THE JOURNAL OF PHYSICAL CHEMISTRY B

Article

pubs.acs.org/JPCB

# Designing High-Affinity Peptides for Organic Molecules by Explicit Solvent Molecular Dynamics

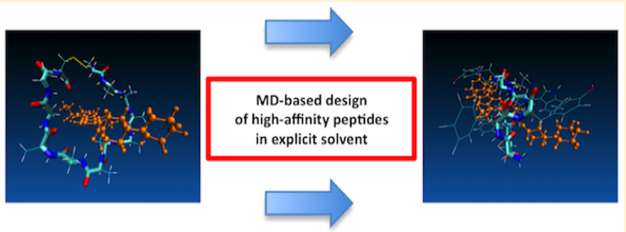
Ivan Gladich,[*,†] Alex Rodriguez,[†] Rolando P. Hong Enriquez,[‡] Filomena Guida,[†] Federico Berti,[§] and Alessandro Laio[*,†]

[†]International School for Advanced Studies (SISSA), Via Bonomea 265, I-34136 Trieste, Italy

[‡]Istituto Italiano di Tecnologia (IIT), Genova, Italy

[§]Dipartimento di Scienze Chimiche e Farmaceutiche, University of Trieste, I-34136 Trieste, Italy

**ABSTRACT:** Short peptides offer a cheap alternative to antibodies for developing sensing units in devices for concentration measurement. We here describe a computational procedure that allows designing peptides capable of binding with high affinity a target organic molecule in aqueous or nonstandard solvent environments. The algorithm is based on a stochastic search in the space of the possible sequences of the peptide, and exploits finite temperature molecular dynamics simulations in explicit solvent to check if a proposed mutation improves the binding affinity or not. The procedure automatically produces peptides which form thermally stable complexes with the target. The estimated binding free energy reaches the 13 kcal/mol for Irinotecan anticancer drug, the target considered in this work. These peptides are by construction *solvent specific*; namely, they recognize the target only in the solvent in which they have been designed. This feature of the algorithm calls for applications in devices in which the peptide-based sensor is required to work in denaturants or under extreme conditions of pressure and temperature.




## INTRODUCTION

Employing specific binders with high affinity and specificity against target molecules is a standard technique for enhancing the sensibility of concentration measurement devices. In this context, peptides offer an invaluable opportunity for developing a cheap device capable of quickly and reliably measuring the concentration of molecules, for example, of a drug or of a metabolite in the blood of the patients.[1−3] Indeed, the number of different properties (affinity toward a target, selectivity, etc.) that can be obtained by combining the natural amino acids is potentially enormous, making conceivable designing peptides capable of recognizing any molecule. Moreover, short peptides can be easily synthesized in large quantities, making the scale-up of the production of the device economically affordable. The opportunity of exploiting short peptides as sensing units has indeed been recognized for a long time.[4] Nakamura and Sugimoto have been able to find small peptides capable of binding organic molecules,[5,6] and some peptides have already been tested for technological applications in food science as receptors for electronic noses.[7−11]

Unfortunately, at variance with antibodies, peptides cannot be designed exploiting the natural selection machinery. The design procedure has to be developed from scratch and cannot be "copied" from living organisms. This implies a major challenge for theoreticians because the number of possible sequences, even for a relatively small peptide with 10 residues, is astronomically large, therefore calling for a computational search[4,12,13] based on importance sampling. However, a simple random walk in the sequence space will not be sufficient to produce reliable solutions. Indeed, any automatic search also has to take into account the structural flexibility of peptides: on the contrary of large proteins, peptides are extremely flexible and their structures can change dramatically even after a single mutation.

Recently, Hong et al.[13] introduced an algorithm capable of designing peptides that bind Efavirenz, a drug used in HIV treatment. In this algorithm, the space of the sequences is explored by Monte Carlo, and viable peptide−ligand conformers are obtained by flexible docking, rolling the peptide on the target molecule and, afterward, relaxing the obtained structures by short molecular dynamics runs. Although this scheme offers a valuable and computationally affordable design solution, flexible docking exploiting a scoring function developed to optimize complexes formed between stable globular proteins and a drug is unavoidably approximate for short peptides. Even more importantly, novel technological applications, such as electronic noses for drug and food detection,[7−11] require designs in very specific environments, ranging from organic solvent to denaturants. To the best of our knowledge, the transferability of a design from one solvent to another has not been tested yet and a solvent specific design procedure is still missing.

In this work, we introduce a novel computational procedure aimed at designing peptides that form thermally stable

**Received:** June 29, 2015
**Revised:** August 27, 2015

DOI: 10.1021/acs.jpcb.5b06227
J. Phys. Chem. B XXXX, XXX, XXX−XXX

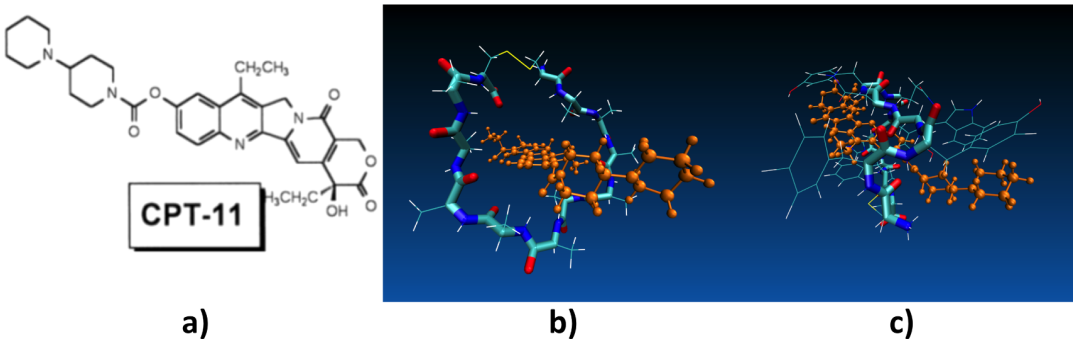



**Figure 1.** Atomic structure of CPT-11. Panels b and c show CPT-11 (in orange) inserted in the peptide at the beginning and the end of the mutation cycle, respectively.

complexes with the target in any required environment. The algorithm builds on the approach introduced by Hong et al. but with several important innovations, that finally make it more reliable and suitable for designing peptides in nonaqueous media. First, the conformational search is carried on by finite temperature molecular dynamics, exploiting a state-of-the-art force field which properly describes the flexibility of both the peptide and the ligand. Second, with the scope of reducing the entropic contribution to the free energy of the unbound state, we now design cyclic peptides with a CYS−CYS bridge between the first and the last residues. Finally, and possibly most importantly, we now perform the design exploiting an explicit atomistic description of the solvent. We will show that this allows designing peptides capable of recognizing the target in practically any kind of solvent, opening a wealth of technological possibilities for developing detection and sensing strategies.

## METHODOLOGY

The algorithm introduced in this work aims at designing short cyclic peptides with high binding affinity toward a target organic molecule, optimizing the peptide amino acid sequence. The main novelty with respect to other algorithms already described in the literature[12−16] is that the geometries of the ligand−peptide complexes are generated by molecular dynamics (MD) in explicit solvent. In order to benchmark this approach, we performed two separated designs, one in water solution and the other in methanol solution, identifying for both solvents peptides capable of binding Irinotecan (CPT-11),[17,18] a chemotherapeutic drug whose atomic structure is reported in Figure 1a. The procedure is schematically described in the flowchart in Figure 2 and consists of three computational blocks: a preparation step, a mutation cycle, and a final validation step, represented, respectively, by green, blue, and red blocks in Figure 2. These steps are described in detail in the rest of this section.

**Preparation Step.** The procedure starts with a cyclic peptide of 10 Ala closed at the end by a disulfide bridge between the two terminal Cys. The CPT-11 molecule is inserted in the middle of the ring, as displayed in Figure 1b. For the design in water, the peptide−ligand complex is solvated in a cubic box containing TIP3P[19] water, while for the design in methanol (hereafter MeOH), in a cubic box containing MeOH[20] molecules. All of the computational details of the MD simulations are provided in the "Computational Details" section below.

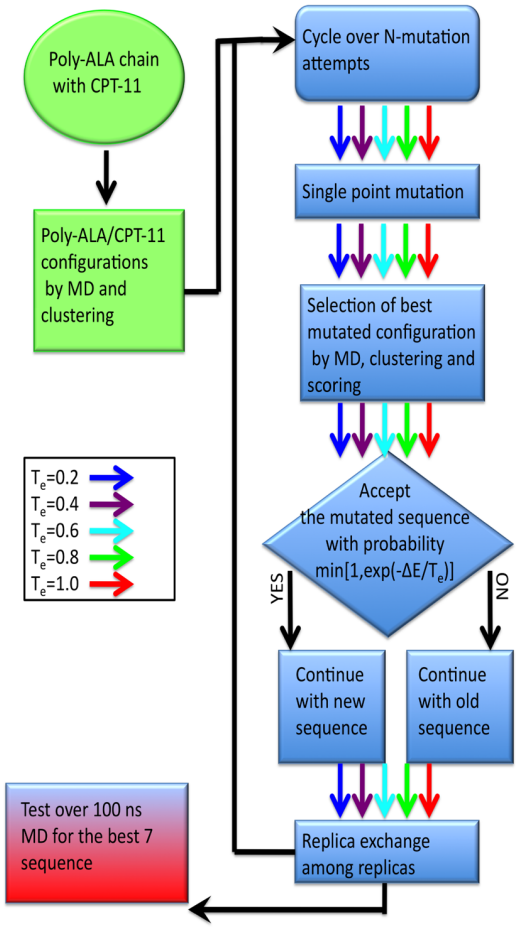



**Figure 2.** Flowchart for the design algorithm. After the preparatory stage (green blocks), five independent mutation cycles at five different effective temperatures $T_e$ are performed, optimizing the peptide sequence and sampling the configurational space by MD (blue blocks). Finally, the thermal stability of the best peptide−ligand complexes is verified by 100 ns NPT-MD simulations (red block).

The initial geometry is first relaxed by an energy minimization of 50 000 steps of steepest descent method, keeping frozen the positions of the peptide backbone and of CPT-11 and allowing the relaxation of the side chains and the solvent. This is followed by 50 ps of constant volume (NVT) simulation, and finally by 10 ns of constant pressure (NPT) simulation. During the two MD runs, the peptide−ligand center-of-mass distance was constrained using a harmonic


DOI: 10.1021/acs.jpcb.5b06227
J. Phys. Chem. B XXXX, XXX, XXX−XXX

potential with a force constant of 700 kJ $mol^{-1}$ $nm^{-2}$, in order to avoid the detachment of the ligand. In order to enhance the sampling of the conformation space, the dynamics was carried out at a temperature of 350 K. We then performed a cluster analysis with the procedure in ref 21 with a cutoff of 0.1 nm, on the conformations generated in the last ns of dynamics, stored every ps. We discarded the clusters including less than 15 conformations. For the centers of the other clusters, we computed the value of the Vina[22] scoring function. The conformation with the lowest score was selected as the starting state for the mutation cycle.

**Mutation Cycle.** The computational optimization of the peptide sequence is based on an exploration of the sequence and configurational space by combined Monte Carlo and MD simulations. We attempt a single random mutation for each step of the cycle and perform a finite temperature MD simulation of the mutated peptide in explicit solvent to generate a meaningful set of conformations. The mutation is accepted or rejected according to a Metropolis criterion based on the binding affinity before and after the mutation. The procedure is similar to the one in ref 13 but with the essential difference that here the conformational search is carried out by finite temperature molecular dynamics in explicit solvent, while in the algorithm in ref 13 it is carried out by flexible docking, followed by a short MD run in a vacuum. We will show that this difference significantly affects the reliability of the algorithm.

The procedure is the following:

1. The peptide is randomly mutated selecting one amino acid from the peptide sequence and replacing it with a different amino acid chosen at random: thus, the new peptide sequence at step $i + 1$, $SEQ_{i+1}$, differs from the sequence at step $i$, $SEQ_i$, for only one amino acid. The terminal CYSs are never mutated in order to conserve the cyclic geometry of the peptide. The atomic coordinates for the mutated peptide, ligand, and solvation box were the ones of the previous step, $CNF_i$, with the only exception of the coordinate of the mutated residue. In order to avoid close contacts, solvent molecules that are closer than 0.2 nm from any heavy atom of the mutated amino acid were removed from the simulation box.
2. The peptide−ligand complex and its solvation box are relaxed with 50 000 steps of steepest descent energy minimization. Then, they are thermalized by 50 ps NVT simulation and a final production NPT run. During the first 10 mutation steps, we keep the constraint on the peptide−ligand distance active, in order to avoid the detachment of the complex. After the 11th step, the constraint is removed. The duration of the production run ($t_{prod}$) is varied during the optimization, in order to maximize computational efficiency. The first 240 mutation steps, in which the binding affinity drops quickly, are performed with $t_{prod}$ = 0.6 ns. The last 120 mutation steps, where it is necessary generating conformations that are highly thermally stable, are performed with $t_{prod}$ = 5 ns. To enhance conformational sampling, all of the MD simulations are performed at 350 K.
3. We then proceed like in the preparation step, namely, we performed for the last 0.4 ns of the production run ($t_{prod}$ = 0.6 ns) a cluster analysis with the procedure of ref 21, with a cutoff of 0.1 nm, discarding the clusters including less than 15 conformations. We then compute the value of the Vina[22] scoring function for the centers of the other clusters. The conformation with the lowest score is selected. We denote by $E_{i+1}$ its score and by $CNF_{i+1}$ the atomic coordinates of the peptide, the ligand, and the solvation box. In the last 120 steps of the design, the ones performed with a production time of $t_{prod}$ = 5 ns, $E_{i+1}$ is computed as the average Vina score in the last 1 ns of the trajectory. This requires docking a large number of peptide−ligand conformers, resulting in a higher computational cost, but it provides a more accurate estimate of the binding affinity. This is beneficial in the final part of the optimization, where it is essential selecting only mutations that allow very stable complexes.
4. The new state ($CNF_{i+1}$, $SEQ_{i+1}$, $E_{i+1}$) is accepted or rejected according to a Metropolis criterion, with a probability of $\min(1, \exp[(E_1 - E_0)/T_e])$. $T_e$ is an efficacious temperature that controls the acceptance rate. In the case of acceptance, the new step started from ($CNF_{i+1}$, $SEQ_{i+1}$, $E_{i+1}$); otherwise, from the previous state ($CNF_i$, $SEQ_i$, $E_i$).

The mutation cycle described above was iterated up to a desired number of mutations. Consistently with ref 13, we strengthened the exploration of the sequence space by performing five simultaneous and independent mutation cycles at five efficacious temperatures $T_e$ = 0.2, 0.4, 0.6, 0.8, and 1.0 kcal/mol. At the end of each mutation step, two replicas $r$ and $r'$ are selected at random and an exchange between their $T_e$'s was attempted according to a parallel tempering scheme:[23] a swap between the states of replica $r$ at $T_r$ ($CNF_r$, $SEQ_r$, $E_r$) and the state of replica $r'$ at $T_{r'}$ ($CNF_{r'}$, $SEQ_{r'}$, $E_{r'}$) is accepted with a probably of $\min(1, \exp[(E_r - E_{r'})]/(1/T_r - 1/T_{r'}))$.

**Validation Step.** After reaching the required number of mutations or the desired binding affinity, the best structures and sequences are selected to benchmark their thermal stability. For such structures, we perform 100 ns NPT simulations at 300 K by monitoring the peptide−ligand center-of-mass distance and the Vina affinity. As we will show in the following, this validation step is essential for judging the quality of the solutions of the algorithm.

**Computational Details.** In all of the MD simulations, the intramolecular and nonbonded interactions are described according to the GAFF force field[24] for CPT-11 and AMBER-99SB[20] for the peptide and MeOH. Consistently with the GAFF practice, the Lorentz−Berthelot combination rule is adopted for the nonbonded interaction expressed in terms of the Lennard-Jones potential. The electrostatic interaction was described by the Coulomb potential between partial charges at individual atom sites. For MeOH and the peptide chain, the partial charges were adopted from the AMBER-99SB force field while the CPT-11 ones were obtained by the AM1-BCC method[25] implemented in Antechamber.[26]

All MD simulations were performed using GROMACS 4.6.7[27] with GPU card support. The MD runs were carried out using a time step of 2 fs. The temperature was controlled by a stochastic velocity rescaling thermostat[28] with a coupling time of 0.1 ps, while, for the NPT run, a Berendsen pressure coupling[29] with a coupling constant of 2 fs was used, keeping the pressure at 1 bar. Standard 3D periodic boundary conditions were adopted with a cutoff distance of 1.0 nm on


DOI: 10.1021/acs.jpcb.5b06227
J. Phys. Chem. B XXXX, XXX, XXX−XXX

the real-space Coulomb and the van der Waals interactions, updating the neighbor list according to the Verlet cutoff scheme. The long-range part of the Coulomb interactions was evaluated using the particle-mesh Ewald method with a relative tolerance of $10^{-5}$, fourth order cubic interpolation, and a Fourier spacing parameter of 0.16. All bonds were constrained using LINCS, while SETTLE was used for constraining the water molecules. With the above setup, we were able to perform about 10 mutations/day for each temperature replica, using an 8 Hyper-threading IntelXeon 2.6 GHz with one Titan Black GPU card and a MD trajectory of $t_{prod}$ = 5 ns.

## RESULTS AND DISCUSSION

Our algorithm aims at designing cyclic peptides of high affinity toward a target molecule. The main innovation introduced in this work is using MD simulations for sampling the most relevant peptide−ligand conformations. This allows designing the peptides in an explicit solvent environment. We here show that this innovation is crucial in order to obtain peptides that are able to form a stable complex with the target. We demonstrate this by comparing our algorithm (MD-algorithm hereafter) with the one in ref 13, that uses flexible docking to generate the conformations (FD-algorithm hereafter). This comparison is reported in Figure 3. The outcome of the mutation cycle, i.e., the binding affinity $E_i$ as a function of mutation iteration $i$, for all effective temperatures $T_e$, is displayed in Figure 3a and c for the FD-algorithm[13] and our new scheme, respectively. We run both algorithms starting from the same initial conformation, a $SA_{10}S$ chain surrounding the ligand (see Figure 1b). At the beginning, the binding affinity is approximately −6 kcal/mol in both cases. Since $SA_{10}S$ is not a good binder for CPT-11, in the first step of the optimization of the MD-algorithm, the peptides show the tendency to detach from the ligand during the MD simulation. We therefore impose a constraint on the peptide−ligand center-of-mass distance for a sufficient number of mutations. After 10 steps, the sequence of the peptide starts adapting to the ligand, and the complex becomes stable for the duration of the MD run used in the design (600 ps). Therefore, the constraint can be removed. Imposing the constraint is not necessary in the FD-algorithm, since in that case MD is used only to avoid close contacts between the ligand and the newly mutated structure, without addressing possible structural thermal instabilities at long times.

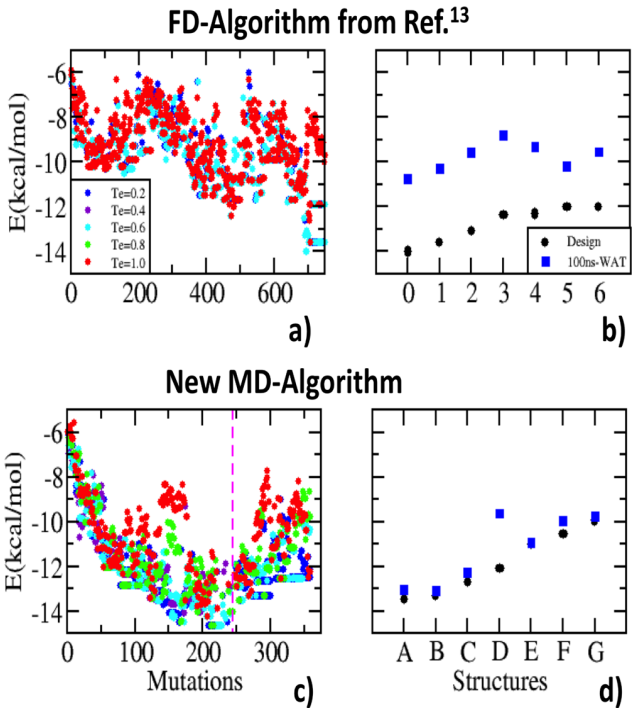



**Figure 3.** Optimization and test of peptides designed according the FD-algorithm of Hong et al. (a and b) and the MD-algorithm presented in this work (c and d). Panels a and c: the Vina scoring function as a function of the mutation step. Different colors correspond to different replicas. Panels b and d: the value of the scoring function for the seven best peptides estimated in the design cycle (black points) and estimated as the average on the last 80 ns of 100 ns MD simulations in explicit solvent (blue squares). The vertical dashed purple line in panel c indicates where the production time, $t_{prod}$, was switched from 0.6 to 5 ns. The sequences are reported in Table 1.

Both algorithms predict a maximum binding affinity of about −14 kcal/mol, which corresponds to the energy plateau observed in Figure 3a and c: The FD-algorithm reaches this value in about 700 mutations, while the new scheme, in less than 300 steps. However, a mutation step in the FD-algorithm is approximately 6−7 times faster than that in the MD-algorithm, since the latter requires performing a MD run in explicit solvent at each step. Figure 3c shows a slightly upward jump in the estimated affinities after 250 mutations. We rationalized this jump by the fact that we switched the production run from 0.6 to 5 ns simulation time after 250 mutations. During the 0.6 ns production runs, the system is at times unable to equilibrate sufficiently, allowing the Monte Carlo procedure to accept some mildly metastable configurations. On the other hand, carrying out all the optimizations using an equilibration time of 5 ns would be too computationally expensive. For this reason, we decided to perform the first part of the design using shorter runs, $t_{prod}$ = 0.6 ns, and, afterward, complete the design with longer runs, allowing a longer time for equilibration.

It is worth noticing that the maximum binding affinity reachable by the design procedure is strongly affected by the starting configuration. Figure 4 shows the result of a design in water in which the CPT-11 molecule was initially placed on the top of the $SA_{10}S$ ring: the maximum binding after 100 mutations is about −8.5 kcal/mol, while the optimization carried out with the target inserted in the ring (Figure 3c) reaches −13 kcal/mol in the same number of iterations. Moreover, we never observed during this design an event in which the target inserts in the ring. We cannot rule out the possibility that, with longer simulation time, the CPT-11 could spontaneously enter the peptide ring, but it seems that this configurational rearrangement is unlikely during our simulation time. This allows concluding that the initial conformation has indeed an important impact on the optimization procedure. For CPT-11, the optimal starting configuration was the one with the ligand inserted in the peptide ring; however, it is reasonable to suppose that the best starting configuration will be, in general, target specific.

The most remarkable difference between the two algorithms is the quality of the solutions they are able to find in terms of thermal stability. We selected the best seven peptides found by the FD- and by the MD-algorithm, which differ among them for at least five amino acids. For these 14 complexes, we then performed a 100 ns MD simulation in water solution, starting


DOI: 10.1021/acs.jpcb.5b06227
J. Phys. Chem. B XXXX, XXX, XXX−XXX

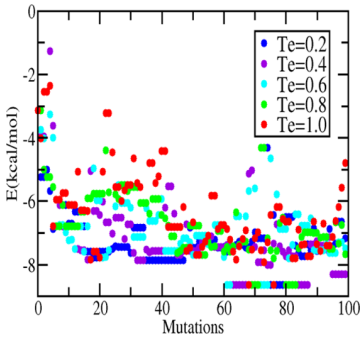



**Figure 4.** Optimization of peptides designed in explicit water according to the MD-algorithm, starting from an initial configuration in which the CPT-11 is not inserted in the ring but it lays on the poly-alanine chain. Similarly to Figure 3c, the figure shows the Vina scoring function as a function of the mutation step and for different temperature replicas. The maximum estimated binding affinity is about −8.5 kcal/mol; the comparison with Figure 3c shows that the performance of the design procedure is dependent from the chosen starting configuration.

from the structures generated by the two algorithms (see the Methodology section). For each complex, we then compute the average binding affinity in the last 80 ns of the simulation. If the peptide−ligand complex is thermally unstable, the peptide tends to detach and the Vina affinity, that by construction always starts from approximately −13 kcal/mol, tends to drift toward higher values. The results of this analysis are illustrated in Figure 3b and d. The black points are the affinities estimated in the mutation cycle, while the blue squares indicate the averages over the MD trajectory. Figure 3b clearly shows that the peptides designed by the FD-algorithm tend to detach from the ligand when the complex is solvated in water and evolved by MD. All the peptides show this upward drift, that is of approximately 4 kcal/mol in the worst cases. This indicates that flexible docking is capable of efficiently generating structures of high estimated binding affinity, but these structures are not stable at finite temperature and in a realistic solvent environment. On the contrary, as shown in Figure 3d, the MD-algorithm generates peptides that under the same conditions are stable, with an average binding affinity very close to the one obtained during the design cycle in almost all cases. The analysis of the peptide sequences in Table 1 shows that peptides designed with the new scheme are richer in aromatic residues than the ones designed by the FD-algorithm, that are on average 6 for the former and 3 for the latter. Apparently, the explicit water description favors the selection of aromatic residues, which are repelled by water and bind toward the aromatic groups of CPT-11, enhancing the binding affinity.

Another major advantage of an algorithm based on molecular dynamics in explicit solvent is that it allows designing peptides capable of binding a target *in any given solvent*. This feature is useful if one exploits the algorithm for designing peptides that can be embedded in a sensor, since in several cases these devices work under very specific physicochemical conditions, for example, in a denaturant. Figure 5 shows that, not surprisingly, peptides which bind well in one type of solvent do not necessarily bind in another one. The figure reports the binding of the peptides designed in water (blue squares) once they are solvated in MeOH (green diamond). The peptide−ligand binding is substantially weaker when they are solvated in an environment (MeOH) which is different from the one of the

**Table 1. Peptide Sequence and Binding Affinities for the Designs toward CPT-11**[a]

| FD-algorithm | sequence | $E_{Design}$ | $E_{100ns}$ |
|---|---|---|---|
| 0 | CTPRKNWGWGYC | −14.0 | −10.7 |
| 1 | CTTRKGWGWGYC | −13.6 | −10.3 |
| 2 | CGWWNPGQGWDC | −13.1 | −9.6 |
| 3 | CVGQWEENWYGC | −12.4 | −8.8 |
| 4 | CGHWYPGQDSDC | −12.3 | −9.3 |
| 5 | CVSWNQFPTFQC | −12.0 | −10.2 |
| 6 | CWEPWEENWMGC | −12.0 | −9.6 |
| **MD-algorithm-WATER** | **sequence** | $E_{Design}$ | $E_{100ns}$ |
| A | CFYWNYQYHVFC | −13.5 | −13.0 |
| B | CNYYNYTWFVYC | −13.3 | −13.1 |
| C | CFFWNYVYWRKC | −12.7 | −12.3 |
| D | CKYWRLVYWFFC | −12.1 | −9.6 |
| E | CFVWIWLYWFRC | −11.0 | −10.9 |
| F | CQFNEHVHWFAC | −10.5 | −9.9 |
| G | CSHRWMHWRHEC | −9.9 | −9.8 |
| **MD-algorithm-MeOH** | **sequence** | $E_{Design}$ | $E_{100ns}$ |
| $\alpha$ | CQYQKYIYYKYC | −13.2 | −12.6 |
| $\beta$ | CWYWALIIYVRC | −13.0 | −12.6 |
| $\gamma$ | CWYTQEMPYDYC | −12.7 | −12.2 |
| $\delta$ | CWYSYNAMYRYC | −12.6 | −11.0 |
| $\epsilon$ | CWNQHYQEYKYC | −12.6 | −12.7 |
| $\zeta$ | CQVQYSIQYYYC | −12.3 | −11.4 |
| $\eta$ | CHVQYYNPYDVC | −11.5 | −11.6 |

[a]For all three designs, i.e., the FD-algorithm of Hong et al. and the MD-algorithm in explicit water and MeOH, the table reports the amino acid sequence of the seven best peptides and their binding affinities, during the mutation cycle ($E_{Design}$) and during the last 80 ns of a 100 ns MD simulation ($E_{100ns}$).

original design (water). In Figure 5b, we report the Vina binding affinity in water (blue curve) and MeOH (green) as a function of time for peptide A, whose primary sequence is reported in Table 1. The figure shows that the affinity is stable in water solution, while it drifts upward in MeOH. The inspection of the trajectory further proves the instability of the peptide−ligand complexes designed in water and solvated in MeOH: Figure 5c shows the backbone of peptide A together with CPT-11 at 0, 25, 50, 75, and 100 ns, showing that the peptide is undergoing severe structural changes. A similar trend has been observed for all the other peptides designed in water and solvated in MeOH reported in Table 1. On the other hand and consistently with what we saw in water, peptides designed in MeOH are stable in MeOH: Figure 5d shows the binding affinity of the seven best peptides designed in MeOH. The Vina binding affinity as a function of time (Figure 5e) and the structures explored by the peptide−ligand complex (Figure 5f) confirm that peptides deigned in MeOH form remarkably stable complexes in the same solvent. The solvent specificity of the design is also confirmed by the composition of the best peptide sequence (see Table 1). The average number of aromatic residues is smaller for the design in MeOH (about 4) than for the design in water (about 6): MeOH is a less polar solvent than water, and thus, aromatic side chains can be more easily exposed to MeOH, competing with the binding toward CPT-11.

The stability of the obtained complexes was further benchmarked carrying out extra 100 ns MD runs at higher temperature (350, 400, and 450 K) for the best structures in water and MeOH, i.e., peptide-A and peptide-$\alpha$ in Table 1. The


DOI: 10.1021/acs.jpcb.5b06227
J. Phys. Chem. B XXXX, XXX, XXX−XXX

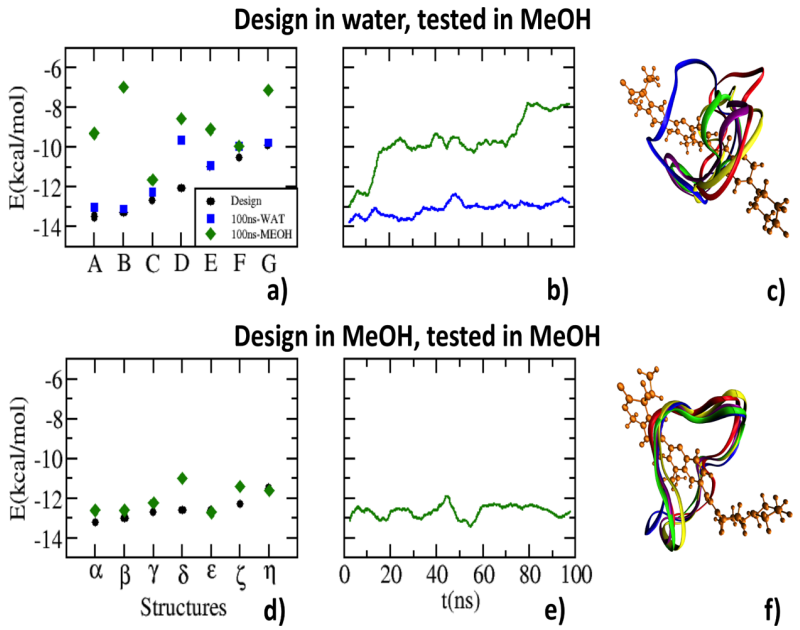



**Figure 5.** Comparison between the design in explicit water and explicit MeOH. Similarly to Figure 3, panel a shows the binding affinities for the seven best peptides designed in water (see Table 1) as they come out from the design (black points), after 100 ns MD in explicit water (blue squares), and after 100 ns MD in explicit MeOH (green diamonds). Panel b reports the Vina score as a function of time for the peptide-A/CPT-11 in water (blue) and MeOH (green). Panel c illustrates the peptide A backbone positions around CPT-11 (in orange) at 0, 25, 50, 75, and 100 ns (blue, purple, green, yellow, and red, respectively). Panels d, e, and f are the equivalent of panels a, b, and c, respectively, for the design in MeOH. The score as a function of time and the structures are shown for the peptide-$\alpha$/CPT-11 complex. The amino acid sequences for peptide-A and peptide-$\alpha$ are reported in Table 1.

Vina scores for these two peptides at different temperatures in water and MeOH are reported in Figure 6a and b, respectively.

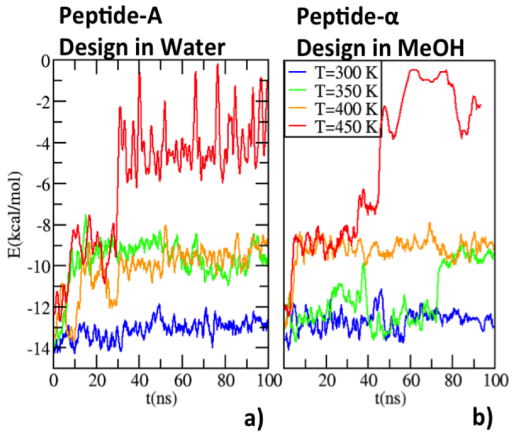



**Figure 6.** Thermal stability at different temperatures for peptide-A and peptide-$\alpha$ in water and MeOH, respectively. Panel a reports the Vina score as a function of time for the peptide-A/CPT-11 complex in water at 300 K (blue), 350 K(green), 400 K (orange), and 450 K (red). Similarly, panel b shows the Vina binding over 100 ns for the peptide-$\alpha$ /CPT-11 complex in MeOH.

As expected, at higher temperature, the contacts between the two molecules are weaker, and some are even broken. However, even at 400 K, the ligand keeps several contacts with the peptide during the entire 100 ns of the simulation, indicating that the complex is remarkably stable. This is remarkable, since complexes formed by peptides with a lower estimated binding affinity are normally unstable even at 300 K, while only the complexes optimized using the procedure described in the manuscript are able to survive systematically 100 ns of MD. In this respect, it should be noted that Vina[22] estimates affinities by counting the number and type of peptide−ligand contacts and assigning, to each of them, an energy value assuming a water environment. For this reason, the docking affinities are not necessarily meaningful in nonaqueous environment. However, we observed a clear correlation between the binding affinity, the number of apolar and polar contacts, and the peptide−drug center-of-mass distance. Moreover, the docking score was not used here with the intent of generating viable conformers but only as a measure of the peptide−ligand stability, screening the MD conformations. For all of these reasons, the docking score should be considered only a rough estimate of the binding affinity, especially in nonaqueous environment. Future work is planned to develop a new scoring function capable of estimating the binding affinity in a nonaqueous environment.

## CONCLUSIONS

Small peptides capable of binding organic molecules have recently emerged as valuable solutions for the development of new biosensors in medical and technological areas. The large number of possible sequences, even in short peptides, call for automatic in silico approaches capable of finding the best sequence for a given target. Some algorithms have been proposed in the literature, but the generation of possible peptide−ligand complex geometries is normally based on docking protocols. The thermal stability in explicit solvent of the obtained peptide−ligand complexes can always be checked a posteriori, or experimentally, but an algorithm capable of finding a solution whose reliability is automatically ensured would be desirable.


DOI: 10.1021/acs.jpcb.5b06227
J. Phys. Chem. B XXXX, XXX, XXX−XXX

In this work, we presented a novel computational procedure for designing high-affinity peptides based on molecular dynamics in a explicit solvent environment. This procedure relies on finite temperature MD simulations, which generate viable peptide−ligand structures properly describing the flexibility of both the peptide and ligand. Designing cyclic peptides allows reducing the entropic contribution to the binding affinity, which is usually poorly described by the scoring functions. All of these new features have enabled peptides to be designed which are thermally stable on a 100 ns time scale in a realistic solvent environment, opening (in principle) the possibility of designing high-affinity peptides for any kind of target in any kind of solvent. The procedure could be further improved by exploiting a scoring function targeted for nonaqueous solvents: future work is currently planned in this direction. Nevertheless, we strongly believe that the computational design scheme presented here already has potential for exciting technological applications, from food science to drug recognition, in which peptide-based sensors are required to work, for example, in denaturants or under extreme conditions of pressure and temperature.

## AUTHOR INFORMATION


### Corresponding Authors

*E-mail: igladich@sissa.it.

*E-mail: laio@sissa.it.


### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We acknowledge financial support from the grant Associazione Italiana per la Ricerca sul Cancro 5 per mille, Rif. 12214, and Fondo per gli Investimenti della Ricerca di Base-Accordo di programma, Rif. RBAP11ETKA

## REFERENCES


(1) Crotti, S.; Enzo, M. V.; Bedin, C.; Pucciarelli, S.; Maretto, I.; Del Bianco, P.; Traldi, P.; Tasciotti, E.; Ferrari, M.; Rizzolio, F.; et al. Clinical Predictive Circulating Peptides in Rectal Cancer Patients Treated with Neoadjuvant Chemoradiotherapy. *J. Cell. Physiol.* **2015**, *230*, 1822−1828.

(2) Wang, W.; Li, M.; Wei, Z.; Wang, Z.; Bu, X.; Lai, W.; Yang, S.; Gong, H.; Zheng, H.; Wang, Y.; et al. Bimodal Imprint Chips for Peptide Screening: Integration of High-Throughput Sequencing by MS and Affinity Analyses by Surface Plasmon Resonance Imaging. *Anal. Chem.* **2014**, *86*, 3703−3707.

(3) Gu, G.; Hu, Q.; Feng, X.; Gao, X.; Menglin, J.; Kang, T.; Jiang, D.; Song, Q.; Chen, H.; Chen, J. PEG-PLA Nanoparticles Modified with APTEDB Peptide for Enhanced Anti-Angiogenic and Anti-Glioma Therapy. *Biomaterials* **2014**, *35*, 8215−8226.

(4) Vanhee, P.; van der Sloot, A.; Verschueren, E.; Serrano, L.; Rousseau, F.; Schymkowitz, J. Computational Design of Peptide Ligands. *Trends Biotechnol.* **2011**, *29*, 231.

(5) Nakamura, C.; Inuyama, Y.; Shirai, K.; Sugimoto, N.; Miyake, J. Detection of Porphyrin using a Short Peptide Immobilized on a Surface Plasmon Resonance Sensor Chip. *Biosens. Bioelectron.* **2001**, *16*, 1095.

(6) Sugimoto, N.; Miyoshi, D.; Zou, J. Development of Small Peptides Recognizing a Monosaccharide by Combinatorial Chemistry. *Chem. Commun.* **2000**, 2295.

(7) Wu, T.; Lo, Y. Synthetic Peptide Mimicking of Binding Sites on Olfactory Receptor Protein for Use in Electronic Nose. *J. Biotechnol.* **2000**, *80*, 63.

(8) Pavan, S.; Berti, F. Short Peptides as Biosensor Transducers. *Anal. Bioanal. Chem.* **2012**, *402*, 3055−3070.

(9) Sinisi, V.; Forzato, C.; Cefarin, N.; Navarini, L.; Berti, F. Interaction of Chlorogenic Acids and Quinides from Coffee with Human Serum Albumin. *Food Chem.* **2015**, *168*, 332−340.

(10) Lim, J. H.; Park, J.; Ahn, J. H.; Jin, H. J.; Hong, S.; Park, T. H. A Peptide Receptor-Based Bioelectronic Nose for the Real-Time Determination of Seafood Quality. *Biosens. Bioelectron.* **2013**, *39*, 244−249.

(11) Sankaran, S.; Khot, L. R.; Panigrahi, S. Biology and Applications of Olfactory Sensing System: A Review. *Sens. Actuators, B* **2012**, *171*, 1−17.

(12) Tinberg, C. E.; Khare, S. D.; Dou, J.; Doyle, L.; Nelson, J. W.; Schena, A.; Jankowski, W.; Kalodimos, C. G.; Johnsson, K.; Stoddard, B. L.; et al. Computational Design of Ligand-Binding Proteins with High Affinity and Selectivity. *Nature* **2013**, *501*, 212−6.

(13) Hong Enriquez, R. P.; Pavan, S.; Benedetti, F.; Tossi, A.; Savoini, A.; Berti, F.; Laio, A. Designing Short Peptides with High Affinity for Organic Molecules: A Combined Docking, Molecular Dynamics, and Monte Carlo Approach. *J. Chem. Theory Comput.* **2012**, *8*, 1121−1128.

(14) Aumentado-Armstrong, T. T.; Istrate, B.; Murgita, R. a. Algorithmic Approaches to Protein-Protein Interaction Site Prediction. *Algorithms Mol. Biol.* **2015**, *10*, 1−21.

(15) Sirin, S.; Pearlman, D. A.; Sherman, W. Physics-based enzyme design: Predicting binding affinity and catalytic activity. *Proteins: Struct., Funct., Genet.* **2014**, *82*, 3397−3409.

(16) Damborsky, J.; Brezovsky, J. Computational Tools for Designing and Engineering Enzymes. *Curr. Opin. Chem. Biol.* **2014**, *19*, 8−16.

(17) Rivory, L. P.; Chatelut, E.; Canal, P.; Mathieuboue, A.; Robert, J. Kinetics of the In-Vivo Interconversion of the Carboxylate and Lactone Forms of Irinotecan (CPT-11) and of Its Metabolite SN-38 in Patients. *Cancer Res.* **1994**, *54*, 6330−6333.

(18) Yang, X. X.; Hu, Z. P.; Chan, S. Y.; Goh, B. C.; Duan, W.; Chan, E.; Zhou, S. F. Simultaneous Determination of the Lactone and Carboxylate Forms of Irinotecan (CPT-11) and Its Active Metabolite SN-38 by High-Performance Liquid Chromatography: Application to Plasma Pharmacokinetic Studies in the Rat. *J. Chromatogr. B: Anal. Technol. Biomed. Life Sci.* **2005**, *821*, 221−228.

(19) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of Simple Potential Functions for Simulating Liquid Water. *J. Chem. Phys.* **1983**, *79*, 926−935.

(20) Hornak, V.; Abel, R.; Okur, A.; Strockbine, B.; Roitberg, A.; Simmerling, C. Comparison of Multiple Amber Force Fields and Development of Improved Protein Backbone Parameters. *Proteins: Struct., Funct., Genet.* **2006**, *65*, 712−725.

(21) Daura, X.; Gademann, K.; Jaun, B.; Seebach, D.; van Gunsteren, W. F.; Mark, A. E. Peptide Folding: When Simulation Meets Experiment. *Angew. Chem., Int. Ed.* **1999**, *38*, 236−240.

(22) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(23) Frenkel, D.; Smit, B. *Understanding Molecular Dynamics*; Academic Press: 2002.

(24) Wang, J. M.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157−1174.

(25) Jakalian, A.; Jack, D. B.; Bayly, C. I. Fast, Efficient Generation of High-Quality Atomic Charges. AM1-BCC model: II. Parameterization and Validation. *J. Comput. Chem.* **2002**, *23*, 1623−1641.

(26) Wang, J.; Wang, W.; Kollman, P. A.; Case, D. A. Automatic Atom Type and Bond Type Perception in Molecular Mechanical Calculations. *J. Mol. Graphics Modell.* **2006**, *25*, 247−260.

(27) Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.

(28) Bussi, G.; Donadio, D.; Parrinello, M. Canonical Sampling through Velocity Rescaling. *J. Chem. Phys.* **2007**, *126*, 014101.




DOI: 10.1021/acs.jpcb.5b06227
*J. Phys. Chem. B* XXXX, XXX, XXX−XXX

582 (29) Berendsen, H. J. C.; Postma, J. P. M.; Vangunsteren, W. F.;
583 Dinola, A.; Haak, J. R. Molecular Dynamics with Coupling to an
584 External Bath. *J. Chem. Phys.* **1984**, *81*, 3684−3690.



DOI: 10.1021/acs.jpcb.5b06227
*J. Phys. Chem. B* XXXX, XXX, XXX−XXX