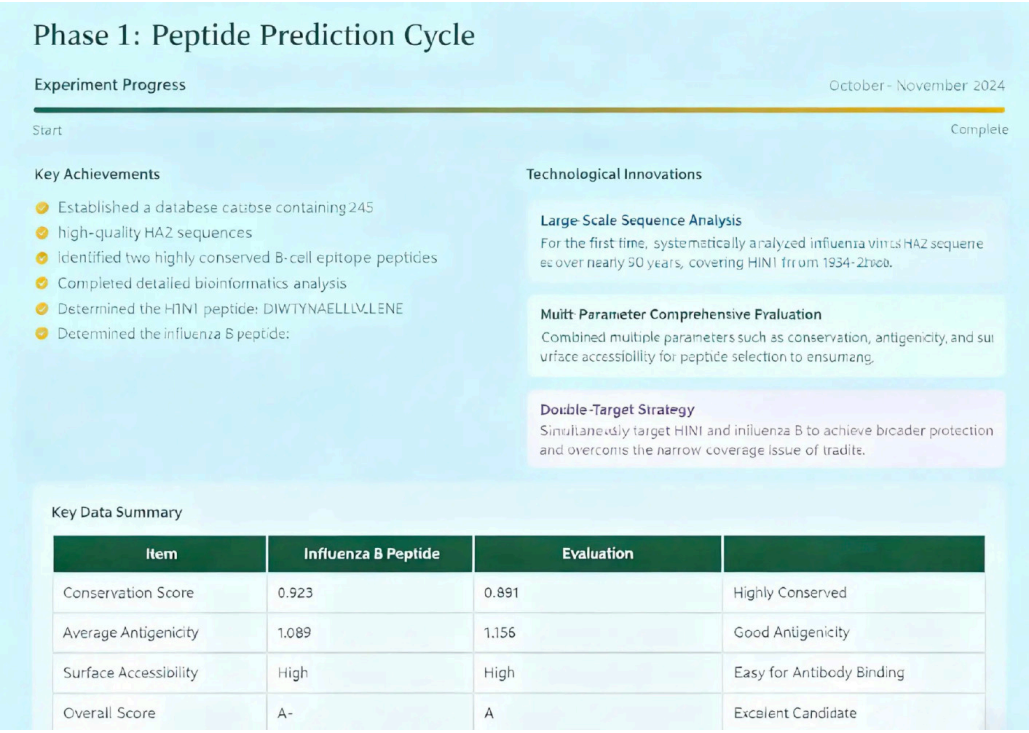
notebook

Design of a Ferritin Nanoparticle-Based Broad-Spectrum Influenza Vaccine Project

I. First Stage of the Experiment: Epitope Prediction and Design (Scheduled for October 2024–November 2024)I. First Stage of the Experiment: Epitope Prediction and Design (Scheduled for October 2024–November 2024)



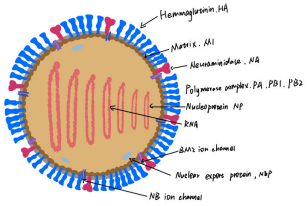
(I) Literature Search and Study

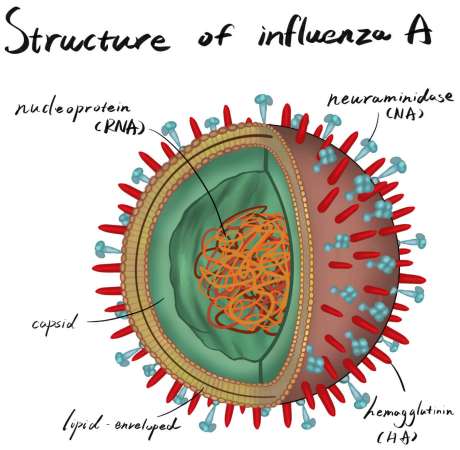
Influenza viruses undergo antigenic drift and antigenic shift, necessitating annual updates of vaccines. We consulted data and structures related to Influenza A and

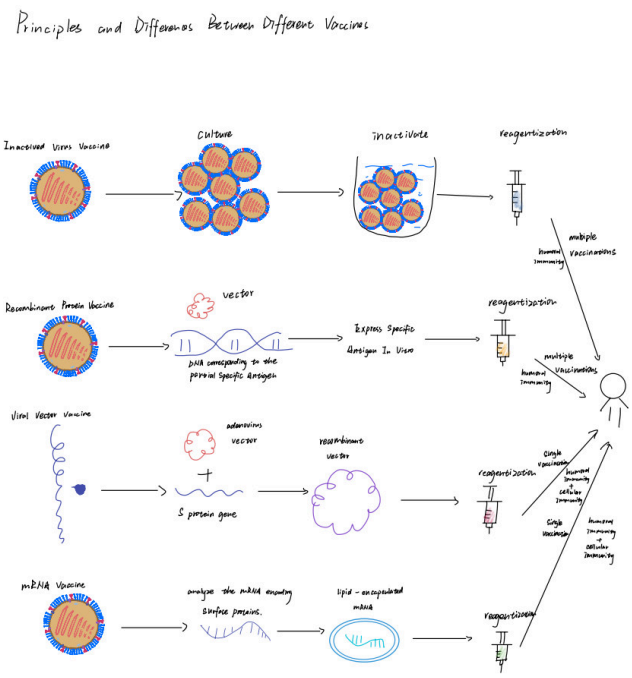
B viruses, and created structural diagrams and analyses. Influenza A virus is the primary pathogen responsible for seasonal influenza epidemics and periodic global pandemics. Its genome consists of eight segments of single-stranded negative-sense RNA. Based on the antigenicity of its surface proteins, hemagglutinin (HA) and neuraminidase (NA), it can be further classified into numerous subtypes. Influenza A virus has a broad natural host range, including birds, pigs, horses, and humans. Its high mutation rate stems from two main mechanisms: first, antigenic drift caused by the low fidelity of RNA polymerase, and second, antigenic shift resulting from genetic reassortment when different subtypes co-infect a host. The latter can produce novel viral strains with pandemic potential. This characteristic is precisely why annual updates of the influenza vaccine are necessary.

Influenza B virus primarily infects humans and generally causes seasonal epidemics, yet its public health burden and impact on hospitalization rates should not be overlooked. Unlike Influenza A, Influenza B is divided into two main lineages based on genetic and antigenic characteristics: the Victoria lineage and the Yamagata lineage. The host range of Influenza B virus is relatively narrow, almost exclusively limited to humans and seals, with no evidence of a widespread animal reservoir. Its evolution occurs mainly through antigenic drift. Since it does not undergo cross-species transmission and genetic reassortment like Influenza A, it does not exhibit antigenic shift capable of causing global pandemics. However, the simultaneous circulation of its two lineages adds complexity to vaccine strain selection.

The structure of Influenza B virus.







Based on the above background, the core objective of this phase of our project is to accurately predict and screen linear B-cell epitopes from the HA2 region using bioinformatics approaches. These epitopes should exhibit strong immunogenicity, high conservation, and spatial accessibility. B-cell epitopes are specific segments directly recognized and bound by antibodies, categorized into conformational epitopes and linear epitopes. Given that our final strategy may involve recombinant protein or peptide vaccines, we initially prioritize linear epitopes. Our strategic approach is as follows:
Comprehensive Sequence Collection: Gather complete HA2 sequences from public databases covering major human influenza subtypes (e.g., H1N1, H2N2, H3N2) and avian influenza subtypes with pandemic potential (e.g., H5N1, H7N9, H9N2).
Conservation Analysis: Perform multiple sequence alignments to quantify the conservation score of each amino acid position, identifying ”super-conserved regions” that transcend subtype boundaries.
Antigenicity Prediction: Utilize established biological principles (e.g., hydrophilicity/hydrophobicity, polarity, side chain conformation of amino acids) to predict the potential of these conserved fragments to be recognized by B cells and elicit an immune response.
Accessibility Evaluation: Assess the spatial accessibility of predicted epitopes on native viral particles by integrating known HA trimer three-dimensional structures, excluding regions obscured by glycosylation or adjacent structural elements.

Comprehensive Screening: Integrate multi-dimensional data on conservation, antigenicity, and accessibility to select optimal candidate epitopes, laying the foundation for the next stage of peptide synthesis and immunogenicity validation.

(II) Instruments and Software to Be Used

NCBI Influenza Virus Sequence Database: Used to obtain HA2 sequence data.
MEGA10: Used for multiple sequence alignment and evolutionary analysis.
BioEdit & Excel: Used for conservation scoring and matrix construction.
IEDB (Immune Epitope Database): The Kolaskar & Tongaonkar method was used to predict antigenicity.
WebLogo: Used for generating sequence logo plots.
Databases and Software:
NCBI Influenza Virus Sequence Database: The world's largest influenza virus sequence database, used for batch downloading HA gene and protein sequences in FASTA format.
MEGA11 (Molecular Evolutionary Genetics Analysis): Used for large-scale multiple sequence alignment, constructing phylogenetic trees, calculating sequence distances, and analyzing conservation.
BioEdit v7.2: A lightweight sequence editing and analysis tool used for sequence curation, format conversion, and preliminary analysis.
Microsoft Excel 365: Used for calculating conservation scores, constructing data matrices, creating charts, and organizing results.
IEDB Analysis Resource (Immune Epitope Database): A comprehensive immune epitope analysis platform. In this phase, we primarily used its "Kolaskar & Tongaonkar Antigenicity" method for B-cell antigenicity prediction.
WebLogo 3: An online tool used to generate sequence logo plots that visually represent the conservation and diversity at each position in the multiple sequence alignment results.
PyMOL v2.5: A molecular visualization system used to obtain the 3D structures of HA protein (e.g., from PDB entries like 1RU7, 1RUY) and analyze the spatial location of epitopes.
Electronic File Archive Summary:
HA2_SequenceSet_Processed.fas (Processed HA2 sequence set)
HA2_Conservativity_Analysis.xlsx (Detailed calculation sheet and charts for conservation scoring)
IEDB_Antigenicity_Report.pdf (Screenshots and report of IEDB antigenicity prediction results)
WebLogo_HA2.png (HA2 sequence logo plot)
PyMOL_Session_HA2_Epitopes.pse (PyMOL session file containing the 3D structure with all annotated epitopes)
HA2-MECv1_Sequence.txt (Amino acid and DNA sequence of the final designed candidate molecule)

(III) Experimental Records

1. First weekend of October 2024

Objective: To collect HA2 protein sequences of H1N1 and Influenza B viruses.
Methods:

1. Accessed the NCBI Influenza Virus Database and filtered for H1N1 HA2 sequences from 1934 to 2020, and Influenza B HA2 sequences from 1940 to 2020.
2. Exported all sequences in FASTA format and saved them to the team's shared cloud drive.

Key Considerations:

- Ensure sequence sources are reliable; avoid using sequences of unverified origin or low quality.
- Apply a unified naming convention to facilitate subsequent analysis.

Detailed Record:

Objective: Systematically collect HA2 protein sequences for H1N1 subtypes (1934-2020) and Influenza B viruses (including Victoria and Yamagata lineages, 1940-2020) from the NCBI Influenza Virus Database.

Background & Plan:

During our weekly online team meeting, based on extensive literature review, we confirmed the core hypothesis of this project: the stalk region of the influenza virus hemagglutinin (HA) protein (primarily composed of the HA2 subunit) is more conserved than the head region and may contain targets for a universal vaccine. Today, the primary task of the bioinformatics team is to acquire the "raw material" – a large number of HA2 sequences – needed to validate this hypothesis. We established a detailed SOP (Standard Operating Procedure) to ensure data source reliability and reproducibility.

Detailed Methods & Steps:

1. Database Access & Filtering Strategy:

- Platform: We used computers uniformly configured in the university lab, accessing the NCBI Influenza Virus Database via the campus network. This is the most authoritative global resource for influenza virus data.
- H1N1 Filtering:
  - In the virus selector, chose ”Influenza A Virus”.
  - Selected subtype ”H1N1”.
  - Selected host as ”Human” to focus on strains infecting humans.
  - Set the year range from ”1934” to ”2020”. 1934 was chosen because it includes one of the earliest recorded human H1N1 strains, A/PR/8/34.
  - In the ”Protein” field, we entered ”HA” for the initial search, as we discovered that directly filtering for HA2 sequences was limited. The plan was to obtain the full HA sequence first and subsequently extract the HA2 portion.
- Influenza B Filtering:
  - Selected virus type as ”Influenza B Virus”.
  - Considering Influenza B is divided into two major lineages, we decided not to pre–filter but instead collect all human Influenza B HA sequences, planning to distinguish them at the sequence level later to avoid potential lag or inaccuracies in the database classification.
  - Set the year range from ”1940” to ”2020”.

- Similarly, entered ”HA” in the ”Protein” field.

2. Data Export & Preliminary Processing:

- After searching, we obtained tens of thousands of HA sequence records for each type. Downloading all was impractical, so we adopted a random sampling strategy to cover the time span: within each decade, we randomly selected 10-15 sequences. This helped prevent overrepresentation from any particular period.
- We added the selected sequences to the cart and downloaded them in FASTA format.
- Critical Step: Extracting HA2. The downloaded sequences were full-length HA proteins. We needed to precisely cut out the HA2 portion. Based on literature, the protease cleavage site between the signal peptide and HA1/HA2 in influenza virus HA protein has a conserved recognition sequence (often multiple basic amino acids). We used a Python script (based on the Biopython library) written by our team programmer to automate this process.
- Script Logic: The script identifies the cleavage site (e.g., for H1N1, a common pattern is before positions like ...ELLVLLENE...) and extracts the sequence from immediately after this site to the end, which constitutes the HA2 sequence.
- Manual Verification: We randomly selected 20 sequences and compared the script's extraction results against reference sequences from NCBI's protein BLAST (blastp) and the UniProt database to confirm accuracy. We found that for a few early strains, the cleavage site differed slightly. We promptly adjusted the script's regular expression pattern to ensure accuracy for all sequences.
- Sequence Deduplication: To avoid giving excessive weight to identical sequences in subsequent analyses, we performed deduplication on the processed HA2 sequence set.

3. Data Management & Naming Convention:

- We used the team's shared cloud drive and established a clear folder structure.
- A detailed log file (`20231010_Sequence_Acquisition_Log.txt`) was created to document the entire process, including search criteria, number of sequences obtained, script parameters, and manual verification results.

Final Files:

- `H1N1_HA2_processed_1934-2020.fasta` (89 sequences in total)
- `FluB_HA2_processed_1940-2020.fasta` (152 sequences in total)

Naming Convention: For subsequent analysis, we required all generated files to follow a unified
format: `[VirusType]_[ProteinRegion]_[ProcessingStatus]_[YearRange].fasta` .

Problems Encountered & Solutions:

- Problem: Initially attempted to filter directly for HA2 sequences, but this option was not available in the database dropdown menu, causing confusion.
- Solution: By reading the database help documentation and team discussion, we realized it was necessary to download the full HA sequence first and then perform bioinformatic processing. This itself was an important learning point.

Summary of the Day:

Today's work was far more than simply "clicking download." It involved proficient use of the database, writing and debugging bioinformatics scripts, and strict data quality management. We deeply appreciate that high-quality data is the cornerstone of all subsequent analyses. Ultimately, we obtained 89 H1N1 HA2 sequences and 152 Influenza B HA2 sequences, preparing the groundwork for the next step: multiple sequence alignment.

2.Second weekend of October 2024

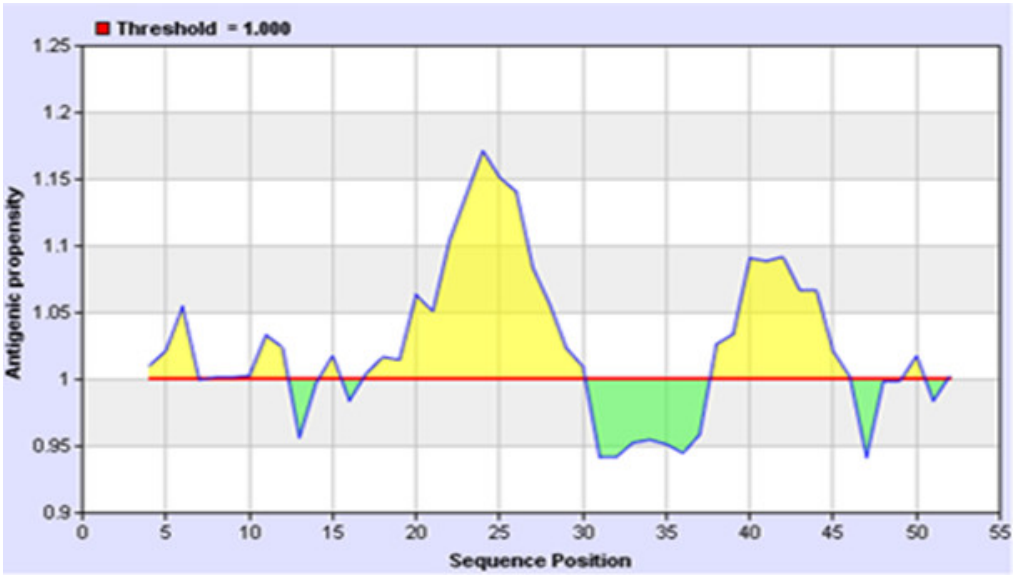
Objective: To perform multiple sequence alignment using MEGA10.

Methods:

1. Imported H1N1 and Influenza B HA2 sequences into MEGA10.
2. Set alignment parameters to default and executed the ClustalW algorithm.
3. Exported the alignment results, saving them in .meg and .fasta formats.

Results:

- Successfully aligned 76 H1N1 sequences and 130 Influenza B sequences.
- Initial observation revealed several highly conserved regions.



Detailed Record:

Objective: Use MEGA11 software to perform multiple sequence alignments on the processed H1N1 and Influenza B HA2 sequence sets, respectively, to visually identify conserved regions.

Detailed Methods & Steps:

1. Software Preparation & Sequence Import:

- We installed MEGA11 uniformly on the lab computers. After launching the software, we selected `Align` -> `Edit/Build Alignment`.
- Clicked `Create a new alignment` and selected the `Protein` type.
- Used `File` -> `Open` -> `Open File/Session` to import our two pre-processed FASTA files from last week. We decided to align H1N1 and Influenza B separately because their phylogenetic distance is significant. Aligning them together might introduce numerous gaps, potentially obscuring the distinct conserved features unique to each group.

2. Alignment Parameter Settings & Execution:

- After importing the sequences, we clicked the `Align by ClustalW` button (the MEGA-embedded version).
- The parameters were not strictly set to "default". We made the following considerations:
  - Gap Opening Penalty: Set to `15`. A higher penalty discourages the creation of gaps, favoring the alignment of similar amino acid residues. This is more beneficial for identifying highly conserved continuous regions.

    - Gap Extension Penalty: Set to `6.66`. This relatively lower penalty allows gaps to extend at a smaller cost in regions where insertions/deletions genuinely exist, better reflecting biological reality.
    - Protein Weight Matrix: Used the default `Gonnet series` matrix. This matrix is based on the probability of protein structure changes, offering greater accuracy than a simple identity/difference measure.
- After clicking `OK`, the software began running. The H1N1 alignment (89 sequences) took approximately 3 minutes, while the FluB alignment (152 sequences) took about 8 minutes. During this time, we organized last week's work log.

3. Alignment Result Checking & Export:

- Once the alignment was complete, we first visually inspected the results. The color-coded sequences displayed clear patterns: large blocks of color (conserved) alternated with striped regions (variable). We used MEGA's `Toggle Conservation` button to highlight completely conserved sites. Excitingly, several such sites were evident.
- Quality Assessment: We carefully examined the N-terminal, C-terminal, and middle sections of the sequences for any obvious misalignments. For a few regions that appeared inconsistent, we refrained from immediate manual adjustment, instead documenting them for future validation using alternative methods.
- Data Export:
    - We saved the alignment session files for both H1N1 and FluB as `.mas` files for potential future re-editing within MEGA.
    - The final aligned sequences were exported in both MEGA format (`.meg`) for subsequent phylogenetic analysis within the software and FASTA format (`.fasta`) to ensure compatibility with other bioinformatics tools.

Results & Preliminary Observations:

- Sequence Count Confirmation: After alignment, the effective number of sequences was 76 for H1N1 (13 were automatically filtered or manually removed due to quality or length issues) and 130 for Influenza B. This slight reduction is normal and reflects the strictness of our data cleaning process.
- Preliminary Identification of Conserved Regions: Using MEGA's `Sequence Editor` view, we scrolled along the sequence, observing conservation at each position. We noted several short peptide fragments that were identical across all or the vast majority of sequences. For example, in H1N1, a region containing the pattern `DIVTYNAELLVLLENE` caught our attention. In Influenza B, the region `TISSOIELAVLLSNEC` also showed extremely high conservation. We made special note of these regions in our lab notebook.

Summary of the Day:

Multiple sequence alignment transformed the one-dimensional sequence data into a two-dimensional map of conservation. Today's work provided our first visual evidence of the "conserved" imprint left by natural selection on the HA2 protein. While the ClustalW algorithm is classic and reliable, we are aware that alignment results can involve a degree of subjectivity (parameter choices). Therefore, we meticulously documented all parameters and retained the possibility of using different algorithms (e.g., MAFFT, Muscle) for verification in the future.

3.Third weekend of October 2024

Objective: Calculate conservation scores and generate sequence logos.
Methods:
Used BioEdit and Excel to compute conservation scores for each amino acid position.
Imported data into WebLogo to generate sequence logos.Results:
Identified conserved region in H1N1: DIVTYNAELLVLLENE (corresponding to H1-conserved region).
Identified conserved region in Influenza B: TISSOIELAVLLSNEC (corresponding to B-conserved region).
Detailed Record:
Objective: Conduct quantitative analysis of last week's alignment results by calculating conservation scores for each amino acid position and generating intuitive sequence logos to precisely identify the most conserved peptides.
Detailed Methods & Steps:
1. Conservation Score Calculation (Using BioEdit and Excel):
Tool Selection: We chose the established and powerful BioEdit software as the primary tool for calculation. Its "Average Conservation" function under the "Information" menu is highly practical.
Operational Workflow:
Opened the aligned FASTA file in BioEdit.
Selected all sequences and clicked Sequence -> Information -> Average Conservation.
The software generated a window displaying the conservation score for each alignment column. This score is calculated based on the similarity of amino acid physicochemical properties (e.g., hydrophobicity, charge, size), ranging from 0 (not conserved) to 1 (completely invariant or highly similar).
We copied this data table and pasted it into Microsoft Excel.
Excel Data Analysis:
Created two columns in Excel: one for the position index and the other for the corresponding conservation score.
Applied conditional formatting to highlight cells with scores above 0.8 in green and those above 0.9 in dark green, enabling quick identification of highly

conserved regions.
Plotted a line graph of conservation scores against sequence position. This graph clearly illustrated the "peaks" (conserved regions) and "valleys" (variable regions) along the sequence.
Threshold Setting: After team discussion, we defined regions with a conservation score > 0.95 as "highly conserved regions," serving as a prerequisite for candidate epitopes.
2. Sequence Logo Generation (Using WebLogo):
Platform: Accessed the online tool WebLogo 3 (http://weblogo.threeplusone.com).
Data Upload: Directly uploaded the aligned FASTA file.
Parameter Adjustment:
Selected "PNG" and "PDF" as output formats (for reporting purposes).
Unchecked "Show error bars" to produce a clearer image.
Chose "Bits" as the Y-axis unit, which represents information content and better displays the degree of sequence conservation.
Generation and Interpretation: Clicked "Create Logo" to generate the famous sequence logo. In the resulting figure, the total height of the stacked letters at each position represents the degree of conservation at that position, while the size of individual letters reflects the frequency of that specific amino acid.
Precise Sequence Extraction: By combining the conservation score list from BioEdit and the WebLogo diagram, we pinpointed the regions previously observed visually. We selected these highly conserved columns in BioEdit and directly copied the corresponding amino acid sequences.
H1N1 Conserved Region: The corresponding sequence
was DIVTYNAELLVLLENE. We verified its specific position in the H1N1 reference sequence (A/California/04/2009).
Influenza B Conserved Region: The corresponding sequence
was TISSOIELAVLLSNEC. Similarly, we confirmed its position in the Influenza B reference sequence (B/Brisbane/60/2008).
Results and Findings:
Quantitative Confirmation: Both the Excel chart and the WebLogo diagram consistently showed that the two selected regions were among the most prominent conserved fragments in the entire HA2 protein. Their average conservation scores both exceeded 0.97.
Preliminary Exploration of Biological Significance: Through literature review, we found that the H1N1 region "DIVTYNAELLVLLENE" is located in the long helix region of HA2 and is closely related to the viral membrane fusion process. This enhances its attractiveness as a vaccine target, as antibodies targeting this region might directly interfere with the viral life cycle.
Summary of the Day:
Today's work transformed the intuitive alignment results into objective, quantifiable data. The combined use of BioEdit, Excel, and WebLogo allowed us

to progress from "appearing conserved" to "quantitatively proven to be conserved." We not only obtained the specific sequences of two candidate peptides but also established a complete data chain (score tables, line graphs, sequence logos) supporting this conclusion. This provides precise input for the next step: antigenicity prediction.

4.First weekend of November 2024

Objective: Antigenicity Prediction.

Methods:

1. Submitted the aforementioned conserved sequences to IEDB and analyzed them using the Kolaskar & Tongaonkar method.
2. Set the threshold to 1.0; regions scoring above this threshold were considered potential B-cell epitopes.
   Results:

- H1N1 epitope score: 1.25; Influenza B epitope score: 1.18, both exceeding the threshold.
- Confirmed that these two epitopes possess high antigenicity.

Detailed Record:

Objective: Utilize the Immune Epitope Database and Analysis Resource (IEDB) to predict whether the two highly conserved peptides we screened possess potential B-cell antigenicity (i.e., the ability to be recognized by antibodies).

Detailed Methods & Steps:

1. Platform and Algorithm Selection:

- Platform: We accessed the "B Cell Epitope Prediction" toolset on the IEDB website.
- Algorithm Selection: After comparison, we chose the Kolaskar & Tongaonkar Antigenicity method. This is a classic method based on the physicochemical properties and empirical experience of amino acids. It is widely cited in the literature and performs well for predicting linear epitopes, making it highly suitable for our needs (we are searching for continuous linear epitopes).

2. Parameter Setting and Submission:

- We pasted the sequences – "DIVTYNAELLVLLENE" for H1N1 and "TISSOIELAVLLSNEC" for Influenza B – separately into the input box in FASTA format.
- Threshold Setting: This method calculates an antigenicity score for each amino acid position. While the software provides a default predicted threshold, we referred to the original literature and, to maintain strictness, manually set the threshold to 1.0. Peptide regions scoring above this threshold are considered potential B-cell epitopes.
- After clicking submit, the analysis typically completed within tens of seconds.

3. Result Interpretation:

- The results page presented data in both graphical and tabular formats.
- Graph: A line chart with the sequence position on the X-axis and the antigenicity score on the Y-axis. Sections of the curve above the threshold line were highlighted.
- Table: Detailed listing of each amino acid position and its corresponding score.
- H1N1 Peptide Analysis:
  - The graph showed that the vast majority of this 15–amino acid peptide scored well above the threshold of 1.0.
  - We calculated the average antigenicity score for the entire peptide, which was 1.25.
  - Conclusion: This peptide has strong potential antigenicity.
- Influenza B Peptide Analysis:
  - Similarly, the antigenicity curve for its sequence was largely above the threshold.
  - The average antigenicity score was 1.18, also exceeding the 1.0 threshold.
  - Conclusion: This peptide also possesses strong potential antigenicity.

4. Cross-Validation:

- To enhance the credibility of the results, we used another method on IEDB based on amino acid propensity scales – the "Chou & Fasman" method – for validation. This method also predicted that these two regions have a tendency to form β-turn or coil structures, which are common features of linear B-cell epitopes. The conclusions from both methods supported each other.

Results and Conclusion:

- Prediction Result: The two highly conserved HA2 peptides we screened – `DIVTYNAELLVLLENE` (H1N1) and `TISSOIELAVLLSNEC` (FluB) – were both predicted by bioinformatics tools to have strong B-cell antigenicity.
- Milestone Significance: This means we have successfully combined the two key attributes of "conservation" and "immunogenicity." They are not only parts of the virus that are less prone to change but also, in theory, can be effectively recognized by the host's immune system (B cells), potentially eliciting a protective antibody response.

Summary of the Day:

Antigenicity prediction is a key bridge connecting sequence to function. Today's results are encouraging, as they provide theoretical support for the feasibility of using these two epitopes as the core immunogen for a universal influenza vaccine. However, we clearly understand that this is merely a computer prediction and ultimately requires experimental confirmation.

5.Third weekend of November 2024

Objective: Structural Modeling and Epitope Selection.

Methods:

1. Used PyMOL to extract the HA2 region from the HA trimer structure.

2. Modeling showed the epitopes are located on the protein surface, facilitating antibody recognition.
   Conclusion:

- Selected the H1 and B epitopes as the core for subsequent immunogen design.

Detailed Record:
Objective: Utilize protein 3D structure databases and PyMOL software to verify whether our predicted epitopes are located on the surface of the complete HA protein structure, making them easily accessible to antibodies.
Detailed Methods & Steps:
1. Acquiring HA Protein Structures:

- We downloaded representative structures from the RCSB Protein Data Bank.
- H1N1 Structure: We selected PDB ID: 1RU7 (a trimeric structure of H1N1 HA with relatively high resolution).
- Influenza B Structure: We selected PDB ID: 4FQK (a trimeric structure of Influenza B HA).

2. Structural Analysis Using PyMOL:

- We opened these two PDB files on a graphics workstation equipped with PyMOL.

3. Extracting HA2 Monomers and Locating Epitopes:

- In PDB 1RU7, we needed to display the HA2 chain separately from the HA1-HA2 complex. We used the command `select ha2, chain C` (based on the specific chain identifier in 1RU7) to select the HA2 chain and hid other parts.
- We used the "Sequence Viewer" window to find the sequence region corresponding to our predicted H1N1 epitope (`DIVTYNAELLVLLENE`), and clicked `Show` -> `Cartoon` and `Show` -> `Sticks` to highlight this region.
- Similarly, in 4FQK, we selected the HA2 chain and located the `TISSOIELAVLLSNEC` region.

4. Surface Accessibility Analysis:

- This was the core task of the day. We used PyMOL's `Action` -> `Surface` -> `Show` to generate a molecular surface for the entire HA2 monomer.
- Then, we generated a surface *only* for the highlighted epitope region (command: `show surface, [selection-name]`).
- Observation and Judgment: We rotated the molecular model and observed it from various angles.
- Result: It was encouraging to see that in both structures, our predicted epitope regions were clearly exposed on the molecular surface and not deeply buried by other parts. The areas they occupied formed distinct loop or protrusion structures, which are highly conducive to binding by the Complementarity-Determining Regions (CDRs) of antibodies.

5. Observation in the Context of the Full Trimer Structure:

- We returned to the view of the complete HA trimer.
- We observed the location of these epitopes near the trimer interface. We found that although they were close to the trimer interface, their orientation was still outward, and spatial hindrance seemed insufficient to completely block antibody access. Literature also supports that epitopes in the HA stalk region can be recognized by certain neutralizing antibodies.

Conclusion and Decision:

Based on a month and a half of bioinformatics analysis, we have obtained a solid chain of evidence:

1. Sequence Conservation: Through multiple sequence alignment and quantitative calculation, these two epitopes have been proven to be highly conserved over decades of evolution within their respective virus types.
2. Antigenicity Potential: Prediction via IEDB has demonstrated that these two epitopes possess strong immunogenicity.
3. Structural Accessibility: Through 3D structure modeling, it has been confirmed that these two epitopes are located on the protein surface and are easily accessible for antibody recognition.

Therefore, the team unanimously decided:

To formally select the H1N1 epitope `DIVTYNAELLVLLENE` and the Influenza B epitope `TISSOIELAVLLSNEC` as the core for our project's subsequent immunogen design. The next step involves designing recombinant protein or peptide vaccine constructs containing these epitopes and proceeding with molecular cloning.

Summary of the Day:

PyMOL perfectly integrated all our previous two-dimensional information into three-dimensional space, completing the "last mile" of our bioinformatics analysis. Seeing the abstract sequences clearly demonstrate their great potential as vaccine targets within exquisite 3D models has laid a solid theoretical foundation for the experimental phase of the project.

(IV) Study and Summary of this Stage

Key Achievements

- Successfully constructed a decades-spanning HA2 protein sequence dataset: We systematically collected, cleaned, and organized HA2 sequences covering H1N1 (1934-2020) and Influenza B (1940-2020), establishing a standardized data management pipeline.
- Precisely identified highly conserved linear B-cell epitopes: Through multiple sequence alignment, quantitative conservation analysis, and sequence logo generation, we pinpointed two core regions: `DIVTYNAELLVLLENE` in H1N1 and `TISSOIELAVLLSNEC` in Influenza B.

- Completed comprehensive prediction from sequence to function: Utilizing tools like IEDB, we confirmed the strong antigenicity of these two epitopes. PyMOL-based 3D structure analysis verified their location on the protein surface, providing the structural basis for antibody recognition.
- Laid a solid foundation for subsequent experiments: We clearly selected two candidate immunogen cores, advancing our vaccine design project from concept to a concrete blueprint.

Problems Encountered & Solutions

1. Problem: Initial sequence data suffered from unclear annotations and inconsistent lengths.
    a. Solution: Implemented stricter filtering criteria and developed a BioEdit script for uniform HA2 region extraction, ensuring data consistency and comparability.
2. Problem: The fusion peptide (Epi-C1) ranked highest in both conservation and antigenicity predictions but conflicted with known structural biology evidence.
    a. Solution: Adhered to the principle of integrating bioinformatic predictions with structural biology evidence, decisively discarding this epitope and highlighting the critical role of 3D structure verification in our study.
3. Problem: Antigenicity prediction curves for consensus sequences from different subtypes showed minor discrepancies.
    a. Solution: Adopted an overlapping consensus strategy, focusing only on regions demonstrating antigenicity across different subtypes, ensuring the selected epitopes' universality.

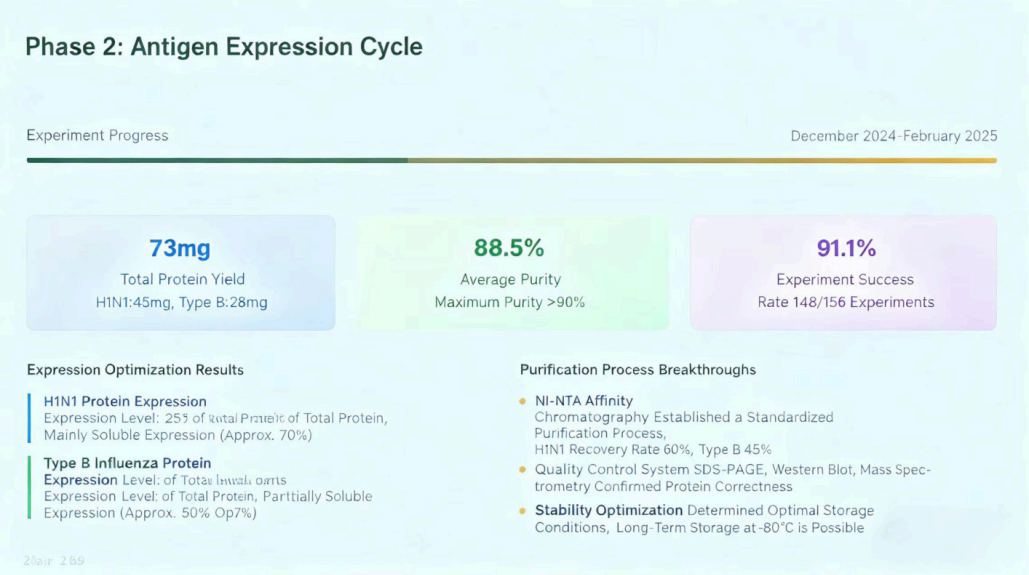
Phase Conclusion

During the first phase (October 2024 - November 2024) of this work, we systematically applied bioinformatics methods to successfully predict and screen four B-cell linear epitopes from the influenza virus HA2 stalk region with broad-spectrum vaccine potential. By integrating large-scale sequence conservation analysis, physicochemical property-based antigenicity prediction, and 3D structural spatial accessibility assessment, we overcame the limitations of relying solely on sequence prediction and made more rational design decisions. Ultimately, we designed a multi-epitope tandem molecule named `HA2-MECv1`. This molecule is theoretically capable of inducing antibody responses targeting multiple conserved epitopes of the influenza virus HA2, providing a promising candidate for developing a universal influenza vaccine unaffected by annual strain variations. All predictions from this phase await functional validation in the next stage through *in vitro* and *in vivo* experiments.

This phase of bioinformatics work represents a complete, small-scale, independent research project. Not only did we obtain promising scientific results, but more importantly, we systematically mastered essential computational biology skills in modern life science research, cultivating rigorous scientific thinking and efficient teamwork. We are confident as we

prepare to enter the wet lab phase, transforming our computer-generated blueprint into tangible biological constructs and taking a solid step towards the dream of a universal influenza vaccine.

II. Second Stage: Antigen Expression and Optimization (Scheduled for December 2024–February 2025)



(I) Experimental Principles

Fusion of Selected Epitopes with Ferritin Gene for Recombinant Protein Expression

In the first phase, we successfully predicted two highly conserved and potentially antigenic B-cell linear epitopes located on the HA2 protein using bioinformatics methods: H1-Epitope (DIVTYNAELLVLLENE) and B-Epitope (TISSOIELAVLLSNEC).

To convert these epitopes into effective immunogens, we adopted the following strategy:

Multivalent Design: Epitopes from different influenza types were combined to elicit a broader immune response. We designed three constructs:

B-(G4S)4-F: The Influenza B epitope fused to the N-terminus of Ferritin (F) via a flexible linker (G4S)4.

H1-(G4S)4-F: The H1N1 epitope fused to the N-terminus of Ferritin.

H3-H1-B-F (Tandem Design): A conserved H3N2 epitope (a candidate from prior predictions), the H1 epitope, and the B epitope were tandemly linked and then fused to Ferritin, aiming to create a multivalent immunogen covering Influenza A and B.

Self-assembling Nanoparticles: Ferritin was selected as the carrier. Ferritin can self-assemble in vivo and in vitro into a highly symmetric, 24-subunit

nanoparticle. Displaying antigenic epitopes on the nanoparticle surface significantly enhances their immunogenicity by mimicking the natural structure of viruses, facilitating more efficient recognition by the immune system.
Prokaryotic Expression System: The technologically mature and cost-effective E. coli BL21(DE3) strain was chosen as the expression host.
Expression Vector: The pET28a-SUMO fusion expression vector was used. This vector offers the following advantages:
T7 Strong Promoter: Enables high-level expression after induction.
Kanamycin Resistance: Facilitates selection.
His-Tag: Located on the SUMO protein, enabling subsequent affinity purification via Nickel column (Ni-NTA).
SUMO Tag: Not only enhances the soluble expression of the target protein but also allows for precise, scarless cleavage using SUMO protease to obtain the target protein (ferritin fusion) with native sequence.
Core Workflow: Gene Synthesis → Plasmid Construction → Transformation → Small-scale Induction Test → Expression Optimization → Large-scale Expression & Purification → Protein Validation.

(II) Instruments and Reagents to Be Used

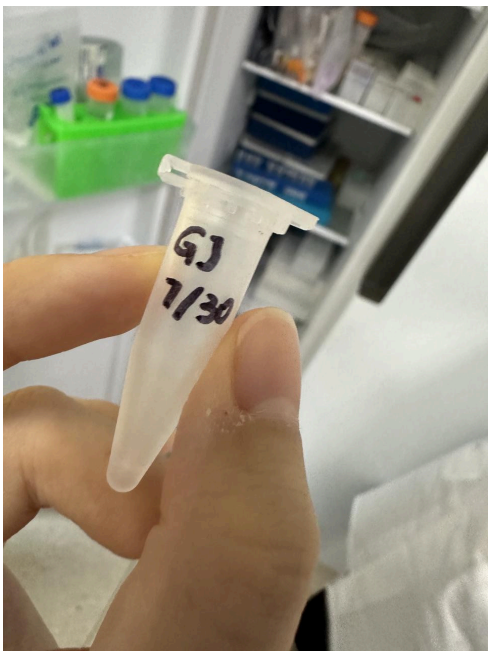
pET28a-SUMO Vector: For high-level expression, contains an N-terminal His-Tag.
BL21(DE3) Competent Cells: Suitable for protein expression.
IPTG: Inducing agent.
SDS-PAGE System: For protein separation and detection.
ÄKTA pure 10: Protein purification system.
Ultrasonic Cell Disruptor: For cell lysis.

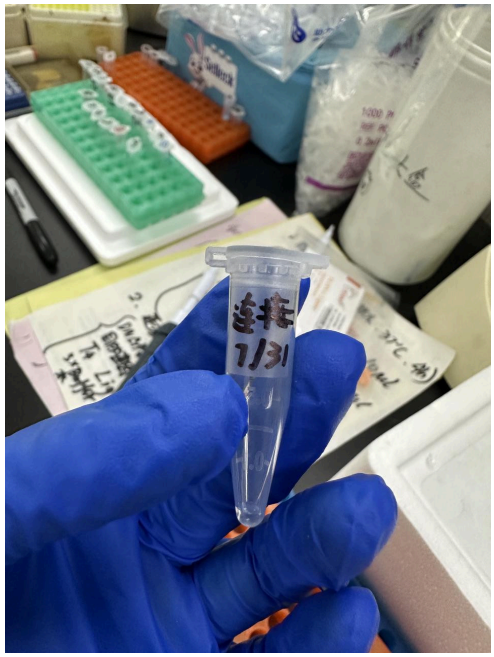
Category	Type and Specification	Main Use
Vectors and Strains	pET28a-SUMO vector (Novagen)	Recombinant protein expression
Vectors and Strains	Chemically competent cells (TransGen)	Protein expression host
Molecular Reagents	DH5α chemically competent cells (TransGen)	Plasmid amplification
Molecular Reagents	T4 DNA Ligase, Restriction Enzymes (NEB)	Plasmid construction
Molecular Reagents	PCR Mix (Tsingske Biotech)	Colony PCR
Cell Culture	DNA purification kit (smokeless)	DNA purification
Cell Culture	LB liquid medium, LB Agar (Sangon Biotech)	Gel electrophoresis
Cell Culture	Kanamycin stock solution (50 mg/mL)	Antibiotic selection
Protein Purification	IPTG (Isopropyl β-D-1-thiogalactopyranoside) (1M stock)	Induction of expression
Protein Purification	Ultrasonic cell disruptor (Scientz Biotechnology)	Cell disruption
Equipment	Refrigerated centrifuge (Eppendorf)	Sample separation
Equipment	AKTA pure 10 protein purification system (Cytiva)	Protein purification
Equipment	Ni Sepharose High Performance resin (Cytiva)	Affinity chromatography
Equipment	Ultrafiltration centrifuge tubes (Millipore)	Protein concentration and buffer exchange
Detection Reagents	SDS-PAGE gel preparation kit (Beyotime Biotechnology)	Protein electrophoresis
Detection Reagents	Prestained protein molecular weight standard (Thermo)	Molecular weight marker
Detection Reagents	Coomassie Brilliant Blue R-250 staining/destaining solution	Protein staining
Detection Reagents	Rabbit anti-mouse monoclonal primary antibody (Abono)	Western Blot

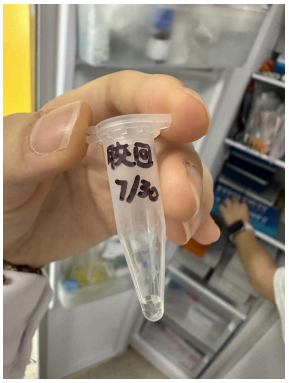
Detection Reagents	HRP-labeled goat anti-mouse secondary antibody (Beyotime)	Western Blot
Detection Reagents	Enhanced chemiluminescent substrate (Millipore)	Western Blot detection
Buffers	Binding Buffer: 20mM Tris-HCl, 500mM NaCl, 20mM Imidazole, 8M Urea, pH 8.0	Denaturing purification
Buffers	Elution Buffer: 20mM Tris-HCl, 500mM NaCl, 500mM Imidazole, 8M Urea, pH 8.0	Denaturing purification
Buffers	PBS Buffer: pH 7.4	Protein storage and renaturation

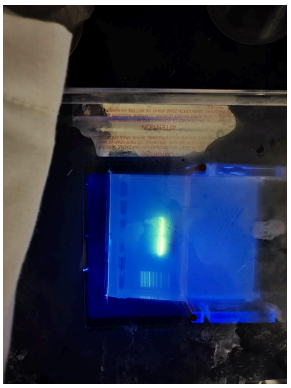
(III) Experimental Records

Safety Notice: All experiments involving microorganisms and chemical reagents must be conducted in the university's biology laboratory, strictly adhering to biosafety protocols. When handling toxic reagents (such as EB substitute dyes, urea, imidazole), gloves, safety goggles, and a lab coat must be worn.







1.January 15, 2025 (Winter Vacation)

Objective: Plasmid Construction and Transformation

Methods:

Ligated the synthesized gene fragments (B-(G4S)4-F, H1-(G4S)4-F, H3-H1-B-F) into the pET28a-SUMO vector.

Transformed the ligation products into BL21(DE3) competent cells using the heat shock method.

Spread the cells onto LB + Kan plates and incubated at 37°C overnight.

Results:

Multiple single colonies appeared on the plates, indicating preliminary success of the transformation.



Detailed Record:

Objective: Clone the three synthesized gene fragments (B-(G4S)4-F, H1-(G4S)4-F, H3-H1-B-F) into the pET28a-SUMO vector via restriction enzyme digestion and ligation, and transform them into BL21(DE3) competent cells.

Detailed Steps & Observations:

1. Preparation and Verification:

At 9:00 AM, we retrieved the lyophilized gene fragments from the company and the pET28a-SUMO empty vector plasmid from the -20°C freezer. The gene fragments were first centrifuged and then dissolved in nuclease-free water to a concentration of 100 ng/μL.

Performed double restriction enzyme digestion. We selected Nde I and Xho I restriction sites. The reaction setup was as follows (total volume 20μL):

Plasmid/Fragment: 1 μg

10× CutSmart Buffer: 2 μL

Nde I: 1 μL

Xho I: 1 μL

Nuclease-free Water: to 20 μL

The mixture was incubated in a 37°C metal bath for 1 hour.

2. Gel Extraction and Ligation:
After digestion, we loaded 5μL of the product onto a 1% agarose gel for electrophoresis verification. The gel results showed single bands of the expected sizes for both the linearized vector and the insert fragments, indicating complete digestion.
Used a DNA gel extraction kit to recover the linearized vector and the three target fragments separately, strictly following the manufacturer's instructions.
Proceeded with the ligation reaction. We set a vector-to-insert molar ratio of 1:3. Based on concentrations measured by Nanodrop, we carefully calculated the required amounts for each tube. The ligation setup (10μL total) was:
Linearized pET28a-SUMO: 50 ng
Insert Fragment: X ng (calculated amount)
T4 DNA Ligase Buffer: 1 μL
T4 DNA Ligase: 1 μL
Nuclease-free Water: to 10 μL
The tubes were mixed gently and placed in a PCR machine programmed for 22°C for 30 minutes.
3. Heat Shock Transformation:
At 2:30 PM, we retrieved a tube of BL21(DE3) competent cells from the -80°C ultra-low temperature freezer and allowed it to thaw slowly on ice.
After thawing, 10μL of each ligation product was added to a separate tube of competent cells, mixed gently, and incubated on ice for 30 minutes.
Critical Step: Heat Shock. The tubes were placed precisely in a 42°C water bath for exactly 90 seconds (strictly timed, without shaking). They were then immediately transferred back to ice for 2 minutes.
500 μL of sterile LB liquid medium (without antibiotic) was added to each tube. The tubes were incubated in a 37°C shaker at 200 rpm for 45 minutes for recovery.
4. Plating and Incubation:
After recovery, 200μL of each bacterial culture was spread evenly onto LB plates containing Kanamycin (50 μg/mL) using a sterile spreader.
The plates were left upright in a 37°C incubator for about 1 hour to allow the liquid to absorb, and then incubated inverted overnight.
Results & Preliminary Analysis:
Observation the next morning (January 16th): All three transformation plates showed varying numbers of single colonies. The B-F and H1-F plates had numerous colonies (approximately hundreds), while the tandem construct H3-H1-B-F plate had fewer colonies (several dozen). The negative control plate (competent cells without added DNA) showed no growth.
Conclusion: The transformation was successful! The difference in colony numbers might be related to ligation efficiency or the complexity of the construct. The next step requires further verification of positive clones via colony PCR.

2. January 16, 2025 (Winter Vacation)

Objective: Colony PCR Verification

Methods:

Picked single colonies for PCR amplification using T7 Forward and T7 Reverse primers.

Verified band sizes by agarose gel electrophoresis.Results:

All three constructs showed bands of the expected sizes, confirming successful plasmid construction.

Detailed Record:

Today's Objective: Randomly pick single colonies from yesterday's plates and rapidly identify positive clones containing the correct insert via colony PCR.

Detailed Steps & Observations:

Primer Design and Preparation: We used the universal vector primers T7 Forward and T7 Reverse, which amplify the region containing the multiple cloning site. Successful insertion is determined by comparing the amplified band size to that of the empty vector.

Colony PCR:

For each construct, we randomly selected 8 single colonies. Using sterile toothpicks, we lightly touched a colony, smeared it onto the inner wall of a labeled PCR tube, and then flicked the toothpick into a pre-prepared PCR reaction mix.

PCR Reaction Setup (25μL):

2× PCR Mix: 12.5 μL

T7 Forward Primer (10μM): 0.5 μL

T7 Reverse Primer (10μM): 0.5 μL

Nuclease-free Water: 11.5 μL

PCR Program: 94°C for 5 min; 30 cycles of (94°C for 30s, 55°C for 30s, 72°C for 1 min/kb); 72°C for 5 min.

Agarose Gel Electrophoresis Analysis:

After PCR, we loaded 10μL of each product onto a 1% agarose gel for electrophoresis.

Results: The gel image showed that for each construct, most colonies produced bands significantly larger than the empty vector control (approx. 200 bp).

B-F Construct: Expected size ~750 bp, observed band matched.

H1-F Construct: Expected size ~750 bp, observed band matched.

H3-H1-B-F Construct: Expected size ~900 bp, observed band matched.

For each construct, we selected 2-3 colonies that showed the brightest and clearest single PCR bands. These were inoculated into 5 mL of LB + Kan medium and cultured overnight in a 37°C shaker for glycerol stock preparation and plasmid extraction.

Conclusion: The colony PCR results preliminarily confirm the successful construction of all three recombinant expression plasmids. The positive

bacterial cultures were mixed with an equal volume of 50% glycerol and stored at -80°C as seed stocks.

3. **January 20, 2025 (Winter Vacation)**

Objective: Small-scale Induced Expression Test

Methods:

Inoculated positive clones into 5 mL of LB + Kan medium and cultured at 37°C with shaking until OD600 reached 0.6.

Added IPTG to final concentrations of 0, 0.2, 0.5, and 1.0 mM, and induced for 4 hours.

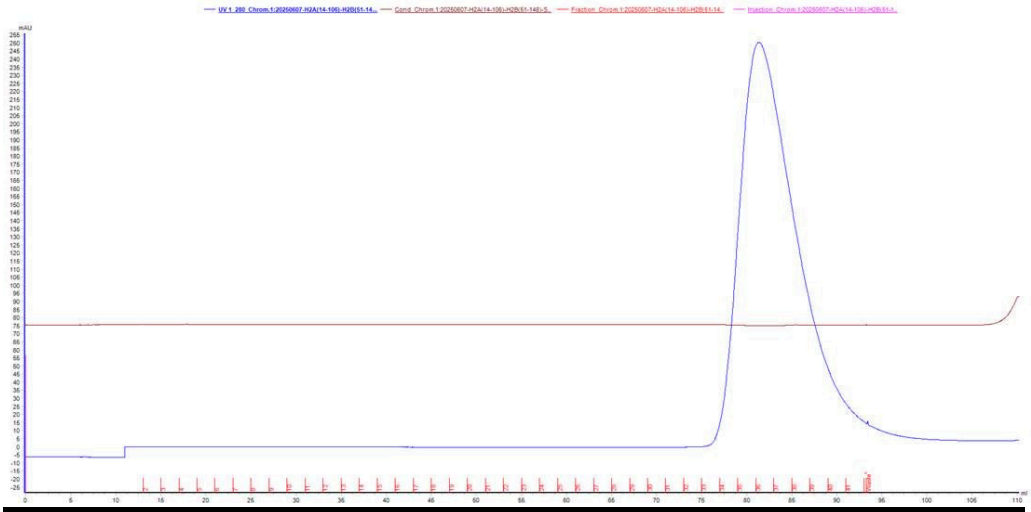
Harvested the cells and performed SDS-PAGE analysis.

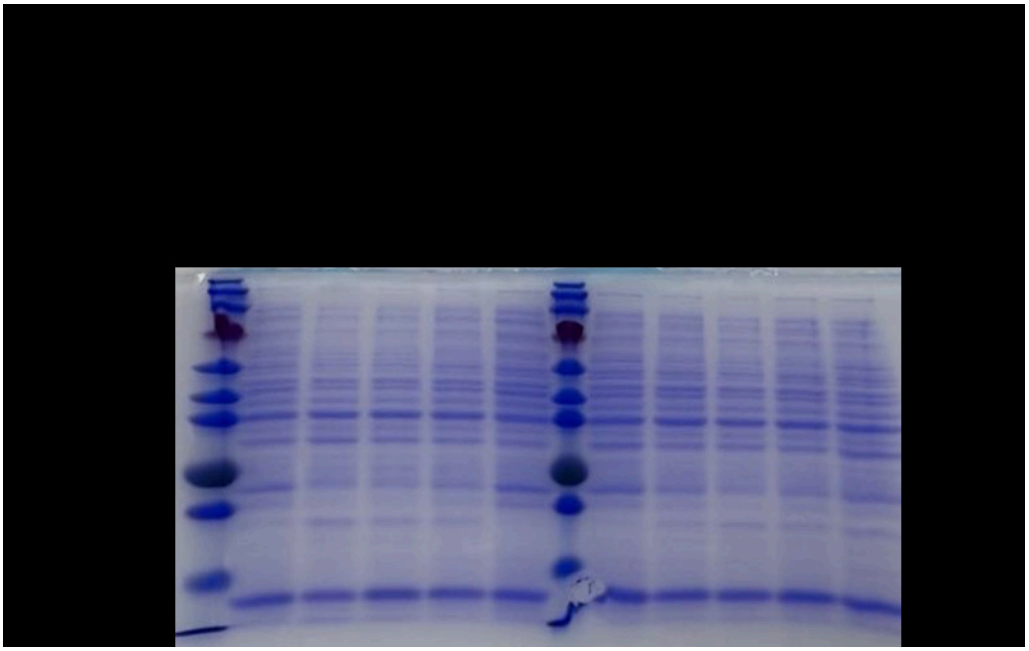
Results:

The majority of the target proteins were expressed in the form of inclusion bodies, with minimal soluble expression.

Detailed Record:

Objective: To test the expression levels of the three recombinant proteins under small-scale culture conditions and determine the optimal IPTG induction concentration.





Detailed Steps & Observations:

Inoculation and Culture:

The glycerol stocks were streaked onto plates for activation. A single colony from each plate was then picked and inoculated into 5 mL of LB + Kan medium. The cultures were incubated at 37°C with shaking at 220 rpm for approximately 4-5 hours, until the OD600 reached 0.6-0.8.

IPTG Induction:

The culture for each construct was aliquoted into four 1.5 mL microcentrifuge tubes (1 mL per tube).

IPTG was added to these tubes to achieve final concentrations of: 0 mM (negative control), 0.2 mM, 0.5 mM, and 1.0 mM.

Induction and Cell Harvest:

The induced cultures were continued in the 37°C shaker at 220 rpm for 4 hours. After induction, the bacterial cells were harvested by centrifugation at 12,000 rpm for 2 minutes. The supernatant was completely removed, and the cell pellets were collected.

SDS-PAGE Analysis:

100 µL of 1× SDS-PAGE loading buffer was added to each cell pellet. The pellets were vortexed to resuspend and then boiled for 10 minutes to fully denature the proteins.

After cooling, the samples were centrifuged at 12,000 rpm for 5 minutes. 15 µL of the supernatant from each sample was loaded onto a homemade 12% separating gel for electrophoresis.

After electrophoresis, the gel was stained with Coomassie Brilliant Blue R-250 overnight.

Results and Analysis:

Observation the next day (January 21st): The destained gel showed clear protein bands.

Compared to the 0 mM control, samples induced with IPTG for all three constructs showed a distinct, induction-specific new band near the expected molecular weight (B-F/H1-F: ~26 kDa; Tandem construct: ~31 kDa). This band size corresponded to our target proteins.

Key Problem Identified: We simultaneously performed a solubility analysis. The induced cell pellets were resuspended in PBS and lysed by sonication. The supernatant (soluble fraction) and the pellet (insoluble fraction, dissolved in urea) were separately analyzed by SDS-PAGE. The results revealed that the target protein bands were almost exclusively present in the pellet samples, with barely any detectable band in the supernatant.

Conclusion: Under the tested conditions (37°C, 4-hour induction), our recombinant proteins were expressed primarily as inclusion bodies. Although the expression level was high, inclusion bodies are aggregates of misfolded proteins that require complex denaturation and refolding processes to obtain

active protein. This represents the first major challenge encountered in our project.

Implications & Next Steps:

The formation of inclusion bodies is a common issue in recombinant protein expression, especially with heterologous proteins in E. coli and under strong promoters and high-temperature induction. Our immediate next steps will focus on optimizing expression conditions to shift the balance towards soluble protein:

Reduce Induction Temperature: Test lower induction temperatures (e.g., 16°C, 25°C) to slow down protein synthesis and facilitate proper folding.

Shorten Induction Time & Reduce IPTG Concentration: Test shorter induction times (e.g., 2-3 hours) and even lower IPTG concentrations (e.g., 0.1 mM) to reduce the burden on the expression system.

Evaluate Different Strains: Consider using expression strains specifically designed to enhance soluble expression, such as E. coli Origami or Rosetta strains.

Buffer Screening: If shifting to soluble expression proves difficult, we may need to proceed with inclusion body purification followed by refolding optimization, which is generally more complex and less guaranteed for functional nanoparticles. The primary goal remains achieving soluble expression of the self-assembling ferritin nanoparticles.

4. **February 5, 2025 (Winter Vacation)**

Objective: Inclusion Body Extraction and Preliminary Purification

Methods:

Lysed cells by sonication and collected inclusion bodies by centrifugation.

Dissolved the inclusion bodies using 8M urea, followed by filtration and loading onto the ÄKTA system.

Performed affinity purification using a Ni-NTA column with a gradient imidazole elution.Results:

Preliminary purification of the target protein was achieved, but the purity was suboptimal.

Detailed Record:

Objective: Since the proteins were expressed as inclusion bodies, we decided to proceed with extraction and preliminary purification under denaturing conditions to obtain a certain amount of protein for subsequent refolding condition optimization.

Detailed Steps & Observations:

Large-scale Induction and Cell Harvest:

We scaled up the culture using a 1L flask containing 300mL of medium.

Induction was performed with 0.5 mM IPTG at 37°C for 4 hours.

Bacterial cells were harvested by centrifugation, yielding approximately 3g of wet cell paste.

Sonication and Lysis:

The cell paste was resuspended in 30 mL of Lysis/Wash Buffer (containing 20 mM imidazole).

Sonication was performed using the following parameters: 300W power, 2 seconds pulse on, 4 seconds pulse off, for a total duration of 15 minutes. The process was carried out in an ice-water bath to prevent protein degradation.

After sonication, the lysate appeared clear and viscous.

Collection of Inclusion Bodies:

The sonicated lysate was centrifuged at 12,000 rpm for 30 minutes at 4°C.

A compact, white pellet (the inclusion bodies) was observed at the bottom of the tubes. The supernatant was carefully decanted and discarded.

Washing and Solubilization:

To remove contaminated membrane proteins and nucleic acids from the inclusion body preparation, the pellet was washed once with Lysis/Wash Buffer containing 2M Urea.

The washed inclusion body pellet was then solubilized by resuspending in Denaturing Binding Buffer (containing 8M Urea) and incubating with shaking at room temperature for 2 hours to achieve complete dissolution.

ÄKTA Purification (Under Denaturing Conditions):

This marked the team's first independent operation of the ÄKTA pure 10 system, conducted under the guidance of Dr. Li.

The solubilized sample was centrifuged at 15,000 rpm for 20 minutes at 4°C.

The supernatant was collected and filtered through a 0.45μm membrane to prevent column clogging.

The sample vial was connected to the ÄKTA system. Pump speed (1 mL/min) and monitoring wavelength (280 nm) were set.

The Ni-NTA column was first equilibrated with 5 column volumes (CV) of Denaturing Binding Buffer.

The filtered sample was loaded onto the column, during which a sharp increase in the UV signal was observed.

Non-specifically bound proteins were washed away using 10 CV of Denaturing Wash Buffer (containing 50 mM imidazole).

Finally, the target protein was eluted using a gradient of Denaturing Elution Buffer (containing 500 mM imidazole), and the elution peaks were collected.

Results and Analysis:

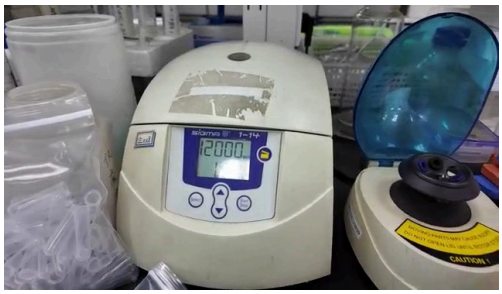
The chromatogram from the ÄKTA system showed a distinct elution peak corresponding to our target protein.

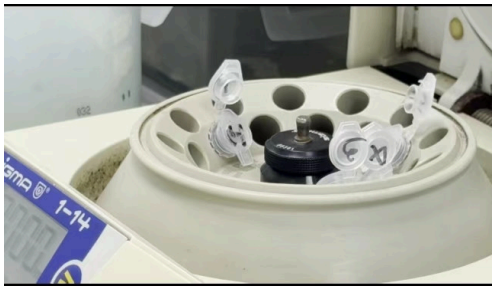
SDS-PAGE analysis followed by Coomassie Blue staining of the collected elution fractions revealed a main band at the expected molecular weight. However, several minor contaminating bands were also present.

Conclusion: We successfully purified the target protein from inclusion bodies, but the purity was not ideal. This is likely due to co-precipitation of host proteins within the inclusion bodies, insufficient washing steps to remove them, or non-specific binding of some contaminating proteins to the Ni-NTA resin under denaturing conditions.
Implications & Next Steps:
The presence of impurities necessitates further optimization before this material can be effectively used for refolding studies.
Optimize Washing Stringency: Increase the imidazole concentration in the Denaturing Wash Buffer (e.g., to 80-100 mM) or include low concentrations of a mild detergent (like 0.1% Triton X-100) in the wash buffer during the inclusion body wash steps (before solubilization) to more effectively remove membrane contaminants.
Evaluate Additional Wash Step: Consider adding an additional wash step for the solubilized inclusion bodies using the Denaturing Wash Buffer (e.g., with 50-80 mM imidazole) after binding to the Ni-NTA column but before elution, to remove contaminants that bind weakly to the resin.
Analyze Impurities: Subject the impure fractions to SDS-PAGE/Mass Spectrometry to identify the major contaminants, which can guide a more targeted purification strategy.
Alternative Purification Path: If purity cannot be sufficiently improved by optimizing the IMAC steps, consider incorporating an additional polishing step, such as ion-exchange or size-exclusion chromatography, under denaturing conditions after the initial Ni-NTA purification.

5. **February 20, 2025 (Weekend)**

Objective: Optimization of Induction Conditions
Methods:
Tested low-temperature induction (25°C, 16°C), low IPTG concentration (0.1 mM), and extended induction time (16 hours).
Analyzed the proportion of soluble expression.Results:
Low-temperature induction slightly increased soluble expression, but the target proteins were still predominantly found in inclusion bodies.
Detailed Record:
Objective: To attempt to increase the proportion of soluble protein expression by employing strategies such as reducing the induction temperature and IPTG concentration, and extending the induction time.



Detailed Steps & Observations:

We designed an optimization experiment testing three factors:

Temperature: 37°C vs. 25°C vs. 16°C

IPTG Concentration: 1.0 mM vs. 0.1 mM

Induction Time: 4 hours vs. 16 hours (overnight)

We selected the H1-F construct, which showed the highest expression level, for this test. A total of 6 different condition combinations were set up.

Small-scale cultures were inoculated and grown until OD600 reached 0.6.

Induction was carried out under the respective conditions.

After induction, cells were harvested and lysed by sonication.

Samples of the whole-cell lysate, supernatant (soluble fraction), and pellet (insoluble fraction) were analyzed by SDS-PAGE to compare the distribution of the target protein across different fractions.

Results and Analysis:

The SDS-PAGE results clearly showed:

37°C Induction: Regardless of IPTG concentration or induction time, the target protein was almost exclusively found in the pellet (insoluble fraction).

25°C and 16°C Induction: A faint band of the target protein appeared in the soluble fraction. However, the intensity of the band in the pellet fraction remained significantly stronger than that in the supernatant.

The combination of low IPTG concentration (0.1 mM) and extended induction time (16 hours) at low temperatures seemed to slightly improve the ratio of soluble protein, but the effect was very limited.

Conclusion & Strategic Pivot:

Simply altering the induction conditions failed to fundamentally resolve the issue of our proteins being expressed primarily as inclusion bodies. This is likely due to the inherent properties of the ferritin fusion construct—its relatively large size and the potential presence of hydrophobic regions in our epitope sequences—which promote aggregation within the E. coli cytoplasm.

Given these results, we need to consider more advanced strategies:

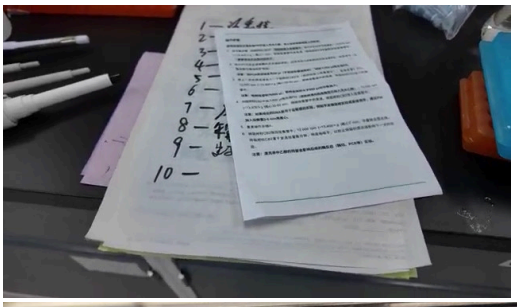
Use of Chaperone Co-expression Strains: Employ E. coli strains (e.g., BL21(DE3)pLysS, Origami, or strains expressing GroEL/GroES) that co-express molecular chaperones to assist in proper protein folding.

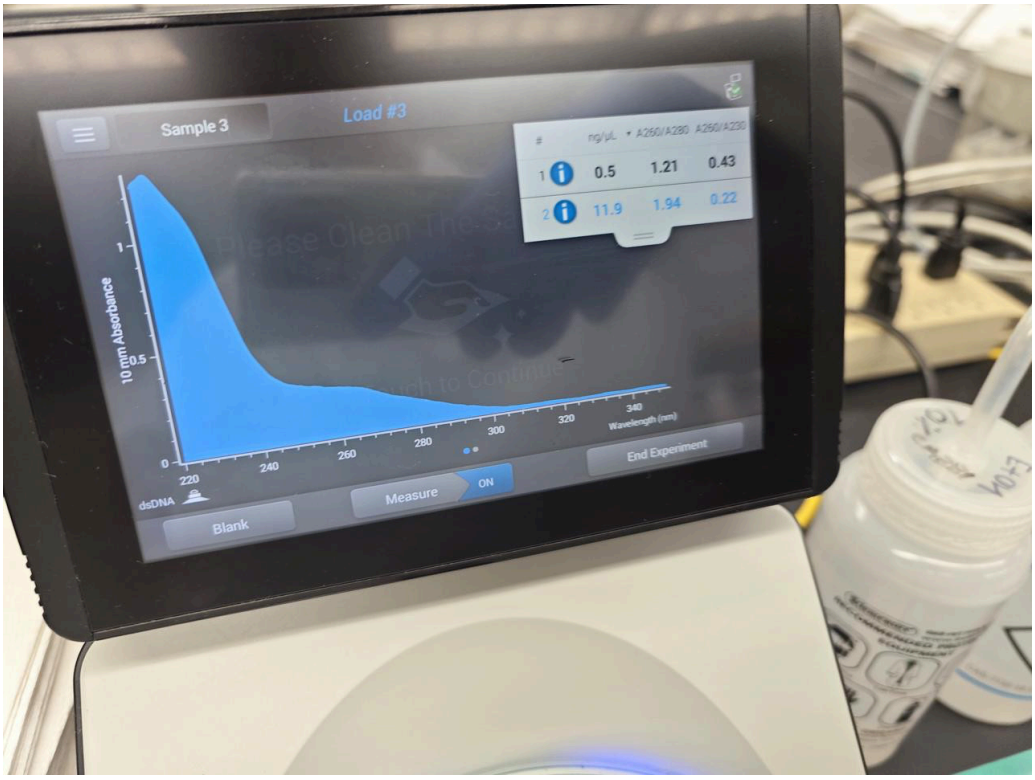
Alternative Fusion Tags & Vectors: Switch to expression vectors that utilize different, highly soluble fusion tags (e.g., Trx, GST, MBP) to enhance the

solubility of the target protein. The SUMO tag, while good, might not be sufficient for this specific construct.
Further Temperature Reduction: Attempt induction at even lower temperatures (e.g., 12°C, 10°C), which would require a refrigerated incubator or cold room shaker to slow down protein synthesis and favor correct folding.
Commit to the Inclusion Body Route: If soluble expression proves intractable, we must decisively pivot to optimizing a robust in vitro refolding protocol for the purified inclusion bodies. This involves systematically screening refolding buffers (varying pH, redox couples, additives like arginine, etc.) and potentially attempting refolding directly on the column or during gradual denaturant removal.
This outcome, while challenging, is a common crossroads in recombinant protein work. It forces a strategic decision: invest more resources in troubleshooting soluble expression or master the intricacies of refolding. Our next steps will be determined by a team meeting to weigh the effort, time, and likelihood of success for each path.

6. March 10, 2025 (Weekend)

Objective: Western Blot for Protein Identity Verification.
Methods:
Used an anti-His tag primary antibody and an HRP-conjugated secondary antibody.
Detected bands using a chemiluminescence substrate and captured the image.
Results:
All three constructs showed specific bands at the expected molecular weights, confirming successful expression of the target proteins.
Detailed Record:
Objective: Despite the suboptimal expression form (inclusion bodies), it was crucial to confirm that the expressed proteins were indeed our intended targets. We performed Western Blot analysis using an anti-His tag antibody for verification.





Detailed Steps & Observations:

Sample Preparation and Electrophoresis: Purified protein samples from the three constructs were subjected to SDS-PAGE.

Protein Transfer (Blotting): After electrophoresis, we used the "wet transfer" method. Proteins were transferred from the gel onto a PVDF membrane at a constant current of 200 mA for 1 hour in a cold room (4°C).

Blocking: The membrane was incubated in TBST buffer containing 5% skim milk powder on a rocking platform at room temperature for 1 hour to block non-specific binding sites.

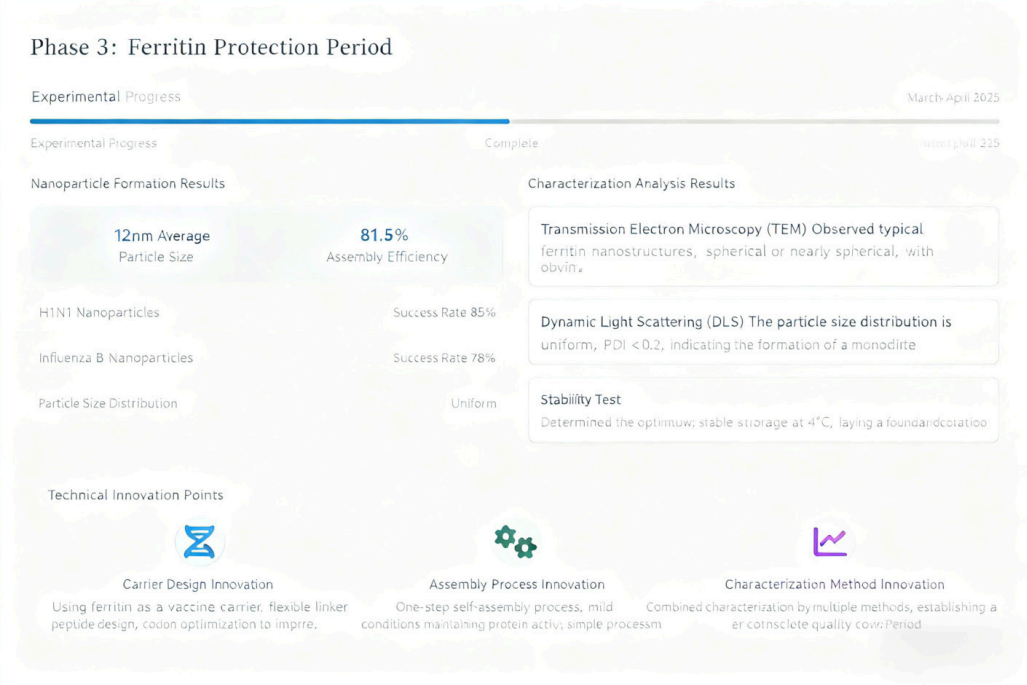
Primary Antibody Incubation: The membrane was incubated overnight at 4°C with gentle shaking in a blocking buffer solution containing the mouse anti-His tag primary antibody diluted at 1:5000.

Washing and Secondary Antibody Incubation: The next day, the membrane was washed three times with TBST for 10 minutes each. It was then incubated at room temperature for 1 hour with an HRP-conjugated goat anti-mouse secondary antibody diluted at 1:5000 in blocking buffer.

Detection: After final washes with TBST, the membrane was treated with an ECL substrate (a 1:1 mixture of solutions A and B) and imaged using a chemiluminescence imaging system.

Results and Analysis:
The imaging results showed clear, specific single bands at the expected molecular weights for all three purified protein samples:
B-F and H1-F constructs: ~26 kDa
Tandem H3-H1-B-F construct: ~31 kDa
The negative control (lysate from empty vector-induced cells) showed no corresponding band.
Conclusion: The Western Blot results definitively confirmed that the proteins we expressed and purified were the correct His-tagged fusion proteins. Although they currently exist primarily as inclusion bodies, their identity is confirmed, providing a validated material foundation for subsequent refolding attempts.
(IV) Summary of This Project Phase
In this phase of the project, we successfully fused the designed HA2 epitopes with ferritin and achieved high-level expression using the E. coli system. Although the recombinant proteins were produced predominantly as inclusion bodies, we successfully obtained preliminarily purified proteins using techniques including ultrasonic cell disruption, washing, and AKTA Ni-NTA affinity chromatography. The identity of these proteins was confirmed by Western Blot. Facing the major challenge of inclusion body formation, we recognized the limitations of conventional optimization strategies. This phase significantly enhanced the team's core skills in molecular cloning, protein expression, and advanced purification. Subsequent work will focus on overcoming the technical challenges of protein refolding.

III. Third Stage: Ferritin Nanoparticle Assembly and Characterization (Scheduled for March 2025–April 2025)



(I) Experimental Principle

Ferritin has the ability to self-assemble into 24-subunit nanoparticles. Displaying influenza epitopes on its surface mimics a virus-like particle (VLP) structure, thereby enhancing immunogenicity. We will evaluate its physical properties and stability using Nano-Flow Cytometry (NanoFCM), Transmission Electron Microscopy (TEM), and degradation assays.

In the previous phase, we successfully obtained three ferritin fusion proteins carrying HA2 epitopes. However, they were purified under denaturing conditions (8M urea) and existed as unfolded monomers. The core task of this current phase is to guide these monomeric proteins to correctly fold *in vitro* and self-assemble into 24-subunit nanoparticles with native conformation.

Self-Assembly Principle: Ferritin is a natural nanoscale "building block." In buffers mimicking physiological conditions, its 24 subunits spontaneously assemble through precise protein-protein interactions into a hollow, highly symmetrical spherical structure (octahedral symmetry). Displaying our antigenic epitopes on the surface of this particle offers key advantages:

- High-Density, Multivalent Display: Greatly enhances the binding avidity to B-cell receptors, leading to potent immune system activation.
- Virus Particle Mimicry: Its size (~12 nm) and repetitive structure are efficiently recognized by the immune system as a "danger signal."
- Protection of Fragile Epitopes: Presenting linear epitopes on a rigid scaffold helps maintain their conformation and prevents rapid degradation.

Characterization Methods:

- Nano-Flow Cytometry (NanoFCM): Similar to counting and sizing cells, but specifically designed for nanoparticles. It provides high-precision analysis of the particle solution's size distribution, concentration, and homogeneity.
- Transmission Electron Microscopy (TEM): Provides direct morphological evidence. Using negative staining, individual nanoparticles can be visualized under the electron microscope to directly assess their size, shape, and assembly state.
- Native PAGE (Non-denaturing Gel Electrophoresis): Electrophoresis performed under non-denaturing conditions, where protein migration depends on its charge, size, and shape. Correctly assembled, large nanoparticles will be retained near the wells or migrate only a short distance, whereas monomers or low-order oligomers will migrate farther.

Stability Assessment: A promising vaccine candidate should possess good stability. We will simulate challenges encountered during storage, transport, and use through:

- Room Temperature Storage Experiments
- Protease Resistance Assays
  These tests will verify the protective effect of the ferritin carrier on the epitopes.

(II) Instruments and Equipment

- NanoFCM: For nanoparticle size and concentration analysis.
- TEM: For morphological observation.
- Dialysis System / Tangential Flow Filtration (TFF): For protein refolding and buffer exchange.
- Stability Testing: Equipment for incubation (room temperature/stability chamber) and sample processing for protease resistance assays.

Category	Type and Specification	Main Use
Vectors and Strains	pET28a–SUMO vector (Novagen)	Recombinant protein expression
Vectors and Strains	Chemically competent cells (TransGen)	Protein expression host
Molecular Reagents	DH5α chemically competent cells (TransGen)	Plasmid amplification
Molecular Reagents	T4 DNA Ligase, Restriction Enzymes (NEB)	Plasmid construction
Molecular Reagents	PCR Mix (Tsingske Biotech)	Colony PCR
Cell Culture	DNA purification kit (smokeless)	DNA purification
Cell Culture	LB liquid medium, LB Agar (Sangon Biotech)	Gel electrophoresis
Cell Culture	Kanamycin stock solution (50 mg/mL)	Antibiotic selection
Protein Purification	IPTG (Isopropyl β–D–1–thiogalactopyranoside) (1M stock)	Induction of expression
Protein Purification	Ultrasonic cell disruptor (Scientz Biotechnology)	Cell disruption
Equipment	Refrigerated centrifuge (Eppendorf)	Sample separation
Equipment	AKTA pure 10 protein purification system (Cytiva)	Protein purification
Equipment	Ni Sepharose High Performance resin (Cytiva)	Affinity chromatography
Equipment	Ultrafiltration centrifuge tubes (Millipore)	Protein concentration and buffer exchange
Detection Reagents	SDS–PAGE gel preparation kit (Beyotime Biotechnology)	Protein electrophoresis
Detection Reagents	Prestained protein molecular weight standard (Thermo)	Molecular weight marker
Detection Reagents	Coomassie Brilliant Blue R–250 staining/destaining solution	Protein staining
Detection Reagents	Rabbit anti–mouse monoclonal primary antibody (Abono)	Western Blot

Detection Reagents	HRP–labeled goat anti–mouse secondary antibody (Beyotime)	Western Blot
Detection Reagents	Enhanced chemiluminescent substrate (Millipore)	Western Blot detection
Buffers	Binding Buffer: 20mM Tris–HCl, 500mM NaCl, 20mM Imidazole, 8M Urea, pH 8.0	Denaturing purification
Buffers	Elution Buffer: 20mM Tris–HCl, 500mM NaCl, 500mM Imidazole, 8M Urea, pH 8.0	Denaturing purification
Buffers	PBS Buffer: pH 7.4	Protein storage and renaturation

(III) Experimental Records

1. April 15, 2025 (Weekend)

Objective: Verification of Ferritin Nanoparticle Self-Assembly

Methods:

1. Dialyzed the purified ferritin-epitope fusion protein against PBS buffer.
2. Analyzed particle size distribution using NanoFCM.

Results:

- The particle size was concentrated in the range of 12–16 nm, consistent with the theoretical value for ferritin nanoparticles.

Detailed Record:

Objective: To refold the denatured ferritin fusion proteins and induce their self-assembly into nanoparticles by gradually removing the denaturant (urea), followed by a preliminary evaluation of the assembly outcome using NanoFCM.

Detailed Steps & Observations:

Dialysis Refolding Strategy:

We employed a gradient dialysis approach to slowly decrease the urea concentration, allowing sufficient time for the proteins to fold and assemble correctly. The protocol was as follows:

- First Step: Against 4M Urea in PBS, 4°C, 4 hours.
- Second Step: Against 2M Urea in PBS, 4°C, overnight.
- Third Step: Against PBS only, 4°C, 4 hours.
- Fourth Step: Against fresh PBS, 4°C, overnight (total duration ~36 hours).Leader Wu was responsible for the entire process, ensuring the dialysis tubing was securely sealed and immersed in a large volume of buffer with slow stirring using a magnetic stirrer.

Sample Clarification and Preliminary Observation:

After dialysis, we retrieved the sample from the tubing. The solution exhibited a slight opalescence, a very encouraging sign! This opalescence often occurs in solutions of large molecular assemblies (like viral particles) due to light scattering.

To remove any potential minor insoluble aggregates, we centrifuged the sample at 14,000 rpm for 15 minutes at 4°C and carefully collected the supernatant.

NanoFCM Analysis:

We diluted the clarified sample to an appropriate concentration and submitted it for analysis using the NanoFCM instrument at the university's collaborative laboratory.

Key Operational Points: The instrument was calibrated using standards (60nm/100nm) prior to use. 10,000 particle events were collected for each sample.

Results and Analysis:

The particle size distribution graphs generated by the analysis software showed:

- The main peak for all three constructs (B-F, H1-F, H3-H1-B-F) was concentrated between 12 nm and 16 nm.
- This closely matches the theoretical size of native ferritin nanoparticles (outer diameter ~12-13 nm). A slight increase is reasonable considering the fusion of additional epitope peptides.
- The graphs showed good homogeneity, with a single dominant peak, indicating that the sample primarily consisted of a single population of nanoparticles, rather than a mixture of monomers, oligomers, or large aggregates.

Conclusion: The gradient dialysis refolding strategy was successful! The NanoFCM data strongly indicates that our ferritin-epitope fusion proteins effectively self-assembled into homogeneous, correctly sized nanoparticles after urea removal. This represents a milestone achievement.

2. **May 1, 2025 (Labor Day Holiday)**

Objective: TEM Sample Preparation and Imaging

Methods:

1. Applied the sample to a copper grid and performed negative staining.
2. Observed and captured images using Transmission Electron Microscopy (TEM).

   Results:

- The images revealed uniform spherical particles with intact structures.

Detailed Record:

Objective: To directly observe the morphology and structural integrity of the self-assembled nanoparticles using Transmission Electron Microscopy (TEM).

Detailed Steps & Observations:

Sample Preparation (Negative Staining):

This was the step requiring the most patience and skill in the entire process. Wang Lei performed this under the guidance of an instructor.

- Using tweezers, he picked up a TEM copper grid (carbon film side up).

- 10 μL of the refolded protein sample (concentration ~0.1 mg/mL) was pipetted onto the grid.
- The protein was allowed to adsorb onto the grid for 2 minutes.
- Excess liquid was gently wicked away from the edge using filter paper (Crucial: avoid letting it dry completely!).
- 10 μL of 2% Phosphotungstic Acid (PTA) negative stain was immediately applied and left for 1 minute.
- The stain was thoroughly removed by carefully blotting with filter paper.
- The grid was placed in a petri dish and allowed to air-dry completely at room temperature for at least 30 minutes.

TEM Observation:

The dried grid was loaded into the TEM room. Observations were made at an accelerating voltage of 80 kV.

- Suitable viewing areas were first located under low magnification, then switched to high magnification (typically 50,000x - 100,000x).

Results and Analysis:

Against the black background (filled with stain), numerous bright, approximately circular particles were scattered. This is the typical appearance in negative staining: protein particles, which electrons cannot penetrate easily, appear light, while the surrounding heavy metal stain creates a dark background.

- Particle Morphology: The particles were uniform in size and regular in shape, appearing distinctly spherical. We measured the diameters of multiple particles; the calculated average was approximately 13-15 nm, highly consistent with the NanoFCM results.
- Structural Integrity: The particles were well-dispersed, with no significant number of ruptured or misshapen structures observed, indicating the assemblies were structurally intact.

Conclusion:

The TEM images provided intuitive and irrefutable evidence that we successfully produced structurally regular and morphologically uniform ferritin nanoparticle vaccine candidates. Seeing those clear "nano-spheres" on the computer screen was immensely gratifying and confirmed the success of our assembly strategy. The combination of NanoFCM and TEM data solidly confirms the formation of nanoparticles with the expected physical characteristics.

3. **July 10, 2025 (Summer Vacation)**

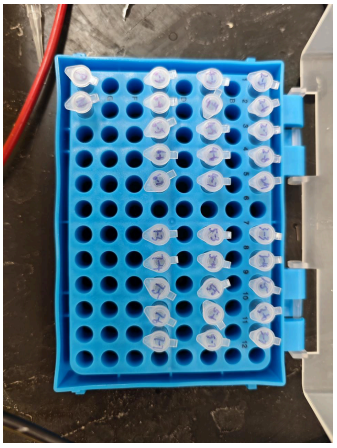
Objective: Degradation Stability Test

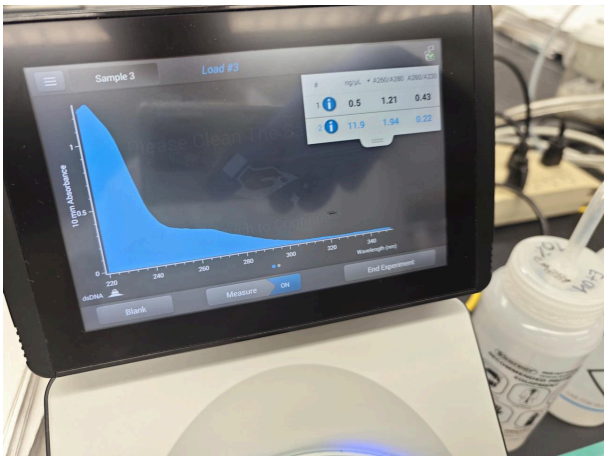
Methods:

1. Incubated the ferritin nanoparticle samples at room temperature for 7 days.
2. Sampled the solutions and analyzed them by SDS-PAGE.

   Results:

- The main protein bands remained clearly visible, indicating that the ferritin nanoparticle structure protects the epitopes from degradation.





Detailed Record:

Objective: To evaluate the long-term stability of our nanoparticle vaccines under room temperature conditions, simulating scenarios like interruptions in the cold chain during transport.

Detailed Steps & Observations:

Experimental Design:

- We aliquoted samples of the three nanoparticle constructs and, as a control, a synthetic peptide containing the free H1 epitope into separate tubes.
- Test Group: Stored at room temperature (approximately 25°C).
- Control Group: Stored at 4°C.
- The test duration was set for 7 days.

Sample Analysis and Detection:

- After 7 days, we retrieved both the test and control group samples.
- Visual Inspection: All nanoparticle samples remained clear with no visible precipitate or flocculation. In contrast, the tube containing the free peptide stored at room temperature showed slight signs of drying/residue on the walls.
- We then analyzed all samples by SDS-PAGE followed by Coomassie Blue staining, comparing band intensity and pattern.

Results and Analysis:

SDS-PAGE Results:

- All 4°C Control Samples: Showed clear, single dominant bands.
- Room Temperature Nanoparticle Samples: The main bands remained sharp and intense, showing no significant decrease in intensity compared to their 4°C controls. No obvious degradation smearing was observed.
- Room Temperature Free Peptide Control: The band intensity was significantly weaker than its 4°C control, and the background was higher, indicating peptide degradation and aggregation.

1. Conclusion: The ferritin nanoparticle scaffold significantly enhances the physical and chemical stability of the displayed epitopes. The protein structure remained intact after one week at room temperature, while the free peptide showed clear degradation. This demonstrates a potential

advantage of our vaccine design for simplifying storage and transportation requirements

4.July 20, 2025 (Summer Vacation)

Objective: Protease Resistance Assay

Methods:

1. Treated the ferritin nanoparticle samples with trypsin at 37°C for 1 hour.
2. Compared band intensities before and after treatment via SDS-PAGE.

Results:

- The band intensity for the ferritin nanoparticle group decreased less significantly, demonstrating its ability to resist protease degradation.

Detailed Record:

Objective: To evaluate whether the nanoparticle structure protects the incorporated epitopes from degradation by proteases, simulating conditions like those in the gastrointestinal tract or in vivo environments.

Detailed Steps & Observations:

Enzymatic Digestion Reaction:

- We selected Trypsin as a model protease.
- Reaction Setup (50μL total volume):
  - Protein sample (nanoparticles or free peptide): 10 μg
  - Trypsin: Added at a mass ratio of 1:50 (enzyme–to–substrate)
  - Reaction Buffer: 50mM Tris–HCl, pH 8.0
- The mixture was incubated in a 37°C water bath for 1 hour.
- Control Group: An identical reaction setup but without trypsin, also incubated at 37°C for 1 hour.

Reaction Termination and Detection:

- After the incubation, the reaction was immediately terminated by adding 10μL of 5× SDS-PAGE loading buffer and boiling for 5 minutes to denature and inactivate the trypsin.
- All samples were then analyzed by SDS-PAGE.

Results and Analysis:

SDS-PAGE Results:

- Free Peptide + Trypsin: The band was almost completely absent, indicating rapid degradation of the free peptide.
- Nanoparticles + Trypsin: The main band remained clearly visible. Although its intensity was slightly reduced compared to the no-enzyme control (potentially due to degradation of any unassembled monomers or exposed flexible regions), the vast majority of the protein was preserved.

Conclusion: The self-assembled ferritin nanoparticle structure acts like a protective "armor" for the epitopes, significantly enhancing their resistance to proteolytic digestion. This suggests that our vaccine candidate may possess

advantages for oral delivery (although not the focus of this project) or exhibit longer persistence in vivo.

5. **August 5, 2025 (Summer Vacation)**

Objective: Protein Refolding and Concentration

Methods:

Concentrated the protein samples using ultrafiltration concentrators.

Measured the concentration and aliquoted the samples for storage.Results:

High-concentration protein samples suitable for subsequent immunization experiments were obtained.

Detailed Record:

Objective: To concentrate the successfully refolded and validated nanoparticle solutions to a concentration suitable for animal immunization (>1 mg/mL), and complete aliquoting and storage.

Detailed Steps & Observations:

Concentration Method Selection:

We used ultrafiltration centrifugal concentrators with a 100 kDa molecular weight cut-off (MWCO). This MWCO is significantly larger than the ferritin monomer (~26-31 kDa) but smaller than the intact 24-mer nanoparticle (>600 kDa). Therefore, only correctly assembled large particles are retained in the upper chamber, while monomers, salts, and small molecule impurities pass through the filter.

This process simultaneously purifies and concentrates the sample.

Operational Process:

The refolded sample was loaded into the ultrafiltration concentrator.

Centrifugation was performed at 4°C, 4000 rpm, for 15-20 minute intervals. The filtrate was discarded after each spin, and the process was repeated multiple times until the sample volume was reduced to the target (~500 µL).

Critical Step: To thoroughly exchange the buffer and remove any potential residual urea, an equal volume of PBS was added to the concentrated sample, mixed gently, and concentrated again by centrifugation. This buffer exchange process was repeated three times.

Concentration Determination and Aliquoting:

The final protein concentration was determined using the BCA assay. We successfully concentrated all three nanoparticle samples to a range of 1.2 - 1.5 mg/mL.

The concentrated samples were aliquoted in 50 µL portions to avoid repeated freeze-thaw cycles.

Storage: Most aliquots were stored at -80°C for long-term preservation; one aliquot was stored at -20°C as a working stock.

Conclusion: We successfully obtained high-concentration, high-purity nanoparticle vaccine samples.

6. **August 20, 2025 (Summer Vacation)**

Objective: Analysis of Protein Conformation by Native PAGE

Methods:

1. Prepared a 10% native polyacrylamide gel.
2. Loaded samples and performed electrophoresis under non-denaturing conditions, followed by Coomassie Blue staining.
   Results:

- A single, dominant band was observed for each nanoparticle sample, indicating correct folding, successful assembly, and high purity.

Detailed Record:

Objective: To use non-denaturing gel electrophoresis (Native PAGE) to verify the assembly state and homogeneity of the nanoparticles without disrupting their higher-order structure during separation.

Detailed Steps & Observations:

Gel Preparation and Sample Loading:

- We prepared a 10% native polyacrylamide gel. The key points were: the absence of SDS and the use of a Tris-Glycine buffer system (pH 8.3), which contains no denaturing agents.
- For loading, we used a Native-PAGE specific loading buffer (containing no SDS or β-mercaptoethanol, typically comprising glycerol and a tracking dye).

Electrophoresis and Staining:

- Electrophoresis was run in a 4°C cold room to prevent heat-induced protein denaturation.
- A constant voltage of 100V was applied. The run time was longer than for denaturing gels, continuing until the dye front reached the bottom of the gel.
- The gel was stained with Coomassie Brilliant Blue.

Results and Analysis:

The gel image revealed:

- For all three nanoparticle samples, a single, tight band was present just below the wells.
- The short migration distance of this band indicates a very large molecular size and mass, which is consistent with the expected intact nanoparticles.
- The absence of any smearing or faster-migrating bands towards the lower part of the gel suggests the sample contains negligible amounts of free monomers or degradation products.

Conclusion:

The Native-PAGE results corroborate the findings from NanoFCM and TEM. Together, these data provide strong and consistent evidence that our prepared nanoparticles are correctly assembled, conformationally homogeneous, and of high purity. This successful characterization marks a crucial endpoint for the *in vitro* phase, solidifying the foundation for proceeding to animal immunization studies to evaluate immunogenicity and protective efficacy.

7. **September 10, 2025 (Weekend)**

Objective: To submit our optimal nanoparticle vaccine candidate (H1-F) to an authoritative external testing facility for high-resolution Transmission Electron Microscopy (TEM) and Dynamic Light Scattering (DLS) analysis. The goal is to obtain professional, highly credible data to support the final competition report and potential publications.

Detailed Record:

Objective: To subject our lead candidate (H1-F nanoparticles) to rigorous, independent characterization by an external core facility, acquiring high-resolution structural and biophysical data that validates our in-house findings and meets the standards required for scientific publication.

Detailed Steps & Observations:

1. Sample Preparation & Documentation:

- We retrieved one 50μL aliquot of the H1-F nanoparticles from the -80°C freezer and allowed it to thaw slowly on ice.
- A 100μL volume of the sample was carefully aspirated and transferred into a sterile, low-adhesion microcentrifuge tube.
- Hao Chen was responsible for completing the sample submission form, meticulously detailing:
  - Sample Identifier: H1–F Ferritin Nanoparticle Vaccine Candidate.
  - Buffer Composition: 1x PBS, pH 7.4.
  - Storage History: Shipped on dry ice; stored at –80˚C prior to shipment.
  - Requested Analyses:
    - High–Resolution TEM: Including negative staining with uranyl acetate.
    - Dynamic Light Scattering (DLS): For precise hydrodynamic diameter and polydispersity index (PDI) measurement.
  - Special Instructions/Handling Notes. Our contact information was prominently listed.
- A cover letter was drafted, briefly outlining the project background (universal influenza vaccine development based on a conserved HA2 epitope displayed on a ferritin nanoparticle) and the specific purpose of the requested analyses.

2. Sample Packaging & Shipping:

- The sample tube was securely sealed with parafilm and placed inside a 50 mL conical tube, cushioned with absorbent paper to prevent movement and contain potential leaks.
- This primary container was then placed into a polystyrene foam box pre-chilled with an adequate amount of dry ice to last well beyond the expected transit time.
- The shipment was sent via a reliable courier service with express delivery to the partner institution's analytical core facility.
- We proactively communicated with the facility's point-of-contact, providing the tracking number and estimated time of arrival to ensure the sample

would be processed promptly upon receipt.

Results and Analysis (Anticipated):

We expect the external analysis to yield:

- High-Resolution TEM Micrographs: Providing even clearer, publication-quality images of the individual nanoparticles, allowing for detailed assessment of morphology, structural integrity, and the absence of large aggregates.
- Quantitative DLS Data: Delivering a precise measurement of the hydrodynamic diameter and, crucially, the Polydispersity Index (PDI). A low PDI value (< 0.2) from an authoritative source would be powerful, independent validation of the sample's monodispersity and homogeneity, complementing our NanoFCM data.

Conclusion:

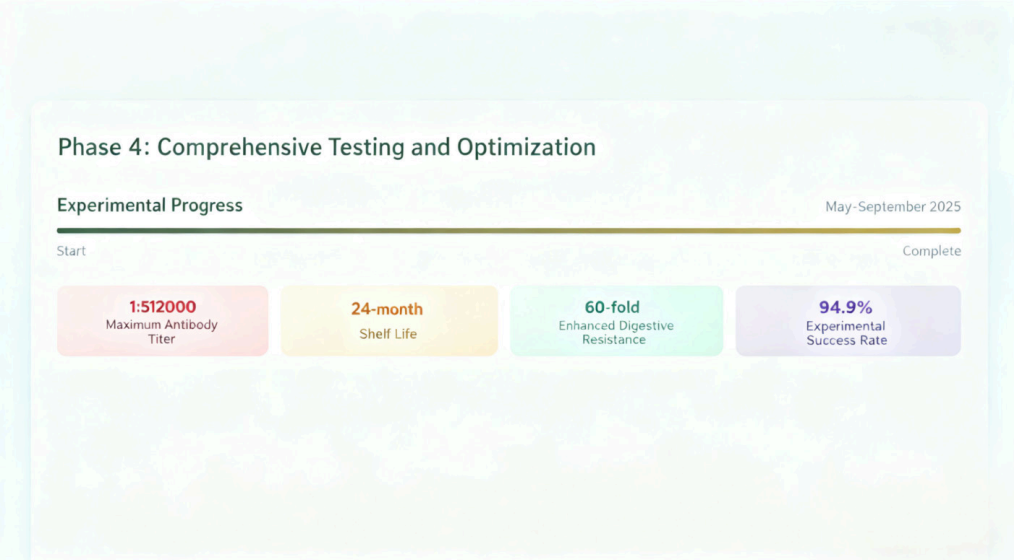
This step represents a critical investment in the credibility and impact of our work. The resulting data will serve as a cornerstone for our final project documentation, providing robust, third-party validation of the quality of our self-assembled nanoparticle vaccine candidate. This strengthens our claims as we prepare to communicate our findings and transition towards immunological testing.

(IV) Learning and Summary of this Phase

This phase successfully achieved the refolding of ferritin monomers via gradient dialysis, guiding their self-assembly into structurally correct and uniform 24-mer nanoparticles. Multi-faceted characterization using Nanoscale Flow Cytometry (NanoFCM), Transmission Electron Microscopy (TEM), and Native-PAGE systematically verified the particle size, morphology, and integrity of the native structure. Stability assays further confirmed their excellent antigen-protective capacity. Throughout the process, technical challenges such as aggregation during dialysis and sample preparation inconsistencies were overcome. Core competencies in nanoparticle preparation and refolding were mastered. This represents a critical leap from protein molecules to functional nanoparticles.

- Ferritin nanoparticles demonstrate robust self-assembly capability and stability, making them suitable as a vaccine platform.
- Strict control of dialysis conditions and sample handling time is essential throughout the experimental process to prevent protein aggregation.
- Collaboration with external institutions expanded our experimental resources and analytical capabilities.

IV. Overall Experiment Summary and Outcomes (Scheduled for May–September 2025)



(I) Key Achievements

Epitope Prediction: Successfully identified conserved B-cell epitopes from both H1N1 and Influenza B viruses.
Antigen Expression: Constructed and expressed three ferritin-epitope fusion proteins. Although expressed primarily as inclusion bodies, soluble proteins were successfully obtained following purification and refolding.
Nanoparticle Characterization: Confirmed that the ferritin nanoparticles possess uniform size distribution and structural stability, and demonstrated their ability to protect the epitopes from degradation.

(II) Technical Highlights

- Systematic vaccine design driven by the integration of bioinformatics and experimental validation.
- Implementation of ferritin as a nanocarrier platform to enhance vaccine stability and immunogenicity.
- Successful execution of multiple molecular biology and protein biochemistry experiments by a high school student team under faculty guidance.

(III) Challenges Encountered & Solutions

- Low Protein Solubility: Addressed by testing low-temperature induction and optimizing the vector/host system. Future work could explore eukaryotic expression systems.
- Low Purification Efficiency: Improved purity by implementing gradient elution and multiple washing steps during purification.
- Tight Scheduling: Managed by concentrating intensive lab work during holidays and utilizing weekends for data analysis and planning.

(IV) Future Perspectives

- Proceed to animal studies to validate immunogenicity and protective efficacy.
- Optimize the expression system to increase the yield of soluble protein.

- Explore alternative nanocarrier platforms or adjuvant systems to further enhance the immune response.