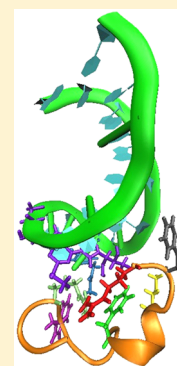


## Amino Acid Signature Enables Proteins to Recognize Modified tRNA

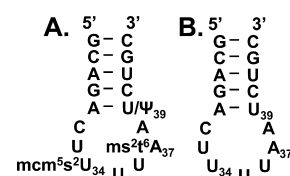
Jessica L. Spears,<sup>†,‡,||</sup> Xingqing Xiao,<sup>§</sup> Carol K. Hall,<sup>§</sup> and Paul F. Agris<sup>\*,†,‡</sup><sup>†</sup>The RNA Institute, University at Albany-SUNY, Life Sciences Research Building, Albany, New York 12222, United States<sup>‡</sup>Department of Biological Sciences, University at Albany-SUNY, Life Sciences Research Building, Albany, New York 12222, United States<sup>§</sup>Department of Chemical and Biomolecular Engineering, North Carolina State University, Engineering Building, Raleigh, North Carolina 27695-7905, United States

**ABSTRACT:** Human tRNA<sup>Lys3</sup><sub>UUU</sub> is the primer for HIV replication. The HIV-1 nucleocapsid protein, NCp7, facilitates htRNA<sup>Lys3</sup><sub>UUU</sub> recruitment from the host cell by binding to and remodeling the tRNA structure. Human tRNA<sup>Lys3</sup><sub>UUU</sub> is post-transcriptionally modified, but until recently, the importance of those modifications in tRNA recognition by NCp7 was unknown. Modifications such as the 5-methoxycarbonylmethyl-2-thiouridine at anticodon wobble position-34 and 2-methylthio-*N*<sup>6</sup>-threonylcarbamoyladenosine, adjacent to the anticodon at position-37, are important to the recognition of htRNA<sup>Lys3</sup><sub>UUU</sub> by NCp7. Several short peptides selected from phage display libraries were found to also preferentially recognize these modifications. Evolutionary algorithms (Monte Carlo and self-consistent mean field) and assisted model building with energy refinement were used to optimize the peptide sequence *in silico*, while fluorescence assays were developed and conducted to verify the *in silico* results and elucidate a 15-amino acid signature sequence (R-W-Q/N-H-X<sub>2</sub>-F-Pho-X-G/A-W-R-X<sub>2</sub>-G, where X can be most amino acids, and Pho is hydrophobic) that recognized the tRNA's fully modified anticodon stem and loop domain, hASL<sup>Lys3</sup><sub>UUU</sub>. Peptides of this sequence specifically recognized and bound modified htRNA<sup>Lys3</sup><sub>UUU</sub> with an affinity 10-fold higher than that of the starting sequence. Thus, this approach provides an effective means of predicting sequences of RNA binding peptides that have better binding properties. Such peptides can be used in cell and molecular biology as well as biochemistry to explore RNA binding proteins and to inhibit those protein functions.



There are three human isoaccepting tRNAs for the amino acid lysine, htRNA<sup>Lys1,2,3</sup>. The three human tRNA<sup>Lys</sup> decode the two lysine codons, AAA and AAG. Two of the isoacceptors, htRNA<sup>Lys1,2</sup><sub>CUU</sub> with the anticodon CUU, decode AAG, but only one, htRNA<sup>Lys3</sup><sub>UUU</sub> with the anticodon UUU, responds to the cognate codon AAA and wobbles to AAG. Besides its important role in protein synthesis, htRNA<sup>Lys3</sup><sub>UUU</sub> serves as the primer of reverse transcription in the replication of the lentiviruses, including human immunodeficiency virus type 1 (HIV-1).<sup>1,2</sup> During the replication of HIV-1, the host cell htRNA<sup>Lys3</sup><sub>UUU</sub> is recognized and bound and its structure destabilized by nucleocapsid protein 7 (NCp7).<sup>3–5</sup> This destabilization allows the relaxed U-rich anticodon stem loop (hASL<sup>Lys3</sup><sub>UUU</sub>), as well as the acceptor stem (Figure 1), to be annealed to the HIV viral RNA. During the subsequent infection, htRNA<sup>Lys3</sup><sub>UUU</sub> is the primer for HIV reverse transcriptase.

htRNA<sup>Lys3</sup><sub>UUU</sub> is one of the most uniquely processed tRNAs having chemically rich post-transcriptional modifications that are important to conformation and function of the tRNA during protein synthesis.<sup>6</sup> Until recently, the role(s) these modifications play in the tRNA's interaction with NCp7 and in viral replication were not known. The naturally occurring modifications, 5-methoxycarbonylmethyl-2-thiouridine (mcm<sup>5</sup>s<sup>2</sup>U<sub>34</sub>), at tRNA's wobble position-34, and 2-methylthio-*N*<sup>6</sup>-threonylcarbamoyladenosine (ms<sup>2</sup>t<sup>6</sup>A<sub>37</sub>), at position-37, 3'-adjacent to the anticodon in the loop of the hASL<sup>Lys3</sup><sub>UUU</sub>, are both chemically rich and constitute a unique combination in



**Figure 1.** Human modified and unmodified ASL<sup>Lys3</sup><sub>UUU</sub>. (A) Human ASL<sup>Lys3</sup><sub>UUU</sub> with all naturally occurring modifications (mcm<sup>5</sup>s<sup>2</sup>U<sub>34</sub>, ms<sup>2</sup>t<sup>6</sup>A<sub>37</sub>, and Ψ<sub>39</sub>). The construct for this study was not modified at position 39. (B) The unmodified hASL<sup>Lys3</sup><sub>UUU</sub> used in this study.

human tRNAs (Figure 1).<sup>7,8</sup> These modifications enhance NCp7's ability to recognize and bind to the RNA, suggesting that these modifications are an important discrimination factor for recognition by NCp7.<sup>9</sup> The presence of these modifications increases NCp7 affinity for hASL<sup>Lys3</sup> almost 10-fold ( $K_d = 0.28 \pm 0.03 \mu\text{M}$  for modified and  $K_d = 2.30 \pm 0.62 \mu\text{M}$  for unmodified ASL).<sup>9</sup> NCp7 is critical to HIV replication because it binds and relaxes the htRNA<sup>Lys3</sup> structure, facilitating annealing of the tRNA to the viral genomic RNA and packaging of the genomic RNA into the viral capsid.<sup>4</sup>

Fifteen- and 16-amino acid peptides have been selected to mimic NCp7's preferential recognition of the fully modified

**Received:** August 26, 2013

**Revised:** January 29, 2014

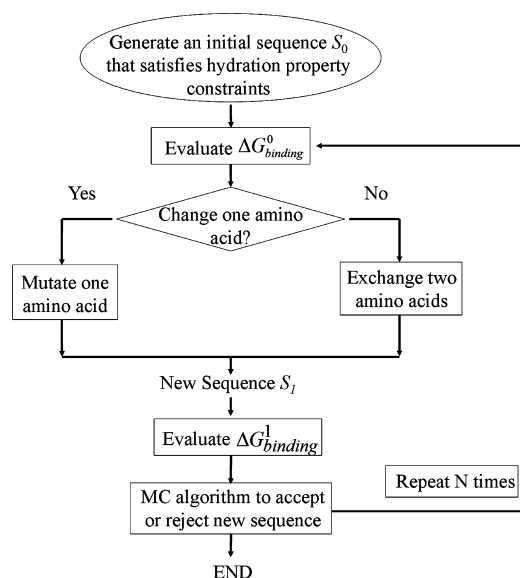
**Published:** January 31, 2014

hASL<sup>Lys3</sup><sub>UUU</sub>.<sup>9,10</sup> These peptides can be used to study modification-dependent protein recognition of RNAs, in general, and recognition and annealing of htrNA<sup>Lys3</sup><sub>UUU</sub> to the HIV viral RNA, specifically. One peptide, P6 (sequence RVTHHAFLGAHRTVG), was also shown to mimic NCp7. P6 not only binds hASL<sup>Lys3</sup><sub>UUU</sub> but also destabilizes the ASL structure as does NCp7.<sup>9</sup> The ability of peptides to mimic NCp7 makes it possible to engineer a peptide with a signature amino acid sequence that can be used as a tool in future studies of protein recognition of RNAs, particularly those with unique modifications chemistries. Herein, we report the development of a signature amino acid sequence for the recognition of htrNA<sup>Lys3</sup><sub>UUU</sub>. An algorithm was developed that optimizes the amino acid sequence by combining self-consistent mean field (SCMF) and Monte Carlo (MC) approaches. The resulting peptides were then validated as binders through empirical experimentation in order to corroborate the computer-aided approach. Amino acid substitutions *in silico* indeed enhanced the modification-dependent binding of the peptide ligand with high affinity and selectivity *in vitro*. The peptide sequences predicted by the algorithms preferentially bound the modified hASL<sup>Lys3</sup><sub>UUU</sub> with affinities at or higher than P6, and with greater specificity. The signature sequence provides insight into peptide and protein recognition of the modified tRNA<sup>Lys3</sup><sub>UUU</sub>.

## EXPERIMENTAL PROCEDURES

**Materials and Reagents.** All materials, buffers, and reagents were of RNA grade quality and RNase free. The modified and unmodified hASL<sup>Lys3</sup><sub>UUU</sub> were chemically synthesized by Thermo Scientific. The modified hASL<sup>Lys3</sup><sub>UUU</sub> was synthesized with the nucleoside phosphoramidites that were 2'-protected with *tert*-butyldimethylsilyl-ether.<sup>9</sup> The unmodified hASL<sup>Lys3</sup><sub>UUU</sub> was synthesized with "ACE" chemistry.<sup>11</sup> All fluorescein labeled peptides were obtained from Sigma-Aldrich (PEPscreen) with sequence verification by mass spectrometry.

**In Silico Evolution of Peptide Sequences.** A random initial sequence that satisfies the constraints on hydration properties is generated (Figure 2). For the search described here, we started with the 15-amino-acid sequence of peptide P6, RVTHHAFLGAHRTVG, found experimentally to bind selectively to the modified hASL<sup>Lys3</sup><sub>UUU</sub>.<sup>9</sup> The peptide backbone conformation is determined via atomistic simulation of the peptide-hASL<sup>Lys3</sup><sub>UUU</sub> complex and then held fixed with respect to the hASL<sup>Lys3</sup><sub>UUU</sub> conformation throughout the search. The binding free energy for the complex is then evaluated. Subsequently, a random number is generated to determine whether to mutate one amino acid or to exchange two amino acids. If one amino acid is to be mutated, one site along the peptide sequence is chosen randomly. The amino acid at that site is then mutated to another amino acid of the same residue type. The best rotamer for the new amino acid is chosen to substitute for the old amino acid in this mutation step. If an exchange step is chosen, two random sites along the chain and their corresponding amino acids are chosen for a mutual exchange attempt. In this exchange step, we calculate the effective potential of all the possible rotamers and perform the self-consistent mean field (SCMF) procedure<sup>12</sup> described below to obtain the best rotamer combination for the exchanged amino acids. Regardless of whether one amino acid was mutated or two amino acids were exchanged, the new generated peptide sequence is evaluated further by calculating the new binding free energy and accepted or rejected according



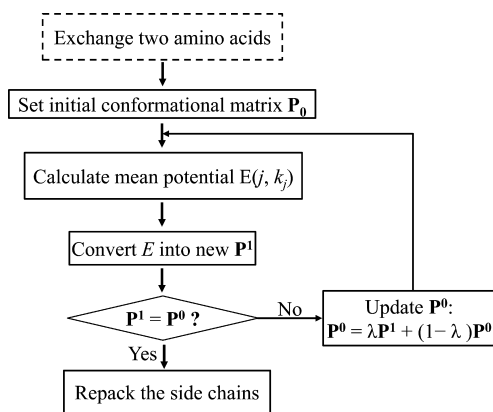
**Figure 2.** Search algorithm flow strategy. An initial peptide sequence is chosen (in this instance peptide P6). Random numbers were generated to determine whether to mutate one amino acid or not ("No" or "Yes"). If yes, then one amino acid from the sequence was randomly changed to an amino acid from the same residue category (Table 1). If no, then two amino acids from the sequence were randomly exchanged regardless of the residue category. The SCMF algorithm was then used to determine the lowest-energy rotamer combination. The MC algorithm was used to accept or reject the newly generated peptide sequence based on the calculation of binding free energy ( $\Delta G^1_{\text{binding}}$ ).

to the Metropolis criterion. After a total of 10,000 evolution steps, the best peptide sequences with the lowest binding free energy are identified.

In our use of the SCMF, a trial exchange between two amino acids at randomly chosen sites is implemented (Figure 3). The conformational probability matrix  $P = P^0$  is set initially for the two amino acids so that all possible rotamers have equal conformational probability. The initial conformational matrix  $P^0$  is then used to calculate the effective potential of each amino acid in each rotamer state. Once the effective potentials for all the rotamer states are known, new conformational probabilities of the rotamers are obtained according to the Boltzmann law so as to constitute a new conformational matrix  $P^1$ . Next, the absolute error between  $P^1$  and  $P^0$  is calculated. If the absolute error is less than  $10^{-3}$ , the best rotamers with the highest conformational probability for the two amino acids are selected from  $P^1$  to repack the side chains. Otherwise, the conformational matrix  $P$  is updated by employing a self-consistent iteration. The updated conformational matrix  $P$  is stored as the old conformational matrix  $P^0$  for the next round evaluation. The conformational matrix is iterated until the absolute error between  $P^1$  and  $P^0$  is less than  $10^{-3}$ . Eventually, the best combination of rotamers is found, thereby repacking the backbone.

The absolute binding energy for a ligand and a receptor is defined to be the difference between the free energy of the complex, and the free energies of the ligand and the receptor prior to binding.<sup>13</sup> It was calculated according to the following equation:

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



**Figure 3.** Self-consistent mean field (SCMF) procedure. A trial exchange between two amino acids is implemented. The conformational probability matrix  $\mathbf{P}^0$  is set initially so that all possible rotamers at any one site have equal probabilities. The effective potential experienced by each rotamer at each site is calculated, and the Boltzmann law is used to determine new conformational probabilities of the rotamers for each amino acid and hence a new conformational probability matrix  $\mathbf{P}^1$ . If the absolute error between  $\mathbf{P}^0$  and  $\mathbf{P}^1$  is less than  $10^{-3}$ , the rotamer combination with the highest conformational probability is selected from  $\mathbf{P}^1$  to repack the side chains. Otherwise, the conformational matrix  $\mathbf{P}$  is updated by employing a self-consistent iteration until the absolute error falls below a certain tolerance.

where  $G_{\text{TOT}}^{\text{complex}}$ ,  $G_{\text{TOT}}^{\text{ligand}}$ , and  $G_{\text{TOT}}^{\text{receptor}}$  represent the total free energies of the complex and the ligand and the receptor in the solution, respectively. The total free energy  $G_{\text{TOT}}$  of the molecular complex in the solution was calculated as follows:

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

where  $U_{\text{INT}}$ ,  $U_{\text{VDW}}$ ,  $U_{\text{ELE}}$ ,  $G_{\text{EGB}}$ , and  $G_{\text{GBSUR}}$  indicate the internal energy (INT), van der Waals energy (VDW), electrostatic energy (ELE), the polar solvation energy (EGB), and the nonpolar solvation energy (GBSUR). The internal energy  $U_{\text{INT}}$  is defined as the potential energy associated with the random, disordered motion of the molecule itself, including the vibration of bonds, bond angles, and the torsion of dihedral angles. The van der Waals energy  $U_{\text{VDW}}$  between two nonbonded atoms adopts a typical 12–6 Lennard–Jones equation. The electrostatic energy  $U_{\text{ELE}}$  between two nonbonded atoms follows the conservative Coulomb's law. The polar solvation energy  $G_{\text{EGB}}$  is calculated based on the generalized Born model, which is an analytical way to evaluate the electrostatic contribution to the solvation free energy. The nonpolar solvation energy  $G_{\text{GBSUR}}$  is approximated by a pairwise potential that occurs at the incompatible interface of two distinct species molecules. In the solution, actually, it is proportional to the solvent-accessible surface area of solute molecules. The expressions for the energies in eq 2 are as follows:

(a) internal energy  $U_{\text{INT}}$

$$\begin{aligned} U_{\text{INT}} &= U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} \\ &= \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 \\ &\quad + \sum_{\text{dihedrals}} (V_n/2)[1 + \cos(n\phi - \delta)] \end{aligned} \quad (3)$$

where  $K_b$  and  $K_\theta$  are the harmonic force constants,  $b_0$  and  $\theta_0$  are the equilibrium bond length and bond angle, respectively,  $b$  and  $\theta$  are the actual bond length and bond angle, respectively,

$V_n$  is an energy constant,  $n$  is the dihedral multiplicity,  $\phi$  is the actual torsion angle, and  $\delta$  is phase shift. All of the values of  $K_b$ ,  $b_0$ ,  $K_\theta$ ,  $\theta_0$ ,  $V_n$ , and  $\delta$  come from the Amber force field.

(b) van der Waals energy  $U_{\text{VDW}}$

$$U_{\text{VDW}} = \sum_i \sum_{j \neq i} \left( \frac{\epsilon_{ij} r_{0ij}^{12}}{r_{ij}^{12}} - \frac{2\epsilon_{ij} r_{0ij}^6}{r_{ij}^6} \right) \quad (4)$$

where  $\epsilon_{ij}$  and  $r_{0ij}$  are the well depth and the characteristic distance for a pair of atoms  $i$  and  $j$ , respectively, and  $r_{ij}$  is the actual distance between two nonbonded atoms  $i$  and  $j$ . Each atom has a depth of the potential well  $\epsilon$ , and a characteristic distance  $r_0$  at which the potential reaches its minimum. For any pair of atoms  $i$  and  $j$ ,  $\epsilon_{ij} = (\epsilon_i \cdot \epsilon_j)^{1/2}$  and  $r_{0ij} = r_{0i} + r_{0j}$ . All of the values of  $\epsilon$  and  $r_0$  come from the Amber force field.

(c) electrostatic energy  $U_{\text{ELE}}$

$$U_{\text{ELE}} = \frac{1}{2} \sum_i \sum_{j \neq i} \frac{q_i q_j}{r_{ij}} \quad (5)$$

where  $q_x$  ( $x = i$  or  $j$ ) represents the charge quantity of atom  $x$ , and  $r_{ij}$  is the actual distance between two nonbonded atoms  $i$  and  $j$ . All of the values of  $q_x$  come from the Amber force field.

(d) polar solvation energy  $G_{\text{EGB}}^{14-17}$

$$\begin{aligned} G_{\text{EGB}} &= -\frac{1}{2} \sum_i \left( 1 - \frac{1}{\epsilon_r} \right) \left[ \frac{q_i^2}{f_{\text{GB}}(r_{ii}, \alpha_i, \alpha_i)} \right] \\ &\quad - \frac{1}{2} \sum_i \sum_{j \neq i} \left( 1 - \frac{1}{\epsilon_r} \right) \left[ \frac{q_i q_j}{f_{\text{GB}}(r_{ij}, \alpha_i, \alpha_j)} \right] \end{aligned} \quad (6)$$

where  $q_x$  ( $x = i$  or  $j$ ) represents the charge quantity of atom  $x$ ,  $\epsilon_r$  is the relative dielectric constant (in this work,  $\epsilon_r = 80.0$  for water),  $r_{ij}$  is the actual distance between atoms  $i$  and  $j$  (thus,  $r_{ii} = 0.00$ ),  $\alpha_x$  ( $x = i$  or  $j$ ) is the effective Born radii of atom  $x$ , and  $f_{\text{GB}}$  is a smoothing function, of which a common form is

$$f_{\text{GB}} = \left[ r_{ij}^2 + \alpha_i \alpha_j \exp \left( \frac{-r_{ij}^2}{4\alpha_i \alpha_j} \right) \right]^{1/2} \quad (7)$$

(e) nonpolar solvation energy  $G_{\text{GBSUR}}$

$$G_{\text{GBSUR}} = \gamma \sum_i \sigma_i + b \quad (8)$$

where  $\gamma$  is surface tension (in this work,  $\gamma = 0.0072$  kcal/mol/Å<sup>2</sup> for the interface between water and biomolecules),  $\sigma_i$  is the solvent-accessible surface area of atom  $i$ , and  $b$  is an offset parameter (in this work,  $b$  is set to 0.00).

**Fluorescein-Labeled Peptides.** Fifteen-amino acid peptides were selected from the sequences predicted *in silico* and were chemically synthesized, each with fluorescein (Flc) at the N-terminus. This set of peptides included the original P1 and P6 sequences to be used as an internal control.<sup>9</sup> The lyophilized peptide set was reconstituted via standard suggestions from the manufacturer (80% DMSO/20% H<sub>2</sub>O, v/v). Concentrated peptide stocks were stored in 25 μL aliquots at −80 °C for later use. Working concentrations were diluted for each experiment and kept on ice or stored at −20 °C.

**Fluorescent Assays.** Fluorescent assays were conducted in phosphate buffer (10 mM Na<sub>2</sub>HPO<sub>4</sub> and 10 mM KH<sub>2</sub>PO<sub>4</sub>, pH 6.8) in low volume 384 well plates. All buffers, peptides, and RNA were pipetted into wells via a liquid handling robot



(Janus, PerkinElmer). All plates were read using a plate reader fitted with fluorescein-specific filters (PerkinElmer EnVision) which were optimized for each plate and peptide before each experiment. Initial validation screens were conducted by obtaining fluorescent signals for each peptide (0.50  $\mu$ M) alone ( $FS_0$ ) and in the presence of a 2-fold excess of modified or unmodified hASL<sup>Lys3</sup><sub>UUU</sub> ( $FS_1$ ). Percent change in the fluorescent signal was calculated (% change =  $100 \cdot (FS_1/FS_0)$ ). A decrease (quench) in fluorescent signal in the presence of RNA indicated a binding event between the peptide and ASL. In control wells, H<sub>2</sub>O, phosphate buffer, and ASLs were individually tested for any inherent fluorescent signal. Water and buffer signals served as blank background signals. The hASL<sup>Lys3</sup><sub>UUU</sub> had a negligible inherent signal, which was taken into account when calculating the overall signal quench. All controls and experimental sets were repeated in triplicate within a single plate, and signals are an average of each triplicate. For the subset of peptides which were further studied to obtain binding constants, the fluorescent binding assay was completed as described above. The peptide fluorescent signal was monitored throughout an increasing ASL concentration (0–3  $\mu$ M). The percent quench in signal was plotted against the hASL concentration. Binding (dissociation) constants ( $K_d$ ) were calculated using the single-site, nonlinear regression function within SigmaPlot. Because the peptide concentration was 0.50  $\mu$ M,  $K_d$  values between 0.05 and 0.13  $\mu$ M are considered approximate ( $\approx$ ). Experiments were performed in triplicate within a single plate, and fluorescent signals are an average of each triplicate. The binding of peptides P6, P27, P31, and P35\* to the unmodified hASL<sup>Lys3</sup><sub>UUU</sub> resulted in little change in fluorescence and could not be satisfactorily fitted with a single-site, nonlinear regression function. Thus, these  $K_d$ s were considered indeterminable. Results could not be curve fitted, and a  $K_d$  could not be calculated from the data (Table 4).

## RESULTS

**Development of an Algorithm That Optimizes a Peptide Sequence for Binding RNA.** Previous studies demonstrated the feasibility of selecting peptides with modification-dependent recognition of tRNAs' anticodon stem and loop domains, ASLs.<sup>18,19</sup> The peptides were selected from completely and partially randomized phage display libraries.<sup>9,10</sup> However, optimizing 15- and 16-amino acid peptide sequences using this approach is not feasible since there are over  $3.3 \times 10^{23}$  possible sequences. Because of the exorbitantly high costs of creating and screening millions of peptides even with the benefit of phage display, we turned to computer algorithms<sup>20</sup> and assisted model building with energy refinement, AMBER, simulations to pare down the number of possibilities before performing *in vitro* assays. We developed a novel optimization strategy<sup>21</sup> that combines MC with SCMF to evolve amino acid sequences. The peptide P6 sequence RVTHHAFLGAHRTVG<sup>9</sup> was the starting point from which an optimized peptide was sought to bind the modified hASL<sup>Lys3</sup><sub>UUU</sub> with the highest specificity and affinity.

The canonical 20 amino acids were categorized into six distinct groups according to hydrophobicity, polarity, size, and charge (Table 1A). These hydration properties were necessary to ensure the peptide did not become too hydrophobic (and thus insoluble) or so hydrophilic that binding to hASL<sup>Lys3</sup> was inhibited. The overall charge of the peptide was chosen to be slightly positive to ensure interaction with the negatively charged ASL. By adjusting the number of amino acids in each

Table 1

| (A) residue categories of 20 canonical amino acids |             | (B) hydration properties of three optimization cases |        |        |        |
|--|-------------|--|--------|--------|--------|
| amino acids  | hydration   | notation   | case 1 | case 2 | case 3 |
| Leu, Val, Ile, Met, Phe, Tyr, Trp                  | hydrophobic | Npho   | 4      | 5      | 3      |
| Glu, Asp   | negatively  | Nneg   | 0      | 0      | 0      |
| Arg, Lys   | positively  | Npos   | 2      | 2      | 1      |
| Ser, Thr, Asn, Gln, His                            | hydrophilic | Npol   | 5      | 6      | 6      |
| Ala, Cys   | other       | Noth   | 2      | 1      | 3      |
| Gly  | neutral     | Ngly   | 2      | 1      | 2      |

category, via  $N_{pho}$ ,  $N_{neg}$ ,  $N_{pos}$ ,  $N_{pol}$ ,  $N_{oth}$ ,  $N_{gly}$ , we maintained hydration properties similar to those of the original P6 sequence while evolving the sequences.<sup>9</sup>

The peptide sequence was optimized using the following computational procedure. The stable structure for the complex between the original P6 sequence and ASL was determined using AMBER. The structure of the fully modified hASL<sup>Lys3</sup><sub>UUU</sub> was taken from the high resolution solution structure,<sup>6</sup> providing a restrained structure to which the peptide would bind *in silico*. Once the stable structure of the peptide P6 with the ASL<sup>Lys3</sup> was determined, the peptide's amino acid sequence was evolved and optimized while keeping the backbone fixed. Each peptide sequence evolved in two types of "moves": (1) a single randomly chosen amino acid in the peptide sequence was mutated to a different amino acid from the same residue category (Table 1); or (2) two randomly chosen amino acids in the peptide sequence were exchanged regardless of the their residue category (Figure 2) using SCMF.<sup>22</sup> SCMF finds the optimal rotamer combination with the lowest binding energy for the two exchanged amino acid residues (Figure 3) based on the preferred, distinct side chain conformations in Lovell's rotamer Library.<sup>23</sup> The sequences were subjected to continued rounds of optimization (Figure 2). By comparing the changes in binding energy before and after each of the two types of moves, the peptide sequence was evolved to those with the lowest binding energies and thus increased binding affinity to the modified hASL<sup>Lys3</sup><sub>UUU</sub>.

### Evolved Peptides Have Increased Specificity for ASL Modifications.

The initial P6 sequence was subjected to an evolution of over several hundred thousands of rounds of 15-amino acid peptide sequences that, based on binding energies, should recognize and bind modified hASL<sup>Lys3</sup><sub>UUU</sub> with a similar or higher affinity than that of P6. Initial results from the *in silico* selection suggested two optimized peptide sequences, P26, R-T-L-H-H-A-L-F-G-A-H-Q-T-V-G and P27, R-W-Q-M-T-A-F-A-H-G-W-R-H-S-G. These sequences exhibited binding energies to the hASL<sup>Lys3</sup><sub>UUU</sub> lower than that of P6 (P26, −22.55 kcal/mol, and P27, −22.07 kcal/mol, respectively, vs P6 −21.26 kcal/mol). On the basis of these initial results, we developed three distinct peptide sequence cases. The three cases varied within the six residue categories (Table 1B). Although different, each of the three cases is still within the overall desired levels of moderate hydration and charge properties (slightly positively charged). P6, the initially evolved sequences P26 and P27, and three of the top 10 sequences from each of the first two cases 1 and 2 have lower binding energies than those of case 3 (Table 2). This is likely due to the increased allowance in cases 1 and 2 for positively charged and hydrophobic residues (Table 2). The binding energies calculated for P26 and P27 and cases 1 and 2, but not 3, are on par with or lower than the binding energy for

**Table 2. Binding Energies for Selected, Optimized Peptide Sequences**

| Peptide Designation and Rank | Peptide Sequence | Binding Energy (kcal/mol) <sup>a</sup> |
|------------------------------|------------------|--|
| <sup>b</sup> P6              | RVTHHAFLGAHRTVG  | -21.26                                 |
| P26                          | RTLHHALFGAHQTVG  | -22.55                                 |
| P27                          | RWQMTAFAHGWRHSG  | -22.07                                 |
| Case 1                       |                  |  |
| P28                          | RWNHCQFWNGWRAQG  | -22.81                                 |
| P35 (P35*)                   | RWNHCQFWNGWRANG  | -22.78                                 |
| P29                          | RWNHQSFWHGWRACG  | -22.64                                 |
| Case 2                       |                  |  |
| P30                          | RWNHSQFWSLWRAHG  | -22.71                                 |
| P31                          | RWQHHSFHPLWRMSG  | -21.86                                 |
| A                            | RWHHHHFSPLWRWHG  | -21.56                                 |
| Case 3                       |                  |  |
| B                            | RHHHHHFGPPWLNCG  | -14.58                                 |
| P32                          | RHHHASFSGPPWLSHG | -14.26                                 |
| P33                          | RHSHAHFGPPWLSHG  | -13.94                                 |

<sup>a</sup>Using the search algorithm, peptide sequence candidates are evolved on the basis of a backbone scaffold of structure. For each binding structure, an explicit binding energy is generated. This search and selection process does not quantify the error in the binding energy. <sup>b</sup>Peptides are numbered in accordance with the previous report of phage display selected sequences.<sup>9</sup> P6 is from the original selection, and P26–P38 were chosen based on predictions *in silico*. P35\* with the sequence RWNHCQFWNGWRANG has a single amino acid serine change from P35. Peptide sequences A and B from cases 2 and 3, respectively, were not selected for chemical synthesis and analysis.

P6. This suggests a potential increase in their binding affinity for modified hASL<sup>Lys3</sup><sub>UUU</sub> versus that of P6.

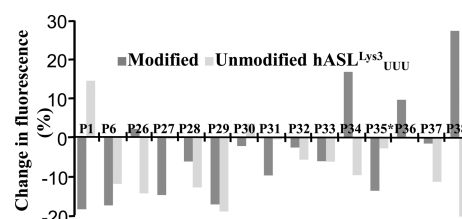
**In Vitro Fluorescence Screens Validate *in Silico* Methods: Qualitative Analysis.** Sequences predicted during the *in silico* optimizations to have the lowest binding energies and thus potentially higher affinity for modified hASL<sup>Lys3</sup> were selected for validation with a fluorescence assay.<sup>9</sup> Fifteen peptides (Table 3) were chemically synthesized with fluorescein at the N-terminus to allow for very sensitive, low volume detection of peptide–RNA binding interactions. P1 and

**Table 3. Synthesized Peptides Based on Predictions *in Silico*<sup>a</sup>**

| Peptide     | Sequence               |
|-------------|------------------------|
| <b>P1</b>   | <b>FSVSFPSPLPAPDRS</b> |
| <b>P6</b>   | <b>RVTHHAFLGAHRTVG</b> |
| P26         | RTLHHALFGAHQTVG        |
| <b>P27</b>  | <b>RWQMTAFAHGWRHSG</b> |
| P28         | RWNHCQFWNGWRAQG        |
| P29         | RWNHQSFWHGWRACG        |
| P30         | RWNHSQFWSLWRAHG        |
| <b>P31</b>  | <b>RWQHHSFHPLWRMSG</b> |
| P32         | RHHHASFSGPPWLSHG       |
| P33         | RHSHAHFGPPWLSHG        |
| P34         | RFQHSNWFSGWKVNG        |
| <b>P35*</b> | <b>RWNHCQFWNGWRANG</b> |
| P36         | RWNGSQWFCAWRANG        |
| P37         | RHTHCAFWGAHRTVG        |
| P38         | RWTHCQFWQGFVRNG        |

<sup>a</sup>Peptides named following the peptides from original phage display library screens.<sup>9</sup> P1 and P6 (bold font) are from the original screen. P6 has been characterized.<sup>9</sup> In addition, the binding of the modified and unmodified hASL<sup>Lys3</sup><sub>UUU</sub> by P1, P27, P31, and P35\* (bold font and shaded row) are characterized in this article.

P6 from phage display selections,<sup>9</sup> the initially evolved sequences, P26 and P27, and the best binders from each of the cases 1, 2, and 3 were synthesized. Variants of these sequences that had one or two amino acid changes were also synthesized (Table 3). During the initial validation assay, changes in the amount of fluorescence were monitored to determine whether the peptide was binding to the modified and/or unmodified hASL<sup>Lys3</sup><sub>UUU</sub> and to what relative degree (Figure 4). In this screening assay, P6 behaved as expected.



**Figure 4.** Fluorescence of chemically synthesized peptides effected by modified and unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. An initial fluorescent signal (FS<sub>0</sub>) of peptide alone (1.5 μM) was obtained. Then, a 2-fold excess of ASL was added to each peptide, and the fluorescent signal (FS<sub>1</sub>) was monitored. The percent change (100·(FS<sub>1</sub>/FS<sub>0</sub>)) is graphed for each of the assayed peptides. Dark gray bars represent the percent change in fluorescence in the presence of the modified hASL<sup>Lys3</sup><sub>UUU</sub>, and light gray bars represent the percent change in the presence of the unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. Sequences for P1–P38 are presented in Table 3

When bound by the modified hASL<sup>Lys3</sup><sub>UUU</sub>, P6 fluorescence was altered more than that when bound by the unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. Other peptides (P27, P31, and P35\*) behaved similarly. Still others demonstrated nonspecific binding in that the degree of fluorescence did not significantly differ between the binding of modified and unmodified hASL<sup>Lys3</sup><sub>UUU</sub> (Figure 4: P30 and P32). Three peptides exhibited a greater degree of change in fluorescence quenching when binding the unmodified RNA versus the modified RNA; their binding could have been negatively affected by modifications (P26, P28, and P37). The stacking interactions of fluorescein with nucleobases can be nonspecific and result in fluorescence quenching that is mistakenly attributed to a selective binding of a ligand to a nucleic acid. We have employed a direct comparison of peptide binding of the modified hASL<sup>Lys3</sup><sub>UUU</sub> to binding of the control that is identical in sequence except for the modified nucleosides in the anticodon loop. If nonspecific stacking interactions were to occur, they were most likely limited to the stacked and identical stem regions of the control and experimental RNAs. Therefore, the observable differences in quenching could be attributed to the peptide ligand's affinity for the modifications.

Three peptides of the 15 exhibited fluorescence enhancement when binding the modified RNA but exhibited quenching when binding the unmodified RNA (Figure 4: P34, P36, and P38), whereas a fourth exhibited the opposite (P1). This result is counter to the better understood binding of P6.<sup>9</sup> The result is also counterintuitive. Fluorescein stacks with the nucleobases of nucleic acids resulting in a reduced fluorescence. The quenching of fluorescein fluorescence is one of its advantageous properties in studying the interactions of fluorescein-coupled ligands with nucleic acids. Without a detailed investigation of the mode of binding by these four peptides, interpretation of these results is difficult at this time.

Three peptides, P27 one of the two initially evolved, P31 from case 1, and P35\*, a variant of P35 from case 2,

preferentially bound the modified hASL<sup>Lys3</sup><sub>UUU</sub> as compared to their binding of the unmodified hASL<sup>Lys3</sup><sub>UUU</sub>, as did the phage selected P6. P27, P31, and P35\* showed a very high selectivity for the modified hASL<sup>Lys3</sup><sub>UUU</sub>. When these peptides were tested for their binding to unmodified hASL<sup>Lys3</sup><sub>UUU</sub>, little or no change in fluorescence was observed (Figure 4; Table 4). However,

**Table 4. Affinity of Peptides for hASL<sup>Lys3</sup><sub>UUU</sub><sup>a</sup>**

| peptide | modified or unmodified hASL <sup>Lys3</sup> <sub>UUU</sub> | K <sub>d</sub> (μM)       |
|---------|--|---------------------------|
| P6      | modified   | 0.50 ± 0.10 <sup>b</sup>  |
|         | unmodified   | ID                        |
| P27     | modified   | ~0.05 ± 0.02 <sup>c</sup> |
|         | unmodified   | ID                        |
| P31     | modified   | 0.58 ± 0.24               |
|         | unmodified   | ID                        |
| P35*    | modified   | 1.87 ± 1.00               |
|         | unmodified   | ID                        |

<sup>a</sup>ID = indeterminable; could not be curve fitted and calculated from the data because of little change in fluorescence with binding of ASL. <sup>b</sup>K<sub>d</sub> value from previous studies.<sup>9</sup> <sup>c</sup>K<sub>d</sub> value is approximate for it is one-tenth the concentration of the fluorescent peptide reporter.

with the addition of the modified hASL<sup>Lys3</sup><sub>UUU</sub>, the fluorescent signals of P27, P31, P35\*, and P6 were quenched. Peptides P32 and P33 had the best binding properties among the case 3 family of peptides. Even so, they had significantly weaker affinities and a lack of specificity for the modified hASL<sup>Lys3</sup><sub>UUU</sub> as compared to those of P27, P31, and P35\* (Figure 4). Thus, peptides P27, P31, and P35\* evolved *in silico* from P6 as having high affinities for the modified hASL<sup>Lys3</sup><sub>UUU</sub> also appear from the screening assay *in vitro* to have higher affinities and higher specificities for the modified RNA.

**Evolved Peptides with Lower Free Energies of Binding Also Have Higher Affinity for Modified hASL<sup>Lys3</sup><sub>UUU</sub>: Quantitative Analysis.** Five of the 15 peptides were chosen for detailed analysis of their binding because the data collected and presented in the initial screen (Figure 4) was indicative of these peptides distinguishing modified from unmodified RNA. Also, these peptides were chosen for detailed analyses because in their binding of the modified hASL<sup>Lys3</sup><sub>UUU</sub>, they exhibited sizable changes in fluorescence that could be accurately quantified (Figure 4 and Table 4). The peptide fluorescence signal was monitored with an increasing hASL concentration, and the percent quench in signal was plotted against the hASL concentration. Binding constants (K<sub>d</sub>) were calculated using the single binding site, nonlinear regression function. Peptides P6, P27, P31, and P35\* having exhibited qualitatively the highest affinity and specificity for the modified substrate (Figure 4) were subjected to a quantitative analysis of their binding to both the modified and the unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. P1 was also subjected to this analysis though the peptide exhibited a decreased fluorescence when binding the modified hASL<sup>Lys3</sup><sub>UUU</sub> and an increased fluorescence in binding the unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. Equilibrium binding constants (as the dissociation constant K<sub>d</sub>; Table 4) were determined from the concentration-dependent fluorescence quenching with the binding of the modified and unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. This data was compared to that for P6 from which these peptides were evolved *in silico* (Table 4). Little to no fluorescence quenching was observed when P27, P31, P35\*, and P6 were titrated with the unmodified hASL<sup>Lys3</sup><sub>UUU</sub>; thus, binding constants could not be extracted from the data (Table

4). In contrast, the fluorescence of P27 was quenched considerably with increasing amounts of the modified hASL<sup>Lys3</sup><sub>UUU</sub>. The peptide bound the RNA with high affinity (K<sub>d</sub> ≈ 0.05 ± 0.10 μM) (Table 4). P27 exhibited the highest affinity for the modified hASL<sup>Lys3</sup><sub>UUU</sub>. P31 bound the modified hASL<sup>Lys3</sup><sub>UUU</sub> with a 10-fold less affinity (K<sub>d</sub> = 0.50 ± 0.10 μM), and its affinity for the hASL was equivalent to that of P6 (Table 4). P35\* bound the modified hASL<sup>Lys3</sup><sub>UUU</sub> with the lowest affinity of these peptides (K<sub>d</sub> = 1.87 ± 1.00 μM). The fluorescence of P1 was altered by both the modified and unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. In binding the modified hASL<sup>Lys3</sup><sub>UUU</sub>, P1 exhibited a quench in fluorescence from which we extracted a modest binding constant (K<sub>d</sub> ≈ 0.13 ± 0.02 μM). However, the fluorescence change in binding the unmodified hASL<sup>Lys3</sup><sub>UUU</sub> resulted in an equivalent affinity (K<sub>d</sub> ≈ 0.15 ± 0.04 μM). Thus, in contrast to P27, P31, P35\*, and P6, P1 lacked specificity for the modified hASL<sup>Lys3</sup><sub>UUU</sub>.

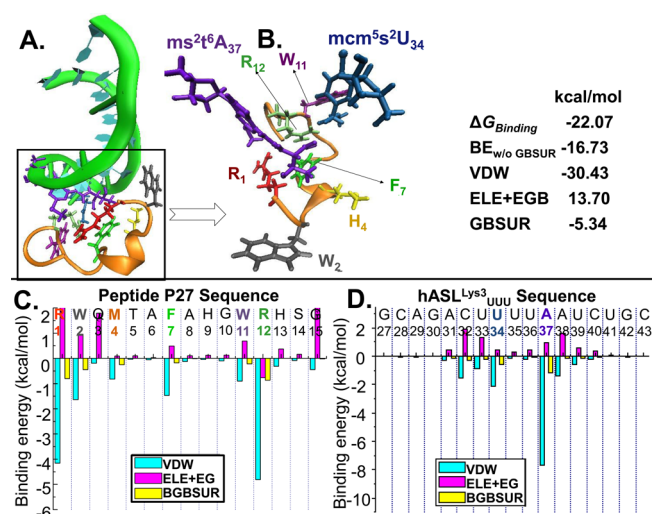
**Computational Analysis of Peptide P27 Binding of the Modified hASL<sup>Lys3</sup><sub>UUU</sub>.** Peptide P27 has the highest affinity coupled with highest specificity for binding of the modified hASL<sup>Lys3</sup><sub>UUU</sub>. The calculated structure bound to the modified hASL<sup>Lys3</sup><sub>UUU</sub> at equilibrium reveals how this affinity and specificity could be achieved and thus suggests future experimentation (Figure 5A). Amino acids throughout P27 are engaged with the extensive chemistries of the two modifications unique to the tRNA<sup>Lys3</sup><sub>UUU</sub>. The two arginines, R<sub>1</sub> and R<sub>12</sub>, bracket the threonylcarbamoyl-group of ms<sup>t</sup><sup>6</sup>A<sub>37</sub> (Figure 5B). At the middle of the peptide, F<sub>7</sub> is closely associated with the hydrophobic methyl of the threonyl-side chain. The imidazole ring of W<sub>11</sub> lies above the methyl-ester of the 5-methoxycarbonylmethyl-moiety of mcm<sup>5</sup>s<sup>2</sup>U<sub>34</sub>. As evidenced by calculations for each of the 15 amino acids, R<sub>1</sub>, R<sub>12</sub>, F<sub>7</sub>, and W<sub>11</sub> contribute to ΔG<sub>Binding</sub> (Figure 5C). The binding energy contributed by each of the nucleosides of the modified hASL<sup>Lys3</sup><sub>UUU</sub> have also been calculated. The binding energy is concentrated in the anticodon loop, as opposed to the stem. However, the two modifications, particularly ms<sup>t</sup><sup>6</sup>A<sub>37</sub>, provide the most significant binding energies.

## DISCUSSION

**An Amino Acid Signature Sequence Important for the Recognition of Uniquely Modified RNA.** Protein recognition of RNA has been well-defined for two families of RNA binding proteins. The Pumilio protein was first identified as a *Drosophila* protein critical for segmentation of the posterior,<sup>24</sup> and the FBF protein was found responsible in *Caenorhabditis* to mediate the sperm/oocyte switch.<sup>25</sup> Other RNA binding domains or RNA recognition motifs have been studied extensively, revealing conserved sequence/structure relationships, yet these structures differ by having subtle distinctions, some of which are dynamic.<sup>26</sup> However, insight into protein recognition of RNA's modified nucleoside chemistries has not been pursued and thus has yet to be characterized.

The primary goal of this study was to demonstrate that a signature amino acid sequence can be identified as binding a uniquely modified RNA with high affinity and specificity. We reached this signature sequence using a combination of computational simulations to obtain optimized amino acid sequences that were then confirmed by binding studies *in vitro*. By comparing peptide sequences which specifically bound the modified hASL<sup>Lys3</sup><sub>UUU</sub> to those which did not, we were able to derive an amino acid signature that should be useful for protein/peptide recognition of RNA with modifications.





**Figure 5.** Peptide P27 binds the modified hASL<sup>Lys3</sup><sub>UUU</sub> with high affinity and specificity. (A) The computed equilibrium binding structure of the modified hASL<sup>Lys3</sup><sub>UUU</sub> bound by P27. The peptide backbone is in gold, and the ribose-phosphodiester backbone of the hASL<sup>Lys3</sup><sub>UUU</sub> is colored in green. (B) Enlargement of the interaction demonstrating the specificity achieved in the binding of the two modifications by the amino acids R<sub>1</sub> (red), F<sub>7</sub> (light green), W<sub>11</sub> (light purple), and R<sub>12</sub> (dark green). The peptide backbone is in gold and the side chains in color. The modifications ms<sup>2</sup>t<sup>6</sup>A<sub>37</sub> (purple) and mcm<sup>5</sup>s<sup>2</sup>U<sub>34</sub> (blue) are bound by amino acids at the beginning, middle, and end of the peptide. The ribose-phosphodiester backbone of the hASL<sup>Lys3</sup><sub>UUU</sub> is not shown. The table characterizes the contributions of different binding modes:  $\Delta G_{\text{Binding}}$ , Gibbs free energy of binding;  $BE_{w/o \text{ GBSUR}}$ , binding energy without GBSUR; VDW, van der Waals energy; ELE, electrostatic energy; EGB, polar solvation energy based on the generalized Born (implicit solvent) model; and GBSUR, nonpolar solvation energy, which is the product of the solvent-accessible surface area of the solute molecules and the interfacial tension between the solute and solvent. (C) Individual contributions of each amino acid to the VDW, ELE + EGB, and GBSUR. The amino acids are colored as in B. (D) Individual contributions of each nucleoside to the VDW, ELE + EGB, and GBSUR. The nucleosides engaged in the interaction with P27 are those of the anticodon loop, particularly the modified nucleosides at U<sub>34</sub> and A<sub>37</sub>. The modified nucleosides are colored as in B.

Focusing primarily on those peptides which showed the highest affinity and specificity for the modified hASL<sup>Lys3</sup><sub>UUU</sub>, the amino acid signature emerged: R-W-Q/N-H-X<sub>2</sub>-F-Pho-X-G/A-W-R-X<sub>2</sub>-G (where X can be most amino acids, and Pho is hydrophobic) (Table 5).

The evolution of peptide sequences *in silico* is rapid relative to screening at the bench. We have developed an algorithm with the potential to simulate a very high number of binding

events for the 15 amino acid peptide theoretical combinations ( $>3.3 \times 10^{23}$ ) to each substrate. In our algorithm, all 20 amino acids are considered. However, we group them for the purpose of describing their hydration properties. There are concessions such as grouping the amino acids by side chain properties to more quickly move through peptide evolution. Our developed algorithm proved to be a powerful tool in accurately predicting peptides which would bind specifically to hASL<sup>Lys3</sup><sub>UUU</sub> modifications. We believe that we can improve the accuracy of *in silico* predictions by developing simulations in tandem to look more closely at nonspecific binding of the peptide to other small RNAs and/or unmodified tRNAs or ASLs. A cross-check performed by a parallel screen assessing binding energies of peptides binding to different ASLs could potentially eliminate nearly all false positives before moving to *in vitro* and/or *in vivo* experiments. The validation screens *in vitro* revealed that while the computer algorithms were not 100% correct in predicting peptide sequences with both high affinity and specificity however, the selection *in silico* was a serious tool for predicting binding trends and quickly screening through many peptide sequence combinations.

**Amino Acid Signature Sequence R-W-Q/N-H-X<sub>2</sub>-F-Pho-X-G/A-W-R-X<sub>2</sub>-G.** The signature amino acid sequence offers an opportunity to investigate the mode or mechanism of binding in future research. The sequence offers clues and surprises to be studied as to why the optimized peptides from cases 1 and 2 bind the modified hASL<sup>Lys3</sup> with high affinity. Interestingly, the 5'-amino terminal sequence is more hydrophilic (R, Q, and H) than the center (F and Pho) or the 3'-carboxyl terminus (G). Conventional thought would have the two positively charged arginine residues (positions 1 and 12) preferentially engaged with the negatively charged phosphate linkages via charge–charge interactions and/or the hydrophilic sugars. Here, the two arginine residues are also involved in interactions with the mcm<sup>5</sup>s<sup>2</sup>U<sub>34</sub> and ms<sup>2</sup>t<sup>6</sup>A<sub>37</sub> due to VDW energy (Figure 5B). The increased number of hydrophobic residues, specifically tryptophan (position 11) and phenylalanine (position 7), contribute to the overall binding specificity through VDW interactions.

The fluorescence changes observed during the binding of the peptides to the hASL<sup>Lys3</sup><sub>UUU</sub> are to a lesser degree than expected for fluorescein stacking or intercalating with nucleobases of a duplex stem. Our results are not consistent with fluorescein labeled peptides recognizing the identical stem regions of the modified and unmodified hASL<sup>Lys3</sup><sub>UUU</sub>. However, the results are consistent with the modest effect on fluorescence that could be expected of a differential interaction of peptides P6, P27, P31, and P35\* with the single stranded nucleosides of the modified hASL<sup>Lys3</sup><sub>UUU</sub> loop. For instance, one could expect that the phenyl-ring of phenylalanine would intercalate within

**Table 5. Amino Acid Signature Derived from Optimized Peptide Sequences**

| peptide                | 1 | 2 | 3   | 4 | 5 | 6 | 7 | 8   | 9 | 10  | 11 | 12 | 13 | 14 | 15 |
|------------------------|---|---|-----|---|---|---|---|-----|---|-----|----|----|----|----|----|
| P1                     | F | S | V   | S | F | P | S | L   | P | A   | P  | P  | D  | R  | S  |
| P6                     | R | V | T   | H | H | A | F | L   | G | A   | H  | R  | T  | V  | G  |
| P27                    | R | W | Q   | M | T | A | F | A   | H | G   | W  | R  | H  | S  | G  |
| P29                    | R | W | N   | H | Q | S | F | W   | H | G   | W  | R  | A  | C  | G  |
| P31                    | R | W | Q   | H | H | S | F | H   | P | L   | W  | R  | M  | S  | G  |
| P35                    | R | W | N   | H | C | Q | F | W   | S | G   | W  | R  | A  | N  | G  |
| signature <sup>a</sup> | R | W | Q/N | H | X | X | F | Pho | X | G/A | W  | R  | X  | X  | G  |

<sup>a</sup>X = any amino acid; Pho = hydrophobic amino acid. Position 3 is either Q or N; position 10 is either G or A.

the 3'-base stack of the anticodon domain. The  $N^6$ -threonylcarbamoyl-group of  $ms^2t^6A_{37}$  is known to enhance base stacking.<sup>6</sup> Phenylalanine has been observed to intercalate between anticodon nucleosides of  $tRNA^{Lys}$  in the cocrystal structure of lysyl-tRNA synthetase and  $tRNA^{Lys}$ .<sup>27</sup> Instead of the expected intercalation, in the computed structure  $F_7$  interacts with the threonyl-side chain contributing to the affinity and specificity of the peptide (Figure 5B). Though the signature sequence and the selected peptide sequences P27 and P31 have the highest affinity and specificity for the modified  $hASL^{Lys3}_{UUU}$  and have two arginines each, there is little sequence homology with RNA binding proteins that are rich in arginine<sup>28–32</sup> or with single-stranded RNA binding proteins.<sup>33–37</sup>

**Potential for Modification-Dependent, RNA Binding Peptides.** The optimization of RNA binding peptides to recognize the unique chemistries of modified nucleosides and the contributions they make to local structure affords the opportunity of inhibiting RNA binding proteins studied *in vitro* and possibly *in vivo*. The benefits of modification-dependent signature peptides are many-fold. First, an amino acid signature peptide that uniquely recognizes a specific RNA modification or combination of modifications becomes a tool in the study of RNA binding proteins that interact with RNA in a modification-dependent manner. Modifications are most often found in the terminal and internal loops of RNA structures.<sup>38</sup> There the modifications negate intraloop hydrogen bonding and can enhance or even decrease the possibility of base stacking.<sup>39</sup> Peptides that recognize the ubiquitous anticodon domain modification  $N^6$ -threonylcarbamoyladenine can be used as a tool to study other modified tRNA–protein interactions, for instance, those between tRNAs and their modification enzymes and/or aminoacyl-tRNA synthetases.

Second, the amino acid signature sequence has the potential to shed light on the residues necessary in protein recognition of modified RNA and hence aid in future studies of modified RNA–protein interactions. Indeed, pBLAST searches for this signature find hits in other RNA-binding proteins such as RNA polymerase sigma factor RpoE and threonyl- and prolyl-tRNA synthetases. The synthetases recognize tRNA substrates having  $t^6A_{37}$ . Whether these amino acids confer specific recognition to substrate modifications has yet to be determined. There are many instances whereby a particular tRNA modification is known to exist, its presence determined by mass spectrometry, 2-dimensional thin layer chromatography, or other appropriate method (RNA MDB), with very few clues about which enzyme(s) are responsible for the modification.

Third, peptides directed against specific modifications could also help in identifying incompletely modified RNAs. Complex modifications typically occur in an organized stepwise fashion.<sup>40</sup> Even the earliest analyses demonstrated that many tRNA modification enzymes often recognize a partially modified RNA.<sup>41,42</sup>

Fourth, knowledge of signature peptide sequences for recognition of specific modified nucleosides could provide clues as to how best to narrow bioinformatic searches for modification enzyme functions. Fifth, optimized peptides can be used as competitive inhibitors of protein–RNA interactions that are related to human disease, symptoms, or causative factors. The enhanced binding activity of the peptides reported here could inhibit the recruitment of  $htRNA^{Lys3}_{UUU}$  and its annealing to the HIV genome, thus inhibiting HIV replication. Sixth, the amino acid signature sequence or the methodology

reported here also could be used to derive a signature in the engineering of new proteins and the design of peptide based therapeutics.<sup>43</sup>

## AUTHOR INFORMATION

### Corresponding Author

\*The RNA Institute University at Albany-SUNY, Life Sciences Research Building, 2033I, 1400 Washington Avenue, Albany, NY 12222. Phone: 518-437-4448. Fax: 518-437-4456. E-mail: pagris@albany.edu.

### Present Address

†(J.L.S.) Bio-Cat Microbials, 689 Canterbury Road, Shakopee, MN 55379.

### Funding

Financial support for this work was provided by the North Carolina Biotechnology Center to P.F.A. and C.K.H. (2008MRG1102) and the National Institutes of Health to P.F.A. (2R01-GM23037-25). It was also supported in part by NSF's Research Triangle MRSEC (DMR-1121107).

### Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We thank the Pittsburgh Supercomputing Center (PSC), the National Institute for Computational Science (NICS), and Texas Advanced Computing Center (TACC) for providing us with computing time.

## ABBREVIATIONS

AMBER, assisted model building with energy refinement; NCp7, HIV-1 nucleocapsid protein 7;  $hASL^{Lys3}_{UUU}$ , anticodon stem and loop of human  $tRNA^{Lys3}$  with anticodon UUU;  $mcm^5s^2U_{34}$ , 5-methoxycarbonylmethyl-2-thiouridine at tRNA position 34;  $ms^2t^6A_{37}$ , 2-methylthio- $N^6$ -threonylcarbamoyladenine at tRNA position 37;  $\Psi_{39}$ , pseudouridine at position 39; SCMF, self-consistent mean field; MC, Monte Carlo;  $\Delta G_{Binding}$ , Gibbs free energy of binding;  $BE_{w/o\ GBSUR}$ , binding energy without GBSUR; VDW, van der Waals energy; ELE, electrostatic energy; EGB, polar solvation energy based on the generalized Born (implicit solvent) model; GBSUR, nonpolar solvation energy, which is the product of the solvent-accessible surface area of the solute molecules and the interfacial tension between the solute and solvent

## REFERENCES

- (1) Ratner, L., Haseltine, W., Patarca, R., Livak, K. L., Starcich, B., Josephs, S., Doran, D. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Ivanoff, L., Petteway, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghayeb, J., Chang, N. T., Gallo, R. C., and Wong-Staal, A. F. (1985) Complete nucleotide sequence of the AIDS virus, HTLV-III. *Nature* 313, 277–284.
- (2) Barat, C., Lullien, V., Schatz, O., Keith, G., and Darlix, J. L. (1989) HIV-1 reverse transcriptase specifically interacts with the anticodon domain of its cognate primer tRNA. *EMBO J.* 8, 3279–3285.
- (3) Fassati, A. (2012) Multiple roles of the capsid protein in the early steps of HIV-1 infection. *Virus Res.* 170, 15–24.
- (4) Levin, J. G., Mitra, M., Mascarenhas, A., and Musier-Forsyth, K. (2010) Role of HIV-1 nucleocapsid protein in HIV-1 reverse transcription. *RNA Biol.* 7, 754–774.
- (5) Chan, B., Wiedemaier, K., Yip, W.-T., Barbara, P. F., and Musier-Forsyth, K. (1999) Intra-tRNA distance measurements for nucleocap-



sid protein-dependent tRNA unwinding during priming of HIV reverse transcription. *Proc. Natl. Acad. Sci. U.S.A.* 96, 459–464.

(6) Vendeix, F. A., Murphy, F. V., 4th, Cantara, W. A., Leszczynska, G., Gustilo, E. M., Sproat, B., Malkiewicz, A., and Agris, P. F. (2012) Human tRNA<sup>Lys3</sup><sub>UUU</sub> is pre-structured by natural modifications for cognate and wobble codon binding through keto-enol tautomerism. *J. Mol. Biol.* 416, 467–485.

(7) Stuart, J. W., Gdaniec, Z., Guenther, R., Marszalek, M., Sochacka, E., Malkiewicz, A., and Agris, P. F. (2000) Functional anticodon architecture of human tRNA<sup>Lys3</sup> includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t<sup>6</sup>A. *Biochemistry* 39, 13396–13404.

(8) Agris, P. F., Vendeix, F. A. P., and Graham, W. D. (2007) tRNA's wobble decoding of the genome: 40 years of modification. *J. Mol. Biol.* 366, 1–13.

(9) Graham, W. D., Barley-Maloney, L., Stark, C. J., Kaur, A., Stolarchuk, C., Sproat, B., Leszczynska, G., Malkiewicz, A., Safwat, N., Mucha, P., Guenther, R., and Agris, P. F. (2011) Functional recognition of the modified human tRNA<sup>Lys3</sup><sub>UUU</sub> anticodon domain by HIV's nucleocapsid protein and a peptide mimic. *J. Mol. Biol.* 410, 698–715.

(10) Eshete, M., Marchbank, M. T., Deutscher, S. L., Sproat, B., Leszczynska, G., Malkiewicz, A., and Agris, P. F. (2007) Specificity of phage display selected peptides for modified anticodon stem and loop domains of tRNA. *Protein J.* 26, 61–73.

(11) Meroueh, M., Grohar, P. J., Qiu, J., SantaLucia, J., Jr, Scaringe, S. A., and Chow, C. S. (2000) Unique structural and stabilizing roles for the individual pseudouridine residues in the 1920 region of *Escherichia coli* 23S rRNA. *Nucleic Acids Res.* 28, 2075–2083.

(12) Mendes, J., Soares, C. M., and Carrondo, M. A. (1999) Improvement of side-chain modeling in proteins with the self-consistent mean field theory method based on an analysis of the factors influencing prediction. *Biopolymers* 50, 111–131.

(13) Gohlke, H., Kiel, C., and Case, D. A. (2003) Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes. *J. Mol. Biol.* 330, 891–913.

(14) Hawkins, G. D., Cramer, C. J., and Truhlar, D. G. (1996) Parametrized models of aqueous free energies of solvation based on pairwise descreening of solute atomic charges from a dielectric medium. *J. Phys. Chem.* 100, 19824–19839.

(15) Jayaram, B., Liu, Y., and Beveridge, D. L. (1998) A modification of the generalized Born theory for improved estimates of solvation energy and pK shifts. *J. Chem. Phys.* 109, 1465–1471.

(16) Jayaram, B., Sproun, D., and Beveridge, D. L. (1998) Solvation free energy of biomacromolecules: parameters for a modified generalized born model consistent with the AMBER Force Field. *J. Phys. Chem. B* 102, 9571–9576.

(17) Onufriev, A., Bashford, D., and Case, D. A. (2000) Modification of the generalized Born model suitable for macromolecules. *J. Phys. Chem. B* 104, 3712–3720.

(18) Agris, P. F., Marchbank, M. T., Newman, W., Guenther, R., Ingram, P., Swallow, J., Mucha, P., Szyk, A., Rekowski, P., Peletskaya, E., and Deutscher, S. L. (1999) Experimental models of protein-RNA interaction: isolation and analyses of tRNA<sup>Phe</sup> and U1 snRNA-binding peptides from bacteriophage display libraries. *J. Protein Chem.* 18, 425–435.

(19) Mucha, P., Szyk, A., Rekowski, P., Weiss, P. A., and Agris, P. F. (2001) Anticodon domain methylated nucleosides of yeast tRNA<sup>Phe</sup> are significant recognition determinants in the binding of a phage display selected peptide. *Biochemistry* 40, 14191–14199.

(20) Huang, P. S., Love, J. J., and Mayo, S. L. (2007) A *de novo* designed protein-protein interface. *Protein Sci.* 16, 2770–2774.

(21) Xiao, X., Hall, C., and Agris, P. F. (2014) The design of a peptide sequence to inhibit HIV replication — A search algorithm combining Monte Carlo and self-consistent mean field techniques. *J. Biomol. Struct. Dyn.*, DOI: 10.1080/07391102.2013.825757.

(22) Koehl, P., and Delarue, M. (1996) Mean-field minimization methods for biological macromolecules. *Curr. Opin. Struct. Biol.* 6, 222–226.

(23) Lovell, S. C., Word, J. M., Richardson, J. S., and Richardson, D. C. (2000) The penultimate rotamer library. *Proteins* 40, 389–408.

(24) Murata, Y., and Wharton, R. P. (1995) Binding of pumilio to maternal hunchback mRNA is required for posterior patterning in *Drosophila* embryos. *Cell* 80, 747–756.

(25) Zhang, B., Gallegos, M., Puoti, A., Durkin, E., Fields, S., Kimble, J., and Wickens, M. P. (1997) A conserved RNA-binding protein that regulates sexual fates in the *C. elegans* hermaphrodite germ line. *Nature* 390, 477–484.

(26) Daubner, G. M., Cléry, A., and Allain, F. H. (2013) RRM-RNA recognition: NMR or crystallography...and new findings. *Curr. Opin. Struct. Biol.* 23, 100–108.

(27) Cusack, S., Yaremchuk, A., and Tukalo, M. (1996) The crystal structures of *T. thermophilus* lysyl-tRNA synthetase complexed with *E. coli* tRNA<sup>Lys</sup> and a *T. thermophilus* tRNA<sup>Lys</sup> transcript: anticodon recognition and conformational changes upon binding of a lysyl-adenylate analogue. *EMBO J.* 15, 6321–6334.

(28) Long, K. S., and Crothers, D. M. (1995) Interaction of human immunodeficiency virus type 1 Tat-derived peptides with TAR RNA. *Biochemistry* 34, 8885–8895.

(29) Hemmerich, P., Bosbach, S., von Mikecz, A., and Krawinkel, U. (1997) Human ribosomal protein L7 binds RNA with an alpha-helical arginine-rich and lysine-rich domain. *Eur. J. Biochem.* 245, 549–556.

(30) Patel, D. J. (1999) Adaptive recognition in RNA complexes with peptides and protein modules. *Curr. Opin. Struct. Biol.* 9, 74–87.

(31) Gosser, Y., Hermann, T., Majumdar, A., Hu, W., Frederick, R., Jiang, F., Xu, W., and Patel, D. J. (2001) Peptide-triggered conformational switch in HIV-1 RRE RNA complexes. *Nat. Struct. Biol.* 8, 146–150.

(32) Godin, K. S., and Varani, G. (2007) How arginine-rich domains coordinate mRNA maturation events. *RNA Biol.* 4, 69–75.

(33) Meehan, H. A., and Connell, G. J. (2001) The hairpin loop but not the bulged C of the iron responsive element is essential for high affinity binding to iron regulatory protein-1. *J. Biol. Chem.* 276, 14791–14796.

(34) Dominski, Z., Erkmann, J. A., Greenland, J. A., and Marzluff, W. F. (2001) Mutations in the RNA binding domain of stem-loop binding protein define separable requirements for RNA binding and for histone pre-mRNA processing. *Mol. Cell. Biol.* 21, 2008–2017.

(35) Battle, D. J., and Doudna, J. A. (2001) The stem-loop binding protein forms a highly stable and specific complex with the 3' stem-loop of histone mRNAs. *RNA* 7, 123–132.

(36) Spangord, R. J., and Beal, P. A. (2001) Selective binding by the RNA binding domain of PKR revealed by affinity cleavage. *Biochemistry* 40, 4272–4280.

(37) Ginisty, H., Amalric, F., and Bouvet, P. (2001) Two different combinations of RNA-binding domains determine the RNA binding specificity of nucleolin. *J. Biol. Chem.* 276, 14338–14343.

(38) Agris, P. F. (1996) The importance of being modified: roles of modified nucleosides and Mg<sup>2+</sup> in RNA structure and function. *Prog. Nucleic Acid Res. Mol. Biol.* 53, 79–129.

(39) Agris, P. F. (2008) Bringing order to translation: Contributions of tRNA anticodon domain modifications. *EMBO Rep.* 9, 629–635.

(40) Deutsch, C., El Yacoubi, B., de Crécy-Lagard, V., and Iwata-Reuyl, D. (2012) Biosynthesis of threonylcarbamoyl adenosine (t<sup>6</sup>A), a universal tRNA nucleoside. *J. Biol. Chem.* 287, 13666–13673.

(41) Colby, D. S., Schedl, P., and Guthrie, C. (1976) A functional requirement for modification of the wobble nucleotide in the anticodon of a T4 suppressor tRNA. *Cell* 9, 449–463.

(42) Shaw, P. A., Tidwell, J. T., Chia, L. L., Randerath, E., and Randerath, K. (1978) Base composition studies on transfer RNA from normal and regenerating rat liver. *Biochim. Biophys. Acta* 518, 457–463.

(43) Mackay, J. P., Font, J., and Segal, D. J. (2011) The prospects for designer single-stranded RNA-binding proteins. *Nat. Struct. Mol. Biol.* 18, 256–261.