6. The method of claim 1, wherein the new interaction term is added to or subtracted from the base model to produce an updated model, further comprising:

(f) repeating (c) using the updated model in place of the base model and adding or subtracting a different interaction term than the one added or subtracted in (c); and

(g) repeating (d) and (e) using the updated model in place of the base model.

7. The method of claim 6, further comprising;

(h) repeating (f) and (g) using a further updated model.

8. The method of claim 1, wherein the ability of the at least one new model to predict activity in (d) is measured by Akaike Information Criterion or Bayesian Information Criterion.

9. The method of claim 1, wherein the sequence is a whole genome, whole chromosome, chromosome segment, a collection of gene sequences for interacting genes, gene, or protein.

10. The method of any of claim 1, wherein the sub-units are chromosomes, chromosome segments, haplotypes, genes, nucleotides, codons, mutations, amino acids, or residues.

11. The method of claim 1, wherein the plurality of biological molecules constitutes a training set of a protein variant library.

12. A computer program product comprising one or more computer-readable non-transitory storage media having stored thereon computer-executable instructions that, when executed by one or more processors of a computer system, cause the computer system to implement a method for identifying biological molecules to affect a desired activity, the method comprising:

(a) receiving sequence and activity data for a plurality of biological molecules;

(b) preparing a base model from the sequence and activity data, wherein the base model predicts activity as a function of the presence or absence of sub-units of the sequence;

(c) preparing at least one new model by adding or subtracting at least one new interaction term to or from the base model, wherein the new interaction term represents the interaction between two or more interacting sub-units;

(d) determining the ability of the at least one new model to predict activity as a function of the presence or absence of the sub-units; and

(e) determining whether to add or subtract the new interaction term to or from the base model based on the ability of the at least one new model to predict activity as determined in (d) and with a bias against including additional interaction terms.

~~12.~~13. A computer system, comprising:

one or more processors;

system memory; and

one or more computer-readable storage media having stored thereon computer-executable instructions that, when executed by the one or more processors, cause the computer system to implement a method for conducting directed evolution of biological molecules, the method comprising:

(a) receiving sequence and activity data for a plurality of biological molecules;

(b) preparing a base model from the sequence and activity data, wherein the base model predicts activity as a function of the presence or absence of sub-units of the sequence;

(c) preparing at least one new model by adding or subtracting at least one new interaction term to or from the base model, wherein the new interaction term represents the interaction between two or more interacting sub-units;

(d) determining the ability of the at least one new model to predict activity as a function of the presence or absence of the sub-units; and

(e) determining whether to add or subtract the new interaction term to or from the base model based on the ability of the at least one new model to predict activity as determined in (d) and with a bias against including additional interaction terms.