

## Video Article

# Multi-target Parallel Processing Approach for Gene-to-structure Determination of the Influenza Polymerase PB2 Subunit

Brianna L. Armour<sup>\*1</sup>, Steve R. Barnes<sup>\*2</sup>, Spencer O. Moen<sup>2</sup>, Eric Smith<sup>3</sup>, Amy C. Raymond<sup>4</sup>, James W. Fairman<sup>5</sup>, Lance J. Stewart<sup>6</sup>, Bart L. Staker<sup>7</sup>, Darren W. Begley<sup>5</sup>, Thomas E. Edwards<sup>8</sup>, Donald D. Lorimer<sup>9</sup>

<sup>1</sup>Protein Crystallization Lab, Emerald Bio

<sup>2</sup>Molecular Biology Lab, Emerald Bio

<sup>3</sup>Scientific Sales Representative, Emerald Bio

<sup>4</sup>Group Leader II, Emerald Bio

<sup>5</sup>Group Leader I, Emerald Bio

<sup>6</sup>Chair of Advisory Board, Emerald Bio

<sup>7</sup>Director of Multi-Target Services, Emerald Bio

<sup>8</sup>Senior Project Leader, Emerald Bio

<sup>9</sup>Project Leader II & SSGCID Site Manager, Emerald Bio

\*These authors contributed equally

Correspondence to: Bart L. Staker at [bstaker@embios.com](mailto:bstaker@embios.com)

URL: <https://www.jove.com/video/4225>

DOI: [doi:10.3791/4225](https://doi.org/10.3791/4225)

**Keywords:** Infection, Issue 76, Structural Biology, Virology, Genetics, Medicine, Biomedical Engineering, Molecular Biology, Infectious Diseases, Microbiology, Genomics, high throughput, multi-targeting, structural genomics, protein crystallization, purification, protein production, X-ray crystallography, Gene Composer, Protein Maker, expression, *E. coli*, fermentation, influenza, virus, vector, plasmid, cell, cell culture, PCR, sequencing

Date Published: 6/28/2013

Citation: Armour, B.L., Barnes, S.R., Moen, S.O., Smith, E., Raymond, A.C., Fairman, J.W., Stewart, L.J., Staker, B.L., Begley, D.W., Edwards, T.E., Lorimer, D.D. Multi-target Parallel Processing Approach for Gene-to-structure Determination of the Influenza Polymerase PB2 Subunit. *J. Vis. Exp.* (76), e4225, doi:10.3791/4225 (2013).

## Abstract

Pandemic outbreaks of highly virulent influenza strains can cause widespread morbidity and mortality in human populations worldwide. In the United States alone, an average of 41,400 deaths and 1.86 million hospitalizations are caused by influenza virus infection each year<sup>1</sup>. Point mutations in the polymerase basic protein 2 subunit (PB2) have been linked to the adaptation of the viral infection in humans<sup>2</sup>. Findings from such studies have revealed the biological significance of PB2 as a virulence factor, thus highlighting its potential as an antiviral drug target.

The structural genomics program put forth by the National Institute of Allergy and Infectious Disease (NIAID) provides funding to Emerald Bio and three other Pacific Northwest institutions that together make up the Seattle Structural Genomics Center for Infectious Disease (SSGCID). The SSGCID is dedicated to providing the scientific community with three-dimensional protein structures of NIAID category A-C pathogens. Making such structural information available to the scientific community serves to accelerate structure-based drug design.

Structure-based drug design plays an important role in drug development. Pursuing multiple targets in parallel greatly increases the chance of success for new lead discovery by targeting a pathway or an entire protein family. Emerald Bio has developed a high-throughput, multi-target parallel processing pipeline (MTPP) for gene-to-structure determination to support the consortium. Here we describe the protocols used to determine the structure of the PB2 subunit from four different influenza A strains.

## Video Link

The video component of this article can be found at <https://www.jove.com/video/4225/>

## Protocol

An overview of the protocol is presented in **Figure 1**.

## Molecular Biology

### 1. Construct Design

Use Gene Composer software to design protein construct and codon engineered synthetic gene sequences. The use of Gene Composer software has been offered in detail elsewhere<sup>3</sup>.

1. Use the Alignment Viewer Module and Construct Design Module to compare protein sequence alignments and define protein construct. Align target amino acid sequence to both the primary and 3D structural elements from homologs in the Protein Data Bank (PDB), if available (Figure 2).
2. Use alignment information to make structure-guided construct designs by choosing new termini based on conservation of the primary structure and 3D structures of homologs.
3. Design insert PCR (iPCR) and vector PCR (vPCR) amplimers (terminal primers).
4. Using Gene Composer's protein-to-DNA algorithm, back-translate the construct amino acid sequence into codon engineered nucleic acid sequence.
5. Use the proper codon usage table (CUT) to optimize sequence for expression in *E. coli*.
6. Virtually clone insert into pET28 vector modified to incorporate a N-terminal 6x Histidine tag and Smt3/SUMO fusion protein that allows for easy purification.
7. Place synthetic gene order with DNA 2.0 and order primers from Integrated DNA Technologies.

### 2. Polymerase Incomplete Primer Extension (PIPE) Cloning

1. Prepare Primers and Genes
  1. Centrifuge the vendor-supplied plates containing primers at 1,000 rpm for 1 min.
  2. Bring primer concentration to 100  $\mu$ M and add 50  $\mu$ l TE buffer.
  3. Dilute primers to 10  $\mu$ M with deionized (DI) water in a 96-well v-bottom plate.
  4. Centrifuge the vendor-supplied gene in a 1.5 ml tube at 1,300 rpm for 1 min.
  5. Using TE buffer, bring the DNA concentration of each tube to 50 ng/ $\mu$ l.
  6. In 1.5 ml tubes, make dilutions of each primer to 10 ng/ $\mu$ l.
  7. Store primers and genes at -20 °C when not in use.
2. Prepare Insert PCR (iPCR)
  1. Thaw a vial of Pfu Master Mix on ice; keep genes and primers at room temperature.
  2. Create a plate map assigning wells to a set of primers and construct.
  3. Add 13  $\mu$ l of DI water into each well of a 96-well PCR plate.
  4. Add 5  $\mu$ l of forward and 5  $\mu$ l of reverse primer to each reaction in the 96-well plate according to the plate map, ensuring to change tips between each well.
  5. Add 2  $\mu$ l of each full length gene to its appropriate well according to the plate map.
  6. Add 25  $\mu$ l of Pfu master mix to each well, ensuring to change tips between each well.
  7. Cycle the reactions using the following PCR conditions:
    - a. 95 °C 2 min
    - b. 95 °C 30 sec
    - c. 50 °C 45 sec
    - d. 68 °C 3 min
    - e. 4 °C  $\infty$
  8. Repeat steps b-d for 25 cycles.
  9. Transfer 10  $\mu$ l of each iPCR reaction to a new 96-well PCR plate.
  10. Add 3  $\mu$ l of 6X load dye to each sample.
  11. Separate each sample on a 1% TAE EtBr agarose gel at 110 V next to a 100-500 bps DNA ladder to confirm fragment amplification.
  12. Store iPCR product at -20 °C when not in use (avoid freeze thaw as much as possible).

### 3. Prepare Vector PCR (vPCR)

1. Start overnight culture of transformed *E. coli* with pET28 vector plasmid.
  1. Inoculate two 5 ml tubes of 2-YT broth with 50  $\mu$ g/ml kanamycin.
  2. Grow cultures overnight at 37 °C in shaker at 220 rpm.
2. Spin down cultures after overnight growth by centrifuging at 3,000 rpm for 15 min.
3. Use a Qiagen QIAprep Spin Miniprep Kit to extract pET28 vector from bacterial pellets according to the manufacturer's instructions.
4. Setup restriction enzyme digestions of extracted pET28 plasmid.
  1. Add 2.2  $\mu$ l of 10X BamHI buffer and 1  $\mu$ l of BamHI and HindIII to 20  $\mu$ l of pET28 vector.
  2. Incubate reaction for 1 hr at 37 °C.
5. Separate digestion product on a gel.
  1. Refer to step 2.2.10.
  2. Cut vector band from gel and purify it using the QIAquick Gel Extraction Kit according to the manufacturer's instructions.

6. Using a NanoDrop, quantify DNA concentration.
7. Dilute cut vector to 10 ng/μl. Store at -20 °C when not in use.
8. Prepare vPCR primers.
  1. Centrifuge IDT supplied oligonucleotides for 1 min at 1,300 rpm.
  2. Bring concentration to 100 μM with DI water.
  3. Prepare 10 μM dilution of both forward and reverse primers in a 1.5 ml tube.
  4. Store primers and primer dilutions at -20 °C.
9. Thaw Pfu Master Mix on ice and thaw template and primers at room temperature.
10. Setup vPCR reactions in a 96-well PCR plate:
  1. In the first row of a 96-well plate combine 60 μl of both forward and reverse vPCR primers and 24 μl of digested pET28 template (10 ng/μl).
  2. Using a 12-tip multichannel pipette, transfer 12 μl of the primer and template master mix to each remaining well of the plate. This should result in 12 μl of primer and template master mix in each well of the plate.
  3. Add 13 μl of DI water to each well.
  4. Add 25 μl of Pfu Master Mix to each well.
11. Cycle the reactions through the PCR conditions used in step 2.2.7.
12. Pool all of the vPCR reactions into a 15 ml Falcon tube.
13. Verify fragment amplification by separating 10 μl of pooled PCR product on a gel (expected length of digested pET28 vector is approximately 6 kb).
  1. Refer to step 2.2.10.
14. Prepare merge plates.
  1. Aliquot 3 μl of vPCR product into each well of a 96-well v-bottom plate.
  2. Store plates at -20 °C until merge with iPCR product.

#### 4. Merge iPCR and vPCR Products

1. Thaw iPCR products and pre-aliquoted vPCR 96-well merge plate at room temperature.
2. Add 3 μl of each iPCR product to its respective well of the merge plate.
3. Transform merge plate into Top Ten chemically competent cells.
4. Add 2 μl of each merge reaction into a single 50 μl tube of vendor-supplied chemically competent cells and proceed with manufacturer's supplied protocol.
5. Prepare overnight cultures for each construct from transformation plate.
6. Aliquot 5 ml TB broth (with 50 μg/ml kanamycin) from a 25 ml sterile reservoir into each well of a deep well block.
7. Using sterile technique, pick an isolated colony from each transformation plate and inoculate the appropriate well of the deep well block.
8. Cover the block with an Airpore cover.
9. Shake block at 220 rpm at 37 °C overnight.
10. Pellet cells by centrifuging the block for 30 min at 4,000 rpm.
11. Pour off the supernatant and pat the top of the block dry with a paper towel.
12. Mini-prep using a Qiagen 96-well vacuum apparatus according to the manufacturer's instructions.

#### 5. Preparing Glycerol Stocks of Successfully Cloned Constructs

1. Transform successfully cloned sequence validated DNA into BL21(DE3) chemically competent cells according to the manufacturer's instructions.
2. For each construct, pick a single isolated colony from the BL21(DE3) transformation and inoculate into 1 ml of 2-YT broth (with 50 μg/ml kanamycin).
3. Shake cultures at 220 rpm for 3-4 hr at 37 °C.
4. Label a 1.5 ml screw cap tube with the unique construct identification number, cell strain, and date. Add 500 μl of 50% glycerol and 500 μl of cell culture and invert several times. Immediately store glycerol stock on dry ice or in a -80 °C freezer.

#### 6. Expression Testing

Lysis Buffer	Wash Buffer	Elution Buffer
50 mM NaH <sub>2</sub> PO <sub>4</sub> , pH 8.0 300 mM NaCl 10 mM Imidazole 1% Tween 20 2 mM MgCl <sub>2</sub> 0.1 μl/ml Benzonase 1 mg/ml Lysozyme	50 mM NaH <sub>2</sub> PO <sub>4</sub> , pH 8.0 300 mM NaCl 20 mM Imidazole 0.05% Tween 20	25 mM Tris, pH 8.0 300 mM NaCl 250 mM Imidazole 0.05% Tween 20

\* Add Benzonase, lysozyme, and protease inhibitor immediately before lysis.

1. Streak a sample from glycerol stock onto kanamycin selective agar and incubate overnight at 37 °C.

2. Start a non-inducing pre-culture in a 96-well round bottom block; inoculate 1.2 ml TB broth (with 50 mg/ml kanamycin) supplemented with 0.5% glucose with a freshly grown *E. coli* isolate. Grow overnight shaking at 220 rpm at 37 °C.
3. After overnight growth, start induction cultures by inoculating 1.2 ml of TB broth (with 50 mg/ml kanamycin) supplemented with Novagen Overnight Express System 1 (according to the manufacturer's protocol) with 40 µl of the pre-culture.
4. Grow the small-scale induction cultures at 20 °C for 48 hr, shaking at 220 rpm.
5. Harvest cells by centrifuging at 4,000 rpm for 15 min, pour off supernatant and store at -20 °C for at least 1 hr prior to processing.
6. In the 96-well block, resuspend the cell pellets in 300 µl lysis buffer.
7. Incubate cells in lysis buffer at room temperature for 30 min followed by mechanical lysis by vigorously shaking for 30 min at room temperature.
8. Clarify the crude lysate by centrifugation for 30 min at 4,000 rpm at 4 °C.
9. Use a multi-channel pipette to transfer 200 µl of the clarified lysate (soluble fraction) to a 96-well flat bottom tray (Qiagen). For each well containing a sample, add 40 µl Ni-NTA magnetic beads (Qiagen).
10. Gently agitate the plate on a rocker for 1 hr at 16 °C.
11. Place the plate on a magnetic post plate (Qiagen) and remove the unbound soluble fraction. Take care to not pipette out any of the Ni-NTA beads.
12. Remove the plate from the post plate and gently resuspended the beads in 200 µl wash buffer. Pipette up and down for 30 sec and then place the plate back on the post plate.
13. Remove the wash buffer and repeat step 6.12.
14. Remove plate from post plate and elute the Ni-NTA bound target protein by washing with 50 µl elution buffer for 5 min.
15. Return flat bottom plate to magnetic post plate and transfer the elution to a fresh 96-well v-bottom plate.
16. Transfer 20 µl of the elution to a fresh 96-well v-bottom plate and react with 1 µl ULP1 protease.
17. According to the manufacturer's protocol, analyze the eluted and eluted+Ulp1 fraction by capillary electrophoresis using a LabChip 90.
18. Alternatively, all fractions from the expression testing can be analyzed via SDS-PAGE.

## 7. Large Scale Fermentation

1. Use a sterile pipette tip to obtain a scrape from a glycerol stock, inoculate 100 ml TB broth (with 50 mg/ml kanamycin) and grow overnight. Shake at 220 rpm and 37 °C.
2. After overnight growth, expand pre-culture by inoculating 1 L of TB broth with EMD autoinduction solutions (see manufacturer's protocol) (with 50 mg/ml kanamycin) in a 2 L baffled flask with 10 ml of the pre-culture (1:100 dilution).
3. Shake the expanded 1 L cultures at 37 °C; change the temperature of the shaking incubator to 20 °C when an optical density of 0.6 (OD<sub>600</sub>) is reached.
4. After overnight growth, take a representative 10 ml aliquot from each construct for expression testing.
5. Harvest cell paste by centrifugation at 5,000 rpm for 15 min and discard supernatant.
6. Freeze cell paste at -80 °C.

### PROTEIN PURIFICATION

#### Buffers:

Lysis Buffer	Buffer A (Equilibration)	Buffer B (Elution)	Sizing Column Buffer
25 mM Tris, pH 8.0 200 mM NaCl 0.5 % Glycerol 0.02 % CHAPS 10 mM Imidazole 1 mM TCEP 50 mM Arginine 5 µl Benzonase 100 mg Lysozyme 3 Protease Inhibitor Tablets (EDTA-free)	25 mM Tris, pH 8.0 200 mM NaCl 10 mM Imidazole 1 mM TCEP 50 mM Arginine 0.25% Glycerol	25 mM Tris, pH 8.0 200 mM NaCl 200 mM Imidazole 1 mM TCEP	25 mM Tris, pH 8.0 200 mM NaCl 1% Glycerol 1 mM TCEP

\* Add Benzonase, lysozyme, and protease inhibitor tablets to each 150 ml sample immediately before lysis.

## 8. Cell Lysis

1. Make 2 L of lysis buffer; do not add lysozyme, protease inhibitor tablets or benzonase (each sample will be lysed separately in 150 ml of lysis buffer).
2. Thaw and resuspend cell paste in lysis buffer at a 1:5 mass:volume ratio by vigorously stirring for 30 min at 4 °C. Break chunks loose from sides of beaker using a clean spatula. During this time period prepare Ni and Dialysis Buffers
3. On ice, lyse the cells using a Misonix sonicator (70% power, 2 sec on/1 sec off pulses for 3 min) and gently swirl container to prevent overheating. Save a small (200 µl) aliquot of crude lysate for future analysis.
4. Clarify the crude lysate by centrifugation at 18,000 x g for 35 min at 4 °C, collect the supernatant and save a small (200 µl) aliquot for future analysis. Store pellet at 4 °C until it is confirmed the protein has been lysed into the soluble fraction.

## 9. Pre-run Protein Maker Setup

1. With the protein maker turned on and the software open, initialize the instrument.
2. Once initialized, attach one 5.0 ml GE Healthcare HisTrap FF Nickel-chelate column (Ni column) on a separate line of the gantry for each of the samples.
3. Run 3-4 column volumes (CV) of equilibration buffer through each column.
4. Prime the equilibration and elution buffer lines.
5. Equilibrate the columns by aspirating buffer A through the column once.

## 10. Nickel 1 (Ni1) Column

1. Wash each column with 20 ml Milli-Q water to remove storage buffer. Run 5 ml buffer B and 25 ml buffer A for equilibration.
2. Load the clarified lysate containing solubilized protein into the columns at a rate of 2 ml/min then follow by a 15 ml wash with buffer A.
3. Elute the bound protein in a step gradient with buffers A and B by the following ratios respectfully: 5 ml 95:5, 5 ml 60:40, 10 ml 0:100. Collect each elution fraction separately.
4. Analyze: eluted fractions, crude lysate, clarified lysate, and flow-through by SDS-PAGE. Pool fractions containing the protein and use a Nanodrop to measure  $A_{280}$  to roughly determine the amount of protein present.

## 11. ULP1 Cleavage

1. Keep a small aliquot (250  $\mu$ l) of the Ni1 column pool for subsequent gel analysis. Bring the rest of the Ni1 pool to 10 ml and add ubiquitin-like protease 1 (ULP1) at 1  $\mu$ l/5 mg of total protein to remove the His-Smt affinity tag.
2. Dialyze the Ni1 pool + ULP1 against 2 L of buffer A for 4 hr at 4 °C in a 10 kDa molecular-weight cutoff (MWCO) on a stir plate at 4 °C.
3. After dialysis, run SDS-PAGE of Ni1 pool and Ni1 pool+ULP1 to determine if ULP1 cleavage was successful.

## 12. Nickel 2 (Ni2) Column

1. Load cleaved protein over the same Ni column and repeat step 9.3 at a reduced flow rate of 1 ml/min. The cleaved off tag will bind to the column and the tagless target protein will now flow-through. Collect the flow-through in a fresh container.
2. Wash Ni column with 3 ml buffer A followed by 5 ml of buffer B to elute all His-tagged and nonspecifically bound protein. Collect each fraction separately.
3. Run SDS-PAGE of Ni2 flow-through, wash and Ni2 elution fractions to verify ULP1 cleavage and that protein is present in the flow-through. Use a Nanodrop to measure  $A_{280}$  to roughly determine the presence of protein.

## 13. Concentrating

1. Concentrate the Ni2 flow-through (and Ni2 elution if protein is present) to 5 ml with an Amicon Ultra 10 kDa MWCO centrifuge tube. Spin in 10 min intervals at 4,000 rpm at 4 °C. Mix with a pipette between each spin to prevent protein from over-concentrating along membrane.

## 14. Size-exclusion Chromatography (SEC)

1. Set up a Sephacryl S-100 10/30 GL column (GE healthcare) by equilibrating with 200 ml SEC buffer at a flow rate of 0.5 ml/min on an AKTApurifier system (GE Healthcare).
2. Prepare 10 ml superloops for use on the SEC column according to the manufacturer's instructions.
3. Using a 5 ml syringe, load samples onto superloops and begin the SEC run.
4. Monitor the UV-absorbance trace at 280 nm while collecting small volume fractions.
5. Run SEC fractions via SDS-PAGE.
6. Pool the SEC fractions showing the highest intensity bands.
7. Concentrate pooled SEC fractions. Refer to step 13.1.
8. Aliquot protein into 100  $\mu$ l samples, flash freeze in liquid nitrogen and store at -80 °C.

### CRYSTALLIZATION

## 15. Protein Crystallization

1. Pre-fill each reservoir of a 96-well Compact Jr crystallization plate (Emerald Bio) with 80  $\mu$ l of crystallization screen (Emerald Bio) of choice.
2. Dilute protein with sizing buffer to 2-20 mg/ml and store on ice.
3. Dispense 0.4  $\mu$ l of protein and 0.4  $\mu$ l of the crystallization screen into each of the 96-wells and cover with crystal clear sealing tape (Manco).
4. Store the plate at 16 °C while checking for protein crystallization periodically over the next few weeks under a dissecting microscope.

## 16. Crystal Harvesting

1. Create a cryoprotectant from the mother liquor and ethylene glycol. Cut the clear tape covering the well with the target protein crystal. To an empty well, add 1.6  $\mu$ l of the corresponding crystallizing condition and combine with 0.4  $\mu$ l of ethylene glycol yielding a final concentration of

20% ethylene glycol and 80% crystallizing condition. Note: to optimize crystal diffraction try different cryoprotectants such as: glycerol, oils, low MW polyethylene glycols, and/or at varying percentages of the cryoprotectant.

2. Before harvesting cool down an ALS-style puck in a dewar filled with liquid nitrogen and cover with lid.
3. Harvest the crystal by placing a CryoLoop with the inner diameter matching the size of the crystal on a Magnetic Crystal Wand (Hampton Research) and scoop it directly from the well solution.
4. Immediately dip the CryoLoop with the harvested crystal into the cryoprotectant then submerge in the ALS-style puck to flash freeze the crystal. Repeat for a desired number of crystals.

## 17. Crystal Screening/Data Collection

1. Once harvesting is complete use a puck wand to place the magnetic cryo puck lid on the ALS puck. With bent tongs, flip the puck upside down.
2. Transfer the puck to a Rigaku ACTOR dewar, screw a Puck Pusher onto the puck, and punch off the lid leaving it in the dewar with pins face up.
3. Using JDirector software, screen each crystal under the following parameters: beam slit set to 0.5 degrees, detector distance set to 50 mm, image step to 70 degrees, and exposure length set to 30 sec.
4. Run Mosflm on the test images you shot with JDirector to determine what the best crystal and strategy is for data collection.
5. Collect a complete dataset based on your results from Mosflm.

## 18. Data Processing/Structure Determination

1. Run XDS/XSCALE<sup>4</sup> to process the dataset.
2. Open the CCP4 suite software.
  1. Run Phaser<sup>5</sup> to calculate a molecular replacement solution using a high homology search model, when available. In this case we used the PDBID 3CW4 as a search model<sup>6</sup>.
  2. Run Refmac<sup>7</sup> to refine your molecular model against the observed reflection collected in the dataset. Final resolution should be based off of the highest shell and determined by the following parameters: R factor > 50%, I/sigma > 2, and completeness > 90%.
3. Build a 3-Dimensional electron density model with the molecular graphics software COOT<sup>8</sup>.
4. Before depositing the structure in the PDB validate it with MolProbity<sup>9</sup> software to verify that quality of structure is suitable for deposition.

## Representative Results

The following results illustrate the expected outcomes of the described protocol, and in the case of PB2, the observed outcomes.

Using Gene Composer, five full-length target amino acid sequences of the influenza virus polymerase subunit PB2 were designed (**Figure 2**). The PB2 sequences were back translated and subjected to many engineering steps<sup>3</sup>, resulting in codon harmonized sequences optimized for expression in *E. coli*. From the iPCR products (**Figure 3b**), a total of thirty-four constructs were successfully cloned into a modified pET28 vector system<sup>10</sup> with an N-terminal 6x His-Smt fusion tag using PIPE cloning<sup>3</sup> as shown in **Figure 3a**. A summary of the cloning workflow is presented in **Figure 4**.

After successful cloning, micro-scale protein expression of each construct was tested in BL21(DE3) *E. coli* cells. Cells were grown in TB medium supplemented with Novagen Overnight Express 1 medium (according to the manufacturer's protocol) for 48 hr at 20 °C in a shaking incubator set at 220 rpm. After growth, cells were harvested and tested for soluble protein expression using capillary electrophoreses with a Caliper LabChip 90. Fourteen of the thirty-four PB2 constructs led to soluble target protein and entered large-scale fermentation. Large-scale cultures of each construct were grown in TB medium supplemented with Novagen's Overnight Express 1 medium according to the manufacturer's protocol. After growth, cells were harvested via centrifugation and stored at -80 °C. Large-scale protein expression of each culture was confirmed via SDS-PAGE analysis (**Figure 5**) before proceeding with large-scale purification.

The Protein Maker was used to conduct parallel purification of the fourteen PB2 constructs. The clarified lysates of all fourteen constructs were run through a nickel-chelate column. After determining which fractions contained target protein by SDS-PAGE, the corresponding fractions were pooled for each sample and the concentration of each was determined by an A<sub>280</sub> reading. Removal of the 6x His-Smt tag was conducted by the addition of ULP1 followed by overnight dialysis and a second nickel column. Confirmation of the His-Smt tag removal was conducted by SDS-PAGE (**Figure 6**), and each sample was concentrated with a 10 kDa Amicon Ultra centrifuge tube. After concentration using the Amicon Ultra centrifuge tubes, each sample was run over a sizing column to achieve crystallographic purity. A second concentration was conducted to increase the protein concentration to a level necessary for crystallization. All fourteen constructs were successfully purified and entered into crystallization trials.

Crystallization was initiated by thawing the previously frozen protein. Crystallization was performed in a climate controlled room at 16 °C with specially designed plates (Emerald Bio) for sitting drop vapor diffusion (**Figure 7**). Initial screening was conducted with four sparse-matrix screens; JCSG+, Pact, Wizard Full, and CryoFull (Emerald Bio), following an extended Newman strategy. 0.4 µl of protein solution was then mixed with 0.4 µl of crystallant (or reservoir solution) from the corresponding reservoir using 96-well Compact Jr crystallization plates (Emerald Bio). Of the fourteen purified samples nine of them yielded crystals suitable for diffraction studies (**Figure 8**). An in-house diffraction data set was collected on five of the nine constructs crystallized at Cu Kα wavelength using a Rigaku SuperBright FR-E+ rotating-anode X-ray generator equipped with Osmic VariMax HF optics and a Saturn 944+ CCD detector (**Figure 9**). Each data set was processed with XDS/XSCALE<sup>4</sup> and scaled to a final resolution. Attempts to solve the structures by molecular replacement were carried out with Phaser<sup>5</sup> from the CCP4 suite<sup>7</sup>. The final models were obtained after refinement in REFMAC<sup>7</sup> and manual rebuilding with Coot<sup>11</sup>. The structures were assessed and corrected for

geometry and fitness with *MolProbity*<sup>9</sup>. A total of four structures of the PB2 subunit were determined (Figure 10) and deposited into the PDB. Figure 11 illustrates the overall outcome at each stage in the MTPP pipeline.

# Gene-to-Structure Pathway

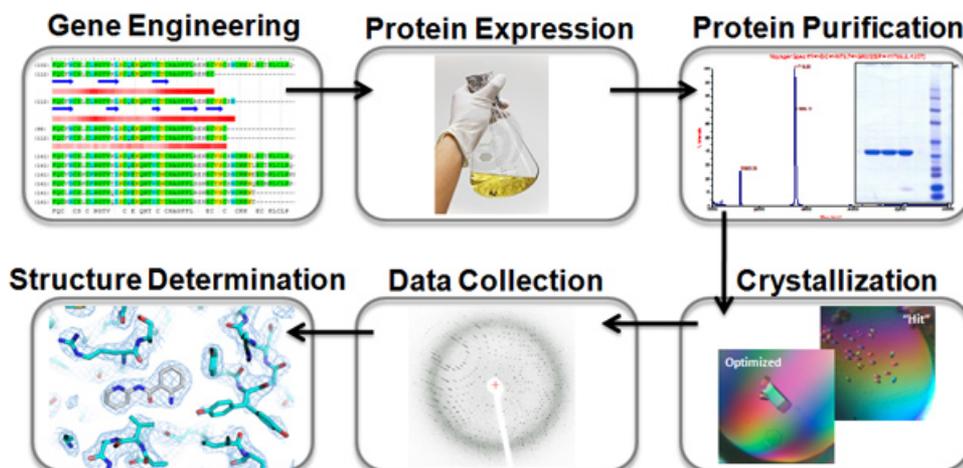


Figure 1. Overview of the SSGCID gene-to-structure pathway for Multi-target parallel processing at Emerald Bio.

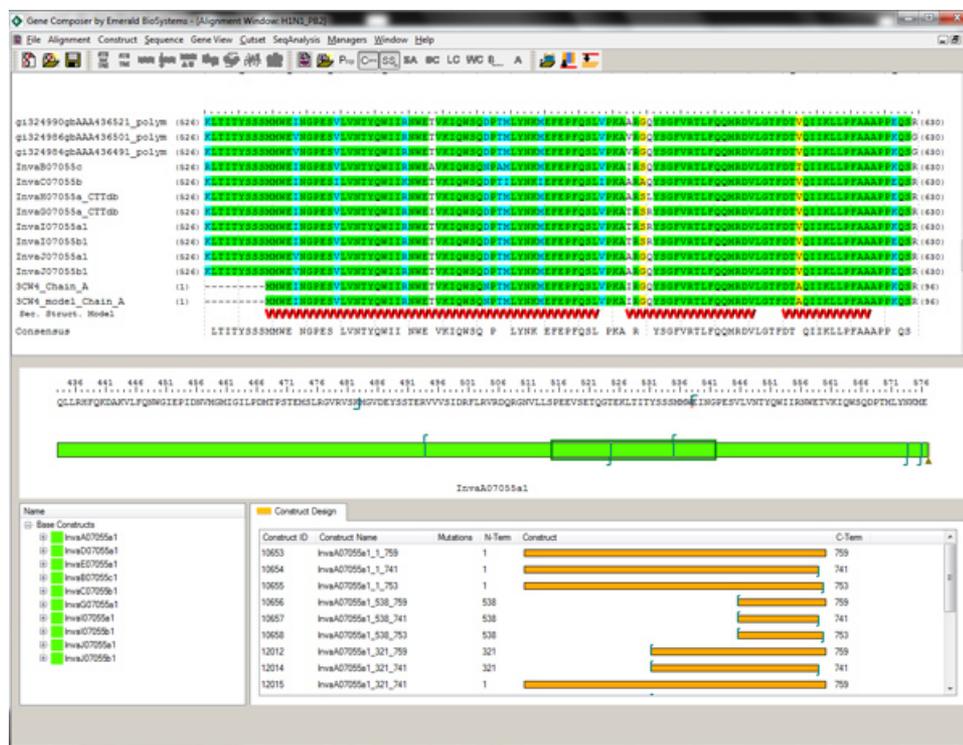
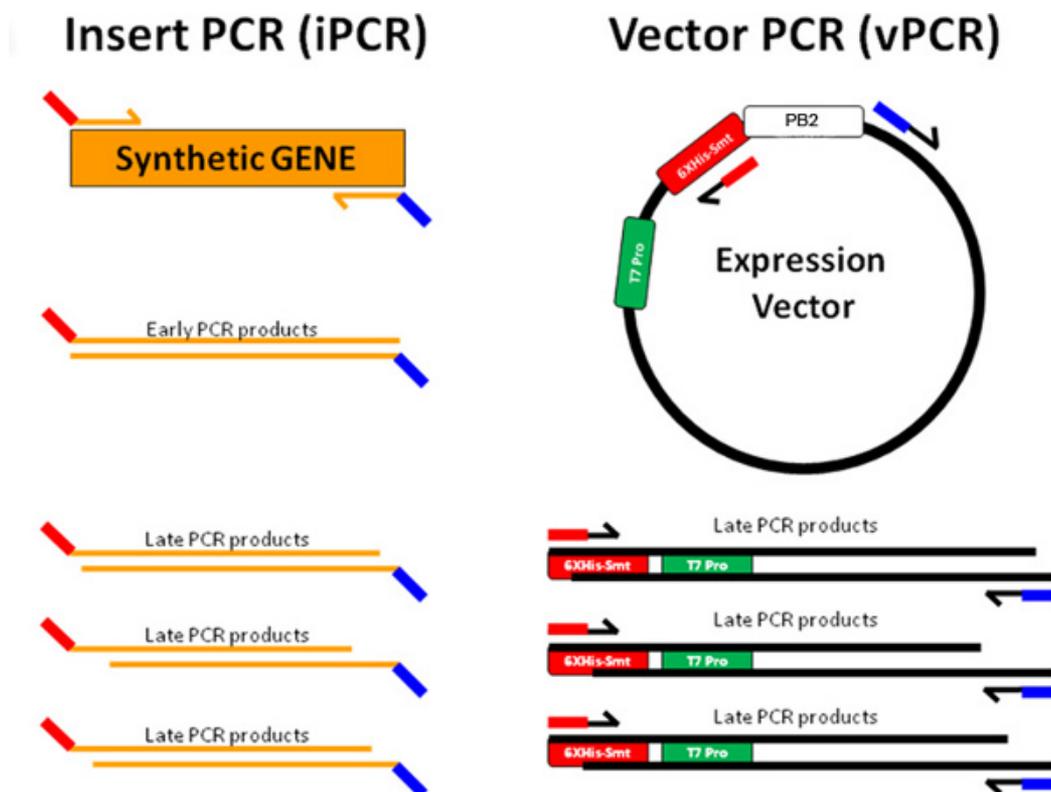
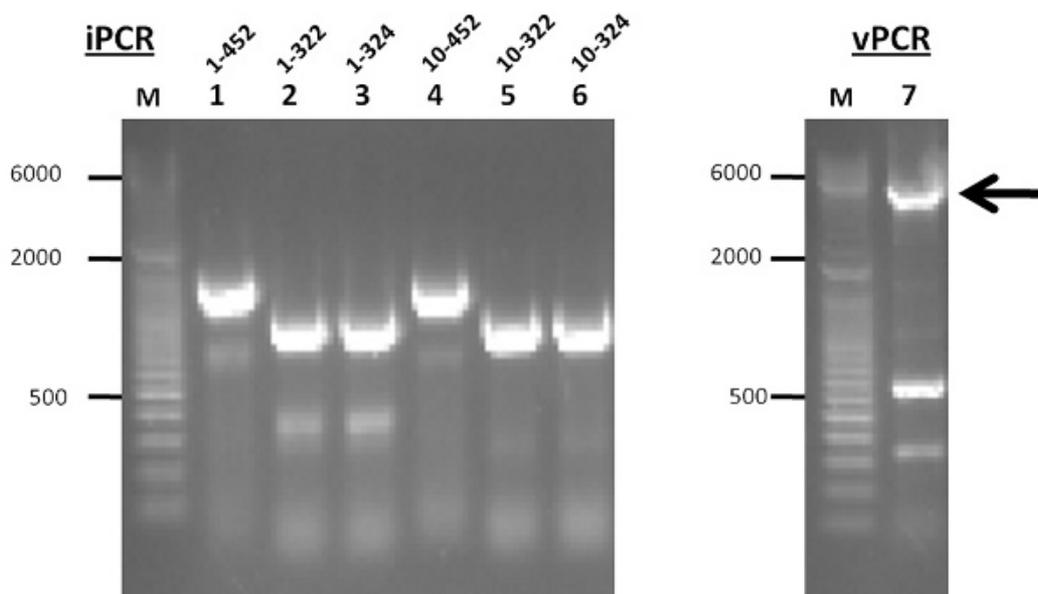


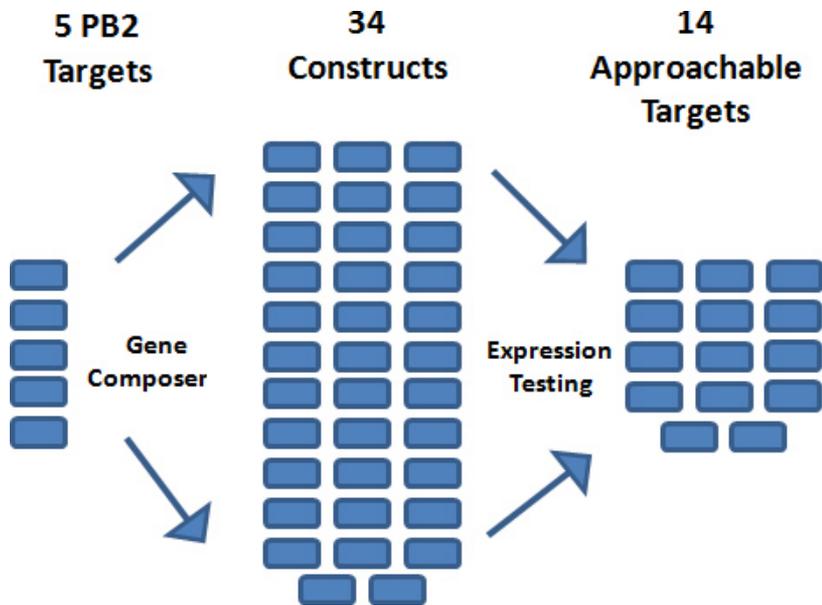
Figure 2. Alignment Viewer and Protein Construct Design Module in Gene Composer software. Amino-acid base construct of target is shown in green (middle window) and the structure guided truncations of alternative constructs are shown in gold (bottom window). An alignment of multiple Flu viral PB2 sequences is shown compared to the sequence and secondary structure elements of the C-terminal domain from PDBID 3CW4. Knowledge of the domain structure and secondary structure elements allows N-terminal truncations to be chosen within the Gene Composer Design Module by right-clicking on the desired amino acid residue. [Click here to view larger figure.](#)



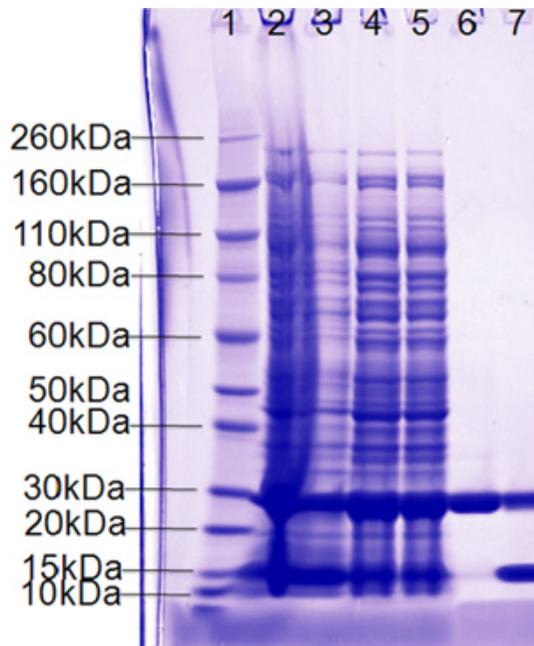
**Figure 3a.** PIPE cloning is illustrated wherein the synthetic gene insert (orange) is amplified by designed forward (red-orange lines) and reverse (orange-blue lines) primers to generate insert PCR material. The expression vector is amplified with reverse (red-black lines) and forward (blue-black lines) primers to generate vector PCR material. The terminal sequences iPCR products are complementary to the terminal sequences of vPCR products (red of iPCR complements red of vPCR and blue of iPCR complements blue of vPCR). This allows the iPCR and vPCR products to anneal to form plasmids that are replicated upon transformation into host BL21(DE3) chemically competent *E. coli* cells.



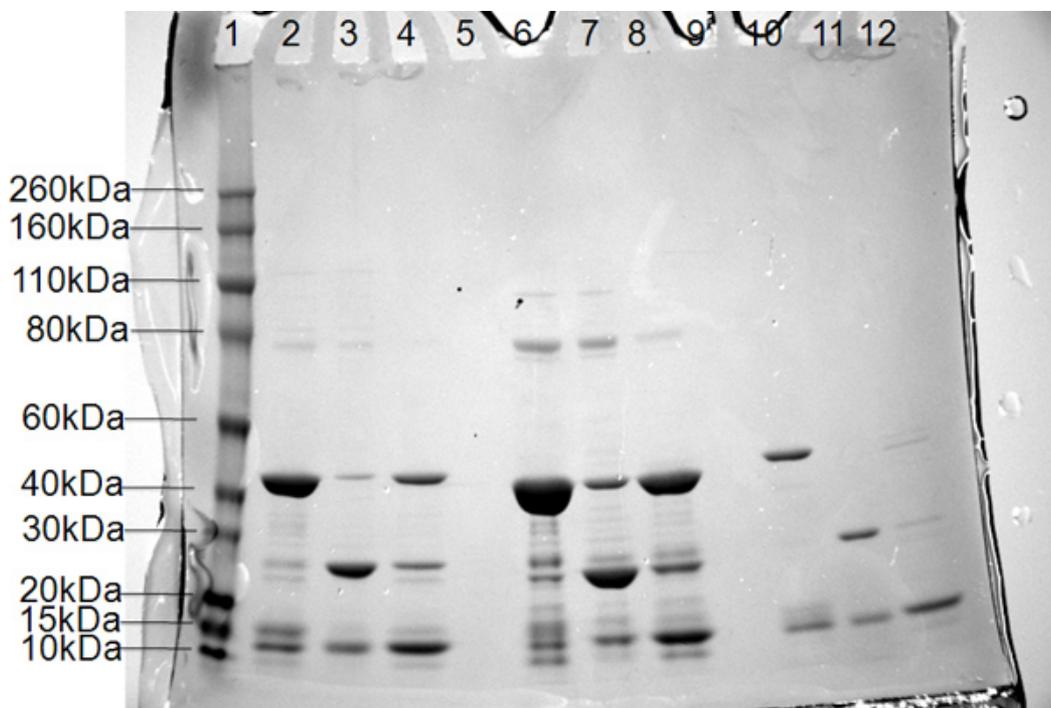
**Figure 3b.** Agarose gel analysis of iPCR products from the PB2 subunit. iPCR failures may be seen as faint or smeary bands, while successful iPCR products are represented by robust bands. iPCR product quality can generally be correlated with cloning success. Molecular weight markers are in kiloDaltons. Figure is reproduced from Raymond *et al.*, 2011<sup>12</sup>.



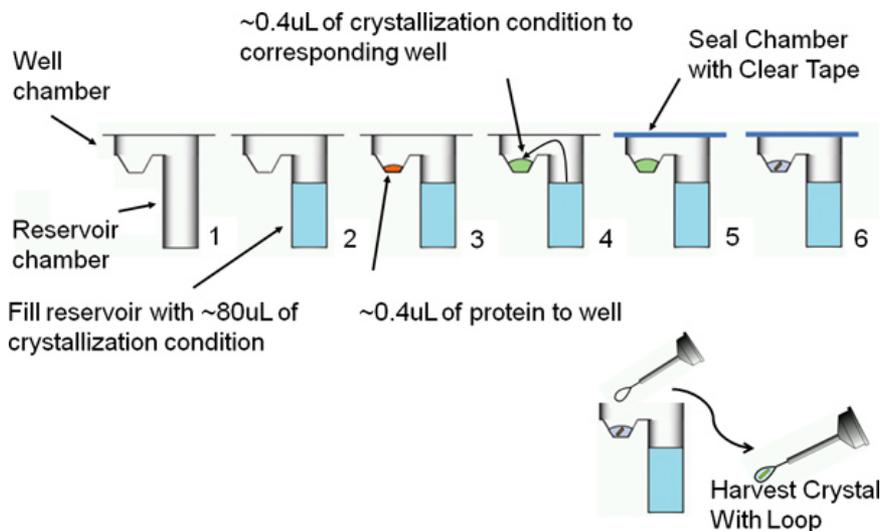
**Figure 4.** Gene engineering steps of target PB2 proteins were performed using Gene Composer software. After the engineered nucleic acid sequence was established for each target, 6-7 alternative protein constructs were designed for each. Multi-target parallel processing in the initial steps of gene design and cloning resulted in 34 constructs, 14 of which were viable targets that produced soluble proteins in *E. coli*.



**Figure 5.** Representative SDS-PAGE analysis of large scale fermentation showing robust protein expression (expected size of 25.76 kDa), roughly 50% soluble (lane 4) and about 50% cleavage of 6x His-Smt tag from eluted protein (lane 7).



**Figure 6. SDS-PAGE results for three constructs of the polymerase PB2 subunit.** Lane 1, molecular-weight markers (labeled on the left in kDa); lanes 2, 6, and 10, pooled protein from Nickel 1 column; lanes 3, 7, and 11, flow-through of cleaved protein in buffer A from Nickel 2; lanes 4, 8, and 12, removal of 6x His-Smt tag in buffer B from Nickel 2.



**Figure 7. A schematic of vapor diffusion by the sitting drop method.** The sitting drop method for protein crystallization falls under the category of vapor diffusion. This method entails a purified sample of protein and precipitant to equilibrate with a larger reservoir containing similar conditions in a higher concentration. As water vaporizes from the protein sample and transfers to the reservoir, the precipitant concentration increases to an optimal level for protein crystallization.

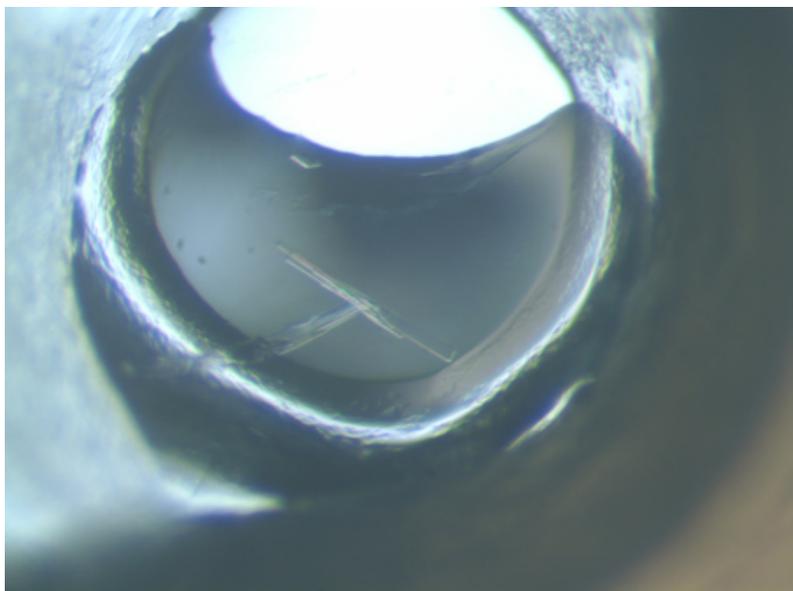


Figure 8. Protein crystal of polymerase PB2 subunit from a strain of the influenza virus.

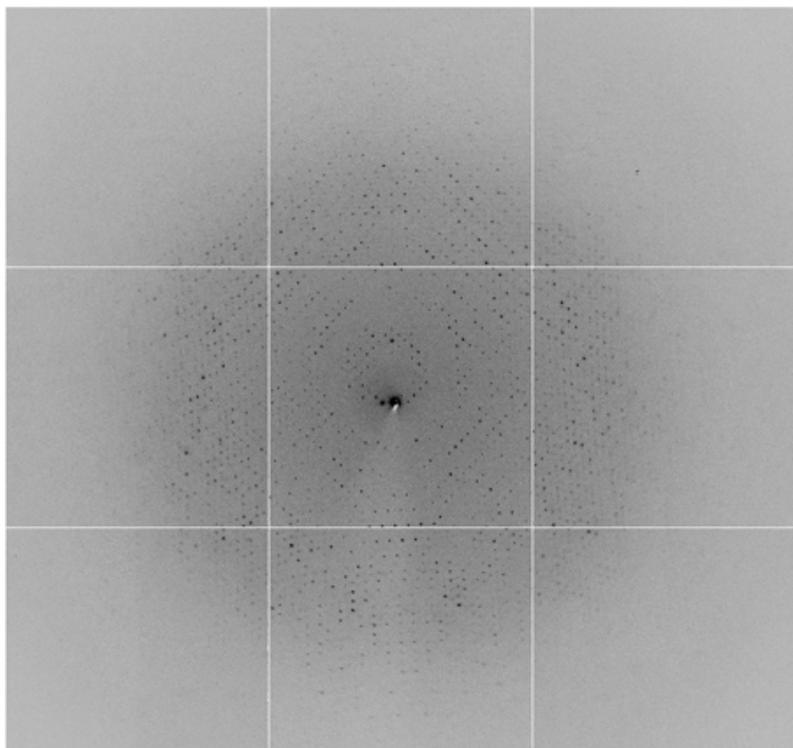
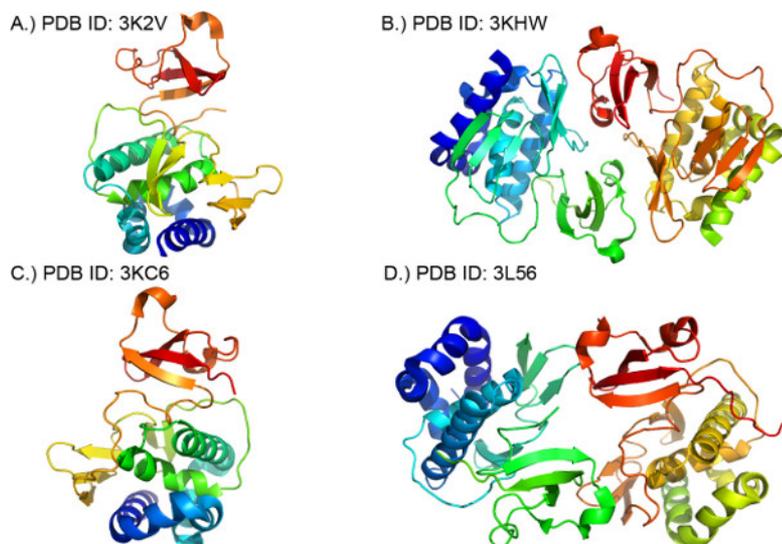


Figure 9. X-ray diffraction image of the polymerase PB2 subunit from a strain of the influenza virus.



**Figure 10. Ribbon diagrams of the molecules in the crystallographic asymmetric unit of 4 PB2 structures.** Secondary structures colored in rainbow pattern with corresponding PDB codes. (a) 3K2V (A/Yokohama/2017/2003/H3N2) (b) 3KHW (A/Mexico/InDRE4487/2009/H1N1) (c) 3KC6 (A/Vietnam/1203/2004/H5N1) (d) 3L56 (A/Vietnam/1203/2004/H5N1).

Strain	Polymerase Subunit	Target	Constructs	Cloned	Soluble	Purified	Crystals	Structure
Avain-1023 H5N1	PB2	InvaA.07055.a	7	7	3	3	2	2
Avain-2017 H2N3	PB2	InvaA.07055.c	7	7	2	2	2	1
Equine-1 H7N7	PB2	InvaA.07055.b	7	7	3	3	3	0
Swine-04 H1N1	PB2	InvaD.07055.a	7	5	3	3	1	1
Swine-InDRE4487 H1N1	PB2	InvaE.07055.a	6	6	3	3	1	1
<b>Total</b>		<b>5</b>	<b>34</b>	<b>33</b>	<b>14</b>	<b>14</b>	<b>9</b>	<b>5</b>

**Figure 11. Outcome analysis for influenza PB2 targets by the methods described.** The structure determination pipeline is illustrated in five steps: Cloning, solubility, purification, crystallization and structure determination.

## Discussion

### Multi-Target Parallel Processing

Structure-based drug design plays an important role in drug discovery. The SSGCID is dedicated to providing the scientific community with three-dimension protein structures from NIAID category A-C pathogens. Making such structural information widely available will ultimately serve to accelerate structure-based drug design.

The first critical step of the MTPP approach is construct design. Multiple constructs of each target protein increases the probability of successful structure determination and increases turnaround. It is inevitable that some protein constructs will fail during stages of the pipeline. Implementing the PIPE cloning method supports the MTPP method by allowing the generation of many constructs in 96-well format without labor intensive purification steps. Pairing PIPE cloning with the ability to analyze protein expression in the same 96-well format (Caliper LabChip 90) further expedites the overall flow. The pairing of these methods allows for quick identification of constructs that produce soluble protein which ensures the success of large-scale protein production and purification.

An essential aspect to the success of the MTPP high-throughput is the Protein Maker (US Patent No. 6818060, Emerald Bio) instrument. The Protein Maker is a 24-channel parallel liquid-chromatography system developed specifically to boost the efficiency of high-throughput protein production and related structural genomic pipeline research applications. Using the previously described protocol for the Protein Maker, the advantages are apparent in comparison to a single line FPLC system. A single person can purify up to 48 targets in parallel within an eight hour period. In contrast, a single person using a single line FPLC system can only purify a maximum of four targets within the same timeframe. The high levels of purity for each target achieved with the Protein Maker are a critical factor in the later success of growing protein crystals for structure analysis.

### Limitations and Troubleshooting

Solving three-dimensional structures by x-ray crystallography is a multi-staged effort with many challenges, one of which is the inability to obtain large amounts of soluble target protein. One strategy that can be implemented to overcome the solubility problem is the use of an alternative expression host as *E. coli* cells are unable to perform several important eukaryotic post-translational modifications. Expression in various yeast, insect and mammalian cell lines that are capable of performing these post-translational modifications are often a suitable alternative. Target proteins are sometimes expressed but completely insoluble in the standard lysis conditions. The Protein Maker can be a valuable resource for the rapid testing of alternative cell lysis conditions as described in Smith *et al.* 2011<sup>13</sup>. This strategy is often necessary to keep targets moving

through the pipeline. In any structural genomics pipeline, standardized protocols may not be suitable for every target that comes through the pipeline and targets may need individual optimization. For example, we have chosen to use 20% ethylene glycol for every cryoprotectant. In cases that this condition is not suitable, alternative cryoprotectants or concentrations may need to be tested.

Due to the unique nature of each individual protein target, the rate-limiting and unpredictable step in determining a structure is crystallization. The MTPP pipeline offsets the commonly low success rate of protein crystallization with optimization from the initial sparse matrix screens. Each initial crystal hit from commercially available sparse matrix screens is further optimized with an E-Screen Builder (Emerald Bio). The optimization screen is designed around the condition of the initial crystal hit, altering the concentrations of the buffers, salts, and additives. Successful optimization screens yield crystals suitable for diffraction studies and structure determination.

The structural genomics program put forth by the National Institute of Allergy and Infectious Disease (NIAID) provides funding to Emerald Bio and three other Pacific Northwest institutions who together are the SSGCID (Emerald Bio, SeattleBiomed, the University of Washington and Pacific Northwest National Laboratory). Each member of the consortium was chosen for their expertise in applying state-of-the-art technologies required for accomplishing the goals of the NIAID structural genomics program. To date, SSGCID has deposited 461 structures into the PDB ranking it as the seventh largest contributor in the world, and in 2011, the most productive. The protocols and methodologies of the SSGCID are provided with the intention of benefiting the scientific community and perpetuating the research of infectious diseases.

## Disclosures

The authors are employees of Emerald Bio, Inc.

## Acknowledgements

The authors would like to thank all members of the SSGCID consortium. Achievement of the SSGCID's goals is made possible by the tremendous efforts of all team members at Emerald Bio. This research was funded under Federal Contract No. HHSN272200700057C from the National Institute of Allergy and Infectious Diseases, the National Institutes of Health and the Department of Health and Human Services.

## References

1. Lowen, A.C., Mubareka, S., Steel, J., & Palese, P. Influenza virus transmission is dependent on relative humidity and temperature. *PLoS Pathogens*. 3 (10), 1470-1476 (2007).
2. Yamada S., *et al.* Biological and structural characterization of a host-adapting amino acid in influenza virus. *PLoS Pathog.* 6, e1001034 (2010).
3. Lorimer, D., Raymond, A., Walchli, J., Mixon, M., Barrow, A., Wallace, E., Grice, R., Burgin, A. & Stewart, L. Gene Composer: database software for protein construct design, codon engineering, and gene synthesis. *BMC Biotechnol.* 9, 36 (2009).
4. Kabsch, W. Integration, scaling, space-group assignment and post-refinement. *Acta Cryst. D.* 66, 125-132 (2010).
5. McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., & Read, R.J. J. Phaser crystallographic software. *Appl. Cryst.* 40, 658-674 (2007).
6. Kuzuhara, T., Kise, D., *et al.* Structural basis of the influenza A virus RNA polymerase PB2 RNA-binding domain containing the pathogenicity-determinant lysine 627 residue. *J. Biol. Chem.* 284, 6855-6860 (2009).
7. Murshudov, G.N., Skubák, P., Lebedev, A.A., Pannu, N.S., Steiner, R.A., Nicholls, R.A., Winn, M.D., Long, F., & Vagin, A.A. REFMAC5 for the refinement of macromolecular crystal structures. *Acta Cryst. D.* 67, 355-367 (2011).
8. Cowtan, K. Recent developments in classical density modification. *Acta Cryst. D.* 66, 470-478 (2010).
9. Chen, V.B., Arendall, W.B., Headd, J.J., Keedy, D.A., Immormino, R.M., Kapral, G.J., Murray, L.W., Richardson, J.S. & Richardson, D.C. MolProbity: all-atom structure validation for macromolecular crystallography. *Acta Cryst. D.* 66, 12-21 (2010).
10. Mossessova, E. & Lima, C.D. Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast. *Mol. Cell.* 5, 865-876 (2000).
11. Emsley, P., Lohkamp, B., Scott, W.G. & Cowtan, K. Features and development of Coot. *Acta Cryst. D.* 66, 486-501 (2010).
12. Raymond, A.C., Haffner, T.E., Ng, N., Lorimer, D., Staker, B.L., & Stewart, L.J. Gene design, cloning and protein-expression methods for high-value targets at the Seattle Structural Genomics Center for Infectious Disease. *Acta Cryst. F.* 67, 992-997 (2011).
13. Smith, E.R., Begley, D.W., Anderson, V., Raymond, A.C., Haffner, T.E., Robinson, J.I., Edwards, T.E., Duncan, N., Gerdtts, C.J., Mixon, M.B., Nollert, P., Staker, B.L., & Stewart, L.J. The Protein Maker: an automated system for high-throughput parallel purification. *Acta Cryst. F.* 67, 1015-1021 (2011).