Title
THE STABILITY OF RIBONUCLEIC ACID IN SOLUTION: MODEL CALCULATIONS

Author
Levine, Mark D.

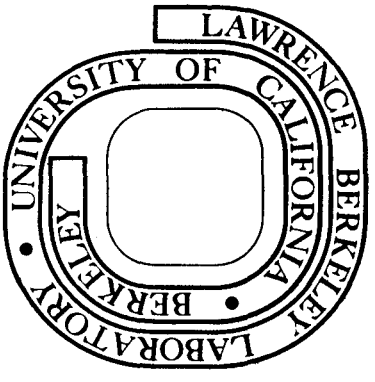
Publication Date
1974-11-01

THE STABILITY OF RIBONUCLEIC ACID IN SOLUTION: MODEL CALCULATIONS

Mark D. Levine
(Ph. D. thesis)

November 1974

Prepared for the U. S. Atomic Energy Commission
under Contract W-7405-ENG-48



LBL-3607
c.2

DISCLAIMER

TABLE OF CONTENTS

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

TABLE OF CONTENTS (Continued)

THE STABILITY OF RIBONUCLEIC ACID IN SOLUTION: MODEL CALCULATIONS

MARK D. LEVINE

ABSTRACT

Model calculations on oligoribonucleic acids were performed in order to gain a quantitative understanding of factors which contribute to the stability of double stranded RNA molecules in solution. The calculations were based on information obtained from melting curves of double stranded oligomers. The work evaluated the enthalpies and entropies of double stranded stacking interactions and of loop formation. Results of the analysis were used to develop a methodology for predicting the secondary structure of RNA molecules consisting of as many as 120 nucleic acid bases.

Chapter I summarizes the thermodynamic information previously obtained by workers who investigated RNA double strand formation. The chapter also discusses the relevance of the present work to an understanding of the biological function of RNA molecules.

Chapter II analyzes the melting behavior of molecules of the form $\begin{matrix} A_n U_n \\ A_n U_n \end{matrix}$, where n = 4, 5, 6, and 7. The analysis leads to the conclusion that it is necessary to include the free energy of the interaction between the ends of the RNA and the solvent in order to achieve satisfactory agreement between theory and experiment.

Chapter III provides an analysis of the following RNA molecules:

$$\begin{matrix} A_nGCU_n \\ U_nCGA_n \end{matrix} \ (n=2,3,4), \quad \begin{matrix} A_nCGU_n \\ U_nGCA_n \end{matrix} \ (n=2,3,4),$$

$$\begin{matrix} A_nCU_n \\ U_nGA_n \end{matrix} \ (n=2,3,4), \quad \begin{matrix} A_4G_2 \\ U_4C_2 \end{matrix}, \quad \begin{matrix} A_4G_3 \\ U_4C_3 \end{matrix}, \quad \begin{matrix} A_4GC \\ U_4CG \end{matrix}$$

The basic conclusions of the chapter are: (1) end effects must be included in order to achieve agreement between experimental and theoretical melting curves; (2) the magnitude of the end effect free energy is comparable for all of the RNA molecules studied; (3) sequence dependence of double stranded stacking interactions is marked if G-C base pairs are present and needs to be included in the calculation if the melting behavior of the double stranded RNA is to be satisfactorily characterized.

Chapter IV investigates the stability of intramolecular loop formation for molecules $A_6C_mU_6$, where m = 4,5,6,8. It is concluded that loop formation for these molecules involves a significant positive enthalpy change, probably associated with loop strain.

Chapter V proposes a methodology for predicting the secondary structure of an RNA molecule, based on a knowledge of its sequence and the thermodynamic analysis of the preceding chapters. The methodology is applied to a number of transfer and 5S RNA molecules.

ACKNOWLEDGEMENTS

I very gratefully acknowledge all those people whom I knew when I was a graduate student who were supportive of my work and ideas and became friends.

CHAPTER 1

INTRODUCTION

It is well known that the addition of heat to a solution of double stranded RNA molecules causes the strands to separate from one another. Because the single and double stranded species differentially absorb ultraviolet light, it is possible to "observe" the separation of the complementary strands. If a solution of double stranded RNA molecules in solution in a cuvette is irradiated with ultraviolet light, preferably at the absorption maximum of approximately 260 nm, and the temperature is gradually increased until and after the strands separate, the resulting information, absorption versus temperature, is known as a melting curve. It is the task of this work to unravel the wealth of information contained in the melting curves of a variety of small RNA molecules and thereby better understand the factors responsible for the stability of RNA double stranded helices in solution.

In the course of our investigation, we will be led from a traditional treatment of the melting phenomena to a somewhat different version of the theory; we will consider small double stranded RNA molecules as well as strands which fold back on themselves to form looped structures. We will be led to attempt a prediction of stabilities of RNA molecules not yet studied or even synthesized and to the determination of possible secondary structures of large and complicated RNA molecules.

A. Thermodynamics of RNA

Before delving into the many details of RNA melting curves, we note what is and is not known about RNA in solution which is relevant to understanding the melting and stability of the molecule. The following provides a summary of this information.

1. ΔH° and ΔS° of Base Pair Formation.

Since lone base pairs are not stable in water or salt solution, our knowledge of the enthalpy and entropy of base pairs derives from double helices composed of several or many base pairs. In the formation of a double helix, all base pairs except the first add to a preexisting helix. To the extent that the double helix is stabilized by interactions between adjacent base pairs stacked on top of each other (and known as double stranded stacking interactions), the free energy of the first base pair is different from that of the other base pairs. Because of this, the entropy and enthalpy corresponding to the addition of a base pair to a double helix are the quantities of greatest interest to us.

Determination of ΔH° and ΔS° for a variety of RNA molecules of differing sequence can give some insight into the forces which stabilize the double strand. At one time, many workers believed that the stability of base pairs came from the hydrogen bonds. It was argued that the increased stability of a G-C base pair over an A-U base pair was due to the additional hydrogen bond of the G-C pair. More recently it has been suggested that, while the hydrogen bonding is responsible for the specificity of base pairing, the stability of the structure comes primarily from

the double stranded stacking interactions. Several arguments supporting the latter hypothesis have been advanced: the energy decrease associated with the hydrogen bond between paired nucleic acid bases is small, since unpaired bases may also form hydrogen bonds with water (and thus the formation of base pairs represents the exchange of one set of hydrogen bonds for another); quantum mechanical calculations indicate that the interaction of the π electron systems of neighboring base pairs is favorable;[2-4] the stability of helical stacks of nucleic acid bases in single stranded RNA and DNA molecules suggests that the same forces stabilize the double helix.

These arguments notwithstanding, experimental evidence which supports one of the two alternative concepts is sparse. The degree of sequence dependence of the thermodynamic functions which we evaluate will help to resolve the matter. If, for example, the stability of a G-C base pair in the sequence $\frac{\text{AGAGA}}{\text{UCUCU}}$ is substantially different from that for such a pair in the sequence $\frac{\text{AAGGA}}{\text{UUCCU}}$, then it is a logical conclusion that the stacking interactions are important determinants of the stability of the base pair. If, on the other hand, the enthalpy and entropy of a base pair are generally sequence invariant, then this lends support to the notion that the hydrogen bond formation is the major stabilizing factor in the RNA double strand.

The thermodynamics of base pair formation in some nucleic acid polymers has been studied directly through use of micro-calorimetry. For the reaction polyA:polyU → polyA + polyU, Krakauer

and Sturtevant measure ΔH° = -7.4 kcal/mole of base pairs in 0.02 M Na^+ (melting temperature = 45°C) and ΔH° = -8.2 kcal/mole in 0.1 M Na^+ (T_m = 58°C).[5] ΔS° for an A-U base pair is 23.2 entropy units and 24.8 entropy units respectively. Neumann and Ackermann obtain values of ΔH° = -9.3 ± 0.5 kcal/mole andΔ S° = 25.2 ± 1.5 entropy units for this reaction extrapolated to a high melting temperature (and high salt concentration) for which polyA exhibits no single stranded stacking.[6] The uniform change of ΔH° with increasing salt may be due to one or both of the factors: (1) the stabilization of the polyelectrolyte backbone of double stranded RNA by counterions and (2) the increase of the temperature at which ΔH° is measured. (At higher temperatures, the single strand is less stacked and ΔH°, which measures the enthalpy difference between the single and double strand, would be expected to increase with temperature.) The relative contribution of the two factors to the measured increase in ΔH° with increasing salt concentration is not known. To the extent that the conformation of polyA:polyU is similar to that of the oligomers herein studied, the values of these thermodynamic parameters are transferable. It will be of interest to note if this is the case.

No calorimetric measurements have been performed on RNA polymers containing G-C base pairs. Kallenbach has noted that the melting temperature of RNA polymers increases regularly with increasing G-C content.[7] This observation will prove useful in our analysis. Klump and Ackermann have measured the heat capacity as a function of temperature for a series of DNA molecules.[8] Their

results suggest that the magnitude of the enthalpy per base pair increases from 7.7 to 8.5 kcal in going from 31% G-C to 72% G-C base pairs. The melting temperatures for RNA are much more sensitive to the fraction of G-C base pairs than for DNA, suggesting that G-C base pairs may have a greater effect on ΔH° for RNA than for DNA. In short, while some direct thermodynamic measurements have been made on RNA polymers, very little is known about the specific interactions which stabilize the double helix.

2. Initiation.

The formation of the first base pair in a double helix, which is called the initiation process, is shown schematically in the first row of Figure 2-1. Initiation involves the collision of complementary mucleic acid strands in such an orientation that the proximity of the strands is maintained long enough to allow base pairs to form. The initial base pair formed cannot participate in a double stranded stacking interaction, since there is no base pair adjacent to it. Therefore, the initiation step is assigned an equilibrium constant, κ, different from the other equilibrium constants for base pair formation.

Little direct information about κ is known. Studies on polynucleotide systems suffer from the limitation that initiation is but one of many steps involved in the formation of the double strand. As a result, estimates of κ for DNA (generally between 10^{-2} and 10^{-5} liter/mole)[9,10] are not very reliable. The few estimates of κ for oligoribonucleotide systems are within this range.[11-15] Because the magnitude of κ depends very much on the model used in the analysis and on the choice of ΔH° for the addition

of a base pair to helix, it is not possible to determine its absolute magnitude very accurately.

Nevertheless, several important questions about κ can be asked. It is of interest to know if κ varies with temperature; i.e., if there is an enthalpy associated with the initiation process. The answer to this question is relevant if one wishes to predict accurately the behavior of RNA molecules at very high temperatures. Does initiation occur preferentially at an A-U or G-C base pair? Although the answer to this question is of interest by itself, it relates also to the stability of RNA molecules. If initiation at a G-C base pair is, let us say, 2 kcal/mole more stable than initiation at an A-U pair, then this is a significant factor in the stability of double helices with one or more G-C base pairs.

3. Loops

When an RNA molecule forms intramolecular base pairs, a hairpin loop is generated. Mismatching of base pairs within a double stranded region results in interior loops. A bulge occurs when one strand is looped out within a double helical region of the nucleic acid. These three types of RNA loops are depicted in Figure 1-1.

A quantitative understanding of the influence of hairpin loops on RNA stability began with the work of Jacobsen and Stockmayer,[16] who estimated the free energy of RNA loops as a function of loop size using a simple hydrodynamic model of the molecule. Their theory related the free energy of loop closure to the probability of the two ends of an unformed loop coming into contact with each other. They predicted a positive free energy of loop closure which increases

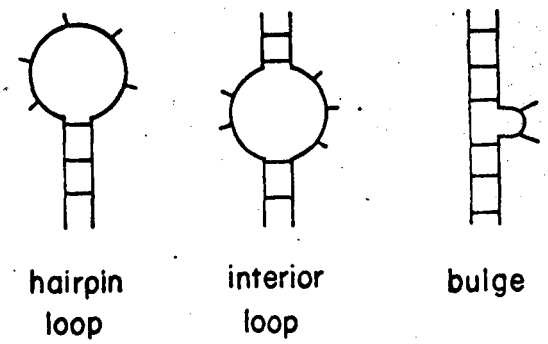


Figure 1-1

with increasing loop size.

Uhlenbeck et al., have studied $A_6C_mU_6$ loop molecules with 4, 5, 6, and 8 C residues in the loop and 6 A-U base pairs in the stem region.[17] They calculate the free energy of loop formation to be between 6.3 and 7.5 kcal/mole for these molecules at 25°C. Their analysis suggests that a positive enthalpy of greater than 20 kcal/mole is associated with loop formation. Gralla and Crothers have studied $A_4GC_mU_4$ loop molecules with 3, 4, and 5 C residues in the loop and 4 A-U base pairs and one G-C base pair in the stem region.[18] They calculate a free energy of greater than 8.4 kcal/mole for the loop with 3 C residues and between 4 and 6 kcal/mole for the two remaining loops. They assume a zero enthalpy for loop formation.

De Lisi and Crothers have done calculations on bulges and interior loops.[19] Gralla and Crothers have melted $A_4GC_mU_4$ molecules (m = 1-4) under conditions in which interior loops are

formed.[20] The analysis from both theory and experiment indicates that interior loops are more stable than hairpin loops by approximately 3 to 4 kcal/mole at 25°C. Bulge loops are thought also to be more stable than hairpin loops. The estimates on stabilities of interior and bulge loops is summarized by Tinoco et al.[21]

The major weakness of the estimates of loop stabilities lies in the lack of sequence dependence of the loop free energies and the paucity of experimental data which has been analyzed, especially for interior loops and bulge loops. In this work, we do not generate significant new information about the stability of loops; rather, we use the information already known to learn more about RNA secondary structure.

B. Applicability of the Helix-Coil Theory to Thermodynamic Analysis of RNA Melting.

1. Helix-Coil Theory Applied to Oligonucleotides

Helix-coil theory has provided an understanding of the thermal transitions of polypeptides, proteins, RNA, and DNA.[22,23] The theory has also seen limited application in the case of small nucleic acids. Published studies include the melting of acid A_n,[11] $A_n + U_n$,[14] A_nU_n,[12] and A_nCGU_n.[24] In these studies, the authors have generally assumed that the helix-coil theory used for polymers is appropriate for oligomers as well. Applequist and Damle have done exactly this. Porschke[14] and Appleby and Kallenbach[15] have added a stacking term to the theory. Martin et al.[12] have used an even simpler version of the theory, in which they approximate the multiple equilibria (shown in Figure 2-1) by just one equilibrium.

In order to answer the questions posed in sections A-1 and A-2 of this chapter, we need a theory which is both internally consistent and able to reproduce experimental findings. We will be led to conclude that the helix-coil theory, in its present form, is not entirely satisfactory in its application to oligomers. It will prove fruitful to investigate several model calculations in order to arrive at a model which is appropriate for oligonucleotide systems. We will conclude that end effects, which may be neglected for polymers, have a significant influence on the melting behavior of oligomers. We will assess these effects quantitatively.

2. Prediction of Secondary Structure

An extremely interesting application of the thermodynamics of RNA is the prediction of RNA secondary structure.† This is of interest primarily because the function of RNA in biological systems is likely to depend on the secondary structure of the molecule. Although we limit our application to transfer RNA and 5S ribosomal RNA, an understanding of secondary structure of other RNA molecules would undoubtably shed light on their biological function. We illustrate with a few examples.

a. tRNA

The relation between structure and function of transfer RNA is an example of the varied way in which RNA molecules can participate in highly specific interactions. The sequence of about

---

†Primary structure is the sequence of nucleic acid bases in the RNA strand. Secondary structure is the arrangement of base pairs in the nucleic acid molecule. Tertiary structure is the three dimensional spatial orientation of double stranded and loop regions in the molecule.

80 bases in tRNA contain sufficient information to allow the following processes, all of which involve a high degree of specificity: (1) recognition of the codon in messenger RNA and binding to it; (2) interaction with a specific activating enzyme so that the correct one of twenty amino acids is loaded onto the 3' end of the molecule; (3) interaction with enzymes which selectively methylate or otherwise alter specific bases (presumably so that these chemically altered nucleic acid bases can perform an as yet unknown biological function); (4) for some tRNA's, interaction with the protein synthesis apparatus so as to suppress mutations. Additionally, tRNA molecules are involved in a series of less specific interactions with the ribosome and a number of protein synthesis factors.

For tRNA, a reasonable model of secondary structure has found general acceptance. This is the cloverleaf model, consisting of four stem regions, three of which are terminated by a hairpin loop, projected out from an interior loop. While a knowledge of the secondary structure of tRNA is not sufficient to illuminate all of its functions, without this knowledge the problem of biological function is almost completely intractable.

b. Ribosomal RNA

The ribosome, on which protein is synthesized, is approximately 50% RNA; each ribosomal subunit contains one large RNA molecule (of 1500 to 3000 bases) and the large subunit contains a small RNA molecule (5S RNA) of about 120 bases. Ribosomal 5S RNA's for several organisms have been sequenced and it is known that the molecules consist of one strand with both single and double

stranded regions. Although the function of ribosomal RNA remains a mystery, the suggestion has been made that unpaired bases in the molecule might help bind transfer RNA and messenger RNA to the ribosome during protein synthesis.[25] It would be of great value to know the secondary structure of ribosomal RNA, in order to test this and other hypotheses.

c. Molecular Evolution of RNA

A number of viruses and bacteriophages contain RNA rather than DNA as the genetic material. Much of the RNA functions as a template for the coat protein and other proteins. Because of the redundant nature of the genetic code, the same protein can be synthesized from RNA templates which differ in composition in approximately every third position. If the sole function of the RNA in these systems is to serve as a template for protein synthesis, it is difficult to understand how a unique RNA sequence would have evolved, as mutations in many of the bases would not change the biological function of the molecule. Since a unique sequence for these RNA molecules does occur, the only explanation is that the nucleic acid bases in the third positions have been selected to serve a necessary function. This function must be determined, wholly or in part, by the secondary structure of the RNA.

In the concluding chapter we develop a methodology for predicting secondary structure from the base sequence of an RNA. We apply this methodology to the secondary structure of 5S and transfer RNA.

CHAPTER 1

REFERENCES

1. S. R. Jaskunas, C. R. Cantor, and I. Tinoco, Jr., Biochemistry, 7, 3164-3178 (1968).

2. H. DeVoe and I. Tinoco, Jr., J. Mol. Biol., 4, 500-517 (1962).

3. R. Rein, P. Claveria, and M. Pollack, Intern'l J. Quant. Chem., II, 129-144 (1968).

4. D. F. Bradley, S. Lifson, and B. Honig, in Electronic Aspects of Biochemistry, Academic Press (1964).

5. H. Krakauer and J. M. Sturtevant, Biopolymers, 6, 491-512 (1968).

6. E. Neumann and T. Ackermann, J. Phys. Chem., 7, 2170-2178 (1969).

7. N. Kallenbach, J. Mol. Biol., 37, 445-466 (1968).

8. H. Klump and T. Ackermann, Biopolymers, 10, 513-522 (1971).

9. D. M. Crothers, N. R. Kallenbach, and B. H. Zimm, J. Mol. Biol., 11, 802 (1965).

10. D. M. Crothers, Biopolymers, 6, 1391-1404 (1968).

11. J. Applequist and V. Damle, J. Am. Chem. Soc., 87, 1450-1458 (1965).

12. F. H. Martin, O. C. Uhlenbeck, and P. Doty, J. Mol. Biol., 57, 201-215 (1971).

13. M. Eigen and D. Porschke, J. Mol. Biol., 53, 123 (1970).

14. D. Porschke, Biopolymers, 10, 1989 (1971).

15. D. W. Appleby and N. R. Kallenbach, Biopolymers, 12, 2093 (1973).

16. H. Jacobson and W. Stockmayer, J. Chem. Phys., 18, 1600 (1950).

17. O. C. Uhlenbeck, P. N. Borer, B. Dengler, and I. Tinoco, Jr., J. Mol. Biol., 73, 483-496 (1973).

18. J. Gralla and D. M. Crothers, J. Mol. Biol., 73, 497 (1973).

19. C. DeLisi and D. M. Crothers, Proc. Nat. Acad. Sci. USA, 68, 2682-2685 (1971).

20. J. Gralla and D. M. Crothers, J. Mol. Biol., 28, 301 (1973).

21. I. Tinoco, Jr., P. N. Borer, B. Dengler, M. D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla, Nature New Biol., 246, 40-41 (1973).

22. B. H. Zimm and J. K. Bragg, J. Phys. Chem., 31, 526-535 (1959).

23. B. H. Zimm, J. Phys. Chem., 33, 1349-1356 (1960).

24. O. C. Uhlenbeck, F. H. Martin, and P. Doty, J. Mol. Biol., 57, 217-229 (1971).

25. J. D. Watson, in The Molecular Biology of the Gene, Benjamin (1970).

CHAPTER 2

MELTING OF DOUBLE STRANDED OLIGONUCLEOTIDES WITH ONLY A-U BASE PAIRS

In this chapter we consider the details of the melting of a series of small double stranded RNA molecules. The molecules of interest are ribonucleic acid oligomers with only A-U base pairs present in the double strand: the series $\begin{matrix} A & U \\ U_n & A_n \end{matrix}$, in which n varies from 4 to 7, and the molecule $\begin{matrix} A_4UAU_4 \\ U_4AUA_4 \end{matrix}$. In the next chapter we will use the results of this analysis to study the molecules $\begin{matrix} A_n & CU \\ U_n & GA_n \end{matrix}$ (n=2,4), $\begin{matrix} A_n GCU_n \\ U_n CGA_n \end{matrix}$ (n=2,4), $\begin{matrix} A_4G_2 \\ U_4C_2 \end{matrix}$, $\begin{matrix} A_4G_3 \\ U_4C_3 \end{matrix}$, and $\begin{matrix} A_4GC \\ U_4CG \end{matrix}$.† These molecules have been synthesized and studied by O. Uhlenbeck, F. Martin, P. Borer, B. Dengler, and D. Koh under the direction of P. Doty and I. Tinoco, Jr.[1,2,3] In all cases, the melting curves were measured in 1 M Na^+ at pH 7. Although the use of high salt was dictated by experimental considerations (in order to increase the melting temperature for the shorter oligomers above 0°C), the choice of 1 M salt was fortuitous for theoretical reasons as well. A high salt concentration screens the negative charges on the phosphate backbone from one another. Since such repulsive interactions are long range (the energy of interaction is proportional to 1/r), at lower salt concentrations very complicated non-nearest neighbor effects render

---

†Henceforth the double stranded oligomer is specified by the sequence of bases in only one strand in the text. Thus A_nU_n means the double stranded oligomer $\begin{matrix} A & U \\ U_n & A_n \end{matrix}$ unless otherwise indicated. When double stranded molecules which are not entirely complementary are referred to, then both strands are usually specified.

the understanding of the melting process very difficult.[4]

We first summarize the equations, based on the helix-coil theory, which have previously been applied to oligonucleic acids. An analysis of the previous calculations is provided which indicates that theory and experiment are not in satisfactory agreement. Calculations on the molecules are presented and they further emphasize the inadequacy of the traditional treatment. Several possible solutions of the problems are suggested and it is concluded that one particular model (which accounts for the effects of the ends of the molecules) is most appropriate for these oligomers (and, by inference, for other small RNA molecules as well). The necessary changes in the equations are made and detailed results are presented and discussed. In all cases, the goal is to so characterize the interactions in the simple double stranded RNA molecules that sufficient information is gained from the calculation to allow the prediction of the stability of more complex RNA molecules.

A. Standard Statistical Thermodynamic Model with only A-U Base Pairs

1. Intermediate States are Considered in the Calculation of Melting Curves

Figure 2.1 is a schematic representation of the intermediate states which are considered important in determining the thermodynamics of the helix coil transition in the standard statistical thermodynamic treatment. For simplicity, the molecule A_3U_3 is shown; the effect of G-C base pairs is treated in Chapter 3. The statistical weight of each of these intermediate states (i.e., the contribu-

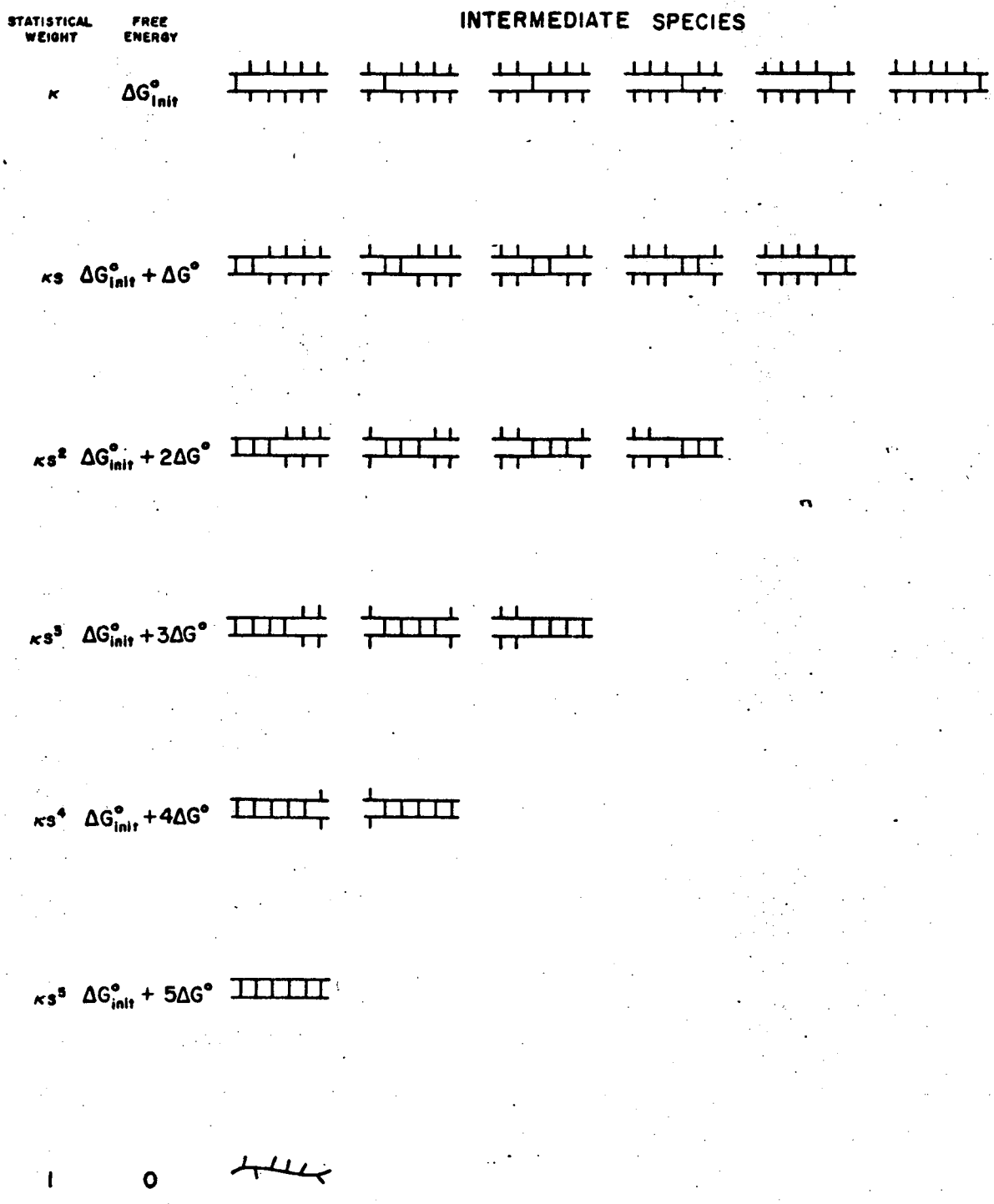

STATISTICAL WEIGHT
FREE ENERGY
INTERMEDIATE SPECIES
κ ΔG°_{init}
κs $\Delta G^\circ_{init} + \Delta G^\circ$
κs^2 $\Delta G^\circ_{init} + 2\Delta G^\circ$
κs^3 $\Delta G^\circ_{init} + 3\Delta G^\circ$
κs^4 $\Delta G^\circ_{init} + 4\Delta G^\circ$
κs^5 $\Delta G^\circ_{init} + 5\Delta G^\circ$
1 0

Figure 2-1

tion of each to the system partition function) as well as the standard free energy is shown. This characterization of intermediate states is similar to that originally derived for DNA and protein polymers.[5] It was first successfully applied to oligomers by Applequist and Damle, who also derived several useful equations which apply to small double stranded nucleic acid molecules.[6]

Two different and distinct equilibria control the melting process, according to Fig. 2.1. These are commonly termed initiation and chain elongation. Initiation, which is the formation of the first base pair between two previously independent strands, is specified by the equilibrium constant

$$\kappa = e^{-\Delta H^\circ_\kappa/RT} \; e^{\Delta S^\circ_\kappa/R} \qquad (2.1)$$

Similarly, the equilibrium constant for the addition of an A-U base pair to a partially formed helical region is given by the expression

$$s = e^{-\Delta H^\circ/RT} \; e^{\Delta S^\circ/R} \qquad (2.2)$$

In these equations, ΔH°_κ and ΔH° refer to the enthalpy for the formation of the first base pair (initiation) and the enthalpy for the addition of an A-U pair to the helix respectively. ΔS°_κ and ΔS° are the entropies associated with the two processes. R is the gas constant and T is the absolute temperature in degrees Kelvin.

The distinction between initiation and elongation equilibria is not merely a formal one. Initiation requires that two strands, which are independent of one another in solution, come together in

such a way that they remain bound long enough to permit the formation of additional base pairs. A large number of collisions may be necessary before this occurs. As such, the entropy for the initiation process is unfavorable (less than zero). The enthalpies for the initiation and elongation equilibria are significantly different as well. As discussed in Chapter 1, the elongation of a helix brought about by the formation of a base pair involves the addition of a double stranded stacking interaction, which stabilizes the double helix. In the initiation step, no double stranded stacking interaction is formed. As a result, $\Delta H^{\circ}_{\kappa}$ is likely to be considerably smaller than ΔH°. Since the initiation step is dominated by a negative entropy term, the free energy associated with initiation is expected to be positive. This explains why RNA double helices with fewer than four base pairs have not been detected in aqueous solution:[7] the stabilization associated with the formation of base pairs is not sufficiently large to overcome the destabilization of initiation until the number of base pairs is four or greater. That κ is less than unity is also consistent with the results of kinetic measurements on oligonucleotides with only A-U base pairs; analysis of the results of these experiments indicate that a complex of at least three base pairs is needed in order for the double helix to be stable enough to lead to the fully formed duplex molecule.[8]

The statistical weights of all intermediate species in Fig. 2.1 at all temperatures can be specified if four thermodynamic functions are known, according to equations 2.1 and 2.2. It is important to note that a number of approximations are implicit in

the use of the statistical weights given in Fig. 2.1 to characterize the intermediate states leading to the formation of the double helical oligomer. The most important of these are summarized:

i. The formation of base pairs is regarded as a two-state process. Partially formed pairs do not contribute to the partition function. Rotations and vibrations are ignored.

ii. The enthalpies and entropies, ΔH° and ΔS°, related to the formation of a base pair are additive. That is to say that only nearest neighbor interactions contribute to the stabilization of the double stranded helix.

iii. ΔH° and ΔS° are independent of temperature throughout the transition region. The effect of single strand stacking, which might give rise to some degree of temperature dependence of these thermodynamic functions, is ignored once its optical contribution is subtracted. (See later discussion.)

iv. Internal loops are ignored. For example, no intermediate species such as $\begin{matrix} AA & AU & UU \\ UU & UA & AA \end{matrix}$ are included in Fig. 2.1. It has been demonstrated that such moieties influence the melting of oligonucleotides negligibly, although they must be considered in polymer calculations.[1] Staggered configurations, in which the complementary strands are out of register, are not included. While these species may be present throughout the melting, their contribution to the overall process can be shown to be vanish-

ingly small. This is due to the fact that for the block oligomers A_nU_n, staggered species can have at most (2n-4)/2 double stranded stacking interactions and are thus not very stable for small n.

v. Certain nearest neighbor interactions are not included in the partition function: end effects are completely ignored; the AU juncture in the middle of the molecule is given the same statistical weight as the AA interaction.

Many of these approximations are critically analyzed in this chapter (section B2) when alternative models for the melting of duplex oligonucleotides are developed.

One further approximation is often made in order to simplify the problem: ΔH°_κ is set to zero so that there is no temperature dependence of κ. The physical reasoning behind this approximation is that the energies of initiation depends on the formation of hydrogen bonds between complementary bases. For every hydrogen bond thus formed, a hydrogen bond between the biological base and water is broken. The net effect on the internal energy (and the enthalpy) of the system should thus be small or negligible. A more detailed discussion of this assumption and its probable validity is found in reference.9. Even if the assumption is slightly in error, it will introduce a very small error in the final calculation.†

---

†A rather large error in ΔH°_κ can be compensated for in the calculation by a rather small increase or decrease in the magnitude of ΔH°. For example, a change of 0.5 kcal/mole in ΔH° corresponds to a change of 3.5 kcal/mole in ΔH°_κ for A_4U_4 and 6.5 kcal/mole in ΔH°_κ for A_7U_7. It is for this reason that the calculation is so

If information from polymers is used, then only two thermodynamic functions need to be solved for. ΔS° is obtained from the fact that for a polymer (in this case an average of polyA:polyU and polyAU:polyAU) the Gibbs free energy is zero at the midpoint of the melting. Thus,

$$\Delta G = \Delta H - T_m^\infty \Delta S = 0 \qquad (2.3)$$

where T_m^∞ is the midpoint of the polymer melting curve. It is also true, from thermodynamics, that

$$\Delta G = \Delta G^\circ - \frac{RT}{N-1} \ln (c_{ds}/c_{ss}^2) \qquad (2.3a)$$

where c_{ds} = the strand concentration of the double strand and c_{ss} = the strand concentration of the single strand in moles/liter. Since for a high polymer the number of double stranded stacking interactions, N-1, approaches infinity, the concentration term in equation 2.3a makes a negligible contribution to the free energy. For this reason, the standard free energy is equal to the free energy near the midpoint of the transition for the polymer and $\Delta S = \Delta S^\circ$ and $\Delta H = \Delta H^\circ$. This is a useful relationship. It means that if the enthalpy and T_m^∞ of the polymer are measured, then the standard enthalpy and standard entropy may be calculated directly. We can then write, for the polymer, that

$$\Delta G = \Delta H^\circ - T_m^\infty \Delta S^\circ = 0$$

---

insensitive to the magnitude of ΔH_K°. It is not possible, as Eigen and Porschke[8] allege, to accurately assess ΔH_K° by means of model calculations on melting curves of a small number of oligomers.

and,

$$\Delta S^\circ = \Delta H^\circ / T_m^\infty \qquad (2.3b)$$

Now we can apply the polymer data to the oligomers. If the reasonable assumption is made that the thermodynamics of base pair formation within a double stranded polyribonucleic acid is the same as for base pair formation within an oligoribonucleic acid molecule, then we can rewrite equation 2.2, using the entropy term of equation 2.3b:

$$s = e^{-\Delta H^\circ / R(1/T - 1/T_m^\infty)} \qquad (2.4)$$

This is the standard relationship for the equilibrium constant which corresponds to the addition of a base pair to a helix. As stated, this equation rests on the assumption of equivalence between the energetics of base pairing in the oligomer and polymer. Experimental measurements of circular dichroism of the oligomers under study supports the idea that the geometry of the base pairs is generally similar for the oligomers and polymers.[3] The magnitude of ΔH°, calculated from equation 2.4, is also consistent with the measured enthalpy of polyA:polyU.[10]

2. Experimental Absorption Profiles must be Corrected for Single Strand Absorption in order to make Comparisons between Theory and Experiment.

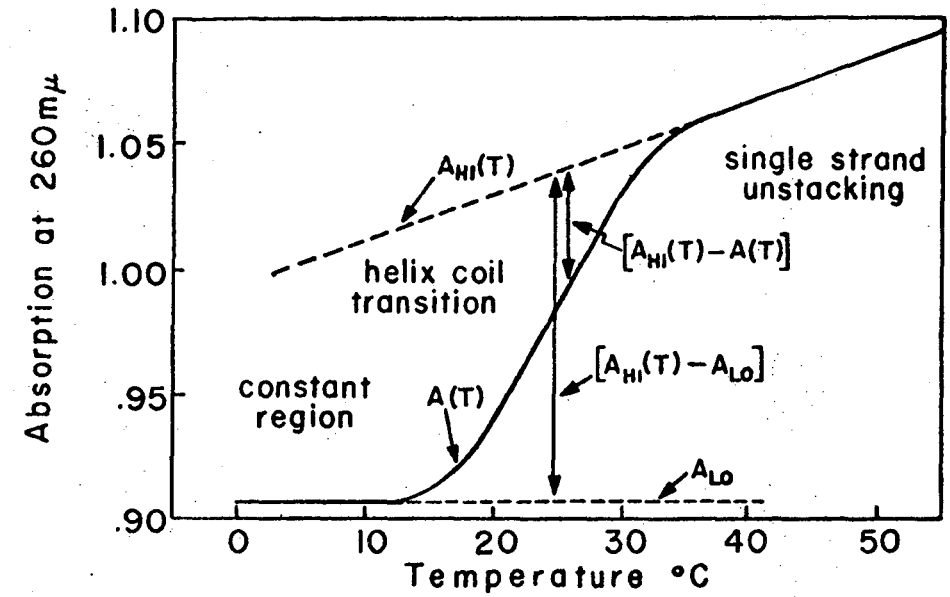
A typical absorption profile for an $A_n U_n$ oligomer is shown in Fig. 2.2a. At low temperatures, absorption changes little if at all with temperature. Throughout the transition from coil to double

helix, the absorption increases rapidly. At higher temperatures a limiting slope greater than zero is attained. This limiting slope must be due to changes within the single strand, since the observed slope is independent of concentration and, for molecules in which the two strands are different entities so that the absorption of the strands can be measured separately, the observed slope is the sum of the contributions from the individual strands. The process responsible for the high temperature absorption change is termed unstacking; it is generally pictured as the weakening of the interactions between bases adjacent to one another on the single strand as the temperature is raised.[11]

Since we are concerned with the helix coil transition, we wish to exclude the optical contribution of single strand unstacking to the melting curve. Martin et al.,[2] have done this by using the relation

$$1\text{-}f = \frac{[A_{hi}(T) - A(T)]}{[A_{hi}(T) - A_{lo}]} \qquad (2.5)$$

where the terms $A(t)$, $A_{hi}(T)$, and A_{lo} are molar absorptions as defined in Fig. 2.2a. f = the fraction of the RNA double strand which is formed at a given temperature. (More precisely, as explained below, f is the average fraction of base stacking interactions present at a given temperature.) The melting temperature of the oligomer, T_m, is defined as the temperature at which f equals 0.5 (i.e., the midpoint of the melting curve). Note that we have defined the experimental curve as an absorption profile; the expression melting curve refers to a plot of 1-f vs. temperature.


1.10
1.05
1.00
.95
.90
Absorption at 260mμ
$A_{HI}(T)$
single strand
unstacking
$[A_{HI}(T)-A(T)]$
helix coil
transition
$[A_{HI}(T)-A_{LO}]$
constant
region
A(T)
A_{LO}
0
10
20
30
40
50
Temperature °C

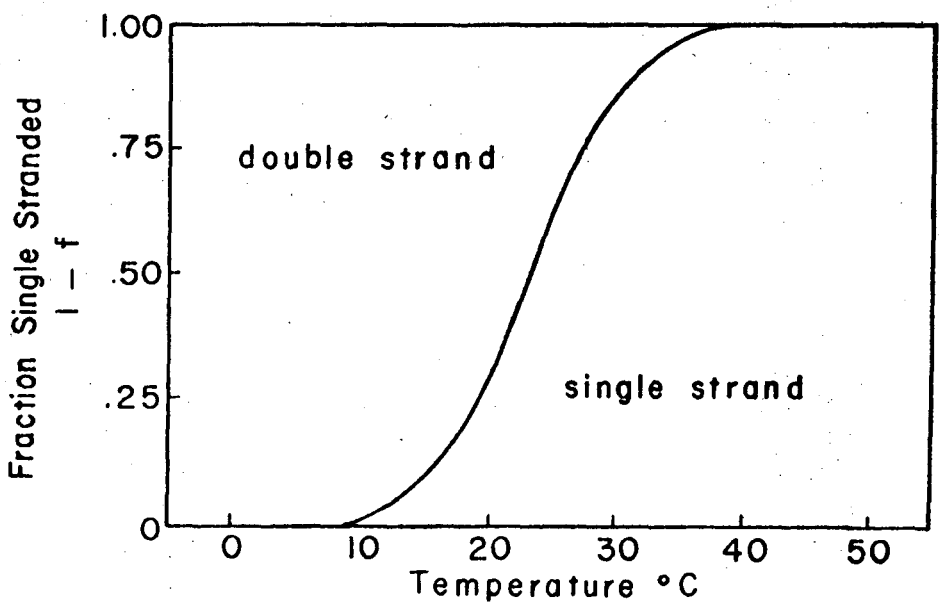

1.00
.75
.50
.25
0
Fraction Single Stranded
1 − f
double strand
single strand
0
10
20
30
40
50
Temperature °C

Figure 2-2

3. Relating Theory to Experiment

In order to relate theory and experiment to one another, it is necessary to use the methods of statistical thermodynamics. This means that a partition function must be derived. For the molecules A_nU_n, or for any nucleic acid molecules with only A-U base pairs, the most general form of the partition function is

$$Q = \sum_{\substack{\text{all} \\ \text{states}}} \kappa s^{i-1} \qquad (2.6)$$

Here, "all states" refers to all intermediate states with i-1=1, 2, ..., N-1 double stranded stacking interactions with no interior loops (Fig. 2.1); N = the number of base pairs in the fully formed duplex oligomer. In this simple case (all A-U base pairs), the partition function may be written

$$Q = \sum_{i=1}^{N} (N+1-i)\ \kappa s^{i-1} \qquad (2.6a)$$

where the sum is over i, the number of base pairs in a given species. As is evident from Fig. 2.1, there are N+1-i species with i base pairs; this accounts for the first term in equation 2.6a. Although equation 2.6a can be summed analytically, the general form of equations 2.6 and 2.6a are more useful for the majority of molecules considered in this work, which have G-C base pairs.

In order to relate theory and experiment, it is necessary to adopt a model for the molecular origin of the increase of absorption with the melting of the double helix (hyperchromicity). The important question is: what is the correspond-

ence between the physical measure of the process, absorption, and the process itself? Two different models have been proposed: one attributes the hyperchromicity to the disruption of base pairs[12] and the other to the breaking of double stranded stacking interactions.[13,14] As shown below, the two models predict different relationships between 1-f, as measured experimentally, and the progress of the helix coil transition.

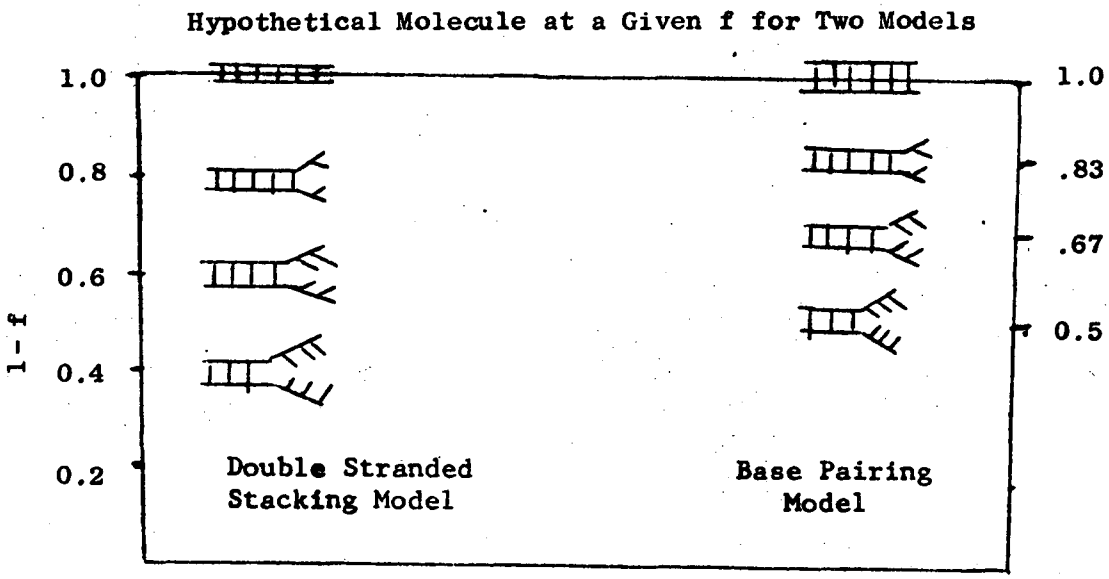


Figure 2-3

Although both of these models represent a simplification of the actual causes of hyperchromicity, calculations based on the double stranded stacking model are in better agreement with experiment than similar calculations based on a base pairing model.[15] For this reason, f is defined as the average fraction of double stranded stacking interactions present at a given temperature. This is equivalent to saying that

f = (fraction of strands which have at least one base pair, ϕ) x (fraction of base stacking interactions/

species with at least one base pair)

Defined in this way, f has the same meaning as in equation 2.5 (with the added assumption that the breaking of the double stranded stacking interaction is responsible for the increased absorption). Mathematically, this may be written

$$f = \phi\kappa \sum_{\text{all states}} \frac{(i-1)s^{i-1}}{(N-1)Q} \tag{2.7}$$

$$= \phi\kappa \sum_{i=1}^{N} \frac{(N-i+1)(i-1)s^{i-1}}{(N-1)Q} \tag{2.7a}$$

Equation 2.7 is the general form with the sum over all states which are shown in Fig. 2.1. Equation 2.7a is specific to the A_nU_n molecules. As before, Q is the partition function, N is the maximum number of base pairs in the fully formed duplex molecule, and the sum is over all i base pairs present in the intermediate species. ϕ is the fraction of strands with at least one base pair. It can be shown that[†]

$$\phi = (2Qc_tk+1-(4Qc_tk+1)^{1/2})/2Qc_tk \tag{2.8}$$

where, c_t = total strand concentration in moles/liter

k = 1 if the duplex molecule is composed of complementary but unlike strands

k = 2 if it is composed of two identical strands, as in the case of the A_nU_n duplexes

---

[†]The derivation of equation 2.8 is as follows. For the case in which the two strands are identical, we can write the chemical equilibrium

$$A + A \rightleftarrows B \tag{2.9}$$

where, A = the single stranded oligomer
B = all duplex molecules (i.e., all molecules with one or more base pairs)

Then, (B) = (species with one base pair) + (species with two pairs) + ... + (species with N base pairs)

From the definition of the partition function, Q, and the fact that the free energy of the single strand is defined as zero (from Fig. 2.1), it follows that

$$Q = \kappa \sum_{i-1}^{N} (N+1-i)s^{i-1} = (B)/(A)^2. \qquad (2.9a)$$

Since the two complementary strands are identical,

$$c_t = 2(B) + (A) \qquad (2.9b)$$

c_t is thus constant throughout the reaction. The quantity of interest is ϕ. From its definition,

$$\phi = (B)/[(B) + 1/2(A)].$$
$$= 2(B)/c_t \qquad (2.9c)$$

Putting 2.9a into 2.9b, we receive

$$c_t = 2(B) + [(B)/Q]^{1/2} \qquad (2.9d)$$

Equation 2.9d can be made into a quadratic in (B), the solution of which is

$$(B) = [(1 + 4Qc_t) - (1 + 8Qc_t)^{1/2}]/8Q$$

Since $\phi = 2(B)/c_t$, it then follows that

$$\phi = \frac{[4Qc_t + 1 - (8Qc_t + 1)^{1/2}]}{4Qc_t} \qquad (2.9e)$$

For the case in which the two strands are not identical, the equilibrium is

$$A_1 + A_2 \rightleftharpoons B \qquad (2.10)$$

In all the experiments referred to in this work, the two single strands are present in the same concentration. Thus,

$$(A_1) = (A_2) = (A)$$

The derivation of ϕ proceeds as above, except that

$$c_t = (B) + (A)$$

and

$$\phi = (B)/[(B) + (A)]$$
$$= (B)/c_t$$

Physically, the reason for the redefinitions in the case of nonidentical strands is that a given strand can combine successfully with only one half of the strands in solution to yield a fully bonded duplex molecule. As defined above, c_t remains constant throughout the reaction. Using these definitions, the result of equation 2.8 is obtained.

The final equation for 1-f is then obtained by putting the expression for ϕ into equation 2.7 or 2.7a:

$$1-f = \frac{[2Qc_t k + 1 - (4Qc_t k + 1)^{1/2}]}{2Qc_t k} \kappa \sum_{i=1}^{N} \frac{(N - i + 1)(i - 1)s^{i-1}}{(N - 1)Q} \qquad (2.11)$$

All terms in this equation are as previously defined. Equation 2.11 thus provides the standard theoretical method of calculating the melting curve for $A_n U_n$ molecules which may then be compared with experimental results plotted according to equation 2.5. The additional assumptions made in order to apply this equation are

i. ΔH°_κ is zero and thus κ does not vary with temperature.

ii. The base pairing for the oligomer is similar to that for the polymer so that ΔS° can be calculated from polymer data.

Thus, in order to solve equation 2.11, two new parameters must be obtained: ΔH° and κ.

4. <u>The All-or None Model Gives Insight into the Problem</u>.

An approximate treatment in which the only states of the multiple equilibria considered are the fully formed duplex and the single stranded species is often used. It has been suggested that this approximation is valid for oligomers with ten or fewer base pairs.[2] Later results of this work call into question its applicability even for very short oligomers. Nonetheless, the theory has the advantage that it leads to very simple equations which are useful in the analysis of the results of the statistical thermodynamic theories.

Defining the equilibrium constant

$$K = (B)/(A)^2 = {}_{\kappa}s^{N-1} \qquad (2.12)$$

for the reaction

$$A + A \underset{}{\overset{K}{\rightleftharpoons}} B,$$

the fraction of double stranded stacking interactions is simply

$$f = \frac{2(B)}{[2(B) + (A)]} = \frac{2(B)}{c_t} \qquad (2.13)$$

c_t and f are as previously defined; (B) is the concentration of fully formed duplexes rather than the concentration of all intermediate states as before. Equation 2.13 is appropriate if the two complementary single strands are identical. If the strands are different (eg., $A_nCU_n + A_nGU_n$), then

$$f = \frac{(B)}{[(B) + (A)]} \qquad (2.13a)$$

where each of the two unlike strands have concentration (A).

Relating f and K (using the identical procedure which led to equation 2.8), we obtain the result

$$f = \frac{(2Kc_tk + 1) - (4Kc_tk + 1)^{1/2}}{2Kc_tk} \qquad (2.14)$$

Equations 2.8 and 2.14 are identical except for the substitution of K for Q in the latter. This is entirely correct: Q represents a sum of the statistical weights of all intermediate double stranded

species whereas K is for only the fully formed duplex. This is, of course, the difference between the all-or-none theory and the statistical thermodynamic theory.

The importance of equation 2.14 is that it leads to the following useful relationships. Substituting $\kappa s^{N-1} = \kappa e^{-\Delta H^\circ/RT} e^{-\Delta S^\circ/R}$ for K in this equation and differentiating with respect to T yields

$$(df/dT)_m = \frac{(N-1)\Delta H^\circ}{6RT_m^2} \tag{2.15}$$

$(df/dT)_m$ is the slope of the melting curve at the midpoint of the transition, T_m, where f = 0.5.

Since, according to equation 2.14, f is constant if $Kc_t \equiv \kappa s^{N-1} c_t$ is constant, it can be shown that

$$1/T_m' = 1/T_m + \frac{R\ln(c'/c)}{(N-1)\Delta H^\circ} \tag{2.16}$$

and

$$1/T_m = 1/T_m^\infty + \frac{R\ln\kappa}{(N-1)\Delta H^\circ} \tag{2.17}$$

A complete derivation of these equations is given in the appendix to reference 6. At this point, it suffices to note that equations 2.15 through 2.17 follow directly from 2.14. T_m^∞ is, as before, the melting temperature of the infinite polymer with a sequence similar or identical to that of the oligomer considered. To use equation 2.17, T_m must often be estimated since infinite polymers of appropriate sequences have generally not been synthesized.

The utility of these equations is that they provide a first approximation to ΔH°, ΔS°, and κ. Martin and Uhlenbeck used these equations as the basis for their analysis of the thermodynamics of the A_nU_n oligomers.[2] For this work, equations 2.15 and 2.17 illustrate an important principle which is not entirely dependent on the use of an all-or-none model: namely that two independent measures of the enthalpy of the reaction are provided by the melting data. Equation 2.15 relates ΔH° to the slope of the melting curve; equation 2.16, to the concentration dependence of a series of melting curves. Both of these properties are obtained at the midpoint of the melting curve, both for reasons of mathematical simplicity and because experimental accuracy is greater at the midpoint of the curve.

That these two properties are measures of ΔH°, regardless of the model which is used to analyze the results, is physically reasonable. The greater the magnitude of the total enthalpy for the reaction,† the greater is the variation of the equilibrium constant ($\propto e^{-(N-1)\Delta H^\circ/RT}$) with temperature and the steeper the melting curve. The physical basis for equation 2.16 is less obvious. The equation states that the greater the total enthalpy,† the less the melting temperature changes with concentration. This can be best understood by noting the effect of changing the concentration of single strands on the melting temperature of the molecule. At f=0.5, the term $\kappa c_t s^{N-1}$

---

† The expression <u>total enthalpy</u> refers to the enthalpy for the formation of all base pairs, $(N-1)\Delta H^\circ$. To avoid the confusion which often occurs in thermodynamic discussions, the expression "increase in enthalpy" or "the greater the magnitude of the enthalpy" means that if ΔH° is negative it will become more negative, if positive, then more positive. The expressions refer to an increase in the <u>absolute value</u> of the enthalpy.

must be constant. (It is either one or two, depending on whether or not the two complementary strands are identical, according to equation 2.14.) This means that, as c_t is increased, s must decrease in order to maintain f constant. The effect is to shift the equilibrium to higher temperature, since $s^{N-1} \propto e^{-(N-1)\Delta H^\circ/RT}$ and ΔH° is less than zero. If the total enthalpy, $(N-1)\Delta H^\circ$ is small, then a large change in temperature is needed to maintain the constancy of $\kappa c_t s^{N-1}$ (i.e., of $\kappa c_t e^{-(N-1)\Delta H^\circ/RT} e^{\Delta S^\circ/R}$). If the total enthalpy is large, then a small change in T will accomplish the same end. In the extreme case of the polymer, where the total enthalpy change for the reaction is enormous (approaches infinity), the melting curve is essentially independent of concentration. The statistical thermodynamic model does not require that $\kappa c_t s^{N-1}$ be constant at f = 0.5, but the constraints (which may be understood by comparing equations 2.14 and 2.11) are similar. It is thus evident that the concentration dependence of melting, as well as the slope of the melting curve, is a measure of the enthalpy of the reaction.

5. The Results of the Statistical Thermodynamic Model are not Consistent: A New Model is Needed.

Equations 2.11 and 2.4 have been used to calculate the melting curves for the series of molecules A_nU_n, n = 4-7. The computer program, listed in Appendix I, was written for and run on a model 6600 Control Data Corporation computer. Both ΔH° and κ were varied so as to achieve the best possible agreement with experiment. The procedure was to choose ΔH° and then vary κ until the deviation between experimental and predicted melting temperatures for all four molecules at three different concentrations was minimized. ΔH° was increased by increments of 0.5 kcal/mole between 5 and 10 kcal/mole.

Table 2.1 illustrates that any choice of ΔH° between -6.5 and -9.0 kcal/mole, with κ chosen appropriately, yields agreement with the melting temperatures to within approximately 1°C. Two conclusions can be drawn: (*i*) the standard statistical thermodynamic theory is in excellent agreement with the experimental melting temperatures for the A_nU_n molecules, since the variation of predicted melting temperatures is almost within experimental error (which is ±0.5°C). The experimental and predicted melting temperatures are summarized in Table 2.7. (*ii*) It is not possible, by fitting T_m data alone, to choose ΔH° and κ unambiguously.

Table 2.1

Accuracy of Calculated Melting Temperatures for A_nU_n (n = 4-7) at 3 Concentrations

ΔH° (kcal/mole)	κ (1/mole)	Standard deviation of T_m
-6.5	2.8×10^{-3}	±1.2°C
-7.0	7.8×10^{-4}	±0.9°C
-7.5	2.2×10^{-4}	±0.75°C
-8.0	6.0×10^{-5}	±0.70°C
-8.5	1.7×10^{-5}	±0.75°C
-9.0	4.8×10^{-6}	±0.9°C

In order to carry the analysis further, we must use more information than the melting temperatures of the A_nU_n oligomers. We have already noted that there are two independent measures of the enthalpy for the reaction: the slope of the melting curve at the midpoint and the concentration dependence of the melting temperature.

Both pieces of information are available from experiment. Since the analytical expressions previously derived are appropriate only for the all-or-none model, a somewhat more complicated procedure must be followed to obtain the value of the enthalpy from these two physical properties within the framework of the statistical thermodynamic model. The procedure followed is listed:

a. For each of the values of ΔH° in Table 2.1 (with κ chosen so as to achieve best agreement with the melting temperature, i.e., using the κ's of column 2, Table 2.1), theoretical melting curves were calculated at three strand concentrations (10^{-4} M, 10^{-5} M, and 10^{-6} M), using equations 2.11 and 2.4. A value of 1-f was calculated for every degree Centigrade between -20°C and +80°C.

b. The computer program extrapolated the temperature at which f = 0.5 from the two temperatures nearest the midpoint of the melting curve. Once T_m was established in this way, a second calculation of the portion of the melting curve about the T_m was performed. 1-f was calculated at five temperature increments of 0.1°C above T_m and five below T_m. In this way, $(df/dT)_m$ was determined very precisely.

c. Since melting curves were calculated for the molecules at three concentrations, the concentration dependence of T_m was calculated directly. To compare theory and experiment, the quantity $(1/T_m - 1/T_m')/\ln(c/c') \equiv \Delta(1/T_m)/\Delta\ln(c)$ was calculated. In general, $c = 10^{-5}$ M and $c' = 10^{-6}$ M were used for the comparison between theory and experiment.

d. The results were plotted as in Fig. 2.4: $(df/dT)_m$ and $\Delta(1/T_m)/\Delta\ln(c)$ versus ΔH° (with κ implicitly varying with the

choice of enthalpy). The vertical line in the figure represents the experimental result for each of the molecules. We have omitted $(df/dT)_m$ for A_6U_6 since the published value[2] is not consistent with the slopes for the other A_nU_n molecules and is not in agreement with the few melting curves of A_6U_6 which are available.

This calculation, the results of which are summarized in Table 2.2, brings out a serious inconsistency in the theory: two significantly different enthalpies for the same process (and for the same base pairing and double stranded stacking interactions) emerge from the calculation: -8.0 ± 0.2 kcal/mole from the concentration dependence of T_m and -6.7 ± 0.3 kcal/mole from the slope of the melting curves.

It is probable that the enthalpy derived from the concentration dependence of T_m is the more nearly correct one, since it is in much better agreement with calorimetric measurements on polyA:polyU (Chapter 1, section A1). But if this value of the enthalpy is used, then all predicted melting curves are considerably sharper than experimental ones. We may use as a measure of the breadth of the transition the quantity $\Delta_{2/3}$, defined as the temperature range over which two-thirds of the melting occurs measured symmetrically about the T_m. Using $\Delta H^\circ = -8.0$ kcal/mole, the predicted breadths are only 60% of the experimental breadths (Table 2.7).

We are thus presented with two inconsistencies: one between theory and experiment and the other within the theory itself. This predicament becomes even more serious when G-C base pairs are

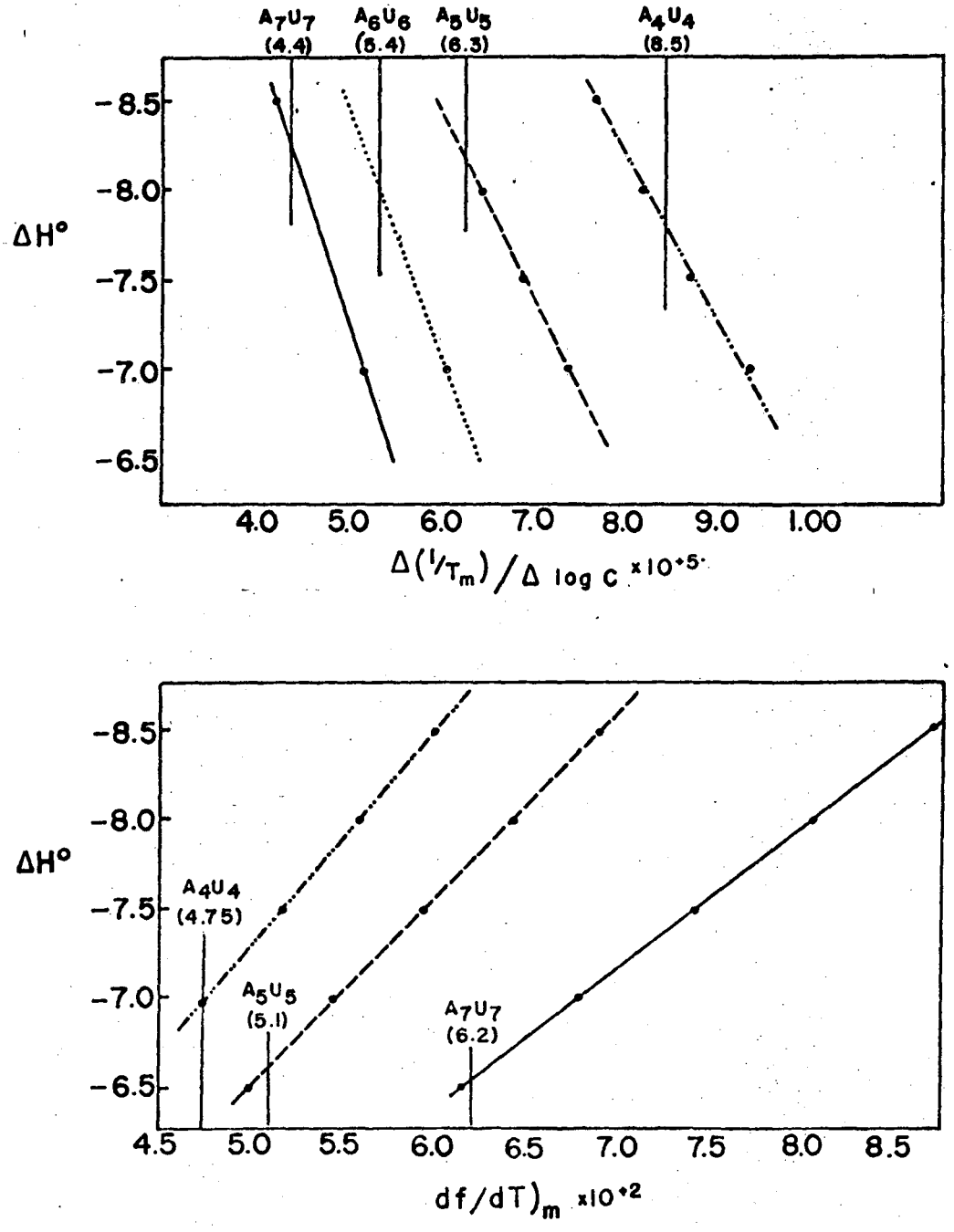


Figure 2-4

Table 2.2

Enthalpy of an A-U Base Pair Calculated from the Standard Statistical Thermodynamic Theory for A_nU_n

Molecule	ΔH° from concentration dependence of T_m	ΔH° from the slope of the melting curve
A_4U_4	-7.8 kcal/mole base pairs	-7.0 kcal/mole base pairs
A_5U_5	-8.15	-6.7
A_6U_6	-7.9	--
A_7U_7	-8.15	-6.5
Average	-8.0 ± 0.2 kcal/mole	-6.7 ± 0.3 kcal/mole

present, as will be pointed out in the next chapter. It is to the resolution of this dilemma that we now turn our attention.

B. Modifications of Existing Theories are Considered.

1. Discussion of Past Theoretical Work

Before embarking on a systematic study to account for the inadequacy of the present theoretical approach, it is relevant to inquire if and how other workers have treated the problem. A limited amount of theoretical work on oligonucleotide systems has been done. The inconsistencies discussed above have been noted by some workers,[2] but a satisfactory solution has not been proposed.

a. Acid Oligo A_n

Applequist and Damle[6] studied the equilibria between single and double stranded oligoribo A_n at pH 4, 0.15 M Na^+. They considered oligomers with eleven or fewer base pairs. The theoretical

approach of section 1 of this chapter is generally similar to Applequist and Damle's treatment of the problem. They also included staggered configurations in the partition function, since these species are important for their system. The data were satisfactorily fit with $\Delta H^\circ = -8.0$ kcal/mole and $\kappa = 2.2 \pm 10^{-3}$ liter/mole.[†]

The criticism can be made that the thermodynamic constants derived from the calculations were not entirely constant from one oligomer to the next. The enthalpy varied from -7.0 kcal/mole for A_8 to -8.8 kcal/mole for A_{10}, with a standard deviation of ±0.95 kcal/mole. Nonetheless, the fit of the melting was generally very good. Specifically, the theoretical and experimental melting curves are in good agreement with respect to the melting temperatures, the slope of the melting curves at the midpoint, and the breadths of the melting curves. In light of our experience with the A_nU_n system, this good agreement is at first glance rather surprising.

Two comments are relevant: first, it is entirely possible that the same inconsistency noted above for A_nU_n applies to acid oligo A_n as well, but that Applequist and Damle were not able to observe it. Their data were limited to one concentration for each oligomer. As such, they fit the calculations only to the shape of the melting curves. Thus, if the concentration dependence gives

---

[†]For our work, we have defined the initiation parameter as κ, which is the statistical weight of the duplex with one base pair. Some workers have defined the statistical weight of the duplex with one base pair as βs, where β is the initiation parameter. Thus, our should generally be compared with others' βs. (s is generally between 2 and 5 at the T_m of the oligomers studied.)

a different ΔH°, they had no way of knowing this. Second, their system is sufficiently different from the A_nU_n system to dictate caution in making a comparison between the two. The double stranded $A_n:A_n$ occurs only at acid pH, in which the adenylic acid residue is protonated. As a result, a favorable interaction between the negatively charged backbone and the adenylic acid residue occurs. Consequently, the double stranded form is less favored at higher salt concentration; this is the opposite of the salt effect for standard Watson-Crick base pairs in neutral solution. It is therefore possible that the physical cause of the inconsistency in the theory is not operative for acid oligo A_n.

It would be of interest to study the concentration dependence of the acid oligo A_n system in order to resolve this question.

b. $A_n + U_n$

Pörschke and Eigen[8] and Pörschke[16] have done extensive studies on the $A_n + U_n$ system, where n ranged between 8 and 18. Except for the fact that their study was limited to very low salt concentrations (0.05 M) to reduce triple strand formation, this system should be very similar to A_nU_n. Their theoretical approach differs from that outlined in section 1 in two essential ways: they have attempted to account for the effect of electrostatic interactions, very prominent at low salt, by adding the term $[(N-1)RT(C_1 - C_2 \log \mu)/N$ to the free energy, and they have included the effect of single strand stacking with the expression $RT \ln(1 + e^{-\Delta G_a^\circ/RT})$, also added to the free energy. (μ is the ionic strength of the solution, C_1 and C_2 are constants, and ΔG_a° is the single

strand stacking free energy.) Agreement between experiment and theory is remarkably good: the calculated melting curves reproduce the experimental ones almost exactly throughout the transition. Also, they have considered the equilibrium at a range of different concentrations. It might appear that the stacking correction, as they have applied it, provides a solution to the problem.

However, a serious objection must be raised to their approach. Agreement at the lower temperature range of the melting curve is obtained by explicitly fitting each predicted curve to experiment in this region. Of the nine parameters used to fit the melting curves, six were estimated from polymer measurements. The remaining three — % hyperchromicity, κ, and $\Delta H^{\circ}_{\kappa}$ (the enthalpy of initiation) — were chosen separately for each molecule at each concentration so that coincidence between theory and experiment was obtained.

Table 2.3 lists the choices of the parameters $\Delta H^{\circ}_{\kappa}$ and κ in columns 3 and 4. $\Delta H^{\circ}_{\kappa}$ varies almost randomly between 1.5 and 17.0 kcal/mole. For example, for N = 8, a change of concentration from 1.2×10^{-3} M to 0.9×10^{-3} M results in a change of almost 8 kcal/mole in the value of this parameter. Similarly, a small change in concentration necessitates varying κ by a factor of 2 to 10 in order to fit the melting curves. There is no physical justification for κ or $\Delta H^{\circ}_{\kappa}$ to vary so markedly and in such a nonregular manner from one oligomer to the next in the series or from one concentration to the next.

We conclude that, while this approach is useful in fitting experimental data, the numbers assigned to the thermodynamic

Table 2.3

Parameters Used by Pörschke to Fit $A_n + U_n$ Melting Curves[16]

Chain length	Concentration (moles/l)	ΔH_K° (kcal/mole)	κ (1/mole)
8	2.3×10^{-3}	3.5	6.2×10^{-4}
	1.2×10^{-3}	13.8	65×10^{-4}
	0.9×10^{-3}	6.0	5.9×10^{-4}
9	1.3×10^{-3}	9.9	11×10^{-4}
	0.9×10^{-3}	2.4	3.5×10^{-4}
10	1.0×10^{-3}	5.0	111×10^{-4}
	0.5×10^{-3}	4.6	6.3×10^{-4}
11	0.8×10^{-3}	7.8	10×10^{-4}
	0.4×10^{-3}	17.1	47×10^{-4}
14	0.9×10^{-3}	9.9	11×10^{-4}
	0.4×10^{-3}	12.0	13×10^{-4}
	0.2×10^{-3}	1.5	3.6×10^{-4}
18	0.9×10^{-3}	10.0	5.7×10^{-4}
	0.1×10^{-3}	6.4	5.1×10^{-4}

parameters do not correspond well to fundamental physical processes. It is likely that the $A_n + U_n$ system is not easily amenable to theoretical analysis in part because of the complex electrostatic interactions which are not very well accounted for by the simple electrostatic correction term which the authors have used.[4]

c. A_nU_n: All-or-None Analysis

Martin et al., who first synthesized the A_nU_n molecules and performed the melting experiments, analyzed their data in terms of the all-or-none model (equations 2.15 through 2.17). In Table 2.4, we show the enthalpy of an A-U base pair calculated from the all-or-none model from the concentration dependence of T_m and slope of the melting curve. Comparing Tables 2.2 and 2.4, we note

Table 2.4

Enthalpy of an A-U Base Pair Calculated from the All-or-None Model for A_nU_n[2]

Molecule	ΔH° from concentration dependence of T_m	ΔH° from the slope of the melting curve
A_4U_4	7.3 kcal/mole	6.6 kcal/mole
A_5U_5	8.0	5.9
A_6U_6	7.7	-
A_7U_7	8.25	5.5
Average	7.8 ± 0.4 kcal/mole	6.0 ± 0.6 kcal/mole

that the discrepancy between the enthalpies calculated by the two methods is even greater for the all-or-none model than if intermediate states are included in the calculation. This is not surprising,

since the inclusion of intermediate states in a calculation should tend to broaden the calculated melting curves. This comparison suggests that partially bonded states should be included in the calculation but that perhaps the statistical weights which have been assigned are in error.

d. $A_n U_n$ and Acid Oligo A_n: Statistical Mechanical Stacking Model

Recently Appleby and Kallenbach[17] proposed a model for base pair formation which involves a statistical mechanical treatment of the single strand stacking process. The theory assumes that each base on the single strand may be in one of two forms, stacked or unstacked, independent of the states of the other bases in the strand. Also, in the partially formed helices, those bases not involved in base pairs are allowed to be either stacked or unstacked on the single strand. The resultant partition function for acid oligo A_n is of the form†

$$Q = \kappa \sum_{i=1}^{N/2} \sum_{j=1}^{N/2} \frac{s^{(N-i-j+1)} (1+\rho)^{2(i+j-2)} G_{ij}}{(1+\rho)^{2(N-1)}} \qquad (2.18)$$

where s and κ are as previously defined, $\rho = e^{-\Delta G^\circ_{ss}/RT} = e^{-(\Delta H^\circ_{ss} - T\Delta S^\circ_{ss})/RT}$; ΔG°_{ss}, ΔH°_{ss}, ΔS°_{ss} are the standard free energy, enthalpy and entropy for the formation of a stacked base from an unstacked base. G_{ij} is the degeneracy factor appropriate for the staggered species of A_n. The term $(1+\rho)^{2(N-1)}$ in the denominator

---

†See reference 17 for the derivation of this equation.

refers to the single strand and implies that its statistical weight may change with temperature as stacked bases are formed or broken. Similarly, the factor $(1 + \rho)^{2(i+j-2)}$ in the numerator counts the single stranded bases in all partially formed intermediate states. A slightly modified partition function was used for the A_nU_n molecules, since Appleby and Kallenbach assumed that U_n does not stack in the single strand.

Thermodynamic parameters were derived by fitting calculated melting curves with the experiental results. The parameters derived from the calculation for A_nU_n are:

$\Delta H^\circ = -7.5$ kcal/mole base pair†

$\Delta H^\circ_{ss} = -3.6$ kcal/mole stack

$\Delta S^\circ_{ss} = -10.3$ entropy units/mole stack

$\kappa = 7.6 \times 10^{-3}$ liters/mole

The approach of Appleby and Kallenbach may be regarded as the only one yet proposed which achieves a measure of success in resolving the inconsistency in the application of helix-coil theory to oligonucleotides. We defer discussion of the results of the calculation until later, so that we may compare them with those of this work.

2. Other Simple Models are Considered

In order to advance the theory, we consider in depth the

---

†ΔH° here is not exactly the same as the enthalpy which we have defined as the addition of a double stranded stacking interaction to a duplex. The authors term it the "apparent enthalpy" of this reaction; because of the single stranded stacking interaction which is included in the term, it is temperature dependent. The value quoted is at 95°C. ΔS° can be calculated from this and it will vary with temperature.

approximations inherent in the earlier application of the helix-coil theory (Chapter 2, sections A1 and A3). These approximations may be classified into those for which sufficient information is not available to dispute them (*i*, *ii*, and *vii*), those which have been proved valid (*iv*), and the remaining ones which can be analyzed by direct calculation (*iii*, *v*, and *vi*).

Approximations *i* and *ii*, which treat the formation of base pairs as a two state process and ignore non-nearest neighbor interactions, both serve to save the theory from becoming hopelessly complex. Their removal would involve the addition of a large number of parameters for which experimental information is completely absent. Their basic justification is that they have been successful in explaining the melting of polymers.

Approximations *vii*, which assumes that the entropy of base pair formation for the oligomer is the same as for the polymer, is a useful simplification which does not profoundly influence the results of the calculation. The reason for this is that the entropy term does not introduce a factor in the partition function which is dependent on temperature. As such, a slight error in this approximation would in no way account for the discrepancy between theory and experiment.

Approximation *iv*, which ignores internal loops, is easily tested on the basis of a reasonable loop weighting function; for short oligomers, it is fully justified.

The remaining approximations are of great interest and considered in some detail.

a. Variation of ΔH° and ΔS° with Temperature: A Test of Approximation *iii*

If ΔH° is dependent on temperature during the melting process, the physical basis is most likely to change in conformation of the single strand with temperature. The calorimetric and optical measurements on polyA:polyU indicate that the double stranded moiety is little affected by temperature after it has been fully formed.[18] The simplest assumption about the single strand is that its enthalpy changes linearly with temperature. This is shown for a given choice of the heat capacity in Fig. 2.5a. Note that the enthalpy of the single strand is negative at low temperatures if one sets the zero enthalpy at higher temperatures. This is consistent with the notion that the stacking of the single strand produces a stabilizing effect.

We can then write, following Fig. 2.5,

$$\Delta H^\circ(T) = H^\circ_{ds}(T) - H^\circ_{ss}(T)$$

where,

$$H^\circ_{ds}(T) = H^\circ_{ds}(T_o) + C_p^{ds}(T-T_o) = H^\circ_{ds}$$

$$H^\circ_{ss}(T) = H^\circ_{ss}(T_o) + C_p^{ss}(T-T_o)$$

The heat capacity for the reaction, ΔC_p, equals $C_p^{ds} - C_p^{ss}$, the heat capacity of the double strand minus that of the single strand. Since the heat capacity of the double strand, is taken to be zero, it follows that

$$\Delta H^\circ(T) = \Delta H^\circ(T_o) + \Delta C_p\ (T-T_o) \qquad (2.19)$$

T_o is an arbitrarily chosen temperature and the subscripts ss and ds refer to the single strand and double strand respectively. Measurements of the heat capacity, ΔC_p, for polyA: polyU have been made, but their accuracy is not good. Krakauer and Sturtevant[10] report $\Delta C_p = -8 \pm 6$ cal/degree mole in 0.018 M Na^+ at 44°C, -17 ± 33 cal/deg. mole in 0.043 M Na^+ at 51°C, and -25 ± 26 cal/deg. mole in 0.1 M Na^+ at 58°C. In K^+ at comparable concentrations, ΔC_p varies between +84 and -66 cal/deg. mole.

The simple assumptions about the heat capacity of the single and double strands fully determine the behavior of the entropy of these moieties with temperature. Since, for an isolated system at constant pressure,

$$\Delta S^\circ = \int \Delta C_p / T \, dT$$

therefore,

$$\Delta S^\circ(T) = \Delta C_p \ln T + Z$$

We can solve for Z, the constant of integration, from the fact that in the limit of the polymer melting temperature, T_m^∞, the enthalpy and entropy are related:

$$\Delta S^\circ(T_m^\infty) = \frac{\Delta H^\circ(T_m^\infty)}{T_m^\infty} = \Delta C_p \ln T_m^\infty + Z$$

Therefore,

$$Z = \frac{\Delta H^\circ(T_m^\infty)}{T_m^\infty} - \Delta C_p \ln T_m^\infty$$

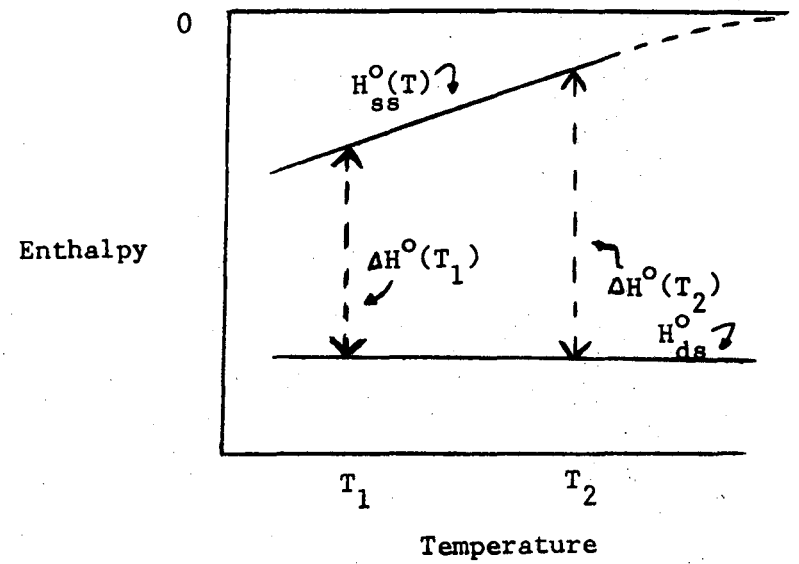


figure 2-5a

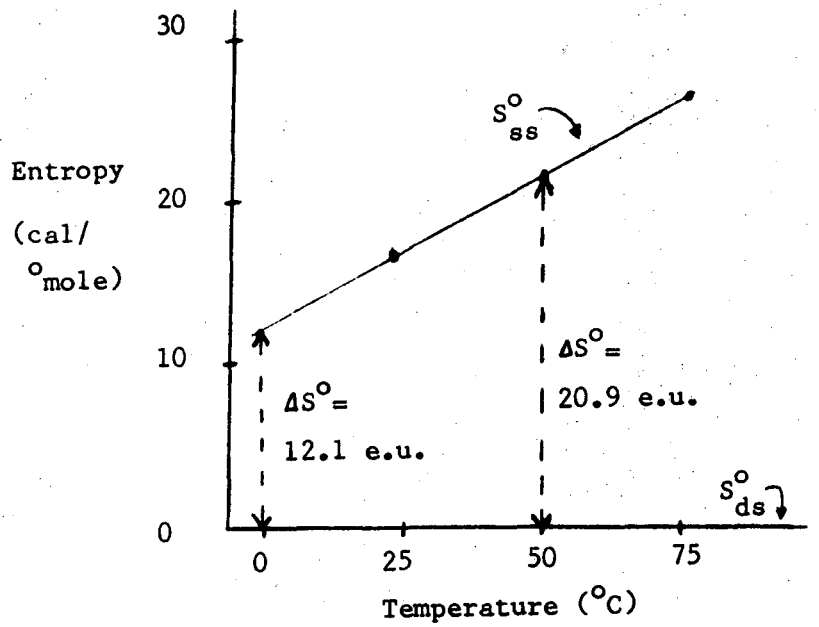


figure 2-5b

Plugging the expression for Z into the original equation for $\Delta S^\circ(T)$, we receive

$$\Delta S^\circ(T) = \Delta C_p \ln(T/T_m^\infty) + \frac{\Delta H^\circ(T_m^\infty)}{T_m^\infty} . \qquad (2.20)$$

We have plotted in Fig. 2.5b the change in entropy with temperature according to equation 2.20, to show the qualitative behavior of $\Delta S^\circ(T)$. Values for $\Delta H^\circ(T_m)$ and ΔC_p were arbitrarily chosen for illustrative purposes: $\Delta H^\circ(T_m) = -8.8$ kcal/mole and $\Delta C_p = -50$ cal/deg. mole; the assumption was explicitly made that the change in entropy with temperature was due solely to changes within the single strand.

In order to test the usefulness of the assumption of a linear heat capacity for rationalizing the experimental results with theory for the melting curves, the following calculation was performed. $\Delta H^\circ(T)$ and $\Delta S^\circ(T)$, from equations 2.19 and 2.20, were used in place of ΔH° and ΔS° in equation 2.2 for s. T_o was set to 25°C. (The results depend on the choice of T_o; we have set it near T_m of the oligomers.) Values of $\Delta H^\circ(T_o)$ equal to -7, -7.5, and -8 kcal/mole were tried. ΔC_p was taken to be -10, -20, -50, and -100 cal/deg. mole. (All except the last choice are well within the experimental range.) Using the methods of section A3, κ was varied to achieve the best fit for the melting temperatures. After the appropriate values of κ were chosen, plots of the predicted slopes of the melting curve, $(df/dT)_m$, and the concentration dependence of the T_m, $\Delta(1/T_m)/\Delta\ln(c)$, versus $\Delta H^\circ(T_o)$ were made, in a manner analogous to that of Fig. 2.4. Selected results are shown in Table

2.5. We also include the results of the calculation with $\Delta C_p = 0$. This is, of course, identical to the earlier calculations in which ΔH° and ΔS° did not vary with temperature.

Table 2.5 makes clear that the choice of a heat capacity within experimental limits does not affect the calculated melting curves greatly. Agreement between theory and experiment is worsened if ΔC_p less than zero is chosen: The discrepancy between the two predicted enthalpies increases from 1.3 kcal/mole for $\Delta C_p = 0$ to 1.6 kcal/mole for $\Delta C_p = -50$ cal/deg. mole. Furthermore, agreement between calculated and experimental melting temperatures becomes significantly worse as ΔC_p becomes increasingly negative, as measured by the standard deviation between predicted and actual melting temperatures: $\sigma_m = \pm 0.7^\circ C$ for $\Delta C_p = 0$ and $\pm 2.3^\circ C$ for $\Delta C_p = -50$ cal/deg. mole.

Although ΔC_p greater than zero improves agreement between theory and experiment very slightly, no physical understanding of stacking is consistent with such a choice. We conclude therefore that a simple variation of the thermodynamic functions with temperature throughout the melting transition does not provide an explanation of the inadequacy of the helix-coil theory when applied to oligonucleotides.

b. Letting $\Delta H^\circ_\kappa \neq 0$ Makes Little Difference: A Test of Approximation *vi*

We have hitherto taken ΔH°_κ equal zero on the assumption that the first base pair, which is devoid of a double strand stacking interaction, has a very small enthalpy associated with it. Although this assumption is reasonable,[9] we have performed calcula-

Table 2.5

Results of Calculation Which Allows the

Enthalpy to Change with Temperature

Molecule	$\Delta H^\circ(T_o)$ from concentration dependence of T_m (in kcal/mole)		
	$\Delta C_p = 0$	$\Delta C_p = -30$	$\Delta C_p = -50$ cal/deg. mole
A_4U_4	-7.8	-8.35	-8.95
A_5U_5	-8.15	-8.35	-8.45
A_6U_6	-7.9	-7.7	-7.6
A_7U_7	-8.15	-7.85	-7.6
Average	-8.0	-8.05	-8.15

Molecule	$\Delta H^\circ(T_o)$ from the slope of the melting curve at T_m (kcal/mole		
	$\Delta C_p = 0$	$\Delta C_p = -30$	$\Delta C_p = -50$ cal/deg. mole
A_4U_4	-7.0	-7.2	-7.5
A_5U_5	-6.7	-6.5	-6.5
A_6U_6	--	--	--
A_7U_7	-6.5	-6.0	-5.6
Average	-6.7	-6.6	-6.5

tions with ΔH°_κ varying between +5 and -5 kcal/mole to show the effect of such a variation.

Choosing ΔH°_κ = -5 kcal/mole, for example, we calculate that for A_5U_5 the concentration dependence yields ΔH° = -7.6 kcal/mole A-U stacking interaction rather than -8.15 kcal/mole when ΔH°_κ = 0. That is to say, the total enthalpy for the entire reaction from the concentration dependence of T_m remains roughly constant. Because ΔH° is decreased, intermediate states contribute slightly more to the partition function, the theoretical melting curve is broadened somewhat, and agreement between theory and experiment is, in principle, improved. In practice, the change in $(df/dT)_m$ is so small that, after the enthalpy of initiation is subtracted, agreement of the enthalpies derived from $(df/dT)_m$ and the concentration dependence of T_m is not observably improved.

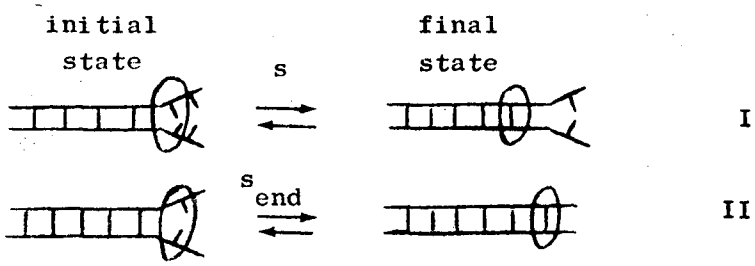
3. A Model with Partially Frayed Ends is Proposed

The final assumption of the theory yet to be tested is that the equilibrium constant for the addition of a base pair to a helix is independent of the position of the pair formed. The question of the A-U juncture in the center of the A_nU_n molecules is similar to discussion of ΔH°_κ above: the effect is small compared with the total enthalpy and entropy of the molecule and cannot be detected by direct calculation.

End effects are a very different matter. Although the ends only account for two of seven to thirteen double stranded stacking interactions in the A_nU_n series, their effect on the shape of the melting curve can be significant. At all temperatures in the calculation based on the standard helix-coil theory, the equilibrium is

dominated by two forms: the fully formed duplex and the single stranded species. If, in fact, intermediate species are more important in the process than the theory suggests, then the calculated melting curves would be broadened considerably. If end effects are present and significant, they could increase the contribution of intermediate species to the melting process, thereby broadening the curve.

That an evaluation of end effects is a logical step for oligomers is suggested by considering the interactions in detail. All recent formulations of the theory have incorporated the idea that the first base pair (the initiation step) should be treated differently from the rest since it is devoid of a stacking interaction. In the same way, the formation of the end base pair is different from the rest because its final environment is significantly perturbed relative to the other base pairs. Consider the equilibria



The initial and final states of the base pair which is formed is circled. In I, the equilibrium constants refer to the ratio: (final state: base pair having on one side a base pair and on the other side free or stacked bases)/(initial state: two bases with one base pair on one side and free or stacked bases on the other). The initial and final states are the same for the formation of all

base pairs in the helix except the first and end pairs. In II, s_{end} refers to: (final state: base pair having on one side a base and on the other side water)/(initial states: consisting of two bases with one base pair on one side and water on the other). It is important to note that even within the framework of the nearest neighbor approximation, the ends are expected to exhibit a different stability from that of the rest of the molecule. Only if a base pair has the same affinity for water as for an adjacent single stranded base will s_{end} equal s.

a. Methodology and results

It remains to test this model with experiment. We let $s_{end} = k_{end}s$, where $k_{end} = e^{-\Delta G^\circ_{end}/RT}$ is the equilibrium constant related to the differential stability of the two end base pairs. (ΔG°_{end} is the increment in free energy due to the end effect.) If $k_{end} < 1$, as it must in order that the weighting of the intermediate states be increased, then the end double stranded stacking interaction is less stable than all similar internal stacking interactions. This is the expected result, since the interaction of the end base pair with water molecules should be destabilizing.[19] The statistical weights of the intermediate states shown in Fig. 2.1 are no longer appropriate for this model; Fig. 2.6 illustrates the new ones.

The method of calculation is straightforward, although somewhat laborious. Equations 2.5 and 2.6 are modified to accomodate the end effect:

$$Q = \sum_{\text{all states } i} \kappa k_{end}^{p} s^{i-1} = \sum_{i-1}^{N} (N+1-i)\kappa k_{end}^{p} s^{i-1} \qquad (2.6b)$$

$$f = \phi\kappa \sum_{\text{all states } i} \frac{(i-1)k_{end}^{p} s^{i-1}}{(N-1)Q} \qquad (2.7a)$$

$$= \phi\kappa \sum_{i-1}^{N} \frac{(N+1-i)(i-1)k_{end}^{p} s^{i-1}}{(N-1)Q} \qquad (2.7b)$$

p = the number of fully formed ends in the given species (0, 1, or 2) and all other quantities are defined earlier. "All states" refers to those states pictured in Fig. 2.6.

The calculation was carried out as follows. For each value of ΔH° chosen, both κ and k_{end} were allowed to vary simultaneously until the best possible fit between experimental and calculated melting temperatures and slopes of the melting curves at T_m were achieved. These two conditions are sufficient to define unambiguous values of κ and k_{end}. In general, approximately 10 trials were adequate to determine κ and k_{end} for a given calculation.†

We obtained, for example, $k_{end} = 0.142$ and $\kappa = 1.42 \times 10^{-3}$ l/mole for $\Delta H^{\circ} = -8$ kcal/mole. In this manner we have fit the slope of the melting curve by varying the relative fraction of

---

†The computer program was written for the teletype machine at the Lawrence Berkeley Laboratory so that the programmer and computer could "interact." Upon receiving the results of a previous calculation, the programmer was able to punch in new trial values. In this way, both κ and k_{end} (or any other parameters) could be varied simultaneously. The program is listed in Appendix I.

Agreement was deemed adequate when $[\Sigma(T_m - T_m(\text{expt}))]/M\ T_m$ and $[\Sigma((df/dT)_m - (df/dT)_m(\text{expt}))]/M\ (df/dT)_m$ were less than 0.5%. That is, approximately equal numbers of the calculated quantities were above and below the experimental values. M is the number of terms in the respective sums. The former sum is over all $A_n U_n$ molecules at three concentrations and the latter is at only one concentration.

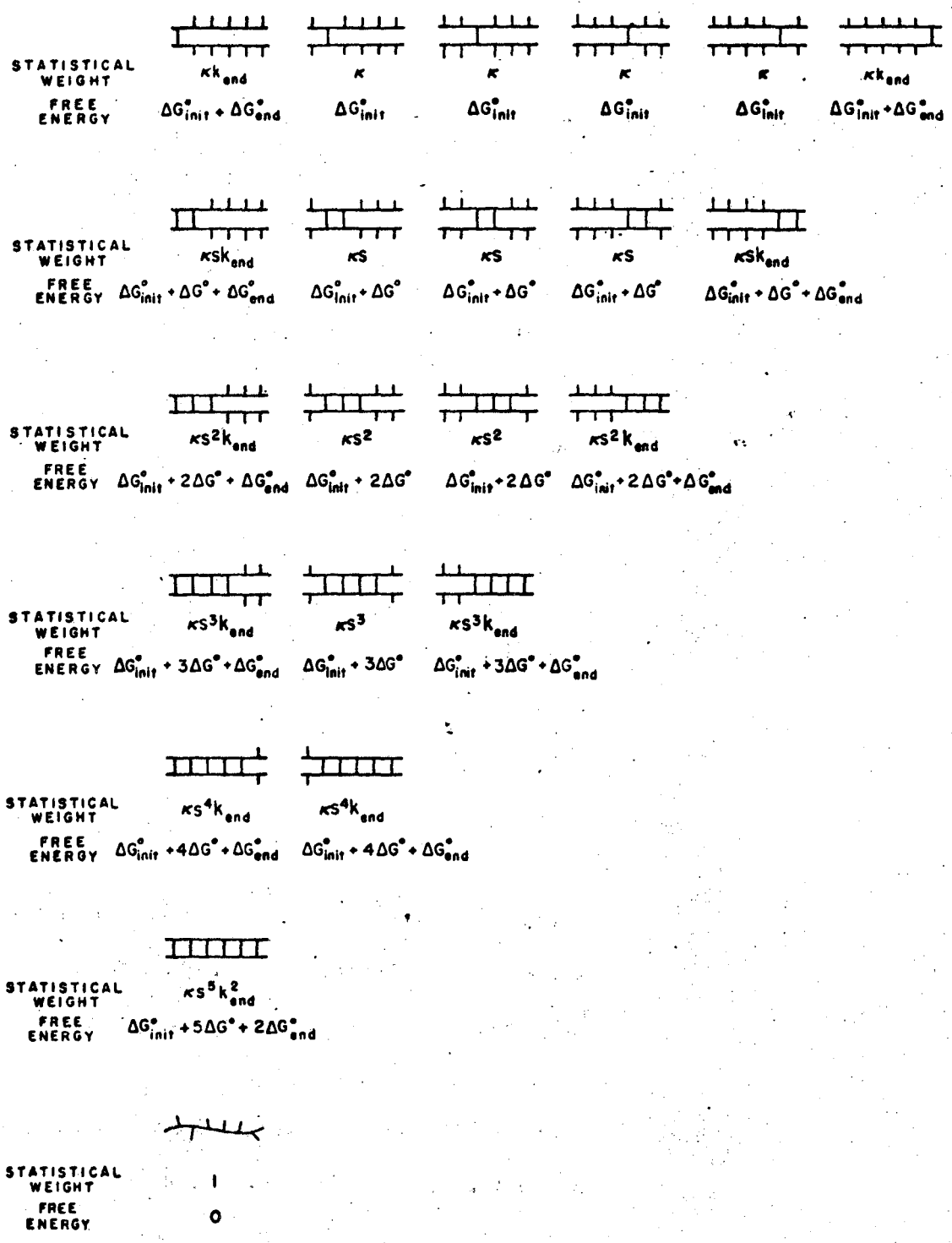


Figure 2-6

intermediate forms with unformed ends. We then asked the question: given the value of k_{end} so determined, what value of ΔH° fits the concentration dependence of T_m? In order to answer this question, we have held k_{end} constant, incremented ΔH° in units of 0.5 kcal/mole and varied κ until the melting temperatures were fit. When the results are plotted as in Fig. 2.4b and the intersection with experimental quantities read from the plot, the enthalpy consistent with $k_{end} = 0.14$ which gives the correct concentration dependence of T_m is then determined. Since the latter enthalpy, from the concentration dependence of T_m, is different from the enthalpy which is consistent with $df/dT|_m$, the procedure is repeated for values of ΔH° between -6.5 and -9.0 kcal/mole (in increments of 0.5 kcal/mole).

We summarize the steps in the calculation:

Step 1: Choose ΔH°. (Begin with ΔH° = -6.5 kcal/mole.)

Step 2: Vary κ and k_{end} until experimental T_m's and $df/dT|_m$'s are fit.

Step 3: Using the value of k_{end} determined in steps 1 and 2, find values of κ which fit each ΔH° from -6.5 to -9.0 kcal/mole (in increments of 0.5 kcal/mole).

Step 4: Plot results of the calculations in step 3 as in Fig. 2.4b: ΔH° vs. $\Delta(1/T_m)/\Delta\ln(c)$. This determines ΔH° which best fits the concentration dependence of T_m at this value of k_{end}.

Step 5: Repeat steps 1 to 4, incrementing initial choice of ΔH° by 0.5 kcal/mole.

The result at step 2 is a set of parameters in which ΔH°, κ, and k_{end} have been determined to fit the melting temperatures and $df/dT|_m$'s of the A_nU_n molecules. We may call this enthalpy ΔH_1°. The result at step 4 is a set of ΔH°, κ, and k_{end} chosen to fit the melting temperatures and the concentration dependence of T_m. We call this enthalpy ΔH_2°. Since k_{end} has the same value for steps 1 and 2 as for steps 3 and 4, ΔH_1° should equal ΔH_2°. Steps 1 through 4 are repeated with different initial choices of ΔH° (which determines a different k_{end}) until this condition is met. The results, a series of ΔH_1° and ΔH_2° values for each of several k_{end} values, are plotted in Fig. 2.7. The magnitude of k_{end} for which ΔH_1° equals ΔH_2° satisfies both the concentration dependence of T_m and the slopes of the melting curves at their midpoint. The results, from Fig. 2.7, are:

Table 2.6

Thermodynamic Parameters for the Frayed End Model of A_nU_n

$\Delta H^\circ = -8.75$ kcal/mole

$k_{end} = 0.058$

$\kappa = 6.55 \times 10^{-4}$ liters/mole

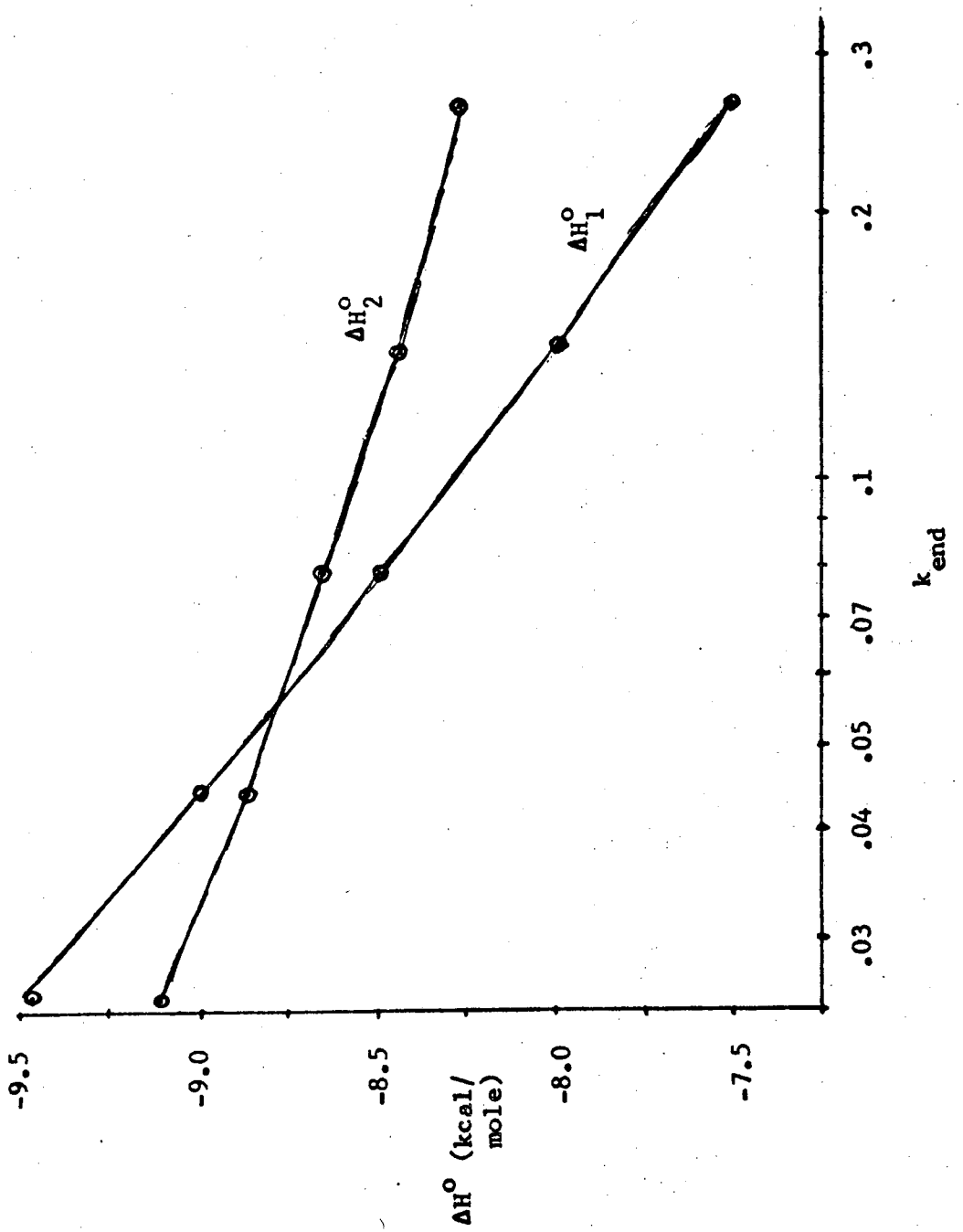


Figure 2-7

Since these three quantities will find much application in the remainder of this work, it is important to ask how accurately they are known. Assuming the suitability of the frayed end model, there are two main sources of error in ΔH°: (1) uncertainty in the enthalpy associated with initiation and with the A_nU_n juncture and (2) the imprecision of experimental determinations of T_m and $df/dT|_m$. We can estimate that the error due to (1) is at most ±6 kcal/mole for the entire molecule or approximately ±0.5 kcal/mole of double stranded stacking interaction. An enthalpy greater than this would show up as a systematic variation in ΔH° for the molecules of varying N in the series.

The accuracy of T_m is ±0.5°C.[2] A systematic error of as much as 2°C would have a small effect on the calculated enthalpy, considerably less than ±1 kcal/mole of molecules. There is however an uncertainty in the experimental values of $df/dT|_m$: the accuracy of $df/dT|_m$ is difficult to assess for the A_nU_n series, since the original melting curves are no longer available; for the molecules A_nGCU_n (n = 2, 3, 4) and A_nCGU_n (n = 2 and 4), melting curves at five or more oligonucleotide concentrations are available. Although $df/dT|_m$ should decrease slightly with increasing oligonucleotide concentration (see equations 2.11 and 2.12), no systematic variation was observed. The uncertainty in $df/dT|_m$ for these molecules was slightly less than 10%. This corresponds to an uncertainty of 0.55 kcal/mole in ΔH°.

It is, of course, possible that these two sources of error will cancel. We estimate that ΔH° is known to within ±0.75 kcal/mole from our analysis.

The largest source of uncertainty in both κ and k_{end} derives directly from that of ΔH°. If $\Delta H^\circ = -8$ kcal/mole, then $\kappa = 1.4 \times 10^{-3}$ and $k_{end} = 1.5 \times 10^{-1}$; if $\Delta H^\circ = -9.5$ kcal/mole, $\kappa = 2.4 \times 10^{-4}$ and $k_{end} = 3 \times 10^{-2}$. Expressed in terms of free energy, this corresponds to ± 0.6 and ± 0.5 kcal/mole for κ and k_{end}, respectively.

b. Discussion of the Results of the Frayed End Model

As already noted, the most important contribution of the frayed end model is that it removes the inconsistency between the two theoretical evaluations of the enthalpy. We now consider the ability of the theory to duplicate known experimental results.

The calculated and corrected experimental melting curve for A_5U_5 are compared in Fig. 2.8, at a strand concentration of 6.5×10^{-6} M. For the lower 60% of the curve (below 1-f = 0.60), the frayed end model simulates the experimental behavior significantly better than the standard statistical thermodynamic model. The slope of the curve at T_m is, of course, much improved by the frayed end model, since the calculation is fit to this quantity. Above 1-f = 0.60, the new model provides only slight improvement. Both calculated curves are steeper than experiment in this region. Later results for A_nGU_n, A_nGCU_n, and A_nCGU_n (see Figs. 3.3, 3.4, and 3.5) indicate that this is not a general phenomenon for the oligonucleotides; it may simply be due to the fact that the availability of so few melting curves for the A_nU_n molecules has forced the choice of atypical experimental curves.

In Table 2.7 we show that the predicted breadths of the transitions for all of the A_nU_n molecules are much improved by

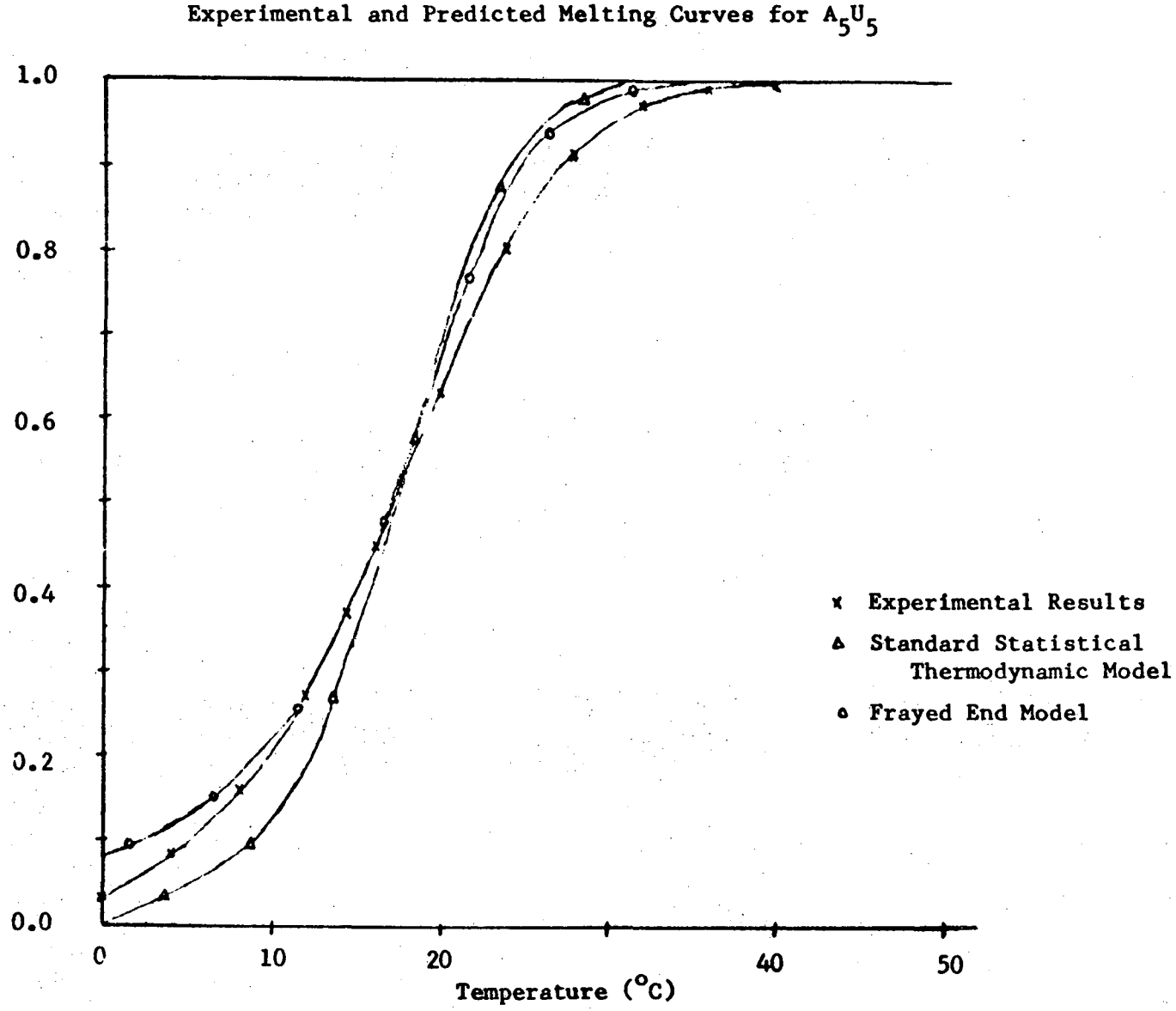


Figure 2-8

the use of the frayed end model. This is due to the increased importance of intermediate states with unformed end base pairs. The predicted breadths are less great than experiment (15% versus 40% for the standard model), but the improvement is marked. For the oligomers with G-C base pairs, the predicted breadths are generally as large as the experimental quantities.

Table 2.7

Breadths of Melting Curves for A_nU_n Molecules

Molecule	$\Delta_{2/3}$[a] experimental[b]	$\Delta_{2/3}$ standard theory	$\Delta_{2/3}$ frayed end model
A_4U_4	23°C	14.5°C	17.5°C
A_5U_5	21	12.5	18.0
A_6U_6	22	11.5	17.0
A_7U_7	17	10.0	15.3

[a] $\Delta_{2/3}$ = temperature range over which two-thirds of the melting occurs measured symmetrically about the T_m.

[b] From Martin and Uhlenbeck.[2]

In Table 2.8 we present the experimental and calculated melting temperatures for all the oligonucleotides at three different concentrations. The standard deviation of the T_m's is ±1.7°C for the frayed end model and ±0.7°C for the earlier calculation. The basic problem is that A_4U_4 is predicted to melt three to four degrees lower than observed. In order to fit all of the molecules in the series as well as possible, the appropriate value of κ shifts

Table 2.8

Experimental and Theoretical Melting Temperatures for A_nU_n Molecules

Molecule	Melting Temperatures		
	$c = 10^{-4}$ M	$c = 10^{-5}$ M	$c = 10^{-6}$ M
	Experimental Results		
A_4U_4	12.2	5.4	-1.1
A_5U_5	23.6	18.0	12.6
A_6U_6	31.9	27.0	22.2
A_7U_7	39.8	35.5	31.3
	Standard Statistical Thermodynamic Model		
A_4U_4	11.2	4.7	-1.5
A_5U_5	24.0	18.4	13.0
A_6U_6	32.8	27.9	23.2
A_7U_7	39.2	34.8	30.6
	Frayed End Model		
A_4U_4	8.9	2.7	-3.2
A_5U_5	23.6	18.0	12.6
A_6U_6	33.7	28.7	23.9
A_7U_7	40.9	36.4	32.1

the other A_nU_n's to slightly higher melting temperatures. Although the poorer agreement of melting temperatures is somewhat disappointing, the problem may be more apparent than real: the offending molecule, A_4U_4, has a measured melting temperature of 12.2°C at $c = 10^{-4}$ M and -1.1°C (extrapolated to $c = 10^{-6}$ M). At such low temperatures, much if not most of the melting curve occurs below 0°C, where it cannot be measured. As such, the experimental T_m's are not well established in this range. In short, although the frayed end model does not do as well as the standard model with respect to reproducing melting temperatures, the results are satisfactory.

An interesting experimental result was obtained for the molecules A_mU_n, where m is greater than n.[2] The T_m of the oligonucleotide increased by approximately 3 to 5°C if an extra A residue, which could not base pair, was added to the end. No information was gained if extra U's were added, since aggregation occurred. This result is consistent with the frayed end model. According to this model, the effect of adding an extra base to the helix is a change in the environment of the end of the molecule such that the end effect should not occur or should be lessened in degree (since only half of the end is protected from the solution). The result of shielding the end is a raise in T_m, since the end effect is destabilizing.

To test this hypothesis, we have calculated the melting curves of the A_mU_n molecules, using the thermodynamic parameters of Table 2.6 but replacing k_{end} with $k_{end}^{1/2}$. This assumes that half of the end effect is wiped out by the additional A residue at each

end of the duplex molecule. The result of this calculation was an increase in T_m of approximately 5 to 6°C. Thus, the disturbance of the interaction at the ends of the molecule provides a plausible, semi-quantitative explanation for the increase in the melting temperature brought about by adding an extra residue to the ends of the molecule. (It is also true that the additional A residue provides extra base pairing possibilities among the intermediate states, allowing the formation of extra intermediate states. However, it is easily shown that these additional states change the melting temperature negligibly.)

The model is not completely satisfactory in explaining the increase in T_m caused by dangling ends. As more extra A residues are added to the molecule, the T_m continues to increase. No simple explanation for this phenomenon is provided by the end effect. However, one may speculate that as the dangling end grows in size, it will tend to isolate the end of its complementary strand from the aqueous environment. If this were so, then the theory would predict that the limiting value of the melting temperature increase would be obtained by replacing k_{end} by $k^{\circ}_{end} = 1$ in the calculation, yielding an increase in T_m of 10 to 12°C. There is not sufficient experimental information to test this prediction; in any case, the prediction should not be taken too seriously, since it is likely that more complicated phenomena than just the end effect are involved, such as the contribution of single strand stacking in the dangling end which would persist even after the duplex is formed.

We conclude by comparing the results of the frayed end

model with those of the single stranded stacking model proposed by Appleby and Kallenbach and discussed in section B.1d of this chapter.[17] We first note that the thermodynamic parameters calculated by the two models are somewhat different:

Frayed End Model	Single Stranded Stacking Model
ΔH° = -8.75 kcal/mole	ΔH° = -7.5 kcal/mole (at 95°C)
	ΔH° = -6.0 kcal/mole (at 30°C)
$\kappa = 6.55 \times 10^{-4}$ 1/mole	$\kappa = 7.6 \times 10^{-3}$ 1/mole

The enthalpy from the frayed end model is in substantially better agreement with the measured enthalpy for the addition of an A-U base pair in polyA:U. From calorimetric measurements, Neumann and Ackermann[18] determined that $\Delta H^\circ = -9.3 \pm 0.5$ kcal/mole; extrapolation of the results of Krakauer and Sturtevant to 1 M salt gives an enthalpy of approximately -9 kcal/mole.[10] No independent determination of κ has been made; however, κ from the single stranded stacking calculation is very high and may not be consistent with the observation of Jaskunas *et al.*,[7] that four or more base pairs are needed to form a stable double helix in solution.

We compare the two models in their fit of the melting data for A_nU_n. Taking the predicted melting temperatures from the plot of $1/T_m$ versus 1/N-1 presented by Appleby and Kallenbach, we observe that the models exhibit nearly identical agreement with the experimental melting temperatures. Both models have the same difficulty with A_4U_4, predicting a T_m slightly lower than deduced from experimental measurements. In both models, almost all T_m's are within a few degrees of experiment. Notwithstanding this fact, the single stranded stacking model does not fit the concentration

dependence of the melting temperature especially well. This is due to the fact that the enthalpy for the addition of an A-U base pair to a duplex is low. In Table 2.9, the results of the single stranded stacking and the frayed end models are compared.

Table 2.9

Concentration Dependence of

T_m for Two Model Calculations

Molecule	$\Delta(1/T_m)/\Delta\log c$ Experiment	$\Delta(1/T_m)/\Delta\log c$ Frayed End Model	$\Delta(1/T_m)/\Delta\log c$ Single Strand Stacking Model
A_4U_4	8.5×10^{-5}	8.0×10^{-5}	10.8×10^{-5}
A_5U_5	6.3×10^{-5}	6.4×10^{-5}	7.9×10^{-5}
A_6U_6	5.4×10^{-5}	5.4×10^{-5}	6.4×10^{-5}
A_7U_7	4.4×10^{-5}	4.6×10^{-5}	5.8×10^{-5}

The final comparison between the two models is the predicted breadth of the melting curves. This is shown in Table 2.10.

Table 2.10

Breadths of Melting

Curves for Model Calculations

Molecule	$\Delta_{2/3}$ Experiment	$\Delta_{2/3}$ Standard Helix-Coil	$\Delta_{2/3}$ Frayed End	$\Delta_{2/3}$ Single Strand Stacking
A_5U_5	21°C	12.5°C	18.0°C	16.5°C
A_7U_7	17°C	10.0°C	15.3°C	12.5°C

The frayed end model is in somewhat better agreement with experimental breadths of melting curves than the single strand stacking model. Both models predict broader curves that the standard helix-coil theory and neither yield curves which are as broad as experiment. For A_5U_5 and A_7U_7, the single stranded stacking model calculates curves having 77% of the observed breadths; the frayed end model predicts melting curves of 88% of the observed breadths.

In summary, the frayed end model fits the experimental data somewhat better than the single stranded stacking model, and with fewer parameters. An additional advantage of the frayed end model is that it is simple enough that it can be extended to duplexes with one or more G-C base pairs. This we do in the following chapter. Nonetheless, it is possible, as Appleby and Kallenbach have suggested that single strand stacking is an important part of the reaction of RNA single strands to form duplex oligoribonucleotides. A complete theory would consider both end effects and single strand stacking or temperature dependence of ΔH° and ΔS°. On the basis of the better agreement between theory and experiment for the frayed end model, it appears that the end effect has the greater influence on the equilibrium.

CHAPTER 2

REFERENCES

1. O. C. Uhlenbeck, F. H. Martin, and P. Doty, J. Mol. Biol., 57, 217-229 (1971).
2. F. H. Martin, O. C. Uhlenbeck, and P. Doty, J. Mol. Biol., 57, 201-215 (1971).
3. P. Borer, Ph.D. Thesis, University of California, Berkeley (1972).
4. R. L. Baldwin, personal communication.
5. B. H. Zimm, J. Chem. Phys., 33, 1349 (1960).
6. J. Applequist and V. Damle, J. Am. Chem. Soc., 87, 1450-1458 (1965).
7. S. R. Jaskunas, C. R. Cantor, and I. Tinoco, Jr., Biochemistry, 7, 3164-3178 (1968).
8. D. Pörschke and M. Eigen, J. Mol. Biol., 62, 361 (1971).
9. I. E. Scheffler, E. L. Elson, and R. L. Baldwin, J. Mol. Biol., 48, 145-171 (1970).
10. H. Krakauer and J. M. Sturtevant, Biopolymers, 6, 491-512 (1968).
11. R. C. Davis and I. Tinoco, Jr., Biopolymers, 6, 223-242 (1968).
12. R. K. Nesbet, Biopolymers, Symposia No. 1, 129-139 (1968).
13. I. Tinoco, Jr., J. Am. Chem. Soc., 82, 4785 (1960).
14. W. Rhodes, J. Am. Chem. Soc., 83, 3609 (1961).
15. H. DeVoe and I. Tinoco, Jr., J. Mol. Biol., 4, 518 (1962).
16. D. Pörschke, Biopolymers, 10, 1989 (1971).
17. D. W. Appleby and N. R. Kallenbach, Biopolymers, 12, 2093 (1973).
18. E. Neumann and T. Ackermann, J. Phys. Chem., 7, 2170-2178 (1969).
19. W. F. Harrington and J. A. Schellman, Compt. Rend. Trav. Lab. Carlsberg Ser. Chim., 30, 31 (1956).

CHAPTER 3

MELTING OF DOUBLE STRANDED OLIGONUCLEOTIDES WITH G-C BASE PAIRS

If we wish to predict stabilities and melting curves of RNA molecules, it is essential that we have knowledge of the thermodynamics of G-C base pairs. Surprisingly little is known about this subject. For polynucleotides, it has been long known that the melting temperature increases with increasing mole fraction of G-C base pairs.[1] This information leads only to a qualitative assessment of the stability of G-C base pairs. On the basis of polymer data and simple assumptions, Tinoco, et al., have taken the free energy of a G-C base pair to be approximately twice that of an A-U base pair at 25°C in 1 M Na^+.[2] This assignment has been accepted by other workers as well.[3]

Coutts has isolated a hairpin loop from t-RNA with a helical region composed of four G-C base pairs.[4] From a van't Hoff plot (which assumes the all-or-none model of the multiple equilibria), he estimated an enthalpy of -11 or -14.5 kcal/mole in 0.11 M Na^+, depending upon whether he attributes the enthalpy to the four G-C base pairs or to the three double stranded stacking interactions. Uhlenbeck, Martin and Doty's analysis of the melting of RNA oligomers with one and two G-C base pairs also indicates that the enthalpy of a G-C base pair is greater than that of an A-U base pair.[5] Gralla and Crothers have analyzed the data of Uhlenbeck et al., using an all-or-none approximation and have calculated that a G-C base pair adjacent to a G-C base pair has a double stranded stacking enthalpy of -12.65 kcal/mole and a G-C base pair

next to an A-U pair has a double stranded stacking enthalpy of -7.45 kcal/mole.[6]

These few words summarize the extent of our present knowledge of the thermodynamics of G-C base pairs in RNA. It is the task of this chapter to refine this knowledge, using the methods of the previous chapter and the experimental data of several workers.

A. Basic Theory for G-C Base Pairs

The intermediate states shown in Figs. 2.1 and 2.5 are also appropriate for molecules with one or more G-C base pairs. What must be changed are the statistical weightings of these states. In a manner exactly analogous to our treatment of A-U base pairs, we let

$$s_G = e^{-(\Delta H^\circ_G - T\Delta S^\circ_G)/RT} \tag{3.1a}$$

where s_G is the equilibrium constant for the addition of a G-C base pair to a growing helix and ΔH°_G and ΔS°_G are the corresponding enthalpy and entropy. It is useful to rewrite this expression as

$$s_G = k_G\, e^{-(\Delta H^\circ - T\Delta S^\circ)/RT} \tag{3.1b}$$

where s, ΔH° and ΔS° relate to the addition of an A-U base pair and are defined as in Chapter 2, and

$$k_G = e^{-[(\Delta H^\circ_G - \Delta H^\circ) - T(\Delta S^\circ_G - \Delta S^\circ)]/RT} \tag{3.2}$$

is the factor in the partition function for the increased stability of a G-C base pair relative to an A-U pair. It should be noted

that k_G is independent of temperature only if the enthalpy for the addition of a G-C base pair is equal to that for the addition of an A-U pair to that helix.

At this point, great care must be taken in the definition of terms. When only A-U base pairs were present, we were able to avoid the question of whether the base pair or the double stranded stacking interaction was the more important source of stabilization of the double stranded region. If the former is correct, then it is entirely adequate to add a factor of $k_G s$ to the partition function for each G-C pair in the helix.

On the other hand, if the double stranded stacking interactions are the more important in stabilizing the double stranded region, then the factor in the partition function which accounts for the increased stability of the helix due to a given G-C base pair must depend on the neighboring base pairs, as discussed in Chapter 1. In order to take this possible sequence dependence of the free energy into account, we modify our formalism. We now write k_{IJ} in place of k_G, where the indices I and J specify the double stranded stacking interaction. For example, for the molecule A_2GCU_2 / U_2CGA_2, the partition function would include factors $k_{AG}s$, $k_{GC}s$, and $k_{CU}s$ for the three double stranded stacking interactions involving a G-C base pair. (Since the AG and CU interactions are the same, the factors may also be written $(k_{AG}s)^2$ and $k_{GC}s$.) The term $k_{AG}s$ refers not to an A-G base pair, which is not known to exist, but rather to the $\begin{matrix} AG \\ UC \end{matrix}$ double stranded stacking interaction. By analogy with the definition of k_G,

$$k_{AG} = e^{-(\Delta H^\circ_{AG} - \Delta H^\circ)/RT} \; e^{+(\Delta S^\circ_{AG} - \Delta S^\circ)/R}$$

$$= k^\circ_{AG} \, e^{-(\Delta H^\circ_{AG} - \Delta H^\circ)/RT} \qquad (3.2a)$$

where ΔH°_{AG} and ΔS°_{AG} are the enthalpy and entropy of the above stacking interaction relative to the single strand. k°_{AG} is the temperature independent (entropic) part of k_{AG}. In all cases, the first subscript refers to the base nearer the free 5'-OH end of the nucleic acid strand.†

Figure 3.1 illustrates the use of the k_{IJ} terms in specifying the statistical weights of the intermediate species of A_2GCU_2, in which the end effect is ignored. Figure 3.2 shows the statistical weights, in which the end effects are included.

Because of the complementary nature of the double stranded nucleic acid, there are ten rather than sixteen different double stranded interaction terms. They are listed in the following array:

$k_{AA} \equiv k_{UU}$	$k_{AC} \equiv k_{GU}$	$k_{AG} \equiv k_{CU}$	k_{AU}
$k_{CA} \equiv k_{UG}$	$k_{CC} \equiv k_{GG}$	k_{CG}	
$k_{GA} \equiv k_{UC}$	k_{GC}		
k_{UA}			

Of these, we have implicitly set $k_{AA} = 1$ in the calculation of the

---

†Using this notation, the term ΔH° of Chapter 2 is now written ΔH°_{AA} and may be equivalently termed ΔH°_{UU}. ΔH°_{AA} is henceforth used.

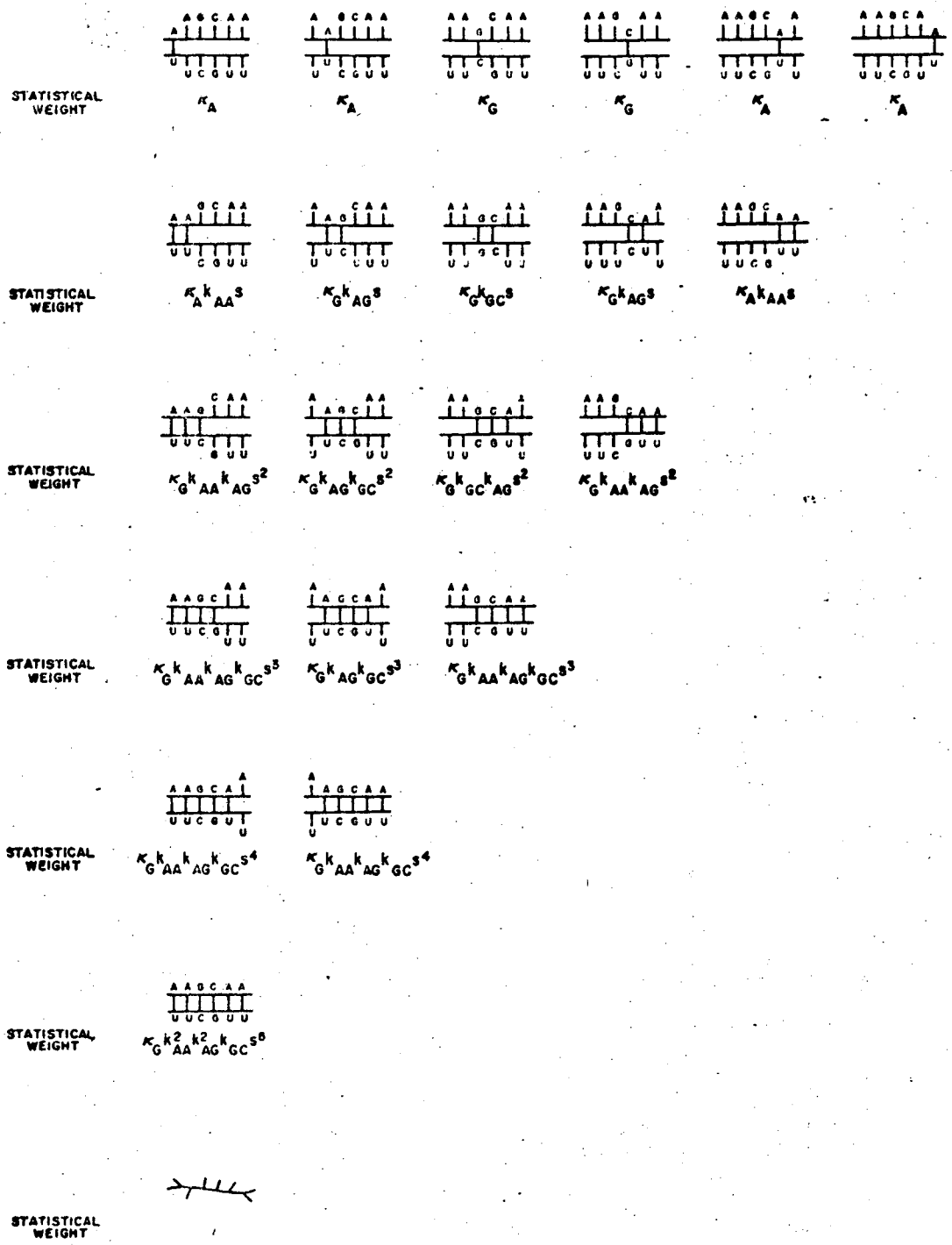

STATISTICAL WEIGHT
κ_A
κ_A
κ_G
κ_G
κ_A
κ_A
STATISTICAL WEIGHT
$\kappa_A k_{AA} s$
$\kappa_G k_{AG} s$
$\kappa_G k_{GC} s$
$\kappa_G k_{AG} s$
$\kappa_A k_{AA} s$
STATISTICAL WEIGHT
$\kappa_G k_{AA} k_{AG} s^2$
$\kappa_G k_{AG} k_{GC} s^2$
$\kappa_G k_{GC} k_{AG} s^2$
$\kappa_G k_{AA} k_{AG} s^2$
STATISTICAL WEIGHT
$\kappa_G k_{AA} k_{AG} k_{GC} s^3$
$\kappa_G k_{AG} k_{GC} s^3$
$\kappa_G k_{AA} k_{AG} k_{GC} s^3$
STATISTICAL WEIGHT
$\kappa_G k_{AA} k_{AG} k_{GC} s^4$
$\kappa_G k_{AA} k_{AG} k_{GC} s^4$
STATISTICAL WEIGHT
$\kappa_G k_{AA}^2 k_{AG}^2 k_{GC} s^6$
STATISTICAL WEIGHT

Figure 3-1

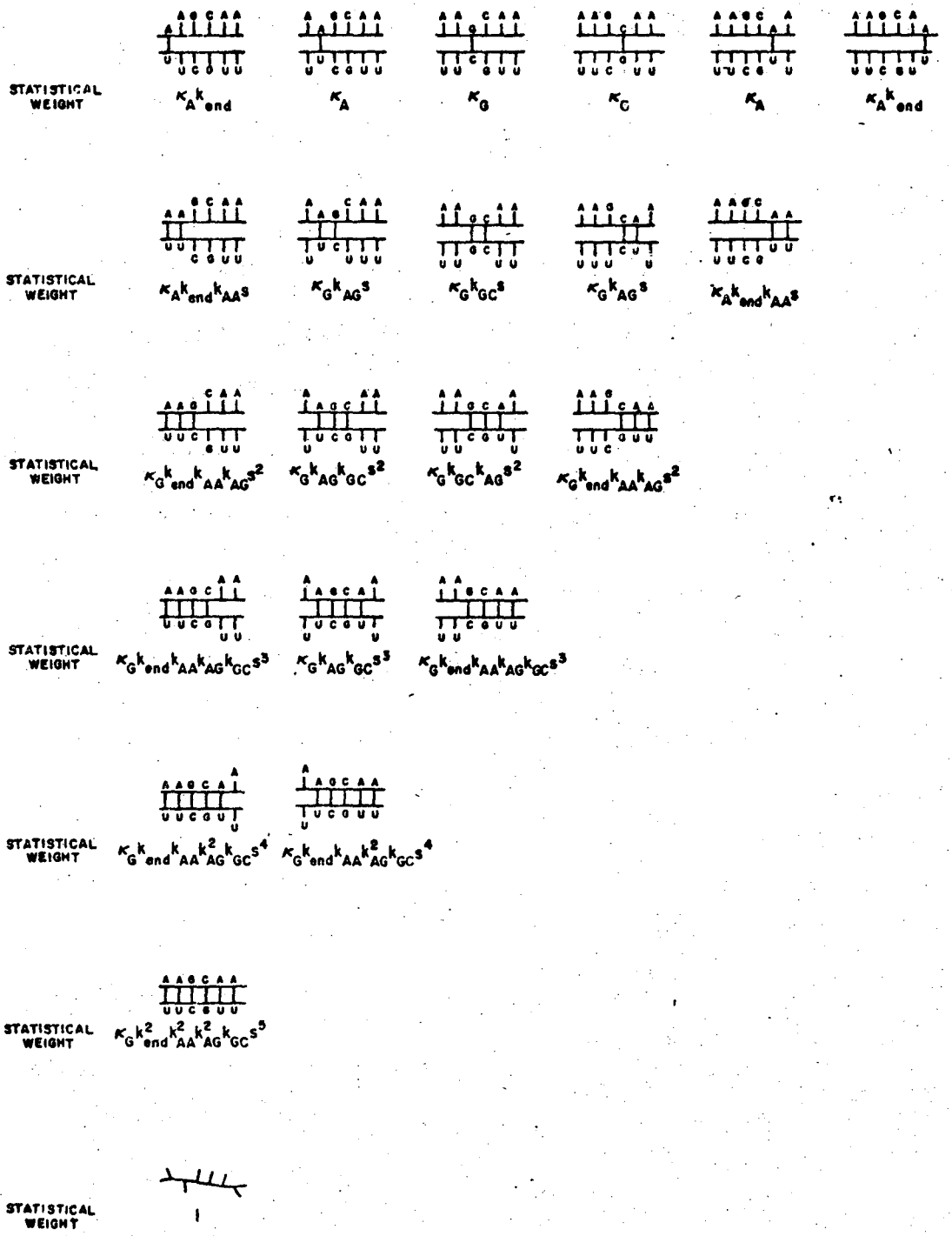

STATISTICAL WEIGHT
$\kappa_A k_{end}$
κ_A
κ_G
κ_C
κ_A
$\kappa_A k_{end}$
STATISTICAL WEIGHT
$\kappa_A k_{end} k_{AA} s$
$\kappa_G k_{AG} s$
$\kappa_G k_{GC} s$
$\kappa_G k_{AG} s$
$\kappa_A k_{end} k_{AA} s$
STATISTICAL WEIGHT
$\kappa_G k_{end} k_{AA} k_{AG} s^2$
$\kappa_G k_{AG} k_{GC} s^2$
$\kappa_G k_{GC} k_{AG} s^2$
$\kappa_G k_{end} k_{AA} k_{AG} s^2$
STATISTICAL WEIGHT
$\kappa_G k_{end} k_{AA} k_{AG} k_{GC} s^3$
$\kappa_G k_{AG} k_{GC} s^3$
$\kappa_G k_{end} k_{AA} k_{AG} k_{GC} s^3$
STATISTICAL WEIGHT
$\kappa_G k_{end} k_{AA} k_{AG}^2 k_{GC} s^4$
$\kappa_G k_{end} k_{AA} k_{AG}^2 k_{GC} s^4$
STATISTICAL WEIGHT
$\kappa_G k_{end}^2 k_{AA}^2 k_{AG}^2 k_{GC} s^5$
STATISTICAL WEIGHT
1

Figure 3-2

$A_n U_n$'s, from the definition of s. We have also assumed that k_{AU} is unity, since we could not determine its value from the available information. (See Chapter 2, section B.3) We are thus left with eight interaction terms, seven of which involve a G-C base pair. This means that within this formalism we must determine or estimate eight enthalpies and eight entropies in order to fully characterize the melting process of oligoribonucleic acid molecules.

We have presented two different notations. In our calculations, we use the latter notation. The reason for this choice is not that we assume stacking interactions to cause the stabilization, but rather because the notation for stacking is inclusive of the base pairing notation. In the limit in which all stabilization is independent of nearest neighbors (i.e., is due only to base pairing), all seven values of k_{IJ} with at least one G-C base pair are equal to k_G of equation 3.2. Therefore, the result of the calculation using the stacking notation will provide a partial answer to the question of the relative thermodynamic importance of stacking and base pairing.

The equilibrium constant for initiation of a G-C base pair is not necessarily equal to that of an A-U pair. We thus give κ a subscript A or G, depending on whether or not any G-C base pairs are present in the species considered. We have assumed that initiation occurs preferentially at G-C base pairs, if any are present; this assumption was not initially made, but later analysis indicates that $\kappa_G > \kappa_A$ within the reliability of the calculation. As in Chapter 2, we have taken ΔH^o_κ to be zero.

To compare the results of the frayed end model with those of the standard statistical thermodynamic model, we have performed

calculations using both models. Equations 3.3 and 3.3a give the partition function for the two models for any double stranded oligonucleotide with complementary base pairs. These equations may be understood both with reference to Figs. 3.1 and 3.2 and also to the discussion below.

For the frayed end model,

$$Q = \sum_{\substack{\text{all} \\ \text{states}}} \left[\left(\prod_{IJ}{}' k_{IJ}^{a_{IJ}} \right) s^{\left(\sum_{I,J}{}' a_{IJ} \right)} \kappa_L \, k_{end}^{p} \right] \qquad (3.3)$$

For the standard statistical thermodynamic model,

$$Q = \sum_{\substack{\text{all} \\ \text{states}}} \left[\left(\prod_{IJ}{}' k_{IJ}^{a_{IJ}} \right) s^{\left(\sum_{I,J}{}' a_{IJ} \right)} \kappa_L \right] \qquad (3.3a)$$

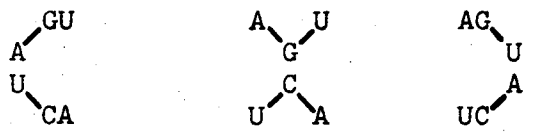
In the first sum, "all states" refers to all intermediate states with 1, 2, 3, . . . , N-1 double stranded stacking interactions with no interior loops (Fig. 3.1). The two products and two sums over I and J are superscripted with a prime to indicate that I takes on the four values A, C, G, and U, while J takes on the following values to avoid overcounting interactions: if I=A, J=A, C, G, U; if I=C, J=A, C, G; if I=G, J=A, C; if I=U, J=A.

The term a_{IJ} is equal to the total number of times a given interaction occurs in the species considered. The subscript L on κ takes on values of G and A, depending on whether or not a G-C base pair is present in the species. p is the number of fully formed ends. The other terms, k_{IJ}, k_{end}, and s have been defined before. For example, in the species $\begin{matrix} \text{AAGGGCUU} & \text{U} \\ \text{UUCCCGAA} & \text{A} \end{matrix}$, $a_{GC} = 1$ and

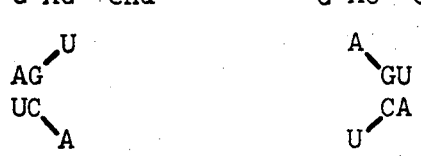
$a_{AA} = a_{CU} = a_{CC} = 2$. All other a_{IJ} are zero and $\sum_{I,J} a_{IJ} = 7$ is the number of double stranded stacking interactions present. In this case, p = 1 and the subscript L is G, since there is at least one G-C base pair. Each "state" or species generates its own set of a_{IJ}, L, and p values, which serve to fully specify the state.

Although equation 3.3 may still appear somewhat imposing, it generates very simple expressions. For example, for the melting of $\begin{matrix} AGU \\ UCA \end{matrix}$,

$$Q = \beta_A k_{end} + \beta_G + \beta_A k_{end}$$



$$+ \kappa_G k_{AG} s k_{end} + \kappa_G k_{AC} s k_{end}$$



$$+ \beta_G k_{AG} k_{AC} s^2 k^2_{end}$$

$$\begin{matrix} AGU \\ UCA \end{matrix}$$

The equation which we use to calculate the melting curve is now, for the frayed end model,

$$1-f = \phi \sum_{\substack{\text{all} \\ \text{states}}} \frac{[(\Pi\Pi' k_{IJ}^{a_{IJ}})(\sum'_{I,J} a_{IJ}) s^{(\sum'_{I,J} a_{IJ})} \kappa_L k_{end}]}{(N-1)Q} \qquad (3.4)$$

and, for the statistical thermodynamic model,

$$1-f = \phi \sum_{\substack{\text{all}\\ \text{states}}} \frac{[(\prod_{IJ} {}'k_{IJ}^{a_{IJ}})(\sum_{I,J}{}' a_{IJ})s^{(\sum_{I,J}{}' a_{IJ})} \kappa_L]}{(N-1)Q} \qquad (3.4a)$$

ϕ is as defined in equation 2.8 and all other terms and limits to the sums and products are the same as for equation 3.3.

B. Application of the Theory to the Molecules A_nGU_n, A_nGCU_n, A_nCGU_n

1. Methodology

In principle, the A_nGU_n molecules require that we evaluate five new thermodynamic functions: κ_G, k_{AG}, k_{AC}, ΔH°_{AG}, and ΔH°_{AC}. From the data available, this is an impossible task. Our approach has been to solve for just two quantities: the product of κ_G, k_{AG}, and k_{AC} (denoted $P_{AG:AC}$) and the sum of ΔH°_{AG} and ΔH°_{AC} (called ΔH°_{AG+AC}).† With one minor problem, to be discussed shortly, this was achieved as follows: we used the values of k_{end} and ΔH°_{AA} from the previous chapter and varied $P_{AG;AC}$ and ΔH°_{AG+AC} so as to fit the melting temperatures at three concentrations (2×10^{-4}, 2×10^{-5}, and 2×10^{-6} M) and $\Delta(1/T_m)/\Delta\ln(c)$ for all members of the series. The procedure is identical to the solution for κ and ΔH°_{AA} molecules: an educated guess of ΔH°_{AG+AC} is made and $P_{AG;AC}$ is then varied until the best fit for all the melting temperatures at the three concentrations is obtained. Then ΔH°_{AG+AC} is incremented by 0.5

---

†Since $P_{AG;CA} = \kappa_G k_{AG} k_{CA} = \kappa_G k^\circ_{AG} k^\circ_{CA} e^{-\Delta H^\circ_{AG+CA}/RT}$, it depends on the absolute temperature. We report its value at 78°C, the melting temperature of polyA:polyU, unless otherwise noted. Because ΔH°_{AG+CA} is calculated, $P_{AG;CA}$ can be determined at any temperature.

kcal/mole and the procedure repeated. This is continued and the results plotted, as in Fig. 2.4, to determine which choice of ΔH°_{AG+AC} is consistent with $\Delta(1/T_m)/\Delta(\ln(c))$. Program OLIGO4 was used as it is written both for molecules with and without G-C base pairs.

The one problem is that it is not possible, by this method, to assign statistical weights to the species

$$\begin{matrix} A_i & & U_n \\ & A_j G & \\ & U_j C & \\ U_i & & A_n \end{matrix} \quad \text{and} \quad \begin{matrix} A_n & & U_i \\ & GU_j & \\ & CA_j & \\ U_n & & A_i \end{matrix} ,$$

where $j \geq 1$. This is not a serious difficulty, since these species do not at any temperature account for more than a few per cent of the partition function. To minimize the error, however, we have broken up $P_{AG;AC}$ into its three terms and ΔH°_{AG+AC} into two in an arbitrary but reasonable manner. For example, if a trial value of ΔH°_{AG+AC} is -10 kcal/2 moles interaction, then we have taken ΔH°_{AG} and ΔH°_{AC} to be -5 kcal/mole each. It is easily shown that any reasonable choice that retains the magnitude of the terms ΔH°_{AG+AC} and $P_{AG;AC}$ gives virtually identical results, since, for example, a 10% error in a quantity which contributes 1% to the partition function is a very small perturbation on the final result. In later discussion, we will undertake a partition of the terms $P_{AG;AC}$ and ΔH°_{AG+AC} in a more quantitative fashion.

2. Results for $A_n GU_n$: The Enthalpy for the AG and AC Double Stranded Interactions is Smaller than for the AA Interaction.

We note that the melting temperature of an oligonucleotide is increased by the substitution of an A-U base pair by a G-C

pair, as shown in Table 3.1. The experimental data, the values of the parameter ΔH°_{AG+AC}, which have been solved for, and the calculated values of T_m, df/dT_m, and the concentration dependence of T_m are presented in Table 3.2.† The enthalpy ΔH°_{AG+AC} for the frayed end model equals -11.50 kcal/2 moles of double stranded stacking interaction. For the model which excludes end effects, we calculate a similar result of ΔH°_{AG+AC} = -9.25 kcal/2 moles. In computing this latter value, we have used the optimum solution for ΔH°_{AA} from the comparable model calculation, namely ΔH°_{AA} = -8 kcal/mole.

Table 3.1

Comparison of Experimental Melting Temperatures of Oligonucleotide Duplexes with Equal Numbers of Base Pairs[a]

Molecule	T_m[b]	Molecule	T_m	Molecule	T_m	Molecule	T_m
A_4U_4	12.4	A_4GU_3	22.7	A_3CGU_3	35.3	A_3GCU_3	42.3
A_5U_5	23.5	A_5GU_4	32.7	A_4CGU_4	42.5	A_4GCU_4	46.8

[a] From reference 5.

[b] All T_m's are at a strand concentration of 10^{-4} M except A_mGU_n, which are at twice this concentration for purposes of comparison. The comparison is made in this manner because the strand complementary to A_mGU_n is not identical to it. (See footnote following equation 2.8.)

The interesting conclusion is that although the substitution of an A-U base pair by a G-C pair in the middle of the helix lowers the free energy of the double stranded form of the nucleic acid,

† See Table 3.12 for $P_{AG;AC}$.

the enthalpic contribution is about 3 kcal/mole less for either the frayed end or stranded model. The extra stability of the G-C base pair surrounded by A-U pairs derives from a more favorable entropy, which more than compensates for the less favorable enthalpy term. For purposes of predicting stabilities of varied nucleic acids, this is an important distinction, as it is the enthalpy which determines the change of the equilibrium constant with temperature. We will later discuss whether this entropy term is related to the new stacking interactions or the more favorable initiation step when G-C base pairs are present.

It remains to comment on the accuracy of the result and to compare the results of the models with and without frayed ends. The melting temperatures (Table 3.2) are not especially well fit by the frayed end model. The standard deviation is ±3.7°C, which is greater than experimental error. For the model without frayed ends, the deviation is ±2.4°C. Part of the problem is that the shortest oligomer, A_3CU_3, is predicted by both models to melt much lower than is observed experimentally. Although we can note, as earlier, that the melting temperatures of this oligomer are not so reliably known as the others, most of the melting curve is obtained at the higher concentrations of the molecule.

The corrected experimental melting curve for A_5GU_4 is compared with curves from two different model calculations in Fig. 3.3. This melting transition is much broader than the transition for the corresponding molecule in the A_nU_n series. The standard model calculation gives little indication of this broadening, predicting a very steep melting curve. The frayed end model, using

Table 3.2

Melting Temperatures, Slopes of the Melting Curves, and Concentration Dependence of T_m for A_nCU_n Molecules

	T_m			$df/dT\|_{T_m}$	$\Delta(1/T_m)/\Delta \ln c$
	$C = 2 \times 10^{-4}$ M	2×10^{-5} M	2×10^{-6} M		
EXPERIMENTAL:					
A_3CU_3	16.8	7.6	~-1.0	3.8×10^{-2}	11.3×10^{-5}
A_4CU_3	22.7	14.7	7.0	3.65	9.4
A_4CU_4	26.4	19.2	12.4	3.6	8.15
A_5CU_4	32.7	26.1	19.8	4.45	7.2
A_5CU_5	36.0	30.1	24.4	5.1	6.3
STANDARD S.M. MODEL: $\Delta H^\circ_{AG+AC} = -9.25$ kcal/2 moles					
A_3CU_3	12.3	3.5	1.1	4.4×10^{-2}	11.2×10^{-5}
A_4CU_3	21.5	13.5	6.0	4.8	9.4
A_4CU_4	28.4	21.2	14.3	5.2	8.15
A_5CU_4	34.0	27.3	21.0	5.6	7.2
A_5CU_5	38.4	32.3	26.4	6.0	6.4
FRAYED END MODEL: $\Delta H^\circ_{AG+AC} = -11.50$ kcal/2 moles					
A_3CU_3	9.8	1.5	-6.2	4.0×10^{-2}	10.5×10^{-5}
A_4CU_3	20.4	12.5	5.3	4.0	9.3
A_4CU_4	28.6	21.2	14.4	4.0	8.2
A_5CU_4	35.2	28.3	21.8	4.2	7.3
A_5CU_5	40.4	33.9	27.9	4.4	6.6

the parameters for the end effect derived from the A_nU_n molecules, does considerably better in this respect. Below 1-f = 0.5 the frayed end model calculation predicts a melting transition sharper than observed. A quantitative comparison of the predicted and observed melting breadths is presented in Table 3.3.

Table 3.3

Molecule	$df/dT\|_m$			$\Delta_{2/3}$ in °C		
	Experiment	Frayed end model	Standard model	Experiment	Frayed end model	Standard model
A_5GU_4	.0445	.042	.056	26.5	20.7	14.2

If we ignore the effect of the ends by applying the standard statistical thermodynamic model, we have no greater success for the A_nGU_n series than for the A_nU_n molecules. To show this in a descriptive way, we have performed calculations which assign an average enthalpy to all the double stranded stacking terms within the A_nGU_n double helix, not distinguishing between the AA, AG, and AC double stranded interactions. The result is that the slope of the melting curve, df/dT_m, predicts an average enthalpy of -5.95 kcal/mole double strand stack, whereas the concentration dependence of T_m yields a value of -7.15 kcal/mole. A similar calculation with the frayed end parameter, k_{end}, set to .0585 gives virtually identical values for the average enthalpy from both physical measurements.

We have hitherto assumed that the magnitude of k_{end} was identical for both A_nU_n and A_nGU_n. Physically, this is as it must

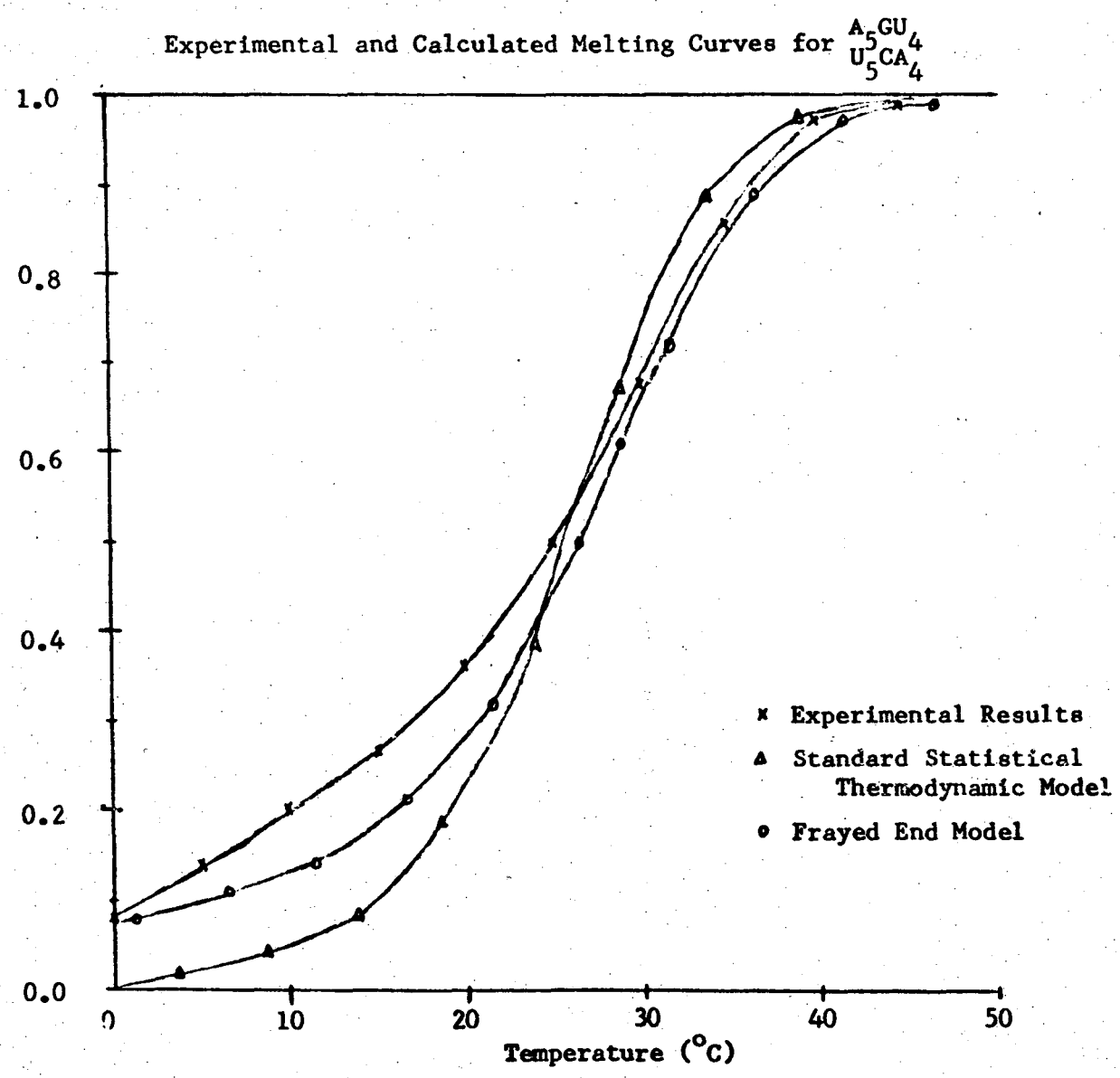


Figure 3-3

be, if the theory is to be consistent. We can, however, solve for the magnitude of k_{end} which best fits the slopes of the melting curves at the T_m. This was done by varying three parameters, $P_{AG;AC}$, ΔH°_{AG+AC}, and k_{end} and fitting these parameters to three sets of experimental data: the melting temperatures, the concentration dependence of T_m, and $df/dT|_m$. The result, when compared with the earlier calculation in which we omitted $df/dT|_m$ and considered k_{end} to be fixed, is in excellent agreement with the earlier solution. The parameters $P_{AG;AC}$ and ΔH°_{AG+AC} are changed by less than 5% and k_{end} is .0645, corresponding to a change in the free energy associated with the end effect from 1.65 kcal/mole to 1.70 kcal/mole at 25°C. This agreement is much better than experimental uncertainties require.

3. Results for A_nGCU_n and A_nCGU_n

The group of molecules A_nGCU_n contains one new double stranded stacking interaction term, k_{GC}. The other interaction term is $k_{AG} \equiv k_{CU}$; we already have some indication of its value from the A_nGU_n molecules. Similarly, for A_nCGU_n, there is one new interaction term, k_{CG}, in addition to $k_{AC} \equiv k_{GU}$, which was also involved in the A_nGU_n molecules. We are thus able to evaluate the terms $\kappa_G k_{AG}^2 k_{GC} \equiv P_{2AG;GC}$, ΔH°_{2AG+GC}, $\kappa_G k_{AC}^2 k_{CG} \equiv P_{2AC;CG}$ and ΔH°_{2AC+CG}. We have again set $k_{end} = .0585$ and $\Delta H^\circ_{AA} = -8.75$ kcal/mole for the frayed end model and $\Delta H^\circ_{AA} = -8.0$ kcal/mole for the standard model.

Results are shown in Figs. 3.4 through 3.6 and Tables 3.4 and 3.5. We have also plotted the population of the most common intermediate species as a function of temperature for A_4CGU_4 in

Fig. 3.5. Most of the conclusions from these figures and tables are generally similar to those derived from the study of the A_nGU_n oligomers. These are summarized as follows: (1) the frayed end model does not do exceptionally well in predicting melting temperatures for molecules of different chain lengths. For the A_nGCU_n oligomers, the melting temperatures are fit within ±2.8°C (versus ±1.0°C for the standard model). For the A_nCGU_n molecules, the standard deviation is ±2.9°C (versus ±1.3°C for the standard model). (2) The shape of the melting curves is greatly improved by including end effects. This permits a more accurate assessment of the double strand stacking enthalpies. (3) ΔH°_{2AG+GC} = -31.3 kcal/3 moles interaction and ΔH°_{2AC+CG} = -27.3 kcal/3 moles interaction, based on a value of k_{end} = 0.0585. If we solve for k_{end} explicitly, as we did for A_nGU_n, the calculated enthalpies are changed only slightly: ΔH°_{2AG+GC} = -29.3 kcal/3 moles and ΔH°_{2AC+CG} = -29.5 kcal/3 moles interaction. The calculated magnitude of k_{end} is 0.12 for A_nGCU_n and 0.025 for A_nCGU_n, corresponding to free energies of +1.3 and +2.2 kcal/mole end interaction at 25°C. Both of these free energies are sufficiently close to 1.6 kcal/mole, the energy of the end interaction calculated for A_nU_n, that we conclude that for these calculations the frayed end model is self-consistent in this respect. (4) Even without relating the enthalpy to specific double stranded interactions, it is now clear that the enthalpies (and also entropies) cannot be assigned to the formation of base pairs. Comparing A_nGU_n and A_nU_n oligomers, we

note that the substitution of one G-C base pair for an A-U pair resulted in a decrease in the absolute value of the enthalpy of the molecule by 6 kcal. The addition of two G-C pairs led to an increase in the absolute value of the enthalpy for the molecule of 1 kcal (for A_nCGU_n) and 5 kcal (for A_nGCU_n) compared to A_nU_n or an increase of 7 kcal (for A_nCGU_n) and 11 kcal (for A_nGCU_n) compared to A_nGU_n. It is evident that, no matter how the energies are finally assigned, the thermodynamics depends strongly on nearest neighbor interactions. It is also clear that the GC and CG double stranded interaction terms are particularly strong ones, the former being especially stabilizing.

C. Application to Other RNA Oligomers

In the previous calculations we have used information from a whole series of molecules, differing only in the number of A-U base pairs. Because of the redundancy of the experimental data, an error or inaccuracy in one or more pieces of data was not a serious matter. From the molecules which follow, we have information about only one (or at most two) members of the series. For this reason, the results in this section are not as accurate as those of section A and B and should be considered somewhat tentative.

1. A_4UAU_4

Although A_4UAU_4 has no G-C base pairs, the formalism developed in this chapter is appropriate to solve for the UA double stranded stacking interaction. Since we know ΔH°_{AA} and κ_A and we have previously assumed that the AA and AU double stranded stacking interactions are equal, we can solve directly for ΔH°_{UA} and k_{UA}. The experimental data[7] are given in Table 3.6 and the results of the

Table 3.4

Melting Temperature, Slope of the Melting Curves, and Concentration Dependence for $^{A_n GCU_n}_{U_n CGA_n}$ Molecules

	T_m			$df/dT\|T_m$	$\Delta(1/T_m)/\Delta \ln$
	$C = 10^{-4}$ M	10^{-5} M	10^{-6} M		
EXPERIMENTAL:					
A_2GCU_2	28.5	19.7	11.3	$(2.7 \pm 0.3) \times 10^{-2}$	10.0×10^{-5}
A_3GCU_3	42.3	33.8	25.8	$(4.0 \pm 0.5) \times 10^{-2}$	8.75
A_4GCU_4	46.8	40.9	35.2	$(4.1 \pm 0.4) \times 10^{-2}$	5.9
STANDARD S.M. MODEL: $\Delta H^\circ_{2AG+GC} = -27.2$ kcal/3 moles					
A_2GCU_2	28.4	19.1	10.3	4.0×10^{-2}	10.5×10^{-5}
A_3GCU_3	40.8	33.2	26.0	4.8	7.85
A_4GCU_4	48.5	42.1	36.0	5.5	6.25
FRAYED END MODEL: $\Delta H^\circ_{2AG+GC} = -31.3$ kcal/3 moles					
A_2GCU_2	25.2	16.3	8.1	2.85×10^{-2}	10.1×10^{-5}
A_3GCU_3	41.4	33.6	26.3	3.0	8.05
A_4GCU_4	51.0	44.4	38.0	3.5	6.45

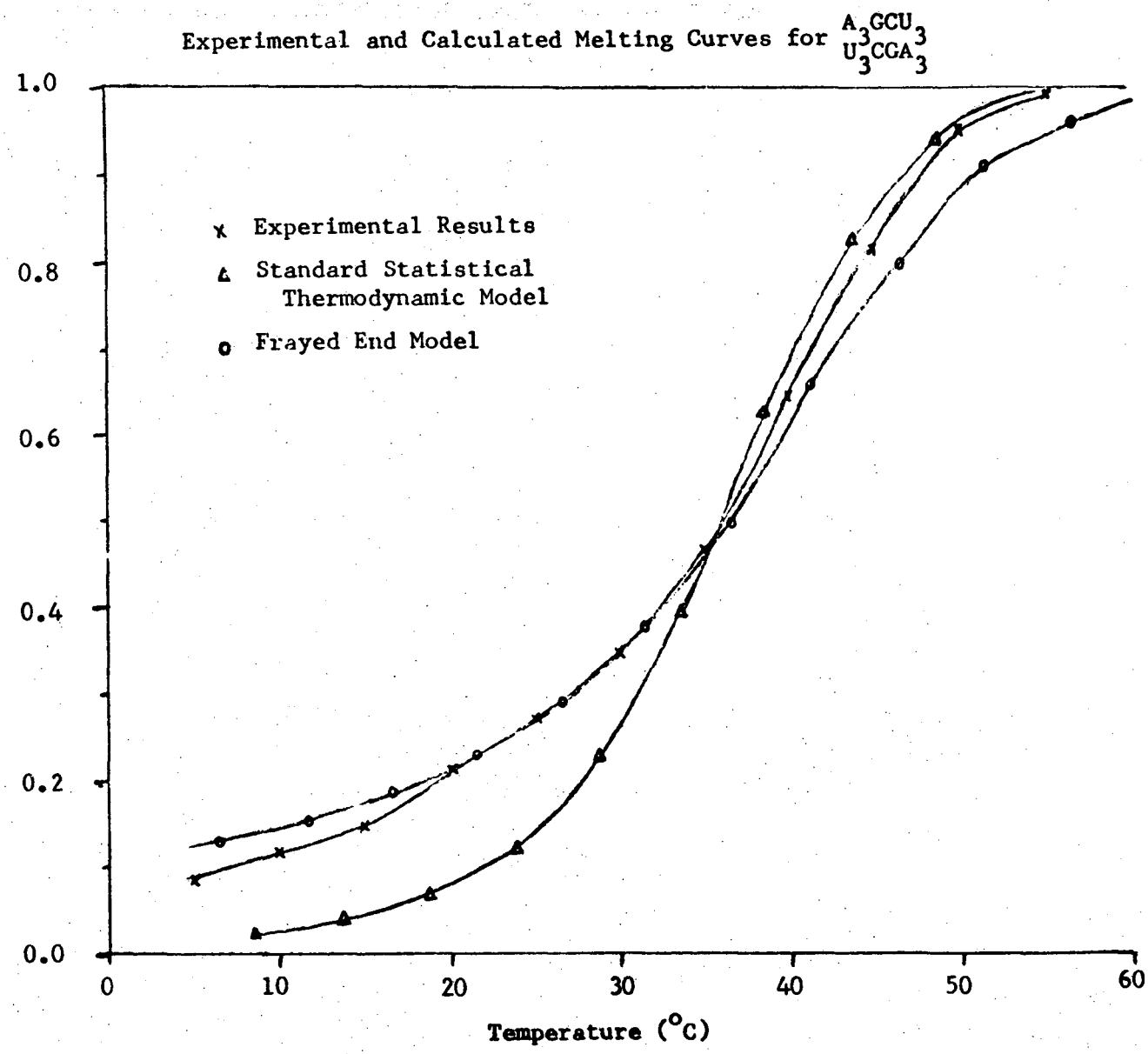

Experimental and Calculated Melting Curves for A_3GCU_3 / U_3CGA_3
1.0
0.8
0.6
0.4
0.2
0.0
1-f
Experimental Results
Standard Statistical Thermodynamic Model
Frayed End Model
0
10
20
30
40
50
60
Temperature (°C)

Figure 3-4

Table 3.5

Melting Temperature, Slope of the Melting Curves, and Concentration Dependence of T_m for $\overset{A}{U}_n^n CGU_n^n GCA_n$ Molecules

	T_m			$df/dT\|_{T_m}$	$\Delta(1/T_m)/\Delta \ln$
	$C=10^{-4}$ M	10^{-5} M	10^{-6} M		
EXPERIMENTAL:					
A_2CGU_2	22.2	11.0	~0.7	$(2.2 \pm 0.3) \times 10^{-2}$	13.3×10^{-5}
A_3CGU_3	35.2	28.8	22.5	--	7.0
A_4CGU_4	42.5	36.6	31.0	$(3.65 \pm 0.3) \times 10^{-2}$	6.0
STANDARD S.M. MODEL: $\Delta H^\circ_{2AC+CG} = -24.35$ kcal/3 moles					
A_2CGU_2	20.5	10.9	2.0	4.0×10^{-2}	11.5×10^{-5}
A_3CGU_3	35.3	27.6	20.3	4.8	8.3
A_4CGU_4	44.3	37.8	31.6	5.5	6.5
FRAYED END MODEL: $\Delta H^\circ_{2AC+CG} = -27.25$ kcal/3 moles					
A_2CGU_2	17.3	8.3	-0.1	3.2×10^{-2}	11.0×10^{-5}
A_3CGU_3	35.9	27.9	20.4	3.2	8.5
A_4CGU_4	47.0	40.1	33.6	3.6	6.8

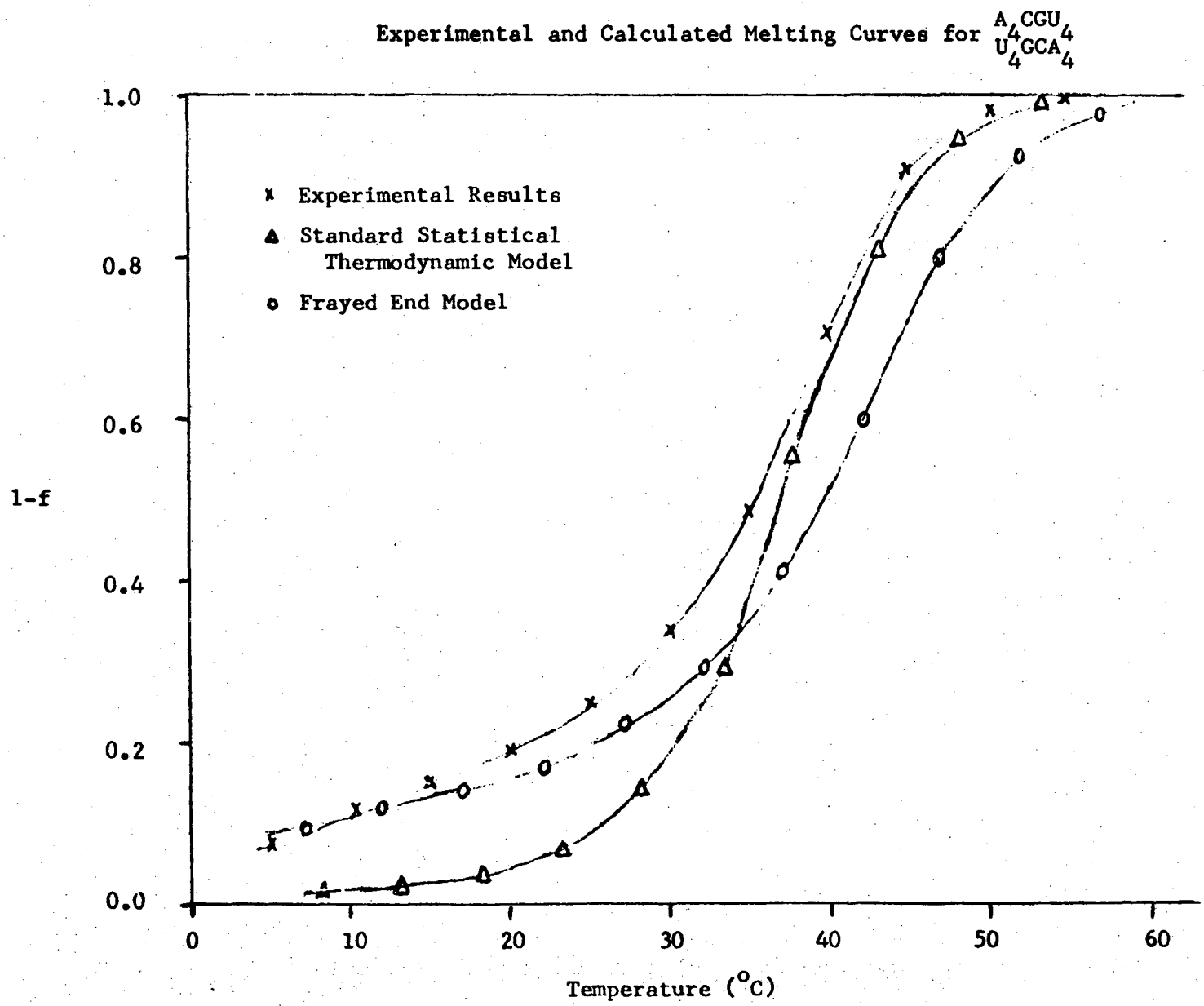

Experimental and Calculated Melting Curves for A_4CGU_4 / U_4GCA_4
Experimental Results
Standard Statistical Thermodynamic Model
Frayed End Model
1-f
1.0
0.8
0.6
0.4
0.2
0.0
0
10
20
30
40
50
60
Temperature (°C)

Figure 3-5

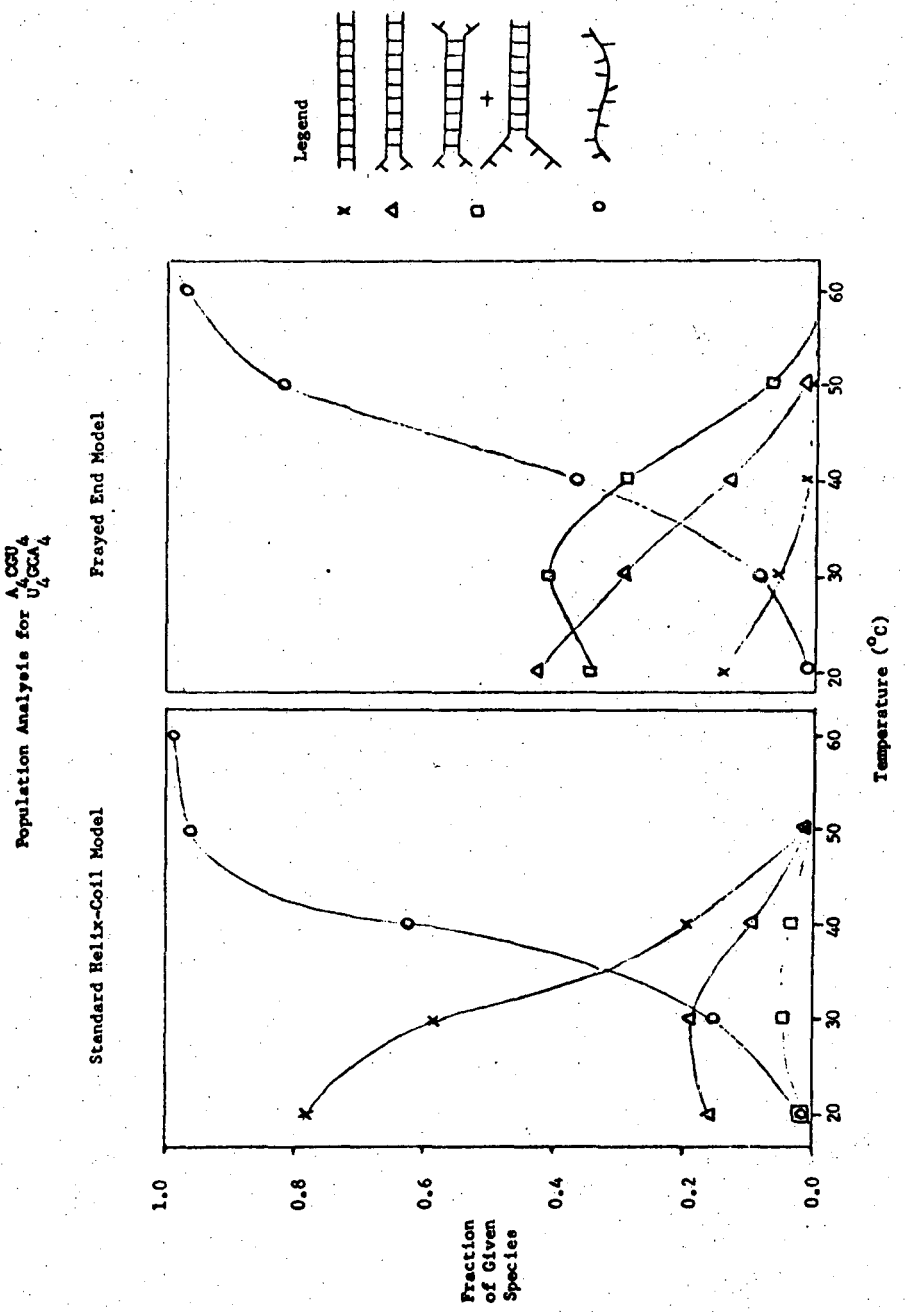

Population Analysis for $A_4^{CCU}U_4GCA_4$
Standard Helix-Coil Model
Frayed End Model
Legend
Fraction of Given Species
Temperature (°C)
1.0
0.8
0.6
0.4
0.2
0.0
20
30
40
50
60

Figure 3-6

calculation in Table 3.7.

Table 3.6

Experimental Data for A_4UAU_4

Melting	$c = 10^{-6}M$	$c = 10^{-5}M$	$c = 10^{-4}M$
Temperatures	16.7°C	21.3°C	27.7°C

Slope of the melting curve: $df/dT|_m = (3.4 \pm 0.2) \times 10^{-2}$

Conc. dependence of T_m: $\Delta(1/T_m)/\Delta(\log c) = 6.44 \times 10^{-5}$

Table 3.7

Results of Calculation for A_4UAU_4

a. Frayed end calculation ($k_{end} = 0.0585$)

Melting	$c = 10^{-6}M$	$c = 10^{-5}M$	$c = 10^{-4}M$
Temperatures	16.7°C	22.1°C	27.7°C

Slope of the melting curve: $df/dT|_m = 5.3 \times 10^{-2}$

Conc. dependence of T_m: $\Delta(1/T_m)/\Delta(\log c) = 6.4 \times 10^{-5}$

b. Standard Helix-Coil Theory

Melting	$c = 10^{-6}M$	$c = 10^{-5}M$	$c = 10^{-4}M$
Temperatures	16.7°C	22.1°C	27.8°C

Slope of the melting curve: $df/dT|_m = 6.6 \times 10^{-2}$

Conc. dependence of T_m: $\Delta(1/T_m)/\Delta(\log c) = 6.4 \times 10^{-5}$

The parameters which give the above results are listed.

Frayed end model:

$$\Delta H^{\circ}_{UA} = -10.75 \text{ kcal/mole}$$

$$k_{UA} = 3.2 \text{ (at } 78^{\circ}\text{C)}$$

Standard helix-coil theory:

$$\Delta H^{\circ}_{UA} = -9.00 \text{ kcal/mole}$$

$$k_{UA} = 2.80 \text{ (at } 78^{\circ}\text{C)}$$

The largest uncertainty in these parameters derives from the assumption that the AA and AU double stranded stacking interactions are equal. Both models predict melting curves which are too steep, although the frayed end model broadens the curve somewhat.

2. A_4G_2 and A_5G_3

The data for these two molecules are limited. A_5G_3 information is based on just one melting curve with a somewhat low hypochromicity (14% at 50°C). For A_4G_2 and its complement, the high value of $\Delta(1/T_m)/\Delta(\log c)$ is suspect, since this quantity is consistently high for the shortest oligomer (with six or seven base pairs) in the A_nU_n, A_nGU_n, and A_nGCU_n series. This may be due to the experimental uncertainty in the melting temperatures in the low temperature range.† If this is the case then $\Delta(1/T_m)/\Delta(\log c)$ should probably be between 10×10^{-5} and 11×10^{-5}, rather than 12×10^{-5} as reported for A_4G_2.

---

†A systematic error of 1°C will change $\Delta(1/T_m)/\Delta(\log c)$ by about 15%; the scatter in the T_m data is much greater than 1°C for the oligomers with six base pairs but not for those with eight or more base pairs.

An additional problem with these molecules is that we now have a terminal G-C base pair: it is no longer necessary or even likely that k_{end} retain its previous value, since the interaction has changed.

Cognizant of these limitations, we have the results shown in Table 3.8 for A_4G_2, with $k_{end} = 0.0585$.

Table 3.8

---

a. Experimental Melting Data for A_4G_2

Melting temperatures: 8.1°C ($c=10^{-5}M$) 17.9°C ($c=10^{-4}M$)

Slope of the melting curve: $df/dT|_m = 2.9 \times 10^{-2}$

$\Delta(1/T_m)/\Delta(\log c) = 12.0 \times 10^{-5}$

b. Calculated Melting Data for A_4G_2 Using $\Delta(1/T_m)/\Delta(\log c) = 12\times10^{-5}$

Parameters assumed: $\Delta H^\circ_{AA} = -8.75$ kcal/mole $k_{end} = 0.0585$

Parameter solved for: $\Delta H^\circ_{AG+CC} = -14.0$ kcal/2 moles

Melting temperatures: 8.1°C ($c=10^{-5}M$) 17.9°C ($c=10^{-4}M$)

Slope of the melting curve: $df/dT|_m = 3.6 \times 10^{-2}$

c. Calculated Melting Data for A_4G_2 Using $\Delta(1/T_m)/\Delta(\log c) = 11\times10^{-5}$

Parameter solved for: $\Delta H^\circ_{AG+CC} = -18$ kcal/2 moles

Melting temperatures: 8.2°C ($c=10^{-5}M$) 17.1°C ($c=10^{-4}M$)

Slope of the melting curve: $df/dT|_m = 3.9 \times 10^{-2}$

d. Calculated Melting Data for A_4G_2 Using $\Delta(1/T_m)/\Delta(\log c) = 10\times10^{-5}$

Parameter solved for: $\Delta H^\circ_{AG+CC} = -21.5$ kcal/2 moles

Melting temperatures: 8.2°C ($c=10^{-4}M$) 16.4°C ($c=10^{-5}M$)

Slope of the melting curve: $df/dT|_m = 4.2 \times 10^{-2}$

It is not possible to interpret the results in Table 3.8 with complete certainty. If $\Delta(1/T_m)/\Delta(\log c) = 12 \times 10^{-5}$ (part b), then the absolute value of the enthalpy of the AG+CC interactions, 14 kcal/2 moles interaction, is 3.5 kcal less than for 2 moles of AA interactions, the predicted slopes of the melting curves are somewhat too steep, and the entropy associated with the interactions is much greater than for all other double stranded stacking interactions heretofore considered. If we take $\Delta(1/T_m)/\Delta(\log c) = 10 \times 10^{-5}$ (part d), the absolute value of the enthalpy of the AG+CC interactions, 21.5 kcal/2 moles interaction, is 4 kcal greater (i.e., more stable) than for 2 moles of AA interaction, the predicted slopes of the melting curves are even more steep, and the entropy associated with the interactions is not greatly different from that of other double stranded stacking interactions.

We have also performed calculations allowing the value of k_{end} to vary at the G-C end of the molecule, in order to fit $df/dT|_m$. It is not possible to change the calculated slope or breadth of the melting curve by this procedure; as k_{end} is decreased, the stability of the GG interaction increases proportionately and the G-C end of the molecule insists on being formed (rather than frayed). As a result, the difference between the frayed end model and the standard model calculation is not great, since only one end of this molecule is frayed.

In order to choose among the three sets of parameters presented in Table 3.8, we have used all three sets to predict the melting temperature of A_5G_3. The results are shown in Table 3.9.

Table 3.9

Calculated Melting Temperature of A_5G_3

(strand concentration = 1.4×10^{-5}M)	
Parameters from Table 3.8	T_m calculated
part b	51.7°C
part c	45.8°C
part d	41.3°C

The experimental T_m for A_5G_3 is 41°C, from the one available melting curve at $c = 1.4 \times 10^{-5}$. On the basis of the prediction of the melting temperature of A_5G_3, the parameters of part d in Table 3.8 are most appropriate and these will be used. Because of the paucity of experimental information, this analysis is only tentative and is likely to be improved when more data are available.

3. U_2CGA_2

For this molecule, the new double stranded stacking interaction is UC≡GA. Because this molecule has only six base pairs and melts at a low temperature, the value of $\Delta(1/T_m)/\Delta(\log c)$ is probably not very accurate. The experimental data are given in Table 3.10 and the results of the calculation in Table 3.11.

Table 3.10

Experimental Data for U_2CGA_2

Melting temperatures: 11.3°C ($c=10^{-4}M$) 1.6°C ($c=10^{-5}M$)

Slope of the melting curves: $df/dT|_m = 2.0 \times 10^{-2}$

Conc. dependence of T_m: $\Delta(1/T_m)/\Delta(\log c) = 12.25 \times 10^{-5}$

Table 3.11

Results of Calculation for U_2CGA_2

---

a. Frayed end calculation ($k_{end} = 0.0585$)

Melting temperatures: 11.3°C ($c=10^{-4}M$) 1.6°C ($c=10^{-5}M$)

Slope of the melting curve: $df/dT|_m = 3.25 \times 10^{-2}$

$\Delta(1/T_m)/\Delta(\log c) = 12.1 \times 10^{-5}$

b. Standard Helix-Coil Calculation

Melting temperatures: 11.3°C ($c=10^{-4}M$) 1.7°C ($c=10^{-5}M$)

Slope of the melting curve: $df/dT|_m = 4.04 \times 10^{-2}$

$\Delta(1/T_m)/\Delta(\log c) = 12.2 \times 10^{-5}$

The resulting enthalpies are $\Delta H^\circ_{CG+2GA} = -22.25$ kcal/3 moles interaction for the frayed end model and $\Delta H^\circ_{CG+2GA} = -21.7$ kcal/3 moles. Although the frayed end model predicts melting curves somewhat sharper than experiment, the optimum solution for the slope of the melting curve yields a free energy for the end effect one kcal/mole larger than the value we have used (from the A_nU_n molecules).

D. Analysis of Results

We summarize in Table 3.12 the thermodynamic parameters which have been solved for in this chapter. They are not especially useful in the form in which they appear in this table. We would like to know the interaction terms for the individual double stranded stacking interactions. This information cannot be obtained rigorously, as we have more unknowns than equations. For the present, we make a few simple approximations in order to derive the thermodynamic parameters of interest to us. It is worth keeping in mind that the parameters in Table 3.12 are those for which we have solved directly.

To obtain the enthalpy terms for the individual double stranded stacking interactions, we assume that: the AG and AC interactions are equal. This approximation cannot be rigorously justified, other than by the fact that it yields results which are neither unreasonable nor internally inconsistent. (As more data are obtained, it will be possible to remove this approximation.) This enables us to solve for all the enthalpy terms; the results are given in Table 3.13. It is of interest that the enthalpy for the GC, CG, and GG interactions are large, being about twice as great as the enthalpy for the AA interaction. It is also of note that both the frayed end and the standard helix coil models give comparable enthalpies for each of the double stranded stacking interactions.

Obtaining k_{IJ} from the product terms in Table 3.13 is a more difficult task. The reason for this is that each product term includes an additional parameter, κ_G, the magnitude of which is not

Table 3.12

Composite Thermodynamic Parameters

Enthalpies		Product terms (at 78°C)		
a. Frayed End Model (k_{end} = 0.0585)				
ΔH°_{AG+AC}	= -11.50 kcal	$P_{AG;AC}$	= $\kappa_G k_{AG} k_{AC}$	= 0.16
$\Delta H^{\circ}_{2AG+GC}$	= -31.25 kcal	$P_{2AG;GC}$	= $\kappa_G k_{AG} k_{GC}$	= 5.3
$\Delta H^{\circ}_{2AC+CG}$	= -27.25 kcal	$P_{2AC;CG}$	= $\kappa_G k_{AC} k_{CG}$	= 2.3
ΔH_{AG+CC}	= -21.5 kcal	$P_{AG;CC}$	= $\kappa_G k_{AG} k_{CC}$	= 1.86
ΔH_{2GA+CG}	= -22.25 kcal	$P_{2GA;CG}$	= $\kappa_G k_{GA} k_{CG}$	= 2.76
b. Standard Helix-Coil Model				
ΔH°_{AG+AC}	= -9.20 kcal	$P_{AG;AC}$	= $\kappa_G k_{AG} k_{AC}$	= 0.017
$\Delta H^{\circ}_{2AG+GC}$	= -27.80 kcal	$P_{2AG;GC}$	= $\kappa_G k^2_{AG} k_{GC}$	= 0.425
$\Delta H^{\circ}_{2AC+CG}$	= -24.3 kcal	$P_{2AC;CG}$	= $\kappa_G k_{AC} k_{CG}$	= 0.165
ΔH°_{AG+CC}	= -22.1 kcal	$P_{AG;CC}$	= $\kappa_G k_{AG} k_{CG}$	= 0.092
ΔH_{2GA+CG}	= -18.4 kcal	$P_{2GA;CG}$	= $\kappa_G k^2_{GA} k_{CG}$	= 0.021

Table 3.13

Enthalpies for Double Stranded Stacking Interactions

(Approximation: $\Delta H^\circ_{AG} = \Delta H^\circ_{AC}$)

a. Frayed End Model

$\Delta H^\circ_{AA} = \Delta H^\circ_{AU}$	$= -8.75$ kcal/mole	ΔH°_{GC}	$= -19.75$ kcal/mole
ΔH°_{UA}	$= -10.75$ kcal/mole	ΔH°_{CG}	$= -15.75$ kcal/mole
$\Delta H^\circ_{AG} = \Delta H^\circ_{AC}$	$= -5.75$ kcal/mole	ΔH°_{CC}	$= -15.75$ kcal/mole
ΔH°_{GA}	$= -3.25$ kcal/mole		

b. Standard Helix-Coil Model

$\Delta H^\circ_{AA} = \Delta H^\circ_{AU}$	$= -8.00$ kcal/mole	ΔH°_{GC}	$= -18.6$ kcal/mole
ΔH°_{UA}	$= -9.00$ kcal/mole	ΔH°_{CG}	$= -15.1$ kcal/mole
$\Delta H^\circ_{AG} = \Delta H^\circ_{AC}$	$= -4.6$ kcal/mole	ΔH°_{CC}	$= -17.5$ kcal/mole
ΔH°_{GA}	$= -3.30$ kcal/mole		

known. The simplest assumption we could make is that $\kappa_G = \kappa_A$. Since we know κ_A from our calculations on the A_nU_n molecules, we are able to solve for the k_{IJ} terms directly, again assuming the thermodynamic equivalence of the AG and GU interactions.

The problem with this approach is that it gives results not fully in accord with experimental measurements on RNA polymers, as we now show. In Fig. 3.7, we reproduce a plot from Kallenbach in which $1/T_m$ is plotted against the fraction G-C base pairs for double stranded RNA polymers in 0.15 M Na^+.[1]

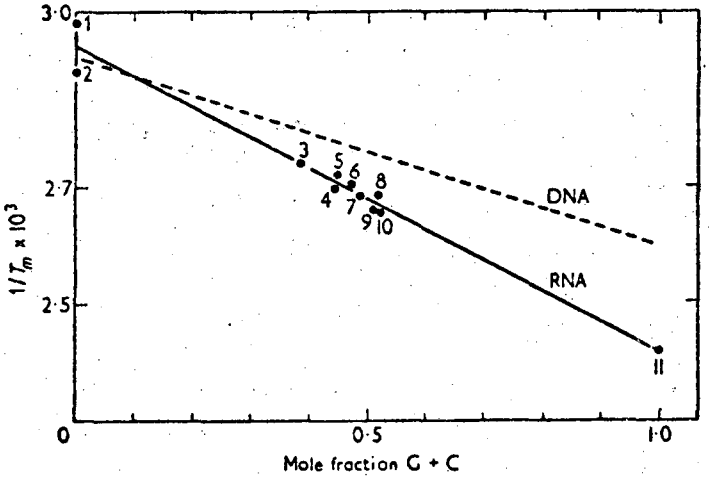


Figure 3-7

If we temporarily assume that A-U and G-C base pairs have equal enthalpies and that only one equilibrium constant, k_G, is necessary to account for the stability of G-C base pairs, as Kallenbach has assumed, then the straight line in Fig. 3.7 is described by the equation

$$\frac{1}{T_m} = \frac{1}{T_m(A\text{-}U)} + \frac{R\nu_G}{\Delta H^\circ} \ln k_G$$

Here, T_m is the melting temperature of an RNA polymer with mole fraction ν_G of G-C base pairs, $T_m(A\text{-}U)$ is the melting temperature of an RNA polymer with only A-U base pairs, and k_G is defined in equation 3.2. This equation cannot be rigorously correct, for we have already observed that the thermodynamics of RNA depends strongly on nearest neighbor interactions. However, the value of k_G derived from Fig. 3.7 should represent a sort of average of the k_{IJ} terms.

The k_{IJ} terms derived from the assumption that $\kappa_G = \kappa_A$ do not fit the slope of the line in Fig. 3.7. Plugging in an average enthalpy for double stranded stacking interactions of -9.5 kcal/mole (from Table 3.13), we determine from the slope of $1/T_m$ versus ν_G that $\bar{k}_G = 10.8$. Assuming $\kappa_G = \kappa_A$, an appropriately averaged k_G equals approximately twice this amount.† Thus, if we assume $\kappa_G = \kappa_A$,

---

†This is done for a random RNA at $\nu_G = 0.5$. The appropriate average is that of $\log k_G$. One must take care in counting interactions properly. For a random RNA polymer at $\nu_G = 0.5$, the GA interaction will occur twice as frequently as the GC interaction, for example. The averaging equation is

$$1/8 \log (k_{GA}k_{AG}k_{GU}k_{UG}k_{GG}k_{AA}) + 1/16 \log (k_{GC}k_{CG}k_{AU}k_{UA}) = \log \bar{k}_G$$

In deriving this equation, we have made the simplification that all ΔH°_{IJ} = the average value of -9.5 kcal/mole. In applying the equation, we assume the unknown term k_{CA} is equal to k_{AG} and k_{AC}.

the slope in Fig. 3.7 is predicted by our calculations to be greater than experiment by a factor of approximately one-third.

It follows that κ_G and κ_A are not equal, if we give consideration to the polymer melting temperatures. We can obtain k_{IJ} and κ_G from the products in Table 3.12 is we normalize the k_{IJ} terms so that their averaged value[†] if 10.7, in agreement with Fig. 3.7. When this is done, the results in Table 3.14 obtain for the frayed end and standard helix coil models.

The important conclusion from this chapter is that the sequence of base pairs in an RNA molecule makes a great deal of difference in the stability of the molecule in solution. We believe that the results in Tables 3.13 and 3.14 are the best presently available in relating the thermodynamic stability of RNA to its double stranded stacking interactions. In Table 3.15, in which we have compiled this thermodynamic information in terms of the free energy (at 25°C), and the standard enthalpy and entropy of the interactions, the large effect of sequence is evident. To make this more vivid, we have calculated the melting temperatures of a series of oligomers of identical base composition but differing sequence, using the data compiled in this chapter. It is readily apparent that the sequence of base pairs has a large effect on the T_m of an RNA molecule, as shown in Table 3.16.

[†]See note previous page.

Table 3.14

k_{IJ} Terms for Double Stranded Stacking Interactions (at 78°C)

a. Frayed End Model ($k_{end} = 0.0585$; $k_G = 1.95 \times 10^{-3}$; $k_A = 6.55 \times 10^{-4}$)

$k_{AA} = k_{AU} = 1.0$

$k_{UA} = 3.2$

$k_{AG} = k_{AC} = 9.15$

$k_{GA} = 10.0$

$k_{GC} = 32.5$

$k_{CG} = 14.2$

$k_{CC} = 104.0$

b. Standard Helix Coil Model ($k_G = 4.25 \times 10^{-5}$; $k_A = 6.0 \times 10^{-5}$)

$k_{AA} = k_{AU} = 1.0$

$k_{UA} = 2.8$

$k_{AG} = k_{AC} = 20$

$k_{GA} = 6.0$

$k_{GC} = 25.3$

$k_{CG} = 9.8$

$k_{CC} = 109.0$

Table 3.15

Standard Free Energy, Enthalpy and Entropy of

Double Stranded Stacking Interactions

Interaction	$\Delta G^\circ(25^\circ C)$	ΔH°	ΔS°
a. Frayed End Model			
AA; AU	-1.3 kcal/mole	-8.75 kcal/mole	-24.9 e. u.
UA	-2.3	-10.75	-28.3
AG; AC	-1.9	-5.75	-13.0
GA	-1.8	-3.25	-4.8
GC	-5.0	-19.75	-49.3
CG	-3.9	-15.75	-39.6
CC	-4.7	-15.75	-36.8
Initiation (at A-U)	+4.4	0	-14.7
Initiation (at G-C)	+3.7	0	-12.5
End effect	+1.7	0	-5.7
b. Standard Helix Coil Model			
AA; AU	-1.2 kcal/mole	-8.00 kcal/mole	-22.8 e.u.
UA	-1.9	-9.00	-23.8
AG; AC	-2.4	-4.60	-7.8
GA	-1.4	-3.30	-6.2
GC	-4.4	-18.60	-47.3
CG	-3.4	-15.10	-39.0
CC	-5.0	-17.50	-41.5
Initiation (at A-U)	+5.8	0	-19.4
Initiation (at G-C	+6.0	0	-20.0

Table 3.16

Calculated T_m's of RNA Molecules of Identical Base Composition and Different Sequences of Base Pairs[a]

a. Two G-C base pairs and four A-U base pairs

A_2GCU_2	T_m = 16.3°C
AGAUCU	-10.1°
GA_2U_2C	-12.8°
GCA_2U_2	12.6

b. Two G-C base pairs and six A-U base pairs

A_3GCU_3	T_m = 33.8°C
A_2GAUCU_2	18.9°
GA_3U_3C	11.3
GCA_3U_3	28.0

c. Three G-C base pairs and four A-U base pairs

A_2GCGU_2	T_m = 40.8°C
AGACUGU	12.2°
GA_2CU_2G	5.6°

d. Four G-C base pairs and four A-U base pairs

A_3GCGU_3	T_m = 51.0°C
$A_2GACUGU_2$	38.2°
GA_3CU_3G	25.3°

e. Four G-C base pairs and four A-U base pairs

A_2GCGCU_2	T_m = 61.6°C
AGCAUGCU	54.9°
ACACGUCU	42.2°
GACAUGUC	27.5°

[a]The complementary strands are not listed. Strand concentration is 10^{-5} M (unless the complementary strand is not identical in which case $c = 2 \times 10^{-5}$ M). The solution is at pH = 7 and is .1 M in NaCl and 10^{-4} M in EDTA.

CHAPTER 3

REFERENCES

1. N. R. Kallenbach, J. Mol. Biol., 37, 445-466 (1968).

2. I. Tinoco, Jr., O. C. Uhlenbeck, and M. D. Levine, Nature, 230, 362-367 (1971).

3. C. DeLisi and D. M. Crothers, Proc. Nat. Acad. Sci. U.S.A., 68, 2682-2685 (1971).

4. S. M. Coutts, Biochim. Biophys. Acta, 232, 94-106 (1971).

5. O. C. Uhlenbeck, F. H. Martin, and P. Doty, J. Mol. Biol., 57, 201-215 (1971).

6. J. Gralla and D. M. Crothers, J. Mol. Biol., 73, 497-511 (1973).

7. P. N. Borer, Ph.D. Thesis, University of California, Berkeley (1972).

CHAPTER 4

THE MELTING OF $A_6C_mU_6$ LOOPS

The remaining information needed to further our understanding of RNA stability pertains to the influence of loops on the energetics of RNA secondary structure. It is important to remember that intramolecular base pair formation cannot occur without the formation of looped regions within the ribonucleic acid. Generally, three different types of loops have been distinguished: hairpin loops, interior loops, and bulges. These three types of loops are shown schematically in Fig. 4.1. In this chapter we subject the hairpin loops which were synthesized and studied by Uhlenbeck, Borer, Dengler, and Tinoco[1] to an analysis similar to that of the preceding chapters. For interior loops and bulges, we have no new data to analyze.

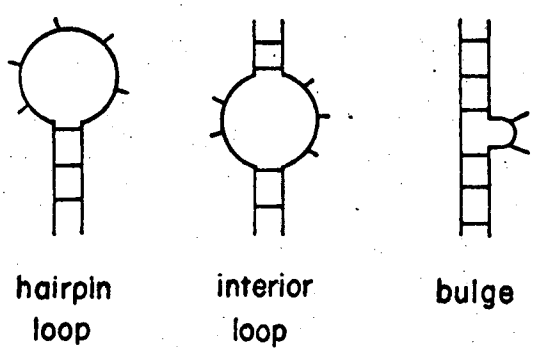


Figure 4-1

A. Theory

The theory needed to calculate and predict the melting behavior of RNA hairpin loops is simpler than that applied to the duplex oligomers, because the melting of the loops is not concentration dependent, since the reaction is intramolecular. Because we are concerned with loops of the form $A_6C_mU_6$, the double stranded or stem region, $\frac{AAAAAA}{UUUUUU}$, is homogeneous with respect to base pairs. This means that such staggered forms as

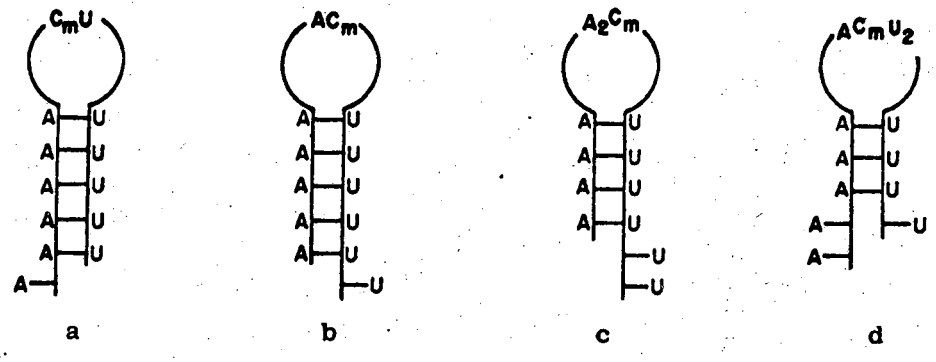


Figure 4-2

are permitted. These staggered forms are included in the list of intermediate states and in the partition function. For loops with inhomogeneous stem regions, staggering of this nature is not likely, as it would lead to noncomplementary base pair formation.

We write the partition function for hairpin loops with N A-U base pairs in the stem region as

$$Q = 1 + \sum_{m=m_o}^{m_o+2N} \sum_{n=1}^{N-(m-m_o)/2} \gamma_m s^{n-1} k_{end}^{p} \qquad (4.1)$$

The term unity is for the single strand contribution, in which no base pairs are present. The inner sum is over intermediate species with a given loop size (of m unbonded bases) and a variable number of base pairs in the stem (from n=1 to $n=N-(m-m_o)/2$). The minimum number of unbonded bases in the loop is specified by m_o. The largest term in the inner sum, $n=N-(m-m_o)/2$, is to be rounded down to the nearest integer value. The equilibrium constant for the end effect, k_{end}, is raised to the power p, where p=1 if an end with two dangling strands is present (as in Fig. 4.2d) and p=0 if an end with one dangling strand is present (as in Fig. 4.2a, b, and c).†

γ_m, called the loop weighting function, is the equilibrium constant for a loop of m unbonded bases held together by one base pair. This is the thermodynamic parameter that we are interested in evaluating from the melting data. Until experimental information

---

†We have made the implicit assumption that both of the terminal bases, A and U, contribute equally to the end effect. It is also possible to assume that only the A residue is involved in the end interaction. The rationale for this is that physical studies have shown that U does not stack well—if at all—in the single strand. This assumption would assign p a value of 1 if an end with dangling A's is present and set p=0 if U's are dangling. Since it is not possible to choose between these alternatives on the basis of present knowledge, we use the first assumption on grounds of simplicity. Conclusions are entirely unchanged if either assumption is made.

on nucleic acid loops became available recently, γ_m could be estimated only from theory. The earliest theoretical treatment of the loop weighting function which is pertinent to RNA hairpin loops assumed that the loop is composed of m+1 rigid, freely moving links whose end-to-end distribution function is Gaussian.[2] This led to the prediction that γ_m is proportional to $(1/m+1)^{3/2}$. Kallenbach, from an analysis of RNA and DNA melting data, has written this equation as

$$\gamma_m = \tau^{-1}(m+1)^{-3/2} \qquad (4.2)$$

where τ has a value of 500 to 1,000.[3]

This approach has the limitations that it fails to consider (1) the effect of excluded volume,† (2) the lack of proportionality between the mean square end-to-end distance $[r^2]$ and the number of links, m+1 for short chains.[4] It is this proportionality which led to the simple form of equation (4.2). (3) The degree to which the chain is non-Gaussian because of intramolecular interactions or interactions with the solvent; and (4) possible enthalpic contributions due to the initiation step (the formation of the first base pair to close the loop) or due to the deformation of the single strand to form a hairpin loop. Because all of these factors, none of which can be quantitatively assessed on the basis of theory, enter into the calculation of a loop weighting function for nucleic

---

†The molecule can occupy only those regions of the solution which are unoccupied. Computer calculations indicate that the exponent on the term (m+1) can be changed to as much as -1.75 or -2 if the effect is included.

acids, it is of special interest to have experimental data with which the theory may be compared.

In order to explicitly include the fourth point above, we write

$$\gamma_m = \gamma_m^\circ \, e^{-\Delta H_m^\circ / RT} \qquad (4.3)$$

where $\gamma_m^\circ = e^{\Delta S_m^\circ / R}$ is the entropic part of the loop weighting function and ΔH_m° is the enthalpy associated with the initiation of the hairpin loop. The total enthalpy of the reaction $A_6C_mU_6 \overset{\leftarrow}{\rightarrow} \underset{UUUUUU}{AAAAAA}C_m$ is

$$\Delta H_{tot} = 5\,\Delta H_{AA}^\circ + \Delta H_m^\circ \, .$$

ΔH_{tot}° is less than zero, since the reaction is exothermic. However, ΔH_m° may be greater than, less than, or equal to zero. If ΔH_m° is greater than zero, then the formation of the looped region must overcome repulsive interactions, associated with strain in the loop of some nature. These repulsive forces are more likely to be important for small loops than for large ones. A consequence of ΔH_m° greater than zero, as the calculation for the $A_6C_mU_6$ loops indicates, is that at higher temperatures the loop region of these molecules is stabilized relative to the loop at lower temperatures, thereby broadening the melting curve.

Equation 4.4, which enables us to calculate 1-f versus temperature for loop molecules with only A-U base pairs in the stem region, assumes that the absorption increase caused by the formation of the double helix is due primarily to the double stranded stacking interaction. The justification for this assumption is the same as

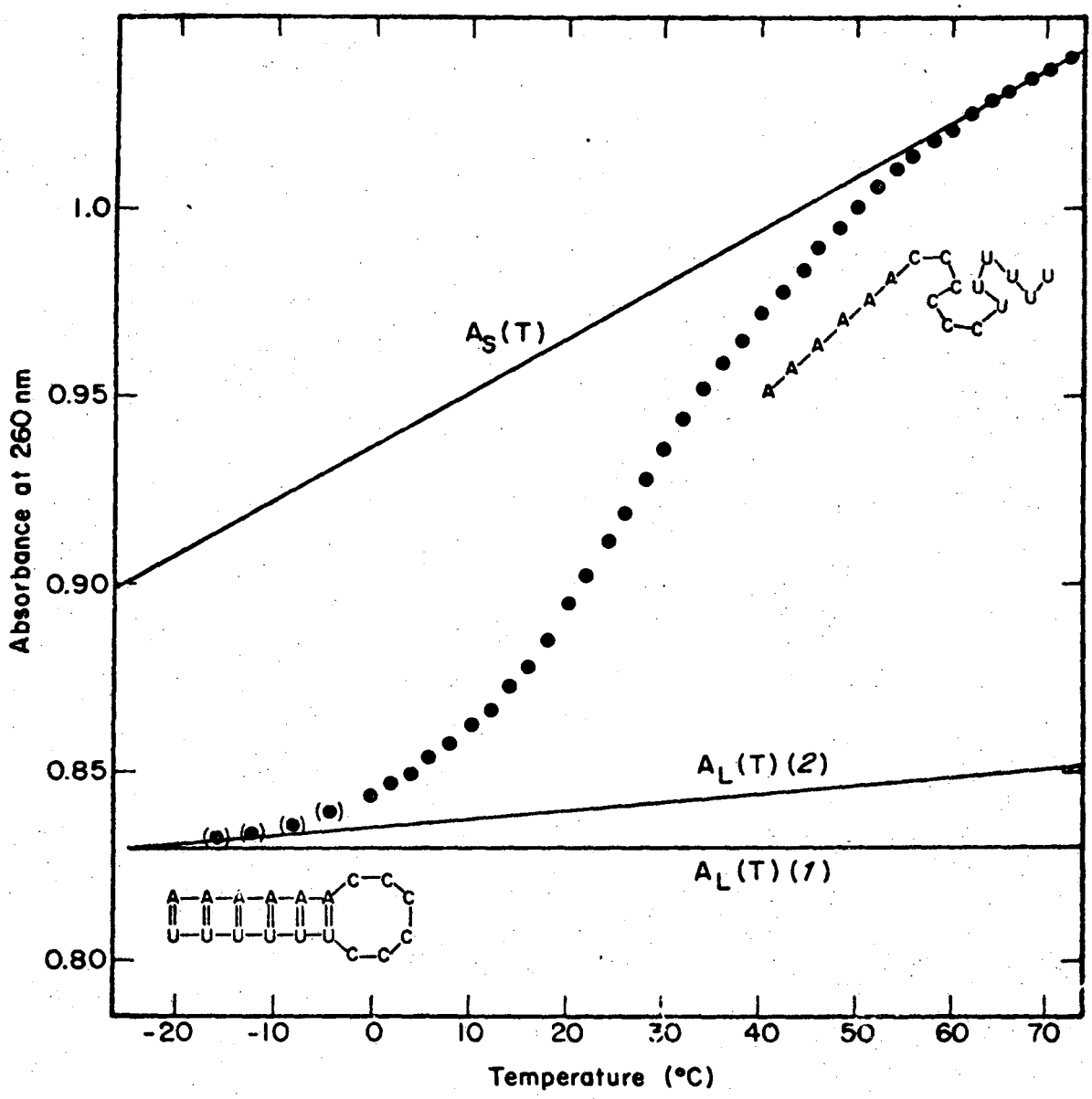
for duplex oligonucleotides. All calculations reported in this

$$1-f \sum_{m=m_o}^{m_o+2N} \sum_{n=1}^{N-(m-m_o)/2} (n-1)\gamma_m s^{n-1} k^p_{end}/(N-1) \qquad (4.4)$$

chapter are based on equation 4.4.

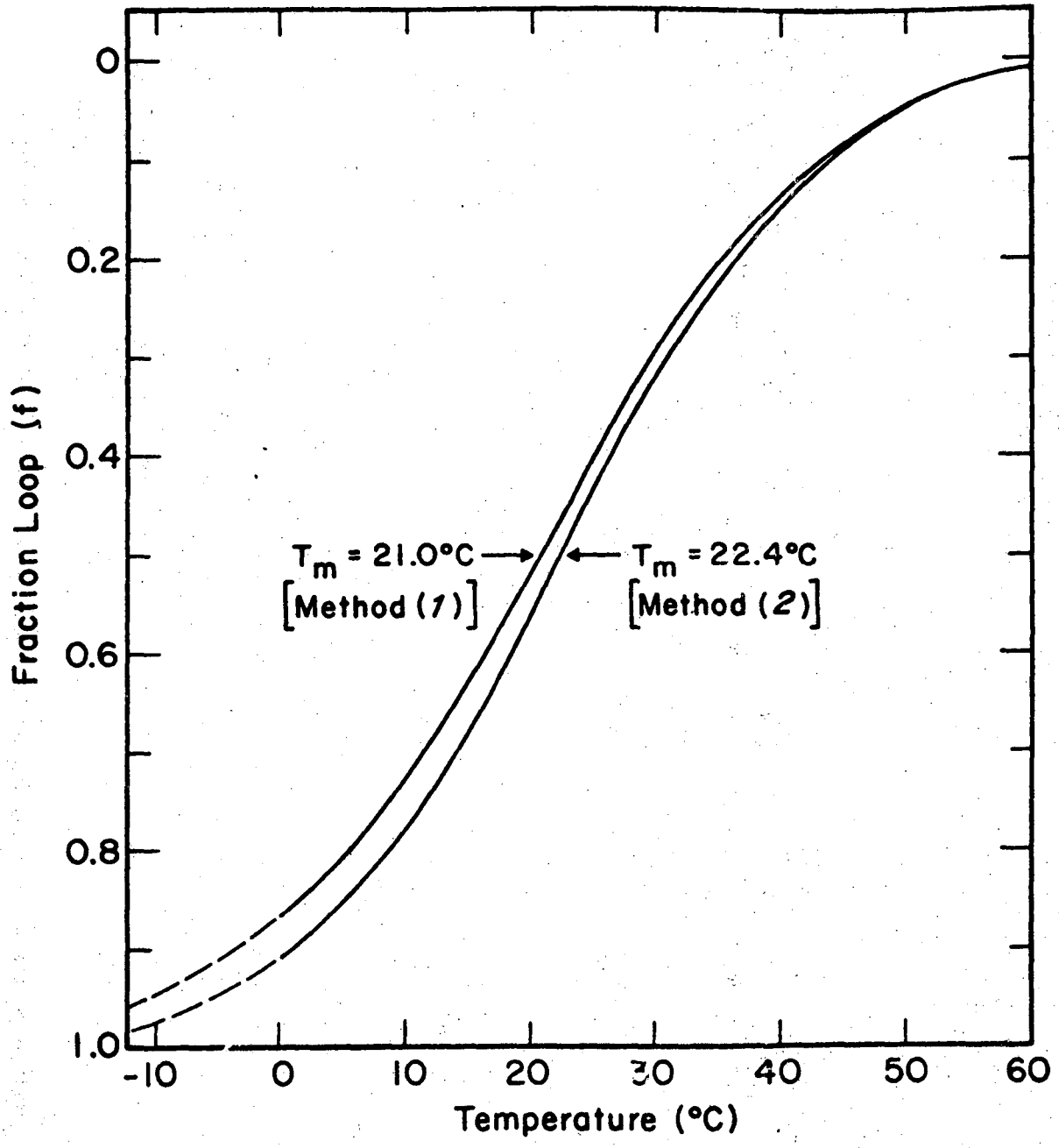
B. The Experimental Results of Uhlenbeck et al., are Summarized.

In Fig. 4.3, we have reproduced the published experimental melting curves for $A_6C_6U_6$. As was the case for duplex oligonucleotides, the high temperature region exhibits a constant slope characteristic of single stranded unstacking. The low temperature region of the melting curve may also possess a nonzero slope, since the C residues in the loop are free to change their geometry even after the stem region is formed. $A_{lo}(T)(1)$ in this figure represents the assumption that the low temperature baseline is constant with temperature; $A_{lo}(T)(2)$, based on data from oligo C, treats the baseline as if it represented the unstacking of single stranded C residues. Because $A_{lo}(T)(1)$ and $A_{lo}(T)(2)$ lead to very similar corrected melting curves (1-f versus T), Uhlenbeck, et al., have chosen to use the first assumption for the sake of simplicity. (See Fig. 4.4.) Because the melting curves cannot be extended significantly below 0°C in 1 M salt solution, it is not possible on the basis of existing information to know the behavior of the low temperature baseline. The synthesis and study of hairpin loop oligonucleotides of greater stability (molecules with long stem regions or with several G-C base pairs) will help resolve this difficulty.



(from ref. 1)

Figure 4-3



(from ref. 1)

Figure 4-4

We show, in Fig. 4.5, the published experimental absorption versus temperature profiles of $A_6C_mU_6$ (m=4,5,6,8). Each of these curves represents an average of about five different concentrations of hairpin loop oligomers in solution. The relevant parameters from these experimental data are summarized in Table 4.1.

Table 4.1

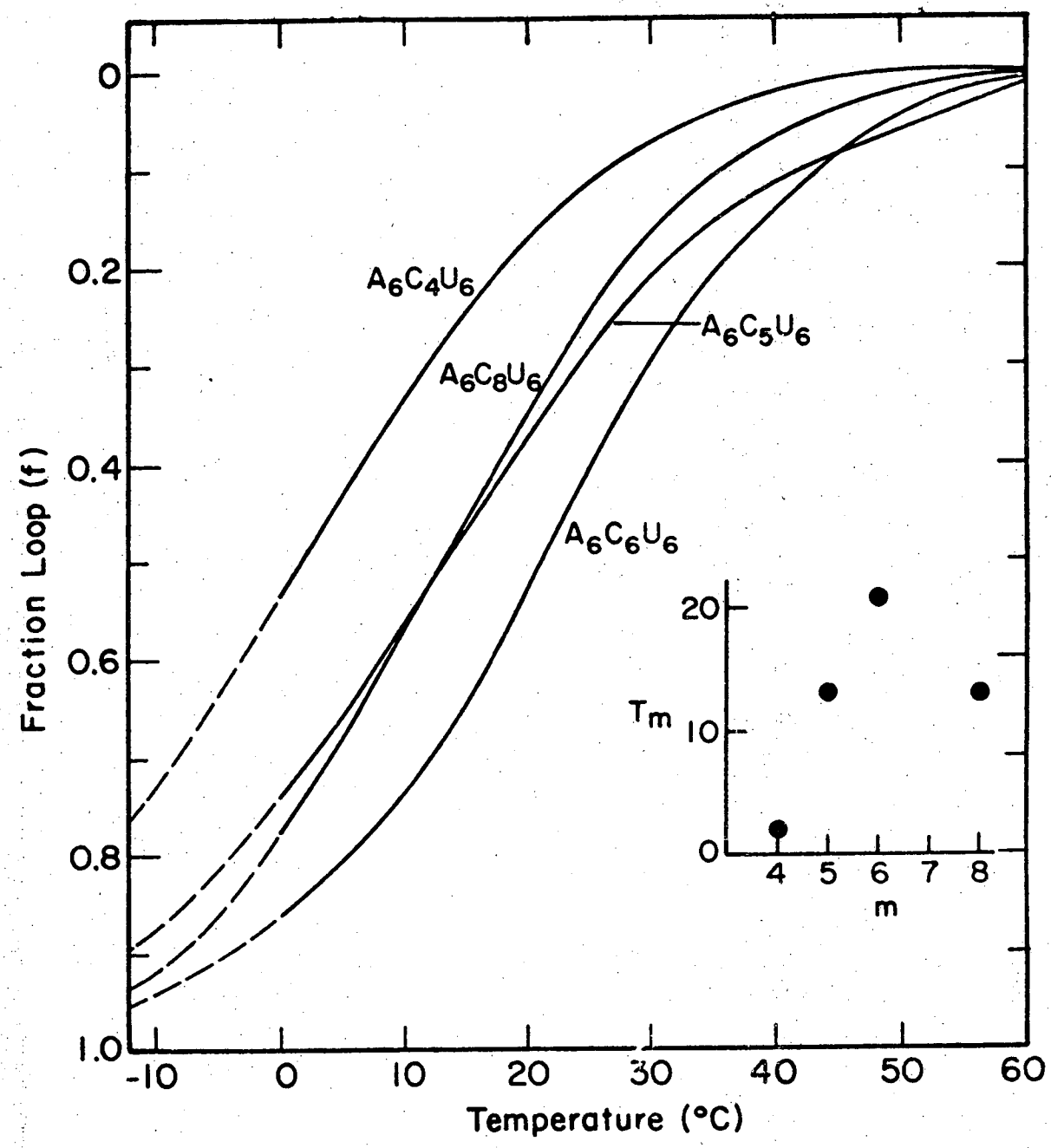
Melting Parameters for RNA Loops

Loop	T_m(°C)	$df/dT\|_m$
$A_6C_4U_6$	2.0	1.98×10^{-2}
$A_6C_5U_6$	13.4	1.96×10^{-2}
$A_6C_6U_6$	21.0	2.37×10^{-2}
$A_6C_8U_6$	13.4	2.16×10^{-2}

The slope of the melting curve at T_m is less great than one would expect for comparable duplex molecules. From the observed $df/dT|_m$'s for A_nU_n duplexes reported in Chapter 2, we might expect $df/dT|_m$ to be about 4×10^{-2} were no loop present. We infer from this that ΔH°_{tot} is not as great as for these RNA molecules as it would be if no interior loop were present.

C. Method of Calculation

We have listed and annotated program LOOP2 in Appendix II. This Fortran program has been used to perform the calculations on the $A_6C_mU_6$ hairpin loop molecules which are discussed in this chapter. Because the melting of a given loop generally depends on the value of γ_m for other loops, program LOOP2 calculates γ°_m and



(from ref. 1)

Figure 4-5

ΔH°_m for all four loops at one time. For example, the thermal behavior of $A_6C_5U_6$ depends most on the quantities γ°_5 and ΔH°_5, but also on γ°_6, ΔH°_6, γ°_7, ..., γ°_{15}, ΔH°_{15} in rapidly decreasing order. As in the earlier calculations on duplex oligonucleotides, the program was written so that the input was fed into the computer from a teletype machine and the output was returned to the teletype so that new estimates of the parameters could be made. The procedure followed for the four $A_6C_mU_6$ molecules is summarized.

1. Program LOOP2, which is based on equation 4.1, was used throughout.

2. The previously established set of thermodynamic quantities ΔH°_{AA}, ΔS°_{AA}, and k_{end} were always used for the stem region. Calculations with $k_{end} = 0.0585$ and $k_{end} = 1$ were made.

3. Initial guesses for γ°_4, γ°_5, γ°_6, γ°_8, and ΔH°_4, ΔH°_5, ΔH°_6, and ΔH°_8 were made. γ°_7 was set equal to $(\gamma^\circ_6 + \gamma^\circ_8)/2$ for lack of better information. ΔH°_9, ..., ΔH°_{20} were set equal to ΔH°_8 for the same reason. We used the relation in equation 4.2 to estimate the magnitudes of γ°_9 through γ°_{20}; namely, $\gamma_I = (I/I+1)^{3/2}\,\gamma_{I+1}$.

4. The predicted melting temperatures and slopes of the melting curves were compared with the experimental values of Table 4.1 and new guesses of γ°_4, ..., ΔH°_8 were made on this basis.

5. Steps (3) and (4) were repeated until experiment and theory were in good agreement: the melting temperatures were fit to within 0.1°C and $df/dT\big|_m$ to within 0.5%.

It is important to note that we have not considered any base composition or sequence dependence of the loop stabilities. The formation for example, would treat a loop of 6 C residues closed by an A-U base pair

as thermodynamically equivalent to a loop of one A, four C's, and one U closed by an A-U base pair. This is a reasonable assumption and in most cases should not lead to serious quantitative errors, since usually the minimum loop (with all C's) dominates the partition function. However, in the frayed end model, staggered species (with at least one A or U in the loop) make a large contribution to the partition function. Because the uncertainty associated with the assumption of sequence independence of the loop weighting function is more important in the frayed end model, estimation of γ_m is less accurate for this model. For the molecule $A_6C_mU_6$, the minimum loop does not contribute as much to the partition function as the loop of six residues, as calculated by any of the models used. As a result, the parameters γ_4° and ΔH_4° cannot be determined very accurately by our methods, unless the thermodynamics does not depend on the base composition and sequence of the loop.

D. Results

The loop weighting functions and loop enthalpies which we have calculated are reported in Table 4.2. As noted earlier, the calculations were repeated with new trial values of the parameters until agreement between theory and experiment was attained for each of the molecules. The reservations noted above for γ_4° and ΔH_4° should be kept in mind.

Table 4.2

Thermodynamic Parameters for RNA Loops

Loop	γ_m(T = 20°C)		ΔH°_m	
	Standard Model (k_{end}=1)	Frayed End Model (k_{end}=.0595)	Standard Model	Frayed End Model
$A_6C_4U_6$	0.13×10^{-5}	0.31×10^{-5}	24.3	24.0
$A_6C_5U_6$	0.45×10^{-5}	0.93×10^{-5}	24.0	22.5
$A_6C_6U_6$	1.10×10^{-5}	3.50×10^{-5}	20.8	17.4
$A_6C_8U_6$	0.54×10^{-5}	1.55×10^{-5}	23.9	22.3

Several observations and conclusions can be made from these results:

1. ΔH°_m for m = 4, 5, 6, and 8 is significantly greater than zero for both model calculations.

2. ΔH°_m for m = 4, 5, 6, and 8 are all of approximately the same magnitude. This is somewhat surprising inasmuch as a positive loop enthalpy is related to strain in loop formation. One would expect the smaller loops to be considerably more strained than larger ones. The near equality of the four enthalpy terms might be interpreted to mean that a certain constant part of the loop is subject to repulsive interactions and that these interactions do not vary greatly with the size of the loop. It is possible that this loop strain is associated with the ends of the loop nearest the stem region or that it has to do with the initiating

base pair. Both of these explanations are consistent with ΔH°_m independent of loop size.

3. The entropy of loop formation, which can be calculated directly from γ°_m, is greater than zero, indicating that the formation of the loop is entropically favored. This is very surprising; one would expect just the opposite, since the residues in the loop are more constrained than the same residues in the single strand. The explanation may have something to do with nucleic acid-solvent interactions. It is possible that the formation of a hairpin loop is accompanied by a net decrease in the binding of water molecules or of ions to the RNA.

4. The loop weighting function goes through a maximum with six or seven bases in the loop. This suggests that it is not entirely fortuitous that tRNA loops often contain this number of bases.

These results make clear that the simple theoretical treatment which treats the loop as being composed of freely moving links is inadequate. It is apparent that the loop forming process is considerably more complicated than the theory assumes.

We also note that the results of the frayed end and standard statistical thermodynamic models are qualitatively similar and differ only slightly quantitatively. This is in part due to the fact that the loop molecules have just one end, rather than two for the duplex molecules. For this reason, it does not matter greatly which set of parameters is used in comparing the free energies of different secondary structures, as we do in the next chapter.

For further discussion of these hairpin loops, the reader is referred to reference 1.

CHAPTER 4

REFERENCES

1. O. C. Uhlenbeck, P. N. Borer, B. Dengler, and I. Tinoco, Jr., J. Mol. Biol., 73, 483-496 (1973).

2. H. Jacobson and W. Stockmayer, J. Chem. Phys., 18, 1600 (1950).

3. N. Kallenbach, J. Mol. Biol., 37, 445-466 (1968).

4. I. E. Scheffler, E. L. Elson, and R. L. Baldwin, J. Mol. Biol., 48, 145-171 (1970).

CHAPTER 5

There is currently much work devoted to determining the primary structure (i.e., the base sequence) of ribonucleic acids. This sequence information is primarily valuable in providing insight into possible biological functions of the molecule. In order to effectively use the sequence information, it is necessary to know the secondary structure (i.e., the most stable arrangement of base pairs, loops and bulges) of the molecule. A variety of methods are available, mostly from physical chemistry, to learn about certain aspects of RNA secondary structure in solution. Measurement of infrared and ultraviolet absorption profiles can yield information about the number of base pairs formed;[1] oligonucleotide binding studies can locate regions of the RNA which are probably not base paired;[2] fluorescence measurements can approximate distances between certain regions of the RNA;[3] nuclear magnetic resonance can indicate whether or not a modified nucleic acid base is base paired.[4] Although all of these techniques hold considerable promise in helping to establish secondary structures of RNA molecules in solution, the information which they provide has hitherto been too scanty to determine which of a large number of secondary structures is most stable for a given RNA molecule.

At the present time, no systematic method of predicting secondary structures from the sequence of the RNA molecule has been developed. The standard procedure is to find a structure which appears to maximize base pairs and to assume that this is the

correct secondary structure. The literature of secondary structure of tRNA, in which a large number of different structures were proposed until the cloverleaf structure was generally agreed upon, suggests the hazardous nature of this endeavor. The secondary structure of tRNA was worked out in large measure because of the large number of different molecules (with different sequences) all of which performed similar biological functions and all of which could be fit into the same general secondary structure.

The purpose of the work discussed in this chapter is to develop a systematic method for predicting the secondary structure of an RNA molecule, once the sequence is known. Although the method which we present is not yet fully developed, we believe it is a significant improvement over the guesswork which is generally employed to predict secondary structure of RNA. As the thermodynamic parameters which govern secondary structure formation are better known, as the assumptions we have made in the model calculations are more completely tested, and as results are compared with information derived from physical experiments on RNA molecules, we believe that it will be possible to almost automatically predict many secondary structures of ribonucleic acid molecules. This work represents one step in this direction.

It may seem that for a sequence of about 80 bases (the size of tRNA) only a few possible secondary structures are likely. This is, however, generally not the case and the number of combinations which must be considered can be extremely large. To illustrate this fact, consider a molecule with just twenty nucleic acid bases, five each of A, U, G, and C. An upper limit to the number of different

secondary structures can be obtained if we ignore all steric constraints imposed by the sequence of the molecule and allow any base to pair with its complement regardless of the other base pairs which are present. (For this simple illustration, we ignore G-U base pairs and we stipulate that a base may participate in only one base pair.) For this case, the problem is a combinatorial one which may be represented by five white and five black boxes, and five white and black balls, all of which are distinguishable. The problem then is to determine the number of ways that the balls can be placed in the boxes, subject to the constraint that balls may only be placed into boxes of their color. The number of different combinations is equal to the number of ways of placing one ball in the boxes ($2\times5^2 = 50$) plus the number of ways of placing two balls in the boxes ($2\times(5\times4)^2 + (5\times5)^2 = 1425$) plus the number of ways of placing three balls in the boxes ($2\times(5\times4\times3)^2 + 2\times(5\times4)^2\times5^2) = 27{,}200$), etc. The total number of different combinations for this molecule with just twenty bases is greater than 10^9! For a molecule with 25 of each of the four bases, there are greater than 10^{63} different ways of forming 50 base pairs. Of course, this illustration greatly overestimates the number of possible secondary structures, since it ignores all steric constraints imposed by the sequence of the molecule for real nucleic acids. Nonetheless, this should serve to illustrate the fact that for a nucleic acid the size of tRNA or larger, an enormous number of different base pairing arrangements leading to a stable secondary structure may be possible.

A. The Model

The calculations reported were performed by program DBL which is listed in Appendix III. The steps involved in the calculation serve to illustrate the method which we have used to predict secondary structures.

1. Thermodynamic Parameters for Base Pairs, Loops, and Bulges are Specified.

We use the thermodynamic parameters that were obtained from the analysis of melting curves in the previous chapters. This specifies double stranded stacking interactions and loop free energies as well as they are presently known. The bulge free energies are taken from Tinoco et al.[5] Gralla and Crothers estimate the free energy of a simple interior loop to be +2 to +3 kcal/mole at room temperature.[6] For the molecules which we are interested in, most interior loops have several stem regions extending from them which must be closed in order to form the loop. (For example, tRNA has four such regions.) The probability of loop closure should be significantly decreased because of the multiple stem regions. For this work, we make the approximation that the interior loop free energy is 6 kcal/mole for an interior loop of any size. This makes the interior loop slightly more stable than most hairpin loops. Improvement of our knowledge of the stability of interior loops is necessary in order to increase the accuracy of these secondary structure calculations. So long as structures with the same number of interior loops are compared, as is usually the case, this lack of information is probably not a serious problem.

The stability parameters which have been used in the

calculation are summarized below; they are all reported at 25°C, as the calculation is performed at this temperature.

Double Strand Stacking Free Energies[†]

$\Delta G^\circ_{AA} = \Delta G^\circ_{AU} = -1.3$ kcal/mole

$\Delta G^\circ_{UA} = -2.3$ kcal/mole

$\Delta G^\circ_{AG} = \Delta G^\circ_{AC} = -2.0$ kcal/mole

$\Delta G^\circ_{GA} = \Delta G^\circ_{AC} = -1.6$ kcal/mole

$\Delta G^\circ_{GC} = -5.0$ kcal/mole

$\Delta G^\circ_{CG} = -3.9$ kcal/mole

$\Delta G^\circ_{CC} = -4.9$ kcal/mole

$\Delta G^\circ_{\text{initiation (A-U)}} = +4.4$ kcal/mole

$\Delta G^\circ_{\text{initiation (G-C)}} = +3.3$ kcal/mole

Hairpin Loop Free Energies

Loop Size (number of links)[††]	ΔG° (25°C)
4	+8.1 kcal/mole
5	+7.1
6	+6.5
7	+5.8
8	+5.9
9	+6.2
10	+6.3
11	+6.4
12	+6.4
13	+6.5

Interior Loop Free Energy:

+6.0

[†]The double stranded stacking free energies are slightly different from Table 3.15, as they are based on an earlier normalization for k_{IJ}. The use of these free energies changes the results insignificantly since they differ from the later free energies by 0.1 kcal or less. We do not count the end effect free energy in these calculations, as it is small and equal for most secondary structures.

[††]Number of links = number of unbonded bases + 1.

Bulge Loop Free Energies[5]

Loop Size (number of links)	ΔG° (25°C)
2	+3.0 kcal/mole
3-4	+4.0
5-8	+5.0
9-21	+6.0
m > 21	+4 + 2 log(m)

The uncertainties introduced into the calculation by the thermodynamic parameters are as follows: double stranded stacking interactions are the best known, and most are probably accurate to +10%. The main improvement over previous work is the sequence dependence of these free energies. As explained in the previous chapter, the free energies of hairpin loops are less well known, because of the uncertainty in our knowledge of the enthalpy of loop formation. However, since most of the loops melt rather close to 25°C, the temperature used in the calculations reported, the uncertainty in the temperature dependence of the free energy is not crucial to the calculations. The two important sources of error are the lack of information about possible sequence dependence of loop free energies and the free energy of interior loops.

One limitation in the methodology is that in the calculation of free energies of a secondary structure, program DBL cannot calculate the free energies exactly by summing over the double stranded stacking and loop free energies, for reasons which will be apparent from later discussion. This necessitates correction of the free energies (or stability numbers, defined as the negative of the free energies) in the final analysis.

2. A Matrix, Called the Base Pairing Matrix, is Formed from the Known Sequence of the Molecule

The base pairing matrix is specified by its elements a_{ij}, where the subscripts i and j refer to the positions of the i th and j th bases in the sequence of the molecule. (The counting begins at the 5' end of the molecule and progresses in the 5' to 3' direction.) The base pairing matrix terms a_{ij} specify whether bases i and j can form a base pair with each other according to the following rules, in which a number other than zero means that a pair can be formed:

a_{ij} = 1 for a G-U base pair

a_{ij} = 2 for an A-U pair

a_{ij} = 4 for a G-C pair

a_{ij} = 0 if no base pair is possible

a_{ij} = 0 if $|i-j|$ is less than or equal to 3

The last constraint stipulates that a hairpin loop must contain at least two bases to be stable. Since a_{ij} equals a_{ji}, the matrix is specified only for j less than or equal to i.

The following molecule with the bases numbered by position serves to illustrate the base pairing matrix. (See Appendix III for the base pairing matrix of a tRNA molecule.)

AUGCCUUACG
12345678910

For this molecule, $a_{1,10}$ = 0 (since A and G do not base pair).

$a_{1,7}$ = 2 (an A-U base pair)

$a_{2,10}$ = 1 (a G-U base pair)

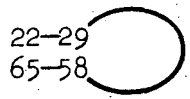
$a_{3,9}$ = 4 (a G-C base pair)

$a_{3,5}$ = 0 (hairpin loop is too small to permit formation of G-C base pair)

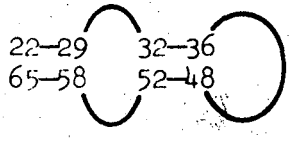
3. Base Pairing Regions (Also Called Base Pairing Vectors or Simply Vectors) are Determined by Scanning the Base Pairing Matrix

This part of program DBL selects all sequences of base pairs which can form a stable double helical region in the RNA molecule. A double stranded region is composed of elements $a_{i,j}$, $a_{i+1,j-1}$, $a_{i+2,j-2}$, . . ., in which all elements are different from zero. This way of specifying a double stranded region derives from the fact that base pairing is anti-parallel and that a given double stranded region must progress sequentially from one base to the next. This means that all possible base pairing regions can be located by noting all sequences of nonzero elements along the +45° diagonals of the base pairing matrix.

For each sequence so noted, the free energy is calculated. This calculation includes an estimation of the free energy of the loop associated with the double stranded region. The loop free energy used in the calculation is likely to be somewhat in error, since the loop might be one from which, in the final structure, one or more additional base pairing regions may be formed. For example, the base pairing denoted (22,29;65,58) refers to the following base pairs and loop in an RNA:



The final secondary structure for this part of the molecule might look like:



It is thus apparent that at this stage of the calculation we can only approximate the size of the loop associated with the double stranded base pairing region.† Because larger loops are more likely to permit additional base pairs than small ones, we have assumed a maximum loop size of twenty bases. (This assumption is made only for convenience; if any loops have more than twenty bases in the final predicted secondary structure, the appropriate correction can be made.)

All base pairing regions with a calculated free energy less than zero (or some arbitrary cutoff, if so desired) are retained and arranged in order of decreasing stability. The calculated free energy is the sum of the double stranded stacking free energies, the initiation free energy, and the loop free energy. Thus, this part of program DBL finds all potentially stable base pairing regions.

One inadequacy in the scheme should be noted. The runs of nonzero integers on the diagonals of the base pairing matrix represents perfect double helical regions. If the helix is interrupted by one or more looped out bases, then the program treats the interrupted helix as two separate base pairing regions. Stable regions with looped out bases might be ignored by this search routine, if the two halves of the double stranded region are individually unstable but together (even with the increase in free energy associated with the looped out bases) form a stable

---

†This is the source of the earlier statement that corrections must be made at the end of the calculation because program DBL cannot calculate free energies exactly.

region. Additional work to overcome this deficiency might well be a fruitful avenue of research.

4. A Vector Exclusion Matrix Specifies the Base Pairing Regions which can be Present Simultaneously in a Stable Structure

This part of the calculation is best explained with an illustration. Consider a molecule in which are found the following stable regions (listed in decreasing order of stability):†

V1: (28,34;49,43)

V2: (4,10;32,26)

V3: (50,55;69,64)

V4: (50,55;72,67)

V5: (32,36;57,53)

V6: (36,39;46,43)

V7: (55,59;66,62)

For these seven vectors, the vector exclusion matrix, e_{ij}, is as follows:

---

†To explain the notation, V1 refers to the region

28–34
49–43

	V1	V2	V3	V4	V5	V6	V7
V1	0	1	0	0	1	1	0
V2	1	0	0	0	(0)	0	0
V3	0	0	0	1	1	0	(0)
V4	0	0	1	0	1	0	(0)
V5	1	(0)	1	1	0	(0)	1
V6	1	0	0	0	(0)	0	0
V7	0	0	(0)	(0)	1	0	0

If $e_{ij} = 0$, then the two vectors i and j can exist together in a stable secondary structure. If $e_{ij} = 1$, then the two vectors overlap such that two or more bases are simultaneously involved in two base pairs—an impossibility, so far as is known for nucleic acid bases. For example, $e_{4,5} = 1$, since bases 53, 54, and 55 would each participate in two base pairs at the same time if V4 and V5 were present simultaneously. An exception to this rule is made for those vectors which overlap in only one position, shown in the above matrix with $e_{ij} = (0)$. In these cases, only one base would have to be a part of two different base pairs. This means that breaking one base pair (an end base pair in the sequence) would allow the structure to exist. For example, $e_{3,7} = (0)$, since only one base, number 55, is involved in two base pairs. This means that either of the following two structures would be permitted:

V3: (50,55;69,64) with V7': (56,59;66,63)

V3': (50,54;69,65) with V7: (55,59;66,62)

(The vector with the ' has been shortened in order to avoid the overlap.) Since both of these two arrangements have a free energy only one double stranded stacking interaction less stable than if

the one overlap did not exit, it would be imprudent to ignore their possible existence. For this reason, we have set e_{ij} in the vector exclusion matrix equal to zero if no base or one base is involved simultaneously in two base pairs. (In the latter case, the free energy of the final secondary structure must be corrected for the base pair which cannot form.)

5. Using the Vector Exclusion Matrix, the Most Stable Sets of Base Pairing Regions (Vectors) are Determined, Thereby Specifying the Preferred Secondary Structure(s).

The problem to be solved is fully specified by the vector exclusion matrix and the stability numbers of the vectors. We must find the set of vectors which are mutually compatible (i.e., for which $e_{ij}=0$ for all vectors in the set) and whose free energy is a minimum. Since there are generally fifty to one hundred or more vectors representing stable base pairing regions to consider for RNA molecules of the size we are interested in (80-150 bases), the number of possible combinations is large; for most RNA molecules of interest, it would require as much as an hour or more of computer time, if all possible combinations were analyzed. It is thus necessary to make one or more simplifying assumptions. We have assumed that at least one of the most stable seven vectors occurs in the final secondary structure. Although this approximation cannot be rigorously justified, it is fully consistent with what is known about the stability of RNA secondary structure. A secondary structure with a large number of small double helical regions would have a correspondingly large number of destabilizing loops, since each double helical region has a loop associated with

it. For this reason, the secondary structure formation strongly favors a few very stable regions over a larger number of less stable ones.

The solution is composed of the following steps:

Step 1: V1, the most stable base paired region, is assumed to be in the preferred secondary structure. (All following steps are repeated with V2 through V7 replacing V1.)

Step 2: A new, reduced vector exclusion matrix is formed, in which all vectors not compatible with V1 are excluded. Only the most stable fifteen vectors consistent with V1 are retained, in order to limit computation time. (In general, about half of the most stable vectors are not compatible with the vector assumed to be in the solution. This means that the new exclusion matrix generally extends out to V30 or so.)

Step 3: All possible combinations of these fifteen vectors are considered, using the reduced exclusion matrix to determine which vectors are compatible with each other. All sets of vectors having a stability number (negative of the free energy) greater than a specified cutoff are retained and reported as output by program DBL. (This cutoff energy will vary from molecule to molecule. It is calculated internally within the program as 90% of the free energy of the most stable secondary structure determined using the above steps, but assuming that at least one of the most stable five vectors is in the final solution and using a reduced exclusion matrix of 5 x 5 elements.)

Step 4: Steps 1 through 3 are repeated for vectors V2 to V7.

For a detailed explanation of the workings of program DBL, which includes examples of the output and a discussion of the internal logic of the program, the reader is referred to Appendix III. Here we illustrate steps 1 through 3 with a simple example. Consider the following vectors with their stability numbers and associated exclusion matrix:

VECTOR	STABILITY NUMBER
V1	11
V2	9
V3	8
V4	7
V5	5.5
V6	5.5
V7	5
V8	4

Exclusion matrix:

	V1	V2	V3	V4	V5	V6	V7
V1	0	0	1	1	0	0	0
V2	0	0	0	1	1	0	1
V3	1	0	0	0	1	1	0
V4	1	1	0	0	1	0	1
V5	0	1	1	1	0	0	0
V6	0	0	1	0	0	0	0
V7	0	1	0	1	0	0	0

Assuming first that V1 is in the solution set (step 1), the reduced exclusion matrix for V1 (step 2) is:

	V2	V5	V6	V7
V2	0	1	0	1
V5	1	0	0	0
V6	0	0	0	0
V7	1	0	0	0

All combinations of vectors are considered to determine which vectors sets are allowed (step 3). The procedure followed in program DBL combines the vectors in the following order: V7; V6; V6,V7; V5; V5,V7; V5,V6; V5,V6,V7; V2; V2,V7; V2,V6; V2,V6,V7; V2,V5; V2,V5,V6; V2,V5,V7; V2,V5,V6,V7. For each of these vector sets, reference is made to the exclusion matrix to determine if they are allowed (i.e., to determine if the vectors are mutually compatible). If so then the stability number is calculated. If it is greater than the cutoff value, it is retained and later outputted along with the vectors which make up the vector set.

For the calculation on a real molecule, this procedure is performed on a 15 x 15 reduced exclusion matrix. The number of different combinations is $\sum_{i=0}^{14} 2^i$ or slightly less than 36,000. Of this large number of combinations, most are forbidden. The calculation is done on a computer relatively quickly, since the determination of whether or not a combination is allowed can be done in a small number of steps.

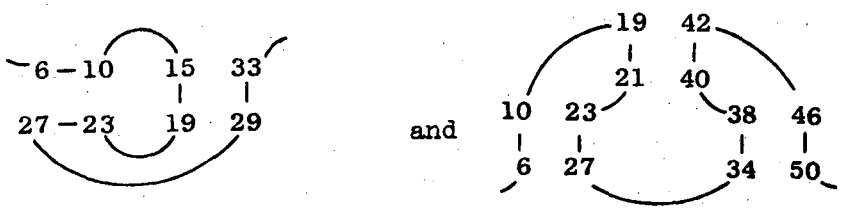
For the problem considered, the results are:

VECTOR SET	STABILITY NUMBER (Including V1)
V7	16
V6	16.5
V6,V7	21
V5	16.5
V5,V7	21.5
V5,V6	22
V5,V6,V7	27
V2	20
V2,V6	25.5

The most stable vector set is V1, V5, V6, V7. The set V1, V2, V6 is 1.5 kcal/mole less stable. When the procedure is repeated assuming each of the six vectors other than V1 to be present in turn in the final solution, no other sets of vectors emerge as possible contenders for the secondary structure. The final choice between the two best sets requires that the stability numbers be recalculated, with corrections made for (1) any changes in loop size by the final vector set, and (2) any base pairs which must be broken because an overlap of one base has been allowed in the vector exclusion matrix.

B. Results

One additional assumption has been made for all RNA molecules which have been calculated using program DBL. In order to limit the number of solution sets, we have forbidden nucleic acid bases in loops to base pair with bases in other regions of the RNA molecule. This eliminates structures of the type:



This is a reasonable assumption to make, since in most instances the steric constraints on the molecule would not allow such types of base pairs to form. It is a necessary assumption in that it eliminates a very large number of secondary structures in which the strands of the RNA are twisted around one another so that the structure cannot be drawn in two dimensions. It is difficult to imagine a mechanism whereby many of these secondary structures would form. It is important to realize that, while this assumption is necessary to limit the number of solution sets and is consistent with what is known about RNA secondary structure, it cannot be fully justified. Program DBL is written to do each calculation both with and without this assumption. It is instructive to note the myriad of "unusual" vector sets which result when the assumption is not invoked.

1. tRNA[†]

The main reason for applying the model to the prediction of tRNA secondary structure is to determine how well it works and to make note of its limitation. There is general agreement that tRNA takes on a secondary structure which looks like a cloverleaf since this is the most stable conformation into which all tRNA molecules can be fit. Because of the similarity of function of all

---

[†]For the sequence of tRNA molecules, the reader is referred to R. Holmquist, T. H. Jukes, and S. Pangburn, J. Mol. Biol., 78 91-116 (1973).

tRNA molecules during protein synthesis, it is undoubtably necessary that all have similar secondary structures. Because tRNA molecules perform biological functions in addition to directing the incorporation of amino acids during protein synthesis, it might be interesting to speculate on other possible stable secondary structures which tRNA might form. Some tRNA molecules are known to exist in stable deactivated forms which probably differ from the native form because of difference in secondary structure.[1]

a. Valine tRNA from brewer's yeast

The sequence of valine tRNA from brewer's yeast is shown below. D = dihydrouridine, P = pseudouridine, and Y is an unknown base. There is no evidence that D and Y participate in base pairs; they are not permitted to form base pairs in the calculation.

Sequence:

GGUUUCGUGGUCPAGDCGGDDAUGGCAPCUGCPUYACACGCAGAACDCCCC

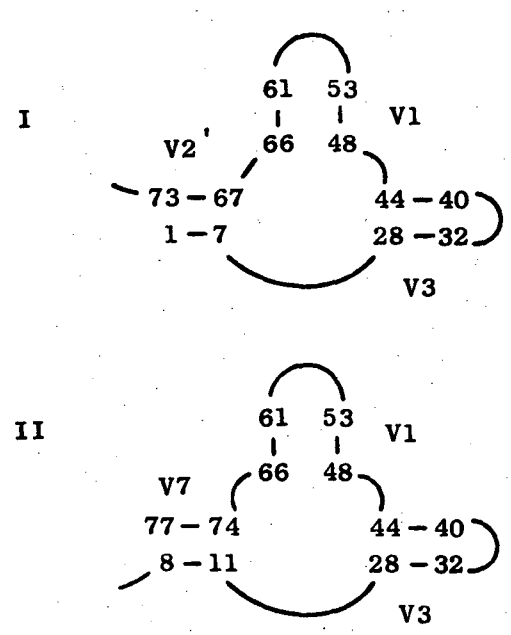
10 20 30 40 50

AGUPCGAUCCUGGGGCGAAAUCACCA

60 70

Only 17 base pairing regions more stable than 0.5 kcal/mole were found for val tRNA. (The free energy of the base pairing region is the sum of the free energies of the double strand stacking, initiation, and loop formation free energies at 25°C.) Altogether 41 base pairing regions with a negative free energy were found, but 20 of these had a free energy of -0.1 kcal/mole or less; these latter regions are unimportant as factors in stabilizing the molecule through secondary structure.

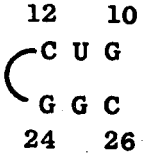
Two secondary structures more stable than the cutoff free energy were found:



In structure I, the prime on V2 indicates that this region has been shortened by one base pair because of an overlap with another vector. Base number 66 was the one involved in two base pairs in the results. The corrected structure pairs base 66 with base 48, forming a G-C base pair. If base 66 were paired with base 8 (the other choice), a G-U base pair would be formed, adding less to the stability of the molecule. Because structures I and II differ by less than 10% in free energy, it would not normally be prudent to predict with assurance which of the two secondary structures would be favored. In this case, however, because of the great similarity between the two — they differ only in one double stranded region — it is safe to conclude that structure I is the favored one.

When the predicted secondary structure is compared

with the cloverleaf structure for val tRNA, we note that the dihydrouridine loop is absent. This results from the fact that the base pairing region associated with the dihydrouridine loop consists of the following sequence:



The free energy of this region is -1 -1 -1.1 +6 = +2.9 kcal/mole.† As noted in Table 5.1, we have taken the free energy for a G-U base pair in any double stranded interaction to be -1 kcal/mole. Unfortunately, the sequence dependence of G-U to base pairs is not known. If the free energy contribution of G-U to each of the two double stranded stacking interactions is -2.5 kcal/mole or greater (i.e., more negative) in the sequence above, then the dihydrouridine loop would be stable.

b. f-Met tRNA from E. coli.

The sequence of f-met tRNA from E. coli is:

CGCGGGGUGGAGCAGCCUGGDAGCUCGUCGGGCUCAUAACCCGAAGGUCG

10 20 30 40 50

UCGGUPCAAAUCCGGCCCCCGCAACCA

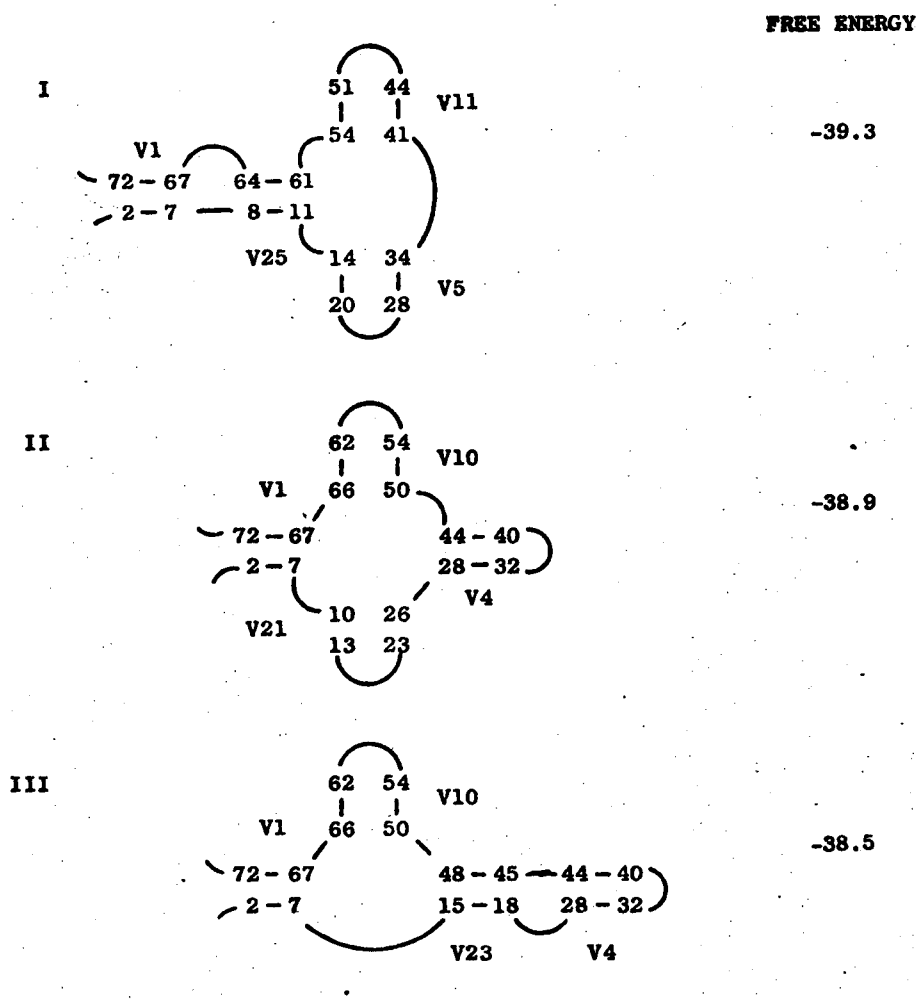
60 70

A large number of possible base pairing regions were found for this molecule: 87 with a negative free energy, 46 more stable than -1 kcal/mole, and 22 more stable than -3 kcal/mole.

---

†The -1.1 kcal/mole is the amount by which initiation at a G-C is favored.

The most stable secondary structures are:



The next best secondary structure is 3 kcal/mole less stable than structure III. The accuracy with which the free energies are known do not make it possible to choose among the three predicted secondary structures, since they differ in free energy by so little. The cloverleaf model is structure II. One of the stabilizing factors which favors structure I over structure II is the bulge of two bases following V1, with a positive free energy of +4 kcal/mole.

instead of a loop which would have a free energy of about +6 kcal/mole. As we have noted, the possible sequence dependence of bulge free energies is not known; the fact that this bulge consists of a C and G base may affect the free energy, although in what manner is not known. Another uncertainty associated with the free energy of this bulge is that it is terminated by a G-U base pair (positions 8,64) which may not be sufficiently stable in this environment to close a bulge.

In summary, it is not possible to unambiguously predict the secondary structure of f-met tRNA because of uncertainty in the free energies, although we can with good assurance limit the choice to one of the three structures above. It is possible that f-met tRNA, under some conditions, might form one of these other secondary structures.

c. Phenylalanine tRNA from yeast is:

GCGGAUUUAGCUCAGDDGGGAGAGCGCCAGACUGAAYAPCUGGAGGUCCU

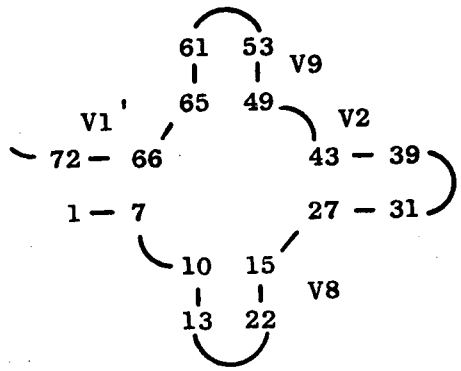
10 20 30 40 50

GUGUPCGAUCCACAGAAUUCGCACCA

60 70

Only 36 base pairing regions with a negative free energy were found. Seventeen of these were more stable than -1 kcal/mole. The procedure produced just one secondary structure, the cloverleaf model, which was corrected for one overlap in V1. Because of the paucity of base pairing regions, this problem is sufficiently simple that it could easily be solved without the use of a computer.

FREE ENERGY



-20.8

d. Phenylalanine tRNA from E. coli.

The sequence of phe tRNA from E. coli is:

GCCCGGAUAGCUCAGDCGGDAGAGCAGGGGAPUGAAAAPCCCCGUGXCCU

10 20 30 40 50

UGGUPCGAUUCCGAGUCCGGGCACCA

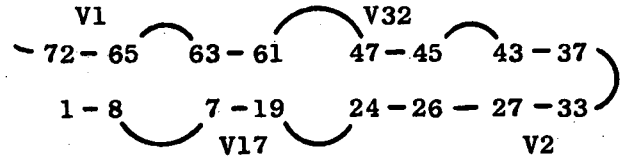
60 70

The identity of the base at position 47 is unknown. We allow it to pair with any base for the purposes of this calculation. (This is accomplished by entering an X for the base, when the sequence of the molecule is read by program DBL.)

Like f-met tRNA, phe tRNA from E. coli has a large number of stable base pairing regions. Some of these are artificial, because we have allowed the base at position 47 to pair with all other bases. 72 base pairing regions were found, with 46 more stable than -1 kcal/mole, 36 more stable than -2 kcal/mole, and 27 more stable than -3 kcal/mole. The most stable structures are:

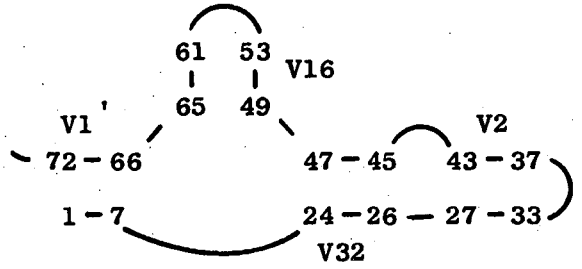
FREE ENERGY

I



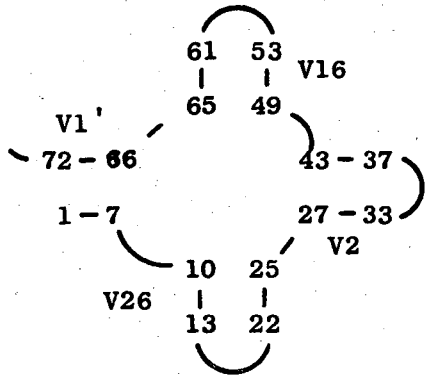
-41.2

II



-40.6

III



-40.0

The next most stable structures are similar to the above structures, but with one double stranded region broken. Thereafter, the best structures are at least 4 kcal/mole less stable than structure III. In this calculation, structure III is the cloverleaf model. Structures I and II are possible only if the unknown base pairs as if it were a cytosine residue. If this is not the case, then the cloverleaf is the favored secondary structure, since base 47 is not base paired in the cloverleaf structure.

e. Alanine tRNA from baker's yeast

The sequence of this molecule is:

GGGCGUGUGGCGCGUAGDCGGDAGCGCGCUCCCUUYGCYPGGGAGAGUCU

10 20 30 40 50

CCGGUPCGAUUCCGGACUCGUCCACCA

60 70

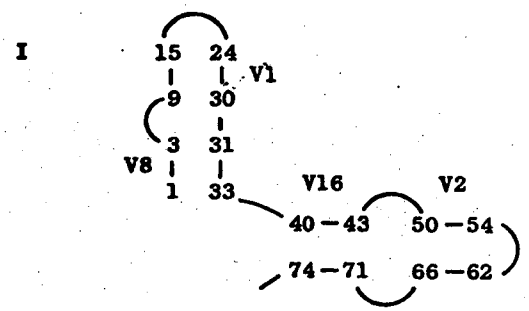
The distribution of base pairing regions is:

20 more stable than -2 kcal/mole

33 more stable than -1 kcal/mole

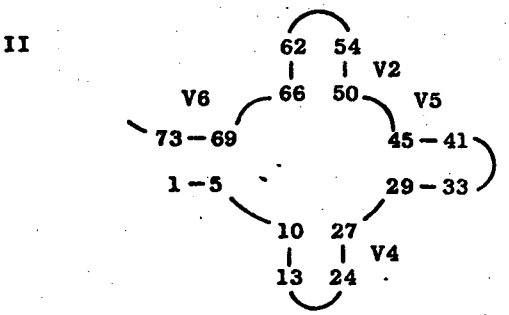
62 more stable than 0 kcal/mole

The calculation found the following secondary structures:

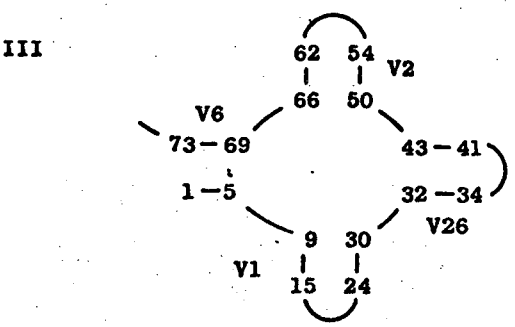


FREE ENERGY

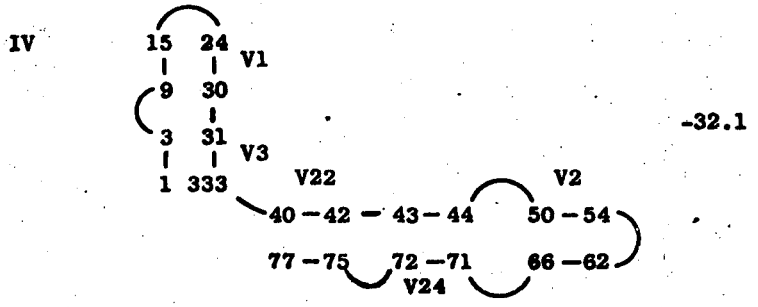
-34.1



-33.9



-32.7



Two different cloverleaf structure (II and III) are indicated. The calculated free energies of the two most stable structures are within 0.2 kcal/mole of one another, so that it is impossible to choose between them. If only double stranded stacking interactions are considered, structure I is less stable than structure II, but the bulge in structure I between bases 3 and 9 is presumed more stable than the corresponding loop in structure II. Structure IV has one additional base pairing region and one more base pair than structures I and II. It is less stable because it possesses an extra bulge.

f. Tyrosine tRNA from baker's yeast.

The sequence of this molecule is:

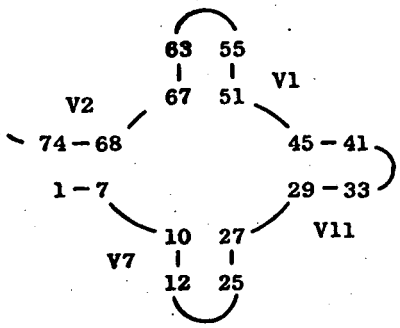
CUCUCGGUAGCCAAGDDGGDDDAAGGCGCAAGACUGPAAAPCUUGAGADC

10 20 30 40 50

GGGCGUPCGACUCGCCCCCGGGAGACCA

60 70

Only 15 base pairing regions more stable than -1 kcal/mole are possible out of the 40 regions which were found. Only one secondary structure is calculated to be more stable than the cutoff free energy and it is the cloverleaf structure.



FREE ENERGY

-30.0

g. Serine tRNA from brewer's yeast.

The sequence of this molecule is:

GGCAACUUGGCCGAGDGGDDAAGGCGAAAGAPUYGAAAPCUUUUGGGCUU

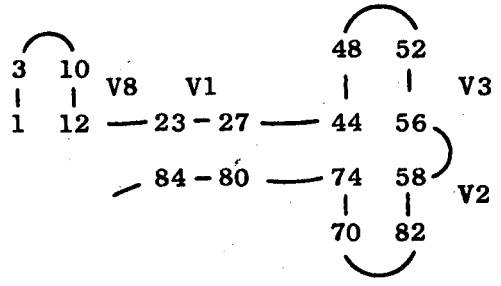
10 20 30 40 50

UGCCCGCGCAGGUPCGAGUCCUGCAGUUGUCGCCA

60 70 80

Sixty-two base pairing regions were found with a negative free energy; 38 of these were more stable than -1 kcal/mole; 28, more stable than -2 kcal/mole; 21, more stable than -3 kcal/mole. The results for this molecule indicate a problem which, although it is not likely to occur often, is sufficiently serious to merit further consideration. As noted earlier, the exclusion matrix permits 0 or 1 overlap between base pairing regions. When two or more overlaps occur between any two regions, they are considered incompatible. For ser tRNA, two overlaps do occur in the cloverleaf structure, since one of the base pairing regions can be extended by two base pairs, thereby increasing its stability but overlapping a neighboring region. Only two structures were found by program DBL for the molecule:

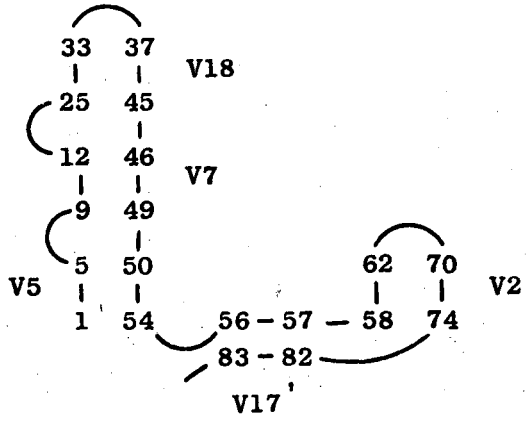
I



FREE ENERGY

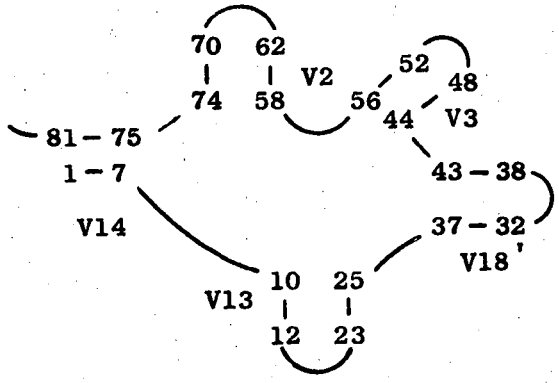
-33.9

II



-28.4

The cloverleaf structure, which is almost as stable as structure I, is shown below. It was not found by the calculation.



FREE ENERGY

-31.1

Consideration of the sequence of the molecule shows that the base pairing region designated V18' can be extended by two base pairs: base 44 can pair with base 26, forming a G-U base pair and base 45 can pair with base 25, forming a G-C base pair. As a result, V18 and V3 are mutually exclusive, and the cloverleaf structure was not found. Program DBL does find the cloverleaf structure (with V18 absent), but its free energy, -28 kcal/mole, is slightly below the cutoff energy and it is not reported. The relatively high cutoff energy results from the particularly stable form, structure I, which is found when the calculation is performed on the reduced exclusion matrix of 5 x 5 elements.

Two related problems are presented by these results: (1) the failure to find the cloverleaf structure for ser tRNA is disturbing, although the reason for this failure is evident. One solution is to allow two overlaps in the vector exclusion matrix. The difficulty with this approach is that the analysis of results is made considerably more complex. An alternative approach, which we have used in analyzing the results of the 5S RNA calculations, is to note carefully all regions of a given structure which are not double stranded. One then uses the base pairing matrix to determine if any additional double stranded regions can be formed which have been omitted because of multiple overlaps. (Two versions of the base pairing matrix are outputted, as shown in appendix III. The second version, which shows all the vectors and their positions in the matrix, is the proper one for this analysis.) The advantage of this procedure is not only that it considers the viability of doubly

overlapping regions but also that it notes additional double stranded regions which fall outside the range of the reduced exclusion matrix. For example, a vector which contributes -2 kcal/mole might be too far down the list of vectors to be considered in the solution. Using the base pairing matrix to determine if any such vectors exist in regions of the molecule which are not double stranded effectively extends the range of the reduced exclusion matrix. In the few instances for which this procedure has yielded additional base pairing regions in a secondary structure, we call attention to the fact. (2) The second disturbing finding related to ser tRNA is that a secondary structure almost 3 kcal/mole more stable than the cloverleaf structure has been found. This is the only instance in which a calculation has suggested that a structure other than the cloverleaf may be as much as 10% more stable than the cloverleaf at room temperature. It is highly unlikely that structure I is the biologically active form of a tRNA, since its spatial orientation is so different from other tRNA molecules. It should, however, be noted that the bases making up the anti-codon in structure I (bases 34-36) are not involved in base pairs and so could, in theory, bind with a codon. They are found in an interior loop in structure I, rather than in a hairpin loop, as is the case with other tRNA secondary structures.

The similarities and differences between structure I and the cloverleaf secondary structure are of interest. Structure I contains nine fewer base pairs than the cloverleaf structure. It also has one fewer loop. Both secondary structures contain regions V2 and V3. Region V8 in structure I has the same base pairs as

region V13 in the cloverleaf structure, and makes almost the same contribution to the free energy of the molecule. This leaves the following double stranded regions of the two structures responsible for the apparent greater stability of structure I:

I: V1

23 27
GGCGA
CCGCU
84 80

Free energy = -4.9-5.0-4.0-1.6
-1.1+6.0 = -10.6

Cloverleaf:

V14

81 75
CUGUUGA
GGCAACU
1 7

43 38
UUUCPA
AAAGAP
27 32
V18'

Free energy = -1-1-1.6-1.3
-2-1.9-1.1+6 = -3.9

Free energy = -1.3-1.3-1.9-1.6
-1.3-1.1+6.5 = -2.0

V1, with just five base pairs, should be significantly more stable than the sum of V14 and V18'. This is primarily due to the fact that neither V14 nor V18' have any GG, GC, or CG interactions. Unless tertiary structure can significantly alter the free energy our present knowledge indicates that structure I should be the favored conformation.

2. Results for 5S RNA

The results discussed above suggest that the methodology for predicting secondary structure is generally reliable when applied to tRNA molecules, which have between 75 and 85 nucleic acid bases. It apparently failed in one instance, and a procedure to safeguard against this failure was suggested. It is evident that future work on the methodology in concert with additional information regarding

loop free energies is needed before we have full confidence in the method. Nonetheless, the method appears generally reliable at this stage in its development. We now apply it to the problem of 5S RNA, which has an as yet unknown biological function and 50% more nucleic acid bases than tRNA.

a. 5S RNA from E. coli

5S RNA from E. coli has 120 bases. Its sequence is:[7]

UGCCUGGCGGCCUUAGCGCGGUGGUCCCACCUGACCCCAUGCCGAACUCA

10 20 30 40 50

GAAGUGAAACGCCGUAGCGCCGAUGGUAGUGUGGGGUCUCCCCAUGCGAG

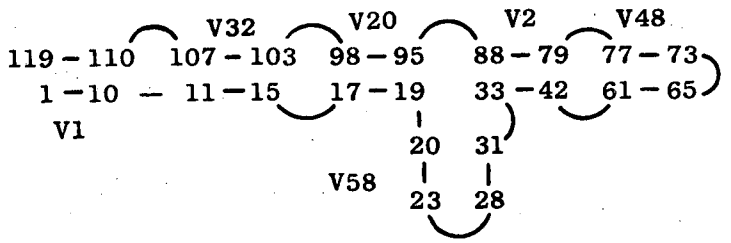
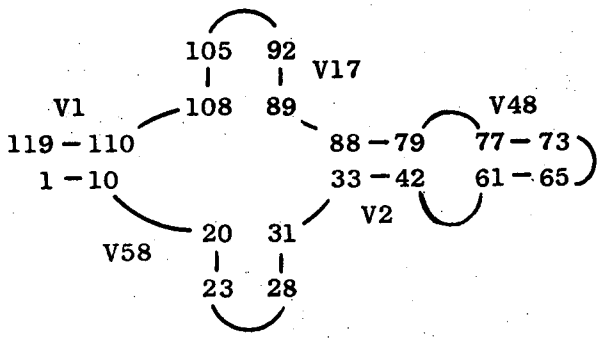
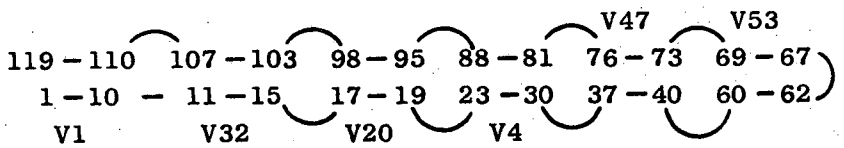
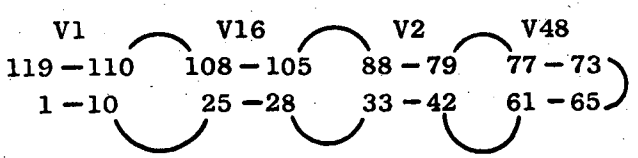
60 70 80 90 100

AGUAGGGAACUGCCAGGCAU

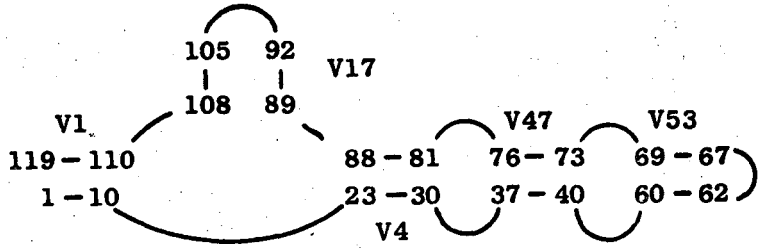
110 120

Although 5S RNA has only about 40 more bases than tRNA, the difficulty of predicting the secondary structure is greatly increased. This is made evident by the fact that program DBL has found 185 base pairing regions with a negative free energy. It was possible to find the most stable secondary structures for a few tRNA molecules by inspection of the sequence of the molecule without resorting to a computer calculation; for 5S RNA, the need to perform a computer calculation is much greater. 68 of the base pairing regions are more stable than -2 kcal/mole, 52 are more stable than -3 kcal/mole, and 39 are more stable than -4 kcal/mole. The use of the reduced exclusion matrix with 15 x 15 elements allows the calculation to consider solutions out to approximately the fiftieth base pairing region. This means that the solution takes account of vectors more stable than -3 kcal/mole. To consider the remaining

vectors, we inspect the base pairing matrix in those regions of the structure in which a string of bases is not base paired, as discussed previously. The resulting secondary structures are:

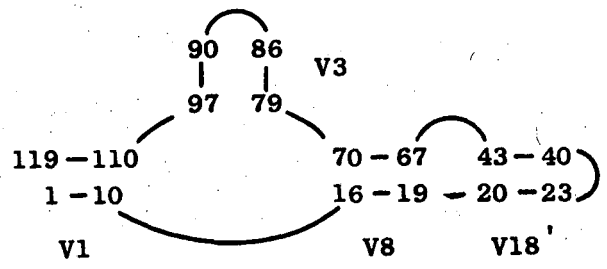
		FREE ENERGY
I		-55.8
II		-53.2
III		-52.5
IV		-50.9

V



-47.1

VI



-46.8

Observations concerning the above results: (1) only a limited number of secondary structures are predicted. This is a necessary condition for the methodology to be useful. We have listed all secondary structures within 10 kcal/mole of structure I. (2) We have scanned the base pairing matrix for vectors not included in the reduced exclusion matrices to determine if any of the structures could be stabilized by added another vector to the structure. The result has been the addition of V58 to structures I and II. This further stabilized the secondary structures by 3 kcal/mole. All the predicted secondary structures have been analyzed in this manner. (3) Because the free energies of structures I, II, and III differ by so little, it is difficult to choose among them. Structure II looks much like a cloverleaf model; structure III is a fully extended model with only one hairpin loop. It will be useful to keep these

three skeletal forms in mind as we analyze other 5S RNA molecules.

b. 5S RNA from Pseudomonas fluorescens

The sequence of this molecule is:[8]

UGUUCUUUGACGAGUAGUAGCAUUGGAACACCUGAUCCCAUCCCGAACUC

10 20 30 40 50

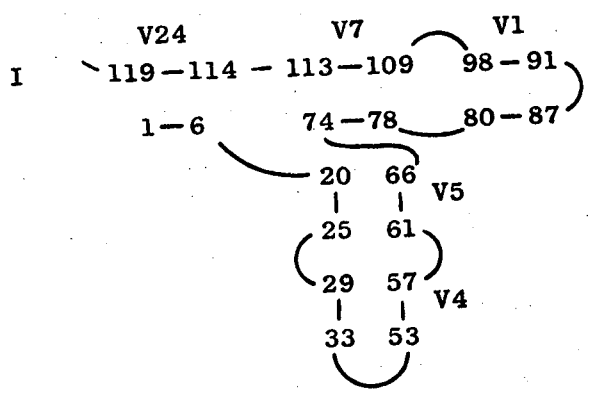
AGAGGUGAAACGAUGCAUCGCCGAUGGUAGUGUGGGGUUUCCCCAUGUCA

60 70 80 90 100

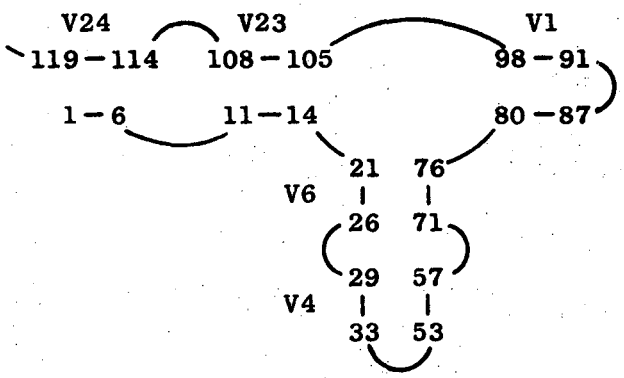
AGAUCUCGACCAUAGAGCAU

110 120

Program DBL found 72 base pairing regions with a negative free energy; 28 were more stable than -2 kcal/mole, 20 were more stable than -3 kcal/mole, and 10 were more stable than -4 kcal/mole. Because of the limited number of base pairing regions, the program was able to consider almost all of them with a negative free energy. The most stable secondary structures are:

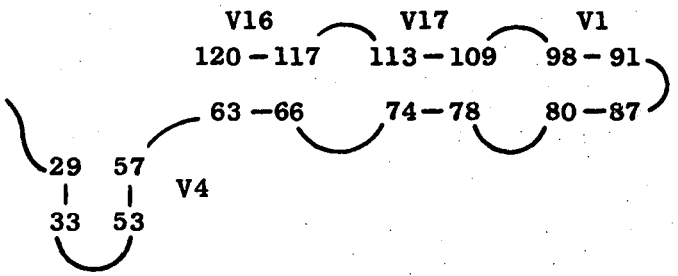


II



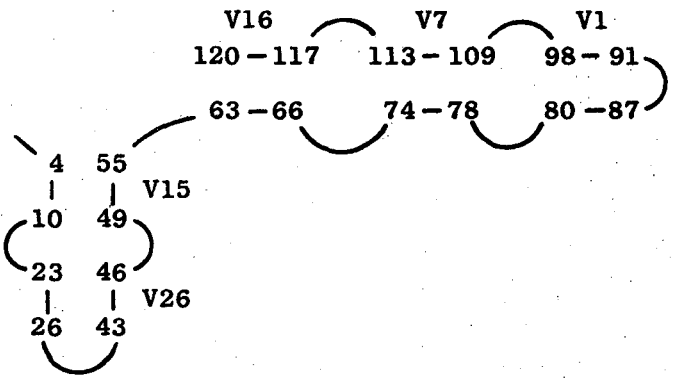
-29.2

III



-27.4

IV



-27.1

There are several other secondary structures reported which are within 10% of the stability of structure I. All of these structures are similar to one of the four most stable forms; they have the same skeletal arrangement with the substitution of a less stable base pairing region for a more stable one. We have for this reason omitted listing them.

Comparing structures I and II with the results for 5S RNA from E. coli, we note that the branching of the base pairing regions is significantly different for the two molecules. The main feature that the secondary structures have in common is that the two ends of the molecule are base paired to one another. Structures III and IV for P. fluorescens 5S RNA are different from the previous secondary structures, in that the ends of the molecule are not bound to each other.

c. 5S RNA from K. B. Cells

The sequence of this molecule is:[9]

GUCUACGGCCAUACCACCCUGAACGCGCCCGAUCUCGUCUGAUCUCGGAA

10 20 30 40 50

GCUAAGCAGGGUCGGGCCUGGUUAGUACUUGGAUGGGAGACCGCCUGGGA

60 70 80 90 100

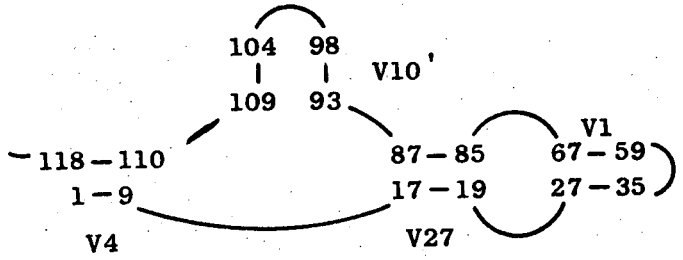
AUACCGGGUGCUGUAGGCUUU

110 120

A total of 157 base pairing regions were found; 63 were more stable than -2 kcal/mole, 48 were more stable than -3 kcal/mole, and 32 were more stable than -4 kcal/mole. The favored secondary structures are:

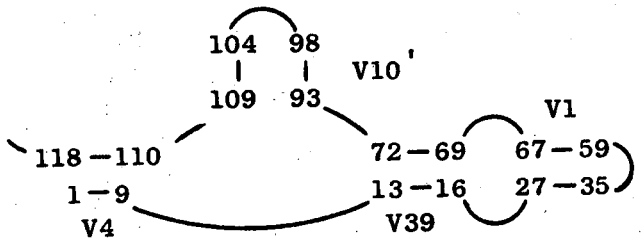
FREE ENERGY

I



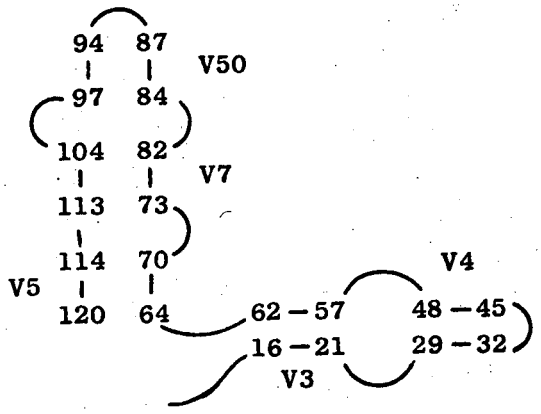
-41.4

II



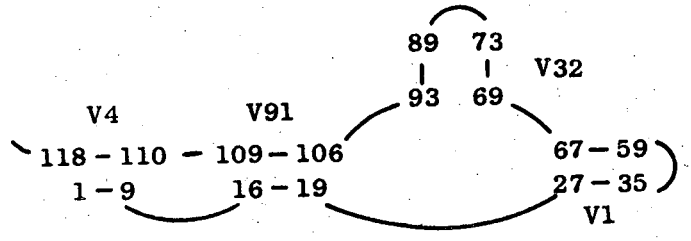
-40.4

III



-40.1

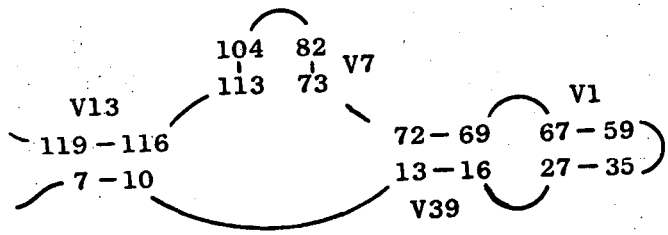
IV



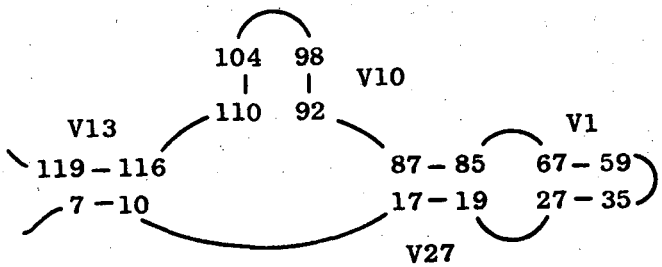
-39.9

FREE ENERGY

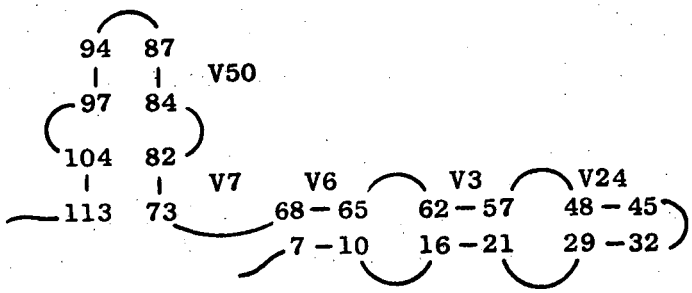
V -38.5



VI -38.1



VII -38.1



Here we have chosen -38 kcal/mole as the cutoff for the free energy. If the cutoff had been chosen at -37 kcal/mole, about an equal number of additional structures would have been included. Most of these structures are very similar to one of the seven structures shown, with a less stable base pairing region replacing a more stable one. Because of the large number of secondary structures of almost equal energy, it is very difficult

in this case to unambiguously predict a secondary structure for the molecule. The solutions do appear to group themselves into three different skeletal arrangements: structures I, II, V, and VI show great similarity, with just one branching hairpin loop following the base pairing region which unites the two ends of the molecule. Structure IV is not greatly different from these structures, with its branching hairpin loop further from the ends of the molecule. Structures III and VII are similar to each other in their skeletal arrangements, with the ends of the RNA not bonded together.

d. 5S RNA from yeast.

The sequence of this RNA is:[10]

GGUUGCGGCCAUACCAUCUAGAAAGCACCGUUCUCCGUCCGAUAACCUGU

10 20 30 40 50

AGUUAAGCUGGUAAGAGCCUGACCGAGUAGUGUAGUGGGUGACCAUACGC

60 70 80 90 100

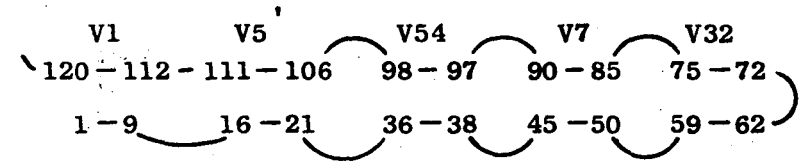
GAAACCUAGGUGCUGCAAUCU

110 120

95 base pairing regions were found. 47 are more stable than -2 kcal/mole; 27, more stable than -3 kcal/mole; 15, more stable than -4 kcal/mole. The results of the calculation are:

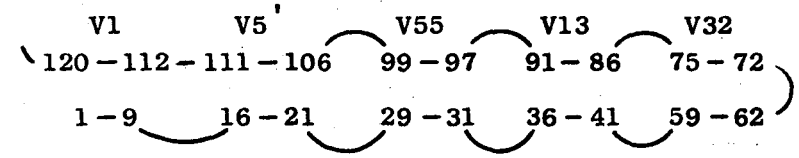
FREE ENERGY

I



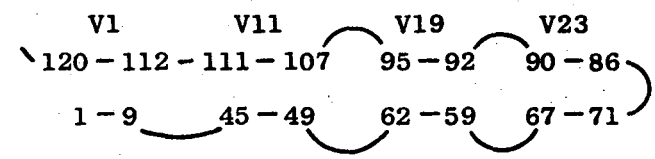
-25.5

II



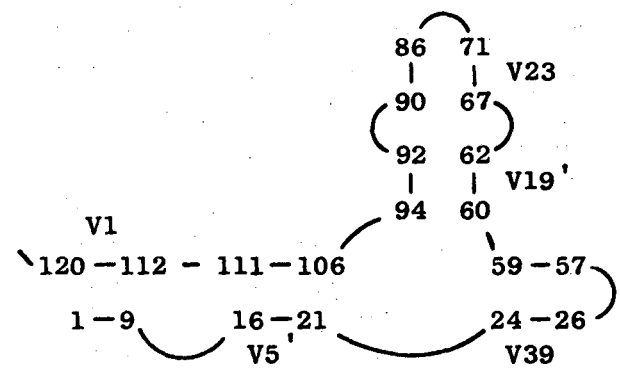
-24.6

III



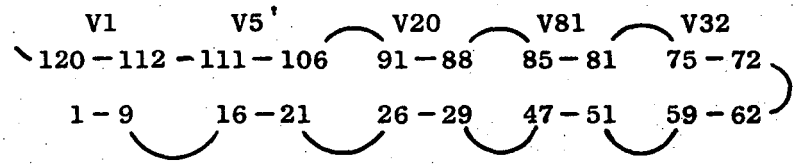
-24.4

IV



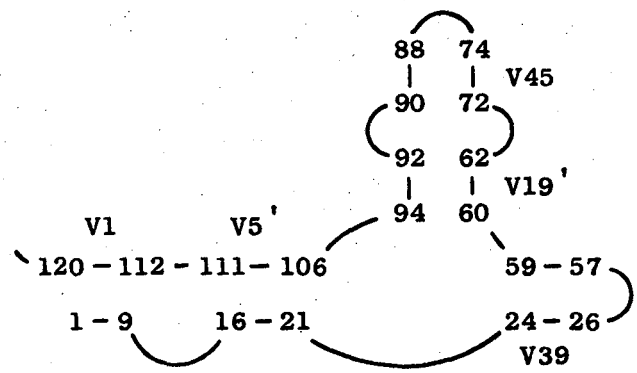
-23.9

V



-23.6

VI



-22.9

We list all structures within 10% of the free energy of structure I. There is one additional structure, equal in energy with and almost identical to structure I, with bases 29 to 31 pairing with 99 to 97 in place of bases 36 to 38. The only difference between this and structure I is in the size of the interior loops. Structures I, II, III, and V have the same skeletal arrangement, being extended RNA structures with only slight differences between them in base pairing regions. Structures IV and VI are similar to each other in their skeletal arrangement, differing in one double stranded region. On the basis of the results of this calculation, it would appear that an extended structure like structure I is the favored secondary structure. It is of interest to note that this 5S RNA has fewer base pairs and is less stable than the others we have considered.

e. 5S RNA from human cells.

The sequence of this RNA is:[11]

GUCUACGGCCAUACCACCCUGAACGCGCCCGAUCUCGUCUGAUCUCGGAA

10 20 30 40 50

GCUAAGCAGGGUCGGGCCUGGUUAGUACUUGGAUGGGAGACCGCCUGGGA

60 70 80 90 100

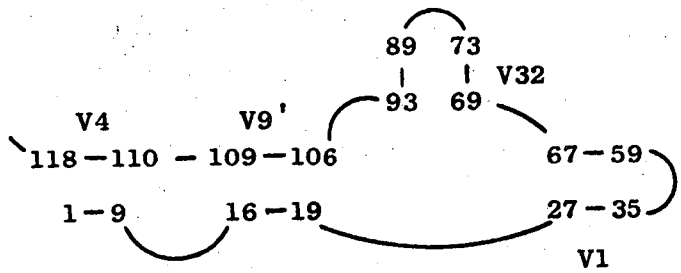
AUACCGGGUGCUGUAGGCUUU

110 120

157 base pairing regions were found: 63 more stable than -2 kcal/mole, 48 more stable than -3 kcal/mole, and 32 more stable than -4 kcal/mole. We report all structures within 10% of the free energy of the most stable structure.

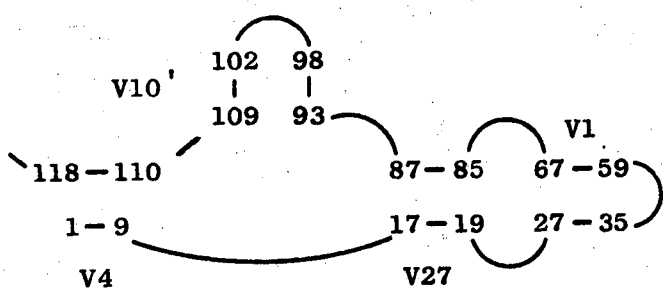
FREE ENERGY

I



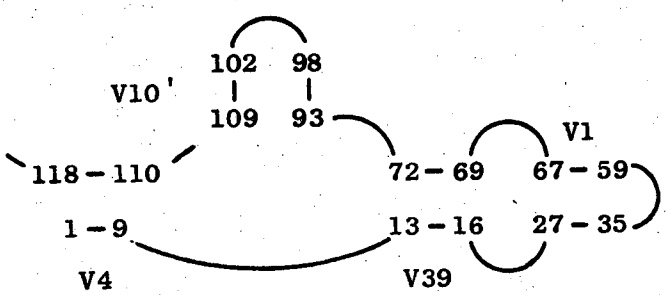
-40.9

II



-40.9

III



-40.8

The next most stable structure is more than 4 kcal/mole less stable than any of the above three secondary structures, all of which have approximately the same free energy.

C. Discussion

1. A Common Model for 5S RNA?

As for tRNA molecules, it is reasonable to expect 5S RNA from different organisms to have similar secondary structures. It is thus of interest to note similarities and differences among the predicted secondary structures of 5S RNA. Two of the 5S RNA's have a fully extended secondary structure with no branching double stranded regions as one or more of the solution structures. The structures in which the two ends of the molecule are not united (for KB, III and VII; for PF, III and IV) can be represented as a variant of the fully extended structure (rather than in the shape of an L, as shown). Only one 5S RNA has a cloverleaf type structure in solution set, with four stem regions extending from an interior loop. The most common structural form, which is represented in the solution sets for all the 5S RNA's, is an extended structure with one stem region branching off from an interior loop. These structures differ from one another in the position of the branching region. For E. Coli. 5S RNA, the branching region has one hairpin loop and occurs after the second double stranded region from the end. For PF, structure I branches after the first double stranded region and has one hairpin and two interior loops; structure II branches after the second double stranded region. Similarly, for KB, yeast, and human 5S RNA, different forms of this general type of secondary structure appear in the solution set.

The basic colculsion is that no one pattern of secondary structures emerges from our analysis, although some structural similarities do appear among the 5S RNA's we have studied.

2. Comparison with other Results

A number of models for 5S RNA from KB and E. Coli. have been proposed.[9, 12-15] Of those models which propose a secondary structure for the entire molecule, the most stable are the models of Cantor[12] and Lewis and Doty[9] for E. Coli. 5S RNA. In both of these models, bases in loops are permitted to base pair and the strands are twisted in and out of one another in a very complicated fashion. For our model calculations, we have excluded such structures from consideration. The Lewis and Doty model has the strongest experimental support, as the authors performed oligonucleotide binding studies to determine which regions of the molecule were single stranded. It is difficult to evaluate the free energy of this structure because of uncertainty in our understanding of the loop free energies in a twisted three dimensional arrangement; we estimate it to be about -53 kcal/mole, somewhat less stable than the calculated free energy of structure I. The binding data of Lewis and Doty are inconsistent with structure I, as they indicate that bases 9 - 13, 25 - 32, 58 - 65, and possibly 95 - 98 are in single stranded regions of the molecule. Of the secondary structures for E. Coli. 5S RNA, only structure VI is consistent with these data.

On the basis of experiments in which 5S RNA from E. Coli. was digested with several ribonucleases and the sequence of the fragments determined, Jordan concluded that regions of the molecule composed of bases 40 to 45, 21, and 61 are single stranded, regions 22 to 44 and 45 to 61 are self-paired, and region 1 - 10 is paired with region 110 - 120.[15] Of the six solution structures reported for E. Coli. 5S RNA, only structure VI is consistent with these results.

This comparison with experimental work suggests a direction for future work in extending and improving these model calculations. The calculations could be combined with the results of oligonucleotide binding, digestion, and other physical studies which yield pieces of information about the secondary structure of the molecule. The calculation could be performed by deleting those regions from the sequence which are known not to base pair.

Further refinements in the calculation await more and better information about loop stabilities.

CHAPTER 5

REFERENCES

1. A. Adams, T. Lindahl, and J. Fresco, Proc. Nat. Acad. Sci. U.S., 57, 1685 (1969).
2. J. Lewis and P. Doty, Nature, 225, 510 (1970).
3. K. Beardsley and C. R. Cantor, Proc. Natl. Acad. Sci. U.S., 65, 39 (1970).
4. J. Crawford, S. Chan, and M. Schweizer, Biochem. Biophys. Res. Comm., 44, 1 (1971).
5. I. Tinoco, jr., P. N. Borer, B. Dengler, M.D. Levine, O. C. Uhlenbeck, D. M. Crothers, and J. Gralla, Nature New Biology, 246, (1973).
6. J. Gralla and D. M. Crothers, J. Mol. Biol., 78, 301 (1973).
7. G. G. Brownlee, F. Sanger, and B. G. Barrell, Nature, 215, 735 (1967).
8. B. Dubruy and S. M. Weissman, J. Biol. Chem., 246, 747 (1971).
9. J. T. Madison, Ann. Rev. Biochem., 37, 131 (1968).
10. J. Hindley and S. M. Page, FEBS Letters, 26, 157 (1972).
11. M. J. Avernal and N. R. Pace, J. Biol. Chem., 249, 4491 (1972).
12. C. R. Cantor, Nature, 216, 513 (1967).
13. H. Boedtker and D. G. Kelling, Biochem. Biophys. Res. Comm., 29, 758 (1967).
14. I. D. Raake, Biochem. Biophys. Res. Comm., 31, 528 (1968).
15. B. R. Jordan, J. Mol. Biol., 55, 423 (1971).

APPENDIX I

PROGRAM OLIGO

Program OLIGO is used in Chapters 2 and 3 to calculate melting curves for double stranded RNA oligomers for different values of ΔH°_{IJ}, k_{IJ}, and κ. The program is written so that results may be outputted onto a teletype machine so that new trial values of the parameters can be inputted as soon as results are reported. More complete output information may be obtained by specifying a complete computer printout. The program sums over all intermediate states shown in Fig. 2.1. Since the logic of program OLIGO is straightforward, we explain only the inputting procedure and the use of the program.

A. Input Data Which is Read into the Program from Cards

Card Number	Symbol(s)	Format	Explanation
1	NUMMCL, ITERMX, DELTA	2I5,F10.5	NUMMCL = the number of molecules; ITERMX = the number of temperatures for which melting curve is calculated; DELTA = temperature increment
2	NCONC, (CT(I), I=1,NCONC)	I5, 7F10.7	NCONC = number of concentrations for which input data is given (usually 3); CT(I) = strand concentration in moles/liter
3	TT(I,I), I=1, NUMMCL	8F10.6	TT(K,I) = melting temperature of I th molecule at K th concentration
4	SLP(I), I=1 NUMMCL	8F10.4	SLP(I) = slope of melting curve at T_m of I th RNA

Card Number	Symbol(s)	Format	Explanation
5	DTMDC(I), I=1 NUMMCL	8F10.5	DTMDC(I) =[$\Delta(1/T_m)$/ $\Delta \log c]\times 10^4$ for I th RNA
6	HIN, HGG, HGC, HCG, HGA, HAG	8F10.6	HIN = ΔH°_{AA}; all other HIJ (where I and J are G, C, A, and U) = ΔH°_{IJ}- ΔH°_{AA}. All enthalpy terms in cal/mole
7	HGU, HUG, HAA, HAU, HUA	8F10.6	HIJ = $\Delta H^\circ_{IJ}-\Delta H^\circ_{AA}$
8	EKGGO, EKGCO, EDCGO, EKGAO, EKAGO	8F10.6	EKIJO = k_{IJ} (78°C), as defined in the text in equation 3.2a
9	EKGUO, EKUGO EKAAO, EKAUO, EKUAO	8F10.6	As above, card 8
10	SAME, STAGG, LOOP, MXIN	2A5, 2I5	If SAME = 5HbSAME, then the two complementary strands in the RNA are identical; the variables STAGG, LOOP, and MXIN are not used in the calculation and may be set to anything
11	N, (WD(I), I= 1,N)	A5, 40A1	N = the number of bases in the molecule; WD(I) = the identity of the I th base; the program assumes that second strand is complementary to the first.
12....			Cards 10 and 11 are repeated for each molecule in the series

B. Input Data Which is Read into Program from Teletype

Datum Number	Symbol(s)	Explanation
1	CLASS	CLASS = 1 is the standard option; for A_nU_n, where $df/dT\|_m$ is not known for A_6U_6, CLASS = 2; for A_nCGU_n, where $df/dT\|_m$ is not known for A_3CGU_3, CLASS = 3
2	OUTPUT	OUTPUT = parameter which controls output to teletype; standard option is OUTPUT = 3, which supplies T_m, $df/dT\|_m$, and $\Delta(1/T_m)/\Delta(\log c)$ as well as agreement of these quantities with experiment for each molecule (D1, D2, D3, D4, D5, D6, as defined; OUTPUT=2 and OUTPUT=1 give less information.
3	HIN	HIN = ΔH°_{AA}
4	DEL	DEL = k_{end}
5	BETA(1)	BETA(1) = κ_A or κ_C, whichever is appropriate for the series of molecules being calculated
6....	EKIJO, HIJ	The remaining cards are varied, depending on which unknown parameters are to be solved for. These cards overwrite the initial estimated of k_{IJ} (78°C) and ΔH°_{IJ} which were read in on cards 6-9.

C. Output to Teletype

Shown below is a typical output (to the teletype) of program OLIGO. It gives the calculated melting temperatures, slopes of the melting curve at T_m, and concentration dependence of the T_m's for each molecule in the series. Also outputted, and of great importance for carrying out the calculation, are the quantities D1, ..., D6. These quantities indicate the degree to which the calculation is in agreement with experiment.

```
3.!
ML
 HIN          =
-8750.!
ML
 DEL          =
.0585!
ML
 BETA(1)      =
.0163!
ML
 EKGAO        =
3.162!
ML
 HGA          =
3000.!
ML
 EKGGO        =
32.5!
ML
 HGG          =
-11000.!
ML
 TM       =   25.17036
 TM       =   16.30747
 DTV      =    1.02636
 SLOPE    =     .02841
 TM       =    8.09308
 DTV      =    1.00897
 TM       =   41.41036
 TM       =   33.60206
 DTV      =     .80917
 SLOPE    =     .02991
 TM       =   26.27793
 DTV      =     .79734
 TM       =   31.02203
 TM       =   44.36144
 DTV      =     .64717
 SLOPE    =     .03500
 TM       =   37.99993
 DTV      =     .64687
 D1       =    2.46119
 D2       =     .05370
 D3       =   12.25437
 D4       =   15.17386
 D5       =     .13157
 D6       =    -.00288
 BETA1    =   16.30000
 CONST    =   53.03074
 OUTPUT       =
```

$$D1 = (T_m(\text{expt.}) - T(\text{calcd.}))^2$$

$$D2 = (T_m(\text{expt.}) - T_m(\text{calcd.}))$$

$$D3 = (df/dT)_m(\text{expt.}) \times 10^3 - df/dT|_m(\text{calcd.}) \times 10^3)^2$$

$$D4 = (df/dT)_m(\text{expt.}) \times 10^3 - df/dT|_m(\text{calcd.}) \times 10^3)$$

$$D5 = (\Delta(1/T_m)/\Delta(\log c)(\text{expt.}) \times 10^4 - \Delta(1/T_m)/\Delta(\log c)(\text{calcd.}) \times 10^4)^2$$

$$D6 = (\Delta(1/T_m)/\Delta(\log c)(\text{expt.}) \times 10^4 - \Delta(1/T_m)/\Delta(\log c)(\text{calcd.}) \times 10^4)^2$$

D2, D4, and D6 indicate whether k_{IJ}, k_{end}, and ΔH°_{IJ} should be increased or decreased for the next trial. D1, D3, and D5 indicate the degree to which calculation and experiment are in agreement.

D. Listing of Program OLIGO

```
      PROGRAM OLIGO(INPUT,OUTPUT,TAPE5=INPUT,TAPETTY=12,TAPE20=TAPETTY,
     1TAPE6)
      COMMON/A/BETA(5),WD1(18),DELT(85)
      COMMON/B/Q1(65),Q2(65),Q1NTOT(65),Q2NTOT(65),F1(65),F2(65)
      COMMON/C/TM,NDEL17(15),NDEL50(15),NDEL83(15),TMLST(40),LM
      COMMON/D/D1(150),D2(150),D3(150),TT(4,10),IC,MOLECL,NCOUNT,SLP(10)
     1,D4(50),D5(50),D6(50),DTMDC(10)
      COMMON/E/SS(90),EKGG(90),EKGC(90),EKCG(90),EKGA(90),EKAG(90),EKGU(
     190),EKUG(90),EKAA(90),EKAU(90),EKUA(90)
      COMMON/F/ITM,TVLST,NHERE,ICLASS,OUTPT
      COMMON/HT/HH(65),EPY(65),UU(65)
      DIMENSION BETAO(3)
      DIMENSION NREM(5)
      DIMENSION BASE(18)
      DIMENSION QN(4,65)
      DIMENSION CT(10)
      DIMENSION B(3,65)
      DIMENSION W(5,20)
      CALL TTYSET
      NHERE=0
      HGA=0.
      HGG=0.
      EKGGO=1.
      TVLST=0.
      TMGC=426.16
      TMAU=351.16
      TMAUV=1./351.16
      TOV=TMAUV
      TBETA=25.+273.16
      TBV=1./TBETA
      RINV=1./1.986
      READ(5,40)  HIN,HGG,HGC,HCG,HGA,HAG
      READ(5,40) HGU,HUG,HAA,HAU,HUA
      READ(5,40)  EKGGO,EKGCO,EKCGO,EKGAO,EKAGO
      READ(5,40)  EKGUO,EKUGO,EKAAO,EKAUO,EKUAO
40    FORMAT(8F10.6)
      WRITE(6,42)  HIN,HGG,HGC,HCG,HGA,HAG
42    FORMAT(*    HIN=*,F8.2,*     HGG=*,F8.2,*     HGC=*,F8.2,*     HCG=*,
     1F8.2,*     HGA=*,F8.2,*     HAG=*,F8.2)
      WRITE(6,43)  HGU,HUG,HAA,HAU,HUA
43    FORMAT(*    HGU=*,F8.2,*     HUG=*,F8.2,*     HAA=*,F8.2,*     HAU=*,
     1F8.2,*     HUA=*,F8.2,//)
      WRITE(6,44)  EKGGO,EKGCO,EKCGO,EKGAO,EKAGO
44    FORMAT(*    EKGGO=*,F8.3,*     EKGCO=*,F8.3,*     EKCGO=*,F8.3,*     E
     1KGAO=*,F8.3,*     EKAGO=*,F8.3)
      WRITE(6,45)  EKGUO,EKUGO,EKAAO,EKAUO,EKUAO
45    FORMAT(*    EKGUO=*,F8.3,*     EKUGO=*,F8.3,*     EKAAO=*,F8.3,*     E
     1KAUO=*,F8.3,*     EKUAO=*,F8.3,//)
      READ(5,61)  NUMMCL,ITERMX,DELTA,FGU
61    FORMAT(2I5,2F10.5)
      READ(5,68)  NCONC,(CT(I),I=1,NCONC)
68    FORMAT(I5,7F10.7)
      DO 66 K=1,NCONC
```

```
      READ(5,67)  (TT(K,I),I=1,NUMMCL)
 67   FORMAT(8F10.6)
 66   CONTINUE
      READ(5,46)  (SLP(I),I=1,NUMMCL)
 46   FORMAT(8F10.6)
      READ(5,47)   (DTMDC(I),I=1,NUMMCL)
 47   FORMAT(8F10.5)
      TINIT0=TINIT
      DELTA0=DELTA
      ITMX0=ITERMX
      CALL TREAD(5HCLASS,CLASS,0)
      ICLASS=CLASS
      CALL TREAD(6HHMODEL,HMODEL,0)
      MODELH=HMODEL
      DO 59 JJTTY=1,250
      DO 15 I=1,25
      D1(I)=D2(I)=D3(I)=D4(I)=D5(I)=D6(I)=0.
 15   CONTINUE
      CALL TREAD(6HOUTPUT,FOUT,0)
      OUTPT=FOUT
      CALL TREAD(3HHIN,HIN,0)
      CV=0.
C     CALL TREAD(2HCV,CV,0)
      CALL TREAD(3HDEL,DEL,0)
      CALL TREAD(7HBETA(1),BETA(1),0)
      BETA(3)=BETA(2)=BETA(1)
      CALL TREAD(5HEKUA0,EKUA0,0)
      CALL TREAD(3HHUA,HUA,0)
 1005 CONTINUE
      WRITE(6,1006)   DEL,BFACT1,BFACT2,BETA0(1),BETA0(2),HIN,MODELH,CV
 1006 FORMAT(///,*  THE PARAMETERS FOR THIS CALCULATION ARE LISTED*,/,
     1*     DEL *,F10.6,/,*     BFACT1*,F10.6,*       BFACT2*,F10.6,/,*
     2BETA0(1)*,F14.8,*       BETA0(2)*,F14.8,/,*     ENTHALPY=*,F10.3,/,
     3*    0 = NO H(T)    1 = H(T)=CV (T-T0)    2= 2 STATE MODEL    MODELH=*
     4,I2,/,*    DH/DT=*,F10.5)
      WRITE(6,1007)  ICLASS,EKGA0,EKGG0,HGA,HGG
 1007 FORMAT(*    CLASS OF OLIGOMERS=*,I3,*    EKGA0=*,F12.6,*    EKGG0=*,
     1F12.6,*    HGA=*,F10.3,*,    HGG=*,F10.3)
      DO 60  MOLECL=1,NUMMCL
      DO 62  IC=1,NCONC
      NCOUNT=0
      CONC=CT(IC)
      WRITE(6,63)  CONC
 63   FORMAT(1H1,*  THE TOTAL STRAND CONCENTRATION IS*,E16.6,///)
      HB=0.
      NCOUNT=NCOUNT+1
      DO 65  ITM=1,2
      LM=IH+ID+IB+IBH-3
      KM=IC+ITM-1
      ITERMX=ITMX0
      DELTA=DELTA0
      TINIT=TT(IC,MOLECL)-ITERMX*DELTA/2.
      IF(ITM.EQ.1)  GO TO 69
      TINIT=TM-0.6
      DELTA=0.1
      ITERMX=11
 69   IF(KM.NE.1)  GO TO 100
      IF(JJTTY.GT.1)  GO TO 99
      READ(5,90)  SAME,STAGG,LOOP,MXIN
 90   FORMAT(2A5,2I5)
      WRITE(6,95) STAGG,SAME,CONC,LOOP,MXIN
 95   FORMAT(*  STAGG=*,A5,*  SAME=*,A5,*   CONC=*,E10.3,*  LOOP=*,I4,
     1*  MXIN=*,I5)
      IDENT=1
      IF(SAME.EQ.5H SAME)  IDENT=2
      NDEBUG=0
```

```
      NSTAG=0
      IF(STAGG.EQ.5HSTAGG)  NSTAG=1
      READ(5,110)  (N,(WD1(I),I=1,N))
  110 FORMAT(I5,40A1)
      NREM(MOLECL)=N
      DO 98 I=1,N
   98 W(MOLECL,I)=WD1(I)
      GO TO 100
   99 N=NREM(MOLECL)
      DO 97  I=1,N
   97 WD1(I)=W(MOLECL,I)
  100 WRITE(6,115)  (WD1(I),I=1,N)
  115 FORMAT(*
     1           *,20A1)
      T=TINIT
      DO 116  I=1,N
  116 BASE(I)=WD1(I)
      DO 120  ITER=1,ITERMX
      H=HIN
      T=T+DELTA
      TINV=1./(T+273.16)
      DELT(ITER)=T
      RTT=RINV*(TINV-TMAUV)
      B(1,ITER)=BETA(1)*EXP(HB*RINV*(TBV-TINV))
      B(2,ITER)=BETA(2)*EXP(HB*RINV*(TBV-TINV))
      B(3,ITER)=BETA(3)*EXP(HB*RINV*(TBV-TINV))
      SS(ITER)=EXP(-HIN*RTT)
      UU(ITER)=SS(ITER)
      EKGG(ITER)=EKGGO*EXP(-HGG*RTT)
      EKGC(ITER)=EKGCO*EXP(-HGC*RTT)
      EKCG(ITER)=EKCGO*EXP(-HCG*RTT)
      EKGA(ITER)=EKGAO*EXP(-HGA*RTT)
      EKAG(ITER)=EKAGO*EXP(-HAG*RTT)
      EKGU(ITER)=EKGUO*EXP(-HGU*RTT)
      EKUG(ITER)=EKUGO*EXP(-HUG*RTT)
      EKAA(ITER)=EKAAO*EXP(-HAA*RTT)
      EKAU(ITER)=EKAUO*EXP(-HAU*RTT)
      EKUA(ITER)=EKUAO*EXP(-HUA*RTT)
      NIT=NNIT=NNNIT=0
      IF(MODELH.NE.1)  GO TO 123
C     THIS IS LINEAR VARIATION OF H WITH T
      HH(ITER)=HIN+CV*(T-T0)
      EPYTM=(HIN+CV*(TMAU-T0-273.16))*TMAUV
      EPY(ITER)=CV*ALOG((T+273.16)*TMAUV) + EPYTM
      SS(ITER)=EXP(-HH(ITER)*RINV*TINV+EPY(ITER)*RINV)
  123 Q1(ITER)=Q1NTOT(ITER)=Q2(ITER)=Q2NTOT(ITER)=0.
      DO 121  INT=1,4
      QN(INT,ITER)=0.
  121 CONTINUE
  120 CONTINUE
      DO 130  JJ=1,N
      NTOT=JJ-1
      DO 140  KK=JJ,N
      ND=NGG=NGC=NCG=NGA=NAG=NGU=NUG=NAA=NAU=NUA=LGU=0
      DO 150  LL=1,JJ
      MM=KK-LL+1
      L=MM
      LASTG=LG
      LASTC=LC
      LASTA=LA
      LASTU=LU
      LA=LU=LG=LC=0
      IF(MM.EQ.1.OR.MM.EQ.N)  ND=ND+1
  227 IF(BASE(L).EQ.1HA)  LA=1
      IF(BASE(L).EQ.1HU) LU=1
      IF(BASE(L).EQ.1HG) LG=1
```

```
      IF(BASE(L).EQ.1HC) LC=1
      IF(LL.EQ.1)  GO TO 150
      NGG=NGG+LASTG*LG+LASTC*LC
      NGC=NGC+LASTC*LG
      NCG=NCG+LASTG*LC
      NGA=NGA+LASTA*LG+LASTC*LU
      NAG=NAG+LASTG*LA+LASTU*LC
      NGU=NGU+LASTU*LG+LASTC*LA
      NUG=NUG+LASTG*LU+LASTA*LC
      NAA=NAA+LASTA*LA+LASTU*LU
      NAU=NAU+LASTU*LA
      NUA=NUA+LASTA*LU
  150 CONTINUE
      INT=N-JJ+1
      NG=2*NGG+NGC+NCG+NGA+NAG+NGU+NUG
      IGC=1
      IF(NG.EQ.2)  IGC=2
      IF(NG.GE.3)  IGC=3
      NTT=NG-NGG+NAA+NUA+NAU
      WRITE(6,221)     NGG,NGC,NCG,NGA,NAG,NGU,NUG,NAA,NAU,NUA,ND,IGC,LGU
  221 FORMAT(*   NGG=*,I2,*   NGC=*,I2,*  NCG=*,I2,*  NGA=*,I2,*  NAG=*,
     1I2,*  NGU=*,I2,*  NUG=*,I2,*  NAA=*,I2,*  NAU=*,I2,*  NUA=*,I2,*
     2ND=*,I2,*  IGC=*,I2,*  LGU=*,I2,//)
      IF(NTT.NE.NTOT)  CALL STOP
      DO 160  ITER=1,ITERMX
      IF(LOOP.GT.0)  GO TO 200
      BT=B(IGC,ITER)
      QF2=(EKGG(ITER)**NGG)*(EKGC(ITER)**NGC)*(EKCG(ITER)**NCG)*(EKGA(IT
     1ER)**NGA)*(EKAG(ITER)**NAG)*(EKGU(ITER)**NGU)*(EKUG(ITER)**NUG)*(E
     2KAA(ITER)**NAA)*(EKAU(ITER)**NAU)*(EKUA(ITER)**NUA)
      QF=QF2*(SS(ITER)**NTOT)*(DEL**ND)*BT
      Q1(ITER)=Q1(ITER)+QF
      Q1NTOT(ITER)=Q1NTOT(ITER)+NTOT*QF
      IF(INT.GT.4)  GO TO 160
      QN(INT,ITER)=QN(INT,ITER)+QF
  160 CONTINUE
      GO TO 140
  200 CONTINUE
C     LEFT OPEN FOR LOOPS
  140 CONTINUE
  130 CONTINUE
  132 VT=NTOT
      VNINV=1./VT
      IF(NTOT.EQ.0)  VNINV=0.
      DO 180  ITER=1,ITERMX
      IF(LOOP.NE.0)  GO TO 250
      BT=B(IGC,ITER)
      QQ=(SS(ITER)**NTOT)*(EKGA(ITER)**NGA)*(EKGG(ITER)**NGG)*BT*(DEL**N
     1D)
      GAMMA=QQ*CONC*IDENT
      AITA=(2.*GAMMA+1.-SQRT(4.*GAMMA+1.))/(2.*GAMMA)
      F2(ITER)=1.-AITA
C     END OF ALL OR NONE
      Q1INV=1./Q1(ITER)
      GAMMA=Q1(ITER)*CONC*IDENT
      AITA=(2.*GAMMA+1.-SQRT(4.*GAMMA+1.))/(2.*GAMMA)
      F1(ITER)=1.-AITA*Q1NTOT(ITER)*Q1INV*VNINV
      AITAM=1.-AITA
      DO 185  J=1,4
      QN(J,ITER)=QN(J,ITER)*AITA/Q1(ITER)
  185 CONTINUE
C     IF(KM.GT.1)  GO TO 252
      WRITE(6,190)  SS(ITER),DELT(ITER),F1(ITER),F2(ITER),AITAM,
     1EKGA(ITER),EKGG(ITER),QN(1,ITER),QN(2,ITER),QN(3,ITER),QN(4,ITER)
  190 FORMAT(F10.4,F10.3,3F11.7,2F8.4,4F11.7)
  252 IF (F1(ITER).GT.(0.11).AND.F1(ITER).LT.(0.23)) GO TO 280
```

```
 260  IF (F1(ITER).GT.(0.37).AND.F1(ITER).LT.(0.63))  GO TO 290
 270  IF (F1(ITER).GT.(0.77).AND.F1(ITER).LT.(0.89)) GO TO 300
      GO TO 180
 280  NIT=NIT+1
      NDEL17(NIT)=ITER
      GO TO 260
 290  NNNIT=NNNIT+1
      NDEL50(NNNIT)=ITER
      GO TO 270
 300  NNIT=NNIT+1
      NDEL83(NNIT)=ITER
      GO TO 180
 250  CONTINUE
C     THE SECOND PLACE FOR THE LOOP CARDS WHEN READY
 180  CONTINUE
      CALL EXTRAP(NIT,NNIT,NNNIT)
      IF(ITM.EQ.2)  CALL DFDT
  65  CONTINUE
  74  CONTINUE
  72  CONTINUE
  71  CONTINUE
  70  CONTINUE
  62  CONTINUE
  60  CONTINUE
      DO 330  I=1,NCOUNT
      X=D1(I)
      D1(I)=SQRT(X)
      X=D3(I)
      D3(I)=SQRT(X)
      X=D5(I)
      D5(I)=SQRT(X)
      WRITE(6,345)   (D1(I),D2(I),D3(I),D4(I))
 345  FORMAT(/,4X,*  VARIANCE OF TM=*,2F12.4,*  VARIANCE OF SLOPE=*,2F12
     1.4)
      CALL TWRITE(2HD1,D1(I),0)
      CALL TWRITE(2HD2,D2(I),0)
      CALL TWRITE(2HD3,D3(I),0)
      CALL TWRITE(2HD4,D4(I),0)
      CALL TWRITE(2HD5,D5(I),0)
      CALL TWRITE(2HD6,D6(I),0)
      BETA(1)=1000.*BETA(1)
      BETA(2)=1000.*BETA(2)
      CALL TWRITE(5HBETA1,BETA(1),0)
      CONST=BETA(1)*(EKGA**2)
      CALL TWRITE(5HCONST,CONST,0)
C     CALL TWRITE(5HBETA2,BETA(2),0)
 330  CONTINUE
  59  CONTINUE
      CALL EXIT
      END
      SUBROUTINE EXTRAP (NIT,NNIT,NNNIT)
      COMMON/A/BETA(5),WD1(18),DELT(85)
      COMMON/B/Q1(65),Q2(65),Q1NTOT(65),Q2NTOT(65),F1(65),F2(65)
      COMMON/C/TM,NDEL17(15),NDEL50(15),NDEL83(15),TMLST(40),LM
      COMMON/D/DR(150),D2(150),D3(150),TT(4,10),IC,MOLECL,NCOUNT,SLP(10)
     1,D4(50),D5(50),D6(50),DTMDC(10)
      COMMON/F/ITM,TVLST,NHERE,ICLASS,OUTPT
      SIXINV=1./6.
      FIVESX=5./6.
      ZERO=0.
      HALF=1./2.
      D1SM=D1LG=1.
      IF (NIT.EQ.0) GO TO 30
      DO 20 I=1,NIT
      IN=NDEL17(I)
      D1=SIXINV-F1(IN)
```

```
      IF (D1.GE.ZERO) GO TO 10
C      HERE F1(ITER) IS GREATER THAN 1/6
      D1=-D1
      IF (D1.GT.D1LG) GO TO 20
      D1LG=D1
      IT17L=IN
      GO TO 20
   10 IF (D1.GT.D1SM) GO TO 20
      D1SM=D1
      IT17S=IN
   20 CONTINUE
      N1SM=D1SM
      N1LG=D1LG
      IF (N1SM.EQ.1.OR.N1LG.EQ.1) NIT=0
   30 D1SM=D1LG=1.
      IF (NNIT.EQ.0) GO TO 60
      DO 50 I=1,NNIT
      IN=NDEL83(I)
      D1=FIVESX-F1(IN)
      IF (D1.GE.ZERO) GO TO 40
      D1=-D1
C     HERE F1(ITER) IS GREATER THAT 5/6
      IF (D1.GT.D1LG) GO TO 50
      D1LG=D1
      IT83L=IN
      GO TO 50
   40 IF (D1.GT.D1SM) GO TO 50
      D1SM=D1
      IT83S=IN
   50 CONTINUE
      N1SM=D1SM
      N1LG=D1LG
      IF (N1SM.EQ.1.OR.N1LG.EQ.1) NNIT=0
   60 D1SM=D1LG=1.
      IF (NNNIT.EQ.0) GO TO 90
      DO 80 I=1,NNNIT
      IN=NDEL50(I)
      D1=HALF-F1(IN)
      IF (D1.GE.ZERO) GO TO 70
      D1=-D1
      IF (D1.GT.D1LG) GO TO 80
      D1LG=D1
      IT50L=IN
      GO TO 80
   70 IF (D1.GT.D1SM) GO TO 80
      D1SM=D1
      IT50S=IN
   80 CONTINUE
      N1SM=D1SM
      N1LG=D1LG
      IF (N1SM.EQ.1.OR.N1LG.EQ.1) NNNIT=0
C     EXTRAPOLATION PROCEDURE  (JUST USING TWO POINTS)
   90 IF (NIT.EQ.0) GO TO 100
      SLOPE=(F1(IT17L)-F1(IT17S))/(DELT(IT17L)-DELT(IT17S))
      T17=(SIXINV-F1(IT17S)+SLOPE*DELT(IT17S))/SLOPE
  100 IF (NNIT.EQ.0) GO TO 110
      SLOPE=(F1(IT83L)-F1(IT83S))/(DELT(IT83L)-DELT(IT83S))
      T83=(FIVESX-F1(IT83S)+SLOPE*DELT(IT83S))/SLOPE
  110 IF (NNNIT.EQ.0) GO TO 120
      SLOPE=(F1(IT50L)-F1(IT50S))/(DELT(IT50L)-DELT(IT50S))
      TM=(HALF-F1(IT50S)+SLOPE*DELT(IT50S))/SLOPE
      DEL23=0.
  120 IF (NIT.EQ.0.OR.NNIT.EQ.0) GO TO 130
      DEL23=T83-T17
  130 IF (NNNIT.EQ.0) TM=0.
      IF(NNNIT.EQ.0)  CALL TWRITE(10HNO TM CALC,NNNIT,1)
```

```
      TV=1./(TM+273.16)
      IF(ITM.EQ.1)  GO TO 146
      IF(OUTPT.EQ.3)  CALL TWRITE(2HTM,TM,0)
      NHERE=NHERE+1
      DTV=TV-TVLST
      DTV=DTV*(10.**4)
      TVLST=TV
      IF(NHERE.EQ.1)  GO TO 146
      D5(NCOUNT)=D5(NCOUNT)+(DTMDC(MOLECL)-DTV)**2
      D6(NCOUNT)=D6(NCOUNT)+(DTMDC(MOLECL)-DTV)
      IF(OUTPT.GE.2)  CALL TWRITE(3HDTV,DTV,0)
      IF(NHERE.EQ.3)  NHERE=0
  146 WRITE(6,690)  TM,DEL23,TV
  690 FORMAT(5X,F10.6,7X,F10.6,6X,E12.4)
      RETURN
      END
      SUBROUTINE DFDT
      COMMON/A/BETA(5),WD1(18),DELT(85)
      COMMON/B/Q1(65),Q2(65),Q1NTOT(65),Q2NTOT(65),F1(65),F2(65)
      COMMON/C/TM,NDEL17(15),NDEL50(15),NDEL83(15),TMLST(40),LM
      COMMON/D/D1(150),D2(150),D3(150),TT(4,10),IC,MOLECL,NCOUNT,SLP(10)
     1,D4(50),D5(50),D6(50),DTMDC(10)
      COMMON/F/ITM,TVLST,NHERE,ICLASS,OUTPT
      DIMENSION SLOPE(25)
      SLPTM=0.
      DO 10  I=1,5
      J=11-I
      SLOPE(I)=(F1(I)-0.5)/(DELT(I)-TM)
      SLOPE(J)=(F1(J)-0.5)/(DELT(J)-TM)
      SLPTM=SLPTM+SLOPE(I)+SLOPE(J)
   10 CONTINUE
      SLPTM=SLPTM/10.
      Q=TT(IC,MOLECL)-TM
      DSQ=Q**2
      WRITE(6,20)  SLPTM,DSQ
   20 FORMAT(*  SLOPE(TM)=*,F12.8,*   DSQ=*,F10.3)
      WRITE(6,30)  (SLOPE(I),I=1,11)
   30 FORMAT(/,*  SLOPE=*,11F10.6)
      D1(NCOUNT)=D1(NCOUNT)+DSQ
      D2(NCOUNT)=D2(NCOUNT)+Q
      GO TO (100,200,300),ICLASS
  100 IF(IC.NE.2)   RETURN
      GO TO 400
  200 IF(IC.GT.1)  RETURN
      IF(MOLECL.EQ.3)  GO TO 440
      GO TO 400
  300 IF(IC.NE.2)  RETURN
      IF(MOLECL.EQ.2)  GO TO 440
  400 D3(NCOUNT)=D3(NCOUNT)+((SLP(MOLECL)-SLPTM)*1000.)**2
      D4(NCOUNT)=D4(NCOUNT)+(SLP(MOLECL)-SLPTM)*1000.
  440 IF(OUTPT.GT.1)  CALL TWRITE(5HSLOPE,SLPTM,0)
      RETURN
      END
      SUBROUTINE TTYSET
      DIMENSION IFET(8)
   77 READ(20,100) I
  100 FORMAT(80X,I1)
      IF(I.NE.1) GO TO 66
      DO 10 J=1,10
      CALL CMOFF
   10 CONTINUE
      GO TO 77
   66 WRITE(20,200)
  200 FORMAT( * I AM READY.  PROCEED,O ONE OF ONES*)
      CALL FET(6LTAPE20,IFET,8)
      IFET(2) = IFET(2).OR.00000010000000000008
```

```
      IFET(8) = IFET(8).OR.400000000000000000008
      CALL FET(6LTAPE20,IFET,-8)
      RETURN
      END
      SUBROUTINE TREAD(M,VAR,N)
      EQUIVALENCE (TEMP,IVAR)
      WRITE(20,100) M
100   FORMAT(1H ,A10,* =    *)
      READ(20,200) VAR
200   FORMAT(F15.5)
      IF(N.EQ.0) RETURN
      IF(VAR) 11,12,13
11    IVAR = IFIX(VAR-0.5)
      VAR = TEMP
      RETURN
12    IVAR = 0
      VAR = TEMP
      RETURN
13    IVAR = IFIX(VAR + 0.5)
      VAR = TEMP
      RETURN
      END
      SUBROUTINE TWRITE(M,VAR,N)
      IF(N.NE.0) GO TO 1
      WRITE(20,100) M,VAR
 100  FORMAT(1H ,A7,*=*,F10.5)
      GO TO 2
1     WRITE(20,200) M,VAR
200   FORMAT(1H ,A10,* = * I10)
2     END FILE 20
      RETURN
      END
```

APPENDIX II

PROGRAM LOOP2

Program LOOP2 is used in Chapter 4 to calculate the melting curves for $A_nC_mU_n$ hairpin loop molecules for different values of γ_m and ΔH°_m. The program is written so that results may be outputted onto a teletype machine so that new trial values of the parameters can be inputted from the teletype. More complete output information (the melting curves and a population analysis of intermediate states) may be obtained from a complete computer printout. The input parameters to the program include the AA double stranded stacking interactions and loop weighting functions and loop enthalpies for loops with 4 to 8 bases. The program considers both staggered and unstaggered configurations of the stem region.

A. Input Data Which is Read into the Program from Cards

Card Number	Symbol(s)	Format	Explanation
1	NUMMCL, ITMXO DELTAO	215, F10.5	NUMMCL = the number of molecules in the series to be calculated; ITMXO = the number of temperatures for which melting curve is calculated; DELTAO = temperature increment
2	TT(I), I = 1, NUMMCL	8F10.6	TT(I) = melting temperature of I th molecule
3	HIN, DEL, TBETA	8F10.6	HIN = ΔH°_{AA}; DEL = k_{end}; TBETA = temperature in °C for which the loop weighting function is reported (30°C)

4	LOOP, MXINO, WNAME(MOLECL)	2I5, A6	LOOP = dummy variable no longer used; MXINO = number of links in smallest loop (i.e., m+1); WNAME = label specifying the name of the molecule.
5	N, WD1(I), I = 1, N	I5, 40A1	N = the number of bases in the molecule; WD1(I) = the identity of the I th base in the stem; only one of the two complementary strands is specified
6 . . .			Cards 4 and 5 are repeated for each molecule to be calculated

B. Input Data Which is Read into Program from Teletype

Datum Number	Symbol(s)	Explanation
1	ALLNO	ALLNO = 1 specifies that the calculation the all-or-none approximation: only the fully bonded and unbonded states are considered
2	OUTPUT	OUTPUT = 1 gives complete output information on computer printout for each trial run
3	PLOT	PLOT = 1 causes melting curve to be plotted by CALCOMP; in this case, the experimental melting curve must be inputted into the program through cards; this is done in subroutine PLOT
4	STAG	STAG = 5HSTAGG causes the calculation to include staggered configurations in the partition function
5	HIN	HIN = ΔH°_{AA}
6	DEL	DEL = k_{end}
7	SKIP	If SKIP = J, the calculation will retain all values of loop weighting functions and loop enthalpies from previous calculation except for the loop with I links

8	SIGMAO(I)	SIGMAO(I) = value of loop weighting function at TBETA °C for loop with I links (i.e., I-1 bases)
9	HSIG(I)	HSIG(I) = enthalpy for loop of I links
10 . . .		SIGMAO(I) and HSIG(I) are repeated for each molecule calculated
11	HSIGLG	HSIGLG = loop enthalpy for all loops of more bases than the loop with stem region fully formed

C. Output

The output to teletype lists the calculated melting temperatures and slopes of the melting curves at T_m. The computer printout gives melting curve and population analysis for each of the molecules calculated. The CALCOMP plot gives plots of both melting curves and population analysis for each of the molecules.

D. Listing of Program LOOP2

```
      PROGRAM  LOOP(INPUT,OUTPUT,TAPE5=INPUT,TAPETTY=12,TAPE20=TAPETTY,
     1       TAPE6,TAPE98,PLOT,TAPE99=PLOT)
      COMMON/A/BETA(5),WD1(18),SS(90),EQK(90),DELT(90)
      COMMON/B/Q1(90),Q2(90),Q1NTOT(90),Q2NTOT(90),F1(90),F2(90)
      COMMON/C/TM,NDEL17(15),NDEL50(15),NDEL83(15),TMLST(40),LM
      COMMON/D/SIGMA0(20),HSIG(20),SIGMA(20,90),HSIG0(20)
      COMMON/PLTT/NUMMCL,ITMX0,DELTA0,MOLECL,DEL,HIN,TINIT,NSTAG,MXINO
     1,NALLNO
      DIMENSION BETA0(3),WNAME(10)
      DIMENSION TT(10)
      COMMON/NN/QN(12,90),QP(12,90)
      DIMENSION NREM(5),W(5,10),MMXX(10)
      CALL TTYSET
      TMAU=351.16
      TMAUV=1./351.16
      RINV=1./1.986
      READ(5,40) NUMMCL,ITMX0,DELTA0
   40 FORMAT(2I5,F10.5)
      READ(5,42)  (TT(I),I=1,NUMMCL)
   42 FORMAT(8F10.6)
      READ(5,51)   HIN,DEL,TBETA
   51 FORMAT(8F10.6)
      TBETA=TBETA+273.16
      TBV=1./TBETA
      DO 59 JJTTY=1,150
      CALL TREAD(5HALLNO,ALLNO,0)
      CALL TREAD(6HOUTPUT,FOUTP,0)
      CALL TREAD(4HPLOT,FPLOT,0)
      CALL TREAD(4HSTAG,STAG,0)
      CALL TREAD(3HHIN,HIN,0)
      CALL TREAD(3HDEL,DEL,0)
      NALLNO=ALLNO
      NPLOT=FPLOT
      NOUTP=FOUTP
      NSTAG=STAG
      DELSQ=SQRT(DEL)
      STAGG=5H
      IF(NSTAG.EQ.1)  STAGG=5HSTAGG
      WRITE(6,55)  HIN,DEL,STAGG
   55 FORMAT(*   ENTHALPY FOR AU FORMATION=*,F12.4,*    DELTA=*,F10.6
     1,10X,A5)
      WRITE(6,52)  NPLOT,NOUTP,NSTAG,NALLNO
   52 FORMAT(*  NPLOT=*,I2,*  NOUTPUT=*,I2,*  NSTAG=*,I2,*  NALLNO=*,I2)
      CALL SIGSET
      WRITE(6,48)  (I,SIGMA0(I),HSIG(I),I=5,9)
   48 FORMAT(2X,5(I2,F13.10,F6.0))
      DO 60  MOLECL=1,NUMMCL
      DO 65  ITM=1,2
      ITERMX=ITMX0



      TINIT=-20.
      TINIT=TT(MOLECL)-ITERMX/2
      DELTA=DELTA0
      IF(ITM.EQ.1)  GO TO 69
      TINIT=TM-1.1
      DELTA=0.1
      ITERMX=21
   69 IF(ITM.NE.1)  GO TO 100
      IF(JJTTY.GT.1)  GO TO 99
```

```
      READ(5,90)  LOOP,MXINO,WNAME(MOLECL)
90    FORMAT(2I5,A6)
      READ(5,110) N,(WD1(I),I=1,N)
110   FORMAT(I5,40A1)
      MMXX(MOLECL)=MXINO
      NREM(MOLECL)=N
      DO 98 I=1,N
98    W(MOLECL,I)=WD1(I)
      GO TO 100
99    N=NREM(MOLECL)
      MXINO=MMXX(MOLECL)
      DO 97 I=1,N
97    WD1(I)=W(MOLECL,I)
100   WRITE(6,115)  WNAME(MOLECL)
115   FORMAT(2X,A6)
      T=TINIT
      DO 120  ITER=1,ITERMX
      H=HIN
      T=T+DELTA
      TINV=1./(T+273.16)
      DELT(ITER)=T
      SS(ITER)=EXP(H*RINV*(TMAUV-TINV))
      NIT=NNIT=NNNIT=0
      Q1(ITER)=Q1NTOT(ITER)=Q2(ITER)=Q2NTOT(ITER)=0.
      DO 121 INT=1,12
      QN(INT,ITER)=0.
      QP(INT,ITER)=0.
121   CONTINUE
      DO 122  ILP=5,20
      SIGMA(ILP,ITER)=SIGMA0(ILP)*EXP(HSIG(ILP)*RINV*(T8V-TINV))
122   CONTINUE
120   CONTINUE
      M=N
      DO 1  IP=1,N
      IF(STAGG.NE.5HSTAGG)  M=IP
      DO 2  JP=IP,M
      I=IP-1
      J=JP-1
      MX=MXINO+I+J
      NMAX=N-J
      ND=0
      W=1.
      IF(I.NE.J)  W=2.
      DO 3  K=1,NMAX
      KM=K-1
      IF(K.EQ.NMAX)  ND=1
      IF(I.EQ.J.AND.K.EQ.NMAX)  ND=2
      NLP=I+J+1
      DO 160  ITER=1,ITERMX
      BT=SIGMA(MX,ITER)
      QF=(SS(ITER)**KM)*BT*W*(DELSQ**ND)
      Q1(ITER)=Q1(ITER)+QF
      Q1NTOT(ITER)=Q1NTOT(ITER)+KM*QF
      LK=N-K+1
      QN(LK,ITER)=QN(LK,ITER)+QF
      QP(NLP,ITER)=QP(NLP,ITER)+QF
160   CONTINUE
3     CONTINUE
2     CONTINUE
1     CONTINUE
140   CONTINUE
130   CONTINUE
126   CONTINUE
125   CONTINUE
      NTOT=N-1
132   VT=NTOT
```

```
      VNINV=1./VT


      IF(NTOT.EQ.0)  VNINV=0.
      DO 180  ITER=1,ITERMX
  250 QQ=(SS(ITER)**NTOT)*SIGMA(MXINO,ITER)*DEL
      F2(ITER)=1./(1.+QQ)
      Q1INV=1./(Q1(ITER)+1.)
      F1(ITER)=1.-Q1NTOT(ITER)*VNINV*Q1INV
      IF(NALLNO.EQ.1)  F1(ITER)=F2(ITER)
      DO 185  J=1,N
      QN(J,ITER)=QN(J,ITER)/Q1(ITER)
      QP(J,ITER)=QP(J,ITER)/Q1(ITER)
  185 CONTINUE
      IF(NOUTP.NE.1)  GO TO 252
  195 WRITE(6,190) SS(ITER),DELT(ITER),F1(ITER),F2(ITER),
     1QN(1,ITER),QN(2,ITER),QN(3,ITER),QN(4,ITER),QN(5,ITER),QP(1,ITER),
     2QP(2,ITER),QP(3,ITER),QP(4,ITER),QP(5,ITER)
  190 FORMAT(2X,14F9.5)
  252 IF (F1(ITER).GT.(0.11).AND.F1(ITER).LT.(0.23)) GO TO 280
  260 IF (F1(ITER).GT.(0.42).AND.F1(ITER).LT.(0.58)) GO TO 290
  270 IF (F1(ITER).GT.(0.77).AND.F1(ITER).LT.(0.89)) GO TO 300
      GO TO 180
  280 NIT=NIT+1
      NDEL17(NIT)=ITER
      GO TO 260
  290 NNNIT=NNNIT+1
      NDEL50(NNNIT)=ITER
      GO TO 270
  300 NNIT=NNIT+1
      NDEL83(NNIT)=ITER
  180 CONTINUE
      CALL EXTRAP(NIT,NNIT,NNNIT)
      PT=ITM*NPLOT
      IF(PT.EQ.1)  CALL PLOT
      IF(ITM.EQ.2) CALL DFDT
   65 CONTINUE
   60 CONTINUE
      CALL TWRITE(3HHIN,HIN,0)
      CALL TWRITE(3HDEL,DEL,0)
      CALL TWRITE(5HNSTAG,NSTAG,1)
   59 CONTINUE
      CALL EXIT
      END
      SUBROUTINE SIGSET
      COMMON/D/SIGMA0(20),HSIG(20),SIGMA(20,90),HSIG0(20)
      CALL TREAD(4HSKIP,SKIP,0)
      NSKIP=SKIP
      DO 5  I=5,9
      IF(I.EQ.8)  GO TO 5
      IF(NSKIP.LT.1)  GO TO 1
      IF(NSKIP.NE.I)  GO TO 5
    1 CALL TWRITE(2H I,I,1)
      CALL TREAD(6HSIGMA0,SIGMA0(I),0)
      CALL TREAD(7HHSIG(I),HSIG(I),0)
      HSIG(I)=1000.*HSIG(I)
    5 CONTINUE
      SIGMA0(8)=(SIGMA0(7)+SIGMA0(9))/2.
      HSIG(8)=(HSIG(7)+HSIG(9))/2.
      CALL TREAD(6HHSIGLG,HSIGLG,0)
      HSIGLG=1000.*HSIGLG
      DO 6  I=10,20
      HSIG(I)=HSIGLG
      X=I
      F=(X-1.)/X
```

```
      F=SQRT(F**3)
      IM=I-1
      SIGMA0(I)=F*SIGMA0(IM)
    6 CONTINUE
      RETURN
      END
      SUBROUTINE EXTRAP (NIT,NNIT,NNNIT)
      COMMON/A/BETA(5),WD1(18),SS(90),EQK(90),DELT(90)
      COMMON/B/Q1(90),Q2(90),Q1NTOT(90),Q2NTOT(90),F1(90),F2(90)
      COMMON/C/TM,NDEL17(15),NDEL50(15),NDEL83(15),TMLST(40),LM
      SIXINV=1./6.
      FIVESX=5./6.
      ZERO=0.
      HALF=1./2.
      D1SM=D1LG=1.
      IF (NIT.EQ.0) GO TO 30
      DO 20 I=1,NIT
      IN=NDEL17(I)
      D1=SIXINV-F1(IN)
      IF (D1.GE.ZERO) GO TO 10
C      HERE F1(ITER) IS GREATER THAN 1/6
      D1=-D1
      IF (D1.GT.D1LG) GO TO 20
      D1LG=D1
      IT17L=IN
      GO TO 20
   10 IF (D1.GT.D1SM) GO TO 20
      D1SM=D1
      IT17S=IN
   20 CONTINUE
      N1SM=D1SM
      N1LG=D1LG
      IF (N1SM.EQ.1.OR.N1LG.EQ.1) NIT=0
   30 D1SM=D1LG=1.
      IF (NNIT.EQ.0) GO TO 60
      DO 50 I=1,NNIT
      IN=NDEL83(I)
      D1=FIVESX-F1(IN)
      IF (D1.GE.ZERO) GO TO 40
      D1=-D1
C     HERE F1(ITER) IS GREATER THAT 5/6
      IF (D1.GT.D1LG) GO TO 50
      D1LG=D1
      IT83L=IN
      GO TO 50
   40 IF (D1.GT.D1SM) GO TO 50
      D1SM=D1
      IT83S=IN
   50 CONTINUE
      N1SM=D1SM
      N1LG=D1LG
      IF (N1SM.EQ.1.OR.N1LG.EQ.1) NNIT=0
   60 D1SM=D1LG=1.
      IF (NNNIT.EQ.0) GO TO 90
      DO 80 I=1,NNNIT
      IN=NDEL50(I)
      D1=HALF-F1(IN)
      IF (D1.GE.ZERO) GO TO 70
      D1=-D1
      IF (D1.GT.D1LG) GO TO 80
      D1LG=D1
      IT50L=IN
      GO TO 80
   70 IF (D1.GT.D1SM) GO TO 80
      D1SM=D1
      IT50S=IN
```

```
   80 CONTINUE


      N1SM=D1SM
      N1LG=D1LG
      IF (N1SM.EQ.1.OR.N1LG.EQ.1) NNNIT=0
C     EXTRAPOLATION PROCEDURE  (JUST USING TWO POINTS)
   90 IF (NIT.EQ.0) GO TO 100
      SLOPE=(F1(IT17L)-F1(IT17S))/(DELT(IT17L)-DELT(IT17S))
      T17=(SIXINV-F1(IT17S)+SLOPE*DELT(IT17S))/SLOPE
  100 IF (NNIT.EQ.0) GO TO 110
      SLOPE=(F1(IT83L)-F1(IT83S))/(DELT(IT83L)-DELT(IT83S))
      T83=(FIVESX-F1(IT83S)+SLOPE*DELT(IT83S))/SLOPE
  110 IF (NNNIT.EQ.0) GO TO 120
      SLOPE=(F1(IT50L)-F1(IT50S))/(DELT(IT50L)-DELT(IT50S))
      TM=(HALF-F1(IT50S)+SLOPE*DELT(IT50S))/SLOPE
      DEL23=0.
  120 IF (NIT.EQ.0.OR.NNIT.EQ.0) GO TO 130
      DEL23=T83-T17
  130 IF (NNNIT.EQ.0) TM=0.
      WRITE (6,690) TM,DEL23
  690 FORMAT  (*   PREDICTED MELTING TEMPERATURE=*,F10.6,*      DELTA(2/3
     13=*,F10.6)
      RETURN
      END
      SUBROUTINE DFDT
      COMMON/A/BETA(5),WD1(18),SS(90),EQK(90),DELT(90)
      COMMON/B/Q1(90),Q2(90),Q1NTOT(90),Q2NTOT(90),F1(90),F2(90)
      COMMON/C/TM,NDEL17(15),NDEL50(15),NDEL83(15),TMLST(40),LM
      DIMENSION SLOPE(25)
      SLPTM=0.
      DO 10  I=1,10
      J=22-I
      SLOPE(I)=(F1(I)-0.5)/(DELT(I)-TM)
      SLOPE(J)=(F1(J)-0.5)/(DELT(J)-TM)
      SLPTM=SLPTM+SLOPE(I)+SLOPE(J)
 10   CONTINUE
      SLPTM=SLPTM/20.
      WRITE(6,20)  SLPTM
 20   FORMAT(* THE SLOPE AT THE MIDPOINT OF THE MELTING CURVE IS*,F12.8)
      CALL TWRITE(2HTM,TM,0)
      CALL TWRITE(5HSLOPE,SLPTM,0)
      RETURN
      END
      SUBROUTINE TTYSET
      DIMENSION IFET(8)
77    READ(20,100) I
100   FORMAT(80X,I1)
      IF(I.NE.1) GO TO 66
      DO 10 J=1,10
      CALL CMOFF
10    CONTINUE
      GO TO 77
66    WRITE(20,200)
200   FORMAT( * I AM READY.  PROCEED,0 ONE OF ONES*)
      CALL FET(6LTAPE20,IFET,8)
      IFET(2) = IFET(2).OR.0000001000000000000B
      IFET(8) = IFET(8).OR.4000000000000000000B
      CALL FET(6LTAPE20,IFET,-8)
      RETURN
      END
      SUBROUTINE TREAD(M,VAR,N)
      EQUIVALENCE (TEMP,IVAR)
      WRITE(20,100) M
100   FORMAT(1H ,A10,* =    *)
```

```
      READ(20,200) VAR
200   FORMAT(F15.5)
      IF(N.EQ.0) RETURN
      IF(VAR) 11,12,13
11    IVAR = IFIX(VAR-0.5)
      VAR = TEMP
      RETURN
12    IVAR = 0
      VAR = TEMP
      RETURN
13    IVAR = IFIX(VAR + 0.5)
      VAR = TEMP
      RETURN
      END
      SUBROUTINE TWRITE(M,VAR,N)
      IF(N.NE.0) GO TO 1
      WRITE(20,100) M,VAR
 100  FORMAT(1H ,A7,*=*,F11.5)
      GO TO 2
1     WRITE(20,200) M,VAR
 200  FORMAT(1H ,A6,* = *,I6)
2     END FILE 20
      RETURN
      END
      SUBROUTINE PLOT
      COMMON/A/BETA(5),WD1(18),SS(90),EQK(90),DELT(90)
      COMMON/B/Q1(90),Q2(90),Q1NTOT(90),Q2NTOT(90),F1(90),F2(90)
      COMMON/D/SIGMA0(20),HSIG(20),SIGMA(20,90),HSIG0(20)
      COMMON/PLTT/NUMMCL,ITMX0,DELTA0,MOLECL,DEL,HIN,TINIT,NSTAG,MXINO
     1,NALLNO
      COMMON/NN/QN(12,90),QP(12,90)
      COMMON/CCPOOL/XMIN,XMAX,YMIN,YMAX,CCXMIN,CCXMAX,CCYMIN,CCYMAX
      DIMENSION FPOP0(90),FPOP1(90),FPOP2(90),FPOP3(90),FPOP4(90),FPOP5(
     190)
      DIMENSION ABEXP(60)
      DIMENSION ABS(60,5),TEXP(60),WNAME(10),DIF(60)
      IF(IHERE.EQ.120)  GO TO 5
      IHERE=120
      WX3=5HDEL =
      WX4=6HA-U EN
      WX5=6HTHALPY
      WX1=2H =
      WX6=6HNO STA
      WX7=6HGGERED
      WX8=6H
      WX9=6HINTERM
      WX10=6HEDIATE
      WX11=6HS
      WX12=6HSTAGGE
      WX13=6HRED SP
      WX14=6HECIES
      WX15=6HARE IN
      WX16=6HCLUDED
      WX17=6H
      WX18=6HLOOP E
      WX19=6HNTHALP
      WX20=3HY =
      WX21=6HSIGMA
      WX22=6H(20 DE
      WX23=4HG) =
      WX24=6HALL OR
      WX25=6H NONE
      WX26=6HMODEL
      HBP=HIN/(10.**3)
      READ(5,100)  IDATA,TFIRST,TDEL
 100  FORMAT(I5,2F10.5)
```

```
      DO 105  IX=1,NUMMCL


      READ(5,110)  WNAME(IX)
  110 FORMAT(12X,A6)
      DO 115  IY=1,IDATA
      READ(5,120)  TEXP(IY),ABS(IY,IX)
  120 FORMAT(2F10.5)
  115 CONTINUE
  105 CONTINUE
      CCXMIN=130.
      CCXMAX=930.
      CCYMIN=200.
      CCYMAX=800.
      CX1=CCXMIN+75.
      CX2=CX1+375.
      CY1=CCYMAX+120.
      CY2=CY1+65.
      HNTT=(CCXMAX-CCXMIN)/2.-55.
      B=(675.-CCYMIN)/(CCYMAX-CCYMIN)+.005
      C=(650.-CCYMIN)/(CCYMAX-CCYMIN)+.004
      D=(625.-CCYMIN)/(CCYMAX-CCYMIN)+.004
      B1=(675.-CCYMIN)/(CCYMAX-CCYMIN) + .004
      B2=(660.-CCYMIN)/(CCYMAX-CCYMIN) + .004
      B3=(645.-CCYMIN)/(CCYMAX-CCYMIN) + .004
      B4=(630.-CCYMIN)/(CCYMAX-CCYMIN) + .004
      B5=(615.-CCYMIN)/(CCYMAX-CCYMIN) + .004
      B6=(600.-CCYMIN)/(CCYMAX-CCYMIN) + .004
    5 XMAX=80.
      XMIN=-20.
      YMIN=0.
      YMAX=1.
      INT=XMAX-XMIN
      INT2=INT/5
      A=XMIN+3.5000
      A1=(640.-CCXMIN)*(XMAX-XMIN)/(CCXMAX-CCXMIN)-1.5+XMIN
      DO 21  I=1,IDATA
   21 ABEXP(I)=ABS(I,MOLECL)
      CALL CCGRID (1,INT,2,6HNOLBLS,1,10,4)
      CALL CCLBL (INT2,10)
      CALL CCLTR (HNTT,80.,0,3,11HTEMPERATURE)
      CALL CCLTR (30.,425.,1,3,7H1-THETA)
      REWIND 98
      WRITE(98,40)  WNAME(MOLECL)
   40 FORMAT(A6)
      CALL CCLTR(CX1,CY1,0,5)
      REWIND 98
      WRITE(98,45)  (WX3,DEL)
   45 FORMAT(A5,F6.4)
      WRITE(98,50)  (WX4,WX5,WX1,HBP)
   50 FORMAT(2A6,A2,F6.1)
      HLOOP=HSIG(MXINO)/(10.**3)
      WRITE(98,47)  WX18,WX19,WX20,HLOOP
   47 FORMAT(2A6,A3,F5.1)
      WRITE(98,48)  WX21,WX22,WX23,SIGMAO(MXINO)
   48 FORMAT(2A6,A4,E10.3)
      IF(NALLNO.EQ.1)  GO TO 62
      IF(NSTAG.EQ.1)  GO TO 60
      WRITE(98,55)  WX6,WX7,WX8
      WRITE(98,55)  WX9,WX10,WX11
   55 FORMAT(3A6)
      GO TO 65
   60 WRITE(98,55)  WX12,WX13,WX14
      WRITE(98,55)  WX15,WX16,WX17
      GO TO 65
```

```
62    WRITE(98,55)  WX24,WX25,WX26
65    CALL CCLTR(CX2,CY2,0,2)
      CALL CCLTR (200.,700.,0,1,6HLEGEND)
      CALL CCLTR (175.,675.,0,1,21H= EXPERIMENTAL POINTS)
      CALL CCPLOT (A,B,1,6HNOJOIN,7,1)
      CALL CCLTR (175.,650.,0,1,19H= CALCULATED POINTS)
      CALL CCPLOT (A,C,1,6HNOJOIN,1,1)
      CALL CCPLOT(DELT,F1,ITMX0,4HJOIN,1,1)
      CALL CCPLOT(TEXP,ABEXP,IDATA,4HJOIN,7,1)
      CALL CCNEXT
      JO=TFIRST-TINIT
      DO 200  I=1,IDATA
      J=JO+(I-1)*(TDEL/DELTAO)
      WRITE(6,201)  DELT(J),TEXP(I)
201   FORMAT(F10.5,* = *,F10.5)
      DIF(I)=F1(J)-ABEXP(I)
200   CONTINUE
      YMIN=-0.25
      YMAX=+0.25
      CALL CCGRID(1,INT,2,6HNOLBLS,2,5,4)
      CALL CCLBL(INT2,10)
      CALL CCLTR(30.,300.,1,3,19H1-THETA (CALC-EXPT))
      CALL CCLTR(HNTT,80.,0,3,11HTEMPERATURE)
      REWIND 98
      WRITE(98,40)  WNAME(MOLECL)
      CALL CCLTR(CX1,CY1,0,5)
      REWIND 98
      WRITE(98,45)  (WX3,DEL)
      WRITE(98,50) (WX4,WX5,WX1,HBP)
      WRITE(98,47)  WX18,WX19,WX20,HLOOP
      WRITE(98,48)  WX21,WX22,WX23,SIGMAO(MXINO)
      IF(NALLNO.EQ.1)  GO TO 162
      IF(NSTAG.EQ.1)  GO TO 160
      WRITE(98,55)  WX6,WX7,WX8
      WRITE(98,55)  WX9,WX10,WX11
      GO TO 165
160   WRITE(98,55)  WX12,WX13,WX14
      WRITE(98,55)  WX15,WX16,WX17
      GO TO 165
162   WRITE(98,55)  WX24,WX25,WX26
165   CALL CCLTR(CX2,CY2,0,2)
      CALL CCPLOT(TEXP,DIF,IDATA,4HJOIN,6,1)
      CALL CCNEXT
2122  CONTINUE
      IF(NALLNO.EQ.1)  RETURN
      YMIN=0.
      YMAX=1.
      CALL CCGRID(1,INT,2,6HNOLBLS,1,10,4)
      CALL CCLBL(INT2,10)
      CALL CCLTR(HNTT,80.,0,3,11HTEMPERATURE)
      CALL CCLTR(30.,300.,1,3,18HSPECIES POPULATION)
      REWIND 98
      WRITE(98,40)  WNAME(MOLECL)
      CALL CCLTR(CX1,CY1,0,5)
      REWIND 98
      WRITE(98,45)  (WX3,DEL)
      WRITE(98,50) (WX4,WX5,WX1,HBP)
      WRITE(98,47)  WX18,WX19,WX20,HLOOP
      WRITE(98,48)  WX21,WX22,WX23,SIGMAO(MXINO)
      IF(NALLNO.EQ.1)  GO TO 262
      IF(NSTAG.EQ.1)   GO TO 260
      WRITE(98,55)  WX6,WX7,WX8
      WRITE(98,55)  WX9,WX10,WX11
      GO TO 265
260   WRITE(98,55)  WX12,WX13,WX14
      WRITE(98,55)  WX15,WX16,WX17
```

```
      GO TO 265


  262 WRITE(98,55)  WX24,WX25,WX26
  265 CALL CCLTR(CX2,CY2,0,2)
      DO 300 I=1,ITMXO
      G=1.-F1(I)
      FPOP0(I)=F1(I)
      FPOP1(I)=G*QP(1,I)
      FPOP2(I)=G*QP(2,I)
      FPOP3(I)=G*QP(3,I)
      FPOP4(I)=G*QP(4,I)
      FPOP5(I)=G*QP(5,I)
  300 CONTINUE
      CALL CCLTR(700.,700.,0,1,6HLEGEND)
      CALL CCLTR(640.,675.,0,1,23HSINGLE STRANDED SPECIES)
      CALL CCLTR(640.,660.,0,1,20HMINIMUM LOOP SPECIES)
      CALL CCLTR(640.,645.,0,1,18HMIN. LOOP + 1 BASE)
      CALL CCLTR(640.,630.,0,1,19HMIN. LOOP + 2 BASES)
      CALL CCLTR(640.,615.,0,1,19HMIN. LOOP + 3 BASES)
      CALL CCLTR(640.,600.,0,1,19HMIN. LOOP + 4 BASES)
      CALL CCPLOT(A1,B1,1,6HNOJOIN,9,1)
      CALL CCPLOT(A1,B2,1,6HNOJOIN,8,1)
      CALL CCPLOT(A1,B3,1,6HNOJOIN,7,1)
      CALL CCPLOT(A1,B4,1,6HNOJOIN,6,1)
      CALL CCPLOT(A1,B5,1,6HNOJOIN,5,1)
      CALL CCPLOT(A1,B6,1,6HNOJOIN,4,1)
      CALL CCPLOT(DELT,FPOP0,ITMXO,4HJOIN,9,4)
      CALL CCPLOT(DELT,FPOP1,ITMXO,4HJOIN,8,4)
      CALL CCPLOT(DELT,FPOP3,ITMXO,4HJOIN,6,4)
      CALL CCPLOT(DELT,FPOP5,ITMXO,4HJOIN,4,4)
      IF(NSTAG.EQ.0)  GO TO 310
      CALL CCPLOT(DELT,FPOP2,ITMXO,4HJOIN,7,4)
      CALL CCPLOT(DELT,FPOP4,ITMXO,4HJOIN,5,4)
  310 CALL CCNEXT
      RETURN
      END
       SUBROUTINE CCLBL(NX1,NY1)
       COMMON/CCPOOL/XMIN,XMAX,YMIN,YMAX,CCXMIN,CCXMAX,CCYMIN,CCYMAX
       COMMON/CCFACT/FACTOR
       ISZERO=0
       XD=XMAX-XMIN           $YD=YMAX-YMIN
       CCXD=CCXMAX-CCXMIN     $CCYD=CCYMAX-CCYMIN
C
      KSIZE=2
      KORIENT = 0
      IF ( NX1 .EQ. 0 )  GO TO 5
C
C      LABEL FROM RIGHT TO LEFT ALONG THE X-AXIS.
C
      XI=XD/FLOAT(NX1)
       DO 2 NX=ISZERO,NX1
       CCX=CCXMAX-CCXD*FLOAT(NX)/FLOAT(NX1)
       X=(CCX-CCXMIN)*XD/CCXD+XMIN
C
C      SET X TO A TRUE ZERO IF X=0. TO WITHIN MACHINE ACCURACY.
C
       IF(ABS (X/XI).LT.1.0E-6)X=0.
      WRITE(98,27) X
    2 CALL CCLTR (CCX-65.*FLOAT(KSIZE)/FACTOR,
     X             CCYMIN-9.*FLOAT(KSIZE)/FACTOR,
     X             KORIENT, KSIZE )
C
    5 IF ( NY1 .EQ. 0 )  RETURN
C
```

```
C       LABEL UPWARD ALONG THE Y-AXIS.
C
       YI=YD/FLOAT(NY1)
        DO 3 NY=ISZERO,NY1
        CCY=CCYMIN+CCYD*FLOAT(NY)/FLOAT(NY1)
        Y=(CCY-CCYMIN)*YD/CCYD+YMIN
C
C       SET Y TO A TRUE ZERO IF Y=0. TO WITHIN MACHINE ACCURACY.
C
        IF(ABS (Y/YI).LT.1.0E-6)Y=0.
       WRITE(98,28)  Y
3       CALL CCLTR(CCXMIN-70.*FLOAT(KSIZE)/FACTOR,CCY,KORIENT,KSIZE)
C
        RETURN
C
 27    FORMAT(F10.0)
 28    FORMAT(F10.2)
        END
```

APPENDIX III

PROGRAM DBL

Program DBL is used in Chapter 5 to predict the secondary structures of an RNA molecule. We present a detailed explanation of the program, which should (1) make clear the various assumptions about thermodynamic parameters made at the time the work was completed. As more data are obtained, especially about loops, several of these assumptions can be removed by changing the internal logic in specific places; (2) clarify difficult points in the logic; (3) explain the output of the program.

A. Input Parameters for Program DBL

Card Number	Symbol(s)	Format	Explanation
1	LAB(I)	12A6	LAB(2) to LAB(12) are printed at beginning of the output as a label. To extend the message to an additional line of output, set LAB(1) = bCONTb.
2	HIN, HGG, HGA, HGU, HGC, HAA	7F10.5	HIN = the enthalpy of the double stranded AA interaction in cal/mole. HGG = ΔH°_{GG} - HIN (The enthalpy of the GG double stranded stacking interaction minus the AA stacking interaction.)
3	HAU, HAG, HUA, HUG, HCG	7F10.5	Symbols have same meaning as for card 2; e.g. HAU = ΔH°_{AU} - HIN.
4	EKGGO, EKGAO, EKGUO, EKGCO, EKAAO	7F10.5	EKGGO $\equiv k^\circ_{GG}$ in the text. (See Chapter 3, equation 3.2a.) EKGGO = $e^{(\Delta S^\circ_{GG}-\Delta S^\circ_{AA})/R}$

			Similarly, EKGAO $\equiv k^{\circ}_{GA} = e^{(\Delta S^{\circ}_{GA}-\Delta S^{\circ}_{AA})/R}$. Note that EKAAO $\equiv k^{\circ}_{AA} = e^{(\Delta S^{\circ}_{AA}-\Delta S^{\circ}_{AA})/R} = 1$.
5	EKAUO, EKAGO, EKUAO, EKUGO, EKCGO	7F10.5	Defined as in card 4.
6	SIGMAO(I), I = 5, 9	7F10.5	SIGMAO(I) = the loop weighting function for a loop with I links at temperature T_o = 20°C
7	HSIG(I), I = 5, 9	7F10.5	HSIG(I) $\equiv \Delta H^{\circ}_I$, the enthalpy of formation of a loop with I links.
8	RGA	F10.5	RGA = κ_G/κ_A. Thus, the initiation factor for a base pairing region with at least one G-C base pair is RGA$\cdot\kappa_A$.
9	INTLP, ENINT	I5, F10.5	INTLP = the minimum number of links in a loop for which the loop is assumed to be an interior loop. ENINT = the standard free energy assigned to an interior loop in cal/mole.
10	NDEBUG	I5	NDEBUG = parameter which determines how much output program prints. Standard option is NDEBUG = 0. If it is necessary to output all intermediate secondary structures, set NDEBUG $\neq$ 0.
11	NUMMCL	I5	NUMMCL = number of different molecules considered in the calculation.

12	LAB(I), I = 1, 12	12A6	The identity of molecule calculated is specified in LAB(I), I = 2, 12. LAB(1) = NONEbb for the standard calculation. LAB(1) = VARY15 for the option which is explained in note under the explanation of subroutine LK15V. (This option takes a series of solutions consisting of three stable vectors and calculates a solution which assumes the presence of these sets of vectors. The information of which sets of vectors to include must be inputted: the proper format for this information is contained in note 3-e.
13	N	I5	Number of biological bases in the RNA
14	EXCLP	12A6	IF EXCLP =bEXCLP, then two calculations are run; the first allows bases in loops to base pair and the second does not permit bases in loop to pair. IF EXCLP = EXCLP2, then only one calculation is performed, for which bases in the loop may not base pair.
15	WD(I), I = 1,N	80A1	WD(I) specifies the identity of the I^{th} base, counting from the 5' end of the molecule. The sequence is specified by the symbols A, U, G, and C, for adenine, uracil, guanine, and cytosine respectively. Since thymidine base pairs like uracil, it may be specified by U. For the odd bases, D = dihydrouracil and it is forbidden to base pair, P = pseudouracil, which is treated like uracil; X = an unknown base, which is permitted to form a base pair with any base; Y = a base which cannot form a base pair.

Cards 12-15 are repeated for each of the NUMMCL RNA molecules.

B. Structure of Program DBL

1. Beginning of Program to Statement 48

This part of the program reads in data and sets up temperature dependence of the free energies. Several things should be noted:

a. After statement 4: TBV is $1/T_o$ in $^oK^{-1}$ in the equation $\sigma = \sigma_o e^{-\Delta H^o_{LOOP}/R(1/T - 1/T_o)}$. TBV is set equal to $1/(20 + 273.16)$, since $\sigma_o^{\dagger}$ is the loop weighting function at 20°C. If a loop weighting function at any temperature other than 20°C is read in as input data, then TBV must be changed within the program.

b. After DO 8: the values of $\sigma_o(I)$, I = 4, 3, 2, are set as follows: $\sigma_o(4) = 0.2\ \sigma_o(5)$; $\sigma_o(3) = 0.2\ \sigma_o(4)$; $\sigma_o(2) = 0.2*\sigma_o(3)$. This is a rather arbitrary quantification of the observation that very small loops are destabilized, presumably because of steric constraints. Later in the program (see note 2.d), loops with three or fewer links (i.e., one or two bases in the loop) are forbidden. In all cases, SIGMAO(I) refers to a loop with I links (i.e., I-1 bases).

c. HSIG(1) = HSIG(2) = ... = HSIG(5): the assumption is made that all loops with less than 5 links have the same enthalpy as the loop with 5 links. This is unlikely to be the case and, if the calculation is to be performed reliably at different temperatures, the assumption ought to be removed (when data are available).

d. ENINT = -ENINT: All energies inputted into the program are inputted as positive numbers. The program converts

---

†We use σ as the loop weighting to conform to the logic of the program. (It is equivalent to σ of Chapter 4.)

the energy of an internal loop into a stability number, defined as the negative of the free energy. Because the free energy of loop formation is positive, the stability number will be negative for loops.

e. SIGMA0(8) = ... : No value of $\sigma_o(8)$ based on experimental was not available. It was assumed equal to the average of $\sigma_o(7)$ and $\sigma_o(9)$.

f. SIGMA0(I) = F*SIGMA0(I-1). Hairpin loops of increasing number of links are assigned a loop weighting function in accordance with the equation $\sigma_o(I) = [(I-1)/I)^{3/2} \cdot \sigma_o(I-1)$. This is the functional relationship appropriate for a chain with links which are unconstrained in their angular orientations. As data become available for larger loops, this assumption might eventually be removed. Alternatively, the loop weighting functions could be assigned a limiting value for all loops above a limiting loop size. (This is effectively the procedure used in the present calculations, since we assume that all loops with greater than INTLP links have a free energy of ENINT kcal/mole. (See note below.)

g. IF(I.G.E.INTLP): Since the analysis of alternative secondary structures has shown that most large loops are internal loops, we approximate all loops with more than INTLP links with a free energy deemed appropriate for internal loops. We assign no temperature dependence nor size dependence to the free energy of these internal loops.

h. GUFAC = ... : A G-U base pair is assumed to have a free energy of -1 kcal/mole when it is involved in any double stranded stacking interaction.

i. FERGA: This is the stability number associated with the initiation of a loop at a G-C base pair. As noted later, if any GC base pair is formed within the base pairing region, it is assumed that initiation takes place at the G-C base pair and the stability number FERGA is added to the total stability of the base pairing region.

2. Statement 48 to 230.

This section takes the primary sequence of the RNA molecule and determines all possible base pairing regions which can be formed, subject to the constraint that only G-C, A-U, and G-U base pairs are allowed and the base pairing regions must be anti-parallel. It calculates two stability numbers for each base pairing region: one includes only the effect of the stacking interactions (and favorable G-C initiation, if present); the second includes the destabilization associated with the loop formation. On the basis of the second stability number, the base pairing regions (called vectors) are listed in order of descending stability.

a. READ(5,5) to READ(5,20): Input parameters for a given RNA molecule are read in.

b. DO 26: 1EXC = 1 and 1EXC = 2 are the two cases considered in which base pairing regions between bases in different loops may and may not form, respectively. Discussion of these options is deferred to note 3.d, where the logic is explained.

c. DO 50 ... 50 CONTINUE: Here, all conceivable base pairs which can form are registered. E(I,J) = 1 means that the I^{th} and J^{th} base taken together may form a G-U base pair; E(I,J) = 2 is an AU base pair; E(I,J) = 4 is a GC base pair. The information

is then summarized in a matrix of all values of E(I,J),J I. For example, for the sequence GAUGCCUAG, the matrix is written.

```
  G A U G C C U A G
G 0
A 0 0
U 1 2 0
G 0 0 1 0
C 4 0 0 4 0
C 4 0 0 4 0 0
U 1 2 0 1 0 0 0
A 0 0 1 0 0 0 1 0
G 0 0 1 0 4 4 1 0 0
```

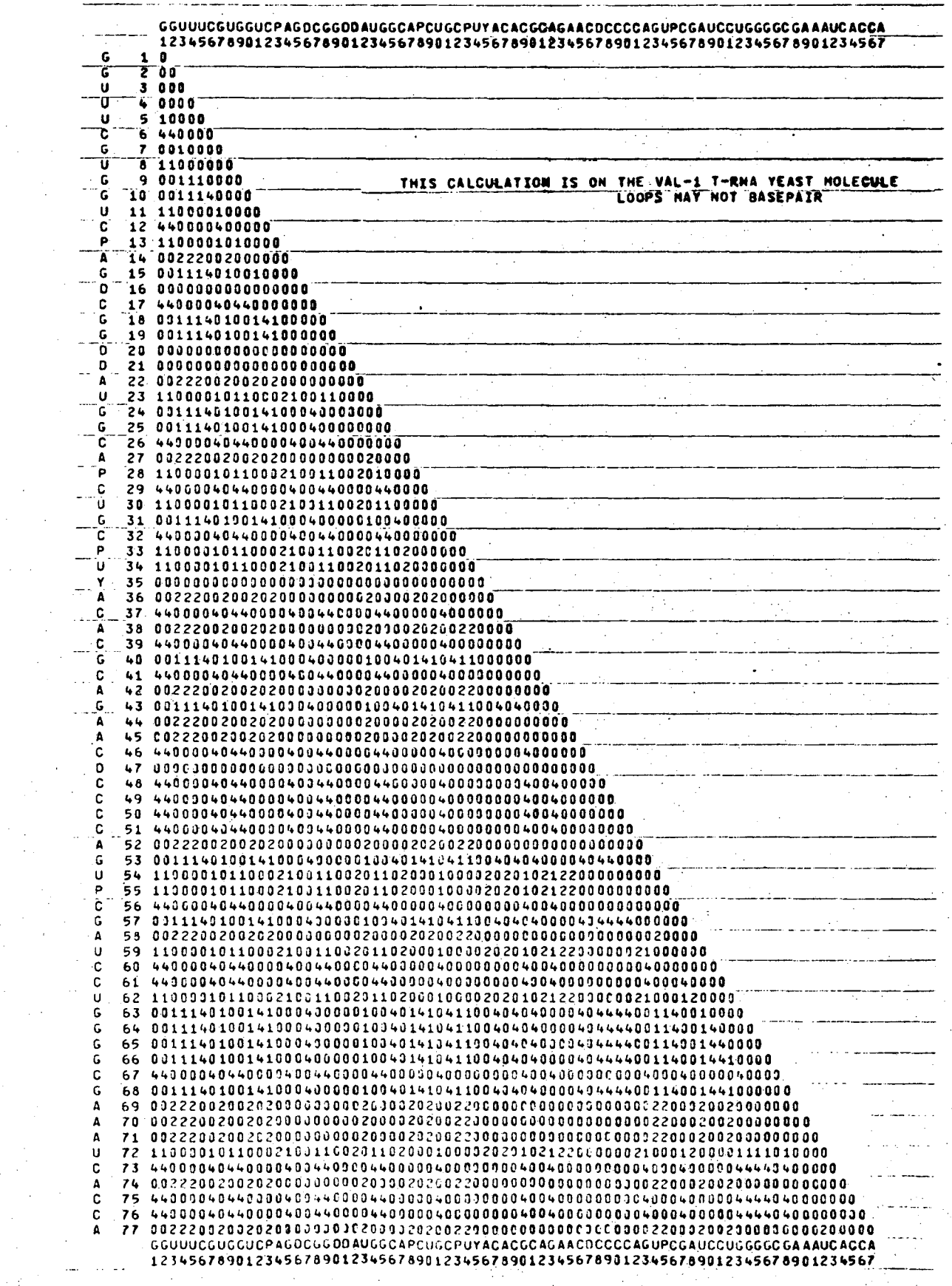
d. IF((I-J).LE.3) E(I,J) = 0: This statement disallows loops with three or fewer links (i.e., 2, 1, or 0 bases). The above matrix is then written:

```
    I  1 2 3 4 5 6 7 8 9
J      G A U G C C U A G
1 G    0
2 A    0 0
3 U    0 0 0
4 G    0 0 0 0
5 C    4 0 0 0 0
6 C    4 0 0 0 0 0
7 U    1 2 0 1 0 0 0
8 A    0 0 1 0 0 0 0 0
9 G    0 0 1 0 4 0 0 0 0
```

For a short sequence, this constraint excludes a significant fraction of base pairs. For a larger sequence, only a small per cent of possible base pairs are excluded. The effect is to set all terms three or less removed from diagonal I = J to zero.

We will illustrate the output of the program for valine tRNA from yeast. The matrix, which establishes all possible base

pairs, is shown below. We call this the base pairing matrix.



e. DO 110 J=1, N-1: N is the number of biological bases in the RNA molecule. This do loop begins the process of searching for stable base pairing regions in the molecule. The principle of the search routine is that anti-parallel base pairing regions will appear in the base pairing matrix only as lines on the 45° diagonal (going from the lower left to the upper right) which have no zero. To understand the reason for this, consider bases 5 to 7 and 58 to 56 in valine tRNA. The base pairs formed are:

5 6 7
U C G
A G C
58 56

As the number in the sequence increases from 5 (for U) to 6 (for C), the number of the complement decreases by 1, from 58 (for A) to 57 (for G). This is the condition for the base pairing region to be on the 45° diagonal, as stated above. An approximation involved in this approach is that, if a bulge occurrs, it will break up a complementary region into two.

The search routine begins with the longest diagonal, i.e., it starts with base 1 and 77, goes to base 2 and 76, 3 and 75, etc. It checks to see if a base pair is allowed and continues until a break occurs, in which a base pair is not allowed. It then calculates the stability number of the base pairing region, henceforth called a vector, by summing up the stability numbers of all double stranded stacking interactions, adding the stability number for initiation at a GC base pair if one is present, and adding the loop stability number. Since the size of the internal loop, if one is present, cannot be determined until the final structure of the

entire molecule is known, the loop contribution is approximated by the number of links between the ends of the vector, going in the 5' to 3' direction. For example, for

```
    5   7
U C G,
A G C
   58  56
```

the loop size is taken to be 56-7 = 49 links. (Because of the fact that we do not distinguish a size dependence of internal loops, this approximation does not in this case make any difference. When the size dependence is known and included, a correction will need to be made in the final assessment of stabilities of the different secondary structures for this approximation.) Once the stability number is calculated, it is compared with a lower limit to see if the vector is to be retained as a possible region in a stable secondary structure. A stability number of zero is taken as the lower limit in the program.

Having completed the first diagonal, the second one is searched in the same manner. This diagonal begins with base 2 and 77, goes to 3 and 76, 4 and 75, etc. The process is continued until all of the diagonals below the first one have been searched for stable vectors. (The DO 115 loop performs the same routine for the remaining diagonals above the first diagonal.)

The logic is easily followed by noting the definition of several variables and indices within the DO 110 loop:

J specifies the diagonal to be searched, and is the index of the first base in the diagonal.

JX is the index for base on the horizontal axis; it increases in increments of one as the diagonal is searched.

IX is the index for the base on the vertical axis; it decreases in increments of one.

JMX and IMX are indices for the preceding base pair, so that the nearest neighbor double stranded stacking stability numbers can be used to calculate the stability of the base pairing vector.

KF is the number of elements in the diagonal which is searched.

S sums the stability numbers associated with the double stranded stacking interactions. It overcounts by allowing an extra stacking interaction at each end of the vector. When the final stability numbers are calculated in subroutine LOOK, this overcounting is corrected.

In SUBROUTINE LOOK,

IFG determines if a GC base pair is present in the vector, so as to add the extra term for initiation at a GC.

SUM1 is the stability number of the vector, not including the loop stability number.

SUM2 is the stability number with loop included.

NCOUNT is the index of the stable vectors.

After DO loop 115, the process of finding all stable vectors is complete. The output is shown below. Base number I1 is paired with base number J1 and base number I2 with base J2. The first stability number is without loop; the second, with.

The program then skips around a bit, as shown by the following comments.

f. 1F(ICT-230) Program goes to 231.

g. Card 231 to 225: Reordering of vectors according to stability numbers.

THIS CALCULATION IS ON THE VAL-1 T-RNA YEAST MOLECULE
LOOPS MAY NOT BASEPAIR

VECTOR NUM	I1	J1	I2	J2	ENERGY	WITH LOOP ENERGY
1	66	48	61	53	19.76	13.84
2	73	1	66	8	12.27	6.27
3	44	28	40	32	11.59	5.67
4	65	48	63	50	10.96	4.46
5	66	49	64	51	10.96	4.46
6	68	39	66	41	10.11	4.11
7	77	8	74	11	9.97	3.97
8	68	30	64	34	9.16	3.16
9	40	6	37	9	8.99	2.99
10	76	1	74	3	8.03	2.03
11	77	23	75	25	7.98	1.98
12	52	23	50	25	7.98	1.98
13	52	8	50	10	7.98	1.98
14	70	54	67	57	7.95	1.68
15	77	62	75	64	7.98	1.63
16	75	7	71	11	7.04	1.04
17	58	5	56	7	6.63	.63
18	56	43	53	46	5.99	.21
19	67	25	66	26	6.16	.16
20	41	25	40	26	6.16	.16
21	67	2	62	7	6.12	.12
22	57	17	55	19	6.06	.06
23	65	11	61	15	6.05	.05
24	33	24	31	26	7.16	.04
25	61	1	60	2	6.04	.04
26	51	1	50	2	6.04	.04
27	50	1	49	2	6.04	.04
28	49	1	48	2	6.04	.04
29	61	18	60	19	6.04	.04
30	61	24	60	25	6.04	.04
31	76	18	75	19	6.04	.04
32	64	48	63	49	6.04	.04
33	66	50	65	51	6.04	.04
34	50	24	49	25	6.04	.04
35	49	24	48	25	6.04	.04
36	61	9	60	10	6.04	.04
37	51	18	50	19	6.04	.04
38	50	18	49	19	6.04	.04
39	49	18	48	19	6.04	.04
40	50	9	49	10	6.04	.04
41	49	9	48	10	6.04	.04

h. Go to 157. Returns to output reordered vectors, as shown below.

i. 1F(ICT-230) After 170 DO loop, in which reordered vectors are outputted, the program goes to 232, where the base pairing matrix is written over and a letter code is assigned to the vectors. A signifies the most stable vector, B the second most stable, etc. The vectors, so coded, are then outputted in a manner similar to the base pairing matrix, so that inspection of the new matrix readily shows where the most stable vectors may be found. Program DBL skips to 51 to output this matrix and then returns to 230, where the first stage of the calculation is completed. The matrix with coded vectors is shown below.

3. Cards 230 to 251

The remaining part of the program determines which vectors are compatible with others to form stable secondary structures for an RNA molecule. The main program is concerned primarily with outputting results and is not discussed further. Subroutines INSPECT, COMPAR, LPK, LPK15, and LPK15V do the work and are clarified below. These subroutines are contained in two DO loops. DO1 repeats the entire calculation, except for the inputting of thermodynamic data, for each of the different RNA molecules whose secondary structure is to be calculated. DO 26 performs the calculation twice for each molecule, once for the case where bases in hairpin loops may base pair and once for the case where they may not base pair.

SUBROUTINE INSPECT

This subroutine determines compatibility between vectors.

a. 1F(NCOUNT.GE.90): The maximum number of vectors considered in the calculation is 90. A maximum (under 200) must be set to avoid overwriting indexed variables. Experience has shown that with the length of molecules suitable for the program, 90 includes all vectors which contribute significantly to the stability of the molecule.

b. DO 10, DO 20: These DO loops set up the comparison of all vectors with each other, to determine which ones exclude the presence of others.

SUBROUTINE COMPARE

This subroutine sets up the vector exclusion matrix, for one of two cases. We explain first the case where bases in loops can base pair and then for the case where they cannot.

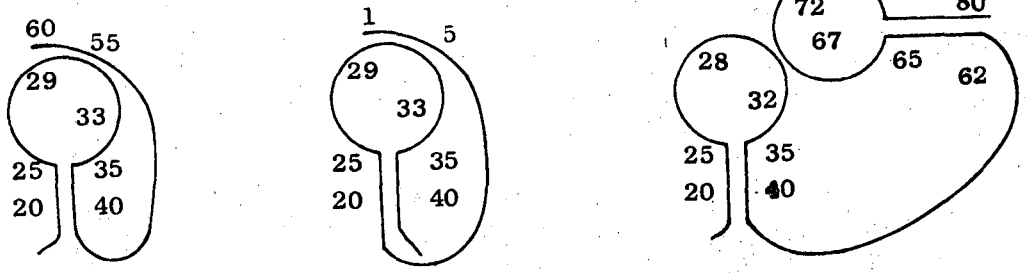
c. If EXCLF $\neq$ 1, program skips to 5 if base pairs between loops are permitted. We explain the logic by considering the case of vector 1 (I1 = 66, J1 = 48, I2 = 62, J2 = 53) compared with vector 8 (I1 = 65, J1 = 48, I2 = 63, J2 = 50). In this instance, the vectors overlap and necessarily exclude each other.

NSSET = IA = 1; program goes to 100 and sets INDX(I) equal to 1 for I = 66, 65, 64, 63, 62, 48, 49, 50, 51, 52, 53; all other values of INDX(I) = 0. Goes to 30, where NLAST = NSSET retains these values of INDX(I), while vector 1 is compared to all other vectors. Program goes through DO loop set up in INSPECT, taking vector 2, 3, ... For vector 8, JNDX(I) = 1 for I = 65, 64, 63, 48, 49, 50; otherwise JNDX(I) = 0. In DO 40, if INDX(I) and JNDX(I)

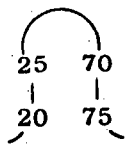
are both 1, then an overlap between the two vectors has occurred. The program is set up to allow one overlap between vectors (ICASE = 2 for all runs), so that the first overlap is simply noted and E(I1, IA) is set to 1. (I1 is the index of the first vector, 1, and IA the index of the second, 8). DO 40 continues and a second overlap is noted. E(I1, IA) = 2 and the two vectors are excluded (i.e., may not appear in the same final secondary structure. If the process were completed without an overlap being found, E(IA, I1) is set to zero. As pointed out, this process is repeated for all sets of two vectors, thus generating a matrix E, which we call the vector exclusion matrix.

If only one overlap occurs (i.e., only one of the indices for the two vectors is the same), then this must of necessity occur at the end of the base pairing region. In this case, one base pair can be eliminated from one of the vectors (in the final analysis), with the appropriate decrease in the stability of the secondary structure accounted for.

d. If EXCLP = 1, base pairing between the loops is forbidden. The logic is contained in the IF statements preceding card 5 in the subroutine. The following base pairing arrangements are forbidden:



However, bases within a loop may base pair with other bases within the same loop. For example, for the loop



base pairing arrangements for bases 26-29 are allowed.

e. SUBROUTINES LPK, LPK15, LPK15V

The procedure followed in these subroutines is discussed extensively in the text. Briefly, subroutine LPK assumes the presence of V1, sets up a reduced exclusive matrix with 5 x 5 elements, determines all allowable (i.e., nonoverlapping) vector sets with one to five vectors, and calculates the energy of the allowable vector sets. The process is repeated, assuming V2, V3, V4, and V5 individually in the final solution. The stability number of the most stable secondary structure multiplied by 0.9 is used as the cutoff for the more complete calculation, performed in subroutine LPK15. In this subroutine, the reduced exclusion matrix is 15 x 15 and one of the first seven vectors is assumed to be in the final solution.

Subroutine LPK15V may be used if one wishes to extend the range of vectors considered in the solution. This is done by taking the best three vectors of the most favorable structures using results of LPK15. In order to use this subroutine, LAB(1) in data card 12 must be set equal to VARY15. Then the following additional input data must be supplied:

Card Number	Symbol	Format	Explanation
16	NDATA	I5	NDATA = the number of sets of three vectors inputted.
17	ICHOSE(I), I=1,3	3I5	ICHOSE(I), I = 1,3 = the indices of the three assumed vectors. The NDATA sets of three vectors are read from NDATA cards which follow the format of card 17.

To illustrate how this subroutine works, assume that the standard calculation has been carried out and all solutions more stable than a cutoff energy are chosen. The most stable three vectors for each of these solutions are inputted into program DBL in cards 16 and 17. For each set of three vectors, a reduced exclusion matrix is formed. Because all vectors excluded by any of the three will not appear in the reduced exclusion matrix, the matrix can extend the calculation to a very large number of vectors. Once the reduced exclusion matrix is set up, the problem is solved in the standard way. Then the process is repeated for the next set of three vectors. (The results reported did not use this method. Instead, the solutions generated were scanned to see if any vectors beyond the range of the program could be added to the structure and stabilize it. Nonetheless, the tool provided by LPK15V could be useful for future analysis, especially if a sizeable number of stable vectors are beyond the range of the exclusion matrices.)

4. A general note on the use of the program: there are two cases in which the calculation does not work. (1) If a vector exclusion matrix runs out of vectors before it is filled, the program will overwrite itself and print nonsense. This has been provided for to some extent in LPK15: the size of the vector exclusion

matrix can be reduced to as small as 10 x 10, but no smaller. (2) If there are so many vectors that the arrays storing their indices are overwritten. The upper limit is 400 vectors. Both problems can be taken care of by changing card 62 in subroutine LOOK. This card is set up to retain all vectors with a negative free energy. For case (1), vectors with a positive free energy can be artifically retained. For case (2), vectors less stable than a cutoff value can be excluded. Note that SUM of card 62 is the stability number, the negative of the free energy.

```
      PROGRAM DBL(INPUT,OUTPUT,PUNCH,TAPE5=INPUT,TAPE6=OUTPUT,TAPE7=PUNC
     1H,TAPE2,TAPE1)
      INTEGER T
      INTEGER U
      INTEGER E
      INTEGER ELF
      INTEGER WORD
      COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
     1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
      COMMON/B/N,NSSET,NLAST,NVEC(24,4),NMAX(24),ENSSET(24)
      COMMON/C/E(200,200)
      COMMON/D/ WD(165),NB(165),B(165),ISP(165)
      COMMON/E/NEXC(200)
      COMMON/F/NI1,NI2,NJ1,NJ2,NI,NJ,NIMAX,NJMAX,I1,IA,NBAD,IBAD,I,ICASE
      COMMON/P/MAXEN,NCO,BEST5,NASUM
      COMMON/X/SUM3(200),SUM4(200),GUTOT,FERGA
      COMMON/Q/FE(5,5),SMIN1,ENLOOP(90),SMIN2,ILOOK
      COMMON/Z/ENMIN
      COMMON/EX/EXCLP
      COMMON/TIME/ITM1,ITM2
      COMMON/BUG/NDEBUG
      DIMENSION LAB(12)
      DIMENSION V(25),T(25,6),U(25,15)
      DIMENSION LB(165)
      DIMENSION          SIGMAO(90),HSIG(90),HE(5,5),SE(5,5)
      DATA WD1/1HU/,WD2/1HA/,WD3/1HG/,WD4/1HC/,WD5/1HX/,WD6/1HD/,WD7/1HY
     1/,WD8/1HP/
    3 READ(5,2)  (LAB (KL),KL=1,12)
    2 FORMAT(12A6)
      WRITE(6,4)  (LAB(KL),KL=2,12)
    4 FORMAT(2X,11A6)
      IF(LAB(1).EQ.6H CONT )  GO TO 3
      TBV=1./(20.+273.16)
      R=1.985
      RINV=1./1.985
      GUTOT=0
      READ (5,31)  HIN, HGG,HGA,HGU,HGC,HAA
      READ (5,31)   HAU,HAG,HUA,HUG,HCG
      READ (5,31)   EKGGO,EKGAO,EKGUO,EKGCO,EKAAO
      READ (5,31)   EKAUO,EKAGO,EKUAO,EKUGO,EKCGO
      READ(5,31)    (SIGMAO(I),I=5,9)
      READ(5,31)    (HSIG(I),I=5,9)
      DO 8  I=1,3
      IM=5-I
      IMP=IM+1
      SIGMAO(IM)=0.2*SIGMAO(IMP)
    8 CONTINUE
      SIGMAO(1)=10.**(-14)
      HSIG(1)=HSIG(2)=HSIG(3)=HSIG(4)=HSIG(5)
      READ (5,31)   TEMP
      TEMP=TEMP+273.16
      TEMPV=1./TEMP
      RT=R*TEMP
      TBV=1./(273.16+20.)
      READ (5,31)   RGA
      READ(5,1257)  INTLP,ENINT
 1257 FORMAT(I5,F10.5)
      WRITE(6,1258)  INTLP,ENINT
 1258 FORMAT(*  IN THIS CALCULATION, ALL LOOPS WITH*,I3,*  OR MORE LINKS
     1 ARE ASSUMED TO BE INTERIOR LOOPS*,/,*  THE FREE ENERGY OF AN INTE
     1RIOR LOOP (IN CALS/MOLE) IS TAKEN TO BE*,F10.5,//)
```

```
      ENINT=-ENINT
      READ(5,32)  NDEBUG
      READ(5,32)  NUMMCL
31    FORMAT(7F10.5)
32    FORMAT(I5)
      SIGMAO(8)=(SIGMAO(7)+SIGMAO(9))/2.
      HSIG(8)=(HSIG(7)+HSIG(9))/2.
      HSIGLG=HSIG(9)
      DO 33  I=10,90
      HSIG(I)=HSIGLG
      X=I
      F=(X-1.)/X
      F=SQRT(F**3)
      IM=I-1
      SIGMAO(I)=F*SIGMAO(IM)
33    CONTINUE
      DO 34  I=1,90
      SIGMA=SIGMAO(I)*EXP(HSIG(I)*RINV*(T8V-TEMPV))
      ENLOOP(I)=RT*ALOG(SIGMA)
      IF(I.GE.INTLP)  ENLOOP(I)=ENINT
      IF(I.GE.INTLP)  HSIG(I)=0.
34    CONTINUE
      RTT=RINV*(TEMPV-(1./351.16))
      S=EXP(-HIN*RTT)
      FE(1,1)=FE(4,4)=RT*ALOG(EKGGO*(EXP(-HGG*RTT))*S)
      FE(1,2)=FE(3,4)=RT*ALOG(EKGAO*(EXP(-HGA*RTT))*S)
      FE(1,3)=FE(2,4)=RT*ALOG(EKGUO*(EXP(-HGU*RTT))*S)
      FE(1,4)    =      RT*ALOG(EKGCO*(EXP(-HGC*RTT))*S)
      FE(2,1)=FE(4,3)=RT*ALOG(EKAGO*(EXP(-HAG*RTT))*S)
      FE(2,2)=FE(3,3)=RT*ALOG(EKAAO*(EXP(-HAA*RTT))*S)
      FE(2,3)      =    RT*ALOG(EKAUO*(EXP(-HAU*RTT))*S)
      FE(3,2)      =    RT*ALOG(EKUAO*(EXP(-HUA*RTT))*S)
      FE(3,1)=FE(4,2)=RT*ALOG(EKUGO*(EXP(-HUG*RTT))*S)
      FE(4,1)      =    RT*ALOG(EKCGO*(EXP(-HCG*RTT))*S)
      HE(1,1)=HE(4,4)=HGG+HIN
      SE(1,1)=SE(4,4)=(HE(1,1)+FE(1,1))*TEMPV
      HE(1,2)=HE(3,4)=HGA+HIN
      SE(1,2)=SE(3,4)=(HE(1,2)+FE(1,2))*TEMPV
      HE(1,3)=HE(2,4)=HGU+HIN
      SE(1,3)=SE(2,4)=(HE(1,3)+FE(1,3))*TEMPV
      HE(1,4)=HGC+HIN
      SE(1,4)=(HE(1,4)+FE(1,4))*TEMPV
      HE(2,2)=HE(3,3)=HAA+HIN
      SE(2,2)=SE(3,3)=(HE(2,2)+FE(2,2))*TEMPV
      HE(2,3)=HAU+HIN
      SE(2,3)=(HE(2,3)+FE(2,3))*TEMPV
      HE(2,1)=HE(4,3)=HAG+HIN
      SE(2,1)=SE(4,3)=(HE(2,1)+FE(2,1))*TEMPV
      HE(3,2)=HUA+HIN
      SE(3,2)=(HE(3,2)+FE(3,2))*TEMPV
      HE(3,1)=HE(4,2)=HUG+HIN
      SE(3,1)=SE(4,2)=(HE(3,1)+FE(3,1))*TEMPV
      HE(4,1)=HCG+HIN
      SE(4,1)=(HE(4,1)+FE(4,1))*TEMPV
      DO 42  I=1,4
42    FE(I,5)=FE(5,I)=HE(I,5)=HE(5,I)=SE(I,5)=SE(5,I)=0.
      FE(5,5)=300.
      WRITE(6,39)  TEMP
39    FORMAT(*   THE CALCULATION IS FOR TEMPERATURE =*,F7.2,//)
      DO 37  I=1,5
      WRITE(6,281)
281   FORMAT(///,*  BELOW, 1=G, 2=A, 3=U, 4=C*,/)
      DO 38  J=1,5
      FERGA=RT*ALOG(RGA)
      WRITE(6,36)  (I,J,FE(I,J),HE(I,J),SE(I,J))
36    FORMAT(*     I=*,I2,*    J=*,I2,*     FREE ENERGY=*,F12.4,*     ENTHAL
```

```
     1PY=*,F12.4,*      ENTROPY=*,F12.4)
38   CONTINUE


37   CONTINUE
     WRITE(6,46)  FERGA
46   FORMAT(*   GC INITIATION ENERGY=*,F10.3,*    LOOP FREE ENERGIES FO
     1R LOOP SIZE 2=5,90*,//)
     WRITE(6,47)  ((IZ,ENLOOP(IZ)),IZ=2,9)
47   FORMAT(2X,10(I2,F10.3))
     DO 48  J1=1,8
     I1=10.*J1
     I2=I1+9
     WRITE(6,47)  ( (IZ,ENLOOP(IZ)),IZ=I1,I2)
48   CONTINUE
     NDIM=200
     DO 1  IMOLCL=1,NUMMCL
     NASUM=1
     ITM1=0
     ITM2=0
     READ(5,5)   (LAB(I8S),I8S=1,12)
     READ(5,10)  N
     READ(5,5)  EXCLP
     NRD=(N/80) + 1
     IF( (N-80*(NRD-1)) .EQ. 0)  NRD=NRD-1
     DO 25  NR=1,NRD
     NMIN=(NR-1)*80 + 1
     NMX=80*NR
     IF(NR.EQ.NRD)  NMX=N
     READ(5,20)  (WD(I),I=NMIN,NMX)
25   CONTINUE
     MAXEX=1
     MINEX=1
     IF(EXCLP.EQ.6HEXCLP )  MAXEX=2
     IF(EXCLP.EQ.6HEXCLP2)  MINEX=MAXEX=2
     DO 26  IEXC=MINEX,MAXEX
     ICT=0
     EXCLP=0
     IF(IEXC.NE.2)  GO TO 29
     EXCLP=1.
     LAB(8)=6H
     LAB(9)=6HLOOPS
     LAB(10)=6HMAY NO
     LAB(11)=6HT BASE
     LAB(12)=6HPAIR
 29  DO 30 I=1,N
     IF(WD(I).EQ.WD1.OR.WD(I).EQ.WD8)  NB(I)=1
     IF(WD(I).EQ.WD1.OR.WD(I).EQ.WD8)  LB(I)=3
     IF(WD(I).EQ.WD2)  NB(I)=2
     IF(WD(I).EQ.WD2) LB(I)=2
     IF(WD(I).EQ.WD3)  NB(I)=4
     IF(WD(I).EQ.WD3)  LB(I)=1
     IF(WD(I).EQ.WD4)  NB(I)=5
     IF(WD(I).EQ.WD4) LB(I)=4
     IF(WD(I).EQ.WD6.OR.WD(I).EQ.WD7)   NB(I)=9
     IF(WD(I).EQ.WD5)   NB(I)=-6
     IF(NB(I).EQ.9.OR.NB(I).EQ.-6)  LB(I)=3
30   CONTINUE
     WRITE(6,40)  ((WD(I),NB(I)),I=1,N)
     DO 50  I=1,N
     E(I,I)=0
     IM=I-1
     DO 60  J=1,IM
     ELF=NB(I)+NB(J)
     E(I,J)=0
     IF(ELF.EQ.3)  E(I,J)=2
```

```
      IF(ELF.EQ.5)  E(I,J)=1
      IF(ELF.EQ.9)  E(I,J)=4
      IF(ELF.LT.0) E(I,J)=4
      IF(ELF.LT.-3)   E(I,J)=2
      IJ=I-J
      IF(IJ.LE.3)  E(I,JI=0
   60 CONTINUE
   50 CONTINUE
      DO 53  I=1,N
      IM=I-1
      DO 54  J=1,IM
      E(J,I)=E(I,J)
   54 CONTINUE
   53 CONTINUE
   51 NLINE=N/125+1
      WRITE(6,6)    (LAB(IBS),IBS=2,12)
      IF((NLINE-1)*125.EQ.N)  NLINE=NLINE-1
      DO 70  NL=1,NLINE
      NTOT=125+(NL-1)*125
      NMN=125*(NL-1)+1
      IF(NL.EQ.NLINE)  NTOT=N-125*(NLINE-1)
      WRITE(6,80)  (WD(I),I=1,NTOT)
      DO 83  I=NMN,NTOT
      ISP(I)=I-(I/10)*10
   83 CONTINUE
      WRITE(6,85)  (ISP(I),I=1,NTOT)
      J1=(NL-1)*125+1
      J2=J1+125
      IF(NL.EQ.NLINE)  J2=NTOT
      DO 90  J=J1,J2
      IF(ICT.EQ.230)  GO TO 91
      WRITE(6,100)  (WD(J),J,(E(I,J),I=1,J))
      GO TO 90
   91 WRITE(6,101)  (WD(J),J,(E(I,J),I=1,J2))
   90 CONTINUE
      WRITE(6,80)    (WD(I),I=1,NTOT)
      WRITE(6,85)  (ISP(I),I=1,NTOT)
   70 CONTINUE
      IF(ICT.EQ.230)  GO TO 230
      DO 92  I=1,N
      DO 93  J=1,N
      IF(E(I,J).EQ.0)  GO TO 94
      E(I,J)=E(I,J)/2
      GO TO 93
   94 E(I,J)=-1
   93 CONTINUE
   92 CONTINUE
      NCOUNT=0
      NM=N-1
      I=N
      DO 110  J=1,NM
  130 CALL LOOK(0,I,J,0)
      S=SMIN1=SMIN2=0.
      ILOOK=1
      IX=I+1
      JX=J-1
      KF=(I-J-1)/2+1
      DO 120  K=1,KF
      IX=IX-1
      JX=JX+1
      JMX=JX-1
      IPX=IX+1
      IF(K.EQ.1)  SMIN1=0.
      IF(K.EQ.1)   GO TO 134
      KX=LB(JX)
      KMX=LB(JMX)
```

```
      IF(E(IX,JX).EQ.0)  KX=5
      IF(E(IPX,JMX).EQ.0)  KMX=5



      S=S+FE(KMX,KX)
      SMIN2=FE(KMX,KX)
      IF(ILOOK.EQ.0)  SMIN1=FE(KMX,KX)
      ILOOK=1
  134 IF(K.EQ.KF)  GO TO 140
      IF(E(IX,JX).LT.0)  CALL LOOK(1,IX,JX,S)
  120 CONTINUE
      GO TO 110
  140 CALL LOOK(1,IX,JX,S)
  110 CONTINUE
      J=1
      DO 115 ILK=1,NM
      I=NM-ILK+1
      CALL LOOK(0,I,J,0)
      S=0.
      SMIN=0.
      ILOOK=1
      IX=I+1
      JX=J-1
      KF=(I-J-1)/2+1
      DO 125  K=1,KF
      IX=IX-1
      JX=JX+1
      JMX=JX-1
      IPX=IX+1
      IF(K.EQ.1)  SMIN1=0.
      IF(K.EQ.1)  GO TO 124
      KX=LB(JX)
      KMX=LB(JMX)
      IF(E(IX,JX).EQ.0)  KX=5
      IF(E(IPX,JMX).EQ.0)  KMX=5
      S=S+FE(KMX,KX)
      SMIN2=FE(KMX,KX)
      IF(ILOOK.EQ.0)  SMIN1=FE(KMX,KX)
      ILOOK=1
  124 IF(K.EQ.KF)  GO TO 145
      IF(E(IX,JX).LT.0)  CALL LOOK(1,IX,JX,S)
  125 CONTINUE
      GO TO 115
  145 CALL LOOK(1,IX,JX,S)
  115 CONTINUE
  150 WRITE(6,155)
      GO TO 159
  157 WRITE(6,158)
      WRITE(6,6)   (LAB(IBS),IBS=2,12)
      ICT=230
  159 WRITE(6,160)
      IF(NCOUNT.GE.15)  GO TO 168
      NCP=NCOUNT+1
      DO 167 IB=NCP,15
      SUM2(IB)=0.
  167 CONTINUE
  168 DO 170  I=1,NCOUNT
      WRITE(6,180)  I,IX1(I),JX1(I),IX2(I),JX2(I),SUM1(I),SUM2(I)
  170 CONTINUE
      IF(ICT-230)  231,232,231
  232 DO 239  I=1,N
      DO 241  J=1,N
      E(I,J)=1H
  241 CONTINUE
  239 CONTINUE
```

```
      LOUNT=1
      DO 233 I=1,NCOUNT
      IM=I-1
      IF(IM.EQ.0)  GO TO 236
      IF(SUM2(I).EQ.SUM2(IM))  GO TO 236
      LOUNT=LOUNT+1
      IF(LOUNT.GE.13)  LOUNT=13
  236 J1=JX1(I)
      I1=IX1(I)
      KOUNT=0
      JMX=JX2(I)
      DO 234 J=J1,JMX
      IX=I1-KOUNT
      KOUNT=KOUNT+1
      GO TO (1001,1002,1003,1004,1005,1006,1007,1008,1009,1010,1011,1012
     1,1013),LOUNT
  238 E(IX,J)=WORD
      E(J,IX)=WORD
      GO TO 234
 1001 WORD=1HA
      GO TO 238
 1002 WORD=1HB
      GO TO 238
 1003 WORD=1HC
      GO TO 238
 1004 WORD=1HD
      GO TO 238
 1005 WORD=1HE
      GO TO 238
      GO TO 238
 1006 WORD=1HF
 1007 WORD=1HG
      GO TO 238
 1008 WORD=1HH
      GO TO 238
 1009 WORD=1HI
      GO TO 238
 1010 WORD=1HJ
      GO TO 238
 1011 WORD=1HK
      GO TO 238
 1012 WORD=1HL
      GO TO 238
 1013 WORD=1HM
      GO TO 238
  234 CONTINUE
  233 CONTINUE
      WRITE(6,6)   (LAB(IBS),IBS=2,12)
      WRITE(6,242)
      GO TO 51
  231 IZ=0
      DO 200 K=1,NCOUNT
      KM=K-1
      BIGSUM=-1.55
      IZ=IZ+1
      DO 190 J=1,NCOUNT
      IF(IZ.EQ.1)  GO TO 195
      DO 210 L=1,KM
      IF(J.EQ.INDX(L))  GO TO 190
  210 CONTINUE
  195 IF(SUM2(J).LE.BIGSUM)  GO TO 190
C     NOTE THAT THE LE IN CARD 195 INTRODUCES THE BIAS OF FAVORING THE LOWER
C     NUMBERED VECTORS IF THE ENERGIES ARE THE SAME  THE MINIMIZATION
C     SCHEME SHOULD AVOID THIS BIAS AND CAN BE TESTED FOR THIS BY CHANGING
C     THE LE TO LT
      BIGSUM=SUM2(J)
```

```
      INDX(K)=J
  190 CONTINUE



  200 CONTINUE
      NCT=0
      NI1=1
      NI2=2
      NJ1=3
      NJ2=4
      DO 220  I=1,NCOUNT
      GO TO 223
  222 NI1=NI1+4
      NI2=NI2+4
      NJ1=NJ1+4
      NJ2=NJ2+4
      NCT=0
  223 NCT=NCT+1
      IF(NCT.EQ.NDIM+1)     GO TO 222
      INDEX=INDX(I)
      E(NI1,NCT)=IX1(INDEX)
      E(NI2,NCT)=IX2(INDEX)
      E(NJ1,NCT)=JX1(INDEX)
      E(NJ2,NCT)=JX2(INDEX)
      SUM4(I)=SUM2(INDEX)
      SUM3(I)=SUM1(INDEX)
  220 CONTINUE
      NCT=0
      NI1=1
      NI2=2
      NJ1=3
      NJ2=4
      DO 225  I=1,NCOUNT
      GO TO 228
  227 NI1=NI1+4
      NI2=NI2+4
      NJ1=NJ1+4
      NJ2=NJ2+4
      NCT=0
  228 NCT=NCT+1
      IF(NCT.EQ.NDIM+1)   GO TO 227
      IX1(I)=E(NI1,NCT)
      IX2(I)=E(NI2,NCT)
      JX1(I)=E(NJ1,NCT)
      JX2(I)=E(NJ2,NCT)
      SUM1(I)=SUM3(I)
      SUM2(I)=SUM4(I)
      NDIF(I)=IX1(I)-IX2(I)+1
      NDIF1=JX2(I)-JX1(I)+1
      IF(NDIF(I).NE.NDIF1)  CALL COMENT(1)
  225 CONTINUE
      GO TO 157
  230 CONTINUE
      REWIND 1
      MAXEN=0
      ICASE=2
      DO 1233 I=1,NCOUNT
      SUM1(I)=SUM2(I)
 1233 SUM2(I)=0.
      CALL INSPECT
      CALL LPK
      IF(LAB(1).EQ.6HNONE  )  CALL LPK15
      IF(LAB(1).EQ.6HVARY15)  CALL LPK15V
      REWIND 1
      WRITE(6,6)   (LAB(IBS),IBS=2,12)
```

```
      WRITE(6,245)  MAXEN,ENMIN,NCO
      MAX=MAXEN/25
      MXLST=25
      IF((25*(MAXEN/25)).LT.MAXEN)    MAX=MAX+1
      DO 251  L1=1,MAX
      IF(L1.EQ.MAX)  MXLST=NDIF(1)
      READ (1)  ((V(MX),(T(MX,IZ),IZ=1,NASUM),(U(MX,IZ),IZ=1,NCO)),MX=1,
     1MXLST)
      DO 254  MX=1,MXLST
      WRITE(6,255)  (V(MX),(T(MX,IZ),IZ=1,NASUM),(U(MX,IZ),IZ=1,NCO))
  254 CONTINUE
  251 CONTINUE
      REWIND 1
   26 CONTINUE
    1 CONTINUE
      CALL EXIT
    5 FORMAT(12A6)
    6 FORMAT(//,*   THIS CALCULATION IS ON THE *,3A6,*MOLECULE*,/,2X,8A6
     1,//)
   10 FORMAT(I5)
   20 FORMAT(80A1)
   40 FORMAT(15(X,A2,I2,2X))
   80 FORMAT(7X,125A1)
   85 FORMAT(7X,125I1)
  100 FORMAT(1X,A1,1X,I3,1X,125I1)
  101 FORMAT(1X,A1,1X,I3,1X,125A1)
  155 FORMAT(//,*   UNREORDERED VECTORS *)
  158 FORMAT(//,*  VECTORS REORDERED ACCORDING TO ENERGY *)
  160 FORMAT(/,*  VECTOR NUM   I1   J1   I2   J2   ENERGYWITH LOOP ENERGY*,/
     1Y*,/)
  180 FORMAT(5X,I3,6X,I3,2X,I3,2X,I3,2X,I3,3X,2F6.2)
  242 FORMAT(///,*  GRAPH SHOWING VECTORS ORDERED ACCORDING TO ENERGY*)
  245 FORMAT(///,*  THE FINAL RESULTS ARE REPORTED .  THERE ARE *,I4,*
     1SETS OF VECTORS WHOSE ENERGY *,/,*  IS GREATER THAN*,F6.2,*.   THI
     2S IS AN N=*,I3,* CALCULATION*,//,                                *
     3  ENERGY    V1  V2  V3  V4  V5  V6  V7  V8  V9 V10 V11 V12 V13 V14
     4 V15 V16*)
  255 FORMAT(5X,F6.2,2X,26(2X,I2))
  400 FORMAT(2X,I4,2X,F6.2,4I5)
      END
      SUBROUTINE TIMER(IPLACE)
      COMMON/TIME/ITM1,ITM2
      DIMENSION ITIME(7)
      CALL STATUS(ITIME(1),ITIME(2))
      IDELT1=ITIME(1)-ITM1
      IDELT2=ITIME(2)-ITM2
      FTIME1=ITIME(1)*0.001
      FTIME2=ITIME(2)*0.001
      FDEL1=IDELT1*0.001
      FDEL2=IDELT2*0.001
      WRITE(6,26)  IPLACE,FTIME1,FTIME2
   26 FORMAT(/,*  IPLACE=*,I3,/,*   CP TIME ELAPSED IS*,F10.5,*    PP TI
     1ME ELAPSED IS*,F10.5)
      WRITE(6,28)   FDEL1,FDEL2
   28 FORMAT(*  CP TIME SINCE LAST CALL IS*,F10.5,*        PP TIME SINCE
     1 LAST CALL IS*,F10.5)
      ITM1=ITIME(1)
      ITM2=ITIME(2)
      RETURN
      END

      SUBROUTINE INSPECT
      COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
     1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
      COMMON/B/N,NSSET,NLAST,NVEC(24,4),NMAX(24),ENSSET(24)
      COMMON/C/E(200,200)
```

```
      COMMON/D/ WD(165),NB(165),B(165),ISP(165)
      COMMON/E/NEXC(200)


      COMMON/F/NI1,NI2,NJ1,NJ2,NI,NJ,NIMAX,NJMAX,I1,IA,NBAD,IBAD,I,ICASE
      INTEGER E
C     THIS SUBROUTINE LOOKS AT THE NCOUNT VECTORS AND DETERMINES THE SEQUENCE
C     OF VECTORS WHICH ARE COMPATIBLE (HAVE NO OVERLAP AS DEFINED IN S.R. BAD)
C     WITH VECTOR NUMBER 1, 2, 3, ..... ,NCOUNT RESPECTIVELY
      IF(NCOUNT.GE.90)  NCOUNT=90
      NBAD=0
      IBAD=0
      NLAST=0
      NCT=0
      DO 10  I1=1,NCOUNT
      E(I1,I1)=0
      NSSET=I1
      INDSS=1
      NI1=IX1(I1)
      NJ1=JX1(I1)
      NI2=IX2(I1)
      NJ2=JX2(I1)
      DO 20  IA=1,NCOUNT
      IF(I1.EQ.IA)  GO TO 20
      NI=IX1(IA)
      NJ=JX1(IA)
      NJMAX=JX2(IA)
      NIMAX=IX2(IA)
      CALL COMPAR
 20   CONTINUE
 10   CONTINUE
      WRITE(6,50)
      GO TO (2,3,4),ICASE
 2    WRITE(6,6)
      GO TO 9
 3    WRITE(6,7)
      GO TO 9
 4    WRITE(6,8)
 9    WRITE(6,60)   (ISP(I),I=1,NCOUNT)
      DO 30  I=1,NCOUNT
      WRITE(6,70)  (I,SUM2(I),NEXC(I),(E(I,J),J=1,NCOUNT))
 30   CONTINUE
      CALL RERITE
  6   FORMAT(*        MATRIX ALLOWS NO OVERLAP   0=ALLOWED VECTORS  1=EXC
     1LUDED VECTORS*,//)
  7   FORMAT(*     MATRIX ALLOWS OVERLAP OF ONE ELEMENT  0=ALLOWED VECTOR
     1S     1=ONE OVERLAP      2=TWO OVERLAPS *,//)
  8   FORMAT(*    MATRIX ALLOWS THE ENDS OF THE VECTORS TO OVERLAP
     1 0=NO OVERLAP    1=ENDS OVERLAP    2=OVERLAP (EXCLUSION)  *,//)
 50   FORMAT(///,*   VECTOR EXCLUSION MATRIX FOR ORDERED VECTORS*,//,*
     1      0=INCLUDED VECTORS    1=EXCLUDED VECTORS  *,//)
 60   FORMAT(*   ENERGY  EXC  *,114I1)
 70   FORMAT(1X,I2,1X,F6.2,2X,I2,2X,114I1)
      RETURN
      END
      SUBROUTINE LOOK(M,IX,JX,S)
      COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
     1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
      COMMON/B/N,NSSET,NLAST,NVEC(24,4),NMAX(24),ENSSET(24)
      COMMON/D/ WD(165),NB(165),B(165),ISP(165)
      COMMON/Q/FE(5,5),SMIN1,ENLOOP(90),SMIN2,ILOOK
      COMMON/X/SUM3(200),SUM4(200),GUTOT,FERGA
      IF(M.EQ.0)  GO TO 1
      MLOOP=IX-JX+2
 12   IF(MLOOP.LE.0)  MLOOP=1
```

```
      IF(MLOOP.GE.90)    MLOOP=90.
      IF1=JXLST
      IF2=JX-1
      IFG=0
      DO 6  IFF=IF1,IF2
      IF(WD(IFF).EQ.1HG.OR.WD(IFF).EQ.1HC)  IFG=1
      IF(IFG.EQ.1)  GO TO 7
    6 CONTINUE
    7 SUM=S-SMIN1-SMIN2+FERGA*IFG+ENLOOP(MLOOP)
      ILOOK=0
   62 IF(SUM.GT.(0.))  GO TO 2
C1    THE ABOVE CARD REPRESENTS THE ENERGY CUTOFF WHICH ELIMINATES FROM
C2    CONSIDERATION THOSE DOUBLE STRANDED REGIONS OF INSUFFICIENT ENERGY
C3    TO BIND
      IXLST=IX-1
      JXLST=JX+1
      S=0.
      RETURN
    2 NCOUNT=NCOUNT+1
      IX1(NCOUNT)=IXLST
      JX1(NCOUNT)=JXLST
      IX2(NCOUNT)=IX+1
      JX2(NCOUNT)=JX-1
      SUM1(NCOUNT)=.001*(SUM-ENLOOP(MLOOP))
      SUM2(NCOUNT)=.001*SUM
   64 GUTOT=0
      IXLST=IX-1
      JXLST=JX+1
      S=0.
      RETURN
    1 IXLST=IX
      JXLST=JX
      RETURN
      END
      SUBROUTINE COMPAR
      COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
     1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
      COMMON/B/N,NSSET,NLAST,NVEC(24,4),NMAX(24),ENSSET(24)
      COMMON/C/E(200,200)
      COMMON/E/NEXC(200)
      COMMON/F/NI1,NI2,NJ1,NJ2,NI,NJ,NIMAX,NJMAX,I1,IA,NBAD,IBAD,I,ICASE
      COMMON/EX/EXCLP
      INTEGER E
      IF(EXCLP.NE.1)  GO TO 5
      EX=0
      IF(NI.GT.NJ1.AND.NI.LT.NI1)  EX=EX+1
      IF(NJ.GT.NJ1.AND.NJ.LT.NI1)  EX=EX+1
      IF(NJMAX.GT.NJ1.AND.NJMAX.LT.NI1)  EX=EX+1
      IF(NIMAX.GT.NJ1.AND.NIMAX.LT.NI1)  EX=EX+1
      IF(EX.EQ.4.OR.EX.EQ.0)  GO TO 5
      E(I1,IA)=2
      RETURN
    5 NDIF1=NJ2
      NDIFA=NJMAX
      IF(NSSET.NE.NLAST)  GO TO 100
   30 NLAST=NSSET
      DO 10 I=1,N
      JNDX(I)=0
   10 CONTINUE
      KCT=0
      DO 20  J=NJ,NDIFA
      I=NI-KCT
      KCT=KCT+1
      JNDX(I)=1
      JNDX(J)=1
   20  CONTINUE
```

```
      DO 40  I=1,N
      IF((INDX(I)+JNDX(I)).EQ.2)   CALL BAD


      IF(IBAD-1)  40,60,60
40    CONTINUE
      IF(NBAD.EQ.1)  GO TO 50
      E(I1,IA)=0
      SUM2(NSSET)=SUM2(NSSET)+SUM1(IA)
      NEXC(NSSET)=NEXC(NSSET)+1
60    IBAD=0
      RETURN
50    E(I1,IA)=1
      NBAD=0
      RETURN
100   DO 110 I=1,N
      INDX(I)=0
110   CONTINUE
      KCT=0
      DO 120  J=NJ1,NDIF1
      I=NI1-KCT
      KCT=KCT+1
      INDX(I)=1
      INDX(J)=1
120   CONTINUE
      SUM2(NSSET)=SUM1(NSSET)
      NEXC(NSSET)=0
      GO TO 30
      END
      SUBROUTINE BAD
      COMMON/F/NI1,NI2,NJ1,NJ2,NI,NJ,NIMAX,NJMAX,I1,IA,NBAD,IBAD,I,ICASE
      COMMON/C/E(200,200)
      INTEGER E
      IF(ICASE-2)  10,20,30
10    E(I1,IA)=1
      IBAD=1
      RETURN
 20   CALL BAD2
      RETURN
30    CALL BAD3
      RETURN
      END
      SUBROUTINE BAD2
      COMMON/F/NI1,NI2,NJ1,NJ2,NI,NJ,NIMAX,NJMAX,I1,IA,NBAD,IBAD,I,ICASE
      COMMON/C/E(200,200)
      INTEGER E
      NBAD=NBAD+1
      IF(NBAD.EQ.2)  GO TO 10
      RETURN
10    E(I1,IA)=2
      NBAD=0
      IBAD=1
      RETURN
      END
      SUBROUTINE BAD3
      COMMON/F/NI1,NI2,NJ1,NJ2,NI,NJ,NIMAX,NJMAX,I1,IA,NBAD,IBAD,I,ICASE
      COMMON/C/E(200,200)
      INTEGER E
      IF(I.EQ.NI.OR.I.EQ.NJ.OR.I.EQ.NIMAX.OR.I.EQ.NJMAX)  GO TO 10
      IF(I.EQ.NI1.OR.I.EQ.NJ1.OR.I.EQ.NI2.OR.I.EQ.NJ2)  GO TO 10
      IBAD=1
      E(I1,IA)=2
      RETURN
10    NBAD=NBAD+1
      IF(NBAD.GE.2)   GO TO 20
```

```
      RETURN
20    WRITE(6,30)  NBAD,I1,IA
30    FORMAT(*   3RD. EXCLUSION CRITERION  END OVERLAPS IN THIS UNUSUAL
     1CASE*,I3,*   TIMES FOR THE VECTORS*,I3,*  AND*,I3)
      RETURN
      END
      SUBROUTINE LPK
      COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
     1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
      COMMON/C/E(200,200)
      COMMON/P/MAXEN,NCO,BEST5,NASUM
      COMMON/BUG/NDEBUG
      DIMENSION NEWE(70,70),ILP(50),IAA(6),LABEL(6),I1POS(32),I2POS(32)
     1,I3POS(32),I4POS(32),I5POS(32)
      DIMENSION S(25),T(25,6),U(25,15)
      INTEGER E
      BEST5=0.0
      MX=0
      IF(NDEBUG.EQ.0)  GO TO 6
      WRITE(6,5)
6     DO 10 I=1,5
      NCT=0
      DO 20  J=1,NCOUNT
      IF(E(I,J).EQ.1)  GO TO 20
      NCT=NCT+1
      INDX(NCT)=J
      IF(I.EQ.J)  NORDER=NCT
20    CONTINUE
      DO 30  I1=1,NCT
      DO 40  I2=1,NCT
      I11=INDX(I1)
      I22=INDX(I2)
      NEWE(I1,I2)=E(I11,I22)
40    CONTINUE
30    CONTINUE
      IF(NDEBUG.EQ.0)  GO TO 82
      WRITE(6,50)  I
      WRITE(6,55)
      DO 60  K=1,NCT
      ILP(K)=INDX(K)-10*(INDX(K)/10)
60    CONTINUE
      WRITE(6,70)  (ILP(K),K=1,NCT)
      DO 80 K=1,NCT
      WRITE(6,90)   (INDX(K),(NEWE(K,L),L=1,NCT))
80    CONTINUE
82    DO 100 L=1,5
      MM=L
      IF(L.GE.NORDER)  MM=L+1
      IAA(L)=MM
100   CONTINUE
      I1X=IAA(1)
      I2X=IAA(2)
      I3X=IAA(3)
      I4X=IAA(4)
      I5X=IAA(5)
      NTWON=0
      DO 110  J1=1,2
      DO 120  J2=1,2
      DO 130  J3=1,2
      DO 140  J4=1,2
      DO 150  J5=1,2
      LABEL(1)=J1-1
      LABEL(2)=J2-1
      LABEL(3)=J3-1
      LABEL(4)=J4-1
      LABEL(5)=J5-1
```

```
      DO 160  MDL=1,5
      ML=IAA(MDL)


      IF(LABEL(MDL).NE.1)  GO TO 160
      NEW=LABEL(1)*NEWE(I1X,ML)+LABEL(2)*NEWE(I2X,ML)+LABEL(3)*NEWE(I3X,
     1ML)+LABEL(4)*NEWE(I4X,ML)+LABEL(5)*NEWE(I5X,ML)
      IF(NEW.GE.1)  GO TO 150
  160 CONTINUE
      NTWON=NTWON+1
      I1POS(NTWON)=LABEL(1)
      I2POS(NTWON)=LABEL(2)
      I3POS(NTWON)=LABEL(3)
      I4POS(NTWON)=LABEL(4)
      I5POS(NTWON)=LABEL(5)
      K1X=INDX(I1X)
      K2X=INDX(I2X)
      K3X=INDX(I3X)
      K4X=INDX(I4X)
      K5X=INDX(I5X)
      SUM2(NTWON)=LABEL(1)*SUM1(K1X)+LABEL(2)*SUM1(K2X)+LABEL(3)*SUM1(K3
     1X)+LABEL(4)*SUM1(K4X)+LABEL(5)*SUM1(K5X)+SUM1(I)
      IF(BEST5.LT.SUM2(NTWON))  BEST5=SUM2(NTWON)
  150 CONTINUE
  140 CONTINUE
  130 CONTINUE
  120 CONTINUE
  110 CONTINUE
      IF(NDEBUG.EQ.0)  GO TO 10
      DO 170  LK=1,NTWON
       WRITE(6,175)  LK
      WRITE(6,180)   I1POS(LK),I2POS(LK),I3POS(LK),I4POS(LK),I5POS(LK),
     1SUM2(LK)
  170 CONTINUE
   10 CONTINUE
    5 FORMAT(//,*   THE ASSUMPTION IS (TEMPORARILY) MADE THAT ONE GIVEN
     1VECTOR IS INCLUDED IN THE FINAL SOLUTION OF THE PROBLEM*,/,*   ALL
     2 COMBINATIONS OF THE (REMAINING BEST) FIVE VECTORS ARE CONSIDERED
     3AND THOSE NOT EXCLUDED BY THE E MATRIX *,/,*   ARE REPORTED.  THE
     4ENERGY IS ALSO REPORTED FOR EACH ALLOWED COMBINATION OF VECTORS*)
   50 FORMAT(////,*  THE VECTOR WHICH IS HERE ASSUMED TO BE IN THE MOST
     1STABLE SET OF VECTORS IS*,I3,/)
   55 FORMAT(/,*  THE NEW VECTOR EXCLUSION MATRIX IS REPORTED*,/)
   70 FORMAT(6X,6012)
   90 FORMAT(2X,I2,2X,6012)
  175 FORMAT(*  VECTOR SET NUMBER*,I3,/,8X,*V1   V2   V3   V4   V5     E
     1NERGY*)
  180 FORMAT(8X,I2,3X,I2,3X,I2,3X,I2,3X,I2,5X,F6.2)
      RETURN
      END
      SUBROUTINE LPK15
      COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
     1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
      COMMON/C/E(200,200)
      COMMON/L/ILP(165),IAA(15),LABEL(15)
      COMMON/P/MAXEN,NCO,BEST5,NASUM
      COMMON/Z/ENMIN
      COMMON/BUG/NDEBUG
      DIMENSION SUM3(200)
      DIMENSION NPOS(25)
      DIMENSION S(25),T(25,6),U(25,15)
      DIMENSION INDNEW(25)
      DIMENSION NEWE(70,70)
      DIMENSION MDOSTP(15)
      INTEGER T
```

```
      INTEGER U
      INTEGER E
      IF(NCOUNT.GT.70)  NCOUNT=70
      SUM1(40)=0.
      IF(NDEBUG.EQ.0)  GO TO 7
      WRITE(6,5)
 7    IF(NCOUNT.GT.15)  GO TO 4
      NCPLUS=NCOUNT+1
      DO 3 I=NCPLUS,15
 3    SUM1(I)=0.
 4    ENMIN=0.9*BEST5
      NUM=7
      NCO=15
      MX=0
      NCOP=NCO+1
      IZCT=0
      DO 10  I=1,NUM
      DO 6  KK=5,15
      MDOSTP(KK)=2
 6    CONTINUE
      NCT=0
      DO 20  J=1,NCOUNT
      IF(E(I,J).EQ.1)  GO TO 20
      NCT=NCT+1
      INDX(NCT)=J
      IF(I.EQ.J)  NORDER=NCT
20    CONTINUE
      NCP=NCT+1
      IF(NCT.GE.15)  GO TO 22
      DO 21  J=NCP,16
21    INDX(J)=40
22    CONTINUE
      IF(NDEBUG.EQ.0)   GO TO 23
      WRITE(6,25)  (NCO,NUM,(INDX(KAPPA),KAPPA=1,NCT))
23    IF(NCT.GE.40)  NCT=39
      DO 30  I1=1,NCT
      DO 40  I2=1,NCT
      I11=INDX(I1)
      I22=INDX(I2)
      NEWE(I1,I2)=E(I11,I22)
40    CONTINUE
30    CONTINUE
      IF(NDEBUG.EQ.0)  GO TO 82
      WRITE(6,50)  I
      WRITE(6,55)
      DO 60  K=1,NCT
      ILP(K)=INDX(K)-10*(INDX(K)/10)
60    CONTINUE
      WRITE(6,70)   (ILP(K),K=1,NCT)
      DO 80  K=1,NCT
      WRITE(6,90)   (INDX(K),(NEWE(K,L),L=1,NCT))
80    CONTINUE
82    DO 100 L=1,NCO
      MM=L
      IF(L.GE.NORDER)  MM=L+1
      IAA(L)=MM
100   CONTINUE
      I1X=IAA(1)
      I2X=IAA(2)
      I3X=IAA(3)
      I4X=IAA(4)
      I5X=IAA(5)
      I6X=IAA(6)
      I7X=IAA(7)
      I8X=IAA(8)
      I9X=IAA(9)
```

```
      I10=IAA(10)
      I11=IAA(11)



      I12=IAA(12)
      I13=IAA(13)
      I14=IAA(14)
      I15=IAA(15)
      NTWON=0
      KT=0
      DO 151  IF=1,NCOP
      IF(INDX(IF).EQ.I)  GO TO 151
      KT=KT+1
      INDNEW(KT)=INDX(IF)
151   CONTINUE
      IF(NDEBUG.EQ.0)  GO TO 157
      WRITE(6,155)
      WRITE(6,156)  (INDNEW(IF),IF=1,NCT)
157   IF(NCT.GT.15)  GO TO 105
      DO 107  KK=NCT,15
      MDOSTP(KK)=1
107   CONTINUE
105   MAX5=MDOSTP(5)
      MAX6=MDOSTP(6)
      MAX7=MDOSTP(7)
      MAX8=MDOSTP(8)
      MAX9=MDOSTP(9)
      MAX10=MDOSTP(10)
      MAX11=MDOSTP(11)
      MAX12=MDOSTP(12)
      MAX13=MDOSTP(13)
      MAX14=MDOSTP(14)
      MAX15=MDOSTP(15)
      DO 110  J1=1,2
      LABEL(1)=J1-1
      DO 120  J2=1,2
      LABEL(2)=J2-1
      DO 130  J3=1,2
      LABEL(3)=J3-1
      DO 140  J4=1,2
      LABEL(4)=J4-1
      DO 150  J5=1,MAX5
      LABEL(5)=J5-1
      DO 260  J6=1,MAX6
      LABEL(6)=J6-1
      DO 270  J7=1,MAX7
      LABEL(7)=J7-1
      DO 280  J8=1,MAX8
      LABEL(8)=J8-1
      DO 290  J9=1,MAX9
      LABEL(9)=J9-1
      DO 300  J10=1,MAX10
      LABEL(10)=J10-1
      DO 310  J11=1,MAX11
      LABEL(11)=J11-1
      DO 320  J12=1,MAX12
      LABEL(12)=J12-1
      DO 330  J13=1,MAX13
      LABEL(13)=J13-1
      DO 340  J14=1,MAX14
      LABEL(14)=J14-1
      DO 350  J15=1,MAX15
      LABEL(15)=J15-1
      DO 160  MDL=1,NCO
      IF(LABEL(MDL).NE.1)  GO TO 160
```

```
      ML=IAA(MDL)
      NEW=LABEL(1)*NEWE(I1X,ML)+LABEL(2)*NEWE(I2X,ML)+LABEL(3)*NEWE(I3X,
     1ML) +LABEL(4)*NEWE(I4X,ML)+LABEL(5)*NEWE(I5X,ML)+LABEL(6)*NEWE(I6X
     2,ML)+LABEL(7)*NEWE(I7X,ML)+LABEL(8)*NEWE(I8X,ML)+LABEL(9)*NEWE(I9X
     3,ML)+LABEL(10)*NEWE(I10,ML)+LABEL(11)*NEWE(I11,ML)+LABEL(12)*NEWE(
     4I12,ML)+LABEL(13)*NEWE(I13,ML)+LABEL(14)*NEWE(I14,ML)+LABEL(15)*NE
     5WE(I15,ML)
      IF(NEW.GE.1)  GO TO 350
  160 CONTINUE
      NTWON=NTWON+1
      K1X=INDX(I1X)
      K2X=INDX(I2X)
      K3X=INDX(I3X)
      K4X=INDX(I4X)
      K5X=INDX(I5X)
      K6X=INDX(I6X)
      K7X=INDX(I7X)
      K8X=INDX(I8X)
      K9X=INDX(I9X)
      K10=INDX(I10)
      K11=INDX(I11)
      K12=INDX(I12)
      K13=INDX(I13)
      K14=INDX(I14)
      K15=INDX(I15)
      SUM3(NTWON)=LABEL(1)*SUM1(K1X)+LABEL(2)*SUM1(K2X)+LABEL(3)*SUM1(K3
     1X)+LABEL(4)*SUM1(K4X)+LABEL(5)*SUM1(K5X)+LABEL(6)*SUM1(K6X)+LABEL(
     27)*SUM1(K7X)+LABEL(8)*SUM1(K8X)+LABEL(9)*SUM1(K9X)+LABEL(10)*SUM1(
     3K10)+LABEL(11)*SUM1(K11)+LABEL(12)*SUM1(K12)+LABEL(13)*SUM1(K13)+L
     4ABEL(14)*SUM1(K14)+LABEL(15)*SUM1(K15)+SUM1(I)
      IF(SUM3(NTWON).LT.ENMIN)  GO TO 345
      IF(NDEBUG.EQ.0)  GO TO 499
      WRITE(6,180)    (NTWON,SUM3(NTWON),(LABEL(KMN),KMN=1,NCO))
  181 IZCT=IZCT+1
      IF(IZCT.LT.30)  GO TO 499
      IZCT=0
      WRITE(6,155)     I
      WRITE(6,156)  (INDNEW(IF),IF=1,NCO)
  499 IF(SUM3(NTWON).LT.ENMIN)  GO TO 350
      DO 500  IF=1,NCO
      JF=IAA(IF)
      NPOS(IF)=LABEL(IF)*INDX(JF)
  500 CONTINUE
      MAXEN=MAXEN+1
      MX=MX+1
      S(MX)=SUM3(NTWON)
      DO 503  IZ=1,NASUM
      T(MX,IZ)=I
  503 CONTINUE
      DO 504  IZ=1,NCO
      U(MX,IZ)=NPOS(IZ)
  504 CONTINUE
      IF(MX.LT.25)   GO TO 350
      WRITE(1)  ((S(MX),(T(MX,IZ),IZ=1,NASUM),(U(MX,IZ),IZ=1,NCO)),MX=1,
     125)
      MX=0
      GO TO 350
  345 NTWON=NTWON-1
  350 CONTINUE
  340 CONTINUE
  330 CONTINUE
  320 CONTINUE
  310 CONTINUE
  300 CONTINUE
  290  CONTINUE
  280 CONTINUE
```

```
  270   CONTINUE
  260   CONTINUE



  150   CONTINUE
  140   CONTINUE
  130   CONTINUE
  120   CONTINUE
  110   CONTINUE
  10    CONTINUE
        NDIF(1)=MX
        MXLST=MX
        WRITE(1)  ((S(MX),(T(MX,IZ),IZ=1,NASUM),(U(MX,IZ),IZ=1,NCO)),MX=1,
       1MXLST)
  5     FORMAT(//,*   THE ASSUMPTION IS (TEMPORARILY) MADE THAT ONE GIVEN
       1VECTOR IS INCLUDED IN THE FINAL SOLUTION OF THE PROBLEM*,/,*   ALL
       2 COMBINATIONS OF THE (REMAINING BEST) FIVE VECTORS ARE CONSIDERED
       3AND THOSE NOT EXCLUDED BY THE E MATRIX *,/,*   ARE REPORTED.  THE
       4ENERGY IS ALSO REPORTED FOR EACH ALLOWED COMBINATION OF VECTORS*)
  25    FORMAT(/,*  ALL COMBINATIONS OF THE FIRST*,I3,* VECTORS ARE CONSID
       1ERED WHERE ONE GIVEN VECTOR OF*,I3,/,*  IS ASSUMED TO BE IN THE FI
       2NAL CONFIGURATION.  WITH THIS ASSUMPTION THE REDUCED EXCULSION MAT
       3RIX*,/,*   CONTAINS THE FOLLOWING VECTORS*,/,40(1X,I2),/,40(1X,I2)
       4,/,40(1X,I2),/,40(1X,I2))
  50    FORMAT(////,*  THE VECTOR WHICH IS HERE ASSUMED TO BE IN THE MOST
       1STABLE SET OF VECTORS IS*,I3,/)
  55    FORMAT(/,*  THE NEW VECTOR EXCLUSION MATRIX IS REPORTED*,/)
  70    FORMAT(6X,60I2)
  90    FORMAT(2X,I2,2X,60I2)
  155   FORMAT(/,5X,*ENERGY     V1  V2  V3  V4  V5  V6  V7  V8  V9 V10 V11
       1 V12 V13 V14 V15     INCLUDED VECTOR=*,I3)
  156   FORMAT(14X,25(2X,I2))
  180   FORMAT(*  VECTOR SET*,I4,/,5X,F6.2,3X,25(2X,I2))
        RETURN
        END
        SUBROUTINE LPK15V
        COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
       1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
        COMMON/C/E(200,200)
        COMMON/L/ILP(165),IAA(15),LABEL(15)
        COMMON/P/MAXEN,NCO,BEST5,NASUM
        COMMON/Z/ENMIN
        DIMENSION SUM3(200)
        DIMENSION NPOS(25)
        DIMENSION INDNEW(25)
        DIMENSION ICHOSE(10),NORDER(10)
        DIMENSION NEWE(165,165)
        DIMENSION MDOSTP(15)
        DIMENSION S(25),T(25,6),U(25,15)
        INTEGER E
        INTEGER T
        INTEGER U
        EQUIVALENCE(E,NEWE)
C1      FINAL VALUE OF NCT = NUMBER OF ELEMENTS IN THE REDUCED EXCLUSION MATRIX
C2      INDX(I),I=1,NCT = ABSOLUTE INDEX (AS OUTPUTTED UNDER HEADING REORDERED
C3      VECTORS)  OF ALL VECTORS WHICH ARE IN REDUCED EXCLUSION MATRIX
C4      IAA(I),I=1,NCO = INDEX (RELATIVE TO POSITION IN REDUCED EXCLUSION MATRIX)
C5      OF FIRST NCO VECTORS WITH THOSE VECTORS ASSUMED TO BE IN THE FINAL
C6      ARRANGEMENT NOT INCLUDED IN THIS LIST
C7       I1X,...,I9X,I10,...INCO CORRESPOND TO IAA(1),.....IAA(9),IAA(10),...
C8      K1X,.....,K10,...KNCO CORRESPOND TO INDX(I1X),.....INDX(I10),.....
C9      K1X, .... = ABSOLUTE INDEX OF THE FIRST NCO VECTORS IN REDUCED EXCLUSION
C10     MATRIX NOT INCLUDING THE CHOSEN (ASSUMED) VECTOR(S)
        ENMIN=0.9*BEST5
        NASUM=3
```

```
      MX=0
      IZCT=0
      READ(5,2)   NDATA
      NCO=15
      NCOP=NCO+1
      WRITE(6,5)   NASUM,NCO
      REWIND 2
      DO 1020 IM=1,NCOUNT
      WRITE(2)  (E(IM,JM),JM=1,NCOUNT)
 1020 CONTINUE
      DO 1  NTOT=1,NDATA
      DO 93  IS=10,15
      MOOSTP(IS)=2
 93   CONTINUE
      NLAST=NCOUNT
      DO 4  J=1,NCOUNT
      INDX(J)=J
 4    CONTINUE
      READ(5,3)  ICHOSE(1),ICHOSE(2),ICHOSE(3)
  3   FORMAT(3I5)
      I1CH=ICHOSE(1)
      I2CH=ICHOSE(2)
      I3CH=ICHOSE(3)
      SUMT=SUM1(I1CH)+SUM1(I2CH)+SUM1(I3CH)
      DO 10  ICH=1,NASUM
      DO 11 ILK=1,NCO
      IF(INDX(ILK).EQ.ICHOSE(ICH))  GO TO 12
 11   CONTINUE
 12   I=ILK
      NCT=0
      DO 20  J=1,NLAST
      JNDX(J)=INDX(J)
      IF(E(I,J).EQ.1)  GO TO 20
      NCT=NCT+1
      INDX(NCT)=INDX(J)
C     NCT LESS THAN OR EQUAL TO J
 20   CONTINUE
      IF(ICH.EQ.1)  GO TO 31
      I11=0
      DO 30  I1=1,NCT
 35   I11=I11+1
      IF(INDX(I1).NE.JNDX(I11))  GO TO 35
      I22=0
      DO 40  I2=1,NCT
 45   I22=I22+1
      IF(INDX(I2).NE.JNDX(I22))  GO TO 45
      NEWE(I1,I2)=E(I11,I22)
C     I1 LE I11 AND I2 LE I22
 40   CONTINUE
 30   CONTINUE
      GO TO 49
 31   DO 41  I1=1,NCT
      DO 42  I2=1,NCT
      I11=INDX(I1)
      I22=INDX(I2)
      NEWE(I1,I2)=E(I11,I22)
C     I1 LESS THAN OR EQUAL TO I11 AND I2 LESS THAN OR EQUAL TO I22
 42   CONTINUE
 41   CONTINUE
 49   WRITE(6,50)  (ICHOSE(IL),IL=1,NASUM)
      WRITE(6,55)
      DO 60  K=1,NCT
      ILP(K)=INDX(K)-10*(INDX(K)/10)
 60   CONTINUE
      WRITE(6,70)   (ILP(K),K=1,NCT)
      DO 80  K=1,NCT
```

```
      WRITE(6,90)    (INDX(K),(NEWE(K,L),L=1,NCT))
80    CONTINUE


      NLAST=NCT
10    CONTINUE
      IF(NLAST.GE.(NCO+NASUM))  GO TO 84
      ND=NLAST-NASUM
      DO 83   IHP=ND,NCO
      MDOSTP(IHP)=1
83    CONTINUE
84    MAX10=MDOSTP(10)
      MAX11=MDOSTP(11)
      MAX12=MDOSTP(12)
      MAX13=MDOSTP(13)
      MAX14=MDOSTP(14)
      MAX15=MDOSTP(15)
      DO 85   IX=1,NCT
      DO 86   IY=1,NASUM
      IF(INDX(IX).EQ.ICHOSE(IY))    NORDER(IY)=IX
86    CONTINUE
85    CONTINUE
      NHOPE=0
      DO 100  L=1,NCO
101   NHOPE=NHOPE+1
      DO 102  IRS=1,NASUM
      IF(NHOPE.EQ.NORDER(IRS))  GO TO 101
102   CONTINUE
      IAA(L)=NHOPE
100   CONTINUE
      I1X=IAA(1)
      I2X=IAA(2)
      I3X=IAA(3)
      I4X=IAA(4)
      I5X=IAA(5)
      I6X=IAA(6)
      I7X=IAA(7)
      I8X=IAA(8)
      I9X=IAA(9)
      I10=IAA(10)
      I11=IAA(11)
      I12=IAA(12)
      I13=IAA(13)
      I14=IAA(14)
      I15=IAA(15)
      NTWON=0
      KT=0
      DO 151  IF=1,NCO
      JF=IAA(IF)
      INDNEW(IF)=INDX(JF)
      IF(IF.GT.(NLAST-NASUM))  INDNEW(IF)=0
151   CONTINUE
      WRITE(6,155)     (ICHOSE(IF),IF=1,NASUM)
      WRITE(6,156)  (INDNEW(IF),IF=1,NCO)
      DO 110  J1=1,2
      LABEL(1)=J1-1
      DO 120  J2=1,2
      LABEL(2)=J2-1
      DO 130  J3=1,2
      LABEL(3)=J3-1
      DO 140  J4=1,2
      LABEL(4)=J4-1
      DO 150  J5=1,2
      LABEL(5)=J5-1
      DO 260  J6=1,2
      LABEL(6)=J6-1
```

```
      DO 270  J7=1,2
      LABEL(7)=J7-1
      DO 280  J8=1,2
      LABEL(8)=J8-1
      DO 290  J9=1,2
      LABEL(9)=J9-1
      DO 300  J10=1,MAX10
      LABEL(10)=J10-1
      DO 310  J11=1,MAX11
      LABEL(11)=J11-1
      DO 320  J12=1,MAX12
      LABEL(12)=J12-1
      DO 330  J13=1,MAX13
      LABEL(13)=J13-1
      DO 340  J14=1,MAX14
      LABEL(14)=J14-1
      DO 350  J15=1,MAX15
      LABEL(15)=J15-1
      DO 160  MDL=1,NCO
      IF(LABEL(MDL).NE.1)  GO TO 160
      ML=IAA(MDL)
      NEW=LABEL(1)*NEWE(I1X,ML)+LABEL(2)*NEWE(I2X,ML)+LABEL(3)*NEWE(I3X,
     1ML) +LABEL(4)*NEWE(I4X,ML)+LABEL(5)*NEWE(I5X,ML)+LABEL(6)*NEWE(I6X
     2,ML)+LABEL(7)*NEWE(I7X,ML)+LABEL(8)*NEWE(I8X,ML)+LABEL(9)*NEWE(I9X
     3,ML)+LABEL(10)*NEWE(I10,ML)+LABEL(11)*NEWE(I11,ML)+LABEL(12)*NEWE(
     4I12,ML)+LABEL(13)*NEWE(I13,ML)+LABEL(14)*NEWE(I14,ML)+LABEL(15)*NE
     5WE(I15,ML)
      IF(NEW.GE.1)  GO TO 350
  160 CONTINUE
      NTWON=NTWON+1
      K1X=INDX(I1X)
      K2X=INDX(I2X)
      K3X=INDX(I3X)
      K4X=INDX(I4X)
      K5X=INDX(I5X)
      K6X=INDX(I6X)
      K7X=INDX(I7X)
      K8X=INDX(I8X)
      K9X=INDX(I9X)
      K10=INDX(I10)
      K11=INDX(I11)
      K12=INDX(I12)
      K13=INDX(I13)
      K14=INDX(I14)
      K15=INDX(I15)
      SUM3(NTWON)=LABEL(1)*SUM1(K1X)+LABEL(2)*SUM1(K2X)+LABEL(3)*SUM1(K3
     1X)+LABEL(4)*SUM1(K4X)+LABEL(5)*SUM1(K5X)+LABEL(6)*SUM1(K6X)+LABEL(
     27)*SUM1(K7X)+LABEL(8)*SUM1(K8X)+LABEL(9)*SUM1(K9X)+LABEL(10)*SUM1(
     3K10)+LABEL(11)*SUM1(K11)+LABEL(12)*SUM1(K12)+LABEL(13)*SUM1(K13)+L
     4ABEL(14)*SUM1(K14)+LABEL(15)*SUM1(K15)+SUMT
      IF(SUM3(NTWON).LT.ENMIN)  GO TO 345
      WRITE(6,180)    (NTWON,SUM3(NTWON),(LABEL(KMN),KMN=1,NCO))
      IZCT=IZCT+1
      IF(IZCT.LT.30)  GO TO 499
      IZCT=0
      WRITE(6,155)     (ICHOSE(IF),IF=1,NASUM)
      WRITE(6,156)  (INDNEW(IF),IF=1,NCO)
  499 IF(SUM3(NTWON).LT.BEST5)  GO TO 350
      DO 500  IF=1,NCO
      JF=IAA(IF)
      NPOS(IF)=LABEL(IF)*INDX(JF)
  500 CONTINUE
      MAXEN=MAXEN+1
      MX=MX+1
      S(MX)=SUM3(NTWON)
      DO 503  IZ=1,NASUM
```

```
      T(MX,IZ)=ICHOSE(IZ)
  503 CONTINUE



      DO 504  IZ=1,NCO
      U(MX,IZ)=NPOS(IZ)
  504 CONTINUE
      IF(MX.LT.25)  GO TO 350
      WRITE(1)  ((S(MX),(T(MX,IZ),IZ=1,NASUM),(U(MX,IZ),IZ=1,NCO)),MX=1,
     125)
      MX=0
      GO TO 350
  345 NTWON=NTWON-1
  350 CONTINUE
  340 CONTINUE
  330 CONTINUE
  320 CONTINUE
  310 CONTINUE
  300 CONTINUE
  290  CONTINUE
  280 CONTINUE
  270  CONTINUE
  260 CONTINUE
  150 CONTINUE
  140 CONTINUE
  130 CONTINUE
  120 CONTINUE
  110 CONTINUE
      REWIND 2
      DO 1010 IM=1,NCOUNT
      READ(2)  (E(IM,JM),JM=1,NCOUNT)
 1010 CONTINUE
    1 CONTINUE
      WRITE(1)  ((S(MX),(T(MX,IZ),IZ=1,NASUM),(U(MX,IZ),IZ=1,NCO)),MX=1,
     1MXLST)
    2 FORMAT(I5)
    5 FORMAT(//,*  THE ASSUMPTION IS MADE THAT*,I2,*  VECTORS ARE INCLUD
     1ED IN THE FINAL SOLUTION OF THE PROBLEM*,/,*   ALL COMBINATIONS OF
     2 THE (REMAINING BEST)*,I3,*  VECTORS ARE CONSIDERED AND THOSE NOT
     3EXCLUDED BY THE E MATRIX*,/,*   ARE REPORTED.          *)
   50 FORMAT(////,*  THE VECTORS WHICH ARE HERE ASSUMED TO BE IN THE MOS
     1T STABLE SET OF VECTORS ARE*,10I3,/)
   55 FORMAT(/,*  THE NEW VECTOR EXCLUSION MATRIX IS REPORTED*,/)
   70 FORMAT(6X,120I1)
   90 FORMAT(2X,I2,2X,120I1)
  155 FORMAT(/,5X,*ENERGY     V1  V2  V3  V4  V5  V6  V7  V8  V9 V10 V11
     1 V12 V13 V14 V15     INCLUDED VECTORS=*,6I3)
  156 FORMAT(14X,25(2X,I2))
  180 FORMAT(*  VECTOR SET*,I4,/,5X,F6.2,3X,25(2X,I2))
      RETURN
      END
      SUBROUTINE RERITE
      COMMON/A/NCOUNT,IX1( 400),IX2( 400),JX1( 400),JX2( 400),SUM1( 400)
     1,NDIF( 400),INDX( 400),SUM2( 400),JNDX( 400)
      COMMON/C/E(200,200)
      COMMON/E/NEXC(200)
      COMMON/D/ WD(165),NB(165),B(165),ISP(165)
      COMMON/F/NI1,NI2,NJ1,NJ2,NI,NJ,NIMAX,NJMAX,I1,IA,NBAD,IBAD,I,ICASE
      INTEGER E
      IF(ICASE.NE.2)  GO TO 100
      DO 10 I=1,NCOUNT
      DO 20 J=1,NCOUNT
      IF(E(I,J).EQ.0)  GO TO 20
      E(I,J)=E(I,J)-1
   20 CONTINUE
```

```
10    CONTINUE
      WRITE(6,50)
       WRITE(6,60)  (ISP(I),I=1,NCOUNT)
      DO 30  I=1,NCOUNT
      WRITE(6,70)  (I,SUM2(I),NEXC(I),(E(I,J),J=1,NCOUNT))
30    CONTINUE
50    FORMAT(*  THE VECTOR EXCLUSION MATRIX IS REWRITTEN FOR THE PURPOSE
     1S OF THE SEARCH ROUTINES*,/,*      0=NO OVERLAP (IE. NO FATAL OVERL
     2AP)     1=OVERLAP (EXCLUSION)*)
60    FORMAT(*    ENERGY  EXC  *,114I1)
 70   FORMAT(1X,I2,1X,F6.2,2X,I2,2X,114I1)
100   RETURN
      END
      SUBROUTINE COMENT(KLM)
      GO TO (1,2,3,4,5,6,7),KLM
1     WRITE(6,10)
      RETURN
2     CONTINUE
      RETURN
3     CONTINUE
      RETURN
4     CONTINUE
      RETURN
5     CONTINUE
      RETURN
6     CONTINUE
      RETURN
7     CONTINUE
      RETURN
10    FORMAT(*  NDIF AND NDIF1 ARE NOT EQUAL AT 225 --SHOULD BE F.E.*)
      END
```

TECHNICAL INFORMATION DIVISION
LAWRENCE BERKELEY LABORATORY
UNIVERSITY OF CALIFORNIA
BERKELEY, CALIFORNIA 94720