METHODS FOR IDENTIFYING EPITOPES AND PARATOPES

CROSS REFERENCE TO RELATED APPLICATIONS

This application claims the benefit of U.S. Provisional Application No. 62/784,617, filed December 24, 2018. The contents of the aforesaid application are hereby incorporated by reference in its entirety.

BACKGROUND

Antibodies bind target antigens with high specificity and affinity. Molecularly, binding is facilitated by the set of amino acids in the antibody (paratope) and the antigen (epitope) which contribute to energetically favorable interactions for binding to occur. Determining the structural features governing antibody-antigen interactions is important for understanding an antibody's mechanism of action and as a reference to aid antibody engineering efforts. X-ray co-crystallography is a leading method to determine the structure of antibody-antigen complexes, detailing both the structural paratope and epitope with high resolution. However, achievement of high resolution co-crystal structures has considerable resource, throughput, and specialized technical expertise requirements. Other methods to characterize paratopes and epitopes provide greater throughput and experimental accessibility but typically come with a tradeoff of resolution. Epitope binning by competition binding or epitope characterization by alanine scanning each provide greater speed and throughput than crystallography but cannot provide the molecular detail nor the comprehensiveness of characterization as in crystallography. Thus, there exists a need in the art for improved methods of identifying epitope and paratope regions between an antibody and its recognized antigen.

SUMMARY

In an aspect, there is provided a method of identifying an epitope on a target polypeptide, the method comprising:

(a) binding an antibody molecule to a plurality of variants of the target polypeptide;

(b) obtaining (*e.g.*, enriching) a plurality of variants exhibiting reduced binding (*e.g.*, reduced binding affinity) to the antibody molecule;

(c) determining (*e.g.*, calculating) an enrichment score for each of the plurality of the obtained (*e.g.*, enriched) variants;

(d) generating an antibody molecule-target polypeptide docking model, wherein the antibody molecule-target polypeptide docking model is constrained according to the enrichment scores; and

(e) identifying a site on the target polypeptide that is capable of being bound by the antibody molecule based on the antibody molecule-target polypeptide docking model;

thereby identifying an epitope on a target polypeptide.

In an example, the disclosure features a method of identifying an epitope on a target polypeptide (*e.g.*, a target polypeptide described herein), the method comprising:

(a) binding an antibody molecule (*e.g.*, an antibody molecule described herein) to a plurality of variants of the target polypeptide;

(b) obtaining (*e.g.*, enriching) a plurality of variants exhibiting altered (*e.g.*, reduced) binding to the antibody molecule;

(c) determining (*e.g.*, calculating) an enrichment score for each of the plurality of the obtained (*e.g.*, enriched) variants;

(d) generating an antibody molecule-target polypeptide docking model, wherein the antibody molecule-target polypeptide docking model is constrained according to the enrichment scores; and

(e) identifying a site on the target polypeptide that is capable of being bound by the antibody molecule based on the antibody molecule-target polypeptide docking model;

thereby identifying an epitope on a target polypeptide.

In an example, the altered binding comprises altered binding affinity, *e.g.*, reduced binding affinity.

In an example, step (a) comprises binding the antibody molecule to a library displaying a plurality of variants of the target polypeptide. In an example, step (a) comprises binding the antibody molecule to a library comprising a plurality of cells expressing (*e.g.*, displaying) a plurality of variants of the target polypeptide. In an example, each of the plurality of cells expresses about one distinct variant of the target polypeptide. In an example, the cell is a eukaryotic cell, *e.g.*, a yeast cell.

In an example, the plurality of variants comprise mutations on one or more surface residues of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of a selected surface residue of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of each of a plurality of selected surface residues of the target polypeptide.

In an example, the plurality of variants comprise single amino acid substitutions, relative to a wild-type amino acid sequence of the target polypeptide. In an example, each of the plurality of variants comprises a single amino acid substitution relative to a wild-type amino acid sequence of the target polypeptide. In an example, the single amino acid substitution occurs at a surface residue of the target polypeptide.

In an example, the altered (*e.g.*, reduced) binding comprises an alteration (*e.g.*, a reduction) of binding detected for the variant and the antibody molecule, relative to the binding detected for a wild-type target polypeptide and the antibody.

In an example, step (b) comprises obtaining (*e.g.*, enriching) variants exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by the wild-type target polypeptide.

In an example, step (b) comprises obtaining (*e.g.*, enriching) cells exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a cell comprising a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by a cell comprising the wild-type target polypeptide.

In an example, step (b) comprises performing one or more, *e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more, enrichments for variants exhibiting reduced binding to the antibody molecule.

In an example, the method further comprises, *e.g.*, prior to step (c), identifying the variants exhibiting altered (*e.g.*, reduced) binding to the antibody molecule, *e.g.*, by sequencing the genes encoding the variants, *e.g.*, by next-generation sequencing.

In an example, step (c) comprises determining the frequency of occurrence for each of the plurality of the obtained (*e.g.*, enriched) variants. In an example, step (c) further comprises aggregating the frequency of occurrence of each variant comprising a distinct mutation at a particular residue and/or weighting (*e.g.*, heavily weighting) variants with higher frequencies of occurrence.

In an example, the enrichment score is specific to a single residue of the amino acid sequence of the target polypeptide. In an example, each enrichment score is specific to a different single residue of the amino acid sequence of the target polypeptide.

In an example, the method further comprises repeating steps (a)-(c) at least once (*e.g.*, once, twice, three times, four times, five times, six times, seven times, eight times, nine times, ten times, or more) with replicates of the plurality of the variants of the target polypeptide, and wherein step (c) further comprises omitting one or more promiscuous mutations, *e.g.*, mutations for which more than 50% of replicates had an enrichment score of greater than 30% and for which more than 75% of replicates had an enrichment score greater than 15%.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding one or more attractive constraints, optionally, wherein the attractive constraint is for a residue having an enrichment score greater than a first preselected value. In an example, the first preselected value is between 20% and 40%, *e.g.*, between 25% and 35%, *e.g.*, about 25%, about 30%, or about 35%. In an example, the attractive constraint comprises a linearly scaled bonus based on the enrichment score.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding a repulsive constraint for a residue having an enrichment score less than a second preselected value. In an example, the second preselected value is between 5% and 20%, *e.g.*, between 10% and 15%, *e.g.*, about 10%, about 12.5%, or about 15%.

In an example, step (d) comprises generating a docked pose between a model of the antibody molecule and a model of the target polypeptide. In an example, step (d) comprises generating a plurality of docked poses between a model of the antibody molecule and a model of the target polypeptide.

In an example, step (d) further comprises scoring the plurality of docked poses according to a docking algorithm, *e.g.*, SnugDock. In an example, step (d) further comprises selecting a subset of the plurality of docked poses having the highest scores, *e.g.*, the highest scoring 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300,

400, 500, 600, 700, 800, 900, 1000 or more docked poses. In an example, step (d) further comprises generating an ensemble docked pose using the selected subset of the plurality of docked poses, and setting the model of the antibody molecule and the model of the target polypeptide in accordance with the ensemble docked pose.

In an example, the model of the antibody molecule comprises an ensemble antibody homology model derived from a plurality of homology models of the antibody.

In an example, step (d) further comprises removing an antibody molecule-target polypeptide docketing model that exhibits a mode of engagement atypical for a known antibody-antigen complex, *e.g.*, according to a structural filter derived from antibody-antigen crystal structure.

In an example, step (d) comprises generating a plurality of antibody molecule-target polypeptide models.

In an example, step (e) comprises identifying a plurality of sites on the target polypeptide that is capable of being bound by the antibody molecule.

In an example, the site comprises or consists of one or more non-consecutive regions on the target polypeptide. In an example, the site comprises or consists of a consecutive region on the target polypeptide.

In another example, the disclosure features a method of identifying an epitope on a target polypeptide (*e.g.*, a target polypeptide described herein), the method comprising:

(a) generating an antibody-target polypeptide docking model, wherein the antibody-target polypeptide docking model is constrained according to a plurality of enrichment scores determined by a method comprising:

(i) binding an antibody molecule (*e.g.*, an antibody molecule described herein) to a plurality of variants of the target polypeptide,

(ii) obtaining (*e.g.*, enriching) a plurality of variants exhibiting altered (*e.g.*, reduced) binding to the antibody molecule, and

(iii) determining (*e.g.*, calculating) enrichment scores for each of the plurality of the enriched variants; and

(b) identifying a site on the target polypeptide that is capable of being bound by the antibody molecule based on the antibody-target polypeptide docking model;

thereby identifying an epitope on a target polypeptide.

In an example, the altered binding comprises altered binding affinity, *e.g.*, reduced binding affinity.

In an example, step (a)(i) comprises binding the antibody molecule to a library displaying a plurality of variants of the target polypeptide. In an example, step (a)(i) comprises binding the antibody molecule to a library comprising a plurality of cells expressing (*e.g.*, displaying) a plurality of variants of the target polypeptide. In an example, each of the plurality of cells expresses about

one distinct variant of the target polypeptide. In an example, the cell is a eukaryotic cell, *e.g.*, a yeast cell.

In an example, the plurality of variants comprise mutations on one or more surface residues of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of a selected surface residue of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of each of a plurality of selected surface residues of the target polypeptide.

In an example, the plurality of variants comprise single amino acid substitutions, relative to a wild-type amino acid sequence of the target polypeptide. In an example, each of the plurality of variants comprises a single amino acid substitution relative to a wild-type amino acid sequence of the target polypeptide. In an example, the single amino acid substitution occurs at a surface residue of the target polypeptide.

In an example, the altered (*e.g.*, reduced) binding comprises an alteration (*e.g.*, a reduction) of binding detected for the variant and the antibody molecule, relative to the binding detected for a wild-type target polypeptide and the antibody.

In an example, step (a)(ii) comprises obtaining (*e.g.*, enriching) variants exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by the wild-type target polypeptide.

In an example, step (a)(ii) comprises obtaining (*e.g.*, enriching) cells exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a cell comprising a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by a cell comprising the wild-type target polypeptide.

In an example, step (a)(ii) comprises performing one or more, *e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more, enrichments for variants exhibiting reduced binding to the antibody molecule.

In an example, the method further comprises, *e.g.*, prior to step (a)(iii), identifying the variants exhibiting altered (*e.g.*, reduced) binding to the antibody molecule, *e.g.*, by sequencing the genes encoding the variants, *e.g.*, by next-generation sequencing.

In an example, step (a)(iii) comprises determining the frequency of occurrence for each of the plurality of the obtained (*e.g.*, enriched) variants. In an example, step (a)(iii) further

comprises aggregating the frequency of occurrence of each variant comprising a distinct mutation at a particular residue and/or weighting (*e.g.*, heavily weighting) variants with higher frequencies of occurrence.

In an example, the enrichment score is specific to a single residue of the amino acid sequence of the target polypeptide. In an example, each enrichment score is specific to a different single residue of the amino acid sequence of the target polypeptide.

In an example, the method further comprises repeating steps (a)(i)-(a)(iii) at least once (*e.g.*, once, twice, three times, four times, five times, six times, seven times, eight times, nine times, ten times, or more) with replicates of the plurality of the variants of the target polypeptide, and wherein step (a)(iii) further comprises omitting one or more promiscuous mutations, *e.g.*, mutations for which more than 50% of replicates had an enrichment score of greater than 30% and for which more than 75% of replicates had an enrichment score greater than 15%.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding one or more attractive constraints, optionally, wherein the attractive constraint is for a residue having an enrichment score greater than a first preselected value. In an example, the first preselected value is between 20% and 40%, *e.g.*, between 25% and 35%, *e.g.*, about 25%, about 30%, or about 35%. In an example, the attractive constraint comprises a linearly scaled bonus based on the enrichment score.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding a repulsive constraint for a residue having an enrichment score less than a second preselected value. In an example, the second preselected value is between 5% and 20%, *e.g.*, between 10% and 15%, *e.g.*, about 10%, about 12.5%, or about 15%.

In an example, step (a) comprises generating a docked pose between a model of the antibody molecule and a model of the target polypeptide. In an example, step (a) comprises generating a plurality of docked poses between a model of the antibody molecule and a model of the target polypeptide.

In an example, step (a) further comprises scoring the plurality of docked poses according to a docking algorithm, *e.g.*, SnugDock. In an example, step (a) further comprises selecting a subset of the plurality of docked poses having the highest scores, *e.g.*, the highest scoring 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more docked poses. In an example, step (a) further comprises generating an ensemble docked pose using the selected subset of the plurality of docked poses, and setting the model of the antibody molecule and the model of the target polypeptide in accordance with the ensemble docked pose.

In an example, the model of the antibody molecule comprises an ensemble antibody homology model derived from a plurality of homology models of the antibody.

In an example, step (a) further comprises removing an antibody molecule-target polypeptide docketing model that exhibits a mode of engagement atypical for a known antibody-antigen complex, *e.g.*, according to a structural filter derived from antibody-antigen crystal structure.

In an example, step (a) comprises generating a plurality of antibody molecule-target polypeptide models.

In an example, step (b) comprises identifying a plurality of sites on the target polypeptide that is capable of being bound by the antibody molecule.

In an example, the site comprises or consists of one or more non-consecutive regions on the target polypeptide. In an example, the site comprises or consists of a consecutive region on the target polypeptide.

In yet another aspect, the disclosure features a method of identifying a paratope on an antibody molecule, the method comprising:

(a) binding the antibody molecule to a plurality of variants of the target polypeptide;

(b) obtaining (*e.g.*, enriching) a plurality of variants exhibiting reduced binding to the antibody molecule;

(c) determining (*e.g.*, calculating) enrichment scores for each of the plurality of the enriched variants;

(d) generating an antibody molecule-target polypeptide docking model, wherein the antibody-target polypeptide docking model is constrained according to the enrichment scores; and

(e) identifying one or more sites on the antibody molecule that is capable of being bound by the target polypeptide based on the antibody-target polypeptide docking model;

thereby identifying a paratope on an antibody molecule.

In an example, the altered binding comprises altered binding affinity, *e.g.*, reduced binding affinity.

In an example, step (a) comprises binding the antibody molecule to a library displaying a plurality of variants of the target polypeptide. In an example, step (a) comprises binding the antibody molecule to a library comprising a plurality of cells expressing (*e.g.*, displaying) a plurality of variants of the target polypeptide. In an example, each of the plurality of cells expresses about one distinct variant of the target polypeptide. In an example, the cell is a eukaryotic cell, *e.g.*, a yeast cell.

In an example, the plurality of variants comprise mutations on one or more surface residues of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of a selected surface residue of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of each of a plurality of selected surface residues of the target polypeptide.

In an example, the plurality of variants comprise single amino acid substitutions, relative to a wild-type amino acid sequence of the target polypeptide. In an example, each of the plurality of variants comprises a single amino acid substitution relative to a wild-type amino acid sequence of the target polypeptide. In an example, the single amino acid substitution occurs at a surface residue of the target polypeptide.

In an example, the altered (*e.g.*, reduced) binding comprises an alteration (*e.g.*, a reduction) of binding detected for the variant and the antibody molecule, relative to the binding detected for a wild-type target polypeptide and the antibody.

In an example, step (b) comprises obtaining (*e.g.*, enriching) variants exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by the wild-type target polypeptide.

In an example, step (b) comprises obtaining (*e.g.*, enriching) cells exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a cell comprising a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by a cell comprising the wild-type target polypeptide.

In an example, step (b) comprises performing one or more, *e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more, enrichments for variants exhibiting reduced binding to the antibody molecule.

In an example, the method further comprises, *e.g.*, prior to step (c), identifying the variants exhibiting altered (*e.g.*, reduced) binding to the antibody molecule, *e.g.*, by sequencing the genes encoding the variants, *e.g.*, by next-generation sequencing.

In an example, step (c) comprises determining the frequency of occurrence for each of the plurality of the obtained (*e.g.*, enriched) variants. In an example, step (c) further comprises aggregating the frequency of occurrence of each variant comprising a distinct mutation at a particular residue and/or weighting (*e.g.*, heavily weighting) variants with higher frequencies of occurrence.

In an example, the enrichment score is specific to a single residue of the amino acid sequence of the target polypeptide. In an example, each enrichment score is specific to a different single residue of the amino acid sequence of the target polypeptide.

In an example, the method further comprises repeating steps (a)-(c) at least once (*e.g.*, once, twice, three times, four times, five times, six times, seven times, eight times, nine times, ten

times, or more) with replicates of the plurality of the variants of the target polypeptide, and wherein step (c) further comprises omitting one or more promiscuous mutations, *e.g.*, mutations for which more than 50% of replicates had an enrichment score of greater than 30% and for which more than 75% of replicates had an enrichment score greater than 15%.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding one or more attractive constraints, optionally, wherein the attractive constraint is for a residue having an enrichment score greater than a first preselected value. In an example, the first preselected value is between 20% and 40%, *e.g.*, between 25% and 35%, *e.g.*, about 25%, about 30%, or about 35%. In an example, the attractive constraint comprises a linearly scaled bonus based on the enrichment score.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding a repulsive constraint for a residue having an enrichment score less than a second preselected value. In an example, the second preselected value is between 5% and 20%, *e.g.*, between 10% and 15%, *e.g.*, about 10%, about 12.5%, or about 15%.

In an example, step (d) comprises generating a docked pose between a model of the antibody molecule and a model of the target polypeptide. In an example, step (d) comprises generating a plurality of docked poses between a model of the antibody molecule and a model of the target polypeptide.

In an example, step (d) further comprises scoring the plurality of docked poses according to a docking algorithm, *e.g.*, SnugDock. In an example, step (d) further comprises selecting a subset of the plurality of docked poses having the highest scores, *e.g.*, the highest scoring 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more docked poses. In an example, step (d) further comprises generating an ensemble docked pose using the selected subset of the plurality of docked poses, and setting the model of the antibody molecule and the model of the target polypeptide in accordance with the ensemble docked pose.

In an example, the model of the antibody molecule comprises an ensemble antibody homology model derived from a plurality of homology models of the antibody.

In an example, step (d) further comprises removing an antibody molecule-target polypeptide docketing model that exhibits a mode of engagement atypical for a known antibody-antigen complex, *e.g.*, according to a structural filter derived from antibody-antigen crystal structure.

In an example, step (d) comprises generating a plurality of antibody molecule-target polypeptide models.

In an example, step (e) comprises identifying a plurality of sites on the antibody molecule that is capable of being bound by the target polypeptide.

In an example, the site comprises or consists of one or more non-consecutive regions on the antibody molecule. In an example, the site comprises or consists of a consecutive region on the antibody molecule.

In still another example, the disclosure features a method of identifying a paratope on an antibody, the method comprising:

(a) generating an antibody-target polypeptide docking model, wherein the antibody-target polypeptide docking model is constrained according to a plurality of enrichment scores determined (*e.g.*, calculated) by a method comprising:

(i) binding the antibody to a plurality of variants of the target polypeptide,

(ii) obtaining (*e.g.*, enriching) variants exhibiting reduced binding to the antibody molecule, and

(iii) determining (*e.g.*, calculating) an enrichment score for each of the plurality of the obtained (*e.g.*, enriched) variants; and

(b) identifying one or more sites on the antibody molecule that is capable of being bound by the target polypeptide based on the antibody-target polypeptide docking model;

thereby identifying a paratope on a target polypeptide.

In an example, the altered binding comprises altered binding affinity, *e.g.*, reduced binding affinity.

In an example, step (a)(i) comprises binding the antibody molecule to a library displaying a plurality of variants of the target polypeptide. In an example, step (a)(i) comprises binding the antibody molecule to a library comprising a plurality of cells expressing (*e.g.*, displaying) a plurality of variants of the target polypeptide. In an example, each of the plurality of cells expresses about one distinct variant of the target polypeptide. In an example, the cell is a eukaryotic cell, *e.g.*, a yeast cell.

In an example, the plurality of variants comprise mutations on one or more surface residues of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of a selected surface residue of the target polypeptide. In an example, the plurality of variants comprise distinct mutations of each of a plurality of selected surface residues of the target polypeptide.

In an example, the plurality of variants comprise single amino acid substitutions, relative to a wild-type amino acid sequence of the target polypeptide. In an example, each of the plurality of variants comprises a single amino acid substitution relative to a wild-type amino acid sequence of the target polypeptide. In an example, the single amino acid substitution occurs at a surface residue of the target polypeptide.

In an example, the altered (*e.g.*, reduced) binding comprises an alteration (*e.g.*, a reduction) of binding detected for the variant and the antibody molecule, relative to the binding detected for a wild-type target polypeptide and the antibody.

In an example, step (a)(ii) comprises obtaining (*e.g.*, enriching) variants exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by the wild-type target polypeptide.

In an example, step (a)(ii) comprises obtaining (*e.g.*, enriching) cells exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a cell comprising a wild-type target polypeptide. In an example, the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by a cell comprising the wild-type target polypeptide.

In an example, step (a)(ii) comprises performing one or more, *e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more, enrichments for variants exhibiting reduced binding to the antibody molecule.

In an example, the method further comprises, *e.g.*, prior to step (a)(iii), identifying the variants exhibiting altered (*e.g.*, reduced) binding to the antibody molecule, *e.g.*, by sequencing the genes encoding the variants, *e.g.*, by next-generation sequencing.

In an example, step (a)(iii) comprises determining the frequency of occurrence for each of the plurality of the obtained (*e.g.*, enriched) variants. In an example, step (a)(iii) further comprises aggregating the frequency of occurrence of each variant comprising a distinct mutation at a particular residue and/or weighting (*e.g.*, heavily weighting) variants with higher frequencies of occurrence.

In an example, the enrichment score is specific to a single residue of the amino acid sequence of the target polypeptide. In an example, each enrichment score is specific to a different single residue of the amino acid sequence of the target polypeptide.

In an example, the method further comprises repeating steps (a)(i)-(a)(iii) at least once (*e.g.*, once, twice, three times, four times, five times, six times, seven times, eight times, nine times, ten times, or more) with replicates of the plurality of the variants of the target polypeptide, and wherein step (a)(iii) further comprises omitting one or more promiscuous mutations, *e.g.*, mutations for which more than 50% of replicates had an enrichment score of greater than 30% and for which more than 75% of replicates had an enrichment score greater than 15%.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding one or more attractive constraints, optionally, wherein the attractive constraint is for a residue having an enrichment score greater than a first preselected value. In an example, the first preselected value is between 20% and 40%, *e.g.*, between 25% and 35%, *e.g.*, about 25%, about 30%, or about 35%. In an example, the attractive constraint comprises a linearly scaled bonus based on the enrichment score.

In an example, the antibody molecule-target polypeptide docking model is constrained by adding a repulsive constraint for a residue having an enrichment score less than a second preselected value. In an example, the second preselected value is between 5% and 20%, *e.g.*, between 10% and 15%, *e.g.*, about 10%, about 12.5%, or about 15%.

In an example, step (a) comprises generating a docked pose between a model of the antibody molecule and a model of the target polypeptide. In an example, step (a) comprises generating a plurality of docked poses between a model of the antibody molecule and a model of the target polypeptide.

In an example, step (a) further comprises scoring the plurality of docked poses according to a docking algorithm, *e.g.*, SnugDock. In an example, step (a) further comprises selecting a subset of the plurality of docked poses having the highest scores, *e.g.*, the highest scoring 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more docked poses. In an example, step (a) further comprises generating an ensemble docked pose using the selected subset of the plurality of docked poses, and setting the model of the antibody molecule and the model of the target polypeptide in accordance with the ensemble docked pose.

In an example, the model of the antibody molecule comprises an ensemble antibody homology model derived from a plurality of homology models of the antibody.

In an example, step (a) further comprises removing an antibody molecule-target polypeptide docketing model that exhibits a mode of engagement atypical for a known antibody-antigen complex, *e.g.*, according to a structural filter derived from antibody-antigen crystal structure.

In an example, step (a) comprises generating a plurality of antibody molecule-target polypeptide models.

In an example, step (b) comprises identifying a plurality of sites on the target polypeptide that is capable of being bound by the antibody molecule.

In an example, the site comprises or consists of one or more non-consecutive regions on the target polypeptide. In an example, the site comprises or consists of a consecutive region on the target polypeptide.

What is claimed is:

1. A method of identifying an epitope on a target polypeptide, the method comprising:

(a) binding an antibody molecule to a plurality of variants of the target polypeptide;

(b) obtaining (*e.g.*, enriching) a plurality of variants exhibiting reduced binding (*e.g.*, reduced binding affinity) to the antibody molecule;

(c) determining (*e.g.*, calculating) an enrichment score for each of the plurality of the obtained (*e.g.*, enriched) variants;

(d) generating an antibody molecule-target polypeptide docking model, wherein the antibody molecule-target polypeptide docking model is constrained according to the enrichment scores; and

(e) identifying a site on the target polypeptide that is capable of being bound by the antibody molecule based on the antibody molecule-target polypeptide docking model;

thereby identifying an epitope on a target polypeptide.

2. A method of identifying a paratope on an antibody molecule, the method comprising:

(a) binding the antibody molecule to a plurality of variants of the target polypeptide;

(b) obtaining (*e.g.*, enriching) a plurality of variants exhibiting reduced binding to the antibody molecule;

(c) determining (*e.g.*, calculating) enrichment scores for each of the plurality of the enriched variants;

(d) generating an antibody molecule-target polypeptide docking model, wherein the antibody-target polypeptide docking model is constrained according to the enrichment scores; and

(e) identifying one or more sites on the antibody molecule that is capable of being bound by the target polypeptide based on the antibody-target polypeptide docking model;

thereby identifying a paratope on an antibody molecule.

3. The method of claim 1 or 2, wherein step (a) comprises:

(i) binding the antibody molecule to a library displaying a plurality of variants of the target polypeptide; and/or

(ii) binding the antibody molecule to a library comprising a plurality of cells expressing (*e.g.*, displaying) a plurality of variants of the target polypeptide.

4. The method of claim 3, wherein:

(i) each of the plurality of cells expresses about one distinct variant of the target polypeptide and/or

(ii) the cell is a eukaryotic cell, *e.g.*, a yeast cell.

5. The method of any of the preceding claims, wherein the plurality of variants comprise:

(i) mutations on one or more surface residues of the target polypeptide;

(ii) distinct mutations of a selected surface residue of the target polypeptide;

(iii) distinct mutations of each of a plurality of selected surface residues of the target polypeptide; and/or

(iv) single amino acid substitutions, relative to a wild-type amino acid sequence of the target polypeptide.

6. The method of any of the preceding claims, wherein each of the plurality of variants comprises a single amino acid substitution relative to a wild-type amino acid sequence of the target polypeptide.

7. The method of claim 5 or 6, wherein the single amino acid substitution occurs at a surface residue of the target polypeptide.

8. The method of any of the preceding claims, wherein the reduced binding comprises a reduction of binding detected for the variant and the antibody molecule, relative to the binding detected for a wild-type target polypeptide and the antibody.

9. The method of any of the preceding claims, wherein step (b) comprises obtaining (*e.g.*, enriching) variants exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a wild-type target polypeptide;

optionally wherein the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by the wild-type target polypeptide.

10. The method of any of the preceding claims, wherein step (b) comprises obtaining (*e.g.*, enriching) cells exhibiting less than about 80% (*e.g.*, less than about 0.01%, 0.1%, 1%, 2%, 3%, 4%, 5%, 6%, 7%, 8%, 9%, 10%, 20%, 30%, 40%, 50%, 60%, 70%, or 80%) of the binding to the antibody molecule exhibited by a cell comprising a wild-type target polypeptide;

optionally wherein the reduced binding is at least about 20% (*e.g.*, at least about 20%, 21%, 22%, 23%, 24%, 25%, 26%, 27%, 28%, 29%, 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 95%, or 100%) of the binding exhibited by a cell comprising the wild-type target polypeptide.

11. The method of any of the preceding claims, wherein step (b) comprises performing one or more, *e.g.*, two, three, four, five, six, seven, eight, nine, ten, or more, enrichments for variants exhibiting reduced binding to the antibody molecule.

12. The method of any of the preceding claims, further comprising, *e.g.*, prior to step (c), identifying the variants exhibiting reduced binding to the antibody molecule, *e.g.*, by sequencing the genes encoding the variants, *e.g.*, by next-generation sequencing.

13. The method of any of the preceding claims, wherein step (c) comprises determining the frequency of occurrence for each of the plurality of the obtained (*e.g.*, enriched) variants;

optionally wherein step (c) further comprises aggregating the frequency of occurrence of each variant comprising a distinct mutation at a particular residue and/or heavily weighting variants with higher frequencies of occurrence.

14. The method of any of the preceding claims, wherein:

(i) the enrichment score is specific to a single residue of the amino acid sequence of the target polypeptide; and/or

(ii) each enrichment score is specific to a different single residue of the amino acid sequence of the target polypeptide.

15. The method of any of the preceding claims, further comprising repeating steps (a)-(c) at least once (*e.g.*, once, twice, three times, four times, five times, or more) with replicates of the plurality of the variants of the target polypeptide, and wherein step (c) further comprises omitting one or more promiscuous mutations, *e.g.*, mutations for which more than 50% of replicates had an enrichment score of greater than 30% and for which more than 75% of replicates had an enrichment score greater than 15%.

16. The method of any of the preceding claims, wherein the antibody molecule-target polypeptide docking model is constrained by adding one or more attractive constraints, wherein the attractive constraint is for a residue having an enrichment score greater than a first preselected value;

optionally wherein the first preselected value is between 20% and 40%, *e.g.*, between 25% and 35%, *e.g.*, about 30%.

17. The method of claim 24 or 25, wherein the attractive constraint comprises a linearly scaled bonus based on the enrichment score.

18. The method of any of the preceding claims, wherein the antibody molecule-target polypeptide docking model is constrained by adding a repulsive constraint for a residue having an enrichment score less than a second preselected value;

optionally wherein the second preselected value is between 5% and 20%, *e.g.*, between 10% and 15%, *e.g.*, about 12.5%.

19. The method of any of the preceding claims, wherein step (d) comprises:

(i) generating a docked pose between a model of the antibody molecule and a model of the target polypeptide; and/or

(ii) generating a plurality of docked poses between a model of the antibody molecule and a model of the target polypeptide.

20. The method of claim 19, wherein step (d) further comprises scoring the plurality of docked poses according to a docking algorithm, *e.g.*, SnugDock.

21. The method of claim 31, wherein step (d) further comprises selecting a subset of the plurality of docked poses having the highest scores, *e.g.*, the highest scoring 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100, 110, 120, 130, 140, 150, 160, 170, 180, 190, 200, 300, 400, 500, 600, 700, 800, 900, 1000 or more docked poses;

optionally wherein step (d) further comprises generating an ensemble docked pose using the selected subset of the plurality of docked poses, and setting the model of the antibody molecule and the model of the target polypeptide in accordance with the ensemble docked pose.

22. The method of any of claims 19-21, wherein the model of the antibody molecule comprises an ensemble antibody homology model derived from a plurality of homology models of the antibody.

23. The method of any of the preceding claims, wherein step (d) further comprises removing an antibody molecule-target polypeptide docketing model that exhibits a mode of engagement atypical for a known antibody-antigen complex, *e.g.*, according to a structural filter derived from antibody-antigen crystal structure.

24. The method of any of the preceding claims, wherein step (d) comprises:

(i) generating a plurality of antibody molecule-target polypeptide models; and/or

(ii) identifying a plurality of sites on the target polypeptide that is capable of being bound by the antibody molecule.

25. The method of claim 1, wherein the antibody-target polypeptide docking model is constrained according to a plurality of enrichment scores determined by a method comprising:

(i) binding the antibody molecule to a plurality of variants of the target polypeptide,

(ii) obtaining (*e.g.*, enriching) a plurality of variants exhibiting reduced binding to the antibody molecule, and

(iii) determining (*e.g.*, calculating) enrichment scores for each of the plurality of the enriched variants.

26. The method of claim 2, wherein the antibody-target polypeptide docking model is constrained according to a plurality of enrichment scores determined (*e.g.*, calculated) by a method comprising:

(i) binding the antibody to a plurality of variants of the target polypeptide,

(ii) obtaining (*e.g.*, enriching) variants exhibiting reduced binding to the antibody molecule, and

(iii) determining (*e.g.*, calculating) an enrichment score for each of the plurality of the obtained (*e.g.*, enriched) variants.

27. An antibody molecule for which the epitope on a target polypeptide or the paratope on the antibody molecule for the target polypeptide is identified according to the method of any of the preceding claims.

28. A nucleic acid molecule encoding one or more chains (*e.g.*, VH and/or VL) of the antibody molecule of claim 27.

29. A vector comprising the nucleic acid molecule of claim 28.

30. A host cell comprising the nucleic acid molecule of claim 28 or the vector of claim 29.

31. A method of making an antibody molecule, comprising culturing the host cell of claim 30 under conditions suitable for expression of the antibody molecule.