Claims

1. A method of assessing a biological sequence capable of a preselected function, comprising:

(a) preselecting a biological molecule, wherein the biological molecule is capable of a function of interest;

(b) preparing a testing sequence database comprising a plurality of sequence fragments of the preselected biological molecule, wherein the preselected sequence fragments are a predetermined length;

(c) identifying in the testing sequence database one or more sequence fragments of the preselected biological sequence that match one or more sequence fragments, respectively, of a second biological molecule having a biological function unrelated to the biological function of interest of the preselected biological molecule, and removing the identified sequence fragments(s) from the testing sequence database:

(d) fragmenting the sequence of one or more test biological molecules into lengths equivalent to the predetermined length of the sequence fragments of the preselected biological molecule in the testing sequence database;

(e) detecting a presence or absence of a sequence match between the sequence of at least one fragment of the fragmented test biological molecules and at least one of the plurality of sequence fragments of the preselected biological sequence, and

(f) acting in response to the detection in (e), wherein the detecting provides an assessment of the test biological molecule, wherein the acting comprises one or more of: preventing synthesis of the test biological molecule, permitting synthesis of the test biological molecule, sequencing one or more polynucleotide molecules, DNA sequencing, DNA molecule design, polypeptide sequence determination, and further sequence identification steps.

2. The method of claim 1, wherein if the presence of a sequence match is detected in (e) the action comprises preventing synthesis of the test biological molecule.

3. The method of any one of claims 1-2, wherein a means of preparing the testing sequence database comprises:

(a) screening the plurality of sequence fragments of the preselected biological sequence molecule against at least one control sequence database, wherein the control sequence database comprises a plurality of control sequence fragments of at least one molecule capable of a function of interest unrelated to the function of interest of the preselected biological molecule;

(b) identifying the presence of a match between a sequence fragment in the plurality of sequence fragments of the preselected biological molecule and a sequence fragment in the control sequence database that is a fragment of the biological molecule identified as capable of a function unrelated to the function of interest of the preselected biological molecule; and

(c) removing from the testing sequence database the sequence fragment of the preselected biological sequence identified as matching the sequence fragment of the biological sequence identified as capable of a function of interest unrelated to the function of interest of the molecule capable of the function of interest.

4. The method of any one of claims 1-3, wherein the preselected biological molecule is a polynucleotide.

5. The method of any one of claims 1-4, wherein the preselected biological molecule comprises a polypeptide.

6. The method of any claims 1-5, wherein the plurality of sequence fragments of the preselected biological molecule: (1) comprises all or a significant portion of possible fragments or at least one essential fragment of the biological molecule that is capable of the function of interest, and (2) does not comprise sequences found in a biological molecule capable of a function unrelated to the function of the preselected biological molecule.

7. The method of any one of claims 1-6, wherein the testing sequence database comprises one or more sequence fragments randomly or pseudorandomly selected from sequences of molecules known to be capable of a function of the preselected molecule's function of interest, and optionally, wherein the randomly or pseudorandomly selected sequence fragments are biased towards sequence regions with greater homology to functionally or phylogenetically related sequences.

8. The method of any one of claims 1-7, wherein the testing sequence database further comprises sequences that are functional equivalents of the plurality of sequence fragments of the preselected biological molecule, optionally wherein a means for identifying the functional equivalents comprises one or both of a computational means and an experimental means.

9. The method of claim 8 wherein a computational means for selecting the functional equivalents included in the testing sequence database comprises using a classifier based on experimental data to evaluate the accuracy of the computational means.

10. The method of claims 8, wherein a means for selecting the functional equivalents included in the testing sequence database comprises a random selection method or a pseudorandom selection method.

11. The method of claim 1, wherein the identities of all sequence fragments of one or both of the testing sequence database and the test biological molecule are protected, optionally wherein a means of the protecting comprises application of a cryptographic hash function, wherein the cryptographic hash function deterministically maps the sequence data to a bit string of fixed size using a one-way function, and wherein the hashing comprises hashing with an oblivious multiparty server system.

12. The method of claim 11, wherein the application of the cryptographic hash function cannot be reversed without a brute-force search of all possible sequence inputs into the testing sequence database.

13. The method of claim 12, wherein the application of the cryptographic hash function further comprises use of one or more information keys that must be accessed to attempt the brute-force search, optionally wherein the application of the cryptographic hash function requires keys from a plurality of independent sources that must cooperate to compute the hash without any one server gaining access to the sequence data, and optionally, wherein the independent sources comprise independent computer servers.

14. A method of identifying a biological sequence capable of a preselected function, comprising:

(a) preselecting a biological molecule, wherein the preselected biological molecule is capable of a function of interest;

(b) preparing a testing sequence database comprising a plurality of sequence fragments of the preselected biological molecule, wherein the preselected sequence fragments are a predetermined length;

(c) fragmenting the sequence of one or more test biological molecules into lengths equivalent to the predetermined length of the sequence fragments the preselected biological molecule in the testing sequence database; and

(d) detecting a presence or absence of a sequence match between the sequence of at least one fragment of the fragmented test biological molecules and at least one of the plurality of sequence fragments of the preselected biological sequence;

wherein a means of preparing the testing sequence database comprises:

(i) screening the plurality of sequence fragments of the preselected biological sequence molecule against at least one control sequence database, wherein the control sequence database comprises a plurality of control sequence fragments of at least one molecule capable of a function of interest unrelated to the function of interest of the preselected biological molecule;

(ii) identifying the presence of a match between a sequence fragment in the plurality of sequence fragments of the preselected biological molecule and a sequence fragment in the control sequence database that is a fragment of the biological molecule identified as capable of a function unrelated to the function of interest of the preselected biological molecule; and

(iii) removing from the testing sequence database the sequence fragment of the preselected biological sequence identified as matching the sequence fragment of the biological sequence identified as capable of a function of interest unrelated to the function of interest of the molecule capable of the function of interest.

15. The method of 14, wherein the preselected biological molecule is a polynucleotide.

16. The method of claim 14, wherein the preselected biological molecule comprises a polypeptide molecule.

17. The method of claim 14, wherein the plurality of sequence fragments of the preselected biological molecule: (1) comprises all or a significant portion of possible fragments of the biological molecule that is capable of the function of interest, and (2) does not comprise sequences not found in a biological molecule capable of a function unrelated to the function of the preselected biological molecule.

18. The method of claim 14, wherein the control sequence database comprises a plurality of control sequence fragments of at least one molecule capable of a function of interest unrelated to the function of interest of the preselected biological molecule.

19. The method of claim 14, wherein the control sequence database comprises a plurality of control sequence fragments of at least one molecule not capable of the function of interest of the preselected biological molecule.

20. The method of claim 14, wherein the testing sequence database comprises one or more sequence fragments randomly selected from sequences of molecules known to be capable of a function of the preselected molecule's function of interest, and optionally wherein the randomly selected sequence fragments are biased towards conserved regions.

21. A testing sequence database prepared by a method of any one of claims 1-20.