



Biochar EDLC Chemistry Analysis Platform V5.0

BioCharge at CSUDH

USER OPERATIONAL MANUAL

A complete guide for users with base chemistry knowledge and beginner-level coding experience

Version 5.0 | February 2026

About This Manual

This manual will walk you through everything you need to get Biochar EDLC Chemistry Analysis Platform V5.0 running on your computer, from installing Python for the very first time, to running advanced machine learning predictions on chemical reactions. No prior programming experience is required.

What this manual covers:

- Installing Python and all required software from scratch
- Understanding what the program does in plain language
- Step-by-step instructions for every menu option
- Troubleshooting common errors
- Glossary of key terms

 This program was built specifically to support biochar supercapacitor research. If you are studying how pyrolysis reactions produce carbon materials useful for energy storage, this tool is designed for you.

Section 1: What Is this Chemistry Platform?

Biochar EDLC Chemistry Analysis Platform V5.0 is a computational chemistry platform built for biochar and supercapacitor research. In plain terms, it is a Python program that lets you:

- Predict what chemical products will form under a chosen condition(s)
- Estimate how well those products could store electrical energy (capacitance)
- Visualize molecules in 3D on your screen
- Analyze large sets of reactions automatically (batch processing)
- Use machine learning to improve predictions over time
- Run quantum chemistry calculations to understand electron behavior

You do not need to write any code to use this program. It runs through a simple numbered menu that you interact with by typing numbers and pressing Enter.

1.1 The Science in Plain Language

Biochar is a carbon-rich solid produced when organic material (like plant waste) is heated without oxygen, a process called pyrolysis. When biochar is made under the right conditions, it can store electrical charge like a tiny battery. This type of energy storage device is called an Electric Double-Layer Capacitor (EDLC).

This platform takes a starting molecule (your feedstock), predicts what it will break down into in a specific condition, for example like electrolysis, and then scores those breakdown products on how useful they would be inside a supercapacitor.

1.2 What Makes V5.0 Special?

Version 5.0 adds two major new capabilities on top of the previous versions:

- **DFT (Density Functional Theory) Calculations:** Quantum-level calculations that can determine a molecule's energy levels, including its HOMO (highest occupied molecular orbital) and LUMO (lowest unoccupied molecular orbital), key data for understanding how electrons move in a material.
- **GNN (Graph Neural Network) Predictions:** An AI model that treats each molecule as a mathematical graph and learns to predict properties like capacitance. The more experimental data you provide, the more accurate it becomes.

Section 2: Installation from Scratch

Follow these steps in order. Do not skip any step. This section assumes you are starting with a brand-new computer that has no Python or chemistry software installed.



Estimated time to complete full installation: 30–60 minutes depending on your internet speed.

2.1 Install Python

Python is the programming language this code is written in. You need to install it before anything else.

1. **Go to the official Python website:** Open your web browser and navigate to: <https://www.python.org/downloads/>
2. **Download Python 3.11 or newer:** Click the large yellow 'Download Python' button. This will be downloaded as an installer file.
3. **Run the installer:** Double-click the downloaded file to start the installation wizard.
4. **CRITICAL — Check 'Add Python to PATH':** On the very first screen of the installer, you will see a checkbox at the bottom that says 'Add python.exe to PATH'. Make sure this box is CHECKED before clicking Install Now. If you miss this step, commands will not work.
5. **Complete the installation:** Click 'Install Now' and wait for it to finish. Click 'Close' when done.

- To verify Python installed correctly, open your terminal (see Section 2.2) and type:
python –version (You should see something like: Python 3.11.x)

2.2 Opening Your Terminal (Command Prompt)

The terminal is a text-based window where you type commands. Think of it as talking directly to your computer. You will use it to install libraries and run the program. (or directly run the program in an IDE like VScode or PyCharm)

On Windows:

6. Press the Windows key on your keyboard
7. Type: Command Prompt
8. Click on the 'Command Prompt' application

9. A black window with text will appear — this is your terminal

On Mac:

10. Press Command + Space to open Spotlight Search
11. Type: Terminal
12. Press Enter
13. A white or dark window will appear — this is your terminal

On Linux:

14. Right-click on your desktop
15. Select 'Open Terminal' or press Ctrl + Alt + T

2.3 Place the Program File

Before installing libraries, organize your files:

16. Create a new folder on your Desktop and name it: ChemistryCode
17. Move the file Chemistry_Code_V5_0_With_DFT_and_GNN.py into that folder
18. In your terminal, navigate to this folder by typing:

On Windows:

```
cd C:\Users\YourName\Desktop\ChemistryCode
```

On Mac or Linux:

```
cd ~/Desktop/ChemistryCode
```



Replace 'YourName' with your actual Windows username. To find it, type: echo %USERNAME% in Command Prompt.

2.4 Install Required Python Libraries

Libraries are collections of pre-written code that give Python new abilities. Chemistry Code V5.0 uses many specialized libraries. Copy and paste each command below into your terminal and press Enter. Wait for each one to finish before typing the next.

 The flag `--break-system-packages` is required on newer Python versions (3.11+). It simply allows installing packages directly without a virtual environment.

Step 1: Core Scientific Libraries

These are the foundational math and data tools:

```
pip install numpy matplotlib pandas scipy requests tqdm psutil seaborn --break-system-packages
```

Step 2: Chemistry-Specific Libraries

```
pip install rdkit chempy chemlib chemplot --break-system-packages
```

Step 3: Machine Learning Libraries

```
pip install scikit-learn --break-system-packages
```

Step 4: AI / Transformer Libraries

These install the large language model tools that predict reaction products:

```
pip install transformers torch --break-system-packages
```

 PyTorch (torch) is a large download — it may take 5–15 minutes. Do not close the terminal while it is running. You will see a progress bar.

Step 5: Graph Neural Network Library (Optional but Recommended)

This unlocks the GNN features (Options 12 and 13 in the menu):

```
pip install torch-geometric --break-system-packages
```

Step 6: Visualization Library

```
pip install plotly pillow openpyxl --break-system-packages
```

Step 7: Reaction Mapping Library

```
pip install rxnmapper --break-system-packages
```

- ✓ Some libraries like rxnmapper may show warnings or minor errors during install — this is usually fine as long as the install completes. The program has fallback behavior when optional libraries are missing.

2.5 Install Quantum Chemistry Software for DFT Calculations (Optional but Recommended)

xTB is a free quantum chemistry program that powers the DFT calculation features, but is also ORCA and Gaussian compatible. Without it, the program will still run with RDKit but will use less accurate estimates for quantum properties. To install xTB, you need Conda (a separate package manager):

19. Download Miniconda from: <https://docs.conda.io/en/latest/miniconda.html>
20. Run the installer and follow the on-screen instructions
21. After installing, open a new terminal and type:

```
conda install -c conda-forge xtb
```

- i** If xTB is not installed, Chemistry Code will automatically fall back to RDKit-based estimates, which are still scientifically useful but less precise than true quantum calculations.

Section 3: Running the Program

3.1 Starting Chemistry Code V5.0

22. Open your terminal

23. Navigate to your ChemistryCode folder:

```
cd Desktop/ChemistryCode
```

24. Start the program:

```
python Chemistry_Code_V5_0_With_DFT_and_GNN.py
```

The program will initialize and print status messages showing which libraries loaded successfully. You will see green checkmarks () for libraries that loaded and warnings () for any that are missing. This is normal, the program works even if some optional libraries are absent.



The first time you run the program, it may download AI models from the internet. This can take 10–30 minutes. Subsequent runs will be much faster because the models are cached locally.

3.2 The Main Menu

Once the program loads, you will see the main menu. Type the number of the option you want and press Enter.

Option	What It Does
1	Interactive Reaction Analysis — Analyze a single reaction with optional rate graphs
2	Batch Processing — Analyze hundreds of reactions from an Excel spreadsheet at once
3	3D Molecule Viewer — Visualize any molecule in 3D in your web browser
4	EDLC Capacitance Analysis — Score molecules for energy storage potential
5	Generate Comprehensive Report — Export a full summary of your analyses
6	View Performance Metrics — See accuracy statistics and processing speed
7	Reaction Rate Analysis — Calculate rate constants and Arrhenius parameters
8	Help & Documentation — Built-in reference guide
9	Chemical Space Analysis — Map your molecules visually in chemical property space

11	DFT Calculations — Quantum chemistry calculations (HOMO/LUMO energies)
12	GNN Property Prediction — AI-based capacitance and conductivity predictions
13	Train GNN on Your Data — Improve AI predictions using your own experimental results
14	System Status — See which features and libraries are available
15	Exit — Close the program

Section 4: Detailed Guide to Each Menu Option

Option 1: Interactive Reaction Analysis

What It Does

This is the core analysis feature. You enter one reactant (feedstock compound) and one catalyst, and the program uses up to four AI transformer models to predict what products form during the reaction. It then scores those products for capacitance potential and can display a kinetic rate graph.

Step-by-Step Instructions

25. Select option 1 from the main menu
26. You will be asked: Enter reactant (compound name or SMILES):
 - You can type a common chemical name like: syringol or vanillin or furfural
 - Or you can type a SMILES string (see Section 5.1) if the compound is not in the database
27. You will be asked: Enter catalyst (e.g., Ni, KOH, ZnCl2):
 - Common metal catalysts: Ni, Cu, Fe, Co, Pd, Zn
 - Common chemical activators: KOH, ZnCl2, H3PO4
28. You will be asked: Enter condition (e.g., pyrolysis, hydrothermal):
 - Most biochar work uses: pyrolysis
 - Other options: electrochemical, ideal, oxidation
29. You will be asked: Enter temperature (e.g., 600C or 800K):
 - Type a temperature with a unit, like: 600C or 873K
30. The program will then predict reaction products using its AI models and display a consensus result
31. You will be asked: Generate rate graph? (y/n): — type y to see a kinetic plot

Understanding the Output

- Predicted Products: The molecules expected to form
- Confidence Score: How much the AI models agreed (higher = more reliable)
- Capacitance Score (0–100): How well the product might store electrical charge
- Composite Score: An overall quality rating combining multiple factors

 If the compound you enter is not found in the database (e.g., rare alkaloids), the program will ask you to enter the SMILES string manually. See Section 5.1 to learn how to find a SMILES string.

Option 2: Batch Processing

What It Does

Instead of analyzing one reaction at a time, batch processing lets you load an Excel spreadsheet (.xlsx) with hundreds of reactions and analyze them all automatically. Results are saved to a new Excel file.

Preparing Your Excel File

Your Excel file must have columns with these exact names (spelling and capitalization matter):

Column Name	What to Put In It
Reactant	Compound name or SMILES string of your feedstock
Catalyst	Name of the catalyst (e.g., Ni, KOH)
Condition	Reaction condition (e.g., pyrolysis)
Temperature	Temperature with units (e.g., 600C)

32. Place your Excel file in the same folder as the Python script (your ChemistryCode folder)
33. Select option 2 from the main menu
34. When asked for a filename, type just the name without the extension (e.g., type reactions if your file is reactions.xlsx)
35. The program will process each row and display a progress bar
36. When complete, results are saved as: yourfilename_results.xlsx

The batch processor is built to handle 300+ reactions without crashing. Memory is automatically cleaned every 5 reactions. Do not close the terminal while it is running.

Option 3: 3D Molecule Viewer

What It Does

This option generates an interactive 3D visualization of any molecule and opens it in your web browser. You can rotate, zoom, and inspect the molecular structure.

37. Select option 3
38. Enter a compound name (e.g., vanillin) or a SMILES string
39. A file called Molecule_3D.html will be created in your folder
40. It will open automatically in your default web browser
41. In the browser, click and drag to rotate the molecule, scroll to zoom

Option 4: EDLC Capacitance Analysis

What It Does

This option provides detailed analysis of how well molecules and metal catalysts perform as EDLC (supercapacitor) materials. It has three sub-options:

- **Sub-option 1 — View Top Candidates:** Shows the highest-scoring molecules from your previous interactive analyses, ranked by their composite capacitance score.
- **Sub-option 2 — Generate Capacitance Plot:** Creates a chart showing all scored molecules and their relative performance.
- **Sub-option 3 — Electrochemical Metal Analysis:** Analyzes individual metals or metal pairs for their electrochemical properties. You can select metals by symbol (e.g., Ni) or by number from a list. The bimetallic analysis tells you how well two metals work together for enhanced charge storage.

Reading the Electrochemical Output

Term	Plain Language Meaning
Reduction Potential (V)	How strongly a metal attracts electrons. Higher = better electron acceptor.
Work Function (eV)	Energy needed to remove an electron from the metal surface. Relevant for charge transfer.
Electrochemical Score /100	Overall rating combining multiple electrochemical factors. Above 70 = excellent.
EDLC vs Pseudocapacitance	EDLC stores charge physically at the surface; Pseudocapacitance stores it via fast surface reactions.
Galvanic Cell Potential (V)	Voltage that forms when two metals are connected. Higher = more energetically favorable pair.
Synergy Score /100	How well two metals complement each other. Above 25 = strong synergy.

Option 5: Generate Comprehensive Report

This generates a full summary report of all analyses you have done in the current session. The report is saved as an Excel file and includes reaction predictions, scoring data, and statistical summaries. Select this option before closing the program to preserve your work.

Option 6: View Performance Metrics

Shows operational statistics for the current session:

- Total number of reactions processed
- Success rate (how often predictions completed without error)
- Average processing time per reaction
- Cache hit rate (how often a previous result was reused, saving time)

You can optionally generate a performance dashboard chart by typing y when prompted.

Option 7: Reaction Rate Analysis

What It Does

This option provides tools for analyzing how fast reactions occur. It has three sub-options:

- **Sub-option 1 — Cumulative Scatter Plot:** Generates a plot from your most recent batch processing results, showing rate data across all reactions over time (in minutes).
- **Sub-option 2 — Quick Rate Calculator:** Enter a temperature and condition type, and the program will calculate the rate constant (k) for that reaction. It also shows the half-life which is the time it takes for half the starting material to be consumed.
- **Sub-option 3 — Arrhenius Analysis:** Plots rate constants across multiple temperatures to determine activation energy (E_a), the energy barrier a reaction must overcome to proceed.



Rate constant (k) has units of s⁻¹. A larger k means a faster reaction. The half-life formula is: t_{1/2} = ln(2) / k. At 600°C with a Ni catalyst, typical pyrolysis k values are in the range of 10⁻³ to 10⁻¹ s⁻¹.

Option 8: Help & Documentation

This opens the built-in help menu within the program. It lists features added in each version, tips for batch processing, a list of supported feedstock compounds (including rare Ailanthus alkaloids not available in PubChem), and notes on the 4-model consensus system.

Option 9: Chemical Space Analysis (V5.0)

What It Does

Chemical space is a concept where each molecule is placed in a multi-dimensional map based on its properties. Molecules that are structurally similar cluster together. This option lets you visualize where your predicted products fall in chemical space, helping you identify patterns. For example, which structural features tend to produce the best supercapacitor materials.

- **Sub-option 1 — Map Batch Results:** Loads a batch results Excel file and generates a PNG image map of the product chemical space, colored by capacitance score.
- **Sub-option 2 — Map Interactive Results:** Uses reactions you analyzed during the current session with Option 1 to generate a chemical space map on the fly.
- **Sub-option 3 — View Most Recent Map:** Shows the path to the most recently generated chemical space image file.
- **Sub-option 4 — List All Maps:** Lists every chemical space image file in your folder with size and timestamp.



At least 3 molecules are needed to generate a chemical space map. For meaningful analysis, 20+ molecules are recommended. The output is a PNG image file you can open with any photo viewer.

Option 11: DFT Calculations (V5.0)

What It Does

DFT (Density Functional Theory) is a quantum chemistry method that calculates a molecule's electronic structure, essentially, how electrons are distributed and how much energy they have. This is important for predicting how well a material will conduct electricity or participate in redox reactions.

42. Select option 11
43. Enter a SMILES string for the molecule you want to analyze
44. The calculation may take 30 seconds to 5 minutes: do not close the terminal
45. Results will show:

Output Term	What It Means
Total Energy (Hartree)	The molecule's total electronic energy. Used to compare stability between structures.
HOMO Energy (eV)	Highest Occupied Molecular Orbital. Related to how easily the molecule donates electrons.
LUMO Energy (eV)	Lowest Unoccupied Molecular Orbital. Related to how easily the molecule accepts electrons.
HOMO-LUMO Gap (eV)	The energy difference between HOMO and LUMO. Smaller gap = better conductor.
Dipole Moment (Debye)	Measure of how unevenly charge is distributed across the molecule.

i If xTB is not installed, you will see a note saying 'These are RDKit estimates.' These estimates are still useful for relative comparisons between molecules but are less accurate than true quantum calculations.

Option 12: GNN Property Prediction (V5.0)

What It Does

A Graph Neural Network (GNN) treats a molecule as a graph where atoms are nodes and chemical bonds are edges. The AI learns patterns from this graph structure to predict physical

properties. Unlike traditional calculations, the GNN improves over time as you provide more training data.

46. Select option 12
47. Enter a SMILES string for the molecule
48. The GNN will predict:
 - Capacitance (F/g): Predicted specific capacitance of the material
 - Conductivity (S/cm): Predicted electrical conductivity
 - Cycling Stability Score: How well the material retains performance over many charge-discharge cycles

 GNN predictions are most accurate after training on your own experimental data (Option 13). Default predictions use a pre-trained baseline model which gives reasonable estimates but is not customized to your specific materials.

 Requires PyTorch and PyTorch Geometric to be installed. If Option 12 shows an error, run: pip install torch torch-geometric --break-system-packages

Option 13: Train GNN on Your Data (V5.0)

What It Does

This is one of the most powerful features: you can teach the AI using your own experimental measurements. Once trained, the GNN predictions in Option 12 will be specific to your materials and conditions.

Preparing Your Training Data

Run batch processing (Option 2) first to generate a results file. Then optionally add a column called Experimental_Capacitance to that Excel file with your measured capacitance values (in F/g).

Training Steps

49. Select option 13
50. The program will list available results files, select the one containing your data
51. Choose what property to train on: predicted capacitance scores or your experimental measurements

52. Enter the number of training epochs (default: 100). More epochs = more training time but potentially better accuracy
53. The trained model is automatically saved to a folder called trained_gnn_models

 You need at least 5 valid data points (molecules with known values) to train. For a reliable model, 20 or more data points are recommended.

Option 14: System Status & Capabilities

This option shows a complete diagnostic report of your installation:

- Which core libraries are loaded (ChemPy, Transformers, etc.)
- Whether PyTorch Geometric is available and which device it is using (CPU or GPU)
- Whether xTB, ORCA, or Gaussian are detected for DFT calculations
- Status of AI transformer models (loaded or available but not yet loaded)
- Whether you have any previously trained GNN models saved
- A list of installation commands for any missing components

Run this option any time you want to verify your setup is working correctly.

Section 5: Key Concepts and Reference

5.1 What Is a SMILES String?

SMILES stands for Simplified Molecular Input Line Entry System. It is a way of describing a molecule's structure using a single line of text that a computer can read.

Examples:

Molecule	SMILES String
Water (H_2O)	O
Ethanol (drinking alcohol)	CCO
Benzene (aromatic ring)	c1ccccc1
Vanillin (vanilla flavoring)	COc1cc(C=O)ccc1O
Furfural (from hemicellulose)	O=Cc1ccco1

How to find a SMILES string:

- Go to: <https://pubchem.ncbi.nlm.nih.gov>
- Search for your compound by name
- On the compound page, look for 'Canonical SMILES' or 'Isomeric SMILES' in the Chemical and Physical Properties section
- Copy the entire string and paste it into the program when prompted
- If all else fails, typing "Canonical SMILES of target molecule" also works too

5.2 Supported Feedstock Compounds

The program has a built-in database for these compound categories:

- **Ailanthus Alkaloids:** Canthine-6-one, ailanthone — NOTE: these are NOT in PubChem, so manual SMILES entry may be required for some
- **Lignin Phenolics:** Syringol, guaiacol, vanillin, isoeugenol, ferulic acid
- **Carbohydrate Furans:** Furfural, HMF (5-hydroxymethylfurfural), levoglucosan
- **Metal Catalysts:** Ni, Cu, Fe, Co, Pd, Al, Zn, Mn, Mg, Ca, K, Na, and others

If your compound is not found automatically, the program will not crash, it will ask you to enter a SMILES string manually. This is by design.

5.3 The Four AI Models Explained

Chemistry Code uses up to four different AI models and combines their predictions into a consensus result:

Model Name	Description
ReactionT5v2	Specialized transformer model trained on the Open Reaction Database. Achieves ~97.5% accuracy on known reactions.
Molecular Transformer	Trained for regioselectivity prediction — predicts which specific carbon or functional group reacts. ~83% accuracy.
MolT5	General-purpose molecular transformer. Good at handling unusual or novel structures.
RXNMapper	Specializes in atom mapping — tracking which atom in the reactant becomes which atom in the product.

The consensus score reflects how many models agreed on the prediction. Agreement among 3–4 models gives high confidence. Disagreement may suggest a less common reaction pathway.

5.4 Understanding Capacitance Scoring

Each predicted product is scored on a scale of 0–100 for its potential as a supercapacitor electrode material:

Score Range	Interpretation
70–100	Excellent candidate — strong EDLC or pseudocapacitive behavior predicted
50–69	Good candidate — suitable for most supercapacitor applications
30–49	Moderate candidate — may work with further optimization
0–29	Poor candidate — consider alternative synthesis routes or catalysts

The composite score combines structural features (from RDKit molecular descriptors), electrochemical parameters (from ChemLib reduction potential data), and thermodynamic

stability estimates. A higher aromatic content, more nitrogen/oxygen heteroatoms, and a larger accessible surface area generally lead to higher scores.

Section 6: Files and Outputs

Chemistry Code V5.0 creates the following files in your working directory (the ChemistryCode folder):

File Name / Pattern	What It Contains
biocharge_v5_log.txt	A running log of all program activity, warnings, and errors. Useful for debugging.
[yourfile]_results.xlsx	Batch processing results including predicted products, SMILES, scores, and kinetics data for each reaction.
[yourfile]_chemical_space.png	Chemical space map image generated from batch results.
interactive_chemical_space_[timestamp].png	Chemical space map generated from interactive analysis sessions.
Molecule_3D.html	3D molecular viewer file — open in any web browser.
trained_gnn_models/	Folder containing your saved trained GNN model files (.pt format).
dft_scratch/	Temporary folder used by DFT calculations. Can be deleted after analysis.



Your results Excel file is the most important output. Always run Option 5 (Generate Report) before closing the program if you want to keep a full summary of your session.

Section 7: Troubleshooting Common Issues

7.1 Installation Problems

⚠ Problem: 'pip' is not recognized as a command
Solution: You may not have added Python to PATH during installation. Reinstall Python and check the 'Add Python to PATH' checkbox on the first installer screen.

⚠ Problem: 'No module named rdkit' error when running the program
Solution: The rdkit library did not install correctly. Try: pip install rdkit --break-system-packages

⚠ Problem: PyTorch installation fails or takes too long
Solution: Try installing the CPU-only version first: pip install torch --index-url https://download.pytorch.org/whl/cpu --break-system-packages

⚠ Problem: 'Permission denied' error during pip install
Solution: On Mac/Linux, add sudo before the command: sudo pip install [package] --break-system-packages

7.2 Runtime Problems

💡 Problem: Compound not found in database
Solution: This is normal for rare compounds. The program will prompt you to enter a SMILES string manually. Look up the SMILES on PubChem (see Section 5.1).

⚠ Problem: Batch processing stops midway
Solution: Check biocharge_v5_log.txt for the error details. Most commonly caused by malformed SMILES strings in your Excel file. Clean your data and try again.

⚠ Problem: AI model download is very slow or fails
Solution: Models are downloaded from Hugging Face. If download fails, check your internet connection. The program will retry automatically. Downloaded models are cached and do not need to be re-downloaded.

! Problem: GNN or DFT options show error messages Solution: These are optional features requiring additional installations. Run Option 14 (System Status) to see exactly what is missing and follow the printed installation commands.

! Problem: The terminal window closes immediately after running Solution: There may be a Python syntax error. Try running: `python -c 'import sys; print(sys.version)'` to confirm Python works. Then run the program again with: `python Chemistry_Code_V5_0_With_DFT_and_GNN.py` and read any error messages that appear.

Section 8: Recommended Workflows

8.1 Exploring a New Feedstock

Use this workflow when you want to investigate a compound you have never tested before:

54. Use Option 1 (Interactive Analysis) to run 3–5 trial reactions with different catalysts and temperatures
55. Use Option 3 (3D Viewer) to visualize promising products
56. Use Option 4, sub-option 1 to see which reactions produced the highest-scoring candidates
57. Use Option 11 (DFT) to investigate the electronic structure of the top candidates
58. Use Option 5 to generate and save a full report

8.2 High-Throughput Screening (Batch Processing)

Use this workflow when you have a large library of reactions to screen:

59. Prepare your Excel file with all reaction conditions
60. Use Option 2 (Batch Processing) to analyze all reactions
61. Use Option 9 (Chemical Space Analysis) to map and visualize the results
62. Identify clusters of high-scoring molecules in the chemical space map
63. Use Option 12 (GNN Prediction) to further evaluate top cluster members
64. Generate a final report with Option 5

8.3 Building a Custom AI Model

Use this workflow after you have collected experimental capacitance data in the lab:

65. Add your measured capacitance values to the Experimental_Capacitance column in your results Excel file
66. Use Option 13 (Train GNN) and select 'Experimental Capacitance' as the training target
67. Train for 100–500 epochs depending on your dataset size
68. Verify the trained model using Option 14 (System Status), it should show your saved .pt model files
69. From now on, Option 12 predictions will use your custom-trained model

Section 9: Glossary

Quick reference for key technical terms used throughout this manual and in the program output.

Term	Definition
Activation Energy (Ea)	The minimum energy needed for a chemical reaction to occur. Lower Ea = faster reaction.
Arrhenius Equation	A formula relating reaction rate to temperature: $k = A \times e^{(-Ea/RT)}$. Used in Option 7.
Biochar	A carbon-rich solid produced by heating organic material (biomass) without oxygen. Used as a supercapacitor electrode material.
Capacitance (F/g)	The ability of a material to store electrical charge per unit mass. Higher = better energy storage.
DFT (Density Functional Theory)	A quantum chemistry method that calculates electron density and molecular properties.
EDLC (Electric Double-Layer Capacitor)	A type of supercapacitor that stores energy through electrostatic charge separation at electrode surfaces.
Feedstock	The starting organic material fed into a pyrolysis process.
GNN (Graph Neural Network)	An AI model that processes molecular graphs (atoms as nodes, bonds as edges) to predict chemical properties.
HOMO	Highest Occupied Molecular Orbital — the electron energy level most important for electron donation.
LUMO	Lowest Unoccupied Molecular Orbital — the energy level most important for electron acceptance.
HOMO-LUMO Gap	The energy difference between HOMO and LUMO. Smaller gap = better electrical conductor.
Pyrolysis	Thermal decomposition of organic material at high temperature (400–900°C) without oxygen.
Pseudocapacitance	A type of charge storage involving fast, reversible redox reactions at the electrode surface (in contrast to pure EDLC).
Rate Constant (k)	A number describing how fast a reaction proceeds. Units: s^{-1} . Larger k = faster reaction.
Reduction Potential (V vs SHE)	Voltage measure of a metal's tendency to gain electrons. Measured vs Standard Hydrogen Electrode.

SMILES	A text notation for describing molecular structure (e.g., CCO = ethanol). Used to input molecules into chemistry software.
Transformer (AI)	A deep learning architecture used in natural language and chemistry prediction. ReactionT5, MolT5 are examples.
Work Function (eV)	Energy required to remove an electron from a metal surface. Related to charge transfer efficiency.

Quick Start Reference Card

Cut out or print this page and keep it next to your computer.

Getting Started in 4 Steps

1

Install Python
python.org/downloads

2

Run pip install
See Section 2.4

3

Run the script
python script.py

4

Choose a menu option
Type 1–15, press Enter

Essential pip Install Commands

```
pip install numpy matplotlib pandas scipy requests tqdm psutil seaborn --break-system-packages  
pip install rdkit chempy chemlib chemplot --break-system-packages  
pip install scikit-learn transformers torch --break-system-packages  
pip install torch-geometric plotly pillow openpyxl rxnmapper --break-system-packages
```

Starting the Program

```
cd Desktop/ChemistryCode  
python Chemistry_Code_V5_0_With_DFT_and_GNN.py  
or run directly in IDE like VScode or PyCharm
```