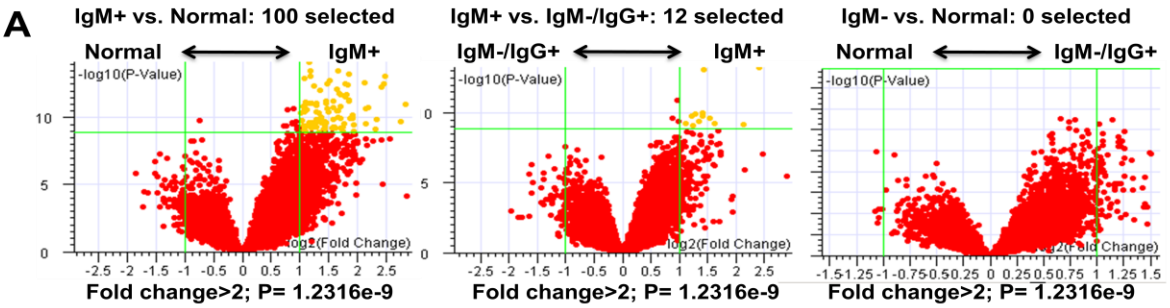
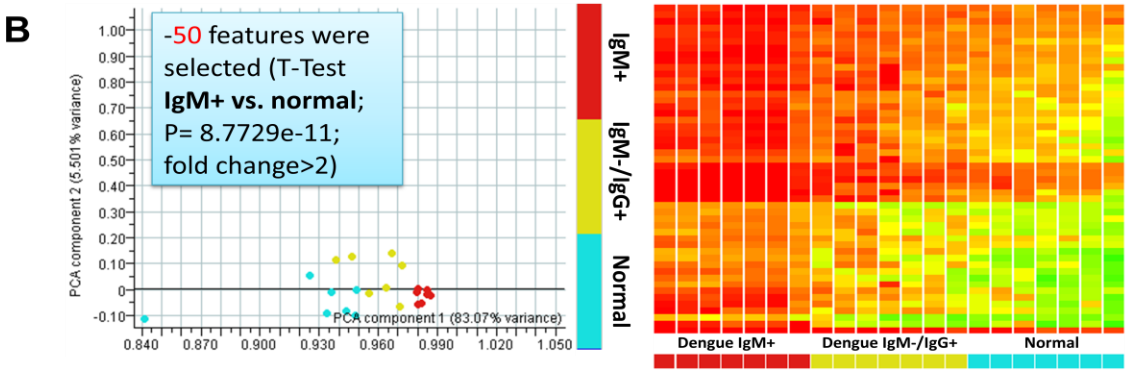


Figure 2.3. Class prediction by the dengue IgG immunosignature. A (top): the training and testing sets selected for the classification. A (bot). the peptide features selected by the training set can distinguish the samples in testing set by conditional PCA. B. The distribution of the selected 50 peptides on volcano plot. C. Classification by Support Vector Machines (SVMs) for the testing set shows 100% accuracy. D. Naïve Bayes 10-folds cross validation for all samples shows 9 7% accuracy.

For the **IgM Immunosignature**, we compared three groups: 7 IgM+ PVD201 samples, 7 IgM-/IgG+ samples and 7 normal samples. As shown in Figure 2.4A, if we set up the selection criteria as fold change>2, P=1.2316e-9, and compare the three groups one with the other, we can select 100 features between IgM+ and normal, 12 features between IgM+ and IgM-/IgG+, and zero between IgM- and normal. Since the more

significant features selected, the more difference there is, IgM+ and normal has the most difference, however, IgM-/IgG+ and normal has the least difference. Hence, IgM-/IgG+ should be between IgM+ and normal. As shown in Figure 2.4B, 50 peptides were selected by the T-test between IgM+ and normal. The PCA and distance clustering results show that IgM-/IgG+ is indeed placed between IgM+ and normal. This indicates that our IgM immunosignature truly reflects the IgM titer measured by standard IgM ELISA assay, which would be the basis for the quantitative immunosignature in the future. Additionally, the Naïve Bayes 10-fold cross validation correctly classify the 7 IgM-/IgG+ samples as dengue rather than normal, indicating the immunosignature has higher detection power than ELISA (Figure 2.4C).





C

NaiveBayes Classification

Use the 50 features selected between IgM+ vs.Normal

a b c <-- classified as	a b <-- classified as
6 1 0 | a = dengue IgM+	14 0 | a = dengue
0 7 0 | b = dengue IgM-	1 6 | b = Norm
0 2 5 | c = Norm	

Figure 2.4. Class prediction by the dengue IgM immunosignature. A. Volcano plot under the same selection criteria (fold change>2, P=1.2316e-9), IgM+ vs. Normal has more significant peptides selected than the other two pairs, indicating more signature difference. B. Conditional PCA and distance clustering shows the relationship of all the 21 samples. IgM-/IgG+ is placed between the IgM+ and normal. C. Naïve Bayes Classification correctly identify the 14 dengue samples, with only one normal sample incorrectly classified as dengue

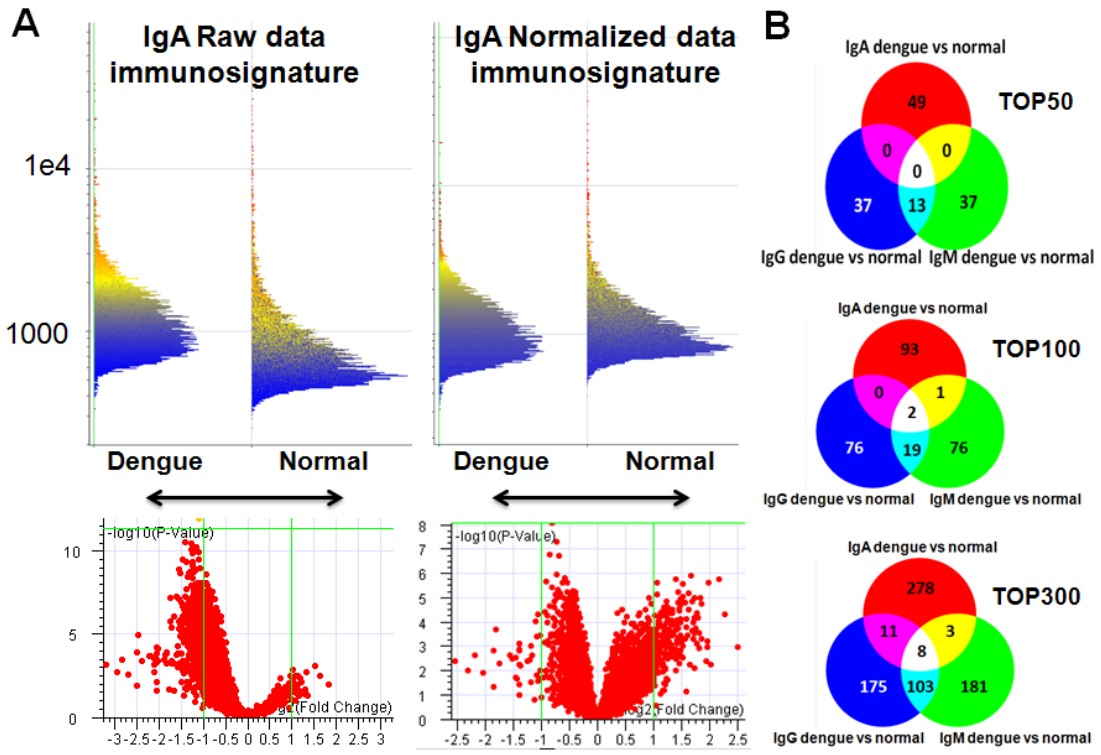


Figure 2.5. The dengue IgA immunosignature. A. IgA immunosignature analyzed by both raw data and median normalized data. PCA: x axis (fold change); y axis (P value). B. Comparison between IgA/IgG/IgM immunosiganture. Selection criteria: IgA and IgM (fold change>2, top 50/100/300 features by P-value); IgG (fold change>1.3, top 100/300 features by P-value; fold change>2. Top 50 features by P-value).

The difference in analyzing the **IgA immunosignature** was that we used the raw data instead of the median normalized data. The reason we cannot use the median normalized data for IgA is shown in Figure 2.5A. We find the median value of dengue IgA immunosignature is significantly higher than that of normal, if we still use the median value for normalization, the major difference will be hidden and even inverted.

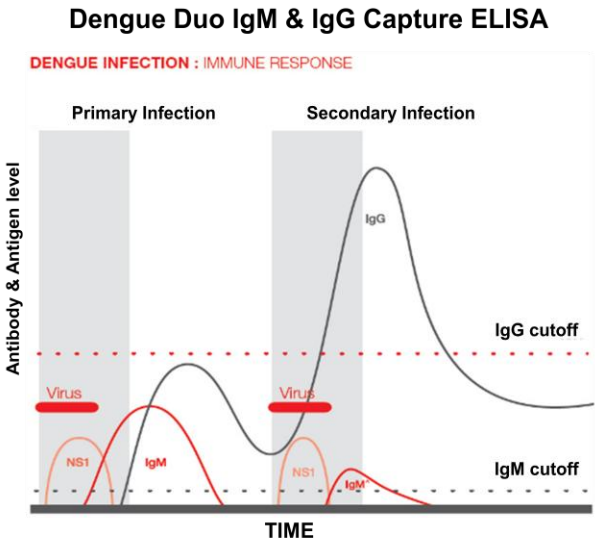
The disadvantage of using raw data is the noise. For example the technical errors, will also be presented. This might explain why when we compare the significant peptides selected by the T-test and fold change, we notice a certain percentage overlapping between IgG with IgM immunosignature, but not with the IgA immunosignature (Figure 2.5B).

2.4.2 10K Ver2 microarray can distinguish dengue primary and secondary infection

In dengue epidemic areas, many people are suffering with the dengue secondary infection rather than the primary infection, and evidence shows that secondary infection with a different serotype might be a risk factor for DHF (dengue hemorrhagic fever, a more serious condition). Distinguishing between primary and secondary dengue infection is always an important issue for clinical diagnosis. Hemagglutination Inhibition (HI) test is the reference test recommended by WHO. However, it is a complicated protocol. More strategies include the ELISA test and IgG affinity test[58], for example, the Panbio Dengue Duo IgM & IgG Capture ELISA. Here, we test the possibility to use the immunosignature data to distinguish the primary and secondary infection.

Based on the Panbio Dengue Duo IgM & IgG Capture ELISA, we can identify three primary dengue samples and 12 secondary dengue samples from the PVD201 panel (Figure 2.6). For the secondary infection, the IgM titer is sometimes undetectable. It's noticeable that three primary samples are all from Ecuador. However, we still have one secondary sample from Ecuador. It's important to include this Ecuador secondary sample to ensure the difference is not caused merely by the country of origin. We can select a list of peptides which can distinguish the selected primary and secondary dengue infection by

IgG/IgM/IgA immunosignature (Figure 2.7). One interesting thing we find is unlike the antibody titer, in which case the secondary is higher than the primary, many significant peptides we pick up on microarray show higher signal in primary than the secondary. I think the reason for this is our microarray is not only a measure for the "titers", but also a measure for the "diversity" of the entire antibodyome. During the primary infection, the antibodies generated are more diverse than the secondary infection, and that's probably



	Country	Panbio ELISA IgG	Panbio ELISA IgM	Prediction*
PVD201-10	Ecuador	0.4	>5.8	primary
PVD201-16	Ecuador	0.7	>5.8	primary
PVD201-20	Ecuador	0.7	>5.8	primary
PVD201-01	Colombia	2.4	>5.8	secondary
PVD201-02	Honduras	1.2	0.2	secondary
PVD201-03	Honduras	1.3	1.1	secondary
PVD201-05	Honduras	4.1	0.8	secondary
PVD201-06	Honduras	4.5	0.7	secondary
PVD201-07	Colombia	7	>5.8	secondary
PVD201-08	Honduras	2.6	0.9	secondary
PVD201-13	Honduras	1.3	0.3	secondary
PVD201-15	Honduras	1.3	0.4	secondary
PVD201-17	Colombia	2.3	>5.8	secondary
PVD201-18	Honduras	2.3	0.3	secondary
PVD201-21	Ecuador	4.3	>5.8	secondary
PVD201-04	Honduras	0.5	0.4	None
PVD201-09	Honduras	0.7	0.3	None
PVD201-11	Honduras	0.5	0.2	None
PVD201-12	Honduras	0.9	0.2	None
PVD201-14	United States	0	0	None
PVD201-19	Honduras	0.6	0.8	None

* Serum collection dates are unknown;

why there's higher signal on primary immunosignature. XIAO: out of place sentence

Figure 2.6. Primary and secondary dengue infection identified by Panbio IgM/IgG assay. Left: the principle of Panbio IgM/IgG assay, the IgG assay can only detect the secondary infection, whose titer is above the cutoff line (equivalent to HAI>1:2560). The samples below the IgG cutoff line, however above the IgM detection limit, are considered as the primary infection. Right: three samples of the PVD201 panel are predicted as the primary infection, 12 are considered as secondary. The other samples are non-infected based on this specific ELISA test.

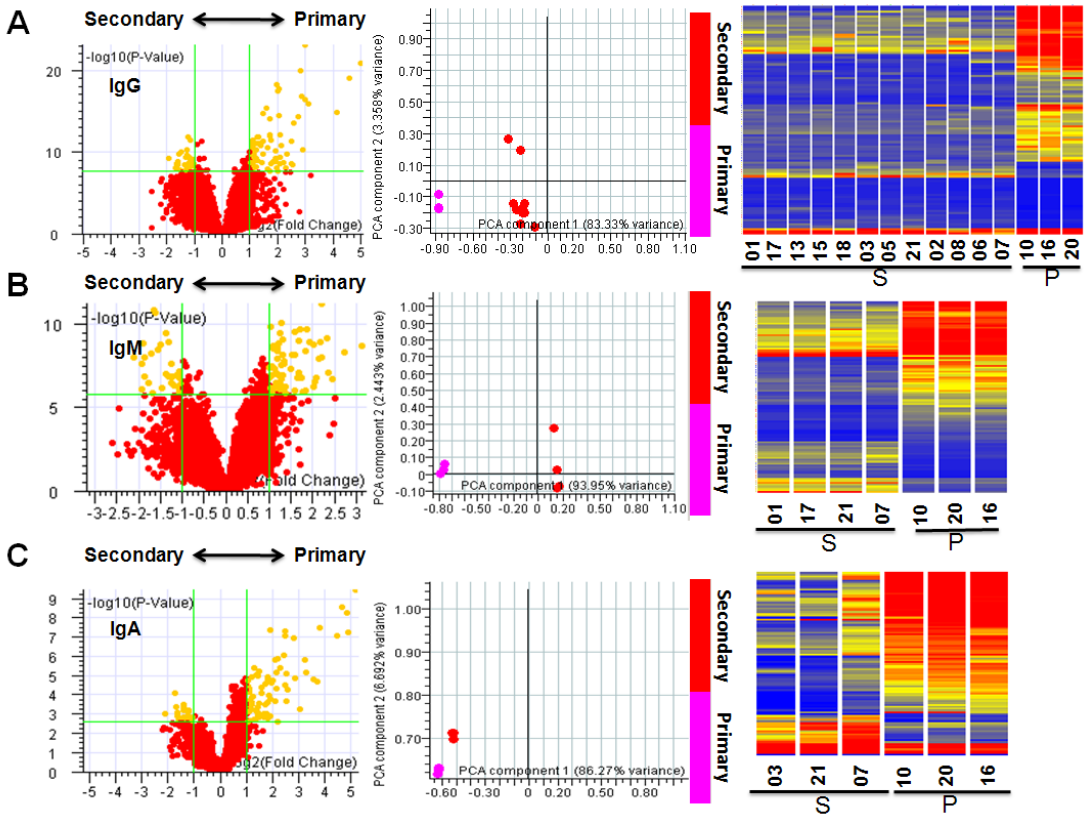


Figure 2.7. Primary and Secondary dengue infection immunosignnature. A. IgG immunosignature, Top 100 peptides are selected (Fold change>2, P=2.607e-8) between the primary samples (PVD201-10,16,20) with the secondary samples (PVD201-01,17,13,15,18,03,05,21,02,08,06,07). B. IgM immunosignature, Top 100 peptides are selected (Fold change>2, P=1.6205e-6) between the primary samples (PVcfD201-10,16,20) with the secondary samples (PVD201-01,17,21,07). C. IgA immunosignature, Top 100 peptides are selected (Fold change>2, P=0.0027) between the primary samples (PVD201-10,16,20) with the secondary samples (PVD201-03,21,07).

Furthermore, as we notice this significant difference between dengue primary and secondary immunosignature, we add this into distinguish the “dengue” and “normal”. As shown in Figure 2.8A left, it shows the distribution of the top 100 peptides (selected to distinguish dengue primary and secondary by IgG immunosignature) on the volcano plot used to selected the significant peptides between “dengue” and “normal” in Figure 2.2,

since previously we did not separate the primary and secondary samples, we notice many peptides of the 100 are not hot hits on the volcano plot (in other words, they were neglected). However, now we realize these peptides are also good features to tell dengue from normal (Figure 2.8A middle). The three groups (primary, secondary and normal) can be separated even better, if we use the features directly selected from the entire 10,000 peptides (Figure 2.8A right). Here is the interesting thing; the secondary IgG immunosignature is closer with the normal immunosignature, rather than the primary immunosignature. This reminds us that we should not make the assumption that the difference between subgroups is less than the major groups, and separate the significant infectious subgroups might be helpful to compare the immunosignature to the normal.

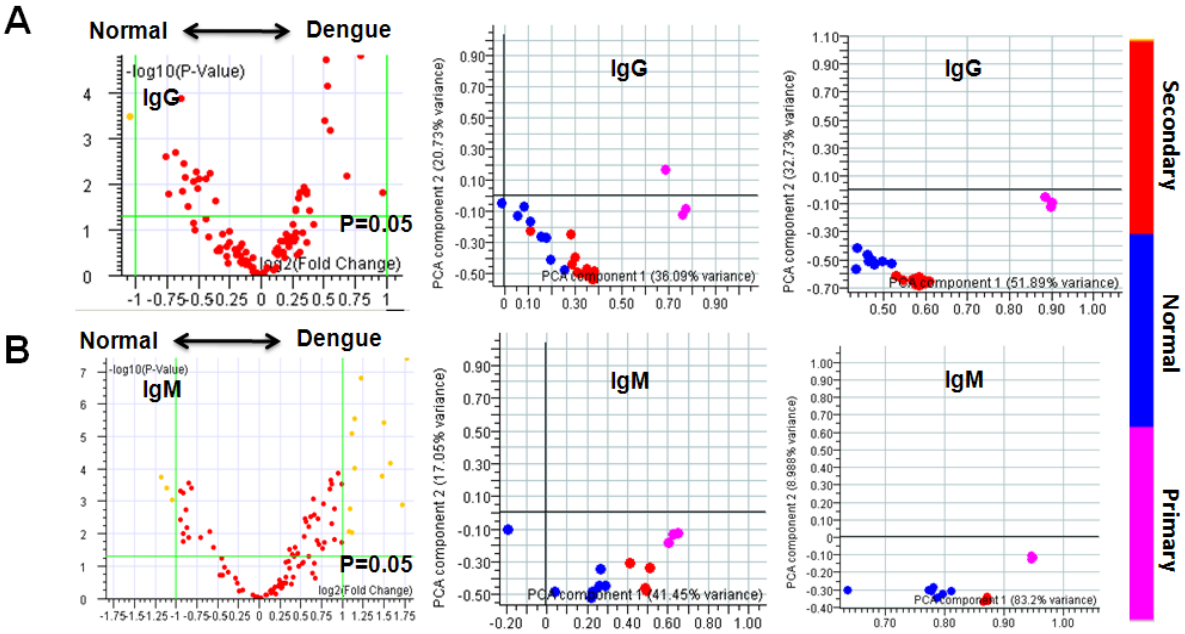


Figure 2.8. Distinguish between dengue primary, dengue secondary and normal samples. A. Left: the distribution of the top 100 peptides (primary vs. secondary by IgG immunosignature) on the volcano plot used to select the significant peptides between "dengue" and "normal". Middle: by one-way ANOVA(P=0.05) 82 features are picked from the top 100 peptides (primary vs. secondary by IgG immunosignature). Right: 104 features are picked by one-way ANOVA (P=5*10e-11) from the entire 10,000 peptides to distinguish "dengue primary", "dengue secondary" and "normal". B. the same as A, for the IgM immunosignature.

Finally, we use the selected top 100 peptides for IgG immunosignature (Figure 2.7A) to test the other six PVD201 samples, which are shown as "NONE" in Figure 2.6 (Right); 5 DS689G samples; 8 DS689M samples and 4 WHO samples. It is interesting to find out if some of these samples are classified as primary, some of them are secondary. The WHO samples (four dengue serotype individuals, denoted as D1,D2,D3,D4) looks like a different group, though closer to the secondary (Figure 2.9).

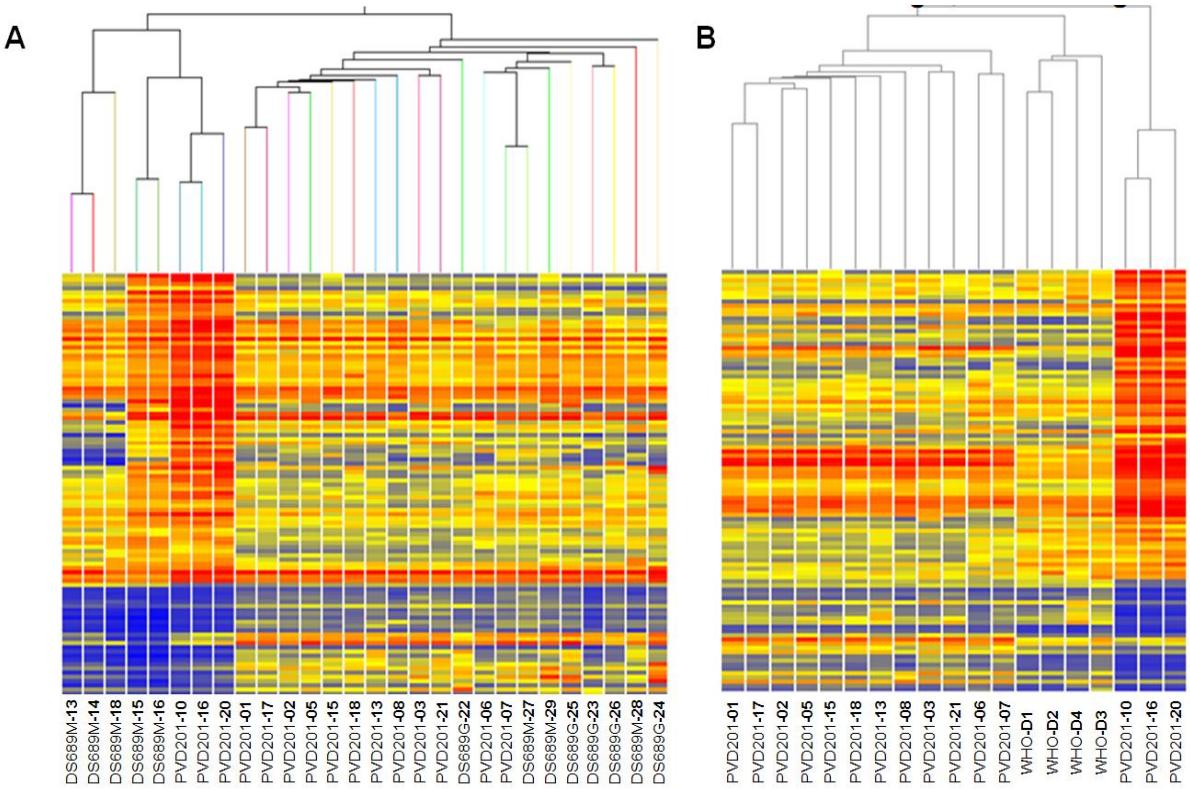


Figure 2.9. Unknown dengue samples are predicted as either primary or secondary. Distance clustering based on the top 100 peptides selected (Fold change>2, P=2.607e-8) between the primary samples (PVD201-10,16,20) with the secondary samples (PVD201-01,17,13,15,18,03,05,21,02,08,06,07). A. PVD201-04,09,11,12,19, DS689G and DS689M samples as the testing set; B. WHO samples as the testing set.

2.4.3 Compare the immunosignature of dengue with the other related infectious diseases

Antibody cross-reactivity is always a big problem for traditional ELISA-based diagnosis of dengue. For example, the commonly used MAC-ELISA is based on the antigens derived from the E protein of the virus, which shows cross-reactivity between other circulating flaviviruses, such as West Nile virus (WNV), St. Louis encephalitis virus (SLE), Japanese encephalitis virus (JEV) and yellow fever virus (YFV). What is more, dengue itself is regarded as an Influenza-Like Illness (ILI), so it could be misdiagnosed as the other ILIs, such as influenza and malaria. To evaluate the cross-reactivity of a diagnostic method is very important, particularly for dengue[36,59].

Here we tested and compared several West Nile Virus (WNV) and malaria samples with the previous dengue samples. 10 WNV, 3 malaria, 15 dengue (training set shown in Figure 2.3A) and 8 normal samples are used here for a simple demonstration. We do observe the difference between dengue and WNV/ malaria by the IgG immunosignature (Figure 2.10). However, we find the raw data of malaria is significantly higher than that of normal, which makes it not possible to use the medium normalization. We consider this significant higher raw data is caused by the overall higher IgG titer in the serum. Recently, we have the idea of adding a series of internal references to detect the general IgG titer in a particular sample; we expect this reference can be used to make better normalizations in the future.

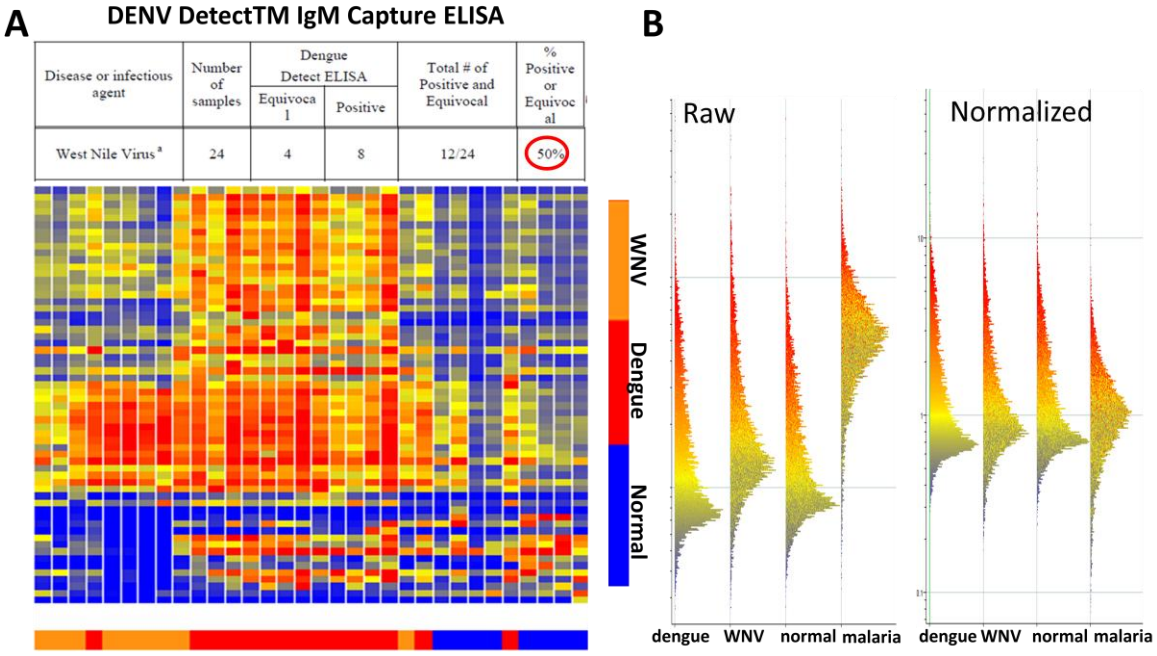


Figure 2.10. Compare dengue immunosignature with WNV and malaria. A. FDA approved IgM ELISA diagnostics has a high cross reactivity with WNV, 10 Feature(s) chosen from 'Dengue vs WNV P=3.56e-7'. 50 Feature(s) chosen from 'Dengue vs Normal P= 2.1239e-5' Pearson correlation clustering; B. Medium normalization is not a good way to analyze the malaria data.

2.4.4 Comparison of different dengue isotopes (NIBSC WHO Standard Panel) by both 10K Ver2 microarray and a custom NS1 subarray

Currently well-accepted techniques to distinguish the four dengue serotypes include the RT-PCR and PRNTs, which are both complicated. A novel serotyping-NS1-ELISA was recently created for the same purpose. It demonstrated that selected Mabs produced by mouse splenocyte and myeloma cell fusions (stimulated by DENV 1/2/3/4 NS1 protein) can specifically recognize their original targets (DENV 1/2/3/4 NS1 protein)[60]. Based on this fact, we select 65 peptides based on the real amino acid sequences of DEN1/2/3/4 NS1 proteins, and create a new sub-array. We want to know

whether this subarray, as well as the 10K random array can be used to distinguish different dengue serotypes.

A

DEN_NS1	den1_1	den1_2	den1_3	den2_1	den2_2	den2_3	den2_4	den2_5	den2_6	den2_7	den2_8	den2_9	den3_1	den3_2	den3_3	den3_4	den3_5	den4_1	den4_2	den4_3	den4_4	den4_5
den1_1	100%	96%	97%	71%	68%	73%	72%	72%	72%	72%	73%	72%	79%	79%	79%	80%	79%	69%	69%	69%	69%	68%
den1_2		100%	98%	72%	70%	74%	73%	73%	72%	73%	73%	73%	79%	80%	79%	81%	80%	69%	70%	69%	69%	69%
den1_3			100%	72%	70%	74%	73%	73%	72%	72%	73%	73%	79%	80%	79%	80%	80%	69%	70%	69%	69%	69%
den2_1				100%	89%	97%	94%	95%	99%	98%	97%	95%	73%	73%	72%	73%	73%	72%	71%	72%	71%	71%
den2_2					100%	92%	89%	90%	90%	90%	91%	90%	70%	71%	69%	71%	71%	69%	69%	69%	68%	68%
den2_3						100%	96%	97%	98%	98%	97%	99%	70%	74%	73%	75%	74%	73%	73%	73%	72%	72%
den2_4							100%	99%	95%	95%	96%	99%	73%	74%	73%	74%	73%	73%	72%	73%	72%	71%
den2_5								100%	95%	95%	97%	99%	73%	74%	73%	74%	73%	73%	72%	73%	72%	72%
den2_6									100%	99%	98%	95%	73%	74%	73%	74%	73%	73%	72%	73%	72%	72%
den2_7										100%	97%	95%	73%	74%	73%	74%	74%	73%	72%	73%	72%	71%
den2_8											100%	97%	74%	75%	73%	75%	74%	73%	73%	73%	73%	72%
den2_9												100%	73%	74%	73%	74%	73%	73%	72%	73%	72%	72%
den3_1													100%	99%	99%	97%	97%	73%	73%	73%	73%	72%
den3_2														100%	98%	98%	99%	74%	73%	74%	73%	73%
den3_3															100%	97%	97%	73%	72%	73%	73%	72%
den3_4																100%	97%	73%	72%	73%	73%	72%
den3_5																	100%	73%	72%	73%	73%	72%
den4_1																		100%	97%	99%	96%	96%
den4_2																			100%	97%	98%	97%
den4_3																				100%	97%	97%
den4_4																					100%	96%
den4_5																						100%

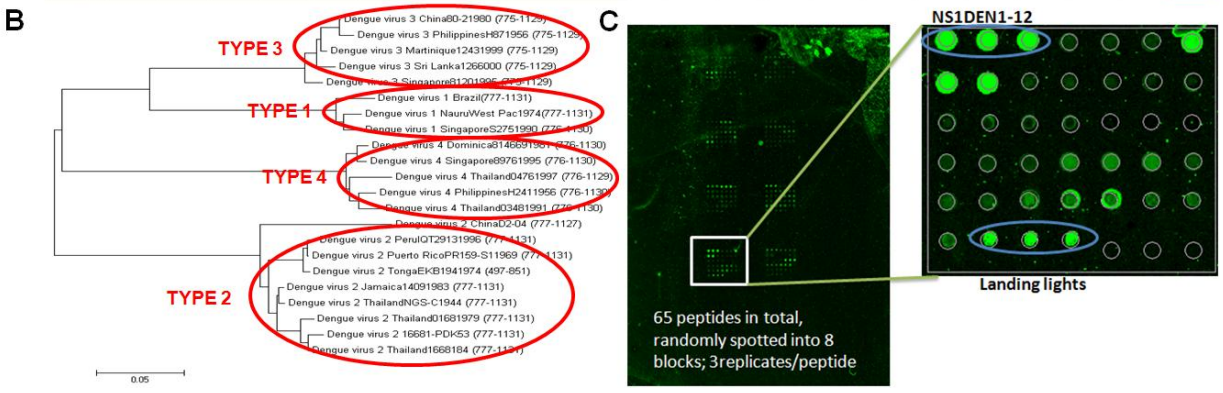


Figure 2.11. The creation of Dengue NS1 subarray. A: NS1 protein identity comparison between 22 dengue strains. B: Phylogeny reconstruction of DEN1/2/3/4 by their NS1 proteins. Neighbor-joining classification by MEGA 5.03. C: The structure of NS1 subarray, 12 peptides selected for DEN1; 25 selected for DEN2; 12 selected for DEN3; 16 selected for DEN4.

Dengue NS1 protein is a secreted extracellular protein (~350a.a), which has been shown to modulate the complement system and to enhance DENV infection. Its primary structure is highly identical between different serotypes (~70%); however, we can still select 65 17mers, which contain the majority of sequence differences. These peptides are

spotted on the slides and screened against 4 WHO dengue serotype standard. The subarray results are compared later with the random 10K array results. 22 dengue NS1 proteins (3 DEN1; 9 DEN2; 5 DEN3; 5 DEN4) are used for the spotting peptide selection. Figure 2.11 shows the similarity between different dengue strains and the distribution of spots on the NS1subarray

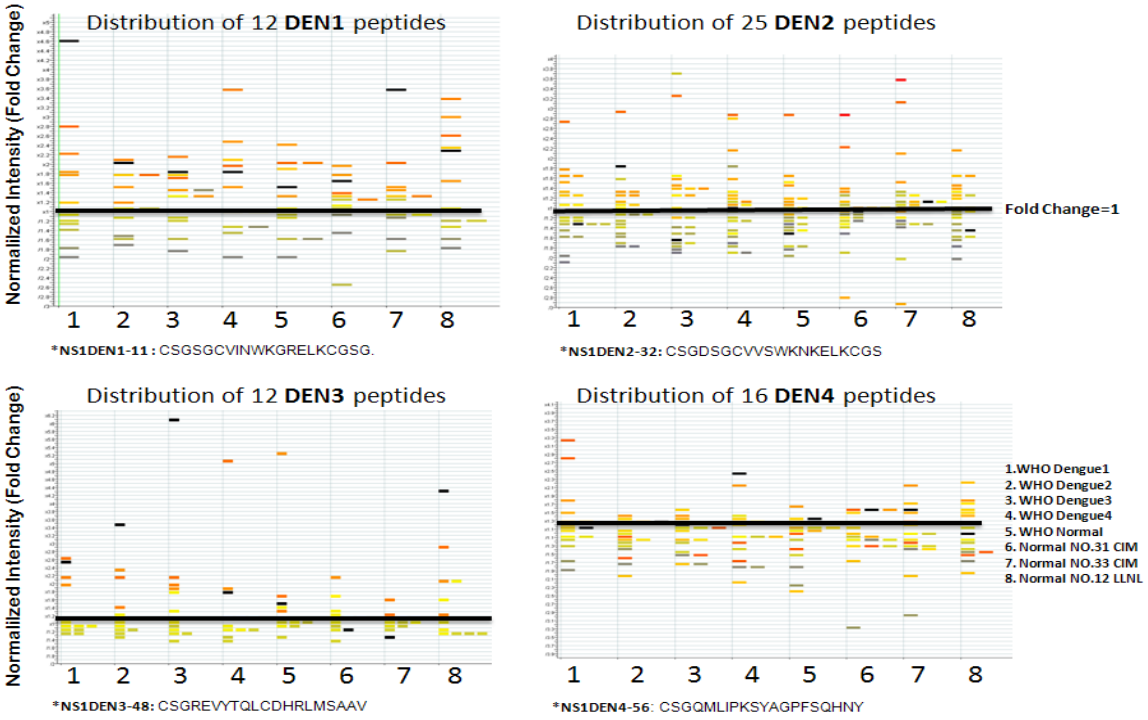


Figure 2.12. Significant peptides selected from NS1 dengue subarray. One significant peptide for DEN1/2/3/4 is selected. We totally tested 8 samples (4 WHO dengue samples and 4 normal samples, listed on the right). For example, NS1DEN3-48 is selected, because it is designed for dengue serotype 3 and it also has higher signal when screened against WHO dengue3 sample. NS1DEN3-48 is the most possible epitope specific for dengue serotype 3.

Since we know the exact origin of each peptide (DEN1/2/3/4), we wanted to see whether the peptides generated from DEN1/2/3/4 have higher signal when screened against its specific serotype. Four WHO standard dengue serum samples (DEN1/2/3/4,

identified by PRNT50) and 4 normal samples (different sources) are screened on the NS1 subarray. Although most of the peptides have similar signals between different dengue serotypes, as well as normal samples, we could still pick up some significant ones. As shown in Figure 2.12, four sets of peptides are viewed separately; the most significant one for each set is marked black and denoted below.

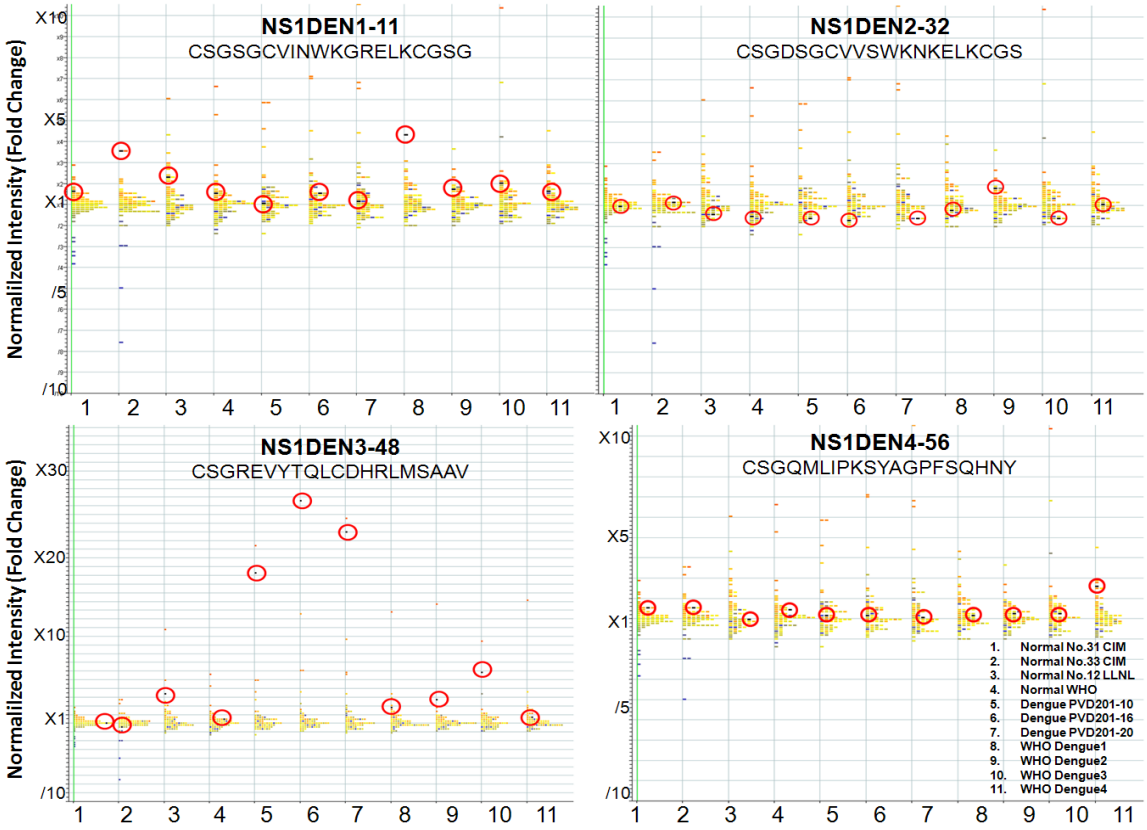


Figure 2.13. Dengue serotype prediction by NS1 subarray. Four peptides (NS1DEN1-11, NS1DEN2-32, NS1DEN3-48, NS1DEN4-56) were used to predict the other three primary PVD201 samples (NO. 10/16/20). All these three samples have relative high signal of NS1DEN3-48.

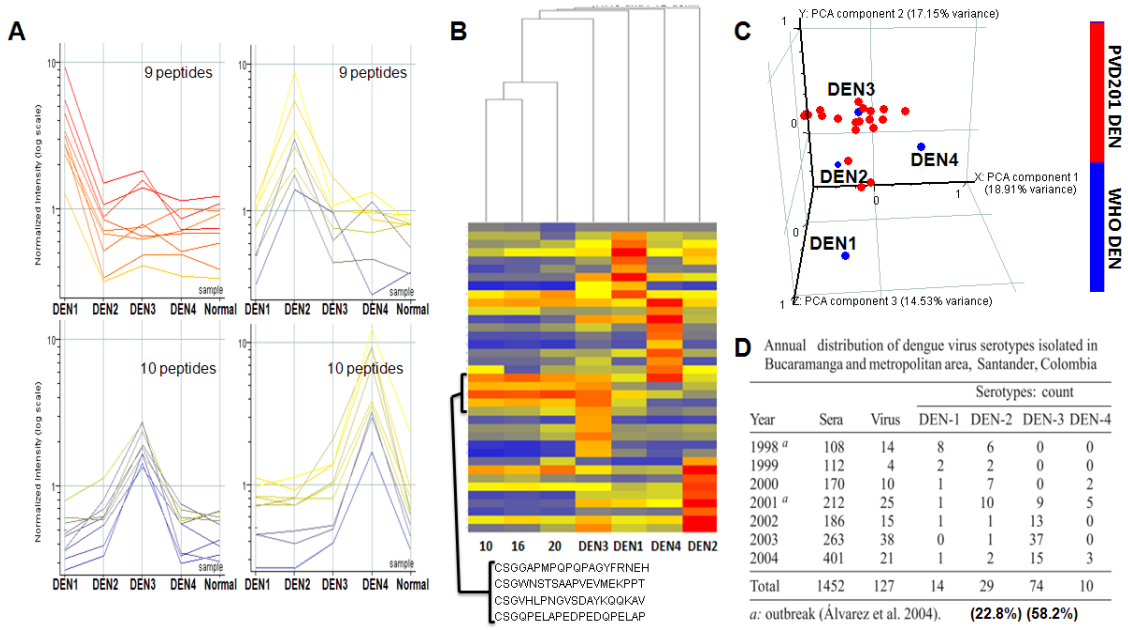


Figure 2.14. Dengue serotype prediction by 10K random microarray. A: 38 random peptides are selected by the "draw expression profile" function in Genespring, each set contains the peptides higher in one serotype, but not the other three serotypes or the WHO normal. B: Heatmap Pearson clustering of three PVD201 primary dengue samples between WHO samples based on the selected 38 peptides. C: PCA plot of all 20 PVD201 dengue samples with WHO samples based on the selected 38 peptides. D: DEN3 are the dominant serotypes in Colombia areas since 2000.

As the next step, we tried to use the four peptides selected above to identify the serotype of three primary PVD201 dengue samples (PVD201-10/16/20). As shown in Figure 2.13, the three PVD primary samples, denoted by 5, 6, 7, have relative high signal on NS1DEN3-48, but not the other three peptides, which indicates they're probably DEN3. We also screened the WHO samples on the 10K random array, as shown in Figure 2.14, we select the top peptides for each serotype (38 in total), and used this peptide list to identify the PVD201 dengue samples. Since we only have one WHO reference sample for each serotype, we do not have the real biological replicates. The feature selection here is based on intensity fold-change, not T-test. Interestingly, we find most of the PVD201 samples are closer to WHO DEN3 on the PCA plot, and the three

primary PVD samples are also classified close to DEN3 on the heatmap. Based on both the NS1 subarray and 10K random array, we predicted the three primary PVD201 dengue samples are DEN3, and most of the secondary PVD201 dengue samples were also probably infected by DEN3 before. This prediction is in accordance with the endemic patterns of dengue disease in the region of the Americas: during the 90s, DEN1 and DEN2 were the most dominant serotypes; since the introduction of DEN3 in 2000, DEN2 and DEN3 became the most frequently reported serotypes[61,62] (Figure 2.14D). When we have more specific dengue serotype samples in the future, we can statistically test whether this serotyping method is specific or accurate enough.

2.5 Discussions: How to decipher the proteome information containing in the 10K immunosignature

In this research, we investigate the immunosignature of different dengue infected serum samples. We have shown the possibility of using 10K microarray to distinguish dengue from normal, dengue primary from secondary, dengue from WNV and different dengue serotypes. One of the most interesting questions is whether the peptides we picked up as significant have a relationship with the real pathogen epitopes. For the traditional epitope microarray created from real pathogen proteome, the high binding of a particular peptide is usually interpreted as the existence of its corresponding antibody in the serum. However, for our microarray, every random peptide has the potency to bind with more than one kind of antibody in the serum; conversely, every kind of antibody has the potency to bind with more than one random peptide on the array. Since the binding relationship on random microarray is not one-to-one, how to decipher the selected

significant peptides is very difficult, even possible. Currently, we do not have a standard way to decipher the data generated from 10K microarray. Several methods are used here to show that the random microarray also contains the real epitope/ mimotope information. However, it is difficult to identify them solely based on the 10K microarray data. Microarrays with higher peptide density (for example, 330K random peptide microarray) or additional mutation microarray (based on selected significant peptides on 10K) is necessary for the further identification.

Among the top 50 significant peptides selected for distinguishing dengue and normal (IgG immunosignature, fold change>2, P=2.1239e-5, Figure 2.3), 46 show higher signal in dengue. These 46 peptides have the highest possibility of containing the primary structure similarity with the real dengue proteome. We first searched these 46 peptides by NCBI blastp (Database: Non-redundant protein sequences; organism: Dengue viruses (taxid:11052)). Another 46 randomly picked up peptides were also searched against blastp as a control. For each entry, the result with the highest E-value is displayed in Table 2.1 below. The molecular weight, isoelectric point and hydrophilicity are calculated by two on-line peptide calculators (GenScript/BACHEM). Compared with the control peptide list, the selected 46 peptides do not have significantly better search matches. However, they do have a higher isoelectric point, higher positive charge and higher hydrophilicity. Since blastp is designed for the query of relative long protein, or short but nearly exact peptide, it might not be very powerful to align our random peptides.

Second, we used GuiTope[63], an application for mapping random sequence peptides to protein sequences, to analyze the 46 peptides. GuiTope is similar with blastp, but with

improved scoring matrix for peptide-protein comparison. We also used a randomly picked up 46 peptides as control. The database we used to compare with is a list of real dengue linear B cell epitope (411 peptides) reported by IEDB. What GuiTope does is it compares every entry from the 46 peptides with the peptides in database, the mostly matched sequences are automatically ranked from the highest score to the lowest (with a

Dengue PVD training set vs. normal selected top 46 higher in dengue										
	Sequence	Score	Expect	Identities	Positives	Gaps	Molecular weight (M_r)	Isoelectric point	Net charge pH 7.0	Hydrophilicity
1	CSGYWNMIPMGKMFPVDYYH	25.7 bits(53)	0.69	5/6(83%)	6/6(100%)	0/6(0%)	2438.9	8.8	1	-0.8
2	CSGFGKWHRIKHIWIDEPSY	23.5 bits(48)	4	6/11(55%)	7/11(63%)	3/11(27%)	2458.83	9.7	2.1	-0.2
3	CSGYSRSSQYEGWHGKRDFM	22.7 bits(46)	8	6/9(67%)	8/9(88%)	0/9(0%)	2380.61	9.6	2	0.2
4	CSGGPGTMNAYGMHHGMLYA	22.3 bits(45)	11	8/15(53%)	8/15(53%)	5/15(33%)	2054.39	8.8	1.1	-0.7
5	CSGWFKYGIHNKRWNKMFDF	21.4 bits(43)	22	8/18(44%)	9/18(50%)	7/18(38%)	2564	10.5	4	-0.3
6	CSGGWGWVVWHYPMSLATKKP	21.0 bits(42)	31	4/4(100%)	4/4(100%)	0/4(0%)	2290.7	10.2	3	-0.6
7	CSGIQHVKPYMWMEKKAFHL	20.6 bits(41)	45	5/6(83%)	5/6(83%)	0/6(0%)	2432.97	10.1	3.1	-0.3
8	CSGDRHKFLQEEQKWWRYMP	20.6 bits(41)	45	6/9(67%)	7/9(77%)	2/9(22%)	2624.01	9.7	2	0.3
9	CSGYQYDRHVWYPWGDDASR	20.6 bits(41)	46	7/10(70%)	7/10(70%)	2/10(20%)	2460.6	7.1	0	-0.1
10	CSGQIWRWVQQTHEPWTHRV	20.2 bits(40)	64	7/14(50%)	9/14(64%)	4/14(28%)	2533.86	10.9	2.1	-0.4
11	CSGRYPGQIKVYNKEHWGGL	20.2 bits(40)	64	7/12(58%)	8/12(66%)	3/12(25%)	2291.62	10	3	-0.1
12	CSGMSYHSLGSFSRHYMPMP	19.7 bits(39)	78	5/6(83%)	5/6(83%)	1/6(16%)	2274.66	9.7	2.1	-0.5
13	CSGSYFSDNAHKHQLKNWSG	19.7 bits(39)	90	6/7(86%)	6/7(85%)	1/7(14%)	2265.45	9.7	2.1	-0.1
14	CSGEKFHYHHKAMNRSYAMI	19.7 bits(39)	90	6/13(46%)	7/13(53%)	6/13(46%)	2409.8	10	3.2	-0.1
15	CSGPPMKIEHWKRVIYWPVV	19.3 bits(38)	120	5/6(83%)	5/6(83%)	0/6(0%)	2424.9	10.2	3	-0.4
16	CSGYAPATAHGWWQMHRFMN	19.3 bits(38)	125	6/10(60%)	6/10(60%)	3/10(30%)	2350.7	9.9	2.1	-0.7
17	CSGRWRNPDPQLTPKTQYLN	19.3 bits(38)	126	5/5(100%)	5/5(100%)	0/5(0%)	2373.68	10.4	3	0.1
18	CSGRDPLFFANKNLKEMYMR	19.3 bits(38)	127	9/16(56%)	10/16(62%)	4/16(25%)	2419.88	10.2	3	0.2
19	CSGYNQMVWTRPQSHYQLSF	18.9 bits(37)	138	4/5(80%)	4/5(80%)	0/5(0%)	2431.74	9.7	2	-0.6
20	CSGDWYTPGYAYNIKKPVNW	18.9 bits(37)	176	5/5(100%)	5/5(100%)	0/5(0%)	2361.67	9.4	2	-0.5
21	CSGYYHQGPRKHLKSNDWIY	18.9 bits(37)	178	4/4(100%)	4/4(100%)	0/4(0%)	2451.75	9.8	3.1	-0.1
22	CSGGGYRQGTGVRTNIMPMY	18.9 bits(37)	178	7/9(78%)	7/9(77%)	1/9(11%)	2147.49	10.3	3	-0.3
23	CSGSNWYFPRYHLMMTHKTA	18.5 bits(36)	235	4/4(100%)	4/4(100%)	0/4(0%)	2429.84	10.1	3.1	-0.6
24	CSGWQNLSWYAGPKHLHSFQ	18.5 bits(36)	244	5/6(83%)	5/6(83%)	0/6(0%)	2345.63	9.8	2.1	-0.7
25	CSGLMWFTKYFTQSKPHYAV	18.5 bits(36)	246	4/4(100%)	4/4(100%)	0/4(0%)	2393.82	10	3	-0.7
26	CSGWQYTRGNQYYPWHNVER	18.5 bits(36)	247	5/8(63%)	6/8(75%)	0/8(0%)	2543.77	9.5	2	-0.4
27	CSGQRAQYHGRMQTFYRGEM	18.5 bits(36)	248	4/5(80%)	5/5(100%)	0/5(0%)	2405.73	10.3	3	0
28	CSGDIAHRKWYQYVDNMWRP	18.0 bits(35)	340	4/6(67%)	5/6(83%)	0/6(0%)	2524.87	9.6	2	-0.1
29	CSGQYKAKMWDPPNYKYQWA	18.0 bits(35)	347	4/5(80%)	4/5(80%)	0/5(0%)	2463.83	9.8	3	-0.2
30	CSGKGSVYRFNDWHRHHTLM	18.0 bits(35)	347	4/5(80%)	4/5(80%)	0/5(0%)	2430.76	10.4	3.2	-0.1
31	CSGKWPHYMNAYYQKDKLPY	18.0 bits(35)	347	5/6(83%)	5/6(83%)	0/6(0%)	2491.88	9.7	3	-0.3
32	CSGYTHGRWMQVRQHPLEMY	17.6 bits(34)	480	7/15(47%)	7/15(46%)	7/15(46%)	2478.87	9.7	2.1	-0.3
33	CSGMPKNGSHPYDYGQSRVL	17.6 bits(34)	481	4/4(100%)	4/4(100%)	0/4(0%)	2195.47	9.6	2	0
34	CSGHKTHRPEFRGPMYELRY	17.2 bits(33)	640	4/4(100%)	4/4(100%)	0/4(0%)	2463.83	10.1	3.1	0.3
35	CSGDYHSHHKFGGHNYKQHL	17.2 bits(33)	651	7/17(41%)	7/17(41%)	10/17(58%)	2351.55	9.5	2.4	-0.1
36	CSGTHYTGYYTMHAPKMGVG	17.2 bits(33)	658	4/4(100%)	4/4(100%)	0/4(0%)	2160.49	9.5	2.1	-0.6
37	CSGKSNLYAKGYHEGGGLYS	17.2 bits(33)	671	5/6(83%)	5/6(83%)	0/6(0%)	2090.3	9.4	2	-0.1
38	CSGKTVARFKFAYPPYGHPD	17.2 bits(33)	671	5/5(100%)	5/5(100%)	0/5(0%)	2240.57	10	3	-0.1
39	CSGGSARGGAPGAARWHFST	17.2 bits(33)	671	8/11(73%)	8/11(72%)	1/11(9%)	1932.11	12.5	3	-0.1
40	CSGYHGHYPMTGHKWGQVVM	16.8 bits(32)	892	4/4(100%)	4/4(100%)	0/4(0%)	2274.64	9.6	2.2	-0.6
41	CSGHPELNSPYMIYPTTSRM	16.8 bits(32)	913	4/4(100%)	4/4(100%)	0/4(0%)	2283.64	8.8	1	-0.3
42	CSGQPRYDYDAFPYFKHVGK	16.3 bits(31)	1230	4/4(100%)	4/4(100%)	0/4(0%)	2377.67	9.4	2	0
43	CSGLLKRPFYYMHHADKPWP	16.3 bits(31)	1276	7/13(54%)	8/13(61%)	1/13(7%)	2445.9	10	3.1	-0.3
44	CSGRADYYRLGEPMFNSWFP	16.3 bits(31)	1295	4/6(67%)	5/6(83%)	0/6(0%)	2395.71	8.8	1	-0.2
45	CSGPQPHYPFYSRVGYDVNQ	15.9 bits(30)	1780	4/4(100%)	4/4(100%)	0/4(0%)	2313.54	8.8	1	-0.3
46	CSGPQMYKHHQGYYHGLYMF	15.1 bits(28)	3363	4/5(80%)	4/5(80%)	0/5(0%)	2446.82	9.4	2.2	-0.7
	AVG	**18.85**	**435.65**				**2367.90**	**9.76**	**2.36**	**-0.27**
	STDEV	**2.09**	**601.16**				**137.85**	**0.74**	**0.79**	**0.29**
Red: acidic residues, like D E and C-terminal -COOH			Green: hydrophobic uncharged residues, like F I L M V W A and P							
Blue: basic residues, like R K H and N-terminal -NH2			Black: other residues, like G S T C N Q and P							

CONTROL(46 randomly selected peptides)										
	Sequence	Score	Expect	Identities	Positives	Gaps	Molecular weight (M_r)	Isoelectric point	Net charge pH 7.0	Hydrophilicity
1	CSGSKDRIPYYMGEYWYERF	24.8 bits(51)	1.3	7/10(70%)	7/10(70%)	2/10(20%)	2549.87	8.7	1	0
2	CSGKMYPLKYSVMYMIMTRI	23.5 bits(48)	3.9	6/6(100%)	6/6(100%)	0/6(0%)	2415.07	10.2	4	-0.5
3	CSGVKMEHESWSEDAWNYDK	23.5 bits(48)	3.9	7/16(44%)	8/16(50%)	4/16(25%)	2400.59	4.7	-2	0.4
4	CSGQGAVQSPYKSWVYVGML	23.1 bits(47)	5.2	5/9(56%)	6/9(66%)	0/9(0%)	2159.52	9.6	2	-0.6
5	CSGPGWKWWYAWVAESKQFW	22.3 bits(45)	11	7/15(47%)	7/15(46%)	3/15(20%)	2501.86	9.7	2	-0.8
6	CSGYQEWMDEFVLYKIMNKW	21.8 bits(44)	16	7/10(70%)	7/10(70%)	1/10(10%)	2570.01	6.5	0	-0.3
7	CSGNWGSEWKKVKAPHNQKP	21.4 bits(43)	23	7/14(50%)	7/14(50%)	4/14(28%)	2280.6	10.6	4	0.3
8	CSGVWFTQQGGAYKHEDFMA	21.4 bits(43)	23	7/9(78%)	8/9(88%)	1/9(11%)	2261.53	7.2	0	-0.3
9	CSGIFDTDITYHLPFAYLKQ	21.4 bits(43)	23	6/7(86%)	6/7(85%)	0/7(0%)	2331.68	7.1	0	-0.5
10	CSGILISYRDPRDLLWRYFF	21.0 bits(42)	32	7/11(64%)	8/11(72%)	3/11(27%)	2519.96	9.7	2	-0.4
11	CSGTMTATDLSHMNLFVQVR	21.0 bits(42)	32	9/19(47%)	10/19(52%)	7/19(36%)	2210.59	9	1	-0.4
12	CSGQHNTYDYIRVHYFKIFD	20.6 bits(41)	45	7/12(58%)	9/12(75%)	1/12(8%)	2505.8	8.8	1.1	-0.3
13	CSGGGKFQIMLVAIWFYTPK	20.6 bits(41)	46	6/8(75%)	7/8(87%)	0/8(0%)	2245.74	10.2	3	-0.7
14	CSGSRFLWDQVDPLIYAKQV	20.6 bits(41)	46	7/11(64%)	8/11(72%)	1/11(9%)	2324.69	8.9	1	-0.3
15	CSGDKLMDFNPLSAETSQKG	20.2 bits(40)	64	6/7(86%)	6/7(85%)	1/7(14%)	2127.38	6.4	0	0.3
16	CSGHIHPRNQDDKMETAFMN	20.2 bits(40)	64	5/5(100%)	5/5(100%)	0/5(0%)	2330.61	7.3	0.1	0.3
17	CSGGYSMKLETVIFSRIAME	20.2 bits(40)	64	8/14(57%)	8/14(57%)	1/14(7%)	2221.65	8.9	1	-0.2
18	CSGWTWKPAEGETEKPIPMN	19.7 bits(39)	88	5/6(83%)	5/6(83%)	0/6(0%)	2260.58	6.5	0	0.2
19	CSGRHETILHTDHLAIARIH	19.7 bits(39)	90	9/13(69%)	9/13(69%)	3/13(23%)	2279.62	9	1.3	-0.1
20	CSGPAVTMEHLWLFLYEQYY	19.7 bits(39)	90	5/5(100%)	5/5(100%)	0/5(0%)	2449.84	5.3	-1	-0.8
21	CSGVFEHTHGAKIMGPTHNT	19.7 bits(39)	90	7/11(64%)	8/11(72%)	1/11(9%)	2123.41	9	1.2	-0.2
22	CSGQWYRVVHYQVGYGMREE	19.7 bits(39)	90	7/15(47%)	8/15(53%)	5/15(33%)	2446.76	8.8	1	-0.2
23	CSGYHLYPHRNHWNSWPTRP	19.3 bits(38)	117	5/10(50%)	6/10(60%)	4/10(40%)	2507.78	10.3	3.2	-0.5
24	CSGEIQEVRTHATLYIILEK	19.3 bits(38)	123	8/10(80%)	8/10(80%)	2/10(20%)	2302.68	7.2	0	0
25	CSGERYQGWGMRAAMNMDGN	19.3 bits(38)	127	6/12(50%)	7/12(58%)	5/12(41%)	2233.52	8.9	1	0.1
26	CSGGQVIHYARIVQWDMYPL	18.5 bits(36)	144	4/4(100%)	4/4(100%)	0/4(0%)	2335.74	8.8	1	-0.7
27	CSGQALNNFVIPKLPYQSLN	18.9 bits(37)	174	5/6(83%)	5/6(83%)	0/6(0%)	2205.57	9.8	2	-0.5
28	CSGTTHIGNDLNVYQTRGFF	18.9 bits(37)	176	7/9(78%)	7/9(77%)	1/9(11%)	2229.46	8.9	1	-0.4
29	CSGQSAIVFAFWHWWIHQET	18.9 bits(37)	177	3/4(75%)	4/4(100%)	0/4(0%)	2432.75	7.3	0.1	-1
30	CSGLSIWPKWPYGQPLYWGL	18.9 bits(37)	178	7/14(50%)	7/14(50%)	4/14(28%)	2350.77	9.6	2	-1
31	CSGNRPMHEIMELYAFQGKR	18.5 bits(36)	234	4/5(80%)	5/5(100%)	0/5(0%)	2366.78	9.8	2	0.1
32	CSGVFPDYVYIFHYYWYMMP	18.5 bits(36)	248	5/5(100%)	5/5(100%)	0/5(0%)	2581.04	7.1	0	-1.3
33	CSGRGAIQHYSGNGWMYNIE	18.5 bits(36)	248	6/10(60%)	6/10(60%)	3/10(30%)	2242.48	8.8	1	-0.4
34	CSGVMQRWPENHFLQTHYDD	18.0 bits(35)	317	4/4(100%)	4/4(100%)	0/4(0%)	2462.71	6	-0.9	-0.1
35	CSGKDYWLGWDHPGRVVTDP	18.0 bits(35)	343	5/9(56%)	6/9(66%)	0/9(0%)	2287.54	7.2	0	0
36	CSGTSRDQMLYYNEFYPLHG	18.0 bits(35)	347	4/5(80%)	5/5(100%)	0/5(0%)	2380.65	7.1	0	-0.3
37	CSGYYLQAVTYYPNHFENPV	18.0 bits(35)	347	5/8(63%)	5/8(62%)	0/8(0%)	2364.62	7.1	0	-0.8
38	CSGTNKYWDQAIMYRKNHNV	17.6 bits(34)	460	5/9(56%)	5/9(55%)	0/9(0%)	2427.75	10	3	-0.1
39	CSGGIFDRVIAMYQVKIDVQ	17.6 bits(34)	478	8/15(53%)	10/15(66%)	3/15(20%)	2241.66	8.9	1	-0.2
40	CSGFWYVSENKAYFTFGTQQ	17.6 bits(34)	483	4/4(100%)	4/4(100%)	0/4(0%)	2362.61	8.8	1	-0.6
41	CSGLIAMTRILYQDSFKIIN	17.2 bits(33)	648	5/6(83%)	5/6(83%)	0/6(0%)	2285.76	9.8	2	-0.4
42	CSGNTHSQAMTQAQIIVPYI	17.2 bits(33)	671	4/4(100%)	4/4(100%)	0/4(0%)	2161.49	8.9	1	-0.6
43	CSGHHMLWGDRTIRMRGMQW	17.2 bits(33)	671	4/5(80%)	5/5(100%)	0/5(0%)	2457.92	12.2	3.1	-0.2
44	CSGYYFEFHANQIYTLNFFS	17.2 bits(33)	671	5/7(71%)	6/7(85%)	0/7(0%)	2450.72	7.1	0	-0.9
45	CSGPNVETQQHFYNLAGFKA	16.8 bits(32)	862	5/6(83%)	5/6(83%)	0/6(0%)	2210.46	8.9	1	-0.3
46	CSGVKKGPTDEYYRSNDQNF	16.3 bits(31)	1295	6/9(67%)	6/9(66%)	2/9(22%)	2307.49	8.8	1	0.5
	AVG	**19.57**	**228.05**				**2341.46**	**8.47**	**1.05**	**-0.32**
	STDEV	**1.95**	**273.30**				**121.57**	**1.51**	**1.26**	**0.40**
Red: acidic residues, like D E and C-terminal -COOH				Green: hydrophobic uncharged residues, like F I L M V W A and P						
Blue: basic residues, like R K H and N-terminal -NH2				Black: other residues, like G S T C N Q and P						

Table 2.1. Selected 46 significant peptides search by blastp.

cut-off value=1.20). As shown in Table 2.2, compared with the control group, the selected 46 peptides have more hits passing the cut-off value, and more 100% identical matches (highlighted in yellow). This indicates that at least some of the 46 selected peptides contain the real dengue epitope information, and the frequency is higher than randomly picked up ones. We cannot exclude the possibility that the other selected peptides from these 46 also have similarity with the real epitopes, which are not in the reported by IEDB.

Dengue vs Normal TOP 46 Higher in Dengue						
NO.	IEDB EPITOPE FOR DENGUE	SCORE	ALIGN LEN	PEP WINDOW	EPITOPE WINDOW	SCORE/LEN
1	TQGSNW	7.939	4	GSNW	GSNW	1.98
2	KLNWFKKGSSIGQ	7.822	4	lkNW	klNW	1.96
3	LKLNWF	7.822	4	lkNW	klNW	1.96
4	QLKLNW	7.822	4	lkNW	klNW	1.96
5	QLKLNWF	7.822	4	lkNW	klNW	1.96
6	QLKLNWFKKGSS	7.822	4	lkNW	klNW	1.96
7	KGMSYSMCTGKFK	7.365	4	GMSY	GMSY	1.84
8	SGNLLFTGHLKCRLRMDKLQLKGMSYSMCTG	7.365	4	GMSY	GMSY	1.84
9	CTGKFKIVKEIAETQHGTIVIRVQYEGDGSPC	7.331	4	QYEG	QYEG	1.83
10	GTIVIRVQYEG	7.331	4	QYEG	QYEG	1.83
11	RVQYEG	7.331	4	QYEG	QYEG	1.83
12	ELRYSWKTWGKAKMLSTELH	7.3	4	ELRY	ELRY	1.83
13	QPTELRYSW	7.3	4	ELRY	ELRY	1.83
14	TELRYSWKT	7.3	4	ELRY	ELRY	1.83
15	TELRYSWKTWGKAKM	7.3	4	ELRY	ELRY	1.83
16	FGKGGI	7.296	4	gfgk	fgkg	1.82
17	FLCKHSMVDRGWGNGCGLFGKGGIVTCAMFT	7.296	4	gfgk	fgkg	1.82
18	GLFGKG	7.296	4	gfgk	fgkg	1.82
19	GLFGKGGIVTCAMFT	7.296	4	gfgk	fgkg	1.82
20	GLFGKGGIVTCAMFTC	7.296	4	gfgk	fgkg	1.82
21	LFGKGG	7.296	4	gfgk	fgkg	1.82
22	AKNKPTLDFELIKTEAKQPAT	7.293	4	Ankn	Aknk	1.82
23	MSSGNL	7.288	4	SgsN	SsgN	1.82
24	ELRYSWKTWGKAKMLSTELH	7.19	4	KAKM	KAKM	1.80
25	TELRYSWKTWGKAKM	7.19	4	KAKM	KAKM	1.80
26	WKTWGKAKM	7.19	4	KAKM	KAKM	1.80
27	GVSGGS	7.174	4	SGGS	SGGS	1.79
28	MRCIGISNRDFVEGVSGGSWVDIVLEHGSC	7.174	4	SGGS	SGGS	1.79
29	SGGSWV	7.174	4	SGGS	SGGS	1.79
30	VSGGSW	7.174	4	SGGS	SGGS	1.79
31	VSGGSWVDIVLE	7.174	4	SGGS	SGGS	1.79
106	VSGGSWVDIVLE	7.258	6	sgWqnl	gsWvdi	1.21

46 Randomly Picked Peptide Sequences						
NO.	IEDB EPITOPE FOR DENGUE	SCORE	ALIGN LEN	PEP WINDOW	EPITOPE WINDOW	SCORE/LEN
1	HKYSWK	8.137	4	ykSW	kySW	2.03
2	DTQGSNWIQKETLVTFKNPHAKKQD	7.939	4	sgNW	gsNW	1.98
3	GSNWIQ	7.939	4	sgNW	gsNW	1.98
4	QGSNWI	7.939	4	sgNW	gsNW	1.98
5	TQGSNW	7.939	4	sgNW	gsNW	1.98
6	TQGSNWIQ	7.939	4	sgNW	gsNW	1.98
7	MRCIGISNRDFVEGVSGGSWVDIVLEHGSC	7.843	4	sgvw	gswv	1.96
8	SGGSWV	7.843	4	sgvw	gswv	1.96
9	VSGGSWVDIVLE	7.843	4	sgvw	gswv	1.96
10	LTTRNGEPHMIVMRQEKGKSLLFKTGDGV	7.215	4	VMqr	VMrq	1.80
11	FLIDGPDTSECPNERRA	7.16	4	GPtd	GPdt	1.79
12	TGYGTVTMECSPRTGLDFNE	7.069	4	VTME	VTME	1.77
13	WYGMEIRPLSEKEENMV	7.066	4	ywGl	wyGm	1.77
14	GLDFNE	6.802	4	Gifd	Gldf	1.70
15	RTGLDF	6.802	4	Gifd	Gldf	1.70
16	RTGLDFN	6.802	4	Gifd	Gldf	1.70
17	TGLDFN	6.802	4	Gifd	Gldf	1.70
18	TGYGTVTMECSPRTGLDFNE	6.802	4	Gifd	Gldf	1.70
19	GLDFNE	6.802	4	Gifd	Gldf	1.70
20	RTGLDF	6.802	4	Gifd	Gldf	1.70
21	RTGLDFN	6.802	4	Gifd	Gldf	1.70
22	TGLDFN	6.802	4	Gifd	Gldf	1.70
23	TGYGTVTMECSPRTGLDFNE	6.802	4	Gifd	Gldf	1.70
24	AWLVHRQWFLDLPLPW	6.778	4	kQfw	rQwf	1.69
25	HRQWFL	6.778	4	kQfw	rQwf	1.69
26	RQWFLD	6.778	4	kQfw	rQwf	1.69
27	VHRQWF	6.778	4	kQfw	rQwf	1.69
28	VHRQWFLDLPLP	6.778	4	kQfw	rQwf	1.69
29	LETRTETWMSSEGAWKHAQRIE	7.941	5	GAyKH	GAwKH	1.59
30	CTGKFKIVKEIAETQHGTIVIRVQYEGDGSPC	7.77	5	IRVhY	IRVqY	1.55
31	GTIVIRVQYEG	7.77	5	IRVhY	IRVqY	1.55
90	EDKAWL	7.286	6	EnKAyf	EdKAwl	1.21

Table 2.2. Comparison between selected 46 peptides with IEDB dengue real epitopes by GuiTope (Rank by SCORE/LEN). Top: Selected 46 peptides have more hits passing the cut-off value (106), the 100% identical ones are highlighted; Bot: Randomly picked up 46 peptides have less hits (90), and only one 100% identical match.

The next step, we set 28 dengue polyprotein sequences as the database, the 46 selected peptides are compared with the database, instead of ranking the hits by SCORE/LEN, we rank them by PRO POS (the starting position of the protein window). The idea here is the epitopes for a particular protein are usually clustered together at some hot spots with close sequence position. As an example shown in Table 2.3A, we

have 5 hits picked up by GuiTope, and they all match with the DEN23 polyprotein around position (317~325), which indicates the high possibility of real epitope around this location (317~325). Totally we can pick up 71 "predicted epitopes", and when these peptides are compared with the IEDB dengue real epitopes database, we find 31 of the 71 peptides match with the real dengue epitopes. The other 40 peptides might be new dengue epitopes (Table 2.3B). Interestingly, the sequence "QYEG/RVQYEG" (shown in the selected 46 peptides: csgysrssQYEGwhgkrdfm; csgqRaQYhGrmqtfyrgem) is picked up by both methods (Table 2.3;2.4), strongly indicating it's a real dengue epitope.

A

Protein: sp|P14339|POLG_DEN23 Genome polyprotein (Fragment) OS=Dengue virus type 2 (isolate Malaysia M3) PE=3 SV=1

Selected Random Peptide	SCORE	ALIGN LEN	PRO POS	PEP POS	PRO WINDOW	PEP WINDOW	SCORE/LEN
QRAQYHGRMQTFYRGEM	7.936	10	257	6	eGaMQTaltG	hGrMQTfyrG	0.7936
GSARGGAPGAARWHFST	8.429999	9	265	8	tGAAeiqmS	pGAArwhfS	0.936666556
WQYTRGNQYYPWHNVER	6.888	13	270	1	iQmssGNIlftgH	wQytrGNqyypwH	0.529846154
QIWRWVQQTHEPWTHRV	8.879	16	309	2	vkefaetqHgtiviRV	iwrwvqqtHepwthRV	0.5549375
YTHGRWMQVRQHPLEMY	8.842999	9	317	3	HGtivirvQ	HGrwmqvrQ	0.982555444
FGKWHRIKHIWIDEPSY	8.502999	15	318	2	GtivirvqyegDgsp	GkwhrikhiwiDeps	0.5668666
YQYDRHVWYPWGDDASR	6.607999	10	323	6	rVqYegdgsp	hVwYpwgdda	0.6607999
YQYDRHVWYPWGDDASR	6.607999	12	323	6	rVqYegdgspck	hVwYpwgddasr	0.550666583
YSRSSQYEGWHGKRDFM	7.331	4	325	6	QYEG	QYEG	1.83275

B

PREDICTED EPITOPES	SCORE	ALIGN LEN	IEDB POS	PEP POS	IEDB WINDOW	PREDICTED EPITOPES WINDOW	SCORE/LEN
RYSWKTWGKAK	21.353	11	3	1	RYSWKTWGKAK	RYSWKTWGKAK	1.941181818
LLQMEDKAW	11.569	6	1	4	MEDKAW	MEDKAW	1.928166667
IVLLQMEDKAW	11.569	6	1	6	MEDKAW	MEDKAW	1.928166667
PGYYTQTAGPWHLG	9.491	5	1	10	PWHLG	PWHLG	1.8982
RTGLDFNE	7.551	4	1	5	DFNE	DFNE	1.88775
TFKNAHAKKQEV	7.352	4	5	1	TFKN	TFKN	1.838
KIVQYENLKY	7.349	4	23	3	VQYE	VQYE	1.83725
RVQYEG	10.982	6	22	1	RVQYEG	RVQYEG	1.830333333
TFKNPHAKKQ	16.453	9	1	2	FKNPHAKKQ	FKNPHAKKQ	1.828111111
QYEG	7.331	4	24	1	QYEG	QYEG	1.83275

Table 2.3. Comparison between selected 46 peptides with 28 dengue polyprotein sequences by GuiTope (Rank by PRO POS). A: An example showing the selection method by PRO POS, the PRO WINDOW with more than three closely related selected random peptide matches is picked up as "predicted epitops"; B: comparison between the "predicted epitopes" with the IEDB dengue real epitopes (10 hits shown here, totally 31 over the cut-off value).

Selected Random Peptide	SCORE	ALIGN LEN	PRO POS	PEP POS	PRO WINDOW	PEP WINDOW	SCORE/LEN
SYFSDNAHKHQLKNWSG	7.939	4	1819	14	wnSG	nwSG	1.98475
YHGHYPMTGHKWGQVVM	7.934	4	2498	9	GkhW	GhkW	1.9835
YYHQGPRKHLKSNDWIY	7.475	4	963	11	skdn	ksnd	1.86875
QPRYDYDAFPYFKHVGK	7.412	4	128	12	kfvh	fkhv	1.853
QRAQYHGRMQTFYRGEM	9.224	5	2131	2	RAyqH	RAqyH	1.8448
QRAQYHGRMQTFYRGEM	9.224	5	2131	2	RAyqH	RAqyH	1.8448
QRAQYHGRMQTFYRGEM	9.224	5	2131	2	RAyqH	RAqyH	1.8448
EKFHYHHKAMNRSYAMI	7.362	4	517	10	MNsr	MNrs	1.8405
EKFHYHHKAMNRSYAMI	7.362	4	468	10	MNsr	MNrs	1.8405
KGSVYRFNDWHRHHTLM	9.194	5	575	1	KGvsY	KGsvY	1.8388
YSRSSQYEGWHGKRDFM	7.331	4	505	6	QYEG	QYEG	1.83275
YSRSSQYEGWHGKRDFM	7.331	4	374	6	QYEG	QYEG	1.83275
YSRSSQYEGWHGKRDFM	7.331	4	325	6	QYEG	QYEG	1.83275
YSRSSQYEGWHGKRDFM	7.331	4	605	6	QYEG	QYEG	1.83275
YSRSSQYEGWHGKRDFM	7.331	4	325	6	QYEG	QYEG	1.83275
QYKAKMWDPPNYKYQWA	8.415	7	894	3	KAKiftP	KAKmwdP	1.202142857

366

Table 2.4. Comparison between selected 46 peptides with 28 dengue polyprotein sequences by GuiTope (Rank by SCORE/LEN). Among the top 366 hits passing the cut-off value (1.20), the sequence QYEG stands out as a significant epitope.

As a conclusion, the work here firstly shows that CIM 10K Ver2 random peptide microarray is a useful diagnostic tool of monitoring the humoral immune response to dengue infection. Both the IgG and IgM immunosignature can be used to distinguish the dengue infection from the healthy, and the primary dengue infection from the secondary infection. It is also possible to use the IgG immunosignature to distinguish dengue from WNV, even between different dengue serotypes (more samples needed for verification). Second, although it is generally difficult to extract the real proteme information from the 10K microarray, we show that the real strong dengue epitopes are still able to stand out as significant. Some peptide sequences selected as significant but not reported by IEDB might be new dengue epitopes, which deserve further verification by ELISA. Third, we created a real dengue epitope subarray based on the its NS1 protein, and tested it on a small scale. There is not a significant improvement to use this subarray to distinguish

dengue infection vesus healthy compared with the 10K random microarray, however, we hope in the furture we can add more peptides from the other dengue proteins to expand the library and thus increase the detecting power.

APPENDIX 1

GLAM2: Search for the common motifs between selected significant peptide features

Based on our previous experience on the immunosignature of polyclonal antibodies of simple peptide antigen (Structural maintenance of chromosomes, SMC1 peptide), the significant features selected on the array share the common short motif, which is part of the real sequence on the original antigen. In order to test whether this is also true in the more complicated case: dengue infection, I use the GLAM2 (Gapped Local Alignment of Motifs) to analyze the similarity among the selected 46 peptide features distinguishing dengue infection and normal. As shown in the figure AP1 and AP2, I tried two different search conditions provided by GLAM2, although at certain positions, there is one or two common amino acids between at least 2 out of the 46 peptides, there is no common motifs longer than three. It appears that GLAM2 is not a powerful tool to decipher the dengue immunosignature.

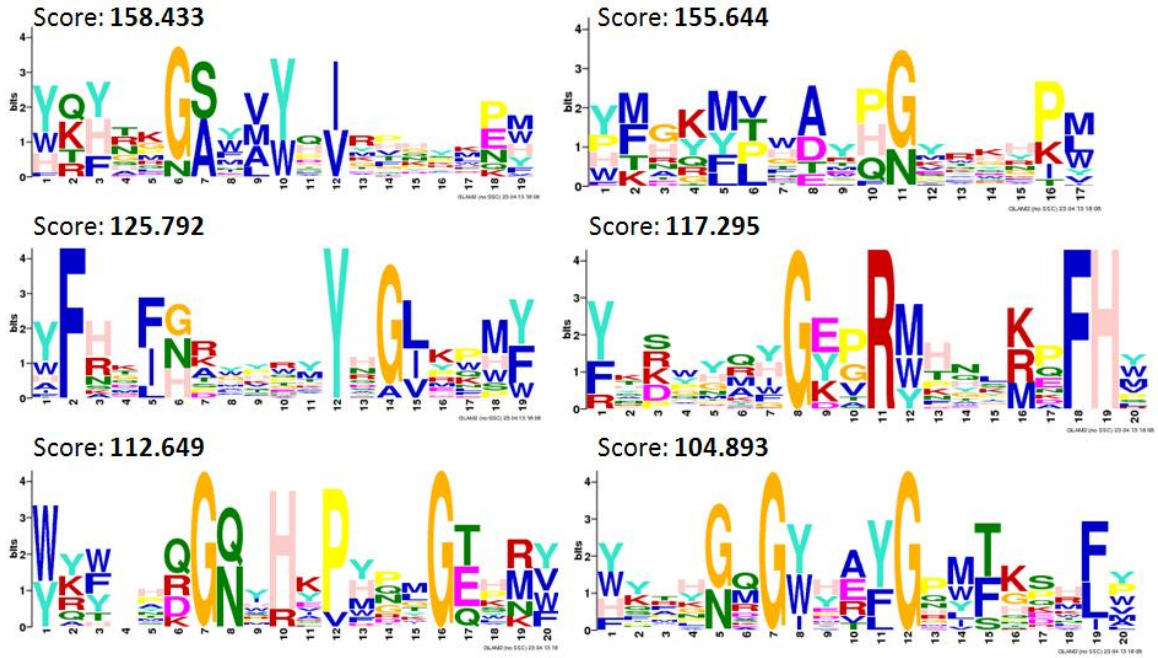


Figure AP1. The 46 selected significant peptides common motif discovery by GLAM2. GLAM2 setting: Minimum aligned sequences 2; Minimum aligned columns 2; Maximum aligned columns 50; Initial aligned columns; 20; Number of alignment replicates 10. The top 6 hits are shown here.

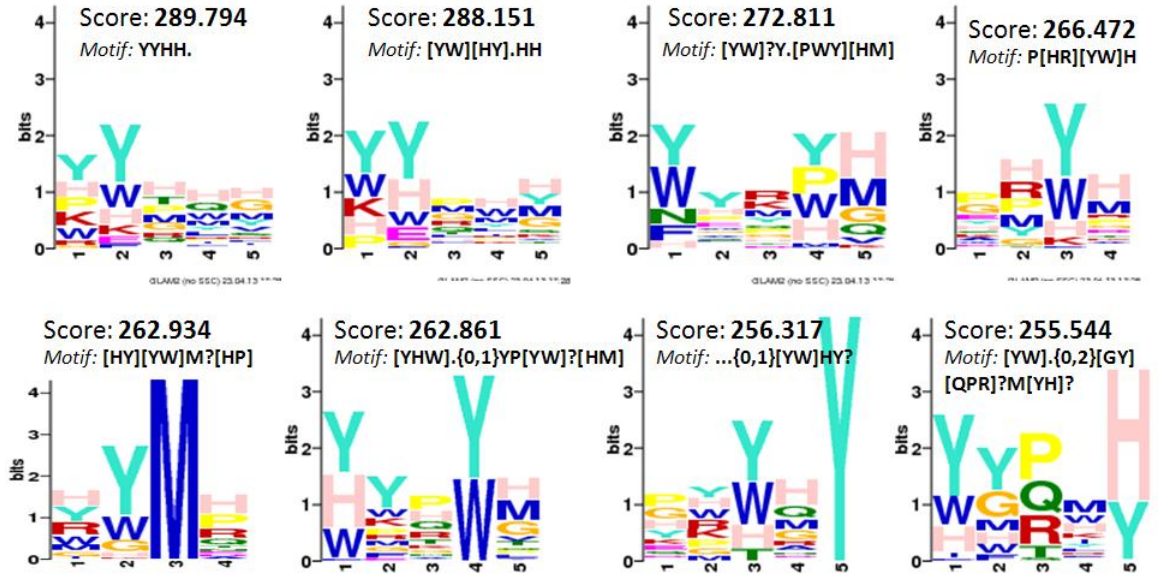


Figure AP2. The 46 selected significant peptides common motif discovery by GLAM2. GLAM2 setting: Minimum aligned sequences 2; Minimum aligned columns 4; Maximum aligned columns 17; Initial aligned columns; 5; Number of alignment replicates 10. The top 6 hits are shown here.

APPENDIX 2

Dengue NS1 subarray peptide list

Peptide Name	NTerm	Sequence	CTerm	MW (daltons)	Amount (mg)
NS1DEN1-1	[H]	CSGADSPKRLSAAIGRAWEE	[NH2]	2103	5.8
NS1DEN1-2	[H]	CSGAKIIGADVQNTTFIIDG	[NH2]	2022	7.8
NS1DEN1-3	[H]	CSGCGSRGPSLRTTTVTGKI	[NH2]	1980	4
NS1DEN1-4	[H]	CSGELDFDLCEGTTVVVDEH	[NH2]	2167	4.4
NS1DEN1-5	[H]	CSGEMIIPKMYGGPISQHNY	[NH2]	2224	3.5
NS1DEN1-6	[H]	CSGIRPQPMEHKYSWKSWGK	[NH2]	2404	9.1
NS1DEN1-7	[H]	CSGKEENLVKSMVSAGSGEI	[NH2]	2024	5
NS1DEN1-8	[H]	CSGKNETWKLARASFIEVKT	[NH2]	2267	8.5
NS1DEN1-9	[H]	CSGPDTPECPDEQRAWNIWE	[NH2]	2332	7.7
NS1DEN1-10	[H]	CSGQISNELNHILLENDMKF	[NH2]	2304	7
NS1DEN1-11	[H]	CSGSGCVINWKGRELKCGSG	[NH2]	2040	8.9
NS1DEN1-12	[H]	CSGTVVVGDVVGILAQGKKM	[NH2]	1960	6.6
NS1DEN2-13	[H]	CSGAKMVPTEPHNQTFLIDG	[NH2]	2144	9.9
NS1DEN2-14	[H]	CSGCHWPKSHTLWINGGLES	[NH2]	2211	7.9
NS1DEN2-15	[H]	CSGDICGIRSVTRLENLMWK	[NH2]	2280	6.7
NS1DEN2-16	[H]	CSGEMDFDFCDGTTVVVTED	[NH2]	2169	6.7
NS1DEN2-17	[H]	CSGEMIIPKNFAGPVSQHNY	[NH2]	2191	4.4
NS1DEN2-18	[H]	CSGMVIPKNIAGPVSQHNNR	[NH2]	2121	4.1
NS1DEN2-19	[H]	CSGGICGIRSVTRLENLMWK	[NH2]	2222	10.2
NS1DEN2-20	[H]	CSGITEWCCRSSTIPPLRIK	[NH2]	2249	6.6
NS1DEN2-21	[H]	CSGKDNRAVHDDMGYWIESA	[NH2]	2253	8.5
NS1DEN2-22	[H]	CSGKEKQDVFCDSKLMSAAI	[NH2]	2159	6.5
NS1DEN2-23	[H]	CSGLNDTWKIEKASFIEVKS	[NH2]	2254	5.8
NS1DEN2-24	[H]	CSGLRPQPTELRYSWKTWGK	[NH2]	2392	8.9
NS1DEN2-25	[H]	CSGPESPSKLASAMRKAHEE	[NH2]	2114	6.5
NS1DEN2-26	[H]	CSGPETAECPNTNRAWNSLE	[NH2]	2178	9.2
NS1DEN2-27	[H]	CSGQITSELNHILSENEVKL	[NH2]	2213	5.5
NS1DEN2-28	[H]	CSGREKEDLCCDSKVMSAAS	[NH2]	2118	11.1
NS1DEN2-29	[H]	CSGREKQDAFCDSKLMSAAI	[NH2]	2159	8.8
NS1DEN2-30	[H]	CSGRPGYYTQTAGPRHLGKL	[NH2]	2161	10.3
NS1DEN2-31	[H]	CSGRPLKEKEENLVTSLVTA	[NH2]	2173	7.2
NS1DEN2-32	[H]	CSGDSGCVVSWKNKELKCGS	[NH2]	2086	8
NS1DEN2-33	[H]	CSGTIMTGDIKGIMQAGTRS	[NH2]	2026	7.7
NS1DEN2-34	[H]	CSGTIMTGDIKGIMQVGKRS	[NH2]	2081	9.1
NS1DEN2-35	[H]	CSGELRYSWKTWGKAKMLSTELH	[NH2]	2711	4.7
NS1DEN2-36	[H]	CSGNRGPSLRTTTASGKLIT	[NH2]	2019	8
NS1DEN2-37	[H]	CSGGEDGCWYGMEIRPLKEK	[NH2]	2257	8
NS1DEN3-38	[H]	CSGADSPKRVATAIAGAWEN	[NH2]	2003	7.7
NS1DEN3-39	[H]	CSGAKIVTAETQNSSFIIDG	[NH2]	2040	6.1
NS1DEN3-40	[H]	CSGCGTRGPSLRATTVSGKL	[NH2]	1950	6
NS1DEN3-41	[H]	CSGCTWPKSHTLWTNGVLES	[NH2]	2205	1.9

NS1DEN3-42	[H]	CSGDMIIPKSLAGPISQHNY	[NH2]	2130	7.8
NS1DEN3-43	[H]	CSGELDFNYCEGTTVVITES	[NH2]	2166	6
NS1DEN3-44	[H]	CSGKNGSWKLEKASLIEVKT	[NH2]	2177	16.5
NS1DEN3-45	[H]	CSGMGCVINWKGKELKCGSG	[NH2]	2056	9.6
NS1DEN3-46	[H]	CSGPSTPECPSASRAWNVWE	[NH2]	2163	9.2
NS1DEN3-47	[H]	CSGQIANELNYILWENNIKL	[NH2]	2334	8.6
NS1DEN3-48	[H]	CSGREVYTQLCDHRLMSAAV	[NH2]	2238	9.6
NS1DEN3-49	[H]	CSGTVVVGDTLGVLEQGKRT	[NH2]	2018	4.7
NS1DEN4-50	[H]	CSGAKIFTPEARNSTFLIDG	[NH2]	2126	9.2
NS1DEN4-51	[H]	CSGEIDFGECPGTTVTIQED	[NH2]	2100	7.1
NS1DEN4-52	[H]	CSGKDQKAVHADMGYWLESS	[NH2]	2211	6.5
NS1DEN4-53	[H]	CSGPDTSECPNERRAWNSLE	[NH2]	2250	9.8
NS1DEN4-54	[H]	CSGPESPARLASAILNAHKD	[NH2]	2036	9.6
NS1DEN4-55	[H]	CSGQITNELNYVLWEGGHDL	[NH2]	2247	8.3
NS1DEN4-56	[H]	CSGQMLIPKSYAGPFSQHNY	[NH2]	2227	3
NS1DEN4-57	[H]	CSGRQGYATQTVGPWHLGKL	[NH2]	2158	6.8
NS1DEN4-58	[H]	CSGSEREENMVKSQVSAGQG	[NH2]	2082	6.4
NS1DEN4-59	[H]	CSGTGCAVSWSGKELKCGSG	[NH2]	1916	7.4
NS1DEN4-60	[H]	CSGTVVAGDVKGVLVKGKRA	[NH2]	1943	8
NS1DEN4-61	[H]	CSGVTQWCCRSCTMPPLRFL	[NH2]	2287	7.2
NS1DEN4-62	[H]	CSGFLIDGPDTSECPNERRA	[NH2]	2166	5.9
NS1DEN4-63	[H]	CSGKFQPESPARLASAILNA	[NH2]	2059	4.8
NS1DEN4-64	[H]	CSGFTTNIWMKFREGSSEVC	[NH2]	2281	7
NS1DEN4-65	[H]	CSGWYGMEIRPLSEKEENMV	[NH2]	2357	9

Table AP1. List of selected dengue NS1 subarray peptides. Totally 65 peptides are selected based on the differences between four dengue serotypes. NS1DEN1(1-12) are for dengue serotype1; NS1DEN2 (13-37) are for dengue serotype2; NS1DEN3 (38-49) are for dengue serotype3; NS1DEN4 (50-65) are for dengue serotype4. Peptides are synthezied by Sigma.

REFERENCES

1. Chiang, C.-Y., Centis, R. & Migliori, G. B. Drug-resistant tuberculosis: past, present, future. Respirology (Carlton, Vic.) 15, 413–32 (2010).

2. Howden, B. P. et al. Evolution of multidrug resistance during Staphylococcus aureus infection involves mutation of the essential two component regulator WalKR. PLoS pathogens 7, e1002359 (2011).

3. Högård, S. Antibiotic resistance - why is the problem so difficult to solve? Infection ecology & epidemiology 2, 1–7 (2012).

4. Peters, B. M., Shirtliff, M. E. & Jabra-Rizk, M. A. Antimicrobial peptides: primeval molecules or future drugs? PLoS pathogens 6, e1001067 (2010).

5. Peschel, A. & Sahl, H.-G. The co-evolution of host cationic antimicrobial peptides and microbial resistance. Nature reviews. Microbiology 4, 529–36 (2006).

6. Wang, G., Li, X. & Wang, Z. APD2: the updated antimicrobial peptide database and its application in peptide design. Nucleic acids research 37, D933–7 (2009).

7. Tossi, a, Tarantino, C. & Romeo, D. Design of synthetic antimicrobial peptides based on sequence analogy and amphipathicity. European journal of biochemistry / FEBS 250, 549–58 (1997).

8. He, J. et al. Novel synthetic antimicrobial peptides against Streptococcus mutans. Antimicrobial agents and chemotherapy 51, 1351–8 (2007).

9. Pini, A. et al. Antimicrobial Activity of Novel Dendrimeric Peptides Obtained by Phage Display Selection and Rational Modification. 49, 2665–2672 (2005).

10. Casadevall, A., Dadachova, E. & Pirofski, L. Passive antibody therapy for infectious diseases. Nature reviews. Microbiology 2, 695–703 (2004).

11. Diehnelt, C. W. et al. Discovery of high-affinity protein binding ligands--backwards. PloS one 5, e10728 (2010).

12. Gupta, N., Belcher, P. E., Johnston, S. A. & Diehnelt, C. W. Engineering a synthetic ligand for tumor necrosis factor-alpha. Bioconjugate chemistry 22, 1473–8 (2011).

13. McEnaney, P. J., Parker, C. G., Zhang, A. X. & Spiegel, D. a Antibody-recruiting molecules: an emerging paradigm for engaging immune function in treating human disease. ACS chemical biology 7, 1139–51 (2012).

14. Parker, C. G., Domaoal, R. a, Anderson, K. S. & Spiegel, D. a An antibody-recruiting small molecule that targets HIV gp120. Journal of the American Chemical Society 131, 16392–4 (2009).

15. CLSI Methods for dilution antimicrobial susceptibility tests for bacteria that grow aerobically. Approved standard-Eight Edition. CLSI document M07-A8. Wayn, PA: Clinical and Laboratory Standards Institute. (2009).

16. Lee, S. et al. and hemolytic activities of a-helical antibiotic peptide , P18 and its analogs. 504–514 (2001).

17. Nakatsuji, T. et al. NIH Public Access. 129, 2480–2488 (2009).

18. Macura, N., Zhang, T. & Casadevall, A. Dependence of macrophage phagocytic efficacy on antibody concentration. Infection and immunity 75, 1904–15 (2007).

19. Poirier, K. et al. Escherichia coli O157:H7 survives within human macrophages: global gene expression profile and involvement of the Shiga toxins. Infection and immunity 76, 4814–22 (2008).

20. Conlon, J. M. et al. Host-defense peptides in skin secretions of the tetraploid frog Silurana epitropicalis with potent activity against methicillin-resistant Staphylococcus aureus (MRSA). Peptides 37, 113–9 (2012).

21. Shai, Y. From innate immunity to de-novo designed antimicrobial peptides. Current pharmaceutical design 8, 715 (2002).

22. Brogden, K. a Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nature reviews. Microbiology 3, 238–50 (2005).

23. Herce, H. D. et al. Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides. Biophysical journal 97, 1917–25 (2009).

24. Domenyuk, V., Loskutov, A., Johnston, S. A. & Diehnelt, C. W. A technology for developing synbodies with antibacterial activity. PloS one 8, e54162 (2013).

25. Brouwer, C. P. J. M., Rahman, M. & Welling, M. M. Discovery and development of a synthetic peptide derived from lactoferrin for clinical use. Peptides 32, 1953–63 (2011).

26. Chen, Y. et al. Helical Antimicrobial Peptides with Enhanced Activities and Specificity / Therapeutic Index. 280, 12316–12329 (2006).

27. Hong, S. Y., Park, T. G. & Lee, K. H. The effect of charge increase on the specificity and activity of a short antimicrobial peptide. Peptides 22, 1669–74 (2001).

28. Bruschi, M. et al. Synthesis, characterization, antimicrobial activity and LPS-interaction properties of SB041, a novel dendrimeric peptide with antimicrobial properties. Peptides 31, 1459–67 (2010).

29. Cao, L. et al. Antibacterial activity and mechanism of a scorpion venom Peptide derivative in vitro and in vivo. PloS one 7, e40135 (2012).

30. Menousek, J. et al. Database screening and in vivo efficacy of antimicrobial peptides against methicillin-resistant Staphylococcus aureus USA300. International journal of antimicrobial agents 39, 402–6 (2012).

31. Rodriguez, P. L. et al. Minimal "Self" peptides that inhibit phagocytic clearance and enhance delivery of nanoparticles. Science (New York, N.Y.) 339, 971–5 (2013).

32. Diagnosis GFOR Dengue: guidelines for diagnosis, treatment, prevention and control. (2009).

33. Rothman, A. L. Immunity to dengue virus: a tale of original antigenic sin and tropical cytokine storms. Nature reviews. Immunology 11, 532–43 (2011).

34. Wahala, W. M. P. B. & Silva, A. M. De The human antibody response to dengue virus infection. Viruses 3, 2374–95 (2011).

35. Peeling RW, Artsob H, Pelegrino JL, Buchy P, Cardosa MJ, Devi S, Enria D a., Farrar J, Gubler DJ, G. M. & Halstead SB, Hunsperger E, Kliks S, Margolis HS, Nathanson CM, Nguyen VC, Rizzo N, V. S. & Y. S. Evaluation of diagnostic tests: dengue. Nature Reviews Microbiology 8, (2010).

36. Hunsperger E a, Yoksan S, Buchy P, Nguyen VC, Sekaran SD, Enria D a, Pelegrino JL, Vázquez S, A. H., Drebot M, Gubler DJ, Halstead SB, Guzmán MG, Margolis HS, Nathanson C-M, Rizzo Lic NR, Bessoff KE, K. S. & RW, & P. Evaluation of commercially available anti-dengue virus immunoglobulin M tests. Emerging infectious diseases 15, 436–40 (2009).

37. Yun Ying Tan Development of ASSURE Dengue IgA Rapid Test for the Detection of Anti-dengue IgA from Dengue Infected Patients. J Glob Infect Dis Jul-Sep, 233–240 (2011).

38. Immunization, Vaccines and Biologicals.

39. Thomas, S. J. et al. Dengue plaque reduction neutralization test (PRNT) in primary and secondary dengue virus infections: How alterations in assay conditions impact performance. The American journal of tropical medicine and hygiene 81, 825–33 (2009).

40. Yan X, Zhong W, Tang A, Schielke EG, H. W. & N. J. Multiplexed Flow Cytometric Immunoassay for Influenza Virus Detection and Differentiation protein detection . A four-plexed assay for influenza virus was developed to demonstrate the potential for multi-. 77, 7673–7678 (2005).

41. Johnson AJ, Cheshier RC, Cosentino G, Masri HP, Mock V, Oesterle R, Lanciotti RS, Martin D a, P. A. & BJ, K. O. & B. Validation of a microsphere-based immunoassay for detection of anti-West Nile virus and anti-St. Louis encephalitis virus immunoglobulin m antibodies. Clinical and vaccine immunology : CVI 14, 1084–93 (2007).

42. Andresen H & Grötzinger C Deciphering the Antibodyome - Peptide Arrays for Serum Antibody Biomarker Diagnostics. 1–12 (2009).

43. Keasey SL, Schmid KE, Lee MS, Meegan J, Tomas P, Minto M, Tikhonov AP, S. B. & U. R. Extensive antibody cross-reactivity among infectious gram-negative bacteria revealed by proteome microarray analysis. Molecular & cellular proteomics : MCP 8, 924–35 (2009).

44. Foong YM, Fu J, Y. S. & U. M. Current advances in peptide and small molecule microarray technologies. Current opinion in chemical biology 16, 234–42 (2012).

45. Gaseitsiwe S, Valentini D, Mahdavifar S, Magalhaes I, Hoft DF, Zerweck J, Schutkowski M, Andersson J, R. M. & MJ, & M. Pattern recognition in pulmonary tuberculosis defined by high content peptide microarray chip analysis representing 61 proteins from M. tuberculosis. PloS one 3, e3840 (2008).

46. Chow SCS, Ho CYS, Tam TTY, Wu C, Cheung T, Chan PKS, Ng MHL, Hui PK, Ng HK, A. D. & L. a W. Specific epitopes of the structural and hypothetical proteins elicit variable humoral responses in SARS patients. Journal of clinical pathology 59, 468–76 (2006).

47. Guo J-P, Petric M, C. W. & M. P. SARS corona virus peptides recognized by antibodies in the sera of convalescent cases. Virology 324, 251–6 (2004).

48. Reddy MM, Wilson R, Wilson J, Connell S, Gocke A, Hynan L, G. D. & K. T. Identification of candidate IgG biomarkers for Alzheimer's disease via combinatorial library screening. Cell 144, 132–42 (2011).

49. Vereda A, Andreae D a, Lin J, Shreffler WG, Ibañez MD, Cuesta-Herranz J, B. L. & S. H. a Identification of IgE sequential epitopes of lentil (Len c 1) by means of peptide microarray immunoassay. The Journal of allergy and clinical immunology 126, 596–601.e1. (2010).

50. Lin J, Bardina L, Shreffler WG, Andreae D a, Ge Y, Wang J, Bruni FM, Fu Z, H. Y. & S. H. a Development of a novel peptide microarray for large-scale epitope

mapping of food allergens. The Journal of allergy and clinical immunology 124, 315–22, 322.e1–3. (2009).

51. List C, Qi W, Maag E, Gottstein B, M. N. & F. I. Serodiagnosis of Echinococcus spp. infection: explorative selection of diagnostic antigens by peptide microarray. PLoS neglected tropical diseases 4, e771. (2010).

52. Maksimov P, Zerweck J, Maksimov A, Hotop A, Gross U, Pleyer U, Spekker K, Däubener W, W. S. & Niederstrasser O, Petri E, Mertens M, Ulrich RG, C. F. & S. G. Peptide microarray analysis of in silico-predicted epitopes for serological diagnosis of Toxoplasma gondii infection in humans. Clinical and vaccine immunology : CVI 19, 865–74 (2012).

53. López-Campos GH, García-Albert L, M.-S. F. & G.-S. a Analysis and management of HIV peptide microarray experiments. Methods of information in medicine 45, 158–62 (2006).

54. Stafford P, Halperin R, Legutki JB, Magee DM, G. J. & J. S. Physical characterization of the "immunosignaturing effect". Molecular & cellular proteomics : MCP 11, M111.011593 (2012).

55. Legutki JB, Magee DM, S. P. & J. S. A general method for characterization of humoral immunity induced by a vaccine or infection. Vaccine 28, 4529–37 (2010).

56. Hughes AK, Cichacz Z, Scheck A, Coons SW, J. S. & S. P. Immunosignaturing can detect products from molecular markers in brain cancer. PloS one 7, e40201 (2012).

57. Nawa, M. et al. Immunoglobulin A Antibody Responses in Dengue Patients : a Useful Marker for Serodiagnosis of Dengue Virus Infection. 12, 1235–1237 (2005).

58. Deparis, X., Labeau, B., Lelarge, J., Morvan, J. & Dussart, P. Discrimination between Primary and Secondary Dengue Virus Infection by an Immunoglobulin G Avidity Test Using a Single Acute-Phase Serum Sample Se. 43, 2793–2797 (2005).

59. Banoo, S. et al. Evaluation of diagnostic tests for infectious diseases: general principles. Nature reviews. Microbiology 4, S21–31 (2006).

60. Puttikhunt, C. et al. The development of a novel serotyping-NS1-ELISA to identify serotypes of dengue virus. Journal of clinical virology : the official publication of the Pan American Society for Clinical Virology 50, 314–9 (2011).

61. Ocazionez, R. E., Cortés, F. M., Villar, L. A. & Gómez, S. Y. Temporal distribution of dengue virus serotypes in Colombian endemic area and dengue

incidence. Re-introduction of dengue-3 associated to mild febrile illness and primary infection. Memórias do Instituto Oswaldo Cruz 101, 725–31 (2006).

62. San Martín, J. L. et al. The epidemiology of dengue in the americas over the last three decades: a worrisome reality. The American journal of tropical medicine and hygiene 82, 128–35 (2010).

63. Halperin, R. F., Stafford, P., Emery, J. S., Navalkar, K. A. & Johnston, S. A. GuiTope: an application for mapping random-sequence peptides to protein sequences. BMC bioinformatics 13, 1 (2012).