What is claimed is:

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

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

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

(ii) updating the subset of ~~test object~~chemical compounds by incorporating the additional subset of ~~test object~~chemical compounds into the subset of ~~test object~~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 ~~test object~~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 ~~test object~~chemical compounds, before application of an instance of the eliminating E), comprises at least 100 million ~~test object~~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 ~~test object~~chemical compound dataset includes a plurality of feature vectors, wherein each feature vector is for a respective ~~test object~~chemical compound in the plurality of ~~test object~~chemical compounds.

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

5. The method of claim 3, wherein the applying B) further comprises selecting one or more ~~test object~~chemical compounds from the plurality of ~~test object~~chemical compounds for the subset of ~~test object~~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 ~~test object~~chemical compounds by selecting one or more ~~test~~

~~object~~chemical compounds from the plurality of ~~test object~~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

~~9. The method of any one of claims 1-7, wherein satisfaction of the one or more predefined reduction criteria comprises~~ (ii) determining that the number of ~~test object~~chemical compounds in the plurality of ~~test object~~chemical compounds has dropped below a threshold number of objects.

~~10.~~9. The method of any one of claims 1-~~0~~9, wherein the target model is a convolutional neural network~~.~~ and

~~11. The method of any one of claims 1-9, wherein~~ 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.

~~12. The method of any one of claims 1-11, wherein~~
~~the at least one target object is a single object, and~~
~~the single object is a polymer.~~

~~13. The method of claim 12 wherein the polymer comprises an active site.~~

~~14.~~10. The method of ~~claim 12 or 13~~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.

~~15.~~11. The method of ~~claim 12~~any one of claims 1-9, wherein ~~the polymer is applied to~~ 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

Formatted: No bullets or numbering

Formatted: No bullets or numbering

~~16. The method of claim 12, wherein the polymer is applied to the target model 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 3.3 Å or better.

Formatted: No bullets or numbering

~~17.~~12. The method of ~~claim 12~~any one of claims 1-9, wherein ~~the polymer is applied to~~ 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.

~~18.~~13. The method of any one of claims ~~1-19~~1-12, wherein the plurality of ~~test object~~chemical compounds, before application of an instance of the eliminating E), comprises at least 500 million ~~test object~~chemical compounds, at least 1 billion ~~test object~~chemical compounds, at least 2 billion ~~test object~~chemical compounds, at least 3 billion ~~test object~~chemical compounds, at least 4 billion ~~test object~~chemical compounds, at least 5 billion ~~test object~~chemical compounds, at least 6 billion ~~test object~~chemical compounds, at least 7 billion ~~test object~~chemical compounds, at least 8 billion ~~test object~~chemical compounds, at least 9 billion ~~test object~~chemical compounds, at least 10 billion ~~test object~~chemical compounds, at least 11 billion ~~test object~~chemical compounds, at least 15 billion ~~test object~~chemical compounds, at least 20 billion ~~test object~~chemical compounds, at least 30 billion ~~test object~~chemical compounds, at least 40 billion ~~test object~~chemical compounds, at least 50 billion ~~test object~~chemical compounds, at least 60 billion ~~test object~~chemical compounds, at least 70 billion ~~test object~~chemical compounds, at least 80 billion ~~test object~~chemical compounds, at least 90 billion ~~test object~~chemical compounds, at least 100 billion ~~test object~~chemical compounds, or at least 110 billion ~~test object~~chemical compounds.

~~19. The method of claim 0, wherein the one or more predefined reduction criteria require the plurality of test objects to have no more than 30 test objects, no more than 40 test objects, no more than 50 test objects, no more than 60 test objects, no more than 70 test objects, no more than 90 test objects, no more than 100 test objects, no more than 200 test objects, no more than 300 test objects, no more than 400 test objects, no more than 500 test objects, no more than 600~~

~~test objects, no more than 700 test objects, no more than 800 test objects, no more than 900 test objects, or no more than 1000 test objects.~~

~~20. The method of any one of claims 1-19, wherein each test object in the plurality of test objects represents a chemical compound.~~

~~21. The method of any one of claims 1-20, wherein the predictive model in the initial trained state comprises an untrained or partially trained classifier.~~

~~22. The method of any one of claims 1-21, wherein the predictive model in the updated trained state comprises an untrained or a partially trained classifier that is distinct from the predictive model in the initial trained state.~~

~~23.~~14. The method of any one of claims ~~1-22~~1-13, wherein the subset of ~~test object~~chemical compounds comprises at least 1,000 ~~test object~~chemical compounds, at least 5,000 ~~test object~~chemical compounds, at least 10,000 ~~test object~~chemical compounds, at least 25,000 ~~test object~~chemical compounds, at least 50,000 ~~test object~~chemical compounds, at least 75,000 ~~test object~~chemical compounds, at least 100,000 ~~test object~~chemical compounds, at least 250,000 ~~test object~~chemical compounds, at least 500,000 ~~test object~~chemical compounds, at least 750,000 ~~test object~~chemical compounds, at least 1 million ~~test object~~chemical compounds, at least 2 million ~~test object~~chemical compounds, at least 3 million ~~test object~~chemical compounds, at least 4 million ~~test object~~chemical compounds, at least 5 million ~~test object~~chemical compounds, at least 6 million ~~test object~~chemical compounds, at least 7 million ~~test object~~chemical compounds, at least 8 million ~~test object~~chemical compounds, at least 9 million ~~test object~~chemical compounds, or at least 10 million ~~test object~~chemical compounds~~.~~ and

~~24. The method of any one of claims 1-23, wherein~~ the additional subset of ~~test object~~chemical compounds comprises at least 1,000 ~~test object~~chemical compounds, at least 5,000 ~~test object~~chemical compounds, at least 10,000 ~~test object~~chemical compounds, at least 25,000 ~~test object~~chemical compounds, at least 50,000 ~~test object~~chemical compounds, at least 75,000 ~~test object~~chemical compounds, at least 100,000 ~~test object~~chemical compounds, at least 250,000 ~~test object~~chemical compounds, at least 500,000 ~~test object~~chemical compounds, at least 750,000 ~~test object~~chemical compounds, at least 1 million ~~test object~~chemical compounds, at

Formatted: No bullets or numbering

least 2 million ~~test object~~chemical compounds, at least 3 million ~~test object~~chemical compounds, at least 4 million ~~test object~~chemical compounds, at least 5 million ~~test object~~chemical compounds, at least 6 million ~~test object~~chemical compounds, at least 7 million ~~test object~~chemical compounds, at least 8 million ~~test object~~chemical compounds, at least 9 million ~~test object~~chemical compounds, or at least 10 million ~~test object~~chemical compounds~~.~~ and

~~25. The method of either claim 23 or 24,~~ wherein the additional subset of ~~test object~~chemical compounds is distinct from the subset of ~~test object~~chemical compounds.

~~26.~~15. The method of claim 1, wherein the F) modifying (iv) the predictive model comprises retraining the predictive model.

~~27.~~16. The method of claim 1, wherein the training (C) further comprises using iii) the ~~at least one target object~~ polymer as an independent variable of the predictive model, in addition to using the at least i) the subset of ~~test object~~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.

~~28. The method of either claim 1 or claim 27, wherein the at least one target object comprises at least two target objects, at least three target objects, at least four target objects, at least five target objects, or at least six target objects.~~

~~29. The method of claim 1, wherein the instance of the plurality of predictive results includes a respective predictive result for each test object in the plurality of test objects.~~

~~30.~~17. The method of any one of claims ~~1-29~~1-16, wherein the modifying F)(iv) further comprises using 3) the ~~at least one target object~~polymer as an independent variable, in addition to using at least 1) the subset of ~~test object~~chemical compounds as independent variables and 2) the corresponding subset of target results as corresponding dependent variables of the predictive model.

~~31.~~18. The method of any one of claims ~~1-30~~1-17, wherein, when the one or more predefined reduction criteria are satisfied, the method further comprises:

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

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

~~32.~~19. The method of any one of claims ~~1-30~~1-17, the method further comprising selecting the subset of ~~test object~~chemical compounds from the plurality of ~~test object~~chemical compounds by:

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

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

~~33.~~20. The method of any one of claims ~~1-32~~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 ~~test object~~chemical compounds and the ~~at least one target object~~polymer, thereby causing the predictive model to provide a respective interaction score for each ~~test object~~chemical compound in the plurality of ~~test object~~chemical compounds~~.~~

~~34. The method of claim 33~~, wherein each respective interaction score corresponds to an interaction between a respective ~~test object~~chemical compound and the ~~at least one target object~~polymer~~.~~

Formatted: No bullets or numbering

~~35. The method of either claim 33 or 34~~, and wherein each respective interaction score is used to characterize the ~~at least one target object~~polymer.

~~36.~~21. The method of claim 1, wherein the eliminating (E) comprises:

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

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

~~37. The method of any one of claims 31, 32, or 36~~ wherein clustering the plurality of ~~test object~~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.

~~38.~~22. The method of claim 1, wherein the eliminating (E) comprises:

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

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

~~39. The method of claim 38~~, wherein the threshold cutoff is a top threshold percentage~~.~~

~~40. The method of claim 39~~, 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.

~~41.~~23. The method of any one of claims ~~1-40~~1-22, wherein each instance of the eliminating (E) eliminates between one tenth and nine tenths of the ~~test object~~chemical compounds in the plurality of ~~test object~~chemical compounds~~.~~

~~42. The method of any one of claims 1-40, wherein each instance of the eliminating (E) eliminates~~ or between one quarter and three quarters of the ~~test object~~chemical compounds in the plurality of ~~test object~~chemical compounds.

Formatted: No bullets or numbering

Formatted: No bullets or numbering

Formatted: No bullets or numbering

43. The method of any one of claims 1-42, wherein the at least one target object is a single target object and the applying, for each respective test object in a subset of test objects from the plurality of test objects, to the respective test object and target object to obtain a corresponding target result B) comprises:

i) obtaining spatial coordinates for the target object;

ii) modeling the respective test object with the target object in each pose of a plurality of different poses, thereby creating a plurality of voxel maps, wherein each respective voxel map in the plurality of voxel maps comprises the test object in a respective pose in the plurality of different poses;

iii) unfolding each voxel map in the plurality of voxel maps into a corresponding vector, thereby creating a plurality of vectors, wherein each vector in the plurality of vectors is the same size;

iv) inputting each respective vector in the plurality of vectors to the target model, wherein the target model includes (a) an input layer for sequentially receiving the plurality of vectors, (b) a plurality of convolutional layers, and (c) a scorer, wherein

the plurality of convolutional layers includes an initial convolutional layer and a final convolutional layer,

each layer in the plurality of convolutional layers is associated with a different set of weights,

responsive to input of a respective vector in the plurality of vectors, the input layer feeds a first plurality of values into the initial convolutional layer as a first function of values in the respective vector,

each respective convolutional layer, other than the final convolutional layer, feeds intermediate values, as a respective second function of (a) the different set of weights associated with the respective convolutional layer and (b) input values received by the respective convolutional layer, into another convolutional layer in the plurality of convolutional layers, and

the final convolutional layer feeds final values, as a third function of (a) the different set of weights associated with the final convolutional layer and (b) input values received by the final convolutional layer, into the scorer;

v) obtaining a corresponding plurality of scores from the scorer, wherein each score in the corresponding plurality of scores corresponds to the input of a vector in the plurality of vectors into the input layer; and

vi) using the plurality of scores to compute the corresponding target result.

44. The method of claim 43, wherein the scorer comprises a plurality of fully-connected layers and an evaluation layer, and wherein a fully-connected layer in the plurality of fully-connected layers feeds into the evaluation layer.

45. The method of claim 43, wherein the scorer comprises a decision tree, a multiple additive regression tree, a clustering algorithm, 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, and ensembles thereof.

46. The method of claim 43, wherein each vector in the plurality of vectors is a one-dimensional vector.

47. The method of claim 43, wherein the plurality of different poses comprises 2 or more poses, 10 or more poses, 100 or more poses, or 1000 or more poses.

48. The method of claim 43, wherein the plurality of different poses is obtained using a docking scoring function in one of markup chain Monte Carlo sampling, simulated annealing, Lamarckian Genetic Algorithms, or genetic algorithms.

49. The method of claim 43, wherein the plurality of different poses is obtained by incremental search using a greedy algorithm.

50. The method of claim 43, wherein the using the plurality of scores to compute the corresponding target result comprises taking a measure of central tendency of the plurality of scores.

51. The method of claim 43, wherein the using the plurality of scores to compute the corresponding target result comprises using the plurality of scores to characterize the respective test object comprises taking a weighted average of the plurality of scores.

~~52. The method of claim 43, wherein a respective convolutional layer in the plurality of convolutional layers has a plurality of filters and wherein each filter in the plurality of filters convolves a cubic input space of N^3 with stride Y, wherein N is an integer of two or greater and Y is a positive integer.~~

~~53. The method of claim 52, wherein the different set of weights associated with the respective convolutional layer are associated with respective filters in the plurality of filters.~~

~~54. The method of claim 43, wherein the scorer comprises a plurality of fully-connected layers and a logistic regression cost layer and wherein a fully-connected layer in the plurality of fully-connected layers feeds into the logistic regression cost layer.~~

~~55.~~24. A computer system for reducing a number of ~~test object~~chemical compounds in a plurality of ~~test object~~chemical compounds in a ~~test object~~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 ~~test object~~chemical compound dataset;

B) applying a target model, for each respective ~~test object~~chemical compound in a subset of ~~test object~~chemical compounds from the plurality of ~~test object~~chemical compounds, to the respective ~~test object~~chemical compound and ~~at least one target object~~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 ~~test object~~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 ~~test object~~chemical compounds thereby obtaining an instance of a plurality of predictive results;

E) eliminating a portion of the ~~test object~~chemical compounds from the plurality of ~~test object~~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~~ instructions further ~~comprises~~comprise:

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

(ii) updating the subset of ~~test object~~chemical compounds by incorporating the additional subset of ~~test object~~chemical compounds into the subset of ~~test object~~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 ~~test object~~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 ~~test object~~chemical compounds, before application of an instance of the eliminating E), comprises at least 100 million ~~test object~~chemical compounds.

~~56.~~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 ~~test object~~chemical compounds in a plurality of ~~test object~~chemical compounds in a ~~test object~~chemical compound dataset comprising:

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

B) applying a target model, for each respective ~~test object~~chemical compound in a subset of ~~test object~~chemical compounds from the plurality of ~~test object~~chemical compounds, to the respective ~~test object~~chemical compound and ~~at least one target object~~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 ~~test object~~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 ~~test object~~chemical compounds thereby obtaining an instance of a plurality of predictive results;

E) eliminating a portion of the ~~test object~~chemical compounds from the plurality of ~~test object~~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~~ instructions further ~~comprises~~comprise:

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

(ii) updating the subset of ~~test object~~chemical compounds by incorporating the additional subset of ~~test object~~chemical compounds into the subset of ~~test object~~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 ~~test object~~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 ~~test object~~chemical compounds, before application of an instance of the eliminating E), comprises at least 100 million ~~test object~~chemical compounds.