What is claimed is:

1. A method for reducing a number of chemical compounds in a plurality of chemical compounds in a chemical compound dataset, the method comprising:

A) obtaining, in electronic format, the chemical compound dataset;

B) applying a target model, for each respective chemical compound in a subset of chemical compounds from the plurality of chemical compounds, to the respective chemical compound and a polymer to obtain a corresponding target result, thereby obtaining a corresponding subset of target results;

C) training a predictive model in an initial trained state using at least i) the subset of chemical compounds as independent variables and ii) the corresponding subset of target results as dependent variables, thereby updating the predictive model to an updated trained state;

D) applying the predictive model in an updated trained state to the plurality of chemical compounds thereby obtaining an instance of a plurality of predictive results wherein the instance of the plurality of predictive results includes a respective predictive result for each chemical compound in the plurality of chemical compounds;

E) eliminating a portion of the chemical compounds from the plurality of chemical compounds based at least in part on the instance of the plurality of predictive results; and

F) determining whether one or more predefined reduction criteria are satisfied, wherein, when the one or more predefined reduction criteria are not satisfied, the method further comprises:

(i) applying the target model, for each respective chemical compound in an additional subset of chemical compounds from the plurality of chemical compounds, to the respective chemical compound and the polymer to obtain a corresponding target result, thereby obtaining an additional subset of target results, wherein the additional subset of chemical compounds is selected at least in part on the instance of the plurality of predictive results;

(ii) updating the subset of chemical compounds by incorporating the additional subset of chemical compounds into the subset of chemical compounds;

(iii) updating the subset of target results by incorporating the additional subset of target results into the subset of target results;

(iv) modifying, after the updating (ii) and the updating (iii), the predictive model by applying the predictive model to at least 1) the subset of chemical compounds as a plurality of independent variables of the predictive model and 2) the corresponding subset of target results as a corresponding plurality of dependent variables of the predictive model, thereby providing the predictive model in an updated trained state; and

(v) repeating the applying (D), eliminating (E), and determining (F), wherein the plurality of chemical compounds, before application of an instance of the eliminating E), comprises at least 100 million chemical compounds.

2. The method of claim 1, wherein

the target model exhibits a first computational complexity,

the predictive model exhibits a second computational complexity, and

the second computational complexity is less than the first computational complexity.

3. The method of either claim 1 or claim 2, wherein the chemical compound dataset includes a plurality of feature vectors, wherein each feature vector is for a respective chemical compound in the plurality of chemical compounds.

4. The method of any one of claims 1-3, wherein the applying B) further comprises randomly selecting one or more chemical compounds from the plurality of chemical compounds to form the subset of chemical compounds.

5. The method of claim 3, wherein the applying B) further comprises selecting one or more chemical compounds from the plurality of chemical compounds for the subset of chemical compounds based on evaluation of one or more features selected from the plurality of feature vectors.

6. The method of claim 3, wherein each feature vector in the plurality of feature vectors is a one-dimensional vector.

7. The method of either claim 3 or claim 4, wherein the applying F)(i) further comprises forming the additional subset of chemical compounds by selecting one or more chemical compounds from the plurality of chemical compounds based on evaluation of one or more features selected from the plurality of feature vectors.

8. The method of any one of claims 1-7, wherein satisfaction of the one or more predefined reduction criteria comprises (i) comparing each predictive result in the plurality of predictive results to a corresponding target result from the subset of target results or

(ii) determining that the number of chemical compounds in the plurality of chemical compounds has dropped below a threshold number of objects.

9. The method of any one of claims 1-0, wherein the target model is a convolutional neural network and

the predictive model comprises a random forest tree, a random forest comprising a plurality of multiple additive decision trees, a neural network, a graph neural network, a dense neural network, a principal component analysis, a nearest neighbor analysis, a linear discriminant analysis, a quadratic discriminant analysis, a support vector machine, an evolutionary method, a projection pursuit, a linear regression, a Naïve Bayes algorithm, a multi-category logistic regression algorithm, or ensembles thereof.

10. The method of any one of claims 1-9, wherein the polymer is a protein, a polypeptide, a polynucleic acid, a polyribonucleic acid, a polysaccharide, or an assembly of any combination thereof.

11. The method of any one of claims 1-9, wherein the target model is applied to the polymer based upon a set of three-dimensional coordinates $\{x_1, \ldots, x_N\}$ for a crystal structure of the polymer resolved at a resolution of 2.5 Å or better Or

a crystal structure of the polymer resolved at a resolution of 3.3 Å or better.

12. The method of any one of claims 1-9, wherein the target model is applied to the polymer based upon spatial coordinates that are an ensemble of three-dimensional coordinates for the polymer determined by nuclear magnetic resonance, neutron diffraction, or cryo-electron microscopy.

13. The method of any one of claims 1-12, wherein the plurality of chemical compounds, before application of an instance of the eliminating E), comprises at least 500 million chemical compounds, at least 1 billion chemical compounds, at least 2 billion chemical compounds, at

least 3 billion chemical compounds, at least 4 billion chemical compounds, at least 5 billion chemical compounds, at least 6 billion chemical compounds, at least 7 billion chemical compounds, at least 8 billion chemical compounds, at least 9 billion chemical compounds, at least 10 billion chemical compounds, at least 11 billion chemical compounds, at least 15 billion chemical compounds, at least 20 billion chemical compounds, at least 30 billion chemical compounds, at least 40 billion chemical compounds, at least 50 billion chemical compounds, at least 60 billion chemical compounds, at least 70 billion chemical compounds, at least 80 billion chemical compounds, at least 90 billion chemical compounds, at least 100 billion chemical compounds, or at least 110 billion chemical compounds.

14. The method of any one of claims 1-13, wherein the subset of chemical compounds comprises at least 1,000 chemical compounds, at least 5,000 chemical compounds, at least 10,000 chemical compounds, at least 25,000 chemical compounds, at least 50,000 chemical compounds, at least 75,000 chemical compounds, at least 100,000 chemical compounds, at least 250,000 chemical compounds, at least 500,000 chemical compounds, at least 750,000 chemical compounds, at least 1 million chemical compounds, at least 2 million chemical compounds, at least 3 million chemical compounds, at least 4 million chemical compounds, at least 5 million chemical compounds, at least 6 million chemical compounds, at least 7 million chemical compounds, at least 8 million chemical compounds, at least 9 million chemical compounds, or at least 10 million chemical compounds and

the additional subset of chemical compounds comprises at least 1,000 chemical compounds, at least 5,000 chemical compounds, at least 10,000 chemical compounds, at least 25,000 chemical compounds, at least 50,000 chemical compounds, at least 75,000 chemical compounds, at least 100,000 chemical compounds, at least 250,000 chemical compounds, at least 500,000 chemical compounds, at least 750,000 chemical compounds, at least 1 million chemical compounds, at least 2 million chemical compounds, at least 3 million chemical compounds, at least 4 million chemical compounds, at least 5 million chemical compounds, at least 6 million chemical compounds, at least 7 million chemical compounds, at least 8 million chemical compounds, at least 9 million chemical compounds, or at least 10 million chemical compounds and

wherein the additional subset of chemical compounds is distinct from the subset of chemical compounds.

15. The method of claim 1, wherein the F) modifying (iv) the predictive model comprises retraining the predictive model.

16. The method of claim 1, wherein the training (C) further comprises using iii) the polymer as an independent variable of the predictive model, in addition to using the at least i) the subset of chemical compounds as a plurality of independent variables of the predictive model, and ii) the corresponding subset of target results as a plurality of dependent variables of the predictive model.

17. The method of any one of claims 1-16, wherein the modifying F)(iv) further comprises using 3) the polymer as an independent variable, in addition to using at least 1) the subset of chemical compounds as independent variables and 2) the corresponding subset of target results as corresponding dependent variables of the predictive model.

18. The method of any one of claims 1-17, wherein, when the one or more predefined reduction criteria are satisfied, the method further comprises:

i) clustering the plurality of chemical compounds, thereby assigning each chemical compound in the plurality of chemical compounds to a cluster in a plurality of clusters; and

ii) eliminating one or more chemical compounds from the plurality of chemical compounds based at least in part on redundancy of chemical compounds in individual clusters in the plurality of clusters.

19. The method of any one of claims 1-17, the method further comprising selecting the subset of chemical compounds from the plurality of chemical compounds by:

i) clustering the plurality of chemical compounds thereby assigning each chemical compound in the plurality of chemical compounds to a respective cluster in a plurality of clusters, and

ii) selecting the subset of chemical compounds from the plurality of chemical compounds based at least in part on a redundancy of chemical compounds in individual clusters in the plurality of clusters.

20. The method of any one of claims 1-19, wherein, when the one or more predefined reduction criterion are satisfied, the method further comprises applying the predictive model to the plurality of chemical compounds and the polymer, thereby causing the predictive model to provide a respective interaction score for each chemical compound in the plurality of chemical compounds

, wherein each respective interaction score corresponds to an interaction between a respective chemical compound and the polymer

, and wherein each respective interaction score is used to characterize the polymer.

21. The method of claim 1, wherein the eliminating (E) comprises:

i) clustering the plurality of chemical compounds, thereby assigning each chemical compound in the plurality of chemical compounds to a respective cluster in a plurality of clusters, and

ii) eliminating a subset of chemical compounds from the plurality of chemical compounds based at least in part on a redundancy of chemical compounds in individual clusters in the plurality of clusters,

wherein clustering the plurality of chemical compounds is performed using a density-based spatial clustering algorithm, a divisive clustering algorithm, an agglomerative clustering algorithm, a k-means clustering algorithm, a supervised clustering algorithm, or ensembles thereof.

22. The method of claim 1, wherein the eliminating (E) comprises:

ranking the plurality of chemical compounds based on the instance of the plurality of predictive results, and

removing from the plurality of chemical compounds those chemical compounds in the plurality of chemical compounds that fail to have a corresponding predictive result that satisfies a threshold cutoff

, wherein the threshold cutoff is a top threshold percentage

, and wherein the top threshold percentage is the top 90 percent, the top 80 percent, the top 75 percent, the top 60 percent, or the top 50 percent of the plurality of predictive results.

23. The method of any one of claims 1-22, wherein each instance of the eliminating (E) eliminates between one tenth and nine tenths of the chemical compounds in the plurality of chemical compounds

or between one quarter and three quarters of the chemical compounds in the plurality of chemical compounds.

24. A computer system for reducing a number of chemical compounds in a plurality of chemical compounds in a chemical compound dataset, the computer system comprising:

one or more processors;

a memory; and

one or more programs, wherein the one or more programs are stored in the memory and are configured to be executed by the one or more processors, the one or more programs including instructions for:

A) obtaining, in electronic format, the chemical compound dataset;

B) applying a target model, for each respective chemical compound in a subset of chemical compounds from the plurality of chemical compounds, to the respective chemical compound and a polymer to obtain a corresponding target result, thereby obtaining a corresponding subset of target results;

C) training a predictive model in an initial trained state using at least i) the subset of chemical compounds as independent variables and ii) the corresponding subset of target results as dependent variables, thereby updating the predictive model to an updated trained state;

D) applying the predictive model in an updated trained state to the plurality of chemical compounds thereby obtaining an instance of a plurality of predictive results;

E) eliminating a portion of the chemical compounds from the plurality of chemical compounds based at least in part on the instance of the plurality of predictive results; and

F) determining whether one or more predefined reduction criteria are satisfied, wherein, when the one or more predefined reduction criteria are not satisfied, the instructions further comprise:

(i) applying the target model, for each respective chemical compound in an additional subset of chemical compounds from the plurality of chemical compounds, to the respective chemical compound and the polymer to obtain a corresponding target result, thereby obtaining an additional subset of target results, wherein the additional subset of chemical compounds is selected at least in part on the instance of the plurality of predictive results;

(ii) updating the subset of chemical compounds by incorporating the additional subset of chemical compounds into the subset of chemical compounds;

(iii) updating the subset of target results by incorporating the additional subset of target results into the subset of target results;

(iv) modifying, after the updating (ii) and the updating (iii), the predictive model by applying the predictive model to at least 1) the subset of chemical compounds as a plurality of independent variables of the predictive model and 2) the corresponding subset of target results as a corresponding plurality of dependent variables of the predictive model, thereby providing the predictive model in an updated trained state; and

(v) repeating the applying (D), eliminating (E), and determining (F), wherein the plurality of chemical compounds, before application of an instance of the eliminating E), comprises at least 100 million chemical compounds.

25. A non-transitory computer readable storage medium and one or more computer programs embedded therein, the one or more computer programs comprising instructions which, when executed by a computer system, cause the computer system to perform a method for reducing a number of chemical compounds in a plurality of chemical compounds in a chemical compound dataset comprising:

A) obtaining, in electronic format, the chemical compound dataset;

B) applying a target model, for each respective chemical compound in a subset of chemical compounds from the plurality of chemical compounds, to the respective chemical compound and a polymer to obtain a corresponding target result, thereby obtaining a corresponding subset of target results;

C) training a predictive model in an initial trained state using at least i) the subset of chemical compounds as independent variables and ii) the corresponding subset of target results as dependent variables, thereby updating the predictive model to an updated trained state;

D) applying the predictive model in an updated trained state to the plurality of chemical compounds thereby obtaining an instance of a plurality of predictive results;

E) eliminating a portion of the chemical compounds from the plurality of chemical compounds based at least in part on the instance of the plurality of predictive results; and

F) determining whether one or more predefined reduction criteria are satisfied, wherein, when the one or more predefined reduction criteria are not satisfied, the instructions further comprise:

(i) applying the target model, for each respective chemical compound in an additional subset of chemical compounds from the plurality of chemical compounds, to the respective chemical compound and the polymer to obtain a corresponding target result, thereby obtaining an additional subset of target results, wherein the additional subset of chemical compounds is selected at least in part on the instance of the plurality of predictive results;

(ii) updating the subset of chemical compounds by incorporating the additional subset of chemical compounds into the subset of chemical compounds;

(iii) updating the subset of target results by incorporating the additional subset of target results into the subset of target results;

(iv) modifying, after the updating (ii) and the updating (iii), the predictive model by applying the predictive model to at least 1) the subset of chemical compounds as a plurality of independent variables of the predictive model and 2) the corresponding subset of target results as a corresponding plurality of dependent variables of the predictive model, thereby providing the predictive model in an updated trained state; and

(v) repeating the applying (D), eliminating (E), and determining (F), wherein the plurality of chemical compounds, before application of an instance of the eliminating E), comprises at least 100 million chemical compounds.