UNIVERSITÀ DEGLI STUDI DI TRIESTE

Antibody Affinity Maturation Using Computational Methods: From an Initial Hit to Small-Scale Expression of Optimized Binders

This is the peer reviewed (post-print) version of the following article:

Original
Antibody Affinity Maturation Using Computational Methods: From an Initial Hit to Small-Scale Expression of Optimized Binders / Medagli, B., Soler, M.A., De Zorzi, R., Fortuna, S.. - 2552:(2023), pp. 333-359. [10.1007/978-1-0716-2609-2_19]

Availability:
This version is available at: 11368/3055241 since: 2023-08-21T07:41:53Z

Publisher:
Springer

Published
DOI:10.1007/978-1-0716-2609-2_19

Terms of use:

Some rights reserved. More details in the PostPrint rights section below.

License:
Digital Rights Management not defined

Publisher copyright

-.-

(Article begins on next page)

22 October 2026

arTs Università degli Studi di Trieste
Archivio della ricerca – postprint

Antibody Affinity Maturation Using Computational Methods: From an Initial Hit to Small-Scale Expression of Optimized Binders

Barbara Medagli, Miguel A. Soler, Rita De Zorzi, and Sara Fortuna

Abstract

Nanobodies (VHHs) are engineered fragments of the camelid single-chain immunoglobulins. The VHH domain contains the highly variable segments responsible for antigen recognition. VHHs can be easily produced as recombinant proteins. Their small size is a good advantage for in silico approaches. Computer methods represent a valuable strategy for the optimization and improvement of their binding affinity. They also allow for epitope selection offering the possibility to design new VHHs for regions of a target protein that are not naturally immunogenic. Here we present an in silico mutagenic protocol developed to improve the binding affinity of nanobodies together with the first step of their in vitro production. The method, already proven successful in improving the low Kd of a nanobody hit obtained by panning, can be employed for the *ex novo* design of antibody fragments against selected protein target epitopes.

Key words Nanobodies (VHH), In silico protocol, Antigen, Binding affinity

1 Introduction

Camelids have a modified form of immunoglobulin, consisting only in the heavy chains (Fig. 1a); their variable fragment (VHH or nanobody; Fig. 1b) ranges from 110 to 150 residues with a conserved immunoglobulin structure, 4 sequence-conserved frameworks, and 3 variable loops or complementary-determining regions (CDRs). The CDR3, in particular, has a variable length conferring each VHH its unique feature and ability to reach active sites and concave regions of target proteins [1–3].

The stability, hydrophilicity, small size (12–15 kDa), and easy production as recombinant proteins make the VHHs good candidates and a viable alternative to the canonical antibodies. In the last two decades, considerable efforts have been undertaken to explore the VHH potential in fields like research, biotechnology, and medicine—both for diagnostics and therapy [4].

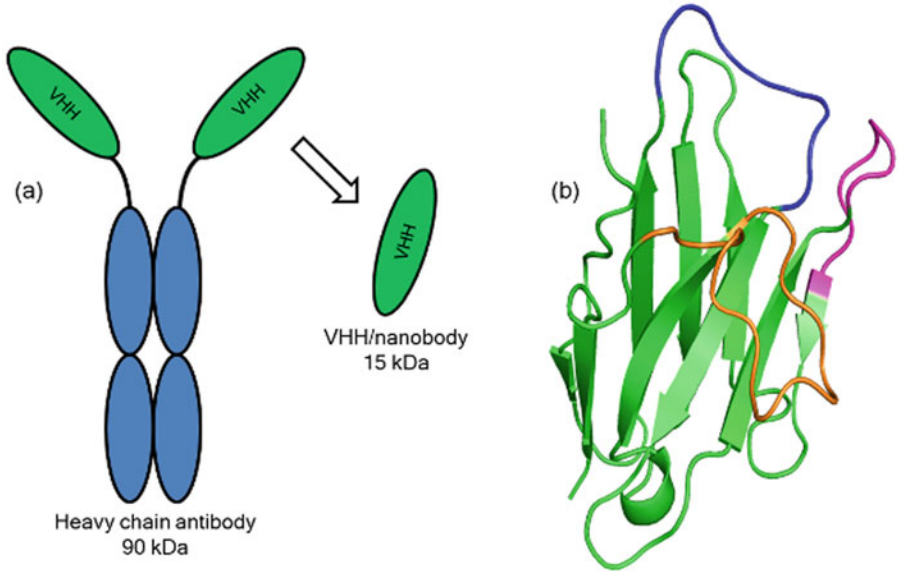

Fig. 1 Nanobody representation. (**a**) Schematic representation of the heavy chain antibody and the variable domain (VHH); (**b**) VHH-D9 structure obtained by homology modelling. VHH framework (green), CDR1 (blue), CDR2 (magenta), and CDR3 (orange)

The best hits can be identified among a library of VHH nanobodies of immune, naïve, or synthetic origins. Immune libraries require animal immunization with the target, cloning of genes corresponding to the VHH domain from peripheral blood lymphocytes, and screening of nanobodies by phage display. When immunization is not successful, e.g., due to a toxic or non-immunogenic target, a naïve pool of camelid antibodies is prepared collecting the peripheral blood lymphocytes of a large number of animals, obtaining a diverse library to be screened against the target by phage display [5]. The synthetic method does not require animal handling, representing a big advantage in terms of cost, time, and space. Synthetic libraries are generated by the accumulation of random mutations on pre-selected nanobodies and the characterization of the entire library considering nanobody stability, specificity, or binding affinity to the antigen [6, 7]. In this context, an emerging number of computational approaches can be exploited to speed up the identification of the best candidate, from the selection of an initial hit to its subsequent optimization [7–10].

Our joint computational/experimental protocol (schematized in Fig. 2) can be used to improve the binding affinity of an initial hit. The protocol for the in silico maturation of the VHH (described below and published in ref. 11) is based on two steps, iteratively repeated: first, a random mutation in the CDR regions is introduced in the VHH, and then the affinity of the mutant toward the target is evaluated and compared with that generated in the

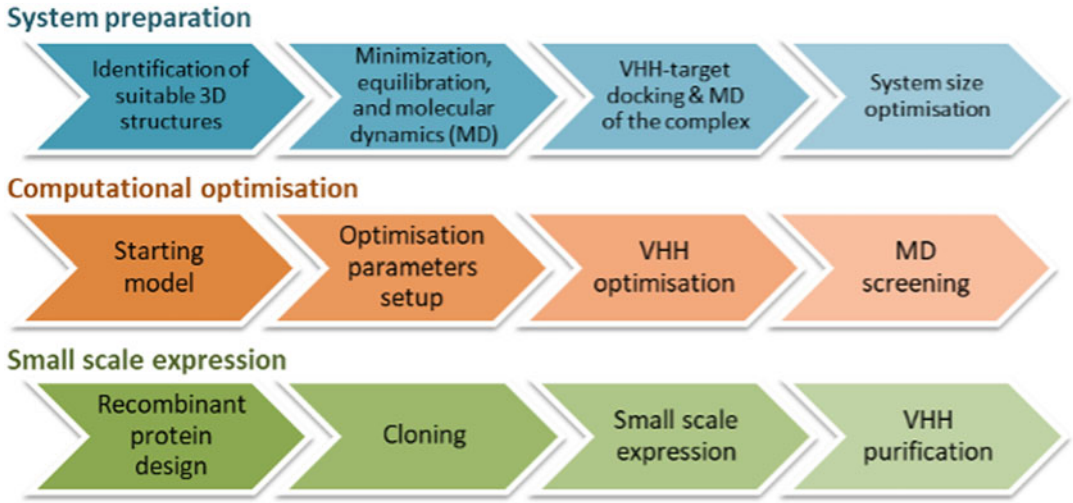

Fig. 2 The procedure covered by this tutorial: from a knowledge of the sequence of an initial hit binder and its target to the small-scale expression of optimized binders

previous step. The algorithm, initially developed for the design of drug-capturing peptides [12–15], was proven successful also for the design of protein binding peptides [16–19]. A further improvement of the protocol allows for the optimization of any peptide-based system by selecting the specific portion where mutations should be introduced [11, 19]. In the subsequent in silico selection process, a set of molecular dynamics (MD) simulations lead to the identification of suitable VHH–target complexes through the evaluation of a set of scoring functions that rank the binding free energy of the generated VHH–target complexes. This second step takes into account also the stability of the new VHHs and their propensity toward self-aggregation or oligomerization [7, 11]. Selected candidates are eventually produced as recombinant proteins for their biophysical characterization and Kd experimental measurement.

Following the presentation of the basic ideas behind the computational method [11], this paper represents a step by step tutorial covering in detail the procedure employed for the optimization of a nanobody, further referred to as VHH-D9, known to bind with low affinity the extracellular domain of the HER2 receptor (HER2-ECD). Codes for computational optimization and screening are reported in Subheading 2.1. Subheading 2.2 covers the materials required for VHH cloning and small-scale expression. As a suitable input, VHH–target complex is crucial to start the optimization procedure; details of its selection and construction are given in Subheading 3.1 (System Preparation). After VHH optimization (Subheading 3.2), selected candidates undergo a computational screening (Subheading 3.3), the best selected for

in vitro production and characterization. Aiming at the production of soluble proteins for further investigation, the constructs are designed (Subheading 3.4) and cloned in a suitable expression vector (Subheading 3.5), and different expression conditions are tested (Subheading 3.6). The whole process is schematically represented in Fig. 2.

2 Materials

2.1 Algorithms

1. BINDesignER [11] is the computational protocol for in silico optimization of antibody fragments. It is implemented in a BASH script and is freely available at https://github.com/migsoler/BINDesignER. As an alternative, the code can be run as a Python script, known as PARCE (Protocol for Amino acid Refinement through Computational Evolution) [19]. In this tutorial, we employed the BASH-based implementation BINDesignER.
2. GROMACS [20] is used for structure minimization and MD simulations. Any version of GROMACS can be used inside BINDesignER by specifying command lines and paths of GROMACS as parameters at the beginning of the design program. The GROMACS package is freely available for download: http://manual.gromacs.org/documentation/.
3. SCWRL4 [21] is employed for the reconstruction of side chains of the mutated residues. Free download is available under Academic License: http://dunbrack.fccc.edu/scwrl4/.
4. Scoring functions [7] are employed to evaluate the binding affinity of the protein–protein complexes. A set of scoring functions should be chosen, installed, and benchmarked (*see* **Note 1**). In the example described here, we employed, Irad, Pie*Pisa, Bluues, Haddock, Bach6, and FireDock [22–26].

2.2 Cloning and Protein Expression

Prepare all solutions using ultrapure water (purified deionized water to a resistivity greater than 18.2 MΩ·cm at 25 °C). Prepare and store all reagents at room temperature (unless otherwise noted). Steps from cloning to small-scale purification require sterile equipment; all media for cell culture are sterilized by 20 min autoclave cycle at 121 °C and 15 psi. Powders are handled according to safety procedures specified by producers.

Enzymes, primers, and nucleotides: Enzymes are used according to the manufacturers' recommendations and stored at −20 °C.

5. NcoI DNA endonuclease and 10× reaction buffer.
6. NotI DNA endonuclease and 10× reaction buffer.
7. T4 DNA ligase and 10× reaction buffer.

8. Taq DNA polymerase and 10× buffer.
9. 10 mM dNTP mix or 100 mM dATP, dGTP, dTTP, dCTP: 10 mM dNTP mixes are prepared in small aliquots in sterile water and stored frozen, to avoid spontaneous hydrolysis.
10. 100 μM T7 promoter primer: 5′- TAATACGACTCACTA TAGGG-3′ in water.
11. 100 μM T7 terminator primer: 5′-GCTAGTTATTGCTCAG CGG-3′ in water.

DNA preparation and cleaning kit: All kits are used and stored according to manufacturers' recommendations.

12. DNA miniprep kit.
13. DNA gel extraction kit.

Reagents and stock solutions: Prepare 1 L volume for all solutions, unless differently reported. Powders are appropriately weighed in a disposable boat or glass beaker and dissolved in 700 ml of water (or 98% EtOH where indicated). For each solution the desired pH is reached at 25 °C by addition of 5 M NaOH or 6 M HCl, using a pH meter in continuous measuring mode. Solutions are transferred in 1 L measuring cylinders and water is added to reach the final volume.

14. Agarose powder.
15. DNA gel staining 10,000–20,000× in DMSO.
16. 50× TAE: 2 M Tris, 1 M acetic acid, 50 mM EDTA. The working concentration is 1×.
17. DNA ladder, used and stored as recommended by the producer.
18. 6× DNA loading buffer: 30% (v/v) glycerol. 0.25% (w/v) bromophenol blue. 0.25% (w/v) xylene cyanol FF. Stored at 4 °C. Working concentration is 1×.
19. 1000× ampicillin: Dissolve 1 g of ampicillin powder in 8 ml of sterile water; add water to reach a volume of 10 ml (100 mg/ml). Divide the solution in 1 ml aliquots, stored at −20 °C.
20. 1000× chloramphenicol: Dissolve 340 mg of chloramphenicol in 8 ml of 98% EtOH; add EtOH to reach a volume of 10 ml (34 mg/ml), stored at −20 °C.
21. 0.5 M IPTG: Dissolve 1.19 g of powder in 8 ml of water; add water to reach 10 ml volume. Filter the solution with 0.22 μm PVDF filter for syringe, divide it in 1 ml aliquots, and store at −20 °C.
22. 25% Arabinose (w/v): Dissolve 2.5 g of powder in 8 ml of water; add water to reach 10 ml; filter with 0.22 μm PVDF filter for syringe.

23. 60% glycerol (v/v): Prepare and sterilize a 60% glycerol solution in water.
24. Ni-NTA agarose resin in 20% EtOH, 50% solution.
25. 100 mM PMSF: Dissolve 174 mg of powder in 8 ml of 98% EtOH; add EtOH to reach a volume of 10 ml; store at 4 °C.
26. 2× nickel buffer: 40 mM Tris, pH 7.9, 1 M NaCl, 10% glycerol (v/v); store at 4 °C. The working solution is 1×.
27. Lysozyme powder or 50 mg/ml lysozyme solution in nickel buffer; if a stock solution is prepared, divide it into 250 μl aliquots and store at −20 °C.
28. 5 mg/ml DNaseI: Prepare the solution following manufacturer's suggestions, divide it into 250 μl aliquots, and store at −20 °C.
29. 2 M $MgCl_2$.
30. Tween 20.
31. Triton X-100.
32. 1 M imidazole, pH 8, stored at 4 °C.
33. TGS 10×: 0.25 M Tris, 1.92 M glycine, and 1% SDS; the working concentration is 1×.
34. Protein ladder: Prepare the solution following manufacturer's suggestions, divide the solution in 25 μl aliquots, and store at −20 °C.
35. 4× SDS loading sample: 200 mM Tris–HCl (pH 6.8), 400 mM DTT, 8% SDS, 0.4% bromophenol blue, 40% glycerol. Divide in aliquots and store at −20 °C. The working concentration is 1×.
36. Buffer for SDS gel electrophoresis or precast SDS-PAGE.

E. coli Strains and DNA

37. DH5α competent cells or any general *recA*− strain.
38. BL21 (DE3) SOX: The strain contains a chloramphenicol-resistant plasmid carrying the sequences coding for prokaryotic disulfide bond isomerase (DsbC) and sulfhydryl oxidase, their expression triggered by the addition of arabinose during culture growth [27].
39. Expression vector: pET14-GFP-6H. The vector encodes for a GFP-6 His tag downstream the C-terminus of the protein of interest. The ampicillin resistance allows the selection of the positively transformed colony.
40. Synthetic genes with appropriate restriction enzymes and with codons optimized for *E. coli* expression.

The cell culture media: Prepared according to recipes from Helmholtz Center Munich Protocols or Cold Spring Harbor Protocols.

41. SOC broth: 2% tryptone, 0.5% yeast extract, 10 mM NaCl, 2.5 M KCl, 10 mM $MgCl_2$, 10 mM $MgSO_4$, and 20 mM glucose.
42. LB broth: 1% tryptone, 0.5% yeast extract, 0.5% NaCl.
43. LB agar: 1% tryptone, 0.5% yeast extract, 0.5% NaCl, 2% agar.
44. TB broth: 1.2% tryptone, 2.4% yeast extract, 0.5% glycerol, 17 mM M KH_2PO_4, 72 mM K_2HPO_4.
45. ZYM 5052 broth: 1% tryptone, 0.5% yeast extract, 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4 (1× M), 2 mM $MgSO_4$, 0.5% glycerol, 0.05% glucose, 0.2% α-lactose (1× 5052), 0.2× trace elements.
46. MDG broth: 25 mM Na_2HPO_4, 25 mM KH_2PO_4, 50 mM NH_4Cl, 5 mM Na_2SO_4 (1× M), 0.5% glucose, 0.25% aspartate, 2 mM $MgSO_4$, 0.2× trace elements.
47. 50× M: 1.25 M Na_2HPO_4, 1.25 M KH_2PO_4, 2.5 M NH_4Cl, 0.25 M Na_2SO_4.
48. 50× 5052: 25% glycerol, 2.5% glucose, 10% α-lactose.
49. 1000× trace metals: 50 mM $FeCl_3$, 20 mM $CaCl_2$, 10 mM $MnCl_2$, 10 mM $ZnSO_4$, 2 mM $CoCl_2$, 2 mM $CuCl_2$, 2 mM $NiCl_2$, 2 mM Na_2MoO_4, 2 mM Na_2SeO_3.
    500× $MgSO_4$: 1 M $MgSO_4$.

Equipment

50. 2, 10, 20, 200, 1000 μl pipette set.
51. Disposable scalpels.
52. Disposable spreaders.
53. 10 and 25 ml disposable serological pipette and pipettor.
54. 90 mm and/or 60 mm Petri dishes.
55. 50 ml conical tubes.
56. 1.5 and/or 2 ml conical tubes.
57. PCR strips and caps.
58. Horizontal and vertical gel casting and running apparatus.
59. Power supply.
60. Nanodrop or SimplyNano spectrophotometer.
61. PCR machine.
62. UV transilluminator.
63. Bench centrifuge with rotors for 2 ml and 50 ml conical tubes.
64. 10 ml 24-deep well plate.

65. Breathable sealer.
66. 15 ml conical tubes.
67. Incubator with an orbital shaker and cooling system; or shaking thermomixer with cooling system and SBS plate holder.
68. Cuvette spectrophotometer.
69. Disposable plastic cuvettes.
70. 50 ml flasks.

3 Methods

3.1 System Preparation

The accuracy of the starting structural model for the initial complex between the VHH and the epitope-containing target is a key aspect for a successful design (*see* **Notes 2–5**). The recommended option is to start from experimental structural models, if available, for both the VHH and its protein target. These can be downloaded from Protein Data Bank (PDB). If PDB structures are not available, models of the protein target, the VHH, or both can be prepared from a suitable template as in **steps 1** and **2**. Otherwise, these two steps can be skipped. In our example, HER2-ECD was modelled by extracting the relevant fragment from the 3N85 PDB structure [28], while the initial VHH was modelled by homology (*see* **Notes 2** and **3**). We applied the following protocol:

1. Identification of a suitable template for homology modelling. A search of the protein sequence is performed using BLAST against the Protein Data Bank. For a suitable VHH template: 80% sequence identity for the framework sequence, CDR loops of the same length, and a high-resolution structure.
2. Mutations in the selected template structure. The framework point mutations and the CDR insertions to obtain the VHH model from the template can be performed using DeepView—Swiss-PdbViewer 4.1 [29].
3. Minimization, equilibration, and molecular dynamics (MD). Each protein (both VHH and target) is placed in a cubic box with a water layer of 0.9 nm and Na^+ Cl^- ions to neutralize the system. In our protocol, the force field employed for all MD and minimization simulations was AMBER99SB-ILDN, with the TIP3P water model. The leapfrog integrator was used with a time step of 2 fs. The long-range electrostatics were calculated by means of the particle mesh Ewald summation. A cutoff of 0.9 nm was used for the electrostatic and van der Waals non-bonded interactions. All covalent bonds of the system were constrained by using LINCS algorithm. To keep constant the temperature in NVT and NPT simulations, the velocity rescaling thermostat was employed, while the isotropic Parrinello–

Rahman barostat at 1 bar was used to control the pressure in the NPT simulations. Before every production phase, a standard 100 ps NVT + 100 ps NPT equilibration protocol was performed. A 250 ns MD simulation of the protein structure in water was carried out using GROMACS (Subheading 2.1, **item 2**). Typical descriptors for analysis of protein structure stability are as follows: (i) the root mean square deviation (RMSD), which measures deviations from the template (in that case the structure at $t = 0$) along the MD trajectory; (ii) the root mean square fluctuation (RMSF), which shows the most fluctuating regions along the protein sequence; and (iii) the potential energy of the protein along the trajectory, which estimates the global stability of the protein secondary structure along the simulated time. Figure 1b shows the model prepared for VHH-D9. The general parameters described above were employed for all subsequent MD simulations. The use of specific parameters will be indicated where appropriate.

4. VHH–protein target docking. The initial protein–protein complex can be generated by a protein–protein docking algorithm such as the HADDOCK webserver "easy interface" [30]. MD simulation output structures for both the VHH and the target are used for the docking. A list of the active residues is required, namely, the residues expected to be at the interface of the VHH–target complex. For the VHH a good choice is represented by the list of residues of the three CDRs (highlighted in Fig. 1b). For our system, previous experiments revealed that VHH-D9 and the trastuzumab (an antibody used in breast cancer therapy) share the same binding site on the Her2-ECD [31] (*see* **Note 4**). Thus, we selected as active residues for VHH-D9 those that compose CDRs, while residues in contact with the trastuzumab in the crystal structure of the complex (579-583, 592-595, and 615-625) were chosen as active for HER2-ECD. The resulting complex is shown in Fig. 3a.
5. MD of the complex. The VHH–target complex structure is minimized and subsequently undergoes a 250 ns MD simulation in water as detailed in Subheading 3.1, **step 3**. To verify whether the complex structure achieves a stable binding conformation, standard observables should be monitored, e.g., root mean square deviation (RMSD), root mean square fluctuation (RMSF), and radius of gyration of both the complex and its constituents. Also, the interface root mean square deviation (IRMSD) calculated along the MD simulation trajectory offers a useful assessment of the simulation progresses (*see* **Note 5**). For instance, the IRMSD curve should show high values at the beginning of the simulation when the VHH target can undergo important rearrangements, while at equilibrium the

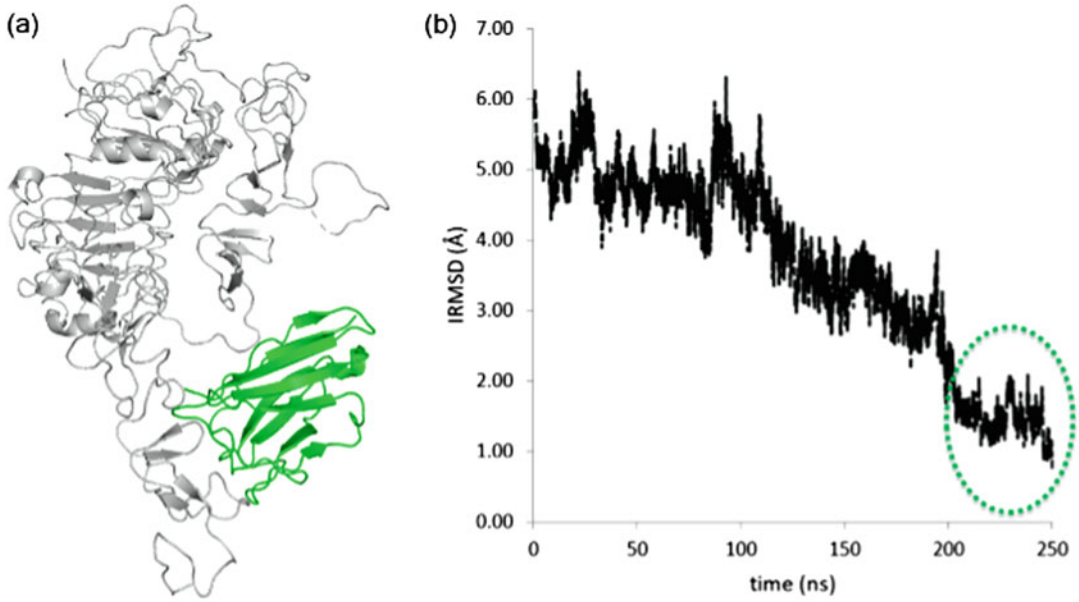

Fig. 3 VHH–target complex optimization: (**a**) docked VHH/target conformation—VHH-D9 (green), Her2-ECD (gray). (**b**) MD simulation over 250 ns of the same complex: the IRMSD fluctuations are related to the complex rearrangement; a green circle highlights the equilibrated binding conformations

IRMSD should reach low and stable values (*see* Fig. 3b). For the following calculations, a good option is to choose a starting conFigureuration of the complex from the pool of equilibrium conFigureurations of the MD simulation. Among these, the representative structure of the most populated cluster can be selected.

6. System size optimization. The computational cost of the VHH design strongly depends on the cost of each optimization step which includes a VHH mutation, the conformational sampling of the new VHH–target complex, and SF evaluation (*see* Fig. 4a). The rate-determining step of the process is the conformational sampling based on short MD runs. The most straightforward solution is to shrink the system size by cleaving the protein target, particularly when formed by different domains (*see* **Note 6**). For instance, in the illustrated example, we started from the whole Her2-ECD (Fig. 3a), but only the binding domain was used in the optimization (Fig. 4c–d). The conformational stability of the VHH–target complex after the cleavage must be tested by carrying out an additional MD simulation and confirming that IRMSD values remain stable along the trajectory.

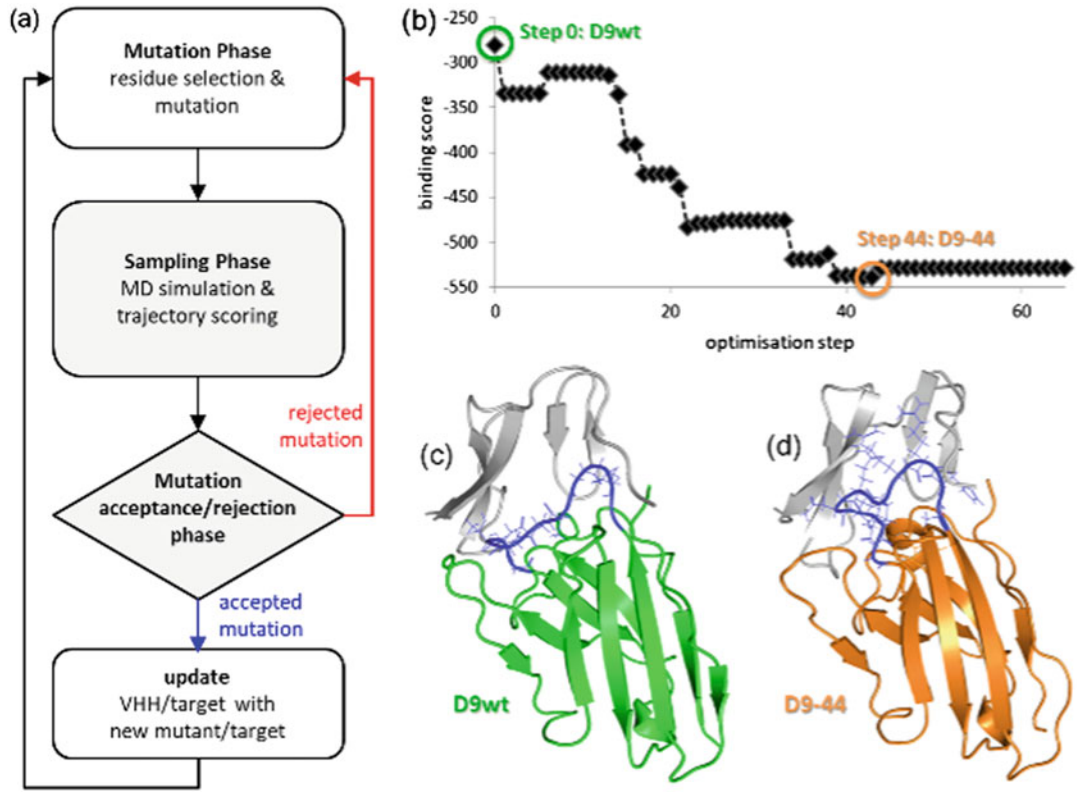

Fig. 4 VHH optimization protocol: (**a**) schematic representation of the design algorithm; (**b**) example of binding score profile obtained during the affinity optimization simulation with a consensus threshold at 3 of the (**c**) VHH D9 wt toward HER2-ECD until achieving the optimal binder, i.e., (**d**) VHH D9-44 with enhanced affinity to the protein target

3.2 In Silico *Optimization*

The protocol for the VHH optimization is an iterative evolutionary process (*see* Fig. 4a). Each iteration consists of the following phases:

- Mutation phase. A random mutation of a randomly selected residue within the CDR regions is performed. The side-chain reconstruction of the mutated residue is carried out using SCRWL4 (Subheading 2.1, **item 3**). Three consecutive minimizations and one equilibration are performed by GROMACS: (i) a local minimization of the mutated and the contiguous side chains; (ii) a local minimization of the surrounding water molecules; and (iii) a global minimization of the system. Afterward, (iv) the whole system is equilibrated by performing a short NVT simulation.
- Sampling phase. An NPT simulation is performed by using GROMACS, and the binding affinity is evaluated along the trajectory by employing a number of scoring functions. Each binding score is averaged along the MD trajectory, i.e., on all VHH–target poses extracted from the trajectory.

- Mutation acceptance/rejection phase. The mutation is accepted or rejected taking into account its effect on the binding affinity of the VHH–target complex. In particular, the gain or loss of binding affinity is evaluated according to each scoring function, and a consensus-making decision is formulated. The algorithm establishes that if the gain of binding affinity upon mutation is predicted by more scoring functions than a certain threshold (a tunable parameter), the mutation is accepted.

7. Before running the optimization algorithm, input files must be accurately prepared, and parameters involved in all three phases must be set up by modifying the initial lines of the BINDesignER script. The coordinates of the equilibrated system must be provided in pdb and gro (GROMACS) formats together with the associated topology files, and restraint files if necessary, as required for any GROMACS simulation (*see* Subheading 3.3, **step 1**). In the coordinates files, the target should be labelled as "chain A," while the binding VHH should correspond to "chain B." The initial files should be renamed as follows: *Start.pdb*, *Start.gro*, *topol-0.top*, *topol_Protein_chain_A-0.itp*, and *topol_-Protein_chain_B-0.itp*. All files, including the BINDesignER script, should be placed in the same starting folder. Other required parameters and input files are listed below:

- Parameters for the mutation phase. Numbers of residues within the segments where mutations should be performed: CDR residue numbers should be listed as they appear in the pdb or gro file. Residue names (three-letter code) that will be used in the mutagenesis: as a standard option, all residue types are considered except for CYS. Path of SCWRL4 executable. Parameters for the minimization runs are provided in a set of mdp file (GROMACS parameter file): standard parameter files are included with BINDesignER, and the parameter options can be modified by the user to adapt to his/her system (*see* below). The radius cutoff to include in the minimization water molecules surrounding the mutated residue (standard option: 0.2 nm) can be set up at the beginning of the script.
- Parameters for the acceptance/rejection phase. Only the threshold value for the consensus algorithm should be set. For the VHH–protein system, it was proved that a threshold of three over six scoring functions (50%) is the optimum option [11]. Thus, this is set as the default value in the simulation. Names of scoring functions to be used must be given, and paths pointing to their location must be specified at the beginning of the script.
- The total number of attempted mutations must be also included at the beginning of the script. As a rule of thumb, the optimization should be performed until all binding scores reach a plateau.

Since the simulation has a restart option, the design can be extended by increasing the total number of attempts.

- Other input files: parameters for the sampling phase. MD parameters must be specified in a set of GROMACS *mdp* files. A suitable set of *mdp* files, optimized for the VHH–target system, are included with BINDesignER: (i) *minim_scmut.mdp* (for the first minimization in vacuo, only the mutated side chain), (ii) *minim_overlap.mdp* (for the second minimization in water solvent, only water molecules surrounding the mutant residue), (iii) *minim.mdp* (for the third, global, minimization), (iv) *md_NVT.mdp* (NVT equilibration), and (v) *md_NPT.mdp* (NPT run). Standard MD parameters are used as described in Subheading 3.1, **step 3**, except for temperature (330 K), simulation length (30 ns), and cutoffs for both electrostatic and van der Waals interactions (0.8 nm). The cutoff has been decreased to adapt it to a shrunken simulation box. Additional parameters control the number of VHH–target poses to be collected for the binding affinity evaluation (in the MD parameter file) and the length of the initial interval to be discarded for analysis. In the default options, poses are sampled from the MD trajectory every 100 ps, and the first 10 ns are discarded. When Haddock scoring function is used, the Haddock input file of the starting VHH–target complex must be set up indicating each histidine protonation state. An input example is included with BINDesignER (files *inpA* and *inpC*).

8. The optimization starts by launching the BINDesignER script. Along (with) the optimization, several temporary files will be written and then deleted. A set of output files will be written in the folder *Teff_${T}* (where *${T}* is the chosen threshold value for the scoring function acceptance) as well as a *log* file.

9. Along the optimization, at every simulation step (*${step}*), the VHH and target coordinates are saved in a pdb file (*evol_binder ${step}.pdb* and *evol_target${step}.pdb*), and its predicted binding affinity (expressed as scoring function averages over the above MD trajectory) is printed as a function of the optimization step (*Edt_${T}.dat*). In Fig. 4b we show an example of binding score profile obtained during the simulation of affinity optimization of VHH-D9 toward HER2 (*see* Fig. 4b). Along with the simulation, accepted mutations in the CDR1 (highlighted in blue in Fig. 4c–d) improved the binding affinity, since values of all six employed scoring functions (one of which is shown in Fig. 4b) decreased with respect to the original binding, indicating a higher binding affinity. During the last 20 steps, no attempted mutation was accepted, suggesting that the optimal candidate (*see* Fig. 4d) for this sequence path had been reached (*see* **Note 7**).

3.3 In Silico Screening

To assess the binding affinity of the designed VHH mutants, the optimum candidates generated by the algorithm (those with the lowest binding scores) are analyzed by long-time MD simulations. Moreover, the stability and aggregation propensity of each selected VHH mutants can be predicted by following the protocol in ref. 32, which is summarized here:

10. Each new VHH–target complex undergoes an MD simulation for at least 250 ns in water at 330 K. The binding affinity of the complexes is analyzed by evaluating the VHH–target poses along the last 100 ns of the trajectory according to selected scoring functions. Binding scores should be stable, fluctuating around a lower average value than the average value obtained for the starting complex.
11. 200 ns MD simulations of the VHH mutants are performed at 300 K in water solvent.
12. Poses from each MD trajectory are clustered by employing the Daura's method in GROMACS [33].
13. Hotspots, i.e., regions prone to interact with other proteins or molecules, are predicted for the most representative structures of each cluster, by employing the BEPPE prediction server [34].
14. Predicted hotspots are used as putative binding sites in the homodimerization docking of each VHH mutant by using the HADDOCK webserver [30]. Docking scores obtained from all hotspot combinations among all cluster centers are used to predict their aggregation propensity. A low aggregation propensity is an indication of higher VHH quality.

3.4 Recombinant Protein Design

Mutants generated by the design algorithm, stable along the computed time, and with low aggregation propensity, are expressed as recombinant proteins in a bacterial host and purified to be further tested (*see* **Note 9**). For each mutant, the gene encoding for the protein of interest is cloned in a suitable expression vector using a restriction enzyme cloning protocol; different expression conditions are tested, to obtain high protein yields.

15. The primary sequence of the new VHHs is retro-translated to a DNA sequence with codons optimized for the *E. coli* expression system, using Sequence Manipulation Suite: Reverse Translate [35].
16. Synthetic genes are purchased as gene fragments that include the restriction sequences of the DNA endonuclease NcoI (C^CATGG) at the 5′ end and NotI (GC^GGCCGC) at the 3′ end. Usually, gene fragments are provided with extra nucleotides flanking the gene of interest. Their presence is required for the enzymatic activity of the restriction enzymes (*see* **Note 10**).

3.5 Cloning

The gene of the mutated VHH protein is cloned as a chimera with a fusion tag. Fusion tags are a major help in protein purification and may also be useful in further applications of the VHH, e.g., binding to a suitable support. A home-modified pET14 (kindly provided by A. De Marco's Lab) is used as expression vector.

In the following protocol, the empty plasmid is replicated to a suitable amount in *E. coli* DH5α. Plasmid and insert (gene fragment) are cleaved in the specific sites by restriction enzymes. The insert is ligated to the vector and *E. coli* competent cells are transformed with the plasmid. Selection of the colonies containing the desired plasmid can be performed taking advantage of antibiotic resistance present in the plasmid. The presence of the inserted gene is evaluated by PCR screen on the colonies.

17. Vector preparation: Thaw *E. coli* DH5α competent cells on ice for 5 min, transfer 20–50 μL into a 1.5 mL tube, and add 5–20 ng of empty pET14 plasmid; mix gently by tapping. The sample is incubated 30 min on ice; heat shock cells at 42 ° C for 45 s in a thermomixer, followed by 2–5 min in ice (*see* **Note 11**).
18. Add 500 μL of LB broth to the cells and incubate at 37 °C for at least 45 min, under 200 rpm agitation (*see* **Note 12**).
19. Transformed cells are inoculated in a 50 mL conical tube, or 50 mL sterile flask, containing 10 mL of LB broth supplemented with 50 μg/mL of ampicillin. The culture is incubated at 37 °C under 200 rpm agitation, O.N. (*see* **Note 13**).
20. After the O.N. incubation, cells are harvested by centrifugation for 10 min at 4400 g. The exhausted media should be discarded after inactivation of any biological contaminant, e.g., with some bleach.
21. Extract the DNA vector from the cell pellet following the instruction of the DNA miniprep kit. To increase plasmid concentration, the plasmidic DNA should be recovered in water in the minimum volume suggested by the manufacturer.
22. Measure DNA concentration using a droplet spectrophotometer. Following instrument specifics, pipette 2 μL of water to generate the blank (or zero), and subsequently, measure the sample concentration using 2 μL of the DNA plasmid.
23. Set up the double digestion reaction following datasheets of the restriction enzymes. Generally, for an optimal reaction, use 2–3 μg of DNA plasmid with 3 U of NcoI and 3 U of NotI in the recommended reaction buffer (i.e., in 40 μL of reaction: 30 μL of plasmid at 70–100-μg/mL concentration, 2 μL NcoI at 1.5 U/μL concentration, 2 μL NotI at 1.5 U/μL concentration, 4 μL of 10× reaction buffer, 2 μL of water). Incubate at 37 °C for 2 h in a water bath.

24. Insert preparation: Resuspend the synthetic gene fragment (insert) in water to reach a concentration of 20 ng/μL, gently pipetting the water in the vial until complete dissolution of the lyophilized DNA. Treat 40 ng of the resuspended DNA fragments with 0.5 U of NcoI and NotI in 20 μL of reaction mixture (*see* **Note 14**). Reaction protocol: 1 h incubation at 37 °C, followed by 20 min at 80 °C for enzyme inactivation. It may be convenient to use a thermal cycler (PCR machine), with a suitable program.
25. Prepare 1% agarose gel in 1× TAE buffer, with a 1× final concentration of DNA gel stain (*see* **Note 15**); make sure the gel is completely solidified before use.
26. Add 8 μL of the 6× DNA loading buffer to the reaction mixture of the vector digestion and gently mix it with a pipette.
27. Load a suitable amount of DNA ladder in the first lane (usually 5 μL, but the amount may vary according to manufacturer's instructions); load the digested DNA vector in the following well; in another lane, load 200 ng of untreated DNA plasmid as control. Run the gel at 80–100 V until the dye line has reached ≈80% of the gel.
28. Analyze the gel with a UV transilluminator. With a clean scalpel or blade, cut out the band corresponding to the number of bases of the linearized plasmid, transfer the excised gel in a new 1.5 mL tube (*see* **Note 16**), and weight it on the appropriate scale (*see* **Note 17**). For extraction of the DNA, follow the recommended procedure of the gel extraction kit.
29. Elute the DNA in water and measure the concentration following the same procedure as at **step 22**.
30. Set up the DNA ligase reaction in a 10 μL volume, mixing linearized plasmid and digested inserts. The molar ratio of plasmid–insert should be kept between 1:3 and 1:4 (when possible, at least 20 fmol of plasmid should be used; *see* **Note 18**), with the T4 DNA ligase amount recommended by the enzyme manufacturer (Table 1). It is advisable to set up a control reaction using water instead of the gene fragment. Table 1 gives an example of the typical amounts for a ligase reaction and its control. Incubate the reaction for 15 min to 1 h at room temperature (*see* **Note 19**).
31. Prepare LB agar plates with 100 μg/mL of ampicillin (*see* **Note 20**).
32. Transform 5 μL of each ligase reaction in 50 μL of *E. coli* DH5α competent cells. Repeat the procedure using the control mixture. Perform the heat shock as described at **step 17**. It is advisable to keep the remaining ligase reaction at 4 °C, for later use.

Table 1
DNA ligase reaction component

	Ligase reaction	Control reaction
Digested DNA vector	2 μL	2 μL
Digested insert	1 μL	/
T4 ligase	0.5 μL	0.5 μL
Reaction buffer 10×	1 μL	1 μL
H_2O	5.5 μL	6.5 μL

The table represents a putative DNA ligase reaction; in the table are listed the reaction component and how to set up the reaction control

33. Add 500 μL of SOC broth and continue as reported at **step 18**.
34. After 45 min incubation at 37 °C, centrifuge the 1.5 mL tubes for 3 min at 500 g. Cell pellet will be visible on the wall of the tube.
35. Discard 400 μL of supernatant medium, and gently resuspend the cells in the residual medium. Transfer the resuspended cells (about 150 μL) on a different LB agar plate for each transformation, including the control. The solution should be dispensed in the center of the plate and uniformly spread using a disposable spreader. Flip the plates upside down and incubate at 37 °C, O.N.
36. Analyze the plates: *E. coli* colonies should be present on the plate corresponding to ligase reactions performed using inserts, but the control plate should be empty (*see* **Note 21**).
37. Bacterial colonies grown on the plate are often too small for a PCR colony screen. It is good practice to prepare a new LB agar plate by restreaking the colonies (restreak plate) obtained from the ligation experiment. With a sterile toothpick, pipette tip or wire loop; pick a single colony and then streak the tip on the new plate; incubate at 37 ° C for 4–5 h, flipping the plate upside down.
38. PCR colony screen. Dispense 5 μL of water for each PCR tube (whether spare or in a strip). With a sterile toothpick or a pipette tip, pick each colony from the restreak plate and dip it in a PCR tube. Make sure that the labelling on the tube corresponds to the plate.
39. Prepare a PCR master mix (number of colonies +2; Table 2). Prepare also a positive control using the empty DNA vector, to compare the PCR product migration, and a negative control, without the template (Table 2). Dispense 15 μl reaction mix in each tube. The DNA polymerase buffer contains salts and dNTPs necessary for the enzyme reaction.

Table 2
PCR colony screening reaction

	PCR	Plasmid control	Negative control	For 24 reactions (N + 2)
Template	5 μL	1 μL (5–20 ng/μL)	/	
T7 promoter primer 10 μM	2 μL	2 μL	2 μL	52 μL
T7 terminator primer 10 μM	2 μL	2 μL	2 μL	52 μL
Taq DNA polymerase (5 U/μl)	0.1 μL	0.1 μL	0.1 μL	2.6 μL
Taq reaction buffer 5×	3 μL	3 μL	3 μL	78 μL
H_2O	7.9 μL	10.9 μL	12.9 μL	205.4 μL

In the table are listed the PCR colony screening reaction component, or 1 or 24 reactions, and positive and negative control

40. Using the PCR machine, set a suitable program, for example, (I) 95 ° C for 5 min; (II) denaturation, 95 °C for 30 s; (III) annealing, 47 ° C for 1 min; and (IV) elongation, 72 ° C per 30 s. Repeat steps II to IV for 20–25 cycles. (V) Last elongation step at 72 °C for 2 min. Length of steps and temperatures are adjusted according to the specific Taq DNA polymerase used for the experiment (*see* **Note 22**).
41. Prepare a 1% agarose gel as previously described (*see* **Note 14**), but use the comb with the highest number of wells (*see* **Note 23**).
42. Dispense 3–4 μL of 6× DNA sample loading buffer in each reaction mix. Load on the gel 5 μL of the DNA ladder in the first lane and 8–9 μL of each PCR in the subsequent lanes. Run the gel as described before (**step 27**).
43. Analyze the gel with a transUV or an imaging system that allows to keep a record of the experiment. Compare the migration of the PCR products with the DNA marker and the control. Select the colony/colonies with the expected fragment size (*see* **Note 24**).
44. Prepare 10 mL of LB broth with the appropriate antibiotic in a 50 mL conical tube. Pick the positive colony with a sterile toothpick or a pipette tip from the restreak plate and inoculate it in the medium. Incubate the inoculum at 37 ° C under agitation, O.N.
45. Harvest cells by centrifuging the tubes and proceed with the miniprep following protocol described at **step 21**. Confirm the positive clones by plasmid sequencing.

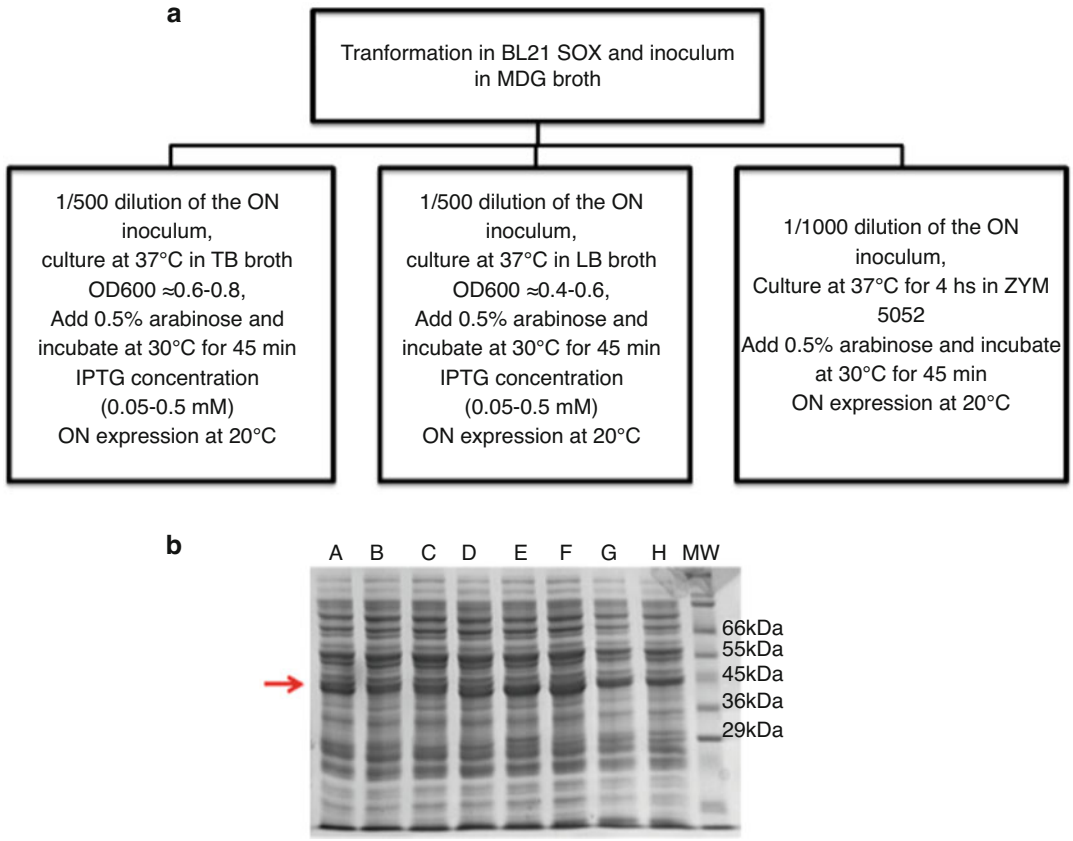

Fig. 5 Small-scale expression. (**a**) Schematic representation of an ideal strategy to explore the expression conditions for the new VHHs. (**b**) SDS-PAGE of VHHs after small-scale expression of the VHH-GFP-6His (expected molecular weight 42 kDa)

3.6 Small-Scale Expression and Purification

The DNA plasmids with the inserted VHH of interest are transformed into an *E. coli* strain suitable for protein expression. Different conditions for culture media, ITPG concentration, and expression temperature can be explored to determine a protocol that maximizes protein yield. Figure 5 is a schematic representation of the suggested approach.

46. Transform 20–50 ng of DNA plasmid in 20–30 μL of *E. coli* BL21 (DE3) SOX (previously thawed on ice). Perform the heat shock. Add 200 μL of LB broth.
47. Inoculate the transformation mix in 2.5 mL of MDG broth supplied with ampicillin (100 μg/mL) and chloramphenicol (34 μg/mL) using a 15 mL conical tube. Incubate the inoculum at 37 ° C O.N., under agitation.
48. For each sample, prepare glycerol stocks by transferring 400 μL of the overnight culture in a 2 mL conical tube and adding 200 μL of 60% sterile glycerol, to a final glycerol concentration

of 20%. Store the glycerol sticks at −80 °C. Glycerol stocks can be used to start expression experiments at a later time, without transformation. The remaining O.N. culture is split to start cultures with different expression conditions.

49. Auto induction media: 3–4 mL of ZYM 5052 is used for each protein. Dispense the media supplied with antibiotics in a 24 deep well plate; add to each well 3–4 μL of the MDG O.N. pre-inoculum (1:1000 dilution). Seal the plate with a breathable sealer, and incubate the plate at 37 °C for 4 h, under agitation (at least 200–250 rpm using an orbital shaker or 600 rpm if you are using a thermomixer). Decrease the temperature of the incubator to 30 °C, add arabinose to a final concentration of 0.5% (60–80 μL of a 25% stock), and incubate at 30 °C for 45 min. Decrease again the temperature to 20 °C and leave the culture O.N. (*see* **Note 25**).

50. IPTG induction: The protocol is the same for the LB and TB broth. Dilute each O.N. pre-inoculum cultures 1:500 in 10–15 mL of broth supplemented with antibiotics in a 50 mL tube and incubate at 37 °C, under 200 rpm agitation. Optical density (OD) can be measured at 600 nm wavelength with a spectrophotometer, after blanking the system with clean broth in a plastic cuvette. After 2–3 h of culture, 1 mL of medium can be collected with a sterile tip or a serological pipette to measure OD_{600}. When the culture reaches OD_{600} ≈0.3–0.4 in LB or OD_{600} ≈0.6–0.7 in TB, decrease the temperature to 30 °C, and add 0.5% arabinose (for 10–12 mL of culture, add 200–240 μL of the 25% stock), and incubate for 45 min. Divide the culture in a 24-deep well plate, using 3–4 mL of culture for each expression test with a different IPTG concentration (e.g., 0.05 , 0.2 , 0.5 , 1 mM IPTG), add the IPTG, and incubate the cultures at 20 °C O.N, under agitation.

51. Transfer the culture in 2 mL conical tube and pellet the cells using a centrifuge. If the centrifuge fits a SBS plate holder, the 24 deep well plate can be used directly, without transferring the culture into tubes. Cell pellet can be stored at −80 °C for later analysis or processed directly.

52. For cell lysis, prepare 1 mL of buffer for each expression test to be analyzed (Table 3). The buffer composition is 20 mM Tris–HCl, pH 7.9, 500 mM NaCl, 5% glycerol, 0.1% Triton X-100, 0.1% Tween 20, 0.5 mg/mL lysozyme, 0.01 mg/ml DNaseI, and 10 mM $MgCl_2$.

53. Resuspend the cell pellet with 1 mL of lysis buffer gently pipetting to mix, until dissolution. Freeze at −80 °C for 10 min and thaw for 20 min. The freezing–thawing sequence improves cell lysis. Keep the thawed suspension under agitation at 4 °C for 1 h. Save 10 μL for SDS-PAGE analysis (total extract, TE).

Table 3
Lysis buffer

1 reaction Lysis buffer	24 reactions Lysis buffer	
500 μL	12 mL	2× nickel buffer
2 μL	48 μL	50 mg/ml lysozyme
2 μL	48 μL	5 mg/ml DNaseI
10 μL	240 μL	100-mM PMSF
5 μL	120 μL	2-M $MgCl_2$
0.1 μL	2.4 μL	Triton X-100
0.1 μL	2.4 μL	Tween 20
480.8 μL	11.54 mL	H_2O

In the table are listed the components of the lysis buffer for 1 and 24 reactions; 1 ml of it will be used for the small-scale expression and purification protocol

54. Prepare 25 μL of Ni-NTA for each sample. For example, if 24 samples are to be processed, use 1.3 mL (corresponding to 24 samples +2) of properly resuspended Ni-NTA solution. Wash the resin twice with water and three times with 1× nickel buffer. The optimal protocol for resin preparation depends on available Ni-NTA resin and should follow the recommendation of the producer. Finally, resuspend the resin in 1.95 mL of nickel buffer to have a 4× resin solution.
55. Centrifuge the cell suspension at high speed and at 4 °C, in 2 mL tubes (or directly in the 24 deep well plate, when a suitable rotor is available). Transfer the supernatant in a 1.5 mL conical tube, paying attention to avoid the transfer of cell debris. Save 10 μL for SDS-PAGE analysis (soluble fraction, S) (*see* **Note 26**).
56. Add to each sample 100 μL of resin solution, paying attention that the resin is well resuspended when pipetting. Incubate under agitation for 1 h.
57. Remove the cell extract by decanting or centrifuging. Wash the resin twice with 10 resin volume (RV, 250 μL) of 20 mM Tris–HCl, pH 7.9, 500 mM NaCl, and 5% glycerol, followed by 10 RV of 20 mM Tris–HCl, pH 7.9, 500 mM NaCl, 5% glycerol, and 5 mM imidazole and 10 RV of 20 mM Tris–HCl, pH 7.9, 500 mM NaCl, 5% glycerol, and 25 mM imidazole.
58. To elute the protein, add 40 μL of 20 mM Tris–HCl, pH 7.9, 500 mM NaCl, 5% glycerol, and 250 mM imidazole. Incubate for 5 min and collect the eluate. Save 30 μL for SDS gel analysis (elution, El).

59. Analyze the sample on SDS-PAGE gel, prepared according to the expected molecular weight of the protein, or using a precast SDS-PAGE. Dilute the total extract and the soluble fraction adding 20 μL of 1× nickel buffer. Add to all the samples 10 μL of SDS loading buffer and heat for 5 min at 95 °C. For each of the expression experiments, load 4 μL of TE in the first lane, followed by 4 μL of S and 15 μL of El fraction. Finally, load 4–5 μL of protein ladder in each gel. Run and stain the gel according to standard procedure.
60. Conditions yielding the highest protein amount can be used to set up large-scale expression (*see* **Note 27**).

4 Notes

1. The choice of scoring functions to evaluate the binding affinity depends on the target system. Additional scoring functions can be introduced in the optimization script and employed for the VHH screening.
2. Communication between the computational team and the experimental team is key to overcome many potential problems that can result in a waste of time and resources. Starting from the choice of the VHH to optimize, make sure the framework is versatile enough to accept mutations and that the construct is suitable for future application.
3. Make sure all tags are included in the starting model for both target and ligand, or at least accounted for. Large tags (such as GFP or Fc tags) could hide some binding sites.
4. Information about the binding mode of the system is necessary to perform the docking between VHH and the protein target. Preliminary experiments or literature can offer the required insight into the binding. When neither is available, a blind docking between binder and target, while resulting in a less accurate approach, can be a sufficient starting point.
5. If the complex conformation comes from an experimental structure, the IRMSD will achieve low and stable values long before the end of the simulation. On the other hand, if the complex is generated by docking protocols, the simulation will show larger rearrangements, and IRMSD analysis helps assessing whether a stable complex is obtained or a longer simulation is required.
6. Depending on the size of the system, i.e., the VHH–target complex inside a solvent (water) box, and the available computational resources, particular care should be taken in establishing the size of the portion that will undergo optimization. Thus, the cleavage of the protein target is a good option as

long as the binding conformation remains unaffected. Restrains can be employed to fix the orientation of the complex.

7. Since BINDesignER is a stochastic algorithm, each new design simulation will explore a different evolutionary path. Thus, performing new design simulations could be advantageous to explore alternative sequence paths.
8. According to the note above, running multiple shorter optimizations allows to generate a diverse pool of candidate binders and is preferable than running one single longer optimization.
9. Multiple VHH candidates should be selected, considering possible problems with recombinant proteins: such as low expression levels, low solubility, folding problems, and proteolysis events.
10. Make sure that the gene is in frame with the DNA plasmid ORF: the ATG sequence of the NcoI site should code for the first methionine residue of the protein (translation start). When purchasing the synthetic gene fragments, verify the company procedures for the synthesis: if no extra tails are provided, then add at least six random bases upstream the NcoI site and six bases downstream the NotI site.
11. Competent cells can be stored at −80 °C as ready-to-use aliquots, or the required volume of cells can be transferred in a new 1.5 mL conical tube. Repeated thawing–freezing cycles, however, will reduce efficiency of the transformation.
12. **Steps 3** and **4** should be performed in a sterile environment, i.e., under a biological fume hood or in close proximity to a flame.
13. The volume of the container, tube or flask, should be at least five times the volume of the culture.
14. Due to the small volume, it may be convenient to prepare a master mix with a volume corresponding to the number of reactions to be performed +1. The master mix contains all the common reagents and an equal amount is dispensed in a vial for each reaction. If a PCR machine is used for incubation, dispense the mix in PCR tubes or strip.
15. For 1% agarose gel preparation: Weight 0.4 g of agarose powder. In a glass flask, add 40 ml of 1× TAE to the powder, gently mix, and cover the flask with a wet tissue paper. Warm up the solution in microwave until agarose is melted. Prepare the cast-on tray gel system with the larger wells available; place the comb in the vessel. Before pouring the gel solution, add the DNA stain following manufacturer's instructions. Pour the solution in the vessel avoiding the formation of bubbles. If present, bubbles can be moved to the side with a pipette tip.

16. Pay attention at reducing as much as possible the exposure of the DNA to the UV light. A prolonged UV exposure causes mutations or damages to the nucleic acid polymer.
17. When weighting the excised gel, consider the unladen weight of the empty 1.5 mL tube.
18. To calculate the required mass of insert (in g) that should be used, multiply the insert-to-vector molar ratio (optimal values: 3–4) by the mass of vector (in g), the insert length (in base pairs, bp), and divide by the vector length (in bp).
19. The ligase buffer contains ATP. A good practice to avoid degradation of the ATP by freezing–thawing cycles is to aliquot the ligase buffer in suitable amounts.
20. Using the microwave, slowly melt the LB agar. Once the solution is liquid, let it cool below 50 °C. When you can keep the bottle in your hands, add the appropriate antibiotic (for the pET14 vector, ampicillin, final concentration 100 μg/mL), mix by gentle agitation, and pour the solution in 90 mm sterile Petri plates (approx. 17–22 ml per plate). Agar plate preparation should be performed in a sterile environment, i.e., under a biological fume hood or in close proximity to a flame. When plates are solidified, store them at 4 °C. Plates can be used up to 3 weeks after preparation.
21. If the same number of colonies is present in the ligation tests and in the control, an incomplete digestion of the plasmid DNA is likely. Alternatively, samples might have been contaminated during the gel extraction. The ratio between colonies in the ligation tests and in the control should be as highest as possible.
22. The Taq polymerase can be error-prone; the concentration of the substrate varies according to the specific enzyme. Annealing and extension times and temperatures depend on the Taq polymerase used. New, optimized DNA polymerases have a processivity of 1 Kb/30 s.
23. Providing that the length of the gel tray allows for an efficient separation, 2 combs per gel may be used to increase the number of samples loaded in a single gel.
24. For a VHH gene insertion, the PCR product should have a length of 360 bp (average length of the VHH domain) plus the upstream and downstream regions included between the T7 promoter and T7 terminal. When tags are fused to the VHH gene, their length should be added to the expected dimension of the PCR product.
25. This step can be performed using a thermomixer with a plate holder and a shaking system.

26. For **steps 53–58**, working at 4 °C or in ice is advisable. A thermomixer with a cooling system can be used for all the incubation steps.
27. The small-scale expression and purification tests give a good hint of optimal expression conditions, but scale-up of the culture can sometimes lead to puzzling results. If this is the case, an intermediate-scale culture (e.g., 50 mL) may help in making sense of the results and identifying the optimal conditions for the large-scale expression.

Acknowledgments

This work has been funded by the Italian Association for Cancer Research (AIRC) through the grant "My First AIRC grant" Rif.18510 (PI: Fortuna) and the CINECA Awards N. HP10C70TG1, 2018, for the availability of high performance computing resources and support.

References

1. Melarkode Vattekatte A, Shinada NK, Narwani TJ et al (2020) Discrete analysis of camelid variable domains: sequences, structures, and in-silico structure prediction. PeerJ 8:e8408. https://doi.org/10.7717/peerj.8408
2. Mitchell LS, Colwell LJ (2018) Comparative analysis of nanobody sequence and structure data. Proteins Struct Funct Bioinforma 86: 697–706. https://doi.org/10.1002/prot.25497
3. Zavrtanik U, Lukan J, Loris R et al (2018) Structural basis of epitope recognition by heavy-chain camelid antibodies. J Mol Biol 430:4369–4386. https://doi.org/10.1016/j.jmb.2018.09.002
4. Arbabi-Ghahroudi M (2017) Camelid single-domain antibodies: historical perspective and future outlook. Front Immunol 8:1–8. https://doi.org/10.3389/fimmu.2017.01589
5. Wang Y, Fan Z, Shao L et al (2016) Nanobody-derived nanobiotechnology tool kits for diverse biomedical and biotechnology applications. Int J Nanomedicine 11:3287–3303. https://doi.org/10.2147/IJN.S107194
6. De Vlieger D, Ballegeer M, Rossey I et al (2018) Single-domain antibodies and their formatting to combat viral infections. Antibodies 8:1. https://doi.org/10.3390/antib8010001
7. Soler MA, Fortuna S, De Marco A, Laio A (2018) Binding affinity prediction of nanobody-protein complexes by scoring of molecular dynamics trajectories. Phys Chem Chem Phys 20:3438–3444. https://doi.org/10.1039/c7cp08116b
8. Kuroda D, Tsumoto K (2020) Engineering stability, viscosity, and immunogenicity of antibodies by computational design. J Pharm Sci 109:1631–1651. https://doi.org/10.1016/j.xphs.2020.01.011
9. Guilliams T, El-Turk F, Buell AK et al (2013) Nanobodies raised against monomeric α-synuclein distinguish between fibrils at different maturation stages. J Mol Biol 425: 2397–2411. https://doi.org/10.1016/j.jmb.2013.01.040
10. Cheng X, Wang J, Kang G et al (2019) Homology modeling-based in silico affinity maturation improves the affinity of a nanobody. Int J Mol Sci 20(17):4187
11. Soler MA, Medagli B, Semrau MS et al (2019) A consensus protocol for the: in silico optimisation of antibody fragments. Chem Commun 55:14043–14046. https://doi.org/10.1039/c9cc06182g
12. Gladich I, Rodriguez A, Hong Enriquez RP et al (2015) Designing high-affinity peptides for organic molecules by explicit solvent molecular dynamics. J Phys Chem B 119: 12963–12969. https://doi.org/10.1021/acs.jpcb.5b06227

13. Hong Enriquez RP, Pavan S, Benedetti F et al (2012) Designing short peptides with high affinity for organic molecules: a combined docking, molecular dynamics, and Monte Carlo approach. J Chem Theory Comput 8: 1121–1128. https://doi.org/10.1021/ct200873y
14. Del Carlo M, Capoferri D, Gladich I et al (2016) In silico design of short peptides as sensing elements for phenolic compounds. ACS Sensors 1:279–286. https://doi.org/10.1021/acssensors.5b00225
15. Guida F, Battisti A, Gladich I et al (2018) Peptide biosensors for anticancer drugs: design in silico to work in denaturizing environment. Biosens Bioelectron 100:298–303. https://doi.org/10.1016/j.bios.2017.09.012
16. Soler MA, Rodriguez A, Russo A et al (2017) Computational design of cyclic peptides for the customized oriented immobilization of globular proteins. Phys Chem Chem Phys 19: 2740–2748. https://doi.org/10.1039/c6cp07807a
17. Russo A, Scognamiglio PL, Enriquez RPH et al (2015) In silico generation of peptides by replica exchange monte carlo: docking-based optimization of maltose-binding-protein ligands. PLoS One 10:1–16. https://doi.org/10.1371/journal.pone.0133571
18. Soler MA, Fortuna S, Scoles G (2015) Computational design of peptides as probes for the recognition of protein biomarkers. Eur Biophys J WITH Biophys Lett 44:S149
19. Ochoa R, Soler MA, Laio A, Cossio P (2020) PARCE: protocol for amino acid refinement through computational evolution. Comput Phys Commun 260:107716
20. Pronk S, Páll S, Schulz R et al (2013) GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. Bioinformatics 29:845–854. https://doi.org/10.1093/bioinformatics/btt055
21. Krivov GG, Shapovalov MV, Dunbrack RL Jr (2009) Improved prediction of protein side-chain conformations with SCWRL4. Proteins 77:778–795. https://doi.org/10.1002/prot.22488
22. Vreven T, Hwang H, Weng Z (2011) Integrating atom-based and residue-based scoring functions for protein-protein docking. Protein Sci 20:1576–1586. https://doi.org/10.1002/pro.687
23. Andrusier N, Nussinov R, Wolfson HJ (2007) FireDock: fast interaction refinement in molecular docking. Proteins Struct Funct Bioinforma 69:139–159. https://doi.org/10.1002/prot.21495
24. Sarti E, Granata D, Seno F et al (2015) Native fold and docking pose discrimination by the same residue-based scoring function. Proteins Struct Funct Bioinforma 83(4):621–630. https://doi.org/10.1002/prot.24764
25. Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125: 1731–1737. https://doi.org/10.1021/ja026939x
26. Fogolari F, Corazza A, Yarra V et al (2012) Bluues: a program for the analysis of the electrostatic properties of proteins based on generalized Born radii. BMC Bioinformatics 13:1–16. https://doi.org/10.1186/1471-2105-13-S4-S18
27. Veggiani G, Giabbai B, Semrau MS et al (2020) Comparative analysis of fusion tags used to functionalize recombinant antibodies. Protein Expr Purif 166:105505. https://doi.org/10.1016/j.pep.2019.105505
28. Fisher RD, Ultsch M, Lingel A et al (2010) Structure of the complex between HER2 and an antibody paratope formed by side chains from tryptophan and serine. J Mol Biol 402: 217–229. https://doi.org/10.1016/j.jmb.2010.07.027
29. Guex N, Peitsch MC (1997) SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling. Electrophoresis 18:2714–2723. https://doi.org/10.1002/elps.1150181505
30. Van Zundert GCP, Rodrigues JPGLM, Trellet M et al (2016) The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J Mol Biol 428: 720–725. https://doi.org/10.1016/j.jmb.2015.09.014
31. Cho H-S, Mason K, Ramyar KX et al (2003) Structure of the extracellular region of HER2 alone and in complex with the Herceptin Fab. Nature 421:756–760. https://doi.org/10.1038/nature01392
32. Soler MA, De Marco A, Fortuna S (2016) Molecular dynamics simulations and docking enable to explore the biophysical factors controlling the yields of engineered nanobodies. Sci Rep 6:1–11. https://doi.org/10.1038/srep34869
33. Daura X, Gademann K, Jaun B et al (1999) Peptide folding: when simulation meets experiment. Angew Chem Int Ed 38:236–240. https://doi.org/10.1002/(sici)1521-3773(19990115)38:1/2<236::aid-anie236>3.3.co;2-d

34. Scarabelli G, Morra G, Colombo G (2010) Predicting interaction sites from the energetics of isolated proteins: a new approach to epitope mapping. Biophys J 98:1966–1975. https://doi.org/10.1016/j.bpj.2010.01.014

35. Stothard P (2000) The sequence manipulation suite: JavaScript programs for analyzing and formatting protein and DNA sequences. BioTechniques 28(1102):1104. https://doi.org/10.2144/00286ir01