ChemComm

COMMUNICATION

View Article Online
View Journal | View Issue

Cite this: *Chem. Commun.*, 2019, **55**, 14043

Received 9th August 2019,
Accepted 29th October 2019

DOI: 10.1039/c9cc06182g

rsc.li/chemcomm

A consensus protocol for the *in silico* optimisation of antibody fragments†

Miguel A. Soler,[ab] Barbara Medagli,[c] Marta S. Semrau,[d] Paola Storici,[d] Gregor Bajc,[e] Ario de Marco,[f] Alessandro Laio*[ag] and Sara Fortuna*[c]

We present an *in silico* mutagenetic protocol for improving the binding affinity of single domain antibodies (or nanobodies, VHHs). The method iteratively attempts random mutations in the interacting region of the protein and evaluates the resulting binding affinity towards the target by scoring, with a collection of scoring functions, short explicit solvent molecular dynamics trajectories of the binder-target complexes. The acceptance/rejection of each attempted mutation is carried out by a consensus decision-making algorithm, which considers all individual assessments derived from each scoring function. The method was benchmarked by evolving a single complementary determining region (CDR) of an anti-HER2 VHH hit obtained by direct panning of a phage display library. The optimised VHH mutant showed significantly enhanced experimental affinity with respect to the original VHH it matured from. The protocol can be employed as it is for the optimization of peptides, antibody fragments, and (given enough computational power) larger antibodies.

The discovery and optimization of antibodies has become an essential tool in nanomedicine and nanotechnology.[1,2] Indeed, the number of new medical applications involving antibodies is constantly growing, as different nude antibodies, antibody-functionalized nanoparticles and effective antibody–drug conjugates (ADCs) have entered clinical trials and have been approved for clinical applications.[3] Antibody engineering has evolved towards improving the production of more efficient immunoagents.

With respect to the past, there is more attention to conjugation procedures aimed at obtaining more homogeneous products. The possibility to control and model the macromolecule functionalization is one of the reasons for which recombinant antibody fragments represent a nimble alternative to full-length IgGs.[2,4] They offer unique opportunities as reagents for super-resolution microscopy[5,6] and are easier to treat with computational approaches.[7–9] Among the different types of antibody fragments, the single domain antibodies, or nanobodies (VHHs), are the smallest fragments that still preserve the binding capacity of the whole original antibody they derive from.[2,10]

The major issue that any computational design approach must deal with is the accuracy in the evaluation of the binder-target binding affinity, an essential feature for selecting the optimum binders.[11,12] The use of scoring functions (SFs) to evaluate the interaction between proteins is a particularly convenient approach.[13] Despite the broad variety of available SFs and the constant development of new ones, no "golden standard" has emerged yet. Indeed, SFs are developed by following different philosophies, employing different experimental databases, and different procedures to compute the interactions between atoms or a group of atoms. The accuracy of SFs is typically system dependent. Since the evaluation of the binding affinity by SFs is often affected by large errors the employment of consensus scores has been proposed.[14,15] This approach often improves the accuracy with respect to individual SFs, but generally relies on the determination of system-dependent weights.

Here we present an iterative mutagenic protocol for the *in silico* maturation of proteic binders based on a parameter-free consensus optimization (Fig. 1a). We believe that this approach could enable the rational design of stronger protein–ligands for the design of novel nanodevices[16] and for single molecule experiments.[17] The protocol differs from formerly proposed algorithms in that it does not rely on structural databases.[8,9,18] Instead, building on previous experience in peptide design,[19,20,21,22]

[a] *International School for Advanced Studies (SISSA), Via Bonomea 265, 34136, Trieste, Italy. E-mail: laio@sissa.it*
[b] *Italian Institute of Technology (IIT), Via Melen 83, B Block, 16152, Genova, Italy*
[c] *Department of Chemical and Pharmaceutical Sciences, University of Trieste, Via L. Giorgieri 1, 34127 Trieste, Italy. E-mail: s.fortuna@units.it*
[d] *Structural Biology Laboratory, Elettra - Sincrotrone Trieste S.C.p.A. SS 14 - km 163,5 in AREA Science Park, 34149 Trieste, Italy*
[e] *Department of Biology, Biotechnical Faculty, University of Ljubljana, Jamnikarjeva 101, 1000 Ljubljana, Slovenia*
[f] *Laboratory of Environmental and Life Sciences, University of Nova Gorica, Vipavska cesta 13, SI-5000, Rožna Dolina (Nova Gorica), Slovenia*
[g] *The Abdus Salam International Centre for Theoretical Physics (ICTP), Strada Costiera 11, 34151 Trieste, Italy*

† Electronic supplementary information (ESI) available: "S1. Starting structure", "S2. Optimization of consensus threshold", "S3. Design of peptides as binders of HER2", "S4. Comparison to Metropolis algorithm", "S5. Sequences", "S6. Binding affinity screening", "S7. Enzyme-linked immunosorbent assay (ELISA)", "S8. Methods". See DOI: 10.1039/c9cc06182g

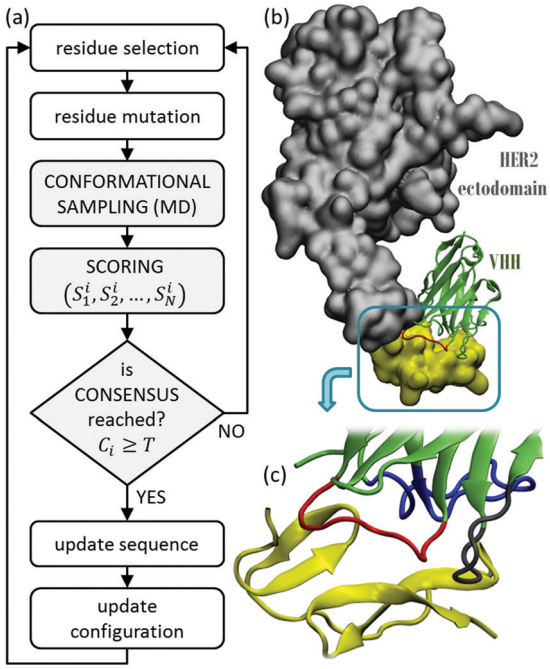

Fig. 1 (a) Algorithm diagram; (b) starting binding conformation between **D9** (green) and the HER2 ectodomain (yellow/gray), and (c) close-up on the HER2 domain employed in the design simulations with highlighted CDR1 (red, to be optimized), CDR2 (gray), and CDR3 (blue).

it relies on the scoring of molecular dynamics (MD) trajectories in explicit solvent. The most important novelty of the protocol is in the consideration of several SFs to evaluate the binding affinity between each binder and its target, making consensus decisions based on a parameter-free combination of all binding scores. As the binding scores are proxies for the binding free energy (*i.e.* the larger the binding affinity, the lower the score) the protocol, implemented in an iterative scheme, allows improving the theoretical binding affinity for all the SFs at the same time.

To benchmark the method we optimize a single CDR of a hit VHH known to bind HER2 (human epidermal growth factor receptor 2) with unsatisfactory dissociation constant (k_d = 900 nM, Fig. 1b and c).[14] The general approach requires the knowledge of the macromolecule structures involved (real or derived). We thus started from a VHH, named D9, modeled by homology modeling and docked to HER2, for which the crystal structure was available.[14,23]

The consensus-based protocol starts from the above conformation i comprising two interacting proteins prepared as in the ESI,† Section S1 (Fig. S1–S4). One first performs a finite temperature MD simulation in water solvent, followed by the computation of N_s trajectory-averaged binding scores $S_1^i, S_2^i, \ldots S_{N_s}^i$. One then subsequently performs: (i) an attempted mutation of one randomly selected residue of the binding region thus generating a new complex i + 1; (ii) a finite temperature MD simulation of the complex i + 1 in water solvent; (iii) the computation of N_s trajectory-averaged binding scores $S_1^{i+1}, S_2^{i+1}, \ldots S_{N_s}^{i+1}$; (iv) the comparison of each k-th binding score obtained for the complex i + 1 with that computed for complex i and the construction of a score vector $\vec{S}_i = \{c_k^i\}$ of dimension N_s where:

$$c_k^i = \begin{cases} 1 & \text{if } S_k^{i+1} - S_k^i < 0 \\ 0 & \text{otherwise} \end{cases} \quad (1)$$

The i-th mutation is accepted if

$$C^i = \sum_{k=1}^{N_s} c_k^i \geq T \quad (2)$$

where T is an integer number between 0 and N_s, and rejected if $C^i < T$. The parameters of this approach are the number N_s of SFs, and the value of T.

In the particular case of VHH optimization we set N_s = 6: in our former work[14] we have identified six different SFs capable of reproducing with good accuracy the experimental binding affinity ranks of VHH sets.[11] Although a higher number of SFs would allow a higher number of combinations to achieve the consensus criterion, this would also require a higher number of concurrent SFs to be computed and optimized. The consensus criterion establishes that one mutation is accepted if at least T binding scores are improved by the mutation, leading to $\binom{N_s}{T}$ possible combinations to achieve this criterion. In the ESI† (Section S2 and Fig. S5, S6) we show that the optimal consensus threshold for our system is T = 3. With lower T the scores do not improve significantly, while with higher T they improve too slowly.

Overall the protocol requires choosing two parameters (N_s and T) which may in principle be different in different systems. The six scoring functions used in this work were selected based on ref. 14. It could well be that for certain types of mutations (*e.g.* charged/uncharged residues), the addition of other SFs to the set might improve the final output. The protocol proposed here is modular as one may easily increase the library of SFs as new ones become available.

A typical optimization is shown in Fig. 2a. Here the tunable parameter T is set to T = 3, optimal consensus threshold for our system and our choice of SFs (ESI,† Fig. S5 and S6). Another example, namely the optimisation of a peptide as a binder towards the same target can be found in the ESI† (Section S3 and Fig. S7, S8) where the same SFs and T = 3 have been employed validating the algorithm on a different system and demonstrating both the algorithm robustness and its general applicability. Being r_k^i the rank of complex i according to the SFs k, we define

$$\hat{r}_k^i = \frac{r_k^i}{N} \quad (3)$$

where N is the total number of accepted mutations obtained in the runs. The average ranking score of complex i is defined as

$$R^i = \frac{1}{N_s} \sum_{k=1,N_s} \hat{r}_k^i; \quad i = 1, N \quad (4)$$

Along the optimization both $\hat{r}_k^i$ and R^i decrease and, after 44 attempted mutations, they reach a stationary state in which the SFs do not improve anymore, but still keep on changing due to

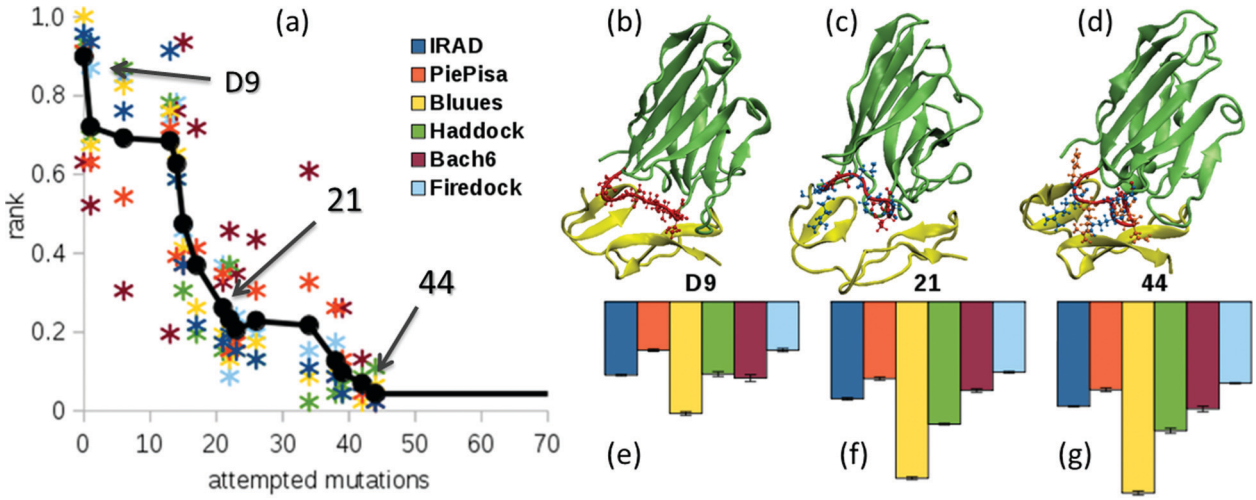

Fig. 2 (a) Global (black circles and lines) and scoring-function specific (star) ranking scores of the bindings between VHH mutants and HER2 as obtained by optimization with $N_s = 6$, $T = 3$. Three selected mutants are indicated by arrows; (b–d) simulation snapshots of the three selected complexes. CDR1 amino acids are highlighted as follows: the same as in **D9** (red), mutated at step 21 with respect to **D9** (blue), mutated at step 44 with respect to **21** (orange); (e–g) binding scores averaged over MD simulations of the selected VHH/HER2 complexes, with an interval of 100–200 ns, where the error bars are standard deviations calculated by block analysis. In this representation the units of the different scoring functions have been rescaled in order to make the relative variations visible.

stochastic fluctuations. The optimization performed with the consensus criterion allowed a lower predicted binding affinity to be reached than an equivalent Monte Carlo (MC) optimization performed by using a single score (details in the ESI,† Section S4 and Fig. S9). The MC optimization brings a gradual decrease of the score which is optimized, but most of the other scores are uncontrolled and do not change significantly. With the same computational effort, our protocol was capable of optimizing all the scores simultaneously.

To choose the optimum binder to undergo experimental validation we selected the lowest ranking solutions along the optimization (mutants **21–44** with sequences in the ESI,† Section S5 and Table S1) and performed 200 ns MD simulations in water solvent at 330 K (further computational details in the ESI,† Sections S6 and S8). Simulation snapshots reveal how successive mutations allow the VHH to extend its side chains into the target structure, thus maximizing the binding interface (Fig. 2b–d). The SFs' computations[14] clearly reveal their lower value with respect to those computed for the initial complex (Fig. S10, ESI†). Their averages calculated over the last 100 ns together with a small standard deviation, further confirm that result (Fig. S11, ESI†): for instance, both the first and last selected mutants, **21** and **44**, are predicted to bind stronger than the original **D9** by all SFs (Fig. 2e–g). We further estimated their yield:[7] while **21** was expected to be prone to aggregation, **44** was expected to be found in the same multimeric state as the original **D9** (see Fig. 3a).

The selected mutants, **21**, **44**, and **D9**, were cloned in a pET 14b expression vector; modified to generate a VHH-Cys-6His tag at the C-terminal domain and expressed in a bacterial host. All proteins were expressed in a soluble form and were purified *via* IMAC followed by SEC. D9 and 44 were shown to be present in the same (monomeric) oligomeric state (Fig. 3b).

The VHH binding ability was then tested using an enzyme-linked immunosorbent assay (ELISA) confirming the VHH/antigen recognition (ESI,† Section S7 and Fig. S12), as well as through surface plasmon resonance (SPR). In SPR the VHHs were used as analytes, after the immobilization of the HER2-Fc *via* amine coupling. In this setup, the experimental K_d calculated for D9 was in the low micromolar range (448 nM, Fig. 3c) and, while **21** did not show any improvement over **D9** and was ill behaved at the SPR, VHH **44** showed a fivefold K_d improvement over **D9** with a K_d in the nanomolar range (98 nM, Fig. 3d).

A number of factors contributed to achieving this result (Fig. 3e and f): (i) most of the original structural pattern of the binding to HER2 was maintained during the optimization: this pattern contains two hydrogen bonds and five out of six hydrophobic contacts; (ii) new strong directional interactions were formed: for example, the two new arginines instead of glycines at positions 25 and 26 now allow for hydrogen bonding with HER2 Cys55 and Glu88 (see Fig. 3f); (iii) van der Waals interactions were maximized passing from six hydrophobic contacts in D9 to ten contacts in the final complex. Overall, the binding affinity increase shifted the original VHH K_d closer to the median of the K_d distribution of the Structural Antibody Database (SAbDab),[8] approximately located at 10 nM. Given the *a priori* lack of structural information (VHHs were built by homology modelling and complexes by docking) the achieved result is in itself remarkable. This highlights the importance of performing a correct sampling of the binding conformations along the design protocol to obtain accurate predictions of the mutants' binding affinities.

In this work we have employed a cleaved domain to represent the target epitope. The length of such a sequence will certainly affect the optimization process but a trade-off must be sought between available computing power (which might require a smaller fragment to be employed) and accuracy of the model (which might require describing explicitly the surrounding protein matrix).

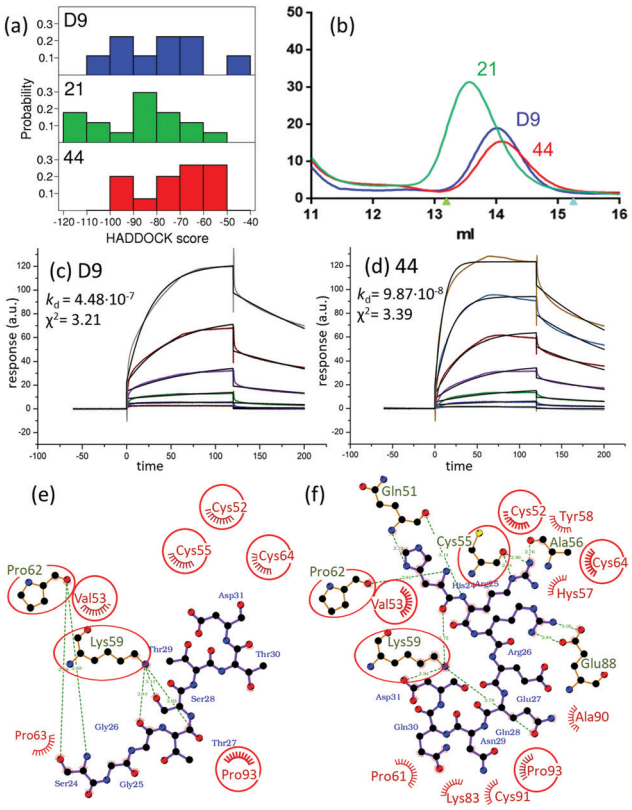

Fig. 3 (a) Histogram of the probability of homodimer binding events for **D9**, **21**, and **44** sorted by their HADDOCK score. More negative scores are associated with aggregation likelihood; (b) their superimposed SEC profiles (**D9** blue lines, **21** green lines, and **44** red lines), and molecular weight protein marker retention volume are also indicated (cytochrome *c*, 12.4 kDa, dark green and aprotinin, 6.4 kDa, cyan); (c and d) SPR profiles and affinity comparison between **D9** and **44**, and curves were fitted by using the steady state affinity method (for **D9**) and the 1 : 1 Langmuir equation (**44**). (e and f) Schematic diagram of the interaction between HER2 and (a) **D9** and (b) **44**: protein residues interacting with the target by van der Waals interactions are highlighted in red, while hydrogen bonds are highlighted in green. Equivalent HER2 binding amino acids are highlighted by circles.

Being aware of the current computational limitations in achieving high accuracy in the computational design, our protocol can be applied to practically any binder-target system, as similar old versions of the design approach have been used for peptide–drug[19,20] and peptide–protein[21,22] systems. The design of a complete antibody using our protocol requires more computational resources than structure-based combinatorial approaches.[8,18] However, high performance computing resources are more and more commonly available, making our approach a viable option if one wants to perform a high accuracy design, based on a quantitative control of the interaction pattern predicted by molecular dynamics.

This work was funded by the Italian Association for Cancer Research (AIRC 18510 and 12214), and the Javna agencija za raziskovalno dejavnost Republike Slovenije (ARRS/N4-0046 and J4-9322). We acknowledge CINECA (HP10BTJPER, 2017) and PRACE (2010PA3749) for HPC access to Marconi Broadwell and Marconi KNL (CINECA, Italy), and Hazelhen (HLRS, Germany).

Conflicts of interest

There are no conflicts to declare.

Notes and references

1 P. Sormanni, F. A. Aprile and M. Vendruscolo, *Chem. Soc. Rev.*, 2018, **47**, 9137.
2 P. Holliger and P. J. Hudson, *Nat. Biotechnol.*, 2005, **23**, 1126.
3 (*a*) H. L. Perez, P. M. Cardarelli, S. Deshpande, S. Gangwar, G. M. Schroeder, G. D. Vite and R. M. Borzilleri, *Drug Discovery Today*, 2014, **19**, 869; (*b*) J. M. Reichert, *mAbs*, 2016, **8**, 197; (*c*) J. M. Reichert, *mAbs*, 2017, **9**, 167; (*d*) H. Y. Yoon, S. T. Selvan, Y. Yang, M. J. Kim, D. K. Yi, I. C. Kwon and K. Kim, *Biomaterials*, 2018, **178**, 597.
4 A. de Marco, *Protein Expression Purif.*, 2018, **147**, 49.
5 T. Pleiner, M. Bates, S. Trakhanov, C.-T. Lee, J. E. Schliep, H. Chug, M. Böhning, H. Stark, H. Urlaub and D. Görlich, *eLife*, 2015, **4**, e11349.
6 B. Traenkle and U. Rothbauer, *Front. Immunol.*, 2017, **8**, 1030.
7 M. A. Soler, A. de Marco and S. Fortuna, *Sci. Rep.*, 2016, **6**, 34869.
8 P. Sormanni, F. A. Aprile and M. Vendruscolo, *Proc. Natl. Acad. Sci. U. S. A.*, 2015, **112**, 9902.
9 M. S. Pooja, M. Bunyarit, W.-Z. Dujduan, K. B. Weyant, I. Kocer, D. C. Butler, A. Messer, F. A. Escobedo and M. P. DeLisa, *Sci. Rep.*, 2018, **8**, 17611.
10 S. Muyldermans, *Annu. Rev. Biochem.*, 2013, **82**, 775.
11 (*a*) T. Vreven, H. Hwang and Z. Weng, *Protein Sci.*, 2011, **20**, 1576; (*b*) S. Viswanath, D. V. S. Ravikant and R. Elber, *Proteins*, 2013, **81**, 592; (*c*) F. Fogolari, A. Corazza, V. Yarra, A. Jalaru, P. Viglino and G. Esposito, *BMC Bioinf.*, 2012, **13**, S18; (*d*) C. Dominguez, R. Boelens and A. M. J. J. Bonvin, *J. Am. Chem. Soc.*, 2003, **125**, 1731; (*e*) E. Sarti, D. Granata, F. Seno, A. Trovato and A. Laio, *Proteins*, 2015, **83**, 621; (*f*) N. Andrusier, R. Nussinov and H. J. Wolfson, *Proteins*, 2007, **69**, 139.
12 (*a*) F. Fogolari, C. J. Dongmo Foumthuim, S. Fortuna, M. A. Soler, A. Corazza and G. Esposito, *J. Chem. Theory Comput.*, 2015, **12**, 1; (*b*) F. Fogolari, A. Corazza, S. Fortuna, M. A. Soler, B. VanSchouwen, G. Brancolini, S. Corni, G. Melacini and G. Esposito, *PLoS One*, 2015, **10**, e0132356.
13 M. M. Gromiha, K. Yugandhar and S. Jemimah, *Curr. Opin. Struct. Biol.*, 2017, **44**, 31.
14 M. A. Soler, S. Fortuna, A. de Marco and A. Laio, *Phys. Chem. Chem. Phys.*, 2018, **20**, 3438.
15 I. A. Guedes, C. S. de Magalhães and L. E. Dardenne, *Biophys. Rev.*, 2014, **6**, 75.
16 (*a*) T. Zou, F. Dembele, A. Beugnet, L. Sengmanivong, S. Trepout, S. Marco, A. de Marco and M.-H. Li, *J. Biotechnol.*, 2015, **214**, 147; (*b*) E. Ambrosetti, P. Paoletti, A. Bosco, P. Parisse, D. Scaini, E. Tagliabue, A. De Marco and L. Casalis, *ACS Omega*, 2017, **2**, 2618.
17 (*a*) W. J. Van Patten, R. Walder, A. Adhikari, S. R. Okoniewski, R. Ravichandran, C. E. Tinberg, D. Baker and T. T. Perkins, *ChemPhysChem*, 2018, **19**, 19; (*b*) J. Vilhena, A. Dumitru, E. T. Herruzo, J. I. Mendieta-Moreno, R. Garcia, P. Serena and R. Pérez, *Nanoscale*, 2016, **8**, 13463.
18 (*a*) G. D. Lapidoth, D. Baran, G. M. Pszolla, C. Norn, A. Alon, M. D. Tyka and S. J. Fleishman, *Proteins*, 2015, **83**, 1385; (*b*) K. C. Entzminger, J.-M. Hyun, R. J. Pantazes, A. C. Patterson-Orazem, A. N. Qerqez, Z. P. Frye, R. A. Hughes, A. D. Ellington, R. L. Lieberman, C. D. Maranas and J. A. Maynard, *Sci. Rep.*, 2017, **7**, 10295; (*c*) R. Chowdhury, M. F. Allan and C. D. Maranas, *Antibodies*, 2018, **7**, 23; (*d*) J. Adolf-Bryfogle, O. Kalyuzhniy, M. Kubitz, B. D. Weitzner, X. Hu, Y. Adachi, W. R. Schief and R. L. Dunbrack, Jr., *PLoS Comput. Biol.*, 2018, **14**, e1006112.
19 M. Del Carlo, D. Capoferri, I. Gladich, F. Guida, C. Forzato, L. Navarini, D. Compagnone, A. Laio and F. Berti, *ACS Sens.*, 2016, **1**, 279.
20 F. Guida, A. Battisti, I. Gladich, M. Buzzo, E. Marangon, L. Giodini, G. Toffoli, A. Laio and F. Berti, *Biosens. Bioelectron.*, 2018, **100**, 298.
21 A. Russo, P. L. Scognamiglio, R. P. Hong Enriquez, C. Santambrogio, R. Grandori, D. Marasco, A. Giordano, G. Scoles and S. Fortuna, *PLoS One*, 2015, **10**, e0133571.
22 M. A. Soler, A. Rodriguez, A. Russo, A. F. Adedeji, C. J. Dongmo Foumthuim, C. Cantarutti, E. Ambrosetti, L. Casalis, A. Corazza, G. Scoles, D. Marasco, A. Laio and S. Fortuna, *Phys. Chem. Chem. Phys.*, 2017, **19**, 2740.
23 R. D. Fisher, M. Ultsch, A. Lingel, G. Schaefer, L. Shao, S. Birtalan, S. S. Sidhu and C. Eigenbrot, *J. Mol. Biol.*, 2010, **402**, 217.

