5212–5217 *Nucleic Acids Research, 1998, Vol. 26, No. 22*

RNA bulge entropies in the unbound state correlate with peptide binding strengths for HIV-1 and BIV TAR RNA because of improved conformational access

Brooke Lustig*, Ivet Bahar[1,2] and Robert L. Jernigan[2]

Department of Chemistry, San Jose State University, San Jose, CA 95192, USA and Applied Sciences Consultants, 621-B River Oaks Parkway, San Jose, CA 95134, USA, [1]Polymer Research Center and Chemical Engineering Department, Bogazici University and TUBITAK Advanced Polymeric Materials Research Center, Bebek 80815, Istanbul, Turkey and [2]Molecular Structure Section, Laboratory of Experimental and Computational Biology, National Cancer Institute, MSC 5677, NIH, Bethesda, MD 20892-5677, USA

Received June 19, 1998; Revised and Accepted September 25, 1998

ABSTRACT

For the binding of peptides to wild-type HIV-1 and BIV TAR RNA and to mutants with bulges of various sizes, changes in the $\Delta\Delta G$ values of binding were determined from experimental K_d values. The corresponding entropies of these bulges are estimated by enumerating all possible RNA bulge conformations on a lattice and then applying the Boltzmann relationship. Independent calculations of entropies from fluctuations are also carried out using the Gaussian network model (GNM) recently introduced for analyzing folded structures. Strong correlations are seen between the changes in free energy determined for binding and the two different unbound entropy calculations. The fact that the calculated entropy increase with larger bulge size is correlated with the enhanced experimental binding free energy is unusual. This system exhibits a dependence on the entropy of the unbound form that is opposite to usual binding models. Instead of a large initial entropy being unfavorable since it would be reduced upon binding, here the larger entropies actually favor binding. Several interpretations are possible: (i) the higher conformational freedom implies a higher competence for binding with a minimal strain, by suitable selection amongst the set of already accessible conformations; (ii) larger bulge entropies enhance the probability of the specific favorable conformation of the bound state; (iii) the increased freedom of the larger bulges contributes more to the bound state than to the unbound state; (iv) indirectly the large entropy of the bound state might have an unfavorable effect on the solvent structure. Nonetheless, this unusual effect is interesting.

INTRODUCTION

RNA–protein interactions are important for many biological processes, including regulation. Understanding the details of the binding process is critical to being able to relate structure to function. Studies of protein–protein and DNA–protein binding have shown that flexibility of the binding partners is often critical for obtaining strong specific binding. These can be viewed as examples of induced fit, sometimes small in scale and sometimes large.

Nucleic acids characteristically form double helices that have only limited flexibility. In addition, RNA has more diverse structural elements, such as bulges and loops, that can be local sites of greater flexibility. Here we will explore the role of bulge and loop flexibility for RNA–protein binding, which is not well understood. One reported effect of flexibility involves strain associated with asymmetric internal loops (1), that can lead to a widening of the major groove to facilitate amino acid binding (2–5).

Specifically, we are going to investigate computationally the flexibility of bulges of various sizes. These calculations will be compared with others' experimental peptide binding studies for HIV (human immunodeficiency virus) and BIV (bovine immunodeficiency virus) that include various bulge insertions and deletions (4–6). Chen and Frankel (6) titrated the TAR binding region of BIV and its various mutants with a 17 residue Tat peptide. Relative amounts of peptide–TAR complex formed for wild-type and various mutant RNAs were determined from gel shift assays. These values can be expressed in terms of relative dissociation constants K_d for the mutant compared with the native. Values have also been reported for competition assays involving 14 and 38 residue Tat peptides (4).

Lattice models have been developed previously to generate coarse grained structures for tRNA, ribozyme RNA and an rRNA fragment (7,8). Each lattice point corresponds to one nucleotide and these points are used to account for the excluded volume. Rigid volume elements representing double helical stems are attached to the lattice points representing nucleotides of the various internal bulges. All possible bulge forms and their various attendant stem positions are generated, accounting for the effects of excluded volume. The resulting structures indicate preferred stem–stem arrangements and other structural features in agreement with experiment. Certain asymmetries and constraints in loop position were found to favor functionally important RNA conformations, including hammerhead ribozyme conformers known to be catalytically active (8). The complete enumeration

*To whom correspondence should be addressed at: Applied Sciences Consultants, 621-B River Oaks Parkway, San Jose, CA 95134, USA. Tel: +1 408 428 0560; Fax: +1 408 924 4945; Email: lustig@batnet.com

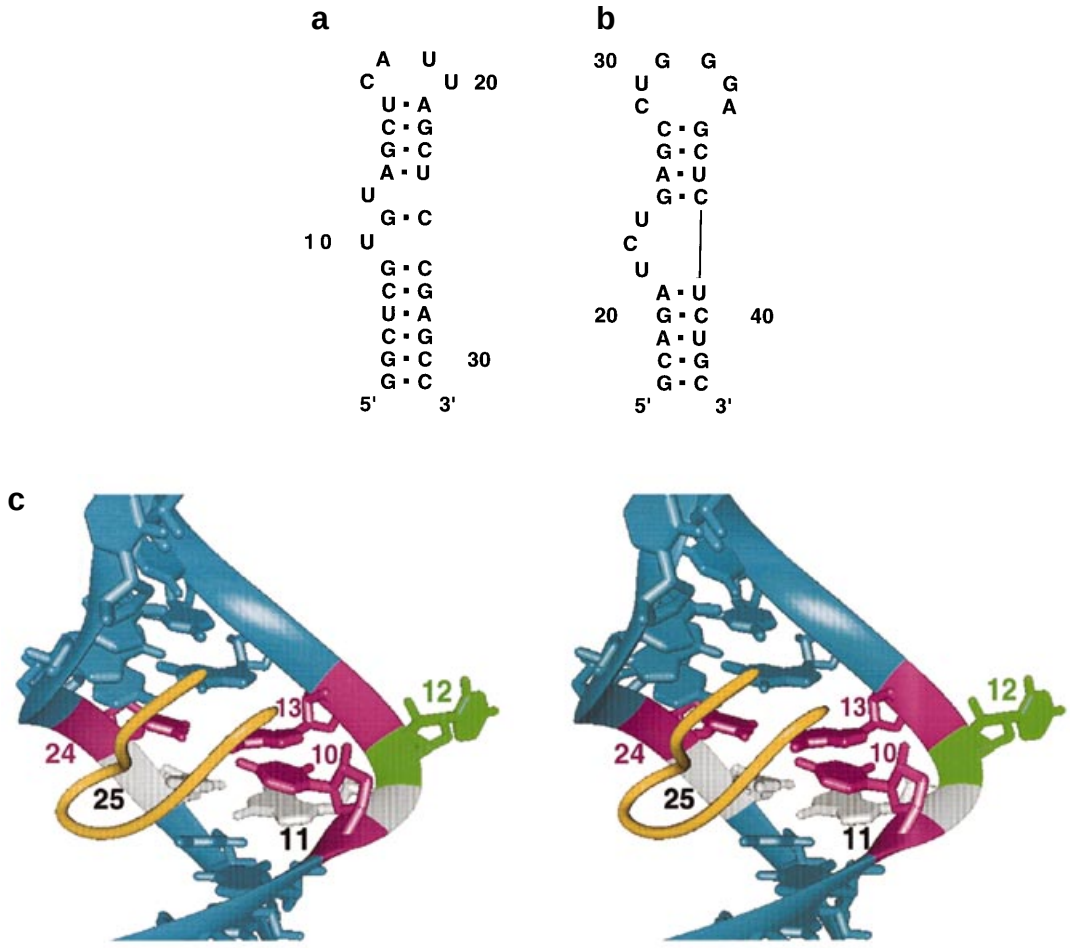

Figure 1. Secondary structures of TAR RNA binding regions. (**a**) For BIV TAR (6). (**b**) For HIV-1 TAR (4,5). (**c**) Three-dimensional structure of BIV Tat–TAR peptide–RNA complex (15,16) showing the unusual non-sequential triplet (bases 10-13-24) in magenta formed by the RNA. Structure shown is from the Brookhaven structure 1mnb.

of all conformations, made possible by this coarse grained lattice, avoids some conformational uncertainties associated with less exhaustive methods.

The relatively large scale conformational changes possible in the bulges would require substantial time to fully sample the ensemble of conformations and this could be important for binding intermediates. This contrasts directly with the smaller scale fluctuations more typically observed in other biomolecular structures. It can mean that the binding to an RNA bulge structure could directly reflect the difficulty of conformational access. Of course the size of the ensemble of generated RNA conformations can be related to conformational entropy.

In this paper we will estimate entropies of the various bulges in two ways: (i) using the number of conformational states generated on a lattice; (ii) from the size of the fluctuations obtained with the Gaussian network model (GNM) recently introduced (9) for describing the conformational flexibility of folded structures (9–11). The main conclusion here will be that decreases in free energy of binding for various bulge insertions are related directly to entropy increases for the unbound RNA. Normally a bound state would have a reduced entropy, but instead here the binding process is enhanced by higher entropies of the unbound RNA. Because the bulge is *likely* a site of flexibility within the binding site of the TAR RNA, it is logical to posit that the stiffness of the structures is diminished by increased flexibility, i.e. by larger bulges, and that access to the bound conformation is thereby enhanced. This is particularly plausible since the structure of the bound state shows a highly distorted bulge in which a base within the bulge forms a base triplet with a stem base pair (Fig. 1). Consequently, for the range of bulge sizes investigated here, access (and adaptability) to the bound RNA conformation might be critical, and this is improved by the greater flexibilities of the larger bulges. Of course, if the bulge were enlarged beyond a certain point, then the probability of the bound state conformation eventually would be reduced. Alternatively, it is possible that the increase in the bulge entropy is even more favorable for the bound state, in comparison with the unbound state.

In order to study the effects of the flexibility of the bulge in BIV and HIV TAR binding targets, we are going to utilize two extremely different *coarse grained* approaches: (i) a lattice method appropriate for investigating large scale changes in conformation; (ii) a GNM appropriate to gauge a range of effects including small scale fluctuations, based on the topology of inter-residue contacts in the examined structure.

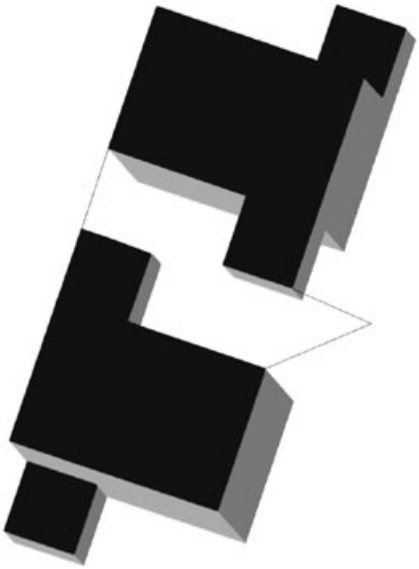

Figure 2. Lattice structure model of mutant BIV TAR with a single deletion in the bulge (–1 nt from bulge).

METHODS AND RESULTS

Lattice models

The lattice calculations are performed on a cubic lattice having cell edges of 6.15 Å, based on the distance between O3′ atoms in sequential nucleotides. The TAR binding regions of BIV and HIV-1 (Fig. 1) are divided into solid double helical stems and individual points representing unpaired nucleotides in the bulge. Each lattice point of the bulge represents an O3′ atom as shown by a sample lattice fold for a one base bulge in Figure 2. Attached to the bulge are two model rigid volume elements that represent the helical stems. Volume exclusion prohibits multiple occupancy of the bulge lattice points and lattice points within the model stems. These model stems are taken as elements of lattice volume fit to an ideal A-form helical stem of 7 bp. This is short enough to permit complete enumeration of all lattice forms of internal bulges.

The modified cubic lattice used here allows virtual bonds connecting O3′ atoms to have lengths of 6.15 or 8.70 Å, for a coordination number of 18, and this lattice offers a simple and appropriate way to exhaustively generate conformations. The two helical axes are allowed to be parallel or perpendicular with respect to one another. As with all lattice methods, the resulting structures are only approximate, but the distributions of conformations should be representative (7,8). The number of RNA lattice folds ranges from 66 844 configurations for the +1 nt added to the HIV-1 TAR bulge to only one possible lattice configuration associated with the –3 nt form that has no bulge, which is fully constrained in this lattice model.

The change in entropy can be estimated from N^{mutant} and N^{wt}, the numbers of mutant and wild-type RNA lattice configurations, as

$$\Delta S \sim C_1 k \ln\left(\frac{N^{mutant}}{N^{wt}}\right) \quad \mathbf{1}$$

where k is the Boltzmann constant and C_1 reflects the details and extent of coarse graining of the lattice but in practice is simply used to scale the results.

Gaussian network model (GNM)

This model gives information about conformational fluctuations based on constraints imposed between close nucleotides (9–11). The entropic component of the change in free energy for the motion $\Delta R_i = |\mathbf{R}_i - \mathbf{R}_i^0|$ of the ith nucleotide away from its equilibrium position ($\mathbf{R}_i^0$) in the folded state is estimated from its mean square fluctuation $<(\Delta R_i)^2>$ as (10)

$$-T\Delta S_i = 3kT(\Delta R_i)^2/2<(\Delta R_i)^2> \quad \mathbf{2}$$

Accordingly, in the GNM description of folded structures any conformational change away from the original equilibrium state (or the native state) is accompanied by a decrease in entropy and, consequently, an increase in free energy. And nucleotides subject to smaller amplitude equilibrium fluctuations exhibit a stronger resistance, of entropic origin, to conformational change. For a given size conformational rearrangement, at a given temperature, the energy cost contributed by the ith residue may thus be expressed as

$$\Delta G_i = C_2/<(\Delta R_i)^2> \quad \mathbf{3}$$

This conforms with the higher free energy requirement for moving residues subject to smaller amplitude fluctuations, for H/D exchanging with the solvent, as observed and predicted for a series of proteins (10). The total molecular entropic contribution to the free energy change is taken to be the sum over all residues.

Correlations with experimental binding energies

All possible lattice structures are enumerated for the various bulge mutants of the BIV and HIV-1 TAR structures shown in Figure 1a and b, respectively. We assume that the structure for the 1 nt deletion (–1 nt) from the bulge of BIV TAR (Fig. 1) is the same as the –2 nt bulge structure of HIV-1 TAR. Similarly, we can assume that the model structures for –2 nt from bulge BIV TAR and –3 nt from bulge HIV-1 TAR are the same.

The change in ΔG of binding resulting from a mutation in the bulge for both peptide–BIV TAR and peptide–HIV TAR can be calculated from the corresponding dissociation constants K_d as

$$\Delta\Delta G = RT \ln\left(\frac{K_d^{mutant}}{K_d^{wt}}\right) \quad \mathbf{4}$$

The K_d values for HIV-1 TAR are an average over the Tfr14 and Tfr38 peptides (4). The K_d value for –1 nt in the bulge of BIV TAR is an average involving two possible –1 nt bulge mutants. The correlation plot (Fig. 3a) compares the change in experimental free energy associated with change in bulge size with that calculated from lattice calculations, for four HIV-1 TAR bulge mutants +1, –1, –2 and –3 nt. They also include a similar correlation involving the –1 nt BIV TAR bulge mutant. It is clear that there is a strong correlation in this plot; the correlation coefficient is 0.98 and the probability of being random is <0.005 (12).

Correlations with GNM calculations

The entropic contribution for the change in free energy associated with mean sqaure fluctuations is calculated using **3** for the HIV-1

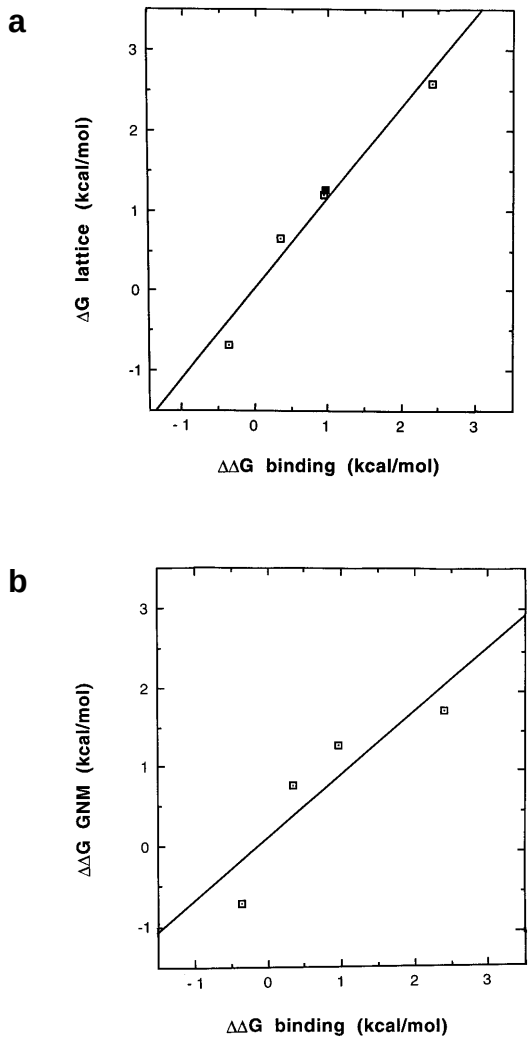

Figure 3. (**a**) Correlations of changes in free energies determined from binding (abscissa) and from lattice calculations (ordinate). Open squares indicate the average $\Delta\Delta G$ for binding at 22°C (4,5) of peptides Tfr14 and Tfr38 with HIV-1 and a series of mutants (from left to right: +1 nt added to the bulge, –1 nt deleted from the bulge, –2 nt from the bulge, –3 nt from the bulge). The closed square indicates the $\Delta\Delta G$ for binding at 4°C (6) of a target peptide with BIV TAR and a mutant (–1 nt from bulge). The ΔG lattice values involve changes in conformational free energies calculated from the various bulge entropies. The linear regression line is $y = 0.26 \times 10^{-3} + 1.13x$ and the correlation coefficient is 0.98 ($P < 0.005$). (**b**) Correlations of changes in free energies determined from HIV-1 binding (abscissa) and from GNM calculations (ordinate). The linear regression line is $y = 0.11 + 0.79x$ and the correlation coefficient is 0.89 ($P < 0.12$). The two sets of calculated data are also highly correlated (see text).

wild-type and its +1, –1, –2 and –3 nt bulge mutants. A comparison of the experimental values with the corresponding calculated change in the ΔG values for the GNM method associated with forming the mutant bulges of HIV-1 is shown in Figure 3b. A good correlation is indicated, with a correlation coefficient of 0.89 and a probability of being random of <0.12 (12). An even stronger correlation is noted when the $\Delta\Delta G$ GNM values are plotted against the ΔG lattice values, with a correlation coefficient of 0.96, which indicates the probability of its being random as <0.05.

The correlations between the two independent calculated changes in the entropy and the experimentally determined changes in binding free energy indicate a significant role for flexibility in the binding of peptides to TAR RNAs. It is interesting to note that these entropies are so similar when calculated with two entirely different computational methods. The GNM results reflect flexibility based on the local density and distribution of contacts, whereas the calculated lattice entropies are a more direct measure of the overall flexibility. But, apparently the overall flexibility is closely related to the local packing density and contact distribution.

DISCUSSION

It is somewhat difficult to know how to characterize the unbound RNA conformationally even though there are NMR determined structures. The multiple structures modeled from the NMR data exhibit rather large excursions within the space of the molecule. Thus the flexible lattice model of the bulge conformations may be appropriate. But, in some sense these conformations are still not extremely different from one another and consequently the fluctuations about a single structure as in the GNM is a contrasting but not altogether inappropriate model. We have utilized these two different approaches to investigate this problem. In order to gain an understanding of the physical basis for the agreement between the experimentally observed $\Delta\Delta G$ values and the ΔG values determined from entropic considerations, we will next consider in more detail the physics and underlying assumptions of the two theoretical approaches presently adopted.

Entropy changes from lattice model calculations

In cases where both the original and final states enjoy a significant conformational freedom, beyond small rearrangements near the equilibrium state, complete enumeration of all accessible conformations on a lattice may be useful for estimating entropic effects. In the present lattice calculations, the conformational space accessible to the unbound bulge mutants has been considered, which provided a measure of their entropy change $\Delta S_{\text{unbound}}$ relative to wild-type TAR bulges or the corresponding contribution to free energy change $\Delta G_{\text{unbound}} = -T\Delta S_{\text{unbound}}$ (see **1** for a lattice estimation of $\Delta S_{\text{unbound}}$).

On the other hand, experimental data refer to $\Delta\Delta G_{\text{binding}} = \Delta G_{\text{bound}} - \Delta G_{\text{unbound}}$, which comprises both entropic and enthalpic effects, from the bound and unbound states. In a strict sense, $\Delta\Delta G_{\text{binding}} = \Delta H_{\text{bound}} - \Delta H_{\text{unbound}} - T\Delta S_{\text{bound}} + T\Delta S_{\text{unbound}}$. In the present analysis, the contribution $T\Delta S_{\text{unbound}}$ is evaluated, exclusively. As expected, $\Delta S_{\text{unbound}}$ is found to gradually decrease as one proceeds from the +1 nt mutant to the –3 nt mutant, giving rise to an increase in $\Delta G_{\text{unbound}}$. Yet, the accompanying change in $\Delta\Delta G_{\text{binding}}$ is positive. This may arise from two effects, both of which indeed correlate with the degree of flexibility of the TAR bulge at the Tat recognition site: (i) enthalpic effects resulting from a tighter binding that is achieved only when the TAR bulge is sufficiently flexible and permits an optimal interaction with the peptide; (ii) an enhancement of the conformational space of the bound state, which may more than counterbalance the effect of entropy increases in the unbound state.

It is interesting to consider the relative importances of entropy and enthalpy to biological macromolecular processes. However, the entropic and enthalpic effects may be somewhat coupled in

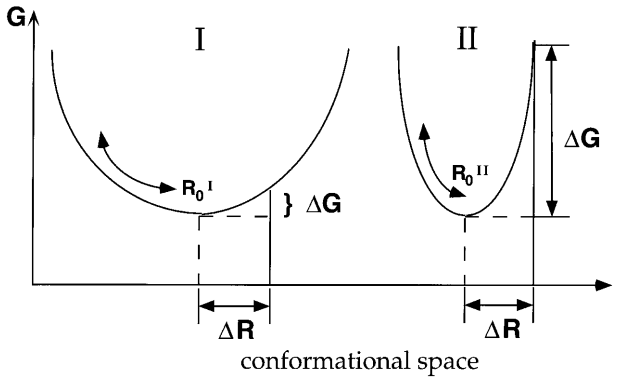

Figure 4. Schematic representation of the change in free energy accompanying a conformational change of fixed magnitude, Δ**R**, near the folded state, for two different equilibrium structures, I and II, differing in their overall conformational flexibility.

that a higher conformational freedom (or entropy) in the unbound state entails the availability of specific conformational states that permit a lower enthalpy association in the bound state. Thus, it is conceivable that conformational entropy or flexibility, as reflected by the smoothness of the energy landscape accessible in the folded state, can play a critical role in a process by providing access to the desired final state, but the present calculations are limited in providing information only about the entropies of the unbound state.

Conformational rearrangements in the GNM

The native state in this model is described by a multidimensional harmonic potential energy surface. There is no specificity in the pairwise interactions, i.e. no sequences, and the relative flexibilities of individual nucleotides, as well as that of the overall structure, are fully determined by the contact topology and are thereby entropic effects (13). Let the equilibrium conformation be designated $\mathbf{R}_0$. This is a minimum free energy state, which may be conveniently described by the schematic representation shown in Figure 4. Therein two different cases, I and II, are displayed, referring to a shallower and steeper energy well, respectively. In conformity with the curvature of the energy landscape, the structure $\mathbf{R}_0^I$ is characterized by larger amplitude equilibrium fluctuations $<(\Delta\mathbf{R}^I)^2>$ compared with those about $\mathbf{R}_0^{II}$. Cases I and II could be representative of +1 and –3 nt bulge mutants, for example.

Suppose, to achieve the bound state, a conformational rearrangement of a given size Δ**R** is required in both cases. As shown in Figure 4, structure I will accommodate this change by a free energy change ΔG^I smaller than ΔG^{II}, that required for structure II to undergo the same extent of deformation. Thus, if in the bound state, the bulge region is required to undergo a distortion of a given size, the free energy cost will be smaller for a structure having larger amplitude fluctuations, or higher flexibility, about the original state (**3**).

It should be noted that the experimental data would reflect the above depicted increase in free energy provided that: (i) relatively small size conformational rearrangements occur upon binding, such that these may be represented as fluctuations near the original global minimum; (ii) the contribution to the observed free energy change from effects other than intramolecular couplings in the RNA bulge, such as those associated with TAR–peptide or peptide–peptide interactions, are negligibly small. The departure between experimental data and theoretical results seen in Figure 3b is reasonable in view of these approximations.

Conclusion

Our interpretation of the correlation between the experimental $\Delta\Delta G$ values and the calculated bulge entropies is simply that the entropies of the unbound RNA reflect the ease of achieving the most specific peptide binding form for the RNA structure. This means that even though binding would reduce the entropy, access to the bound RNA conformation and/or adaptability of the original structure to assume conformational states required for optimal binding (through a possible induced fit mechanism) are more critical and would be enhanced by the larger flexibility in a bulge manifested in its higher entropy. It is interesting to note that a recent study on the binding specificities of HIV and BIV Tat proteins to a series of hybrid TARs also indicated that the presence of larger, more flexible bulges increases the peptide affinity (14). This effect was also attributed to the fact that a larger bulge may provide a more favorable bound conformation, perhaps by widening the major groove or allowing more flexibility to the recognition site.

In the present case, the two different estimates of bulge entropy both exhibit a strong correlation with binding strengths. This is in spite of the fact that neither calculation explicitly accesses the bound conformation. However, the fact that both of these estimates of entropy show correlations with the experimental data indicates that it really is the flexibility of the bulge that can enhance binding. Insofar as the energy of the bound state is concerned, it is likely that, in the case of mutants enjoying sufficient conformational flexibility, the bulge in the bound state assumes a relatively low energy conformation (or an ensemble of conformations) among those already available in the free state, whereas in the case of mutants with highly restrictive conformational space, binding is presumably achieved at the cost of assuming a highly strained state, only. The energy requirement for sustaining such a strained conformation may more than counterbalance the entropic effect associated with the smaller reduction in the number of available conformations compared with the more flexible mutants. Alternatively the flexibility remaining in the bound state might be relatively more favorable than the loss of freedom to the unbound state.

Overall, the original higher flexibility of the TAR bulge may give rise to a reduced strain and optimal intermolecular interaction in the bound state and the resulting relatively high probability of the bound conformation may possibly contribute to the observed correlation.

The NMR structures of the BIV TAR–peptide complex indeed indicate how the bulge can rearrange to achieve a highly specific peptide binding site at a widened major groove, by forming a non-sequential base triplet that is likely to represent a structure mutually induced when the peptide forms a β-hairpin (15–18). Other more complex explanations of the present phenomena remain possible, especially indirect ones wherein the solvent (water and/or ions) could have free energies opposing the free energy changes of the RNA bulges.

REFERENCES

1 Peritz,A.E., Kierzek,R., Sugimoto,N. and Turner,D.H. (1991) *Biochemistry*, **30**, 6428–6436.
2 Weeks,K.M. and Crothers,D.M. (1993) *Science*, **261**, 1574–1577.
3 Tan,R. and Frankel,A.D. (1995) *Proc. Natl Acad. Sci. USA*, **92**, 5282–5286.
4 Weeks,K.M. and Crothers,D.M. (1992) *Biochemistry*, **31**, 10281–10287.
5 Weeks,K.M. and Crothers,D.M. (1991) *Cell*, **66**, 577–578.
6 Chen,L. and Frankel,A.D. (1994) *Biochemistry*, **33**, 2708–2715.
7 Lustig,B., Covell,D.G. and Jernigan,R.L. (1994) *J. Biomol. Struct. Dyn.*, **12**, 145–161.
8 Lustig,B., Lin,N.H., Smith,S.M., Jernigan,R.L. and Jeang,K.-T. (1995) *Nucleic Acids Res.*, **23**, 3531–3538.
9 Bahar,I., Atilgan,A.R. and Erman,B. (1997) *Folding Des.*, **2**, 173–181.
10 Bahar,I., Wallqvist,A., Covell,D.G. and Jernigan,R.L. (1998) *Biochemistry*, **37**, 1067–1075.
11 Bahar,I. and Jernigan,R.L. (1998) *J. Mol. Biol.*, in press.
12 Bevington,P.R. (1969) *Data Reduction and Error Analysis for the Physical Sciences*. McGraw-Hill, New York, NY, p. 311.
13 Bahar,I., Atilgan,A.R., Demirel,M.C. and Erman,B. (1988) *Phys. Rev. Lett.*, **80**, 2733–2736.
14 Smith,C.A., Crotty,S., Harada,Y., Frankel,A.D. (1988) *Biochemistry*, **37**, 10808–10814.
15 Puglisi,J.D., Tan,R., Calnan,B.J., Frankel,A.D. and Williamson,J.R. (1992) *Science*, **257**, 76–80.
16 Puglisi,J.D., Chen,L., Blanchard,S. and Frankel,A.D. (1995) *Science*, **270**, 1200–1203.
17 Ye,X., Kumar,R.A. and Patel,D.J. (1995) *Chem. Biol.*, **2**, 827–840.
18 Greenbaum,N.L. (1996) *Structure*, **4**, 5–9.