Computational Evolution Protocol for Peptide Design

Rodrigo Ochoa, Miguel A. Soler, Ivan Gladich, Anna Battisti, Nikola Minovski, Alex Rodriguez, Sara Fortuna, Pilar Cossio, and Alessandro Laio

Abstract

Computational peptide design is useful for therapeutics, diagnostics, and vaccine development. To select the most promising peptide candidates, the key is describing accurately the peptide–target interactions at the molecular level. We here review a computational peptide design protocol whose key feature is the use of all-atom explicit solvent molecular dynamics for describing the different peptide–target complexes explored during the optimization. We describe the milestones behind the development of this protocol, which is now implemented in an open-source code called PARCE. We provide a basic tutorial to run the code for an antibody fragment design example. Finally, we describe three additional applications of the method to design peptides for different targets, illustrating the broad scope of the proposed approach.

Key words Peptide design, *In silico* antibody maturation, Molecular dynamics, Consensus scoring functions, Sensor technology, Evolutionary algorithm, Antibody design, Affinity optimization

1 Introduction

The design of synthetic peptides is unanimously considered of enormous potential for biomedical applications, in the emerging field of nanomedicine [1–3] as well as in medicinal chemistry [4]. Their versatility enables their use as alternatives to antibodies in targeted drug delivery and biomarker detection [5, 6]. Indeed, like antibodies, they can be mounted on detection devices or on nanoparticles to form ordered capturing arrays [7–10]. They can display pharmacological activity [11–15] and can be employed as modulators of protein/protein interactions [16, 17], with lower adverse effects and a higher binding specificity with respect to traditional drugs [18]. All these applications rely on the possibility to identify suitable hits.

The state of the art of peptide design is strongly rooted on biotechnology. Phage display library screening is used to assess

interactions between different types of macromolecules, including peptides [13, 19]. With this technique, it is possible to massively screen potential peptide binders. If a binding partner is known, a suitable sequence corresponding to the minimal sub-domain responsible for binding can be extracted from the partner itself [17]. However, these approaches require specialized infrastructures and are expensive. A more cost-effective alternative, which has aroused in the last years, is *computational design*. Due to enormous advances in computer power and to a better understanding of the chemical properties of natural amino acids, it is nowadays possible to rationally design peptides or proteins with a high probability of being active *in vitro* and *in vivo* [20].

An advantage of computational techniques is that they can describe the binding mechanisms at the atomistic level, allowing for a rational supervision of their properties. For instance, they enable controlling at the molecular level the binding site on the target protein, enhancing the selectivity properties of the designed binders [21]. However, all these benefits do not come for free. Computational design of peptides and proteins requires the efficient exploration of the sequence space, the accurate description of the bound (and unbound) conformations, and an accurate prediction of the peptide–target binding affinities (Fig. 1).

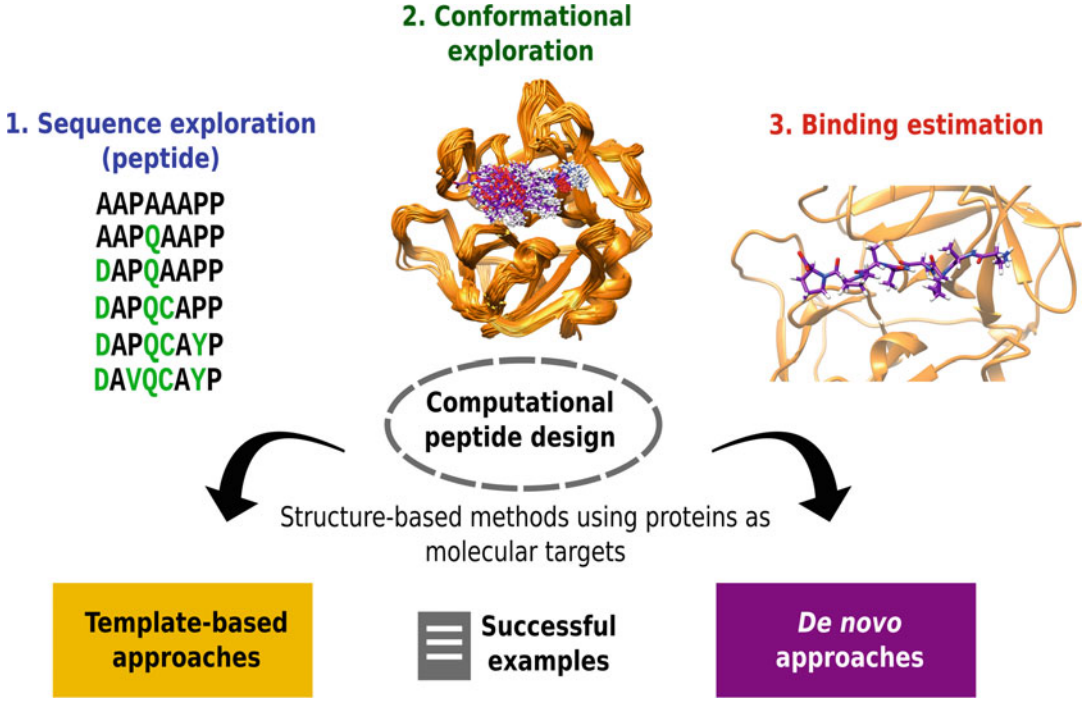

Fig. 1 The challenges in computational peptide design: exploring efficiently the sequence space, the bound conformations, and predicting the binding affinity of the protein–peptide complexes

Different strategies have been developed to face the challenges associated with computational peptide design. For example, the design can be performed using an *in silico* panning method with structural information [22], using a genetic algorithm [23] for sequence optimization. In this approach, conformational optimization and binding energy estimation are performed by a docking program. In [24], the authors use a Gaussian Network Model [25] for identifying the binding site and from that an approximate position of the peptide backbone. Then, they systematically attempt docking 400 dipeptides at the positions determined by this procedure in order to maximize the interaction energy, checking simultaneously the quality of the peptide conformation by characterizing the $\phi - \psi$ propensities of the dipeptides.

These approaches can be classified as template-based protocols. They are very computationally efficient but require the prior knowledge of the structure of a template. Reversely, *de novo* methods are computationally more expensive but can be used also when a template is not available (Fig. 1). This is the case of the pepspec protocol [26] included in the Rosetta software suite [27]. The pepspec tool follows a strategy of "anchor and grow" flexible backbone docking by starting from one key residue and optimizing from this point peptide sequences and structures [28]. Another example is the VitAl approach [29], which generates the peptides by sequentially docking a pair of residues and selecting the best fit by scoring the binding energies with AutoDock. PepComposer [30] retrieves patches from a database similar to the query and peptide fragments that interact with these patches. It then merges these fragments into an initial proposal that is further optimized using a set of iterative mutations and controlled backbone movements. Our design algorithm, called PARCE (Protocol for Amino acid Refinement through Computational Evolution), belongs to this second group.

PARCE, like most other *de novo* computational approaches, generates successive single-point mutations on the peptide or protein binder sequence. Each mutation is then accepted or rejected by analyzing the behavior of the complex using explicit solvent molecular dynamics trajectories. This makes the approach much more computationally expensive than the design schemes based on docking, but at the same time it enables describing the conformational changes induced by the binding with a level of accuracy which is only limited by the quality of the force field used in the simulation.

Chapter Overview *In the following sections, we describe in detail the design protocol implemented in PARCE [31]. We then provide a detailed tutorial and manual-like example for the design. Afterward, three additional applications of the protocol are presented. Finally, we discuss the open problems that require further development for the code and the field.*

The evolution of PARCE until its current version can be summarized in the timeline of Fig. 2. The original idea of this design approach was proposed by Laio and collaborators in 2012 [32], as an *in silico* mutagenesis platform for the optimization of amino acid-based binders. The approach can be used to design not only peptides but also antibody fragments, or other proteins whose amino acid sequence requires accurate engineering for applications in biosensing, biomedicine, and bioengineering. The protocol explores the sequence space of peptides bound to proteins or small molecules, using a Monte Carlo approach that integrates various simulation and prediction techniques [33].

The method, already in its original formulation [32], was based on a sequence of single-point mutations. One of the first applications was the design of peptides capable of binding with high affinity to an organic molecule in a denaturating solvent [34]. The idea of performing MD in explicit solvent was originally motivated exactly by this need, but it was afterward adopted in general, also when the design is performed in water. In Ref. 34, the quality of the complexes was estimated by computing the average value over the trajectory of a single suitable scoring function (Vina [35] for this case). This approach was refined and specialized to the design of protein binders in Ref. 36. The process is repeated many times, with the aim to evolve the original sequence toward novel sequences with predicted better affinities toward their targets [21, 32, 34, 36–38]. In 2019, Ref. 39 introduced another key idea: the estimation of the suitability of an attempted mutation to be carried out by a consensus mechanism using a set of binding

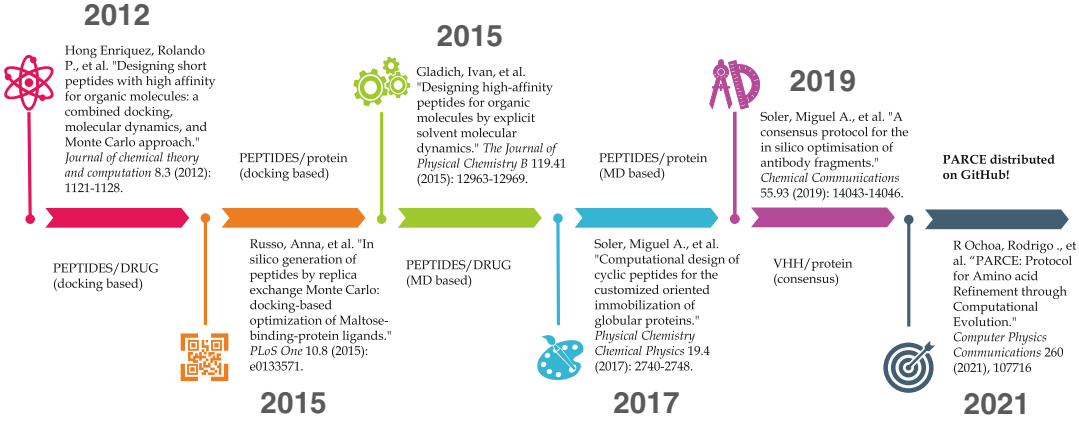

Fig. 2 The PARCE timeline describing the main development milestones and the publications supporting the progress of the peptide design protocol

scoring functions [40]. This makes the results much less dependent on the accuracy and the quality of a single scoring function. The approach has been successfully used to design peptides and protein fragments capable of binding to protein targets [21, 36, 39].

The PARCE code is distributed as an open-source software (https://github.com/PARCE-project/PARCE-1) and enables designing peptides or proteins capable of binding with higher affinity to a generic target, as long as this can be accurately described by a classical force field. The method, in its current formulation, combines several computational biophysics and bioinformatics tools, in order to achieve an equilibrium between accuracy and computational efficiency. In general, in a design run, one obtains several peptide candidates, whose number can be increased by performing several statistically independent runs (if enough computational resources are available). This increases the pool of sequences for further filtering and validation and is an advantage against more deterministic or brute force alternatives. Moreover, since the code is open source, it is possible to adapt it according to the research project needs. A graphical representation of the protocol and the required dependencies are shown in Fig. 3. In the following, we explain its main steps.

2.1 Mutation Protocol

The core of the algorithm is an iterative sequence optimization. At every iteration step, a single-point mutation on the peptide sequence is generated by selecting at random a position along the peptide chain and by replacing the selected residue by a random amino acid (*i.e.*, a mutation). A key element of the algorithm is generating a reliable structure of the mutant. If the mutated side chain is placed incorrectly, it will likely make severe steric clashes with other side chains or with the target. To heal these clashes, it would be necessary to perform very long MD equilibrations, which are not affordable. In PARCE, the configuration of the mutated amino acids can be generated either with the programs Scwrl4 [41] or with FASPR [42]. These approaches were selected based on a study that assessed if a mutation protocol is able to predict amino acid rotamers similar to those that would be generated in a long MD run [43]. After performing the mutation, a first minimization of just the predicted side chain is performed. In order to avoid clashes between the mutated amino acid and the surrounding water molecules, a second minimization of the new amino acid and the water molecules surrounding it within 2 Åis carried out. Finally, a minimization of the full system is performed followed by an NVT equilibration of typically 100 ps. Other standard parameters employed in the equilibration are described in the next section. After the minimization and equilibration steps/phases, the new system is then sampled by performing an MD simulation.

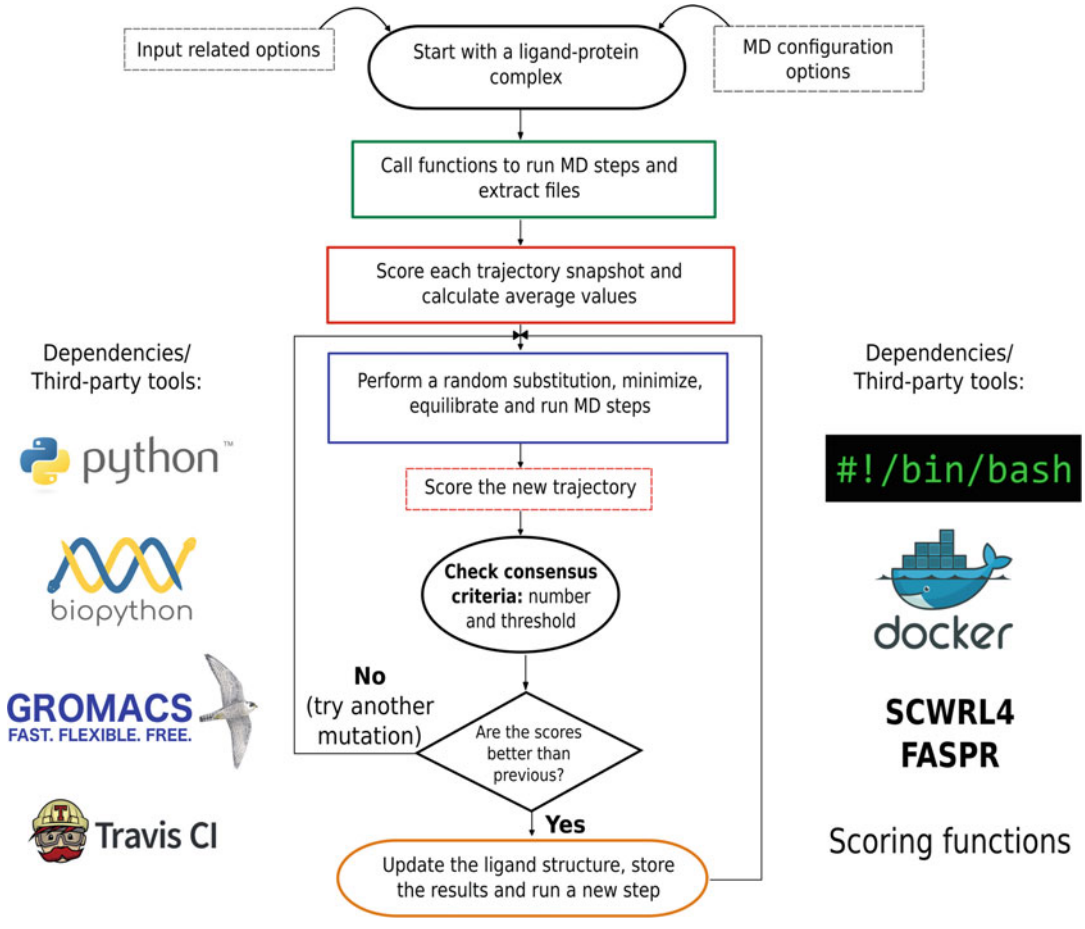

Fig. 3 Schematic representation of the PARCE pipeline. It includes four main phases: a single-point mutation, the conformational sampling of the new protein–binder complex, the scoring of the conformations of the new complex, and the acceptance or rejection of the mutation. The protocol is iterated to improve the binder sequence. Several-open source dependencies are required to run the protocol

2.2 Conformational Sampling with Molecular Dynamics

For each mutation, an MD simulation is run to sample the conformations of the complex. This step can be seen as the "fingerprint" of this design approach, the feature that makes it different from most other design strategies. The specific force filed can be chosen based on the experience of the user, and the MD setup can be adapted to the physico-chemical conditions of the environment in which the binding should happen. For example, for a design of peptides capable of binding a protein in water solution at ambient conditions, one can use the Amber99SB-ILDN protein force field [44], a TIP3P water model [45], a modified Berendsen thermostat [46], and a Parrinello–Rahman barostat [47]. In general, the complex formed by the peptide–protein and its target is solvated in a cubic box with periodic boundaries at a distance of at least 8 Å from any atom of the complex. By default, Na^+ and Cl^- counterions are included in the solvent to make the box neutral, but the

concentration and the ion type can be easily changed to take into account a specific ionic strength. In general, the electrostatic interactions are calculated by using the Particle Mesh Ewald (PME) method, with 1.0 nm short-range electrostatic and van der Waals cutoffs [48], and the equations of motion are solved with the leap-frog integrator [49], using a timestep of 2 fs.

2.3 Scoring and Mutation-Acceptance Strategies

After performing the mutated peptide–protein (or peptide–ligand) conformational sampling, the trajectory is scored with a chosen set of scoring functions used for protein–protein, protein–peptide, or protein–ligand affinity predictions. The mutation can be accepted or rejected based on three different strategies, outlined in the following.

2.3.1 Monte Carlo Optimization

The most simple optimization strategy is based on Monte Carlo and on the use of a single scoring function for estimating the binding affinity. At each step of the mutation cycle, the peptide chain is randomly mutated selecting one amino acid from the sequence and replacing it with a different amino acid. The protocol offers the possibility to select the amino acid positions in the peptide chain that are eligible for mutations, as well as the list of possible amino acids selected for the replacement. For example, in Ref. 34, the design is performed on cyclic peptides: the terminal CYS positions were never mutated in order to conserve the cyclic geometry, while GLY was removed from the amino acid list used for the replacement, avoiding undesired mobility in the new peptide chain.

After each mutation step, meaningful conformations of the mutated peptide/target complex in explicit solvent are generated by finite temperature MD, employing the methodology described in the previous section. The binding affinity of the mutated peptide toward the target ligand is then estimated using a single scoring function. In Ref. 34, a cluster analysis is performed over the last part of the trajectory (the last 1 ns of a 5 ns NPT production run) to extract statistically relevant conformations of the peptide–ligand complex. Poorly populated clusters were discarded (clusters with less than 15 conformations), while for the central structure of the remaining clusters, the peptide–ligand affinities were scored employing the Vina scoring function [35]. In the PARCE implementation, the cluster analysis is not performed anymore, and the binding affinity is simply estimated as the average value of the scoring function on the whole MD trajectory, neglecting only its first part (whose length can be set by the user).

The new peptide sequence at step k is accepted or rejected based on the Metropolis criterion, with a probability

$$\min\left(1, \exp[(E_k - E_{k-1})/T_e]\right), \tag{1}$$

where E_{k-1} is the estimated binding affinity before the mutation, E_k is the binding affinity after the mutation, and T_e is an efficacious temperature that controls the acceptance rate. If the sequence is accepted, a new mutation cycle is started from the mutated sequence; otherwise, the former sequence from step $k-1$ is used as starting point for a new mutation attempt. The mutation cycle described above is iterated up to a desired number of mutations.

2.3.2 Replica Exchange Optimization

The exploration of the sequence space can be increased using a replica exchange scheme by running simultaneous and independent mutation cycles at many different efficacious temperatures. At the end of each step, a swap between two randomly selected replicas (*e.g.*, r and r') is attempted. The swap is accepted according to a parallel tempering scheme,

$$\min(1, \exp)\left((E_r - E_{r'})\left(\frac{1}{T_r} - \frac{1}{T_r}\right)\right), \tag{2}$$

where E_r and E'_r are the peptide/target binding energy in replicas r and r', and T_r and T'_r are the efficacious temperatures. If the swap is accepted, the replica indexes are swapped. The replica exchange scheme is not currently implemented in PARCE.

2.3.3 Consensus Optimization

The two optimization approaches described above attempt optimizing the binding affinity estimated by *a single scoring function*. If, for example, one estimates the binding affinity with Vina, the evaluation is based on counting the number and type of peptide–ligand contacts and providing, for each of them, an energy value assuming that the complex is fully solvated in an aqueous environment [35]. For this reason, binding affinities were not necessarily meaningful in non-aqueous environments and the scoring was used with the only intent to screen the most viable peptide–ligand complex observed during the MD trajectory.

These limitations motivated us to improve the scoring scheme of the protocol with a consensus optimization scheme. In this strategy, the mutation is accepted following a consensus-based approach using N scoring functions. If a particular number n of scoring functions agrees on an improvement of the binding affinity of the mutated peptide B, with respect to the one prior to the mutation, *i.e.*, peptide A, then the final consensus will accept the attempted mutation [39]. Formally, the consensus score C is defined as

$$C = \sum_{k=1}^{N} c_k \,, \tag{3}$$

where c_k for the scoring function k is

$$c_k = \begin{cases} 1, & S_k^B - S_k^A < 0 \,, \\ 0, & otherwise, \end{cases} \tag{4}$$

where S_k^I is the value of the average score for peptide I. It should be noted that all employed scoring functions are defined as binding energies, so that lower values mean higher binding affinities. The criterion to evaluate if a consensus among the scoring functions is achieved is based on the comparison of C to a predefined threshold T (with a value between 1 and N). If $C \geq T$, the mutated sequence is accepted. The scores are estimated as an average over all the snapshots of the trajectory.

The next section presents a tutorial on installing and running PARCE and an example guide on designing a nanobody paratope region bound to a protein fragment.

3 Tutorial

3.1 Installing and Running PARCE

PARCE can be downloaded from https://github.com/PARCE-project/PARCE-1 and installed under any Linux operating system. A `README` file with instructions is included in the repository. The code has been initially optimized for Debian and Ubuntu OS server distributions. We note that all the dependencies required to run PARCE are open-source software, but some of them, such as Scwrl4 [41], require academic licenses. In such cases, it is recommended to install these packages following the developer's documentation to integrate their paths to the code. To guarantee that the additional tools and dependencies are functioning, a set of tests is provided in the repository. A docker container is also available in case the user wants to skip the installation of third-party tools.

After installing PARCE, one has to set up the configuration file that contains instructions to start the system and launch the protocol. It describes the path and the characteristics of the input files, as well as the necessary parameters to run the design protocol. An explanation of the input parameters is provided in Table 1.

Before running the protocol, we recommend performing an equilibrated MD simulation of the initial complex. Then, the protocol is run by the command:

```
python3 run_protocol . py [- h ] - c CONFIG_FILE
```

The design protocol results are summarized in the output file called `mutation_report.txt`, which contains details per mutation step, like the type of mutation, the average scores, the binder sequence, and if the mutation was accepted or not. The mutation is defined by the syntax: `[old amino acid][binder chain][position][new amino acid]`. An example of a mutation is AB2P, which means that an alanine located in the position number 2 of the chain B is replaced by a proline.

Table 1
Parameters provided by the user in the configuration file

Parameter	Explanation
Folder	Name of the folder that has all the input and output files of the protocol
src_route	Route of the PARCE folder where the src folder is located
Mode	The design mode, which has three possible options, including start and restart modes
peptide_reference	The sequence of the peptide, or protein fragment that will be modified
pdbID	Name of the structure that is used as input
Chain	Chain id of the peptide/protein in the structural complex
sim_time	Time in nanoseconds that will be used to sample the complex after each mutation
num_mutations	Number of mutations that will be attempted
residues_mod	These are the specific positions of the residues that want to be modified.
md_route	Path to the folder containing the input files used during the previous MD sampling of the system
md_original	Name of the system file located in the folder containing the previous MD sampling
score_list	List of the scoring functions that will be used to calculate the consensus.
half_flag	Flag that controls which part of the trajectory is used to obtain the average score.
Threshold	Threshold used for the consensus scoring.
mutation_method	Protocol to perform the single-point mutations
scwrl_path	Provide the path to Scwrl4 in case it is not installed in a PATH folder.
gmxrc_path	Provide the path to GMXRC for Gromacs

In addition, the report file includes failed attempts based on minimization or equilibration problems in MD. To overcome these issues, the protocol automatically attempts a number of mutations using the last accepted structure. If the simulation keeps failing after a certain number of attempts (defined by the user in the input file with the keyword `try_mutations`), a new mutation will be attempted but using the complex that was accepted previous to the current one. If the problem persists more than the number of `try_mutations`, the design run is stopped. If the protocol is successful, the number of attempted mutations is decided by the key word `try_mutations`.

PARCE has an MIT license that allows for the distribution of the code and its improvement through new functionalities, for example, for adding new scoring functions. We note that the computational resources required for running PARCE are determined by the complexity of the system, since the design is based on running MD. HPC versions of the code are available upon request.

Fig. 4 The structure of the VHH antibody fragment. The peptides to be optimized correspond to the complementary determining regions (highlighted in red). The framework sequence (yellow) is not mutated throughout the whole process

3.2 Tutorial Example: The Optimization of Anti-HER2 Antibody Fragments

The human epidermal growth factor receptor 2 (HER2) is a transmembrane protein whose overexpression is associated with specific classes of breast cancer and is thus a widely recognized biomarker employed for monitoring cancer progression, as well as a key pharmacological target for cancer therapy [50, 51].

For this particular example, the goal is to design a novel antibody fragment of camelid origin (or VHH, Fig. 4) capable of detecting HER2 in a patient's biological fluids [8, 10]. The idea is to optimize a peptide, or a set of peptides, already embedded into an existing protein to recognize the target. In particular, we aim to design peptide fragments that are part of the antibody binding domain, also known as complementary determining regions (CDRs). This process, called antibody maturation, is usually performed *in vivo*, by animal immunization. Using an *in silico* process reduces the use of animals for binder discovery. A further advantage of the computational design is that it enables choosing *a priori* the binding site on the target protein. The selection of the binding site (or epitope) to be targeted is of paramount importance for the development of new nanodevices, for targeted therapies, and for drug design [52].

An example of VHH optimization performed by PARCE is described in Ref. 39. Here, we show how to set up the design and how it is possible to employ a different set of scoring function to obtain an *ex novo* designed antibody fragment for an arbitrarily selected binding site on HER2.

3.2.1 Design Methodology

To get started with any design, it is necessary to have a reasonable starting model of the initial complex. This can be done using either a crystallographic complex or a conformation obtained by docking

the binder to the target. If there are no experimental 3D structures available, these can be constructed by homology modelling. We also remind that, when working with a new system and before getting started with the optimization, all scoring functions should be benchmarked over the particular system [40, 53].

As a second step, it is necessary to identify the residues that should be mutated. In the case of an antibody fragment, these can be the residues belonging to either one or two or all three CDRs (highlighted in Fig. 4). Only this selected region will be optimized, leaving the sequence of the rest of the protein unchanged throughout the whole process. The input files for the design are the starting topologies for the MD and the configuration file. Examples of these files, aiming to reproduce the results of Ref. 39, can be found in the folder `design_input/protein_protein`. The `CONFIG_FILE` contains the input parameters shown in Table 1.

The individual peptide residues to be optimized should be explicitly listed in the `CONFIG_FILE`. For instance, for the optimization of a single antibody fragment loop, the `config_vhh.txt` would read

```
residues_mod: 54,55,56,57,58,59,60,61
```

Instead, to optimize all the VHH residues highlighted in Fig. 4, namely residues 29–25 corresponding to the first CDR, 55–61 corresponding to the second CDR, and 101–109 corresponding to the third CDR, one should write

```
residues_mod: 29,30,31,32,33,34,35,55,56,57,58,59,60,61,101,
              102,103,104,105,106,107,108,109
```

simply listing all residues even if they belong to different regions of the system.

While the example of the optimization of a single CDR can be found in Ref. 39, here we show how the same process can lead to the optimization of all three VHH CDRs.

3.2.2 Design-Optimization Results

A typical optimization path is reported in Fig. 5a, where each score is a proxy measure of the binding affinity between the two components of the system, in that case, the whole VHH and its target. It is important to note that, even if only the selected residues are mutated, the score is calculated over the entire complex. An optimization is considered concluded when all scoring functions reach a plateau.

For a collective view of the whole optimization process, a rank-based analysis can be used. First, one computes the rank r_k^i associated with each sequence i according to the score obtained with a single scoring function k. Accordingly, r_k^i can be normalized as

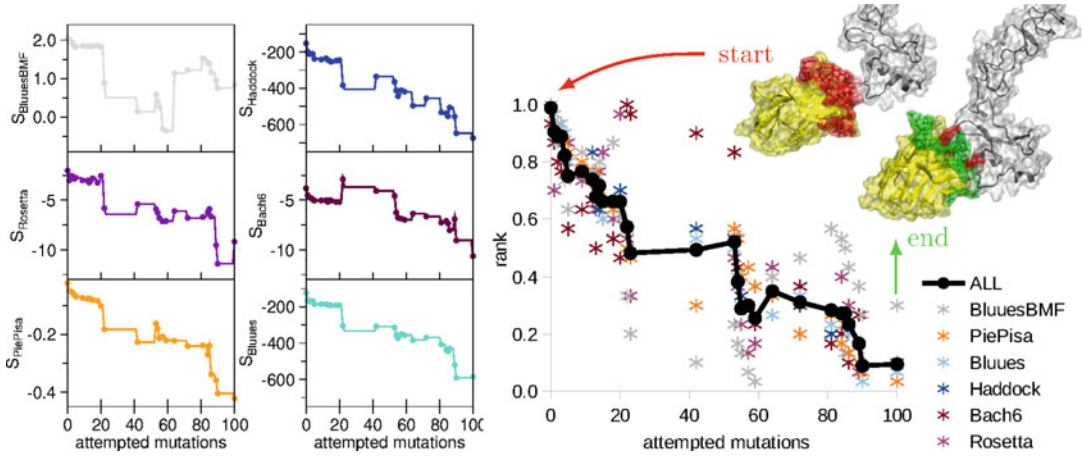

Fig. 5 Design of an antibody fragment (VHH) bound to the HER2 terminal domain. (**a**) Evolution of the six scoring functions during the design. The dots in the curve represent the mutations that were accepted. The scoring functions used are BMF-Bluues [75, 76] (gray), Rosetta [27] (magenta), PiePisa [73] (orange), Haddock [70] (blue), Bach6 [77, 78] (mauvre), and Bluues [76] (cyan). (**b**) Ranking of the configurations: both the single scoring function normalized ranks ($\hat{r}_k^i$, stars) and the global normalized rank R^i (black line) for each peptide *i*. In the insets, starting and final configuration of the VHH/HER2 complexes. Color code: HER2 (gray), VHH framework (yellow), starting residues (red), and optimized residues (green)

$$\hat{r}_k^i = \frac{r_k^i}{N} \ , \tag{5}$$

where N is the total number of accepted mutations obtained in the runs. From the collection of $\hat{r}_k^i$ (indicated by stars in Fig. 5b), a global ranking score R^i for each sequence is defined (black dots in Fig. 5b) as

$$R^i = \sum_{k=1,\,N_s} \frac{\hat{r}_k^i}{N_s} ; \quad i = 1, N, \tag{6}$$

where N_s is the number of scoring functions. If the ranks of a certain sequence *i* are consistently low for all the scoring functions, then R^i is small.

In the particular case illustrated in Fig. 5, R^i decreases when more mutations are performed, as expected. By comparing the initial and the final configurations of the system, the former with sequence associated with $\max(R^i)$ and the latter with $\min(R^i)$ (insets in Fig. 5b), it is possible to see how the initial VHH evolves into a final VHH by changing its orientation to maximize its contacts with the target, defining a larger contact area between the two. The optimized VHHs, or better a selection of the lowest ranking sequences, will then need to undergo extensive MD simulations and stability tests [54]. VHHs passing all the computational tests will then be ready to be expressed in bacterial cells [55].

In the next section, several additional examples of peptide design are presented.

4 Additional Peptide Design Examples

4.1 Drug-Binding Peptide Design in Different Environments

Reference 34 was the first works performing peptide design in explicit solvent with our scheme. It reports the design of high-affinity cyclic peptides toward Ironotecan (CPT-11). CPT-11 is a chemotherapy drug, and its choice was motivated by the need of engineering sensors for therapeutic drug monitoring in denaturant solvent (e.g., methanol), which were afterwards validiated experimentally [37].

Compared to the original protocol of 2012 [32], three important innovations were introduced: (1) the conformational search for viable peptide–ligand conformations during the mutation cycle was carried on by finite-temperature molecular dynamics and not by flexible docking in vacuum, (2) cyclic peptides were adopted for the design, and (3) the design was performed with the peptide–ligand complexes fully solvated in a simulation box with an explicit atomistic description of the solvent molecules [34]. Computationally intense design in explicit solvent was made possible thanks to the advent of GPU-based computing, which started to be efficiently implemented in commonly used MD packages in those years.

The protocol adopted was basically the same employed in the current version of PARCE and described in Subheading 2, using in particular Replica Exchange optimization with a single scoring function (Vina [35]). Two independent designs were performed, one in water and one in methanol. The procedure started from a deca-alanine cyclized by a disulfide bridge between two terminal cysteines. CPT-11 was initially inserted within the cyclic peptide, and one randomly selected amino acid of the peptide chain was mutated at each step. The terminal cysteines were not selected for the mutation in order to conserve the cyclic geometry. After each mutation, MD simulations of 5 ns were performed for the peptide–ligand complex fully solvated in water (or methanol), and relevant peptide–ligand structures were selected from the last part of the MD trajectory by cluster analysis (*see* Subheading 2.3.1).

For the selected structures, the peptide–ligand affinities were estimated using the Vina scoring function, and the mutation was accepted or rejected according to a Metropolis criterion. To further enhance the exploration of the sequence space, a replica exchange scheme with 5 effective temperatures was employed, as described in Subheading 2.3.2 (Fig. 6a). After ~400 mutations, the best seven peptide–ligand complexes, in terms of binding energies, were selected and their stability further assessed by longer (at least 100 ns) MD simulations in explicit solvent at different temperatures (i.e., from 300 K up to 450 K).

The designed peptides revealed a solvent specificity, namely peptides designed in aqueous environment do not necessarily bind the ligand in a different solvent, such as methanol. This is

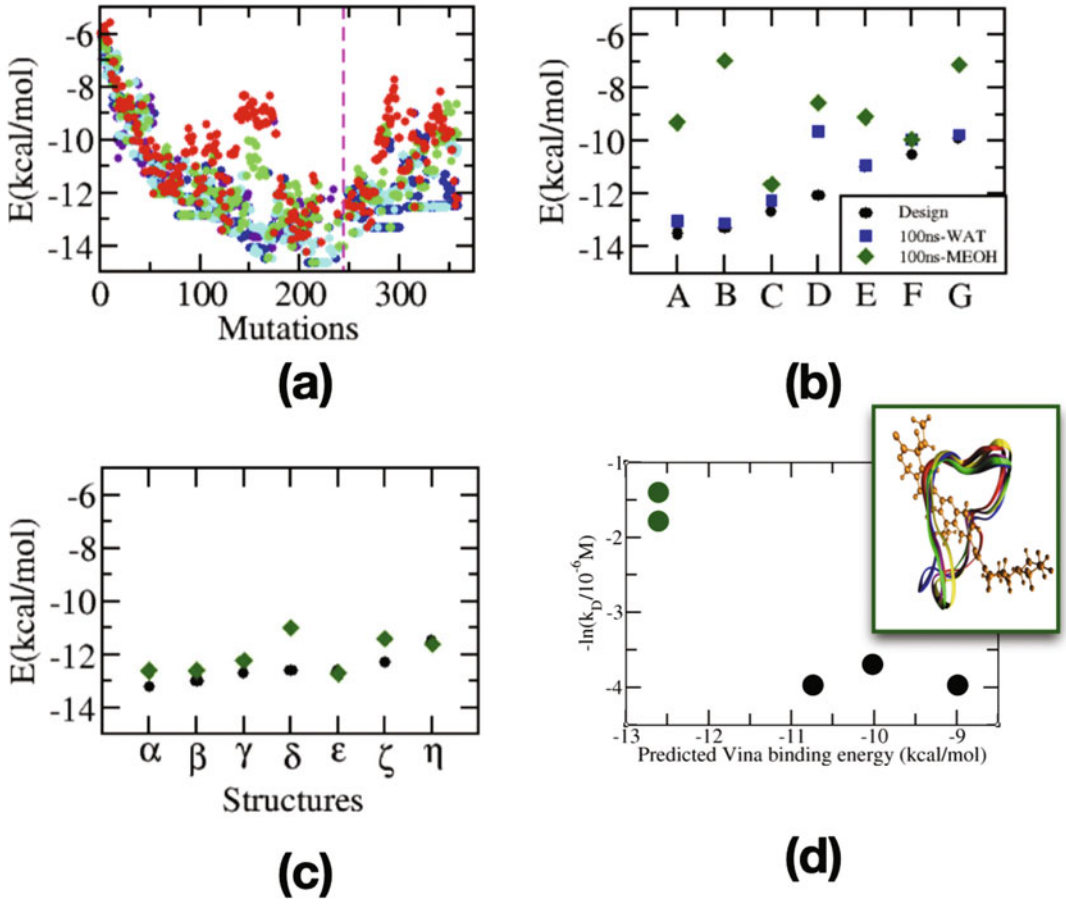

Fig. 6 Design of peptides for CPT-11 in aqueous and methanol solutions. (**a**) The Vina scoring as a function of the mutation steps for the design in water. The five different colors report the binding affinities observed at the five effective temperatures employed during the procedure (see main text for details). (**b**) The best seven peptides (A-G), in terms of binding affinity toward CPT-11, from the design in water. Black dots are the binding energies as predicted during the mutation cycle, while in square blue after 100 ns MD at 300 K in water. The green diamond displays the binding affinity of the A-G peptides after 100 ns MD in methanol. (**c**) as panel (**b**) for the seven best peptides ($\alpha - \eta$) designed in methanol. (**d**) The experimental dissociation constant, k_D in methanol solution vs. the computationally predicted binding energies. The green dots show results for two peptides designed in methanol, black dots for three peptides designed in vacuum using the flexible docking approach [32]. The experimental values were taken from Ref. 37. In the green inlet, the peptide backbone around CPT-11 at 0, 25, 75, and 100 ns MD simulations for one of the peptides designed in methanol. Panels (**a**)–(**c**) adapted from Ref. 34, copyright 2015 American Chemical Society

evident in Fig. 6b: once peptides designed in water are solvated in methanol, the binding becomes weaker and, occasionally, some peptides can even detach from the ligand [34]. This solvent specificity is a consequence of the explicit description of the solvation environment during the design. Peptides created in water are, indeed, richer of aromatic residues than those designed in methanol: once peptides designed in water are immersed in methanol, the

aromatic side chains are more easily exposed to the solvent, competing with the binding toward the ligand.

The high affinity of the designed peptides in methanol has been confirmed experimentally in a follow-up paper using surface plasmon resonance and fluorescence spectroscopy [37]. The peptides displayed an experimental micromolar affinity toward CPT-11 in methanol solution, and MD simulations revealed peptide–drug complexes more stable in solvent than those designed in vacuum using flexible docking (Fig. 6d). Interestingly, the designed peptides were selective toward the target and unable to bind SN-38, an active metabolite of CPT-11 lacking of the carbamate and piperidyl-piperidine groups [37].

A similar procedure was also adopted to design peptides that bind chlorogenic acid (GCA), a compound present in coffee blends, in water solution [56]. Electrochemical measurements and circular dichroism and fluorescence spectroscopy certified the high affinity of the design, showing a remarkable peptide selectivity toward CGA and not to other related phenolic compounds [56].

4.2 Peptides for Protein Recognition

The protocol introduced in Ref. 34 was subsequently employed for the design of peptides for protein recognition [9, 10, 21, 38, 54, 57]. While the approach was still relying on a single scoring function, namely Vina [35], it allowed for an unprecedented versatility in the choice of the binding site. In particular, after having successfully designed linear peptides for a well-defined protein pocket with the docking based code [32, 33], the new approach introduced in Ref. 34 allowed designing ligands for surface-exposed binding sites (Fig. 7), which are generally regarded as "undraggable."

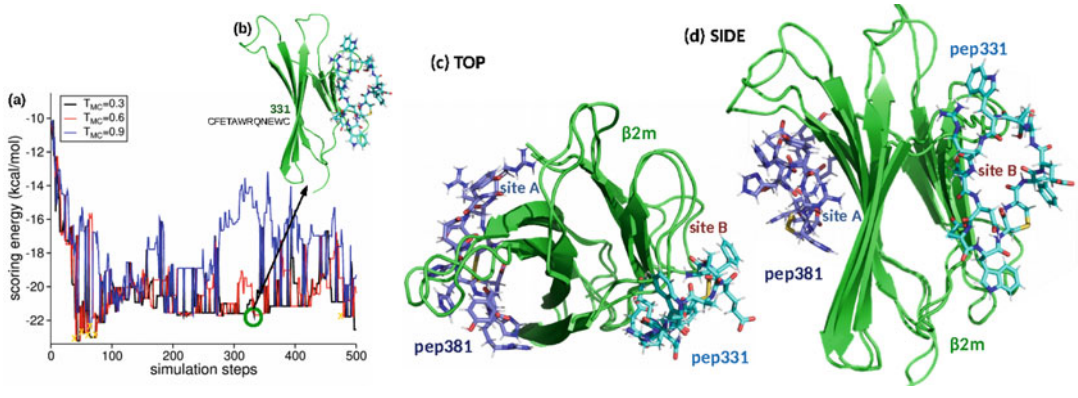

Fig. 7 Design of peptides for B2M *in vacuum*. (**a**) The Vina score as a function of the mutation steps. The three different colors report the binding affinities observed at the three effective temperatures employed during the procedure (see main text for details). Yellow crosses indicate peptides that underwent computational and experimental screening. The best peptide/protein complex is shown in (**b**). (**c**–**d**) Top and side view of the two computationally designed peptides discussed in the text. Adapted from Ref. 21

To show the versatility of PARCE in terms of choice of the target binding site, we selected two sites on opposite sides of a globular protein that does not possess pockets: the beta-2-microglobulin. Due to the large system size, the design was performed in vacuo, followed by a screening in explicit solvent using MD simulations and a final experimental validation.

The first binding site chosen for the design was a surface-exposed site, which is known to interact with the human histocompatibility antigen [54, 57]. Among the generated peptides, five were experimentally tested giving dose–response surface plasmon resonance (SPR) signals with dissociation constants in the micromolar range. The result was confirmed by means of isothermal titration calorimetry and nuclear magnetic resonance, showing that the approach is capable of designing binders for an arbitrarily selected binding site. We then identified another site on the opposite side of B2M and attempted to generate a second peptide (Fig. 7a). Once again SPR confirmed the dissociation constant to be in the micromolar range. Competition experiments further confirmed the two peptides to bind to non-overlapping binding sites, thus confirming the theoretical predictions (Fig. 7b and c).

We further showed that this design approach can be exploited for bottom-up design of smart nanodevices. Indeed, the peptides designed in these projects were employed as sensing elements to build a self-assembled nanochip capable of capturing a target protein by means of preselected binding sites [21], allowing for the immobilization of the chosen protein in a predefined orientation [54].

4.3 MHC Class II Peptide-Binder Design

The major histocompatibility complex (MHC) class II is a complex of encoding proteins responsible for regulating the immune system in humans [58] through the interaction with antigen proteins and peptide subunits. Different experimental and computational strategies have been implemented to predict affinities of peptides toward relevant alleles within the population [59], as a strategy for the development of more efficient vaccines [60]. The field known as immunoinformatics has provided an extensive set of tools, mainly sequence-based strategies, for predicting the affinity between a peptide and MHC class I or class II molecules [61]. However, structural information is also crucial to rationally study peptides bound to the MHC class II binding interface, which has been characterized by a large groove located between the solvent exposed α and β structural subunits [62] (Fig. 8). Specific interactions created between some protein pockets and core amino acids of the peptides contribute to the molecular affinity [63]. The latest has been correlated with immunogenic properties, as well as

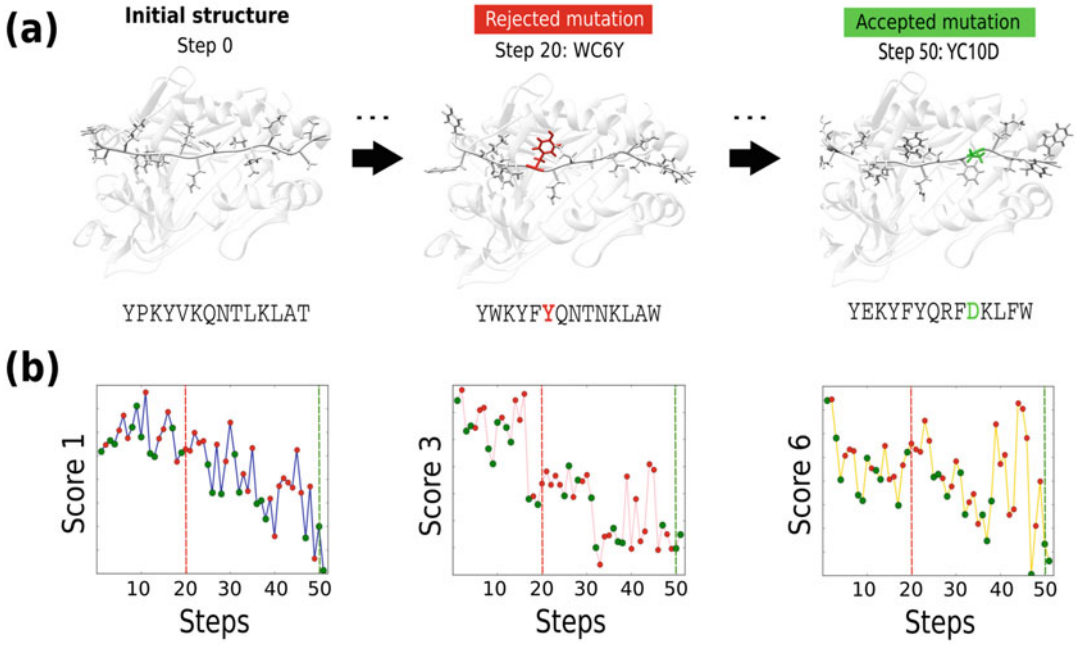

Fig. 8 Summary of the scoring strategy used in the design protocol. (**a**) The structure of MHC class II in complex with the peptide at step 0, and examples of a rejected mutation at the 20th step (colored in red) and of an accepted mutation at the 50th step (colored in green). (**b**) Representation of the accepted mutations (green circles) and the rejected (red circles), with the rejected and accepted examples depicted by dash lines

other events during the MHC editing process [64]. This motivates the use of structure- and dynamic-based approaches such as PARCE to engineer peptides with better affinities for this molecular receptor.

In this example, the starting complex for the design was the MHC class II allele DRB1:01*01 bound to a peptide of 14 amino acids, that is part of an influenza virus antigen (YPKYVKQNTLK LAT) (Fig. 8). This sequence has a reported bioactivity of $IC_{50} = 130nM$ from a curated dataset of peptide binders against multiple MHC class II alleles [65]. As reference, we used the crystal structure 1DLH [66] from the Protein Data Bank (PDB) [67] that has a missing tyrosine at the N-terminal flanking region. The missing amino acid was modelled using the Rosetta Remodel functionality [68] using the full protein–peptide complex as a template. The side chains of the complex were relaxed using Rosetta with the protein backbone fixed. The refined protein–peptide structure was equilibrated by an MD simulation of 100 nanoseconds (ns), with previous minimization and NVT/NPT equilibration, using Gromacs v5.1 [69]. Despite the linear conformation of the peptide in the bound state, the complex remains stable during the simulation,

mostly due by the hydrogen bonds between the receptor and peptide backbone atoms. The final snapshot of the MD was used as the starting conformation for the design.

We then applied the PARCE protocol explained in Subheading 2. Specifically, we configured the protocol to mutate randomly any amino acid of the peptide. We iterated the mutation process and sampled each mutated protein–peptide complex for 5 ns, at a temperature of 350 K. A high temperature was chosen to allow a more efficient exploration of the conformational space. All the protein atoms located at a distance greater than 12 Å from the peptide were restrained to keep the system stable at the selected temperature. The design was performed by using the consensus scheme described in Subheading 2.3.3, using six scoring functions that were previously benchmarked for this specific system [53]: Haddock [70], Vina [35], a combination of DFIRE and GOAP (DFIRE-GOAP) [71, 72], Pisa [73], FireDock [74], and BMF-BLUUES [75, 76]. The threshold parameter T (Subheading 2.3.3) was set equal to 3, following Ref. 39. This means that if 3 or more scoring functions predict better scores for the new mutation, the mutation is accepted. During the design, 100 mutations were attempted, with an acceptance ratio of around 20–30%. The evolution of the scores with accepted and rejected mutations is shown in Fig. 8. As expected, the mutations minimize the majority of the scoring functions scoring functions.

This specific example illustrates the usefulness of the consensus strategy with respect to the standard design strategy, which is based on a Monte Carlo optimization of a single scoring function [34] (*see* Subheading 4.1). Indeed, each scoring function is typically affected by errors, but very often the errors of different scoring functions are uncorrelated and might compensate. The consensus criterion allows complementing the different empirical and physics-based terms of the scoring functions [39]. As shown in Fig. 9, the scoring functions are, on average, minimizing their value through the trial of multiple mutations. Nonetheless, due to the nature of the stochastic search and the definition of the consensus score, the single scoring functions can also increase.

We note that to explore the sequence space, it might be beneficial to run multiple replicas of the protocol starting from the same initial complex but with different random seeds. In this case, the peptides from the different runs can be combined following the same re-ranking procedure using the average rank from all the scoring functions calculated from MD simulations (similarly to that described in Subheading 3.2.2).

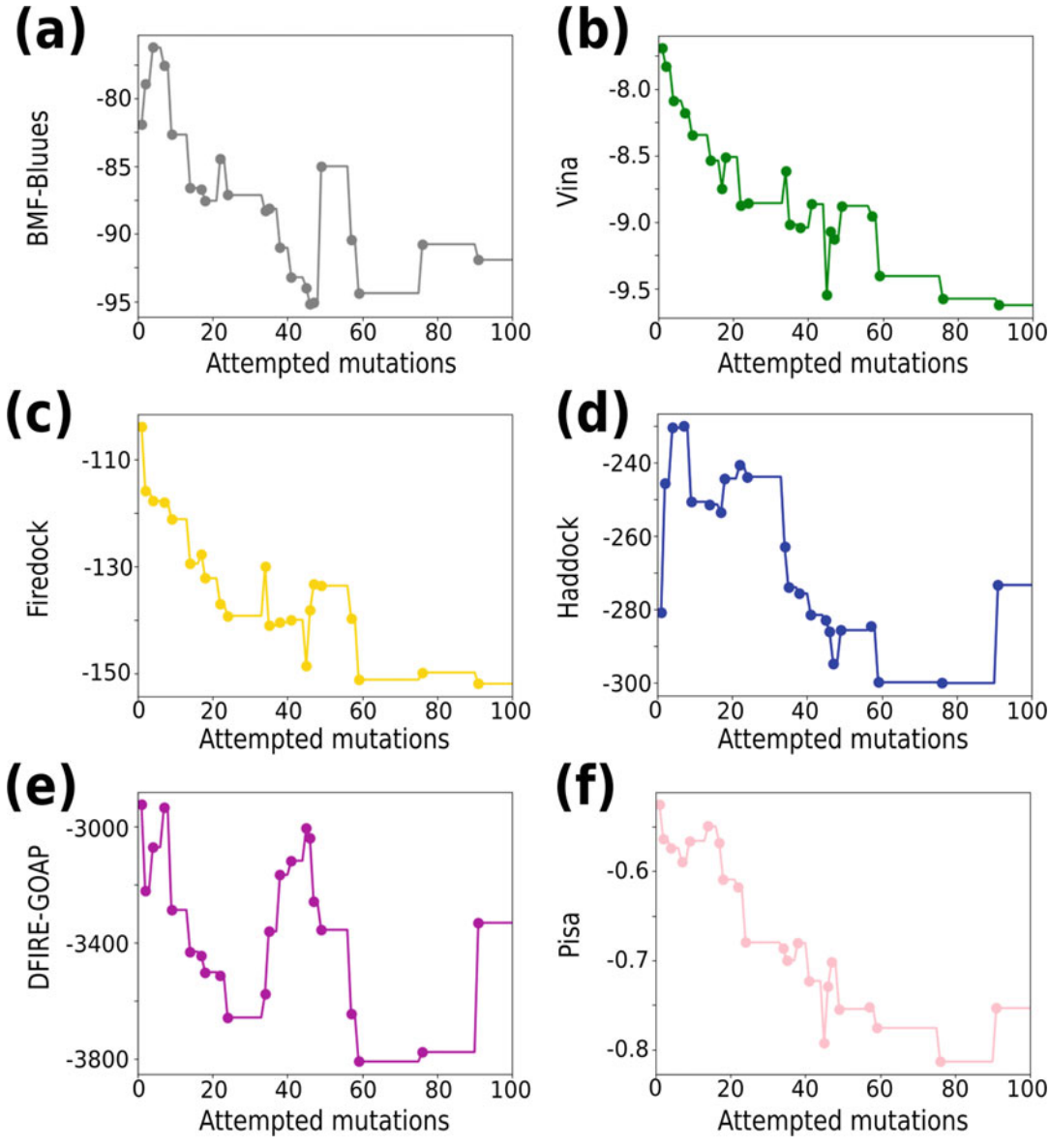

Fig. 9 Evolution of the scoring functions for the design of peptides bound to the MHC class II. We used six scoring functions to calculate the consensus. The dots in the curve represent the mutations that were accepted. The scoring functions used were (**a**) BMF-Bluues [75, 76], (**b**) Vina [35], (**c**) Firedock [74], (**d**) Haddock [70], (**e**) DFIRE-GOAP [71, 72], and (**f**) Pisa [73]

5 Concluding Notes and Perspectives

The exponential increase of computational resources enables the use of novel strategies to complement, assist, or even replace traditional experimental methods for designing and screening novel peptide ligands for applications ranging from biomarker detection, drug delivery, drug design, and vaccine development.

This chapter presented the methods that our team has developed to address this exciting challenge. We developed, implemented, tested, and validated a modular algorithm for the *ex novo* optimization of amino acid based binders, named PARCE [31]. It enables the optimization of the peptide sequence to maximize its (predicted) binding affinity toward a molecular target.

The protocol, initially introduced as an evolutionary algorithm based on iterative docking [32, 33], evolved into a comprehensive open-source design protocol, embedding a number of functionalities that have been tested and improved during the years, thus enhancing its outreach. Indeed, the key of PARCE's success relies on its modularity: it has been designed so that when novel more accurate approaches are available, these can be easily embedded into the existing code.

The explicit description of the solvation environment during the design procedure was crucial for selecting successful candidates that are solvent-specific and target-selective. This procedure and the ongoing improvement of force fields for MD have opened the possibility of exploring new design conditions (*e.g.*, binding in nonstandard solvents and under extreme pressures and temperatures), which may be hardly accessible with other computational approaches. Another determinant for successful designs was the inclusion of multiple scoring functions, in the form of a consensus criterion. This enabled, for example, the *in silico* unsupervised maturation of an antibody fragment [39]. All these improvements are expected to push forward the limits of peptide design by reaching affinities analogous to those reached by nature.

PARCE is only limited by the accuracy of the predictors it relies upon, which can be updated as new techniques become available. We are thus looking forward for novel advances in structure prediction and free energy evaluations. However, we note that accurate predictors typically involve more computational resources. For the case of PARCE, MD simulations involve costs which are orders of magnitude higher than docking methodologies. Nevertheless, we believe that using MD helps improving the quality of the design. We foresee that the continuous growth in computing power will make the trade-off between computational cost and accuracy more and more unbalanced toward accuracy.

Acknowledgements

R.O and P.C. were supported by MinCiencias, Ruta N, University of Antioquia, Colombia, and the Max Planck Society, Germany. N.M. was supported by the Alternatives Research & Development Foundation (Annual Open Grant, PI: S.F., M.A.S.). S.F. would like to acknowledge the Italian Association for Cancer Research (AIRC) through the grant "My First AIRC grant," Rif.18510, and the CINECA Awards N. HP10B3JT25, 2020, for the availability of high performance computing resources and support.

Conflict of Interest The authors declare that they have no competing interest.

References

1. Kim BY, Rutka JT, Chan WC (2010) Nanomedicine. N Engl J Med 363(25):2434–2443
2. Zhang X-X, Eden HS, Chen X (2012) Peptides in cancer nanomedicine: drug carriers, targeting ligands and protease substrates. J Controll Release 159(1):2–13
3. Chung EJ (2016) Targeting and therapeutic peptides in nanomedicine for atherosclerosis. Exp Biol Med 241(9):891–898
4. Brayden DJ, Hill T, Fairlie D, Maher S, Mrsny R (2020). Systemic delivery of peptides by the oral route: formulation and medicinal chemistry approaches. Adv Drug Deliv Rev
5. Kurrikoff K, Aphkhazava D, Langel Ü (2019) The future of peptides in cancer treatment. Curr Opin Pharmacol 47:27–32
6. Deutscher, S. (2019). Phage display to detect and identify autoantibodies in disease. N Engl J Med 381(1):89–91
7. Cretich M, Damin F, Pirri G, Chiari M (2006) Protein and peptide arrays: recent trends and new directions. Biomol Eng 23(2–3):77–88
8. Ambrosetti E, Paoletti P, Bosco A, Parisse P, Scaini D, Tagliabue E, De Marco A, Casalis L (2017). Quantification of circulating cancer biomarkers via sensitive topographic measurements on single binder nanoarrays. ACS Omega 2(6):2618–2629
9. Adedeji AF, Ambrosetti E, Casalis L, Castronovo M (2018a) Spatially resolved peptide-DNA nanoassemblages for biomarker detection: a synergy of DNA-directed immobilization and nanografting. In: DNA nanotechnology. Springer, New York, pp 151–162
10. Adedeji AF, Ambrosetti E, Casalis L, Castronovo M (2018b) Spatially resolved peptide-DNA nanoassemblages for biomarker detection: a synergy of dna-directed immobilization and nanografting. In: DNA nanotechnology. Springer, New York, pp 151–162
11. Ciemny M, Kurcinski M, Kamel K, Kolinski A, Alam N, Schueler-Furman O, Kmiecik S (2018) Protein–peptide docking: opportunities and challenges. Drug Discov Today 23(8):1530–1537
12. Diller DJ, Swanson J, Bayden AS, Jarosinski M, Audie J (2015) Rational, computer-enabled peptide drug design: principles, methods, applications and future directions. Future Med Chem 7(16):2173–2193
13. Fosgerau K, Hoffmann T (2015) Peptide therapeutics: current status and future directions. Drug Discovery Today 20(1):122–128
14. La Manna S, Di Natale C, Florio D, Marasco D (2018) Peptides as therapeutic agents for inflammatory-related diseases. Int J Mol Sci 19(9):2714
15. Lee AC-L, Harris JL, Khanna KK, Hong J-H (2019) A comprehensive review on current advances in peptide drug development and design. Int J Mol Sci 20(10):2383
16. Sillerud LO, Larson RS (2005) Design and structure of peptide and peptidomimetic antagonists of protein-protein interaction. Curr Protein Peptide Sci 6(2):151–169
17. Russo A, Aiello C, Grieco P, Marasco D (2016) Targeting "undruggable" proteins: design of synthetic cyclopeptides. Curr Med Chem 23(8):748–762
18. Vlieghe P, Lisowski V, Martinez J, Khrestchatisky M (2010) Synthetic therapeutic peptides: science and market. Drug Discov Today 15(1–2):40–56
19. Leurs U, Lohse B, Ming S, Cole PA, Clausen RP, Kristensen JL, Rand KD (2014) Dissecting the binding mode of low affinity phage display peptide ligands to protein targets by

hydrogen/deuterium exchange coupled to mass spectrometry. Anal Chem **86**(23): 11734–11741

20. Fjell CD, Hiss JA, Hancock REW, Schneider G (2011) Designing antimicrobial peptides: form follows function. Nat Rev Drug Discov 2 (Mic):31–45
21. Adedeji Olulana, AF, Soler MA, Lotteri M, Vondracek H, Casalis L, Marasco D, Castronovo M, Fortuna S (2021) Computational evolution of beta2-microglubulin binding peptides for nanopatterned surface sensors. Int J Mol Sci 22(2):812
22. Yagi Y, Terada K, Noma T, Ikebukuro K, Sode K (2007) In silico panning for a non-competitive peptide inhibitor. BMC Bioinform 8(1):11
23. Mitchell M (1998) An introduction to genetic algorithms. MIT Press, Cambridge
24. Besray Unal E, Gursoy A, Erman B (2010) Vital: Viterbi algorithm for de novo peptide design. PLoS One 5(6):e10926
25. Haliloglu T, Seyrek E, Erman B (2008) Prediction of binding sites in receptor-ligand complexes with the Gaussian network model. Phys Rev Lett 100(22):228102
26. Raman S, Vernon R, Thompson J, Tyka M, Sadreyev R, Pei J, Kim D, Kellogg E, DiMaio F, Lange O, Kinch L, Sheffler W, Kim B-H, Das R., Grishin NV, Baker D (2009) Structure prediction for casp8 with all-atom refinement using Rosetta. Proteins Struct Funct Bioinform 77(S9):89–99
27. Alford RF, Leaver-Fay A, Gonzales L, Dolan EL, Gray JJ (2017) A cyber-linked undergraduate research experience in computational biomolecular structure prediction and design. PLOS Comput Biol 13(12):e1005837
28. King CA, Bradley P (2010) Structure-based prediction of protein-peptide specificity in Rosetta. Proteins Struct Funct Bioinform 78(16):3437–3449
29. Unal EB, Gursoy A, Erman B (2010) Vital: Viterbi algorithm for de novo peptide design. PLOS One 5(6):1–15
30. Obarska-Kosinska A, Iacoangeli A, Lepore R, Tramontano A (2016) PepComposer: computational design of peptides binding to a given protein surface. Nucleic Acids Res 44(W1): W522–W528
31. Ochoa R, Soler M, Laio A, Cossio P (2020) PARCE: protocol for amino acid refinement through computational evolution. Comput Phys Commun 260:107716
32. Hong Enriquez RP, Pavan S, Benedetti F, Tossi A, Savoini A, Berti F, Laio A (2012) Designing short peptides with high affinity for organic molecules: a combined docking, molecular dynamics, and Monte Carlo approach. J Chem Theor Comput 8(3): 1121–1128
33. Russo A, Scognamiglio PL, Enriquez RPH, Santambrogio C, Grandori R, Marasco D, Giordano A, Scoles G, Fortuna S (2015) In silico generation of peptides by replica exchange Monte Carlo: docking-based optimization of maltose-binding-protein ligands. PLoS One 10(8):1–16
34. Gladich I, Rodriguez A, Hong Enriquez RP, Guida F, Berti F, Laio A (2015) Designing high-affinity peptides for organic molecules by explicit solvent molecular dynamics. J Phys Chem B 119(41):12963–12969
35. Trott O, Olson AJ (2010) Autodock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J Comput Chem 31(2):455–461
36. Soler MA, Rodriguez A, Russo A, Adedeji AF, Dongmo Foumthuim CJ, Cantarutti C, Ambrosetti E, Casalis L, Corazza A, Scoles G, Marasco D, Laio A, Fortuna S (2017) Computational design of cyclic peptides for the customized oriented immobilization of globular proteins. Phys Chem Chem Phys 19(4): 2740–2748
37. Guida F, Battisti A, Gladich I, Buzzo M, Marangon E, Giodini L, Toffoli G, Laio A, Berti F (2017) Peptide biosensors for anticancer drugs: design in silico to work in denaturizing environment. Biosens Bioelectron 100: 298–303
38. Chi LA, Vargas MC (2020) In silico design of peptides as potential ligands to resistin. J Mol Model 26:1–14
39. Soler MA, Medagli B, Semrau MS, Storici P, Bajc G, de Marco A, Laio A, Fortuna S (2019) A consensus protocol for the in silico optimisation of antibody fragments. Chem Commun 55(93):14043–14046
40. 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(5):3438–3444
41. Peterson LX, Kang X, Kihara D (2014) Assessment of protein side-chain conformation prediction methods in different residue environments. Proteins Struct Funct Bioinform 82(9):1971–1984
42. Huang X, Pearce R, Zhang Y (2020) FASPR: an open-source tool for fast and accurate protein side-chain packing. Bioinformatics 36: 3758–3765

43. Ochoa R, Soler MA, Laio A, Cossio P (2018) Assessing the capability of in silico mutation protocols for predicting the finite temperature conformation of amino acids. Phys Chem Chem Phys 20(40):25901–25909
44. Lindorff-Larsen K, Piana S, Palmo K, Maragakis P, Klepeis JL, Dror RO, Shaw DE (2010) Improved side-chain torsion potentials for the Amber ff99SB protein force field. Proteins Struct Funct Bioinform 78(8): 1950–1958
45. Jorgensen WL, Chandrasekhar J, Madura JD, Impey RW, Klein ML (1983) Comparison of simple potential functions for simulating liquid water. J Chem Phys 79(2):926–935
46. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126(1):014101
47. Parrinello M, Rahman A (1980) Crystal structure and pair potentials: a molecular dynamics study. Phys Rev Lett 45(14):1196–1199
48. Di Pierro M, Elber R, Leimkuhler B (2015) A stochastic algorithm for the isobaric-isothermal ensemble with Ewald summations for all long range forces. J Chem Theor Comput 11(12): 5624–5637
49. Janežič D, Merzel F (1995) An efficient symplectic integration algorithm for molecular dynamics simulations. J Chem Inf Comput Sci 35(2):321–326
50. Hicks DG, Kulkarni S (2008) HER2+ breast cancer: review of biologic relevance and optimal use of diagnostic tools. Am J Clin Pathol 129(2):263–273
51. Oh D-Y, Bang Y-J (2020) HER2-targeted therapies-a role beyond breast cancer. Nat Rev Clin Oncol 17(1):33–48
52. Sawant MS, Streu CN, Wu L, Tessier PM (2020) Toward drug-like multispecific antibodies by design. Int J Mol Sci 21(20):7496
53. Ochoa R, Laio A, Cossio P (2019) Predicting the affinity of peptides to major histocompatibility complex class II by scoring molecular dynamics simulations. J Chem Inf Model 59(8):3464–3473
54. 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 6:34869
55. Medagli B, Soler MA, de Zorzi R, Fortuna S (2021) Antibody affinity maturation using computational methods: from an initial hit to small scale expression of optimised binders. In: Computer-aided antibody design. Springer, in press
56. Del Carlo M, Capoferri D, Gladich I, Guida F, Forzato C, Navarini L, Compagnone D, Laio A, Berti F (2016) In silico design of short peptides as sensing elements for phenolic compounds. ACS Sensors 1(3):279–286
57. Soler M, Fortuna S, Scoles G (2015) Computational design of peptides as probes for the recognition of protein biomarkers. In: 10th European-biophysical-societies-association (EBSA) European biophysics congress, vol 44. Springer, New York, pp 149–149
58. Negroni MP, Stern LJ (2018) The N-terminal region of photocleavable peptides that bind HLA-DR1 determines the kinetics of fragment release. PLoS One 13(7):e0199704
59. Peters B, Nielsen M, Sette A (2020) T cell epitope predictions. Ann Rev Immunol 38(1): 123–145
60. Purcell AW, McCluskey J, Rossjohn J (2007) More than one reason to rethink the use of peptides in vaccine design. Nat Rev Drug Discov 6(5):404–414
61. Wang P, Sidney J, Dow C, Mothé B, Sette A, Peters B (2008) A systematic assessment of MHC class II peptide binding predictions and evaluation of a consensus approach. PLoS Comput Biol 4(4):e1000048
62. Bjorkman PJ (2015) Not second class: the first class II MHC crystal structure. J Immunol 194(1):3–4
63. Unanue ER, Turk V, Neefjes J (2016) variations in MHC class II antigen processing and presentation in health and disease. Annu Rev Immunol 34(1):265–297
64. Weaver JM, Sant AJ (2009) Understanding the focused CD4 T cell response to antigen and pathogenic organisms. Immunol Res 45(2–3): 123–143
65. Wang P, Sidney J, Kim Y, Sette A, Lund O, Nielsen M, Peters B (2010) Peptide binding predictions for HLA DR, DP and DQ molecules. BMC Bioinform 11(1):568
66. Stern LJ, Brown JH, Jardetzky TS, Gorga JC, Urban RG, Strominger JL, Wiley DC (1994) Crystal structure of the human class II MHC protein HLA-DR1 complexed with an influenza virus peptide. Nature 368(6468): 215–221
67. Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, Shindyalov IN, Bourne PE (2000) The protein data bank. Nucleic Acids Res 28(1):235–242
68. Huang P-S, Ban Y-EA, Richter F, Andre I, Vernon R, Schief WR, Baker D (2011) RosettaRemodel: a generalized framework for flexible backbone protein design. PLoS One 6(8): e24109

69. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: algorithms for highly efficient, load balanced, and scalable molecular simulations. J Chem Theor Comput 4: 435–447
70. Dominguez C, Boelens R, Bonvin AMJJ (2003) HADDOCK: a protein–protein docking approach based on biochemical or biophysical information. J Am Chem Soc 125(7): 1731–1737
71. Yang Y, Zhou Y (2008) Specific interactions for Ab Initio folding of protein terminal regions with secondary structures. Proteins Struct Funct Genet 72(2):793–803
72. Zhou H, Skolnick J (2011) GOAP: a generalized orientation-dependent, all-atom statistical potential for protein structure prediction. Biophys J 101(8):2043–2052
73. Krissinel E, Henrick K (2007) Inference of macromolecular assemblies from crystalline state. J Mol Biol 372(3):774–797
74. Andrusier N, Nussinov R, Wolfson HJ (2007) FireDock: fast interaction refinement in molecular docking. Proteins Struct Funct Bioinform 69(1):139–159
75. Berrera M, Molinari H, Fogolari F (2003) Amino acid empirical contact energy definitions for fold recognition in the space of contact maps. BMC Bioinform 4(1):8
76. Fogolari F, Corazza A, Yarra V, Jalaru A, Viglino P, Esposito G (2012) Bluues: a program for the analysis of the electrostatic properties of proteins based on generalized Born radii. BMC Bioinform 13(Suppl 4):S18
77. Cossio P, Granata D, Laio A, Seno F, Trovato A (2012) A simple and efficient statistical potential for scoring ensembles of protein structures. Sci Rep 2:1–8
78. Sarti E, Zamuner S, Cossio P, Laio A, Seno F, Trovato A (2013) Bachscore. A tool for evaluating efficiently and reliably the quality of large sets of protein structures. Comput Phys Commun 184(12):2860–2865