WHAT IS CLAIMED IS:

1. A method of discovering a test compound that has a first biological property, the method comprising:

at a computer system comprising at least one processor and a memory storing at least one program for execution by the at least one processor, the at least one program comprising instructions for:

A) obtaining a first training dataset, in electronic form, wherein:

the first training dataset comprises, for each respective compound in a first plurality of compounds, (i) information regarding a chemical structure of the respective compound and (ii) one or more biological properties, in a plurality of biological properties, of the respective compound,

the first plurality of compounds comprises 100 or more compounds, and

the plurality of biological properties includes the first biological property;

B) training an untrained or partially untrained neural network encoder and an untrained or partially untrained classifier by performing a first procedure that comprises:

(i) for each respective compound in the first plurality of compounds, (a) projecting the information regarding the chemical structure of the respective compound into a latent representation space in accordance with a first plurality of weights associated with the untrained or partially untrained neural network encoder to obtain a corresponding projected representation of the respective compound, and (b) inputting the corresponding projected representation of the respective compound into the untrained or partially untrained classifier to obtain a classification of the respective compound in accordance with a second plurality of weights associated with the untrained or partially untrained classifier; and

(ii) updating the first plurality of weights and the second plurality of weights by comparing the classification of each respective compound in the first plurality of compounds to the one or more biological properties of the respective compound in the first training dataset thereby obtaining a trained neural network encoder and a trained classifier;

C) obtaining a second training dataset, in electronic form, wherein the second training dataset comprises, for each respective compound in a second plurality of compounds, information regarding a chemical structure of the respective compound and wherein the second plurality of compounds comprises 100 or more compounds;

D) training an untrained or partially untrained decoder by performing a second procedure that comprises:

(i) for each respective compound in the second plurality of compounds, (a) projecting the information regarding the chemical structure of the respective compound into a latent representation space in accordance with the first plurality of weights associated with the trained neural network encoder to obtain a corresponding projected representation of the respective compound, and (b) inputting the corresponding projected representation of the respective compound into the untrained or partially untrained decoder to obtain a chemical structure of the respective compound in accordance with a third plurality of weights associated with the untrained or partially untrained decoder; and

(ii) updating the third plurality of weights by comparing the chemical structure of each respective compound outputted by the untrained or partially untrained decoder to the actual chemical structure of the respective compound from the second training dataset thereby obtaining a trained decoder; and

E) using the trained neural network encoder, trained classifier, and trained decoder to identify a test compound that has the first biological property, wherein the test compound is not present in the first and second training set.

2. The method of claim 1, wherein the information regarding a chemical structure of the respective compound in the first plurality of compounds is a chemical structure of the respective compound or a high dimensional vector representation based upon a chemical structure of the respective compound.

3. The method of claim 1, wherein the E) using comprises:

interpolating a projected representation of a first compound and a projected representation of a second compound, produced by the trained neural network encoder, wherein the first and second compound have the first molecular property thereby obtaining an interpolated projection;

inputting the interpolated projection into the trained decoder thereby obtaining a plurality of candidate compounds;

for each respective candidate compound in all or a portion of the plurality of candidate compounds:

(i) obtaining a corresponding projected representation for the respective candidate compound by inputting a chemical structure of the candidate compound into the trained neural network encoder; and

(ii) obtaining a classification of the respective candidate compound by inputting the corresponding projected representation of the respective candidate compound into the trained classifier, wherein, when the trained classifier indicates that the corresponding projected representation of the respective candidate compound has the first biological property, the respective candidate compound is deemed to have the first biological property.

4. The method of claim 3, the method further comprising verifying a first compound in the plurality of candidate compounds has the first biological property by a third procedure that comprises:

subjecting the first compound to a wet lab assay that verifies that the respective candidate compound has the first biological property.

5. The method of claim 4, the method further comprising:

synthesizing the first compound.

6. The method of claim 1, the method further comprising verifying the trained neural network encoder, trained classifier, and trained decoder by a third procedure that comprises:

obtaining a first compound, not present in the first or second training dataset, that has the first biological property and has a known chemical structure;

obtaining a projected representation for the first compound by inputting a chemical structure of the first compound into the trained neural network encoder;

inputting the projected representation of the first compound into the trained classifier to verify that the trained classifier identifies the first compound as having the first biological property; and

inputting the projected representation of the first compound into the trained decoder to verify that the trained decoder reconstructs the chemical structure of the first compound.

7. The method of any one of claims 1-6, wherein

(i) the information regarding the chemical structure of the respective compound is a molecular structure of the respective compound,

(ii) the method further comprises:

forming a featurization of the chemical structure; and

incorporating the featurization of the chemical structure into a multi-dimensional vector space, and

(iii) the projecting the information regarding the chemical structure of the respective compound into the latent representation space in accordance with the first plurality of weights associated with the untrained or partially untrained neural network encoder comprises inputting the multi-dimensional vector space of the chemical structure into the untrained or partially untrained neural network encoder.

8. The method of claim 7, wherein the featurization of the chemical structure is a tensor.

9. The method of claim 8, wherein the tensor is a one-dimensional vector or a two-dimensional matrix.

10. The method of claim 7, wherein the featurization of the chemical structure is an extended circular fingerprint, or a molecular graph of a plurality of one-hot-encoded vectors.

11. The method of claim 7, wherein the multi-dimensional vector space is an N-dimensional space, wherein N is an integer between 20 and 80.

12. The method of claim 11, wherein N is 50.

13. The method of claim 7, wherein the incorporating the featurization of the chemical structure into the multi-dimensional vector space for the chemical structure comprises inputting the featurization of the chemical structure into a spatial graph convolutional network (GCN).

14. The method of claim 13, wherein the GCN is a graph attention network (GAT) or a graph substructure index-based approximate graph (SAGA).

15. The method of claim 7, wherein the incorporating the featurization of the molecular structure into the multi-dimensional vector space for the chemical structure comprises an

application of a spectral graph convolution (SGC) to the featurization of the chemical structure.

16. The method of claim 15, wherein the application of the SGC to the featurization of the chemical structure uses Chebyshev polynomial filtering.

17. The method of claim 7, wherein the forming the featurization of the chemical structure comprises:

converting the chemical structure to a simplified molecular-input line-entry system (SMILES) string, and

converting the SMILES string into a molecular graph representation that comprises an adjacency matrix and a feature matrix.

18. The method of any one of claims 1-17, wherein the first biological property is selected from the group consisting of: an indication as to whether a compound activates a cell state, an indication as to whether a compound inhibits a cell state, an affinity for a biological target, an EC50 of the compound for inhibiting a biological state, an IC50 of the compound for inhibiting a biological state, an ED50 of the compound for inhibiting a biological state, an LD50 of the compound for inhibiting a biological state, and a TD50 of the compound for inhibiting a biological state.

19. The method of claim 18, wherein the cell state is characterized by an up-regulation or down-regulation of one or more respective genes in a plurality of genes associated with the cell state.

20. The method of claim 18, wherein the cell state is a diseased state.

21. The method of claim 18, wherein the cell state is characterized by an upregulation or a down-regulation of one or more biological pathways.

22. The method of claim 18, wherein the cell state is characterized by an upregulation or a down-regulation of one or more biological pathways in a plurality of biological pathways.

23. The method of claim 18, wherein the cell state is characterized by an upregulation or a down-regulation of one or more of cellular-components.

24. The method of claim 23, wherein the one or more cellular-components comprises a plurality of genes, optionally measured at the RNA level.

25. The method of claim 23, wherein the one or more cellular-components are quantified using single-cell ribonucleic acid (RNA) sequencing (scRNA-seq), scTag-seq, single-cell assay for transposase-accessible chromatin using sequencing (scATAC-seq), CyTOF/SCoP, E-MS/Abseq, miRNA-seq, CITE-seq, or any combinations thereof, or summaries of the same, including combinations, such as linear combinations, representing activated pathways in the single-cell cellular-component expression datasets.

26. The method of claim 23, wherein the one or more cellular-components comprises a plurality of proteins.

27. A computer system, comprising one or more processors and memory, the memory storing instructions for performing a method for discovering a test compound that has a first biological property, the method comprising:

A) obtaining a first training dataset, in electronic form, wherein:

the first training dataset comprises, for each respective compound in a first plurality of compounds, (i) information regarding a chemical structure of the respective compound and (ii) one or more biological properties, in a plurality of biological properties, of the respective compound,

the first plurality of compounds comprises 100 or more compounds, and

the plurality of biological properties includes the first biological property;

B) training an untrained or partially untrained neural network encoder and an untrained or partially untrained classifier by performing a first procedure that comprises:

(i) for each respective compound in the first plurality of compounds, (a) projecting the information regarding the chemical structure of the respective compound into a latent representation space in accordance with a first plurality of weights associated with the untrained or partially untrained neural network encoder to obtain a corresponding projected representation of the respective compound, and (b) inputting the corresponding projected representation of the respective compound into the untrained or partially untrained classifier

to obtain a classification of the respective compound in accordance with a second plurality of weights associated with the untrained or partially untrained classifier; and

(ii) updating the first plurality of weights and the second plurality of weights by comparing the classification of each respective compound in the first plurality of compounds to the one or more biological properties of the respective compound in the first training dataset thereby obtaining a trained neural network encoder and a trained classifier;

C) obtaining a second training dataset, in electronic form, wherein the second training dataset comprises, for each respective compound in a second plurality of compounds, information regarding a chemical structure of the respective compound and wherein the second plurality of compounds comprises 100 or more compounds;

D) training an untrained or partially untrained decoder by performing a second procedure that comprises:

(i) for each respective compound in the second plurality of compounds, (a) projecting the information regarding the chemical structure of the respective compound into a latent representation space in accordance with the first plurality of weights associated with the trained neural network encoder to obtain a corresponding projected representation of the respective compound, and (b) inputting the corresponding projected representation of the respective compound into the untrained or partially untrained decoder to obtain a chemical structure of the respective compound in accordance with a third plurality of weights associated with the untrained or partially untrained decoder; and

(ii) updating the third plurality of weights by comparing the chemical structure of each respective compound outputted by the untrained or partially untrained decoder to the actual chemical structure of the respective compound from the second training dataset thereby obtaining a trained decoder; and

E) using the trained neural network encoder, trained classifier, and trained decoder to identify a test compound that has the first biological property, wherein the test compound is not present in the first and second training set.

28. A non-transitory computer-readable medium storing one or more computer programs, executable by a computer, for performing a method for discovering a test compound that has a first biological property, the computer comprising one or more processors and a memory, the one or more computer programs collectively encoding computer executable instructions that perform a method comprising:

A) obtaining a first training dataset, in electronic form, wherein:

the first training dataset comprises, for each respective compound in a first plurality of compounds, (i) information regarding a chemical structure of the respective compound and (ii) one or more biological properties, in a plurality of biological properties, of the respective compound,

the first plurality of compounds comprises 100 or more compounds, and

the plurality of biological properties includes the first biological property;

B) training an untrained or partially untrained neural network encoder and an untrained or partially untrained classifier by performing a first procedure that comprises:

(i) for each respective compound in the first plurality of compounds, (a) projecting the information regarding the chemical structure of the respective compound into a latent representation space in accordance with a first plurality of weights associated with the untrained or partially untrained neural network encoder to obtain a corresponding projected representation of the respective compound, and (b) inputting the corresponding projected representation of the respective compound into the untrained or partially untrained classifier to obtain a classification of the respective compound in accordance with a second plurality of weights associated with the untrained or partially untrained classifier; and

(ii) updating the first plurality of weights and the second plurality of weights by comparing the classification of each respective compound in the first plurality of compounds to the one or more biological properties of the respective compound in the first training dataset thereby obtaining a trained neural network encoder and a trained classifier;

C) obtaining a second training dataset, in electronic form, wherein the second training dataset comprises, for each respective compound in a second plurality of compounds, information regarding a chemical structure of the respective compound and wherein the second plurality of compounds comprises 100 or more compounds;

D) training an untrained or partially untrained decoder by performing a second procedure that comprises:

(i) for each respective compound in the second plurality of compounds, (a) projecting the information regarding the chemical structure of the respective compound into a latent representation space in accordance with the first plurality of weights associated with the trained neural network encoder to obtain a corresponding projected representation of the respective compound, and (b) inputting the corresponding projected representation of the respective compound into the untrained or partially untrained decoder to obtain a chemical structure of the respective compound in accordance with a third plurality of weights associated with the untrained or partially untrained decoder; and

(ii) updating the third plurality of weights by comparing the chemical structure of each respective compound outputted by the untrained or partially untrained decoder to the actual chemical structure of the respective compound from the second training dataset thereby obtaining a trained decoder; and

E) using the trained neural network encoder, trained classifier, and trained decoder to identify a test compound that has the first biological property, wherein the test compound is not present in the first and second training set.

29. A method of discovering a candidate compound that has a first biological property, the method comprising:

at a computer system comprising at least one processor and a memory storing at least one program for execution by the at least one processor, the at least one program comprising instructions for:

obtaining a first projected representation of a first compound that is assigned the first biological property by inputting a chemical structure of the first compound into a trained neural network encoder, wherein the first projected representation has N dimensions, wherein N is an integer between 20 and 80;

using the first projection to obtain one or more candidate projections;

inputting each candidate projection in the one or more candidate projections into a trained decoder thereby obtaining a plurality of candidate compounds, wherein the first compound is not present in the plurality of candidate compounds;

for each respective candidate compound in the plurality of candidate compounds:

(i) obtaining a corresponding projected representation for the respective candidate compound by inputting a chemical structure of the candidate compound into the trained neural network encoder, wherein the corresponding projected representation has N dimensions; and

(ii) obtaining a classification of the respective candidate compound by inputting the corresponding projected representation of the respective candidate compound into the trained classifier, wherein, when the trained classifier indicates that the corresponding projected representation of the respective candidate compound has the first biological property, the respective candidate compound is deemed to have the first biological property.

30. The method of claim 29, the method further comprising:

obtaining a second projected representation of a second compound that has the biological property by inputting a chemical structure of the second compound into the trained neural network encoder, and wherein

the using the first projection to obtain one or more candidate projections comprises interpolating the first projection and the second projection thereby obtaining the one or more candidate projections.

31. A computer system, comprising one or more processors and memory, the memory storing instructions for performing a method of discovering a candidate compound that has a first biological property, the method comprising:

obtaining a first projected representation of a first compound that is assigned the first biological property by inputting a chemical structure of the first compound into a trained neural network encoder, wherein the first projected representation has N dimensions, wherein N is an integer between 20 and 80;

using the first projection to obtain one or more candidate projections;

inputting each candidate projection in the one or more candidate projections into the trained decoder thereby obtaining a plurality of candidate compounds, wherein the first compound is not present in the plurality of candidate compounds;

for each respective candidate compound in the plurality of candidate compounds:

(i) obtaining a corresponding projected representation for the respective candidate compound by inputting a chemical structure of the candidate compound into the trained neural network encoder, wherein the corresponding projected representation has N dimensions; and

(ii) obtaining a classification of the respective candidate compound by inputting the corresponding projected representation of the respective candidate compound into the trained classifier, wherein, when the trained classifier indicates that the corresponding projected representation of the respective candidate compound has the first biological property, the respective candidate compound is deemed to have the first biological property.

32. A non-transitory computer-readable medium storing one or more computer programs, executable by a computer, for performing a method of discovering a candidate compound that has a first biological property, the computer comprising one or more processors and a

memory, the one or more computer programs collectively encoding computer executable instructions that perform a method comprising:

obtaining a first projected representation of a first compound that is assigned the first biological property by inputting a chemical structure of the first compound into a trained neural network encoder, wherein the first projected representation has N dimensions, wherein N is an integer between 20 and 80;

using the first projection to obtain one or more candidate projections;

inputting each candidate projection in the one or more candidate projections into the trained decoder thereby obtaining a plurality of candidate compounds, wherein the first compound is not present in the plurality of candidate compounds;

for each respective candidate compound in the plurality of candidate compounds:

(i) obtaining a corresponding projected representation for the respective candidate compound by inputting a chemical structure of the candidate compound into the trained neural network encoder, wherein the corresponding projected representation has N dimensions; and

(ii) obtaining a classification of the respective candidate compound by inputting the corresponding projected representation of the respective candidate compound into the trained classifier, wherein, when the trained classifier indicates that the corresponding projected representation of the respective candidate compound has the first biological property, the respective candidate compound is deemed to have the first biological property.

33. The method of claim 29, wherein the first biological property is a compound function.

34. The method of claim 29, further comprising:

subjecting the respective candidate compound to a wet lab assay that verifies that the respective candidate compound has the first biological property.

35. The method of claim 34, the method further comprising synthesizing the respective candidate compound.