FINAL SCRIPT: APPROVED FOR FILMING

Submission ID #: 68933
Scriptwriter Name: Sulakshana Karkala
Project Page Link: https://review.jove.com/account/file-uploader?src=21024613

Title: Pharmacophore Modeling for Targets with Extensive Ligand Libraries: A Case Study on SARS-CoV-2 Mpro

Authors and Affiliations:
L. J. Córdova-Bahena[1,2], S. M. Pérez-Tapia[3,4,5], M. A. Velasco-Velázquez[1]

[1]School of Medicine, Universidad Nacional Autónoma de México (UNAM)
[2]Investigadores por México, Secretaría de Ciencia, Humanidades, Tecnología e Innovación (SECIHTI)
[3]Research and Development in Biotherapeutics Unit (UDIBI), National School of Biological Sciences, National Polytechnic Institute
[4]National Laboratory for Specialized Services of Investigation, Development and Innovation (I+D+i) for Pharma Chemicals and Biotechnological products, LANSEIDI-FarBiotec-CONAHCYT
[5]Immunology Department, National School of Biological Sciences, National Polytechnic Institute

Corresponding Authors:
Marco A. Velasco-Velázquez marcovelasco@unam.mx
Email Addresses for All Authors:
L. J. Córdova-Bahena luisbahena@unam.mx
S. M. Pérez-Tapia mayra.perez@udibi.com.m
Marco A. Velasco-Velázquez marcovelasco@unam.mx

Author Questionnaire

1. **Microscopy**: Does your protocol require the use of a dissecting or stereomicroscope for performing a complex dissection, microinjection technique, or something similar? **No**

2. **Software:** Does the part of your protocol being filmed include step-by-step descriptions of software usage? **Yes, all done**

3. **Filming location:** Will the filming need to take place in multiple locations? **No**

4. **Testimonials (optional):** Would you be open to filming two short testimonial statements **live during your JoVE shoot**? These will **not appear in your JoVE video** but may be used in JoVE's promotional materials. **Yes**

Current Protocol Length
Number of Steps: 09
Number of Shots: 24

Introduction

Videographer: Obtain headshots for all authors available at the filming location.

1.1. **Luis Córdova-Bahena:** Our research explores pharmacophore modeling with large ligand libraries. In this study, we focus on SARS-CoV-2 Mpro to identify structural patterns that guide drug discovery.

1.1.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B.roll:4.1*

What are the most recent developments in your field of research?

1.2. **Luis Córdova-Bahena:** Recent advances in pharmacophore modeling use larger and more diverse datasets, which improve accuracy in identifying key molecular features for drug discovery.

1.2.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B.roll:3.4*

What technologies are currently used to advance research in your field?

1.3. **Luis Córdova-Bahena:** Molecular docking, molecular dynamics simulations, and machine learning are currently used to advance pharmacophore modeling and improve drug discovery outcomes.

1.3.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera.

What are the current experimental challenges?

1.4. **Luis Córdova-Bahena:** A key challenge is managing large and diverse ligand datasets while improving feature extraction accuracy and reducing computational costs.

1.4.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B.roll:3.5*

What significant findings have you established in your field?

1.5. **Luis Córdova-Bahena:** We have shown that extensive ligand libraries improve pharmacophore feature accuracy, enabling more reliable predictions and better identification of potential drug candidates.

1.5.1. INTERVIEW: Named Talent says the statement above in an interview-style shot, looking slightly off-camera. *Suggested B.roll:4.4*

Videographer: Obtain headshots for all authors available at the filming location.

Testimonial Questions (OPTIONAL):

Videographer: Please capture all testimonial shots in a wide-angle format with sufficient headspace, as the final videos will be rendered in a 1:1 aspect ratio. Testimonial statements will be presented live by the authors, sharing their spontaneous perspectives.

How do you think publishing with JoVE will enhance the visibility and impact of your research?

1.6. **Luis Cordova Bahena, Senior Researcher at UNAM** : (authors will present their testimonial statements live)

Can you share a specific success story or benefit you've experienced—or expect to experience—after using or publishing with JoVE? (This could include increased collaborations, citations, funding opportunities, streamlined lab procedures, reduced training time, cost savings in the lab, or improved lab productivity.)

1.7. **Luis Cordova Bahena, Senior Researcher at UNAM** : (authors will present their testimonial statements live)

Protocol

2. Construction of Consensus Pharmacophore Models from Ligand-Target Complexes Using PyMOL

Demonstrator: Luis Córdova-Bahena

2.1. To begin, launch the PyMOL software **[1].** Align all protein–ligand complexes on the software **[2]**. For each aligned complex, extract the ligand conformer and save it as a separate file in SDF *(S-D-F)* format **[3]**.

2.1.1. WIDE: Talent launches the PyMOL software on a computer.

2.1.2. SCREEN: 2.1.mov. 00:34-00:54

2.1.3. SCREEN: 2.1.mov. 01:29-01:49,02:28-02:44

2.2. Upload each ligand file individually to Pharmit using the **Load Features** option **[1]**. After loading, use the **Save Session** option to download the pharmacophore file in the JSON *(J-son)* format **[2]**. Store all downloaded pharmacophore JSON files together in a single folder for later use in the Google Colab *(Co-lab)* environment **[3]**.

2.2.1. SCREEN: 2.2.mov. 00:00-00:12

2.2.2. SCREEN: 2.2.mov. 00:12-00:19

2.2.3. SCREEN: 2.2.mov. 01:14-01:30

3. Construction of Consensus Pharmacophore Models from Ligand-Target Complexes Using ConPhar

3.1. Launch Google Colab in a web browser and create a new notebook **[1]**. Install Conda and PyMOL **[2]**, then run the cell by clicking the **play** icon or pressing Shift and Enter to verify successful execution. A green bar will appear above the cell if completed correctly **[2]**.

3.1.1. SCREEN: 3.1.1.mov 00:04-00:15

3.1.2. SCREEN: 3.1.2-&-3.1.2.mov 00:03-00:10

3.1.3. SCREEN: 3.1.2-&-3.1.2.mov 02:30-03:01

3.2. Now, install a the ConPhar *(Con-Far)* Python package and import the required modules **[1]**. Run the cell by clicking the **Play** icon to confirm successful installation. A confirmation message will appear upon successful execution **[2]**.

3.2.1. SCREEN: 3.2.1-&-3.2.2.mov 00:00-00:06

3.2.2. SCREEN: 3.2.1-&-3.2.2.mov. 00:07-00:25, 00:31-00:33

3.3. Next, create a new folder for storing pharmacophore JSON files **[1]**. Open the folder, right-click inside it, and select **Upload** to upload the JSON files **[2]**.

3.3.1. SCREEN: 3.3.1-&-3.3.2.mov. 00:00-00:11

3.3.2. SCREEN: 3.3.1-&-3.3.2.mov. 00:12-00:30

3.4. To parse and consolidate pharmacophoric features, extract the features from all uploaded files **[1].** Click on the Play icon to run the cell **[2]** and generate a consolidated DataFrame containing the features from each individual ligand **[3]**.

3.4.1. SCREEN: 3.4.1-&-3.4.2-&-3.4.3-.mov 00:00-00:06

3.4.2. SCREEN: 3.4.1-&-3.4.2-&-3.4.3-.mov. 00:07-00:10

3.4.3. SCREEN: 3.4.1-&-3.4.2-&-3.4.3-.mov. 00:11-00:15

3.5. Display all pharmacophoric descriptors **[1].** Then run the cell to display clustering patterns **[2]**.

3.5.1. SCREEN: 3.5.1-&-3.5.2.mov 00:00-00:06

3.5.2. SCREEN: 3.5.1-&-3.5.2.mov 00:07-00:20

3.6. Now, save the pharmacophore model in the PyMOL format**.** Run the cell to generate a .pse *(dot-P-S-E)* file **[1]**. Then, save the model as a JSON file by running the corresponding code cell **[2]**.

3.6.1. SCREEN: 3.6.1-&-3.6.2.mov 00:30

3.6.2. SCREEN: 3.6.3.mov. 00:03-00:12

3.7. Generate feature-clustered outputs and dendrograms using the provided code **[1]**. Click on Play and export the consensus pharmacophore data to a CSV *(C-S-V)* file by executing the export command **[2]**. Use the generated consensus pharmacophore model for virtual screening **[3]**.

3.7.1. SCREEN: 3.7.1-(1).mov 00:06-00:22

3.7.2. SCREEN: 3.7.2.mov. 00:00-00:11

3.7.3. SCREEN: 3.7.3.mov. 00:08-00:32

Results

4. Results

4.1. One hundred main protease or Mpro *(m-pro)* complexes co-crystallized with different non-covalent inhibitors were aligned to visualize ligand diversity within the binding pocket **[1]**. Each ligand was individually extracted from the aligned complexes to generate separate files for virtual screening **[2]**. The extracted ligands were uploaded to the Pharmit server and used to generate structured JSON files for pharmacophore modeling **[3]**.

4.1.1. LAB MEDIA: Figure 9 *Video Editor: Please highlight A*

4.1.2. LAB MEDIA: Figure 9 *Video Editor: Please highlight B*

4.1.3. LAB MEDIA: Figure 9 *Video Editor: Please highlight C*

4.2. The JSON files yielded 1,450 pharmacophoric features grouped into 110 clusters **[1]**, including 23 aromatic **[2]**, 30 hydrogen bond acceptors **[3]**, 16 hydrogen bond donors **[4]**, 36 hydrophobic **[5]**, and 5 anion clusters **[6]**.

4.2.1. LAB MEDIA: Figure 10 *Video Editor: Please highlight A*

4.2.2. LAB MEDIA: Figure 10B *Video Editor: Please highlight image labeled "Aromatic"*

4.2.3. LAB MEDIA: Figure 10B *Video Editor: Please highlight image labeled "HBA"*

4.2.4. LAB MEDIA: Figure 10B *Video Editor: Please highlight image labeled "HBD"*

4.2.5. LAB MEDIA: Figure 10B *Video Editor: Please highlight image labeled "Hyd"*

4.2.6. LAB MEDIA: Figure 10B *Video Editor: Please highlight image labeled "Anion"*

4.3. To build the consensus model, only the most densely populated clusters were retained: **[1]**, while anion clusters were excluded due to insufficient members **[2]**.

4.3.1. LAB MEDIA: Figure 10B *Video Editor: Please highlight images labeled "Aromatic, HBA, HBD, Hyd"*.

4.3.2. LAB MEDIA: Figure 10B *Video Editor: Please highlight image labeled "Anion"*

4.4. The final consensus pharmacophore model consisted of 11 features comprising 3 aromatic, 4 hydrogen bond acceptors, 2 hydrogen bond donors, and 2 hydrophobic features **[1]**.

4.4.1. LAB MEDIA: Figure 11 *Video Editor: Please highlight A. Sequentially highlight the purple, orange, grey and green circles when VO says " aromatic", "*

hydrogen bond acceptors", "hydrogen bond donors" and " Hydrophobic deatures"

4.5. After removing the least representative aromatic feature, Aro 1 *(arrow-one)* , two candidate compounds were identified **[1]**. The two-dimensional chemical structure of compound 101267741 *(one-zero-one-two-six-seven-seven-four-one)* was visualized to confirm its compatibility with the pharmacophore model **[2]**.

4.5.1. LAB MEDIA: Figure 11 *Video Editor: Please highlight B*

4.5.2. LAB MEDIA: Figure 11 *Video Editor: Please highlight C*

4.6. Compound 101267741 fit deeply into the Mpro binding pocket, occupying the S1 and S2 subpockets more effectively **[1]** than the reference ligand 38a *(Thirty-eight-A)* , which spread toward the S1' *(S-one-dash)* pocket **[2]**.

4.6.1. LAB MEDIA: Figure 11D (left panel). *Video editor: Highlight the green compound conformer*

4.6.2. LAB MEDIA: Figure 11D (middle panel). *Video editor: Highlight the purple compound conformer*

4.7. Compound 101267741 formed 11 intermolecular interactions—7 hydrogen bonds and 4 hydrophobic contacts **[1]**.

4.7.1. LAB MEDIA: Figure 11D (right panel).

Pronunciation Guide:

1. **SARS-CoV-2**
   Pronunciation link: https://dictionary.cambridge.org/us/pronunciation/english/sars-cov-2
   IPA: /ˌsɑːrz.koʊ.viːˈtuː/
   Phonetic spelling: SARZ-koh-vee-TOO
2. **Pharmacophore**
   Pronunciation link: https://www.howtopronounce.com/pharmacophore
   IPA: /ˌfɑːr.məˈkɔːr/ or /ˌfɑːr.məˈkoʊr/
   Phonetic spelling: far-muh-KOR
3. **Mpro**
   (short for "main protease")
   IPA (approximate): /ˈɛm.proʊ/
   Phonetic spelling: EM-proh
4. **Ligand**
   IPA: /ˈlaɪ.gənd/
   Phonetic spelling: LYE-guhnd
5. **Conformer**
   IPA: /kənˈfɔːr.mər/
   Phonetic spelling: kun-FOR-mer
6. **Dendrogram**
   IPA: /ˈdɛn.drəˌgræm/
   Phonetic spelling: DEN-druh-gram
7. **Aromatic**
   IPA: /ˌær.əˈmætɪk/
   Phonetic spelling: air-uh-MAT-ik
8. **Hydrophobic**
   IPA: /ˌhaɪ.drəˈfoʊ.bɪk/
   Phonetic spelling: hy-drə-FOH-bik
9. **Hydrogen bond acceptor**
   - hydrogen: /ˈhaɪ.drə.dʒən/ → HY-dro-jen
   - acceptor: /əkˈsɛp.tər/ → uh-SEP-tur
     Combined: HY-dro-jen bond uh-SEP-tur
10. **Hydrogen bond donor**
    - donor: /ˈdoʊ.nər/ → DOH-ner
      Combined: HY-dro-jen bond DOH-ner
11. **Cluster / clustering**
    - cluster (noun): IPA: /ˈklʌs.tər/ → KLUHS-ter
    - clustering (verb / gerund): /ˈklʌs.tər.ɪŋ/ → KLUHS-ter-ing
12. **Diverse**
    IPA: /daɪˈvɜrs/
    Phonetic spelling: dye-VERS
13. **Extract / extraction**

- extract (verb): /ɪkˈstrækt/ → ik-STRAKT
- extraction (noun): /ɪkˈstræk.ʃən/ → ik-STRAK-shun

14. **Intermolecular**
    IPA: /ˌɪn.tər.məˈlɛk.jə.lər/
    Phonetic spelling: in-ter-muh-LEK-you-lar
15. **Virtual screening**
    - virtual: /ˈvɜr.tʃu.əl/ → VUR-choo-uhl
    - screening: /ˈskriː.nɪŋ/ → SCREEN-ing