

Deep Learning Analysis of binding behavior of virus displayed peptides to AuNPs

Haebom Lee¹, Jun Jo¹, Yong Oh Lee¹,
Nuriye Korkmaz Zirpel¹, and Leon Abelmann^{1,2}

¹ KIST Europe, Saarbrücken, Germany

² University of Twente, Enschede, The Netherlands

Abstract. Filamentous fd viruses have been used as biotemplates to develop nano sized carriers for biomedical applications. Genetically modified fd viruses with enhanced gold binding properties have been previously obtained by displaying gold binding peptides on viral coat proteins. In order to generate a stable colloidal system of dispersed viruses decorated with AuNPs avoiding aggregation, the underlying binding mechanism of AuNP-peptide interaction should be explored. In this paper, we therefore propose a macro scale self-assembly experiment using 3D printed models of AuNP and the virus to extend our understanding of Au binding process. Moreover, we present our image analysis algorithm which combines image processing techniques and deep learning to automatically examine the coupling state of the particles.

Keywords: Deep learning, Genetic engineering, Phage display, Gold

1 Introduction

Non-toxic and cost effective stable carrier systems are important for drug delivery and cancer therapy applications. Although AuNPs are able to be synthesized chemically using reducing agents with defined geometries and sizes, the functionalization of particles in order to obtain targetable agents still remains as a challenge. Until now AuNPs have been functionalized with antibodies, biotin molecules or cell targeting moieties using chemical reaction processes [3]. An alternative way of functionalization avoiding chemical modifications is to use specific Au binding moieties like short peptides which would selectively bind Au. Au binding peptides can be tagged with cell targeting sequences and the resulting AuNP-peptide-targeting unit complex can be used for drug delivery, cancer therapy or imaging studies [7]. Au binding peptides or in general material specific peptide sequences can be identified by phage display method through biopanning where 3-5 rounds of selection are performed before the identification of binding sequences [2, 5].

After identification of peptides, it is crucial to test their functionalities first. Specific and selective binding of the peptides to targeted objects such as AuNPs or Au surfaces can be confirmed using quartz crystal microbalance measurements, electron microscopy analyses or spectrophotometric techniques. We have

previously generated genetically engineered fd viruses expressing various short peptide sequences and investigated their Au binding affinities using biochemical and microscopical analyses [6, 10]. Genetically modified viruses showed enhanced Au binding properties. These viruses were tested as potential biotemplates for nanowire synthesis taking the advantage of improved Au binding characteristics. However, to obtain single dispersed virus templated nanowires was challenging due to the aggregation problem which we face after the metallization process. In order to control the aggregation of viral particles during Au binding process, we need to understand how they interact with the particles and which factors may affect the binding behavior.

In order to obtain a cost effective, well defined peptide functionalized nanoparticle based system, it is extremely important to understand the complete binding mechanism of these peptides to Au surfaces. Until now, molecular dynamic (MD) studies using Monte Carlo simulations have been performed in order to explore the Au binding mechanism of peptides and which factors would potentially affect the attachment [1]. Classical MD analyses do not include entropy changes. Generally, enthalpy changes were considered for calculations not the Gibbs free energies. Moreover, MD simulation mostly deal with atomic level interactions.

In this paper, we propose a new experimental approach with big data analysis using deep neural networks. The major advantage of using deep learning rather than analogue MD simulations is that no information will be lost as the complexity of the analyzed system is increased. First, macro scale self-assembly experiments will be conducted using 3D printed scaled-up virus and AuNP mimicking solid models. AuNP-peptide molecular interactions are simulated using magnetic forces in liquid. This will enable us to visualize a reflected 3D system where virus filaments can move freely in a liquid system representing the wet experimental conditions. Using the macro scale self-assembly reactor chamber, binding patterns of AuNPs to virus displayed peptides can be evaluated at different conditions including various AuNP and virus concentration, peptide display position, number of displayed peptides and temperature.

Images recorded during the macro scale self-assembly experiments will be analyzed using deep machine learning techniques in order to reveal the underlying Au binding mechanism of Au binding peptides. Moreover, the most optimal binding condition will be obtained which will be further applied in real-time experiments for drug delivery and targeting studies. For the presented work here, we used a self-assembly system incorporating one virus template and one AuNP. To best our knowledge, this is the first attempt to apply deep machine learning for analyzing binding mechanism of recombinantly expressed Au specific peptides to AuNPs.

The contributions of this paper follow: (1) we propose novel self-assembly experiment with scaled up particle design and 3D printing, and (2) we show the capability of the proposed deep neural network for Au binding pattern analysis.

The paper is organized as follows: (1) self-assembly experiment (2) binding pattern analysis method (3) experiments (4) conclusion and future work.

2 Self-Assembly Experiment

To investigate and simulate the AuNP-virus interaction, which is induced by Au binding peptides displayed on virus coat proteins (Fig. 1a), a 3D self-assembly experiment is designed using scaled-up solid models of virus and Au particles employing magnetic intermolecular forces.

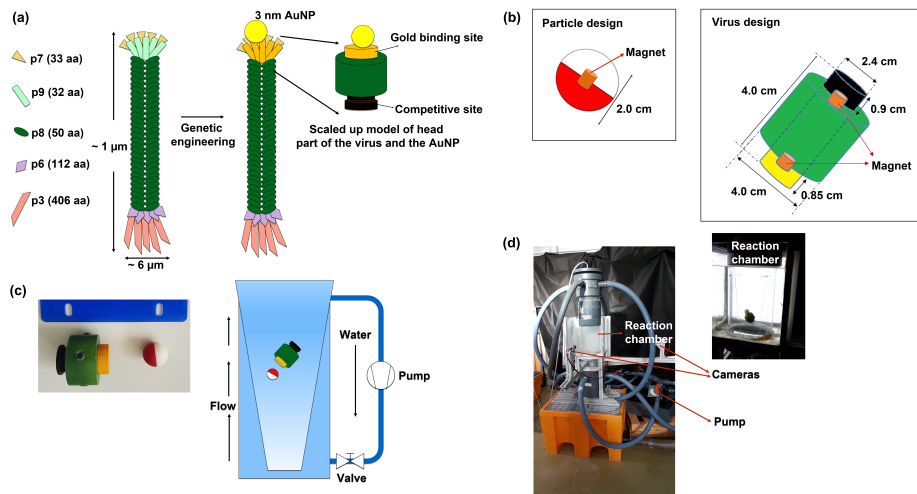


Fig. 1. (a) Schematic drawing of an fd filamentous virus composed of a single stranded circular DNA enclosed in a protein cage. Head part of the virus is genetically modified to display Au binding peptides. The part is considered as a cylindrical object in our 3D scaled-up model. (b) Yellow part stands for the Au binding site, whereas black part indicates the competitive binding site with less binding energy. (c) 3D printed models of viral template and the AuNP. Solid objects are inserted into (d) the self-assembly reactor chamber and recorded using two cameras under turbulent flow.

Fig. 1b shows the simplified designs of virus and Au particle. Note that there are holes for magnets in order to simulate AuNP-peptide short range molecular interactions. Extended cylindrical part in yellow represents the head part of the fd virus where Au binding peptide is displayed. There is also an extruding cylindrical part in black as the competitive binding site. The center to center distance between the magnet in the black part and the AuNP mimicking spherical particle is arranged to be 19.0 mm when they are combined, which is 0.5 mm longer than the distance between the magnet of the yellow cylindrical part and the gold particle. As a result, we expect to have stronger binding of AuNP mimicking sphere on the yellow part. Particles are designed using a free 3D modeling tool "Openscad" (<http://www.openscad.org/>) and 3D printed using ABS polymer (Fig. 1c). Permanent magnets, that are axially magnetized cylindrical NdFeB magnets (5 mm in length and 4 mm in diameter, Supermagnete, grade N42, Webcraft GmbH), are glued inside each hole of the cylindrical subunits and assembled together.

Following the particle design and 3D printing, we organize a self-assembly trial experiment in a flow reactor chamber as illustrated in Fig. 1c-d. The macroscopic self-assembly reactor is based on that of Hageman et al. [4]. A water pump forms a turbulent flow against the gravitational force inside the conical reaction chamber and is introducing turbulence as the source of disturbing energy. We use a MAXI.2 40T pump moving water through four inlets into the cone shaped inset (Fig. 1c-d) keeping the objects in the center of the observation area. Two calibrated, synchronized cameras (Mako G-131, Allied Vision) are perpendicularly positioned at the front and right sides (Fig. 1d) of the reaction chamber for image recording.

3 Binding Pattern Analysis

In this section, we explain our novel method that combines image processing techniques and a convolutional neural network (CNN) for determining various binding patterns (see Fig. 2) in the images captured during the self-assembly experiment.

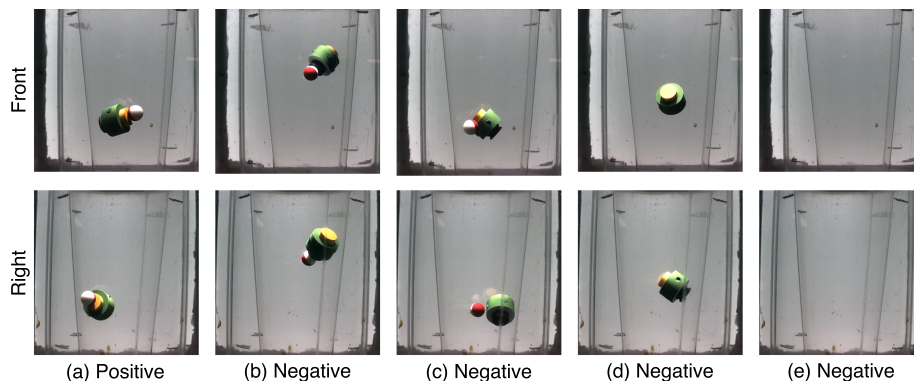


Fig. 2. Examples of analysis results for various input images: (a) two particles are combined at the Au binding site; (b) connected, but at the competitive binding site; (c) particles are separated; (d) and (e) one or no particle is captured in the image.

3.1 Image Processing

Image processing techniques for finding meaningful features in images are widely exploited in bioinformatics area [9]. Our analysis algorithm also employs image processing step to remove uninformative background and to extract meaningful region of interest (ROI) from the captured images. We first calculate the difference between a test image and a background image, and apply a threshold to obtain a black and white (BW) image. On that image, we run a naive shape-based analysis which detects and counts the number of connected components. This metric is critical for diagnosing two cases, as there must be two components

in either of front or right image if the particles are separated (Fig. 2c), and no components if no particle is captured (Fig. 2e).

As illustrated in Fig. 2, the analysis should recognize the case as positive, only when the red part of gold particle and the yellow part of virus are connected. This condition leads us to extract red and yellow pixels from the ROI and check whether they are combined in one component. Extracted image as well as aforementioned metrics are delivered to the CNN, to make the final decision. The entire image processing steps and their intermediate results are depicted in Fig 3.

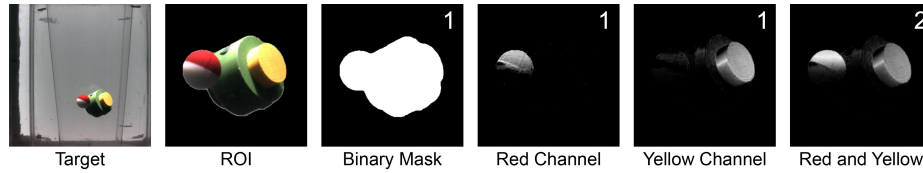


Fig. 3. The pipeline of our image processing procedure. From a target image, region of interest is extracted and further investigated through its binary image, red channel, yellow channel, and a combined image of red and yellow channels. Numbers on images indicate the number of connected components in each image. The rightmost combined image and four metrics are the input of the CNN.

3.2 Utilizing Synthetic Data

Images acquired from the self-assembly experiments should be labeled before being used as a training data for our CNN. Instead of conducting arduous manual classification, we exploit synthetic images that can be generated from a 3D game engine. Utilizing synthetic data to train a neural network is gaining popularity in modern deep learning applications [8] as it provides flexibility and control at the same time and can generate ground truth data with a lower cost.

Our aim is to achieve maximum similarity between the real and the synthetic data. For this reason, the 3D Openscad object design data is exploited again to construct a basic scene of the simulation. Then we applied arbitrary motions to the virtual particles. Furthermore, we locate the virtual lamps and cameras cautiously, to imitate our self-assembly experiment environment.

Fig. 4 shows the rendered synthetic images and corresponding image processing results. We generate thousands of synthetic images for various cases and use them to train and test our CNN classifier.



Fig. 4. Images generated from a 3D game engine and their subsequent image processing results.

3.3 Deep Learning

We extract several metrics from the input images and deliver them as an additional input to the network. For a pair of input images (front and right), we calculate four metrics from each image: the number of connected components in ROI’s BW, Red channel, Yellow channel, and Red+Yellow channel images.

Fig. 5 shows the overall architecture of our CNN. The network mainly consists of three convolution blocks and two fully-connected blocks, as well as flattening and concatenating layers that bridge the two types of blocks. Each convolution block is composed of a convolution layer with ReLU activation, followed by a max pooling layer, and a dropout layer. The first convolution layer filters the input of size $1280 \times 1024 \times 2$ using 16 kernels of size 10×10 with a horizontal stride of 10 pixels and a vertical stride of 8 pixels. The result image of size $128 \times 128 \times 16$ is then shrunk to $64 \times 64 \times 16$ in the following max pooling layer. The second convolution layer filters it using 32 kernels of size 3×3 with a stride of 2 pixels. Through a similar process, the input of the third convolution layer becomes of size $16 \times 16 \times 32$ and at the end of the block, the input is flattened to a vector of length 1024 from $4 \times 4 \times 64$. The vector is then concatenated with the second input, a vector of eight metrics, and delivered to a fully-connected layer of 512 neurons. Finally, the vector of length 512 goes through a sigmoid function to generate the output.

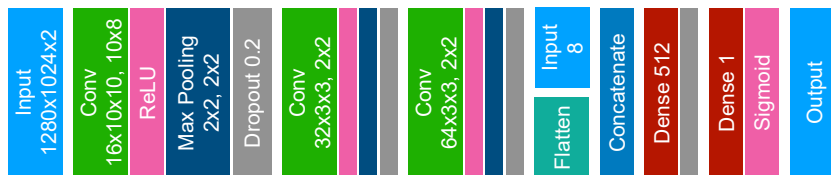


Fig. 5. Overall architecture of our CNN classifier. Layers with same color have the same setting unless otherwise stated.

4 Experiment

In the self-assembly experiment, we set an upward flow in the reactor chamber of $6.5 \pm 0.1 \text{ cm/s}$ using a valve setting with an equivalent energy of $10 \pm 2 \mu\text{J}$. Note that the actual disturbing energy in the cone is expected to be somewhat lower due to the external factors.

To measure the binding energy between the virus model and the gold particle, we exploit a dipole model as illustrated in Fig. 6 since the distance between the magnets is large compared to their size (see Fig. 1). When the sphere is connected to the yellow side of the cylinder, the distance between the magnets is $19.3 \pm 0.2 \text{ mm}$ and the energy is $-94 \pm 3 \mu\text{J}$. When it is connected to the thick black side, the distance is $19.8 \pm 0.2 \text{ mm}$ and the energy consequently becomes higher: $-87 \pm 3 \mu\text{J}$. The energy difference between the two states is $7 \pm 6 \mu\text{J}$.

We conducted the self-assembly experiment for 4 hours and recorded images of size 1280×1024 at 1 fps , leading to over 14,000 pairs of front and right

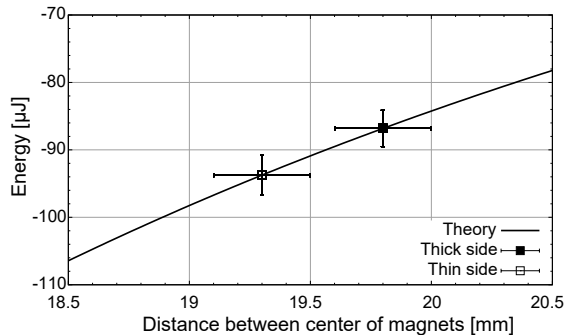


Fig. 6. Calculated energy between two cylindrical magnets of diameter 4 mm and height 5 mm as a function of the distance between the magnets. The two points indicated are the distances between the magnets when the sphere is on the thick (black) side or thin (yellow) side of the main cylinder.

images. We labeled 3,200 pairs of them in order to measure the performance of our classifier in various aspects. The averaged accuracies of classification results over 10-fold cross validation in different settings are presented in the table below. Specifically, we trained our CNN model in two different ways: training with a portion of labeled real data, training only with synthetic data. Then we evaluated accuracy of the classifier for the training set, the test set, and for the unseen real data. As expected, the classifier showed higher accuracy, 91.1%, on the labeled pairs when we trained it with real data. However, the potential of synthetic data based learning is also noticeable, as it achieved 87.7% accuracy on unseen real data. We also measured the accuracy of an image processing-based naive classifier for a comparison, 76.1% and 76.8% for synthetic and real data, respectively.

When the classifier was trained with real data, it analyzed about 22.5% of input data as positives, where the actual portion is 22.7%. Furthermore, the classifier collected 22.2% of unlabeled real data as positives. Since we can assume the unlabeled data has a similar portion of positives as the labeled, the result shows the consistency of our classifier.

	Training	Test	Real Data
Trained w/ Real Data	97.0%	97.0%	91.1%
Trained w/o Real Data	97.9%	97.8%	87.7%

5 Conclusion and Future Work

Considering increasing significance of particle-peptide based hybrid materials for a broad range of applications including biomedical and catalysis studies, it is important to understand the binding mechanism of peptides. To this end, we devised and conducted macro scale self-assembly experiments with 3D printed objects simulating a genetically modified virus displaying a gold binding peptide and a gold particle. In order to mimic molecular interactions, we inserted small

sized magnets into the objects and performed the self-assembly experiments in a reactor chamber under turbulent flow.

Images captured during the self-assembly experiment are then analyzed using our classifier which combines image processing technique and convolutional neural network. We also introduced how we utilized synthetic data to train the classifier in an efficient way. Various measurements over the classifier showed its capability as a reliable combined state analysis algorithm, by achieving 91.1% accuracy which is 14.3 p.p. better than a naive classifier.

In future works, we want to conduct further self-assembly experiments with increased number of Au particles. Moreover, we hope to extend our experimental approach in order to investigate effect of AuNP and virus concentration, peptide display position, number of displayed peptides, and temperature on the self-assembly process. As a consequence, we expect more sophisticated synthetic data generation and more flexible classifier in our future work.

References

1. Chiappini, M., Eiser, E., Sciortino, F.: Phase behaviour in complementary dna-coated gold nanoparticles and fd-viruses mixtures: a numerical study. *The European Physical Journal E* 40(1), 7 (2017)
2. Gagic, D., Ciric, M., Wen, W.X., Ng, F., Rakonjac, J.: Exploring the secretomes of microbes and microbial communities using filamentous phage display. *Frontiers in microbiology* 7, 429 (2016)
3. Giljohann, D.A., Seferos, D.S., Daniel, W.L., Massich, M.D., Patel, P.C., Mirkin, C.A.: Gold nanoparticles for biology and medicine. *Angewandte Chemie International Edition* 49(19), 3280–3294 (2010)
4. Hageman, T., L othman, P., Dirnberger, M., Elwenspoek, M., Manz, A., Abelmann, L.: Macroscopic equivalence for microscopic motion in a turbulence driven three-dimensional self-assembly reactor. *Journal of Applied Physics* 123(2), 024901 (2018)
5. Huang, J.X., Bishop-Hurley, S.L., Cooper, M.A.: Development of anti-infectives using phage display: biological agents against bacteria, viruses, and parasites. *Antimicrobial agents and chemotherapy* 56(9), 4569–4582 (2012)
6. Korkmaz, N.: Recombinant bacteriophages as gold binding bio-templates. *Colloids and Surfaces B: Biointerfaces* 112, 219–228 (2013)
7. Kumar, A., Ma, H., Zhang, X., Huang, K., Jin, S., Liu, J., Wei, T., Cao, W., Zou, G., Liang, X.J.: Gold nanoparticles functionalized with therapeutic and targeted peptides for cancer treatment. *Biomaterials* 33(4), 1180–1189 (2012)
8. Richter, S.R., Vineet, V., Roth, S., Koltun, V.: Playing for data: Ground truth from computer games. In: *European Conference on Computer Vision*. pp. 102–118. Springer (2016)
9. Teixeira-Castro, A., Dias, N., Rodrigues, P., Oliveira, J.F., Rodrigues, N.F., Maciel, P., Vilaça, J.L.: An image processing application for quantification of protein aggregates in *caenorhabditis elegans*. In: *5th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB 2011)*. pp. 31–38. Springer (2011)
10. Zirpel, N.K., Arslan, T., Lee, H.: Engineering filamentous bacteriophages for enhanced gold binding and metallization properties. *Journal of colloid and interface science* 454, 80–88 (2015)