Wet-Lab Validation Workflow for De Novo Antibody Design

The training data used to develop Clickmab’s de novo antibody design model were generated from Biocytogen’s fully human shared-light-chain RenLite® mouse platform. All computationally designed antibodies are fully human and utilize a common light chain. To facilitate experimental validation, the final computational report provides the designed antibody sequences in three components： heavy chain HCDR3, VH region upstream of HCDR3 (FR1–FR3), and shared light chain. Subsequent gene synthesis, display screening, and recombinant expression are used to obtain designed antibodies that bind the specified epitope.

1. Gene Synthesis of Linearized scFv Fragments

After completing computational design, download the generated HCDR3 sequences and their corresponding heavy-chain FR1–FR3 regions from the final computational report. The report also includes the shared light chain sequence. These three parts are assembled into an scFv fragment, as shown in Figure 1.

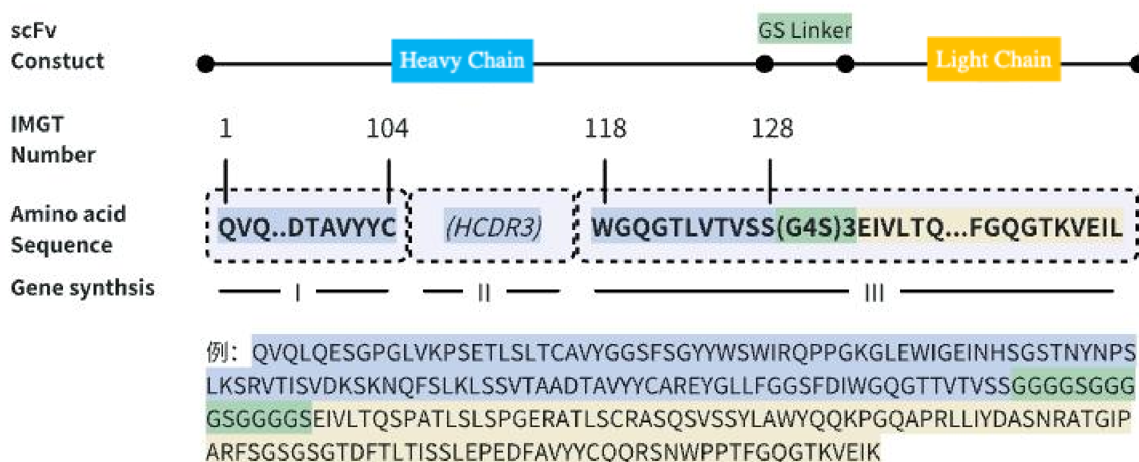

Figure 1. Linearized scFv Assembly Diagram

The scFv fragment can be synthesized in three parts corresponding to the numbered segments in Figure 1:

- **I: Heavy-Chain FR1–FR3 Region (including CDR1 and CDR2)**
- **II: Heavy chain HCDR3**
- **III: Heavy chain FR4–GS Linker–VK**

An example amino acid sequence of an assembled linearized scFv is shown below:

QVQLQESGPGLVKPSETLSLTCAVYGGSFSGYYWSWIRQPPGKGLEWIGEINHSGSTNYNPSLKSRVTISVDKS
KNQFSLKLSSVTAADTAVYYCAREYGLLFGGSFDIWGQGTTVTVSSGGGGSGGGGSGGGGSEIVLTQSPATLS
LSPGERATLSCRASQSVSSYLAWYQQKPGQAPRLLIYDASNRATGIPARFSGSGSGTDFTLTISSLEPEDFAVYY
CQQRSNWPPTFGQGTKVEIK

Restriction sites and conserved bases must be added to both ends during synthesis for downstream vector construction. Additionally, approximately 20 bp homology arms should be added between different fragments for overlap extension PCR (OE-PCR) assembly. The detailed process is as follows:

1. **Codon-optimize the three parts** for the expression host *E. coli* (Note: phage display is used as the example here; other display systems such as yeast display are also compatible — optimize for *Yeast* in that case. Optimize each part separately; optimizing the full assembled sequence for large-scale HCDR3 libraries is very time-consuming).

2. **Synthesize Part I** (heavy-chain FR1–FR3 region). Computation typically recommends 10–20 high-matching germline genes for conventional gene synthesis. Note that a restriction site must be added to the 5′ end for downstream vector construction (e.g., SfiI; for phagemid vector pCANTAB5e, SfiI and NotI restriction sites are used).

3. **Synthesize Part II** (heavy chain HCDR3). Computation typically generates and filters candidates at three scales: 10,000, 200,000, and 1,000,000. Select the appropriate scale based on your needs — larger library sizes generally improve the probability of identifying functional binders, but also increase synthesis costs. Given the high diversity and relatively short length (<100 nt) of heavy chain HCDR3, pooled oligonucleotide synthesis is recommended. Homology arms for assembly are added to both ends: the 5′ arm is the last 7 amino acids of heavy chain framework FR3, i.e., **DTAVYYC** (this terminal sequence is identical across all Part I variants); the 3′ arm corresponds to the first seven amino acids of FR4 derived from the most frequently used human antibody heavy-chain J gene, i.e., **WGQGTLV**. After synthesis, amplify the single-stranded oligopool to double-stranded DNA using the 5′ and 3′ homology arm primers (primers F1 and R1 in Figure 2). Note that oligopool synthesis typically carries a random error rate of approximately 1/2000. Excessive amplification cycles introduce bias and increased errors, so stop amplification as soon as sufficient material for the downstream overlap extension step is obtained.

4. **Synthesize Part III** (heavy chain FR4–GS Linker–VK). This part consists of the heavy-chain FR4 region derived from the most frequently used human antibody J gene, a GS linker, and the shared light chain — all fixed sequences — synthesized by conventional gene synthesis. Note that the 5′ end must include the homology arm **WGQGTLV** for assembly, and the 3′ end must include a restriction site for vector construction (e.g., NotI for pCANTAB5e).

5. **Assemble the linearized scFv by overlap extension PCR**, joining Parts I, II, and III through their shared homology arms. Use the 5′ primer containing the SfiI site (from Part I) and the 3′ primer containing the NotI site (from Part III) — primers F and R in Figure 2. As with the oligopool amplification, excessive cycling causes bias and mutations; amplify only to the amount required for downstream restriction digestion and library construction — typically 15–30 μg is sufficient.

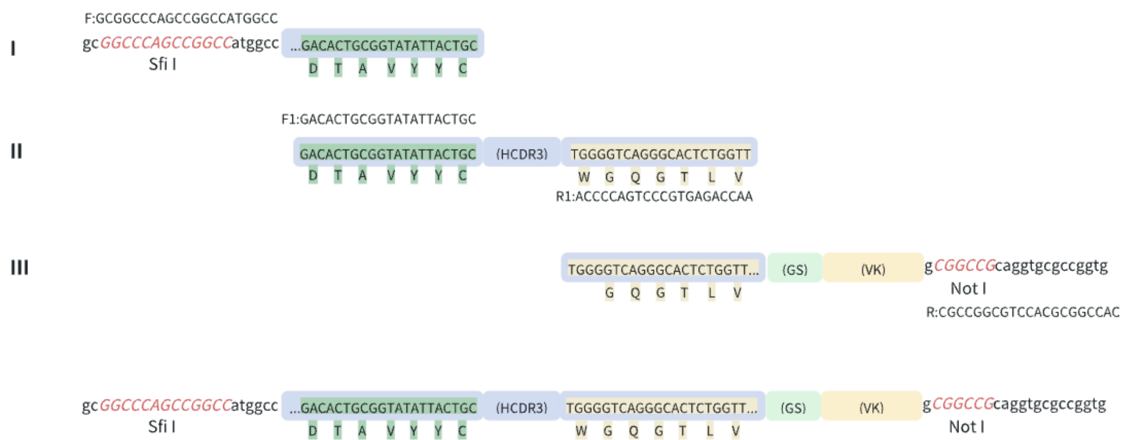

Figure 2. Overlap Extension PCR (OE-PCR) for Linearized scFv Fragment Assembly

6. **Quality control of the linearized scFv library**. To ensure screening quality, assess the proportion of correctly assembled scFv constructs in the linearized library. Frameshift mutations, stop codons, and point mutations can arise during oligopool synthesis and assembly; under normal conditions, a correct assembly rate of ≥60% is required to ensure downstream screening quality (a small proportion of mis-assembled constructs has relatively little impact on display screening). For Sanger sequencing QC: ligate the linearized fragment into a vector by restriction digestion or blunt-end ligation, plate the transformants, pick 96 single colonies for Sanger sequencing, and calculate the correct assembly proportion — a simple, fast, and cost-effective approach. Alternatively, NGS-based assessment (amplicon library preparation, sequencing, and analysis of the linearized library) provides a comprehensive evaluation but is more time-consuming and expensive.

2. Phage Library Construction and Screening

After gene synthesis of the linearized scFv library, construct a phage display library and perform phage display screening using standard protocols, briefly described below:

1. **Double-digest the phagemid vector and linearized scFv fragment** with SfiI/NotI. Gel-purify the relevant bands, then ligate the vector and linearized fragment.

2. **Electroporate to construct the phage library**. Ensure that the electroporated library size covers 10–100× the theoretical diversity — for example, if the theoretical library diversity is 10^6, the phage library size should reach 10^7–10^8.

3. **Biopanning**. Design the panning strategy based on available materials. Running multiple parallel panning strategies in combination is recommended (e.g., solid-phase, solution-phase, or alternating formats). If cell lines are available, cell-based panning can also be performed simultaneously.

4. **Set up a positive control library**. Include a positive control during panning by constructing phage display particles from a known binder or benchmark antibody, then spiking them into the screening library at a defined ratio. If the positive-control antibody has very high affinity, it can be spiked into

the screening library at a lower frequency (e.g., 1:10,000). Antibodies with lower affinity may require a higher spike-in frequency.

5. **Screening rounds and detection**. After ≥3 rounds of panning, check the titer and perform phage ELISA to assess enrichment. If enrichment is not apparent, additional rounds may be performed. If clear enrichment is observed, proceed with single-clone screening and/or NGS sequencing to identify positive clones.
   - **5.1 Single-clone screening**: Perform single-clone phage display and screening on enriched libraries. Sanger-sequence phage-ELISA-positive clones to obtain the sequences of identified binders. Note that some sequences that are phage-ELISA-positive may fail to express or lose binding activity upon eukaryotic expression — if the experimental workflow is correct, the false-positive rate should be <10%.
   - **5.2 NGS sequencing**: The enriched library may also be analyzed by next-generation sequencing (NGS). Since the screening library uses a shared light chain antibody format, only the heavy chain variable region needs to be amplified and sequenced. After sequencing, analyze highly enriched heavy chain sequences — these are likely to have binding activity, but must be confirmed by eukaryotic recombinant expression.

3. Recombinant Expression of Designed Antibodies

Antibody clones identified through phage display screening require further recombinant expression and validation to confirm binding activity.

1. **Gene synthesis**: Synthesize the variable region sequences of identified binding antibodies and construct expression vectors — typically in full-length IgG format (e.g., human IgG1 with a kappa light chain), though other formats (scFv, Fab, etc.) and Fc variants can be used as needed. Common expression vectors include pcDNA3.4 and pTT5.
2. **Recombinant expression**: Express in CHO or HEK293 eukaryotic expression systems. For initial screening and characterization, 4 mL CHO expression is sufficient (under normal conditions, ≥200 ng purified antibody can be obtained from 4 mL CHO expression with SDS/SEC purity ≥90%; some antibodies may have lower expression levels). If expression is low but ≥50 ng purified antibody is obtained, this is sufficient for BLI or ELISA to confirm binding activity. Note that expression level and physicochemical properties (e.g., purity) can serve as a rough indicator of antibody developability.
3. **Binding activity assay**: After obtaining purified antibody, select the appropriate binding assay based on the target antigen. For soluble antigens: ELISA, BLI, or SPR. For multi-pass transmembrane targets: FACS using an overexpression cell line or tumor cell line to assess binding activity.

4. **Epitope verification**: If a benchmark antibody or ligand that binds the designed epitope is available, competitive ELISA or epitope binning can be used to rapidly determine whether the identified antibodies bind to the specified epitope. For higher-resolution confirmation, structural biology methods such as Cryo-EM or X-ray crystallography can be used to verify that the designed antibodies bind to the intended epitope.