Claims

1. A system for predicting binding affinity of one molecule to one target protein, the system comprising:

at least one electronic datastore configured to store records reflecting input data;

a data encoder module configured to perform a method comprising:

constructing a data structure from the stored input data in the least one electronic datastore, the data structure providing a geometric representation of the input data, wherein the constructing comprises:

augmenting the stored input data by generating additional instances of the stored input data through application of one or more pre-defined transformations of the stored input data wherein the one or more pre-defined transformations of the stored input data include a combination of rotation, translation, and mirror operators in any combination of the three X, Y, and Z planes of the stored input data thereby resulting in geometric data; and

truncating each instance of the stored input data in the geometric data to fit within a bounding box; and

populating the data structure by encoding data relating to the one molecule and the one target protein in the geometric data for each instance of the stored input data fitted to the bounding box in the geometric data, wherein the populating comprises:

partitioning the instance of the stored input data fitted to the bounding box in the geometric data into a corresponding voxel grid by transforming the instance of the input data fitted to the bounding box into a discrete set of regularly-spaced cubic voxels using a spacing that yields an integer number of voxels, wherein each respective voxel in the respective voxel grid comprises one or more input channels that encode a type of atom represented by the respective voxel in the respective voxel grid; and

a predictive model module configured to:

apply a predictive model to the data structure to generate an indicator of a binding affinity for the one molecule to the one target protein represented by the data structure, wherein the predictive model is a deep neural network or convolutional

10. A computer-implemented method for predicting binding affinity of one molecule to one target protein, the method comprising:

storing, in at least one electronic datastore, records reflecting input data;

constructing, at a data encoder module, a data structure from the stored input data in the least one electronic datastore the data structure providing a geometric representation of the input data, wherein the constructing comprises:

augmenting the stored input data by generating additional instances of the stored input data through application of one or more pre-defined transformations of the stored input data wherein the one or more pre-defined transformations of the stored input data include a combination of rotation, translation, and mirror operators in any combination of the three X, Y, and Z planes of the stored input data thereby resulting in geometric data, and

truncating each instance of the stored input data in the geometric data to fit within a bounding box; and

populating, at the at least one processor, the data structure by encoding data relating to the one molecule and the one target protein in the geometric data by, for each instance of the stored input data fitted to the bounding box in the geometric data, wherein the populating comprises:

partitioning the instance of the stored input data fitted to the bounding box in the geometric data into a corresponding voxel grid by transforming the instance of the input data fitted to the bounding box into a discrete set of regularly-spaced cubic voxels using a spacing that yields an integer number of voxels, wherein each respective voxel in the respective voxel grid comprises one or more input channels that encode a type of atom represented by the respective voxel in the respective voxel grid; and

applying, at the at least one processor, a predictive model to the data structure to generate an indicator of a binding affinity for the one molecule to the one target protein represented by the data structure, wherein

the predictive model is a deep neural network or convolutional neural network comprising a convolutional filter layer comprising a set of function computation