Claims

1. A method of assessing 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 plurality of sequence fragments are a predetermined length;

(c) fragmenting the sequence of a test biological molecule into lengths equivalent to the predetermined length of the plurality of sequence fragments of the preselected biological molecule in the testing sequence database;

(d) detecting, in an automated manner, a presence or absence of a sequence match between at least one fragment of the fragmented test biological molecule and at least one of the plurality of sequence fragments of the preselected biological molecule in the testing sequence database, and

(e) acting in response to the detection in (d), 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 the acting of (e) comprises preventing synthesis of the test biological molecule.

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

(i) screening the plurality of sequence fragments of the preselected biological molecule against at least one control sequence database, wherein the at least one 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 a match between a sequence fragment in the plurality of sequence fragments of the preselected biological molecule and a control sequence fragment in the at least one control sequence database; and

(iii) removing from the testing sequence database the sequence fragment of the preselected biological molecule identified as matching the control sequence fragment.

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 different from the preselected biological molecule's function of interest, and optionally, wherein the one or more 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.

9. 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

the protecting comprises applying a cryptographic hash function that deterministically maps sequence data to a bit string of fixed size using a one-way function.

10. The method of claim 9, wherein applying the cryptographic hash function cannot be reversed without a brute-force search of all possible sequence inputs into the testing sequence database.

11. The method of claim 10, wherein applying the cryptographic hash function further comprises use of one or more information keys that must be accessed to attempt the brute-force search.

12. 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 plurality of sequence fragments are a predetermined length, and wherein preparing the testing sequence database comprises:

(i) screening the plurality of sequence fragments of the preselected biological molecule against at least one control sequence database, wherein the at least one 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 a match between a sequence fragment in the plurality of sequence fragments of the preselected biological molecule and a control sequence fragment in the at least one control sequence database; and

(iii) removing from the testing sequence database the sequence fragment of the preselected biological molecule identified as matching the control sequence fragment;

(c) fragmenting the sequence of a test biological molecule into lengths equivalent to the predetermined length of the plurality of sequence fragments of the preselected biological molecule in the testing sequence database; and

(d) detecting, in an automated manner, a presence or absence of a sequence match between at least one fragment of the fragmented test biological molecule and at least one of the plurality of sequence fragments of the preselected biological molecule in the testing sequence database.

13. The method of claim 12, wherein the preselected biological molecule is a polynucleotide.

14. The method of claim 12, wherein the preselected biological molecule comprises a polypeptide molecule.

15. The method of claim 12, wherein the plurality of sequence fragments of the preselected biological molecule: (1) comprises all or a significant portion of possible fragments of the preselected 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.

16. The method of claim 12, wherein the at least one 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.

17. The method of claim 12, wherein the testing sequence database comprises one or more sequence fragments randomly selected from sequences of molecules known to be capable of a function different from the preselected biological molecule's function of interest, and optionally wherein the one or more randomly selected sequence fragments are biased towards conserved regions.