Amino Acid Signature Enables Proteins to Recognize Modified tRNA

Jessica L. Spears,†‡§ Xingqing Xiao,§ Carol K. Hall,§ and Paul F. Agris*†‡

†The RNA Institute, University at Albany-SUNY, Life Sciences Research Building, Albany, New York 12222, United States
‡Department of Biological Sciences, University at Albany-SUNY, Life Sciences Research Building, Albany, New York 12222, United States
§Department of Chemical and Biomolecular Engineering, North Carolina State University, Engineering Building, Raleigh, North Carolina 27695-7905, United States

ABSTRACT: Human tRNA^{Lys}_{UUU} is the primer for HIV replication. The HIV-1 nucleocapsid protein, NCp7, facilitates htRNA^{Lys}_{UUU} recruitment from the host cell by binding to and remodeling the tRNA structure. Human tRNA^{Lys}_{UUU} 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-s^N-threonylcarbamoyladenosine, adjacent to the anticodon at position-37, are important to the recognition of htRNA^{Lys}_{UUU} 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_2-F-Pho-X-G/A-W-R-X_2-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^{Lys}_{UUU}. Peptides of this sequence specifically recognized and bound modified htRNA^{Lys}_{UUU} 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^{Lys}_{UUU}. The three human tRNA^{Lys}_{UUU} decode the two lysine codons, AAA and AAG. Two of the isoacceptors, hrrNA^{Lys}_{UUU} with the anticodon CUU, decode AAG, but only one, hrtRNA^{Lys}_{UUU} with the anticodon UUU, responds to the cognate codon AAA and wobbles to AAG. Besides its important role in protein synthesis, htrNA^{Lys}_{UUU} serves as the primer of reverse transcription in the replication of the lentiviruses, including human immunodeficiency virus type 1 (HIV-1). During the replication of HIV-1, the host cell htrNA^{Lys}_{UUU} is recognized and bound and its structure destabilized by nucleocapsid protein 7 (NCp7). This destabilization allows the relaxed U-rich anticodon stem loop (hASL^{Lys}_{UUU}), as well as the acceptor stem (Figure 1), to be annealed to the HIV viral RNA. During the subsequent infection, htrNA^{Lys}_{UUU} is the primer for HIV reverse transcriptase.

htRNA^{Lys}_{UUU} 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. 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^5s^U_{34}), at tRNA’s wobble position-34, and 2-methylthio-s^N-threonylcarbamoyladenosine (m^5tA_{37}), at position-37, 3′-adjacent to the anticodon in the loop of the hASL^{Lys}_{UUU} are both chemically rich and constitute a unique combination in human tRNAs (Figure 1). 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. The presence of these modifications increases NCp7 affinity for hASL^{Lys}_{UUU} almost 10-fold (K_d = 0.28 ± 0.03 μM for modified and K_d = 2.30 ± 0.62 μM for unmodified ASL). NCp7 is critical to HIV replication because it binds and relaxes the htrNA^{Lys}_{UUU} structure, facilitating annealing of the tRNA to the viral genomic RNA and packaging of the genomic RNA into the viral capsid.

Fifteen- and 16-amino acid peptides have been selected to mimic NCp7’s preferential recognition of the fully modified
hASL_{3}, UUU-9,10 These peptides can be used to study modification-dependent protein recognition of RNAs, in general, and recognition and annealing of htRNA_{3}, UUU to the HIV viral RNA, specifically. One peptide, P6 (sequence RVTHHAFLGAHRTVG), was also shown to mimic NCp7. P6 not only binds hASL_{3}, UUU but also destabilizes the ASL structure as does NCp7.9 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 hirtRNA_{3}, UUU. 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_{3}, UUU 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_{3}, UUU.

**EXPERIMENTAL PROCEDURES**

Materials and Reagents. All materials, buffers, and reagents were of RNA grade quality and RNase free. The modified and unmodified hASL_{3}, UUU were chemically synthesized by Thermo Scientific. The modified hASL_{3}, UUU was synthesized with the nucleoside phosphoramidites that were 2',3'-protected with tert-butyldimethylsilyl-ether.9 The unmodified hASL_{3}, UUU was synthesized with “ACE” chemistry.11 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 to the hASL_{3}, UUU complex, and the free energies of the ligand and the receptor are calculated according to the following equation:

\[ \Delta G_{\text{binding}} = G_{\text{complex}} - G_{\text{ligand}} - G_{\text{receptor}} \]

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 (\(\Delta G_{\text{binding}}\)).

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 conformation 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.13 It was calculated according to the following equation:

\[ \Delta G_{\text{binding}} = G_{\text{complex}} - G_{\text{ligand}} - G_{\text{receptor}} \]
where $G_{\text{complex}}$, $G_{\text{ligand}}$, and $G_{\text{receptor}}$ 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:

$$G_{\text{TOT}} = U_{\text{INT}} + U_{\text{VDW}} + U_{\text{ELE}} + G_{\text{EGB}} + G_{\text{GBSUR}}$$

(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 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 $U_{\text{INT}}$

$$U_{\text{INT}} = U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}}$$

$$= \sum_{\text{bonds}} K_b (b - b_0)^2 + \sum_{\text{angles}} K_\theta (\theta - \theta_0)^2$$

$$+ \sum_{\text{dihedrals}} (V_\phi/2)[1 + \cos(n\phi - \delta)]$$

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

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

$$U_{\text{VDW}} = \sum_i \sum_{j \neq i} \left( \frac{\epsilon_{ij}(\sigma_{ij}^2)}{r_{ij}^6} - \frac{2\epsilon_{ij}}{r_{ij}^3} \right)$$

(4)

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

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

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

(5)

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

(d) polar solvation energy $G_{\text{EGB}}$

$$G_{\text{EGB}} = -\frac{1}{2} \sum_i \left( 1 - \frac{1}{e_i} \right) \left[ \text{f}_{\text{GB}}(r_{ii}, \alpha_x, \alpha_x) \right]$$

$$- \frac{1}{2} \sum_i \sum_{j \neq i} \left( 1 - \frac{1}{e_i} \right) \left[ \text{f}_{\text{GB}}(r_{ij}, \alpha_x, \alpha_x) \right]$$

(6)

where $e_i (x = i \text{ or } j)$ represents the charge quantity of atom $x$, $e_i$ is the relative dielectric constant (in this work, $e_i = 80.0$ for water), $r_{ij}$ is the actual distance between atoms $i$ and $j$ (thus, $r_{ii} = 0.000$), $\alpha_x (x = i \text{ or } j)$ is the effective Born radii of atom $x$, and $\text{f}_{\text{GB}}$ is a smoothing function, of which a common form is

$$\text{f}_{\text{GB}} = \left[ r_{ij}^2 + a \alpha_x \exp \left( -r_{ij}^2 / 4a \alpha_x \right) \right]^{1/2}$$

(7)

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

$$G_{\text{GBSUR}} = \gamma \sum_i \sigma_i + b$$

(8)

where $\gamma$ is surface tension (in this work, $\gamma = 0.0072$ kcal/mol/Å² 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. The lyophilized peptide set was reconstituted via standard suggestions from the manufacturer (80% DMSO/20% H2O, 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 Na2HPO4 and 10 mM KH2PO4, pH 6.8) in low volume 384 well plates. All buffers, peptides, and RNA were pipetted into wells via a liquid handling robot.
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.9,10 The peptides were selected from completely and partially randomized phage display libraries.9,10 However, optimizing 15- and 16-amino acid peptide sequences using this approach is not feasible since there are over 3.3 × 10^{25} 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 algorithms20 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 strategy3 that combines MC with SCMF to evolve amino acid sequences. The peptide P6 sequence RVTHHAF0LGAHRTVG3 was the starting point from which an optimized peptide was sought to bind the modified hASL^{3,3}UUU 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^{3,3}UUU 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 category, via $N_{\text{pos}}, N_{\text{neg}}, N_{\text{pho}}, N_{\text{neg}}$, $N_{\text{dia}}, N_{\text{gly}}$, we maintained hydration properties similar to those of the original P6 sequence while evolving the sequences.9

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^{3,3}UUU was taken from the high resolution solution structure,6 providing a restrained structure to which the peptide would bind in silico. Once the stable structure of the peptide P6 with the ASL^{3,3} 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 their residue category (Figure 2) using SCMF.22 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.23 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^{3,3}UUU.

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^{3,3}UUU 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-L-A-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^{3,3}UUU 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 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 |

---

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

| Peptide Designation and Rank | Peptide Sequence | Binding Energy (kcal/mol) |
|-----------------------------|------------------|--------------------------|
| 1P6                         | RVTHAPGGHKTG     | -21.26                   |
| P26                         | RTLLHALPGAQQTG   | -22.55                   |
| P27                         | RWQMTAPARKRSG    | -22.07                   |

Case 1

- P28: RWNHQCPWGNRQGG (-22.81)
- P35 (P35*): RWNHQCPWGNRANG (-22.78)
- P29: RWNHQCPWGNRAG (-22.64)

Case 2

- P30: RWNHQCPWNLWHRHG (-22.71)
- P31: RWHHSAPGPPWLSHG (-21.86)
- A: RWHHHFSPSLWRHG (-21.56)

Case 3

- B: RHHHHHPGGPWNC (14.58)
- P32: RHHHASPGPPWLSHG (-14.26)
- P33: RHHASHPGPPWLSHG (-13.94)

“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. Peptides are numbered in accordance with the previous report of phage displayed selected sequences. P6 is from the original selection, and P26–P38 were chosen based on predictions in silico. P35* with the sequence RWNHQCPWGNRANGLH has a single amino acid 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 hASL133UUU 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 hASL133UUU were selected for validation with a fluorescence assay. 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 P6 from phage display selections,9 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 hASL133UUU and to what relative degree (Figure 4). In this screening assay, P6 behaved as expected.

Table 3. Synthesized Peptides Based on Predictions in Silico

| Peptide | Sequence |
|---------|----------|
| P1      | FSVSPFSLAPAPDRS |
| P6      | RVTHAPGLAHRTV  |
| P26     | RTLLHALPGAQQT  |
| P27     | RWQMTAPARKRSG  |
| P29     | RWNHQCPWGNRAG  |
| P30     | RWNHQCPWNLWHRHG|
| P31     | RWQHSPFHLWRMSG |
| P32     | RHHHASPGPPWLSHG|
| P33     | RHHASHPGPPWLSHG|
| P34     | RPQHHSNFSGWKNVG|
| P35*    | RWNHQCPWGNRANG |
| P36     | RWNHQCPWGNRANG |
| P37     | RWQHSPFHLWRMSG |
| P38     | RWQHSPFHLWRMSG |

“Peptides named following the peptides from original phage display library screens.9 P1 and P6 (bold font) are from the original screen. P6 has been characterized.9 In addition, the binding of the modified and unmodified hASL133UUU by P1, P27, P31, and P35* (bold font and shaded row) are characterized in this article.

When bound by the modified hASL133UUU, P6 fluorescence was altered more than that when bound by the unmodified hASL133UUU. 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 hASL133UUU (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 and unmodified hASL133UUU 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/or unmodified hASL133UUU and light gray bars represent the percent change in the presence of the unmodified hASL133UUU. Sequences for P1–P38 are presented in Table 3

Figure 4. Fluorescence of chemically synthesized peptides effected by modified and unmodified hASL133UUU. An initial fluorescent signal (FS0) of peptide alone (1.5 μM) was obtained. Then, a 2-fold excess of ASL was added to each peptide, and the fluorescent signal (FS1) was monitored. The percent change (100·FS1/FS0) is graphed for each of the assayed peptides. Dark gray bars represent the percent change in fluorescence in the presence of the modified hASL133UUU and light gray bars represent the percent change in the presence of the unmodified hASL133UUU.
preferentially bound the modified hASL<sub>Ly3</sub>UUG as compared to their binding of the unmodified hASL<sub>Ly3</sub>UUG as did the phage selected P6, P27, P31, and P35<sup>+</sup> showed a very high selectivity for the modified hASL<sub>Ly3</sub>UUG. When these peptides were tested for their binding to unmodified hASL<sub>Ly3</sub>UUG little or no change in fluorescence was observed (Figure 4; Table 4). However, with the addition of the modified hASL<sub>Ly3</sub>UUG, the fluorescent signals of P27, P31, P35<sup>+</sup>, and P6 were quenched. Peptides P32 and P33 had the best binding properties among the case 3 modiﬁed peptides P27, P31, and P35<sup>+</sup>(Figure 4). Thus, peptides P27, P31, and P35<sup>+</sup> evolved in silico from P6 as having high affinities for the modified hASL<sub>Ly3</sub>UUG also appear from the screening assay in vitro to have higher affinities and higher speciﬁcities for the modiﬁed RNA.

Table 4. Affinity of Peptides for hASL<sub>Ly3</sub>UUG<sup>a</sup>

| peptide | modified or unmodiﬁed hASL<sub>Ly3</sub>UUG | K<sub>i</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<sup>+</sup> | 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>i</sub> value from previous studies. <sup>c</sup>K<sub>i</sub> value is approximate for it is one-tenth the concentration of the ﬂuorescent peptide reporter.

In contrast, the ﬂuorescence of P27 was quenched considerably with increasing amounts of the modiﬁed hASL<sub>Ly3</sub>UUG. The peptide bound the RNA with high afﬁnity (K<sub>i</sub> = 0.05 ± 0.10 μM) (Table 4). P27 exhibited the highest afﬁnity for the modiﬁed hASL<sub>Ly3</sub>UUG. P31 bound the modiﬁed hASL<sub>Ly3</sub>UUG with a 10-fold less afﬁnity (K<sub>i</sub> = 0.50 ± 0.10 μM), and its afﬁnity for the hASL was equivalent to that of P6 (Table 4). P35<sup>+</sup> bound the modiﬁed hASL<sub>Ly3</sub>UUG with the lowest afﬁnity of these peptides (K<sub>i</sub> = 1.87 ± 1.00 μM). The ﬂuorescence change in binding the unmodiﬁed hASL<sub>Ly3</sub>UUG resulted in an equivalent binding constant (K<sub>i</sub> = 0.13 ± 0.02 μM). However, the ﬂuorescence change in binding the unmodiﬁed hASL<sub>Ly3</sub>UUG resulted in an equivalent binding constant (K<sub>i</sub> = 0.15 ± 0.04 μM). Thus, in contrast to P27, P31, P35<sup>+</sup>, and P6, P1 lacked speciﬁcity for the modiﬁed hASL<sub>Ly3</sub>UUG.

Computational Analysis of Peptide P27 Binding of the Modified hASL<sub>Ly3</sub>UUG. Peptide P27 has the highest afﬁnity coupled with highest speciﬁcity for binding of the modiﬁed hASL<sub>Ly3</sub>UUG. The calculated structure bound to the modiﬁed hASL<sub>Ly3</sub>UUG 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<sub>Ly3</sub>UUG. The two arginines, R<sub>1</sub> and R<sub>12</sub>, bracket the threonylcarbamoyl-group of ms<sup>4</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 mc<sup>5</sup>S<sub>U</sub><sub>14</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<sub>Ly3</sub>UUG 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>4</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 posteria and the FBF protein was found responsible in Caenorhabditis to mediate the sperm/oocyte switch. 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. 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<sub>Ly3</sub>UUG 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>Lys<sub>3</sub></sup><sub>UUU</sub> with high affinity and specificity. (A) The computed equilibrium binding structure of the modified hASL<sup>Lys<sub>3</sub></sup><sub>UUU</sub> bound by P27. The peptide backbone is in gold, and the ribose-phosphodiester backbone of the hASL<sup>Lys<sub>3</sub></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 m<sub>5</sub>C<sub>2</sub>S<sub>2</sub>T<sub>6</sub>A<sub>37</sub> (purple) and m<sub>cm</sub>5S<sub>2</sub>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>Lys<sub>3</sub></sup><sub>UUU</sub> is not shown. The table characterizes the contributions of different binding modes: ΔG<sub>binding</sub>, Gibbs free energy of binding; BE<sub>GBS</sub>/<sub>WS</sub><sub>B</sub>, 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>Lys<sub>3</sub></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>7</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>Lys<sub>3</sub></sup><sub>UUU</sub> 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 m<sub>cm</sub>5S<sub>2</sub>U<sub>34</sub> and m<sub>5</sub>C<sub>2</sub>S<sub>2</sub>T<sub>6</sub>A<sub>37</sub> due to VDW energy (Figure 5B). The increased number of hydrophilic 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>Lys<sub>3</sub></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>Lys<sub>3</sub></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>Lys<sub>3</sub></sup><sub>UUU</sub> loop. For instance, one could expect that the phenyl-ring of phenylalanine would intercalate within events for the 15 amino acid peptide theoretical combinations (>3.3 × 10<sup>23</sup>) 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>Lys<sub>3</sub></sup><sub>UUU</sub> modifications. We believe that we can improve the accuracy of in <i>silico</i> 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 <i>in vitro</i> and/or <i>in vivo</i> experiments. The validation screens in <i>sitio</i> revealed that while the computer algorithms were not 100% correct in predicting peptide sequences with both high affinity and specificity however, the selection in <i>silico</i> was a serious tool for predicting binding trends and quickly screening through many peptide sequence combinations.

### 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>*</sup> | R | W | Q/N | H | X | X | F | Pho | X | G/A | W | R | X | X | G |

<sup>*</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 N6-threonylcarbamoyl-group of ms2t6A37 is known to enhance base stacking.6 Phenylalanine has been observed to intercalate between anticodon nucleosides of tRNAlys in the cocrysal structure of lysyl-tRNA synthetase and tRNAlys.27 Instead of the expected intercalation, in the computed structure F2 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 hASLlys3 UUU and have two arginines each, there is little sequence homology with RNA binding proteins that are rich in arginine.28–32 or with single-stranded RNA binding proteins.33–37

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.38 There the modifications negate intraloop hydrogen bonding and can enhance or even decrease the possibility of base stacking.39 Peptides that recognize intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39 There the modifications negate intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39 There the modifications negate intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39 There the modifications negate intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39 There the modifications negate intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39 There the modifications negate intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39 There the modifications negate intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39 There the modifications negate intraloop hydrogen bonding and can enhance or decrease the possibility of base stacking.39

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 RNA substrates having t6A37. 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.40 Even the earliest analyses demonstrated that many tRNA modification enzymes often recognize a partially modified RNA.41,42

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 htRNAlys 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.43

REFERENCES

(1) Ratner, L., Haseltine, W., Patarca, R., Livak, K. L., Starich, B., Josephs, S., Doran, D. R., Rafalski, J. A., Whitehorn, E. A., Baumeister, K., Irvanoff, L., Petretty, S. R., Pearson, M. L., Lautenberger, J. A., Papas, T. S., Ghrayeb, 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 nucleoside protein in HIV-1 reverse transcription. RNA Biol. 7, 754–774.

(5) Chan, B., Wiedermair, 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) Vendex, 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<sub>51UUU</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<sub>51UUU</sub> includes disruption of intraloop hydrogen bonding by the naturally occurring amino acid modification, t<sup>A</sup>. *Biochemistry* 39, 13396–13404.

(8) Agris, P. F., Vendex, 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-Malone, L., Stark, C. J., Kaur, A., Stolaruchuk, C., Sproat, B., Leszczynska, G., Malkiewicz, A., Safwat, N., Mucha, P., Guenther, R., and Agris, P. F. (2001) Functional recognition of the modified human tRNA<sub>51UUU</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 phase 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 tRNA. *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 solution 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 form theory for improved estimates of solvation energy and pH shifts. *J. Chem. Phys.* 109, 1465–1471.

(16) Jayaram, B., Sprous, 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 suited 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., Peletesksaya, E., and Deutscher, S. L. (1999) Experimental models of protein-RNA interaction: isolation and analyses of tRNA<sup>U</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>U</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.