Published online March 16, 2004

1746–1755 *Nucleic Acids Research, 2004, Vol. 32, No. 5*
DOI: 10.1093/nar/gkh331

Secondary structure and stability of the selenocysteine insertion sequences (SECIS) for human thioredoxin reductase and glutathione peroxidase

Andres Ramos[1], Andrew N. Lane[1,2], David Hollingworth[1] and Teresa W.-M. Fan[3,4,*]

[1]Division of Molecular Structure, National Institute for Medical Research, The Ridgeway, Mill Hill, London NW7 1AA, UK, [2]James Graham Brown Cancer Center, University of Louisville, 529 S. Jackson Street, Louisville, KY 40202, USA, [3]Department of Chemistry, University of Louisville, 2320 S. Brook Street, Louisville, KY 40208, USA and [4]Department of Land, Air and Water Resources, University of California, One Shields Avenue, Davis, CA 95616-8627, USA

Received December 4, 2003; Revised January 27, 2004; Accepted February 17, 2004

ABSTRACT

We have used high resolution NMR and thermodynamics to characterize the secondary structure and stability of the selenocysteine insertion sequences (SECIS) of human glutathione peroxidase (58 nt) and thioredoxin reductase (51 nt). These sequences are members of the two classes of SECIS recently identified with two distinct structures capable of directing selenocysteine incorporation into proteins in eukaryotes. UV melting experiments showed a single cooperative and reversible transition for each RNA, which indicates the presence of stable secondary structures. Despite their large size, the RNAs gave well resolved NMR spectra for the exchangeable protons. Using NOESY, the imino protons as well as the cytosine amino protons of all of the Watson–Crick base pairs were assigned. In addition, a number of non-canonical base pairs including the wobble G.U pairs were identified. The interbase-pair NOEs allowed definition of the hydrogen-bonded structure of the oligonucleotides, providing an experimental model of the secondary structure of these elements. The derived secondary structures are consistent with several features of the predicted models, but with some important differences, especially regarding the conserved sequence motifs.

INTRODUCTION

Selenium is an essential trace element required for normal growth and cellular functioning (1–3). Both deficiency and excess of Se lead to diseases or toxicity, and the margin between deficiency and toxicity is very narrow (4–6). The nutritional requirement of Se is associated with several critical enzymes involved in cellular redox processes. These include thioredoxin reductase (TrxR) and glutathione peroxidase (GPX) amongst others. These selenoproteins and TrxR, in particular, have been implicated in tumor suppression in mammals possibly via their effect(s) on redox control of cell cycle proteins, growth-promoting transcription factors and/or apoptosis (7–15).

The biologically active form of selenium in proteins is the twenty-first amino acid selenocysteine (SeC), which forms an essential part of the active site of mammalian TrxR and GPX (10). The biosynthesis of selenoproteins such as TrxR and GPX involves the insertion of SeC during translation by decoding the UGA stop codon in the appropriate context using the special charged $tRNA^{sec}$. This tRNA is charged initially with serine, and is then converted enzymatically to SeC using selenophosphate. Decoding UGA requires interaction of the translation machinery with additional factors. In prokaryotes, the general features of the decoding system have been elucidated. The multidomain translation elongation factor SelB recognizes a conserved GU sequence within a stem–loop structure immediately 3′ to the UGA codon, which helps recruit the charged $tRNA^{sec}$ (16–19).

In eukaryotes, the decoding system is more complex and involves at least two proteins, one recognizing the tRNA (mSelB/eEF^{Sec}), and the other(s) (e.g. SBP-2) interacting with a sequence in the 3′ untranslated region (UTR) of the mRNA transcript more than a kilobase downstream from the stop codon (20). This region of the 3′UTR, called the selenocysteine insertion sequence (SECIS), has a minimal functional size of 50–60 nt (21–24). This sequence is longer than the bacterial counterpart (18). Only a few short sequences, such as AA, AUGA and GU, are conserved between the SECIS of different selenoprotein genes. However, all SECIS elements appear to have extensive internal sequence complementarity such that Watson–Crick base-paired stem–loop structures are expected to form. Indeed, RNA folding algorithms

*To whom correspondence should be addressed. Tel: +1 502 852 6448; Fax: +1 502 852 8149; Email: teresa.fan@louisville.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors

consistently predict structures that contain two helical regions and an apical loop (21,24–27). The conserved AA residues lie in the 5′ side of the apical loop, and AUGA and UG bases lie close together on opposite sides of a helical stem. Modeling work in conjunction with chemical and enzymatic digestion experiments have also indicated that the SECIS element may contain novel G.A mismatches at the base of the helical stem, of the type found in GNRA hairpin loops (26). The size of the apical loop and the location of the conserved AA residues led to a proposal that there are two functional forms of SECIS, which can be experimentally interconverted by appropriate mutations (21). More recently, the conservation of these bases and the ability to form stem–loop structures of a certain length has been used to search for selenoproteins in a variety of genomes (27). This study found several new mammalian selenoprotein genes containing SECIS elements that, in some instances, were predicted to have significantly different secondary structures (27).

However, no direct structural characterization of SECIS has been reported. The elucidation and comparison of SECIS secondary structures is crucial for understanding the regulation of SeC incorporation. We have synthesized milligram quantities of the functional SECIS sequences for both human GPX and human TrxR. These two sequences are predicted to belong to two distinct classes of SECIS elements (21) and have the potential to form tandem G.A mismatched base pairs (26). Although the SECIS elements are large by NMR standards (51 and 58 nt), the imino proton region of the spectrum should be highly informative regarding the presence of stable Watson–Crick base pairs and other H-bonded structures (28–30). Two-dimensional NMR spectroscopy at high (14.1 T and 18.8 T) magnetic field strengths was employed to define the features of the secondary structure of the two SECIS, including the formation of the mismatched base pairs, which was compared with their thermodynamic stability obtained from UV melting and hydrodynamic properties by analytical ultracentrifugation.

MATERIALS AND METHODS

RNA synthesis

The SECIS RNA sequences for the TrxR and GPX genes are as follows: a 58 nt human cytoplasmic glutathione peroxidase SECIS (24)

```
          10        20        30        40        50
5′ ggc cua uga ggg ugu uuc cuc uaa acc uac gag gga gga aca ccu gau cuu
   aca ggc c
```

and a 51 nt human thioredoxin reductase SECIS (25)

```
          10        20        30        40        50
5′ guc cau gaa guc acc agu cuc aag ccc aug ugg uag gcg gug aug gaa caa
```

RNA was enzymatically synthesized *in vitro* via a run-off T7 transcription reaction, using complementary DNA templates in 10 1.5-ml reactions using 1.6 μM DNA template following established procedures (31–33). HPLC-purified oligodeoxynucleotide templates were purchased from Interactiva (Germany) (1 μmol scale). The two DNA strands were annealed by incubating at 65°C for 5 min followed by slow cooling to room temperature. The templates were then incubated at 37°C with the four rNTPs, PEG-8000, T7 RNA polymerase and transcription buffer containing Tris, spermidine, Triton X100 and DTT. The RNA transcripts were purified on 15% acrylamide gels containing 7 M urea in TBE buffer. The RNA products in the gel were visualized by UV shadowing, and the bands cut from the gel were macerated, electroeluted under sterile conditions at 4°C using an Elutrap electroeluter (Schleicher & Schuell) and dialyzed against decreasing concentrations of NaCl (1–0.1 M) in 10 mM sodium phosphate buffer at pH 7. The dialysate was lyophilized and subsequently exchanged to an appropriate buffer by microdialysis for NMR analysis (see below), gel electrophoresis and analytical ultracentrifugation. The last procedure was carried out on a Beckman Optima XLA instrument recording the A_{260} of ~1 μM samples at 20°C, as previously described (34). The yield of purified RNA from 10 reactions was at least 5 mg for both SECIS. Two independent syntheses were carried out for both RNA species, and the products were found by NMR and analytical electrophoresis to be essentially identical. This indicates that the folding of the two RNAs was reproducible and reversible.

NMR spectroscopy

RNA samples were equilibrated in 10 mM Na phosphate, 100 mM NaCl pH 6 at a concentration of ~0.9 mM. NMR spectra in H_2O were recorded at 600 and 800 MHz on Varian Inova NMR spectrometers at 5 or 10°C. Watergate (35) was used for solvent suppression for H_2O samples. For NOESY spectra, mixing times of 100 and 200 ms were employed. Samples were then lyophilized and redissolved in 100% D_2O for further analysis of non-exchangeable protons using TOCSY (50 ms mixing time) and NOESY (mixing times of 50–50 ms). ^{31}P NMR spectra were recorded at 14.1 T using a 5 mm switchable pfg probe.

UV melting

Thermal unfolding studies were carried out in a Cary300 UV spectrophotometer equipped with a computer-controlled six-place Peltier cell holder. Temperature was measured directly in the cuvette. Solutions at an absorbance of 0.5–1 unit at 260 nm were degassed and introduced into the cuvette, and allowed to incubate at 15°C for 20 min. The solution was then heated at 0.25, 0.5 or 1°C/min until after the melting transition. All solutions were scanned at least twice, and independently measured three times to determine the reproducibility. T_m was determined from the derivative of the A_{260} versus T plot, and also by fitting the absorbance profile $A(T)$ to the equations:

$$A(T) = \varepsilon_F r_t + \Delta\varepsilon u \quad \mathbf{1}$$

where ε_F is the absorption coefficient of the folded form, $\Delta\varepsilon$ is the difference in the absorption coefficients for the unfolded and folded forms, u is the concentration of the unfolded form and r_t is the total concentration of RNA.

For a simple unfolding reaction characterized by an equilibrium constant K, equation **1** can be written as

$$A(T) = \varepsilon_F r_t + \Delta\varepsilon K r_t/(1+K). \quad \mathbf{2}$$

Assuming an insignificant difference in heat capacity between the folded and unfolded states, the equilibrium constant K is

related to the enthalpy change ΔH and temperature T by the van't Hoff equation

$$K(T) = K_0\exp\{(\Delta H/R)(1/T_0 - 1/T)\} \quad \mathbf{3}$$

where T_0 is a reference temperature, which for a single transition can be taken as the melting temperature T_m, at which temperature $K_0 = 1$.

This provides the enthalpy change ΔH for the unfolding reaction. Since equations **2** and **3** did not fully account for the data, the possibility of sloping baselines was considered. Assuming essentially linear baselines, the absorption coefficients were treated as

$$\varepsilon(T) = \varepsilon_0 + a(T - T_0)$$

where a is the slope of the line. This formalism means that the hyperchromicity can be obtained at $T = T_m$ in a straightforward manner.

RESULTS

Initial characterization of the two SECIS

The purified RNA products were analyzed on analytical 15% acrylamide gels in the presence or absence of 7 M urea. A single band was observed with ethidium bromide staining for each product. The products were also characterized by analytical ultracentrifugation. Assuming a partial specific volume v of 0.53 ml/g (34), the TrxR SECIS gave an apparent molecular weight of 19 ± 1 kDa which is close to the chemical molecular weight of 16.5 kDa. The discrepancy between these two values reflects the uncertainty of the partial specific volume of the RNA oligonucleotides. In addition, no non-ideality or evidence of concentration-dependence was observed, indicating that the oligonucleotide was monomeric. In the case of GPX RNA, the apparent molecular weight, assuming again a partial specific volume v of 0.53 ml/g was 20.6 ± 1 kDa, in agreement with a chemical molecular weight of 19 kDa.

UV melting analysis of the two SECIS

Nucleic acids with secondary structures (even short stem–loop structures) show cooperative transitions that can be detected by the hyperchromicity associated with the disruption of the base pairing and unstacking (36). Under a given set of conditions, the melting temperature is characteristic of the number of base pairs involved in the cooperative transition, the fraction of GC base pairs and penalties for mismatches (28,37–39). Thus we performed UV melting studies on the two SECIS to assess their relative stability and to provide an indicator of the extent of secondary structures that are present.

Figure 1 shows the UV melting profile of the two SECIS under the same buffer conditions. The melting transition was fully reversible in each, and was independent of heating rate in the range 0.25–1°C/min. The TrxR SECIS showed a single cooperative transition centered at 60 ± 1°C. The fit to the absorbance profile assuming flat baselines was poor. However, by including linearly sloping baselines, a much better fit was obtained (Fig. 1A). The curves were highly reproducible, giving the average thermodynamic values shown in Table 1. The significant enthalpy change and hyperchromicity

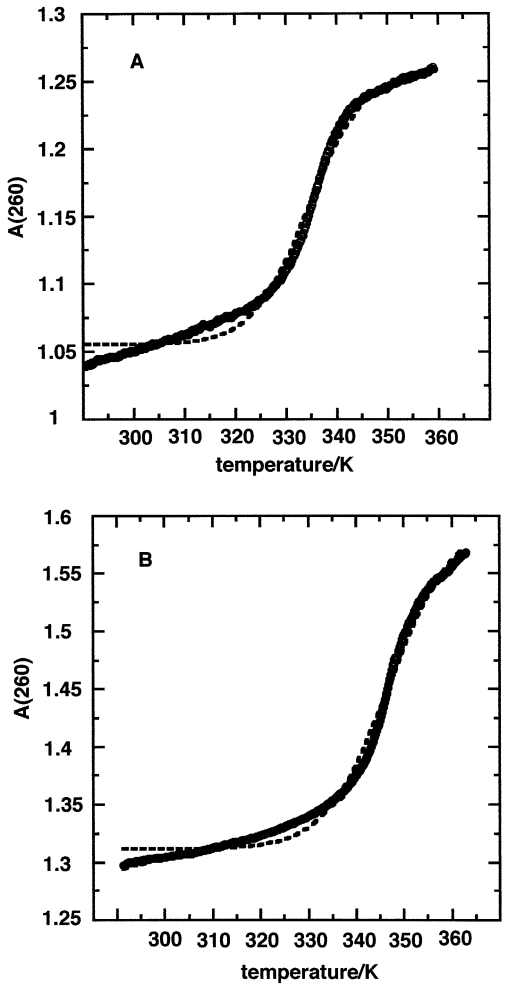

Figure 1. UV melting of SECIS. The heating rate was 0.25°C/min with a sampling interval of 0.06°. The circles are experimental points, the dashed line is a fit to a model without sloping baselines and the solid line is a fit to a model with sloping baselines. (**A**) TrxR SECIS, T_m = 334 K. (**B**) GPX SECIS, T_m = 347 K.

demonstrates that the molecule has a stable folded structure involving stacked base pairs. The magnitude of the enthalpy change is consistent with the presence of ~7–8 Watson–Crick base pairs assuming an average enthalpy change per base pair of ~37 kJ/mol (39). This calculation may underestimate the number of pairs involved, as the structure may contain mismatches and internal loops which generally decrease the observed total enthalpy change (39).

The GPX SECIS showed a single cooperative transition at ~74°C (Fig. 1B). As for TrxR, the transition was reversible and independent of concentration and heating rate. The derived thermodynamic parameters are given in Table 1 and are consistent with the presence of ~8–10 stable base pairs in the folded GPX. Thus the GPX SECIS is more stable than the TrxR SECIS as shown by the higher T_m value and the free energy difference at 37°C.

^{1}H NMR spectroscopy of TrxR SECIS

The chemical shifts of imino protons are good indicators of the nature of hydrogen bonding interactions, and therefore

Table 1. Thermodynamics of SECIS melting

Molecule	$\Delta A/A_0$	T_m (K)	ΔH (kJ/mol)	ΔG (37°C) (kJ/mol)
TrxR	0.14 ± 0.02	333.5 ± 0.9	282 ± 28	19.9 ± 1.6
GPX	0.17 ± 0.02	347.2 ± 0.4	274 ± 32	28.5 ± 1.7

secondary structure. Figure 2 shows common base-pairing schemes found in RNA and proposed for the SECIS studied here. The chemical shift ranges of the imino protons observed for each structure is also shown.

The low-field region of a 1D NMR spectrum of the TrxR SECIS recorded at 800 MHz is shown in Figure 3A. At 10°C and pH 6, the spectrum was well resolved with mainly sharp resonances between 11 and 14.5 p.p.m., which are typical of Watson–Crick or G.U wobble base pairing. The TrxR SECIS showed 16 resonances in this region. Between 9 and 11 p.p.m., there are six to seven broad peaks that became more intense at 5°C (data not shown), indicating exchange broadening at higher temperature and absence of stable H-bonding. This is typical of either unpaired imino protons or imino protons H-bonded to oxygen. Hence a total of 22–23 imino protons were observed, which is very close to the maximum possible number of imino protons in the 51 nt SECIS (i.e. 26).

To discriminate between G.C and U.A base pairs, and other base pairs such as the G.U wobble pair, additional information is needed. The distinction can be made based on NOE intensities corresponding to characteristic short proton–proton distances. In a G.C base pair, NOEs are expected between the GN1H (11–13 p.p.m.) and N4 (6–8.5 p.p.m.) amino protons of the complementary C (Fig. 2). In an A.U base pair, the NOE is between the UN3H (13–15 p.p.m.) and the C2H (7–8 p.p.m.) of the complementary A. Figure 3B shows the imino-aromatic region of the NOESY spectrum of the TrxR SECIS. The two downfield imino proton resonances at 14.4 and 13.6 p.p.m. showed NOEs to sharp resonances in the 7–8 p.p.m. region (Ade C2Hs) and can be assigned to UN3H involved in A.U base pairs. A broader resonance at 14.05 p.p.m. also showed a single strong NOE to a resonance in the 7–8 p.p.m. region, indicating the presence of a third less stable AU base pair. Resonances in the region 11.5–13.4 p.p.m. showed strong NOEs to pairs of resonances at 8–8.5 p.p.m. and 6.5–7 p.p.m. characteristic of the non-equivalent amino protons in CG base pairs. These amino protons also gave intense mutual NOEs (not shown). Thus it is possible to count the number of stable GC (7) and AU Watson–Crick base pairs (3). Sharp imino proton resonances between 11 and 12 p.p.m. not accounted for by Watson–Crick base pairs are likely to be non-canonical base pairs such as G.U, which usually do not show NOEs to pairs of amino protons.

Figure 3C shows two strong imino–imino proton cross-peaks in the spectral region 11–12 p.p.m.. These protons, not accounted for by G.C or A.U base pairs, showed NOEs only to amino protons in the 6.5–7 p.p.m. range (Fig. 3B). This signature is characteristic of the G.U wobble base pair. Hence the TrxR SECIS contains at least seven G.C, three A.U and two G.U base pairs, accounting for 14 imino protons. The two remaining imino protons, at 12.0 and 13.1 p.p.m. (Fig.3A),

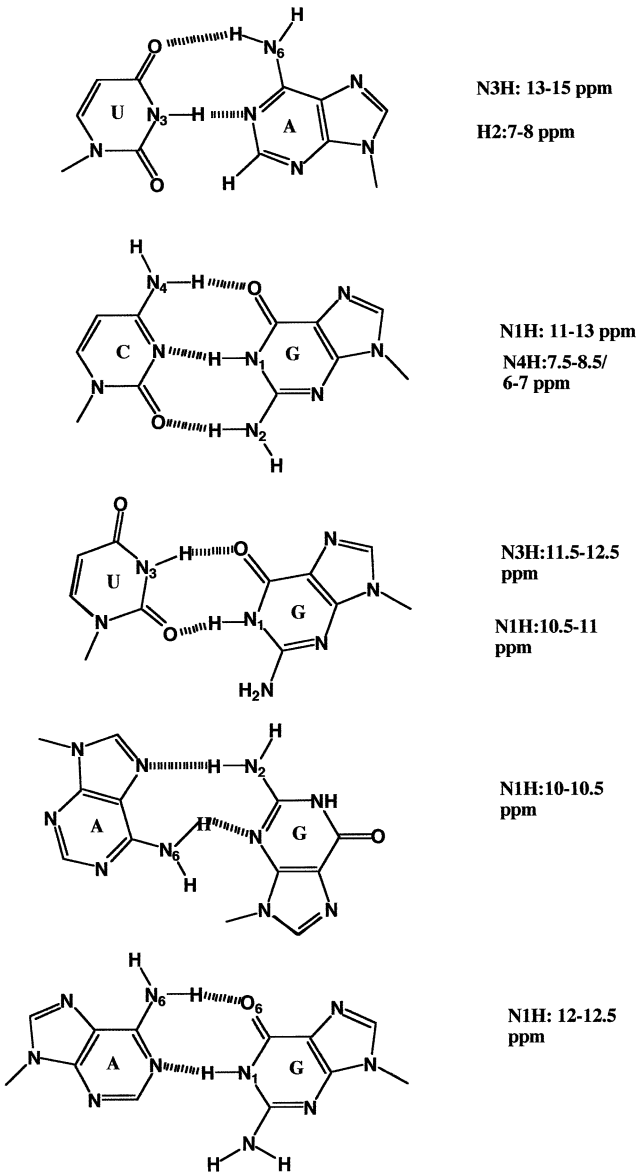

Figure 2. Base pairing and NMR chemical shifts for RNA. From top to bottom: Watson–Crick CG; Watson–Crick UA; wobble G.U; sheared A.G; wobble G.A. Chemical shift ranges were obtained from the literature (28–30,39,40).

were broad, suggesting exchange, and showed no unambiguous NOEs. However, their chemical shifts are not consistent with free unpaired G or U (usually found at ~10 p.p.m.), suggesting that they may participate in weakly H-bonded structures such as wobble G.A base pairs but not the sheared G.A base pair (cf. Fig. 2).

Sequential assignments were obtained using the NOEs between the imino protons of neighboring base pairs, as shown in Figure 3C. For example, imino resonances at 12.4, 13.3, 13.6, 11.85, 14.4, 11.95 and 11.44 are sequentially connected by NOEs (Fig. 3C) and therefore define a mini-helix of sequence GGUGU(G.U) which can be fitted in the lower stem of the predicted TrxR SECIS structure (Fig. 4B). Further connectivities from the helix were either not present or ambiguous. However, NMR data showed that the first G.C

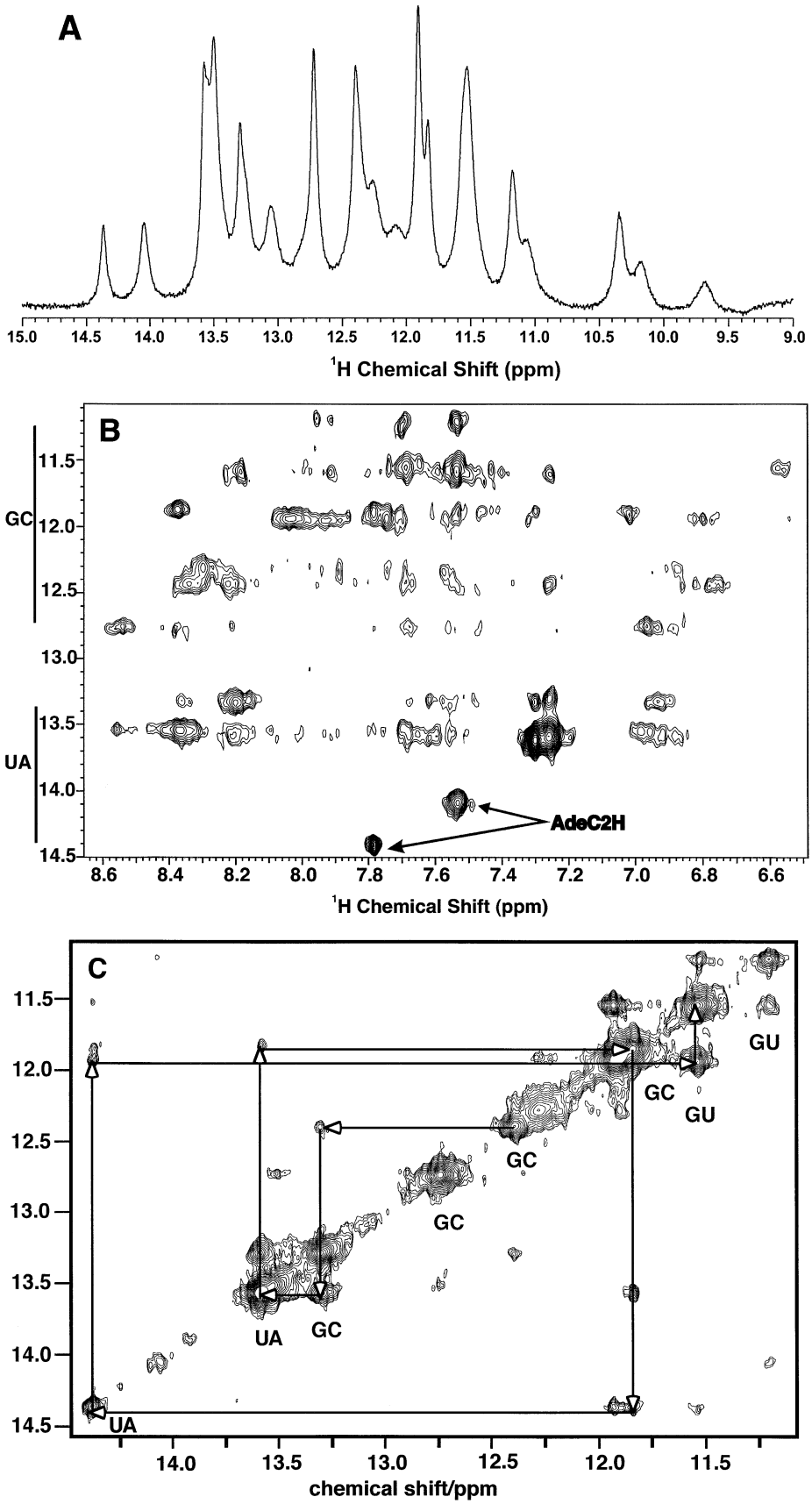

Figure 3. NMR spectra of TrxR SECIS at 18.8 T. Spectra were recorded at 5°C. (**A**) 1D spectrum of the low field region. (**B**) The imino-amino region of the NOESY spectrum recorded with a mixing time of 200 ms. (**C**) The imino region of the same NOESY spectrum as in (B).

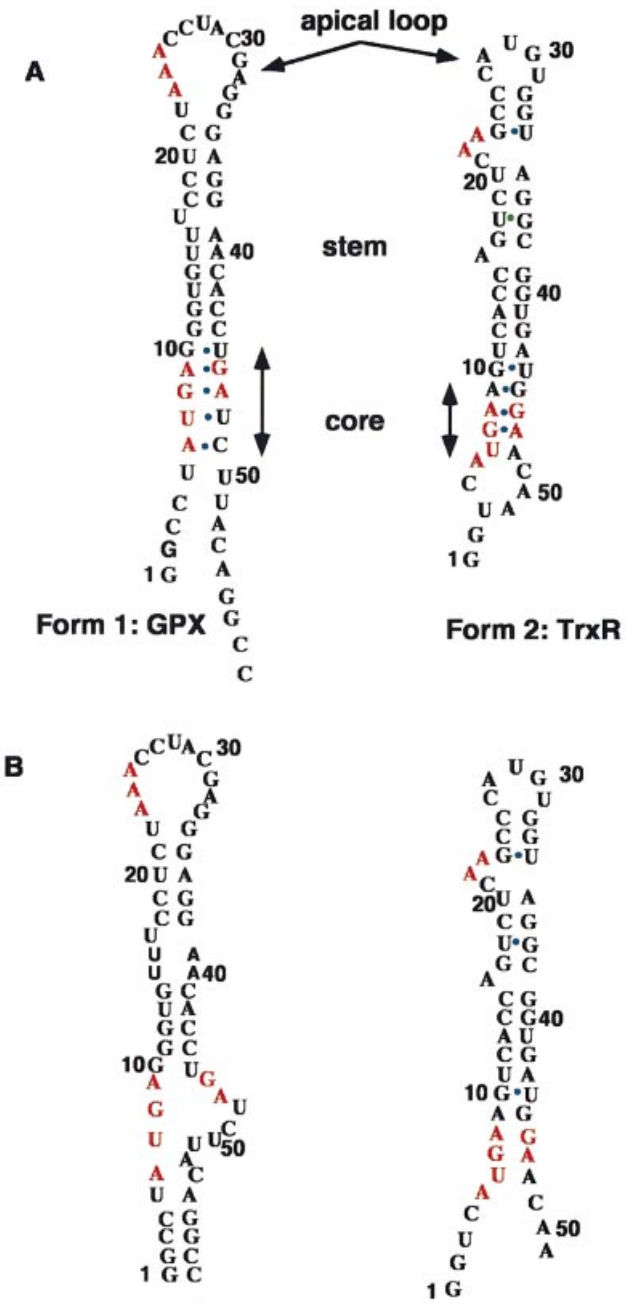

Figure 4. Predicted and NMR-based secondary structure of GPX and TrxR SECIS. The conserved sequences are shown in red. Mismatches are designated with a blue dot. The predicted structures are based on the proposals of Grundner-Culemann *et al*. (21) who defined the general structures and denoted two forms, 1 and 2, of which the GPX and TrxR SECIS are representative members. (**A**) Predicted secondary structures; (**B**) NMR-based secondary structures.

base pair of this stretch was stable, as was the terminal G.U wobble base pair. This indicates that additional bases or base pairs may be stacked at each side of the helix, which would reduce the end fraying usually responsible for ready chemical exchange. Within the context of the model, this would indicate that the bases prior to G10.U44 interact with the helix, and also that the A16 base may bulge out to allow stacking of the two helices at the C15–G39/G17–C38 junction. The NOESY

Table 2. Chemical shifts of imino protons in TrxR SECIS

Base	δ_{HN} (p.p.m.)	Type	AH2	CN4H
G10	11.54	GU		
U44	11.93	GU		
U11A43	14.4	AU	7.70	
C12G42	11.85	GC		8.45, 7.15
A13U41	13.6	AU	7.31	
C14G40	13.3	GC		8.23, 7.00
C15G39	12.4	GC		8.35, 6.85
G24U34	11.2, 11.55 G.U			
U20A35	14.05	AU	7.54	
C25G33	12.7[a]	GC		8.60, 7.05
C26G32	13.5[a]	GC		8.4, 7.02
G	12.3	GC		8.3, 6.95
G	11.9	GC		8.1, 6.85
6–7 G,U	10–12 unpaired			

[a]These shifts may be reversed.

spectrum also showed an isolated sequential correlation between two other G.C base pairs, plausibly C25–G33 and C26–G32. Helices that are 2 bp long are not stable and, for the reasons described, we expect the two G.C base pairs above to have bases stacked at either end. Finally, the A.U base pair at 14.05 p.p.m. showed NOEs to a G.U base pair at 11.2/11.5 p.p.m.. The only possible match in the predicted secondary structure is found if U20–A35 stacks against G24–U34. This would indicate that the two conserved adenines are bulging out from the stem.

Excluding possible sheared G.A mismatches (cf. 26) in the lower stem, the model shown in Figure 4 predicts five unpaired NH resonances. However, the NMR spectrum showed evidence for about six or seven such protons, suggesting that there are additional unpaired G and U residues, such as the U at the bottom of the core sequence (Fig. 4B).

Therefore we were able to sequentially assign the H-bonded residues involved in secondary structure elements and the number of residues that were not involved in secondary structures, such as in the apical loop region. The assignments are given in Table 2, and they provide direct evidence for the formation of two of the predicted helices and indirect evidence for the formation of the third as well as for interhelical base stacking and for splaying out of two functionally relevant adenines.

Further information is provided by the TOCSY spectrum in H_2O and TOCSY and NOESY spectra in D_2O. In the TOCSY experiment, the expected number of H6–H5 cross-peaks for U and C was observed (see Supplementary Material). However, their intensity was variable, which may reflect the presence of unpaired U and C (see above). Only a few H1′–H2′ cross-peaks were visible (not shown) which suggests that a small number of residues have C2′-endo character and are either very mobile or are involved in non-canonical structures, probably in the apical loop and unpaired extremes of the stem. This result also indicates that the majority of the sugars of TrxR SECIS are essentially C3′-endo and therefore engaged in canonical A-form RNA helices.

^{1}H NMR spectroscopy of GPX SECIS

Figure 5A shows the 1D NMR spectrum of GPX SECIS in H_2O at 5°C. There were two main groups of resonances: sharp

peaks between 12 and 14 p.p.m., which arise from stably H-bonded imino protons, and much broader resonances between 10 and 12 p.p.m. that correspond to weakly or non-H-bonded imino protons. Integration of these peaks revealed that there were 15–16 imino resonances in the 12–14 p.p.m. region, and six to seven between 10 and 12 p.p.m., for a total of 21–23 imino protons. However, the number of proton resonances in the 10–12 p.p.m. region is probably underestimated as the degree of exchange with solvent was substantial. The maximum number of NH protons expected for the GPX sample is 29, which suggests that some resonances were too exchange-broadened to be observed. To a first approximation, the GPX SECIS contains only ~15–16 stable base pairs.

Figure 5B shows the imino-aromatic region of the NOESY spectrum of the GPX SECIS. This shows that there were only two stable A.U base pairs, and nine or ten stable G.C base pairs. However, some of the GN1H resonances were strongly overlapped. This is evident in the 1D spectrum (Fig. 5A), where the peaks at 12.20, 12.64 and 13.25 p.p.m. integrated to twice the area of the well separated UN3H resonances at 13.8 and 13.95 p.p.m. which each comprised a single proton. These two U resonances showed two NOEs to GC base pairs (Fig. 5C), and are useful anchor points for the assignment of two short helical stretches. One helix can be traced by starting at the single G.C peak at 12.5 p.p.m., with imino-proton cross-peak correlations 12.5–13.8–13.1–12.62 p.p.m.. This sequence is GUGG. Similarly, a second sequence is 12.62–13.95–12.2–12.62 p.p.m., which also corresponds to GUGG. However, because magnetization can be transferred between imino protons that are 2 bp apart by spin diffusion, the overlap of the GN1H resonances at 12.2 and 12.64 p.p.m. makes this assignment ambiguous. Thus an alternative possibility is that the cross-peak between 12.2 and 13.95 is a long-range NOE between i and $i + 2$ base pairs and the direct correlations for this second stem (13.95–12.62–12.2) define a UGG sequence. Although both cross-peaks remained at a shorter mixing time (100 ms), the direct pathway is more likely (as indicated in Table 3). However, this still does not formally rule out spin diffusion in such a large molecule. A third set of correlations (12.8–13.3–13.42) defined a third stem that is unambiguously a GGG helix. The first two sequences can be fitted in the helical regions of the RNA stem as predicted by the model of Grundner-Culemann *et al*. (21) (Fig. 4A). The GGG sequence is unique and must arise from the bottom stem below the core, i.e. involving G2C57, C3G56, C4G55 (Fig. 4B). The two GUGG stretches (or GUGG and GGU, if spin diffusion is responsible for the 12.2–13.95 cross-peak described above) would correspond to the mini-helices prior to the apical loop. The assignments of the exchangeable protons for GPX SECIS are given in Table 3.

No correlation was visible between these two mini-helices, indicating that the bulging U17 disrupts the integrity of the stems. However, the fact that G38 N3H was visible indicates that some base stacking was present. We found no evidence of stable G.U base pairs, such as the predicted G10U45 (Fig. 4A). Furthermore, as we observed only two stable A.U base pairs, other potential A.U base pairs must be unstable, such as those at the end of the helical regions.

The broad resonances in the 10–12 p.p.m. region probably arose primarily from unpaired G and U imino protons, of which some are in the apical loop and others are in the conserved AUGA/GA motif in the core. As the 12–14 p.p.m. region only accounted for 15–16 protons and some H-bonded GN1H were not accounted for in the mini-helices assigned above, there may be other non-canonical base pairs present. The absence of correlations and their exchange with solvent implies that they were at best weakly H-bonded. The data are not consistent with the sheared G.A mismatch motif (cf. Fig. 2) in the core region, but other G.A base pairing cannot be ruled out. For example, the wobble G.A base pair, which has the GN1H at ~12 p.p.m., is a possibility (39,40). However, the predicted internal loop formed by these tandem wobble G.A base pairs would probably be very unstable.

^{31}P NMR of TrxR and GPX SECIS

In both normal DNA and RNA duplexes, the range of phosphodiester torsion angles is relatively uniform, giving rise to a small dispersion in chemical shifts, typically ~0.6–1 p.p.m. (39,41). However, mismatches and tight loops can cause substantial changes in chemical shifts of backbone phosphates (30,37,39,41,42). The ^{31}P NMR spectra of TrxR and GPX SECIS (see Supplementary Material) were largely unresolved envelopes with a width of ~1.5 p.p.m. and ~1.3 p.p.m. respectively (including the downfield 5′ phosphate). This appears to rule out the changes in phosphodiester backbone torsions associated with tight turns and G.A sheared base pairs flanked by G.U base pairs such as the ones that have been suggested for the SECIS core region (26), but is consistent with other non-standard base-pairing schemes such as G.A wobble base pairs flanked by G.U base pairs (39).

DISCUSSION

Despite the large sizes of TrxR and GPX SECIS, high magnetic field strength NMR has enabled us to determine the secondary structure of these two RNAs. Our results lend support to the notion that there are two forms of SECIS (21) that differ in the size of the apical loop and the relative location of the conserved AA sequence (Fig. 4B). The NMR data also indicate the formation of one long and two short helical stems in the TrxR SECIS and of three short helical stems in the GPX SECIS. The UV melting data showed that the TrxR SECIS underwent a single melting transition whose T_m and enthalpy change (Table 1) are consistent with that expected for the melting of a duplex region of the sequence GGUGU(G:U). The very short helices comprising only 2 bp are likely to melt at lower temperature with low cooperativity. The GPX SECIS also melted in a single cooperative transition at a significantly higher temperature than TrxR SECIS, and with a larger hyperchromicity. This suggests the presence of more base pairs and/or greater GC content in the cooperative unit. The sloping baseline in both cases was substantial, indicating the possibility of significant fraying of weakly stable structures at temperatures well below the T_m such as from the ends (TrxR), the apical loop or the core sequences.

In both SECIS models (Fig. 4A), the conserved GA sequences would be closed by a G.U base pair stacking on the neighboring G.A base pairs. Although in the TrxR SECIS a stable G.U base pair did form at the boundary between the lower stem and the core, other features of the model are difficult to reconcile with our data. The TrxR predicted core

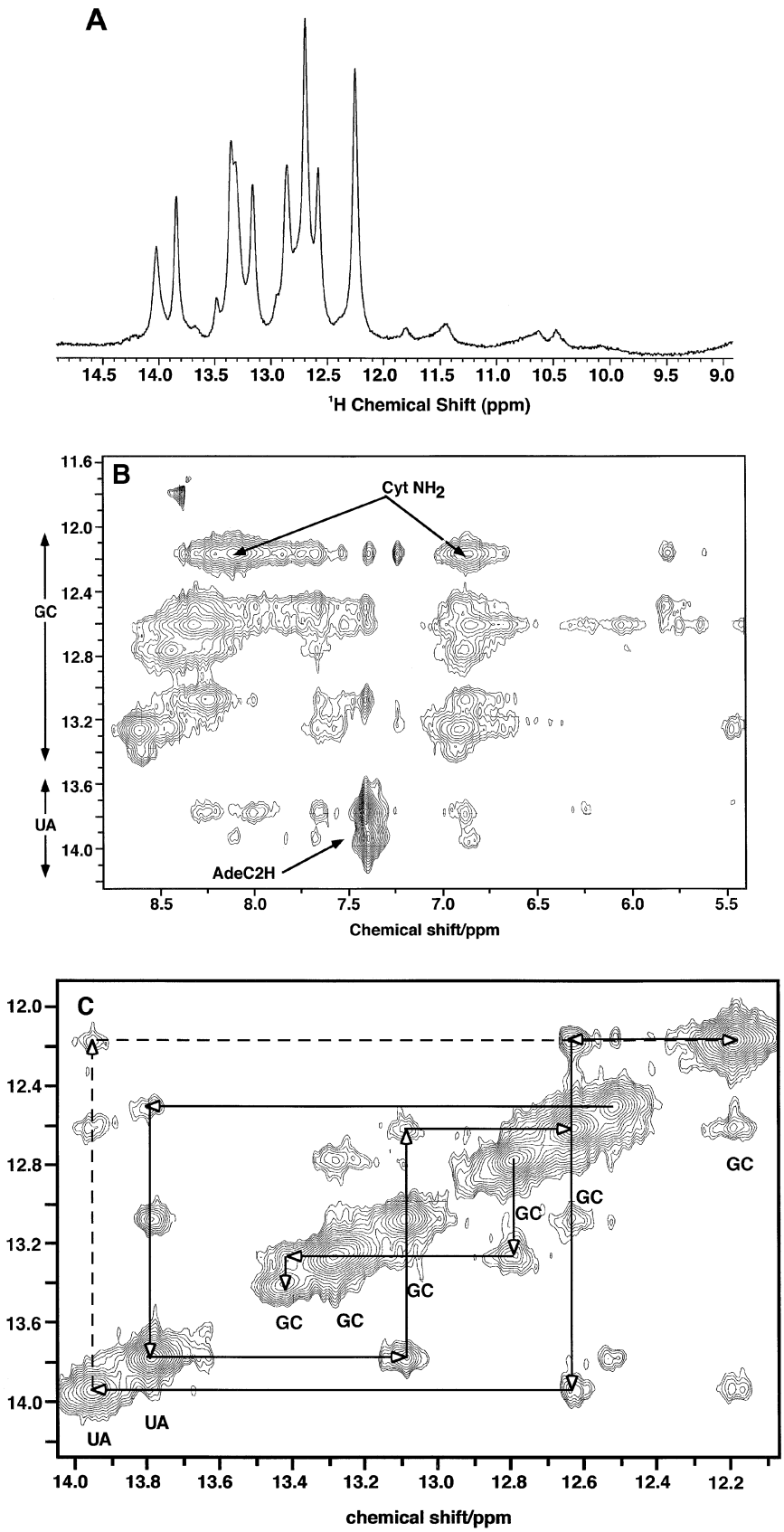

Figure 5. NMR spectra of GPX SECIS. Spectra were recorded at 5°C. (**A**) 1D spectrum at 18.8 T. (**B**) The imino-amino region of the NOESY spectrum recorded with a mixing time of 200 ms at 14.1 T. (**C**) The imino region of the same NOESY spectrum as in (B). The dashed line shows the potential spin diffusion pathway as described in the text.

Table 3. Chemical shifts of imino protons in GPX SECIS

Base	δ_{HN} (p.p.m.)	Type	AH2	CN4H
G2C57	13.42	GC		8.63, 6.95
G56C3	13.29	GC		8.65, 6.93
G55C4	12.78	GC		8.45, 6.88
G11C44	12.18	GC		8.13, 6.85
G12C33	12.62	GC		8.31, 6.85
U13A42	13.95	AU	7.40	
G14C41	12.19	GC		8.13, 6.85
G38C18	12.64	GC		8.35, 6.82
G37C19	13.09	GC		8.3, 6.88
U20A36	13.79	AU	7.39	
G35C21	12.53	GC		
G	13.29	GC		
G	12.88	GX		
>7 G,U	10–12 unpaired			

sequence will require a novel structure with three successive G.A base pairs (denoted by green dots in Fig. 4A) which were suggested to be of the sheared type (Fig. 2) (26) as found in GNRA loops (43) and tandem G.A mismatches (28,38,39). We have found no evidence for this kind of base pair. The presence of GN1H resonances in the 12.5–13 p.p.m. region (Fig. 3A) other than those involved in G.C base pairs is consistent with the presence of wobble G.A base pairs. Furthermore, sheared tandem G.A mismatches closed by a G.U base pair typically have a large ^{31}P NMR shift dispersion (39), whereas the ^{31}P NMR dispersion in the TrxR SECIS was moderate. This is again consistent with wobble tandem G.A base pairs closed by a G.U base pair.

For GPX SECIS, the NMR data are compatible with those obtained from nuclease digestion and chemical modification experiments (26). The latter data provide evidence for the presence of single strands in the large apical loop and structured regions in the stem and core sequences, as inferred from accessibility to base-modifying agents and nucleases. Also consistent with the NMR data is the accessibility of several of the base pairs, especially those adjacent to the apical loop and possible mismatches in the AUGA region.

In addition, the NMR data for GPX SECIS are consistent with one of the low energy GPX structures predicted by M-fold, i.e. a GC-rich stem followed by the core sequence which forms an internal loop and then by another base-paired stem (cf. Fig. 4B). Correlations between the RNA H-bonded imino protons confirm the presence of the GC-rich stem and the clear delineation of the core region. Although formation of the two AG base pairs is possible (Fig. 4A), the resulting lengthy run of unstable mismatches would be thermodynamically unstable.

The flexible apical loop and core structures of the mammalian SECIS are expected to be involved in the binding to proteins participating in the translational machinery. For example, the conserved AUGA sequence was reported to be required by SBP-2 binding. Such interaction differs from that present in prokaryotes, where a structured apical tetraloop is essential to the SelB recognition while no conserved AA or conserved G.A mismatch structure is involved (18). Moreover, the mammalian SECIS binding protein SBP-2 has little sequence homology with SelB or EFTu, but is predicted to have an RNA binding domain (44–46).

The differential interaction of SBP-2 with various SECIS RNA core regions (44–48) could provide a functional explanation for the ranking of different RNAs in the SEC incorporation process. However, there remain important issues concerning the structural and sequence features present in the apical part and the length of the stem in the SECIS. The optimal length of the SECIS stem has been shown to vary between 8 and 12 nt (21). In some SECIS, the stem is interrupted either by a bulging base (as is the case for both TrxR and GPX) or by a mismatch. Such a feature could also influence interactions with translational proteins and therefore SeC insertion efficiency.

Clearly, a full structure determination with isotopically labeled material is needed to obtain a more detailed structural insight into the precise roles of these elements.

SUPPLEMENTARY MATERIAL

Supplementary Material is available at NAR Online.

ACKNOWLEDGEMENTS

We thank Dr G. Kelley and Dr T.A. Frenkiel for advice and Dr A. Pastore for encouragement. NMR spectra were recorded at the MRC NMR Centre, Mill Hill, and at the James Graham Brown NMR Center, University of Louisville. This work was supported in part by the MRC (UK), the Brown Foundation (USA) and NCI grant 1R01CA101199. A.N.L. gratefully acknowledges the Kentucky Challenge for Excellence Program.

REFERENCES

1. Stadtman,T.C. (1974) Selenium biochemistry. *Science*, **183**, 915–921.
2. Powis,G., Oblong,J.E., Gasdaska,P.Y., Berggren,M., Hill,S.R. and Kirkpatrick,D.L. (1994) The thioredoxin/thioredoxin reductase redox system and control of cell growth. *Oncol. Res.*, **6**, 539–544.
3. Arner,E.S.J. and Holmgren,A. (2000) Physiological functions of thioredoxin and thioredoxin reductase. *Eur. J. Biochem.* **267**, 6102–6109.
4. Levander,O.A. and Beck,M.A. (1997) Interacting nutritional and infectious etiologies of Keshan disease—insights from Coxsackie virus B-induced myocarditis in mice deficient in selenium or vitamin E. *Biol. Trace Elem. Res.*, **56**, 5–21.
5. Hoffman,D.J. (2002) Role of selenium toxicity and oxidative stress in aquatic birds. *Aquatic Toxicol.*, **57**, 11–26.
6. Schrauzer,G.N. (2000) Selenomethionine: a review of its nutritional significance, metabolism and toxicity. *J. Nutr.*, **130**, 1653–1656.
7. Ganther,H.E. (1999) Selenium metabolism, selenoproteins and mechanisms of cancer prevention: complexities with thioredoxin reductase. *Carcinogenesis*, **20**, 1657–1666.
8. Ma,X.R., Karra,S., Guo,W., Lindner,D.J., Hu,J.D., Angell,J.E., Hofmann,E.R., Reddy,S.P.M. and Kalvakolanu,D.V. (2001) Regulation of interferon and retinoic acid-induced cell death activation through thioredoxin reductase. *J. Biol. Chem.*, **276**, 24843–24854.
9. Ma,X.R., Karra,S., Lindner,D.J., Hu,J.B., Reddy,S.P.M., Kimchi,A., Yodoi,J. and Kalvakolanu,D.D. (2001) Thioredoxin participates in a cell death pathway induced by interferon and retinoid combination. *Oncogene*, **20**, 3703–3715.
10. Tamura,T. and Stadtman,T.C. (1996) A new selenoprotein from human lung adenocarcinoma cells: purification, properties and thioredoxin reductase activity. *Proc. Natl Acad. Sci. USA*, **93**, 1006–1011.
11. Xanthoudakis,S. and Curran,T. (1992) Identification and characterization of Ref-1, a nuclear protein that facilitates Ap-1 DNA-binding activity. *EMBO J.*, **11**, 653–665.
12. Ueno,M., Masutani,H., Arai,R.J., Yamauchi,A., Hirota,K., Sakai,T., Inamoto,T., Yamaoka,Y., Yodoi,J. and Nikaido,T. (1999) Thioredoxin-

dependent redox regulation of p53-mediated p21 activation. *J. Biol. Chem.*, **274**, 35809–35815.
13. Karimpour,S., Lou,J.Y., Lin,L.L., Rene,L.M., Lagunas,L., Ma,X.R., Karra,S., Bradbury,C.M., Markovina,S., Goswami,P. *et al*. (2002) Thioredoxin reductase regulates AP-1 activity as well as thioredoxin nuclear localization via active cysteines in response to ionizing radiation. *Oncogene*, **21**, 6317–6327.
14. Saitoh,M., Nishitoh,H., Fujii,M., Takeda,K., Tobiume,K., Sawada,Y., Kawabata,M., Miyazono,K. and Ichijo,H. (1998) Mammalian thioredoxin is a direct inhibitor of apoptosis signal-regulating kinase (ASK) 1. *EMBO J.*, **17**, 2596–2606.
15. Erden-Inal,M., Sunal,E. and Kanbak,G. (2002) Age-related changes in the glutathione redox system. *Cell Biochem. Funct.*, **20**, 61–66.
16. Leinfelder,W., Stadtman,T.C. and Böck,A. (1989) Occurrence *in vivo* of selenocysteyl-tRNA(SERUCA) in *Escherichia coli*. Effect of sel mutations. *J. Biol. Chem.*, **264**, 9720–9723.
17. Forchhammer,K., Leinfelder,W. and Böck,A. (1989) Identification of a novel translation factor necessary for the incorporation of selenocysteine into protein. *Nature*, **342**, 453–456.
18. Fourmy,D., Guittet,E. and Yoshizawa,S. (2002) Structure of prokaryotic SECIS mRNA hairpin and its interaction with elongation factor SelB. *J. Mol. Biol.*, **324**, 137–150.
19. Zinoni,F., Heider,J. and Böck,A. (1990) Features of the formate dehydrogenase mRNA necessary for decoding of the UGA codon as selenocysteine. *Proc. Natl Acad. Sci. USA*, **87**, 4660–4664.
20. Berry,M.J., Tujebajeva,R.M., Copeland,P.R., Xu,X.M., Carlson,B.A., Martin,G.W., Low,S.C., Mansell,J.B., Grundner-Culemann,E., Harney,J.W. *et al*. (2001) Selenocysteine incorporation directed from the 3′UTR: characterization of eukaryotic EFsec and mechanistic implications. *Biofactors*, **14**, 14–24.
21. Grundner-Culemann,E., Martin,G.W., Harney,J.W. and Berry,M.J. (1999) Two distinct SECIS structures capable of directing selenocysteine incorporation in eukaryotes. *RNA*, **5**, 625–635.
22. Shen,Q., Wu,R., Leonard,J.L. and Neuberger,P.E. (1998) Identification and molecular cloning of a human selenocysteine insertion sequence-binding protein. *J. Biol. Chem.*, **273**, 5443–5446.
23. Wu,R., Shen,Q. and Neuberger,P.E. (2000) Recognition and binding of the human selenocysteine insertion sequence by nucleolin. *J. Cell Biol.*, **77**, 507–516.
24. Shen,Q., Leonard,J.L. and Neuberger,P.E. (1995) Structure and function of the selenium translation element in the 3′-untranslated region of human cellular glutathione peroxidase. *RNA*, **1**, 519–525.
25. Gasdaska,J.R., Harney,J.W., Gasdaska,P.Y., Powis,G. and Berry,M.J. (1999) Regulation of human thioredoxin reductase expression and activity by 3′-untranslated region selenocysteine insertion sequence and mRNA instability elements. *J. Biol. Chem.*, **274**, 25379–25385.
26. Walczak,R., Westhof,E., Carbon, P and Krol,A. (1996) A novel RNA structural motif in the selenocysteine insertion element of eukaryotic selenoprotein mRNAs. *RNA*, **2**, 367–379.
27. Kryukov,G.V., Castellano,S., Novoselov,S.V., Lobanov,A.V., Zehtab,O., Guigo,R. and Gladyshev,V.N. (2003) Characterization of mammalian selenoproteomes. *Science*, **300**, 1439–1443.
28. Wu,M., McDowell,J.A. and Turner,D.H. (1995) A periodic table of symmetrical tandem mismatches in RNA. *Biochemistry*, **34**, 3204–3211.
29. Ramos,A. and Varani,G. (1997) Structure of the acceptor stem of *Escherichia coli* tRNA(Ala): role of the G3 center dot U70 base pair in synthetase recognition. *Nucleic Acids Res.*, **25**, 2083–2090.
30. Lane,A.N., Ebel,S. and Brown,T. (1994) Properties of multiple G.A mismatches in stable oligonucleotide duplexes. *Eur. J. Biochem.*, **220**, 717–727.
31. Batey,R.T., Inada,M., Kujawinski,E., Puglisi,J.D. and Williamson,J.R. (1992) Preparation of isotopically labeled ribonucleotides for multidimensional NMR-spectroscopy of RNA. *Nucleic Acids Res.*, **20**, 4515–4523.
32. Nikonowicz,E.P., Sirr,A., Legault,P., Jucker,F.M., Baer,L.M. and Pardi,A. (1992) Preparation of C-13 and N-15 labeled RNAs for heteronuclear multidimensional NMR studies. *Nucleic Acids Res.*, **20**, 4507–4513.
33. Ramos,A. (1999) NMR studies of specificity in RNA protein recognition. PhD Thesis, University of Cambridge.
34. Bonifacio,G., Brown,T., Conn,G.L. and Lane,A.N. (1997) Comparison of the electrophoretic and hydrodynamic properties of DNA and RNA oligonucleotides. *Biophys. J.*, **73**, 1532–1538.
35. Piotto,M., Saudek,V. and Sklenar,V. (1992) Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J. Biomol. NMR*, **2**, 661–665.
36. Serra,M.J., Lyttle,M.H., Axenson,T.J., Schadt,C.A. and Turner,D.H. (1993) RNA hairpin loop stability depends on closing base-pair. *Nucleic Acids Res.*, **21**, 3845–3849.
37. Ebel,S., Brown,T. and Lane,A.N. (1994) Thermodynamic stability and solution conformation of tandem G.A mismatches in RNA and RNA.DNA hybrid duplexes. *Eur. J. Biochem.*, **220**, 703–715.
38. Walter,A.E., Wu,M. and Turner,D.H. (1994) The stability and structure of tandem GA mismatches in RNA depend on closing base-pairs. *Biochemistry*, **33**, 11349–11354.
39. Schroeder,S.J. and Turner,D.H. (2001) Thermodynamic stabilities of internal loops with GU closing pairs in RNA. *Biochemistry*, **40**, 11509–11517.
40. Lane,A.N., Jenkins,T.C., Brown,D. and Brown,T. (1991) Thermodynamics, kinetics and conformations in the syn-anti equilibrium in an A·G mismatch determined by solution NMR. *Biochem. J.*, **279**, 269–281
41. Legault,P. and Pardi,A. (1994) P-31 chemical-shift as a probe of structural motifs in RNA. *J. Magn. Reson. B*, **103**, 82–86.
42. Gorenstein,D.G. (1984) *Phosphorus P-31 NMR: Principles and Applications*. Academic Press, Orlando, FL.
43. Heus,H.A. and Pardi,A. (1991) Structural features that give rise to the unusual stability of RNA hairpins containing GNRA loops. *Science*, **253**, 191–194.
44. Low,S.C., Grundner-Culemann,E., Harney,J.W. and Berry,M.J. (2000) SECIS-SBP2 interactions dictate selenocysteine incorporation efficiency and selenoprotein hierarchy. *EMBO J.*, **19**, 6882–6890.
45. Fletcher,J.E., Copeland,P.R., Driscoll,D.M. and Krol,A. (2001) The selenocysteine incorporation machinery: interactions between the SECIS RNA and the SECIS-binding protein SBP2. *RNA*, **10**, 1442–1453.
46. Copeland,P.R., Fletcher,J.E., Carlson,B.A., Hatfield,D.L. and Driscoll,D.M. (2000) A novel RNA binding protein, SBP2, is required for the translation of mammalian selenoprotein mRNAs. *EMBO J.*, **19**, 306–314.
47. Allmang,C., Carbon,P. and Krol,A. (2002) The SBP2 and 15.5 kD/Snu13p proteins share the same RNA binding domain: identification of SBP2 amino acids important to SECIS RNA binding. *RNA*, **8**, 1308–1318.
48. Zavacki,A.M., Mansell,J.B., Chung,M., Klimovitsky,B., Harney,J.W. and Berry,M.J. (2003) Coupled tRNASec-dependent assembly of the selenocysteine decoding apparatus. *Mol. Cell.*, **11**, 773–781.