JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Designing Peptide Sequences in Flexible Chain Conformations to Bind RNA: A Search Algorithm Combining Monte Carlo, Self-Consistent Mean Field and Concerted Rotation Techniques

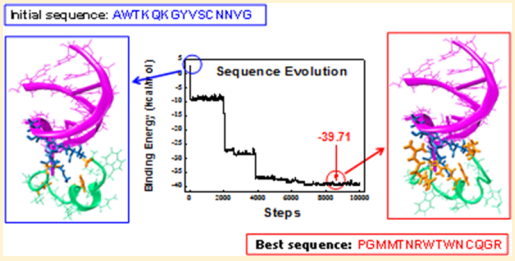
Xingqing Xiao,[†] Paul F. Agris,[‡] and Carol K. Hall*[,†]

[†]Chemical and Biomolecular Engineering Department, North Carolina State University, Raleigh, North Carolina 27695-7905, United States

[‡]The RNA Institute, University at Albany, State University of New York, Albany, New York 12222, United States

Supporting Information

ABSTRACT: A search algorithm combining Monte Carlo, self-consistent mean field, and concerted rotation techniques was developed to discover peptide sequences that are reasonable HIV drug candidates due to their exceptional binding to human $tRNA_{UUU}^{Lys3}$, the primer of HIV replication. The search algorithm allows for iteration between sequence mutations and conformation changes during sequence evolution. Searches conducted for different classes of peptides identified several potential peptide candidates. Analysis of the energy revealed that the asparagine and cysteine at residues 11 and 12 play important roles in "recognizing" $tRNA^{Lys3}$ via van der Waals interactions, contributing to binding specificity. Arginines preferentially attract the phosphate linkage via charge−charge interaction, contributing to binding affinity. Evaluation of the RNA/peptide complex's structure revealed that adding conformation changes to the search algorithm yields peptides with better binding affinity and specificity to $tRNA^{Lys3}$ than a previous mutation-only algorithm.



1. INTRODUCTION

As the sixth leading cause of death in the world, the human immunodeficiency virus (HIV) has received a great deal of attention. HIV triggers acquired immune deficiency syndrome (AIDS) and, if untreated, eventually damages the immune system.[1] Although HIV cannot be cured, its progression can be slowed through daily doses of a cocktail of antiretroviral drugs.[2] Today, we are faced with a virus that is rapidly developing single and multidrug resistance. Efforts to find novel therapeutics and a cure for HIV proceed on many fronts, including continuing approaches to interfere with HIV replication. However, there are also novel approaches to abrogate the development of resistance and interfere with replication. At the beginning of its replication cycle, the HIV-1 virus uses a host $tRNA^{Lys3}$ as a primer to initiate HIV reverse transcription.[3−7] The human $tRNA^{Lys3}$ is recruited by the virus and then annealed to the viral RNA. The annealing is mediated by the viral nucleocapsid protein[8] and other proteins. The uniqueness of post-transcriptional modifications to this tRNA's anticodon stem and loop domain (ASL^{Lys3}) contributes to the virus's selection of this tRNA among all tRNAs of the cell as a primer of reverse transcription. During subsequent infections of other cells, the viral RNA genome[9] is reverse transcribed into DNA and integrated into the host DNA.[10] New viral RNAs are transcribed from the integrated DNA, and proteins are translated for the production of new capsids. This suggests that blocking the recruitment of the specific primer $tRNA^{Lys3}$ by the HIV virus prior to this cycle might interfere with or even stop HIV reverse transcription.[11] The fact that resistance has been found in mutations of proteins, yet not a single human HIV isolate has been found to use a host tRNA other than $tRNA^{Lys3}$, suggests that blocking $tRNA^{Lys3}$ recruitment could deter resistance.

Our goal is to design a peptide chain that can mimic the ability of the viral nucleocapsid protein to bind selectively to the primer $tRNA^{Lys3}$ with the ultimate purpose of blocking the recruitment of $tRNA^{Lys3}$ by the HIV virus. Agris and co-workers have been working toward this goal for several years. They have synthesized many different peptide sequences, 15 and 16 amino acids in length, by using Peptide Phage Display Libraries.[12−14] Through testing the affinity and specificity of the peptide chains to the anticodon stem and loop (ASL) of $tRNA^{Lys3}$, a relatively good 15-mer peptide called P6 (sequence: RVTHHAFL-GAHRTVG) was identified in their experiments. Since synthesis and testing of all possible candidate sequences is not practical, we have been working to develop a protein design strategy that identifies other potential peptide candidates in a fast and effective way; the merits of these candidate peptides can be verified experimentally in the Agris group.[14]

Computationally based strategies[15,16] can be used to search the broad sequence space to design proteins for particular purposes, e.g. novel enzymes,[17] protein−DNA interactions,[18] and immune epitopes.[19] Most protein design strategies are

Received: September 10, 2014
Published: January 7, 2015

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

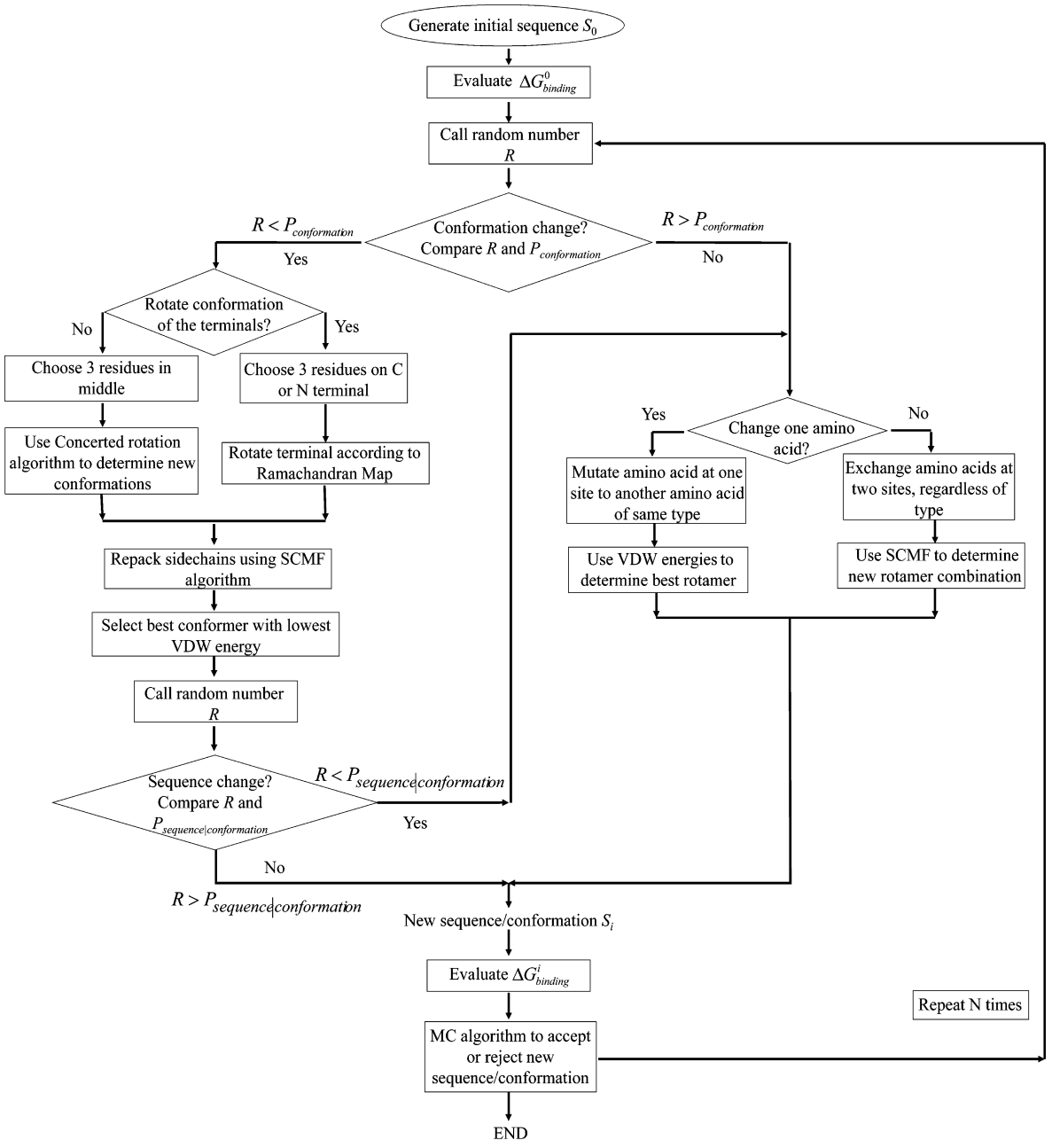


Figure 1. Flow sheet for the hybrid search algorithm.

based on assuming a fixed backbone.[20,21] Usually, the standard Monte Carlo (MC) technique or its extensions are used to search sequence space.[22] An alternative method is the statistical, computationally assisted design strategy (SCADS) developed by Saven and co-workers,[23,24] which is based on using self-consistent theory to sample different peptide sequences on a fixed backbone. Other search algorithms, such as dead-end elimination (DEE)[25] and genetic algorithms (GA)[26] are also applied widely to sequence-design problems.

Accommodating backbone flexibility is challenging in protein design, although several successes have been achieved in protein docking,[27,28] conformation prediction,[29−31] and interface modeling.[32,33] In protein loop modeling, the conformation for a short loop (<5 residues)[34,35] is changed by slight movement of the loop's backbone, and the conformation for a medium loop (6−7 residues) is changed via fragment database searching. If the loop is long, a fragment buildup strategy is used to select several small peptide folding conformations from the protein fragment database so as to build up a large loop architecture.[36] The fragment buildup strategy is, however, not suitable for generating local perturbations of the conformation of a short peptide chain for sampling.

Multifunctional Web-based design tools[37,38] have been developed to facilitate protein design. The most popular protein design tool is Rosetta, a software suite based on the MC technique.[39−41] The Rosetta program exhibits broad applicability for predicting and designing the components, structures, and functionalities of proteins. It is especially convenient for protein−protein interface design and protein docking. The Rosetta program provides the user with two basic ways to design a protein structure: change the backbone conformation or change the sequence. When designing a sequence using a

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

rotamer library, the search approach is run assuming a fixed backbone conformation; when optimizing the backbone conformation within the context of an atomic-resolution model, the search approach is run assuming a fixed sequence. Additionally, an approach that iterates between a sequence search and a conformation search can be executed in Rosetta,[42−44] but the backbone adjustments are mostly confined to small perturbations in ϕ and ψ angles.[45]

Thus far, only a few studies of the design of proteins for interaction with DNA or RNA have been reported.[46−49] In these efforts the investigators start with a known peptide−DNA (or RNA) complex and then develop algorithms and/or models that allow them to "redesign" the interface starting from a random peptide sequence. Success is measured by how faithfully the redesigned peptide sequence compares to the native sequence. Thyme et al.[50] redesigned the interface of the I-AniI protein−DNA complex on a fixed backbone by embedding a motif rotamer library into a computational peptide sequence optimization protocol. A good match was obtained between the designed and the native protein interfaces, confirming the validity of the computational sequence optimization protocol. Chen et al.[51] built a distance and orientation-dependent hydrogen-bonding potential to describe the interactions between amino acids and nucleosides based on observations of high-resolution crystal structures for protein−DNA and protein−RNA complexes. The energy resulting from the hydrogen-bonding potential was used as the objective function in their computational protein-design approach. By using the objective function, they successfully discriminated five native protein−RNA complexes from 10 000 incorrectly docked decoys.

In this paper, we develop a novel hybrid search algorithm based on evaluating the binding free energy to evolve a peptide sequence that binds selectively to the post-transcriptionally modified anticodon stem and loop domain of $tRNA^{Lys3}$, ASL^{Lys3}. The novel hybrid search algorithm uses a combination of Monte Carlo (MC), self-consistent mean field (SCMF) theory, and the concerted rotation move (CONROT) techniques. It is an extension of our previously published search algorithm that was based on MC/SCMF, which searched for sequences on a fixed backbone conformation.[52] The advantage of the new hybrid search algorithm presented here is that it allows us to evolve the peptide sequence and optimize the backbone conformation at the same time. Since the hybrid search algorithm samples the conformations in both backbone space and sequence space, it can identify peptide sequences with stronger binding specificity to ASL^{Lys3} than when searching sequence space alone. In this paper, we apply the hybrid search algorithm to the design of a 15-residue peptide with high binding affinity to ASL^{Lys3} using the conformation of peptide P6 as the starting conformation. Analysis of the contributions to the free energy of the complex between the peptide sequence and the ASL^{Lys3} loop reveals that the hydrophilic amino acids asparagine at site 11 and cysteine at site 12 play an important role in "recognizing" ASL^{Lys3} due to the van der Waals (VDW) energy, thereby contributing to its binding specificity. The positively charged amino acids preferentially attract the sugar ring/phosphate linkage of ASL^{Lys3} due to the charge−charge interaction and contribute to its binding affinity.

This paper begins with a full description of the novel hybrid search algorithm. A review of the concerted rotation (CONROT) method can be found in the Supporting Information. This is followed by a comparison of the evolution results for different parameter sets and a description of the best peptide sequences obtained by implementing the hybrid search algorithm. Subsequently, an analysis of structure and contributions to the free energy of the ASL^{Lys3}−peptide complexes is given. Finally, we present our conclusions.

2. METHOD

We extend our previous algorithm, a search through sequence space to find the best binder,[52] to include a search through conformation space to take the backbone flexibility into account. The procedure for the resulting hybrid search algorithm is shown in Figure 1. There are two main functional modules: one is for conformation changes, and the other is for sequence mutations. Two probability parameters, $P_{\text{conformation}}$ and $P_{\text{sequencelconformation}}$, are used to control the process of evolution so that the peptide has either a conformation change, a sequence change, or both simultaneously. In order to design peptides that are drug candidates, we also introduce some constraints on the allowed hydration properties of the evolved peptides. Details of the hydration property constraints are described later. The outline of the strategy is as follows:

(1) Generate an initial peptide sequence S_0 that meets the hydration property constraint.

(2) Calculate the binding free energy (without GBSUR, the nonpolar solvation energy) $\Delta G^0_{\text{binding}}$ for the complex composed of the ASL^{Lys3} and the initial peptide chain S_0.

(3) Compare the conformation probability ($P_{\text{conformation}}$) with a random number (R) in order to determine which module to call: the conformation change module or the sequence mutation module.

(4) If $R > P_{\text{conformation}}$, the sequence of the peptide is mutated. There are two ways to do this: either mutate one amino acid or exchange two amino acids. When one amino acid is mutated, another amino acid of the same residue type (see below) is randomly chosen to substitute for the old one, resulting in the generation of a new attempted sequence. In contrast, when two amino acids are exchanged, they are randomly chosen regardless of the residue types of the amino acids, again resulting in the generation of a new attempted sequence. Skip to step 7 to evaluate the binding capability of the new sequence.

(5) If $R < P_{\text{conformation}}$, the conformation of the peptide backbone is changed. There are two ways to do this. The first way is to use the concerted rotation (CONROT) method to displace three consecutive residues (viz. nine consecutive skeletal atoms) in the middle of the peptide chain. The second way is to move one of the two ends (N- and C-terminus). Any attempts to twist the skeletal bonds on the three consecutive residues at the end of the peptide chain are permissible as long as the torsion angles (ϕ and ψ) satisfy the Ramachandran plot.[53−55] After either type of move, there will be many possible conformations for the side chains. Self-consistent mean field (SCMF) theory is employed to repack the side chains. Through calculating the VDW energy of the repacked conformer, the best attempted conformer is selected and is then subjected to further evaluation.

(6) After step 5, the functional module to mutate the sequences is conditionally launched by comparing the conditional probability that the sequence is changed after a conformation change move ($P_{\text{sequencelconformation}}$) and another random number (R). If $R < P_{\text{sequencelconformation}}$, we execute the sequence mutation and go to step 4 again. If not, this new

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

attempted conformer will get a final evaluation for its binding capability at step 7.

(7) The new attempted sequence/conformation S_i is evaluated, this time by calculating the binding free energy (without GBSUR) $\Delta G^i_{\text{binding}}$. The Metropolis step is used to accept or reject this attempted sequence/conformation S_i. These seven steps are repeated hundreds of thousand of times to evolve good sequence candidates.

We restrict our search for candidate peptide sequences to those peptides that are reasonable drug candidates. This means that they should be soluble in water and exhibit the desired hydration properties, having to do with hydrophobicity, net charge, polarity of the peptide chain, etc. A description of the rationale for imposing hydration constraints is given in our previous paper.[52] The constraints on the hydration property of the peptide chain are specified before launching the hybrid search algorithm and then held fixed throughout the sequence evolution process. We have classified the 20 natural amino acids into six residue types according to their hydrophobicity, polarity, size, and charge.[56] The first column in Table 1 gives

Table 1. Six Residue Types for the 20 Amino Acids

hydrophobic	Leu, Val, Ile
	Met
	Phe
	Tyr, Trp
negatively charged	Glu, Asp
positively charged	Arg, Lys
hydrophilic	Ser, Thr
	Asn, Gln
	His
other	Ala
	Cys
	Pro
glycine	Gly

the amino acid type, and the second column lists the amino acids of that type. By adjusting the number of amino acids in each residue type in the peptide, the peptide's hydration properties can be controlled. Here, we consider 15-mer peptide chains with three different sets of hydration properties:[52] cases one, two, and three, as shown in Table 2. These are listed

Table 2. Three Cases with Different Hydration Properties

	case one	case two	case three
$N_{\text{hydrophobic}}$	4	5	3
$N_{\text{negative-charge}}$	0	0	0
$N_{\text{positive-charge}}$	2	2	1
$N_{\text{hydrophilic}}$	5	6	6
N_{other}	2	1	3
N_{glycine}	2	1	2

according to the number of hydrophobic $N_{\text{hydrophobic}}$, negatively charged $N_{\text{negative-charge}}$, positively charged $N_{\text{positive-charge}}$, hydrophilic $N_{\text{hydrophilic}}$, other amino acids N_{other}, and glycine N_{glycine} residues along the 15 amino acid chain.

The search algorithm requires an initial configuration for the complex between the peptide chain and ASL^{Lys3}. We use molecular dynamics atomistic simulation with the AMBER 10 package to get the initial location and conformation for the complex. The peptide P6 (sequence: RVTHHAFLGAHRTVG)[13,14] which was found in Agris' recent experimental work to exhibit relatively good binding behavior to ASL^{Lys3} was put into a truncated octahedral box with an 8 Å buffer of TIP3P water around the peptide chain in each direction, the primary purpose being to determine its folded structure. Both ASL^{Lys3} and the folded peptide chain were then placed together to solvate in a periodic box containing TIP3P water. The ASL^{Lys3} and peptide complex were simulated at 298 K for 60 ns in order to get a stable binding conformation. The resulting complex shown in Figure 2 is the same as complex 2

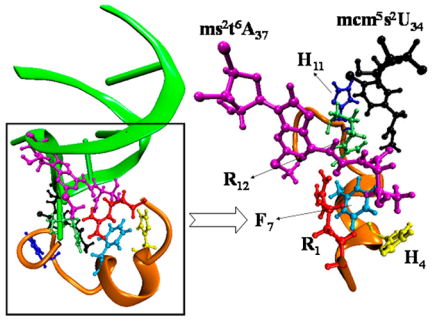


Figure 2. Snapshots of the initial binding conformation for the complex in the hybrid search algorithm. The ASL^{Lys3} is represented by the green ribbon; the P6 peptide sequence RVTHHAFLGAHRTVG is represented by the orange ribbon. Several important amino acids and nucleotides are shown in distinct colors. The complex is the same as complex 2 in our previous paper, which is extracted from a 60 ns atomistic simulation and is presumed to be at a global minimum in the binding free energy.

in ref 57, which is presumed to be at a global minimum in the free energy (binding energy = −24.05 kcal/mol; binding energy without GBSUR = −19.72 kcal/mol) that results from a 60 ns atomistic simulation. This is the initial binding structure for the search process in this paper. Another binding conformation, complex 1, discovered via atomistic simulations in ref 57, was not considered in this work, because it had a higher (less favorable) binding energy than complex 2. Interestingly, however, the sequences that were evolved from complex 1 had lower binding energies than those evolved from complex 2, which suggested to us that the conformational flexibility afforded by complex 1 compared to complex 2 made the search for sequence candidates more successful. This motivated us to add conformational change moves to the search algorithm, which is what we do in this paper. A detailed discussion of the configurations for the complexes and an analysis of their free energy are given in our previous paper.[57]

Here, we briefly introduce other aspects of the search algorithm. Prior to the evolution, we generate a random starting sequence that satisfies the hydration properties required for each case. The starting conformation is the same as that for P6 (determined in ref 57). If the case has the same hydration properties as the P6 peptide, we randomly mutate the amino acids on P6 to other amino acids of the same residue type, or randomly exchange the locations of some amino acids regardless of their residue types. No energy evaluation is involved in the mutation and the exchange of the amino acids here. If the case does not have the same hydration properties as the P6 peptide, we randomly mutate some of the residues on P6 to achieve a peptide that has the requisite hydration properties. Subsequently, we follow the above strategy to randomly mutate and exchange the amino acids on the chain to

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

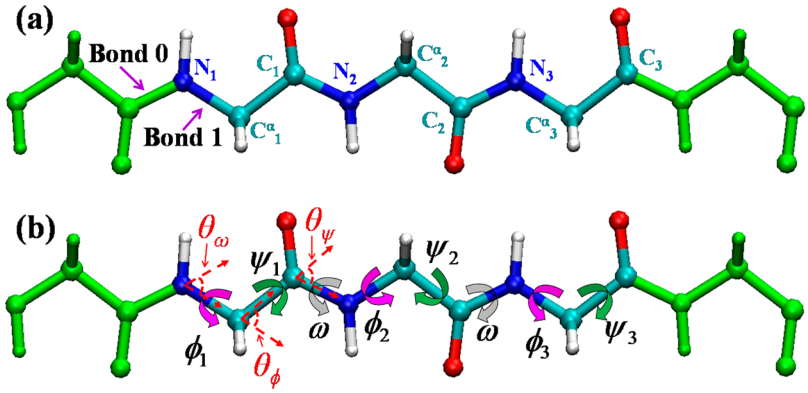


Figure 3. Three consecutive residues (multicolored beads) in the middle of the peptide chain subjected to the CONROT move and two other residues at the ends (green beads) kept fixed. The side chains on the peptide are not shown for clarity. The hydrogen atoms (white), nitrogen atoms (blue), carbon atoms (cyan), and oxygen atoms (red) are shown. (a) Nine skeletal atoms are labeled for identification. The first bond ($N_1-C^{\alpha}_1$) is designated as **Bond 1**; the bond preceding **Bond 1** is designated as **Bond 0**. (b) The dihedral angles (ϕ, ψ, ω) and the bond angles (θ_ω, θ_ϕ, θ_ψ) are marked.

generate a random starting sequence suitable for this case. In the search algorithm, the SCMF technique[58,59] is employed to search for appropriate rotamer combinations during the residue exchange moves. The side-chain conformations are chosen from the rotamer library of Lovell et al.[60]

The binding free energy is defined to be the difference between the free energy of the complex and the free energies of the ligand (here, the peptide chain) and of the receptor (here, the ASL^{Lys3}) prior to binding. It can be calculated according to

$$\Delta G_{\text{binding}} = G_{\text{TOT}}^{\text{complex}} - G_{\text{TOT}}^{\text{ligand}} - G_{\text{TOT}}^{\text{receptor}} \quad (1)$$

The free energy in each term of eq 1 has the following contributions:

$$G_{\text{TOT}} = U_{\text{INT}} + U_{\text{VDW}} + U_{\text{ELE}} + G_{\text{SOL}} \quad (2)$$

where U_{INT}, U_{VDW}, U_{ELE}, and G_{SOL} are the internal energy (INT), van der Waals energy (VDW), electrostatic energy (ELE), and solvation energy (SOL); the latter contains the polar solvation energy (EGB) and the nonpolar solvation energy (GBSUR).

At each step of the search algorithm, we calculate the binding free energy (without GBSUR)[58,61−65] to evaluate the binding capability of the new trial sequence, then employ the Metropolis sampling method to accept this new attempt or not. The GBSUR contribution is neglected because it is very small, does not change very much during the entire evolution process, and results in little to no significant effect. Additionally, the calculation of GBSUR is time-consuming. During the process of sequence evolution, we choose the best rotamer or rotamer combinations in sequence change moves and the best conformer in conformation change moves based on the evaluation of the VDW energy. That is because the atomic overlaps can be monitored directly by the VDW energy. Details can be found in our previous work.[52]

The CONROT technique[66,67] is employed to displace the backbone conformation of any three consecutive nonterminal residues, i.e., residues in the middle of the peptide chain. The skeletal dihedral angles which describe the individual rotations of the bonds ($N-C_\alpha$), ($C_\alpha-C$), and ($C-N$) in the backbone scaffold are denoted by (ϕ, ψ, ω), respectively, and the skeletal bond angles with an apex at (N, C_α, and C) are specified by (θ_ω, θ_ϕ, and θ_ψ), respectively. Through measuring the torsion angles (ϕ, ψ, ω), we can determine the backbone conformation of the peptide. Figure 3a gives a representation of a short fragment containing three consecutive nonterminal residues (viz., nine consecutive skeletal atoms) that are subject to a CONROT move. For convenience, we have labeled the nine consecutive skeletal atoms in order to identify them. The different torsion angles {ϕ_1, ψ_1, ω, ϕ_2, ψ_2, ω, ϕ_3} along the backbone are indicated in Figure 3b.

In the CONROT move, we change the torsion angles {ϕ_1, ψ_1, ω, ϕ_2, ψ_2, ω} of the three consecutive residues and leave the positions of the remaining residues on the backbone unchanged, as shown in Figure 3. Since the backbone atoms ($C_\alpha-C-N-C_\alpha$) adopt the *trans* conformation, the skeletal dihedral angle ω is always equal to π. Given a change in ϕ_1, the "driver angle," the other three torsion angles {ψ_1, ϕ_2, ψ_2} can be expressed as functions of ϕ_1 using the CONROT technique. For any given ϕ_1, solution sets for {ψ_1, ϕ_2, ψ_2} may or may not exist. If the solution sets for (ϕ_1, ψ_1, ϕ_2, ψ_2) exist, and each pair of (ϕ, ψ) does not violate the Ramachandran plot for the general case, we rotate these skeletal bonds according to the solution set, resulting in the change of backbone conformation. More details on how to obtain the solution set (ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3, ψ_3) are given in the Supporting Information.

Supporting Information Figure S3 exhibits a flow diagram to explain how to conduct the CONROT method in order to produce many new candidate conformations. Each time a conformational change move is selected, we generate several trial conformational candidates. Any attempts are permissible as long as the torsion angles (ϕ and ψ) of peptide backbone satisfy the Ramachandran plot. After repacking side chains on the trial conformation candidates, the best potential conformer is selected to compare to the old conformer. Later, the Metropolis criterion is employed to accept or reject this new conformer. If the binding energy of the new conformer is lower than that of the old conformer, we always accept the new conformer. But if the binding energy of the new conformer is higher than the old one, we would accept the new conformer with a certain probability.

3. RESULTS AND DISCUSSION

a. Sequence Evolution. The sequence of moves in the hybrid Monte Carlo (MC)/self-consistent mean field (SCMF)/concerted rotation move (CONROT) search algorithm is controlled by two probability parameters: $P_{\text{conformation}}$

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

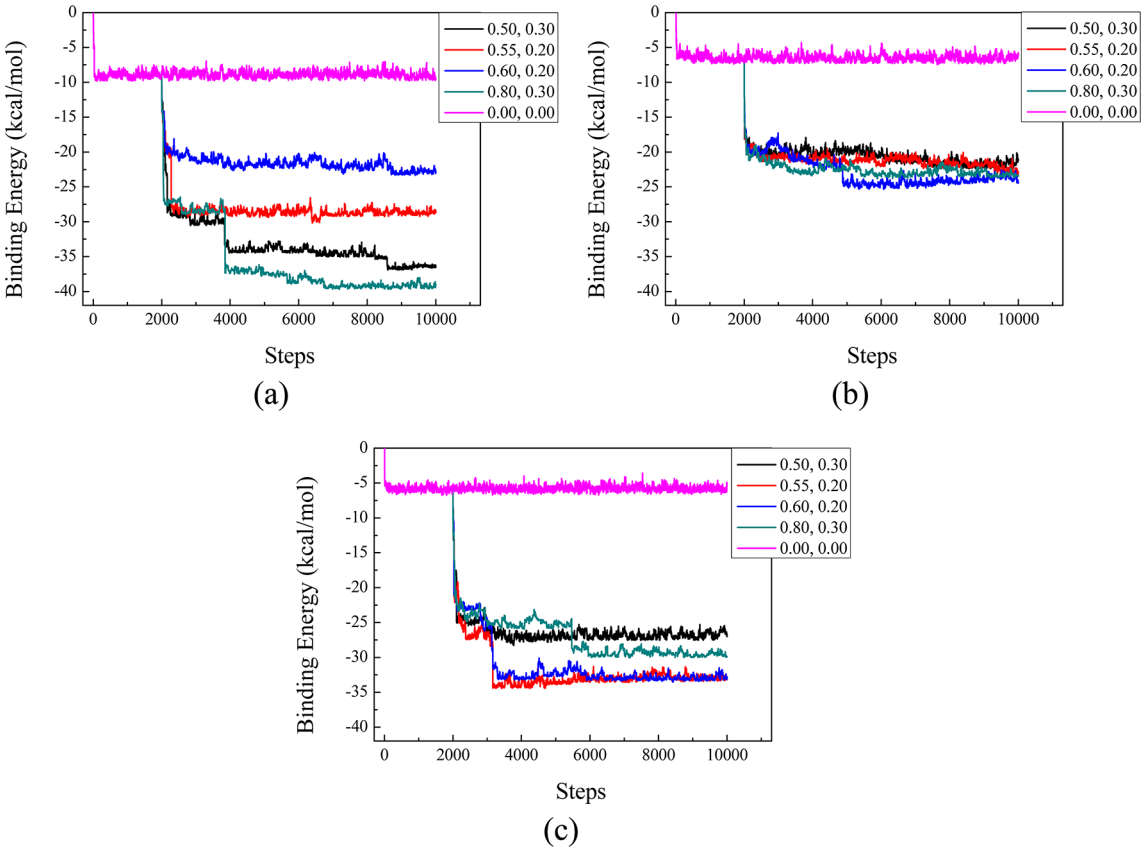


Figure 4. Binding energy profiles at various values of ($P_{conformation}$, $P_{sequencelconformation}$). (a) Case one, (b) case two, and (c) case three.

and $P_{sequencelconformation}$, which determine the probability of making a conformation change move and the probability of making a sequence change move after making a successful conformation change, respectively. On the basis of the value of $P_{conformation}$, we can either make a sequence change move alone (right side of flow diagram in Figure 1) or we can make a conformation change move that may or may not be followed by a sequence change move (left side of flow diagram in Figure 1). The conditional probability $P_{sequencelconformation}$ determines whether a sequence change move will occur after the conformation change move. For example, setting $P_{conformation}$ = 0.00 allows for a sequence change move alone with no further attempts to change the backbone conformation. Setting ($P_{conformation}$, $P_{sequencelconformation}$) = (0.60, 0.20) means that there is a 60% probability to change the peptide's conformation and a 40% probability to change the peptide's sequence alone; once a successful conformation change move has been made, there still remains a 20% probability to change the old sequence to a new sequence. A series of searches at different values of $P_{conformation}$ and $P_{sequencelconformation}$ were performed in order to avoid the local searches on peptide sequence. There are 10 000 steps in each search, wherein each step contains at least 15 attempts to mutate the amino acids or to change the backbone conformation. Overall, more than 150 000 attempts were made for each search. The first 2000 steps in the search procedure are limited to sequence-mutation moves (the conformation is set to the fixed initial configuration), while the later 8000 steps involve execution of both types of moves based on the values of $P_{conformation}$ and $P_{sequencelconformation}$.

The binding energy profiles have been analyzed (Figure 4) in regard to the number of search steps at different values of $P_{conformation}$ and $P_{sequencelconformation}$ for different sets of hydration properties: cases one, two, and three as listed in Table 2. The values of the energies at ($P_{conformation}$, $P_{sequencelconformation}$) ≠ (0.00, 0.00) are much lower than the energies at ($P_{conformation}$, $P_{sequencelconformation}$) = (0.00, 0.00) (Figure 4). This indicates that the evolved sequences with conformational changes are much better than those with only sequence mutations. The sequences with the lowest energies for each ($P_{conformation}$, $P_{sequencelconformation}$) and each hydration property and their corresponding binding energies are listed in Table 3. It is clear that the evolved peptide sequences at ($P_{conformation}$, $P_{sequencelconformation}$) ≠ (0.00, 0.00) are greatly improved relative to those at ($P_{conformation}$, $P_{sequencelconformation}$) = (0.00, 0.00). The global minimum in each column is highlighted in bold, exhibiting the best peptide sequence for each hydration property case. For example, the lowest binding energy in case

Table 3. Lowest Binding Energy (kcal/mol) for Each ($P_{conformation}$, $P_{sequencelconformation}$) Shown for the Three Cases[a]

$P_{conformation}$, $P_{sequencelconformaiton}$	case one	case two	case three
0.50, 0.20	−35.32	−25.01	−29.73
0.50, 0.30	−36.92	−22.44	−28.27
0.55, 0.20	−30.18	−23.42	**−34.47**
0.55, 0.30	−36.73	−24.74	−30.63
0.60, 0.20	−23.32	**−25.35**	−33.51
0.60, 0.30	−34.44	−21.93	−30.61
0.80, 0.20	−35.98	−23.40	−31.01
0.80, 0.30	**−39.71**	−23.79	−30.07
0.00, 0.00	−9.83	−7.48	−6.79

[a]The best search result in each case is highlighted in bold.

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

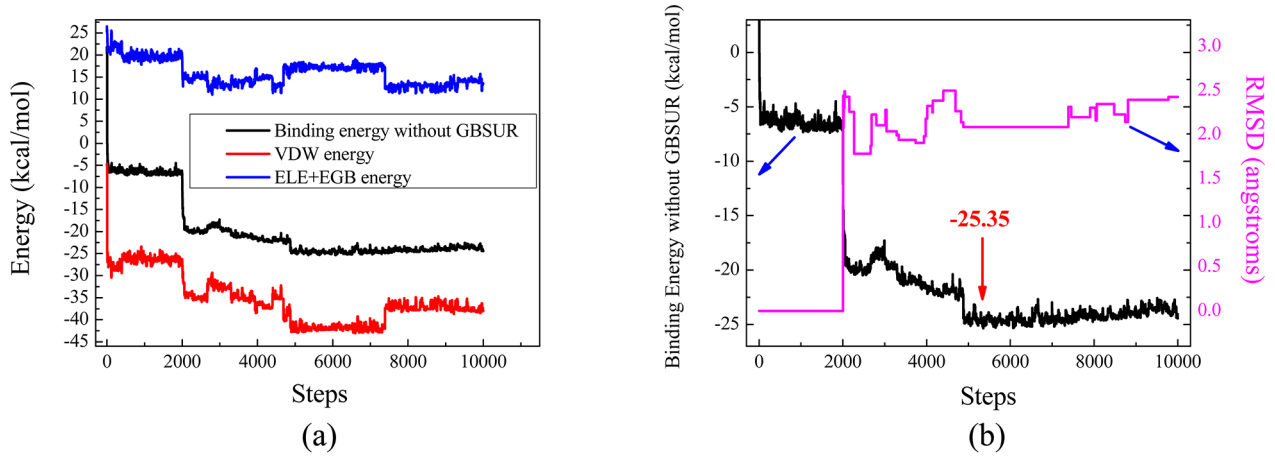


Figure 5. Analysis of energy contributions in case two at ($P_{conformation}$, $P_{sequence|conformation}$) = (0.60, 0.20): (a) binding energy without GBSUR, the VDW energy and the (ELE+EGB) energy vs evolution steps, (b) binding energy without GBSUR and RMSD vs evolution steps. The persistent change of RMSD profile reflects a relative good acceptance probability in conformation change move.

Table 4. Five Top-Ranked Results of the Peptide Sequences Obtained by the Hybrid Search Algorithm

	sequences for case one ($P_{conformation}$ = 0.80, $P_{sequence|conformation}$ = 0.30)															
	sites															
rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	binding energy without GBSUR (kcal/mol)
1	P	G	M	M	T	N	R	W	T	W	N	C	Q	G	R	−39.71
2	P	G	M	M	S	S	R	W	H	W	N	C	Q	G	R	−39.69
3	P	G	N	M	S	L	R	W	S	W	N	C	Q	G	R	−39.69
4	P	G	M	M	T	T	R	W	T	W	N	C	Q	G	R	−39.68
5	P	I	G	M	S	H	R	W	T	W	N	C	Q	G	R	−39.67
initial sequence	T	W	A	K	Q	K	G	Y	V	S	C	N	N	V	G	2.30
	sequences for case two ($P_{conformation}$ = 0.60, $P_{sequence|conformation}$ = 0.20)															
	sites															
rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	binding energy without GBSUR (kcal/mol)
1	R	G	S	I	S	M	R	W	T	S	N	C	Q	I	Y	−25.35
2	R	G	S	V	N	M	R	W	T	N	N	C	Q	I	Y	−25.35
3	R	G	S	M	S	F	R	W	H	T	N	C	Q	I	Y	−25.35
4	R	G	S	I	S	M	R	W	T	N	N	C	Q	I	Y	−25.35
5	R	G	S	S	S	N	R	W	I	M	N	C	Q	I	Y	−25.34
initial sequence	S	S	A	R	Y	T	F	V	R	S	H	T	M	F	G	21.80
	sequences for case three ($P_{conformation}$ = 0.55, $P_{sequence|conformation}$ = 0.20)															
	sites															
rank	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	binding energy without GBSUR (kcal/mol)
1	P	G	G	M	S	S	R	W	H	H	N	C	Q	W	P	−34.47
2	P	G	G	M	T	Q	R	W	S	H	N	C	Q	W	P	−34.45
3	P	G	T	M	T	T	R	W	T	H	N	C	P	W	G	−34.44
4	P	G	Q	M	S	T	R	W	G	P	N	C	Q	W	N	−34.44
5	P	G	T	M	G	Q	R	W	S	H	N	C	Q	W	P	−34.44
initial sequence	P	P	T	T	F	S	G	K	Q	S	A	T	M	Y	G	23.14

two is −25.35 kcal/mol at ($P_{conformation}$, $P_{sequence|conformation}$) = (0.60, 0.20), while that in case three is −34.47 kcal/mol at ($P_{conformation}$, $P_{sequence|conformation}$) = (0.55, 0.20).

Structural and energetic analysis of the complex formed by the peptide chain and ASL^{Lys3} can help us better understand the mechanism of binding. For example, consider case two at ($P_{conformation}$, $P_{sequence|conformation}$) = (0.60, 0.20), the best binder for that case; Figure 5a shows the binding energy without GBSUR, the VDW energy, and the sum of the ELE (electrostatic energy) and EGB (polar solvation energy) contributions to the binding free energy. The RMSD (root-mean-square deviation) has been evaluated relative to the number of steps in the search, along with the binding energy without GBSUR (Figure 5b). It is apparent that the sharp drop in the binding energy as the sequence evolves is due mainly to the decline in VDW energy, while the sum of (ELE+EGB) energy shows little change (Figure 5a). The binding energy without GBSUR changes in lockstep with the changes in the RMSD (Figure 5b). Interestingly, the first time the binding energy has a major drop is also the time when the peptide's conformation undergoes its first major fluctuation. This means that the conformation changes make the peptide more accessible to the ASL^{Lys3}; thereby resulting in a notable improvement of binding capability. Furthermore, such improve-

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

ment is a result of the decrease of VDW energy (Figure 5a) and enhances molecular recognition greatly. Additionally, a persistent change in the RMSD profile also reflects a relative good acceptance probability in conformation change moves in the hybrid search algorithm.

We have ranked the five top-rated sequences for all three cases resulting from the search and their corresponding binding energies (Table 4). For instance, since case one's lowest binding energy (see Table 3) is −39.71 kcal/mol at ($P_{conformation}$, $P_{sequence|conformation}$) = (0.80, 0.30), Table 4 lists the next four top-ranked peptide sequences at (0.80, 0.30). Also shown as the bottom line in each section of the table is the starting sequence and its binding energy without GBSUR. Examination of these top-ranked peptide sequences yields commonalities in all three cases. Some similar, even identical, amino acids occupy the same sites in the three cases, especially at sites 7, 8, 11, 12, and 13. A positively charged arginine (R) with its long side chain is at site 7; a hydrophobic tryptophan (W) is at site 8; and three hydrophilic amino acids, i.e., asparagine (N), cysteine (C), and glutamine (Q), are at sites 11, 12, and 13, respectively. Since these sites always point toward their proximate nucleotides on ASL^{Lys3}, the amino acid side chains located at these sites have a good spatial opportunity to contact the ASL^{Lys3}. Detailed discussion of this point is given in a later section on the energy analysis.

b. Energy Analysis. The binding energy, the binding energy without GBSUR, the VDW energy, the sum of (ELE+EGB) energies, and the GBSUR energy for the three best peptide sequences in the three cases have been compared (Table 5). Examination of the energies in Table 5 shows that

Table 5. Contributions to the Energy for the Best Peptide Sequences in the Three Cases

cases	binding energy	binding energy without GBSUR	VDW	ELE +EGB	GBSUR
one[a]	−46.47	−39.71	−34.33	−5.38	−6.76
two[b]	−32.19	−25.35	−42.15	16.80	−6.84
three[c]	−40.92	−34.47	−33.75	−0.72	−6.45

[a] PGMMTNRWTWNCQGR. [b] RGSISMRWTSNCQIY. [c] PGGMSSRWHHNCQWP.

the peptide sequences in the three cases exhibit notable differences in the VDW energy and the ELE+EGB energy. The different peptide's hydration properties strongly affect the charge−charge (ELE+EGB) interaction as a result of the different number of the hydrophilic or positively charged amino acids on the peptide chain. A strong VDW interaction (a relatively short-range force) means that the structures are bound together tightly. However, an excessively tight binding structure easily leads to a repulsive (positive) charge−charge (ELE+EGB) energy, thereby hindering the binding.

As the new algorithm's ability to optimize the conformation appears to boost the binding capability of the peptide chain, a question arises: is it the conformation of the main chain (N−C_α−C) on the peptide that advances the binding capability or is it the conformation of the side chains? To answer this question, Figure 6 shows maps of the VDW and ELE+EGB interactions between the main chain (backbone) of the peptide and the bases on ASL^{Lys3} in case one when there is no conformational change, panels a and c, and when there is a conformational change, panels b and d. In comparing the VDW (Figure 6a) and the ELE+EGB (Figure 6c) energies of the old and new (Figure 6b and d) peptide conformations, we observe that there is a small decrease in the VDW (Figure 6b) and the ELE+EGB (Figure 6d) energies at sites 10, 11, 12, 13, and 15. This means that the interactions between the main chain near the C-terminus of the peptide and the bases of the ASL^{Lys3} are strengthened when conformation changes are allowed. A decrease in energy implies an improvement of the binding capability of the peptide. Although the new conformation leads to a decrease in the VDW and the ELE+EGB energies between the peptide backbone and the modified ASL^{Lys3} loop, this improvement is not sufficient enough to account for the improvement of the binding capability of the entire peptide chain. For example, in case one of Table 3, the binding free energy decreases from −9.83 kcal/mol when there is no conformation change to −39.71 kcal/mol when conformation changes are allowed. However, the decrease in the VDW and the ELE+EGB energies at the C-terminus for this case (approximately −5.00 kcal/mol in total) is not enough to account for the decrease in the total free energy. We conclude from this example that, although the change of the backbone conformation in the hybrid search algorithm advances the binding capability of the main chain of the peptide, the major improvement must come from the side chains because it does not come from the main chain.

To better understand the interactions between the side chains and the ASL^{Lys3}, we have compared a set of energy maps for case one (Figure 7, left-side panels a, c, e, and g) with that of case three (Figure 7 right-side panels b, d, f, and h), referring to the VDW energy and the sum of the ELE+EGB energies. We first focus on the interactions (Figure 7a,b,c,d) between the side chains and the bases of ASL^{Lys3}. As can be seen from the energy maps in Figure 7a,b, the hydrophilic amino acids at the C-terminus of the peptide interact strongly with the modified anticodon loop domain, especially with the two modified nucleotides via VDW interactions. For example, the asparagine at site 11 and the cysteine at site 12 have an intense preference for the special anticodon loop, $mcm^5s^2U_{34}-U_{35}-U_{36}-ms^2t^6A_{37}$. As is well-known, the unique order of the bases and the unique chemistries of these two natural modifications within the anticodon loop of $tRNA^{Lys3}$ play important roles in the virus' recruitment of the tRNA and the tRNA's annealing to the virus' primer binding site. The observation that the asparagine at site 11 and the cysteine at site 12 interact strongly with the anticodon loop implies that the two hydrophilic amino acids "recognize" ASL^{Lys3}, thereby improving binding specificity. Next, we focus on the energy interactions (Figure 7e,f,g,h) between the side chains and the sugar ring/phosphate linkage of the ASL^{Lys3}. The positively charged amino acids preferentially attract the sugar ring/phosphate linkage as indicated by the charge−charge (ELE+EGB) interaction, enhancing binding affinity. For example, arginine with its positive charge attracts the phosphate linkages in ASL^{Lys3}, as shown in Figure 7g,h, providing a general binding capability. Other amino acids such as proline at site 1, methionine at site 4, and tryptophan at sites 8 and 14 in the peptide sequence also contribute to the binding to some extent, as shown in Figure 7c,d,e,f. It is noted that to have good binding the sequence not only needs the key amino acids but also needs a good folded conformation, which can effectively promote and enhance the binding specificity and affinity for the key amino acids.

c. Conformation Analysis. The complexes formed by ASL^{Lys3} and the peptide chain obtained in the hybrid search algorithm with and without the conformation changes are

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

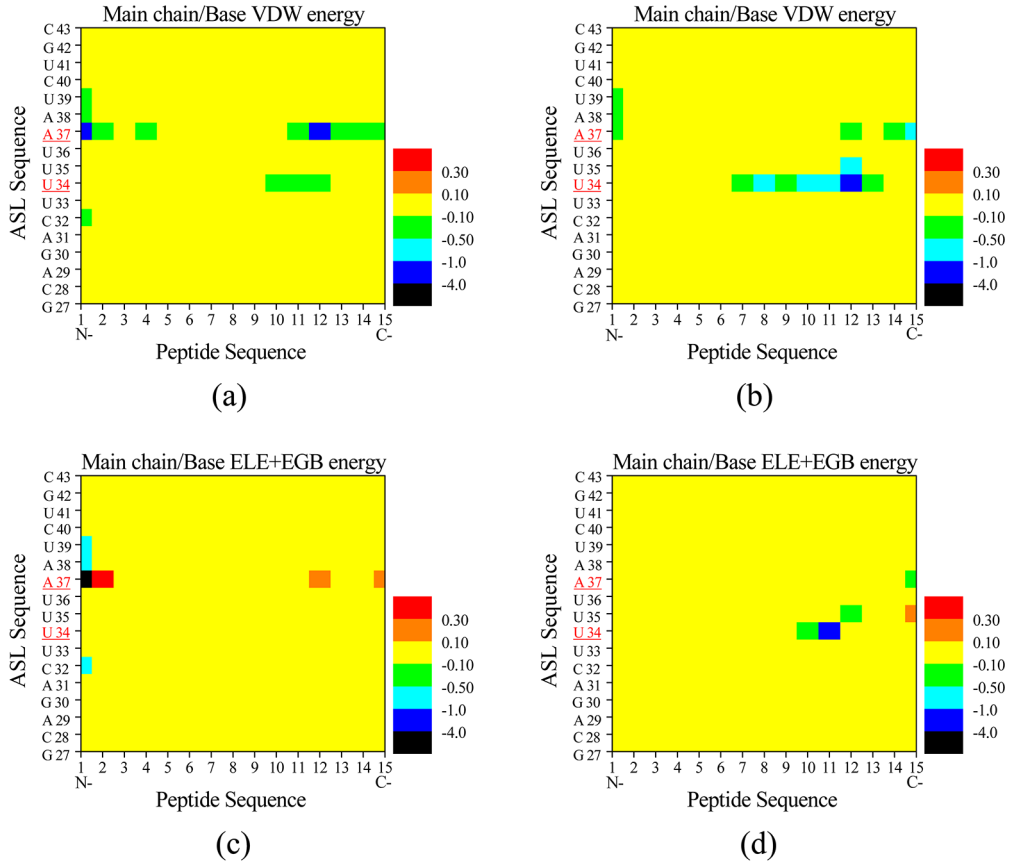


Figure 6. A map of the contributions to the binding energy for interactions between the main chain (backbone) on the peptide and the bases on ASL for case one with and without conformational changes. (a) VDW energy and (c) ELE+EGB energy without conformational change moves; (b) VDW energy and (d) ELE+EGB energy with conformational change moves. The *x* axis represents the sites along the peptide chain; the *y* axis represents the base sequence along ASL, and the color bar on the right scales the value of the energies.

shown in Figure 8. The red ribbon is the initial conformation, and the blue ribbon represents a new folded conformation of the peptide chain. It can be seen that the helix in the middle of the peptide remains at its original position, but both ends move freely in a β-strand configuration. The helix region stacks on the C_{32}, U_{33}, $ms^2t^6A_{37}$, and A_{38} of the ASL^{Lys3}, serving as a strong "anchor" to provide binding affinity. In contrast, the strand regions prefer interacting with the $mcm^5s^2U_{34}$–U_{35}–U_{36}–$ms^2t^6A_{37}$ region of the anticodon loop domain, serving as a strong "recognizer" to provide binding specificity. When performing the hybrid search algorithm to evolve the sequences, we found that the helix region is usually stable and retains its folded structure, but the strand region is always flexible and easily adjusts its conformation. This is consistent with the experimental observations by Xia et al.[68] who used a combination of fluorescence up-conversion and transient absorption and found that the complex formed by the antiterminator N protein and the stem-loop RNA hairpin exists in a dynamic equilibrium. Experimentally, the N-terminal helical domain of the bound peptide always stacks with the RNA, but the C-terminal helical domain undergoes a change of conformation between stacked and unstacked states. Zhang et al.[69] utilized site-directed spin labeling to examine the conformation distributions at the interface between a peptide and a stem-loop RNA element. They observed that the C-terminal fragment of the bound peptide tends to adopt multiple discrete conformations in the complex.

To obtain a better understanding of the advantages of the new hybrid search algorithm relative to the old one, we compared properties for the original P6 sequence and two best sequences from case one, viz. PGMMTNRWTWNCQGR and PHWRTTGWMNNCRMG, which are obtained from this new search algorithm (with conformational change) and our old search algorithm (without conformational change; see ref 52), respectively. Their conformational properties, including $\langle R_g^2 \rangle$ (mean square radius of gyration), the SASA (solvent accessible surface area), the GBSUR energy (nonpolar solvation energy), the VDW energy, the ELE+EGB energy, and the binding energy without GBSUR, have been listed (Table 6). It is clear that the peptide resulting from the new search algorithm (with conformational changes) has a lower binding energy when compared to the original P6 peptide and the peptide resulting from the old search algorithm (without conformational changes). Allowing conformational changes results in an increase of the $\langle R_g^2 \rangle$ of the peptide's main chain from 44.25 to 48.88, and an increase in the corresponding SASA from 1621.89 Å^2 (P6 peptide) and 1989.02 Å^2 (old binding conformation) to 2158.44 Å^2 (new binding conformation). This indicates that the folded chain has elongated its structure and exposed more previously hidden surface area to ASL^{Lys3}. This, of course, causes an increase in the molecular interactions between peptide chain and the ASL^{Lys3}, as is verified by the fact that the VDW and the ELE+EGB energies become much lower when the conformation is changed. The decrease in the VDW energy from −30.65 kcal/mol (P6 peptide) and −27.59 kcal/

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

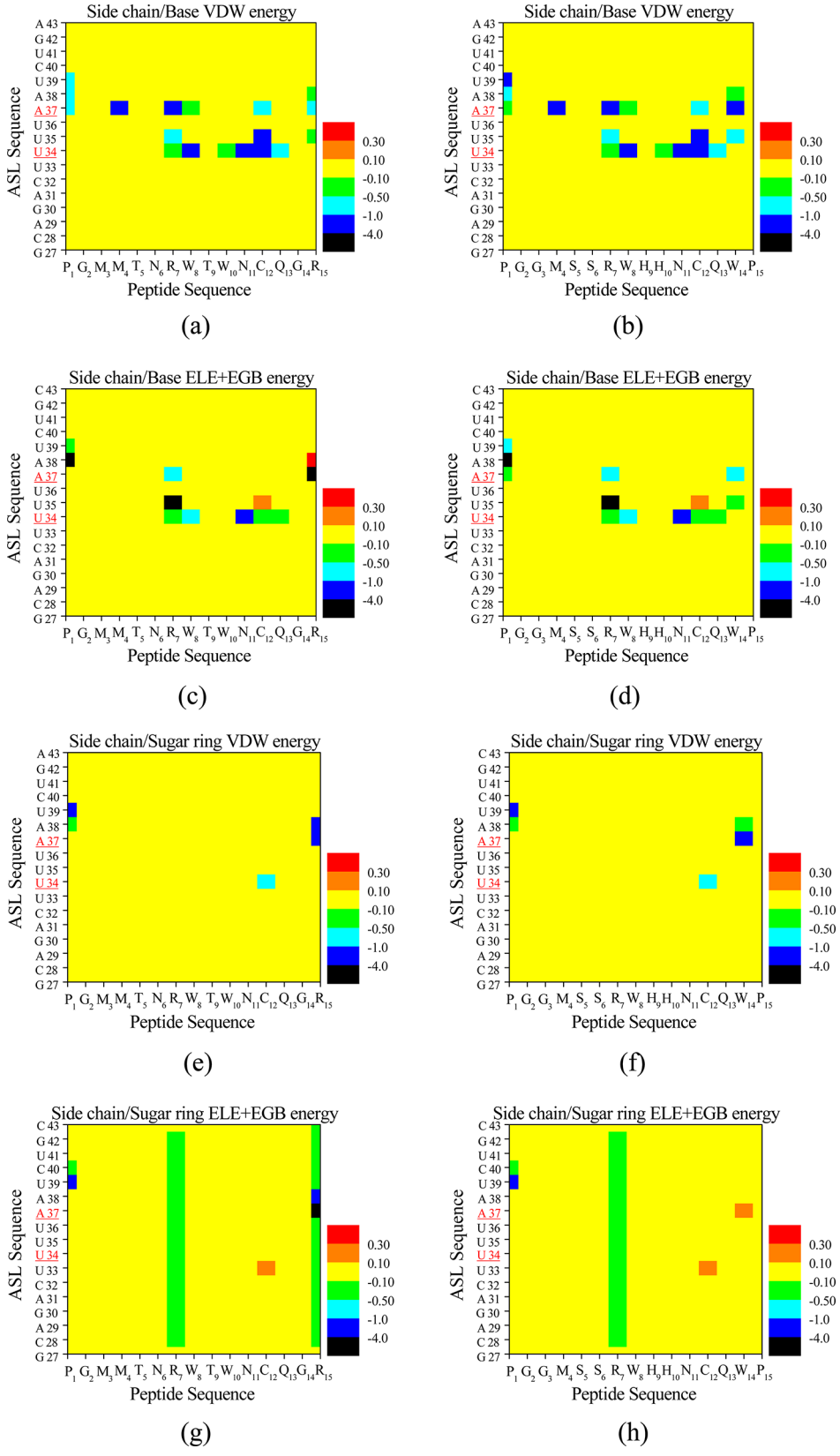


Figure 7. A map of the contributions to the binding energy for interactions between the side chains on the peptide and the nucleotides on ASL for case one (left-hand side) and case three (right-hand side) with conformation changes. (a) VDW energy and (c) ELE+EGB energy involving the peptide side chain and the base of ASL^{Lys3} for case one. (e) VDW energy and (g) ELE+EGB energy involving the peptide side chain and the sugar ring and phosphate linkage of ASL^{Lys3} for case one. (b) VDW energy and (d) ELE+EGB energy involving the peptide side chain and the base of ASL^{Lys3} for case three. (f) VDW energy and (h) ELE+EGB energy involving the peptide side chain and the sugar ring and phosphate linkage of ASL^{Lys3} for case three. The *x* axis represents the residue sequence along the peptide chain; the *y* axis represents the nucleotide sequence along ASL, and the color bar on the right scales the value of the energies.

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

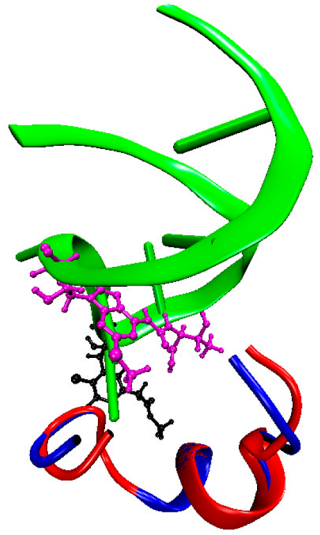


Figure 8. Snapshot of the complex formed by ASL^{Lys3} and the peptide chain in case one when conformational changes are included in the algorithm. ASL^{Lys3} is represented by the green ribbon, and the two modified nucleotides ($mcm^5s^2U_{34}$ and $ms^2t^6A_{37}$) are colored in black and magenta, respectively. The blue ribbon shows a new binding configuration of the peptide chain, and the red one shows the initial configuration of the peptide chain. The other nonmodified nucleotides on ASL^{Lys3} and all of the side chains on the peptide are not shown for clarity.

Table 6. Comparison of the Conformational Properties of the P6 Peptide Sequence and the Best Peptide Sequences Resulting from the Search Algorithms with Conformational Changes (PGMMTNRWTWNCQGR) and without Conformational Changes (PHWRTTGWMNNCRMG) in Case One

	P6 peptide sequence	best sequence without conformation changes	best sequence with conformation changes
$\langle R_g^2 \rangle$	44.25	44.25	48.88
SASA (Å^2)	1621.89	1989.02	2158.44
GBSUR (kcal/mol)	−4.33	−5.63	−6.76
VDW (kcal/mol)	−30.65	−27.59	−34.33
ELE+EGB (kcal/mol)	10.93	17.75	−5.38
binding energy without GBSUR (kcal/mol)	−19.72	−9.83	−39.71

mol (old binding conformation) to −34.33 kcal/mol (new binding conformation) indicates improved recognition of the peptide for the ASL^{Lys3}. The sizable decrease in the ELE+EGB energy from 10.93 kcal/mol (P6 peptide) and 17.75 kcal/mol (old binding conformation) to −5.38 kcal/mol (new binding conformation) also results in a significant improvement in the binding capability of the peptide to the ASL^{Lys3}, as shown in the binding energy without GBSUR (Table 6). We conclude that the new hybrid search algorithm is able to sample effectively the conformational space and to find better conformations and sequences than the old search algorithm.

d. Comparison of Sequence Evolution Starting from the P6 Peptide and a Random Peptide. It is of interest to see how the new search algorithm performs using P6 as the starting sequence as opposed to the random starting sequence already described. Generally speaking, our rationale for starting searches using a randomly generated sequence was to avoid potential traps in local energy minimum. This was based on previous experience on other projects which showed that we could get to lower energy states by starting with a random sequence than with the sequence that we wish to improve upon, so long as the complex remained fixed in good candidate conformation. Here, we compare the results obtained using the original P6 peptide and a randomly generated peptide as starting sequences in the new search algorithm where both sequence and conformation can vary. Figure 9 shows the

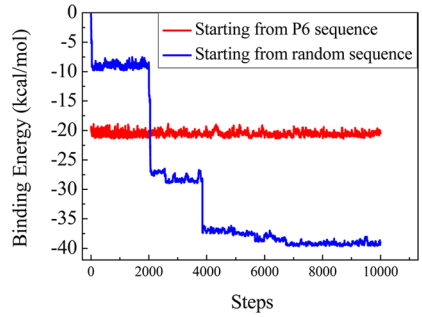


Figure 9. Comparison of sequence evolution when starting from the P6 peptide (red line) and a random peptide (blue line).

sequence evolution results starting from P6 and from a random sequence. The initial conformation in both cases is the same: the complex between the P6 peptide and the ASL^{Lys3} was obtained from a 60 ns atomistic MD simulation.[57] It can be seen from the figure that when we start with the P6 peptide sequence, the evolution process is trapped into a deep local energy-minimum well. Few of the conformation change moves are accepted, and hence the energy profile shows only slight fluctuation over time. Evidently, the peptide backbone conformation moves become too hard to make when the evolution process starts from such a low energy point. In contrast, when the randomly generated peptide is draped over the P6 backbone, the energy initially increases, avoiding an initial low energy point, and allowing both the sequence and conformation moves to be executed more frequently. As a result, the energy of the best peptide starting with the random sequence (binding energy = −46.47 kcal/mol; binding energy without GBSUR = −39.71 kcal/mol) is much lower than the energy of the best peptide starting from P6 (binding energy = −26.60 kcal/mol; binding energy without GBSUR = −21.41 kcal/mol).

4. CONCLUSION

By adding conformational moves in a novel hybrid search algorithm, we extended and enhanced our previous Monte Carlo (MC)/self-consistent mean field (SCMF) algorithm for identifying peptide sequences that bind to the naturally modified human ASL^{Lys3} with high affinity and selectivity. Our ultimate goal was to design a peptide sequence that could serve as a drug candidate that binds the anticodon stem and loop domain of the human $tRNA^{Lys3}$ so as to interfere with the replication of the HIV-1 virus during infection. The solution structure by NMR and crystal structure of the fully modified human $tRNA^{Lys3}$ had been determined, facilitating this endeavor.[70]

A brief description of the search algorithm for finding peptides with the highest binding affinity and specificity is

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

given. The SCMF technique is used to repack the side chains and determine the best combination of rotamers. The concerted rotation (CONROT) method is used to move the peptide during the search for conformation candidates. Two probability parameters, $P_{conformation}$ and $P_{sequencelconformation}$, are used to control the process of evolution so that the peptide has either a conformation change, a sequence change, or both simultaneously. Thus, a large number of trial sequences and conformations can be generated by performing the new hybrid search algorithm. We calculate the binding free energy (without GBSUR) for these new attempted sequences and conformations, and then employ the MC technique to accept or reject the attempted sequences and conformations based on the Metropolis sampling method.

The novel hybrid search algorithm with adjustable probability parameters ($P_{conformation}$, $P_{sequencelconformation}$) was used to find the best binding peptides for three different cases of peptide hydration properties, and then the best binding structures and binding energies for the evolved peptide sequences were analyzed. The new sequences had a better binding capability to ASL^{Lys3} than the peptide sequences that were evolved without allowing conformation changes. The reason for the improvement is that conformation change moves in the hybrid search algorithm allow us to find more conformers that are accessible to the binding receptor, thereby enhancing the likelihood of binding contacts between peptide chain and ASL^{Lys3}. We also found that the improvement in the binding energy compared to the old one is mainly due to the decrease of the ELE+EGB energy, giving the new sequences enhanced ability to access to ASL^{Lys3}. Furthermore, by plotting maps of the various contributions to the binding energy, we observed that the hydrophilic amino acids at the C-terminus of the peptide, especially asparagine at site 11 and cysteine at site 12, play important roles in selectively "recognizing" the modified nucleosides of ASL^{Lys3} via the VDW interaction, thus improving the binding specificity. In addition, the positively charged amino acids on sites 7 and 15 in case one, on sites 1 and 7 in case two, and on site 7 in case three preferentially attract the sugar ring/phosphate linkage due to the ELE+EGB interaction, improving the binding affinity. Finally, we examined conformational properties, such as $\langle R_g^2 \rangle$, the SASA, the GBSUR, the VDW, and the ELE+EGB energies in order to compare the evolved peptide sequences obtained via the new and old search algorithms. Overall, the novel hybrid MC/SCMF/CONROT search algorithm generates sequences with better binding affinity and specificity to ASL^{Lys3} than the old search algorithm.

The validity of the hybrid search algorithm will be examined further by synthesizing and testing peptide sequences selected in this manner against ASL^{Lys3}. Our preliminary data appear to support this search algorithm.[14] Some candidate sequences designed via an early version of the algorithm[52] were selected to test their binding capabilities with the target modified ASL^{Lys3} and the decoy unmodified ASL^{Lys3} in experiments. By measuring fluorescent signals in the assays, we found that most designed sequences are able to specifically bind to the modified ASL^{Lys3} rather than to the unmodified ASL^{Lys3}. One peptide sequence (RWQMTAFAHGWRHSG), even exhibited a 10-fold higher affinity than the P6 peptide. Peptide selectivity will be further increased by eliminating those peptides that also bind the unmodified ASL^{Lys3} with significant affinity, but without specificity. The approaches detailed here are applicable to both linear and circular peptides that have become favored as therapeutic agents.[71−73] Our contribution in this work is to develop a search algorithm that can identify potential peptide drugs for HIV in a fast and effective way. It is important to point out that any potential peptide candidates that we identify must be thoroughly evaluated *in vitro* and then *in vivo*. Beyond that are many steps including evaluation of the toxicity of the peptide drugs to humans.

■ ASSOCIATED CONTENT

Supporting Information

Detailed descriptions of the calculations of the torsion angles (ϕ_1, ψ_1, ϕ_2, ψ_2, ϕ_3, ψ_3) in CONROT move. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: hall@ncsu.edu.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We acknowledge Dr. Doros N. Theodorou for valuable and helpful discussions on his Concerted Rotation method. Financial support for this work was awarded by the North Carolina Biotechnology Center to P.F.A. and C.K.H. (2008MRG1102), by the National Science Foundation to C.K.H. (CBET-0835794) and to P.F.A. (MCB1101859), and by the National Institutes of Health USA to C.K.H. (EB006006) and to P.F.A. (GM-23037). This work was also supported in part by the NSF's Research Triangle MRSEC, DMR-1121107. This work used the Extreme Science and Engineering Discovery Environment (XSEDE), which is supported by National Science Foundation grant number ACI-1053573. We thank the National Institute for Computational Science (NICS), Texas Advanced Computing Center (TACC), and San Diego Supercomputer Center (SDSC) for providing us computing time.

■ REFERENCES

(1) Norris, P. J.; Rosenberg, E. S. *AIDS* **2001**, *15*, S16.
(2) Werb, D.; Mills, E. J.; Montaner, J. S. G.; Wood, E. *Lancet Infect. Dis.* **2010**, *10*, 464.
(3) Kleiman, L.; Caudry, S.; Boulerice, F.; Wainberg, M. A.; Parniak, M. A. *Biochem. Biophys. Res. Commun.* **1991**, *174*, 1272.
(4) Marquet, R.; Isel, C.; Ehresmann, C.; Ehresmann, B. *Biochimie* **1995**, *77*, 113.
(5) Tisné, C.; Roques, B. P.; Dardel, F. *Biochimie* **2003**, *85*, 557.
(6) Barraud, P.; Gaudin, C.; Dardel, F.; Tisné, C. *Biochimie* **2007**, *89*, 1204.
(7) Puglisi, E. V.; Puglisi, J. D. *J. Mol. Biol.* **2011**, *410*, 863.
(8) Tisné, C. *Curr. HIV Res.* **2005**, *3*, 147.
(9) Watts, J. M.; Dang, K. K.; Gorelick, R. J.; Leonard, C. W.; Bess, J. W., Jr; Swanstrom, R.; Burch, C. L.; Weeks, K. M. *Nature* **2009**, *460*, 711.
(10) Isel, C.; Ehresmann, C.; Marquet, R. *Viruses* **2010**, *2*, 213.
(11) Guo, M.; Shapiro, R.; Morris, G. M.; Yang, X.; Schimmel, P. *J. Phys. Chem. B* **2010**, *114*, 16273.
(12) Eshete, M.; Marchbank, M. T.; Deutscher, S. L.; Sproat, B.; Leszcynska, G.; Malkiewicz, A.; Agris, P. F. *Protein J.* **2007**, *26*, 61.
(13) Graham, W. D.; Barley-Maloney, L.; Stark, C. J.; Kaur, A.; Stolyarchuk, K.; Sproat, B.; Leszczynska, G.; Malkiewicz, A.; Safwat, N.; Mucha, P.; Guenther, R.; Agris, P. F. *J. Mol. Biol.* **2011**, *410*, 698.
(14) Spears, J. L.; Xiao, X.; Hall, C. K.; Agris, P. F. *Biochemistry* **2014**, *53*, 1125.
(15) Lippow, S. M.; Tidor, B. *Curr. Opin. Biotechnol.* **2007**, *18*, 305.

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752

(16) Samish, I.; MacDermaid, C. M.; Perez-Aguilar, J. M.; Saven, J. G. *Annu. Rev. Phys. Chem.* **2011**, *62*, 129.
(17) Jiang, L.; Althoff, E. A.; Clemente, F. R.; Doyle, L.; Röthlisberger, D.; Zanghellini, A.; Gallaher, J. L.; Betker, J. L.; Tanaka, F.; Barbas, C. F., III; Hilvert, D.; Houk, K. N.; Stoddard, B. L.; Baker, D. *Science* **2008**, *319*, 1387.
(18) Ashworth, J.; Havranek, J. J.; Duarte, C. M.; Sussman, D.; Monnat, R. J., Jr.; Stoddard, B. L.; Baker, D. *Nature* **2006**, *441*, 656.
(19) Ofek, G.; Guenaga, F. J.; Schief, W. R.; Skinner, J.; Baker, D.; Wyatt, R.; Kwong, P. D. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 17880.
(20) Dahiyat, B. I.; Mayo, S. L. *Science* **1997**, *278*, 82.
(21) Voigt, C. A.; Gordon, D. B.; Mayo, S. L. *J. Mol. Biol.* **2000**, *299*, 789.
(22) Simons, K. T.; Bonneau, R.; Ruczinski, I.; Baker, D. *Proteins* **1999**, *37*, 171.
(23) Zhou, J.; Saven, J. G. *J. Mol. Biol.* **2000**, *296*, 281.
(24) Tang, J.; Kang, S.; Saven, J. G.; Gai, F. *J. Mol. Biol.* **2009**, *389*, 90.
(25) Desmet, J.; Maeyer, M. D.; Hazes, B.; Lasters, I. *Nature* **1992**, *356*, 539.
(26) Jones, D. T. *Protein Sci.* **1994**, *3*, 567.
(27) Wang, C.; Bradley, P.; Baker, D. *J. Mol. Biol.* **2007**, *373*, 503.
(28) Chaudhury, S.; Gary, J. J. *J. Mol. Biol.* **2008**, *381*, 1068.
(29) Georgiev, I.; Keedy, D.; Richardson, J. S.; Richardson, D. C.; Donald, B. R. *Bioinformatics* **2008**, *24*, i196.
(30) Mandell, D. J.; Kortemme, T. *Curr. Opin. Chem. Biol.* **2009**, *20*, 420.
(31) Hallen, M. A.; Keedy, D. A.; Donald, B. R. *Proteins* **2013**, *81*, 18.
(32) Correia, B. E.; Ban, Y. A.; Friend, D. J.; Ellingson, K.; Xu, H.; Boni, E.; Bradley-Hewitt, T.; Bruhn-Johannsen, J. F.; Stamatatos, L.; Strong, R. K.; Schief, W. R. *J. Mol. Biol.* **2011**, *405*, 284.
(33) Karanicolas, J.; Corn, J. E.; Chen, I.; Joachimiak, L. A.; Dym, O.; Peck, S. H.; Albeck, S.; Unger, T.; Hu, W.; Liu, G.; Delbecq, S.; Montelione, G. T.; Spiegel, C. P.; Liu, D. R.; Baker, D. *Mol. Cell* **2011**, *42*, 1.
(34) Davis, I. W.; Arendall, W. B.; Richardson, D. C.; Richardson, J. S. *Structure* **2006**, *14*, 265.
(35) Smith, C. A.; Kortemme, T. *J. Mol. Biol.* **2008**, *380*, 742.
(36) Rohl, C. A.; Strauss, C. E. M.; Chivian, D.; Baker, D. *Proteins* **2004**, *55*, 656.
(37) Hou, T.; Xu, X. *J. Mol. Graphics Model.* **2001**, *19*, 455.
(38) Mandell, D. J.; Kortemme, T. *Nat. Chem. Biol.* **2009**, *5*, 797.
(39) Gray, J. J.; Moughon, S.; Wang, C.; Schueler-Furman, O.; Kuhlman, B.; Rohl, C. A.; Baker, D. *J. Mol. Biol.* **2003**, *331*, 281.
(40) Wang, C.; Vernon, R.; Lange, O.; Tyka, M.; Baker, D. *Protein Sci.* **2010**, *19*, 494.
(41) Lewis, S. M.; Kuhlman, B. A. *PLoS One* **2011**, *6*, e20872.
(42) Sammond, D. W.; Bosch, D. E.; Butterfoss, G. L.; Purbeck, C.; Machius, M.; Siderovski, D. P.; Kuhlman, B. *J. Am. Chem. Soc.* **2011**, *133*, 4190.
(43) Jha, R. K.; Leaver-Fay, A.; Yin, S.; Wu, Y.; Butterfoss, G. L.; Szyperski, T.; Dokholyan, N. V.; Kuhlman, B. *J. Mol. Biol.* **2010**, *400*, 257.
(44) Guntas, G.; Purbeck, C.; Kuhlman, B. *Proc. Natl. Acad. Sci. U. S. A.* **2010**, *107*, 19196.
(45) Murphy, G. S.; Mills, J. L.; Miley, M. J.; Machius, M.; Szyperski, T.; Kuhlman, B. *Structure* **2012**, *20*, 1086.
(46) Havranek, J. J.; Duarte, C. M.; Baker, D. *J. Mol. Biol.* **2004**, *344*, 59.
(47) Morozov, A. V.; Havranek, J. J.; Baker, D.; Siggia, E. D. *Nucleic Acids Res.* **2005**, *33*, 5781.
(48) Ashworth, J.; Havranek, J. J.; Duarte, C. M.; Sussman, D.; Monnat, R. J., Jr; Stoddard, B. L.; Baker, D. *Nature* **2006**, *441*, 656.
(49) Ashworth, J.; Baker, D. *Nucleic Acids Res.* **2009**, *37*, e73.
(50) Thyme, S. B.; Baker, D.; Bradley, P. *J. Mol. Biol.* **2012**, *419*, 255.
(51) Chen, Y.; Kortemme, T.; Robertson, T.; Baker, D.; Varani, G. *Nucleic Acids Res.* **2004**, *32*, 5147.
(52) Xiao, X.; Hall, C. K.; Agris, P. F. *J. Biomol. Struct. Dyn.* **2014**, *32*, 1523.
(53) Ramachandran, G. N.; Ramakrishnan, C.; Sasisekharan, V. *J. Mol. Biol.* **1963**, *7*, 95.
(54) Ramachandran, G. N.; Sasisekharan, V. *Adv. Protein Chem.* **1968**, *23*, 284.
(55) Lovell, S. C.; Davis, I. W.; Arendall, W. B., III; de Bakker, P. I. W.; Word, J. M.; Prisant, M. G.; Richardson, J. S.; Richardson, D. C. *Proteins* **2003**, *50*, 437.
(56) Cheon, M.; Chang, I.; Hall, C. K. *Proteins* **2010**, *78*, 2950.
(57) Xiao, X.; Agris, P. F.; Hall, C. K. *J. Biomol. Struct. Dyn.* **2015**, *33*, 14.
(58) Koehl, P.; Delarue, M. *J. Mol. Biol.* **1994**, *239*, 249.
(59) Koehl, P.; Levitt, M. *J. Mol. Biol.* **1999**, *293*, 1161.
(60) Lovell, S. C.; Word, J. M.; Richardson, J. S.; Richardson, D. C. *Proteins* **2000**, *40*, 389.
(61) Hawkins, G. D.; Cramer, C. J.; Truhlar, D. G. *J. Phys. Chem.* **1996**, *100*, 19824.
(62) Jayaram, B.; Liu, Y.; Beveridge, D. L. *J. Chem. Phys.* **1998**, *109*, 1465.
(63) Jayaram, B.; Sprous, D.; Beveridge, D. L. *J. Phys. Chem. B* **1998**, *102*, 9571.
(64) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712.
(65) Gohlke, H.; Kiel, C.; Case, D. A. *J. Mol. Biol.* **2003**, *330*, 891.
(66) Dodd, L. R.; Boone, T. D.; Theodorou, D. N. *Mol. Phys.* **1993**, *78*, 961.
(67) Ulmschneider, J. P.; Jorgensen, W. L. *J. Am. Chem. Soc.* **2004**, *126*, 1849.
(68) Xia, T.; Wan, C.; Roberts, R. W.; Zewail, A. H. *Proc. Natl. Acad. Sci. U. S. A.* **2005**, *102*, 13013.
(69) Zhang, X.; Lee, S. W.; Zhao, L.; Xia, T.; Qin, P. Z. *RNA* **2010**, *16*, 2474.
(70) Murphy, F. V., IV; Ramakrishnan, V.; Malkiewicz, A.; Agris, P. F. *Nat. Struct. Biol.* **2004**, *11*, 1186.
(71) Pechar, M.; Pola, R.; Laga, R.; Braunová, A.; Filippov, S. K.; Bogomolova, A.; Bednárová, L.; Vaněk, O.; Ulbrich, K. *Biomacromolecules* **2014**, *15*, 2590.
(72) Garcia, A. E.; Camarero, J. A. *Mol. Pharmacol.* **2010**, *3*, 153.
(73) Hall, C. K.; Agris, P. F.; Xiao, X. Peptide Therapeutic Against HIV. Provisional Patent Application No. 61/893,600, October 21, 2013.

DOI: 10.1021/ct5008247
J. Chem. Theory Comput. 2015, 11, 740−752