

Claims

1. A method for selecting a polypeptide from a library of polypeptides, comprising the steps of
 - a. providing a first nucleic acid library, wherein each member of said first nucleic acid library comprises a polypeptide-encoding sequence encoding a member of a first polypeptide library;
 - b. providing a second nucleic acid library, wherein said second nucleic acid library comprises a plurality of members, wherein each member comprises a tag-encoding sequence encoding a detection tag, wherein said detection tag:
 - i. is characterized by an amino acid sequence different from the amino acid sequence of any other detection tag encoded by said second nucleic acid library;
 - ii. is characterized by a molecular mass of between 200 and 5000 Da; and
 - iii. comprises a first severable element;
 - iv. is characterized by a hydrophobicity value between -27 and 128;
 - c. inserting said polypeptide-encoding sequence comprised in said member of said first nucleic acid library into a member of said second nucleic acid library, thereby creating a tagged nucleic acid library encoding a tagged polypeptide library, wherein each member of said tagged polypeptide library comprises a polypeptide and a detection tag separated from said polypeptide by said first severable element;
 - d. obtaining a plurality of nucleic acid sequences from said tagged nucleic acid library, wherein each of said plurality of nucleic acid sequences comprises a polypeptide-encoding sequence and a tag-encoding sequence;
 - e. predicting a mass spectrometry fragmentation pattern for each detection tag encoded by a tag-encoding sequence obtained in step d;
 - f. expressing said tagged polypeptide library from said tagged nucleic acid library;
 - g. selecting a member of said tagged polypeptide library in a selection step, yielding a selected polypeptide;
 - h. severing said first severable element, thereby separating said detection tag from said selected polypeptide, yielding an isolated detection tag;
 - i. identifying said isolated detection tag by
 - i. recording a fragmentation pattern of said isolated detection tag by mass spectrometry;

- ii. matching said fragmentation pattern obtained in step i with said fragmentation patterns predicted in step e, thereby identifying said isolated detection tag;
 - j. selecting from said plurality of nucleic acid sequences obtained in step d a nucleic acid sequence comprising a tag-encoding sequence encoding said detection tag identified in step i, thereby identifying the member of said tagged polypeptide library associated with said detection tag identified in step i.
- 5 2. The method according to claim 1, wherein said isolated detection tag is characterized by a hydrophobicity value of between -1 and 70.
- 10 3. The method according to any one of the above claims, wherein said isolated detection tag comprises a sequence element I selected from a collection of sequence elements I, wherein said sequence element I consists of 5 to 10 amino acids, independently of each other selected from A, S, T, N, Q, D, E, V, L, I, F, Y, W, G and P.
- 15 4. The method according to claim 1 or 2, wherein said isolated detection tag consists of
 - a. a sequence element III, wherein said sequence element III is GS;
 - b. said sequence element I, wherein said sequence element I consists of 5 to 10 amino acids, independently of each other selected from A, S, T, N, Q, D, E, V, L, I, F, Y, W, G and P; and
 - c. said sequence element II selected from SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) and SEQ ID NO 05 (WQEGGR);
- 20 wherein the order of said sequence elements from N-terminus to C-terminus is sequence element III, sequence element I, sequence element II.
- 25 5. A collection of polypeptides, wherein each member of said collection of polypeptides is associated with at least one detection tag, and wherein said detection tag
 - a. is characterized by an amino acid sequence different from the amino acid sequence of any other detection tag encoded by said plurality of expression vectors;
 - b. is characterized by a molecular mass of between 200 and 5000 Da;
 - c. is separated from said member of said collection of polypeptides by a first severable element;
 - d. is characterized by a hydrophobicity value between -27 and 128.
- 30 6. The collection of polypeptides according to claim 5, wherein said isolated detection tag is characterized by a hydrophobicity value of between -1 and 70.
7. The collection of polypeptides according to any one of claims 5 or 6, wherein said detection tag comprises

- a. a sequence element I, wherein said sequence element I consists of 5 to 10 amino acids, independently of each other selected from A, S, T, N, Q, D, E, V, L, I, F, Y, W, G and P; and
 - b. a sequence element II selected from SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) and SEQ ID NO 05 (WQEGGR).
8. A detection tag consisting of 4 to 20 amino acids, wherein said detection tag
 - a. comprises only one amino acid having a positively charged side chain;
 - b. is characterized by a molecular mass of between 200 and 5000 Da;
 - c. is characterized by a hydrophobicity value between -27 and 128.
9. The detection tag according to claim 8, wherein said detection tag consists of
 - a. a sequence element I, wherein said sequence element I consists of 5 to 10 amino acids, independently of each other selected from A, S, T, N, Q, D, E, V, L, I, F, Y, W, G and P; and
 - b. a sequence element II selected from SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) and SEQ ID NO 05 (WQEGGR).
10. A collection of detection tags comprising at least 96 detection tags according to claim 8, wherein each detection tag consists of 4 to 20 amino acids and is characterized by an amino acid sequence different from the amino acid sequence of any other detection tag comprised in said collection of detection tags.
11. A collection of plasmid vectors comprising at least 96 plasmid vectors, wherein each member of said collection of plasmid vectors comprises a tag-encoding nucleic acid sequence encoding a detection tag, wherein each detection tag consists of 4 to 20 amino acids and is characterized by an amino acid sequence different from the amino acid sequence of any other detection tag encoded by said collection of plasmid vectors; and wherein said encoded detection tag is characterized by a hydrophobicity value between -27 and 128.
12. The collection of plasmid vectors according to claim 11, wherein said detection tag consists of
 - a. a sequence element I, wherein said sequence element I consists of 5 to 10 amino acids, independently of each other selected from A, S, T, N, Q, D, E, V, L, I, F, Y, W, G and P; and
 - b. a sequence element II selected from SEQ ID NO 01 (WR), SEQ ID NO 02 (WLR), SEQ ID NO 03 (WQSR), SEQ ID NO 04 (WLTVR) and SEQ ID NO 05 (WQEGGR).
13. A method of protein detection comprising
 - a. providing a nucleic acid library encoding a polypeptide library

wherein said polypeptide library comprises a plurality of members and each member is associated with a detection tag, and wherein said detection tag

- i. is characterized by an amino acid sequence different from the amino acid sequence of any other detection tag encoded by said nucleic acid library;
- ii. is characterized by a molecular mass of between 200 and 5000 Da; and
- iii. is separated from said member of said polypeptide library by a first severable element;
- iv. is characterized by a hydrophobicity value between -27 and 128;

b. providing a database comprising

- i. a plurality of nucleic acid and/or amino acid sequences, wherein said plurality of sequences comprises the sequences of all members of said nucleic acid library, and wherein each of said sequences comprises a sequence specifying a polypeptide and a sequence specifying a detection tag;
- ii. a predicted mass spectrometry fragmentation pattern for each detection tag encoded by said nucleic acid library;

c. expressing said polypeptide library from said nucleic acid library;

d. selecting a member of said polypeptide library in a selection step, yielding a selected polypeptide;

e. severing said first severable element, thereby separating said detection tag from said selected polypeptide, yielding an isolated detection tag;

f. identifying said isolated detection tag by

- i. recording a fragmentation pattern of said isolated detection tag by mass spectrometry;
- ii. matching said fragmentation pattern obtained in step i with said fragmentation patterns predicted in said database, thereby identifying said isolated detection tag;

g. selecting from said plurality of sequences comprised in said database a sequence specifying said detection tag identified in step f, thereby identifying the member of said polypeptide library associated with said detection tag identified in step f.

14. The method according to claim 13, wherein each member of said polypeptide library or wherein each detection tag is associated with an affinity tag, or an affinity tag selected from the group comprising a His-tag, a CBP-tag, a CYD-tag, a Strep-tag, a StreplI-tag, a FLAG-tag, a HPC-tag, a GST-tag, an Avi-tag, a biotinylation tag, a Myc-tag, a 3xFLAG tag and a MBP-tag.

15. A method of associating a polypeptide with a unique detection tag, comprising the steps of

- 5
- a. providing a first nucleic acid library, wherein each member of said first nucleic acid library comprises a polypeptide-encoding sequence encoding a member of a first polypeptide library;
 - b. providing a second nucleic acid library, wherein each member of said second nucleic acid library comprises a tag-encoding sequence encoding a detection tag, wherein said detection tag:
 - i. is characterized by an amino acid sequence different from the amino acid sequence of any other detection tag encoded by said second nucleic acid library;
 - 10 ii. is characterized by a molecular mass of between 200 and 5000 Da;
 - iii. is characterized by a hydrophobicity value between -27 and 128;
 - c. inserting said polypeptide-encoding sequence comprised in said member of said first nucleic acid library into a member of said second nucleic acid library, wherein
 - i. said first nucleic acid library has a size of 5 to 100.000 and
 - 15 ii. said second nucleic acid library has a size of 10^3 to 10^{11} ,thereby generating a plurality of polypeptide-tag combination plasmids;
 - d. selecting a subset of said plurality of polypeptide-tag combination plasmids, thereby generating a tagged nucleic acid library encoding a tagged polypeptide library.
- 20 16. The method according to claim 14, wherein said affinity tag is separated from said detection tag by a second severable element, and said second severable element is severed prior to step f.