^{1}H-NMR STUDIES OF SEQUENCE AND CONFORMATION USING SHORT RNA OLIGOMERS

By

PAUL ANTHONY HADER, B.Sc.

A Thesis

Submitted to the School of Graduate Studies in

Partial Fulfilment of the Requirements

for the Degree

Doctor of Philosophy

McMaster University

September, 1982

RNA SEQUENCE STUDIES

DOCTOR OF PHILOSOPHY (1983) McMASTER UNIVERSITY
(Biochemistry) Hamilton, Ontario

TITLE:

^{1}H-NMR Studies of Sequence and Conformation Using Short RNA Oligomers

AUTHOR:

Paul Anthony Hader, B.Sc. (McMaster University)

SUPERVISOR:

DR. Thomas Neilson

NUMBER OF PAGES:

xv , 190

ABSTRACT

The sequence and conformation of short RNA strands was investigated by proton nuclear magnetic resonance spectroscopy. This physical technique gives conformational information about RNA strands in solution at the nucleoside level. The short oligoribonucleotides used in these studies were chemically synthesized by the phosphotriester method developed by T. Neilson and associates, which is capable of preparing a wide variety of sequences in the quantities required for NMR spectroscopic experiments, and the many in vitro termination assays. The work presented in this thesis was the first to demonstrate that NMR spectral information can be used to sequence RNA oligomers; and that sequence and conformation around termination codons can redefine their activity *in vitro*.

A set of parameters has been developed that provides a new and accurate mathematical approach for the assignment of the NMR chemical shifts for heterobase ring and anomeric protons in oligoribonucleotides. The set of empirical shift parameters was derived from the chemical shifts of RNA oligomers and can also be applied to DNA species. The accuracy of the parameters is such that they can be used to accurately determine the sequence of RNA oligomers.

The specificity of *in vitro* prokaryotic chain

termination was examined as a function of the sequence neighbouring UAA. RF-1 dependent termination was assayed from the release of N-acetyl or N-formyl-methionine from initiation complexes containing N-acetyl- or N-formyl-met-tRNA:ribosomes:AUG or AUG $(N)_m$. In the AUG case, a second oligomer was added to cause release.

When added to the AUG-bound intermediate, hexamer, UAAUAG, containing tandem termination codons, was slightly more effective in the release of N-acetyl-met-tRNA by RF-1 than UAA, UAAN (where N is A, G, C, or U), and UAAUGA whereas UAAUAA, was least effective. Oligomers, AUGUA, AUGUAA, AUGUUA, AUGUUAA, AUGUUAU, and $AUG(U)_mUAA(A)_{18-25}$ (where m = 1-5) stimulated binding of N-acetyl-met-tRNA and release of N-acetyl-met by RF-1, whereas AUGCUA, AUGCUAA and $AUGUAA(A)_n$, were less effective.

The sequence UA acts as release signal for RF-1, its activity varying with mRNA sequence context. The conformation surrounding UA is responsible which is dependent upon the nature of adjacent nucleoside residues. This UA recognition is competed for by cognate aminoacyl-tRNAs.

Some of the results presented in this thesis have appeared in the literature:

1. Alkema, D., Bell, R., Hader, P. A., and Neilson, T. (1981) in "Biomolecular Stereodynamics" Ed., Sarma, R. H.,

Adenine Press, NY.

2. Alkema, D., Bell, R., Hader, P. A., and Neilson, T. (1981) J. Am. Chem. Soc. 103, 2866-2868.

3. Alkema, D., Hader, P. A., Bell, R. A., and Neilson, T. (1981) Biochemistry 21, 2109-2117.

4. Hader, P. A., Neilson, T., Alkema, D., Kofoid, E. C., and Ganoza, M. C. (1981) FEBS Letters 136, 65-69.

5. Hader, P. A., Alkema, D., Bell, R. A., and Neilson, T. (1982) J. Chem. Soc., Chem Commun. 1017.

ACKNOWLEDGEMENTS

I would like to thank Dr. Thomas Neilson for his guidance and enthusiasm. I am grateful to the members of my supervisory committee, Dr. Russell Bell and Dr. Gerhard E. Gerber for their helpful advice.

I particularly wish to thank my wife Karen for performing most of the termination assays and for helping to put this thesis in presentable form. Special thanks also go to Drs. D. Alkema, J. Coddington, and D. W. Hughes (the "Top Guns"), for their collaboration on the NMR work.

I thank the Natural Sciences and Engineering Research Council of Canada for the financial support of a Postgraduate Scholarship.

TABLE OF CONTENTS

ABBREVIATIONS

A[a]	Adenosine
Å	angstrom (10^{-8} cm)
$AgNO_3$	silver nitrate
C	cytidine
c	total concentration
CD	circular dichroism
CH_3OH	methanol
Cu/Zn	copper-zinc couple
dA	2'-deoxyadenosine
dC	2'-deoxycytidine
dG	2'-deoxyguanosine
DMF	N,N-dimethylformamide
DNA	deoxyribonucleic acid
DSS	4,4-dimethyl-4-silapentane-1-sulfonate
E. coli	Escherichia coli
f[^{35}S]Met-tRNA	N-formyl-[^{35}S]-methionyl transfer RNA
G	guanosine
g	gauche
10,000g	10,000 times the force of gravity
^1H	hydrogen (proton)
Hz	hertz (cycles/second)
J	coupling constant
KOH	potassium hydroxide
MST	mesitylenesulfonyl 1,2,4-triazole
mRNA	messenger RNA

NH_4OH	ammonium hydroxide
NMR	nuclear magnetic resonance
NOE	nuclear overhauser enhancement
OD	optical density
ppm	parts per million
PPO	2,5-diphenyloxazole
POPOP	p-bis-[2-5-phenyloxazolyl]-benzene
R_f	ratio of distance travelled by solute to that of solvent
RNA	ribonucleic acid
rRNA	ribosomal RNA
T	thymidine
t	trans
thp	tetrahydropyranyl
tlc	thin layer chromatography
Tm	melting temperature of a duplex at which 50% of the strands exist in the duplex form
trac	trityloxacetyl
tRNA	transfer RNA
U	uridine
UV	ultraviolet

---

[a]Monoucleotides and oligoribonucleotides are abbreviated in the standard format (IUPAC-IUB Commission on Biochemical Nomenclature, 1970). pA represents 5'-adenylic acid, Cp represents 3'-cytidylic acid and CpA represents cytidyly (3'-5') adenosine. Unless otherwise noted, oligoribonucleotides are numbered sequentially from the 5' terminus.

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Chemical synthesis of oligoribonucleotides has been applied to various problems of biochemical interest. The objective of this work is to demonstrate how chemically synthesized oligoribonucleotides may be used to investigate aspects of conformation and sequence in short RNA oligomers. Specifically, it will be shown that; (i) when RNA oligomers are in a highly flexible conformational state the ^{1}H-NMR chemical shifts can be used to sequence the oligomer, (ii) the sequence and conformation of oligomers can effect the termination process in protein biosynthesis.

1.1 The Chemical Synthesis of Oligoribonucleotides.

Chemically synthesized oligoribonucleotides have played a major role in many biochemical studies. Synthetic oligoribonucleotides were used in experiments that elucidated the triplet nature of the genetic code (Nirenberg and Leder, 1964). Chemically synthesized oligodeoxynucleotides have been used to synthesize complete genes (Khorana, et al., 1976; Edge, et al., 1981; News and Views, 1981). Chemically synthesized DNA fragments are also used to induce site specific mutations (Ciampi, et al., 1982).

Synthetic oligoribonucleotides have been used as substrates by the RNA ligase enzyme to build the precistronic region of bacteriophage Qβ A protein (Neilson, et al., 1979),

and to alter tRNAs so that structure function relationships could be investigated (Bruce and Uhlenbeck, 1978, 1981; Gumport and Uhlenbeck, 1980). Various short oligomers that mimic regions of mRNA have been synthesized to investigate sequence effects on translational initiation in vitro (Ganoza, et al., 1978; Neilson, et al., 1979; Ganoza, et al., 1982).

Successful chemical synthesis of oligoribonucleotides remains a challenging problem. For instance no automated procedure for RNA synthesis has been developed, while at least two such proceedures are available for DNA synthesis (Alvarado-Urbina, et al., 1981; Gait and Sheppard, 1977; Gait, et al., 1982; Tundo and Venturello, 1979).

Before a specific 3' to 5' internucleotide phosphodiester linkage can be formed, other equally reactive sites on the nucleosides must be made chemically inert by prior reaction with blocking compounds. Then after the internucleotide linkages are formed, these blocking groups must be removed without chemically modifying the nucleosides, or disrupting the phosphodiester linkages. Work in the field has been divided between the development of appropriately blocked nucleoside precursors and of suitable phosphorylating methods (Ikakura, et al., 1977; Reese, et al., 1978).

Basically, two approaches to the chemical synthesis of oligoribonucleotides have been used, the phosphodiester

method and the phosphotriester method. Of the two methods the phosphodiester method which has been used successfully, particularly for the synthesis of oligodeoxynucleotides is largely falling into disuse. The approach has several pitfalls: the unprotected phosphate linkage is subjected to branching or cleavage during chain extension; the charged phosphodiester intermediates become increasingly insoluble in organic solvents, which limits the scale of preparation; and the phosphodiester linkage is very sensitive to acid and base, which can lead to problems during final purifications. Protection of the phosphodiester linkage provided a solution to all of these problems.

1.1.1 Blocking Groups Used in Oligoribonucleotide Synthesis.

The various functional sites that are required to be masked at various times during synthesis are the primary 5'-hydroxyl, the secondary 3' (in ribose, 2')-hydroxyl, the primary amino groups of cytosine, adenine and guanine, and the phosphodiester. Blocking groups must be easily introduced onto the nucleoside, and be removable under conditions which do not alter other blocked sites. Conditions for removal of the blocking groups should not alter the phosphodiester or glycosidic linkages of the free oligonucleotide.

1.1.2 Blocking of the Primary Amino Groups of the Heterobases.

Acyl groups, whose use were introduced in the mid-sixties are the method of choice used to protect the primary amino function. The acyl group is stable to neutral, acidic and mildly basic pH conditions, used to deprotect other functions selectively. The acyl group is removed by ammonolysis (conc. NH_4OH/CH_3OH, 1:1 (v/v)).

1.1.3 The 2'-Hydroxyl Blocking Group.

During the deblocking of the 2'-hydroxyl function of oligoribonucleotides it is essential that the conditions do not disturb the phosphodiester linkages. Mainly phosphodiester isomerization must be avoided in all phases of oligoribonucleotide synthesis, since it is not possible to separate the desired product from those containing some 2'-5' linkages. The tetrahydropyranyl group has been a popular choice as a 2' blocking group. The group is easily removed in three to four hours upon treatment with 0.01 M HCl at room temperature, conditions which do not allow phosphodiester isomerization. The only drawback is that the point of attachment on the tetrahydropyranyl ring is asymmetric leading to a diastereomeric mixture of the modified nucleoside; however, the diastereomeric mixtures

can be separated by silica gel chromatography. Another factor in favour of the tetrahydropyranyl group is that it is bulky, and prevents 3'-3' phosphodiester formation, thus eliminating the need for separate 3'-hydroxyl blocking (Neilson, 1969).

1.1.4 Blocking of the Terminal Primary 5'-Hydroxyl Group.

The blocking group chosen to protect the primary 5'-hydroxyl plays the major role in block synthesis of longer oligoribonucleotides. In a block synthesis, trimers for example are covalently joined to yield a hexanucleotide. The incoming block must have a free 5'-hydroxyl, and all other reactive groups blocked. Thus, in block synthesis it is necessary to remove the 5' blocking group selectively, and in high yield leaving all other protecting groups intact. The trityloxyacetyl (trac) group is ideal, since it can be removed under mild base conditions (Werstiuk and Neilson, 1973). The trac group's selectivity towards primary hydroxyls makes it a useful group in synthesis of oligodeoxyribonucleotides; however the commonly used primary blocking group is the monomethoxytrityl (MMTr), or dimethoxytrityl, which can be specifically removed with mild acid (Narang, et al., 1980).

1.1.5 The Phosphotriester Blocking Group.

The triester group used in synthesis of

oligoribonucleotides for this work was the 2,2,2-trichlorethyl group. The group is stable to the mild base conditions used to selectively deprotect 5' blocking groups.

The 2,2,2-trichlorethyl group is removed by reductive cleavage of the O-alkyl bond by treatment with copper-zinc couple in anhydrous dimethylformide. Cleavage at the O-alkyl bond is extremely important since it removes the possibility of internucleotide bond cleavage, which can occur if blocking group removal is via the O-phosphoryl bond, a side reaction that can occur with the para chlorophenol blocking group.

The versatile phosphotriester approach to the chemical synthesis of oligoribonucleotides, as developed by Neilson and associates is perhaps best illustrated by considering that all 64 trinucleoside diphosphates and at least 100 other oligomers of defined sequence have been prepared by this method (Neilson, et al., 1982).

1.2 Conformational Properties of Single Stranded Oligoribonucleotides and Polyribonucleotides.

The present understanding of the conformational properties of single stranded polynucleotides has evolved from solid state (X-ray crystallography), semi-empirical conformational energy calculations, and solution state studies employing UV, CD, fluorescence and NMR as tools. The

seven dihedral torsion angles with rotational freedom in a nucleotidyl unit (Figure 1), makes complete understanding of RNA conformation very difficult. By comparison, polysaccharides and polypeptides have only two bonds with rotational freedom in their backbone structures, and even in these cases the factors governing conformation are not completely understood (Gutte, et al., 1979).

1.2.1 Solid State Structure of RNA Strands.

The use of X-ray crystallography in the determination of the conformations adopted by single and multistranded nucleic acids is classified into two branches based on sample preparation; single crystals and fibres. Molecules within a single crystal form a three dimensional array. Mathematical least squares type analysis of the diffraction pattern yields the unique position of each atom in the crystal (Kim and Quigley, 1979). The resolution of this technique depends on the size of the molecule and the number of reflections recorded. Resolution for an oligoribonucleotide may reach 0.01 angstrom for a dinucleoside monophosphate (Kallenbach and Berman, 1977), but is generally in the range 2-3 angstrom for oligomers (Conner, et al., 1982).

Polynucleotides do not form single crystals instead they form fibres. The fibres form ordered structures which will give results similar in quality to single crystals

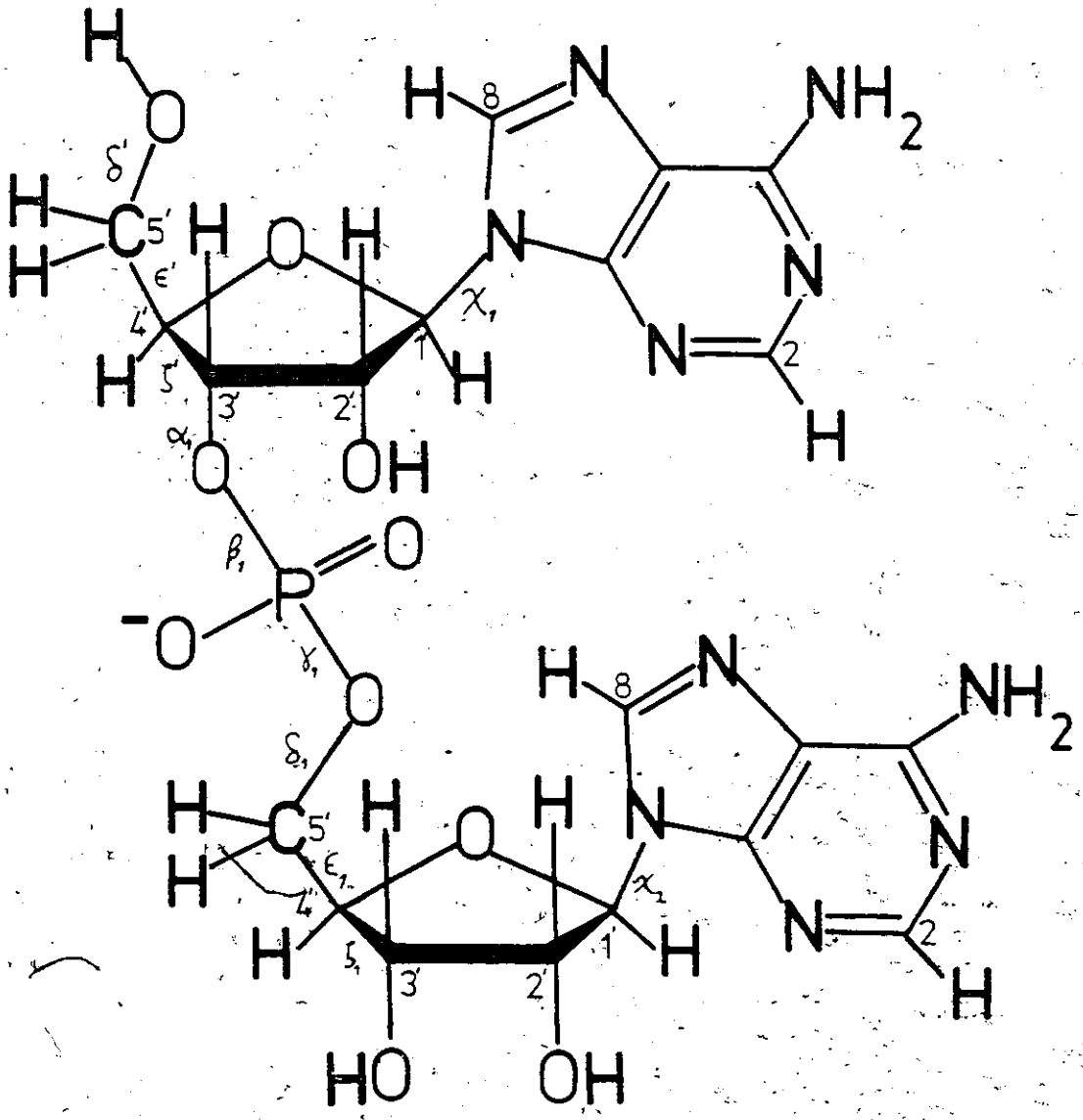


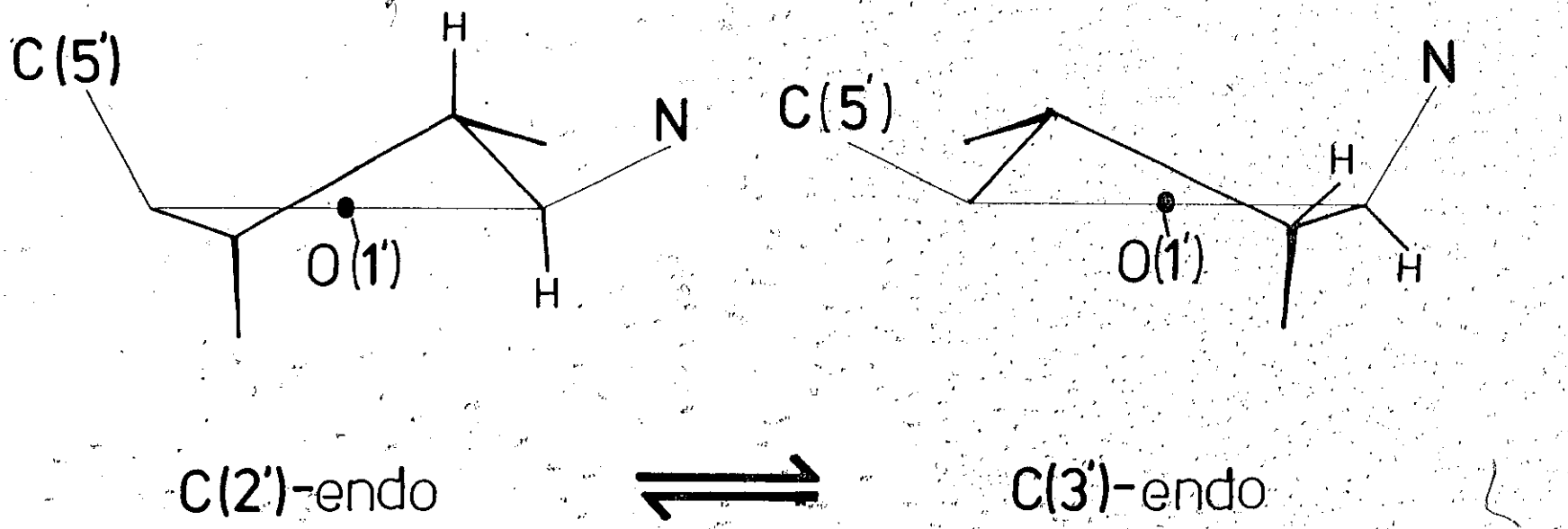
Figure 1. Structure of 3'-5' ApA, the numbering scheme and proposed IUPAC-IUB nomenclature for the various torsion angles (Sarma, 1980).

(Conner, et al., 1982). Usually the minimum amount of information realized from fibre crystallography defines the helical parameters of pitch, radius, rise per residue, and helical rotation. Combining these data with the more intimate details derived from oligomer crystals, it is possible to generate a detailed conformation from the polymer (Kallenbach and Berman, 1977).

Single strand and double stranded nucleic acid helices belong to one of three classes; A, B, or Z (Dickerson, et al., 1982). Most of the torsion angle differences between the helical classes are small, especially between A and B helices; however, the helix types are overall quite different. The base tilt and base overlaps, important components of base stacking, are different for A, B, and Z. The A helix contains the C3'-endo sugar pucker, the B helix the C2'-endo pucker, and the Z helix an alternating C3'-, C2'-endo pucker (Figure 2). The glycosyl torsion angle is also quite different for the three helices (Davies and Zimmerman, 1980; Conner, et al., 1982).

Studies on double stranded RNAs demonstrated that the ribose ring is in the C3'-endo conformer (Arnott, et al., 1967; Arnott and Hukins, 1972; Holbrook, et al., 1981). In an RNA-DNA hybrid, an A type hybrid is formed even though DNA usually exists in B type duplexes (Milman, et al., 1967; Broyde and Hingerty, 1978). X-ray crystallographic evidence of B type RNA helices has not been found, and along with other factors such as hydration, it is likely that the 2'-hydroxyl is partly

Figure 2. The two major conformations of the sugar ring.



$$10(10 - J_{1',2'}) = \%\ C(3')\text{-endo}$$

responsible for the observed C3'-endo preference giving the A helix (Kallenbach and Berman, 1977; Clementi and Corongiu, 1981).

X-ray studies on nucleotides and nucleosides led to the idea of a rigid conformation for the nucleoside torsion angles (Sundaralingam, 1969). The sugar pucker presumably can exist only as C3'-endo or C2'-endo. This requirement means that there is very little flexibility about the C3'-C4' bond (ψ'), also the other torsion angles show very little variance, and it has been proposed that most of the different conformations for ribonucleotides result from rotation about the P-O bonds (ω, ω') and the C3'-C4' (ψ') bond (Figure 3), (Sundaralingam, 1975, 1981; Lee, et al., 1976; Evans and Sarma 1976).

Recent studies which compared experiment with theory demonstrated that the pseudorotation of ribose and deoxyribose sugar moieties is restricted. Both the computed potential energies and the cumulative X-ray data show that the pentose oscillates between n and s puckerings via the e quadrant of intermediate states rather than over the entire range of pseudorotational space (Figure 4). Furthermore theory that best reproduces NMR coupling constants suggest that the potential barrier in the e (nomenclature of Olson, 1982), quadrant lies between 2 and 5 kcal/mol (Olson, 1982; Olson and Sussman 1982). These energies are high enough to suppress intermediate forms (suggested by the lack of intermediates found in crystals), but low enough to permit the rapid

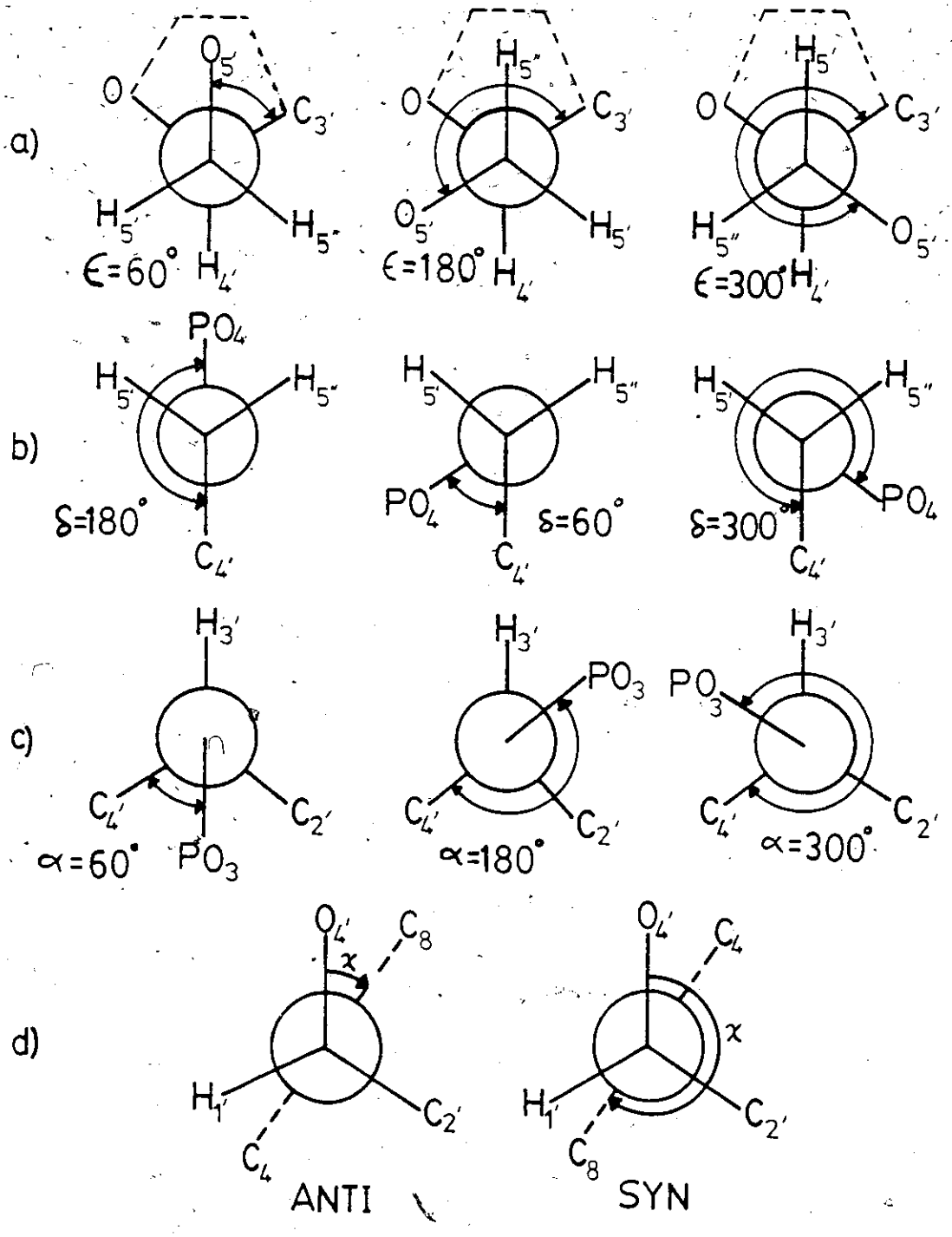


Figure 3. Schematic representation of the bond angles about the a) C(4')-C(5') (ε) bond, b) C(5')-O(5') (δ) bond, c) C(3')-O(3') (α) bond and d) the glycosidic linkage (χ) (Sarma, 1980).

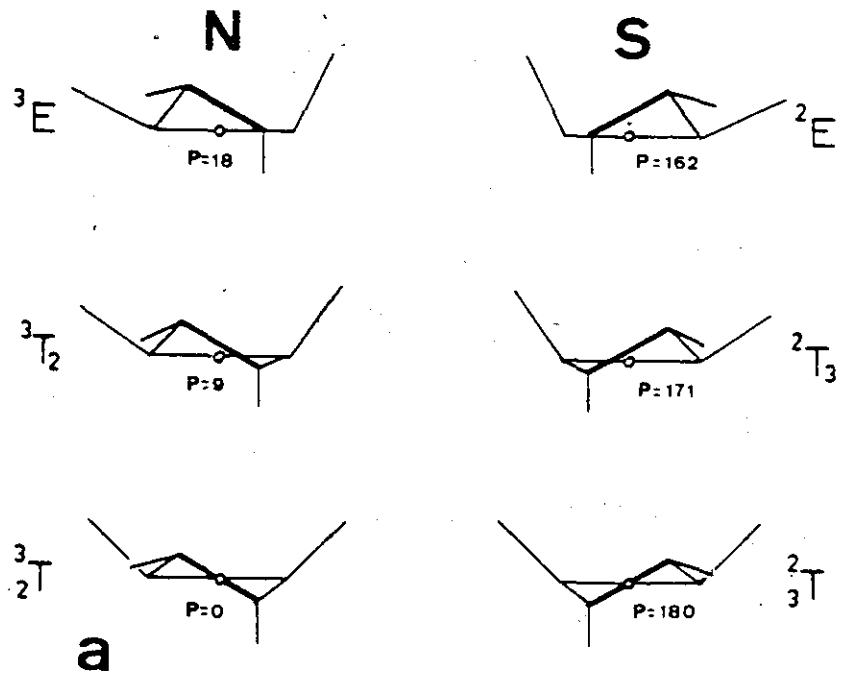


Figure 4a. Schematic presentation of the main conformations N and S of the sugar moiety of nucleosides in solution. The phase angles P are indicated for each conformation.

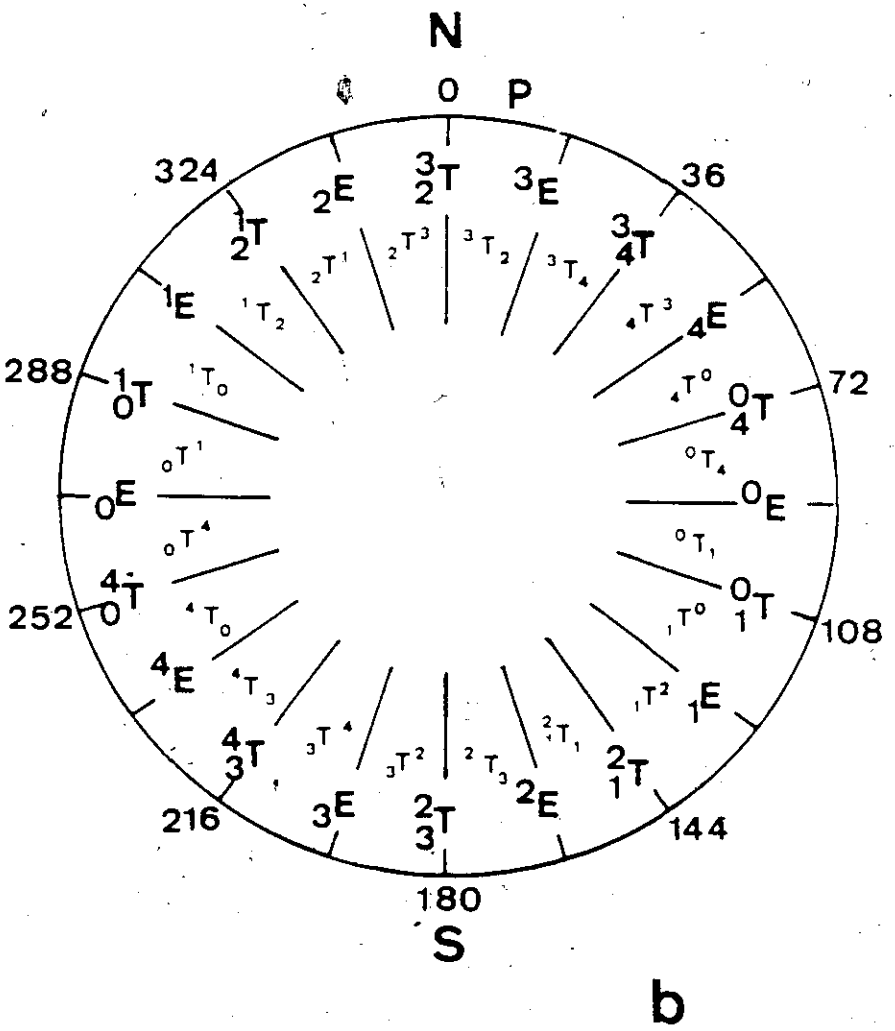


Figure 4b. Pseudorotation cycle indicating the path of pseudorotation through all possible sugar conformations. Each sugar conformation along this path is characterized by the phase angle P indicated.

interconversion of n and s states (suggested by NMR studies). Nevertheless the rigid nucleoside concept, especially in solution, remains controversial at the present time.

1.2.2 Torsion Angles and Conformation Derived from Minimum Energy Calculation.

Potential energy functions have been developed to estimate the pseudorotational motions of ribose and 2'-deoxyribose sugars. These calculatiions are semiempirical in nature, and the total potential energy is usually calculated as a sum of nonbonded, torsional, and valence angle strain contributions (Olson, 1978, 1982; Olson and Sussman, 1982). In general torsion angles (χ,ψ, ψ',ϕ, ϕ',ω, ω') are varied while other bond lengths, bond angles and atomic coordinates are derived from X-ray (Levitt and Warshel, 1978). Minor deviations of valence bond angles from standard X-ray values are found to perturb the normal motions of the furanose drastically (Olson and Sussman 1982). Minimum energy conformations are derived from starting conformations arrived at by using various combinations of the preferred values for the seven dihedral angles, known from X-ray and NMR studies. Energy calculations proceed while the torsion angles are varied from their starting values through the pseudorotational cycle.

Recent calculations have confirmed the hindered ring flexibilities suggested by solid-state and solution studies of low molecular weight nucleic acid analogues. These

results also suggest that the distinct conformational differences of ribose and 2'-deoxyribose stem from the combinations of nonbonded forces and gauche interactions associated with the two sugars. Steric interactions combined with the strong tendencies of the O-C-C-O bond sequences to adopt gauche rotational arrangements induce the ribose to adopt n- or s-type puckerings. A proton in the 2'-deoxyribose introduces a strong imbalance in the gauche interactions, and the single O-C-C-O sequence in 2'-deoxyribose pushes the n/s equilibrium firmly into the s-type puckerings. Thus the gauche effect appears to be a major determinant of ring puckering (Olson, 1982b).

1.2.3 NMR Studies on the Solution Conformations of Oligonucleotides

NMR spectroscopy is at present the most powerful technique available for studying the solution conformations of oligonucleotides or any molecule for that matter. The magnetic resonance of ^{1}H, ^{13}C, ^{31}P, and associated chemical shift and coupling constants allow the researcher to monitor on an atomic level specific regions of the molecule while changing surrounding conditions, specifically, changes in the sugar-phosphate backbone and heterobase stacking. Measurement of coupling constants which vary in magnitude with the conformation about a covalent bond allows the calculation directly or indirectly

of the various conformations (Altona, 1975; Guschlbauer,1980; de Leeuw, and Altona, 1982).

Much of the investigations of mononucleosides, mononucleotides, and mononucleoside diphosphates are directed towards an understanding of the sugar ring conformation and the effects of phosphate on the overall conformation of the mononucleoside in solution. Ultimately the information on RNA and DNA building blocks should be useful in understanding the conformations possible in the polymers. Problem in determining solution conformations from the NMR data is that, in ribofuranose only the $^3J_{1',2'}$, $^3J_{2',3'}$, and $^3J_{3',4'}$, coupling constants are available whereas the solution conformation is fully described by five parameters. Therefore to elucidate the solution conformation at least two degrees of freedom must be eliminated (de Leeuw and Altona, 1982). Generally the two degrees of freedom are eliminated by assumptions based on observed X-ray and/or NMR data. For example Guschlbauer has assumed that there is equal puckering in the N and S conformer ($\Phi_N = \Phi_S$) and that the phase angles are inter-related by symmetry ($P_S = 180^\circ - P_N$). The first assumption seems justified considering the average puckerings found in X-ray structures. The second assumption is not tenable for the ribose ring as it does not have C_{2v} symmetry; that is, the N and S conformers are not mirror images. For this reason a

different approach has been suggested to reliably estimate the mole fraction of S conformer in solution. In this approach it is assumed that the solution geometry is similar to the solid state geometry when $^{3}J_{1'2'} + {}^{3}J_{3'4'}$ equals 9.4 + 0.5 Hz (de Leeuw and Altona, 1982).

Dinucleoside mono- and diphosphates have been useful compounds for studies on the backbone conformation of single stranded oligonucleotides in solution. By comparison of the solution conformational dynamics of the dinucleoside with the appropriate mononucleotides an understanding of the way in which a neighbouring unit affects its linked partner can be ascertained (Young and Krugh, 1975).

Dinucleotides can be considered to be the building blocks of a polymer in the sense that a trinucleoside consists of two dinucleotides sharing a common neighbouring group. The approach is only valid if the properties observed are the result of interactions mainly between neighbouring mononucleotidyl moities (Ts'o, et al., 1969). Ring currents for the bases and shift effects calculated as a function of interplanar distance between bases suggests that shielding effects are a nearest-neighbour phenomenon (Giessner-Prettre, et al., 1975; Giessner-Prettre and Pullman, 1976). Empirically derived proton chemical shift parameters suggest that next-nearest as well as nearest-neighbour effects are important (Everett, et al., 1980; Bell, et al., 1981).

Short RNA and DNA strands (2-14 base residues), used as models of longer natural RNA molecules and DNA molecules have been extensively investigated (Alkema, et al., 1981a,b,c; Bubienko, et al., 1981; Cornelis, et al., 1979; Kan, et al., 1973; Kondo, et al., 1975; Krugh, et al., 1976; Haasnoot, et al., 1980; Patel, 1976, 1977, 1979; Patel et al., 1982a,b,c,d; Stone, et al., 1981; and references therein). These studies have been aimed at understanding the biophysical consequences of sequence, on conformation, duplex formation and stabilization (Tinoco, et al., 1971, 1973; Ninio, 1979) as well as the biochemical consequences that mismatched bases, unmatched bases, sequence and helix type, may have on mutation and in the case of DNA, the effect exerted on gene expression (Lindahl, 1981; Nordheim, et al., 1981; Noller and Woese, 1981; Patel, et al., 1982e and f). Similar studies using complete tRNA molecules have also related structure to function (Reid, 1979; 1981).

Studies of short RNA strands have suggested that a number of factors can be involved in duplex formation. Strong base stacking, which is sequence dependent, is the major contributing factor (Alkema et al., 1981 a,b,c; Turner et al., 1981).

The greater base stacking (Figure 5) ability of an adenosine residue over a uridine residue explains why the RNA oligomer AGACU forms a double strand RNA duplex, while AGUCU does not (Alkema et al. 1982). Furthermore, base

sequence plays a role, through base stacking, since oligomer CAAUG does not form a stable RNA duplex (Romaniuk et al., 1979).

Base stacking apparently accounts for the formation, by the trinucleoside diphosphate, GCA of a mini RNA duplex GCA:GCA. However, the adenosine residue, in this case, is not involved in Watson-Crick base pairing (Figure 6) but is involved in extensive intra-strand stacking which contributes to the stability of this duplex (Tm = 33°C, Tm is a measure of duplex stability). The Tm equals that of the tetramer duplex formed by UGCA, which contains 4 Watson-Crick pairs (Alkema et al., 1981a).

Sequence and base stacking effects of bases at the ends of helical regions also contribute to duplex stability. Oligomer AAGCUA forms a duplex (Tm of 48°C) which is considerably more stable than that formed by the parent oligomer AGCU (Tm of 33°C).

The "dangling" adenosine residues enhance the stability of the duplex by extending base stacking thoughout the duplex. The adenosine residue at the 3'-end makes the most significant contribution since AGCUA:AGCUA has a Tm of 45°C (Alkema et al., 1981a). This result is consistent with temperature jump studies which demonstrated that the duplex formed by yeast $tRNA^{Phe}$ and E. Coli $tRNA^{Glu}$ at their complementary anticodon regions was 10^6 fold stronger than that expected for a three base pair duplex

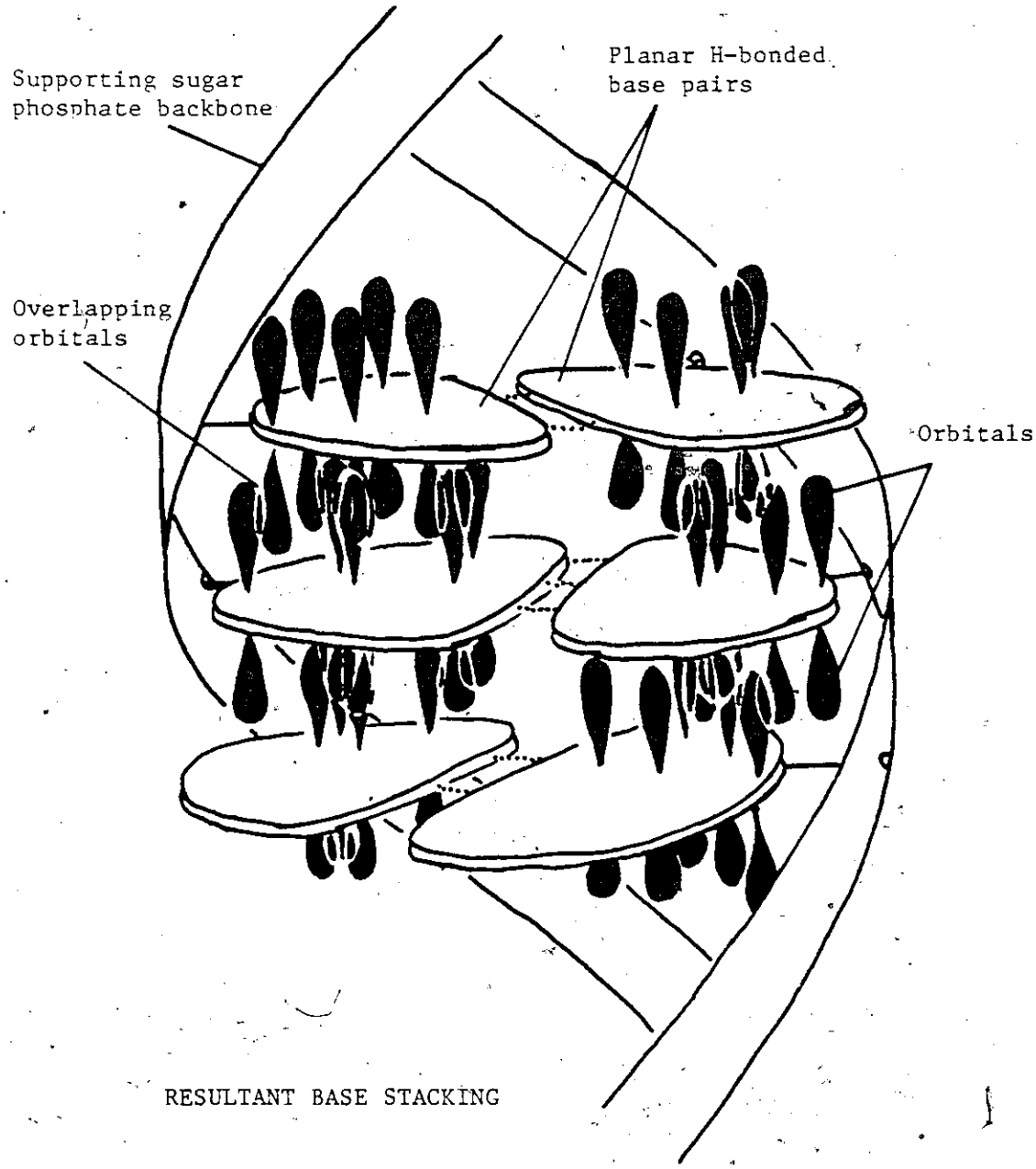


Figure 5. Diagramatic representation of the vertical electronic interactions, between parallel bases, inherent in base stacking.

(Grosjean, et al., 1975). From fragmentation studies, it was concluded that a 1000 fold portion of the increased stability was due to the additional base stacking in the unpaired regions next to the double helix (Grosjean, et al., 1976a, b). Continued temperature jump studies by Grosjean and coworkers culminated in a proposed physical basis for ambiguity in genetic coding interactions (Grosjean, et al., 1977, 1978a,b).

The duplex formed by AAGCUA, models the D-loop stem region of tRNA molecules (Kim, et al.,1973, 1974; Clark, 1977). We proposed that the invariant adenosine residues at positions 14 and 21 in the tRNA (Sprinzl, et al., 1980) may be involved in stabilization of the weak helical region of the D-stem (Alkema et al., 1981c). This idea is consistent with the finding that in yeast $tRNA^{Phe}$ the acceptor stem melts first followed by the D-stem (Johnston and Redfield, 1981).

Studies of short RNA strands specifically GCA, AGC, UGCA, GCAA, AGCU, AAGCU, AGCUA, and AGCUAA have led to the proposal that base stacking is directional and the stacking direction which is sequence dependent, stabilizes duplexes when the stacking direction is 5' to 3' (Alkema et al., 1981a). Since a change in stacking direction can cause a change in duplex stability, this may play a role in tRNA:mRNA translocation where a change in stacking direction of the anticodon loop takes place during the codon-anticodon

interaction (Ghosh and Ghosh, 1970; Woese, 1970; Lake, 1977; Urbanke and Maass, 1978; Geerdes, et al., 1980a, b; Labuda and Porshke, 1980).

An extension of the idea that sequence affects duplex stability in a directional sense has led Borer and coworkers to conclude that RNA helix conformation in solution differs from that in the solid state, where crystal structures suggest that there is no correlation between base-pair sequence and RNA conformation (Holbrook, et al., 1981). This suggestion was based on proton NMR investigations of short RNA oligomers which are considered better approximations of naturally occurring RNAs than mononucleotides or dinucleotides (Cruz, et al., 1982). The suggestion is that duplex regions that are made up of alternating purine-pyrimidine are most stable because of a change in the helix winding angle W, that leads to the possibility of inter-strand base stacking. This model has also been applied to suggest a physical basis for the codon-anticodon interaction (Borer, et al., 1982).

1.2.3.1 Identification of resonances in ^{1}H NMR Spectra.

Much of the advantage of using proton NMR as a technique to monitor the solution conformation of small RNA and DNA molecules is that regions of the molecule can be monitored. The disadvantage is that a NMR spectrum has a high information content and in molecules made up of

repeating units individual resonances become increasingly difficult to identify. Complete and accurate identification of resonances is needed if the function of biological molecules is to be established based on structure and conformation. For instance, the direction of base stacking could not be determined without accurate assignment of the spectral resonances to their protons (Alkema, et al., 1981a). Also the model for base overlap proposed by Borer and coworkers (Cruz, et al., 1982) depends critically upon chemical shift assignments especially since resonances from base pairs at the helix ends are not used in the fitting process as they are influenced by rapid fraying and possibly end-to-end aggregation (Bubienko, et al., 1981; Patel and Hilbers, 1975; Romaniuk, 1979). Borer and coworkers point out that one of the reasons that DNA spectral data fitting to B-helix configurations is unsuccessful is because there is a lack of shielding data on DNA oligomer duplexes where resonances are uniquely assigned.

Presently the incremental assignment procedure is used extensively to identify resonances in RNA or DNA oligomers. This procedure however requires one to have the spectra of each sequentially intermediate strand (Borer et al., 1975). Also, in some cases unambiguous assignments cannot be made without reference to other oligomer spectra (Everett, et al., 1980). Proton NMR-NOE experiments can also be used to identify resonances; however the experiments are

time consuming and generally require high sensitivity instruments (Petersheim and Turner, 1982b). In this thesis I present a new method which has been used successfully to identify proton resonaces in both RNA and DNA oligomers.

1.3 Factors in Prokaryotic Translational Termination.

Translation of mRNA can be divided into three stages comprising the initiation, propagation, and termination reactions of protein chain assembly. Each of these stages has been investigated with model systems where the requirements for partial reactions can be studied more easily. In particular, the use of artificial mRNA templates has been a major tool. The studies have led to the discovery of numerous factors involved in each step of the protein synthesis pathway (Ganoza, 1977; Metzler, 1977; Lodish, 1976; Hall, et al., 1982).

In the initiation reaction, nonribosomal proteins designated IF-1, IF-2, and IF-3, in the presence of GTP are required for optimum formation of a 30S ribosomal subunit:fMet-tRNA$_f$ complex. The joining of the 50S ribosomal particle to this 30S intermediate may require still other, less well-characterized protein(s). The joining event results in ejection of the initiation factors (Metzler, 1977).

Chain elongation begins when the soluble proteins

EF-Tu and EF-Ts in concert with GTP bind aminoacyl-tRNAs to ribosomes at a site which is presumably adjacent to the bound initiator-tRNA. Positioning of both the initiator-tRNA and aminoacyl-tRNAs on ribosomes is accompanied by GTP hydrolysis. This hydrolytic reaction is required before peptide bond synthesis can be catalyzed by the peptidyl transferase, a part of the 50S subunit that includes proteins L7 and L12. Peptide bond formation is thermodynamically favoured, thus the energy derived from GTP cleavage may be utilized for alignment of aminoacyl-tRNAs, ejection of the initiation and elongation factors, or breaking of the codon-anticodon interaction after the 50S particle has catalyzed peptide bond formation. Peptide synthesis involves movement of ribosomes relative to mRNA. Protein, EF-G stimulates both ejection of deacylated tRNA from ribosomes as well as translocation of the peptidyl-tRNA:mRNA into the 'donor' position of the apparatus. This process, possibly mechanical, occurs with concomitant GTP hydrolysis, and is cyclic (Woese, 1970; Lake, 1977; Holschuh, et al., 1980, 1981)

The end product of peptide elongation is a completed protein esterified to the 3'-OH end of the ribose moiety of the terminal adenosine of tRNA. The release proteins are involved in the recognition of termination signals, and induce the hydrolysis of peptidyl-tRNA by the peptidyl transferase. These proteins have been designated RF-1 (UAA,

UAG, specific) and RF-2 (UAA, UGA, specific). A third protein RF-3 apparently stimulates the activity of RF-1 and RF-2 (Caskey, 1979; Capecchi, and Klein, 1969). After protein cleavage occurs on polysomes, ribosome releasing factor, RR, dissociates ribosomes from mRNA (Ryoji, et al., 1981).

This detailed information on the partial reactions of protein synthesis suggests that mechanisms responsible for regulating translation of natural messengers are well understood. Certainly this is the impression obtained by reading any recent text in biochemistry (for example see Metzler, 1977)

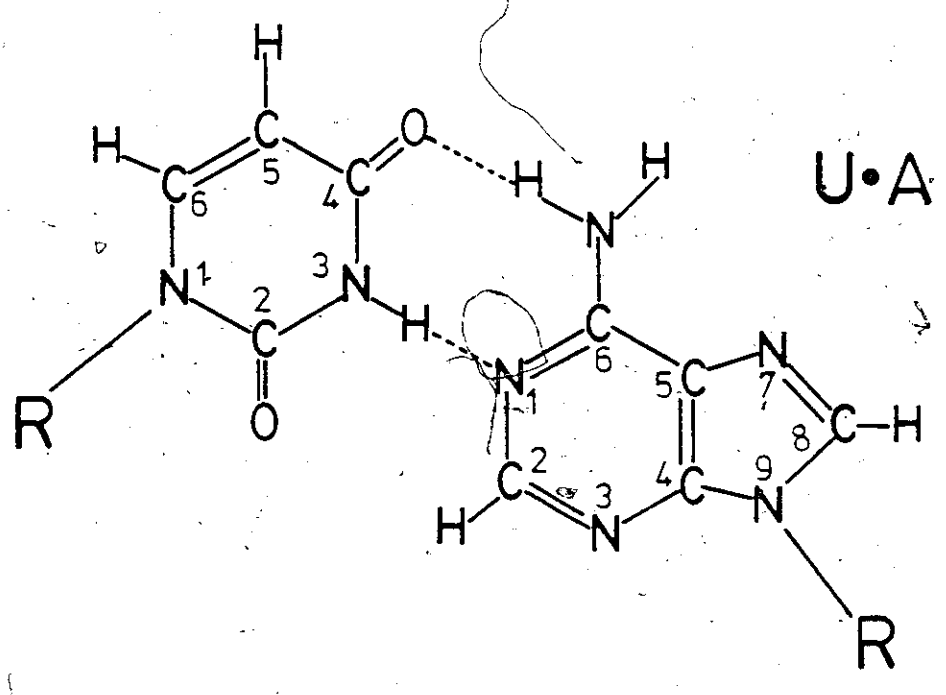
Natural mRNAs contain special untranslated regions that punctuate genes and are critical in determining the frequency and order of gene expression. Focussing on protein chain termination, it is clear that termination must yield proteins with the correct C-terminal end; however, there are a number of factors that remain unresolved (Bienz and Kulbi, 1982).

In vivo, nonsense mutations generate in-phase UAG, UAA, and UGA, codons. These codons cause premature termination in nonsuppressing cells (Sup^-), in suppressing cells (Sup^+), mutant tRNAs insert an amino acid at the nonsense site. If single nonsense codons suffice to terminate protein synthesis, then it is difficult to understand why strong suppressors do not affect growth

considering the many proteins made by polygenic mRNAs. The single terminators at Qβ coat protein, MS2 replicase or A protein, and A and B cistrons of tryptophan mRNA can be suppressed. Possibly single stop codons are important in evolution of proteins of altered function, the Qβ read-through protein is a prime example (Hofstetter, et al., 1974; Geller and Rich, 1980).

The standard release assay upon which much of the understanding of termination is based fails to respond to a protein termed "rescue", which inhibits termination at UGA, and UAA stop codons (Ganoza, et al, 1973; Ganoza, 1977). The "rescue" protein also affects early translation (Herrington, et al., 1975). Thus at least one other factor is involved in translational initiation and termination.

The 3'-end of 16S ribosomal RNA, $_{HO}$UUA..., has been proposed to be involved with recognition of termination codons UAA, UAG, and UGA. The stop codon UAA is readthrough with low efficiency because base pairing (Figure 6), with UUA is complete eliminating suppressor tRNA competition (Shine and Dalgarno, 1974). Removing the 49 nucleotides from the 3'-end of the 16S rRNA eliminated the termination response; however, it also eliminated aminoacyl-tRNA binding (Caskey, et al., 1977). The data are inconclusive as the removal of the 49 nucleotides alters more than one ribosomal function. Furthermore removal of specific ribosomal proteins near the termination codon recognition site also



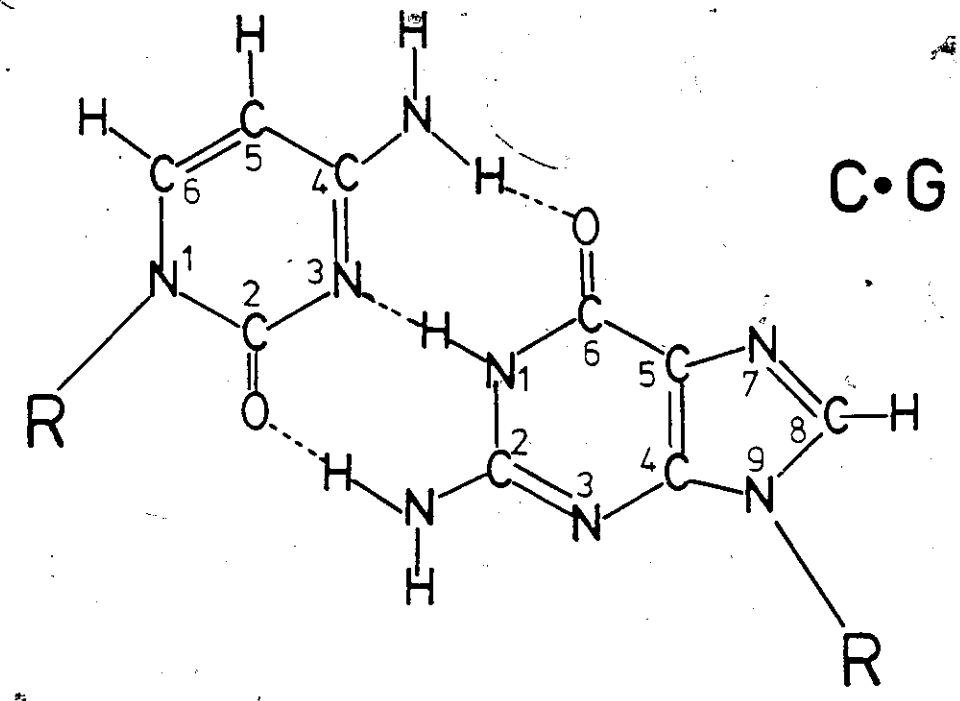


Figure 6. Representation of the U·A and C·G base pairs depicting the hydrogen bonding and the numbering system for the bases.

reduces or eliminates the termination reaction (Armstrong and Tate, 1978; 1980; Stoffler et al., 1982). It has been shown that release factor competes with tRNAs for termination codons presumably at the aminoacyl site (Ganoza and Tomkins, 1970; Ganosa, et al., 1982).

Other evidence that translational termination and initiation are linked stems from the similar results of recent experiments. In an in vitro experiment where ribosome releasing factor was absent, ribosomes reinitiated translation "in phase" after the termination codon UAG, without fMet-tRNA. These data do not agree with theories of "phaseless and sterile" travel between cistrons (Ryoji, et al., 1981). Lysis gene expression of RNA phage MS2 in vivo has been demonstrated to require a frameshift during coat protein translation which leads to termination at either of two stop codons (UAA), just preceding the lysis cistron. This termination is required for translation of the lysis gene for if the UAA codon is removed, translation is abolished. Furthermore the shorter the distance between termination and initiation codons the higher the probability for reinitiation (Kastelein, et al., 1982; Model, 1982). A frameshift mechanism has also been shown to relieve premature termination in a yeast mitochondrial gene (Fox and Weiss-Brummer, 1980), suggesting this mechanism may commonly occur.

Experimental data suggests that ribosomes prefer

specific nucleotides at the 5' side of the initiation triplet (Ganoza et al., 1978; 1982). It is likely that several nucleotides are involved in the selection of AUG, GUG, or UUGs as start codons. The Shine-Dalgarno region (a purine-rich sequence) 5' to initiator codons binds the 3'-end of the 16S ribosomal RNA (5' GAUCACCUCCUUA$_{OH}$ 3') and helps in this selection (Steitz, 1979). It has been suggested that the termination codons UAA, UAG, and UGA, which occur between genes, and are abundant in intercistronic regions (Steitz and Jakes, 1975) serve a dual function, terminating translation and possibly priming the translational apparatus for initiation (Ganoza, 1977). On inspection of regions around initiator codons, only the termination codons UAA and UGA are present suggesting that these two particular stop codons function in initiation (Atkins, 1979). Interestingly the stop codon UAA is implicated in reinitiation (Kastelein, et al., 1982), and these two codons are suppressed by "rescue" (Ganoza, 1977).

Initial studies to determine if initiation and termination reactions might be interlinked, probed the possible entrance of N-formylmethionyl-tRNA onto the ribosome when AUG was preceded or followed by a covalently joined termination codon. Apparently AUGUAA bound greater than 70% of the added N-fmet-tRNA$_f$ to ribosomes. No binding was observed with UAAAUG, and this hexamer did not inhibit formation of the initiation complex with AUG or with

AUG$(N)_m$ (Ganoza et al., 1978). Also, UAAAUG stimulated hydrolysis of ribosomal-bound N-fmet-$tRNA_f$ in the presence of release factors as effectively as UAA (Ganoza, 1977).

These data, taken together, show that the sequence neighbouring termination codons can alter the termination response both in vitro and in vivo. Termination codons, and the termination reaction, also seem to play a role in translational initiation. In this thesis, data are presented that demonstrates; that, the presence of an initiator codon, AUG, can alter the sequence specificity of the fmet-tRNA:mRNA:70S ribosome:RF-1 complex such that a termination reaction is elicited in the presence of a partial termination codon.

EXPERIMENTAL

2. Experimental Procedures.

2.1 Materials and Reagents.

All reagents and solvents used were commercial reagent grade and were used without further purification unless otherwise noted. Methylene chloride (BDH) was further purified by passing through silica gel (60-200 mesh, Baker Analyzed Reagent). Mesitylene-sulfonyl chloride (Aldrich) was recrystallized from hot petroleum ether (40-60°C), pyridine, dimethyl formamide and p-dioxane (Baker Analyzed Reagents) were further dried over Fisher 4Å molecular sieves prior to use.

Silica gel G plates (Analtech) were used for thin layer chromatography of protected nucleosides and coupling reaction mixtures. The chromatography solvent system was 10% methanol in methylene chloride. Sample detection was accomplished by spraying the plates with 1% ceric sulfate ($H_4Ce(SO_4)_4$) in 10% sulfuric acid and heating to 160-180°C. Most compounds appeared as black-brown spots, except those containing a trac group which appeared as a bright yellow-brown spot.

Silica gel (60-200 mesh, Baker Analyzed Reagent) for column chromatography was dried at 60-80°C over NaOH for 48 hours prior to use. Columns used for purification of

protected nucleosides and oligoribonucleotides were prepared in methylene chloride and compounds eluted via step gradients of methanol in methylene chloride.

Whatman #1 and #40 papers were used for the descending paper chromatography of free oligoribonucleotides. The solvent was ethanol/NH_4OAc (1M, pH 7.3) 1:1 (v/v). Deblocked oligoribonucleotides were detected by short wave ultra-violet light.

The columns used in the HPLC purification of oligoribonucleotides were Analtex Ultrasphere ODS 5 µ (25cm x 4.6mm ID) (Beckman, U.S.A.) and Bondapak C18 (30cm x 3.9mm ID; 3000 plates min.) (Waters Scientific Ltd. Canada). HPLC grade acetonitrile and methanol were purchased from Fisher Scientific or BDH Chemicals, and used without further processing. The water used for HPLC applications was purified by filtration through a Milli-R04 purification system (Millipore). The ammonium acetate buffers (Fisher HPLC grade) (0.1M, pH 7.0) were filtered through a 0.5 µ filter (Millipore) and degassed by stirring under vacuum prior to use.

2.2 Protected Nucleosides and Coupling Reagents.

Nucleoside derivatives were synthesized and characterized by their melting points and ^1H NMR spectra. These syntheses were routine and carried out as published;

HOUtOH (Griffin et al., 1966), HOC^{bz}tOH (Neilson & Werstiuk, 1971), HOG^{bz}tOH and HOA^{bz}tOH (Gregoire & Neilson, 1978), TracUtOH (Werstiuk & Neilson, 1972), TracA^{bz}tOH, TracC^{bz}tOH and TracG^{bz}tOH (Werstiuk & Neilson, 1973).

Coupling reagents 2,2,2-trichlorethylphosphate and MST were also synthesized and characterized by published procedures (England & Neilson, 1976; Katagiri et al., 1974) respectively.

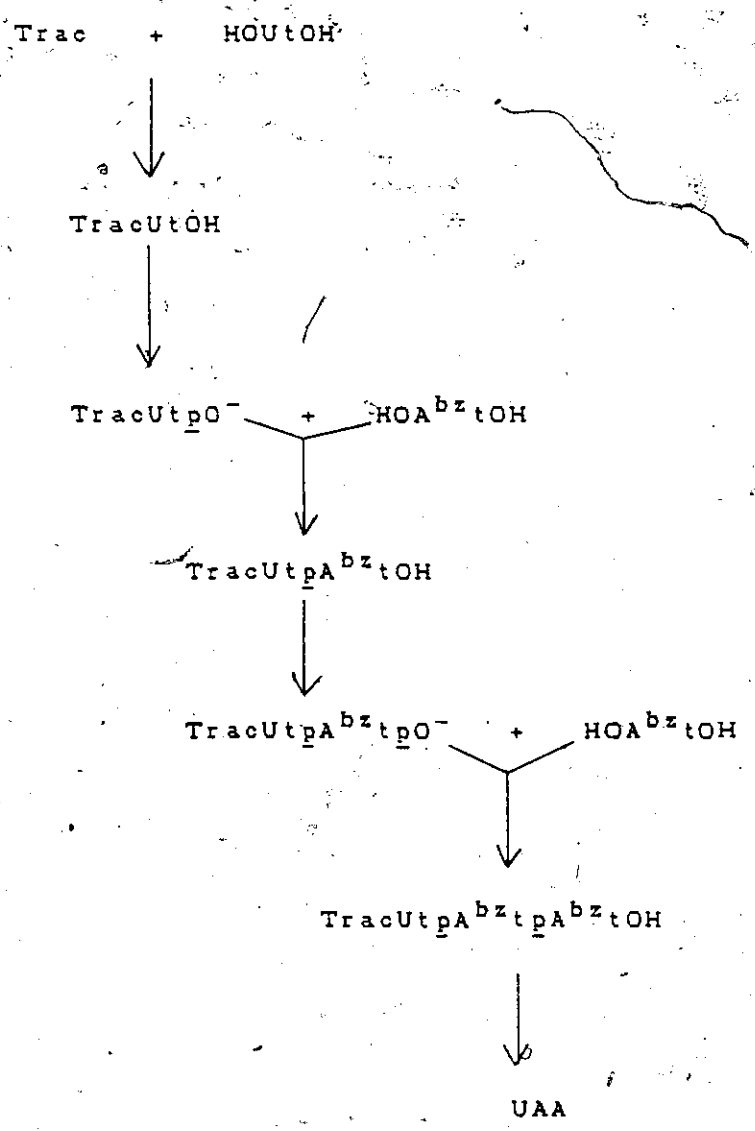
2.3 Preparation of oligoribonucleotides.

Complete details for the sythesis of sequences prepared for use in termination, parameter and sequence studies are presented in Tables 1a and 1b. All of the protected oligoribonucleotides were constructed using published techniques (England & Neilson, 1976; Werstiuk & Neilson, 1976). The stepwise synthesis of chemically protected UAA (Figure 7) is described to illustrate the general procedure.

2.3.1 2'-O-Tetrahydropyranyl-5'-O-triphenylmethoxy-acetyluridylyl (3'-(2,2,2-trichlorethyl)-5') N-benzoyl-2'-O-tetrahydropyranyladenosine.

The blocked nucleoside TracUtOH (1.9 g, 3.025 mmol, 1 equiv.) was dissolved in anhydrous pyridine (ca. 20 mls)

Figure 7. Chemical Synthesis of Chemically Protected UAA.



and further dried by evaporation in vacuo from anhydrous pyridine (3 x 20 mls) and sealed under dry nitrogen.

The incoming monophosphate ester 2,2,2-trichlorethylphosphate (England & Neilson, 1976) (1.4 g, 6.1 mmol, 2.0 equiv.) was evaporated in vacuo from anhydrous pyridine (3 x 30 mls) to convert it to the pyridinium salt (as evidenced by formation of a white precipitate). The coupling agent MST (3.0 g, 12.1 mmol, 4.2 equiv.) was added to the pyridine suspension (ca. 40 mls) and the solution was warmed (30-40°C, 1-2 h) under dry nitrogen.

The solution of the activated phosphate was added to the pyridine solution of TracUtOH, the total volume was reduced in vacuo (ca. 30 mls), sealed under dry nitrogen and let stand at room temperature in the dark. The tlc of the reaction mixture indicated the phosphorylation was ca. 85-90% complete (R_f 0.8 → 0.3) after 48 hours. The reaction was quenched by the addition of ice (ca. 3-5 g). After 15 min, the reaction mixture was poured into ice water (ca. 80 mls) and extracted with methylene chloride (4 x 50 mls) and evaporated in vacuo to a light yellow-brown foam.

TracUtpO$^-$ (ca. 3.0 mmol) was evaporated from anhydrous pyridine (3 x 40 mls, final volume ca. 25 mls) and MST (1.5 g, 6.1 mmol, 2.0 equiv.) was added and the reactivation mixture sealed under dry nitrogen. The solution was warmed (30-35°C) for 1 hour. An anhydrous,

pyridine solution (ca. 20 ml) of high R_f isomer of $HOA^{bz}tOH$ (1.37 g, 3.0 mmol, 1.1 equiv.) was then added, the final volume reduced *in vacuo*, and the reaction mixture sealed under dry nitrogen. After 5 days of reaction at room temperature in the dark, tlc indicated the reaction was ca. 80-90% complete (R_f 0.3 → 0.7). The reaction was quenched with ice (ca. 3-5 g), poured into ice water (ca. 80 ml) and extracted with methylene chloride (4 x 100 ml). The combined extracts were washed with water (1 x 50 ml) and evaporated *in vacuo* to dryness. Traces of pyridine were removed by co-distillation with toluene (1 x 10 ml) and a toluene-methylene chloride mixture (2 x 10 ml) to give a light brown foam. TracUtpA^{bz}tOH was purified by chromatography on 30 g silica gel. Elution with 2.5-3% methanol-methylene chloride yielded TracUtpA^{bz}tOH (798 mg, 0.78 mmol, 21%).

2.3.2 2'-O-Tetrahydropyranyl-5'-O-triphenylmethoxyacetyluridyly (3'-(2,2,2-trichlorethyl)-5') N-benzoyl-2'-O-tetrahydropyranyladenylyl (3'-(2,2,2-trichlorethyl)-5' N-benzoyl-2'-O-tetrahydropyranyladenosine.

The blocked dinucleoside monophosphate TracUtpA^{bz}OH (586 mg, 0.55 mmol, 1 equiv.) was dissolved in anhydrous pyridine and further dried by evaporation *in vacuo* from anhydrous pyridine (3 x 20 ml) and sealed under

dry nitrogen.

The 2,2,2-trichlorethylphosphate (253 mg, 1.1 mmol, 2 equiv.) was converted to its pyridinium salt by repeated evaporation in vacuo from anhydrous pyridine (3 x 20 mls) and activated by MST (552 mg, 2.2 mmol, 4 equiv.). After 2.5 hours, the activated solution was added to the pyridine solution of TracUtpA^{bz}OH, the volume reduced in vacuo (ca. 15-20 mls) and sealed under dry nitrogen. After 72 hours at room temperature, tlc indicated the reaction was complete (R_f 0.7 → 0.2). Ice was added, the reaction was then poured into ice water (ca. 50 mls) and repeatedly extracted with methylene chloride (4 x 50 mls). The combined extracts were washed with water (1 x 50 mls) and evaporated in vacuo to a yellow-brown oil.

TracUtpA^{bz}tpO$^-$ (ca. 0.55 mmol) was evaporated from anhydrous pyridine (3 x 20 mls) and sealed under dry nitrogen (final vol. ca. 10 mls). The phosphate was reactivated with MST (276 mg, 1.1 mmol, 2 equiv.). After 2 hours, an anhydrous pyridine solution of the high R_f isomer of HOA^{bz}tOH (275 mg, 0.61 mmol, 1.1 equiv.) was added. After 3 days, tlc indicated the reaction was 60-70% complete (R_f 0.2 → 0.6). A further 50 mgs of MST were added and the reaction mixture resealed under dry nitrogen. After a further 24 hours, tlc indicated the reaction was complete (R_f 0.6). Ice (ca. 3 g) was added, the reaction was poured into ice water (ca. 50 mls) and extracted with

methylene chloride (4 x 50 mls). The combined extracts were washed with water (1 x 50 mls) and evaporated to a yellow-brown foam.

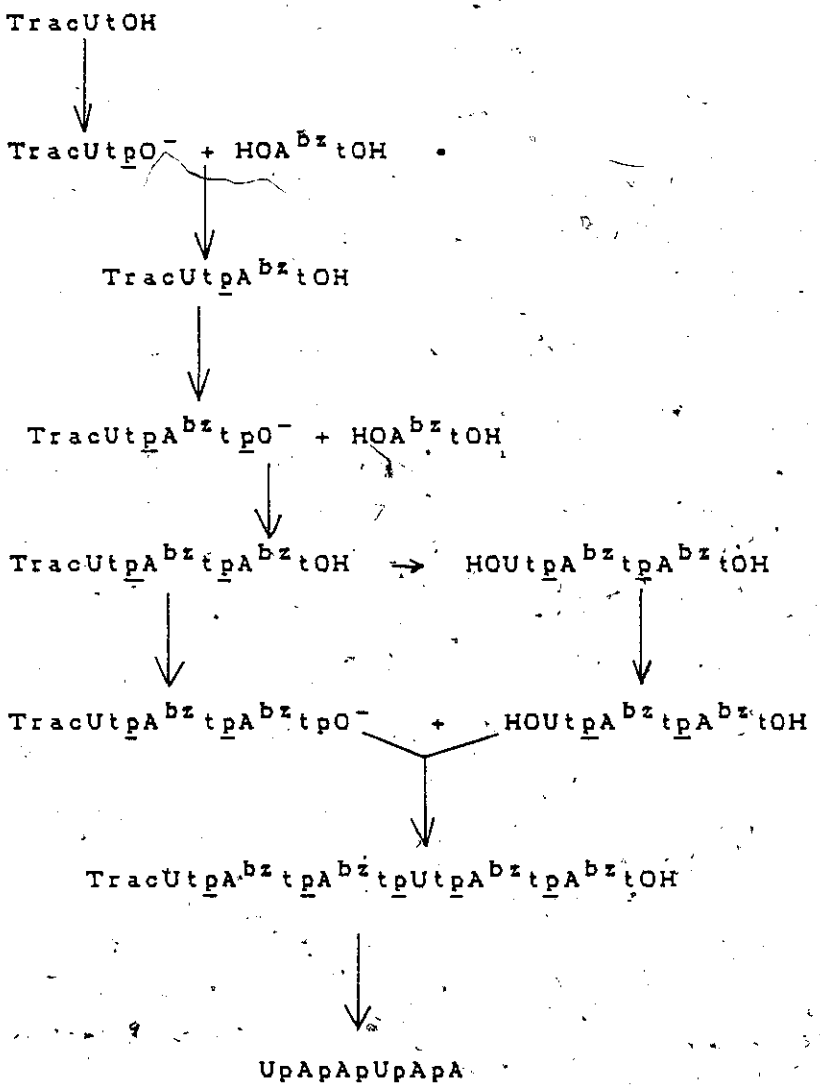
Chromatography of the foam on a 10 g silica gel column yielded TracUtpA^{bz}tpA^{bz}tOH (522 mg, 0.27 mmol, 59%) upon elution with 3-4% methanol-methylene chloride.

2.4 Preparation of Oligoribonucleotides Via a Block Synthesis.

Stepwise assembly is convenient for oligomers up to five residues in length. For longer oligomers, fragment coupling is less tedious and gives better overall yields (Werstiuk & Neilson, 1972).

The method is especially useful in constructing oligomers such as UpApApUpApA where a repeated shorter unit is contained within the sequence (in this case it is UpApA). The chemical coupling of two oligomer blocks proceeds under the same conditions as the coupling of two nucleoside residues and proceeds in a 5' to 3' direction (Figure 8). For the block synthesis to be reasonable, the specific removal of the 5' Trac blocking group must be accomplished in high yield. This step in the synthetic procedure and the subsequent coupling are described.

Figure 8. Block Synthesis of Chemically Protected UAAUAA.



2.4.1 2'-O-Tetrahydropyranyluridylyl (3'-(2,2,2-trichlorethyl)-5') N-benzoyl-2'-O-tetrahydropyranyladenylyl (3''-(2,2,2-trichloethyl)-5') N-benzoyl-2'-O-tetrahydropyranyladenosine.

TracUtpA^{bz}tpA^{bz}tOH (212 mg, 0.14 mmol, 1.0 equiv.) was dissolved in 30 mls of reaction solvent (1% v/v NH_4OH/CH_3OH). The reaction was followed by tlc (R_f 0.60 → 0.25 + 1.0). Specific removal of the triphenylmethoxyacetyl group was evidenced by the appearance of a yellow spot at the solvent front. After 3 hours, the reaction was 85-90% complete as evidenced by intensities of spots on tlc plates.

The reaction was quenched by evaporation in vacuo to dryness. Chromatography of the white foam on a 12 g silica gel column yielded HOUtpA^{bz}tpA^{bz}tOH (136 mg, 0.11mmol, 80.2%) upon elution with 4% methanol-methylene chloride.

2.4.2 2'-O-Tetrahydropyranyl-5'-O-triphenylmethoxy-acetyluridylyl (3'-(2,2,2-trichlorethyl)-5' N-benzoyl-2'-O-tetrahydropyranyladenylyl (3'-(2,2,2-trichlorethyl)-5') N-benzoyl-2'-O-tetrahydropyranyladenylyl(3'-(2,2,2-trichlorethyl)-5') 2'-O-tetrahydropyranyluridylyl (3'-(2,2,2-trichloethyl)-5') N-benzoyl-2'-O-tetrahydropyranyladenylyl (3'-(2,2,2-trichloethyl)-5') N-benzoyl-2'-O-tetrahydropyranyladenosine.

TracUtpA^{bz}tpA^{bz}tOH (300 mg, 0.156 mmole, 1.4 equiv.) was dissolved in 30 mls of dry pyridine, and remaining moisture removed by evaporation in vacuo from anhydrous pyridine (3 X 20 mls) and sealed under dry nitrogen. The 2,2,2-trichlorethylphosphate (92 mg, 0.398 mmole, 2 equiv.) was converted to its pyridinium salt by repeated evaporation in vacuo from anhydrous pyridine (3 X 20 mls) and activated by MST (201 mg, 0.801 mmole, 4 equiv.). After 2-3 hours, the activated solution was added to the pyridine solution of TracUtpA^{bz}tpA^{bz}tOH, the volume reduced in vacuo (ca, 15-25 mls) and sealed under dry nitrogen. After 72 hours at room temperature, tlc indicated that the reaction was complete (R_f 0.6 → 0.2). The reaction was quenched by the addition of ice, the mixture was then poured into ice water (ca. 40-50 mls) and repeatedly extracted with methylene chloride (4 X 50 mls).

The combined extracts were washed with water (1 X 50 mls) and evaporated in vacuo to a yellow-brown oil.

TracUtpA^{bz}tpA^{bz}tpO^{-} (ca. 0.156 mmole) was evaporated from anhydrous pyridine (3 X 20 mls) and sealed under dry nitrogen (final vol. ca. 15 mls). The phosphate was reactivated with MST (100 mg, 0.398 mmole, 2 equiv.). After 2-3 hours, an anhydrous pyridine solution of HOUtpA^{bz}tpA^{bz}tOH (136 mg, 0.091 mmole, 1.0 equiv.) was added. After 2 days, tlc indicated the reaction was 60-70% complete (R_f 0.2 + 0.5 → 0.60). A further 50 mg of MST were added and the reaction mixture resealed under dry nitrogen. After a further 48 hours, tlc indicated the reaction was complete (R_f 0.65). Ice (ca. 3 g) was added, the reaction was poured into ice water (ca. 50 mls) and extracted with methylene chloride (4 X 50 mls). The combined extracts were washed with water (1 X 50 mls) and evaporated to a yellow-brown foam.

Chromatography of the foam on a 10 g silica gel column yielded TracUtpA^{bz}tpA^{bz}tpUtpA^{bz}tpA^{bz}tOH (245 mg, 0.068 mmole, 75%) upon elution with 5-6% methanol-methylene chloride.

2.5 Deprotection of Chemically "Blocked" Oligoribonucleotides.

Protected oligonucleotide (20 mg) was dissolved in

1.0 ml DMF and Cu/Zn couple (ca. 10-20 mg) was added, and left to react for 1-2 hours at 50°C. Reaction (removal of phosphate blocking group, 2,2,2-trichloroethanol) was judged complete by tlc ($R_f \rightarrow 0.0$). Methanolic ammonia 2.0 mls (methanol:cNH_3, 1:1 (v/v)) was added, the reaction vessel was tightly sealed, and the reaction left to stir at room temperature for 2 days. This effected removal of the Trac, and benzoyl blocking groups. The Cu/Zn couple was filtered and washed with 1 N ammonia. Chelex-100 (NH_4^+ form) was added to the combined filtrate and washings and stirred for 1 hour. The chelex was filtered and washed with 1 N ammonia. The combined filtrate and washings were evaporated in vacuo to dryness. The partially deprotected oligomer was purified by descending paper chromatography on Whatman #1 in ethanol:1 M NH_4OAc (pH 7.3) (1:1, v/v). The desired band (R_f = 0.8 → 0.9) was cut-out, desalted by soaking in absolute ethanol (1 hour) and dried in anhydrous diethyl ether (20 min) and eluted with glass distilled water. The pH of the eluate was adjusted to 2.0 with 2 N HCl, and left to stand at room temperature for 48 hours to facilitate removal of the tetrahydropyranyl group. The solution was then neutralized with concentrated ammonia, evaporated in vacuo to dryness, and purified on Whatman #40 in ethanol:1 M NH_4OAc (pH 7.3) (1:1, v/v).

The desired band was cut-out, desalted as described

and eluted with glass distilled water. The data for the deprotection of the oligoribonucleotides are presented in Table 2. The sequences were characterized by their 70°C ^{1}H NMR spectra.

The HPLC procedure involves the following modification: After the combined filtrate and washings were evaporated *in vacuo* to dryness, the partially deprotected oligomer (thp blocking groups not removed) was purified by HPLC chromatography. The dried oligomer was redissolved in a minimum amount of water, generally 200 µl and 100 µl put into each of two 300 µl sample vials (Waters). Two washings of 100 µl each are added to the samples. The entire 200 µl sample containing ca. 10 mg of oligomer and impurities is injected onto a reverse phase C-18 column and eluted by linear gradient HPLC. Initial solvent 5% v/v Acetonitrile/0.1 M NH_4OAc (pH 7.0) to 30% v/v Acetronitrile/0.1 M NH_4OAc (pH 7.0) in 30 min, flow rate 1.0 ml/min at 1500 psi.

Fractions containing the partially blocked oligomer were collected and pooled. The pH was adjusted to 2.0 by the addition of 2 N HCl. After 48 hours the sample is neutralized with 7.5 N ammonia, evaporated *in vacuo* to dryness, and redissolved in a minimum amount of water (200-400 µl). The samples (200 µl) containing ca. 10 mg of oligomer and impurities are injected onto a reverse phase C-18 column and eluted by linear gradient HPLC. Initial

solvent is 0.1 M NH_4OAc (pH 7.0) to 3% v/v CH_3CN/0.1 M NH_4OAc (pH 7.0) in 30 min, at a flow rate of 1.0 ml/min, at 1500 psi.

2.6 Proton Nuclear Magnetic Resonance of Oligoribonucleotides.

The 90 MHz proton spectra were obtained in the Fourier transform mode of a Bruker WH-90 spectrometer equipped with quadrature detection. Probe temperatures were maintained to within $\pm 1^{\circ}C$ by a Bruker B-ST 100/700 variable temperature unit, and were calibrated by thermocouple measurements. The samples were lyophilized twice from D_2O and then dissolved in 100% D_2O (Aldrich) or 100% D_2O buffer (1.0 M NaCl, 10 mM NaH_2PO_4, pH 7.2). t-Butyl alcohol-OD was used as an internal reference and the chemical shifts are reported in ppm relative to DSS. The field frequency lock was provided by the deuterium signal of D_2O. The spectra were recorded over a 1,000 Hz sweep width into 8 K data points (4.096 s aquisition time). The pulse width was 3.5 μs (67.5° pulse angle). The spectra of shorter sequences (trimers and tetramers) were obtained in 100-200 transients, while those of larger sequences (hexamers, heptamers) required 500 to 1,000 transients.

The 250.1335 MHz proton NMR spectra were obtained in

the Fourier transform mode of a Bruker WM-250 spectrometer equipped with quadrature detection. Probe temperatures were maintained to within $\pm 1^{\circ}C$ by a Bruker B-VT 1000 variable temperature unit, and were calibrated by thermocouple measurements. Samples were prepared as described for the 90 MHz spectra. Spectra were recorded over a 2,400 Hz sweep width into 16 K data points (3.4079 s aquisition time). Pulse width was 11.0 μs (=90.0° pulse angle). Most spectra were obtained with the HDO peak suppressed by irradiation of the peak (35L decoupler power). Many spectra were resolution enhanced by guassian multiplication of the FID by a negative line broadening (-1 Hz), and a guassian factor of 0.200. Spectra were obtained in 64 to 256 transients, in most cases.

2.7 Ribosome:Oligoribonucleotide Complex Formation.

Seventy S ribosomes were isolated from midlog E. coli Q13 cells (Ganoza and Barraclough, 1975; Ganoza, et al., 1978), with the following modifications: Twice washed ribosomes in 0.01 M Tris-HCl, pH 7.4; 0.001 M $MgCl_2$ and 0.5 M NH_4Cl, at a concentration of 10 mg/ml were centrifuged 10 minutes at 30,000 g. The supernatant was spun for 160 min at 110,000 g. The resulting pellet was suspended in 0.01 M Tris-HCl, pH 7.4; 0.01 M $MgCl_2$ and 0.5 M NH_4Cl (buffer 1), at a concentration of 10 mg/ml, then centrifuged as above. The ribosomal pellet was

suspended in Buffer 1 at a concentration of 100-200 mg/ml and stored at 0°C.

The f[^{35}S]met-tRNA$_f$ was prepared as previously described using E. coli B tRNA (Ganoza and Nakamoto, 1966). The specific activity of formylated f[^{35}S]met-tRNA$_f$ was determined by isotope dilution to be 108 x 10^3 dpm/pmole. The extent of formylation was analyzed on aliquots of f[^{35}S]met-tRNA$_f$ after 10 min hydrolysis in 0.33 N KOH at 37°C. After neutralizing, 0.25 ml of 0.1 N HCl was added and the f[^{35}S]met was extracted into ethyl acetate (Caskey, et al., 1968).

Partially purified activating enzymes, obtained from E. coli B supernatant fractions (Ganoza, et al., 1976) were used to aminoacylate pure tRNA$_f^{met}$. The acylated tRNA was extracted with phenol in 0.1% SDS (Pace, et al., 1968). The [^{3}H]met-tRNA was acetylated and purified as described (Haenni and Chapeville, 1966). The specific activity of the N-Acetyl-[^{3}H]met-tRNA was approximately 19.5 x 10^4 dpm/pmole.

Unless otherwise stated, ribosome:fmet-tRNA:oligoribonucleotide complexes were formed in an incubation volume of 0.045 ml containing 100 µg of four times washed E. coli Q13 ribosomes, 300 pmoles of oligoribonucleotide and approximately 2.5 pmoles of f[^{35}S]met-tRNA or N-acetyl[^{3}H]met-tRNA. The reaction was incubated at 24°C for 15 min and halted immediately by

plunging tubes into ice. If binding of fmet-tRNA or N-acetyl-met-tRNA was to be measured, approximately 5 ml of 0.01 M $MgCl_2$, 0.05 M Tris-HCl, pH 7.4 and 0.1 M NH_4Cl was added to each tube and reactions were filtered on 0.45 μ filters (Millipore), washed with this buffer, dried and counted in toluene, PPO and POPOP (Nirenberg and Leder, 1964).

2.7.1 Peptide Synthesis.

For dipeptide synthesis, reaction mixtures contained partially purified initiation factors and EF-T. Initiation factors were isolated as described (Suttle, et al., 1973). EF-Tu and EF-Ts were prepared according to (Wurmbach and Nierhaus, 1979; or Arai, et al., 1972). Reaction mixtures contained an approximate 70-fold excess of hexanucleotide to ribosomes. Reaction conditions were as described (Erbe and Leder, 1968), except that incubations (0.045 ml), contained 5 ug of hexanucleotide, 2.6 μg IF-1, 2.5 μg IF-2, 1.2 μg EF-T, 0.01 M $MgCl_2$, and about 4 pmoles each of f[^{35}S]met-$tRNA_f$ and unlabelled leu-tRNA and 100 μg of 70S Q13 ribosomes. When indicated, 14 μg of release factor was added. After 20 min of incubation at 24°C, an aliquot (0.03 ml) was withdrawn and hydrolyzed in 0.5 M NH_4OH for 30 min at 37°C. About 0.05 ml of the reaction mixture was spotted onto Whatman 3 mm filter papers and dried under

N_2. Electrophoresis was carried out at 125 mA for 60 min (200 V/cm) in pyridine:glacial acetic acid:water (1:10:189). Formylated dipeptide standards were stained by spraying the paper with 0.1 N Na dichromate in glacial acetic acid (1:1 v/v) and then with 0.1 N $AgNO_3$. One cm strips were cut and counted using toluene, PPO, and POPOP as scintillation fluid.

2.7.2 Translational Termination In Vitro.

After ribosome:fmet-tRNA:oligoribonucleotide complex formation, the reactions were stopped immediately by plunging the reaction tubes into ice, and then 0.030 ml aliquots were added containing 14 µg of RF-1. The release factor was prepared from E. coli K12 cells (Ganoza and Tompkins, 1970). Unless otherwise specified, the termination reaction proceeded at 24°C for thirty min and was halted by the addition of 0.25 ml of 0.1 M HCl and 1.6 ml ethyl acetate. The amount of N-acetyl-[^{3}H]methionine or of f[^{35}S]methionine hydrolyzed (released), was analyzed after extraction of 1 ml aliquots from the ethyl acetate layer and counted (Caskey, et al., 1968).

RESULTS AND DISCUSSION

3. Proton NMR Studies of Sequence and Conformation using Short RNA Oligomers.

3.1 Chemical Synthesis of Oligoribonucleotides.

All oligoribonucleotides used in this work were prepared by the general phosphotriester synthesis developed by Neilson and coworkers (England & Neilson, 1976; Werstiuk & Neilson, 1976). The method allows large scale preparations of oligoribonucleotides, suitable for use in these studies.

The preparative data for the protected oligoribonucleotides are contained in Tables 1a,b,c. Table 2 gives chromatography data and the yields of the pure, free oligoribonucleotides after chemical deprotection. The data illustrate that the chromatographic characteristics of RNA oligomers depends on the length, base composition and sequence. HPLC of RNA oligomers demonstrates that oligomer length, base composition, and sequence affects the retention time of both 2'-o-tetrahydropyranol derivatives and free oligoribonucleotides (Table 3). Oligomer length in general has the greatest affect on retention time while base composition and sequence more moderately affect retention time. These trends are consistent with those found for RNA and DNA oligomers of similar length and base composition (McFarland and Borer, 1979; Fritz, et. al., 1978).

Table 1a. Summary of preparation of Protected Oligoribonucleotides[a].

Reactants						Products		
	Quantity			Quantity			Quantity	
Comp.	mg	mmole	Comp.	mg	mmole	Comp.	mg	% Yield
U	1900	3.025	A	1370	3.011	UA	798	21
U	500	.796	G	412	.875	UG	587	57
UU	820	.713	A	350	.769	UUA	720	56
UA	586	.459	A	275	.605	UAA	522	59
UA	540	.423	G	240	.510	UAG	540	66
UG	567	.439	A	221	.486	UGA	560	66
UUA	390	.214	UUA	250	.165	UUAUUA	320	55
UUU	150	.090	UUA	165	.110	UUUUUA	75	25
UAA	300	.156	UAA	136	.084	UAAUAA	245	75
UAA	323	.168	UAG	230	.140	UAAUAG	268	51
UAA	215	.112	UGA	125	.076	UAAUGA	200	70
AUG	192	.089	CUA	130	.081	AUGCUA	135	45
AUG	100	.052	CUUA	80	.037	AUGCUUA	131	83
AUG	208	.096	CUAA	170	.076	AUGCUAA	110	33

[a]Column 1 contains the 5'-trityloxacetyl reactants and U stands for tracUtOH; column 4 contains the incoming nucleosides and A stands for HOBz-AtOH; column 7 contains the trityloxyacetyl product and UA stands for tracU-p-Bz-AtOH. Pyridinium mono-2,2,2-trichloroethyl phosphate (2 equiv) activated by 4 equiv of MST in anhydrous pyridine is used in each phosphorylation step. The coupling step to the incoming nucleoside is driven by 1.2 equiv of MST.

Table 1b. Summary of preparation of Protected Oligoribonucleotides[a].

Reactants						Products		
	Quantity			Quantity			Quantity	
Comp.	mg	mmole	Comp.	mg	mmole	Comp.	mg	% Yield
G	435	.564	A	282	.620	GA	398	49
GA	213	.150	U	54	.165	GAU	106	36
G	980	1.270	G	659	1.398	GG	740	41
GG	247	.172	A	86	.188	GGA	142	40
GG	247	.172	G	89	.188	GGG	126	35
GG	247	.172	U	62	.188	GGU	118	35
G	435	.564	C	267	.620	GC	350	44
GC	300	.215	G	122	.258	GCG	78	18
GC	208	.149	C	71	.164	GCC	184	61
G	1300	1.322	U	477	1.454	GU	840	49
GU	210	.162	A	82	.180	GUA	113	36
GU	210	.162	G	92	.195	GUG	120	38
GU	210	.162	C	79	.182	GUC	148	48
GU	210	.162	U	60	.183	GUU	96	33
U	898	1.430	U	518	1.579	UU	1000	61
UU	500	.433	A	206	.453	UUA	300	39
UU	232	.217	G	105	.222	UUG	204	52
UU	200	.174	C	140	.325	UUC	217	70

[a]Column 1 contains the 5'-trityloxacetyl reactants and G stands for tracBz-GtOH; column 4 contains the incoming nucleosides and A stands for HOBz-AtOH; column 7 contains the trityloxyacetyl product and GA stands for tracBz-Gt-p-Bz-AtOH. Pyridinium mono-2,2,2-trichloroethyl phosphate (2 equiv) activated by 4 equiv of MST in anhydrous pyridine is used in each phosphorylation step. The coupling step to the incoming nucleoside is driven by 1.2 equiv of MST.

Table 1c. Specific Deblock of 5'-Trityloxyacetyl Oligomers[a].

Reactants	Quantity		Products	Quantity	
	mg	mmole		mg	% Yield
TracUAA	212	0.110	HOUAA	136	76
TracUAG	350	0.180	HOUAG	230	79
TracUGA	350	0.180	HOUGA	125	42
TracUUA	250	0.139	HOUUA	130	62
TracUUA	312	0.173	HOUUA	250	96
TracCUA*	1000	0.525	HOCUA	436	53
TracCUAA*	980	0.383	HOCUAA	574	66
TracCUUA	194	0.080	HOCUUA	80	47
TracUUAU	340	0.146	HOUUAU	208	70

[a]Ammonia (0.15 M) in Methanol was reactant solvent (10 mg Trac compound in 1 ml).

*Prepared by Dirk Alkema.

Table 2. Experimental Data of Free Oligoribonucleotides.

Compound[a]	R_f[b]	A_{260} Units	Yield (%)[c]
UA	0.61	140	43
UG	0.49	175	73
UUA	0.51	168	38
UUG	0.56	135	40
UUC	0.54	100	34
UUU	0.62	118	37
UAA	0.28	310	89
UAG	0.33	239	82
UGA	0.41	154	47
GUU	0.35	163	43
GUC	-	120	40
GUG	0.21	54	20
GUA	-	132	37
GCC	-	117	41
GCG	-	128	27
GGU	0.35	76	26
GGG	0.06	18	7
GGA	0.10	137	36
AUC	0.39	97	28
UAAA	0.25	168	44
UAAG	0.29	37	17
UAAC	0.29	175	50
UAAU	0.31	135	37
UUUUUA	0.25	113	32
UUAUUA	0.13	54	25
UAAUAA	0.10	87	24
UAAUAG	0.07	93	28
UAAUGA	0.09	134	38
AUGUA	0.11	95	31
AUGCUA	0.13	62	18
AUGCUUA	0.10	30	8
AUGCUAA	0.05	64	16
AUGUUAU	0.06	35	21
AUGUUAA	0.09	34	19

[a]Refers to the free, chemically deblocked oligoribonucleotide.

[b]Chromatography system: 1.0 M ammonium acetate/ethanol (50/50, v/v) on Whatman #40 paper.

[c]Calculated from UV spectrophotometric data assuming a 90% hypochromicity factor.

Table 3. Experimental HPLC data of partially blocked[a] and free Oligoribonucleotides.

Compound	RT[b,c] (min) partially blocked	RT[d] (min) free oligomer
GUA		12.1
GUG	18.8	
GUC		11.6
GCC	14.5	
GGU	20.7	8.6[e]
AAC	20.5	19.2

[a]Refers to the 2'-O-tetrahydropyranal derivatives.

[b]Linear Gradient HPLC, 5% v/v acetonitrile/0.1 M NH_4OAc (pH 7.0) to 30% v/v , in 30 min; flow rate 1 ml/min at 1500 psi.

[c]Approximately 10 mgs in a 200 µl pulse.

[d]Gradient HPLC, 0% v/v acetonitrile/0.1 M NH_4OAc (pH 7.0) to 3% v/v, in 30 min; flow rate 1.0 ml/min at 1500 psi.

[e]Gradient HPLC, 0% v/v acetonitrile/0.1 M NH_4OAc (pH 7.0) to 5% v/v, in 20 min; flow rate 1.0 ml/min at 1500 psi.

3.2 Characterization by Proton NMR Spectroscopy of Biologically Active Synthetic Oligoribonucleotides.

The oligomers UAA and UAAX, where X is A, G, C, or U were characterized by their 90 MHz proton magnetic resonance spectra (^{1}H-NMR). Table 4 lists the chemical shifts of all the non-exchangeable base protons and the ribose H-1' protons for these oligomers at 70 ± 1°C. Table 5 lists the chemical shift data for the tandem stop codons UAAUAA, UAAUAG, and UAAUGA. Table 6a and 6b lists the chemical shift data for the sequences AUG(N)$_n$. The shift data at 70 ± 1°C of the dinucleoside monophosphates (dimers) and the trinucleoside diphosphates (trimers) used in these studies are contained in Appendix 1.

Since many of the oligoribonucleotide sequences studied were sythesized sequentially from the 5'-terminus to the 3'-terminus, and the ^{1}H-NMR spectrum of each intermediate was obtained under identical conditions to those of the final product, the incremental analysis method was used in the assignment of the spectrum of the final product (Borer et. al., 1975). Changes in the spectra from one intermediate to the next are due to the appearance of the resonances of the additional nucleotide and its shielding effects on the previously present protons. Shielding effects are observed at the nearest-neighbour and next-nearest-neighbor positions. The addition of a

Table 4. Chemical Shifts[a] of the oligoribonucleotides UAAX at 70°C in D_2O[b]

Proton	UAA	UAAU	UAAC	UAAG	UAAA
U(1)H-6	7.659	7.689	7.696	7.669	7.679
A(2)H-8	8.289	8.317	8.330	8.241	8.292
A(2)H-2	8.126	8.121	8.121	8.185	8.086
A(3)H-8	8.311	8.336	8.330	8.274	8.292
A(3)H-2	8.178	8.189	8.181	8.137	8.103
U(4)H-6		7.753			
C(4)H-6			7.795		
G(4)H-8				7.899	
A(4)H-8					8.292
A(4)H-2					8.162
U(1)H-1'	5.737	5.764	5.765	5.749	5.756
A(2)H-1'	5.963	5.960	5.975	5.910	5.924
A(3)H-1'	6.042	6.058	6.035	5.965	5.971
U(4)H-1'		5.849			
C(4)H-1'			5.834		
G(4)H-1'				5.803	
A(4)H-1'					6.0[illegible]4
U(1)H-5	5.773	5.779	5.777	5.777	5.766
U(4)H-5		5.749			
C(4)H-5			5.910		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 5. Chemical shifts[a] of the tandem stop codons UAAUAA, UAAUAG, and UAAUGA in D_2O[b].

Proton	UAAUAA	UAAUAG	UAAUGA
U(1)H-6	7.652	7.659	7.670
A(2)H-8	8.261	8.273	8.262
A(2)H-2	8.068	8.094	8.074
A(3)H-8	8.252	8.273	8.246
A(3)H-2	8.036	8.081	8.024
U(4)H-6	7.639	7.659	7.628
A(5)H-8	8.227	8.273	
A(5)H-2	8.059	8.111	
G(5)H-8			7.865
A(6)H-8	8.261		8.285
A(6)H-2	8.132		8.123
G(6)H-8		7.915	
U(1)H-1'	5.731	5.749*	5.735
A(2)H-1'	5.916	5.926	5.910
A(3)H-1'	5.954	5.973	5.954
U(4)H-1'	5.745	5.757*	5.752
A(5)H-1'	5.904	5.973	
G(5)H-1'			5.712
A(6)H-1'	6.005		6.027
G(6)H-1'		5.814	
U(1)H-5	5.757	5.771	5.743
U(4)H-5	5.708	5.724	5.674

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

*Assignments could be interchanged.

Table 6a. Chemical shifts of the Oligoribonucleotides $AUG(N)_m$ in D_2O[b] at 70°C.

Proton	AUGUUU[c]	AUGUA	AUGUUA[d]	AUGUUAA	AUGUUAU
A(1)H-8	8.237	8.244	8.305	8.278	8.250
A(1)H-2	8.140	8.170	8.257	8.145	8.176
U(2)H-6	7.728	7.724	7.748	7.728	7.719
G(3)H-8	7.921	7.929	7.977	7.934	7.939
U(4)H-6	7.773	7.715	7.779	7.749	7.742
U(5)H-6	7.793		7.748	7.677	7.694
A(5)H-8		8.333			
A(5)H-2		8.170			
U(6)H-6	7.823				
A(6)H-8			8.427	8.235	8.362
A(6)H-2			8.235	8.068	8.184
A(7)H-8				8.296	
A(7)H-2				8.145	
U(7)H-6					7.753
A(1)H-1'	5.987	6.001	6.041	6.002	6.007
U(2)H-1'	5.754[e]	5.833*	5.835*	5.830*	5.827[f]
G(3)H-1'	5.812[e]	5.833*	5.844	5.830*	5.838[f]
U(4)H-1'	5.880	5.740	5.889	5.815	5.854
U(5)H-1'	5.897*		5.835*	5.770	5.800
A(5)H-1'		6.051			
U(6)H-1'	5.897*				
A(6)H-1'			6.090	5.948	6.076
A(7)H-1'				6.025	
U(7)H-1'					5.865
U(2)H-5	5.728[f]	5.785	5.810	5.783	5.782
U(4)H-5	5.741[f]	5.875	5.828*	5.815	5.810*
U(5)H-5	5.832		5.828*	5.783	5.768
U(6)H-5	5.863				
U(7)H-5					5.810*

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.
[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.
[c]Chemical shifts for AUGUUU are reported at 31.7°C.
[d]Chemical shifts for AUGUUA are for the oligomer in D_2O only.
[e,f]Assignments could be interchanged.
*Denotes coincident peaks.

Table 6b. Chemical shifts of the Oligoribonucleotides $AUG(N)_m$ in D_2O[b] at 70°C.

Proton	AUGCUA	AUGCUUA	AUGCUAA	AUGUCU
A(1)H-8	8.276	8.281	8.274	8.264
A(1)H-2	8.241	8.215	8.160	8.199
U(2)H-6	7.757*	7.756	7.741	7.726
G(3)H-8	7.961	7.962	7.935	7.952
C(4)H-6	7.757*	7.830	7.671	
U(4)H-6				7.772
U(5)H-6	7.757*	7.775	7.741	
C(5)H-6				7.810
U(6)H-6		7.712		7.797
A(6)H-8	8.397		8.241	
A(6)H-2	8.263		8.092	
A(7)H-8		8.399	8.298	
A(7)H-2		8.260	8.160	
A(1)H-1'	6.032	6.030	5.989	6.017
U(2)H-1'	5.852[e]	5.851	5.835*	5.839*
G(3)H-1'	5.858[e]	5.840	5.835*	5.839*
U(4)H-1'				*5.875
C(4)H-1'	5.913	5.976	5.847	
U(5)H-1'	5.858	5.918	5.775	
C(5)H-1'				*5.875
U(6)H-1'		5.816		*5.875
A(6)H-1'	6.084		5.954	
A(7)H-1'		6.115	6.033	
U(2)H-5	5.817	5.807[f]	5.788*	5.796
U(4)H-5				5.858
C(4)H-5	5.896	5.914	5.880	
U(5)H-5	5.817	5.837[f]	5.788*	
C(5)H-6				6.024
U(6)H-5		5.791		5.911

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.
[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.
[e,f]Assignments could be interchanged.

3'-terminal pyrimidine residue causes an observable next-next-nearest-neighbour effect (Table 4).

The resonances of most of the trimers could be assigned by comparison to the resonances of dimers, as in the incremental analysis method, with the following guidelines. Resonances of the non-exchangeable base and ribose H-1' protons could be divided into two groups; a low-field group (8.4-7.5ppm) containing AH-8, AH-2, GH-8, CH-6 and UH-6 resonances; and a high-field group (6.1-5.3ppm) containing CH-5, UH-5 and the ribose H-1' signals (Figures 9 and 10). Within these two groups several "rules" of assignment (at 70 $\pm$ 2°C) hold for all the oligomers so far investigated. For the low-field group, the signals from low to high field were in the order AH-8, AH-2, GH-8, and U,CH-6. Within the higher-field group, the AH-1' signal was the lowest field signal, followed in general by the CH-5 and the remaining H-1' and UH-5 signals. The pyrimidine H-5 and H-6 protons displayed temperature invariant $^{3}J_{5,6}$ couplings of 7.6 Hz (C) and 8.2 Hz (U). Ribose H-1' resonances had splittings $^{3}J_{1'2'}$, (mostly due to coupling with H-2') which were temperature sensitive and largest in magnitude (4.5-6.0 Hz) at high temperatures (Everett, et al.,1980)

In some cases, assignments could not be made unambiguously by just considering the spectra of incremental intermediates. This is particularly important in the

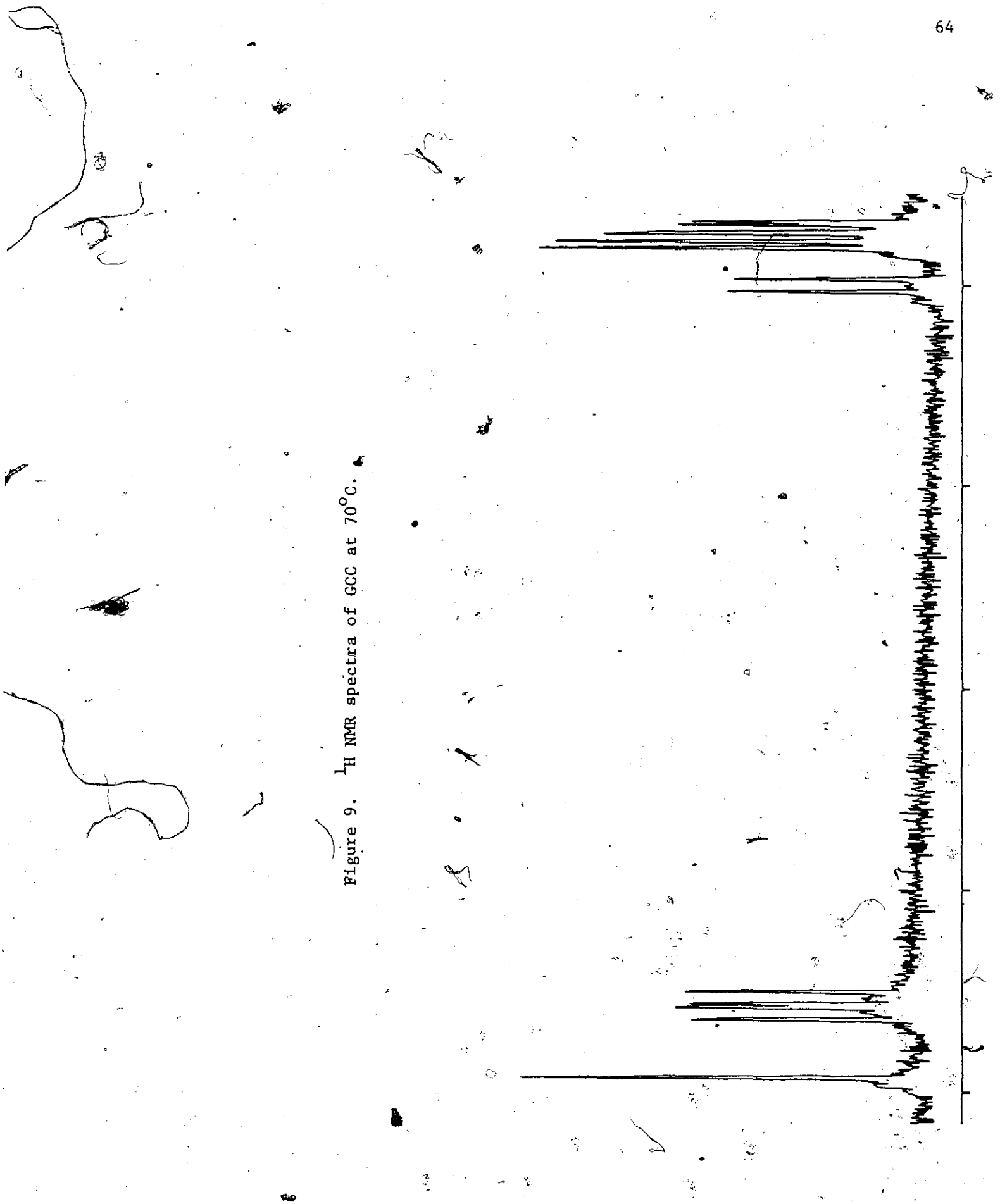
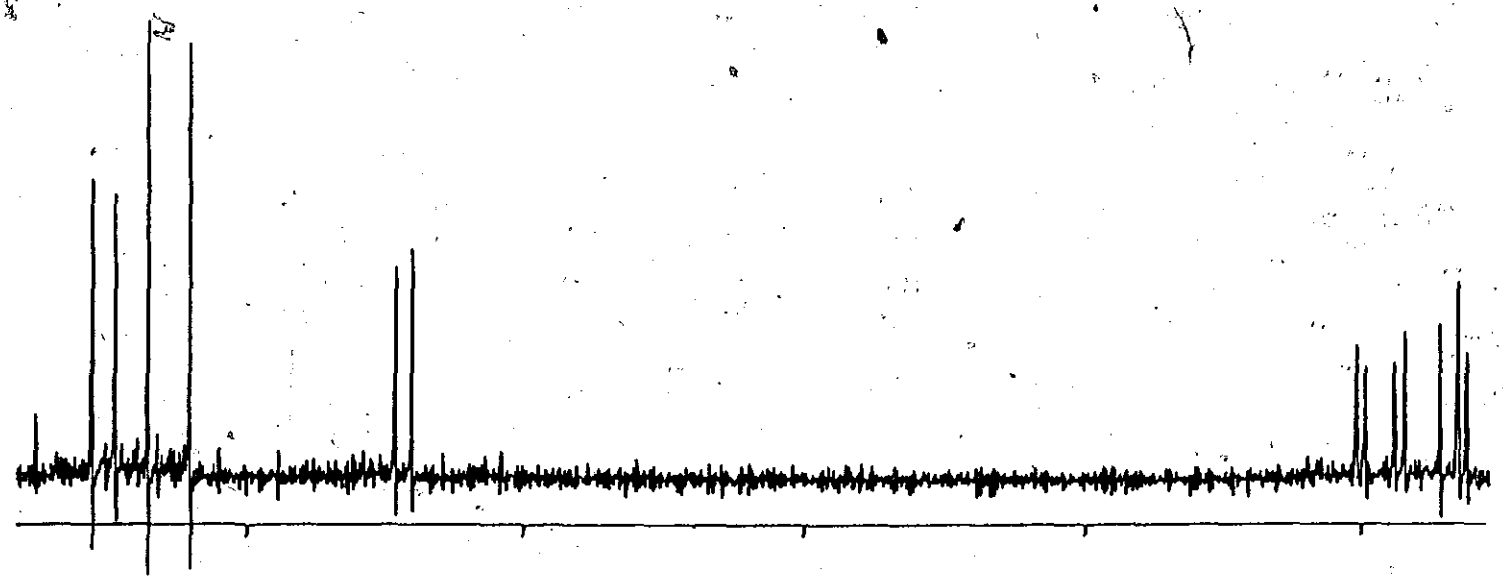


Figure 9. ^{1}H NMR spectra of GCC at 70°C.

Figure 10. ^{1}H NMR spectra of AAC at 70°C



assignment of the spectrum of oligomers which were synthesized by a block coupling procedure. In these cases assignments were made by considering the assignments in related oligomers, and by considering observed shielding trends (Everett, et. al., 1980; Bell, et. al., 1981; Hader, et. al., 1981). For example, the assignment of the spectrum of UAAUGA was facilitated by considering the assignments of resonances in the spectra of oligomers UAA, UAAU, UAAUAA, AUGA, and UGA. The rationale being that neighbours more than three bases away from a particular residue do not contribute significantly to the observed chemical shift at 70 °C (Hader, et. al., 1981). With this in mind the chemical shifts of the U(1) in UAA,UAAU, or UAAUAA, should be good approximations of the chemical shifts of U(1) in UAAUGA. This is in fact found. The U(1)H-6 (7.659ppm) in UAA and the U(1)H-6 (7.689) in UAAU suggests that the UH-6 resonance at 7.670ppm in the spectrum of UAAUGA corresponds to the U(1)H-6 proton, while the resonance at 7.628ppm is assigned to U(4)H-6 considering (U(2)H-6 at 7.655 ppm in AUG), and that the proton would be further shielded by additional 5' and 3' next-nearest neighbour adenosine residues. Other assignments were made considering similar arguments.

3.3 Prediction of Proton Chemical Shifts in Oligoribonucleotides.

The proton NMR chemical shifts of the non-exchangeable base and ribose H-1' protons in oligoribonucleotides CAU, AUA, UAU, AUG and CAUAUG (Table 7) were recorded at 70°C to avoid multistrand aggregation (Ts'o, et al.,1969; Martin, et al., 1971) and to avoid the different conformations presumably present at lower temperatures which may contribute to the observed chemical shift changes by allowing non-neighbouring bases to come into close proximity (Lee and Tinoco, 1980; Stone, et al., 1981). Thus at high temperatures (>70°C) it is assumed that RNA molecules of any sequence have a similar distribution between conformational microstates; that is, under these conditions the oligoribonucleotides are largely unstacked and exhibit high conformational flexibility. Evidence supporting this conclusion is that in all oligomers so far investigated the $^{3}J_{1'2'}$ coupling constants observed at all of the H-1' signals are large at 70°C (greater than 4.5Hz), reflecting a low degree of base stacking (Lee, et al.,1976; Ezra, et al.,1977; Dhingra & Sarma,1978). Therefore at 70°C the changes in chemical shifts of the various protons on a given base residue are considered to be caused exclusively by the ring anisotropy

of neighbouring bases (Giessner-Prettre, et al.,1976). For the chemical shift changes to be useful in predicting the chemical shifts of a proton in a given sequence the ring current anisotropic effects would have to be different not only for purines and pyrimidines, but also between each base type. This is in fact observed. The addition of a nucleotide pX to the 3'-end of a dinucleoside monophosphate produces an X-dependent shielding of the dimer protons, which decreased in the order A>G>C>U (Everett, et al.,1980).

To demonstrate how the observed chemical shifts of protons in trimers can be used to predict the chemical shifts of protons in longer oligomers, the predicted and observed chemical shifts of the oligomer CAUAUG will be compared. The predicted chemical shifts of the oligomer CAUAUG were determined by averaging the chemical shifts of overlapping trimers which together would make up the longer oligomer. By averaging the chemical shifts of protons in related trimers nearest neighbour and next-nearest neighbour ring current effects can be approximated. The effects for both 3' and 5' neighbours are accounted for in this manner. For example, the resonances of the protons in the third uracil residue of CAUAUG are predicted by averaging the chemical shifts observed in the trimers CAU, AUA and UAU for the uracil residue. The 5' nearest (adenosine) and next-nearest (cytosine) neighbour ring current shieldings.

are reflected in the chemical shifts of the uracil protons in CAU. Similarly, the uracil protons in trimer AUA reflect a 5' nearest neighbour effect of adenosine and a 3' nearest neighbour effect of adenosine. In trimer UAU a 3' nearest (adenosine) and next-nearest (uracil) neighbour are reflected in the chemical shifts of the uracil U(1), protons.

Thus, by averaging the observed shifts of the uracil protons in the trimers CAU, AUA and UAU the chemical shifts of the protons in the third uracil residue of CAUAUG are found. Similar averages can be calculated for the other base residue protons (see Table 7). The predicted chemical shifts of the protons are found to be in excellent agreement with those observed.

3.4 Calculation of Proton Chemical Shifts in Oligoribonucleotides.

With the assumption that at 70°C oligoribonucleotides of any sequence have the same average conformation in solution, one can compile a set of proton NMR chemical shift parameters that can be used to calculate the chemical shifts in other oligomers (Everett, et al.,1980; Bell, et al.,1981) A set of parameters derived from the 16 dimers and 64 trimers has been used to calculate in terms of a generalized sequence IpJpKpLpM, the chemical

Table 7. Chemical Shifts (in ppm) of the Nonexchangeable Aromatic and Anomeric Protons in the Oligoribonucleotides CAU, AUA, UAU, AUG and CAUAUG at 70 ± 1 °C.

	Nucleotide Sequence						
	Observed				Predicted[a]	Obsd.	
Proton	123 CAU	234 AUA	345 UAU	456 AUG	123456 CAUAUG	123456 CAUAUG	Diff.[b]
C(1)H-6	7.687				7.687	7.643	0.044
A(2)H-8	8.382	8.241			8.312	8.326	0.014
A(2)H-2	8.250	8.157			8.204	8.154	0.050
U(3)H-6	7.744	7.722	7.698		7.721	7.685	0.036
A(4)H-8		8.320	8.388	8.266	8.325	8.310	0.015
A(4)H-2		8.174	8.241	8.198	8.204	8.131	0.073
U(5)H-6			7.766	7.725	7.746	7.685	0.061
G(6)H-8				7.957	7.957	7.945	0.012
C(1)H-1'	5.786				5.786	5.764	0.022
A(2)H-1'	6.088	5.988			6.038	6.002	0.036
U(3)H-1'	5.858	5.892	5.768		5.841	5.804	0.037
A(4)H-1'		6.038	6.100	6.019	6.052	6.002	0.050
U(5)H-1'			5.821	5.837	5.829	5.822	0.007
G(6)H-1'				5.837	5.837	5.832	0.005
C(1)H-5	5.925				5.925	5.894	0.031
U(3)H-5	5.745	5.778	5.839		5.787	5.736	0.051
U(5)H-5			5.782	5.767	5.775	5.751	0.024

[a]Predicted values are determined by averaging the shift values observed in related trinucleoside diphosphates.
[b]Differences equal to or less than 0.04 ppm are considered not to be significant as differences in temperature and concentration between trimer data can lead to changes in chemical shifts equal to 0.02 ppm. Differences greater than 0.04 ppm could be caused by long range ring current effects of bases, not accounted for in trinucleoside diphosphates where only nearest and next-nearest neighbour effects can be observed. Evidence for this is the fact that proton chemical shifts in tetranucleoside triphosphates are better approximations of those in longer sequences.

shift of a proton on base K assuming only first and second neighbour effects. The calculation is considered as Equation 1.

$$\delta^{\text{Proton}}_{K} = \delta^{\text{Proton}}_{K,\text{mon}} + \Delta^{3'}_{1,pL} + \Delta^{3'}_{2,pM} + \Delta^{5'}_{1,Jp} + \Delta^{5'}_{2,Ip} \quad \ldots\ldots 1.$$

The Δ_1 terms are nearest (first)-neighbour shielding effects, and the Δ_2 terms are next-nearest (second)-neighbour shieldings. The resonance of the proton in the nucleoside at 70°C is denoted $\delta_{K,mon}$. There are four types of first-neighbour effects giving four $\Delta^{3'}{}_1$ values (L = A,G,C, or U) for each proton on K. Similarly there are four values for each of the empirical $\Delta^{3'}{}_2$, $\Delta^{5'}{}_1$, and $\Delta^{5'}{}_2$ parameters. Thus a table with sixteen values should allow the prediction of the chemical shift for a particular proton on a given residue.

If the hypothesis about the average conformation is reasonable, then a workable assumption would be that the same empirical parameters $\Delta^{3'}{}_{1,2}$ and $\Delta^{5'}{}_{1,2}$ could be used to calculate the chemical shifts for a H-1' proton on any of the four bases. Parameters derived from the 64 trinucleoside diphosphates are to a first approximation correct; however, calculated chemical shifts fit observed values more exactly if parameters derived for a H-1' proton

on a particular base (for example a H-1' on adenosine), are used only in predicting the chemical shifts of this proton (see Sec. 4.5). Similarly parameters derived for the pyrimidine H-5's, pyrimindine H-6's and purine H-8's are to a first approximation equal in magnitude. Including the data for the adenosine H-2 proton there are 5 sets of 16 $\Delta^{3'}_{1,2}$ and $\Delta^{5'}_{1,2}$ empirical parameters. Thus proton chemical shift information from the trinucleoside diphosphates can be compiled into eighty (80), parameters that with the chemical shifts of the base and ribose H-1' protons of the four major ribonucleosides (Table 8; there are eleven protons, see Figure 6), allow to a first approximation the prediction of the chemical shifts for each proton in any sequence.

For example the chemical shift parameters $\Delta^{3'}_{1}$ for protons on an adenosine residue influenced by a 3'-neighbour guanosine can be quantitated by comparing the proton chemical shifts in adenosine (A), to the proton chemical shifts in the dinucleoside monophosphate ApG, where $\Delta^{3'}_{1}$ for the adenosine H-8 proton would be found by subtracting the chemical shift value in the monomer from the value in the dimer Equation 2.

Table 8. Chemical Shift Assignments[a] and Coupling Constants for the Nucleosides at 70°C in D_2O[b].

	A		G		C		U	
Proton		$J_{5,6}$ $J_{1'2'}$		$J_{5,6}$ $J_{1'2'}$		$J_{5,6}$ $J_{1'2'}$		$J_{5,6}$ $J_{1'2'}$
	ppm	Hz	ppm	Hz	ppm	Hz	ppm	Hz
AH-8	8.313	-						
AH-2	8.260	-						
AH-1'	6.065	5.62						
GH-8			7.961	-				
GH-1'			5.888	5.62				
CH-6					7.781	7.34		
CH-1'					5.894	3.66		
CH-5					6.073	7.32		
UH-6							7.798	8.06
UH-1'							5.887	4.88
UH-5							5.907	8.06

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

$$\Delta^{3'}_{1} = \delta^{AH-8}_{ApG} - \delta^{AH-8}_{A} \qquad \ldots\ldots\ldots 2.$$

δ^{AH-8}_{ApG}, is the chemical shift of the AH-8 in dimer ApG, and δ^{AH-8}_{A}, is the chemical shift of the AH-8 proton in the mononucleoside adenosine.

The $\Delta^{3'}_{1}$ parameter for the influence of 3'-neighbour guanosine can also be determined by comparison of sequences such as XpA and XpApG. Second neighbour parameters $\Delta^{3'}_{2}$ and $\Delta^{5'}_{2}$ values are obtained by comparison of sequences such as ApX and ApXpG for $\Delta^{3'}_{2}$ values and XpA and GpXpA for $\Delta^{5'}_{2}$ values, where X is any of the four major ribonucleosides. For example the $\Delta^{3'}_{2}$ value for an adenosine H-8 proton is given by equation 3.

$$\Delta^{3'}_{2} = \delta^{A(1)H-8}_{ApCpG} - \delta^{A(1)H-8}_{ApC} \qquad \ldots\ldots\ldots 3.$$

This relationship is derived from expressions (4) and (5) by elimination of the terms corresponding to the chemical shift of the GH-8 in the mononucleoside, and the $\Delta^{3'}_{1}$ term (see Figure 11). In this way the parameter values are always obtained from two experimental numbers (Hader, et

Table 9. Proton NMR Chemical Shift Parameters as Determined From a set of Trinucleoside Diphosphates.

Proton	Base	$\Delta^{3'}_{1}$	$\Delta^{3'}_{2}$	$\Delta^{5'}_{1}$	$\Delta^{5'}_{2}$
H-8	A	-.081	-.048	-.011	-.050
	G	-.058	-.028	.016	-.029
	C	-.009	-.002	.053	-.006
	U	.002	.002	.061	.005
H-2	A	-.127	-.075	-.048	-.081
	G	-.058	-.036	-.028	-.055
	C	-.027	-.013	.005	-.005
	U	-.008	-.007	.009	-.012
H-6	A	-.086	-.043	-.020	-.047
	G	-.065	-.029	.012	-.026
	C	-.006	.002	.049	-.011
	U	.004	-.005	.041	.011
H-5	A	-.064	-.031	-.115	-.046
	G	-.042	-.031	-.087	-.018
	C	-.032	-.027	-.023	-.009
	U	-.012	-.001	.006	.022
H-1'	A	-.094	-.046	-.014	-.043
	G	-.054	-.029	.030	-.020
	C	-.015	-.009	.027	-.008
	U	.004	-.004	.046	.004

Parameters are in ppm (a negative sign indicates shielding and a positive sign deshielding), and are derived from spectra recorded at 70 ± 1°C in D_2O; 0.01 M NaH_2PO_4, 1.0 M NaCl (pD 7.0).

al., 1982).

Individual parameters are then combined according to type (e.g. H-1', H-5, etc.) and averaged. Entries in Table 9, are the averages of four or more values. This procedure gave experimental justification for the assumptions, and while there were variations in individual parameters derived from the different oligoribonucleotides, the majority of the parameters are within a range of ±0.007 ppm (probable error of the mean).

Application of the set of parameters is straight-forward and is illustrated by computation of the predicted chemical shift of A(2)H-8 in the pentaribonucleotide,CpApApUpG. The shift, given by equation (1) is expected at 8.274ppm which agrees favorably with the observed shift at 8.255ppm (difference 0.019ppm). Calculated and observed ^{1}H chemical shifts of the protons in ACGU and CAUG, are contained in Table 10. The oligoribonucleotides contain the same base ratio, but they have different sequences and the observed and calculated chemical shifts are unique to each sequence. Agreement of 0.04ppm or better between calculated and observed shift is regarded as acceptable (footnote e, Table 10), and as can be noted, the agreement is very good. This is further emphasized by the agreement between the predicted chemical shifts and those observed in the oligoribonucleotides CCGGp,CCGGAp,ACCGGp,CCGGUp,and ACCGGUp (Table 11) where the

Table 10. Calculated ^1H Chemical Shifts[a,b] of the Protons in ACGU, and CAUG and their Differences[c] from Observed[d] Values.

ACGU				CAUG			
Proton	Cal.	Obs.	Diff.[e]	Proton	Cal.	Obs.	Diff.
A(1)H-8	8.287	8.264	-0.023	C(1)H-6	7.705	7.665	-0.040
A(1)H-2	8.181	8.180	-0.001	A(2)H-8	8.340	8.354	0.014
C(2)H-6	7.714	7.679	-0.035	A(2)H-2	8.025	8.196	-0.009
G(3)H-8	7.971	7.942	-0.029	U(3)H-6	7.719	7.699	-0.020
U(4)H-6	7.804	7.771	-0.033	G(4)H-8	7.970	7.962	-0.008
A(1)H-1'	6.036	6.006	-0.030	C(1)H-1'	5.793	5.773	-0.020
C(2)H-1'	5.820	5.845	0.025	A(2)H-1'	6.062	6.043	-0.019
G(3)H-1'	5.860	5.845	-0.015	U(3)H-1'	5.801	5.822	0.021
U(4)H-1'	5.908	5.881	-0.029	G(4)H-1'	5.874	5.835	-0.039
C(2)H-5	5.898	5.861	-0.037	C(1)H-5	5.999	5.905	-0.094
U(4)H-5	5.775	5.795	0.020	U(3)H-5	5.740	5.746	0.006

[a] Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b] 1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

[c] Differences are observed-calculated values.

[d] The observed shift assignments were made independently using the method of incremental analysis; i.e. comparison of the spectra of AC, ACG, and ACGU; and of the spectra of CA, CAU, and CAUG.

[e] The value (0.042) is a maximum additive error that could be obtained when a calculated value is determined for a proton on a residue which is influenced by a complete set of neighbours.

Table 11. Observed[a] and Calculated[b] Chemical Shifts for Nonexchangeable Protons at High Temperature (°C).

	CCGGp		CCGGAp		ACCGGp		ACCGGUp	
	(73±1)		(74±1)		(72±)		(84±1)	
Proton	calc.	obsd.	calc.	obsd.	calc.	obsd.	calc.	obsd.
5'								
AH-8					8.323	8.299	8.323	8.298
AH-2					8.204	8.205	8.204	8.230
AH-1'					6.059	6.049	6.059	6.052
CH-6	7.762	7.764	7.762	7.757	7.751	7.724	7.751	7.728
CH-5	5.986	6.002	5.986	5.982	5.870	5.871	5.870	–
CH-1'	5.865	5.880	5.865	5.848	5.846	5.848	5.846	–
CH-6	7.751	7.732	7.751	7.738	7.710	7.700	7.710	7.704
CH-5	5.948	5.984	5.948	5.982	5.901	5.943	5.901	–
CH-1'	5.824	5.849	5.824	5.848	5.776	5.802	5.776	–
GH-8	7.949	7.932	7.888	7.914	7.949	7.925	7.951	7.904
GH-1'	5.841	5.782	5.789	5.760	5.841	5.848	5.853	–
GH-8	7.978	7.998	7.883	7.893	7.978	7.970	7.981	7.970
GH-1'	5.909	5.900	5.796	5.732	5.909	5.878	5.909	–
UH-6							7.780	7.780
UH-5							5.759	5.866
UH-1'							5.887	5.978
AH-8			8.300	8.372				
AH-2			8.144	8.215				
AH-1'			6.065	6.104				
3'								

[a] In ppm from TSP; 1.0 M NaCl, 99.8% D_2O.

[b] Calculated from parameters determined by Hader et al., 1982; ppm from DSS; 1.0 M NaCl, 100% D_2O, 70±1°C.

Taken from Petersheim and Turner, 1982a.

^{1}H chemical shifts of these oligomers were recorded in a different laboratory using a different buffer (Petersheim & Turner, 1982a).

To eliminate some of the drudgery involved in calculating the chemical shifts of all the aromatic and anomeric protons in a given oligoribonucleotide a short Fortran program ASSIGN has been written which determines the proton chemical shifts from the oligomer sequence. The algorithm determines the neighbours for a given base residue in the sequence, from the 3' to the 5' end. The parameter value for the specific base neighbour and proton is then added to the chemical shift value for the proton on the residue in question. Residues near or at the end of a sequence are considered to have neighbours with parameters that do not contribute to the chemical shift; that is, the parameter value is zero. The information returned to the user is a printout of the sequence and the calculated proton chemical shifts for the protons on each residue, in the order H-1', H-8 or H-6 and H-2 or H-5. A listing of the program is contained in Appendix 2.

3.5 Calculation of Proton Chemical Shifts in Oligoribonucleotides using an Extended Set of Parameters.

In the derivation of the set of shift parameters to predict chemical shifts of oligoribonucleotides, three assumptions were made. To improve the accuracy of the calculated chemical shifts the assumptions were modified. For oligomers which are self complementary and have reasonably high Tms the assumption that they will be completely destacked at 70°C is not correct and predicted chemical shifts will poorly represent observed shifts. In these cases the calculated values reflect the chemical shifts that would be observed at higher temperatures where the oligomer is destacked and in a random conformation (for example see AAGCUA Table 12).

The assumption that all anomeric protons (H-1') are "equivalent" in the sense that their chemical shifts are affected similarly by neighbouring bases regardless of the nucleosidyl moiety to which they are bound can be relaxed. By considering all protons separately the accuracy of chemical shift prediction can be improved. For example the effect of a 3'-nearest-neighbour adenosine on a guanosine H-1' proton is to shield it by -0.134 ppm (negative sign indicates shielding), similarly the adenosine H-1' proton is shielded -0.135 ppm by a 3'-nearest-neighbour adenosine. In contrast the same adenosine shields a cytosine H-1' by -0.058 ppm and a uridine H-1' by -0.069 ppm. The

Table 12. Chemical Shifts[a] of XAGCYZ Series of RNA Oligomers at 70°C in D_2O[b]

Proton	123 AGC	1234 AGCU	12345 AGCUA	A12345 AAGCUA[c]	A1234 AAGCU
A(A)H-8				8.196	8.201
A(A)H-2				8.114	8.084
A(1)H-8	8.242	8.248	8.229	8.250	8.224
A(1)H-2	8.180	8.176	8.144	8.128	8.119
G(2)H-8	7.926	7.929	7.929	7.891	7.844*
C(3)H-6	7.744	7.737	7.704	7.693	7.713*
U(4)H-6		7.768*	7.698*	7.705	7.767*
A(5)H-8			8.343	8.338	
A(5)H-2			8.186*	8.189*	
A(A)H-1'				5.900	5.891
A(1)H-1'	5.974	5.978	5.965	5.958	5.930*
G(2)H-1'	5.812*	5.791*	5.798*	5.766*	5.710*
C(3)H-1'	5.886	5.903	5.869*	5.869*	5.869
U(4)H-1'		5.892	5.817*	5.834*	5.869
A(5)H-1'			6.063	6.064	
C(3)H-5	5.876	5.850	5.839	5.884*	5.788*
U(4)H-5		5.821	5.793	5.806	5.806*

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

[c]The spectrum of Oligomer AAGCUA was recorded at 85°C.

*Indicates Predicted value is outside of the observed chemical shift by greater than 0.040 ppm.

Table 13. Shielding Parameters for the Influence of Neighbouring Nucleotide residues (X), on the ^1H Chemical Shifts of Adenosine Aromatic and Anomeric Protons in Oligoribonucleotides.

Proton	X	$\Delta_1^{3'}$	$\Delta_2^{3'}$	$\Delta_3^{3'}$	$\Delta_1^{5'}$	$\Delta_2^{5'}$	$\Delta_3^{5'}$
H-8	A	-.093	-.059	-.016	-.011	-.053	-.014
	G	-.059	-.034	-.010	.016	-.029	-.018
	C	-.013	-.001	.003	.064	.010	.005
	U	.001	.000	.002	.074	.011	.015
H-2	A	-.127	-.075	-.028	-.048	-.081	-.016
	G	-.058	-.036	-.009	-.028	-.055	-.029
	C	-.027	-.013	-.002	.005	-.005	.001
	U	-.008	-.007	.005	.009	-.012	.008
H-1'	A	-.129	-.056	-.015	-.028	-.050	-.014
	G	-.081	-.035	-.005	.023	-.025	-.015
	C	-.020	-.006	.005	.022	-.003	.001
	U	-.001	-.008	-.001	.038	.001	.008

Parameters are in ppm (a negative sign indicates shielding and a positive sign deshielding) and are derived from spectra recorded at $70\pm1^{\circ}C$ in D_2O containing 0.01 M sodium phosphate buffer and 1.0 M sodium chloride (pD 7.0).

Table 14. Shielding Parameters for the Influence of Neighbouring Nucleotide residues (X), on the ^{1}H Chemical Shifts of Guanosine Aromatic and Anomeric Protons in Oligoribonucleotides.

Proton	X	$\Delta_1^{3'}$	$\Delta_2^{3'}$	$\Delta_3^{3'}$	$\Delta_1^{5'}$	$\Delta_2^{5'}$	$\Delta_3^{5'}$
H-8	A	-.069	-.037	-.010	-.010	-.047	-.020
	G	-.056	-.022	.00	.015	-.028	-
	C	-.004	.004	.001	.041	-.021	.003
	U	-.004	.003	.000	.048	-.002	-.003
H-1'	A	-.126	-.056	-.008	-.041	-.058	-.010
	G	-.074	-.031	-	-.002	-.029	-
	C	-.018	-.009	.010	.004	-.013	.004
	U	.000	-.002	-.002	.020	-.006	-.004

Parameters are in ppm (a negative sign indicates shielding and a positive sign deshielding) and are derived from spectra recorded at 70±1°C in D_2O containing 0.01 M sodium phosphate buffer and 1.0 M sodium chloride (pD 7.0).

Table 15. Shielding Parameters for the Influence of Neighbouring Nucleotide residues (X), on the ^{1}H Chemical Shifts of Cytosine Aromatic and Anomeric Protons in Oligoribonucleotides.

Proton	X	$\Delta_1^{3'}$	$\Delta_2^{3'}$	$\Delta_3^{3'}$	$\Delta_1^{5'}$	$\Delta_2^{5'}$	$\Delta_3^{5'}$
H-6	A	-.097	-.043	-.020	-.021	-.039	-.022
	G	-.076	-.034	.001	.015	-.019	-.006
	C	-.014	.004	.004	.059	-.003	-.007
	U	-.002	-.009	.010	.048	.022	-
H-5	A	-.088	-.048	-.020	-.132	-.041	-.037
	G	-.060	-.037	-.007	-.117	-.010	-.028
	G	-.052	-.054	-.006	-.031	-.002	-.002
	U	-.021	-.009	-.007	.013	.031	-
H-1'	A	-.052	-.035	-.010	.013	-.028	-.019
	G	-.022	-.030	-.005	.066	-.017	-.016
	C	-.026	-.014	.003	.039	-.005	-.004
	U	.003	.001	.007	.074	.014	-

Parameters are in ppm (a negative sign indicates shielding and a positive sign deshielding) and are derived from spectra recorded at 70±1°C in D_2O containing 0.01 M sodium phosphate buffer and 1.0 M sodium chloride (pD 7.0).

Table 16. Shielding Parameters for the Influence of Neighbouring Nucleotide residues (X), on the ^{1}H Chemical Shifts of Uridine Aromatic and Anomeric Protons in Oligoribonucleotides.

Proton	X	$\Delta_1^{3'}$	$\Delta_2^{3'}$	$\Delta_3^{3'}$	$\Delta_1^{5'}$	$\Delta_2^{5'}$	$\Delta_3^{5'}$
H-6	A	-.075	-.042	-.002	-.018	-.054	-.016
	G	-.054	-.024	.004	.008	-.032	-.011
	C	.003	.000	.034	.039	-.018	.006
	U	.009	.000	.013	.033	.000	.016
H-5	A	-.039	-.014	-.005	-.097	-.050	-.018
	G	-.023	-.025	.001	-.056	-.026	-.009
	C	-.011	.001	.000	-.015	-.015	-.007
	U	-.002	.008	.002	-.001	.013	.011
H-1'	A	-.070	-.038	-.001	-.001	-.034	-.022
	G	-.040	-.018	.006	.034	-.010	-.003
	C	.003	-.006	.026	.043	-.011	-.002
	U	.014	-.005	.025	.053	.008	.020

Parameters are in ppm (a negative sign indicates shielding and a positive sign deshielding) and are derived from spectra recorded at 70±1°C in D_2O containing 0.01 M sodium phosphate buffer and 1.0 M sodium chloride (pD 7.0).

differences between purine and pyrimidine shieldings is a trend found throughout the chemical shifts of oligoribonucleotides used to determine the shielding parameters (Tables 13-16), and reflects theoretically calculated ring current anisotropy (Giessner-Prettre, et al., 1976).

Other differences between shielding effects of neighbouring nucleotide residues demonstrate that the calculated chemical shifts will more correctly reflect the observed chemical shifts if the eleven protons (aromatic and anomeric), on the four major nucleosides are considered separately. For instance the effect of a 5'-nearest-neighbour guanosine is to shield a cytosine H-5 proton by -0.147 ppm, while the same guanosine shields a uridine H-5 proton by -0.056 ppm, a difference of 0.091 ppm. An average of these numbers used to predict the H-5 proton chemical shifts for both uridine and cytosine causes incorrect predictions.

The first approximation which calculates chemical shifts assuming only first and second neighbours contribute to the observed chemical shifts can be refined by allowing for third neighbour shielding effects. The evidence for the third neighbour shieldings can be observed in tetranucleoside triphosphates (tetramers), and are significant for the purine residues (Table 13).

The extended set of empirical chemical shift parameters includes data from the 16 dimers, the 64 codons

Figure 11. Equations used in the determination of Chemical Shifts

Relationship 3 is derivable from expressions 4 and 5 for $\delta_{ApC}^{A(1)H-8}$ in the dimer and trimer by elimination of δ_{A}^{AH-8} and $\Delta^{3'}$.

$$\delta_{ApC}^{A(1)H-8} = \delta_{A}^{AH-8} + \Delta_{1}^{3'} \qquad 4.$$

$$\delta_{ApCpG}^{A(1)H-8} = \delta_{A}^{AH-8} + \Delta_{1}^{3'} + \Delta_{2}^{3'} \qquad 5.$$

In this way the parameters are always obtained from two experimental numbers.

For the generalized sequence HpIpJpKpLpMpN the chemical shift of a proton on residue K, δ_{K}^{Proton}, is given by equation (6).

$$\delta_{K}^{Proton} = \delta_{Mon,\,K}^{Proton} + \Delta_{1}^{3'} + \Delta_{2}^{3'} + \Delta_{3}^{3'} + \Delta_{1}^{5'} + \Delta_{2}^{5'} + \Delta_{3}^{5'} \qquad 6.$$

Where δ_{K}^{Proton}, is the chemical shift of the proton on the mononucleoside K, (K = U, C, G, A in actual cases), and $\Delta^{3'}_{1,2,3}$ and $\Delta^{5'}_{1,2,3}$ are the 3' and 5' shieldings of the first, second and third neighbouring nucleotide residues.

(XpYpZ), as well as data from 36 tetramers. This extended set of shielding values includes 264 averaged shieldings as opposed to 80 values in the first approximation (where certain protons were considered equivalent). Calculated chemical shifts using the extended parameter set more accurately reflect the observed chemical shifts of a particular RNA oligomer. The chemical shifts are calculated by a modified equation (1), which includes the additional parameters $\Delta^{3'}_{3}$ and $\Delta^{5'}_{3}$ to quantitate the 3' and 5' third neighbour effects (see Figure 11). For example the $\Delta^{3'}_{3}$ shieldings are determined by subtracting the chemical shifts of the adenosine residues in AXYG from those in AXY, where XY are residues common to both sequences but can vary. In this way the 3'-third-neighbour effect of a guanosine residue on adenosine protons is found.

Since calculations of chemical shifts in oligomers longer than tetramers can be tedious, a modified version of the Fortran program ASSIGN has been written which uses the extended set of parameters, and will calculate the predicted shifts from the oligomer sequence.

3.6 Sequencing of Short RNA Oligomers

The method developed by Sanger and coworkers for the sequencing of short RNA strands (less than 20 nucleotides in length) has been used to establish most known RNA sequences especially those of tRNA molecules. The technique has the

advantage of allowing one to determine a sequence with as little as 0.5ug of sample. The sequence of the oligonucleotide is determined by the angular mobility shifts. In the first dimension, the relative mobilities of two oligonucleotide intermediates in a homologous partial digestion series which differ by a single nucleotide will depend both on the pKa of the nucleotide by which they differ and on the size and base composition of the sequence common to them. The second dimension resolves an oligonucleotide mixture on the basis of size, shorter oligonucleotides moving faster than longer ones. Also a difference of one purine residue between two successive intermediates in a series of partial degradation products causes a larger (a factor of 1.5-2.5) mobility shift than a difference of one pyrimidine residue. The end result in two dimensions is that the stepwise removal of nucleotides in a homologous partial degradation series can be "read" as angular mobility shifts between successively shorter fragments. The mobility shifts of all four nucleotides are characteristic and identifiable (Silberklang, et al., 1979).

The most common complications in "reading" two dimensional homochromatography patterns arise from the presence of extraneous spots in the autoradiogram or from the presence of ambiguous or "unusual" mobility shifts. Extraneous spots may either result from radioactive impurities present in the oligonucleotide as eluted from the

Figure 12. ^{1}H NMR spectrum of GGU at $70^{o}C$.

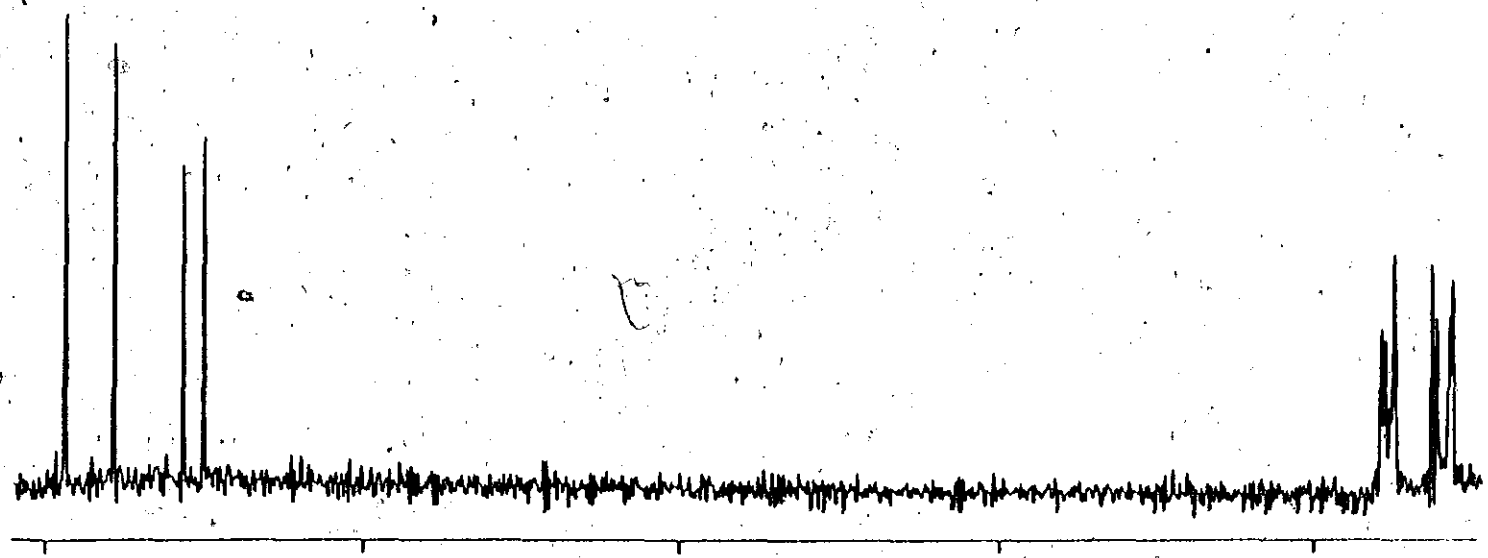
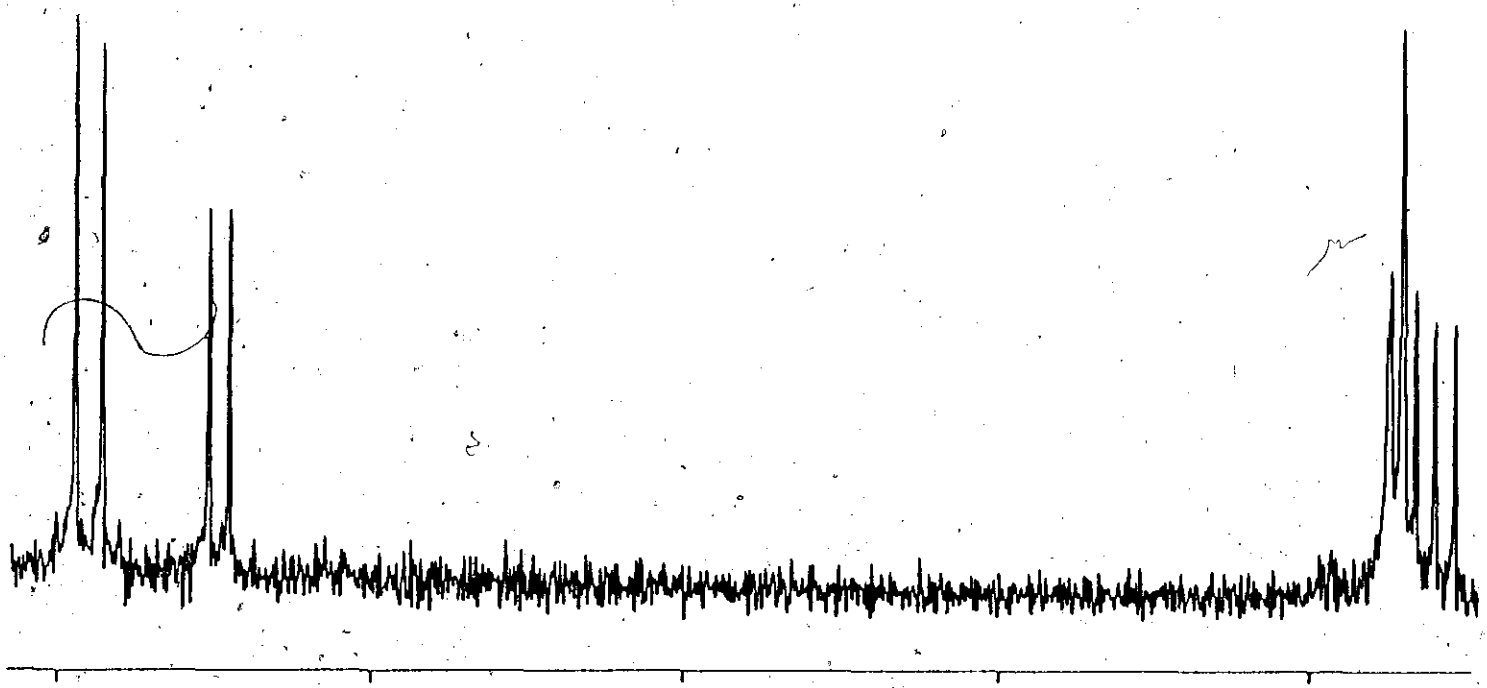


Figure 13. ^{1}H NMR spectrum of GUG at $70^{\circ}C$



finger print, or be generated during partial digestion by contaminating endonucleases. Unusual mobility shifts in the two-dimensional homochromatograph patterns of partial digests may indicate the presence of modified nucleosides (Silberklang, et al., 1979).

I present, for the first time, a spectroscopic method for determining the sequence of an RNA oligomer. This method is not intended to replace well established methods but to complement them because it is independent methodologically.

In the previous chapters a method has been presented that allows one to determine from a given RNA sequence the expected chemical shifts of the aromatic and anomeric protons. The resonances of these protons exhibit changes in their particular chemical shifts which depend on the bases immediately surrounding a particular residue in a given sequence. For instance the spectra of GUG and GGU are quite different (Figures 12 and 13). Furthermore, the sequence effects on chemical shifts of protons on a particular residue do not extend significantly beyond the third-(next-next-nearest) neighbour when the oligomer is completely destacked (Hader, et al., 1981). This fact allows the calculation of the chemical shifts of the protons in longer oligomers from the chemical shift information of short oligomers such as trimers.

Therefore at high temperatures and low

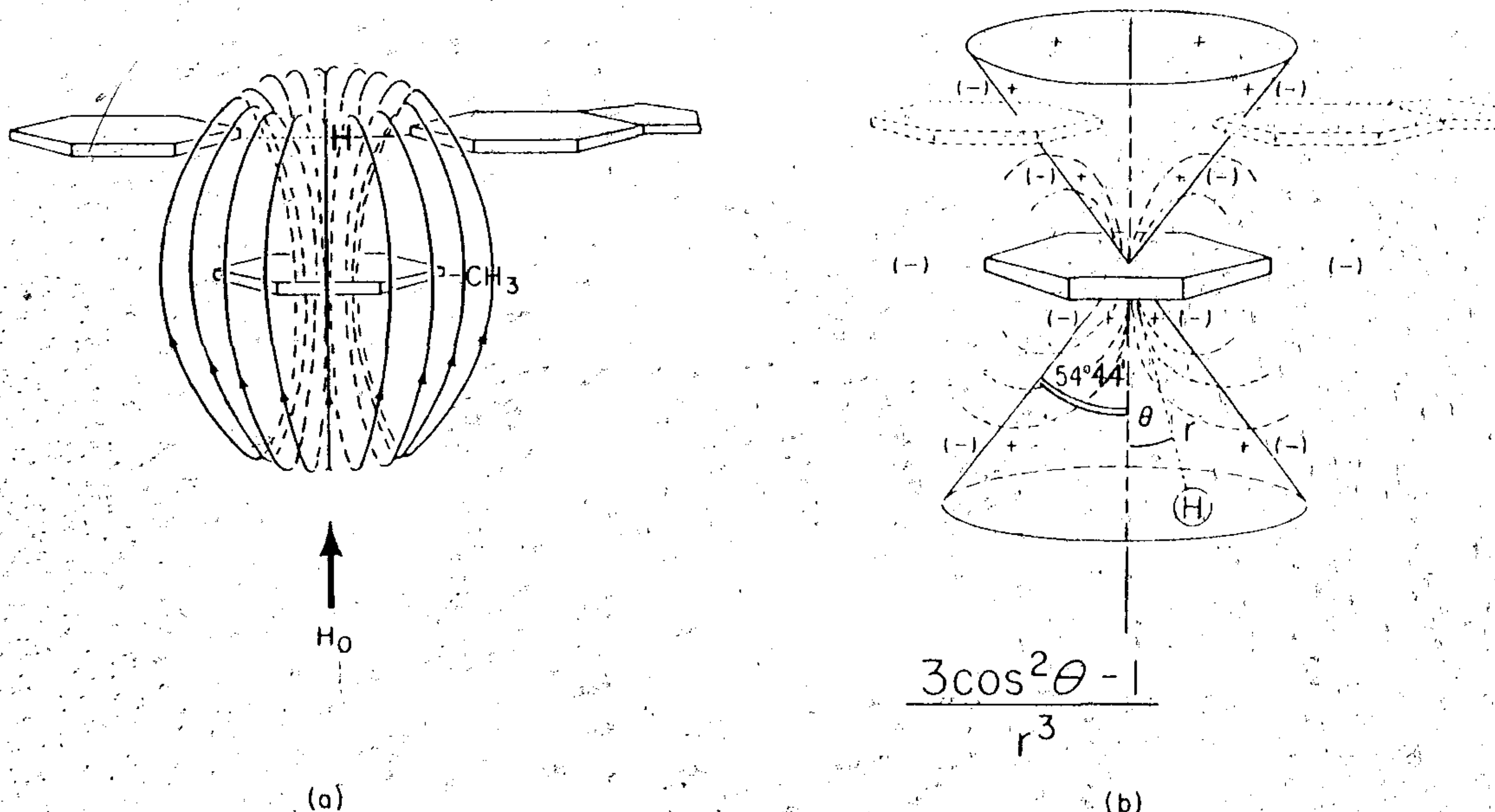


FIGURE 14. (a) Diagrammatic conceptualisation of the induced magnetic field from the ring current of a pyrimidine stacked below an adjacent base pair. (b) Illustration of the distance and angle dependence of the ring current shift for a proton located at the point H.

concentrations, changes in chemical shift of the various protons on a given nucleoside residue are considered to be caused mainly by ring current anisotropy (Figure 14) of neighbouring residues (Giessner-Prettre, et al., 1976). For the chemical shift changes to be useful in sequencing, the ring current anisotropic effects must be different not only for purines and pyrimidines but also between each base type. This is in fact observed, the addition of a nucleotide pX to the 3'-end or Xp to the 5'-end of a dimer produces an X-dependent shielding of the dimer protons, which decreased in the order A>G>C>U (Everett, et al., 1980; Hader et al., 1981).

To demonstrate how the chemical shift parameters can be used to calculate shifts of individual protons, and hence determine the sequence of oligoribonucleotides, the predicted chemical shifts of the oligomers CAUAUG, AUGCUA, UCAUGA, and AAGCUU were determined by solving the appropriate chemical shift equation for the protons. The predicted and observed chemical shifts for the sequences (Tables 17a and 17b) were compared and found to be in good agreement. Values different by less than ±0.042 ppm are considered accurate predictions. Although all four sequences contain the same base ratio, 2A:1G:1C:2U, the calculated chemical shifts for each sequence are unique. Furthermore, the calculated chemical shifts correspond to the observed chemical shifts which are also unique for a given oligomer

Table 17a. Calculated ^1H Chemical Shifts[a,b] of the Protons in CAUAUG, and AUGCUA and their Differences[c] from Observed[d] Values.

CAUAUG				AUGCUA			
Proton	Cal.	Obs.	Diff.	Proton	Cal.	Obs.	Diff.
C(1)H-6	7.643	7.705	0.062	A(1)H-8	8.276	8.286	0.010
A(2)H-8	8.326	8.309	-0.017	A(1)H-2	8.214	8.205	-0.009
A(2)H-2	8.154	8.157	0.003	U(2)H-6	7.757	7.730	-0.027
U(3)H-6	7.685	7.710	0.025	G(3)H-8	7.961	7.976	0.015
A(4)H-8	8.310	8.295	-0.015	C(4)H-6	7.757	7.775	0.018
A(4)H-2	8.131	8.120	-0.011	U(5)H-6	7.757	7.741	-0.016
U(5)H-6	7.685	7.738	0.053	A(6)H-8	8.397	8.373	-0.024
G(6)H-8	7.945	7.970	0.025	A(6)H-2	8.236	8.249	0.013
C(1)H-1'	5.764	5.793	0.029	A(1)H-1'	6.032	6.042	0.010
A(2)H-1'	6.002	6.033	0.031	U(2)H-1'	5.852	5.801	-0.051
U(3)H-1'	5.804	5.767	-0.037	G(3)H-1'	5.858	5.880	-0.022
A(4)H-1'	6.002	6.028	0.026	C(4)H-1'	5.913	5.875	-0.038
U(5)H-1'	5.822	5.813	-0.009	U(5)H-1'	5.858	5.773	-0.085
G(6)H-1'	5.832	5.874	0.042	A(6)H-1'	6.084	6.099	0.015
C(1)H-5	5.894	5.999	0.105	U(2)H-5	5.817	5.706	-0.111
U(3)H-5	5.736	5.716	-0.020	C(4)H-5	5.896	5.895	-0.001
U(5)H-5	5.751	5.773	0.022	U(5)H-5	5.817	5.793	-0.024

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

[c]Differences are observed-calculated values.

[d]The observed shift assignments were made independently using the method of incremental analysis; i.e. comparison of the spectra of AU, AUG, AUGC and AUGCUA; and of the spectra of UG, AUG, UAUG, AUAUG, and CAUAUG (Hader *et al.*, 1981).

Table 17b. Calculated ^{1}H Chemical Shifts[a,b] of the Protons in UCAUGA, and AAGCUU and their Differences[c] from Observed[d] Values.

UCAUGA				AAGCUU			
Proton	Cal.	Obs.	Diff.	Proton	Cal.	Obs.	Diff.
U(1)H-6	7.709	7.745	0.036	A(1)H-8	8.196	8.188	0.008
C(2)H-6	7.663	7.744	0.081	A(1)H-2	8.103	8.083	0.020
A(3)H-8	8.339	8.344	0.005	A(2)H-8	8.237	8.231	0.006
A(3)H-2	8.182	8.191	0.009	A(2)H-2	8.128	8.103	0.025
U(4)H-6	7.678	7.665	-0.013	G(3)H-8	7.867	7.902	-0.035
G(5)H-8	7.889	7.875	-0.014	C(4)H-6	7.717	7.781	-0.064
A(6)H-8	8.320	8.330	0.010	U(5)H-6	7.769	7.851	-0.082
A(6)H-2	8.160	8.199	0.039	U(6)H-6	7.791	7.836	-0.045
U(1)H-1'	5.911	5.829	-0.082	A(1)H-1'	5.941	5.929	0.012
C(2)H-1'	5.833	5.827	-0.006	A(2)H-1'	5.941	5.979	-0.038
A(3)H-1'	6.016	6.074	0.058	G(3)H-1'	5.741	5.827	-0.086
U(4)H-1'	5.813	5.749	-0.064	C(4)H-1'	5.893	5.879	0.014
G(5)H-1'	5.787	5.761	-0.026	U(5)H-1'	5.880	5.886	-0.006
A(6)H-1'	6.057	6.098	0.041	U(6)H-1'	5.893	5.921	-0.028
U(1)H-5	5.806	5.810	0.004	C(3)H-5	5.827	5.860	-0.033
C(2)H-5	5.958	5.985	0.027	U(4)H-5	5.827	5.832	-0.005
U(4)H-5	5.748	5.687	-0.061	U(5)H-5	5.884	5.892	-0.008

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

[c]Differences are observed-calculated values.

[d]The observed shift assignments were made independently using the method of incremental analysis; i.e. comparison of the spectra of AGC, AGCU, AAGCU, and AAGCUU; and of the spectra of UG, AUG, CAUG, UCAUG and UCAUGA (Hader et al., 1981).

Moreover a fitting of the observed chemical shifts, from the oligomer AUGCUA, for instance, to the chemical shifts of all possible sequences containing the same base ratio determines that the shifts calculated for the sequence AUGCUA best fit the observed set of chemical shifts thus determining the oligomers sequence to be AUGCUA.

In the general case, to determine the sequence of an oligomer from its proton NMR spectrum the procedure would include the following steps. The spectrum of the oligomer to be sequenced would be obtained as previously described. The oligoribonucleotide base ratio is immediately available from the spectrum, because resonances from protons, specfic for a given base, are found consistently in similar regions of the spectrum. This information is also used to assign resonances to protons by type. For instance, in the spectrum of AUGCUA, the resonances at 8.276 and 8.397 ppm are AH-8 protons, the resonances at 8.236 and 8.214 ppm are AH-2 protons, the resonance at 7.961 ppm is a GH-8 proton, while the doublet centered at 7.757 ppm corresponds to three protons, two UH-6 and a CH-6. In the case of CH-6 and UH-6 protons the proton type is discerned by the $^{3}J_{5,6}$ coupling which is 7.6 Hz for cytosine H-6 and H-5 protons, and 8.2 Hz for uridine H-6 and H-5 protons. Similarly the doublet resonances in the higher field group are assigned by type as AH-1' then CH-1', UH-1' and GH-1' and CH-5 and UH-5. The H-1' and H-5 resonances are discerned by their coupling

Table 18. Comparison of Sequence Predicted by Fitting of the Observed Chemical Shifts to those Calculated for Oligomers of the Same Base Ratio.

Sequence		Fitting Criteria		
Actual	Predicted	# >0.042 (ppm)	Sum of Diff (>0.042)	Absolute average Diff.
AAGCUU	AAGCUU	0	0.00	0.020
CAUAUG	CAUAUG	3	0.076	0.025
AUGCUA	AUGCUA	1	0.014	0.017
UCAUGA	CAUGUA	2	0.077	0.021
	UCAUGA	6	0.213	0.039
AUGUCU	AUGUCU	0	0.00	0.021
AUGUA	AUGUA	2	0.077	0.026
AUGA	AUGA	0	0.00	0.018
CAUG	CAUG	1	0.004	0.019
ACGU	ACUG	2	0.054	0.030
	ACGU	3	0.065	0.032
AGCU	AGCU	0	0.0	0.023
AGUCU	AGUCU	2	0.043	0.023
AGCCU	AGCCU	2	0.036	0.021
AGGCU	AGGCU	2	0.047	0.024
AGACU	AGACU	1	0.015	0.019

Fitting criteria values were determined by using the program SEQUEN (see Appendix).

constants. In general U and G, H-1' resonances are arbitrarily assigned. At this point resonances are assigned to a base type (A,G,C,U), but the positions of the bases in the oligomer are not known.

To determine the order of the bases in the oligomer the chemical shifts of the resonances are compared to those generated for a given sequence. Resonances then become fully assigned reflecting not only residue type but its position in the sequence, based on minimizing the difference between the observed and predicted chemical shift. The correct sequence is found by minimizing three criteria which reflect the best fit. The first criterion minimized is the number of calculated and observed chemical shifts that differ by greater than 0.042 ppm. A second related criterion minimized is the sum of difference between observed and calculated values that do differ by greater than 0.042 ppm. When this number is minimized the overall fit between calculated and observed chemical shifts is better. The final criterion is the absolute average difference, which reflects the average difference between the calculated and observed chemical shifts. This value is also a minimum when the set of calculated chemical shifts best fit the observed set. Table 18 gives a sample of the sequences which have been confirmed by this method. In most cases, the known sequence was also the sequence which was determined from the fitting process. Oligomers which were

not correctly sequenced point to some of the limitations of the method.

Oligomer length can be a limitation for two reasons. The number of signals in a particular spectrum that overlap lead to inaccurate assignment of resonances to proton type. This problem becomes critical for oligomers longer than ten residues. For oligomers containing more than fifteen nucleotidyl residues the calculated chemical shifts must be very accurate for unique sequence determination because of the enormous number of sequences possible for a given base ratio (see Table 19).

Oligomer UCAUGA, when sequenced by this method, gives CAUGUA, and the difference between the two sequences, in terms of fitting values is significant. The reason for this, is that oligomer UCAUGA is self complementary and has a relatively high Tm. Thus the chemical shift data at 70 °C used for the sequence determination is still influenced by interstrand stacking interactions, and therefore gives an incorrect sequence determination. In situations where a unique sequence cannot be determined other spectral information, such as sigmoidal chemical shift changes with temperature, indicating base pairing which limits the possible sequences or information from a proton NMR NOE experiment could be used to arrive at a unique sequence (Petershiem and Turner, 1982b).

The advantages of the method are that it is fast,

Table 19. The Number of Possible Sequences for a Base Ratio as a Function of the Oligomer Length.

Nucleotydal Residues	Protons (Maximum)[a]	Sequences Possible[b]
4	11	24
5	14	60
6	17	180
7	20	630
8	22	2,520
9	25	7,560
10	28	25,200
11	31	92,400
12	33	369,600
13	36	1,201,200
14	39	4,204,200
15	42	15,765,750
20	55	1.17327×10^{10}

[a]Aromatic and anomeric protons on each of the four major bases A, G, C, U.

[b]Calculated for the base ratio which gives the maximum number of permutations.

one step, does not destroy the sample, is independent methodologically from the existing chemical and enzymatic procedures, and can involve little interpretation, since the sequence determination process is computerized. Generally a ^1H NMR spectrum of an oligomer (300-500 µg/400 µL) can be obtained in 10 to 20 minutes. Although the present sample size is large, being two or three orders of magnitude larger than other sequencing techniques (Silberklang, et al., 1979), advancements in NMR instrumentation, especially probe sensitivity, will dramatically reduce the sample size in the near future (Levy and Craik, 1981). Sample size can also be reduced with the use of microsample NMR tubes, which reduce the needed volume to 60 µL and sample to 45-75 µg. The present sample size is convenient for checking oligomers from chemical synthesis.

Oligomers that are longer than six residues will in general have many possible sequences for a given base ratio. For example the base ratio 2A:2G:1C:2U, gives 630 sequences. To determine the correct sequence by manual inspection is too time consuming. To shorten this procedure a Fortran program SEQUEN has been written. The program uses the program ASSIGN as a subroutine to generate the chemical shifts of the sequences that are generated by a subroutine from the given base ratio. The program compares the calculated chemical shifts to the observed set of chemical shifts and calculates the fit in terms of the three values previously discussed. The sequence can then be determined

by quick visual inspection.

3.7 Prediction of Chemical Shifts in Short DNA Strands.

The improved technology to obtain significant quantities of DNA oligomers by both chemical (Gait, et al., 1982) and enzymatic means has led in recent years to the increased use of NMR spectroscopy as a tool in determining the detailed molecular behaviour of these biologically important compounds. The ability to assign the chemical shifts of protons in individual nucleotide units of a DNA oligomer is essential if NMR is to be used to its full advantage (Cruz, et al., 1982; Patel, et al., 1982e). For the first time I report the successful application of an extended set of RNA parameters to the prediction of the chemical shifts of similar protons in DNA oligomers.

It is particularly difficult to make chemical shift assignments for individual protons in oligomers which are composed of similar repeating units. DNA with four monomeric nucleotides (dA,dG,dC,T) is a typical example of this kind of oligomer. Recent papers reporting on the ^{1}H NMR spectra of DNA oligomers up to the dodecamer level, at the highest available field strengths have been quite unable to assign the majority of the resolved heterobase and anomeric protons (Patel, et al., 1982a,b,c,d). The only

practical approach to making chemical shift assignments of DNA compounds to date has been the method of incremental analysis (Borer, et al., 1975). The disadvantage of this method is that it requires the spectroscopist to have available in sufficient quantity the precursor fragments of the oligomer under study.

In constructing a set of shift parameters for calculating the chemical shifts of RNA oligomers in the unstacked state, it was reasoned that, while DNA and RNA duplexes have significant differences (Pardi, et al., 1981), at higher temperatures (where strands are destacked), DNA single strands should be very similar to the corresponding RNA single strands. A wide range of conformational states are expected and the averaged magnetic effects of neighbouring nucleotides on a given proton in a mobile DNA strand should be similar in magnitude to those in RNA strands.

Accordingly, the ^1H chemical shifts of the base and anomeric protons of dA, dG, dC, and T, were recorded in buffered D_2O solutions (Table 20). These values applied with the extended set of RNA shift parameters (see Section 4.5), were used to calculate the expected chemical shifts of protons in a number of DNA oligomers. Calculations were carried out using equation (7), where $\delta_{K,mon}$ is the chemical shift of the appropriate monomeric deoxynucleoside protons and the parameters $\Delta^{3'}_{1,2,3}$ and $\Delta^{5'}_{1,2,3}$

Table 20. ^1H Chemical Shifts[a] in ppm for Deoxynucleosides at 70°C.

Proton	dA	dG	dC	T
AH-8	8.280			
AH-2	8.212			
AH-1'	6.431			
GH-8		7.945		
GH-1'		6.263		
CH-6			7.779	
CH-5			6.065	
CH-1'			6.237	
TH-6				7.587
T-Me				1.887
TH-1'				6.245

[a]Shifts are referenced to DSS, using the methyl absorption of t-BuOD at 1.321 ppm and are recorded in 100% D_2O containing 0.01 M sodium phosphate and 1.0 M sodium chloride (pD 7.0).

represent the shift contributions from the three neighbours on the 3' and 5' sides respectively. The parameters were derived from the chemical shifts of RNA species (see Section 4.5).

Since thymidine does not exist in RNA strands, parameters for the uridine H-5 proton were used to predict the chemical shifts of C-5 methyl protons. This assignment was justified on the grounds that the averaged magnetic environment of the three methyl protons should be equivalent to a point slightly beyond the position of the uridine H-5 proton but in the plane of the pyrimidine ring. The thymidine methyl would therefore experience magnetic effects similar to a uridine H-5 from neighbouring residues.

To test the application of the RNA parameters to DNA oligomers, compounds were chosen for which high temperature (>70 °C) chemical shift data was available and where reasonable chemical shift assignments have been made. The detailed results for dACATGT (Tran-Dinh, et al., 1982), dCGCGAATTCGCG (Patel, et al., 1982a), and dAATTGCAATT (Lancelot, et al., 1981) are shown in Tables 21 and 22. A few of the protons show differences between the predicted and observed values that are greater than 0.10 ppm but the majority are much below this level. The mean absolute differences of 0.041 ppm for dACATGT, 0.043 for dCGCGAATTCGCG, and 0.051 for dAATTGCAATT, are remarkable when it is noted that the shifts of these three oligomers

Table 21. Calculated and Observed ^{1}H NMR Shifts of d-ACATGT and d-CGCGAATTCGCG[a]

ACATGT				CGCGAATTCGCG[c]			
Proton	Cal.	Obsd.	Diff.	Proton	Cal.	Obsd.	Diff.
A(1)H-8	8.222	8.22	0.002	C(1)H-6	7.720	7.62	0.10
C(2)H-6	7.685	7.63	0.055	G(2)H-8	7.963	7.94	0.023
A(3)H-8	8.270	8.30	-0.030	C(3)H-6	7.681	7.70	-0.019
T(4)H-6	7.491	7.40	0.091	G(4)H-8	7.860	7.77	0.090
G(5)H-8	7.949	7.94	0.009	A(5)H-8	8.250	8.27	-0.020
T(6)H-6	7.582	7.57	0.012	A(6)H-8	8.248	8.14	0.108
A(1)H-2	8.116	8.14	-0.024	T(7)H-6	7.529	7.50	0.029
C(2)H-5	5.844	5.86	-0.016	T(8)H-6	7.574	7.55	0.024
A(3)H-2	8.099	8.18	-0.081	C(9)H-6	7.755	7.70	0.055
T(4)Me	1.739	b		G(10)H-8	7.969	7.94	0.029
T(6)Me	1.885	b		C(11)H-6	7.745	7.70	0.055
A(1)H-1'	6.361	6.31	0.051	G(12)H-8	7.964	7.94	0.024
C(2)H-1'	6.178	6.12	0.058	C(1)H-5	5.942	5.97	-0.028
A(3)H-1'	6.370	6.35	0.020	C(2)H-5	5.842	5.90	-0.059
T(4)H-1'	6.165	6.06	0.105	A(5)H-2	8.037	8.02	0.017
G(5)H-1'	6.176	6.13	0.046	A(6)H-2	8.057	8.00	0.057
T(6)H-1'	6.224	6.24	-0.016	T(7)Me	1.741	1.75	-0.009
				T(8)Me	1.796	1.783	0.013
				C(9)H-5	5.931	6.01	-0.079
				C(10)H-5	5.911	5.94	-0.029
Mean Absolute Diff.			0.041				
				Mean Absolute Diff.			0.042

[a]For spectra recorded in aqueous solutions relative to DSS as standard and at temperatures of 70°C or greater.

[b]Chemical shifts and assignments not quoted by authors (Patel, 1976).

[c]Patel, et al., 1982a.

Table 22. Calculated and Observed ^{1}H NMR Shifts of d-AATTGCAATT[a].

	AATTGCAATT[b]		
Proton	Cal.	Obsd.	Diff.
A(1)H-8	8.196	8.14*	0.056
A(2)H-8	8.261	8.30*	-0.039
T(3)H-6	7.549	7.50	0.049
T(4)H-6	7.501	7.44	0.061
G(5)H-8	7.931	8.07	-0.139
C(6)H-6	7.699	7.63	0.069
A(7)H-8	8.247	8.17	0.077
A(8)H-8	8.300	8.30	0.00
T(9)H-6	7.537	7.49	0.047
T(10)H-6	7.555	7.57	-0.015
A(1)H-2	8.079	8.130	-0.051
A(2)H-2	8.120	8.05	0.070
T(3)Me	1.745	1.76	-0.015
T(4)Me	1.764	1.83	-0.066
C(6)H-5	5.829	5.91	-0.081
A(7)H-2	8.030	8.05	-0.020
A(8)H-2	8.122	8.05	0.072
T(9)Me	1.756	1.76	-0.004
T(10)Me	1.817	1.85	-0.033
Mean Absolute Difference			0.051

[a]For spectra recorded in aqueous solutions relative to DSS as standard and at temperatures of 70°C or greater.

[b]Lancelot, et al., 1981.

*Assignments have been reversed from those given by the authors.

were recorded in different laboratories and under conditions of slightly different buffer concentrations, temperature, and pH. Moreover, each of the RNA parameters used has a probable experimental error of the order of 0.007 ppm and for a base which is influenced by six neighbours, the combined error could be up to 0.042 ppm.

Shift predictions have been made for other DNA oligomers that have high temperature ^{1}H NMR spectra reported in the literature and the mean absolute differences between the predicted and assigned shifts are shown in Table 23. The agreement between calculated and assigned chemical shifts is excellent. Upon examining the data one notes the very limited number of H-1' protons that have been assigned. Frequently these anomeric protons are well resolved resonances and the parameter set enables reasonable assignments to be made with confidence and efficiency.

Thymidine methyl resonances showed good agreement between calculated and observed values with a mean absolute difference of 0.032 ppm for the 16 examples and the rationale for employing the parameters for uridine H-5 is justified.

The RNA parameter set may also be used in the reverse sense to sequence short DNA strands. If the ^{1}H NMR chemical shifts and base composition are given, the base sequence which best fits the NMR data can be derived. Computer trials have indicated that the current parameter set accurately selects the correct sequence for short DNA

Table 23. Differences Between Calculated and Observed ^1H NMR Shifts of DNA Oligomers[a].

Oligomer	Mean Difference	Reference
dCGCG	0.016	Patel, 1976
dGGCC	0.018	Patel, 1977
dCCGG	0.022	Patel, 1977
dCGCGCG	0.047	Patel, 1976
dCCAA	0.069	Cheng, et al., 1982
dTTGG	0.072	Cheng, et al., 1982
dCCAAG	0.059	Cheng, et al., 1982
dCCA	0.029	Cheng, et al., 1982
dCTTGG	0.041	Cheng, et al., 1982
dTGG	0.066	Cheng, et al., 1982
dATGT	0.040	Neuman, et al., 1982
dTATATA	0.056	Phillips, Roberts, 1980

[a]For spectra recorded at 70°C or greater.

oligomers up to the octamer level but that longer oligomers have base ratios that lead to an enormous number of permutations, and unique sequence prediction is limited by the accuracy of the parameters.

3.8 The Effect of Base Sequence and Conformation on Protein Chain Termination In Vitro.

Protein chain termination occurs when the translational machinery encounters a nonsense codon UAA, UAG, or UGA (AGA or AGG in mitochondria (Anderson, et al., 1981)). In the simplest biochemical assay, the initiator of translation, fmet-tRNA, is bound to ribosomes with AUG and the separate nonsense codons UAA, UAG, or UGA are added in conjunction with either of two release proteins: RF-1 which responds to UAA and UAG, and RF-2 which responds to UAA and UGA (Caskey, 1979). Under these conditions N-formyl-met is hydrolyzed from fmet-tRNA. The release factors RF-1 and RF-2, have been highly purified from E. coli cells using chromatography and assaying for activity using the release of N-formyl-met (Caskey, 1979), or by the proteins ability to promote hydrolysis of ribosome-bound peptidyl-tRNA presumably adjacent to UAA (Ganoza, Ganoza and Nakamoto, 1966; and Ganoza and Tompkins, 1970) or UAG (Klein and Capecchi, 1971).

Release protein RF-1, competes with tRNA mediated suppression of UAG (Ganoza and Tompkins, 1970; Beaudet and

Caskey, 1970) and binds to ribosomes in the presence of nonsense codons (Scolnick and Caskey, 1969). Both RF-1 and RF-2 also stimulate the binding of nonsense codons (Scolnick and Caskey, 1969). Experiments in which antibodies to these proteins were used to remove them from crude extracts suggest that they participate in termination of bacteriophage encoded proteins (Capecchi and Klein, 1970). Thus complexes which include ribosomes, a nonsense codon, and a release factor, are proposed as intermediates in protein chain termination.

If a nonsense codon is inserted into different locations of the bacteriophage T_4 rIIb cistron, so that the sequence around each nonsense triplet varied, dramatic changes in the efficiency of tRNA-mediated suppresion results, suggesting that termination in vivo, may employ a signal longer than a single triplet (Fluck, et al., 1977).

Read-through of UGA can, in more than one case, result in the generation of two proteins essential to viability of a number of phages. Codon, UGA, can be recognized in two ways in strains which do not harbour a nonsense suppressor-tRNA: It can either specify termination, or it can be translated as tryptophan, generating a longer protein essential to function (Hofstetter, et al., 1974; Yates, et al., 1977; Weiner and Weber, 1973; Geller and Rich, 1980). Since the generation of two proteins from the same cistron by this mechanism is uncommon, the base

sequence neighbouring this nonsense codon may control the efficiency of this termination mechanism. Recently, a report suggested that the base following UGA, e.g. UGAA, might determine the level of tRNA-mediated suppression and presumably the extent to which translational termination occurs in vivo, since release factors compete with supressor tRNAs (Engelberg-Kulka, 1981).

Tandem termination codons occur at the end of a number of bacteriophage cistrons but only 10% of all E. coli proteins appear to use contiguous termination triplets in vivo (Lu and Rich, 1971). In such cases, tandem nonsense codons may ensure that termination occurs at the end of the cistron (Nichols, 1970).

In this thesis I address two issues: Whether bases surrounding the nonsense codon UAA can influence in vitro translational termination; and whether UAA must be in the triplet reading frame with AUG for this reaction to take place. This finer point was considered important because UAA can trigger termination when added as a separate triplet, a situation where the spatial relationship between codons AUG and UAA is uncertain. This investigation was accomplished by chemically synthesizing RNA polymers having the UAA codon at variable locations 3' to AUG within the same polymer.

Since all amino acids are found at the C-termini of E. coli proteins without bias (Spahr, et al., quoted in Lu

and Rich, 1971), additional signals might be expected to reside either "downstream" or directly on the 3' side of the stop triplet(s). The ability of UAAA, UAAG, UAAC, or UAAU to stimulate release of N-formyl[^{35}S]methionine from N-formyl[^{35}S]met-tRNA:ribosome:AUG complexes was tested. The results show that each tetranucleotide terminates to almost the same extent and rate as the UAA reference codon when partially purified preparations of RF-1 were used (Figure 15). Therefore one base at the 3' side of UAA does not affect the in vitro termination reaction, indicating that the E. coli recognition system differs from that of eukaryotes, because the latter responds to tetranucleotides better than to triplets (Caskey, 1979).

The capacity of the termination assay to respond to tandem nonsense codons was also studied. In contrast to the tetranucleotides, the hexanucleotides, UAAUAG, UAAUAA, and UAAUGA, stimulated RF-1 mediated release to markedly different extents (Figure 16). Hexamer, UAAUAG, stimulated the reaction somewhat better than UAAUGA, but UAAUAA, was at least five-fold less effective than the other two hexanucleotides or triplet UAA. These difference were maintained when the reaction time was extended up to 60 minutes. The lower level of termination promoted by RF-1 with UAAUAA, could not be explained by the presence of inhibitors in the oligonucleotide preparations, as mixing experiments with UAAUAA and UAA failed to reveal inhibitors

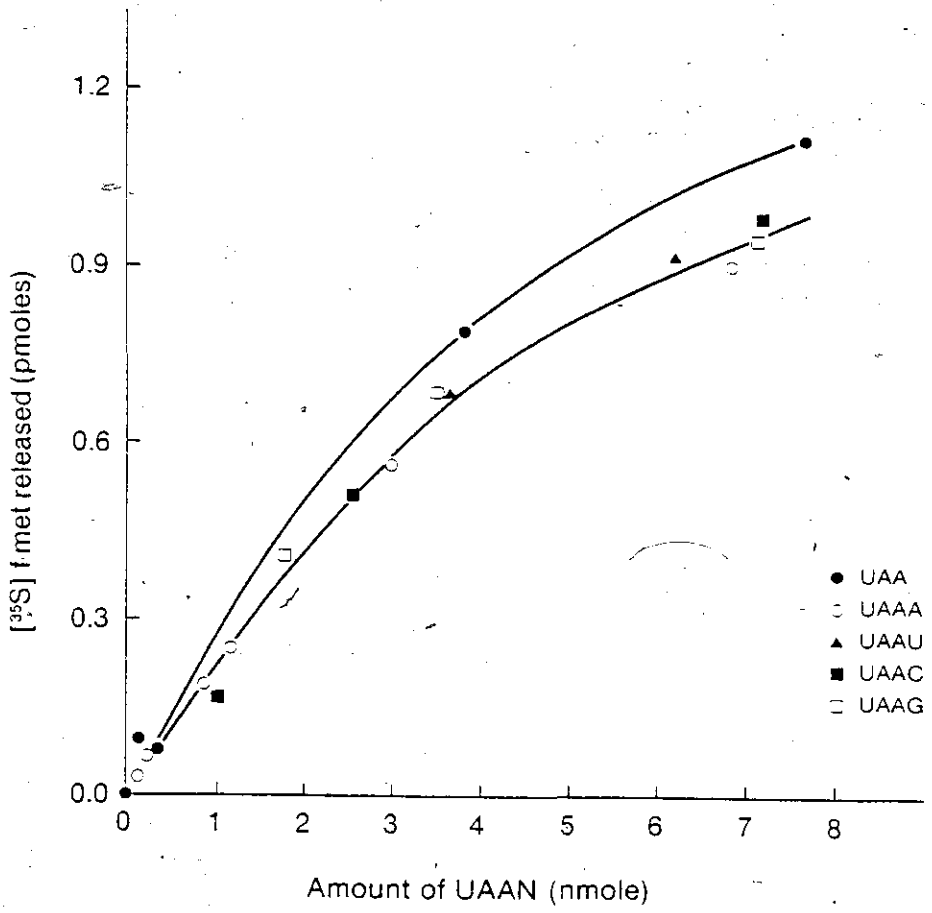


Figure 15. Stimulation of [^{35}S]fmet release by polymers of the form UAAN.

Reactions were carried out as described in Experimental. Incubations contained 300 pmoles of AUG, 100 ug of four times washed E. coli Q13 ribosomes, and 7.0 pmoles of [^{35}S]fmet-tRNA. Termination reaction mixtures (30 min at 24°C) contained 14 ug of RF-1 and varying levels of UAAN. Unspecific release observed with AUG in the presence of RF-1 and no UAAN was subtracted.

in the hexanucleotide. It is of interest that UAAUAA has not yet been observed in prokaryotic termination regions (Caskey, 1979; Steege and Soll, 1979). It seems likely that the conformation of this hexamer restricts recognition of UAA. A previous observation that the $AUGUAA(A)_{25}$ oligomer was unable to stimulate efficient release of N-formyl-methionine (Ganoza and Thach, 1969) while $AUG(U)_3UAA(A)_n$ successfully released fmet-phe, can also be explained by the proposal that the conformation adopted by the oligomer on ribosomes, may influence the termination reaction.

The effect of sequence and conformation cannot be examined using triplet UAA, and since tetranucleotide set, NUAA, is ineffective in termination (Ganoza, unpublished results), two types of longer oligomers were assembled. The first was $AUG(U)_mUAA(A)_n$ having AUG at the 5'-end which was covalently linked to the terminator, UAA, by zero to five bases, and the second was a series of short oligomers (5-7 bases), each starting with AUG; AUGUA, AUGUAA, AUGUUA, AUGCUA, AUGCUAA, AUGCUUA, AUGUUAA, AUGUUAU, AUGAUG, AUGUUU, AUGAGC, AUGGCU, and AUGUCU.

The ability of these oligomers to trigger termination was tested. Each of the first set bound fmet-tRNA to approximately the same extent (Table 24). The bound fmet-tRNA was active in peptide bond synthesis, quantitatively yielding an fmet-puromycin product (Ganoza;

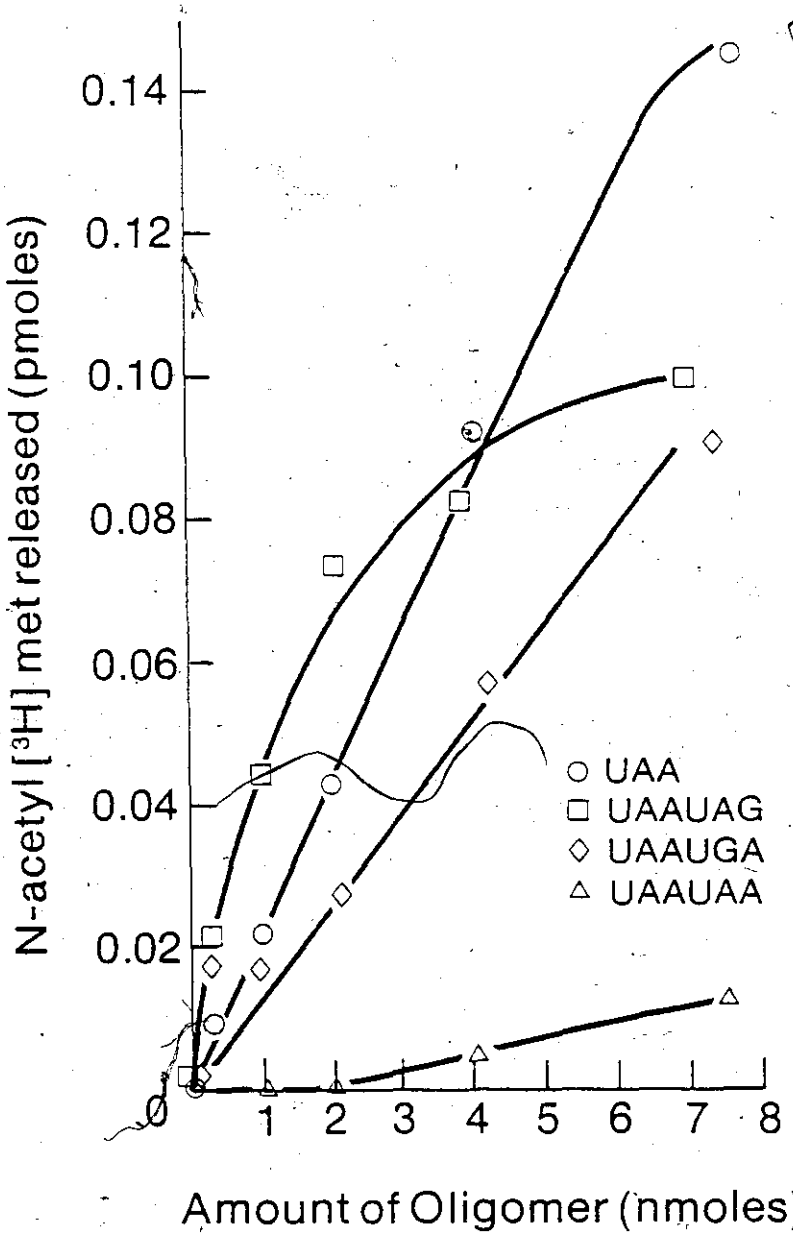


Figure 16. Stimulation of N-acetyl[^{3}H]met release by tandem stop polymers, UAAUAA, UAAUAG and UAAUGA.

Conditions for initiation complex formation were as described in Experimental with 300 pmoles AUG and 2.5 pmoles of N-acetyl[^{3}H]met-tRNA. Termination reaction mixtures contained 14 ug of RF-1 and varying levels of UAA or of UAAUAA, UAAUAG or UAAUGA. Incubation was for 5 min at 24°C. Unspecific release obtained with AUG in the presence of RF-1 and no UAA or UAA-containing polymer, was subtracted.

Table 24. Release Reaction with mRNAs of Formula $AUG(U)_m UAA(A)_n$.

Oligomer	f[^{35}S]met-tRNA Bound dpm x 1000	f[^{35}S]met Released dpm x 1000
$AUGAAA(A)_n$	36.7	2.4^{3*}
$AUGUAA(A)_n$	35.6	2.8^{3}
$AUG(U)_1UAA(A)_n$	34.0	9.8^{4}
$AUG(U)_3UAA(A)_n$	41.7	9.1^{4}
$AUG(U)_4UAA(A)_n$	41.6	9.2^{5}
$AUG(U)_5UAA(A)_n$	43.6	5.2^{5}

Polymers of composition $AUG(A)_n$ or $AUG(U)_mUAA(A)_n$ were synthesized by use of primer-dependent polynucleotide phosphorylase as described (Ganoza, 1978). Initiation complex formation and termination reactions were preformed as described in experimental using 86,300 dpm f[^{35}S]met-tRNA$_f$ and 220 pmoles of each polymer.

*Numbers of experiments, binding experiments were duplicates. Values are averages of samples which were within 10% of each other, and are not corrected for unspecific deacylation.

unpublished result). Oligomers which lacked nonsense codons, i.e. $AUG(A)_n$, did not hydrolyze fmet-tRNA and as previously observed $AUGUAA(A)_n$ did not chain terminate (Ganoza and Thach, 1969). Significant hydrolysis of fmet-tRNA did occur with $AUG(U)_1UAA(A)_n$ and $AUG(U)_4UAA(A)_n$, where the termination codon is out-of-phase from AUG by one base. Oligomer, $AUG(U)_3UAA(A)_n$, containing in-phase UAA also chain terminated, whereas $AUG(U)_5UAA(A)_n$, was far less effective. This hydrolysis in response to spaced and/or out-of-phase nonsense codons was further investigated using shorter oligoribonucleotides containing UAA juxtaposed to AUG, or separated by one or more nucleotides.

The level of ribosomal complex formation with each oligoribonucleotide was measured (Table 25). Oligomers, AUGUA, AUGUUA, AUGCUA, AUGAUG, AUGUUU, and AUGAGC, all bound N-acetylmet-tRNA with equal efficiency, supporting the previous conclusion that the base immediately following AUG does not affect ribosome-oligonucleotide interactions as much as the nucleotide 5' to the initiator (Ganoza, et al., 1978, 1982). Lower levels of AUGCUAA bound to ribosomes indicating that the sequence following AUG may be important.

Somewhat unexpected results were observed when the fraction of N-acetyl-met released was compared. About half of the ribosome-bound N-acetylmet-tRNA was hydrolyzed with AUGUUAA or AUGUUAU. Nearly 40% of the bound substrate was

Table 25. Binding and Release Reactions with Various AUG-Containing Polymers[a]

Polymer Bound	N-Acetyl [^{3}H]met-tRNA Bound x 100 (pmoles)	N-Acetyl [^{3}H]met Released x 100 (pmoles)	%Release
AUG	27.9	0.0	0.0
AUGUAA	14.7	3.5	23.8
AUGUA	19.0	7.3	38.4
AUGUUA	19.3	7.1	37.4
AUGUUAA	14.0	6.7	47.9
AUGUUAU	13.7	6.5	47.4
AUGCUA	17.2	2.1	12.2
AUGCUUA	14.0	0.0	0.0
AUGCUAA	11.8	1.9	16.1
AUGAUG	22.2	0.0	0.0
AUGUUU	19.7	0.2	0.1
AUGGCU	19.2	0.2	0.1
AUGUCU	14.7	0.6	0.4
AUGAGC	20.4	0.2	0.1

[a]Values are the average of triplicates which were within 10% of each other. Over 80% of the bound label reacted with puromycin with AUG, and each penta- or hexaribonucleotide programmed-ribosome. Non-enzymatic hydrolysis (about 0.1 pmoles) fmet-tRNA was subtracted from each.

hydrolyzed when RF-1 was added to ribosomes programmed by AUGUUA, and about 20% hydrolysis was observed with AUGCUAA or AUGUAA. The hexamer AUGUAA was more effective than oligomer AUGUAA$(A)_n$ (Table 25). Oligomer AUGCUA, stimulated ~2% of the hydrolytic reaction while AUGUA stimulated hydrolysis nearly 40%, implying that the UA part of the sequence is partly responsible for the specificity of these in vitro termination reactions. Hexamers, AUGAUG, AUGUUU, AUGGCU, AUGUCU, and AUGAGC, were inactive, indicating that the in vitro assay for termination responds only to oligomers that contain at least part of a termination codon (Table 25). Since AUGUUA, AUGUUAA, AUGUUAU, and AUGUA, stimulated release to approximately the same extent, the common UA portion of the oligomers appears to be equally effective whether it is positioned adjacent to, or one base away from, AUG bound N-acetylmet-tRNA.

By comparison, AUGUAA, caused less response than observed for AUGUA. The result is interesting because AUGUAA, contains an intact nonsense codon whereas AUGUA, does not. Furthermore an additional adenosine residue 3' to AUGUA, decreased the amount of release considerably and extra adenosine residues in oligomer AUGUAA$(A)_n$, depressed the reaction even further (Table 24). In contrast, the oligomers, AUGUUA, AUGUUAA and AUGUUAU, stimulated approximately the same level of release indicating that addition of either adenosine or uridine

residues 3' to UA does not affect the abilities of these oligomers to stimulate the termination reaction. A lower level of release was observed with AUGCUA and AUGCUAA indicating that cytidine 5' to UA does restrict the reaction. The oligomer AUGCUUA, did not detectably cause a termination reaction indicating that when the UA region is two bases from AUG it is not positioned correctly to cause a response.

Since the sequence, UA, when covalently linked to AUG, caused hydrolysis of N-formylmethionine (or N-acetylmethionine), free UA and UUA were compared with UAA (Figure 17). Neither UA nor UUA stimulated the model termination reaction verifying earlier results that only nonsense triplets stimulate hydrolysis (Caskey, et al., 1968).

The time course of the termination reaction programmed by AUGUAA, AUGUUA, and a mixture of the free triplets, AUG and UAA, is shown in Figure 18. As expected, different concentrations of UAA trigger release of N-acetylmethionine from N-acetylmet-tRNA:ribosome:AUG:RF-1 complexes. The expected lag of about twenty minutes in reactions programmed with low levels of free UAA was observed (Caskey, et al., 1968; Capecchi, 1967). This lag was markedly diminished in reactions containing AUGUAA, whereas reactions programmed by AUGUUA, proceeded at the fastest rate without detectable time lapse, indicating that

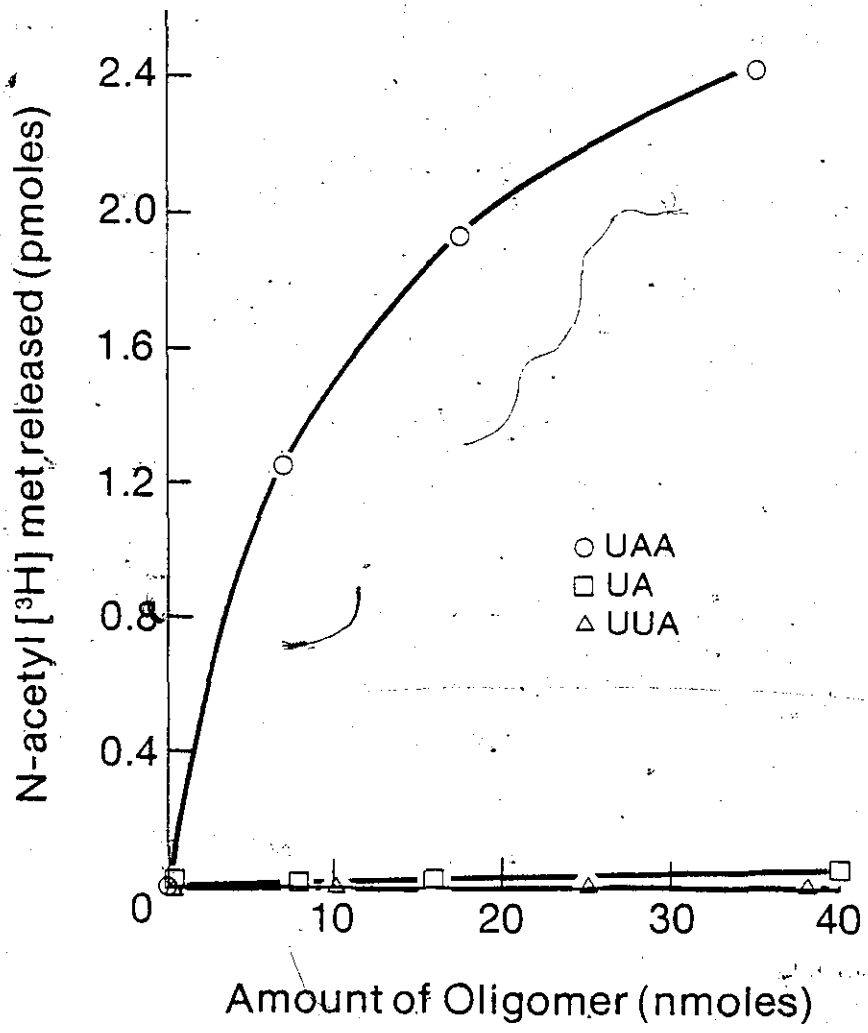


Figure 17. Lack of stimulation of N-acetyl[^{3}H]met release by UA and UUA.

Conditions for initiation complex formation were as described in Experimental with 300 pmoles AUG and 2.5 pmoles of N-acetyl[^{3}H]met-tRNA. Termination reaction mixtures contained excess RF-1 and varying levels of UAA, UA or UUA, and were incubated for 30 min at 24°C. Unspecific release obtained with AUG in the presence of RF-1 was subtracted.

the N-acetylmet-tRNA:ribosome:mRNA:RF-1 complex reacts more efficiently when the nonsense codon occurs within a mRNA (Figure 18). The faster rate of release with AUGUUA, suggests a lack of specificity on the part of RF-1. The possibility of a protein contaminant of RF-1, causing this hydrolysis seems unlikely as this artifact would have to have very similar physical properties to RF-1 and would have to co-purify through at least four purification steps (Ganoza and Tompkins, 1970).

Release factor RF-1 appears to recognize UA which is part of the UAA or UAG sequence, and the sequence around UA is very important for this termination. A serious consequence of this lack of specificity is that RF-1 could potentially elict release of growing peptides from peptidyl-tRNAs in the neighbourhood of any UA sequence. Since this dinucleotide occurs at too high a frequency in mRNA, the possibility that aminoacyl-tRNAs could effectively compete with the release factor during chain propagation was examined. It was found that fmet-leu-tRNA is synthesized from AUGUUA, in the presence of proteins, IF-1, IF-2 (initiation factors), and EF-Tu, EF-Ts (elongation factors), as previously reported for hexamer, AUGUUU (Erbe and Leder, 1968). Addition of RF-1 decreased the level of fmet-leu-tRNA synthesized and this effect was more pronounced when UAA was added to ensure that RF-1 had bound to ribosomes (Table 26; Ganoza et al., 1982).

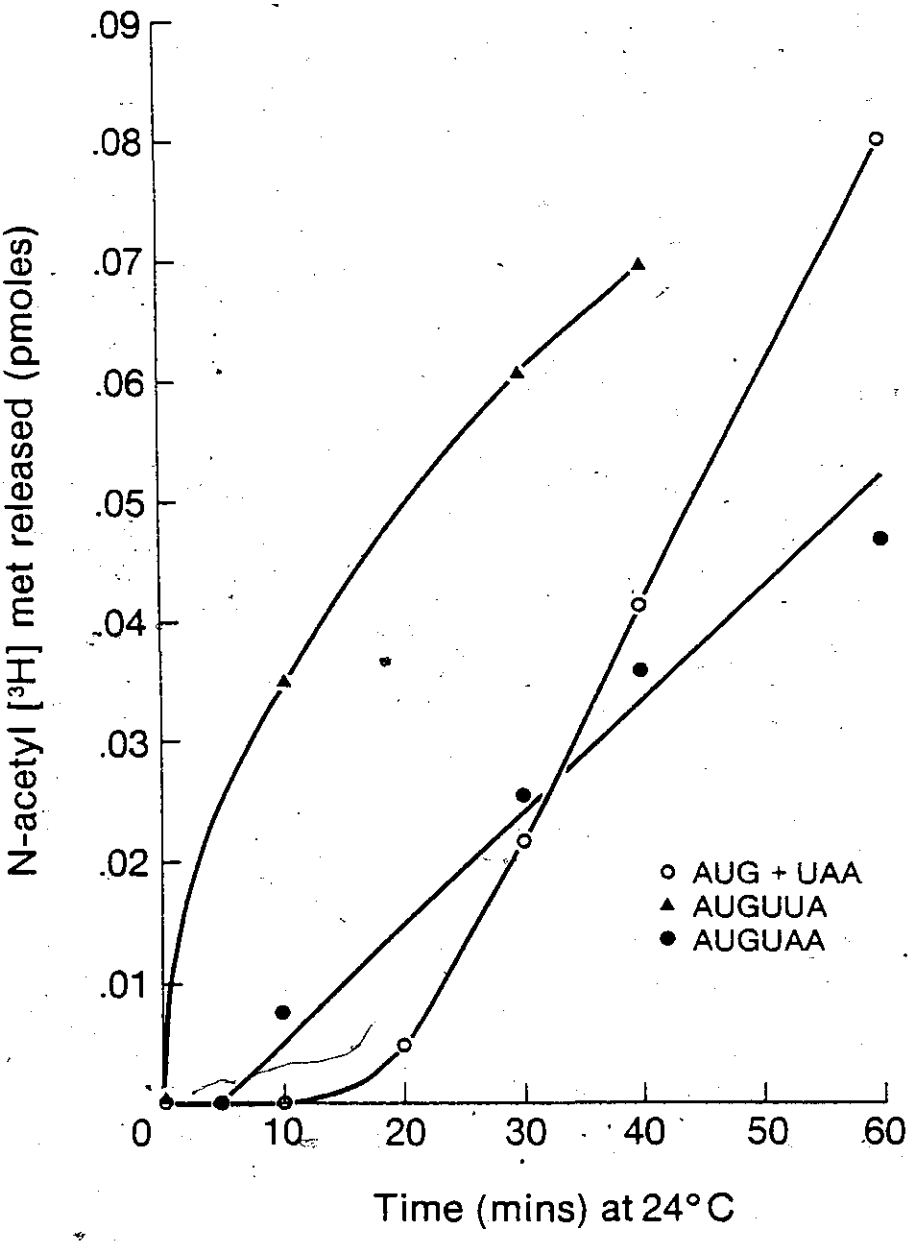


Figure 18. Time course of release of N-acetyl[^{3}H]met with AUGUAA or AUGUUA or with the free UAA codon.

Initiation complex formation was carried out as described in Experimental with 300 pmoles of oligomer and 2.5 pmoles of N-acetyl[^{3}H]met-tRNA. The acetylated tRNA was bound to ribosomes with AUG or AUGUAA or AUGUUA as described. Release was initiated by adding 14 ug RF-1 in each case and 7.6 nmoles UAA where indicated. Unspecific release observed in the presence of RF-1 was subtracted.

Table 26. Effect of Release Factor on the Synthesis and Hydrolysis (Release) of f-met-leu Programmed by AUGUUA.

Additions	Total f[^{35}S]met Recovered (dpm)	f[^{35}S]-met-leu Synth. (dpm)	Peptide Hydrolysis f[^{35}S]met (dpm)	Peptide Hydrolysis f[^{35}S]-met-leu (dpm)
1. IF-1, IF-2 EF-Tu, EF-Ts	56000	10140	7060	800
2. IF-1, IF-2 EF-Tu, EF-Ts+RF-1	51600	5000	7800	1300
3. IF-1, IF-2, EF-Tu EF-Ts+RF-1+UAA	39000	2100	7100	1200
4. RF-1	30600	-	12493	190*
5. None	30600	-	944	190*

Dipeptide synthesis was analyzed electrophoretically. Recovery of total radioactivity was approximately 85% of added counts. Release of f[^{35}S]met with added factors was due to a trace of fmet-tRNA hydrolase in the IF-1 preparations. Experiments 4 and 5 used AUG and f[^{35}S]met-tRNA bound to ribosomes as substrates. Non-specific hydrolysis may account for about 10% of the f[^{35}S]met-leu released. Traces of EF-G in RF-1 could account for the remaining hydrolysis of fmet-leu in experiments 2 and 3.

*Represents background radioactivity.

Since a significant fraction of RF-1 is normally found loosely bound to ribosomes, and since the release factor appears to be in competition with leu-tRNA^{leu} for the sequence UUA, it seems release factor could have a dual function: namely, termination when nonsense codons are in-phase, or cleavage of nascent peptide chains when UA occurs in the proper sequence context and the required aminoacyl-tRNA is not immediately available to recognize the UA-containing triplet, thus acting as a regulator of protein synthesis.

In this thesis it is reported for the first time that the sequence UA, is recognized by RF-1 to trigger chain termination in vitro. This ability of UA to signal release depends upon the nature of nucleotides adjacent, both 3' and 5', to this sequence. The result is interesting, because RF-1 was considered to be specific for the triplet stop codons, UAA and UAG (Caskey, 1979). Indeed, when UA is added as a free dinucleotide to fmet-tRNA:ribosome:AUG complex, termination is not observed with RF-1. Apparently RF-1 exhibits different specificity when potential terminator sequences are covalently linked to AUG within the same polynucleotide, as in mRNA. Under these circumstances, within certain base context, (i) UA functions as a termination signal and (ii) UA containing terminator signals can be read out of the AUG-aligned reading frame.

A model based on the conformation adopted by

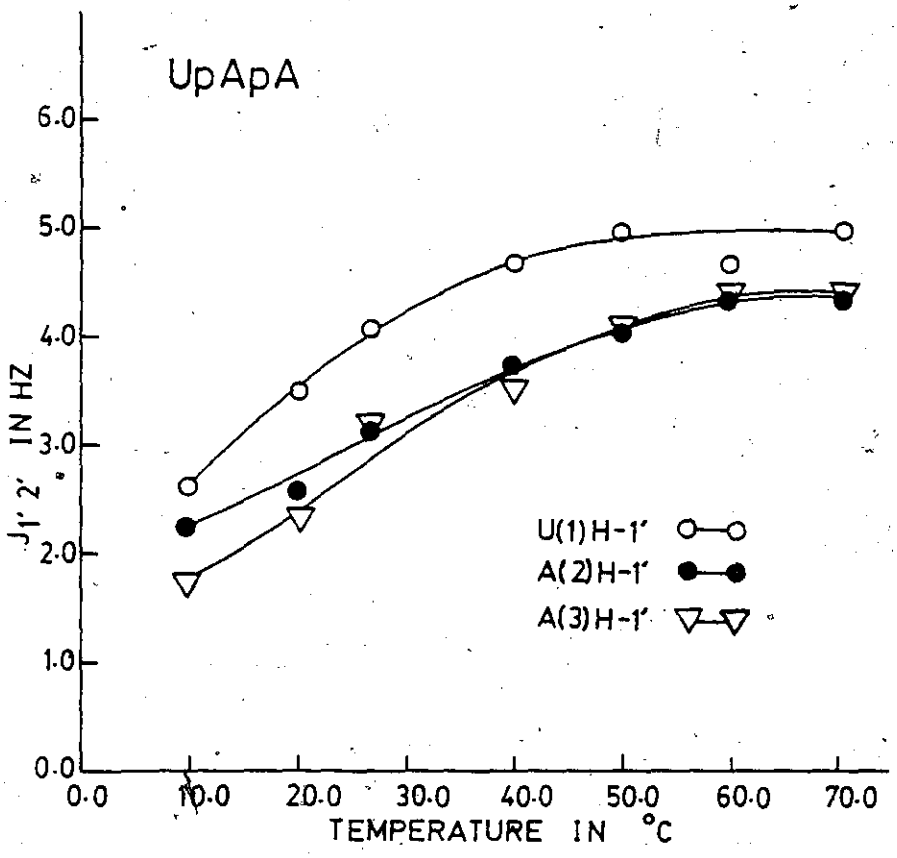


Figure 19. Base-stacking in UAA as monitored by the anomeric coupling constant ($J_{1',2'}$), as a function of temperature.

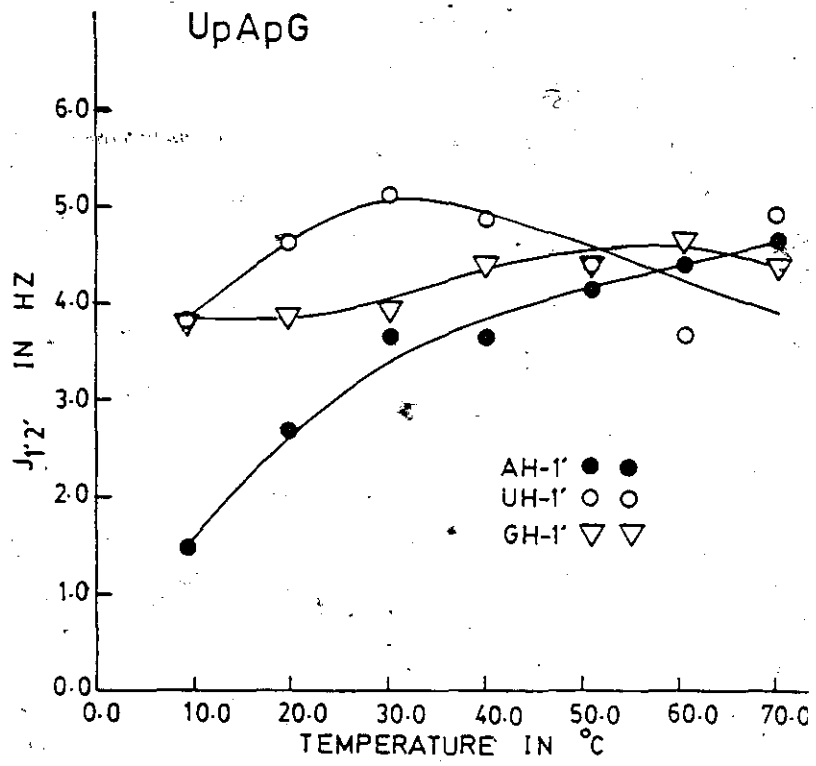


Figure 20. Base-stacking in UAG as monitored by the anomeric coupling constant ($J_{1'2'}$), as a function of temperature.

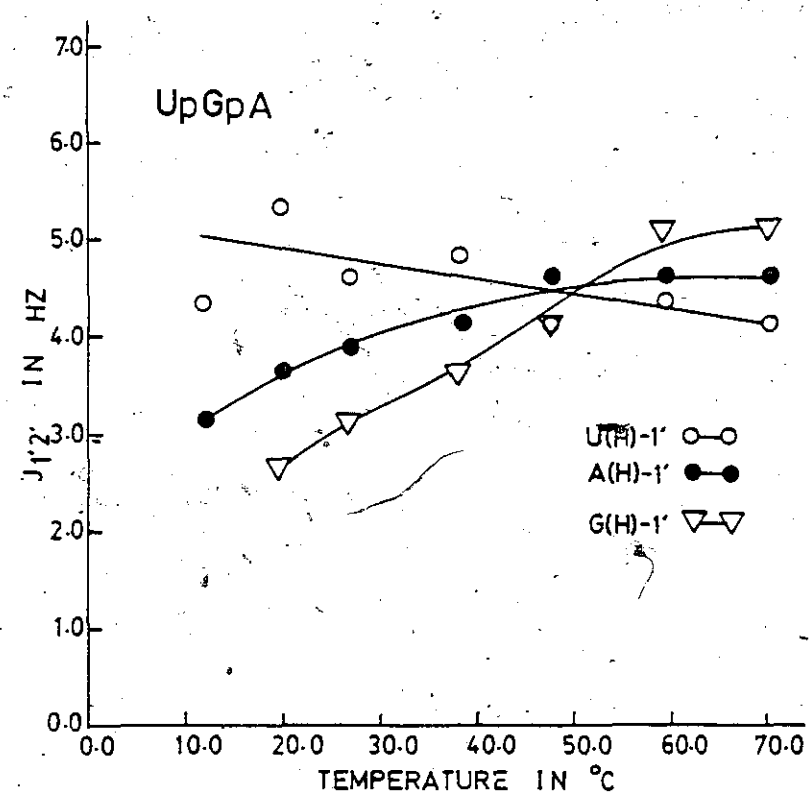


Figure 21. Base-stacking in UGA as monitored by the anomeric coupling constant ($J_{1',2'}$), as a function of temperature.

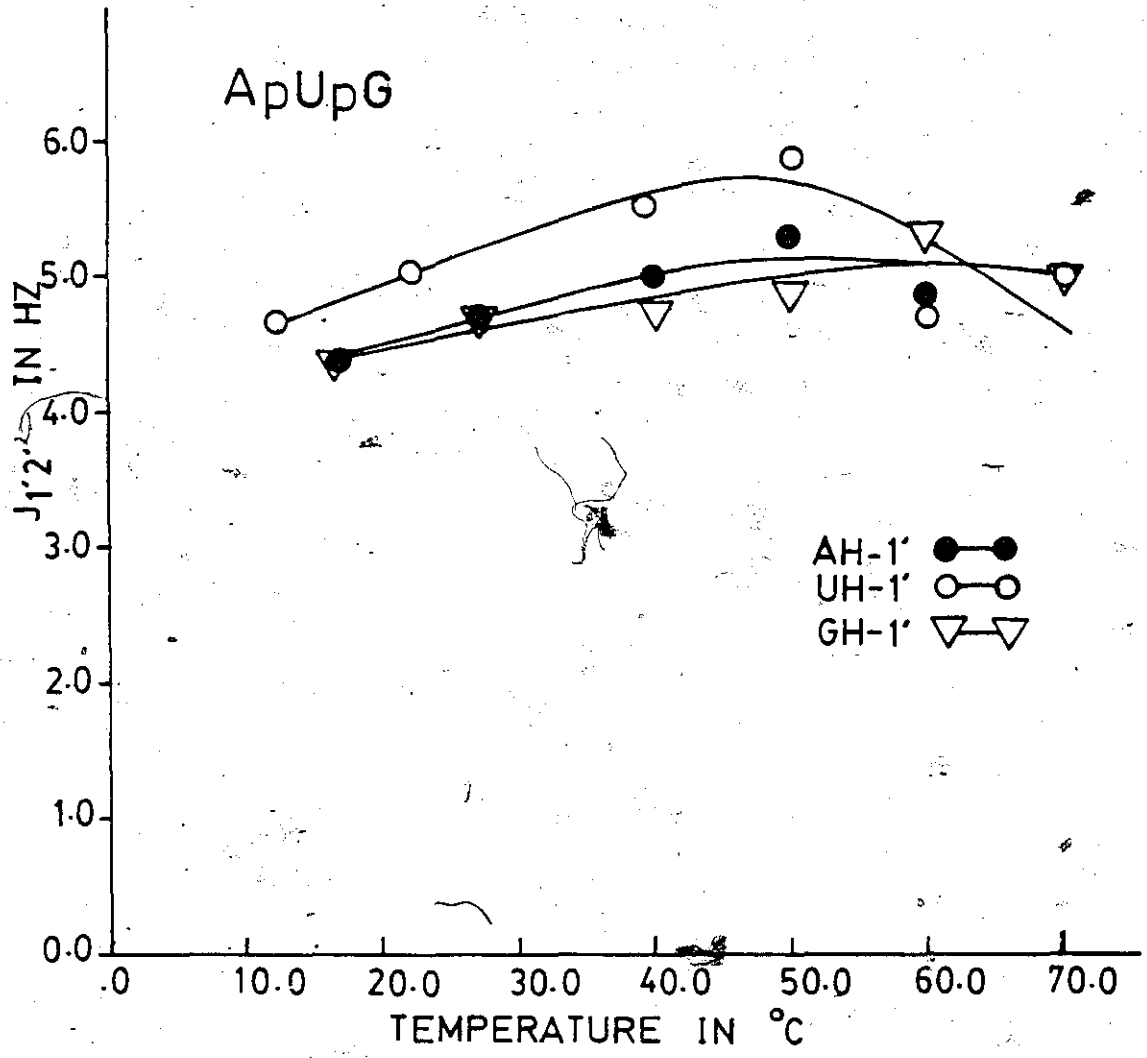


Figure 22. Base-stacking in AUG as monitored by the anomeric coupling constant ($J_{1',2'}$), as a function of temperature.

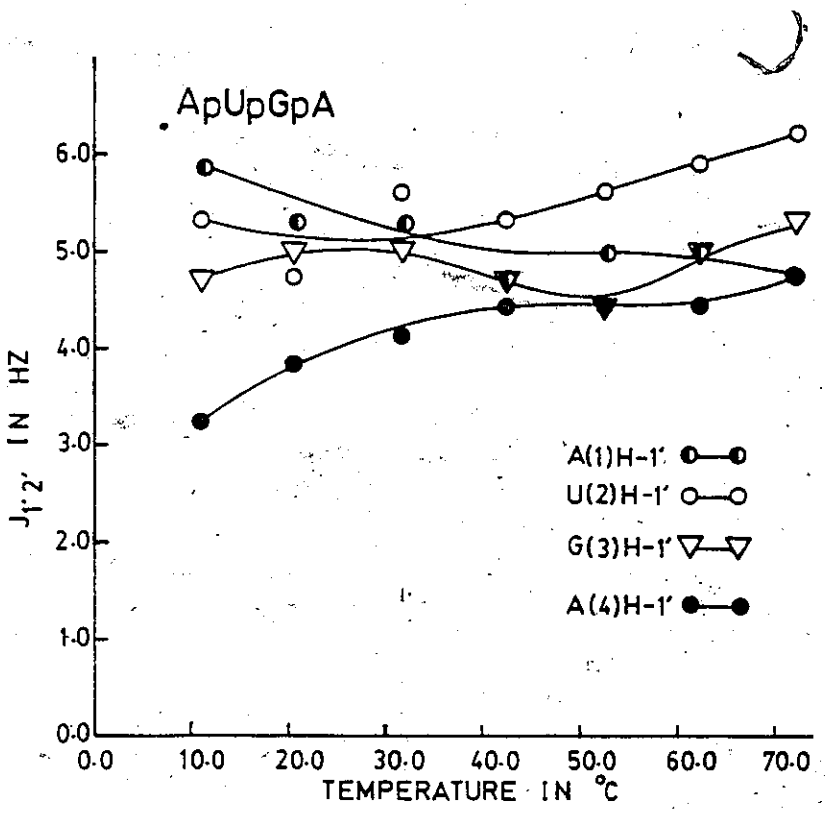


Figure 23. Base-stacking in AUGA as monitored by the anomeric coupling constant ($J_{1'2'}$), as a function of temperature.

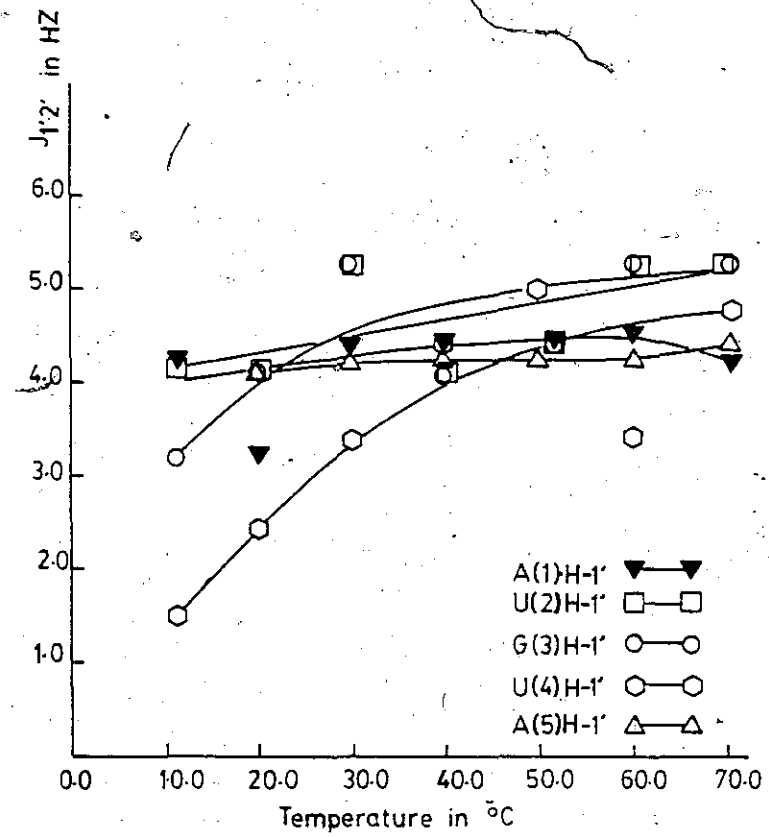


Figure 24. Base-stacking in AUGUA as monitored by the anomeric coupling constant ($J_{1'2'}$), as a function of temperature.

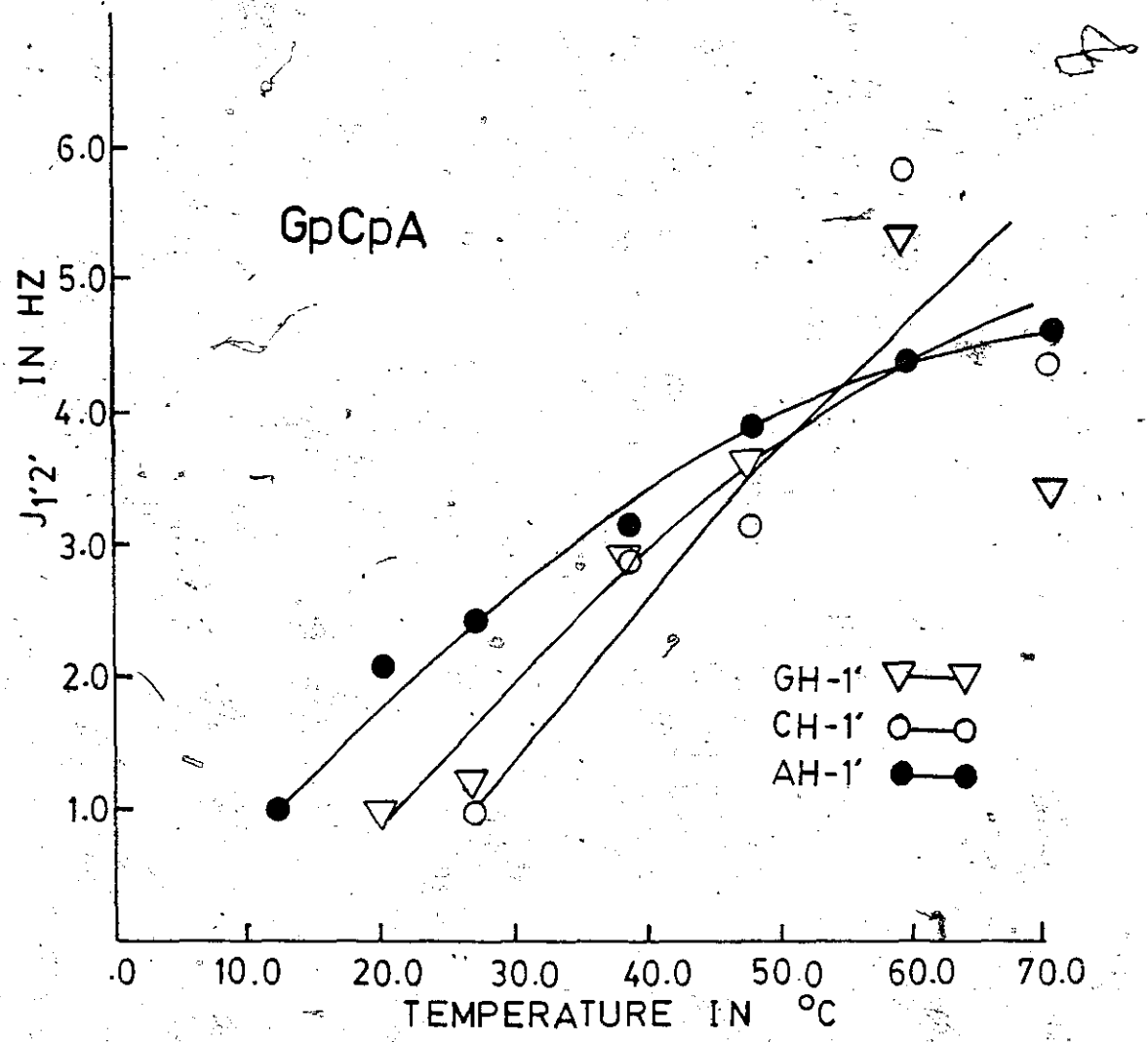


Figure 25. Base-stacking in GCA as monitored by the anomeric coupling constant ($J_{1',2'}$), as a function of temperature.

Table 27. 3' Endo Conformation[a] as a percent for the Biologically Significant Trinucleoside Diphosphates UAA, UAG, UGA, AUG, and Oligomers UAAA, and AUGUA at various Temperatures.

Oligomer Nucleotide	Temperature °C.						
	70.0	60.0	50.0	40.0	30.0	20.0	10.0
UpApA							
Up	50.2	53.1	50.2	53.1	59.0	64.8	73.6
pAp	56.0	56.0	59.0	61.9	67.8	73.6	76.6
pA	56.0	56.0	59.0	64.8	67.8	76.5	82.4
UpApG							
Up	51.2	63.4	56.0	51.2	48.7	53.6	61.6
pAp	53.6	56.0	58.5	63.4	63.4	73.1	85.0
pG	56.4	53.6	56.0	56.1	60.9	61.6	61.6
UpGpA							
Up	58.5	56.1	58.5	51.2	53.6	46.3	56.1
pGp	48.7	48.7	58.5	63.4	68.3	73.1	68.3
pA	53.6	53.6	53.6	58.5	60.9	63.4	68.3
ApUpG							
Ap	39.0	45.0	47.0	47.0	48.0	50.0	55.0
pUp	52.0	59.0	48.0	50.0	56.0	49.0	53.0
pG	50.0	47.0	45.0	45.0	40.0	49.0	53.0
ApUpGpUpA							
Ap	56.1	51.2	53.6	53.6	53.6	66.2	56.1
pUp 2	46.3	46.3	53.6	58.5	45.6	58.5	58.5
pGp	46.3	46.3	53.6	58.5	45.6	58.5	67.3
pUp 4	51.0	62.5	50.1	53.6	63.7	73.1	85.0
pA	53.6	56.1	56.1	56.1	56.1	58.5	>85.0
UpApApA							
Up	68.3	43.9	50.5	78.0	70.0	85.0	>85.0
pAp 2	53.6	63.4	70.7	70.7	71.7	78.0	>85.0
pAp 3	58.5	63.4	70.7	70.7	71.7	78.0	>85.0
pA	58.5	60.9	63.4	82.9	85.0	>85.0	>90.0

[a]Calculated as % 3'-endo = 10 (10 - $J_{1',2'}$)
Temperatures are accurate to within 2°C.

oligoribonucleotides in aqueous solution may explain these results. The conformations adopted by a oligomer depend to a large extent on the degree of base stacking within the oligomer, which in turn is dependent upon the equilibrium between 2'-endo and 3'-endo ribofuranoside. Proton NMR spectroscopy allows one to monitor this sugar conformation equilibrium by measuring the coupling constant ($^3J_{1'2'}$) of the ribose H-1' proton. The $^3J_{1'2'}$ coupling constant has been demonstrated to reflect the amount of base stacking in short oligomers, and hence solution conformation (Altona, 1975). For example on complete base stacking, the equilibrium shifts to the 3'-endo conformer and the anomeric H-1' doublet resonance collapses to a singlet.

The change with respect to temperature of the $^3J_{1'2'}$ coupling constant for the anomeric resonances of the oligomers UAA, UAG, UGA, AUG, AUGA, AUGUA, and GCA, over the temperature range 10-70°C, clearly show that the "rate" of base stacking is not the same for each base in the oligoribonucleotides (Figures 19-25). Also the oligomers exhibit very different overall stacking as a function of temperature. The trimer GCA which exhibits base pairing (Alkema, et al., 1981a,b), is in a completely stacked conformation at 20°C, whereas trimer AUG is largely in an unstacked conformation. Calculations of the percentage 3'-endo conformer, illustrate that the termination codons

UAA, UAG, and UGA exhibit a higher degree of base stacking than the initiator codon AUG (Table 27).

Since local conformations adopted by short sequences within a native RNA are part of the recognition signal within protein-RNA interactions it is plausible that conformation can affect the termination event. Assuming that the conformation adopted by the trimer approximates the conformation of the same trimer region in a longer oligomer, a conformational hypothesis can be presented to explain the in vitro termination results. For example the amount of stacking in AUG is not altered significantly upon addition of a pA residue to the 3'-end of the molecule forming AUGA. However, the oligomer AUGA can also be considered as the extension of a UGA termination codon. In this case the effect on stacking is significant. Interestingly the triplet UGA is ambiguous, it codes for both tryptophan and translational termination. Furthermore release factor RF-2, can compete with tryptophan for the UGA site (Ganoza, personal communication). If the protein-RNA interaction favours a low degree of stacking at the RNA site, then it is possible that changes in base stacking are required to terminate protein synthesis at UGA codons.

The most interesting aspect of the result was that the in vitro termination reaction responded to the presence of UA in the immediate vicinity of an AUG, initiation codon, thus redefining the specificity of release protein, RF-1.

That RF-1 recognizes UAA and UAG suggests that the interaction at the third base in the termination codon is reminiscent of the wobble type of pairing exhibited in tRNA-mRNA interactions. A truncated termination codon (UAA, UAG → UA) held in a favorable position on the ribosomes stimulates the termination event, indicating that the third base interaction is not essential for the ribosome:tRNA:mRNA:release factor complex. This idea is not entirely unfounded, Smrt et al., (1970), measured fmet release from ribosome:fmet-tRNA:AUG:release factor complexes induced by a series of UAA and UAG derivatives containing substituted uridine residues. The response pattern suggested that the N_3 proton and C_4 carbonyl moieties of the pyrimidine ring are essential for terminator codon recognition. Substituents in the C_5 position do not impair function and strikingly N_3, and C_4 are the positions required for hydrogen bonding in Watson-Crick base pairs.

The observations that AUGUA, AUGUUA, AUGUUAA, and AUGUUAU, stimulate the release reaction to about the same extent, can now be explained. Each oligomer is bound to the ribosome:tRNA complex through its AUG portion. If the 3' unbound remainder of each oligomer is assumed to adopt the same conformation that it would in solution, AUGUUA, AUGUUAA, and AUGUUAU, would exhibit conformations where the

guanosine (G-3), and adenosine (A-6) residues stack forcing the intervening uridine residue (U-4), to loop-out. In this way the UA region in sequences AUGUA, AUGUUA, AUGUUAA, and AUGUUAU, would fall in approximately the same region on the ribosome (Figure 26). Since hexamer AUGUAA, only effects approximately one-half of the termination response obtained with AUGUUAA, the UAA or UA region must be masked. Base stacking in AUGUAA between the guanosine (G-3) and adenylyladenosine residues could force the uridine (U-4), residue to loop-out (Cheng, et al., 1978; Lee and Tinoco, 1980), or alternatively the adjacent adenylyladenosine residue increases the amount of stacking throughout the molecule, causing in either case the oligomer to be less extended, thus displacing the important UA portion from its optimum position on the ribosome. If this is correct, AUGUA, should exhibit a more extended conformation than AUGUAA. The highly flexible conformation for AUGUA, is supported by variable temperature NMR analysis which indicates that this pentamer exhibits a large anomeric coupling constant even at 0 °C (Figure 24).

Oligomers, AUGUUA and AUGUUAA, promote termination to an approximately equivalent degree. The adenosine residue 3' to UA only slightly enhances the termination reaction, evidence that the third base interaction is not required by the ribosome:fmet-tRNA:mRNA:RF-1 complex.

That a truncated termination codon can effect the

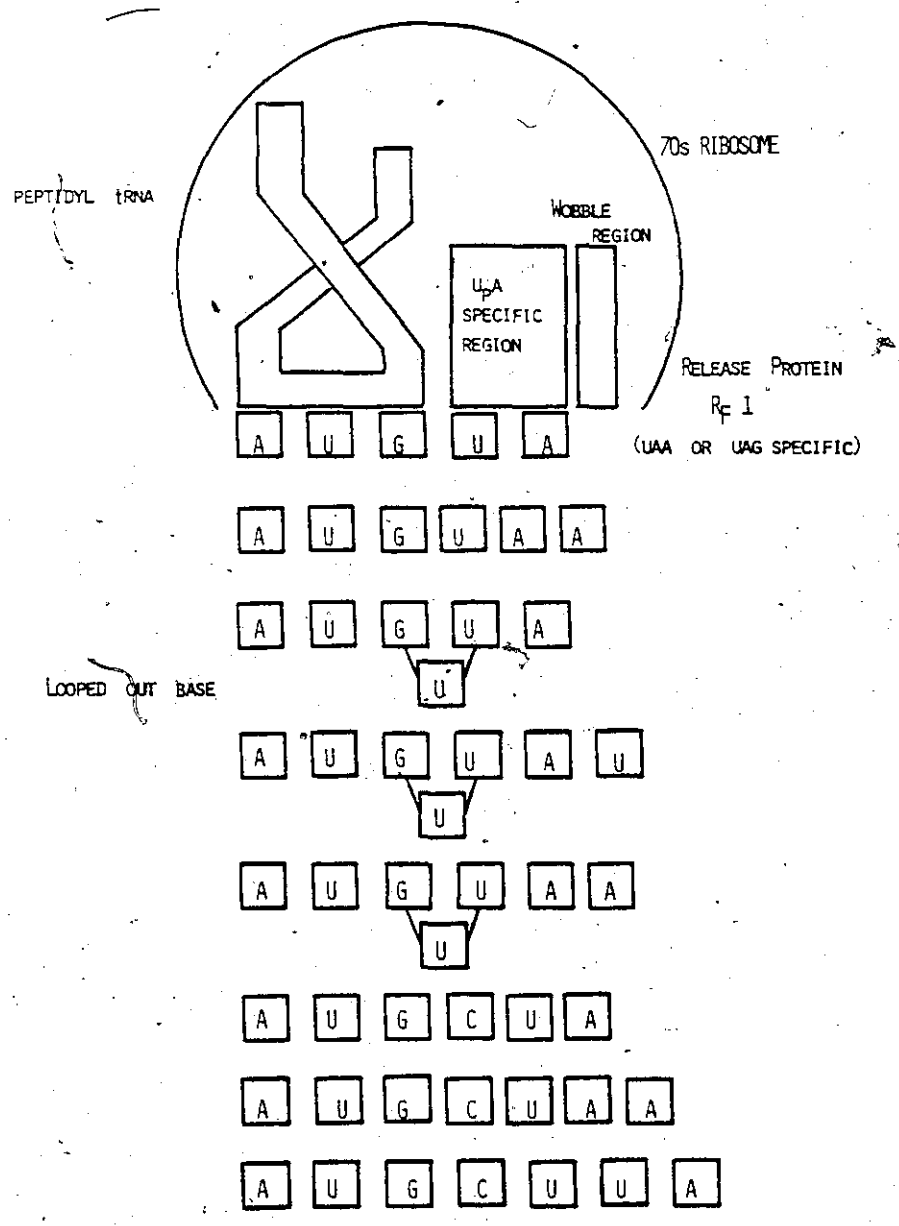


Figure 26. Schematic representation of how base stacking within various oligomers affects the position of the UA "termination signal" in the in vitro termination complex.

release reaction fails to explain why the termination event is poorly directed by the model messengers, AUGCUA and AUGCUAA. These oligomers are expected to exhibit more highly stacked conformations in comparison to the other oligomers studied because of the greater ability of a cytidine residue relative to a uridine residue to base stack (Alkema, et al., 1982; Everett, et al., 1980; Giessner-Prettre, et al., 1975). The strength of the stacking interaction within these oligomers does not allow the cytidine residue to loop-out, thereby positioning the sequence UA incorrectly and so preventing termination. Similarly in AUGCUUA, the UA is positioned two bases away from AUG, in an unfavourable "configuration" for the release factor to effect a termination response.

An explanation for the results obtained with the tandem stop codons is possible considering conformation. In triplet UAA, base stacking between the adenosine residues weakly extends to the uridine residue (Table 27). Apparently this conformation for the UA region is best recognized by RF-1. In hexamer, UAAUAA, base stacking through four adenosine residues extends to uridine (1), and loops out uridine (4). The more compressed UAA, as in AUGUAA, or AUGUAA$(A)_n$, is less suitable for RF-1 recognition. In UAAUAG, the guanosine residue does not stack as well as an adenosine, consequently the UAG, region approximates that of UA and the hexamer operates as two UA

units with the resultant enhanced release (compared to UAA) by RF-1. In the case of UAAUGA, only one UAA, or UA, region is available for recognition by RF-1, while the relatively weaker stacking of a guanosine residue inhibits continuous stacking throughout the hexamer. Somewhat lower release is expected by a UAA-containing hexamer compared to UAA, on steric and entropic grounds.

Base stacking interactions clearly explain the lack of release with AUGUAA$(A)_n$, and the stimulation of termination by intervening uridine residues, which can disrupt the strong stacking by the poly A tail. The results predicted by the model (Figure 26), are not only in good agreement with those obtained in the present *in vitro* system but are consistent with recent *in vivo* studies where it was suggested that the presence of an adenosine residue next to a termination codon enhanced suppression by tRNAs because of enhanced base stacking (Engelberg-Kulka, 1981). The suggestion that sequence effects are mediated by a conformational change may also explain the observation that the efficiency of the primary events in tRNA-mediated suppression can be changed at least 50-fold by the reading context (Salser, *et al*., 1969).

Examination of the products of synthesis and release suggests that one or more of the proteins required for dipeptide synthesis may actively prevent release by the out-of-phase UA "codon". Note that RF-1 inhibits the

AUGUUA, programmed synthesis of fmet-leu but does not proportionally increase the level of fmet released (Table 26). Elongation factor, EF-T, appears to decrease UA-dependent hydrolysis, however further study is required to determine if EF-Tu, EF-Ts or the associated subunits are involved or whether other proteins, such as "rescue" (Van der Meer and Ganoza, 1975), play a role in suppressing the UA mediated hydrolysis (Ganoza and Buckingham, unpublished results).

It is possible that UA stimulates release of fmet from ribosome:fmet-tRNA:mRNA:RF-1 complexes but not a polypeptide from ribosome:aa-tRNA:mRNA:RF-1 complexes. This would suggest that the fmet-tRNA somehow alters the ribosome release factor interaction such that release factor specificity is altered. This situation may be biologically relevant nonetheless, because the intiation and termination events seem to be interlinked (Ganoza, 1977; Atkins, 1979; Kastelein, et al., 1982). However, it has been reported that essentially identical requirements are needed for peptidyl-tRNA and fmet-tRNA in the release reaction (Menninger, 1971).

Regardless of the explanation the observed lack of specificity is disturbing, in that, RF-1 could potentially cleave peptidyl-tRNA during chain growth if the supply of charged aminoacyl-tRNAs were limiting, or if the translation rate decreased for any other reason. The selective

utilization within mRNAs of UA "termination signals", which depend on base context, may allow protein synthesis to have many controllable stop signals. Nature could, through release factors, achieve a measure of fine tuning of gene expression at the translational level.

CONCLUSION

Much of the understanding of the conformational dynamics of DNA and RNA in solution is based on studies which use ^{1}H NMR spectroscopy as the technique to determine the factors which govern conformation. This physical technique is advantageous because it allows one to monitor regions of the DNA or RNA strand at the molecular level. The information obtained from an NMR experiment is dramatically diminished however if resonances cannot be assigned. Currently assigning the chemical shifts of anomeric and aromatic protons is laborious. The incremental analysis method (Borer, et al., 1975) is presently the only convenient assignment scheme but it requires the investigator to have available a series of related oligomers of increasing length, a situation that most often is not practical. Furthermore reference to oligomers outside a sequential series is often necessary for unambiguous assignment (Everett, et al., 1980). Considering the important role that ^{1}H NMR has played (Davies, 1978; Kearns, 1977; Reid and Hurd, 1977), and should continue to play (Jones, 1980; Schmidt and Edelheit, 1981) in understanding biochemical structure-function relationships of these molecules, an accurate and convenient assignment scheme would be of great value. The set of

empirical chemical shift parameters determined in this work allows one to assign accurately and quickly the resonances of RNA and DNA oligomers containing up to twenty nucleotide units. Accurate assignments using these parameters have been made for the following RNA oligomers; AGCU, CAUG, AUGA, AUGUA, AAGCUU, CAUAUG, AUGCUA, AUGUCU, AGUCU, AGCCU, AGGCU, and AGACU. Assignments have also been determined for the following DNA oligomers; dACATGT, dCGCGAATTCGCG, dAATTGCAATT, dCGCG, dGGCC, dCCGG, dCGCGCG, dCCAA, dTTGG, dCCAAG, dCCA, dCTTGG, dTGG, dATGT, and dTATATA. The accuracy of the assignments based on the shift parameters is very good. By using the calculated chemical shifts in a reverse sense the information from an ^1H NMR spectra can be used to determine the sequence of an RNA or DNA oligomer. Data are presented to demonstrate that this technique is accurate for RNA oligomers containing up to six nucleotide units. The use of NMR to sequence RNA oligomers complements the present enzymatic and chemical techniques because it is independent methodologically. The method should be useful for confirming the sequence of RNA or DNA oligomers to be used in physical studies where quantities are not a limitation.

The studies of sequence effects on translational termination have indicated that base context around termination codons affects their suppression, presumably by

altering the competition between release factors and tRNAs for the termination site (Salser, et al., 1969; Engelberg-Kulka, 1981). The results presented in this work provide further evidence that release factors and tRNAs compete for codons on a mRNA.

Bases immediately following a termination codon UAA did not significantly alter the in vitro termination reaction. Tandem termination codons UAAUAG and UAAUGA were slightly more and slighty less effective than UAA in the termination reaction, while oligomer UAAUAA was almost ineffective in the termination assay.

The most compelling aspect of this work is that release factor specificity in vitro is altered when the termination codon eliciting the response is covalently linked to AUG forming a small messenger RNA. Furthermore oligomers AUGUA, AUGUUA, and AUGUUAU, cause a termination event but do not contain an intact termination codon. They contain only a common UA sequence, a truncated termination codon, thus the UA region must act as a termination signal for RF-1. This reaction is quite specific because model messengers that do not contain a UA sequence do not cause a termination event. Model messengers AUGCUA, AUGCUAA, and AUGUAA were less effective in stimulating the termination reaction indicating that base context around the UA termination signal influences its recognition. Model messengers UAAUAA, AUGUAA, and AUGUAA$(A)_n$ each contain

significant adenosine residues and each stimulates the termination reaction poorly. Thus considering the base context of the individual messengers and their activities in the termination reaction I conclude that the UA termination signal must be held in a favourable location on the ribosome and must not be involved in extensive base stacking for optimum recognition by release factor. These same factors must also influence terminator codon suppression by tRNAs and termination by release factors during the synthesis of proteins in intact organisms.

APPENDICES

1. The aromatic and anomeric protonchemical shifts of the RNA dinucleoside monophosphates and trinucleoside diphosphates.

2. Listing of the Fortran program ASSIGN, used to calculate the chemical shifts of short RNA and DNA strands from their sequence.

3. Listing of the Fortran program SEQUEN, used to calculate the fitting values for sequence determination.

Table 1. Chemical Shifts[a] of the Diribonucleotide UU and Triribonucleotides UUX in buffered D_2O[b].

Proton	U	UU	UUU	UUC	UUG	UAA
U(1)H-6	7.798	7.814	7.811	7.811	7.778	7.757
U(2)H-6		7.824	7.825	7.837	7.764	7.740
U(3)H-6			7.833			
C(3)H-6				7.860		
G(3)H-8					7.998	
A(3)H-8						8.385
A(3)H-2						8.245
U(1)H-1'	5.887	5.882	5.871	5.861	5.844	5.827
U(2)H-1'		5.924	5.894	5.929	5.876	5.846
U(3)H-1'			5.925			
C(3)H-1'				5.941		
G(3)H-1'					5.892	
A(3)H-1'						6.094
U(1)H-5	5.907	5.895	5.909	5.987	5.876	5.861
U(2)H-5		5.895	5.883	5.868	5.863	5.831
U(3)H-5			5.909			
C(3)H-5				6.086		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 2. Chemical shifts[a] of the Diribonucleotide UC and the Triribonucleotides UCX in D_2O[b].

Proton	U	UC	UCU	UCC	UCG	UCA
U(1)H-6	7.798	7.806	7.805	7.806	7.772	7.710
C(2)H-6		7.832	7.835	7.832	7.752	7.701
U(3)H-6			7.814			
C(3)H-6				7.832		
G(3)H-8					7.995	
A(3)H-8						8.385
A(3)H-2						8.251
U(1)H-1'	5.887	5.864	5.846	5.838	5.826	5.862
C(2)H-1'		5.939	5.945	5.927	5.911	5.891
U(3)H-1'			5.908			
C(3)H-1'				5.927		
G(3)H-1'					5.884	
A(3)H-1'						6.086
U(1)H-5	5.907	5.882	5.889	5.858	5.850	5.817
C(2)H-5		6.058	6.043	6.018	5.998	5.971
U(3)H-5			5.873			
C(3)H-5				6.045		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 3. Chemical shifts[a] of the Diribonucleotide UG and the Triribonucleotides UGX in D_2O[b].

Proton	U	UG	UGU	UGC	UGG	UGA
U(1)H-6	7.798	7.723	7.724	7.731	7.701	7.674
G(2)H-8		8.004	8.009	8.015	7.949	7.917
U(3)H-6			7.806			
C(3)H-6				7.809		
G(3)H-8					7.973	
A(3)H-8						8.329
A(3)H-2						8.194
U(1)H-1'	5.887	5.816	5.809	5.812	5.818	5.761
G(2)H-1'		5.907	5.908	5.896	5.844	5.765
U(3)H-1'			5.908			
C(3)H-1'				5.932		
G(3)H-1'					5.875	
A(3)H-1'						6.069
U(1)H-5	5.907	5.828	5.830	5.832	5.820	5.812
U(3)H-5			5.830			
C(3)H-5				5.958		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 4. Chemical shifts[a] of the Diribonucleotide UA and the Triribonucleotides UAX in D_2O[b].

Proton	U	UA	UAU	UAC	UAG	UAA
U(1)H-6	7.798	7.692	7.701	7.580	7.682	7.662
A(2)H-8		8.381	8.391	8.325	8.312	8.287
A(2)H-2		8.264	8.242	8.195	8.193	8.128
U(3)H-6			7.763			
C(3)H-6				7.766		
G(3)H-8					7.949	
A(3)H-8						8.310
A(3)H-2						8.176
U(1)H-1'	5.887	5.765	5.771	5.788	5.756	5.739
A(2)H-1'		6.100	6.102	6.057	6.019	5.964
U(3)H-1'			5.840			
C(3)H-1'				5.873		
G(3)H-1'					5.847	
A(3)H-1'						6.042
U(1)H-5	5.907	5.834	5.791	5.640	5.795	5.777
U(3)H-5			5.791			
C(3)H-5				5.918		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 5. Chemical shifts[a] of the Diribonucleotide - CU and the Triribonucleotides CUX in D_2O[b].

Proton	C	CU	CUU	CUC	CUG	CUA
C(1)H-6	7.781	7.811	7.784		7.766	7.752
U(2)H-6		7.826	7.836		7.766	7.752
U(3)H-6			7.822			
C(3)H-6						
G(3)H-8					7.977	
A(3)H-8						8.390
A(3)H-2						8.244
C(1)H-1'	5.894	5.861	5.916		5.865	5.837
U(2)H-1'		5.901	5.916		5.865	5.826
U(3)H-1'			5.916			
C(3)H-1'						
G(3)H-1'					5.888	
A(3)H-1'						6.072
C(1)H-5	6.073	6.033	6.010		5.995	5.963
U(2)H-5		5.873	5.860		5.831	5.820
U(3)H-5			5.900			
C(3)H-5						

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 6. Chemical shifts[a] of the Diribonucleotide CC and the Triribonucleotides CCX in D_2O[b].

Proton	C	CC	CCU	CCC	CCG	CCA
C(1)H-6	7.781	7.786	7.759		7.750	7.737
C(2)H-6		7.832	7.751		7.750	7.737
U(3)H-6			7.803			
C(3)H-6						
G(3)H-8					7.984	
A(3)H-8						8.404
A(3)H-2						8.251
C(1)H-1'	5.894	5.852	5.895		5.884	5.861
C(2)H-1'		5.900	5.921		5.819	5.811
U(3)H-1'			5.949			
C(3)H-1'						
G(3)H-1'					5.884	
A(3)H-1'						6.099
C(1)H-5	6.073	6.020	6.004		5.976	5.939
C(2)H-5		6.020	6.011		5.954	5.905
U(3)H-5			5.878			
C(3)H-5						

[a] Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b] 1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 7. Chemical shifts[a] of the Diribonucleotide CG and the Triribonucleotides CGX in D_2O[b].

Proton	C	CG	CGU	CGC	CGG	CGA
C(1)H-6	7.781	7.691	7.682	7.693	7.662	7.656
G(2)H-8		8.004	7.996	7.992	7.968	7.934
U(3)H-6			7.787			
C(3)H-6				7.788		
G(3)H-8					7.942	
A(3)H-8						8.343
A(3)H-2						8.214
C(1)H-1'	5.894	5.823	5.804	5.805	5.803	5.795
G(2)H-1'		5.894	5.885	5.865	5.811	5.755
U(3)H-1'			5.895			
C(3)H-1'				5.911		
G(3)H-1'					5.873	
A(3)H-1'						6.080
C(1)H-5	6.073	5.975	5.947	5.931	5.946	5.920
U(3)H-5			5.811			
C(3)H-5				5.931		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 8. Chemical shifts[a] of the Diribonucleotide CA and the Triribonucleotides CAX in D_2O[b].

Proton	C	CA	CAU	CAC	CAG	CAA
C(1)H-6	7.781	7.660	7.687	7.679	7.640	7.620
A(2)H-8		8.377	8.382	8.369	8.303	8.277
A(2)H-2		8.260	8.250	8.222	8.189	8.218
U(3)H-6			7.744			
C(3)H-6				7.741		
G(3)H-8					7.926	
A(3)H-8						8.294
A(3)H-2						8.186
C(1)H-1'	5.894	5.779	5.786	5.769	5.752	5.743
A(2)H-1'		6.093	6.088	6.072	6.003	5.953
U(3)H-1'			5.858			
C(3)H-1'				5.850		
G(3)H-1'					5.832	
A(3)H-1'						6.033
C(1)H-5	6.073	5.958	5.925	5.894	5.897	5.865
U(3)H-5			5.745			
C(3)H-5				5.870		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 9. Chemical shifts[a] of the Diribonucleotide GU and the Triribonucleotides GUX in D_2O[b].

Proton	G	GU	GUU	GUC	GUG	GUA
G(1)H-8	7.961	7.963	7.955	7.957	7.929	7.908
U(2)H-6		7.800	7.801	7.798	7.742	7.732
U(3)H-6			7.786			
C(3)H-6				7.807		
G(3)H-8					7.973	
A(3)H-8						8.339
A(3)H-2						8.191
G(1)H-1'	5.888	5.896	5.885	5.875	5.866	5.850
U(2)H-1'		5.896	5.922	5.906	5.846	5.813
U(3)H-1'			5.878			
C(3)H-1'				5.915		
G(3)H-1'					5.869	
A(3)H-1'						6.061
U(2)H-5		5.817	5.825	5.816	5.803	5.806
U(3)H-5			5.866			
C(3)H-5				6.034		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Chemical shifts of Oligomers GUX were recorded at 250.1335 MHz.

Table 10. Chemical shifts[a] of the Diribonucleotide GC and the Triribonucleotides GCX in D_2O[b].

Proton	G	GC	GCU	GCC	GCG	GCA
G(1)H-8	7.961	7.966	7.967	7.964	7.935	7.920
C(2)H-6		7.780	7.789	7.766	7.712	7.709
U(3)H-6			7.789			
C(3)H-6				7.804		
G(3)H-8					7.965	
A(3)H-8						8.348
A(3)H-2						8.191
G(1)H-1'	5.888	5.874	5.913	5.860	5.834	5.812
C(2)H-1'		5.913	5.913	5.898	5.894	5.883
U(3)H-1'			5.861			
C(3)H-1'				5.898		
G(3)H-1'					5.855	
A(3)H-1'						6.061
C(2)H-5		5.910	5.885	5.857	5.861	5.837
U(3)H-5			5.841			
C(3)H-5				5.998		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to $\pm$0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Chemical shifts of oligomers GCC, GCG, and GCA, were recorded at 250.1335 MHz.

Table 11. Chemical shifts[a] of the Diribonucleotide GG and the Triribonucleotides GGX in D_2O[b].

Proton	G	GG	GGU	GGC	GGG	GGA
G(1)H-8	7.961	7.896	7.895	7.903	7.874	7.860
G(2)H-8		7.971	7.973	7.977	7.921	7.906
U(3)H-6			7.770			
C(3)H-6				7.774		
G(3)H-8					7.952	
A(3)H-8						8.309
A(3)H-2						8.176
G(1)H-1'	5.888	5.798	5.797	5.820	5.767	5.761
G(2)H-1'		5.877	5.884	5.866	5.809	5.749
U(3)H-1'			5.881			
C(3)H-1'				5.895		
G(3)H-1'					5.853	
A(3)H-1'						6.051
U(3)H-5			5.799			
C(3)H-5				5.914		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Chemical shifts of oligomers GGU, GGG, and GGA were recorded at 250.1335 MHz.

Table.12. Chemical shifts[a] of the Diribonucleotide GA and the Triribonucleotides GAX in D_2O[b].

Proton	G	GA	GAU	GAC	GAG	GAA
G(1)H-8	7.961	7.874	7.888	7.883	7.867	7.849
A(2)H-8		8.329	8.333	8.320	8.278	8.252
A(2)H-2		8.213	8.208	8.185	8.166	8.099
U(3)H-6			7.737			
C(3)H-6				7.718		
G(3)H-8					7.925	
A(3)H-8						8.275
A(3)H-2						8.148
G(1)H-1'	5.888	5.733	5.745	5.745	5.719	5.699
A(2)H-1'		6.075	6.080	6.062	6.007	5.959
U(3)H-1'			5.843			
C(3)H-1'				5.855		
G(3)H-1'					5.827	
A(3)H-1'						6.019
U(3)H-5			5.742			
C(3)H-5				5.849		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Chemical shifts of oligomers GAC, and GAA were recorded at 250.1335 MHz.

Table 13. Chemical shifts[a] of the Diribonucleotide AU and the Triribonucleotides AUX in D_2O[b].

Proton	A	AU	AUU	AUC	AUG	AUA
A(1)H-8	8.313	8.313	8.306	8.304	8.271	8.245
A(1)H-2	8.260	8.234	8.224	8.219	8.198	8.159
U(2)H-6		7.759	7.769	7.762	7.729	7.723
U(3)H-6			7.769			
C(3)H-6				7.798		
G(3)H-8					7.963	
A(3)H-8						8.321
A(3)H-2						8.159
A(1)H-1'	6.065	6.062	6.065	6.049	6.023	5.993
U(2)H-1'		5.834	5.866	5.845	5.838	5.825
U(3)H-1'			5.866			
C(3)H-1'				5.896		
G(3)H-1'					5.847	
A(3)H-1'						6.041
U(2)H-5		5.766	5.783	5.764	5.771	5.777
U(3)H-5			5.846			
C(3)H-5				6.012		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Table 14. Chemical shifts[a] of the Diribonucleotide AC and the Triribonucleotides ACX in D_2O[b].

Proton	A	AC	ACU	ACC	ACG	ACA
A(1)H-8	8.313	8.316	8.304	8.304	8.264	8.239
A(1)H-2	8.260	8.210	8.197	8.187	8.170	8.138
C(2)H-6		7.737	7.723	7.700	7.688	7.676
U(3)H-6			7.758			
C(3)H-6				7.776		
G(3)H-8					7.952	
A(3)H-8						8.324
A(3)H-2						8.169
A(1)H-1'	6.065	6.054	6.039	6.036	6.013	5.985
C(2)H-1'		5.848	5.862	5.802	5.868	5.839
U(3)H-1'			5.835			
C(3)H-1'				5.881		
G(3)H-1'					5.834	
A(3)H-1'						6.041
C(2)H-5		5.873	5.840	5.807	5.858	5.826
U(3)H-5			5.804			
C(3)H-5				5.960		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Chemical shifts of oligomers ACU, and ACC were recorded at 400.135 MHz.

Table 15. Chemical shifts[a] of the Diribonucleotide AG and the Triribonucleotides AGX in D_2O[b].

Proton	A	AG	AGU	AGC	AGG	AGA
A(1)H-8	8.313	8.238	8.233	8.242	8.212	8.185
A(1)H-2	8.260	8.186	8.180	8.180	8.167	8.103
G(2)H-8		7.942	7.943	7.926	7.906	7.898
U(3)H-6			7.751			
C(3)H-6				7.744		
G(3)H-8					7.935	
A(3)H-8						8.286
A(3)H-2						8.126
A(1)H-1'	6.065	5.967	5.967	5.974	5.946	5.922
G(2)H-1'		5.842	5.837	5.812	5.785	5.743
U(3)H-1'			5.863			
C(3)H-1'				5.886		
G(3)H-1'					5.837	
A(3)H-1'						6.030
U(3)H-5			5.775			
C(3)H-5				5.876		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Chemical shifts of the oligomer AGU, were recorded at 400.135 MHz.

Table 16. Chemical shifts[a] of the Diribonucleotide AA and the Triribonucleotides AAX in D_2O[b].

Proton	A	AA	AAU	AAC	AAG	AAA
A(1)H-8	8.313	8.210	8.225	8.235	8.191	8.172
A(1)H-2	8.260	8.112	8.118	8.100	8.091	8.043
A(2)H-8		8.297	8.294	8.276	8.243	8.210
A(2)H-2		8.167	8.193	8.175	8.142	8.072
U(3)H-6			7.695			
C(3)H-6				7.719		
G(3)H-8					7.901	
A(3)H-8						8.239
A(3)H-2						8.120
A(1)H-1'	6.065	5.906	5.923	5.928	5.894	5.872
A(2)H-1'		6.032	6.024	5.997	5.963	5.906
U(3)H-1'			5.824			
C(3)H-1'				5.817		
G(3)H-1'					5.800	
A(3)H-1'						5.987
U(3)H-5			5.710			
C(3)H-5				5.841		

[a]Chemical shifts are in ppm relative to DSS using tert-Butylalcohol-OD as an internal reference and are accurate to ±0.005 ppm.

[b]1.0 M NaCl, 1.0 mM NaH_2PO_4, pD 7.0; concentrations 5-12 mM approximately.

Chemical shifts of the oligomer AAC, were recorded at 400.135 MHz.

Fortran Program ASSIGN.

```
      PROGRAM ASSIGN(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
      REAL N3P1(5,5,5),N3P2(5,5,5),N3P3(5,5,5),N5P1(5,5,5),N5P2(5,5,5)
      REAL N5P3(5,5,5),P(20,3), H1(4),H8(4),H2(4)
      INTEGER B(4), LENGTH, IA(20), A(20),BASE(20),NUMBER
      INTEGER  K(20), II(20)
        READ(5,1000) (B(K), K=1,4)
 1000   FORMAT(4A1)
*
        READ(5,700) (H1(I),H8(I),H2(I), I=1,4)
  700   FORMAT(3F10.6)
      DO 2 I= 1,5
        DO 3 J=1,4
          DO 3 K = 1, 5
          READ(5,15)N3P1(I,J,K),N3P2(I,J,K),N3P3(I,J,K),N5P1(I,J,K),
     CN5P2(I,J,K),N5P3(I,J,K)
          WRITE(6,55)N3P1(I,J,K),N3P2(I,J,K),N3P3(I,J,K),N5P1(I,J,K),
     CN5P2(I,J,K),N5P3(I,J,K)
    3   CONTINUE
    2 CONTINUE
   55   FORMAT(1H0,6F10.4)
   15   FORMAT(6F10.5)
        READ(5,1100) NUMBER
 1100   FORMAT(I5)
        DO 2000 KK = 1, NUMBER
*
*     SET ALL PROTON VALUES TO ZERO.
*
      DO  1 IJ = 1, 20
        DO  1 JJ = 1, 3
          P(IJ,JJ) = 0.0
    1 CONTINUE
*
*     READ IN THE LENGHT OF THE SEQUENCE TO BE DETERMINED.
*
      READ(5,25) LENGTH
   25 FORMAT(I5)
*
*     HERE ONE READS IN THE THE SEQUENCE FOR WHICH THE CHEMICAL
*     SHIFTS ARE TO BE DETERMINED.
*
        READ(5,35) (A(I), I= 1,LENGTH)
   35   FORMAT(20A1)
*
        WRITE(6,40)
   40   FORMAT(1H1/1H0,56X,16HS E Q U E N C E  )
*
        WRITE(6,45)( A(I), I = 1, LENGTH)
   45   FORMAT(1H0//1H0,55X,20A3)
*
*     THE FACTOR WHICH DETERMINES FIRST AND SECOND BASE TYPE
*     IS DETERMINED AT THIS POINT.
      IA(1)=1
      DO 5 I = 1, LENGTH
        BASE(I) = 1
          DO 5 J = 1,4
            IF(A(I).EQ.B(J)) IA(I+1) = J + 1
```

```
      PROGRAM ASSIGN     73/173  TS TRACE
                  5   CONTINUE
                      DO 10 I = 1, LENGTH
      60                  DO 80 J = 1, 4
                          IF(A(I).NE.B(J)) GO TO 80
                            DO 88 K = 1,3
                              II(K+1) = I+K
                                  IF(II(K+1).GT.LENGTH) II(K+1) =0
      65                          II(K+4) = I-K
                              IF(II(K+4).LE.0) II(K+4) = 0
                  88        CONTINUE
                            DO  90 JK = 2 , 7
                              K(JK) = IA( II(JK) + 1 )
      70            90      CONTINUE
                    F(I,1)=H1(J)+N3P1(1,J,K(2))+N3P2(1,J,K(3))+N3P3(1,J,K(4))
                   C+N5P1(1,J,K(5))+N5P2(1,J,K(6))+N5P3(1,J,K(7))
                      IF (J.LE.2) GO TO 180
                    F(I,2)=H8(J)+N3P1(3,J,K(2))+N3P2(3,J,K(3))+N3P3(3,J,K(4))
      75           C+N5P1(3,J,K(5))+N5P2(3,J,K(6))+N5P3(3,J,K(7))
                    F(I,3)=H2(J)+N3P1(4,J,K(2))+N3P2(4,J,K(3))+N3P3(4,J,K(4))
                   C+N5P1(4,J,K(5))+N5P2(4,J,K(6))+N5P3(4,J,K(7))
                    GO TO 10
                  180   CONTINUE
      80            F(I,2)=H8(J)+N3P1(2,J,K(2))+N3P2(2,J,K(3))+N3P3(2,J,K(4))
                   C+N5P1(2,J,K(5))+N5P2(2,J,K(6))+N5P3(2,J,K(7))
                      IF(J.NE.1) GO TO 10
                    F(I,3)=H2(J)+N3P1(5,J,K(2))+N3P2(5,J,K(3))+N3P3(5,J,K(4))
                   C+N5P1(5,J,K(5))+N5P2(5,J,K(6))+N5P3(5,J,K(7))
      85            80    CONTINUE
                  10    CONTINUE
                    WRITE(6,510) (BASE(I),J=1,LENGTH)
                  510   FORMAT(1H0///1H0,10X,6H BASE  ,29X,10I6)
                        WRITE(6,515)(A(I),I=1,LENGTH)
      90            515     FORMAT(1H0,50X,10A6)
                    WRITE(6,520)( F(I,1),I = 1, LENGTH)
                  520   FORMAT(1H0///1H0,10X,6H H-1P  ,29X,10F6.3)
                    WRITE(6,525)( F(I,2), I = 1, LENGTH)
                  525     FORMAT(1H0//1H0,10X,6H H-8,H-6  ,27X,10F6.3)
      95            WRITE(6,530)( F(I,3), I = 1, LENGTH)
                  530     FORMAT(1H0//1H0,10X,6H H-2,H-5  ,27X,10F6.3)
                  200   CONTINUE
                                WRITE(6,600)
                  600           FORMAT(1HC,14H END OF OUTPUT    )
     100            STOP
                    END

-*ENTRY POINTS--

         1370   ASSIGN          38   INPUT#          548   OUTPUT#          38   TAPE5#
          548   TAPE6#
```

--EXTERNALS--

INPCI. INPCR. MMURPV. OUTCI. OUTCR. STOP.

--STATEMENT LABELS--

1	ID	0B	.2	ID	0B	.3	ID	0B	.5	ID	0B			
.10	D	6400	.15	F	7530	.25	F	7570	.35	F	7610			
.40	F	7630	.45	F	7700	.55	F	7500	.80	D	6350			
.89	ID	0B	.90	ID	0B	.180		5530	.510	F	7730			
.515	F	7770	.520	F	10020	.525	F	10070	.530	F	10140			
.600	F	10210	.700	F	7460	.1000	F	7440	.1100	F	7550			
.2000	D	7320												

--VARIABLE MAP--

A	I	U	11520	20	ASSIGN	-		1370	ENTRY		
B	I	U	11760	4	BASE	I		27430		20	
H1	R	U	12040	4	H2	R	U	27370		4	
H8	R	U	11460	4	I	I		12100			
IA	I		21210	20	II	I		27670		20	
IJ	I		21450		INPCI.	-			EXTERNAL.		
INPCR.	-			EXTERNAL.	INPUT#	-		38	ENTRY		
J	I		12110		JJ	I		21460			
JK	I		30130		K	I		12120		20	
KK	I		12220		LENGTH	I		25410			
MMURPV.	-			EXTERNAL.	NUMBER	I		12030			
N3P1	R	U	21470	125	N3P2	R	U	23440		125	
N3P3	R	U	25420	125	N5P1	R	U	12360		125	
N5P2	R	U	14330	125	N5P3	R	U	16300		125	
OUTCI.	-			EXTERNAL.	OUTCR.	-			EXTERNAL.		
OUTPUT#	-		540	ENTRY	P	R		20250		60	
STOP.	-			EXTERNAL.	TAPE5#	-		38	ENTRY		
TAPE6#	-		540	ENTRY							

70600 PROGRAM-UNIT LENGTH 60 SYMBOLS

440008 CM STORAGE USED 1.E42 SECONDS

Fortran Program SEQUEN.

```
      PROGRAM SEQUEN(INPUT,OUTPUT,TAPE5=INPUT,TAPE6=OUTPUT)
C*
C*       PROGRAM SEQUEN IS A ROUTINE WRITTEN TO FIND  A GIVEN RNA
C*       OR DNA SEQUENCE FROM PROTON N.M.R. SPECTRA DATA.
C*
C*       THE BASE RATIO IS READ IN IN THE ORDER U,C,A,G.
C*       NEXT THE CHEMICAL SHIFT VALUES ARE READ IN ALONG WITH THEIR
C*       PARTICULAR BASE TYPE OR KIND. FOR INSTANCE AN H-8  VALUE IS PREVIOUSLY
C*       DETERMINED  TO BE AN ADENOSINE OR A GANOSINE.
C*
C*       THE FUNCTION OF THE MAIN PROGRAM IS TO FIT THE CALCULATED VALUES
C*       FOR A GIVEN SEQUENCE TO THE OBSERVED VALUES.
C*       THE SEQUENCE WHICH IS CONSIDERED THE CORRECT SEQUENCE IS THE
C*       ONE THAT HAS THE LOWEST COUNT AND DIFFERENCE VALUES.
C*
C*     WRITTEN BY PAUL A. HADER  1981
C*
      COMMON VEC(20),N3P1(5,5,5),N3P2(5,5,5),N3P3(5,5,5),
     CN5P1(5,5,5),N5P2(5,5,5),N5P3(5,5,5),H1(4),H8(4),H2(4),P(20,3)
      REAL H1,H8,H2,N3P1,N3P2,N3P3,N5P1,N5P2,N5P3,P,PROTON(20,3)
      REAL PRODIFF(20), MIN , DIFFER
      INTEGER VEC,BBASE(20),COUNT,N,VV(4),BASE(20),M,FACT
         READ(5,700) (H1(I),H8(I),H2(I), I=1,4)
  700    FORMAT(3F10.6)
      DO 2 I= 1,5
        DO 3 J=1,4
          DO 3 K = 1, 5
          READ(5,15)N3P1(I,J,K),N3P2(I,J,K),N3P3(I,J,K),N5P1(I,J,K),
     CN5P2(I,J,K),N5P3(I,J,K)
          WRITE(6,55)N3P1(I,J,K),N3P2(I,J,K),N3P3(I,J,K),N5P1(I,J,K),
     CN5P2(I,J,K),N5P3(I,J,K)
    3   CONTINUE
    2 CONTINUE
   55  FORMAT(1H0,6F10.4)
   15  FORMAT(6F10.5)
C*
       DO 222 LP = 1, 10
C*
C*
C*       READ IN THE BASE RATIO IN THE ORDER U , C , A , G .
C*       FOR DNA READ IN THE BASE RATIO IN THE ORDER T,C,A,G .
C*
      READ(5,1000) (VV(I), I=1,4)
 1000  FORMAT(4I3)
      WRITE(6,1005) (VV(I), I=1,4)
 1005  FORMAT("1","NUMBER OF BASES:",2X,I3,"U",I3,"C",I3,"A",I3,"G"//)
      N=VV(1)+VV(2)+VV(3)+VV(4)
        IF(N.EQ.0) STOP
      READ(5,500)(BBASE(I),I=1,N)
  500  FORMAT(20A1)
      DO 555  K = 1, 3
        IF( N .GT. 8 ) GO TO 33
       READ(5,550)(PROTON(J,K),J=1,N)
  550 FORMAT(8F10.5)
        GO TO  34
   33     READ(5,550) (PROTON(J,K),J = 1,8)
          READ(5,550)(PROTON(J,K),J = 9, N)
```

```
    34    CONTINUE
    555   CONTINUE
    *
    *        SUBROUTINE SEQUENCE GIVES THE FIRST SEQUENCE  WITH THE BASES
    *        IN ASSENDING ORDER U,C,A,G.
                        CALL SEQUENC(VV,N)
    *
    *        FUNCTION FACT DETERMINES THE NUMBER OF POSSIBLE SEQUENCES
    *        FROM THE GIVEN BASE RATIO.
    *
          M=FACT(N)/(FACT(VV(1))*FACT(VV(2))*FACT(VV(3))*FACT(VV(4)))
                   WRITE(6,35) M
    35.                FORMAT(1H0,"NUMBER OF SEQUENCES = ",10X,I10)
           DO 115 I = 1, M
             TOTAL = 0.00
           DIFFER = 0.0
           COUNT  = 0
    *
    *        SUBROUTINE ASSIGN DETERMINES THE CALCULATED CHEMICAL SHIFTS FOR
    *        A GIVEN BASE SEQUENCE.
    *
             CALL ASSIGN(N,BASE)
              DO 20 KK = 1, 3
                 DO 25 II = 1,   N
                        MIN   = 0.5000
                 DO 30   JJ = 1 , N
             PRODIFF(JJ) = 0.5000
    *
          IF(BBASE(II).EQ.BASE(JJ)) PRODIFF(JJ)=ABS(PROTON(II,KK)-P(JJ,KK))
    *
                   IF(PRODIFF(JJ).LT.MIN) MIN = PRODIFF(JJ)
    *
    30    CONTINUE
                       TOTAL = TOTAL + MIN
    *
    *     ONCE A FIT IS FOUND THE CALCULATED VALUE IS REMOVED FROM FURTHER
    *     CONSIDERATION.
    *
                       KOUNT  = 0
              DO   60 L = 1 , N
                       IF(MIN  .EQ. PRODIFF(L)) GO TO 70
             GO TO 60
    70              KOUNT  = KOUNT + 1
                   IF(KOUNT  .LT. 2 ) P(L,KK)  = 0.000
    60      CONTINUE
    *
                   IF( MIN .LT. 0.043000) GO TO 25
    *
                   COUNT   = COUNT + 1
                   IF ( COUNT .GT. 9 ) GO TO 110
    *
                       TEMPORY = ( MIN - 0.04200 )
            DIFFER = DIFFER + TEMPORY
    25    CONTINUE
    20    CONTINUE
                   AVEDIFF = TOTAL/( N*3 - VV(4))
```

```
                      WRITE(6,200) (VEC(J), J = 1,N)
  200                 FORMAT(1H0,20A1)
                    WRITE(6,100) COUNT, DIFFER, AVEDIFF
  100   FORMAT(1H+,30X,"COUNT = ",I10,5X,"DIFFERENCE = ",F10.4,5X,
     C"ABSOLUTE AVE. DIFF. =",F10.4)
  110               CONTINUE
*
*       SUBROUNTINE SEQ GIVES EACH NEW SEQUENCE AFTER THE LAST ONE HAS
*       CALCULATED AND FITTED TO THE OBSERVED VALUES.
*
                    CALL SEQ(N)
*
  115   CONTINUE
*
  222   CONTINUE
*
        STOP
        END
```

--COMMON BLOCKS--

```
     1512B  //
```

--ENTRY POINTS--

```
       3B   INPUT#          54B   OUTPUT#        137B   SEQUEN          3B   TAPE5#
      54B   TAPE6#
```

--EXTERNALS--

```
ASSIGN    FACT      INPCI.    INPCR.    MMURPV.   OUTCI.    OUTCR.    SEQ
SEQUENC   STOP.
```

--STATEMENT LABELS--

```
    .2   ID    0B        .3   ID    0B        .15  F   6458      .20  D   5738
   .25   D   570B        .30  ID    0B        .33      3768      .34      4270
   .35   F   665B        .55  F   6428        .60  D   5500      .70      5368
  .100   F   674B        .110      6208        .115 D   6238      .200 F   6728
  .222   D   6260        .500 F   6618        .550 F   6638      .555 ID    0B
  .700   F   6408        .1000 F  6478        .1005 F  6518
```

--VARIABLE MAP--

```
ABS       R                INTRINSIC          ASSIGN    R                SUBROUTINE
AVEDIFF   R   1246B                           BASE      I   10718B             20
BBASE     I U 1215B              20           COUNT     I   10428
DIFFER    R   1117B                           FACT      I                FUNCTION
H1        R   14028 //           4           H2        R   14128 //           4
H8        R   14068 //           4           I         I   10378
II        I   1243B                          INPCI.    -                EXTERNAL.
```

VARIABLE MAP

INPCR.	-		EXTERNAL.			INPUT#	-	30	ENTRY
J	I	10400				JJ	I	11160	
K	I	10410				KK	I	10320	
KOUNT	I	12440				L	I	12450	
LP	I	12420				M	I	12410	
MIN	R	12140				MMURPV.	-		EXTERNAL.
N	I	10670				N3P1	R	240	// 125
N3P2	R	2210	//	125		N3P3	R	4160	// 125
N5P1	R	6130	//	125		N5P2	R	10100	// 125
N3P3	R	12050	//	125		OUTCI.	-		EXTERNAL.
OUTCR.	-		EXTERNAL.			OUTPUT#	-	540	ENTRY
P	R	14160	//	60		PRODIFF	R	10430	20
PROTON	R U	11200		60		SEQ	R		SUBROUTINE
SEQUEN	-	1370	ENTRY			SEQUENC	R		SUBROUTINE
STOP.	-		EXTERNAL.			TAPE5#	-	30	ENTRY
TAPE6#	-	540	ENTRY			TEMPORY	R	11150	
TOTAL	R	10700				VEC	I	00	// 20
VV	I	10330		4					

70250 PROGRAM-UNIT LENGTH 72 SYMBOLS

430008 CM STORAGE USED 1.719 SECONDS

```
      FUNCTION FACT        73/173  TS TRACE                FTN 4.8+552        83/02/23. 20.53.16        PAGE      1

                           INTEGER FUNCTION FACT(L)
                            FACT = 1
                              DO 10 I = 1, L
                        10      FACT = FACT * I
                           5  RETURN
                              END

 --STATEMENT LABELS--
          ..10   ID         0B

 --VARIABLE MAP--
      FACT          -        17B ENTRY                    I             240
      L             I A      0B                           VALUE.        238

     25B  PROGRAM-UNIT LENGTH              5 SYMBOLS
    41000B CM STORAGE USED                 .049 SECONDS
```

```
      SUBROUTINE SEQUENC    73/173  TS TRACE                FTN 4.8+552      83/02/23. 20.53.16      PAGE    1

                            SUBROUTINE SEQUENC(VV,II)
                           IMPLICIT INTEGER (A-Z)
                            COMMON VEC(20)
                           COMMON /VECTOR/ V(20)
                           DIMENSION NUC(5), VV(4)
                            INTEGER VEC,N,V
                            DATA NUC/" ","T","C","A","G"/
                             DO 50 II = 1, 20
                            V(II) = 0
     10              50    CONTINUE
                           IF(N.EQ.0) RETURN
                              JJ=0
                             DO 10 I = 1,4
                               JMIN = JJ + 1
                               JJ = JJ +VV(I)
                               DO 10 J = JMIN,JJ
                        10   V(J) = I
                     C     ORDER THE SEQUENCE INTO LOWEST NUMBER TERM
                             DO 4    I = 1, 20
     20                       VEC(I) = NUC(1+V(I))
                         4   CONTINUE
                           RETURN
                           END

--COMMON BLOCKS--

          240  /VECTOR/              240  //

--STATEMENT LABELS--

          .4    ID        0B            .10   ID        0B            .50   ID        0B

--VARIABLE MAP--

      I          I          56B                        II         I          60B
      J          I          57B                        JJ         I          66B
      JMIN       I          55B                        N          I A         0B
      NUC        I          61B          5             SEQUENC    -          51B ENTRY
      V          I           0B /VECTOR/  20            VEC        I           0B //      20
      VV         I A         0B           4

    137B  PROGRAM-UNIT LENGTH          14 SYMBOLS
    10000B CM STORAGE USED             .143 SECONDS
```

```
      SUBROUTINE  ASSIGN (LENGTH,BASE)
      COMMON VEC(20),N3P1(5,5,5),N3P2(5,5,5),N3P3(5,5,5),
     CN5P1(5,5,5),N5P2(5,5,5),N5P3(5,5,5),H1(4),H8(4),H2(4),P(20,3)
      REAL H1,H8,H2,N3P1,N3P2,N3P3,N5P1,N5P2,N5P3, P
      INTEGER VEC,B(4),LENGTH, IA(20), A(20), BASE(20)
      INTEGER  K(20), II(20)
       DATA B/"A","G","C","T"/
*     SET ALL PROTON VALUES TO ZERO.
*
      DO  1  IJ = 1, 20
           DO   1 JJ = 1, 3
           P(IJ,JJ) = 0.0
    1 CONTINUE
      DO 55  IK = 1, LENGTH
           A(IK) = VEC(IK)
        BASE(IK) = A(IK)
   55  CONTINUE
*
*     THE FACTOR WHICH DETERMINES FIRST AND SECOND BASE TYPE
*     IS DETERMINED AT THIS POINT.
*
      IA(1)=1
      DO 5 I = 1, LENGTH
           DO 5 J = 1,4
              IF(A(I).EQ.B(J)) IA(I+1) = J + 1
    5 CONTINUE
      DO 10 I = 1, LENGTH
           DO 80 J = 1, 4
           IF(A(I).NE.B(J)) GO TO 80
             DO 88 K = 1,3
               II(K+1) = I+K
                  IF(II(K+1).GT.LENGTH) II(K+1) =0
               II(K+4) = I-K
             IF(II(K+4).LE.0) II(K+4) = 0
   88        CONTINUE
           DO  90 JK = 2 , 7
             K(JK) = IA( II(JK) + 1 )
   90      CONTINUE
     P(I,1)=H1(J)+N3P1(1,J,K(2))+N3P2(1,J,K(3))+N3P3(1,J,K(4))
    C+N5P1(1,J,K(5))+N5P2(1,J,K(6))+N5P3(1,J,K(7))
      IF (J.LE.2) GO TO 180
     P(I,2)=H8(J)+N3P1(3,J,K(2))+N3P2(3,J,K(3))+N3P3(3,J,K(4))
    C+N5P1(3,J,K(5))+N5P2(3,J,K(6))+N5P3(3,J,K(7))
     P(I,3)=H2(J)+N3P1(4,J,K(2))+N3P2(4,J,K(3))+N3P3(4,J,K(4))
    C+N5P1(4,J,K(5))+N5P2(4,J,K(6))+N5P3(4,J,K(7))
     GO TO 10
  180  CONTINUE
     P(I,2)=H8(J)+N3P1(2,J,K(2))+N3P2(2,J,K(3))+N3P3(2,J,K(4))
    C+N5P1(2,J,K(5))+N5P2(2,J,K(6))+N5P3(2,J,K(7))
     IF(J.NE.1) GO TO 10
     P(I,3)=H2(J)+N3P1(5,J,K(2))+N3P2(5,J,K(3))+N3P3(5,J,K(4))
    C+N5P1(5,J,K(5))+N5P2(5,J,K(6))+N5P3(5,J,K(7))
   80    CONTINUE
   10 CONTINUE
      RETURN
```

END

--COMMON BLOCKS--

1,5120 //

--STATEMENT LABELS--

.1	ID	0B	.5	ID	0B	.10	0	3070	.55	ID	0B
.80	0	3040	.89	ID	0B	.90	ID	0B	.180		2220

--VARIABLE MAP--

A	I	3330	20	ASSIGN	-	3160	ENTRY	
B	I	3570	4	BASE	I A	0B		20
H1	R	14020 //	4	H2	R	14120 //		4
H4	R	14060 //	4	I	I	3630		
IA	I	4130	20	II	I	4370		20
IJ	I	4110		IK	I	4630		
J	I	3640		JJ	I	4120		
JK	I	4640		K	I	3650		20
L_NGTH	I A	0B		N3P1	R	240 //		125
N3P2	R	2210 //	125	N3P3	R	4160 //		125
N5P1	R	6130 //	125	N5P2	R	10100 //		125
N5P3	R	12050 //	125	P	R	14160 //		60
V_C	I	0B //	20					

21770 PROGRAM-UNIT LENGTH 33 SYMBOLS

430000 CM STORAGE USED .989 SECONDS

```
                      SUBROUTINE SEQ(N)
      IMPLICIT INTEGER (A-Z)
      COMMON VEC(20)
                      COMMON /VECTOR/ V(20)
        DIMENSION NUC(5)
      INTEGER  VEC,V,NUC,N,X,H
        DATA  NUC/" ","T","C","A","G"/
        DO 2 I = 2 , N
          L = N-I+2
        IF(V(L).GT.V(L-1)) GO TO 3
    2 CONTINUE
      RETURN
    3  X=1+V(L-1)
      DO 8 H=X,4
         DO 8 I = L, N
            IF (V(I).EQ.H) GO TO 9
    8 CONTINUE
    9 X=V(L-1)
         V(L-1)=H
         V(I)=X
      DO  5 I = L, N
         DO 5 J = I, N
         IF(V(I).LT.V(J)) GO TO 5
           X = V(I)
         V(I) = V(J)
        V(J) = X
    5 CONTINUE
         DO 6 I = 1 , 20
    6      VEC(I) = NUC(1+V(I))
      RETURN
      END
```

--COMMON BLOCKS--

24B	/VECTOR/	24B //

--STATEMENT LABELS--

.2	D	148	.3	218	.5	D	718	.6	I9	08				
.8	D	348	.9	428										

--VARIABLE MAP--

I	I	1158			J	I	1168		
L	I	1178			N	I	08		
NUC	I	1208		5	SEQ	-	1118	ENTRY	
V	I	08	/VECTOR/	20	VEC	I	08	//	20
H	I	1258			X	I	1268		

177B PROGRAM-UNIT LENGTH 16 SYMBOLS

41000B CM STORAGE USED .283 SECONDS

REFERENCES

Alkema, D., Bell, R.A., Hader, P.A. and Neilson, T. (1981a) in "Biomolecular Stereodynamics", Ed., Sarma, R.H., Vol. 1, Adenine Press, New York pp 417-428.

Alkema, D., Bell, R.A., Hader, P.A. and Neilson, T. (1981b) J.A.C.S. 103, 2866-2868.

Alkema, D., Bell, R.A., Hader, P.A. and Neilson, T. (1981c) Febs Letters 136(1), 70-71.

Alkema, D., Hader, P.A., Bell, R.A. and Neilson, T. (1982) Biochemistry 21, 2109-2117.

Altona, C. (1975) in "Structure and Conformation of Nucleic Acids and Protein-Nucleic Acid Interactions", Ed.,Sundaralingam, M. and Rao, S.T., University Park Press, Baltimore, p 613.

Alvarado-Urbina, G., Sathe, G.M., Liu, W.-C., Gillen, M.F., Duck, P.D., Bender, R. and Ogilvie, K.K. (1981) Science 214, 270-274.

Anderson, S., Bankier, A.T., Barrell, B.G., de Bruijn,

M.H.L. and Coulson, A.r., et al. (1981) Nature 290, 457-465.

Arai, K.I., Kawakita, M. and Kaziro, Y. (1972) J. Biol. Chem. 247, 7029-7037.

Armstrong, I.L. and Tate, W.P. (1978) J. Mol. Biol. 120, 155-166.

Armstrong, I.L. and Tate, W.P. (1980) Febs Letters 109, 228-232.

Arnott, S., Dover, S.D. and Wonacott, A.J. (1969) Acta Crystallog. B. 25, 2192.

Arnott, S., and Hukins, D.W.L. (1972) Biochem. Biophys. Res. Comm. 47, 1504-1509.

Atkins, J.F. (1979) Nucleic Acids Research 7, 1035-1041.

Beaudet, A.L. and Caskey, C.T. (1970) Nature 227, 38-40.

Bell, R.A., Everett, J.R., Hughes, D.W., Alkema, D., Hader, P., Neilson, T. and Romaniuk, P.J. (1981) Biopolymers 20, 1383-1398.

Bienz, M., and Kubli, E. (1981) Nature 294, 188.

Borer, P.N., Bubienko, E., Cruz, p. (1982) Submitted to Nature.

Borer, P.N., Kan, L.S. and Ts'o, P.O.P. (1975) Biochemistry 14, 4847.

Borer, P.N. and McFarland, G.D. (1979) Nucleic Acids Research 7, 1067-1080.

Broyde, S., and Hingerty, B. (1978) Nucleic Acids Research 5, 2729.

Bruce, A.G. and Uhlenbeck, O.C. (1978) Nucleic Acids Research 5(10), 3665-3677.

Bruce, A.G. and Uhlenbeck, O.C. (1981) Submitted to Biochemistry.

Bubienko, E., Uniack, M.A. and Borer P.N. (1981) Biochemistry 20, 6987-6994.

Capecchi, M.R. (1967) Biochem. Biophys. Res. Commun. 28, 773-777.

Capecchi, M.R. and Klein, H.A. (1969) Cold Spring Harbour

Symp. Quant. Biol. 34, 469-477.

Capecchi, M.R. and Klein, H.A. (1970) Nature 226, 1029-1033.

Caskey, C.T. (1979) in "Molecular Mechanisms of Protein Biosynthesis", Ed., Weissbach, H. and Pestka, S., Academic Press Inc., New York pp 443-465.

Caskey, C.T., Bosch, L. and Konecki, D.S. (1977) J. Biol. Chem. 252, 4435-4437.

Caskey, C.T., Tomkins, R., Scolnick, E., Caryk, T. and Nirenberg, M. (1968) Science 162, 135-138.

Cheng, D.M., Dhingra, M.M. and Sarma, R.H. (1978) Nucleic Acids Research 5, 4399.

Cheng, D.M., Kan, L.-S. Leutzinger, E., Jayaraman, K., Miller, P.S. and Ts'O, P.O.P. (1982) Biochemistry 21, 621-630.

Ciampi, M.S., Melton, D.A. and Cortese, R. (1982) Proc. Natl. Acad. Sci. U.S.A. 79, 1388-1392.

Clark, B.F.C. (1977) Prog. Nucl. Acid, Res. Mol. Biol. 20, 1-19.

Clementi, E. and Corongiu, G. (1981) in "Biomolecular Stereodynamics", Ed., Sarma, R.H., Vol. 1, Adenine Press, New York pp 209-260.

Conner, B.N., Takano, T., Tanaka, S., Itakura, K. and Dickerson, R.E. (1982) Nature 295, 294-299.

Cornelis, A.G., Haarsroot, H.J., denHartog, J.F., Rooij, M., vanBoom, J.H. and Cornelis, A. (1979) Nature 281, 235-236.

Cruz, P., Subienko, E. and Borer, P.N. (1982) Nature 298, 198-200.

Davies, D.B. (1978) Prog. Nucl. Magn. Reson. Spectrosc. 12, 135.

Davies, D. and Zimmerman, S. (1980) Nature 283, 11-12.

Dhingra, M.M. and Sarma, R.H. (1978) Nature 272, 798.

Dickerson, R.E., Drew, H.R., Conner, B.N., Wing, R.M., Fratini, A.V. and Lopka, M.L. (1982) Science 216, 475-484.

Edge, M.D., Greene, A.R., Heathcliffe, G.R., Meacock, P.A., Schuch, W., Scanlon, D.B., Atkinson, T.C., Newton, C.R., and Markham, A.F. (1981) Nature 292, 756-762.

Engelberg-Kulka, H., (1981) Nucleic Acids Research 9, 983-991.

England, T.E. and Neilson, T. (1976) Can. J. Chem. 54, 1714.

Erbe, R.W. and Leder, P. (1968) Biochem. Biophys. Res. Commun. 31, 798-803.

Evans, F.E. and Sarma, R.H. (1976) Nature 263, 567-572.

Everett, J.R., Hughes, D.W., Bell, R.A., Alkema, D., Neilson, T. and Romaniuk, P.J. (1980) Biopolymers 19, 557-573.

Ezra, F.S., Lee, C.-H., Kondo, N.S., Danyluk, S.S. and Sarma, R.H. (1977) Biochemistry 16, 1977-1987.

Fluck, M.M., Salser, W. and Epstein, R.H. (1977) Molec. gen. Genet. 151, 137-149.

Fox, T.D. and Weiss-Brummer, B. (1980) Nature 288, 60-63.

Fritz, H., Belagaje, R., Brown, E.L., Fritz, R.H., Jones, R.A., Lees, R.G. and Khorana, H.G. (1978) Biochemistry 17, 1257.

Gait, M.J., Matthes, H.W.D., Singh, M., and Titmas, R.C. (1982) J. Chem. Soc. Chem. Comm. 37-40.

Gait, M.J. and Sheppard, R.C. (1977). Nucleic Acids Research 4, 1135-1158.

Ganoza, M.C. (1966) Cold Spring Harbour Symp. Quant. Biol. 31, 273-278.

Ganoza, M.C. (1977) Can. J. Biochem. 55, 267-281.

Ganoza, M.C. and Barraclough, N. (1975) FEBS Letters 53, 159-163.

Ganoza, M.C., Barracough, N. and Wong, J.T.-F. (1976) Eur. J. Biochem. 65, 613-622.

Ganoza, M.C., Buckingham, K., Hader, P.A. and Neilson, T. Manuscript in Preparation.

Ganoza, M.C., Fraser, A. and Neilson, T. (1978) Biochemistry 17, 2769-2775.

Ganoza, M.C. and Nakamoto, T.N. (1966) Proc. Natl. Acad. Sci. U.S.A. 55, 162-169.

Ganoza, M.C. and Thach, R. (1969) Cold Spring Harbour Symp. Quant. Biol. 34, 488-490.

Ganoza, M.C. and Tompkins, J.K.N. (1970) Biochem. Biophys. Res. Comm. 40(6), 1455-1463.

Ganoza, M.C., Sullivan, P., Cunningham, C., Hader, P., Kofoid, E.C. and Neilson, T. (1982) J. Biol. Chem. 257, 8228-8232.

Ganoza, M.C., Van derMeer, J., Debreceni, N., and Phillips, S.L. (1973) Proc. Natl. Acad. Sci. USA 70, 31-35.

Geerdes, H.A.M., VanBoom, J.H. and Hilbers, C.W. (1980a) J. Mol. Biol. 142, 195-217.

Geerdes, H.A.M., VanBoom, J.H. and Hilbers, C.W. (1980b) J. Mol. Biol. 142, 219-230.

Geller, A.I. and Rich, A. (1980) Nature 283, 41-46.

Ghosh, K. and Ghosh, H.P. (1970) Biochem. Biophys. Res. Comm. 40, 135-143.

Giessner-Prettre, C., Pullman, B., Borer, P.N., Kan, L.-S. and Ts'o, P.O.P. (1976) Biopolymers 15, 2277-2286.

Giessner-Prettre, C., and Pullman, B. (1976) Biochem. Biophys. Res. Comm. 70, 578-581.

Gregoire, R.J. and Neilson, T. (1978) Can. J. Chem. 56, 487.

Griffin, B.E., Jarman, M., Reese, C.B., Sulston, J.E. and Trentham, D.R. (1966) Biochemistry 5, 3638.

Grosjean, H.J., Henau, S. and Crothers, D.M. (1978a) Proc. Natl. Acad. Sci. U.S.A. 75, 610-614.

Grosjean, H., Sankoff, D., Min Jou, W., Fiers, W. and Cedergren, R.J. (1978b) J. Mol. Evol. 12, 113-119.

Grosjean, H., Soll, D. and Crothers, D. (1977) Arch. internat. Physiol. Biochim. 85, 414-415.

Grosjean, H., Soll, D.G. and Crothers, D.M. (1976a) J. Mol. Biol. 103, 499-519.

Grosjean, H., Soll, D.G. and Crothers, D. (1976b) Arch. internat. Physiol. Biochim. 84, 163-164.

Grosjean, H., Soll, D.g. and Crothers, D. (1975) Arch. internat. Physiol. Biochim. 83, 970-971.

Gumport, R.I. and Uhlenbeck, O.C. (1980) in "Gene Amplification and Analysis, Analysis of Nucleic Acid Structure by Enzymatic Methods", Ed., Chirikjian, J.G. and Papas, T.S., Vol. 2, Elsevier North Holland Inc., New York.

Guschlbauer, W. (1980) Biochim. Biophys. Acta 610, 47-55.

Gutte, B., Daumigen, M. and Wittschieber, E. (1979) Nature 281, 650-655.

Haasnoot, C.A.G., denHartog, J.H.J., deRooij, J.F.M., vanBoom, J.H. and Altona, C. (1980) Nucleic Acids Research 8, 169-181.

Hader, P.A., Alkema, D., Bell, R.A. and Neilson, T. (1982) J. Chem. Soc., Chem. Commun. 1017.

Hader, P.A., Neilson, T., Alkema, D., Kofoid, E.C. and Ganoza, M.C. (1981) Febs Letters 136(1), 65-69.

Haenni, A.L. and Chapeville, F. (1966) Biochem. Biophys. Acta 144, 135-148.

Hall, M.N., Gabay, J., Debarbouille, M. and Schwartz, M. (1982) Nature 295, 616-618.

Herrington, M.B., Doherty, M.J. and Ganoza, M.C. (1975) Nature 256, 678-679.

Hofstetter, H., Monstein, H.J., and Weissman, C. (1974) Biochem. Biophys. Acta 374, 238-251.

Holbrook, S.R., Sussman, J.L. and Kim, S. (1981) Science 212, 1275.

Holschuh, K., Bonin, J. and Gassen, H.G. (1980) Biochemistry 19, 5857-5864.

Holschuh, K., Riesner, D. and Gassen, H.G. (1981) Nature 293, 675-677.

Ikakura, K., Hirose, T., Crea, R., Riggs, A., Heyneker, H., Boliver, F., and Boyer, H. (1977) Science 198, 1056.

Johnston, P.D. and Redfield, A.G. (1981) Biochemistry 20,

3996-4006.

Jones, C.R. (1980) J.A.C.S. 102, 5933.

Kallenbach, N.R. and Berman H.M. (1977) Quarterly Review of biophysics 10, 138.

Kan, L.S., Barrett, J.C. and Ts'o, P.O.P. (1973) Biopolymers 12, 2409.

Kastelein, R.A., Remaut, E., Fiers, W. and van Duin, J. (1982) Nature 295, 35-41.

Katagiri, N., Itakura, K. and Narang, S.A. (1974) J. Chem. Soc. Chem. Commun., 325.

Kearns, D.R. (1977) Ann. Rev. Biophys. Bioeng. 6, 477.

Khorana, H.G., Agarwal, K.L., Besmer, P., Buechi, H., Caruthers, M.H., Cashion, P.J., Fridkin,M., Jay, E., Kleppe, K., Kleppe,R., Kumar, A., Loewen,P.C., Miller, R.C., Minamoto, K., Panet, A., RajBhandary, U.L., Ramamoorthy, B., Sekiya,T., Takaya, T. and van de Sande, J.H. (1976) J. Biol. Chem. 251, 565 et seq.

Kim, S.-H. and Quigley, G.J. (1979) Methods in Enzymology

LIX, 111.

Kim, S.-H., Quigley, G.J., Suddath, F.L., McPherson, A., Sneden, D., Kim, J.J., Weinzierl, J. and Rich, A. (1973) Science 179, 285-288.

Kim, S.-H., Suddath, F.L., Quigley, G.J., McPherson, A., Sussman, J.L., Wang, A.H.J., Seeman, N.C. and Rich, A. (1974) Science 185, 435-440.

Klein, H.A. and Capecchi, M.R. (1971) J. Biol. Chem. 246, 1055-1061.

Kondo, N.S., Ezra, F. and Danyluk, S.S. (1975) FEBS Letters 53(2), 213.

Krugh, T.R., Laing, J.W. and Young, M.A. (1976) Biochemistry 15, 1224-1228.

Labuda, D. and Porschke, D. (1980) Biochemistry 19, 3799-3805.

Lake, J.A. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 1903-1907.

Lancelot, G., Mayer, R., Thuong, N.T., Chassignol, M. and

Helene, C. (1981) Biochimie 63, 785-790.

Lee, C. and Tinoco, I. jr. (1980) Biophysical Chemistry 11, 283-294.

Lee, C.-H., Ezra, F.S., Kondo, N.S., Sarma, R.H. and Danyluk, S.S. (1976) Biochemistry 15, 3627-3639.

de Leeuw, F.A.A.M. and Altona, C. (1982) J.C.S. Perkin II, 375-384.

Levitt, M. and Warshel, A. (1978) J.A.C.S. 100(9), 2607-2613.

Levy, G.C. and Craik, D.J. (1981) Science, 214, 291-299.

Lindahl, T. (1981) Nature, 290,363.

Lodish, H. (1976) Annu. Rev. Biochemistry 46, 39-72.

Lu, P. and Rich, A. (1971) J. Mol. Biol. 58, 513-531.

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

McFarland, G.D. and Borer, P.N. (1979) Nucleic Acids

Research 7(4), 1067-1080.

Menninger, J.R. (1971) Biochem. Biophys. Acta 240, 237-243.

Metzler, D.E. (1977) in "Biochemistry: The Chemical Reactions of Living Cells." New York, NY, Academic Press, Inc., pp 932-937.

Milman, G., Langridge, R. and Chamberlin, M.J. (1967) Proc. Natl. Acad. Sci. U.S.A. 57, 1804.

Model, p. (1982) Nature 295, 15.

Narang, S.A., Brousseau, R., Hsiung, H.M. and Michniewicz, J.J. (1980) Methods in Enzymology 65, 610-621.

Neilson, T., J. Chem. Soc. Chem. Comm. (1969) 1139.

Neilson, T., Alkema, D., Bell, R.A., Hader, P.A. and Hughes, D.W. (1982) Unpublished Results.

Neilson, T., Gregoire, R.J., Fraser, A.R., Kofoid, E.C., Ganoza, M.C. (1979) Eur. J. Biochem. 99, 429-437.

Neilson, T. and Werstiuk, E.S. (1971) Can. J. Chem. 49, 493.

Neuman, J.M., Huynh-Dinh, T., Kan, S.K., Genissel, B., igolen, J. and Tran-Dinh, S. (1982) Eur. J. Biochem. 121, 317-323.

News and Views (1981) Nature 292, 667.

Nichols, J.L. (1970) Nature 225, 147-151.

Ninio, J. (1979) Biochimie 61, 1133-1150.

Nirenberg, N.W. and Leder, P. (1964) Science 145, 1399-1407.

Noller, H.F. and Woese, C.R. (1981) Science 212, 403-411.

Nordheim, A., Pardue, M.L., Lafer, E.M., Moller, A., Stollar, B.D. and Rich, A. (1981) Nature 294, 417-422.

Olson, W.K. (1978) Biopolymers 17, 1015-1040.

Olson, W.K. (1982a) J.A.C.S. 104, 278-286.

Olson, W.K. and Sussman, J.L. (1982b) J.A.C.S. 104, 270-278.

Pace, N.R., Hanura, F. and Spiegelman, S. (1968) Methods in Enzymology 12B, 540-555.

Pardi, A., Martin, F.H. and Tinoco, I. jr. (1981) Biochemistry 20, 3986-3996.

Patel, D.J. (1976) Biopolymers 15, 533-558.

Patel, D.J. (1977) Biopolymers 16, 1635-1656.

Patel, D.J. (1979) Accounts, Chem. Res. 12, 118.

Patel, D.J., Canuel, L.L. and Pohl, F.M. (1979) Proc. Natl. Acad. Sci. U.S.A. 76, 2508-2511.

Patel, D.J. and Hilbers, C.W. (1975) Biochemistry 14, 2651-2660.

Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Itakura, K., Breslauer, K.J. (1982a) Biochemistry 21, 428-436.

Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Dallas, J., Itakura, K., Breslauer, K.J. (1982b) Biochemistry 21, 437-444.

Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Itakura, K., Breslauer, K.J. (1982c) Biochemistry 21, 445-450.

Patel, D.J., Kozlowski, S.A., Marky, L.A., Rice, J.A., Broka, C., Itakura, K., Breslauer, K.J. (1982d) Biochemistry 21, 451-455

Patel, D.J., Kozlowski, S.A., Nordheim, A. and Rich, A. (1982e) Proc. Natl. Acad. Sci. U.S.A. 79, 1413-1417.

Patel, D.J., Pardi, A. and Itakura, K. (1982f) Science 216, 581-595.

Petersheim, M. and Turner, D.H. (1982a) Submitted to Biochemistry.

Petersheim, M. and Turner, D.H. (1982b) Submitted to Biochemistry.

Phillips, D.J. and Roberts, G.C.K. (1980) Biochemistry 19, 4795-4801.

Reese, C.B., Titmas, R.C. and Yau, L. (1978) Tetrahedron Letters, 2727.

Reid, B.R. (1979) Methods in Enzymology LIX, Nucleic Acids and Protein Synthesis (part G), 1-57.

Reid, B.R. (1981) in "Topics In Nucleic Acid Structure", Ed., Neidle, S., John Wiley and Sons Inc., New York pp 113-139.

Reid, B.R. and Hurd, R.E. (1977) Acc. Chem. Res. 10, 396.

Romaniuk, P.J. (1979) Ph.D. Thesis, McMaster University.

Romaniuk, P.J., Hughes, D.W., Gregoire, R.J., Bell, R.A. and Neilson, T. (1979) Biochemistry 18, 5109-5116.

Ryoji, M., Berland, R., and Kaji, A. (1981) Proc. Natl. Acad. Sci. U.S.A. 78, 5973-5977.

Salser, W., Fluck, M. and Epstein, R. (1969) Cold Spring Harbour Symp. Quant. Biol. 513-520.

Sarma, R.H. (1980) in "Nucleic Acid Geometry and Dynamics." Sarma, R.H., Ed., Elmsford, N.Y., Pergamon Press, p 1.

Schmidt, P.G. and Edelheit, E.B. (1981) Biochemistry 20, 79.

Scolnick, E.M. and Caskey, C.T. (1969) Proc. Natl. Acad. Sci. U.S.A. 64, 1235-1241.

Silberklang, M., Gillum, A.M. and Rajbhandary, U.L. (1979) Methods in Enzymology LIX, Nucleic Acids and Protein Synthesis

(part G), pp 58-109.

Shine, J. and Dalgarno, L. (1974) Proc. Natl. Acad. Sci. U.S.A. 71, 1342-1346.

Smrt, J., Kemper, W., Caskey, T. and Nirenberg, M. (1980) J. Biol. Chem. 245, 2753-2757.

Sprinzel, M., Grueter, F., Spelzhaus, A., and Gauss, D.H. (1980) Nucleic Acids Research 8, r-1.

Steege, D.A. and Soll, D.G. (1979) in "Biological Regulation and Development, Gene Expression", Ed., goldberger, R.F., Vol. 1, Plenum Press, New York pp 433-485.

Steitz, J.A. (1979) in "Biological Regulation and Development, Gene Expression", Ed., Goldberger, R.F., Plenum Press, New York pp 349-399.

Steitz, J.A. and Jakes, K. (1975) Proc. Natl. Acad. Sci. U.S.A. 72, 4734-4738.

Stoffler, G., Tate, W.P. and Caskey, T.C. (1982) J. Biol. Chem. 257, 4203-4206.

Stone, M.P., Johnson, D.L. and Borer, P.N. (1981) Biochemistry 20, 3604-3610.

Sundaralingam, M. (1969) Biopolymers 7, 821.

Sundaralingam, M. (1975) in "Structure and Conformation of Nucleic Acid Interactions", Ed., Sundaralingam, M. and Rao, S.T., Baltimore University Park Press, Baltimore, Md., pp 487.

Sundaralingam, M. (1981) in "Biomolecular Stereodynamics", Ed., Sarma, R.H., Vol. 1., Adenine Press, New York pp 301-326.

Suttle, D.P., Haralson, M.A. and Ravel, J.M. (1973) Biochem. Biophys. Res. Commun. 51, 276.

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

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

Tran-Dinh, S., Neuman, J.M., Huynh-Dinh, T., Simmonot, G. and Igolen, J. (1982) C. R. Acad. Sc. Paris, 294, Serie III, 5-8.

Tran-Dinh, S., Neuman, J.M., Huynh-Dinh, T., Igolen, J. and Kan, S.K. (1982) Org. Magn. Reson. 18, 148-152.

Ts'O, P.O.P., Kondo, N.A., Schweizer, M.P. and Hollis, D.P. (1969) Biochemistry 8, 997.

Tundo, P. and Venturello, P. (1979) J.A.C.S. 101, 6066.

Turner, D.H., Petersheim, M., DePrisco-Albergo, D., Dewey, T.G. and Freir, S.M. (1981) in "Biomolecular Stereodynamics", Ed., Sarma, R.H., Vol. 1, Adenine Press, New York pp 429-438.

Urbanke, C. and Maass G. (1978) Nucleic Acids Research 5, 1551-1559.

Van der Meer, J. and Ganoza, M.C. (1975) Eur. J. Biochem. 54, 229-237.

Wang, A.H.-J., van Boom, J.H. (1979) Nature 282, 680.

Weiner, A.M. and Weber, K. (1973) J. Mol. Biol. 80, 837-855.

Werstiuk, E.S. and Neilson, T. (1972) Can. J. Chem. 50, 1283.

Werstiuk, E.S. and Neilson, T. (1973) Can J. Chem. 51, 1889.

Werstiuk, E.S. and Neilson, T. (1976) Can. J. Chem. 54, 2689-2696.

Woese, C. (1970) Nature 226, 817-820.

Wurmbach, P. and Nierhaus, K.H. (1979) Methods in Enzymology 60, 593-603.

Yates, J.L., Gette, W.R., Furth, M.E. and Nomura, M. (1977) Proc. Natl. Acad. Sci. U.S.A. 74, 689-693.

Young, M.A. and Krugh, T.R. (1975) Biochemistry 14, 4841-4847.