Semi-rational Directed Evolution of Electrically Conductive Pili Based on Adaptive Particle Swarm Optimization Design

Abstract

The type IV pili of *Geobacter metallireducens* exhibit an exceptional electrical conductivity of up to 277 S/cm at pH 7, making them the most conductive biological nanomaterials known to date.The overlapping Π-Π orbitals of aromatic moieties imparts metallic conductivity to these revolutionary "green" nanomaterials.We build an antibody detection system, and the conductivity of this pili is the key to maintaining stability and accuracy of our system.We use optimization methods of mathematical modeling combined with protein directed evolution to attempt to optimize the structure of this pili to improve its conductivity.We find 5 possible mutation sites（V9, A20, S34, D54, T56）may improve the maximum conductivity of e-pili. Combing with an innovative screening method based on the microbial fuel cell system,we may eventually obtain a pili with higher conductivity.

Keywords: electrically conductive pili，semi-rational Directed Evolution，molecular dynamic simulation，particle Swarm Optimization,microbial fuel cell system

Introduction

Geobacter is a class of rod-shaped Gram-negative anaerobic bacteria belonging to the δ-Proteobacteria class [1,2]. They are important dissimilatory iron-reducing bacteria [3]. In order to survive in harsh anaerobic environments, certain species of *Geobacter* can utilize extracellular metal ions as electron acceptors for respiration, maintaining their own metabolism. This extracellular electron transfer process relies on a unique type of conductive appendage known as **"electrically conductive pili"（e-pili）**[4]. Electrons are transmitted through fibrous conductive structures on the cell surface, often with nanoscale diameters but lengths extending to the micrometer range, capable of facilitating electron transfer[5,6].

For instance, the type IV pili of ***Geobacter metallireducens*** exhibit an exceptional electrical conductivity of up to 277 S/cm at pH 7 [7], making them the most conductive biological nanomaterials known to date. Within these conductive bacterial nanowires, there is a high

density of aromatic groups, and **the overlapping Π-Π orbitals of aromatic moieties** imparts metallic conductivity to these organic materials [8].

This type of material is considered a **revolutionary "green" nanomaterial**. The energy required to produce these protein nanowires through microbial processes is over two orders of magnitude less than the energy needed to produce an equivalent mass of silicon or carbon nanowires. Moreover, the production process does not generate any toxic chemicals [9]. Additionally, this material is relatively robust while being biodegradable and exhibits excellent biocompatibility [10].

This year， we used this bio-nanomaterial derived from *Geobacter metallireducens***,** as the "variable resistor" in our disease biomarker detection system. Through our design, biomarkers can selectively bind to e-pili. This specific binding alters the charge distribution on the surface of the pili, consequently modifying their conductivity. By measuring the conductivity changes of the pili before and after binding and combining it with a standard curve, we can easily achieve quantitative detection.

In our system, the conductivity of the e-pili is the most important parameter. Higher conductivity allows our detection system to become more stable and accurate. Additionally, higher conductivity may also imply more possibilities for future applications of this bio-nanomaterial.

However, traditional X-ray and nuclear magnetic resonance techniques are difficult to use for resolving the structure of conductive bacterial filaments due to their insolubility, surface attachments, and multiphase nature[11]. As of now, there is no complete structural resolution of ***Geobacter metallireducens****'s* pili available. Furthermore, the structure of bacterial filaments is complex, and the assembly process is not fully understood. Traditional rational protein design approaches are challenging to optimize the conductivity of e-pili.

Therefore, we plan to utilize the optimization methods of mathematical modeling combined with protein directed evolution to attempt to optimize the structure of e-pili to improve its conductivity.

Methods and Materials

Molecular dynamic simulation

Obtaining the structure of pilus proteins involves challenges such as static snapshots, signal interference, and atomic position conflicts in methods like nuclear magnetic resonance imaging, X-ray techniques, and homology prediction. To address these issues, we employ molecular dynamics to optimize pilus subunits before assembly, introduce fuzzy constraints during assembly, and once again use molecular dynamics to optimize the final pilus assembly structure for the best polymer structure. Using molecular dynamics for optimization during assembly offers significant advantages: protein simulations occur in a realistic biological environment,

accounting for the dynamics of the organism. Molecular dynamics simulations (MD) can optimize the structure of type IV pilus proteins and annotate structural details.

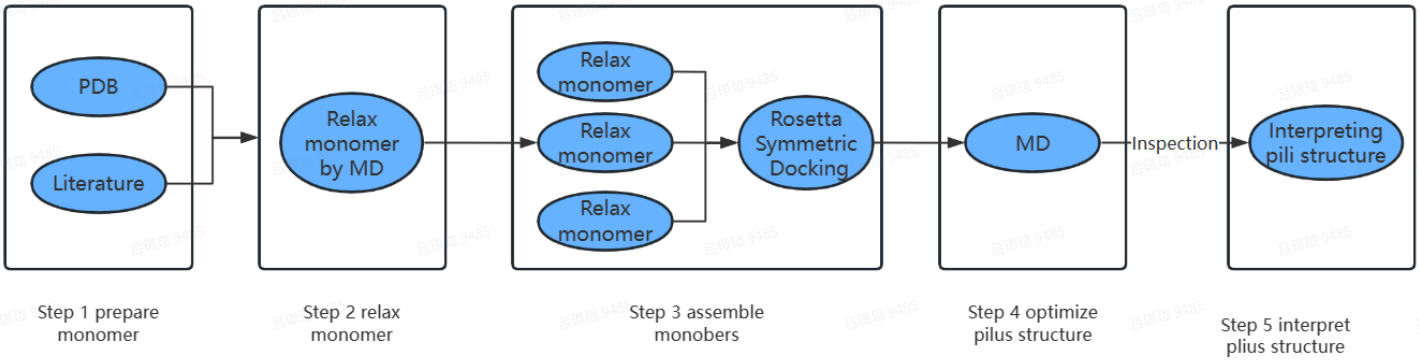


Fig1. Here's the translation of the process flow for pilus optimization you described, divided into five steps: Prepare subunits, Optimize subunits, Rosetta helical assembly, Optimize pilus structure, Annotate pilus structure.

Step1：PDB File Preprocessing

Protein structure PDB files often contain various information that is irrelevant to Amber, and sometimes Amber cannot correctly read standard residues or types of hydrogen atoms in the protein. Therefore, protein structures need to be cleaned.

This process involves three main steps:

1. Removal of all water molecules from the protein structure. PDB crystal structures typically include solvent molecules, namely water, which need to be completely removed.
2. Removal of hydrogen atoms from all amino acid residues, retaining only the coordinates of heavy atoms. The LEaP program in AmberTools will rebuild the positions of hydrogen atoms and convert them into Amber-recognizable atom labels (Amber Masks).
3. Ensuring that residue atoms are renumbered sequentially to obtain a 'formatted' PDB structure file.

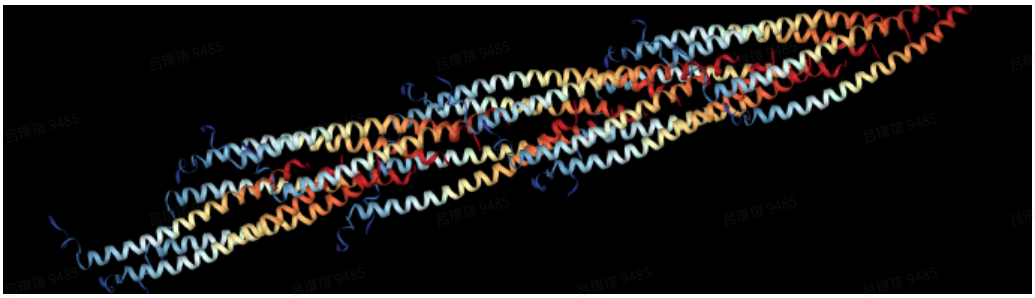
4. # Remove all non-protein residues, all water molecules, and add missing heavy atoms to standard amino acid residues
5. pdb4amber -i RXRa_pro.pdb -o RXRa_pro_dry.pdb -p --dry --add-missing-atoms
6. # The H atom of all amino acid residues is removed, thus retaining only the coordinates of the heavy atoms
7. pdb4amber -i RXRa_pro_dry.pdb -o RXRa_pro_noH.pdb -y

```
# Format the structure
pdb4amber -i RXRa_pro_noH.pdb -o RXRa_pro_processed.pdb
```

Step2: Rosetta helical assembly:

Sub-Step 1: Assemble the pilus model of the bacterium using the initial subunits and random assembly parameters; each symmetry parameter is randomized according to a Gaussian distribution.

Sub-Step 2: For each local energy minimum in the energy landscape, perform a restricted backbone docking calculation within a narrower range of symmetry parameters around the minimum obtained in Step 1. Then, proceed with a clustering process, where models are rotated around their symmetry axis and moved along their symmetry axis to determine the lowest root mean square deviation (rmsd) based on Cα positions, with a cutoff value of 2.00Å.



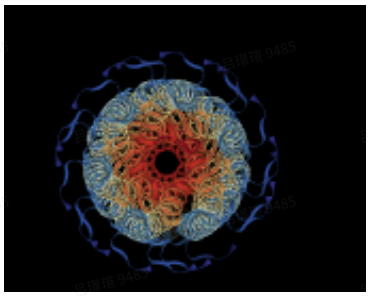


Fig2:3D structure of pilus　Fig3: Side view of pilus

Step3: Molecular Dynamics Simulation

Sub-step1：File Preparation

To maintain the system's charge neutrality, an appropriate number of Na+ and Cl- ions are added. Next, the protein is placed in a cubic TIP3P water box with a minimum distance of at least 10 angstroms (Å) between the box boundaries and the protein surface. Finally, all missing hydrogen atoms are added using the LEaP module in Amber23 software.

Force Field Selection: ff19SB、Water Model: TIP3P

```
source leaprc.protein.ff19SB
source leaprc.water.tip3p
pdb = loadPDB pili.pdb
solvateBox pdb TIP3PBOX 10.0
addIons2 pdb Na+ 0
saveAmberParm pdb protein.prmtop protein.inpcrd
```

Place the PDB file of the pilus protein in the appropriate directory and create an input file named "input.in" to specify the parameters for the simulation and energy minimization. Load the ff14SB force field and TIP3P water model, then dissolve the protein in a 10.0 Å cubic box based on the size of the protein. Sodium ions are added to the system to neutralize the charge.

Sub-step 2: Energy Minimization

The algorithm of PMEMD in Amber12 software is used for energy optimization, heating, and equilibration steps during the MD process. Energy minimization consists of two steps: first, energy minimization of water molecules and hydrogen atoms while the protein is restrained (harmonic restraint: 20 kcal mol−1 Å−2); then, energy minimization of the entire system. Each step of energy minimization involves 50,000 steps of steepest descent and 20,000 steps of conjugate gradient algorithms. The non-bonded cutoff is set to be greater than 8 angstroms.

```
Minimization
&cntrl
 imin=1,
 irest=0,
 maxcyc=10000,
 ncyc=5000,
 ntpr=100,
 cut=8.0
/
```

Sub-step 3: Heating and NVT Equilibration

After energy minimization, the system undergoes a heating process, raising the temperature from 0K to 300K. Once the system reaches a temperature of 300K, a 500-picosecond canonical ensemble MD simulation is conducted (harmonic restraint: 5 kcal mol−1 Å−2).

```
Heating and NVT
&cntrl
 imin=0, irest=0,
 nstlim=50000, dt=0.002,
 ntp=1, taup=2.0,
 ntb=2,ntc=2,ntf=2,
```

```
 ntx=1,cut=8.0,
 ntpr=1000, ntwr=1000, ntwx=1000,
 tempi=0.0, temp0=300, ntt=3, gamma_ln=2.0,
/
&wt TYPE='TEMP0',ISTEP1=0, ISTEP2=25000,
VALUE1=0.0, VALUE2=300.0,
/
&wt TYPE='END'
/
```

Sub-step 4: Molecular Dynamics Simulation

During the equilibration step (200 picoseconds) and subsequent molecular dynamics simulations (isothermal-isobaric NPT ensemble; subunits: 20 nanoseconds; flagellum: 40 nanoseconds), the SHAKE algorithm is used for constraint calculations. The time step is set to 2 femtoseconds, the temperature is maintained at 300K, and periodic boundary conditions are applied to avoid boundary effects. Coordinate files are recorded every picosecond, and later analysis is performed using the Cpptraj module in Amber23.

```
Molecular Dynamics Simulation
&cntrl
 imin=0,
 temp0=300, ntt=3, gamma_ln=2.0,
 ntp=1, taup=2.0,
 ntb=2, ntc=2, ntf=2,
 ntx=5, irest=1,
 ntpr=1000, ntwr=1000, ntwx=1000,
 nstlim=50000000, dt=0.002,
/
```

Optimization Design

To establish an optimization model for the conductivity of e-pili, we need to consider three basic elements of optimization separately: **Decision variables**, **Objective function**, and **Constraints**.

Decision variables

The decision variables are factors that decision-makers can control.

In this model, the variables we can control are the amino acid sequences of the e-pili. By changing the types and positions of amino acids in the sequence, we can attempt to alter the overall conductivity of the pili. The most critical indicator affecting pili conductivity is the spacing of Π residues (aromatic amino acids capable of producing Π-Π interactions). There are two methods to modify this spacing:

1)The quantity and positioning of aromatic amino acids within the e-pili are important factors determining their conductivity. For example, in comparison to the wild type *G. sulfurreducens*'s e-pili, a mutant variant Y27A exhibits less than one-fifth of the conductivity. This is because in Y27A, tyrosine, an aromatic amino acid, is replaced by alanine, a non-aromatic amino acid, leading to an increased spacing between the Π residues in that specific region. This alteration impacts the transfer of electrons, resulting in reduced conductivity.

2)The hydrophobicity of amino acids in the e-pili can also influence the overall structure and spacing of Π residues within the e-pili. For instance, in the mutant W51W57, replacing the phenylalanine and tyrosine at the c-terminal of the pili with tryptophan enhances the hydrophobicity of the pili and reduces the diameter of the pili. This results in a more compact structure, leading to a reduction in the spacing between internal Π residues. As a result, the conductivity of W51W57 has been shown to be approximately 2000 times higher than that of the wild type in some studies[6].

Due to the complexity of the structural model for type IV pili assembly and the lack of suitable methods to predict the impact of hydrophobicity changes in individual pili subunits on the overall assembled pili structure, our focus in this model is primarily on **method one**: understanding the impact of introducing new aromatic amino acids or altering the positions of existing aromatic amino acids on the pili conductivity.

Furthermore, it's worth noting that the assembly of e-pili requires a complex and conserved type IV pilus assembly system. This assembly system has certain requirements for the conservation of amino acid sequences within the pili. Modifying too many amino acids may not only result in the collapse of the original pili structure but also prevent proper assembly within cells.

This implies that our algorithm needs to find the optimal modification strategy under the constraint of minimizing the number of modifications required.

Objective function

The objective function is a function used to describe the goals pursued by decision-makers.

In our model, our ultimate goal is to obtain e-pili with higher conductivity characteristics. The conductivity characteristics of the e-pili are closely related to the distance between the centroid of the aromatic ring on the aromatic amino acid residues.

Based on this theory, we have designed the following two objective functions:

1）Average distance between the centroid of aromatic rings of Π residues on the shortest electron transfer chain in e-pili:

This objective function is used to describe the maximum conductivity of conductive pili under ideal conditions.

We can assume that in an ideal state, electrons will move along the shortest propagation path composed of Π residues in e-pili. At this point, the conductive pili has the highest conductivity within this structure.

This function describes the upper limit of the conductivity of pili under given structural conditions. Optimization based on this objective function may help us obtain the pili with higher conductivity.

Dijkstra's Shortest Path Algorithm

Dijkstra's algorithm is a popular graph search and shortest path algorithm used in computer science and mathematics. It is named after its inventor, Dutch computer scientist Edsger W. Dijkstra. The algorithm is designed to find the shortest path between a starting node (or vertex) and all other nodes in a weighted graph, where each edge has a numerical weight or cost associated with it.

Inputs:

Source node, Target node, Distance matrix, Adjacency matrix

Initialization:

Distance array: An array of distances from the source node, initialized as follows:

Set the distance from the source node to itself (dist(source)) to 0.

Set the distance from all other nodes (v) to infinity (∞) for nodes v not equal to the source node.

Q: A queue of all nodes in the graph.

S: An empty set to indicate which nodes the algorithm has visited.

Proceeding:

While queue Q is not empty, follow these steps:

Step1: Pop the node v from the queue Q that is not already in set S and has the smallest dist(v) value.

Step2: Add node v to set S to indicate that it has been visited.

Step3: Update the dist values of adjacent nodes of the current node v as follows:

For each new adjacent node u, if dist(v) + Distance_matrix[u, v] < dist(u), then update dist(u) to the new minimal value.

The algorithm stops after the target node is visited.

2）Average length between neighboring nodes of the Minimum Spanning Tree for all Π residues.

This objective function is used to describe the overall conductivity of e-pili under the influence of external environmental factors.

Under the influence of environmental factors such as pH changes or surface adsorption, other aromatic amino acids in e-pili, beyond the original shortest electron propagation path, may also participate in the electron transfer process as the structure of the e-pili changes.

This objective function describes the spacing between all Π residues inside the pili, reflecting the overall conductivity of the pili under different environmental influences. Optimization based on this objective function may improve the conductivity performance of our pili under different environmental conditions and make the conductivity curve smoother.

Prim's Minimum Spanning Tree Algorithm

Prim's algorithm is a greedy algorithm used in graph theory to find the Minimum Spanning Tree (MST) of a weighted, connected graph. A Minimum Spanning Tree of a graph is a subset of the graph's edges that connects all the vertices while minimizing the total sum of edge weights. This tree contains no cycles and includes every vertex of the original graph.

Inputs:

Distance matrix

Initialization:

An arbitrary starting node.

MST: Initialize an empty set to represent the Minimum Spanning Tree (MST).

Key array: Create an array to keep track of the minimum edge weight from each node to the MST. Initialize all values with infinity except for the starting node, which is set to 0.

Visited array: Create an array to keep track of visited nodes. Initialize all values to false.

Proceeding:

While the MST set does not include all nodes, follow these steps:

Step1: Select a node (u) that is not yet in the MST but has the minimum key value.

Step2: Add node (u) to the MST set.

Step3: Update key values: For each node v adjacent to the newly added node u, if Distance_matrix[u,v] < Key(v), update the Key(v) with the new weight.

Repeat steps 1-3 until all nodes are included in the MST

Note:

As a tree structure satisfies:

Number of edges = Number of nodes - 1

The final objective function is obtained from:

(Total length of the MST) / (Number of nodes - 1).

Constraints

Constraints are the limitations that decision variables must satisfy.
In this model, the constraints are the possible coordinates of the centroid of aromatic rings in aromatic amino acids. All decision variables (the added aromatic amino acids) can only occur at specific positions determined by the structure of e-pili themselves.

Since we only focus on the changes in conductivity caused by the quantity and position of aromatic amino acids , we can initially assume that adding or modifying a small number of aromatic amino acids has minimal impact on the overall structure of the e-pili. This means that we do not need to rebuild the molecular dynamics model for each iteration. Instead, we can use the spatial coordinate information of the original structure to roughly determine the range of the centroid coordinates for the new aromatic rings.

In our model, we approximate the centroid coordinates of the newly added aromatic ring by using the average coordinates of all atoms in the side chain at the modified position.

Since the purpose of constructing this model is to discover potential mutation sites, we can combine the semi-rational directed evolution method to screen and obtain the actual optimal results. Therefore, the accuracy requirement for this model is not as strict as that of rational protein design.

Model optimization

After clarifying the three elements of optimizing the model, we will now use particle swarm algorithm to optimize the electron transfer chain of the pili.

Graph structure construction

We create a new data class for pi-residues to store all their information, with the average of the x, y, and z coordinates of all the carbon atoms in each pi-residue as its center-of-mass coordinates in the three-dimensional space. We then construct a graph with each pi-residue as a vertex and obtain the corresponding adjacency and distance matrices based on the center-of-mass coordinates. Among them, we set a value k to constrain that only pi-residues with a distance less than k are capable of transferring electrons, otherwise, we set their adjacency value to 0 and their distance to infinity.

Particle Swarm Optimization

Particle Swarm Optimization (PSO) can be described as follows: Suppose the search space is L-dimensional and there are N particles in the population. The ith particle in the population can be represented as an L-dimensional vector $X_i = (x_{i1}, x_{i2}, \cdots, x_{il})$, and the best position it experiences is denoted as $P_{best} = (p_1, p_2, \cdots, p_l)$.Each position of the particle represents a potential solution to the requirement, which is substituted into the objective function to obtain its fitness value. The optimal position searched so far by the whole population is denoted as $G_{best} = (g_1, g_2, \cdots, g_l)$.

Next, we do the following iterations for each particle in the population :

$$V_i^{(t+1)} = \omega V_i^{(t)} + c_1 r_1 (P_{best}^{(t)} - X_i^{(t)}) + c_2 r_2 (G_{best}^{(t)} - X_i^{(t)})$$
$$X_i^{(t+1)} = X_i(t) + V_i(t+1)$$

The above formula consists of three parts. The first part is the current speed of the particle, with the positive value ω being the Inertia Weight, which indicates the current state of the particle. The second part is the Cognition Modal, which indicates the cognition of this generation of particles about their own state (c_1 is Self-recognition Factor). The third part is the Social Modal, which indicates the sharing of information between particles through generations (c_2 is Social-recognition Factor).

Adaptive Particle Swarm Optimization

APSO makes three main improvements on the basis of PSO:

1.Evolutionary State Estimation

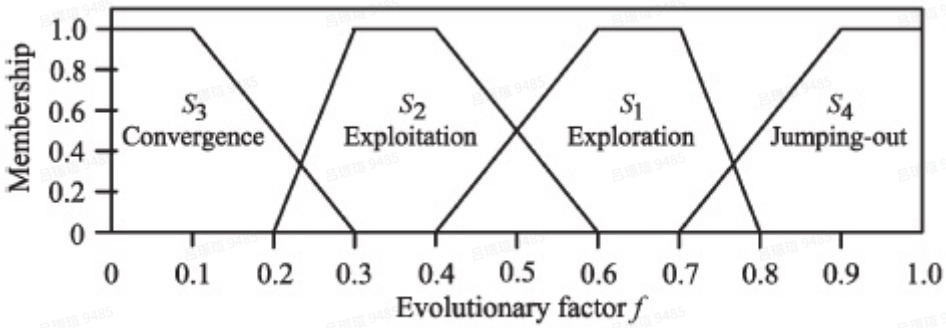
ESE calculates the distributional information of the population of each generation in the space of independent variables. We first calculate the average Euclidean distance from particle i to all the other particles.

$$d_i = \frac{1}{N-1} \sum_{j=1, j \neq i}^{N} \sqrt{\sum_{k=1}^{L} (x_{ik} - x_{jk})^2}$$

Accordingly, we define the Evolutionary Factor f that measures the optimizing state of the population:

$$f = \frac{d_g - d_{min}}{d_{max} - d_{min}} \in [0, 1]$$

The Evolutionary factor we calculated above can help to decide which evolutionary state the population is in. The different evolutionary states of the population affect our different adaptive control strategies for the parameters (here we are referring to ω, c_1, c_2). We apply the obtained f to define four different evolutionary states S_1, S_2, S_3, S_4 and their fuzzy membership function, representing the states of exploration, exploitation, convergence, and jumping out respectively.



2.Parameters self-adaptive control strategy

We use a sigmoid mapping to allow ω to change with the Evolutionary factor f

$$\omega(f) = \frac{1}{1 + 1.5e^{-2.6f}} \in [0.4, 0.9]$$

We know $\omega(f)$ is a monotonically increasing function.Therefore, a larger f in S_1 and S_4 will result in a larger value of ω , which facilitates a global search. On the contrary, S_2 and S_3 are more suitable for a local search.

c_1 represents self-recognition factor, which pulls the particles to their own historical optimal position; c_2 represents social-recognition factor, which pushes the population to converge to the current global optimal region and helps fast convergence. These coefficients is initialized to 2.0 and are adaptively controlled according to the evolutionary state with the following strategies:

(1) S_1 - Increase c_1 and decrease c_2 in the exploration state, which helps particles to explore individually and achieve their own historical optimal positions, rather than crowding around the current best particles that may be associated with the local optimum.

(2) S_2 - Slightly increase c_1 and slightly decrease c_2 in the exploitation state.

(3) S_3 - Slightly increase c_1 and slightly increase c_2 in the convergence state.

(4) S_4 - Decrease c_1 and increase c_2 in the jumping-out state.

3.Elitist Learning Stratrgy

The jumping out effect done by the ELS mechanism on G_{best} is necessary for the global optimal nature of this algorithm, which randomly selects a dimension of G_{best} , which is represented by the g_d d of the dth dimension:

$$g_d = g_d + (X_d^{max} - X_d^{min}) \cdot Gaussian(\mu, \sigma^2)$$

APSO's adaptive tuning of parameters can improve the performance of the PSO algorithm in terms of both accelerated convergence and global search. In addition, its ELS strategy will maintain the diversity of the population in order to allow the algorithm to jump out of potential local optima.

Discrete Mapping

In the optimization process, we defined the independent variable as the sequence number of regular residues in a monomer that needs to be adjusted to pi-residues. Suppose there are N non-pi residues in each monomer and the number that needs to be changed to pi-residues in each monomer is d , then our independent variable X can be denoted as
$X = (x_1, x_2, \cdots x_d),\ x_i \in \{1, 2, \cdots N\}$
Before performing the optimization, we sorted all the residue sequences of each monomer according to their L1 norm from the origin to ensure that the spatial distances of residues with similar sequence numbers are also relatively closer spatially, ensuring sequence numbers and the geometrical positions are consistent with each other.
However, APSO still deals with only the continuous problem, so we need to discretize the continuous particles $Y = (y_1, y_2, \cdots y_d),\ y_i \in [0, 1]$ to correspond to the sequence number of residues in each monomer. The method we use is to stick to the continuous optimization in the interval [0, 1] and divide [0, 1] into N equally spaced sub-intervals of length $\frac{1}{N}$. Therefore, we can map Y to X one on one by letting $x_i = [y_i \times N]$, which means keeping the integer part of Ny_i .

Chassis microbe

The selection of chassis microbes is also a crucial issue in our project. *G.metallireducdens.* is a strictly anaerobic that is challenging to industrially cultivate with existing technologies. In addition, compared to commonly used engineering strains, *G.metallireducdens* has a slow growth rate, low protein expression, and difficulties in regulation. These limitations clearly restrict the application potential of this conductive pili.

Vibrio natriegens is a gram-negative marine bacterium known for its extremely fast growth rate, with a generation time of less than 10 minutes, making it the fastest growing bacterium known to date[12,13]. As a newly emerging chassis microbe, *V. natriegens* has numerous advantages such as substrate versatility, high metabolic rates, lack of pathogenicity to humans, ease of genetic manipulation, and efficient expression of heterologous proteins, demonstrating promising applications in the field of synthetic biology.

In nature, a large amount of Vibrio species with Type IV pili can be found[14,15]. By using NCBI Nucleotide BLAST, we searched the genome of V. natriegens ATCC 14048 (taxid:691) and found that it does indeed possess a Type IV pilus assembly system (T4aP). This indicates that compared to E. coli, V. natriegens may be a more suitable chassis microorganism for the production of conductive pili.

plasmids

Pilus producting plasmid-pPilin

Before producing different varients of e-Pili, we wanted to establish a protocol that could stably harvest pilus. Therefore, we first constructed a plasmid to express original e-Pili from Geobacter

metallireducens and test the function of our expressing vector and harvest protocols. 6×His tag was added to the C terminus of PilA, the monomer that can assemble into e-Pilin in G.metallireducens, so the harvested e-Pili can be easily further purified using Ni-NTA resin. Recombined PilA is under the control of pBAD which can work in V.natriegens properly according to assays[3]. PilA-C is a protein that interacts with monomer PilA-N,i.e. PilA, and helps in the process of pilus assembly, so PilA-C is also constructed in this plasmid[15]. The resultant e-Pili producing plasmid, was named pPili.

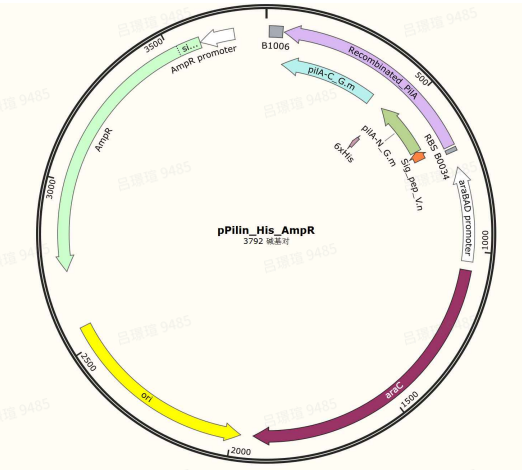


Fig 4:Plasmid structure of pPilin

CRISPRi plasmid-pdCas9-sg

The fact that Vibrio natriegens possesses Type IV pili (T4aP) is a major advantage for it to be the microbial chassis of our project. However, it also presents new challenges. The assembly of the pili in Vibrio parahaemolyticus may potentially conflict with the introduced T4aP of Geobacter metallireducens, resulting in the failure in optimal production of conductive pili. Therefore, we have decided to use the CRISPRi system to suppress the expression of the Vibrio parahaemolyticus’ pili gene.

The reason why CRISPR/Cas9 was not chosen for direct knockdown of Vibrio natriegens’ own pili is that it has been reported in the literature that DNA double-stranded breaks caused by the CRISPR/Cas9 system can be extremely toxic to Vibrio spp because non-homologous end-joining (NHEJ) activity is almost undetectable in the microorganism[16].

We designed four different sgRNAs that direct to different sequences near the PilA transcription start site. In order to select the sgRNA with the best inhibitory effect, we also designed a sfGFP-based reporter plasmid(see in report plasmid).

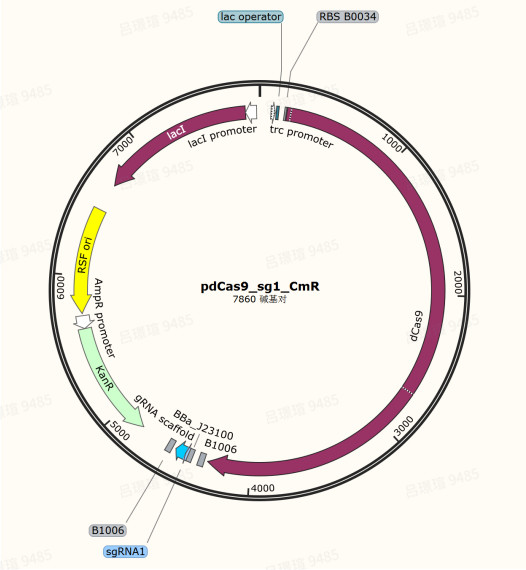


Fig 5:Plasmid structure of pdCas9-sg

report plasmid-pGFP

The reporter plasmid based on sfGFP uses pUC19 as the backbone. The expression of recombinant sfGFP is controlled by the promoter of the monomer PilA (V.PilA) of Vibrio natriegens. Additionally, a segment of the N-terminal sequence of V.PilA(start codon removed) is attached to the N-terminus of the recombinant sfGFP. The four sgRNAs we have selected all target the transcription start site (TSS) region near V.PilA, including the promoter and the N-terminal coding sequence. Therefore, the intensity of green fluorescence from this reporter plasmid can indicate the effectiveness of the CRISPRi system. We underwent a series of adjustments in the design of the reporter plasmid and finally obtained this pGFPcut version using reversed PCR and Gibson assembly, which has shown promising results.

electroporation

Considering the relatively low tranfer efficiency of V.natriegens compared to commenly used E.coli strains[17], we chose electroporation to tranfer our plasmid into V.natriegens. The parameters we used for V.natriegens electroporation were 790 V, 4 ms.The plasmids were first expanded in DH5α before electroporation.

Pilus production and harvest

Pius was induced by 2% arabinose solution.20 agar plates supplemet wth ampicillin and Kanamycin were spreaded with 100 µL of 2% arabinose soultion.After the arabinose was absorbed by the plates, 300 µL of overnight V.PC culture was spread on the plates.

After culturing at 30 °C for 48 h, the biofilm on the plates were scrapped from the surface and transferred into a 15 mL centrifuge tube with 1.5 mL of LB + v2 salts medium(LB supplement with 204 mM NaCl,4.2 mM KCl, 23.14 mM MgCl2, 17.6 mM K2HPO4). The cells were collected by

centrifugation and resuspended in 150 mM ethanolamine buffer (pH 10.5). The pili were sheared from the cells in a Waring blender at low speed for 1 min. Cells were removed by centrifugation at 13,000 × g. The pili in the supernatant were precipitated with 10% ammonium sulfate overnight, followed by centrifugation at 13,000 × g. The precipitate was resuspended in ethanolamine buffer, and additional debris was removed with centrifugation at 16000 × g. The pili were collected with a second 10% ammonium sulfate precipitation and subsequent centrifugation at 13,000 × g. The final pili preparation was resuspended in ethanolamine buffer and stored at 4°C[18]. Harvested pilus solution was further purified by Ni-NTA resin for analysis.

Protein analysis and characterization

The green fluorescence was detected using microplate reader.The excitation light was of wavelength 488 nm and the emitted light was of wavelength 507 nm.The fluorescence strength of strains trsnferred or not transferred with pGFP were compared using GraphPad Prism 6.

Purified pilus was analysed with tricine-tris SDS-PAGE to monitor the protein concentration before further experiments.

Microbial fuel cell

After obtaining a mutant library of e-pili, we need to devise an appropriate screening strategy to quickly identify mutants within the library that may exhibit higher electrical conductivity. Unlike directed evolution for enzymes, there is currently no mature approach for directed evolution of structural proteins like e-pili. This means that we need to design a completely new screening method.

Our initial plan was to express the pili heterologously in engineered bacteria, isolate and purify them, and then measure their electrical conductivity individually using a conductivity meter in vitro. However, this method is inefficient, and due to the challenges in standardizing the isolation and purification of pili, there may be variations between batches, leading to significant errors in the measured electrical conductivity. This phenomenon has been reported in several studies.

Furthermore, the assembly mechanism of Type IV pili is relatively conservative, and modifying the amino acids within them is likely to disrupt assembly and lead to loss of their conductive properties. This implies that the screening system we use must be capable of quickly eliminating pili with electrical conductivities below a certain threshold from the library, leaving only those with potentially high electrical conductivity for further in vitro testing. This approach aims to increase screening efficiency while minimizing experimental errors associated with large-scale operations.

We plan to start with the original function of e-pili and use *Geobacter sulfurreducens* as the chassis organism. Based on the improved H-type microbial fuel cell model previously developed by researchers [19,20], we will construct a bioelectrochemical system to initially screen for pili with higher electrical conductivity by observing the maximum current produced by the strains in microbial fuel cells.

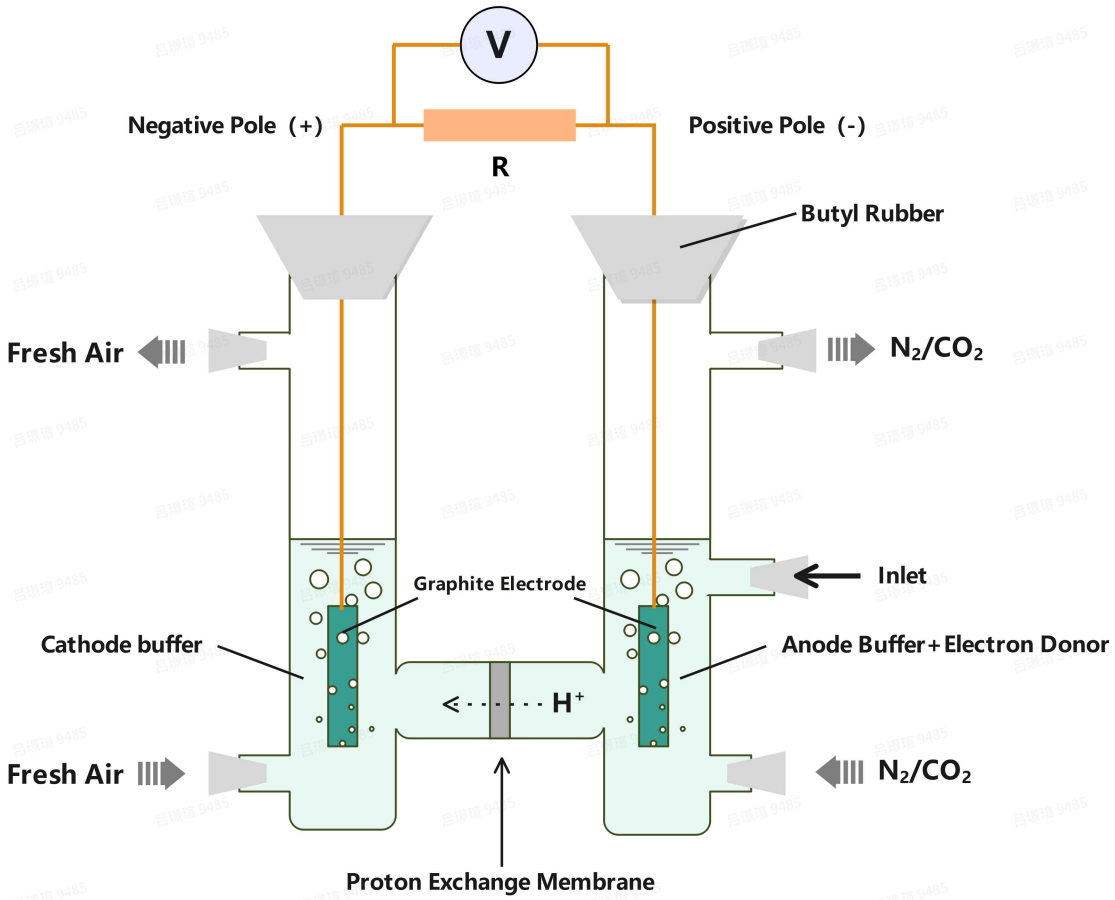


Fig 6: Microbial fuel cell

The cell is inoculated onto the anode, which uses a graphite electrode. A voltage regulator is used to maintain an anodic potential. A resistor is used to connect the anode and cathode of the cell. After the bacterial concentration in the anode chamber reaches OD600 = 0.2, the anode solution is completely replaced with freshwater medium containing only 10 mM sodium acetate (electron donor) and without 40 mM fumarate (electron acceptor). The colonies will then switch to the graphite electrode as the electron acceptor. When the current reaches approximately 1 mA, the system is switched to continuous flow mode, with the medium flowing through the anode chamber at a dilution rate of 0.15/h, continuously providing an electron donor. The cathode chamber uses an electrolyte solution similar to the anode chamber (same composition and pH) and replaces bicarbonate with the same molar concentration of TRIS buffer to prevent changes in CO2 content and pH due to air introduction. The anode chamber needs to be continuously flushed with sterile N2/CO2 gas (80:20) to remove oxygen throughout the reaction, while the cathode chamber is continuously supplied with sterilized air.

Results

V.natriegens electroporation

The three plasmids were tranferred into V.natriegens seperately. Strain transferred with pPilin, pdCas9-sg pGFP were named as V.PL, V.CR and V.GF respectively. Then pPilin and pGFPcut were transferred into V.CR respectively.The results of colony PCR are shown in the figure below.

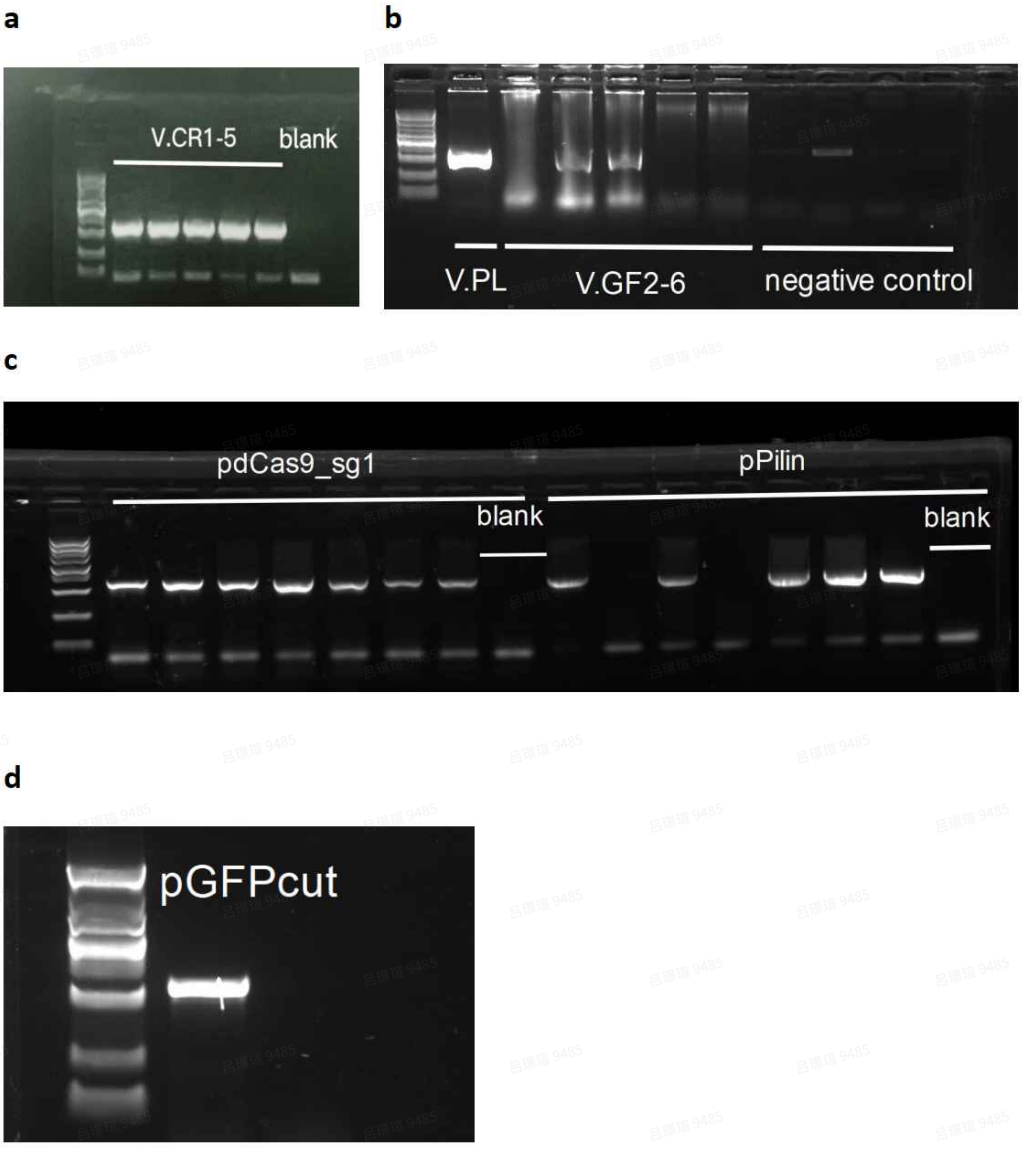


Fig 7:Colony PCR results. a,V. natriegens transferred with pdCas9-sg. b,V.natriegens transferred with pPilin and pGFP. c,V.CR transferred with pPilin. d,V.CR transferred with pGFPcut

Green fluorescence measurement

Overnight culture of V.CR-cut and V.CR were used for green fluorescence measurement with microplate reader.The excitation light was 488 nm and emitted light was 535 nm.the The result was as follows.The fluorescence strength was significantly stronger in strain transferred with pGFPcut(designated as V.CR-cut).

	V. CR-cut	ATCC14048
T1	2820994	2460322
T2	2874411	2181941
T3	2663322	2099766
T4	2733929	2181769
T5	2767116	2195564
T6	2829154	2281576

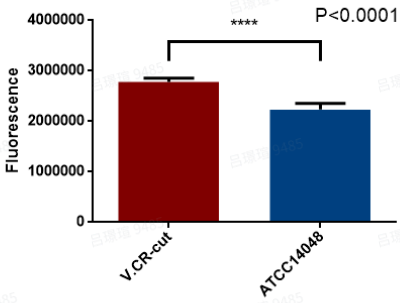


Fig 8:Comparison of fluorescence strength between strain V.CR-cut and wild type V.natriegens ATCC14048.

Pilus harvest and purification

Harvested pilus solution was purified by Ni-NTA resin. Flo-through solution,washing solution and elution solution were analysed by tricine-tris SDS-PAGE. There was a band near 25kD(marked with asterisk) in elution solution.

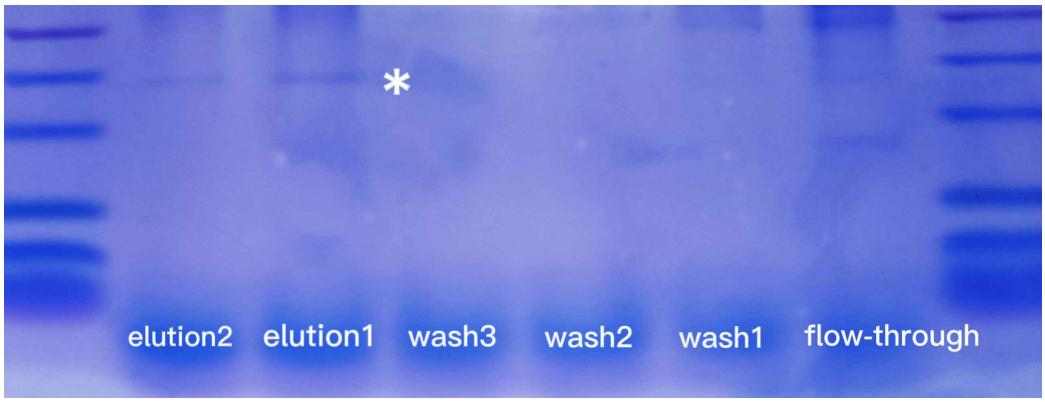


Fig 9:Tricine-tris SDS-PAGE of pilus

Molecular dynamic simulation

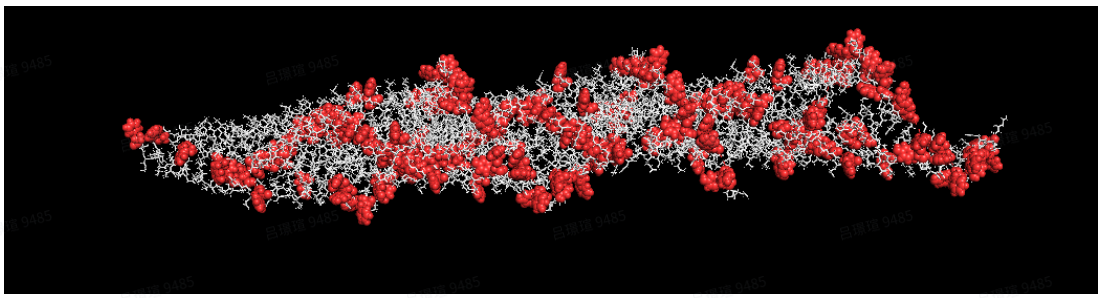


Fig 10： All aromatic amino acids are labeled

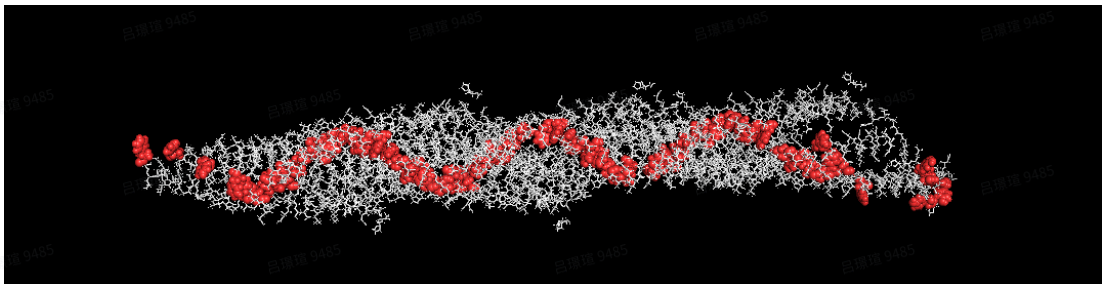


Fig 11： Electron transfer chain

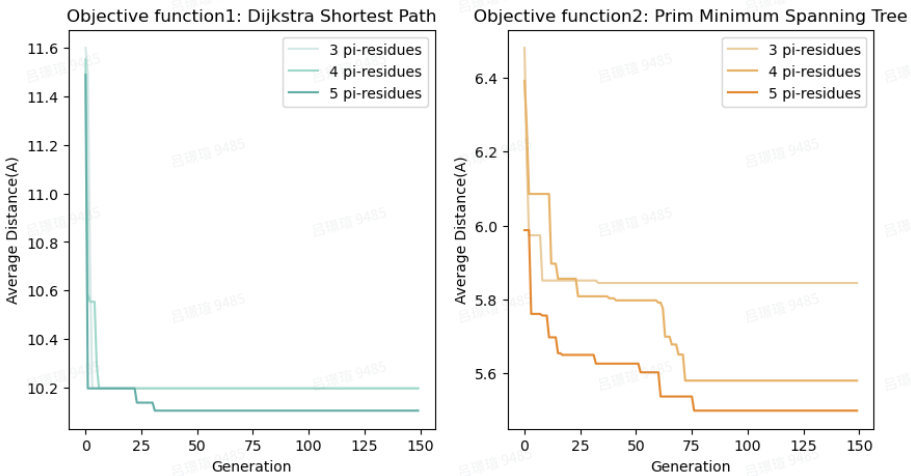
We optimized the overall structure of the pilus using a flexible prediction method. Aromatic amino acids (PHE, TYR, TRP) from different subunits formed an electron transfer chain based on a helical structure. Visual processing of this was carried out using PyMOL software. As shown in the figure, a complete electron transfer chain is displayed.

Optimization Results

Observations show that adding more residues can lead to a significant reduction in the average distance in the shortest path and the minimum spanning tree, which can result in an increase in conductivity. We recorded the optimal extra residue numbers when adding 5 residues:

Objective function1(Dijkstra Shortest Path): **9, 20, 34, 54, 56**

Objective function2(Minimum Spanning Tree): **6, 29, 54, 55, 60**



Discussion

The type IV pili of *Geobacter metallireducens* exhibit an exceptional electrical conductivity of up to 277 S/cm at pH 7, making them the most conductive biological nanomaterials known to date.The overlapping Π-Π orbitals of aromatic moieties imparts metallic conductivity to these revolutionary "green" nanomaterials.We build an antibody detection system, and the conductivity of this pili is the key to maintaining stability and accuracy of our system.We use optimization methods of mathematical modeling combined with protein directed evolution to attempt to optimize the structure of this pili to improve its conductivity.

Through algorithm optimization, we have identified 5 possible mutation sites（9, 20, 34, 54, 56）to improve the maximum conductivity of e-pili,and another 5 possible mutation sites(6, 29, 54, 55, 60) to improve the conductivity performance of our pili under different environmental conditions.

Although we have meticulously optimized the structure of pili using the flexible prediction method mentioned above, we still observe some discrepancies between the optimized subunits and the actual structure. We are looking for ways to further enhance this method and obtain improved pili structures. Additionally, certain pili structures exhibit breakpoints, which could potentially alter the overall optimal electron transfer chain. Our team will continue to explore how to optimize these breakpoints to achieve better pili structures.

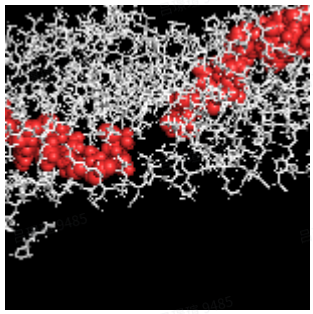


Fig 12：Breakpoints in the electron transport chain

We have established the strain and to produce our expected e-Pili and the methods to harvest then purify it Although due to time restriction, we didn't further verify the properties of our harvested pilus solution or test the mutations. We will, in the future, optimize our system to produce e-Pili more efficiently and characterize the variants we designed in this project.

References

1. Liu X, Tremblay P, Malvankar N S, et al. A Geobacter sulfurreducens Strain Expressing Pseudomonas aeruginosa Type IV Pili Localizes OmcS on Pili but Is Deficient in Fe (III) Oxide Reduction and Current Production[J]. Applied and environmental microbiology. 2014, 80(3): 1219-1224. https://doi.org/10.1128/AEM.02938-13
2. Sun D, Wang A, Cheng S, et al. Geobacter anodireducens sp. nov., an exoelectrogenic microbe in bioelectrochemical systems[J]. Int J Syst Evol Microbiol. 2014, 64(Pt 10): 3485-3491. https://doi.org/10.1099/ijs.0.061598-0
3. Smith J A, Lovley D R, Tremblay P. Outer cell surface components essential for Fe (III) oxide reduction by Geobacter metallireducens[J]. Applied and environmental microbiology. 2013, 79(3): 901-907 https://doi.org/10.1128/AEM.02954-12
4. Malvankar, N. S., & Lovley,. R. (2014). Microbial nanowires for bioenergy applications. Current Opinion in Biotechnology, 27, 88–95. https://doi.org/10.1016/j.copbio.2013.12.003

5. Gorby, Y. A. *et al.* Electrically conductive bacterial nanowires produced by Shewanella oneidensis strain MR-1 and other microorganisms. *Proc Natl Acad Sci USA* 103**,** 11358–63 (2006). https://doi.org/10.1073/pnas.0604517103
6. Reguera, G. *et al.* Extracellular electron transfer via microbial nanowires. *Nature* 435, 1098–101 (2005). https://doi.org/10.1038/nature03661
7. Tan Y, Adhikari RY, Malvankar NS, Ward JE, Woodard TL, Nevin KP, Lovley DR. 2017. Expressing the Geobacter metallireducens PilA in Geobacter sulfurreducens yields pili with exceptional conductivity. mBio 8:e02203-16. https://doi.org/10.1128/mbio.02203-16
8. Malvankar NS, Vargas M, Nevin K, Tremblay P-L, Evans-Lutterodt K, Nykypanchuk D, Martz E, Tuominen MT, Lovley DR. 2015. Structural basis for metallic-like conductivity in microbial nanowires. mBio 6(2):e00084-15. doi:10.1128/mBio.00084-15. https://doi.org/10.1128/mbio.00084-15
9. Lovley DR. 2017. e-Biologics: Fabrication of sustainable electronics with ‘green’ 650 biological materials. mBio 8:e00695-17https://doi.org/10.1128/mbio.00695-17
10. Thomas G. Schuhmann Jr., Jun Yao, Guosong Hong, Tian-Ming Fu, and Charles M. Lieber Syringe-Injectable Electronics with a Plug-and-Play Input/Output Interface Nano Letters 2017 *17* (9), 5836-5842 DOI: 10.1021/acs.nanolett.7b03081 https://doi.org/10.1021/acs.nanolett.7b03081
11. Xiao K, Shu C, Yan Q, et al. Predicting Homogeneous Pilus Structure from Monomeric Data and Sparse Constraints[J]. Biomed Res Int. 2015, 2015: 817134. https://doi.org/10.1155/2015/817134
12. Eagon RG. *Pseudomonas natriegens*, a marine bacterium with a generation time of less than 10 minutes. J Bacteriol, 1962, 83: 736-737.https://doi.org/10.1128/jb.83.4.736-737.1962
13. Weinstock MT, Hesek ED, Wilson CM, et al. *Vibrio natriegens* as a fast-growing host for molecular biology. Nat Methods, 2016, 13(10): 849-851.https://doi.org/10.1038/nmeth.3970
14. Aagesen, A.M., Häse, C.C. Sequence Analyses of Type IV Pili from *Vibrio cholerae*, *Vibrio parahaemolyticus*, and *Vibrio vulnificus* . *Microb Ecol* **64**, 509–524 (2012). https://doi.org/10.1007/s00248-012-0021-2
15. Berry JL, Pelicic V. Exceptionally widespread nanomachines composed of type IV pilins: the prokaryotic Swiss Army knives. FEMS Microbiol Rev. 2015 Jan;39(1):134-54. https://doi.org/10.1093/femsre/fuu005
16. Lee HH, Ostrov N, Wong BG, Gold MA, Khalil AS, Church GM. Functional genomics of the rapidly replicating bacterium Vibrio natriegens by CRISPRi. Nat Microbiol. 2019 Jul;4(7):1105-1113.https://doi.org/10.1038/s41564-019-0423-8
17. Xu J, Yang S, Yang L. Vibrio natriegens as a host for rapid biotechnology. Trends Biotechnol. 2022 Apr;40(4):381-384.https://doi.org/10.1016/j.tibtech.2021.10.007

18. Tan Y, Adhikari RY, Malvankar NS, Ward JE, Woodard TL, Nevin KP, Lovley DR. Expressing the Geobacter metallireducens PilA in Geobacter sulfurreducens Yields Pili with Exceptional Conductivity. mBio. 2017 Jan 17;8(1):e02203-16. https://doi.org/10.1128/mbio.02203-16
19. Nevin, K. P. *et al.* Anode biofilm transcriptomics reveals outer surface components essential for high density current production in Geobacter sulfurreducens fuel cells. *PLoS One* **4,** e5628 (2009). https://doi.org/10.1371/journal.pone.0005628
20. Bond, D. R. & Lovley, D. R. Electricity Production by Geobacter sulfurreducens Attached to Electrodes. *Appl. Environ. Microbiol.* **69,** 1548–1555 (2003).https://doi.org/10.1128/AEM.69.3.1548-1555.2003