論文 / 著書情報
Article / Book Information

題目(和文)	
Title(English)	Prediction of Material Properties of Model Organic Biomaterials by Machine Learning
著者(和文)	Kwaria Rudolf Jason
Author(English)	Rudolf Jason Kwaria
出典(和文)	学位:博士(理学), 学位授与機関:東京工業大学, 報告番号:甲第11580号, 授与年月日:2020年9月25日, 学位の種別:課程博士, 審査員:林 智広,曽根 正人,和田 裕之,柘植 丈治,石田 忠
Citation(English)	Degree:Doctor (Science), Conferring organization: Tokyo Institute of Technology, Report number:甲第11580号, Conferred date:2020/9/25, Degree Type:Course doctor, Examiner:,,,,
学位種別(和文)	博士論文
Type(English)	Doctoral Thesis

AUGUST 2020

DOCTOR DISSERTATION

PREDICTION OF MATERIAL PROPERTIES OF MODEL ORGANIC BIOMATERIALS BY MACHINE LEARNING

TOKYO INSTITUTE OF TECHNOLOGY

SCHOOL OF MATERIALS AND CHEMICAL TECHNOLOGY

DEPARTMENT OF MATERIALS SCIENCE AND ENGINEERING

HUMAN CENTERED SCIENCE AND BIOMEDICAL ENGINEERING

HAYASHI LABORATORY

Rudolf Jason Kwaria

Academic advisors:

Assoc. Prof. Tomohiro Hayashi

Table of Contents

List of Tables

List of Figures

List of Equation

Chapter 1. Introduction

In the field of biomaterials, there has been many materials such as ceramics, polymers, and metals being developed as medical devices such as dental implants, artificial blood vessels, and biosensors.[1-6] Developing these materials into various medical devices requires changing the properties of these materials in which the adjusted property will vary between each usage of medical device. For certain biomedical devices, the hydrophilicity of the material involved is important to ensure adhesion with its surrounding tissue. [7],[8] For other materials, in order to prevent blood clotting, the material used for the biomedical device needs to have low or even prevent the adsorption of fibrinogen into the material. [9]

Researcher has developed many ways to design material with specific properties. One of the most classical method to design material is by manually changing the structure of the material experimentally and then check the property of the material. If the material with desired property has been successfully obtained, then the designing process can be stopped. This method is called trial-and-error method (Figure 1.1) and although many novel materials have been discovered using trial-and-error method, this method is usually conducted by non-targeted fashion with unknown criteria (Figure 1.2) about which material structure must be changed to achieve the desired property. Although some empirical criteria for hydrophilicity and fibrinogen adsorption exists[10, 11], there has been lack of quantitative criteria to make more targeted material design. This material designing method is inefficient with respect to the time and resources.

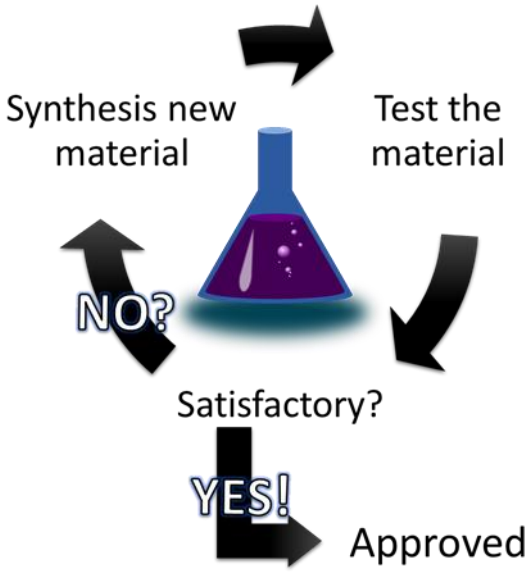


Figure 1. 1 Material design process by trial-and-error

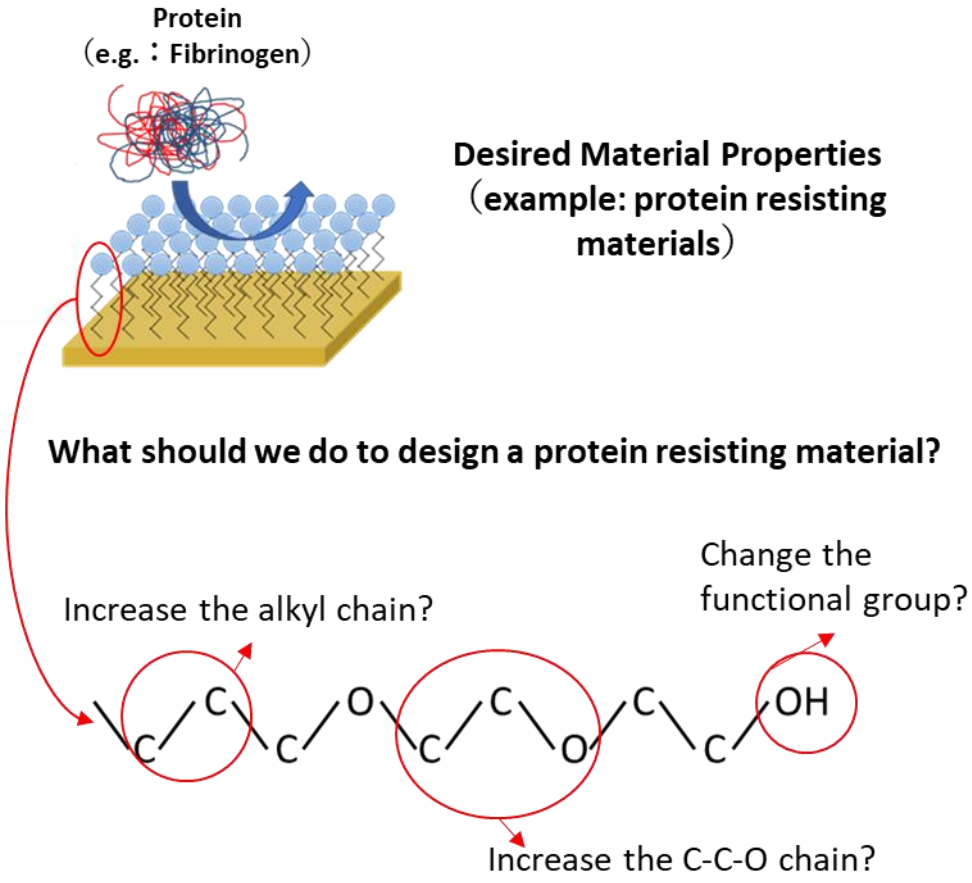


Figure 1. 2 Unknown criteria in designing material with specific properties which leads to trial-and-error material design process

In order to solve this problem a method that can make criteria for material design as well as predict material properties without an actual synthesis is necessary. To do this, researchers starts developing ways to design material in a more targeted fashion. In the case of solid materials, the theoretical calculation using density functional theory (DFT) or molecular simulation using atom interactions (Monte Carlo and molecular dynamics) are often used. Using the above theoretical methods, elemental composition, crystal structure, and other structural properties can be freely changed, and the resulting material property can be calculated in theoretical fashion. Using this method, it is possible to simulate many kinds of materials without having to manually synthesize the material which could reduce the cost of material design.

However, the material properties used in biomaterial design could not be easily calculated using the above theoretical calculation. For example, to simulate the adhesion of proteins or cells into materials, the proteins and the cells must be carefully modelled. However, modelling proteins and cells for simulation are computationally heavy and demanding. Moreover, with the cost of computation being expensive, the time required to model protein and cells would be too long.

Material informatics can supposedly solve this problem. Material informatics is an emerging alternative method to design novel materials. [12, 13] It has made remarkable success in various field of material design. Since its initiation in “Material Genome Initiative” in the United States, this method has made remarkable success in various fields to design batteries, catalysts, and semiconductors.[14-17]. In material informatics approach, a wide variety of mathematical and algorithms (also known as machine learning algorithm) are used to analyse the correlation between chemical structures of materials and the material properties. This includes many complex

material properties such as electronic structure, conductivity, catalytic activity, and magnetism. After the correlation is understood, the algorithm would be able to correctly predict the material properties from the analysed correlation. Moreover, with the analysed correlation, the algorithm would be able to make quantitative criteria to design material with specific property.

Although using material informatics can seemingly solve many problems in biomaterial design, this method relies heavily on the availability of data to train the algorithm. Using past experimental results in the form of papers and publications, it is supposedly possible to construct data to be used in material informatics method. The next question goes to the possibility for the applicability of the previously reported results to predict the material properties. It is equally important to understand whether it is possible or not to apply this method on a data constructed by experimental results conducted with different experimental methods.

In this work, we use self-assembled monolayers (SAMs) as our study platform. Since 1990, SAMs has been widely employed as a platform to model organic surfaces to study interactions between protein and materials. Its easiness to prepare and flexibility in controlling the surface property by changing the terminal groups of SAM with different molecules allows many types of SAMs to be made and studied. [18] At present, there have been more than several hundreds of SAMs with different WCA and fibrinogen adsorption reported which makes it ideal to be used as platform for material informatics study.

In this research, we discuss the possibility of data-driven approach to predict material properties from the structure of the molecules constituting the material. We are mainly focusing on the hydrophilicity, which is shown by water contact angle (WCA),

and fibrinogen adsorption. We constructed databases consisting information on WCA and fibrinogen adsorption and analyse the correlation between the molecular structure with WCA and fibrinogen adsorption using machine learning algorithm (ML) called Artificial Neural Network (ANN) algorithm. We will discuss the possibility of predicting WCA and fibrinogen adsorption using this method and the challenges that must be overcome to apply this method.

We will also investigate the possibility of using material informatics method to make quantitative criteria. We applied this to make quantitative criteria to design material with specific WCA as well as fibrinogen adsorption in the form of ranking the importance of specific material structure towards the targeted material property. In particular, we discuss the agreement of our quantitative criteria with the empirically established publications.

Lastly, we will demonstrate how to screen and design SAMs exhibiting desired WCA and fibrinogen adsorption

1.1 Outline of this work

Chapter 1 briefly introduced the background of this topics and about material informatics in general. The attention was focused on the application of material informatics to optimize material design processes. In order to explain the algorithm used for this material informatics method in detail, **Chapter 2** describes detailed theory and mathematics of the algorithm used in this research. **Chapter 3** provides more detail about application of this method to predict simple material property like WCA as well as to analyze the importance of material descriptors towards WCA.

After we managed to make prediction for simple material property like WCA, in **Chapter 4**, we tried more complex material property which is fibrinogen adsorption.

During this phase, we recognized the problem that may arise with the prediction of fibrinogen adsorption in which the experimental method applied to gather the data. Therefore, in this chapter, we make comparison of the prediction when the ANN is trained by dataset gathered from single laboratory and when the ANN is trained by dataset gathered from various laboratory.

Finally, a short summary for this investigation and the overall conclusion is provided in **Chapter 5**.

Chapter 2: Artificial Neural Network Model

Machine learning algorithm is the backbone of data-science related research. Being related to data-science, material informatics also utilized machine learning to study the pattern of material data and use the knowledge to either predict the properties of the material or calculate the importance of the molecular descriptors to make quantitative guide for material design.

Among so many available machine learning algorithms, Artificial Neural Network (ANN) algorithm has gained reputation to solve many complex problems due to its prediction accuracy and ease of training. Moreover, the architecture of this algorithm allows it to be modified according to the needs of the research. Outside of material science research, ANN has been used for image recognition as well as detecting meaning of sentence through natural language processing (NLP). In material science, ANN has been used to solve both theoretical problems such as optimization of functional of density functional theory (DFT) and practical problems such as predicting thermal conductivity of materials from basic properties.

In order to be able to utilize this algorithm better, the knowledge regarding this algorithm is necessary. In this chapter, I will discuss the properties of ANN algorithm such as the mathematical background behind the algorithm. I will begin by explaining the structure of the ANN algorithm. In this chapter, I will also explain the mathematics of the algorithm which includes the activation function in the ANN algorithm and method of reducing prediction error employed in ANN algorithm.

2. 1 Structure of ANN algorithm

Before we discuss about mathematics of ANN algorithm, we must have an understanding about the structure of the ANN and what inspires the structure of ANN.

ANN is an algorithm inspired by natural nerve system of so many living beings. One of the cells involved in nerve systems is called neuron. Inside the nerve systems, hundreds of neurons work together by gaining information from one or more neurons, process the information inside the neurons, and then pass the processed information to other neurons. Every transfer of information is done through the connection between one neuron and the others

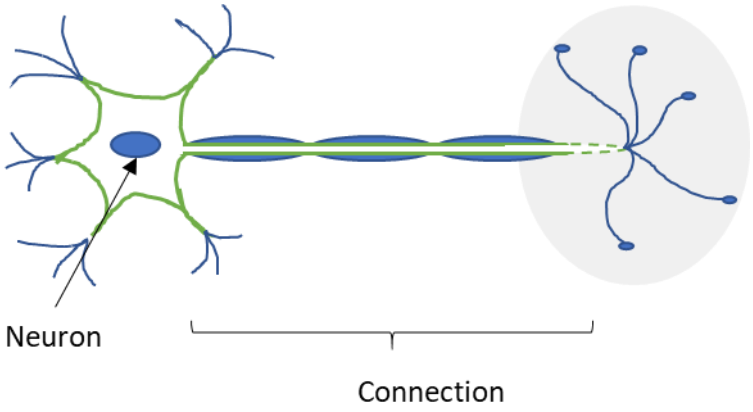


Figure 2. 1 Schematic of Biological Neuron

Figure 2. shows the structure of the ANN algorithm for prediction of WCA or fibrinogen adsorption which were used in this research. This model is the most basic ANN model which is called multilayer feed-forward ANN.

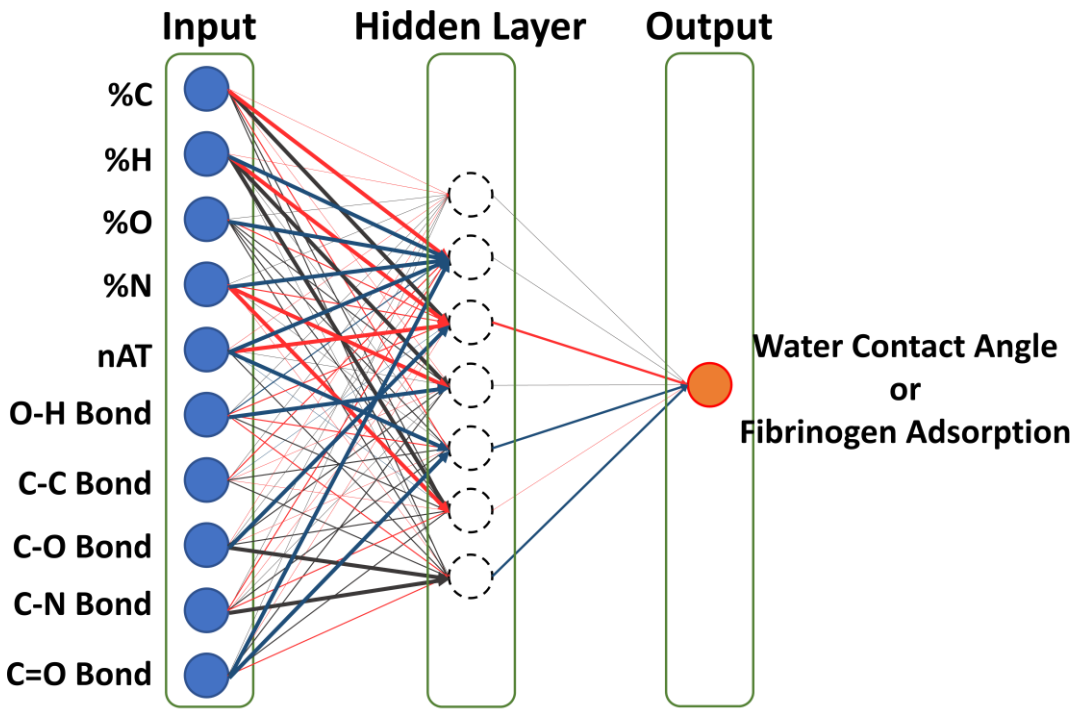


Figure 2. 2 ANN Model to predict material properties based on the structure descriptors. The red and blue arrows illustrate positive and negative weights after the training, respectively.

Just like the nerve system, ANN works through the transfer of information of several elements in its systems. Each circle in the ANN model, which is called neuron, acts just like the neurons in the nerve system which is to process the information. The processed information is then transferred into the next of neurons through the connectors which acts like the connection between neuron cells for further process.

In a standard multilayer feed-forward ANN, the neurons inside the ANN were divided into three layers. The first layer is called input layer. Inside this layer, all neuron represents the independent variables of our problem. In the case of prediction of material properties using structure descriptors, neurons inside the input layer represent the structure descriptors of the materials. The third layer of the multilayer feed-forward ANN is called output layer and it contains the dependent variable of our problem. In the case of prediction of material properties using structure descriptors, neurons inside

the output layer represents the material properties. The layer in-between is called hidden layer. There could be more than one hidden layer with several neurons inside each layer. This layer serves as an element that adds complexity into the model. If the model that needs to be solved is complicated, the number of hidden layers and neurons inside the hidden layers needs to be increased.

The number of hidden layers and hidden neurons should be adjusted manually by researchers depending on the research objective and the nature of the database.

In this simple model, each of the neurons connected into one or more neurons in the next layer without any connection between neuron inside same layer, hence the algorithm is called feed-forward ANN.

2.2. Mathematics of ANN Algorithm

2.2.1 Activation function of ANN nodes

In order to understand how ANN algorithm works, we need to understand how information is processed and transferred from one neuron into other neurons in ANN structure.

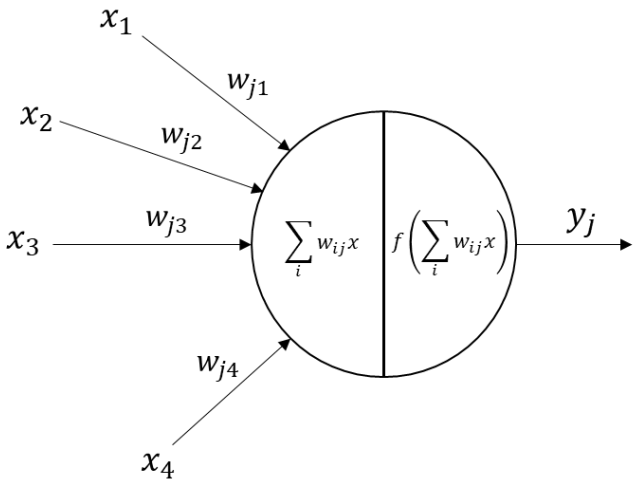


Figure 2. 3 Schematic description of the relationship between input and output vectors of one neuron

Figure 2.3 shows how information is being processed in one ANN neuron. x and y represents the information that comes in (henceforth will be referred as input vector) and out (henceforth will be referred as output vector) of a single neuron respectively while w represents the strength of connection between one neuron and other neurons, also known as the connection weight. The relationship between x and y with the connection weight w can be described as follows

$$y_j = f\left(\sum_i w_{ij}x_i\right) \quad \text{Eq. 1}$$

In which $f(x)$ is an activation function used to process the information in a single neuron. In this research where the ANN algorithm has 3 layered neural network, there are two types of activation function. One is the activation function employed by neurons in the hidden layer, and the other one is the activation function employed by neurons in the output layer. The neurons inside the input layer do not have activation function as its task is to transfer the information as it is to the neurons in the next layer.

For the neurons inside the hidden layer, the activation function employed is tan-sigmoid function which is shown in expression below:

$$f\left(\sum_i w_{ij}x_i\right) = \frac{1}{1+e^{-(\sum_i w_{ij}x_i)}} \qquad \text{Eq. 2}$$

By using tan-sigmoid function, the ANN can model many kinds of system including non-linear systems and make prediction accurately.

For the neurons inside the output layer, the activation function is linear function which is shown in expression below:

$$f\left(\sum_i w_{ij}x_i\right) = \sum_i w_{ij}x_i \qquad \text{Eq. 3}$$

In which the activation function is the sums the product of connection weight w and the input vector x.

With each neuron having their own activation function, the overall activation function of the ANN becomes the combination of the activation function of each neuron. This means, increasing the number of neurons in the ANN either by adding more features in the input layer or adding more neurons in the hidden layer will make the calculation process become more complex, therefore, slowing down the calculation.

On the other hand, increasing the number of neurons, particularly in the hidden layer, might enable the ANN to model more sophisticated problems.

2.2.2 Adjusting Connection Weight of the ANN

Result of ANN algorithm is a prediction of the actual material properties. Therefore, there is a possibility that the prediction result is different than the actual material properties. This difference between the prediction and the actual material properties is called error.

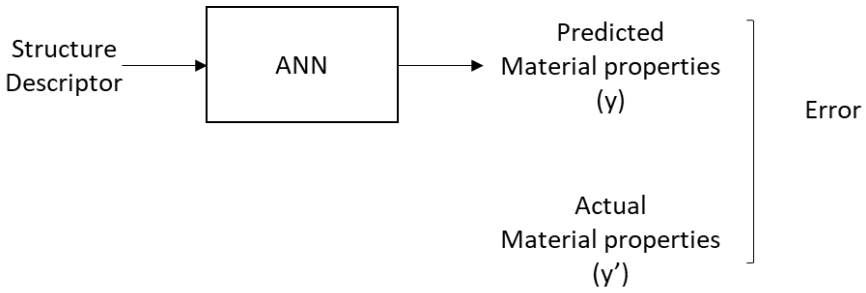


Figure 2. 4 The illustration of error

Figure 2.4 illustrates the error or difference between the prediction and the actual values. In order to make an accurate prediction, ANN must be able to minimize the error between the predicted material properties and the actual material properties.

The error function of a dataset with sample size of n that must be minimized by the ANN is expressed with formula below:

$$J = \sum_{i=1}^{n} (y_i - y'_i)^2 \qquad \text{Eq. 4}$$

In which n,y, and y'in both equations are the total number of data, the expected value, and the value as predicted by the ANN, respectively.

Looking back at the activation function of the ANN, we can see that the activation function of the ANN considers the connection weight of the ANN (w), which represents the strength of connection between one node and the other. Therefore, it is important to adjust the connection weight of the ANN accordingly so that the error of prediction is minimum. The process of minimization of error is when the ANN algorithm learns the pattern of the database to find the appropriate value for connection weight between each neuron which is why the ANN algorithm is called machine learning algorithm. The adjustments itself can be done using an optimization algorithm called Stochastic Gradient Descent (SGD) algorithm.

The adjustment of connection weight of the ANN starts with the machine learning algorithm makes an initial randomized guess of connection weight. Next, the SGD algorithm considers the error of prediction by the ANN using the initial randomized guess of connection weight and use the error value to adjusts the connection weight of the ANN. This process is done iteratively until the error reaches its minimum value. Figure 2.5 illustrates how SGD algorithm works.

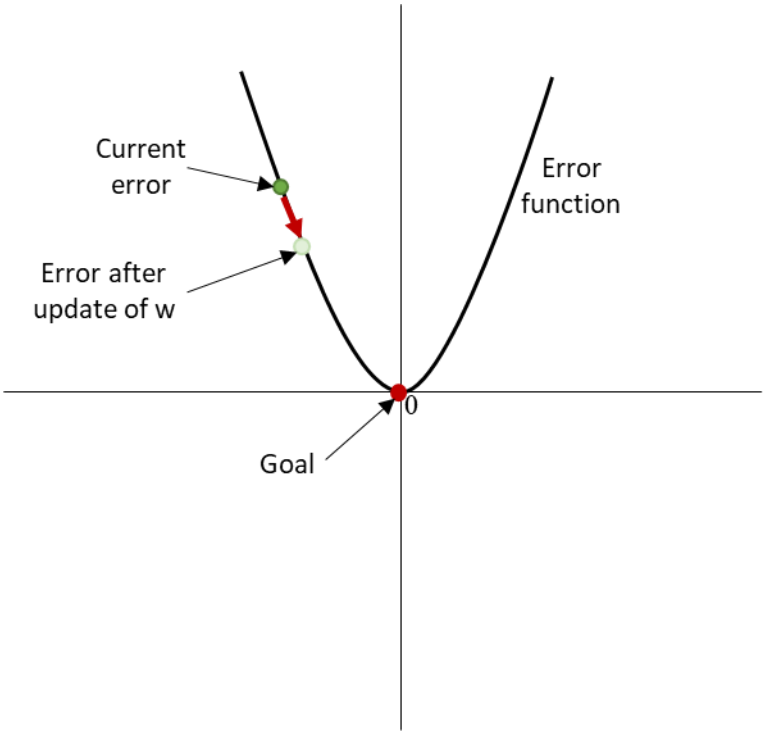


Figure 2. 5 Minimization of the error value

The formula to optimize weight through SGD algorithm iteratively is expressed as follows:

$$w_{ij}(t+1) = w_{ij}(\mathrm{t}) - \eta \frac{\partial J}{\partial w_{ij}(t)} \qquad \text{Eq. 5}$$

In which the $w_{ij}(t+1)$ and $w_{ij}\,(t)$ is the connection weight after and before being updated respectively. η is a parameter called learning rate. The learning rate can be adjusted accordingly by the user in range of [0,1] to control how much w_{ij} will change each iteration. If the value of η is too low, the ANN will take longer time to optimize its connection weight. However, if the value of η is too large, the ANN might *overshoot* the minimum error value just as illustrated in figure 2.5 below. In order to prevent *overshooting* and to save computation time, the value of learning rate should be optimized manually by researcher.

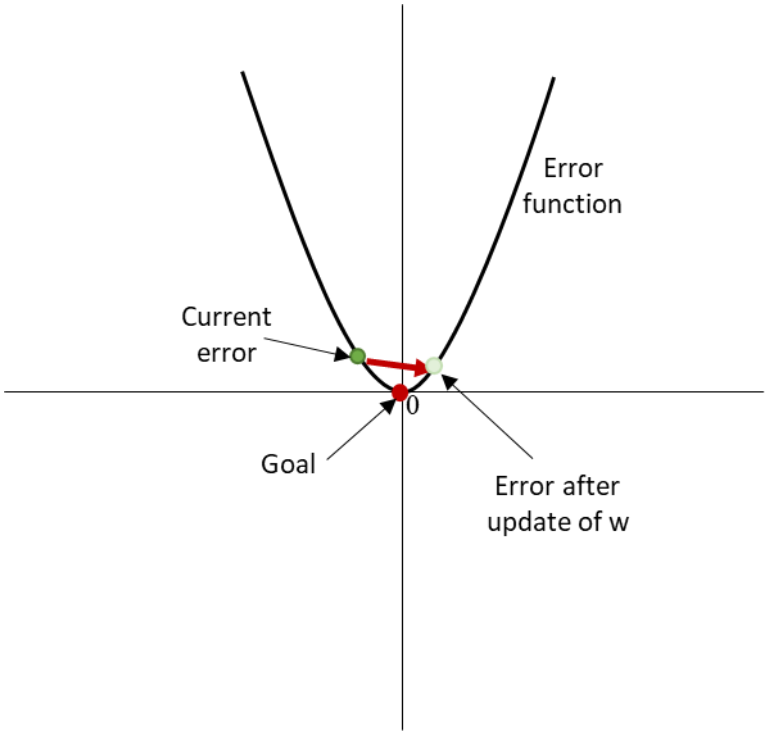


Figure 2. 6 Illustration of how SGD algorithm *overshoots* the minimum error because of large learning rate

2.3 Hyperparameter

Although ANN algorithm can learn from a dataset and adjust accordingly to make an accurate prediction, there are several parameters, also known as hyperparameter, that needs to be adjusted manually by researchers working closely with ANN algorithm. By adjusting these hyperparameters, the researcher can tune the computation speed as well as the desired accuracy of the ANN.

In this part, I will list and discuss how to adjust the hyperparameters of the ANN so the ANN can make accurate prediction.

2.3.1 Number of Hidden Neurons

Number of hidden neurons is related to the hidden layer of the ANN. In a three-layered feed-forward ANN algorithm, the hidden neurons refer to all neurons located in a layer between input layer and output layer.

The number of hidden neurons relates to the complexity of the ANN algorithm. By increasing the number of hidden neurons inside the hidden layer, we can enable the ANN to learn more complex data.

While putting more neurons inside the hidden layer can increase ANN capabilities, putting too many neurons inside the hidden layer might slow down the computation time. Moreover, putting too many neurons inside the hidden layer might make the ANN algorithm to overfit; a condition in which the prediction accuracy is good when predicting the training dataset but have low accuracy when predicting data outside the training dataset. Overfitting is an indicator that the ANN could not make generalized prediction and this condition must be prevented to make an accurate prediction.

To prevent this from happening, the number of hidden neurons must be adjusted manually by researcher. This can be done by training several ANNs with different number of hidden layers then pick the ANN with best accuracy.

2.3.2 Learning Rate

In SGD optimizer, learning rates is a hyperparameter which is influential in updating the connection weight of the ANN during the training phase.

As can be seen in equation 5, learning rate, which ranges from [0, 1.0] is a coefficient that influences how much the connection weight should change during each iteration during the training phase. Putting a large learning rate can make the training time faster, thus, shorts the computation time. However, this can also lower the accuracy of the ANN as can be seen in figure 2.6.

On the other hand, putting a very small value can improve the prediction accuracy. However, since the value is very small, the ANN training time might take longer. This means, the computation time will become slower.

The value of the learning rate can be manually adjusted by the researcher. However, during the adjustment phase, the researcher needs to consider the computation time and the accuracy of the ANN. Similar to number of hidden neurons, deciding the appropriate value for learning rate can be done training several ANNs with different learning rate then pick the learning rate that yields ANN with best performance.

Chapter 3. Prediction and Importance Analysis of Water Contact Angle (WCA)

3.1 The role of WCA in Biomedical Devices

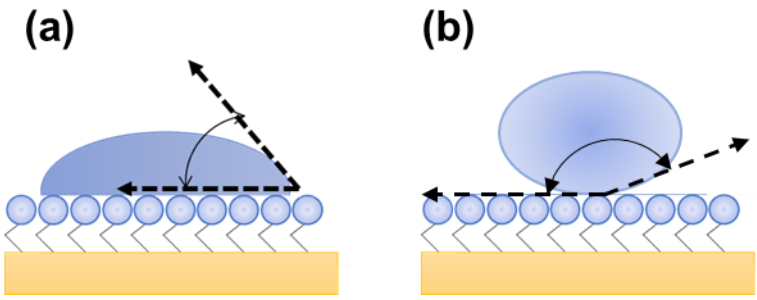


Figure 3. 1 WCA of (a) hydrophilic SAM and (b) hydrophobic SAM

WCA shows the degree of hydrophilicity of a surface and is one of the surface factors that must be considered when selecting or designing implant biomaterials. This is especially true when the biomaterials need to adhere firmly to the biological tissue. For example, in the case of dental implant, low WCA allows the implant to be quickly integrated to the surrounding hard tissue [7],[8] .Therefore, it is important to be able to design material with desired WCA. One of the methods to manipulate the WCA is by manipulating the surface polarity of the material.

Although empirical guide to design material with specific WCA exists, [10] the quantitative guide that can enable us to make an even more targeted material design remains unclear.

Fortunately, with so many research regarding WCA of materials have been conducted, the data of WCA of materials is already widely available, with the data about WCA of various SAMs is widely available. This makes WCA a good starting place to apply material informatics method with ANN algorithm.

In this part, we will discuss the possibility of data-driven approach using ANN algorithm in predicting WCA of SAMs as well calculate the importance of structure descriptors towards WCA that can serve as quantitative guide to design material with specific hydrophilicity. We will also demonstrate the application of ANN algorithm to predict WCA of SAMs using several selected SAMs outside of the dataset that is used to train the ANN algorithm.

3.2 Experimental

3.2.1 Data Gathering

A dataset of gold-based based SAM was created using data gathered from the past experiments conducted by Hayashi Laboratory and several publications [19-28] to train the ANN models. In this research, all SAMs were fabricated on a well-conditioned gold surface with rms roughness of around 0.3 nm.

In order to increase the data for ANN training, we used not only the WCA data of single-typed SAMs but also the data of mix-SAMs.

In the preprocessing method, we need to further clean the dataset that will be used for the ANN training. During this step, we should consider:

1. Missing data
2. Double input
3. The difference of WCA caused by difference in experimental method
4. Existence of outliers

ANN algorithm could not function well if there are missing data in the dataset. To prevent this, the missing data should be treated. For missing data, the database was searched in order to find if a value has not inputted. If any SAM without WCA data is found, then the data should be looked at once again in the corresponding paper that reports the specific SAM.

The double input problem should be addressed to prevent double training of ANN using same data, as well as to prevent bias in the dataset. For double input, the database should be checked so that the database does not contain same type of SAM. This problem could occur if there are two or more publications which reported the same type

of SAM. When this problem happened, the duplicates were deleted so that only one SAM of that specific type is left in the database.

Difference in experimental method could cause WCA of same SAM to be different when reported by different laboratories. This problem could affect the accuracy of the prediction of the ANN. For this problem, further observation of the dataset was conducted to find WCA reported using various experimental methods. The WCA data taken would be the WCA of SAMs that were experimented using experimental method that is used by majority of the data.

Outliers are unusual value in the dataset which might distort the statistical analysis. This distortion could affect the prediction results of ANN. For this reason, the outlier should be deleted from the dataset. According to statistics [Walpole], an outlier is defined by statistics to be any data point more than 1.5 interquartile ranges (IQRs) below the first quartile or above the third quartile. The data points that fulfills the definition were deleted from the dataset.

3.2.2 Selection of structure descriptors

To describe the molecular structure of the SAMs, several descriptors that describes the chemical structure of most of the SAMs in the dataset were selected as inputs for the ANN. These descriptors were generated by Dragon 7, a molecular structure calculation software. Since all the molecules in the datasets have a linear shape, the structure descriptors used were total number of atoms, the atomic composition and the number of covalent bonds constituting the terminal group (Table 3.1).

Table 3. 1 List of structure descriptors to describe chemical structure

structure descriptor	notation

Percentage of Carbon atom	%C
Percentage of Hydrogen atom	%H
Percentage of Oxygen atom	%O
Percentage of Nitrogen atom	%N
Total number of atoms	nAT
Total number of O-H bonds	O-H Bond
Total number of C-C bonds	C-C Bond
Total number of C-O bonds	C-O Bond
Total number of C-N bonds	C-N Bond
Total number of C=O bonds	C=O Bond

For single-type SAMs, these descriptors were used as it is without further modification. However, for mix-SAMs, by considering these factors:

1. SAMs in the dataset were linear in shape
2. Each of the mix-SAMs does not consist more than two components

the values of the descriptors were calculated according to the fraction of SAMs exists in the mix-SAM using formula below:

$$Descriptor_{tot} = \sigma_1 Descriptor_1 + \sigma_2 Descriptor_2 \quad \text{Eq. 6}$$

In which the σ_1 and σ_2 describes the fraction of the SAM 1 and SAM 2 of the mix-SAM respectively.

3.2.3 ANN Training

3.2.3.1 Optimizing the hyperparameter of ANN

All machine learning algorithm contains several hyperparameter that needs to be optimized to reduce the error of prediction. In ANN, the most important hyperparameter is number of hidden neurons inside the hidden layer. To optimize this hyperparameter, we used grid-search method. In this method, several combinations of learning rate and number of hidden neurons were tested. After the prediction results were validated using Leave-one-out Cross Validation (LOOCV), the combination of learning rate and hidden neuron with highest accuracy were selected as the hyperparameter.

To select the learning rate, by considering the computation time of the ANN, we tested combination of 20 learning rates which ranges from 0.001 up to 0.02 with 0.001 interval between each learning rates.

Moreover, for hidden layer, we tested the number of hidden neurons from 1 hidden neuron up to 15 hidden neurons.

To measure the accuracy of the ANN trained with specific number of hidden neurons, we used mean -squared-error (MSE). The value of MSE were calculated using formula below:

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(y_i - y'_i)^2 \qquad \text{Eq. 7}$$

in which n,y, and y'in both equations are the total number of data, the expected value, and the value as predicted by the ANN, respectively. Best accuracy will be given by the combination of hidden neurons and learning rate that yields minimum MSE.

3.2.3.3 Validation of the result of ANN prediction

To evaluate the prediction capability of ANN, the prediction results of WCA were tested using LOOCV[29] in which for each dataset containing n number of data, *n* ANNs were trained using (n-1) of the data. After the training, each trained ANN was used to predict the properties of the missing data.

The accuracy of each ANN prediction was then measured MSE and correlation coefficient (R). The MSE used for ANN evaluation is the same MSE value used to optimize the hyperparameter. As for the expression used to calculate R, the formula is shown below:

$$R = \frac{\sum_{i=1}^{n}[(y_i-\bar{y})(y'_i-\overline{y'})]}{\sqrt{\sum_{i=1}^{n}(y_i-\bar{y})^2(y'_i-\bar{y})^2}} \qquad \text{Eq. 8}$$

in which n,y, and y'in both equations are the total number of data, the expected value, and the value as predicted by the ANN, respectively. For MSE value, the smaller value of MSE would indicate the better ANN prediction quality. On the other hand, the value of R ranges between -1 to 1. The better the prediction quality, the closer the value will be to 1.

3.2.4 Importance Analysis

After the ANN is validated using LOOCV method, the importance of each descriptors was calculated to measure the effect and correlation of each descriptor towards WCA. Since the connection weight of each neurons in the ANN reflects the importance of structure descriptors, the importance of each descriptors would be calculated based on the connection weight of each neurons in the ANN using method proposed by Olden,et. al[30]

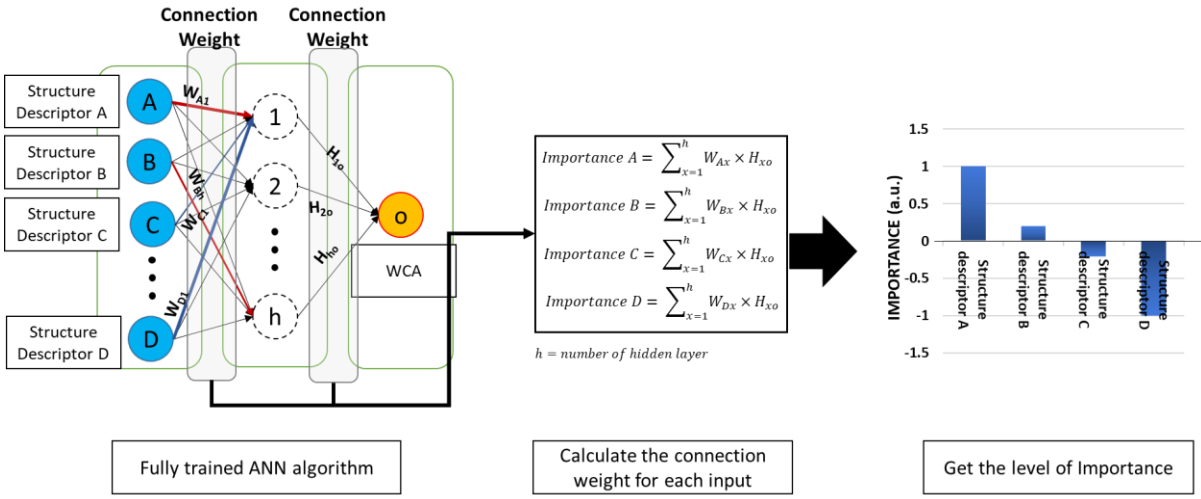


Figure 3. 2 shows the calculation of descriptors importance using ANN's connection weight.

After the ANN were trained by the database for WCA prediction, the connection weight of each descriptors' node and the nodes in the hidden layer as well as the connection weight of each node in the hidden layer and the output node were used to calculate the importance of each descriptors using the formula below:

$$Importance\ A = \sum_{x=1}^{h} W_{Ax} \times H_{xo} \qquad \text{Eq. 9}$$

With the W_{Ax} and H_{xo} representing the connection weight of descriptors node towards the hidden layer node and the hidden layer node towards the output node respectively.

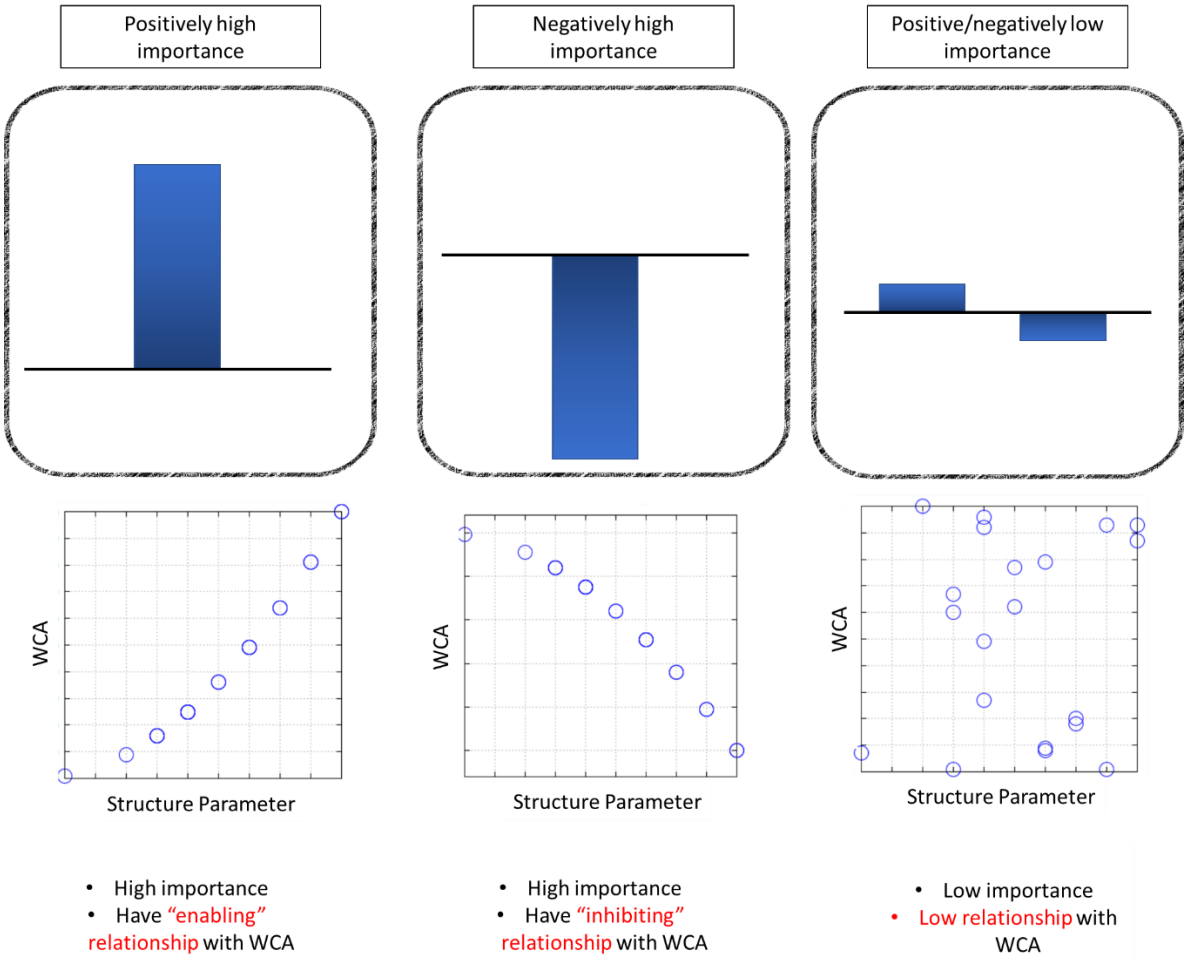


Figure 3. 3 Positive and Negatively important descriptors and their correlation towards WCA

The result of the importance analysis is explained in figure 3.3 above. There are three results after the calculation which are:

1. Positively high importance structure descriptors
2. Negatively high importance structure descriptors
3. Low importance structure descriptors

For the positively high and the negatively high importance, the structure descriptors have high degree of importance. However, if the positively high importance increases the WCA when the value is bigger, the negatively high importance decreases WCA

when the value is bigger. On the other hand, descriptors with low importance have low impact on the WCA.

The connection weight of the ANN after being trained by the material database might differ based on the initial guess of connection weight given by the machine learning algorithm random seed. In order to provide more stability to the importance value, using a method proposed by Oña, et. al. [31], 2000 initial guess of connection weight were tried and the final importance value of 2000 different ANNs trained with 2000 different initial connection weight were averaged to calculate the final importance value.

3.2.5 Prediction on Imaginary SAMs

After the validation using LOOCV on the training dataset, the ANN were used to predict WCA of several SAMs that does not exist in the training dataset. These SAMs are mentioned in the table below:

Table 3. 2 SAMs outside of training dataset

EG SAMs

EG-OH SAM
$HS-(CH_2)_8-O-C-C-OH$
$HS-(CH_2)_8-(O-C-C)_2-OH$
$HS-(CH_2)_8-(O-C-C)_3-OH$
$HS-(CH_2)_8-(O-C-C)_4-OH$

EG-Ome SAM
$HS-(CH_2)_8-O-C-C-O-CH_3$
$HS-(CH_2)_8-(O-C-C)_2-O-CH_3$
$HS-(CH_2)_8-(O-C-C)_3-O-CH_3$
$HS-(CH_2)_8-(O-C-C)_4-O-CH_3$

Non-EG SAMs

OH SAM
$HS-(CH_2)_9-OH$
$HS-(CH_2)_{10}-OH$
$HS-(CH_2)_{11}-OH$
$HS-(CH_2)_{12}-OH$

NH_2 SAM
$HS-(CH_2)_9-NH_2$
$HS-(CH_2)_{10}-NH_2$
$HS-(CH_2)_{11}-NH_2$
$HS-(CH_2)_{12}-NH_2$

CH_3 SAM
$HS-(CH_2)_7-CH_3$
$HS-(CH_2)_8-CH_3$
$HS-(CH_2)_9-CH_3$
$HS-(CH_2)_{10}-CH_3$

COOH SAM
$HS-(CH_2)_9-COOH$
$HS-(CH_2)_{10}-COOH$
$HS-(CH_2)_{11}-COOH$
$HS-(CH_2)_{12}-COOH$

The prediction results of these SAMs will be used as a basis to further measure the capability of ANN to be used as a tool for material design.

3.3 Result and Discussion

3.3.1 Data Gathering

A dataset of gold-based based SAM was created using data gathered from the past experiments conducted by Hayashi Laboratory and several publications [19-28] to train the ANN models. After the WCA data were gathered from these sources, we managed to gather 152 data which would go to the preprocessing step.

Among these 152 data, we observed the database to find the existence of missing data. Since every paper that were gathered reports data on WCA, we used every SAMs in those papers and made sure that the WCA dataset does not contain missing value. From this process, we retained all 152 data points.

Next, we considered the possibility that same SAMs were reported more than once in the database. We found out that several papers reported SAMs which were already reported in another paper (e.g. paper by Sascha, et. al reported WCA of EGOH SAM which was reported by Hayashi, et. al). We removed 7 SAMs in this process and retained 145 SAMs in our WCA dataset.

In the third step, we considered the difference of reported WCA due to difference in experimental method. Further observation at the paper shows that all papers use same method in measuring WCA but express their data in several ways. Several papers use either the value of advancing, receding, or both WCA. For most of the data, we picked the average value between advancing and receding WCA as the WCA value to be used in ANN training. Furthermore, for papers that only use either advancing or receding value, we used those value as it is for ANN training.

Lastly, we also addressed the existence of the outlier in the dataset. As previously mentioned, an outlier is defined by statistics to be any data point more than

1.5 IQRs below the first quartile or above the third quartile. Using this definition and the value in our dataset, we need to see if there are any data points that has WCA less than -55.25 degree or more than 182.75 degree. The first definition is theoretically impossible because the minimum WCA possible is 0 degree which means the outlier for WCA database is any WCA value more than 182.75 degree.

Further observation at the dataset shows that for WCA prediction do not have outlier in their WCA data and can be used to train the ANN.

After these preprocessing procedures were done, we have 145 data of WCA of SAMs for ANN training. The final dataset can be seen in the appendix section.

3.3.2 ANN Training

3.3.2.1 Hyperparameter optimization

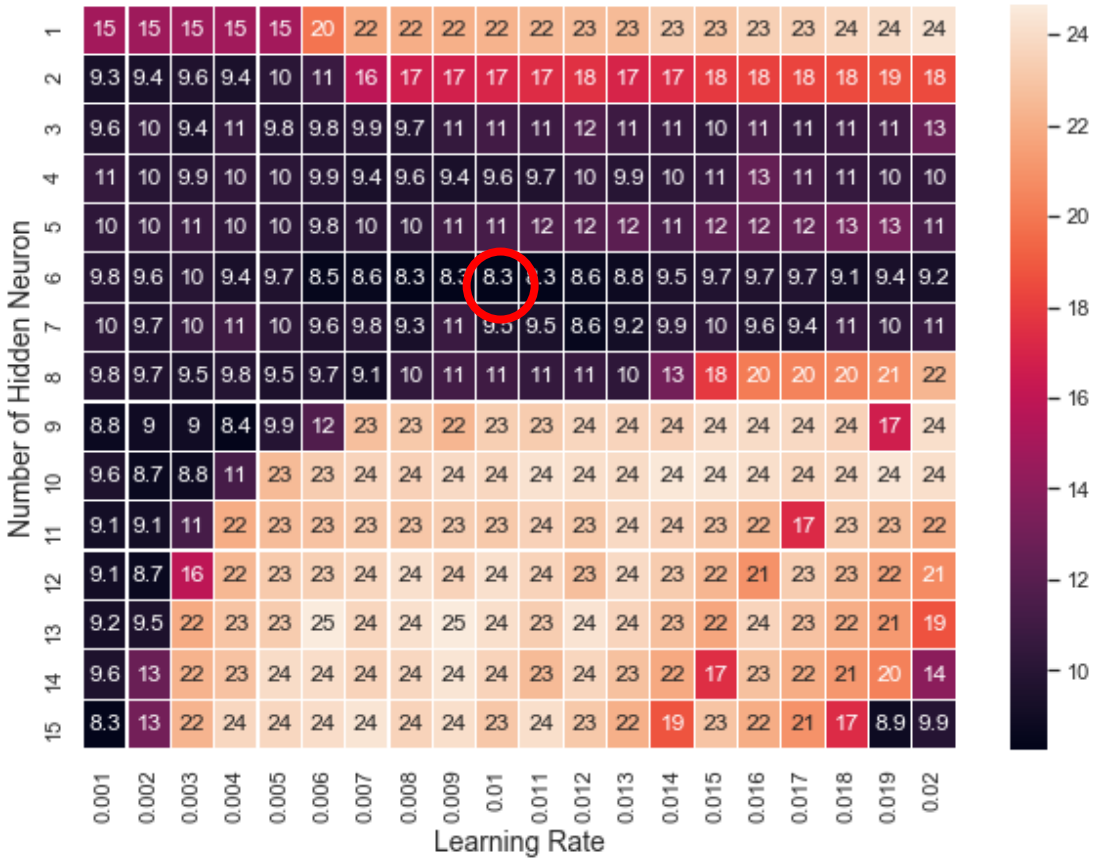


Figure 3. 4 Plot of MSE of ANN from several combination of hidden neuron and learning rate for WCA prediction

Figure 3.4 shows several combinations of learning rate and number of hidden neurons of ANN for prediction of WCA of SAM. The MSE were plotted as a heat-map with darker color indicates lower MSE and lighter color indicates higher MSE. We can also see the value of MSE for each combination of learning rate and number of hidden neurons as written in the heat-map.

By looking at the figure, we can see that 6 is the number of hidden neurons that allows ideal performance of the ANN. Among them we can see several combination of learning rates from 0.008 up to 0.011 in which the MSE does not change very much.

Between them, 0.01 is the best performing learning rates. Although the MSE in the figure were rounded, the actual values are 8.33, 8.28, and 8.31 for 0.009 learning rate, 0.01 learning rate, and 0.011 learning rate respectively. For this reason, we selected the 0.01 learning rate as the hyperparameter to be coupled with 6 hidden neurons.

For the rest of the study regarding WCA, the ANN with 6 hidden neurons and 0.01 learning rate will be used.

3.3.2.2 ANN's Validation

Figure 3.5 shows the validation results of the WCA predictions by ANN using LOOCV. Each dot in the figure represents a type of SAMs. The x axis and y axis represent the WCA and the predicted WCA of each SAM. The diagonal line in the middle shows where the dots would be placed if the prediction is completely accurate.

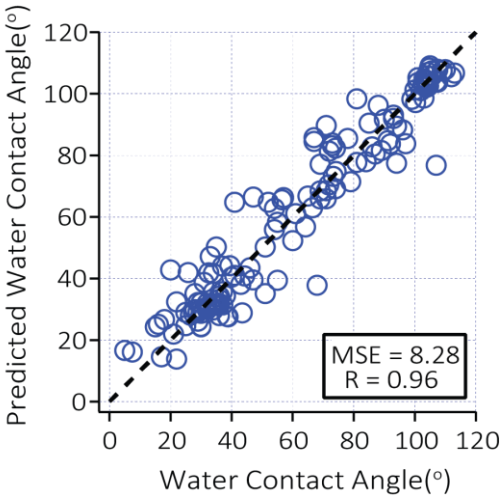


Figure 3. 5 WCA predicted by the trained ANN plotted against corresponding experimental values. A dashed line is a plot of y = x for eye-guide.

Results in figure 3.5 shows that most of the WCA of the SAMs are accurately predicted with most of the dots aligned to the center line. The accuracy of the prediction is further supported by the small MSE and the R value that is close to 1.

With the LOOCV results shows accurate prediction of the ANN, it is safe to say that the ANN can make accurate prediction of the WCA and is viable to be used for the following tasks:

1. Measuring importance of molecular descriptors
2. Predicting value outside the training dataset

Moreover, with WCA being a surface property that can be measured with high reproducibility and little effect from measurement conditions, we can use WCA data from any paper as a dataset to extend the applicable domain of the ANN.

3.3.3 Importance Analysis

Before the importance of each descriptors was calculated, it is important to see if 2000 different initial weight is statistically enough to describe the importance of each descriptors. To do this, we made a histogram of importance value for each descriptor.

Figure 3. 6 shows the histogram of importance value for each descriptor. By looking at the histogram, it is clearly visible that the ANN gives different importance value for different initial guess of connection weight. However, after using 2000 different initial guess, we started to see the distribution of importance value starts to make a shape similar standard distribution. These results show that 2000 samples are statistically enough to describe the importance of each descriptors. For that reason, the importance value used to describe the importance of molecular descriptors towards WCA would be the average value of importance of 2000 different initial guess of connection weight.

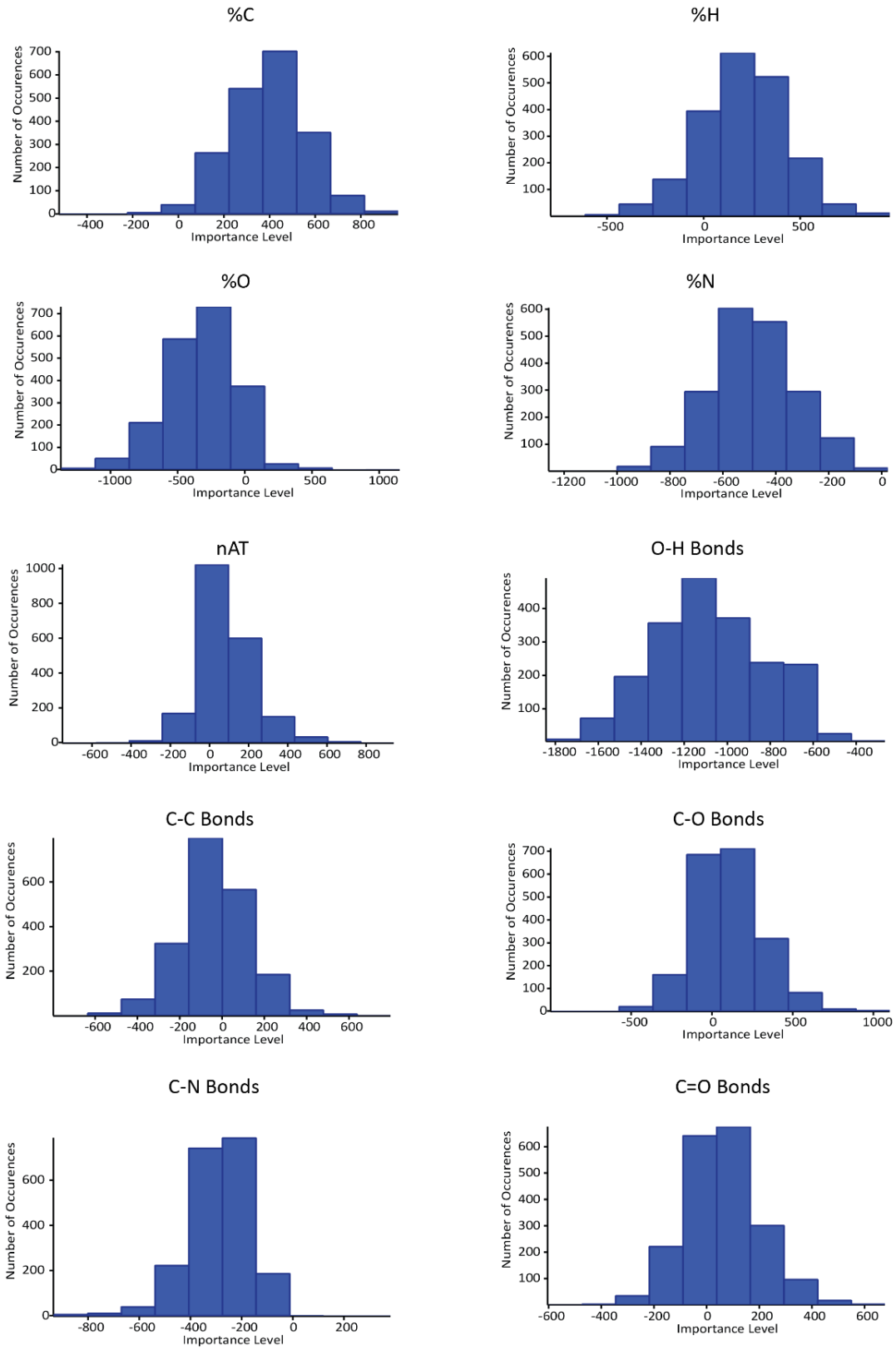


Figure 3. 6 Histogram of importance value of structure descriptors towards WCA across 2000 of initial guess of connection weight of ANN

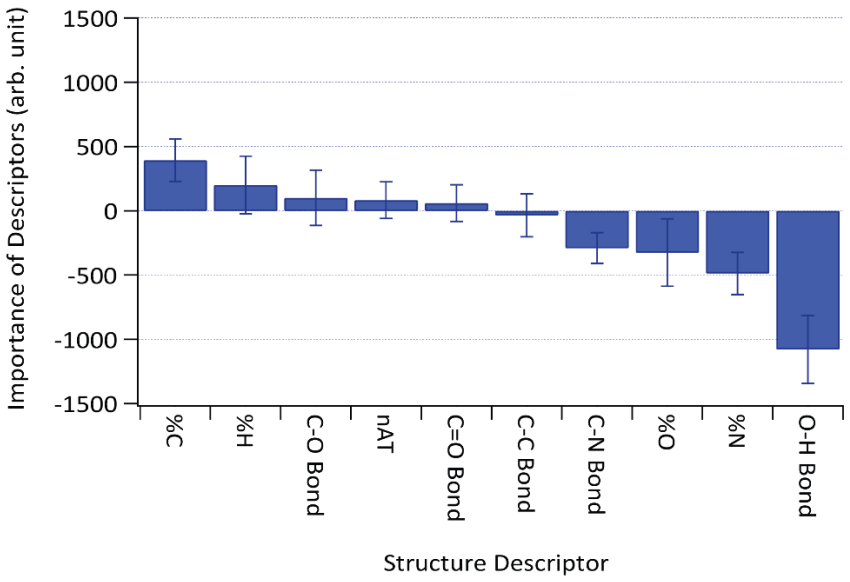


Figure 3. 7

Figure 3.7 show the degree of importance of each structure descriptor towards WCA. The structure descriptors were sorted from the positive importance towards the negative importance. For example, from the figure, it is clear that several structure descriptors such as O-H Bond and %N have negative importance with O-H bond has negatively high importance.

To describe the physical meaning behind these values, we made further observation in our WCA dataset. Further observation in our database, O-H bonds are mostly existing in hydroxyl terminated SAMs and nitrogen atoms are mostly existing in amide SAMs (-NH). Combining these knowledges, it is safe to infer that O-H bonds and %N represents the number of hydroxyl groups incorporated within SAM and number of amide group incorporated within SAM respectively.

By observing the characteristic of these functional groups, both hydroxyl group and amide group are polar functional group.

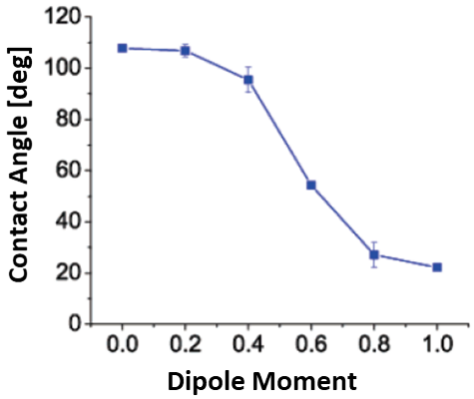


Figure 3. 8 Plot of dipole moment vs WCA as reported by Giovambattista, et. al [10]

Figure 3. 8 shows the plot of dipole moment and WCA as reported by Giovambattista, et. al (2017) [10]. According to this research, as dipole moment, which proportionally related to number of polar groups within a material, increases, the WCA will decrease. This result shows validates the negative importance of polar groups which includes hydroxyl group and amide group.

Furthermore, as previously mentioned in the experimental section, the magnitude of the value of the importance shows the degree of importance of each descriptor. From figure 3.7, it is visible that O-H Bond has importance value over 600 to the negative direction while C-C bond which represents number of alkyl chain has importance value less than 300 in the positive direction. This suggests that number of hydroxyl group is more important than number of alkyl chain in determining WCA of SAM. This result is also in agreement with results reported by Bain et al. (1989) and Folkers et al. (1992) who also found that there is a lack of correlation between the length of alkyl chain to WCA. [25, 27].

Since this data was an average data of importance of ANN with 2000 different initial connection weights, the importance of each structure descriptor also has standard

deviation which is represented by the error bar. The value of the error bar suggests that there is possibility that the value may change or fluctuate when the ANN is trained using different random initial weight. However, the fluctuation of the importance analysis is not so big and does not change the overall conclusion of this research.

3.3.5 Prediction on Imaginary SAMs

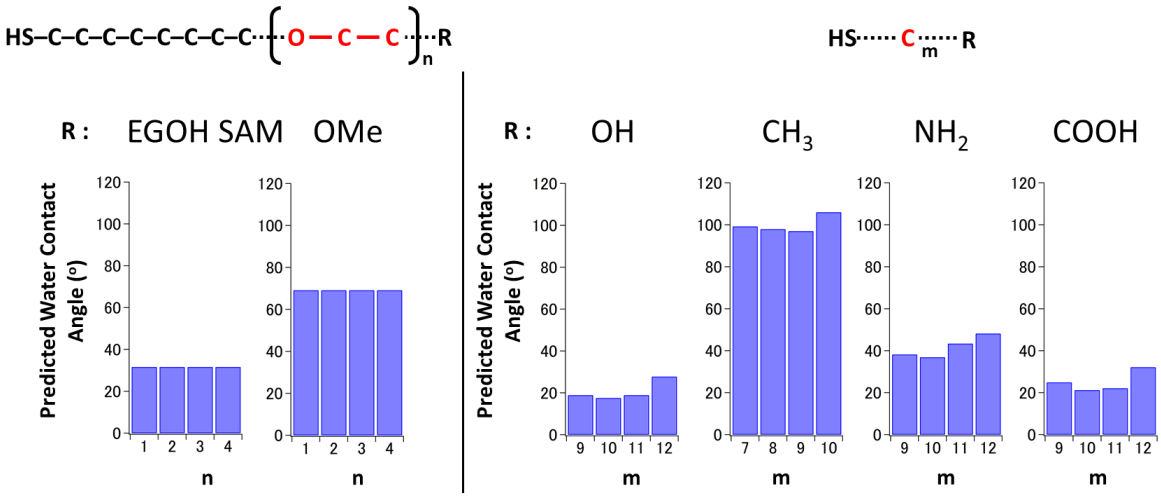


Figure 3. 9 Prediction of WCA for SAMs outside of training dataset. For EGOH and OMe SAMs, the EG units were changed while for OH, CH_3, NH_2, and COOH SAMs, the number of carbon atoms in their alkyl chains are changed.

Figure 3.9 above showed the prediction of WCA of several SAMs with different functional group, chain length, and EG-group length which were not included in the original dataset to train the ANN.

From these results, it can be clearly seen that neither the change in alkyl chain nor the change in EG-chain affects the WCA of SAMs. There are some slight changes such as in the case of CH_3 SAMs and NH_2 SAMs. However, these changes are not regular occurrences.

These prediction results combined with the previously discussed validation results of the ANN proved the capabilities of the ANN to predict WCA even for SAM that is outside of the training dataset.

Chapter 4. Prediction and Importance Analysis of Fibrinogen Adsorption

4.1 Fibrinogen Adsorption on Biomedical Devices

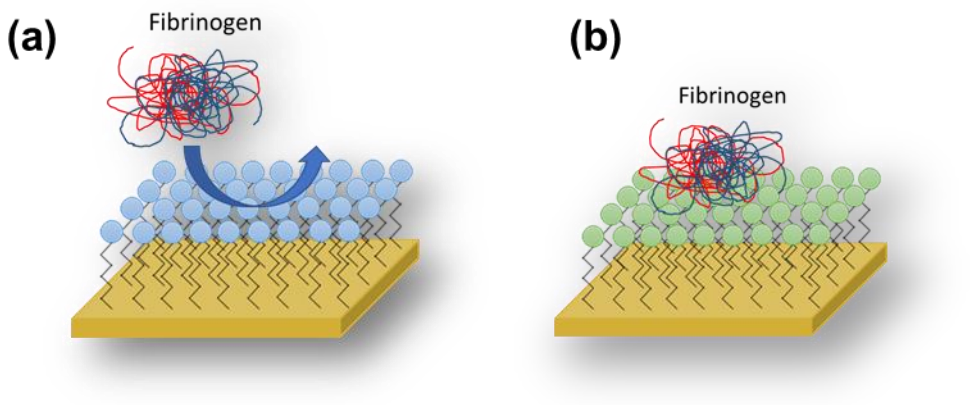


Figure 4. 1 Fibrinogen on top of SAM. By changing the structure of the material, it is possible to make SAM that is repelling (a) or adsorbing (b) fibrinogen

Fibrinogen adsorption is one of the most important when designing biomaterial for medical application especially because of its correlation with platelet adhesion and blood clotting. Several biomedical devices such as artificial blood vessels require low degree of fibrinogen adsorption to prevent blood coagulation and improve blood compatibility. [9]

However, designing material with specific fibrinogen adsorption remains challenging. Although an effort to make criteria of protein resistant materials has been conducted [11], the criteria is largely empirical rather than quantitative. Moreover, an effort to model protein adhesion on the top of material requires modelling a very complex protein which is computationally expensive.

With so many experimental data of fibrinogen adsorption using various SAMs as studying platform being widely available and with ANN algorithm being successful in predicting as well as making quantified criteria for WCA of material, it is supposedly possible to apply the same thing to design material with specific fibrinogen adsorption. However, in contrast with WCA, the data of fibrinogen adsorption could have big discrepancy with regards to the experimental method used to obtain the data. The question remains whether this discrepancy can pose a problem in applying data-driven method in material science. In this chapter, we will discuss the possibility of applying data-driven method using ANN algorithm to predict as well as making quantified criteria to design material with specific fibrinogen adsorption.

4.2 Experimental

4.2.1 Database Gathering

For fibrinogen adsorption prediction, dataset of gold-based based SAM was created using data gathered from the past experiments conducted by Hayashi Laboratory and several publications [19-21, 24] to train the ANN models. In this research, all SAMs were fabricated on a well-conditioned gold surface with rms roughness of around 0.3 nm.

To study the effect of the difference of experimental methods on ANN predictions of fibrinogen adsorption, we created two different datasets. One dataset contains data from a single laboratory with standardized measurement condition [19, 21] (henceforth referred to as single-lab dataset) while the other contains data from various laboratories with different measurement conditions (henceforth referred to as multi-lab dataset).

In order to increase the data for ANN training, we used not only the fibrinogen adsorption data of single-typed SAMs but also the data of mix-SAMs.

In the preprocessing step, the dataset is cleaned to make sure that ANN training can be done smoothly. During this step, several points need to be considered:

1. Missing data
2. Double input
3. Existence of outliers

Although in the WCA prediction we included the difference of experimental method as one of the criteria to clean the dataset, since in this section we want to see the effect of the difference of experimental method in multi-lab dataset and the data in

single-lab dataset were already taken under the same experimental condition, we did not consider difference of experimental method as a criteria to clean the dataset.

Just like with the prediction of WCA, the ANN algorithm could not function well if the dataset contains missing data which means that the missing data needs to be treated. To solve the problem with missing data, any SAMs with missing data is searched once more in the corresponding papers and if the paper does not contain the fibrinogen adsorption data of the specified SAM, the data will be omitted from the dataset.

To prevent bias in the dataset as well as to prevent double training of the ANN using same data, the dataset should not have more than 1 data for a specific SAM. However, since a specific SAM might be researched by more than one researcher, there are possibilities that the same type of SAM is included twice in the dataset. When this problem occurs, the duplicates were deleted so that only SAM of that specific type is left in the database.

Lastly, we also considered the existence of outliers in the dataset. We used the same definition of outliers used in the preprocessing of the dataset from WCA prediction which is any data point more than 1.5 IQRs below the first quartile or above the third quartile. The data points that fulfills the definition were omitted from the dataset.

4.2.2 Selection of structure descriptors

To describe the molecular structure of the SAMs, several descriptors that describes the chemical structure of most of the SAMs in the dataset were selected as inputs for the ANN. These descriptors were generated by Dragon 7, a molecular structure calculation software. Since all the molecules in the datasets have a linear shape, the structure

descriptors used were total number of atoms, the atomic composition and the number of covalent bonds constituting the terminal group (Table 4.1).

Table 4. 1 List of structure descriptors to describe chemical structure

structure descriptor	**notation**
Percentage of Carbon atom	%C
Percentage of Hydrogen atom	%H
Percentage of Oxygen atom	%O
Percentage of Nitrogen atom	%N
Total number of atoms	nAT
Total number of O-H bonds	O-H Bond
Total number of C-C bonds	C-C Bond
Total number of C-O bonds	C-O Bond
Total number of C-N bonds	C-N Bond
Total number of C=O bonds	C=O Bond

For single-type SAMs, these descriptors were used as it is without further modification. However, for mix-SAMs, by considering these factors:

1. SAMs in the dataset were linear in shape
2. Each of the mix-SAMs does not consist more than two components

the values of the descriptors were calculated according to the fraction of SAMs exists in the mix-SAM using formula below described in Eq. 6.

4.2.3 ANN Training

4.2.3.1 Optimizing the hyperparameter of ANN

Just like the study in WCA prediction, there are several hyperparameter that needs to be optimized to reduce the error of prediction. Since we are using ANN to predict fibrinogen adsorption, we need to optimize the number hidden neurons inside the hidden layer as well as the learning rates of the ANN. To optimize these hyperparameter, we used grid-search method once more

To select learning rate, by considering the computation time of the ANN, we tested combination of 20 learning rates which ranges from 0.01 up to 0.3 with 0.01 interval between each learning rates.

As for the hidden layer, the number of hidden neurons from 1 hidden neuron up to 15 hidden neurons were tested.

To measure the accuracy of the ANN trained with specific number of hidden neurons, we used mean -squared-error (MSE). The value of MSE were calculated using formula described in Eq. 7.

4.2.3.2 Validation of the result of ANN prediction

To evaluate the prediction capability of ANN, the prediction results of WCA were tested using LOOCV,[29] in which for each dataset containing n number of data, *n* ANNs were trained using (n-1) of the data. After the training, each trained ANN was used to predict the properties of the missing data.

The accuracy of each ANN prediction was then measured MSE and R value as used in the WCA prediction. The MSE used for ANN evaluation is the same MSE value used to optimize the hyperparameter.

4.2.4 Importance Analysis

After both the ANN trained by single-lab and multi-lab dataset are validated using LOOCV method, the importance of each descriptors was calculated to measure the effect and correlation of each descriptor towards fibrinogen adsorption. Since the connection weight of each neurons in the ANN reflects the importance of structure descriptors, the importance of each descriptors would be calculated based on the connection weight of each neurons in the ANN using method proposed by Olden,et. al[30]

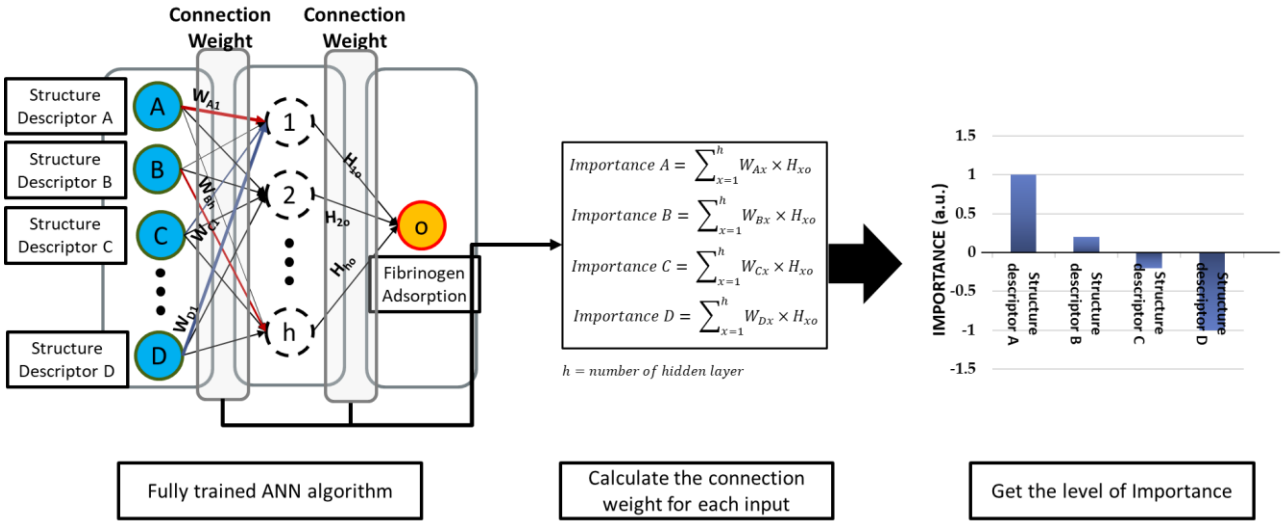


Figure 4. 2 Calculating the importance of descriptors towards fibrinogen adsorption using connection weight of ANN

Figure 4.2 shows the calculation of descriptors importance using ANN's connection weight. Similar to the WCA prediction, after the ANN were trained by the database to predict fibrinogen adsorption using both the multi-lab and single-lab dataset, the connection weight of each descriptors' node and the nodes in the hidden layer as well as the connection weight of each node in the hidden layer and the output node were

used to calculate the importance of each descriptors using the formula previously described in Eq. 9.

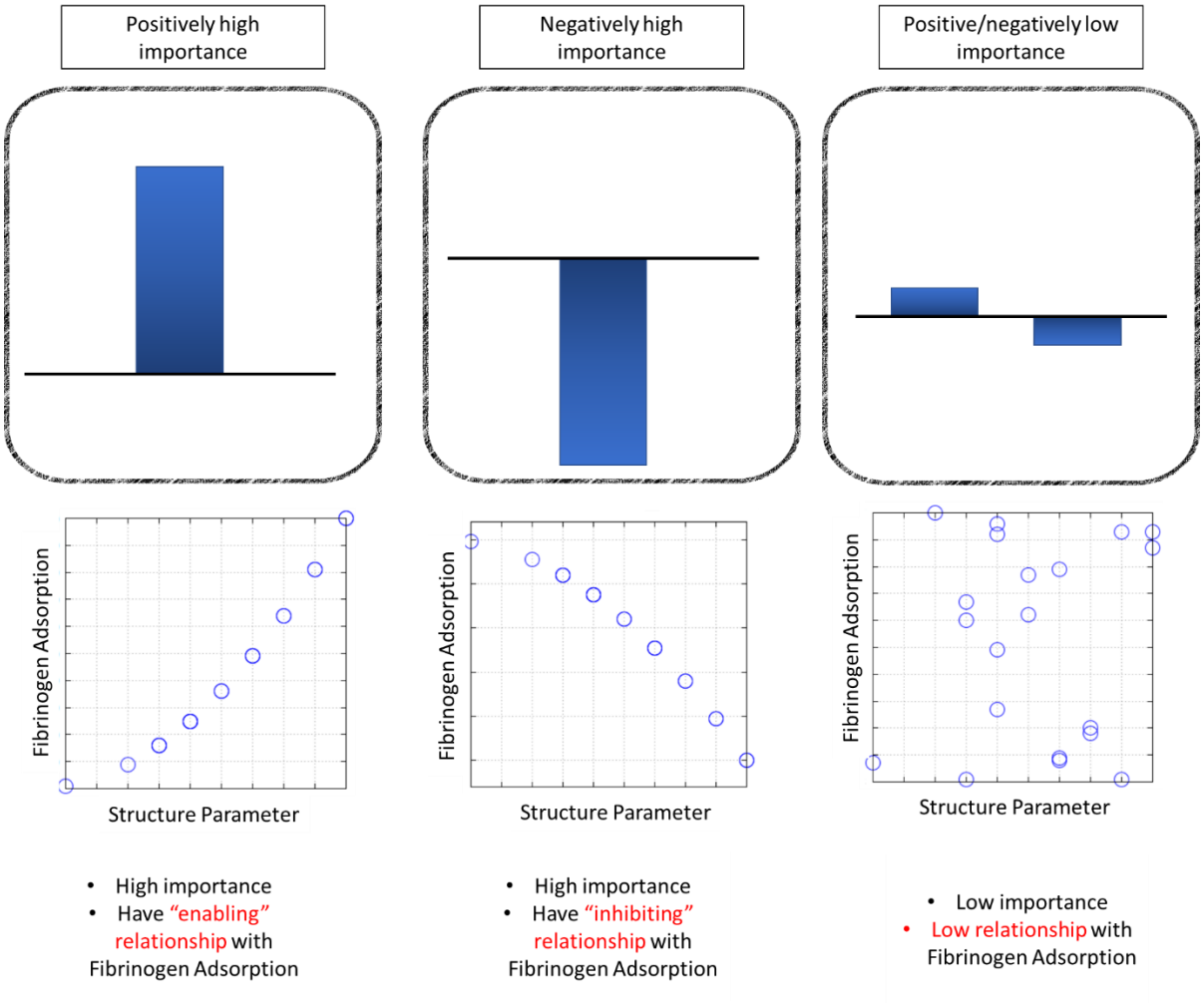


Figure 4. 3 Positive and Negatively important descriptors and their correlation towards Fibrinogen Adsorption

The result of the importance analysis is explained in figure 4.3 above. There are three results after the calculation which are:

1. Positively high importance structure descriptors
2. Negatively high importance structure descriptors
3. Low importance structure descriptors

For the positively high and the negatively high importance, the structure descriptors have high degree of importance. However, if the positively high importance increases the fibrinogen adsorption when the value is bigger, the negatively high importance decreases WCA when the value is bigger. On the other hand, structure descriptors with low importance has low impact on the fibrinogen adsorption.

Just like the importance analysis of WCA, the connection weight of the ANN after being trained by the material database might differ based on the initial guess of connection weight given by the machine learning algorithm random seed. In order to provide more stability to the importance value, using a method proposed by Oña, et. al. [31], 2000 initial guess of connection weight were tried and the final importance value of 2000 different ANNs trained with 2000 different initial connection weight were averaged to calculate the final importance value.

Here, we will also discuss the effects of different experimental method on the importance analysis result by comparing the importance analysis result of multi-lab and single-lab dataset.

4.2.5 Extrapolation prediction of SAMs

After the validation using LOOCV on the training dataset, the ANN were used to predict fibrinogen adsorption of several SAMs that does not exist in the training dataset. These SAMs are mentioned in the table below:

Table 4. 2 SAMs outside of training dataset

EG SAMs

EG-OH SAM
HS-$(CH_2)_8$-O-C-C-OH
HS-$(CH_2)_8$-$(O\text{-}C\text{-}C)_2$-OH
HS-$(CH_2)_8$-$(O\text{-}C\text{-}C)_3$-OH
HS-$(CH_2)_8$-$(O\text{-}C\text{-}C)_4$-OH

EG-Ome SAM
HS-$(CH_2)_8$-O-C-C-O-CH_3
HS-$(CH_2)_8$-$(O\text{-}C\text{-}C)_2$-O-CH_3
HS-$(CH_2)_8$-$(O\text{-}C\text{-}C)_3$-O-CH_3
HS-$(CH_2)_8$-$(O\text{-}C\text{-}C)_4$-O-CH_3

Non-EG SAMs

OH SAM
HS-$(CH_2)_9$-OH
HS-$(CH_2)_{10}$-OH
HS-$(CH_2)_{11}$-OH
HS-$(CH_2)_{12}$-OH

NH_2 SAM
HS-$(CH_2)_9$-NH_2
HS-$(CH_2)_{10}$-NH_2
HS-$(CH_2)_{11}$-NH_2
HS-$(CH_2)_{12}$-NH_2

CH_3 SAM
HS-$(CH_2)_7$-CH_3
HS-$(CH_2)_8$-CH_3
HS-$(CH_2)_9$-CH_3
HS-$(CH_2)_{10}$-CH_3

COOH SAM
HS-$(CH_2)_9$-COOH
HS-$(CH_2)_{10}$-COOH
HS-$(CH_2)_{11}$-COOH
HS-$(CH_2)_{12}$-COOH

The ANN that will be used to predict these ANNs were picked based on the result of the validation. If the validation results of the ANN were bad, then the ANN will not be used for extrapolation prediction.

The prediction results of these SAMs will be used as a basis to further measure the capability of ANN to be used as a tool for material design.

4.3 Result and Discussion

4.3.1 Data Gathering

To understand the effect of difference of measurement condition across different laboratories on prediction quality, we made 2 dataset which were single-lab dataset and multi-lab dataset. By including data of mixed-SAMs on which incorporated by so many papers included in single-lab dataset and multi-lab dataset, we gathered 72 and 136 data for single- and multi-lab dataset for fibrinogen adsorption respectively. These numbers of data were used in the later step which is the preprocessing step.

During the first step which is cleaning the data from missing data, the missing value on both the single-lab and multi-lab dataset were checked and needs to be either filled immediately if the fibrinogen adsorption value for the specified SAMs were found in the corresponding paper or omitted if the fibrinogen adsorption value could not be found. Checking at the dataset, both the dataset for single-lab and multi-lab dataset do not contain missing value. Thus, all of the datapoints were retained and the single-lab dataset and multi-lab dataset has 72 data and 136 data respectively.

Next, the data is also cleaned from data that were inputted more than once. Here, although duplicates were not found in single-lab dataset, there are several double inputs that were found in multi-lab dataset.

The paper that were used as source of data in multi-lab dataset, specifically, data reported by Herrwerth, et. al. and Hayashi, et. al. contains several SAMs that were researched in both laboratories. Among them, we found three identical SAMs which were reported by these papers which were: EG3-OMe, EG3-OH, and EG6-OH. Since the value of the fibrinogen adsorption reported by both of these papers are also identical,

we opted to omit the duplicates of these SAMs. After this step, we have 72 data and 133 data for single-lab dataset and multi-lab dataset respectively.

Lastly, we also addressed the existence of the outlier in the dataset. As previously mentioned, an outlier is defined by statistics to be any data point more than 1.5 IQRs below the first quartile or above the third quartile. Using this definition and the value in our dataset, we need to see if there are any data points that has fibrinogen adsorption less than -2.07 microgram/cm2 or more than 3.68 microgram/cm2 and less than -2.72 microgram/cm2 or more than 4.77 microgram/cm2 for mixed-lab dataset and single-lab dataset respectively.

Further observation at the dataset shows that for mixed-lab dataset and single-lab dataset do not have outlier in their fibrinogen adsorption data and can be used to train the ANN.

After the dataset is being cleaned, the numbers of data in the single- and multi-lab datasets for fibrinogen adsorption are 72 and 133, respectively. The final dataset can be seen in the appendix section.

4.3.2 ANN Training

4.3.2.1 Hyperparameter Optimization

Using grid-search method to tune ANN parameter for fibrinogen adsorption of multi-lab dataset and single-lab dataset, we managed to have the combination of number of hidden neurons and learning rate for both the multi-lab dataset and single-lab dataset.

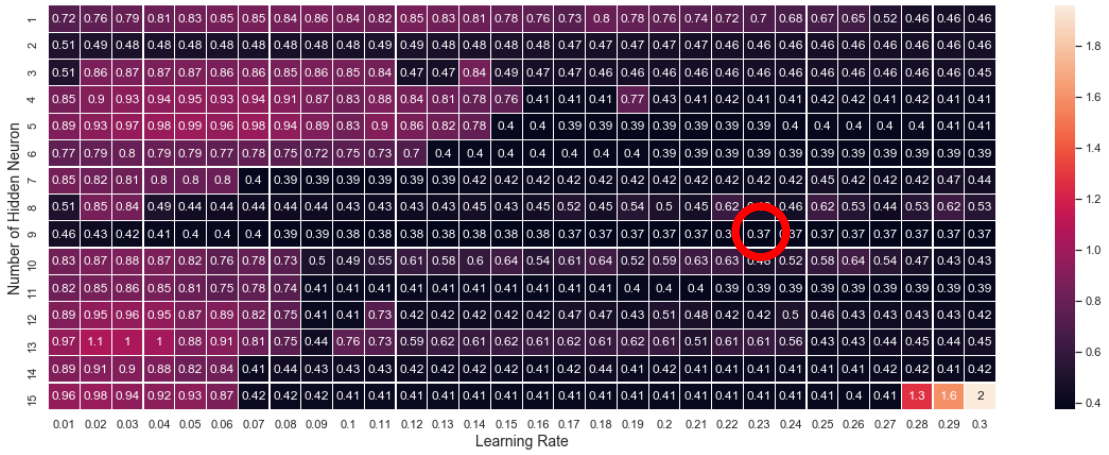


Figure 4. 4 Plot of MSE of ANN from several combination of hidden neuron and learning rate for fibrinogen adsorption prediction of multi-lab data

Figure 4.4 shows several combinations of learning rate and number of hidden neurons of ANN for prediction of fibrinogen of SAM in multi-lab dataset. The MSE were plotted as a heat-map with darker color indicates lower MSE and lighter color indicates higher MSE. We can also see the value of MSE for each combination of learning rate and number of hidden neurons as written in the heat-map.

By looking at the figure, we can see that 9 is the number of hidden neurons that allows ideal performance of the ANN. Among them we can that most of the learning rates used in combination with 9 hidden neurons does not have major effects on the MSE. However, among them, 0.23 is the best performing learning rates when the MSE is compared with precision of 5 significant digits. Although the difference is negligible, since the computation time does not change much between these learning rates, 0.23 were picked as the learning rates.

For the rest of the study regarding fibrinogen adsorption of multi-lab dataset, the ANN with 9 hidden neurons and 0.23 learning rate will be used as the prediction model.

Regarding the fibrinogen adsorption of single-lab dataset, the MSE plot can be seen in figure 4.5 below.

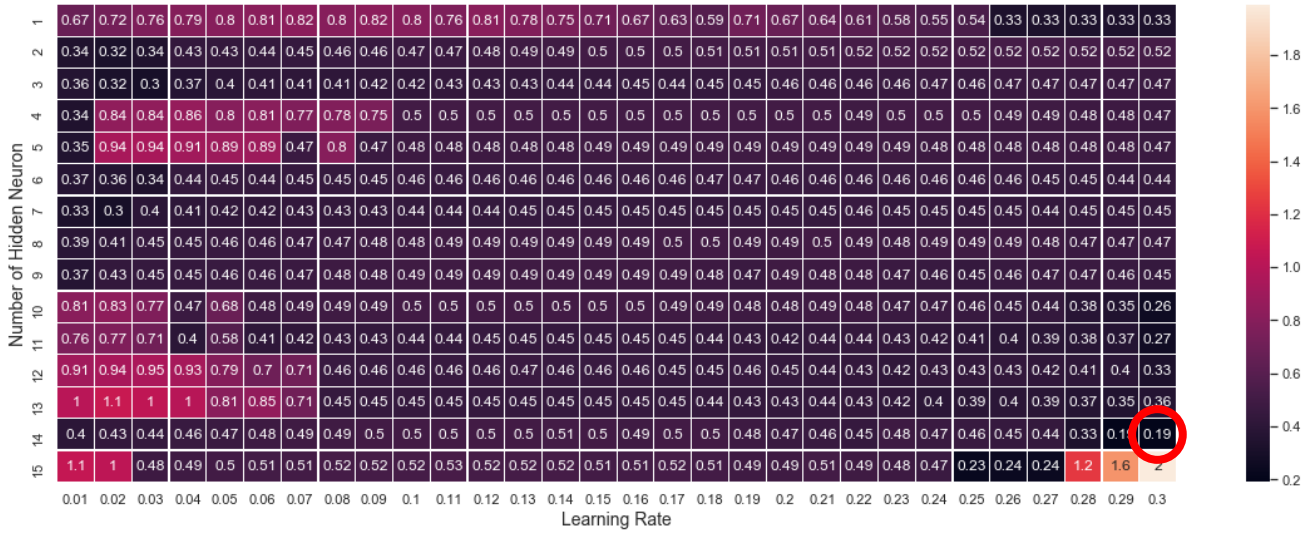


Figure 4. 5 Plot of MSE of ANN from several combination of hidden neuron and learning rate for fibrinogen adsorption prediction of single-lab data

By looking at the figure, there are multiple hidden neurons that shows similar quality with other hidden neurons. For example, using 6 hidden neurons up to 9 hidden neurons does not change the quality of the prediction very much. Thus, it is necessary for us to also look at the learning rate.

Looking at the rightmost part of the table, it is visible that 0.3 learning rate gives mostly darker color which indicates low MSE. From this, it is safe to assume that 0.3 is the ideal learning rate for this model.

Among the 0.3 learning rate, we now look at the number of hidden neurons that gives the best prediction performance. After carful observation, it is found out that 14 hidden neurons is the best combination with 0.3 learning rate which, among the other 450 combination of learning rate and number of hidden neuron, give the minimum MSE of 0.19.

Thus, for the rest of the study regarding fibrinogen adsorption of single-lab dataset, the ANN with 14 hidden neurons and 0.30 learning rate will be used.

4.3.2.2 Validation of ANN

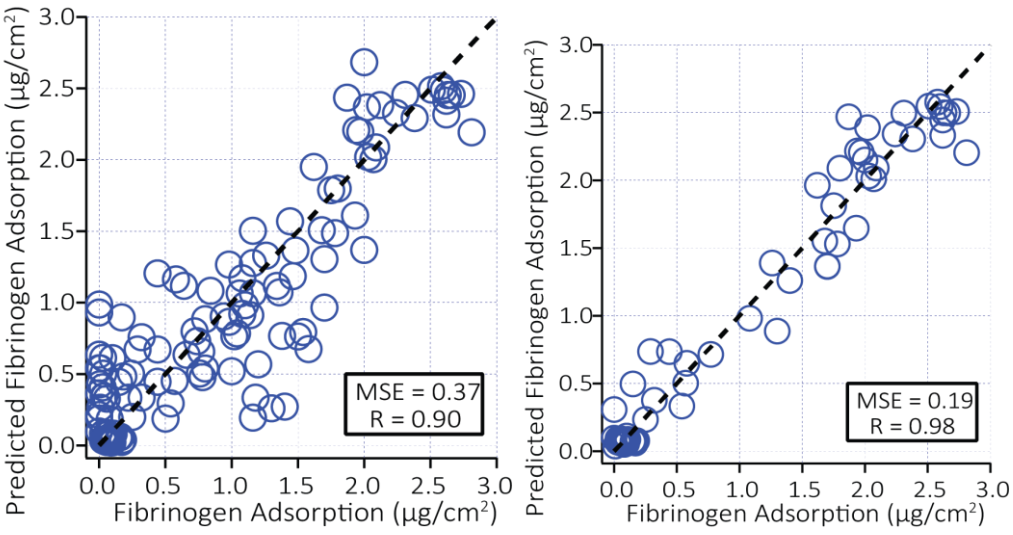


Figure 4. 6 LOOCV results of amounts of adsorbed fibrinogen predicted by ANN. The ANNs were trained with (a) the multi-lab and (b) single-lab datasets. Dashed lines are plots of $y = x$ for eye-guide

Figure 4.6 shows the validation results of ANN trained by multi-lab and single-lab dataset. For the ANN trained by multi-lab dataset, the MSE and R value are 0.37

and 0.9 respectively and for the ANN trained by single-lab dataset, the MSE and R value are 0.19 and 0.98 respectively.

Both the MSE and R value of these ANN suggested that the accuracy of the ANN trained by single-lab dataset is higher than the ANN trained by multi-lab dataset.

A possible explanation for this difference in accuracy could be that the different measurement condition employed to measure the adsorption of fibrinogen across different laboratories makes it very difficult to make a dataset of fibrinogen of SAM with big reproducibility. There are several difference of measurement conditions which will be explained. The difference of measurement condition includes:

1. Measurement equipment (QCM, SPR, etc.)
2. The expression of data
3. Treatment of SAMs before the measurement

The measurement equipment relates to the equipment used to measure the amount of fibrinogen adsorbed into the surface of the SAMs. There are several measuring equipment that can be used to measure the amount of fibrinogen adsorbed into the surface of the SAMs which the calibration of the equipment and how the equipment works might affect the data. For example, Whitesides, et. al and Herrwerth, et. al [11, 20] measure the fibrinogen adsorption using surface plasmon resonance (SPR) while Hayashi, et.al and Sekine, et. al measure the fibrinogen adsorption using Quartz Crystal Microbalance with Dissipation monitoring (QCM-D). Although both measures the fibrinogen thickness, they have different method in which SPR measures using optical method and QCM-D measures the thickness using acoustic method in which both will have several consequences in theoretical setup and hardware setups.

From theoretical perspective, there are different range of thickness of fibrinogen adsorption that can be measured accurately for both methods. Since the data reported by Whitesides and Herrwerth are expressed as fibrinogen adsorption relative to C6 SAMs, it is very difficult to verify the thickness of the SAMs used in their findings. However, there is possibility that the accuracy of the result is difficult to be compared when being put into the same dataset with other laboratory such as data reported by Hayashi and Sekine which shows the exact value of fibrinogen adsorption using QCM-D equipment. Moreover, both SPR and QCM-D has experimental capabilities and hardware limitation. For example, SPR generally use smaller sample compared with QCM-D which might also affect the calculation results. Other than this, there might be more underlying problem such as hardware and equipment settings that makes it impossible for both of this equipment to show identical results. These underlying problems caused several measurement errors in the fibrinogen adsorption data which affects the predictions of ANN. These errors caused by measuring equipment does not happen when the data comes from single laboratory in which the equipment used is the same and the calibration technique is standardized.

Moreover, in relation to the previous point, different measurement equipment caused different laboratories to use different expressions to report their data. For example, Whitesides, et. al and Herrwerth, et. al reported their data relative to a specific SAM such as Undecanethiol ($C_{11}H_{24}S$) and Hexanedecanethiol ($C_{16}H_{34}S$) respectively. On the other hand, Hayashi, et. al and Sekine, et. al reported their fibrinogen adsorption data based on exact density of protein adsorbed in the vicinity of SAMs. The problem arise because for the data to be able to be used to train ANN algorithm, the fibrinogen adsorption data must be expressed in the same way which means the data from different

laboratories must be converted so that the expression is the same across every data. However, there might be some information lost during the conversion process. Moreover, since Whitesides, et. al and Herrwerth, et. al uses different SAMs as their relative point, it becomes increasingly difficult to convert the value of their fibrinogen adsorption. These problems of expression resulted in difficulty in training of ANN algorithm which causes the error of LOOCV prediction to be high. On the contrary, since the expression of data used by Hayashi, et. al and Sekine, et. al are standardized, these problems did not occur in the prediction using single-lab dataset.

Lastly, the treatment of SAMs might be different across different laboratories. This treatment of SAMs could results in different data although the data were taken using same measurement equipment. For example, Whitesides, et. al reported two different fibrinogen adsorptions after the SAMs were exposed to the solution of protein for 3 minutes and 30 minutes. On the other hand, Herrwerth, et. al reported the fibrinogen adsorptions of SAMs after the SAMs were exposed to the same solution of protein for 15 minutes. In data reported by Whitesides, et. al, we can see that that the amount of fibrinogen adsorbed by SAMs will change depends on the time of exposure. Thus, it is very possible that this different of exposure time employed by Whitesides and Herrwerth might affect the data used for ANN training. On the other hand, data reported by Hayashi, et. al. and Sekine, et. al were taken with same treatment of SAMs which results in better prediction for single-lab dataset.

On the other hand, when the ANN is trained by single-lab dataset which contains data that are taken under standardized condition, the validation results by the ANN shows high accuracy. The similar results were obtained when validating WCA of SAMs

in which ANN could make prediction of WCA with high accuracy when trained using WCA dataset that has less different in measurement conditions.

These results show that it is important for the data used to train the ANN to have taken under standardized measurement condition. In case of fibrinogen adsorption, the data of fibrinogen adsorption must have standardized measurement condition. There are several ways to take a data with standardized condition such as:

1. Using data taken from the same laboratory
2. Using data taken from different laboratory that has the same or similar measurement condition

4.3.3 Importance Analysis

Just like with importance analysis on WCA dataset, it is important to see if 2000 different initial weight is statistically enough to describe the importance of each descriptors. To see if the data is statistically enough, we made a histogram of importance value for each descriptor.

Figure 4.7 and 4.8 below show the histogram of importance value for each descriptor for the multi-lab dataset and the single-lab dataset respectively. By looking at both sets of histograms, it is visible that the ANN gives different importance value for different initial guess of connection weight. However, by using 2000 initial guesses, the distribution of importance value starts to make shapes similar to standard distribution which shows that 2000 samples are statistically enough to describe the importance of each descriptors. There are several exceptions such as importance distribution of %O and CN Bond for multi-lab dataset as well as the importance distribution of C=O Bond for single-lab dataset which shows that the distribution is

bimodal or shows two different peaks. However, in those cases, the second peak is small compared to the first peak which means it can be neglected. In conclusion, after using 2000 samples of initial guesses, the samples are statistically enough to describe the importance. Hence, the importance value used for each descriptor would be the average value of these 2000 importance values for both the multi-lab and single-lab dataset.

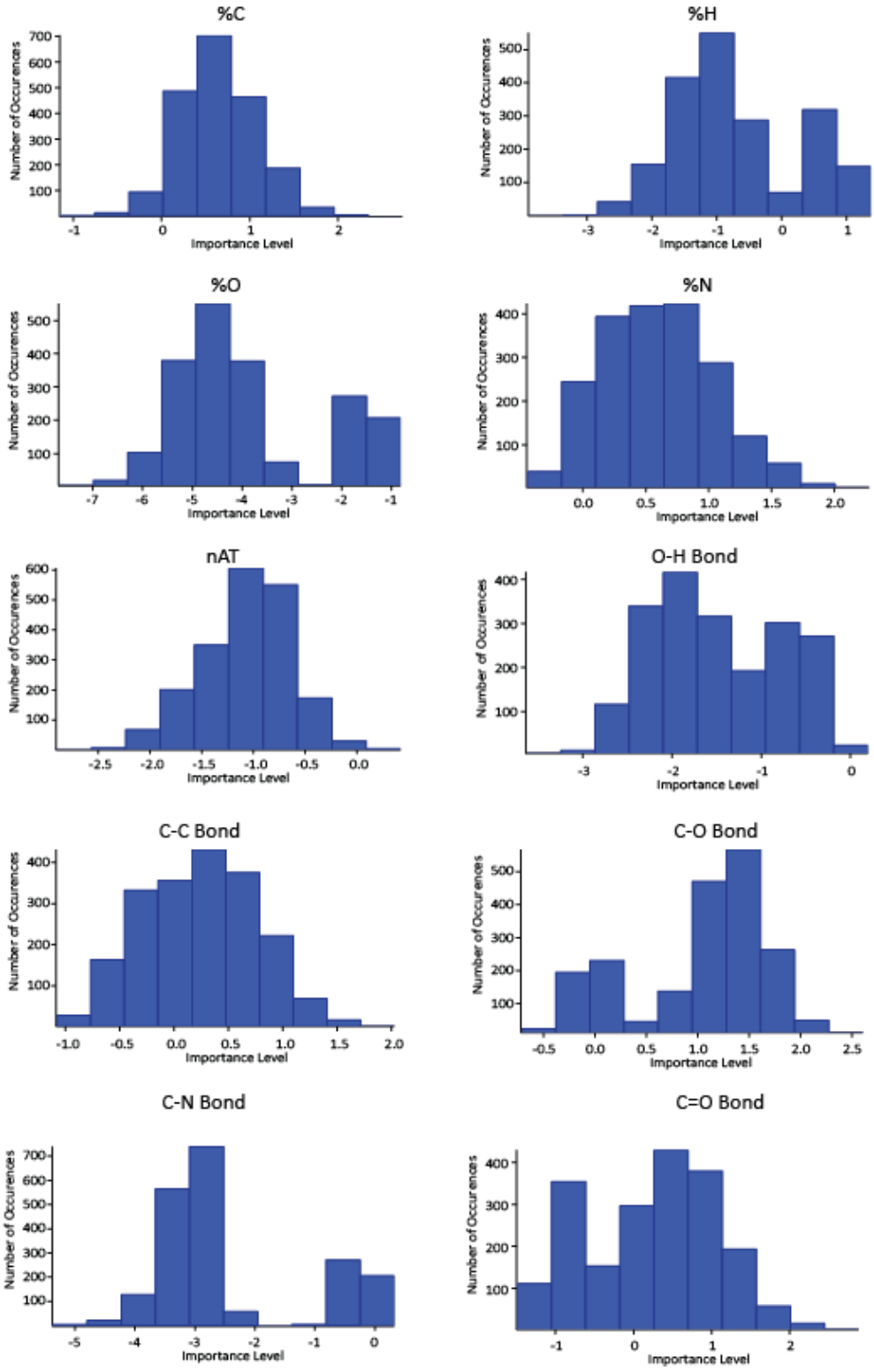


Figure 4. 7 Histogram of importance value of structure descriptors towards fibrinogen adsorption across 2000 of initial guess of connection weight of ANN trained by multi-lab dataset

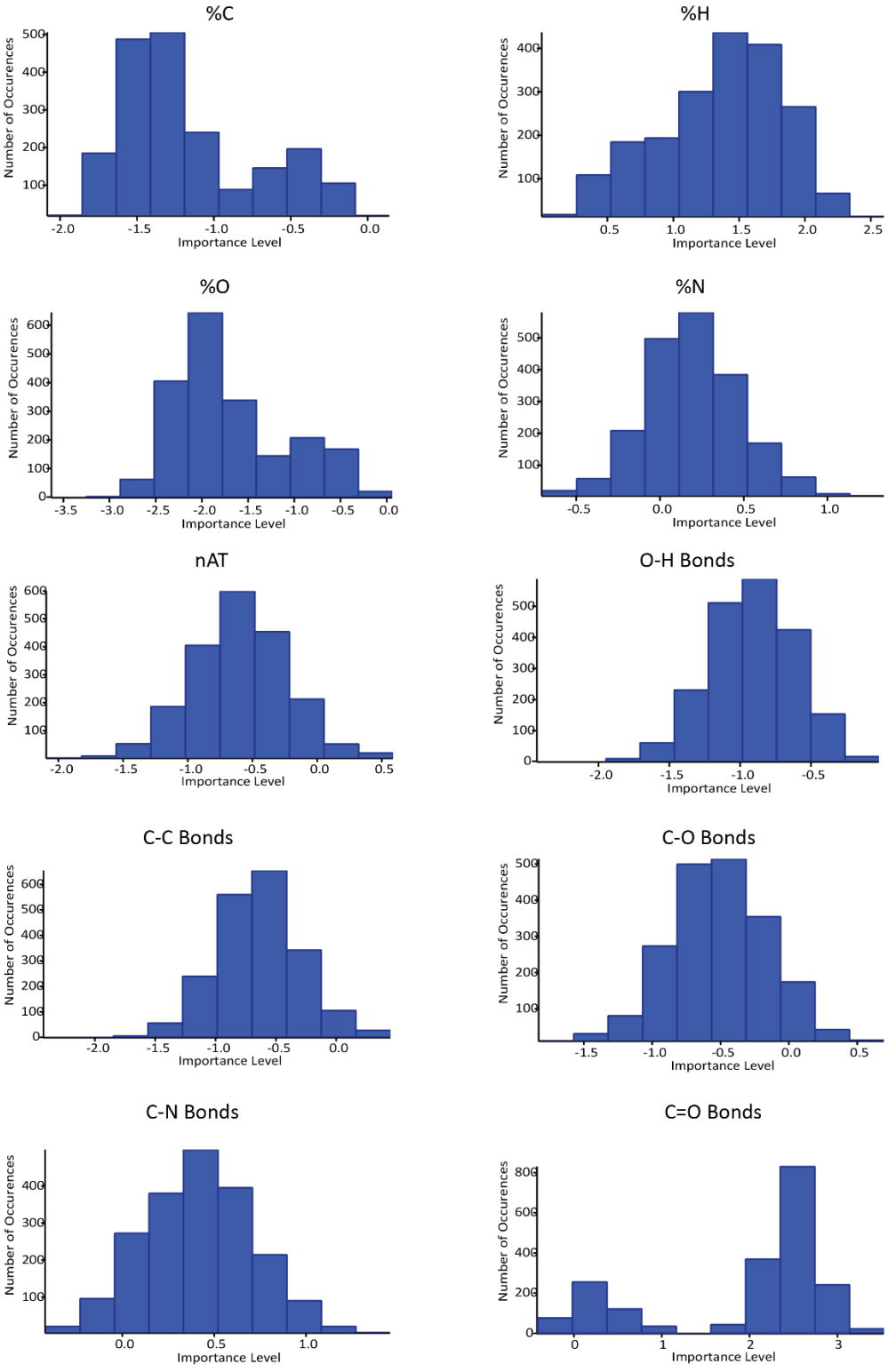
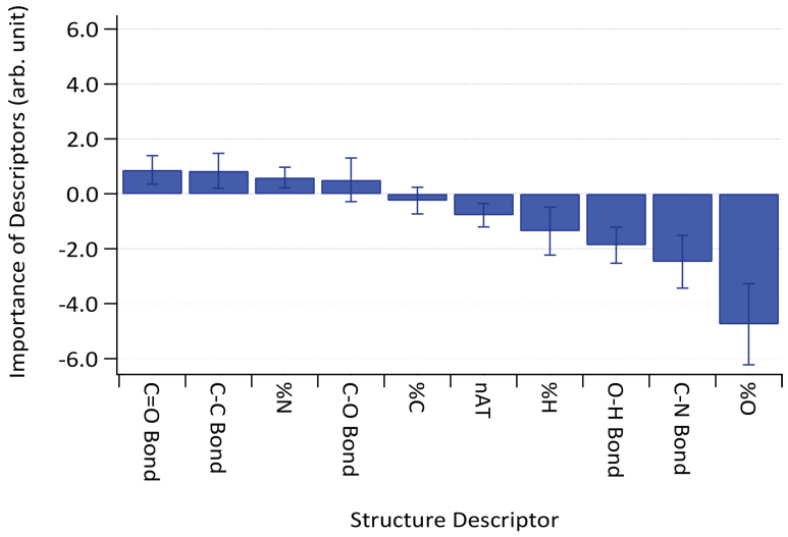
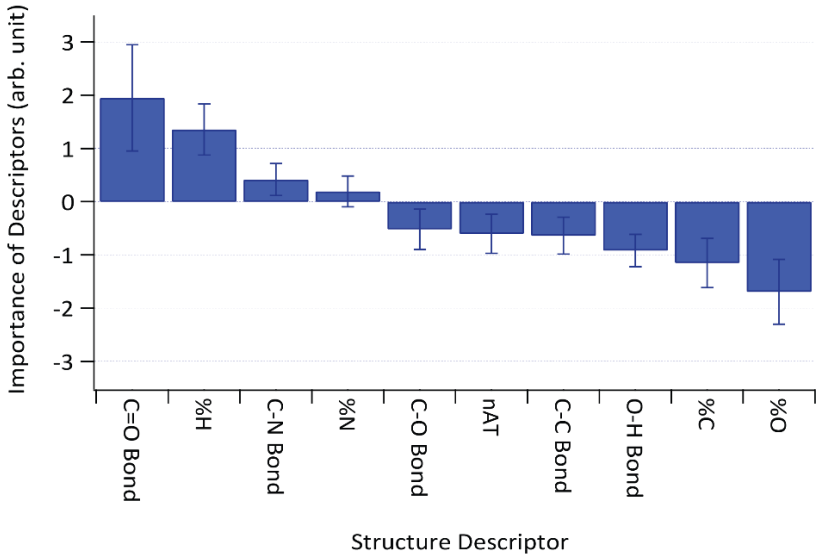


Figure 4. 8 Histogram of importance value of structure descriptors towards fibrinogen adsorption across 2000 of initial guess of connection weight of ANN trained by single-lab dataset



(a)



(b)

Figure 4. 9 analysis of fibrinogen adsorption. Figure (a) shows the importance analysis result using multi-lab data and figure (b) shows the importance analysis result using single-lab data

Figure 4.9 above shows the average importance of structure descriptors towards fibrinogen adsorption for both the multi-lab and single-lab dataset using. Just like the average. These importance results were average values of importance value given by several ANNs trained using 2000 different initial connection weights.

From these results, we can see and compare the importance value produced by ANN trained by multi-lab dataset and single-lab dataset.

For the importance from single-lab dataset, we can see several structure descriptors with high importance towards fibrinogen adsorption such as:

1. C=O Bond which according to the database represents the carboxyl group
2. %N and C-N Bond which represents the amide group
3. %O and C-O Bond which represents Ethylene Glycol (EG)

These descriptors can be further classified into hydrogen bond donating groups (C=O Bond, %N, and C-N Bond) and hydrogen bond accepting groups (%O and C-O Bond)

To check to validity of the importance analysis results, we compared our importance analysis with the well-known rule for fibrinogen adsorption on SAMs which is Whitesides' rule [11]. According to Whitesides' rule, protein resistant SAMs has these following criteria:

1. Contain polar functional group and hydrogen bond accepting groups
2. Do not contain hydrogen bond donating group and have no net charge

These criteria could be translated into these following points:

1. Contain polar functional group: structure descriptors that corresponds to polar functional groups should have negative importance towards fibrinogen adsorption as it inhibits protein adsorption
2. Contain hydrogen bond accepting group: structure descriptors that corresponds to hydrogen bond accepting groups should have negative importance towards fibrinogen adsorption as it inhibits protein adsorption
3. Do not contain hydrogen bond donating group: structure descriptors that corresponds to hydrogen bond donating group should have positive importance towards fibrinogen adsorption as the mere existence of hydrogen bond donating group will increase protein adsorption
4. Have no net charge: Whitesides' rule states that non-fouling materials have no net charge or neutrally charged. This means, structure descriptors that represents group which charge is not neutrally charged will have positive importance.

For the purpose of discussion, we will explain the analysis from the single-lab dataset first. Looking at the importance analysis results produced by ANN trained by single-lab dataset, the structure descriptors which represents hydrogen bond accepting group such as %O and C-O Bond which represents EG group shows negative importance. On the other side, structure descriptors which represents hydrogen bond donating groups such as C=O Bond, %N, and C-N Bond shows positive importance.

Moreover, Whitesides' rule also states the importance of polar functional group in reducing protein adsorption which means that descriptors which represents polar functional groups should have negative importance. That means, several polar functional groups in our set of descriptors which are: C-O Bond, O-H Bond, C-N Bond, and C=O Bond which corresponds to EG group, hydroxyl group, amide group, and

carboxyl group, respectively should have negative importance. Looking at the importance analysis result for single-lab dataset, we can see that C-O Bond and O-H Bond have negative importance which means that the result is in agreement with Whitesides' rule. However, we also see several exceptions such as C=O Bond and C-N bond. A possible explanation for this is: C=O Bond and C-N bond are also representing hydrogen bond donating group. This means, despite these groups are polar functional groups, because of their nature as hydrogen bond donating groups, they still promote protein adsorption. On the other hand, O-H Bond which is also both a hydrogen bond donating group and polar functional group, has its nature as a polar functional group as the more dominant factor in deciding fibrinogen adsorption. From this, we can see that ANN is able to put importance value even when the functional groups in question have conflicting nature.

Moreover, regarding the charge-neutrality of the surface, we considered the net charge of SAMs when we measure the fibrinogen adsorption. In single-lab dataset, the fibrinogen adsorption was measured in water which means the surrounding pH was 7. Looking at the importance analysis results of fibrinogen adsorption by ANN trained by single-lab dataset, we can see that C=O Bond which represents carboxyl group has positive importance. Since hydroxyl group has isoelectric points of 5.1 to 5.8, it just makes sense if this SAMs is negatively charged at pH of 7. This result suggests that the non-neutrality of SAMs which was caused by the existence of the carboxyl group promotes fibrinogen adsorption which is shown by the importance analysis result.

Other descriptors that should be observed is the %C which exists in almost all of the functional groups. Here, the %C is negatively important which is make sense because the dataset also contains EG SAMs. With EG consists of carbon, increasing

the length of EG chain will also increase the %C. As we know that EG groups reduce fibrinogen adsorption, it makes sense that increase in %C also reduce fibrinogen adsorption. This is reflected in the important analysis results with %C having negative important.

We also observed the positive importance of %H. Hydrogen exists in every functional group of SAMs included in this dataset. However, different functional group will have different percentage of hydrogen. For example, using CH_3 to terminate the functional group of SAM will increase the number of hydrogen atom by 3 while using OH to terminate the functional group of SAM will only increase the hydrogen atom by 1. From the previous discussion, we know that OH bond, a polar functional group, will decrease the fibrinogen adsorption of SAM. However, CH_3 is a non-polar group which means adding this into the SAM will not decrease the fibrinogen adsorption.

Moreover, we can also consider amide groups (NH_2) to further discuss the role of %H. In the previous discussion, we know that amide group is a hydrogen bond donating group which has positive importance towards fibrinogen adsorption. We also know that by terminating the functional group of SAM by NH_2 will add 2 hydrogen atom which increases the hydrogen atom percentage.

By comparing these three terminal groups (CH_3, OH, and NH_2), it can be concluded that %H represents many functional groups that promotes fibrinogen adsorption. This is reflected in the importance analysis result which shows that %H as a descriptor with positive importance.

Moreover, regarding the C-C bond, we can see that the importance of C-C bond is low which suggests that C-C bond does not influence the fibrinogen adsorption. This is

because the C-C bond usually exists as the alkyl chain which acts as a spacer between the head group of the SAM and the terminal group of the SAM. As a result, the increase of number of alkyl chain which only increase the length of the spacer does not contributes much towards the change in fibrinogen adsorption.

Looking at the result of importance analysis result of fibrinogen adsorption by ANN trained by single-lab dataset, we can see that the importance of each molecular descriptor agrees with Whitesides' rule.

Now, we will look at the importance analysis result of fibrinogen adsorption by ANN trained by multi-lab dataset. Once more, we tried to describe the descriptors based on the data in the database. The descriptors were classified as follows:

1. %N and C-N Bond represents the amide group in multi-lab dataset because majority of the data with C-N Bond contains amide group
2. %O and C-O Bond represents Ethylene Glycol (EG) in multi-lab dataset

Looking at the results and comparing it to Whitesides' rule, we see results that does not falls in agreement with Whitesides' rule. For example, although %O and C-O Bond both represents EG groups which is hydrogen accepting, we can see that these two descriptors have importance value that is far from each other. While %O is still negatively important, the C-O bond is positively important although it can be considered low importance.

Moreover, it should be noted that because the ANN could not validate the prediction results of multi-lab dataset with high accuracy, the possibility that the connection weights placed for neurons on ANN trained by multi-lab dataset are wrong is high.

From this result, it is too early to say that the result of the importance analysis trained by multi-lab dataset agrees with Whitesides' rule. Rather, considering several disagreements of importance in several descriptors that should be connected with each other suggested that there is possibility that the importance value of fibrinogen adsorption by ANN trained by multi-lab dataset is wrong.

However, looking at both results, we can conclude that, when trained using the right data, the ANN is able to calculate the importance of each structure descriptors and presents the importance quantitatively. Moreover, the results can also be used to know the direction of the importance. Through this capability, we can make a clear and quantitative guide for material design in which we will be able to focus on the most important part to design material with specific fibrinogen adsorption.

Although we have succeeded in producing results of importance analysis that represents the well-known empirical criteria, we do realize that using molecular descriptors as the input means that there are various mechanism underlying the protein resistance that could not be explained. That is because the goal of the current experiment is to make a guide that can be easily used for material design using the structural descriptor as the input. However, we also expect that the importance analysis between physicochemical properties (e.g., WCA, zeta potential, molecular packing density, pKa) and protein adsorption will provide valuable findings to explain the mechanism underlying protein resistance.

4.3.5 Prediction of Fibrinogen Adsorption on Imaginary SAMs

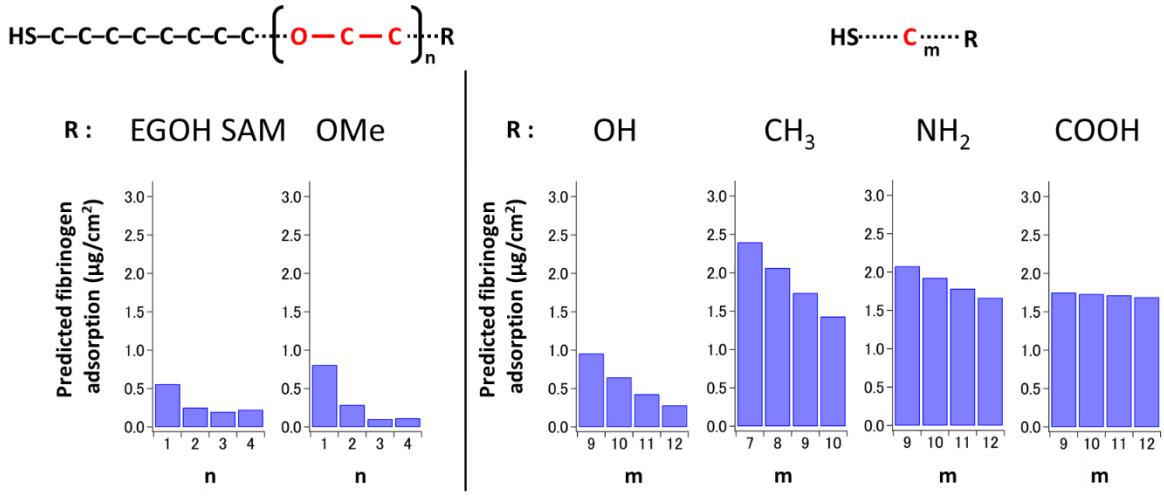


Figure 4. 10 Prediction of Fibrinogen Adsorption for SAMs outside of training dataset. For EGOH and OMe SAMs, the EG units were changed while for OH, CH_3, H_2, and COOH SAMs, the number of carbon atoms in their alkyl chains are changed.

Figure 4.10 shows the prediction of fibrinogen adsorption from several imaginary SAMs or SAMs that is outside of the training dataset. Here, we can see that changing the length of alkyl chain unit by adding carbon atoms does not affect the fibrinogen adsorption so much as shown in several picked data as shown by CH_3, OH, NH_2, and COOH SAMs. On the contrary, as shown by the data of EGOMe and EGOH SAMs, the addition of EG unit drastically affects the fibrinogen adsorption of SAMs. This can be seen that simply by adding single EG unit already put the fibrinogen of EGOH SAMs and EGOH SAMs into 0.5 and 0.75 respectively with further addition keeps on decreasing the fibrinogen adsorption until it becomes almost fibrinogen resistant.

There are some interesting phenomena in the prediction of CH_3 and OH SAMs where there are trends suggesting that by increasing the length of alkyl chain by adding

carbon atom seemingly decreases the fibrinogen adsorption. This does not agree with our empirical prediction as C-C bond should not affect the fibrinogen adsorption at all. This result may arise from the nature of EG unit which also incorporates C-C bond similar to alkyl chain. This similarity manifested itself in which: as EG unit increases, C-C bond also increases. This trend was interpreted by the ANN algorithm as SAMs that fibrinogen adsorption decreases as C-C bond increases which is also reflected in the importance analysis that suggests the negative importance of C-C bond although the importance value itself is very small when compared with C-O bond. This is clearly the artifact of ML algorithm in this work and to overcome this, we need to optimize the database by setting a new descriptor to distinguish the alkyl chain and EG unit.

Despite the small artifact arising from the database, these results demonstrated that the trained ANNs reproduced our general understanding on SAMs, i.e., the fibrinogen adsorption of the SAM is not determined by the length of the alkyl chain, but by the terminal groups.

Chapter 5: Conclusion

The target of this research was to discuss the possibility of data-driven based approach using ANN algorithm to predict material properties from the structure of the molecules constituting the material and make quantitative criteria that can serve as a guide for future design of material. This was achieved by demonstrating the prediction of several material properties of SAMs (i.e. WCA and fibrinogen adsorption) as well as ranking the importance of structure descriptors and then validate the results with established sources.

Chapter 1 briefly discussed the background, purpose, and significance of this work. Currently, design of biomaterial that has desirable properties (i.e. specific WCA and fibrinogen adsorption) for medical application is still being done in trial-and-error manner. This is mainly because of lack of quantitative guide to design the biomaterial. Although using theoretical approach like DFT and molecular simulation can seemingly make a guideline for material design and solve this problem, modelling and simulating the biomaterial theoretically is still largely difficult because of the complexity of the system that needs to be modelled. To solve this problem, we are employing data-driven approach using ANN algorithm trained by the previously published experimental data of properties of SAMs to make quantitative guide to design material with specific WCA and fibrinogen adsorption, as well as predicting the WCA and fibrinogen adsorption. However, it is also important to discuss the possibility of applying this method when the data, particularly on fibrinogen adsorption data, were gathered from different sources.

Chapter 2 discusses the basic theory and mathematics behind ANN algorithm. The mathematics can serve as a detailed guide to understand how the algorithm used in this

research works. This can also serve as a guide to better understand the experimental method employed in this research.

In Chapter 3, the research is being focused on WCA prediction and importance analysis. With the data of WCA of various SAMs already widely available, we discuss the possibility of using these data for data-driven approach using ANN algorithm to make quantified guidelines to design material with specific WCA. By looking at the high accuracy of the validation results, we can conclude that WCA data shows little effect from measurement conditions and that it is possible to use data from any papers to increase the scope of applicability of this method. Next, we analyse the importance of several material descriptors towards WCA and validate the result with well-established knowledge regarding surface polarity. We found out that the importance analysis result agrees well with the well-established knowledge which proof the validity of this method. Moreover, the result of prediction of several imaginary SAMs suggests that the prediction result makes sense particularly with our knowledge that neither increasing the length of alkyl chain by adding carbon atom nor adding EG unit to the molecule change the WCA of the surface. These results suggest that data-driven approach using ANN is applicable to design material with specific WCA.

Chapter 4 discusses the possibility of applying data-driven approach using ANN for fibrinogen adsorption prediction of SAMs. Here, we found out that different experimental approach employed in each lab caused large discrepancy of data of fibrinogen adsorption. To study the effect of this discrepancy on the prediction result as well as the importance analysis result, we made two separate datasets which is multi-lab and single-lab dataset to train the ANN for prediction of fibrinogen adsorption as well as importance analysis of structure descriptors. From the result, we found out that prediction of ANN trained by multi-lab dataset shows large error when compared with

prediction of ANN trained by single-lab dataset. This result suggests that discrepancy of data affects the prediction results which means that while we can use data gathered from any paper for WCA prediction, the data for fibrinogen adsorption should be selected carefully. Moreover, we also study the effect of this discrepancy on importance analysis. The result of importance analysis of the ANN trained by multi-lab dataset and single-lab dataset is further analysed. For the single-lab dataset, we found many qualities that is in agreement with well-known criteria of protein resistant SAMs as established by Whitesides, et. al. Meanwhile, for the multi-lab dataset, we found several qualities that is not in agreement with well-known criteria such as %O and C-O bond that both represents the number of EG unit in SAMs molecule have different direction of importance which does not make sense.

Overall, it is safe to assume that, in order to apply data-driven method to predict material properties as well as establishing quantified criteria, we need a dataset consisting of data taken under the same measurement condition just like the single-lab dataset to train the ANN algorithm.

Just like in the case of WCA, we demonstrate the capability of the ANN to make several predictions of SAMs outside of the training dataset. The results made sense and agrees well with previously published data on fibrinogen adsorption which further validate this method.

Acknowledgement

First, I would like to sincerely thank Professor Tomohiro Hayashi for his advices, guidance, patience, and most importantly his mentorship during my entire doctoral course. He helped me not only on my research but also supported me in my job-hunting period.

I also want to thank my lab mates for spending their time to help me on experiments as well as their continuous advices and helps for both scientific and non-scientific problems. Moreover, I would also like to thank Kazue Taki, the secretary of the lab for all the help especially in the administration of this work.

I would like to thank my friend, Marcel Satria, my friend of the same grade in Tokyo Institute of Technology who has passed away in 2018. He has always been cheering me up when I was feeling down and I am very grateful for his supports. I pray that he will find happiness in heaven.

Special thanks to my entire choir community, Chamber Choir Vox Gaudiosa. Together with them I am able to not only having extracurricular activities where I can cultivate my interpersonal skills and hobbies, and also very helpful for my Japanese language study.

I would also like to thank Indonesia Endowment Fund for Education (LPDP), Ministry of Finance of the Republic of Indonesia, and Ministry of Education, Culture, Sports, Science and Technology (MEXT) who has supported my study and research in Japan for my master study and doctor study respectively.

Lastly, I would like to thank my family - my mother, father, and brother- for their love and lifelong supports. Without their supports, I would not be here today.

Copyright Permission

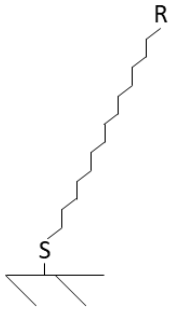
This thesis includes the contents from a published journal article as the following:

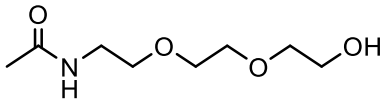
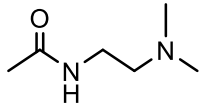
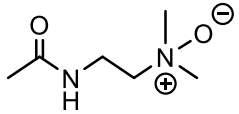
"Data-driven prediction of protein adsorption on self-assembled monolayers toward material screening and design."
***ACS Biomaterials Science & Engineering* 2020,** *in print*.

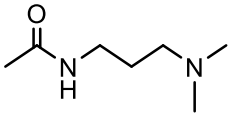
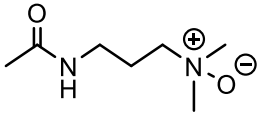
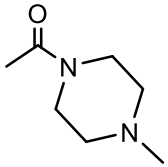
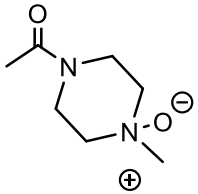
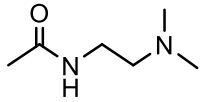
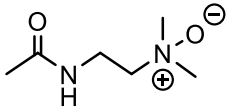
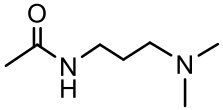
Appendix1

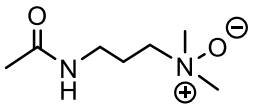
WCA Data

Dobrzanska SAMs[32]



No.	R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
1.		30.14	61.64	5.48	1.37	73.00	1.00	18.00	2.00	5.00	1.00	43.00
2.		30.30	63.64	1.52	3.03	66.00	0.00	16.00	5.00	0.00	1.00	34.00
3.		29.85	62.69	2.99	2.99	67.00	0.00	16.00	5.00	0.00	1.00	25.00

4.		30.43	63.77	1.45	2.90	69.00	0.00	17.00	5.00	0.00	1.00	37.00
No.	R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
5.		30.00	62.86	2.86	2.86	70.00	0.00	17.00	5.00	0.00	1.00	28.00
6.		31.34	62.69	1.49	2.99	67.00	0.00	17.00	6.00	0.00	1.00	37.00
7.		30.88	61.76	2.94	2.94	68.00	0.00	17.00	6.00	0.00	1.00	29.00
8.		30.43	63.77	1.45	2.90	69.00	0.00	16.00	6.00	0.00	1.00	38.00
9.		30.00	62.86	2.86	2.86	70.00	0.00	16.00	6.00	0.00	1.00	30.00
10.		30.56	63.89	1.39	2.78	72.00	0.00	17.00	6.00	0.00	1.00	36.00

11.		30.14	63.01	2.74	2.74	73.00	0.00	17.00	6.00	0.00	1.00	28.00

Abe SAMs[28]

No.	Structure	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
12.	$HS-(CH_2)_{17}-CH_3$	31.58	66.67	0.00	0.00	57.00	0.00	17.00	0.00	0.00	0.00	107.00

Sekine SAMs

[21]

SAM1	**$HS(CH_2)_5CH_3$**											
SAM2	**$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
13.	0	28.57	66.67	0.00	0.00	21.00	0.00	5.00	0.00	0.00	0.00	105.00
14.	12.60	28.78	65.36	1.96	0.00	25.66	0.13	6.01	0.00	0.88	0.00	91.20
15.	16.75	28.84	65.02	2.46	0.00	27.20	0.17	6.34	0.00	1.17	0.00	74.20
16.	17.08	28.84	65.00	2.50	0.00	27.32	0.17	6.37	0.00	1.20	0.00	73.30
17.	20.40	28.88	64.76	2.86	0.00	28.55	0.20	6.63	0.00	1.43	0.00	71.80
18.	28.64	28.96	64.25	3.63	0.00	31.60	0.29	7.29	0.00	2.00	0.00	64.20
19.	38.00	29.04	63.78	4.34	0.00	35.06	0.38	8.04	0.00	2.66	0.00	47.10
20.	55.38	29.14	63.11	5.34	0.00	41.49	0.55	9.43	0.00	3.88	0.00	39.40
21.	71.60	29.22	62.65	6.03	0.00	47.49	0.72	10.73	0.00	5.01	0.00	37.10
22.	72.87	29.22	62.62	6.08	0.00	47.96	0.73	10.83	0.00	5.10	0.00	36.10
23.	78.73	29.24	62.48	6.28	0.00	50.13	0.79	11.30	0.00	5.51	0.00	35.70

24.	83.57	29.26	62.37	6.44	0.00	51.92	0.84	11.69	0.00	5.85	0.00	35.70
25.	90.61	29.28	62.24	6.65	0.00	54.53	0.91	12.25	0.00	6.34	0.00	34.70
26.	91.78	29.29	62.21	6.68	0.00	54.96	0.92	12.34	0.00	6.42	0.00	33.80
27.	100	29.31	62.07	6.90	0.00	58.00	1.00	13.00	0.00	7.00	0.00	30.00

Arima SAMs[22]

SAM1	**$HS(CH_2)_{10}CH_3$**											
SAM2	**$HS(CH_2)_{11}OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
28.	0	30.56	66.67	0.00	0.00	36.00	0.00	10.00	0.00	0.00	0.00	109.00
29.	10	30.47	66.48	0.28	0.00	36.10	0.10	10.00	0.00	0.10	0.00	102.00
30.	20	30.39	66.30	0.55	0.00	36.20	0.20	10.00	0.00	0.20	0.00	93.00
31.	50	30.14	65.75	1.37	0.00	36.50	0.50	10.00	0.00	0.50	0.00	60.00
32.	63	30.03	65.52	1.72	0.00	36.63	0.63	10.00	0.00	0.63	0.00	46.00
33.	87	29.83	65.09	2.36	0.00	36.87	0.87	10.00	0.00	0.87	0.00	36.00
34.	93	29.79	64.99	2.52	0.00	36.93	0.93	10.00	0.00	0.93	0.00	30.00
35.	100	29.73	64.86	2.70	0.00	37.00	1.00	10.00	0.00	1.00	0.00	22.00
SAM1	**$HS(CH_2)_{10}CH_3$**											
SAM2	**$HS(CH_2)_{11}NH_2$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
36.	5	30.47	66.62	0.00	0.14	36.10	0.00	10.00	0.05	0.00	0.00	103.00
37.	9	30.40	66.58	0.00	0.25	36.18	0.00	10.00	0.09	0.00	0.00	99.00
38.	15	30.30	66.53	0.00	0.41	36.30	0.00	10.00	0.15	0.00	0.00	93.00
39.	28	30.09	66.41	0.00	0.77	36.56	0.00	10.00	0.28	0.00	0.00	90.00
40.	59	29.59	66.14	0.00	1.59	37.18	0.00	10.00	0.59	0.00	0.00	73.00
41.	90	29.10	65.87	0.00	2.38	37.80	0.00	10.00	0.90	0.00	0.00	55.00
42.	100	28.95	65.79	0.00	2.63	38.00	0.00	10.00	1.00	0.00	0.00	51.00

SAM1	**$HS(CH_2)_{10}CH_3$**											
SAM2	**$HS(CH_2)_{10}COOH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
43.	12	30.56	66.00	0.67	0.00	36.00	0.12	10.00	0.00	0.12	0.12	103.00
44.	26	30.56	65.22	1.44	0.00	36.00	0.26	10.00	0.00	0.26	0.26	88.00
45.	50	30.56	63.89	2.78	0.00	36.00	0.50	10.00	0.00	0.50	0.50	74.04
46.	60	30.56	63.33	3.33	0.00	36.00	0.60	10.00	0.00	0.60	0.60	54.00
47.	83	30.56	62.06	4.61	0.00	36.00	0.83	10.00	0.00	0.83	0.83	36.00
48.	90	30.56	61.67	5.00	0.00	36.00	0.90	10.00	0.00	0.90	0.90	30.00
49.	100	30.56	61.11	5.56	0.00	36.00	1.00	10.00	0.00	1.00	1.00	21.00

Zhu SAMs [26]

No.	Structure	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
50.	$HS\text{-}(CH_2)_{11}\text{-}(O\text{-}CH_2\text{-}CH_2)_3\text{-}OCH_3$	29.51	62.30	6.56	0.00	61.00	0.00	13.00	0.00	8.00	0.00	56.50
51.	$HS\text{-}(CH_2)_{11}\text{-}(O\text{-}CH_2\text{-}CH_2)_6\text{-}OH$	29.11	60.76	8.86	0.00	79.00	1.00	16.00	0.00	13.00	0.00	29.00

Bain SAMs[25]

SAM1	**$HS(CH_2)_{10}CH_3$**											
SAM2	**$HS(CH_2)_{10}CH_2OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
52.	0	30.56	66.67	0.00	0.00	36.00	0.00	10.00	0.00	0.00	0.00	113.00
53.	5	30.51	66.57	0.14	0.00	36.05	0.05	10.00	0.00	0.05	0.00	101.00
54.	20	30.39	66.30	0.55	0.00	36.20	0.20	10.00	0.00	0.20	0.00	81.00
55.	32	30.29	66.08	0.88	0.00	36.32	0.32	10.00	0.00	0.32	0.00	78.00
56.	47	30.16	65.81	1.29	0.00	36.47	0.47	10.00	0.00	0.47	0.00	52.00
57.	79	29.90	65.24	2.15	0.00	36.79	0.79	10.00	0.00	0.79	0.00	29.00
58.	14	30.56	65.98	0.68	0.00	41.04	0.28	11.40	0.00	0.28	0.00	92.00
SAM1	**$HS(CH_2)_{10}CH_3$**											
SAM2	**$[S(CH_2)_{10}CH_2OH)]_2$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
59.	19	30.56	65.78	0.89	0.00	42.84	0.38	11.90	0.00	0.38	0.00	81.00
60.	36	30.56	65.20	1.47	0.00	48.96	0.72	13.60	0.00	0.72	0.00	68.00
61.	49	30.56	64.84	1.83	0.00	53.64	0.98	14.90	0.00	0.98	0.00	55.00
62.	58	30.56	64.63	2.04	0.00	56.88	1.16	15.80	0.00	1.16	0.00	47.00
63.	100	30.56	63.89	2.78	0.00	72.00	2.00	20.00	0.00	2.00	0.00	15.00

SAM1	**$[S(CH_2)_{10}CH_3)]_2$**											
SAM2	**$HS(CH_2)_{10}CH_2OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
64.	0	31.43	65.71	0.00	0.00	70.00	0.00	20.00	0.00	0.00	0.00	110.00
65.	31	31.10	65.55	0.52	0.00	59.77	0.31	16.90	0.00	0.31	0.00	74.00
66.	35	31.05	65.53	0.60	0.00	58.45	0.35	16.50	0.00	0.35	0.00	69.00
67.	45	30.92	65.46	0.82	0.00	55.15	0.45	15.50	0.00	0.45	0.00	57.00
68.	60	30.68	65.34	1.20	0.00	50.20	0.60	14.00	0.00	0.60	0.00	41.00
69.	77	30.34	65.17	1.73	0.00	44.59	0.77	12.30	0.00	0.77	0.00	33.00
SAM1	**$[S(CH_2)_{10}CH_2OH)]_2$**											
SAM2	**$[S(CH_2)_{10}CH_3)]_2$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
70.	4	31.39	65.64	0.11	0.00	70.08	0.08	20.00	0.00	0.08	0.00	101.00
71.	14	31.30	65.45	0.40	0.00	70.28	0.28	20.00	0.00	0.28	0.00	89.00
72.	28	31.18	65.19	0.79	0.00	70.56	0.56	20.00	0.00	0.56	0.00	71.00
73.	35	31.12	65.06	0.99	0.00	70.70	0.70	20.00	0.00	0.70	0.00	61.00
74.	70	30.81	64.43	1.96	0.00	71.40	1.40	20.00	0.00	1.40	0.00	41.00
75.	78	30.74	64.28	2.18	0.00	71.56	1.56	20.00	0.00	1.56	0.00	31.00
76.	82	30.71	64.21	2.29	0.00	71.64	1.64	20.00	0.00	1.64	0.00	29.00
77.	94	30.61	64.00	2.62	0.00	71.88	1.88	20.00	0.00	1.88	0.00	16.00

Hayashi SAMs[19]

No.	Structure	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
78.	$HS-(CH_2)_7-CH_3$	29.63	66.67	0.00	0.00	27.00	0.00	7.00	0.00	0.00	0.00	112.00
79.	$HS-(CH_2)_{11}-OH$	29.73	64.86	2.70	0.00	37.00	1.00	10.00	0.00	1.00	0.00	17.00
80.	$HS-(CH_2)_{11}-NH_2$	28.95	65.79	0.00	2.63	38.00	0.00	10.00	1.00	0.00	0.00	35.00
81.	$HS-(CH_2)_{11}-COOH$	30.77	61.54	5.13	0.00	39.00	1.00	11.00	0.00	1.00	1.00	18.00
82.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)-OH$	29.55	63.64	4.55	0.00	44.00	1.00	11.00	0.00	3.00	0.00	33.00
83.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_2-OH$	29.41	62.75	5.88	0.00	51.00	1.00	12.00	0.00	5.00	0.00	32.00
84.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OH$	29.31	62.07	6.90	0.00	58.00	1.00	13.00	0.00	7.00	0.00	32.00
85.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-O-CH_3$	29.51	62.30	6.56	0.00	61.00	0.00	13.00	0.00	8.00	0.00	69.00
86.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-NH_2$	28.81	62.71	5.08	1.69	59.00	0.00	13.00	1.00	6.00	0.00	25.00
87.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-COOH$	30.00	60.00	8.33	0.00	60.00	1.00	14.00	0.00	7.00	1.00	22.00
88.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OH$	29.11	60.76	8.86	0.00	79.00	1.00	16.00	0.00	13.00	0.00	30.00

Folkers SAMs[27]

SAM1	**$HS(CH_2)_{10}CH_2OH$**											
SAM2	**$HS(CH_2)_{21}CH_3$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O(%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
89.	0.00	29.73	64.86	2.70	0.00	37.00	1.00	10.00	0.00	1.00	0.00	5.00
90.	8.00	30.03	65.12	2.33	0.00	39.56	0.92	10.88	0.00	0.92	0.00	27.50
91.	12.00	30.17	65.23	2.15	0.00	40.84	0.88	11.32	0.00	0.88	0.00	38.50
92.	25.00	30.56	65.56	1.67	0.00	45.00	0.75	12.75	0.00	0.75	0.00	51.00
93.	40.00	30.92	65.86	1.20	0.00	49.80	0.60	14.40	0.00	0.60	0.00	66.50
94.	65.00	31.40	66.26	0.61	0.00	57.80	0.35	17.15	0.00	0.35	0.00	86.00
95.	80.00	31.63	66.45	0.32	0.00	62.60	0.20	18.80	0.00	0.20	0.00	94.00
96.	90.00	31.76	66.57	0.15	0.00	65.80	0.10	19.90	0.00	0.10	0.00	100.00
97.	95.00	31.82	66.62	0.07	0.00	67.40	0.05	20.45	0.00	0.05	0.00	107.50
98.	99.00	31.87	66.66	0.01	0.00	68.68	0.01	20.89	0.00	0.01	0.00	105.50
99.	100.00	31.88	66.67	0.00	0.00	69.00	0.00	21.00	0.00	0.00	0.00	109.00

SAM1	**$HS(CH_2)_{10}CH_2OH$**											
SAM2	**$HS(CH_2)_{21}CH_2OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O(%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
100.	0.00	29.73	64.86	2.70	0.00	37.00	1.00	10.00	0.00	1.00	0.00	7.50
101.	7.50	29.96	64.98	2.53	0.00	39.48	1.00	10.83	0.00	1.00	0.00	26.50
102.	10.00	30.02	65.01	2.48	0.00	40.30	1.00	11.10	0.00	1.00	0.00	38.85
103.	31.00	30.51	65.26	2.12	0.00	47.23	1.00	13.41	0.00	1.00	0.00	43.50
104.	47.50	30.80	65.40	1.90	0.00	52.68	1.00	15.23	0.00	1.00	0.00	44.25
105.	67.50	31.08	65.54	1.69	0.00	59.28	1.00	17.43	0.00	1.00	0.00	40.00
106.	77.50	31.20	65.60	1.60	0.00	62.58	1.00	18.53	0.00	1.00	0.00	34.00
107.	100.00	31.43	65.71	1.43	0.00	70.00	1.00	21.00	0.00	1.00	0.00	20.00

SAM1	**$HS(CH_2)_{10}CH_3$**											
SAM2	**$HS(CH_2)_{21}CH_3$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O(%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
108.	0	30.56	66.67	0.00	0.00	36.00	0.00	10.00	0.00	0.00	0.00	104.75
109.	0.05	30.68	66.67	0.00	0.00	37.65	0.00	10.55	0.00	0.00	0.00	105.50
110.	0.1	30.79	66.67	0.00	0.00	39.30	0.00	11.10	0.00	0.00	0.00	104.25
111.	0.18	30.95	66.67	0.00	0.00	41.94	0.00	11.98	0.00	0.00	0.00	104.50
112.	0.23	31.04	66.67	0.00	0.00	43.59	0.00	12.53	0.00	0.00	0.00	103.75
113.	0.4	31.30	66.67	0.00	0.00	49.20	0.00	14.40	0.00	0.00	0.00	102.95
114.	0.48	31.40	66.67	0.00	0.00	51.84	0.00	15.28	0.00	0.00	0.00	103.75
115.	0.63	31.57	66.67	0.00	0.00	56.79	0.00	16.93	0.00	0.00	0.00	102.65
116.	0.75	31.69	66.67	0.00	0.00	60.75	0.00	18.25	0.00	0.00	0.00	104.40
117.	0.83	31.76	66.67	0.00	0.00	63.39	0.00	19.13	0.00	0.00	0.00	103.30
118.	0.95	31.85	66.67	0.00	0.00	67.35	0.00	20.45	0.00	0.00	0.00	104.30
119.	1	31.88	66.67	0.00	0.00	69.00	0.00	21.00	0.00	0.00	0.00	107.50

SAM1	**$HS(CH_2)_{10}CH_3$**											
SAM2	**$HS(CH_2)_{21}CH_2OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O(%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
120.	**0.00**	**30.56**	**66.67**	**0.00**	**0.00**	**36.00**	**0.00**	**10.00**	**0.00**	**0.00**	**0.00**	**107.50**
121.	2.00	30.59	66.63	0.05	0.00	36.68	0.02	10.22	0.00	0.02	0.00	104.00
122.	10.00	30.71	66.50	0.25	0.00	39.40	0.10	11.10	0.00	0.10	0.00	100.25
123.	19.00	30.83	66.37	0.45	0.00	42.46	0.19	12.09	0.00	0.19	0.00	96.00
124.	29.00	30.94	66.25	0.63	0.00	45.86	0.29	13.19	0.00	0.29	0.00	85.00
125.	44.00	31.08	66.09	0.86	0.00	50.96	0.44	14.84	0.00	0.44	0.00	79.00
126.	77.00	31.31	65.84	1.24	0.00	62.18	0.77	18.47	0.00	0.77	0.00	54.00
127.	95.00	31.41	65.74	1.39	0.00	68.30	0.95	20.45	0.00	0.95	0.00	32.50
128.	99.00	31.42	65.72	1.42	0.00	69.66	0.99	20.89	0.00	0.99	0.00	25.75

Herrwerth SAMs[20]

No.	Structure	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	WCA
129.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)-OMe$	29.79	63.83	4.26	0.00	47.00	0.00	11.00	0.00	4.00	0.00	71.00
130.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_2-OH$	29.41	62.75	5.88	0.00	51.00	1.00	12.00	0.00	5.00	0.00	33.00
131.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_2-OMe$	29.63	62.96	5.56	0.00	54.00	0.00	12.00	0.00	6.00	0.00	69.00
132.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OH$	29.31	62.07	6.90	0.00	58.00	1.00	13.00	0.00	7.00	0.00	34.00
133.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OMe$	29.51	62.30	6.56	0.00	61.00	0.00	13.00	0.00	8.00	0.00	65.00
134.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OEt$	29.69	62.50	6.25	0.00	64.00	0.00	14.00	0.00	8.00	0.00	84.00
135.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OPr$	29.85	62.69	5.97	0.00	67.00	0.00	15.00	0.00	8.00	0.00	94.00
136.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OBu$	30.00	62.86	5.71	0.00	70.00	0.00	16.00	0.00	8.00	0.00	107.00
137.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OH$	29.11	60.76	8.86	0.00	79.00	1.00	16.00	0.00	13.00	0.00	32.00
138.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OMe$	29.27	60.98	8.54	0.00	82.00	0.00	16.00	0.00	14.00	0.00	67.00
139.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OEt$	29.41	61.18	8.24	0.00	85.00	0.00	17.00	0.00	14.00	0.00	87.00
140.	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OPr$	29.55	61.36	7.95	0.00	88.00	0.00	18.00	0.00	14.00	0.00	97.00
141.	$HS-(CH_2)_{11}-(O-CH(CH_3)-CH_2)_2-OMe$	30.00	63.33	5.00	0.00	60.00	0.00	14.00	0.00	6.00	0.00	73.00
142.	$HS-(CH_2)_{11}-(O-CH(CH_3)-CH_2)_3-OMe$	30.00	62.86	5.71	0.00	70.00	0.00	16.00	0.00	8.00	0.00	72.00
143.	$HS-(CH_2)_{11}-(O-CH(CH_3)-CH_2)_4-OMe$	30.00	62.50	6.25	0.00	80.00	0.00	18.00	0.00	10.00	0.00	71.00
144.	$HS-(CH_2)_{11}-(O-CH_2-CH_2-CH_2)_3-OH$	29.85	62.69	5.97	0.00	67.00	1.00	16.00	0.00	7.00	0.00	34.00
145.	$HS-(CH_2)_{11}-(O-CH_2-CH_2-CH_2)_3-OMe$	30.00	62.86	5.71	0.00	70.00	0.00	16.00	0.00	8.00	0.00	67.00

Appendix 2

Multi-lab Data

Hayashi SAMs[19]

No.	Structure	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
1	$HS-(CH_2)_7-CH_3$	29.63	66.67	0.00	0.00	27.00	0.00	7.00	0.00	0.00	0.00	2.00
2	$HS-(CH_2)_{11}-OH$	29.73	64.86	2.70	0.00	37.00	1.00	10.00	0.00	1.00	0.00	0.58
3	$HS-(CH_2)_{11}-NH_2$	28.95	65.79	0.00	2.63	38.00	0.00	10.00	1.00	0.00	0.00	1.80
4	$HS-(CH_2)_{11}-COOH$	30.77	61.54	5.13	0.00	39.00	1.00	11.00	0.00	1.00	1.00	1.70
5	$HS-(CH_2)_{11}-(O-CH_2-CH_2)-OH$	29.55	63.64	4.55	0.00	44.00	1.00	11.00	0.00	3.00	0.00	0.10
6	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_2-OH$	29.41	62.75	5.88	0.00	51.00	1.00	12.00	0.00	5.00	0.00	0.08
7	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OH$	29.31	62.07	6.90	0.00	58.00	1.00	13.00	0.00	7.00	0.00	0.00
8	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-O-CH_3$	29.51	62.30	6.56	0.00	61.00	0.00	13.00	0.00	8.00	0.00	0.01
9	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-NH_2$	28.81	62.71	5.08	1.69	59.00	0.00	13.00	1.00	6.00	0.00	1.40
10	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-COOH$	30.00	60.00	8.33	0.00	60.00	1.00	14.00	0.00	7.00	1.00	1.30
11	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OH$	29.11	60.76	8.86	0.00	79.00	1.00	16.00	0.00	13.00	0.00	0.00

Sekine SAMs[21]**:**

SAM1	**$HS(CH_2)_5CH_3$**											
SAM2	**$HS\text{-}(CH_2)_{11}\text{-}(O\text{-}\ CH_2\text{-}\ CH_2)_3\text{-}OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
12	4.43	28.66	66.14	0.78	0.00	22.64	0.04	5.35	0.00	0.31	0.00	2.58
13	5.29	28.67	66.05	0.92	0.00	22.96	0.05	5.42	0.00	0.37	0.00	2.60
14	5.46	28.67	66.03	0.95	0.00	23.02	0.05	5.44	0.00	0.38	0.00	2.51
15	5.85	28.68	65.99	1.01	0.00	23.16	0.06	5.47	0.00	0.41	0.00	2.73
16	6.30	28.69	65.95	1.08	0.00	23.33	0.06	5.50	0.00	0.44	0.00	2.66
17	6.35	28.69	65.94	1.09	0.00	23.35	0.06	5.51	0.00	0.44	0.00	2.63
18	6.93	28.70	65.88	1.18	0.00	23.56	0.07	5.55	0.00	0.48	0.00	2.31
19	7.55	28.71	65.82	1.27	0.00	23.80	0.08	5.60	0.00	0.53	0.00	2.62
20	8.19	28.72	65.76	1.36	0.00	24.03	0.08	5.66	0.00	0.57	0.00	1.87
21	9.55	28.74	65.63	1.56	0.00	24.53	0.10	5.76	0.00	0.67	0.00	2.02
22	10.17	28.75	65.57	1.64	0.00	24.76	0.10	5.81	0.00	0.71	0.00	2.62
23	10.37	28.75	65.55	1.67	0.00	24.84	0.10	5.83	0.00	0.73	0.00	2.24
24	10.95	28.76	65.50	1.75	0.00	25.05	0.11	5.88	0.00	0.77	0.00	2.38
25	12.64	28.78	65.35	1.97	0.00	25.68	0.13	6.01	0.00	0.88	0.00	2.81
26	13.20	28.79	65.31	2.04	0.00	25.88	0.13	6.06	0.00	0.92	0.00	1.94
27	13.42	28.79	65.29	2.07	0.00	25.97	0.13	6.07	0.00	0.94	0.00	1.97
28	15.45	28.82	65.12	2.31	0.00	26.72	0.15	6.24	0.00	1.08	0.00	2.09
29	16.54	28.83	65.04	2.44	0.00	27.12	0.17	6.32	0.00	1.16	0.00	2.03
30	16.80	28.84	65.02	2.47	0.00	27.22	0.17	6.34	0.00	1.18	0.00	2.07
31	18.29	28.85	64.91	2.63	0.00	27.77	0.18	6.46	0.00	1.28	0.00	1.62
32	20.64	28.88	64.74	2.88	0.00	28.64	0.21	6.65	0.00	1.44	0.00	1.75

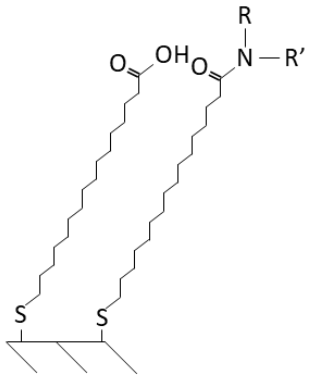
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
33	23.23	28.91	64.57	3.14	0.00	29.59	0.23	6.86	0.00	1.63	0.00	1.93
34	24.99	28.93	64.46	3.31	0.00	30.25	0.25	7.00	0.00	1.75	0.00	1.68
35	25.21	28.93	64.45	3.33	0.00	30.33	0.25	7.02	0.00	1.77	0.00	1.78
36	28.08	28.95	64.28	3.58	0.00	31.39	0.28	7.25	0.00	1.97	0.00	1.70
37	28.28	28.96	64.27	3.60	0.00	31.46	0.28	7.26	0.00	1.98	0.00	1.26
38	35.43	29.02	63.90	4.15	0.00	34.11	0.35	7.83	0.00	2.48	0.00	1.08
39	41.05	29.06	63.64	4.54	0.00	36.19	0.41	8.28	0.00	2.87	0.00	0.77
40	41.38	29.06	63.63	4.56	0.00	36.31	0.41	8.31	0.00	2.90	0.00	0.44
41	41.52	29.06	63.62	4.57	0.00	36.36	0.42	8.32	0.00	2.91	0.00	0.29
42	46.55	29.09	63.42	4.87	0.00	38.22	0.47	8.72	0.00	3.26	0.00	0.57
43	47.77	29.10	63.37	4.94	0.00	38.68	0.48	8.82	0.00	3.34	0.00	0.15
44	50.84	29.12	63.26	5.11	0.00	39.81	0.51	9.07	0.00	3.56	0.00	0.32
45	51.69	29.12	63.23	5.15	0.00	40.13	0.52	9.14	0.00	3.62	0.00	0.54
46	56.67	29.15	63.07	5.40	0.00	41.97	0.57	9.53	0.00	3.97	0.00	0.25
47	70.03	29.21	62.69	5.97	0.00	46.91	0.70	10.60	0.00	4.90	0.00	0.18
48	74.06	29.23	62.59	6.12	0.00	48.40	0.74	10.92	0.00	5.18	0.00	0.17
49	76.31	29.24	62.53	6.20	0.00	49.23	0.76	11.10	0.00	5.34	0.00	0.10
50	78.07	29.24	62.49	6.26	0.00	49.89	0.78	11.25	0.00	5.47	0.00	0.10
51	78.75	29.24	62.48	6.28	0.00	50.14	0.79	11.30	0.00	5.51	0.00	0.08
52	79.45	29.25	62.46	6.31	0.00	50.40	0.79	11.36	0.00	5.56	0.00	0.09
53	81.05	29.25	62.43	6.36	0.00	50.99	0.81	11.48	0.00	5.67	0.00	0.06
54	81.30	29.25	62.42	6.37	0.00	51.08	0.81	11.50	0.00	5.69	0.00	0.09
55	82.47	29.26	62.40	6.40	0.00	51.52	0.82	11.60	0.00	5.77	0.00	0.05
56	85.11	29.27	62.34	6.49	0.00	52.49	0.85	11.81	0.00	5.96	0.00	0.08
57	85.48	29.27	62.34	6.50	0.00	52.63	0.85	11.84	0.00	5.98	0.00	0.04

No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
58	85.92	29.27	62.33	6.51	0.00	52.79	0.86	11.87	0.00	6.01	0.00	0.03
59	86.93	29.27	62.31	6.54	0.00	53.16	0.87	11.95	0.00	6.08	0.00	0.09
60	87.59	29.27	62.29	6.56	0.00	53.41	0.88	12.01	0.00	6.13	0.00	0.03
61	88.98	29.28	62.27	6.60	0.00	53.92	0.89	12.12	0.00	6.23	0.00	0.10
62	89.43	29.28	62.26	6.61	0.00	54.09	0.89	12.15	0.00	6.26	0.00	0.07
63	90.08	29.28	62.25	6.63	0.00	54.33	0.90	12.21	0.00	6.31	0.00	0.07
64	91.48	29.29	62.22	6.67	0.00	54.85	0.91	12.32	0.00	6.40	0.00	0.07
65	92.49	29.29	62.20	6.70	0.00	55.22	0.92	12.40	0.00	6.47	0.00	0.15
66	92.56	29.29	62.20	6.70	0.00	55.25	0.93	12.41	0.00	6.48	0.00	0.10
67	93.32	29.29	62.19	6.72	0.00	55.53	0.93	12.47	0.00	6.53	0.00	0.16
68	93.81	29.29	62.18	6.74	0.00	55.71	0.94	12.51	0.00	6.57	0.00	0.16
69	96.55	29.30	62.13	6.81	0.00	56.72	0.97	12.72	0.00	6.76	0.00	0.07
70	97.17	29.30	62.12	6.82	0.00	56.95	0.97	12.77	0.00	6.80	0.00	0.03
71	97.60	29.30	62.11	6.84	0.00	57.11	0.98	12.81	0.00	6.83	0.00	0.11
72	97.88	29.30	62.10	6.84	0.00	57.22	0.98	12.83	0.00	6.85	0.00	0.06

Herrwerth SAMs[20]**:**

No.	Structure	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
73	$HS-(CH_2)_{11}-(O-CH_2-CH_2)-OMe$	29.79	63.83	4.26	0.00	47.00	0.00	11.00	0.00	4.00	0.00	0.44
74	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_2-OH$	29.41	62.75	5.88	0.00	51.00	1.00	12.00	0.00	5.00	0.00	0.00
75	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_2-OMe$	29.63	62.96	5.56	0.00	54.00	0.00	12.00	0.00	6.00	0.00	0.00
76	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OH$	29.31	62.07	6.90	0.00	58.00	1.00	13.00	0.00	7.00	0.00	0.00
77	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OMe$	29.51	62.30	6.56	0.00	61.00	0.00	13.00	0.00	8.00	0.00	0.00
78	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OEt$	29.69	62.50	6.25	0.00	64.00	0.00	14.00	0.00	8.00	0.00	1.20
79	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OPr$	29.85	62.69	5.97	0.00	67.00	0.00	15.00	0.00	8.00	0.00	1.58
80	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OBu$	30.00	62.86	5.71	0.00	70.00	0.00	16.00	0.00	8.00	0.00	1.54
81	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OH$	29.11	60.76	8.86	0.00	79.00	1.00	16.00	0.00	13.00	0.00	0.00
82	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OMe$	29.27	60.98	8.54	0.00	82.00	0.00	16.00	0.00	14.00	0.00	0.00
83	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OEt$	29.41	61.18	8.24	0.00	85.00	0.00	17.00	0.00	14.00	0.00	1.02
84	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OPr$	29.55	61.36	7.95	0.00	88.00	0.00	18.00	0.00	14.00	0.00	1.38
85	$HS-(CH_2)_{11}-(O-CH(CH_3)-CH_2)_2-OMe$	30.00	63.33	5.00	0.00	60.00	0.00	14.00	0.00	6.00	0.00	1.04
86	$HS-(CH_2)_{11}-(O-CH(CH_3)-CH_2)_3-OMe$	30.00	62.86	5.71	0.00	70.00	0.00	16.00	0.00	8.00	0.00	0.98
87	$HS-(CH_2)_{11}-(O-CH(CH_3)-CH_2)_4-OMe$	30.00	62.50	6.25	0.00	80.00	0.00	18.00	0.00	10.00	0.00	0.84
88	$HS-(CH_2)_{11}-(O-CH_2-CH_2-CH_2)_3-OH$	29.85	62.69	5.97	0.00	67.00	1.00	16.00	0.00	7.00	0.00	0.00
89	$HS-(CH_2)_{11}-(O-CH_2-CH_2-CH_2)_3-OMe$	30.00	62.86	5.71	0.00	70.00	0.00	16.00	0.00	8.00	0.00	0.00

Ostuni SAMs [24]**:**

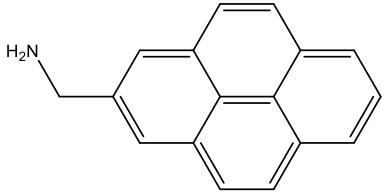


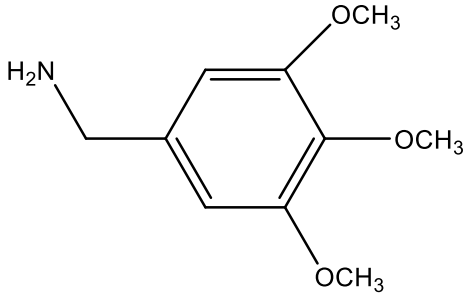
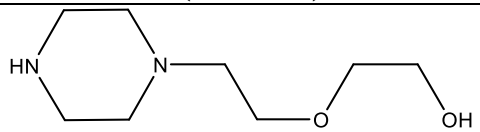
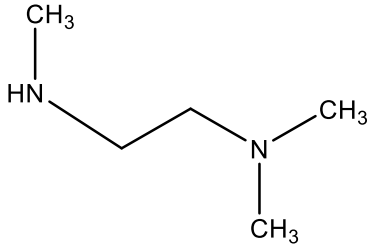
Schematic of Mixed SAMs in paper published by Ostuni, et. al.

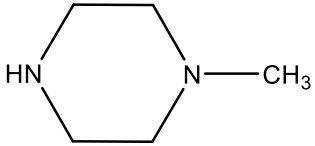
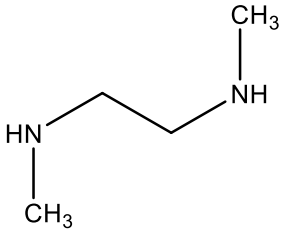
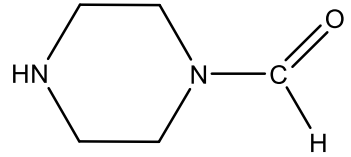
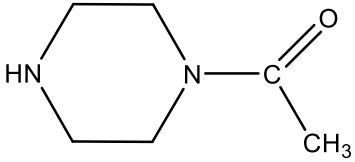
The SAM is a 1:1 mixture of -CONR'R and -CO_2H/CO_2^- groups

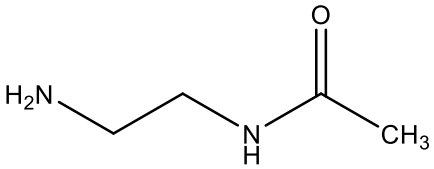
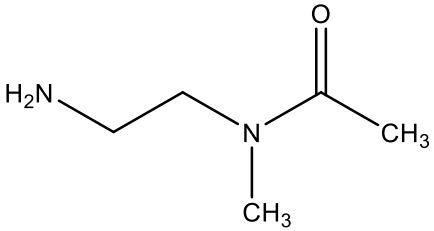
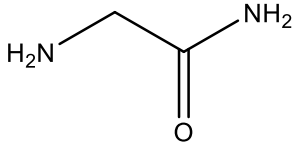
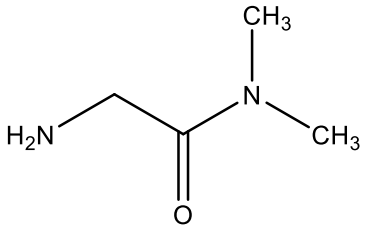
The -CONR'R part of the SAMs were synthesized by reacting the A SAMs

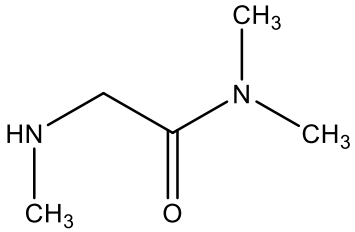
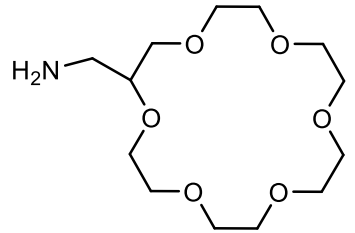
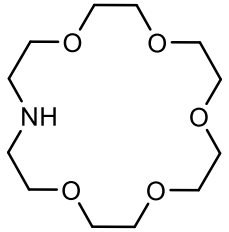
with amines of the form HNR'R[24]

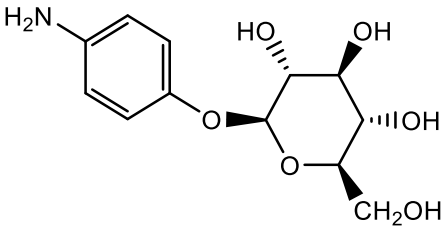
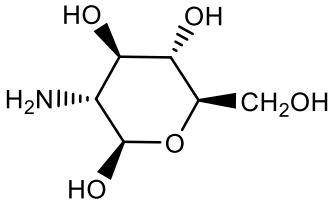
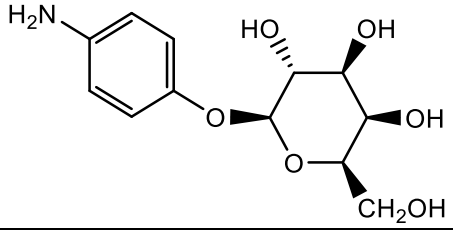
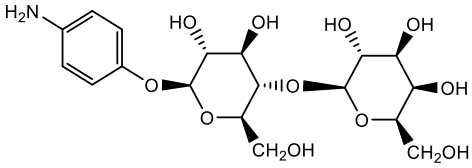
No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
90	$H_2N(CH_2)_{10}CH_3$	31.62	63.97	2.21	0.74	68.00	0.50	20.00	1.00	0.50	1.00	2.00
91		37.69	57.69	2.31	0.77	65.00	0.50	25.00	1.00	0.50	1.00	2.12

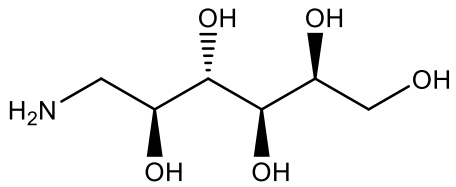
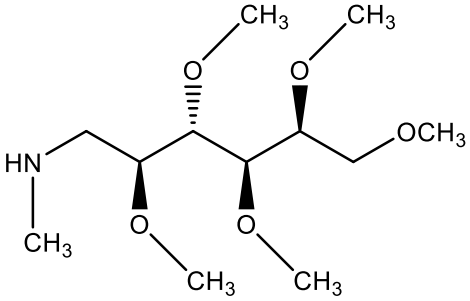
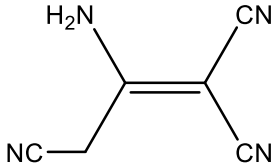
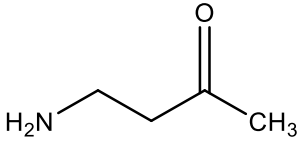
No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
92		32.81	60.16	4.69	0.78	64.00	0.50	18.50	1.00	3.50	1.00	1.50
93	$H_2NCH_2CH_2OCH_3$	30.97	62.83	3.54	0.88	56.50	0.50	15.50	1.00	1.50	1.00	1.06
94	$H_2NCH_2CH_2OH$	30.91	62.73	3.64	0.91	55.00	1.00	15.50	1.00	1.00	1.00	0.94
95	$H(NCH_2CH_2OCH_3)_2$	30.89	62.60	4.07	0.81	61.50	0.50	16.00	1.50	2.50	1.00	1.00
96	$H_2N(CH_2CH_2O)_3CH_3$	30.71	62.20	4.72	0.79	63.50	0.50	16.50	1.00	3.50	1.00	0.03
97	$H_2N(CH_2CH_2O)_3H$	30.65	62.10	4.84	0.81	62.00	1.00	16.50	1.00	3.00	1.00	0.03
98	$H_2N(CH_2CH_2O)_6CH_3$	30.41	61.49	6.08	0.68	74.00	0.50	18.00	1.00	6.50	1.00	0.01
99	$H_2N(CH_2CH_2O)_6H$	30.34	61.38	6.21	0.69	72.50	1.00	18.00	1.00	6.00	1.00	0.01
100		31.01	62.02	3.88	1.55	64.50	1.00	17.00	3.00	2.00	1.00	0.07
101		30.83	63.33	2.50	1.67	60.00	0.50	15.50	3.00	0.50	1.00	0.80

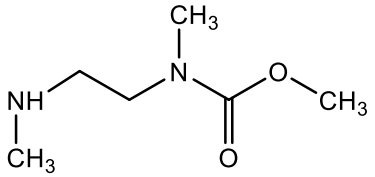
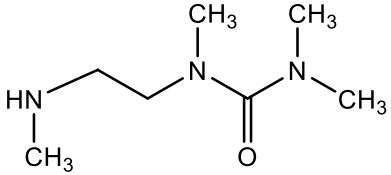
No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
102	$H_2N(CH_2CH_2O)_2CH_2CH_2NH_2$	30.40	62.40	4.00	1.60	62.50	0.50	16.50	1.50	2.50	1.00	0.74
103		31.36	62.71	2.54	1.69	59.00	0.50	16.00	3.00	0.50	1.00	0.32
104		30.77	63.25	2.56	1.71	58.50	0.50	15.50	2.50	0.50	1.00	0.17
105	$HN(CH_3)_2$	31.19	63.30	2.75	0.92	54.50	0.50	15.00	1.50	0.50	1.00	1.08
106		31.62	61.54	3.42	1.71	58.50	0.50	16.00	3.00	0.50	1.50	0.78
107		31.67	61.67	3.33	1.67	60.00	0.50	16.50	3.00	0.50	1.50	0.76

No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
108		31.03	62.07	3.45	1.72	58.00	0.50	16.00	2.00	0.50	1.50	0.80
109		31.15	62.30	3.28	1.64	61.00	0.50	16.00	3.00	0.50	1.50	0.24
110		30.91	61.82	3.64	1.82	55.00	0.50	15.50	1.50	0.50	1.50	1.16
111	 $H_2N(Gly)_1N(CH_3)_2$	31.03	62.07	3.45	1.72	58.00	0.50	15.50	2.50	0.50	1.50	0.66
112	$H_2N(Gly)_3N(CH_3)_2$	31.75	60.32	4.76	1.59	63.00	0.50	17.50	2.50	0.50	2.50	0.44

No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
113	 $H(CH_3)N(Sar)_1N(CH_3)_2$	31.09	62.18	3.36	1.68	59.50	0.50	15.50	3.00	0.50	1.50	0.18
114	$H(CH_3)N(Sar)_3N(CH_3)_2$	31.78	60.47	4.65	1.55	64.50	0.50	17.50	3.00	0.50	2.50	0.04
115	$H(CH_3)N(Sar)_4N(CH_3)_2$	32.09	59.70	5.22	1.49	67.00	0.50	18.50	3.00	0.50	3.00	0.03
116	$H(CH_3)N(Sar)_5N(CH_3)_2$	32.37	58.99	5.76	1.44	69.50	0.50	19.50	3.00	0.50	3.50	0.03
117		30.82	60.96	6.16	0.68	73.00	0.50	18.50	1.00	6.50	1.00	1.18
118		30.99	61.27	5.63	0.70	71.00	0.50	18.00	1.50	5.50	1.00	0.22

No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
119		32.59	58.52	6.67	0.74	67.50	2.50	20.50	1.00	4.50	1.00	1.34
120		30.65	60.48	6.45	0.81	62.00	2.50	17.50	1.00	3.50	1.00	1.10
121		32.59	58.52	6.67	0.74	67.50	2.50	20.50	1.00	4.50	1.00	0.98
122		31.85	56.69	9.55	0.64	78.50	4.00	23.00	1.00	8.00	1.00	0.64

123		30.16	61.11	6.35	0.79	63.00	3.00	17.50	1.00	3.00	1.00	1.36
No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
124		30.56	61.81	5.56	0.69	72.00	0.50	17.50	1.50	5.50	1.00	0.06
125		33.63	58.41	2.65	3.54	56.50	0.50	17.00	3.00	0.00	1.50	1.16
126	$HN(CH_2CH_2CN)_2$	32.48	60.68	2.56	2.56	58.50	0.50	17.00	2.50	0.50	1.00	1.16
127	$HN(CH_2CN)_2$	32.43	60.36	2.70	2.70	55.50	0.50	16.00	2.50	0.50	1.00	1.46
128	$H_2NCH_2CH_2CN$	31.82	61.82	2.73	1.82	55.00	0.50	16.00	1.50	0.50	1.00	1.44
129		31.58	62.28	3.51	0.88	57.00	0.50	16.50	1.00	0.50	1.50	1.48
130	$H_2NC(CH_2CH_2CH_2OH)_3$	30.88	62.50	4.41	0.74	68.00	2.00	19.50	1.00	2.00	1.00	1.14

No.	HNR'R	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
131		30.89	61.79	4.07	1.63	61.50	0.50	15.50	3.00	1.50	1.50	1.16
132	$H(CH_3)NCH_2CH(OCH_3)_2$	31.03	62.93	3.45	0.86	58.00	0.50	15.50	1.50	1.50	1.00	0.72
133		30.71	62.20	3.15	2.36	63.50	0.50	15.50	4.50	0.50	1.50	0.50

Appendix 3

Single-lab data

Hayashi SAMs[19]

No.	Structure	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
1	$HS-(CH_2)_7-CH_3$	29.63	66.67	0.00	0.00	27.00	0.00	7.00	0.00	0.00	0.00	2.00
2	$HS-(CH_2)_{11}-OH$	29.73	64.86	2.70	0.00	37.00	1.00	10.00	0.00	1.00	0.00	0.58
3	$HS-(CH_2)_{11}-NH_2$	28.95	65.79	0.00	2.63	38.00	0.00	10.00	1.00	0.00	0.00	1.80
4	$HS-(CH_2)_{11}-COOH$	30.77	61.54	5.13	0.00	39.00	1.00	11.00	0.00	1.00	1.00	1.70
5	$HS-(CH_2)_{11}-(O-CH_2-CH_2)-OH$	29.55	63.64	4.55	0.00	44.00	1.00	11.00	0.00	3.00	0.00	0.10
6	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_2-OH$	29.41	62.75	5.88	0.00	51.00	1.00	12.00	0.00	5.00	0.00	0.08
7	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-OH$	29.31	62.07	6.90	0.00	58.00	1.00	13.00	0.00	7.00	0.00	0.00
8	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-O-CH_3$	29.51	62.30	6.56	0.00	61.00	0.00	13.00	0.00	8.00	0.00	0.01
9	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-NH_2$	28.81	62.71	5.08	1.69	59.00	0.00	13.00	1.00	6.00	0.00	1.40
10	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_3-COOH$	30.00	60.00	8.33	0.00	60.00	1.00	14.00	0.00	7.00	1.00	1.30
11	$HS-(CH_2)_{11}-(O-CH_2-CH_2)_6-OH$	29.11	60.76	8.86	0.00	79.00	1.00	16.00	0.00	13.00	0.00	0.00

Sekine SAMs[21]**:**

SAM1	**$HS(CH_2)_5CH_3$**											
SAM2	**$HS\text{-}(CH_2)_{11}\text{-}(O\text{-}\ CH_2\text{-}\ CH_2)_3\text{-}OH$**											
No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
12	4.43	28.66	66.14	0.78	0.00	22.64	0.04	5.35	0.00	0.31	0.00	2.58
13	5.29	28.67	66.05	0.92	0.00	22.96	0.05	5.42	0.00	0.37	0.00	2.60
14	5.46	28.67	66.03	0.95	0.00	23.02	0.05	5.44	0.00	0.38	0.00	2.51
15	5.85	28.68	65.99	1.01	0.00	23.16	0.06	5.47	0.00	0.41	0.00	2.73
16	6.30	28.69	65.95	1.08	0.00	23.33	0.06	5.50	0.00	0.44	0.00	2.66
17	6.35	28.69	65.94	1.09	0.00	23.35	0.06	5.51	0.00	0.44	0.00	2.63
18	6.93	28.70	65.88	1.18	0.00	23.56	0.07	5.55	0.00	0.48	0.00	2.31
19	7.55	28.71	65.82	1.27	0.00	23.80	0.08	5.60	0.00	0.53	0.00	2.62
20	8.19	28.72	65.76	1.36	0.00	24.03	0.08	5.66	0.00	0.57	0.00	1.87
21	9.55	28.74	65.63	1.56	0.00	24.53	0.10	5.76	0.00	0.67	0.00	2.02
22	10.17	28.75	65.57	1.64	0.00	24.76	0.10	5.81	0.00	0.71	0.00	2.62
23	10.37	28.75	65.55	1.67	0.00	24.84	0.10	5.83	0.00	0.73	0.00	2.24
24	10.95	28.76	65.50	1.75	0.00	25.05	0.11	5.88	0.00	0.77	0.00	2.38
25	12.64	28.78	65.35	1.97	0.00	25.68	0.13	6.01	0.00	0.88	0.00	2.81
26	13.20	28.79	65.31	2.04	0.00	25.88	0.13	6.06	0.00	0.92	0.00	1.94
27	13.42	28.79	65.29	2.07	0.00	25.97	0.13	6.07	0.00	0.94	0.00	1.97
28	15.45	28.82	65.12	2.31	0.00	26.72	0.15	6.24	0.00	1.08	0.00	2.09
29	16.54	28.83	65.04	2.44	0.00	27.12	0.17	6.32	0.00	1.16	0.00	2.03
30	16.80	28.84	65.02	2.47	0.00	27.22	0.17	6.34	0.00	1.18	0.00	2.07
31	18.29	28.85	64.91	2.63	0.00	27.77	0.18	6.46	0.00	1.28	0.00	1.62
32	20.64	28.88	64.74	2.88	0.00	28.64	0.21	6.65	0.00	1.44	0.00	1.75

No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
33	23.23	28.91	64.57	3.14	0.00	29.59	0.23	6.86	0.00	1.63	0.00	1.93
34	24.99	28.93	64.46	3.31	0.00	30.25	0.25	7.00	0.00	1.75	0.00	1.68
35	25.21	28.93	64.45	3.33	0.00	30.33	0.25	7.02	0.00	1.77	0.00	1.78
36	28.08	28.95	64.28	3.58	0.00	31.39	0.28	7.25	0.00	1.97	0.00	1.70
37	28.28	28.96	64.27	3.60	0.00	31.46	0.28	7.26	0.00	1.98	0.00	1.26
38	35.43	29.02	63.90	4.15	0.00	34.11	0.35	7.83	0.00	2.48	0.00	1.08
39	41.05	29.06	63.64	4.54	0.00	36.19	0.41	8.28	0.00	2.87	0.00	0.77
40	41.38	29.06	63.63	4.56	0.00	36.31	0.41	8.31	0.00	2.90	0.00	0.44
41	41.52	29.06	63.62	4.57	0.00	36.36	0.42	8.32	0.00	2.91	0.00	0.29
42	46.55	29.09	63.42	4.87	0.00	38.22	0.47	8.72	0.00	3.26	0.00	0.57
43	47.77	29.10	63.37	4.94	0.00	38.68	0.48	8.82	0.00	3.34	0.00	0.15
44	50.84	29.12	63.26	5.11	0.00	39.81	0.51	9.07	0.00	3.56	0.00	0.32
45	51.69	29.12	63.23	5.15	0.00	40.13	0.52	9.14	0.00	3.62	0.00	0.54
46	56.67	29.15	63.07	5.40	0.00	41.97	0.57	9.53	0.00	3.97	0.00	0.25
47	70.03	29.21	62.69	5.97	0.00	46.91	0.70	10.60	0.00	4.90	0.00	0.18
48	74.06	29.23	62.59	6.12	0.00	48.40	0.74	10.92	0.00	5.18	0.00	0.17
49	76.31	29.24	62.53	6.20	0.00	49.23	0.76	11.10	0.00	5.34	0.00	0.10
50	78.07	29.24	62.49	6.26	0.00	49.89	0.78	11.25	0.00	5.47	0.00	0.10
51	78.75	29.24	62.48	6.28	0.00	50.14	0.79	11.30	0.00	5.51	0.00	0.08
52	79.45	29.25	62.46	6.31	0.00	50.40	0.79	11.36	0.00	5.56	0.00	0.09
53	81.05	29.25	62.43	6.36	0.00	50.99	0.81	11.48	0.00	5.67	0.00	0.06
54	81.30	29.25	62.42	6.37	0.00	51.08	0.81	11.50	0.00	5.69	0.00	0.09
55	82.47	29.26	62.40	6.40	0.00	51.52	0.82	11.60	0.00	5.77	0.00	0.05
56	85.11	29.27	62.34	6.49	0.00	52.49	0.85	11.81	0.00	5.96	0.00	0.08
57	85.48	29.27	62.34	6.50	0.00	52.63	0.85	11.84	0.00	5.98	0.00	0.04

No.	Percentage of SAM2 (%)	%C (%)	%H (%)	%O (%)	%N (%)	nAT	O-H Bond	C-C Bond	C-N Bond	C-O Bond	C=O Bond	Fibrinogen Adsorption
58	85.92	29.27	62.33	6.51	0.00	52.79	0.86	11.87	0.00	6.01	0.00	0.03
59	86.93	29.27	62.31	6.54	0.00	53.16	0.87	11.95	0.00	6.08	0.00	0.09
60	87.59	29.27	62.29	6.56	0.00	53.41	0.88	12.01	0.00	6.13	0.00	0.03
61	88.98	29.28	62.27	6.60	0.00	53.92	0.89	12.12	0.00	6.23	0.00	0.10
62	89.43	29.28	62.26	6.61	0.00	54.09	0.89	12.15	0.00	6.26	0.00	0.07
63	90.08	29.28	62.25	6.63	0.00	54.33	0.90	12.21	0.00	6.31	0.00	0.07
64	91.48	29.29	62.22	6.67	0.00	54.85	0.91	12.32	0.00	6.40	0.00	0.07
65	92.49	29.29	62.20	6.70	0.00	55.22	0.92	12.40	0.00	6.47	0.00	0.15
66	92.56	29.29	62.20	6.70	0.00	55.25	0.93	12.41	0.00	6.48	0.00	0.10
67	93.32	29.29	62.19	6.72	0.00	55.53	0.93	12.47	0.00	6.53	0.00	0.16
68	93.81	29.29	62.18	6.74	0.00	55.71	0.94	12.51	0.00	6.57	0.00	0.16
69	96.55	29.30	62.13	6.81	0.00	56.72	0.97	12.72	0.00	6.76	0.00	0.07
70	97.17	29.30	62.12	6.82	0.00	56.95	0.97	12.77	0.00	6.80	0.00	0.03
71	97.60	29.30	62.11	6.84	0.00	57.11	0.98	12.81	0.00	6.83	0.00	0.11
72	97.88	29.30	62.10	6.84	0.00	57.22	0.98	12.83	0.00	6.85	0.00	0.06

References

1. Shao, Y., et al., *Graphene based electrochemical sensors and biosensors: a review.* Electroanalysis: An International Journal Devoted to Fundamental and Practical Aspects of Electroanalysis, 2010. **22**(10): p. 1027-1036.
2. Wink, T., et al., *Self-assembled monolayers for biosensors.* Analyst, 1997. **122**(4): p. 43R-50R.
3. Ramakrishna, S., et al., *Biomedical applications of polymer-composite materials: a review.* Composites science and technology, 2001. **61**(9): p. 1189-1224.
4. Zhang, S., G. Wright, and Y. Yang, *Materials and techniques for electrochemical biosensor design and construction.* Biosensors and Bioelectronics, 2000. **15**(5-6): p. 273-282.
5. Xue, L. and H.P. Greisler, *Biomaterials in the development and future of vascular grafts.* Journal of vascular surgery, 2003. **37**(2): p. 472-480.
6. Saini, M., et al., *Implant biomaterials: A comprehensive review.* World Journal of Clinical Cases: WJCC, 2015. **3**(1): p. 52.
7. Le Guéhennec, L., et al., *Surface treatments of titanium dental implants for rapid osseointegration.* Dental materials, 2007. **23**(7): p. 844-854.
8. Elias, C.N., et al., *Relationship between surface properties (roughness, wettability and morphology) of titanium and dental implant removal torque.* Journal of the mechanical behavior of biomedical materials, 2008. **1**(3): p. 234-242.
9. Horbett, T.A., *Principles underlying the role of adsorbed plasma proteins in blood interactions with foreign materials.* Cardiovascular Pathology, 1993. **2**(3): p. 137-148.
10. Giovambattista, N., P.G. Debenedetti, and P.J. Rossky, *Effect of surface polarity on water contact angle and interfacial hydration structure.* J Phys Chem B, 2007. **111**(32): p. 9581-7.
11. Chapman, R.G., et al., *Surveying for surfaces that resist the adsorption of proteins.* Journal of the American Chemical Society, 2000. **122**(34): p. 8303-8304.
12. Rajan, K., *Materials informatics: The materials "gene" and big data.* Annual Review of Materials Research, 2015. **45**: p. 153-169.
13. Liu, Y., et al., *Materials discovery and design using machine learning.* Journal of Materiomics, 2017. **3**(3): p. 159-177.
14. Hu, X.S., S.E. Li, and Y.L. Yang, *Advanced Machine Learning Approach for Lithium-Ion Battery State Estimation in Electric Vehicles.* Ieee Transactions on Transportation Electrification, 2016. **2**(2): p. 140-149.
15. Dey, P., et al., *Informatics-aided bandgap engineering for solar materials.* Computational Materials Science, 2014. **83**: p. 185-195.
16. Shandiz, M.A. and R. Gauvin, *Application of machine learning methods for the prediction of crystal system of cathode materials in lithium-ion batteries.* Computational Materials Science, 2016. **117**: p. 270-278.
17. Fujimura, K., et al., *Accelerated Materials Design of Lithium Superionic Conductors Based on First‐Principles Calculations and Machine Learning Algorithms.* Advanced Energy Materials, 2013. **3**(8): p. 980-985.
18. Love, J.C., et al., *Self-assembled monolayers of thiolates on metals as a form of nanotechnology.* Chemical reviews, 2005. **105**(4): p. 1103-1170.
19. Hayashi, T., et al., *Mechanism underlying bioinertness of self-assembled monolayers of oligo(ethyleneglycol)-terminated alkanethiols on gold: protein adsorption, platelet adhesion, and surface forces.* Phys Chem Chem Phys, 2012. **14**(29): p. 10196-206.

20. Herrwerth, S., et al., *Factors that determine the protein resistance of oligoether self-assembled monolayers− internal hydrophilicity, terminal hydrophilicity, and lateral packing density.* 2003. **125**(31): p. 9359-9366.
21. Sekine, T., et al., *Evaluation of factors to determine platelet compatibility by using self-assembled monolayers with a chemical gradient.* Langmuir, 2015. **31**(25): p. 7100-7105.
22. Arima, Y. and H. Iwata, *Effect of wettability and surface functional groups on protein adsorption and cell adhesion using well-defined mixed self-assembled monolayers.* Biomaterials, 2007. **28**(20): p. 3074-82.
23. Dobrzanska, D.A., et al., *Oxidation of tertiary amine-derivatized surfaces to control protein adhesion.* Langmuir, 2013. **29**(9): p. 2961-70.
24. Chapman, Robert G., et al. "Surveying for surfaces that resist the adsorption of proteins." *Journal of the American Chemical Society*,2000. 122.**34**: p. 8303-8304.
25. Bain, C.D., H.A. Biebuyck, and G.M. Whitesides, *Comparison of self-assembled monolayers on gold: coadsorption of thiols and disulfides.* Langmuir, 1989. **5**(3): p. 723-727.
26. Zhu, B., et al., *Chain‐length dependence of the protein and cell resistance of oligo (ethylene glycol)‐terminated self‐assembled monolayers on gold.* Journal of Biomedical Materials Research: An Official Journal of The Society for Biomaterials, The Japanese Society for Biomaterials, and The Australian Society for Biomaterials and the Korean Society for Biomaterials, 2001. **56**(3): p. 406-416.
27. Folkers, J.P., P.E. Laibinis, and G.M. Whitesides, *Self-assembled monolayers of alkanethiols on gold: the adsorption and wetting properties of monolayers derived from two components with alkane chains of different lengths.* Journal of adhesion science and technology, 1992. **6**(12): p. 1397-1410.
28. Abe, K., H. Takiguchi, and K. Tamada, *Dynamic contact angle measurement of Au (111)− Thiol self-assembled monolayers by the Wilhelmy plate method.* Langmuir, 2000. **16**(5): p. 2394-2397.
29. Efron, B., *The jackknife, the bootstrap, and other resampling plans*. Vol. 38. 1982: Siam.
30. Olden, J.D., M.K. Joy, and R.G. Death, *An accurate comparison of methods for quantifying variable importance in artificial neural networks using simulated data.* Ecological Modelling, 2004. **178**(3-4): p. 389-397.
31. de Oña, J. and C. Garrido, *Extracting the contribution of independent variables in neural network models: a new approach to handle instability.* Neural Computing and Applications, 2014. **25**(3-4): p. 859-869.
32. Dobrzanska, D.A., et al., *Oxidation of Tertiary Amine-Derivatized Surfaces To Control Protein Adhesion.* Langmuir, 2013. **29**(9): p. 2961-2970.