WHAT IS CLAIMED IS:

1. A method for designing a stabilized antibody or antigen binding fragment thereof, which comprises a stabilized V_H domain and a stabilized V_L domain derived from a V_H domain and a V_L domain of a candidate antibody, wherein said method comprises:

(a) providing a curated reference set of amino acid sequences which belong to immunoglobulin superfamily polypeptides;

(b) aligning said reference set of amino acid sequences to generate an aligned set;

(c) calculating covariation between two or more amino acid positions in the aligned set to generate a covariation dataset;

(d) identifying non-covarying amino acids in said candidate antibody V_H domain or said candidate antibody V_L domain, wherein said non-covarying amino acids fail to satisfy one or more covariations in the covariation dataset;and

(e) substituting one or more of said non-covarying amino acids in the candidate antibody V_H domain or candidate antibody V_L domain with amino acids found at the corresponding amino acid positions in the aligned set which satisfy covariations in the covariation dataset, thereby making a stabilized antibody and antigen binding fragment thereof with an improved biophysical property.

2. The method of claim 1, wherein said biophysical property is selected from the group consisting of thermal stability, pH unfolding profile, stable removal of glycosylation, solubility, biochemical function, and combinations thereof.

3. The method of claim 1, wherein said stabilized antibody or antigen binding fragment thereof is selected from the group consisting of a domain antibody, a humanized antibody, a human antibody, a non-human monoclonal antibody, a chimeric antibody, a bispecific antibody, a scFv molecule, a scFv-containing antibody, a domain-deleted antibody, and a combination of any of said antibody fragments.

4. The method of claim 1, wherein said amino acids in the covariation dataset in (b) are part of a structural feature selected from the group consisting of a disulfide bond, a salt bridge, a portion of a ligand binding pocket or surface, a network of van der Waals interactions, a network of hydrogen bond interactions, a network of charge-charge interactions, and a combination of two or more of said structural features.

5. The method of claim 1, wherein each amino acid sequence of said aligned set in (b) has less than 95% identity with the other sequences in said aligned set.

6. The method of claim 1, wherein the construction of said curated reference set of amino acid sequences comprises:

(a) collecting a set of Ig-fold three-dimensional structures, wherein said structures contain an Ig fold selected from the group of Ig fold classes consisting of a V-class fold, an I-class fold, a C1-class fold, a C2-class fold, and a combination of said Ig fold classes;

(b) filtering said set of Ig-fold structures by discarding Ig-fold structures containing sequences with breaks, 100% sequence identity, aberrant length, or displaying misfolded topologies;

(c) constructing structure alignments, wherein said filtered set of Ig-fold structures are

superimposed upon one another;

(d) obtaining sequence alignments from said structure alignments, wherein amino acids from the sequence of one structure are matched to amino acids from the sequence of a second structure based on the shortest distance between α-carbons of the polypeptide backbones;

(e) constructing Hidden Markov Models (HMMs) based upon the structure-based sequence alignments for one of the Ig-fold classes;

(f) searching a protein sequence database using one or more of said Ig-fold class-specific HMMs, wherein sequences in the protein sequence database matching said HMMs are retrieved; and,

(g) validating the Ig-fold class assignment of the protein sequences retrieved in (f) using an annotated database of protein domains, wherein the candidate Ig-fold protein sequences found in (f) are retained only if their assignment to a Ig-fold class in the annotated database of protein domain is statistically significant.

7. The method of claim 1, wherein said aligning further comprises:

(aa) removing redundant or highly similar amino acid sequences from the sequence alignment obtained in (b); and

(bb) removing columns in the alignment that are not match states in the HMM profile.

8. The method of claim 1, wherein the calculation of (c) comprises one or more features selected from the group consisting of (i) identification of gaps as a distinct amino acid type, (ii) a weighing function, wherein said weighing function is not a Henikoff diversity weighing function, (iii) a filtering function for covarying pairs wherein said function does not

use Sequence Average Identities (SAI), (iv) an event cut-off, wherein covarying pairs are not reported unless they are observed a minimum number of times, and wherein said event cut-off is about 2 or more events, and (v) a combination of two or more of said features.

9. The method of claim 1, wherein the calculation of (c) comprises a calculation of the statistical significance of variation using a χ^2 analysis.

10. The method of claim 9, wherein values of said χ^2 are calculated using an event based count formula.

11. The method of claim 10, wherein said event based count formula is:

$$\chi^2 = \frac{[c(i,j) - (p(i) \cdot p(j) \cdot c(t))]^2}{p(i) \cdot p(j) \cdot c(t)}$$

wherein

p(i) and *p(j)* are the amino acid frequencies of any two amino acid types of interest at positions *i* and *j*, respectively, in the aligned set of sequences;

c(i,j) is the number of times *p(i)* and *p(j)* are observed in the same sequence; and,

c(t) is the number of total sequences in the alignment;

and wherein amino acid frequencies are defined as the number of times an amino acid type is observed at a specific position in an alignment divided by the total number of sequences in the alignment.

12. The method of claim 1, wherein the calculation of (c) comprises a calculation of the correlation coefficient (ϕ) performed using the formula:

$$\phi(a_i b_j) = \frac{(a_i b_j * \bar{a}_i \bar{b}_j) - (a_i \bar{b}_j * \bar{a}_i b_j)}{\sqrt{(a_i b_j + \bar{a}_i b_j) * (a_i \bar{b}_j + \bar{a}_i \bar{b}_j) * (a_i b_j + a_i \bar{b}_j) * (\bar{a}_i b_j + \bar{a}_i \bar{b}_j)}}$$

wherein

$a_i b_j$ is the number of times amino acids of type a or b are found in the same sequence at positions i and j, respectively;

$\bar{a}_i \bar{b}_j$ is the number of times both amino acid types are absent from the same sequence;

$a_i \bar{b}_j$ is the number of times amino acid a is found present while amino acid b is absent; and,

$\bar{a}_i b_j$ is the number of times amino acid a is absent while amino acid b is present.

and wherein the correlation coefficient (ϕ) measures the statistical strength of the covariation.

13. The method of claim 1, wherein the calculation of covariation includes the generation of covariation scores only for covariations that satisfy a threshold level of statistical significance.

14. The method of claim 13, wherein the covariation scores are generated only for covariations above or below a threshold χ^2 value or ϕ value.

15. The method of claim 14, wherein amino acid positions of the candidate

sequence are assigned positive specific covariation scores for positive covariations with ϕ correlation coefficients of about +0.25 to about +1.0.

16. The method claim 14, wherein amino acid positions of the candidate sequence are assigned negative specific covariation scores for negative covariations with ϕ correlation coefficients of about -0.25 to about -1.0.

17. The method of claim 1, further comprising:

(aa) providing structural models of a template antibody V_H or V_L domain and the candidate antibody V_H or V_L domain;

(bb) identifying protein-protein interface amino acids in the template antibody V_H or V_L domain that are important for stability;

(cc) identifying scaffolding amino acids which covary with the interface amino acids of (bb);

(dd) substituting one or more interface amino acids or scaffolding amino acids in the candidate antibody V_H or V_L domain with the corresponding interface amino acids or scaffolding amino acids identified in (bb) and (cc).

18. The method of claim 17, wherein the template antibody V_H or V_L domain contains an Ig superfamily fold.

19. The method of claim 17, wherein the protein-protein interface amino acids are located at the V_H/V_L interface of a candidate V_H or V_L domain.

20. The method of claim 1, further comprising:

(1) calculating consensus-based scores for one or more amino acid positions of said candidate antibody V_H or V_L domain;

(2) combining said consensus-based scores with the data in the covariation dataset; and,

(3) selecting amino acid substitutions predicted to stabilize the candidate antibody V_H domain or V_L domain, wherein said selection is based on the combination of consensus-based scores and covariation data.

21. The method of claim 20, wherein the calculation of consensus-based scores comprises the steps of:

(i) providing a curated reference set of polypeptide sequences which belong to the Immunoglobulin superfamily;

(ii) aligning said reference set sequences to generate an aligned set; and,

(iii) calculating the test amino acid frequency for each amino acid position within the candidate antibody V_H or V_L domain, wherein said frequency is calculated by summing the number of times the amino acid at said amino acid position is present at the corresponding position in the aligned set , and dividing the summed value by the total number of sequences within the reference set.

22. The method of claim 21, comprising the further steps of:

(a) calculating a consensus sequence, wherein the amino acid at each position within the sequence corresponds to the most common amino acid at said position in the aligned set;

(b) calculating the consensus amino acid frequency for each amino acid position within the

consensus polypeptide sequence, wherein said sequence is calculated by summing the number of times the amino acid at said amino acid position is present at the corresponding position in the aligned set, and dividing the summed value by the total number of sequences within the reference set; and,

(c) dividing the test amino acid frequency by the consensus amino acid frequency to give a consensus score.

23. An antibody or antigen binding fragment thereof, which is designed by the method of any one of claims 1-22.

24. The antibody or fragment thereof of claim 23, which comprises a stabilized scFv molecule.

25. The antibody or fragment thereof of claim 23, wherein said stabilized scFv molecule comprises an amino acid sequence derived from a candidate scFv molecule comprising SEQ ID NO: 4.

26. A polynucleotide encoding said antibody or antigen binding fragment thereof of claim 23.

27. A host cell comprising said polynucleotide of claim 26.

WHAT IS CLAIMED IS:

1. A method for producing a stabilized antibody or antigen binding fragment thereof, which comprises a stabilized V_H domain and a stabilized V_L domain derived from a V_H domain and a V_L domain of a wild-type candidate antibody, wherein said method comprises:

(a) providing a curated reference set of amino acid sequences which belong to immunoglobulin superfamily polypeptides;

(b) aligning said reference set of amino acid sequences to generate an aligned set;

(c) calculating covariation between two or more amino acid positions in the aligned set to generate a covariation dataset;

(d) identifying non-covarying amino acids in said wild-type candidate antibody V_H domain or said wild-type candidate antibody V_L domain, wherein said non-covarying amino acids fail to satisfy one or more covariations observed in the covariation dataset;

(e) substituting one or more of said non-covarying amino acids in the wild-type candidate antibody V_H domain or wild-type candidate antibody V_L domain with one or more replacement amino acids, wherein each replacement amino acid is an amino acid found at the corresponding amino acid position in the aligned set which satisfies one or more covariations in the covariation dataset, and wherein the replacement amino acid establishes one or more covariations in the candidate antibody;

(f) expressing in a host cell the antibody or antigen binding fragment thereof comprising at least one substitution, and,

(g) testing at lest one biophysical property of the antibody or antigen binding fragment thereof comprising at least one substitution to determine whether at least one biophysical property is improved with respect to the wild-type candidate antibody,

thereby making a stabilized antibody and antigen binding fragment thereof with an improved biophysical property.

2. The method of claim 1, wherein said biophysical property is selected from the group consisting of thermal stability, pH unfolding profile, stable removal of glycosylation, solubility, biochemical function, and combinations thereof.

3. The method of claim 1, wherein said stabilized antibody or antigen binding fragment thereof is a domain antibody, a humanized antibody, a human antibody, a non-human monoclonal antibody, a chimeric antibody, a bispecific antibody, a scFv molecule, a scFv-containing antibody, or a domain-deleted antibody comprising a stabilized V_H domain and a stabilized V_L domain derived from a V_H domain and a V_L domain of a wild-type candidate antibody.

4. The method of claim 1, wherein said amino acids in the covariation dataset in (b) are part of a structural feature selected from the group consisting of a disulfide bond, a salt bridge, a portion of a ligand binding pocket or surface, a network of van der Waals interactions, a network of hydrogen bond interactions, a network of charge-charge interactions, and a combination of two or more of said structural features.

5. The method of claim 1, wherein each amino acid sequence of said aligned set in (b) has less than 95% identity with the other sequences in said aligned set.

6. The method of claim 1, wherein the construction of said curated reference set of amino acid sequences comprises:

(a) collecting a set of Ig-fold three-dimensional structures, wherein said structures contain an Ig fold selected from the group of Ig fold classes consisting of a V-class fold, an I-class fold, a Cl-class fold, a C2-class fold, and a combination of said Ig fold classes;

(b) filtering said set of Ig-fold structures by discarding Ig-fold structures containing sequences with breaks, 100% sequence identity, aberrant length, or displaying misfolded topologies;

(c) constructing structure alignments, wherein said filtered set of Ig-fold structures are superimposed upon one another;

(d) obtaining sequence alignments from said structure alignments, wherein amino acids from the sequence of one structure are matched to amino acids from the sequence of a second structure based on the shortest distance between α-carbons of the polypeptide backbones;

(e) constructing Hidden Markov Models (HMMs) based upon the structure-based sequence alignments for one of the Ig-fold classes;

(f) searching a protein sequence database using one or more of said Ig-fold class-specific HMMs, wherein sequences in the protein sequence database matching said HMMs are retrieved; and,

(g) validating the Ig-fold class assignment of the protein sequences retrieved in (f) using an annotated database of protein domains, wherein the candidate Ig-fold protein sequences found in (f) are retained only if their assignment to a Ig-fold class in the annotated database of protein domain is statistically significant;

7. The method of claim 1, wherein said aligning further comprises:

(aa) removing redundant or highly similar amino acid sequences from the sequence alignment obtained in (b); and

(bb) removing columns in the alignment that are not match states in the HMM profile.

8. The method of claim 1, wherein the calculation of (c) comprises one or more features selected from the group consisting of (i) identification of gaps as a distinct amino acid

type, (ii) a weighing function, wherein said weighing function is not a Henikoff diversity weighing function, (iii) a filtering function for covarying pairs wherein said function does not use Sequence Average Identities (SAI), (iv) an event cut-off, wherein covarying pairs are not reported unless they are observed a minimum number of times, and wherein said event cut-off is about 2 or more events, and (v) a combination of two or more of said features.

9. The method of claim 1, wherein the calculation of (c) comprises a calculation of the statistical significance of variation using a χ^2 analysis.

10. The method of claim 9, wherein values of said χ^2 are calculated using an event based count formula.

11. The method of claim 10, wherein said event based count formula is:

$$\chi^2 = \frac{[c(i,j) - (p(i) \cdot p(j) \cdot c(t))]^2}{p(i) \cdot p(j) \cdot c(t)}$$

wherein

p(i) and *p(j)* are the amino acid frequencies of any two amino acid types of interest at positions *i* and *j*, respectively, in the aligned set of sequences;

c(i,j) is the number of times *p(i)* and *p(j)* are observed in the same sequence; and,

c(t) is the number of total sequences in the alignment;

and wherein amino acid frequencies are defined as the number of times an amino acid type is observed at a specific position in an alignment divided by the total number of sequences in the alignment.

12. The method of claim 11, wherein the calculation of (c) comprises a calculation of the correlation coefficient (ϕ) performed using the formula:

$$\phi(a_i b_j) = \frac{(a_i b_j * \bar{a}_i \bar{b}_j) - (a_i \bar{b}_j * \bar{a}_i b_j)}{\sqrt{(a_i b_j + \bar{a}_i b_j) * (a_i \bar{b}_j + \bar{a}_i \bar{b}_j) * (a_i b_j + a_i \bar{b}_j) * (\bar{a}_i b_j + \bar{a}_i \bar{b}_j)}}$$

wherein

$a_i b_j$ is the number of times amino acids of type *a* or *b* are found in the same sequence at positions *i* and *j*, respectively;

$\bar{a}_i \bar{b}_j$ is the number of times both amino acid types are absent from the same sequence;

$a_i \bar{b}_j$ is the number of times amino acid *a* is found present while amino acid *b* is absent; and,

$\bar{a}_i b_j$ is the number of times amino acid *a* is absent while amino acid *b* is present.

and wherein the correlation coefficient (ϕ) measures the statistical strength of the covariation.

13. The method of claim 11, wherein the calculation of covariation includes the generation of covariation scores only for covariations that satisfy a threshold level of statistical significance.

14. The method of claim 13, wherein the covariation scores are generated only for covariations above or below a threshold χ^2 value or ϕ value.

15. The method of claim 14, wherein amino acid positions of the candidate sequence are assigned positive specific covariation scores for positive covariations with ϕ correlation coefficients of about +0.25 to about +1.0.

16. The method claim 14, wherein amino acid positions of the candidate sequence are assigned negative specific covariation scores for negative covariations with ϕ correlation coefficients of about -0.25 to about -1.0.

17. The method of claim 1, further comprising:

(aa) providing structural models of a template antibody V_H or V_L domain and the candidate antibody V_H or V_L domain;

(bb) identifying protein-protein interface amino acids in the template antibody V_H or V_L domain that are important for stability;

(cc) identifying scaffolding amino acids which covary with the interface amino acids of (bb);

(dd) substituting one or more interface amino acids or scaffolding amino acids in the candidate antibody V_H or V_L domain with the corresponding interface amino acids or scaffolding amino acids identified in (bb) and (cc).

18. The method of claim 17, wherein the template antibody V_H or V_L domain contains an Ig superfamily fold.

19. The method of claim 17, wherein the protein-protein interface amino acids are located at the V_H/V_L interface of a candidate V_H or V_L domain.

20. The method of claim 1, further comprising:

(1) calculating consensus-based scores for one or more amino acid positions of said candidate antibody V_H or V_L domain;

(2) combining said consensus-based scores with the data in the covariation dataset; and,

(3) selecting amino acid substitutions predicted to stabilize the candidate antibody V_H domain or V_L domain, wherein said selection is based on the combination of consensus-based scores and covariation data.

21. The method of claim 20, wherein the calculation of consensus-based scores comprises the steps of:

(i) providing a curated reference set of polypeptide sequences which belong to the Immunoglobulin superfamily;

(ii) aligning said reference set sequences to generate an aligned set; and,

(iii) calculating the test amino acid frequency for each amino acid position within the candidate antibody V_H or V_L domain, wherein said frequency is calculated by summing the number of times the amino acid at said amino acid position is present at the corresponding position in the aligned set, and dividing the summed value by the total number of sequences within the reference set.

22. The method of claim 21, comprising the further steps of:

(a) calculating a consensus sequence, wherein the amino acid at each position within the sequence corresponds to the most common amino acid at said position in the aligned set;

(b) calculating the consensus amino acid frequency for each amino acid position within the consensus polypeptide sequence, wherein said sequence is calculated by summing the number of times the amino acid at said amino acid position is present at the corresponding position in the aligned set, and dividing the summed value by the total number of sequences within the reference set; and,

(c) dividing the test amino acid frequency by the consensus amino acid frequency to give a consensus score.