Optimal Strategies for the Chemical and Enzymatic Synthesis of Bihelical Deoxyribonucleic Acids

Gary J. Powers,* Russell L. Jones,[1] George A. Randall,[2] Marvin H. Caruthers,[3] J. H. van de Sande,[4] and H. G. Khorana

Contribution from the Department of Chemical Engineering, Carnegie-Mellon University, Schenley Park, Pittsburgh, Pennsylvania 15213, and the Department of Chemistry and Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139.
Received March 6, 1974

Abstract: A synthesis strategy embodied in the computer program, DINASYN, is derived for the production of sequence-defined macromolecules. This strategy is especially useful for DNA with four repetitive units. The DINASYN synthesis program allows for the generation, evaluation, and selection of optimal synthetic plans. The strategy covers both the enzymatic and chemical synthetic steps in preparation of macromolecular DNA. Reductions in synthetic effort of up to 50% can be realized using these strategies.

Sequence-specific polynucleotides can be used to investigate many unsolved biochemical problems. Among these are regulation of gene expression, how proteins interact and recognize specific DNA or RNA sequences, and the nucleic acid sequence of gene regulatory regions. The synthesis of sequence-specific high-molecular-weight DNA is now well established.[5] The methodology involves a combination of chemical and enzymatic techniques. These methods have been extended to the total synthesis of *E. coli* tyrosine suppressor tRNA which is almost complete.[6] The availability of this gene should be very useful for studies on control of gene expression (initiation and termination of transcription) and for studies on the structure-function relationship of tRNA. Already fragments of this gene have been used to obtain the primary sequence of control regions associated with this suppressor tRNA.[7] More recently, the synthesis of *lac* operator DNA has been initiated.[8] This DNA (approximately 31 base pairs) should prove useful for examining how control proteins (*lac* repressor) interact and therefore recognize regulatory DNA (*lac* operator DNA). Some medically useful proteins are difficult to obtain from biological specimens. Furthermore, some are too complicated to prepare in sufficient quantity and in active form by presently available synthesis techniques. Therefore, Nussbaum and coworkers are synthesizing part of the gene (a minigene) for ribonuclease B.[9] This minigene can then be used to test whether proteins could best be prepared by first synthesizing the corresponding DNA gene, transcribing the gene into mRNA, and finally translating the mRNA into protein. Clearly sequence-specific, synthetically prepared DNA is proving useful for problems related to molecular biology and biochemistry.

Present techniques for chemical and enzymatic synthesis of DNA require considerable investment in time and chemicals. For example, synthesis of the yeast alanine tRNA gene required 5 calendar years and approximately 20 man years. In order to realistically use sequence-defined DNA to answer pertinent biological questions, methods must be developed to reduce the complexity of polynucleotide synthesis. One approach is to systematically plan the synthesis so as to optimize yield and reduce the time involved in synthesis. In this paper, we outline a computer program (DINASYN) for determining the optimal reaction path for synthesizing any deoxypolynucleotide. This program was evaluated against the actual time required for synthesizing the yeast alanine tRNA gene. Results indicate that by optimizing yield and minimizing synthesis time with the DINASYN program, an efficient procedure for synthesizing genes or other defined DNA is possible.

*Send correspondence to this author at Carnegie-Mellon University.

DNA Synthesis Problem

The synthesis problem for bihelical deoxyribonucleic acid is shown diagramatically in Figure 1. The target molecule, bihelical DNA, is to be synthesized from 5′-deoxycytidylic acid, 5′-deoxyguanylic acid, 5′-deoxyadenylic acid, and 5′-thymidylic acid. In order to carry out this task several operations must be performed. First short deoxypolynucleotides containing from 4 to 20 monomer units must be chemically synthesized. These short deoxypolynucleotides are then covalently joined with the T4 ligase to form hydrogen-bonded, complementary duplexes. For each chemical or enzymatic step, product must be separated from starting materials and analyzed as to chemical composition.

Given that any reaction pathway to a DNA can be evaluated, a strategy has to be developed for finding the pathway which maximizes the yield of product but minimizes the effort involved. Many possible pathways from starting material to product exist. Table I gives the number of possible reaction paths to form a single strand of DNA as a function of length. As an example, consider the tetranucleotide d-ACGT. Five reaction paths are possible for chemical synthesis. These are illustrated in Figure 2. They are not equivalent. Some pathways require more starting materials and investment of time than others. These factors must be considered as part of the synthesis program.

The generation of reaction paths can proceed utilizing any of three basic search strategies: (1) antithetic (working backward) method; (2) synthetic (working forward) method; or (3) hybird method.

The antithetic method has been proposed as an effective method for generating and searching for optimal reaction paths.[10] In the antithetic method, one begins with the desired molecule. The reactions which could give rise to the desired molecule are considered, and all precursors one reaction step "away" are generated. All precursors which could lead to these molecules are generated in the same manner. This procedure is repeated until starting materials (*i.e.*, the mononucleotides) are encountered.

The main problem with the antithetic method is that if one works backward, generating all possible precursors, the breadth of the precursor tree becomes overwhelming. In addition, it is not possible to accurately evaluate the pathways generated in the antithetic approach until the starting materials are encountered. This is due to the fact that evaluation of a pathway depends upon how much time and start-

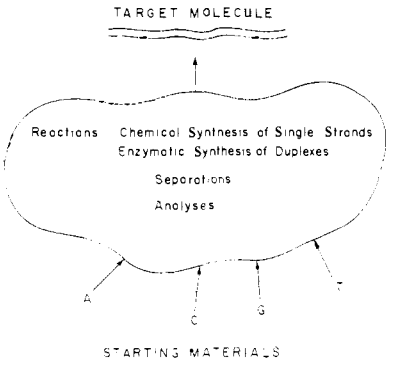


Figure 1. DNA synthesis problem.

Table I. Total Number of Reaction Paths for Forming a Nucleotide Sequence of a Given Length

Length	No. of paths	Length	No. of paths
2	1	11	16,796
3	2	20	10^{10}
4	5	50	10^{27}
5	14	100	10^{60}
6	42	150	10^{84}
7	132	200	10^{113}
8	429	500	10^{286}
9	1,430	1000	10^{575}
10	4,862		

ing material must be utilized for synthesis of each precursor. These values are not known until the beginning of the path is generated. Hence to find the optimal reaction path, it is necessary to evaluate all the reaction paths. The numbers of pathways given in Table I indicate that exhaustive search is clearly not applicable to the synthesis of even small DNA molecules (length less than 100 nucleotides). Heuristics are commonly invoked to reduce the breadth of the synthesis tree. Heuristics are rules of thumb which have, in the past, proved to be useful means for reducing the scope of the search. The rules do not guarantee optimal solutions in a mathematical sense. Several general rules based on simplicity of intermediates have been advanced to reduce the size of the search.[10]

The synthetic method[11,12] solves the synthesis problem by working forward from the starting materials to the target molecule. In other words, the starting materials are transformed, by repeated application of the operators, into the target molecule. Thus planning is done as the synthesis is executed. This approach has one great advantage: the search for the optimum path can be guided by the principle of optimality. The principle of optimality[13] is based on the intuitively obvious principle that an optimal set of decisions has the property that whatever the first decision is, the remaining decisions must be optimal with respect to the outcome which results from the first decision.

In DNA synthesis, the key is the fact that the best way to produce any DNA molecule is to select the optimum way of carrying out the immediately preceding reaction on the best of all possible precursors. For example, if time is used to evaluate reaction pathways, the minimum time to produce any molecule is made up of the minimum times to join two or more groups which have also been made in the minimum time. Hence, the optimum way to make any given molecule depends on the optimum way to produce its precursors. The only place to start this process is at the starting materials. This search technique, also called dynamic programming, has been used in a wide range of decision problems in inventory control, material allocation, process control, and chemical process design. The synthetic approach has considerable computational advantage when compared with the antithetic generation with heuristic evaluation. The principle of optimality greatly reduces the size of the search while still maintaining mathematical guarantees of optimality. The actual size of the reduction is discussed in a following section.

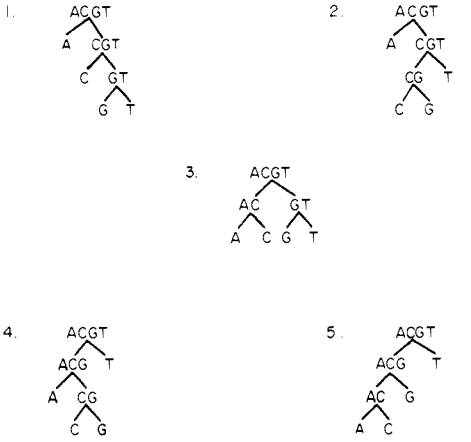


Figure 2. Five reaction paths for the synthesis of a tetranucleotide.

It is possible to combine the antithetic and synthetic methods into various hybrid techniques. Corey[10] has suggested a method in which the synthesis planning works backward from the target and forward from the starting materials, meeting in the middle of the reaction paths. Alternatively, a "critical" intermediate precursor could be identified which decomposes the overall synthesis problem into two subproblems: one problem is to get from starting materials to the critical intermediate, and the second is to generate reaction paths from the intermediate to the desired molecule. These subproblems could be solved by either the antithetic or synthetic method or be decomposed by identifying additional critical intermediates.

Evaluation of Chemical and Enzymatic Procedures

Outline of the Problem. In order to determine what constitutes a desirable reaction path, chemical and enzymatic procedures must be evaluated. The evaluation of reaction paths introduces several problems. First, what are the criteria for evaluation? Second, how can these criteria be combined to yield a single evaluation criterion? Third, can we predict, in general, the values for these criteria?

When carrying out a chemical or enzymatic reaction, three areas of activity can be identified: (1) reaction, (2) separation, and (3) analysis. Within each of these areas, evaluation criteria must be identified. For the reaction, the yield of product, unit cost of raw materials, and time required for a successful reaction must be evaluated. The separation is evaluated on the basis of the yield of separation, the purity of products, and the time required for a successful separation. An evaluation of the analysis is based on the time required for a successful analysis. Such success is directly related to each reaction path and how many tests must be performed to obtain an unambiguous analysis.

Therefore in the laboratory synthesis of DNA's, the most important feature is the time required to achieve a successful synthesis. In most cases, the cost of raw materials is small compared with the cost of time invested in the synthesis. In addition, it is possible to determine time values for each of the other criterion. For example, if low yields are encountered for some reaction paths, the same reaction can be run a second time or only once with more starting material. Both alternatives require more time. Hence the time value of yield can be determined. If a difficult separation is

encountered, then more time will be required to carry out the separation. The additional time may be consumed in operating the separation procedure more slowly or by rerunning the partially purified mixture through a second separation. The analysis may require several tests to determine the exact structure of an ambiguous compound. Hence a greater amount of time is needed. For operations which are chemically impossible, an infinite amount (or at least a very high amount) of time can be assigned. Therefore, a uniform means of evaluating a DNA synthetic pathway can be determined, if the time required to achieve the desired synthesis is considered. Thus, the problem is to evaluate chemical and enzymatic reactions in time units.

Evaluation of Chemical Procedures. The general plan for synthesizing duplex DNA of defined sequence is shown in Figure 3. There are two major sets of reactions for transforming the four mononucleotides A, C, G, and T into a finished DNA duplex. These are chemical syntheses of deoxyoligonucleotides with defined sequence and enzymatic joining of the deoxyoligonucleotides.

The chemical synthesis of defined deoxyoligonucleotides can be divided into several steps.[11] (1) Protection of the amino functions in the heterocyclic base of a nucleoside or nucleotide. Cytosine is protected with the anisoyl group (An), adenosine with the benzoyl group (Bz), and guanine with either the isobutyroyl (iB) or methylbutyroyl (mB) group. These groups remain throughout and are removed only upon completion of the chemical synthesis. (2) Protection of the 5′-hydroxyl of nucleosides with a bulky, acid-sensitive methoxytrityl group (MMTr). This group is removed only after the chemical synthesis of the deoxyoligonucleotide is complete. (3) Protection of nucleotide 5′-phosphates with the cyanoethyl group (CE). This group is useful in synthesis of very short deoxyoligonucleotides (two to four units). It is removed after every synthetic step. (4) Protection of the 3′-hydroxyl on nucleotide 5′-phosphates and oligonucleotides with an acetyl group. This group is also removed after every synthesis step. (5) Repeated condensation of appropriately protected mononucleotides or oligonucleotides using activating agents such as dicyclohexylcarbodiimide or various aromatic sulfonyl chlorides.

Two types of condensation reactions are generally used. Equation 1 illustrates the preparation of a dinucleotide containing a 5′-phosphate. Starting materials are the 5′-cyanoethyl phosphoryl- and 3′-*O*-acetyl-protected mononucleotides. After condensation, saponification to remove the cyanoethyl- and acetyl-protecting groups, purification, and acetylation, the dinucleotide is ready for use in the second type of condensation reaction. Equation 2 illustrates this reaction. The dinucleotide d-MMTrC^{An}pA^{Bz}-OH containing the acid-labile, bulky monomethoxytrityl group on the 5′-hydroxyl is condensed with the dinucleotide d-pC^{An}-pA^{Bz}-OAc to form the tetranucleotide. After removal of the acetyl group with alkali, the tetranucleotide is ready for further elongation of the sequence.

$$\text{d-CEpC}^{An} + \text{d-pA}^{Bz}\text{-OAc} \xrightarrow[\text{2. OH}^-;\ \text{3. column chromatography};\ \text{4. acetic anhydride}]{\text{1. DCC}} \text{d-pC}^{An}\text{pA}^{Bz}\text{-OAc} \quad (1)$$

$$\text{d-MMTrC}^{An}\text{pA}^{Bz}\text{-OH} + \text{d-pC}^{An}\text{pA}^{Bz}\text{-OAc} \xrightarrow[\text{2. OH}^-;\ \text{3. column chromatography}]{\text{1. DCC}} \text{d-MMTrC}^{An}\text{pA}^{Bz}\text{pC}^{An}\text{pA}^{Bz} \quad (2)$$

Two general condensation reactions are therefore possible depending on whether a trityl or cyanoethyl group is used to protect the 5′ end of the growing nucleotide chain. Table II gives the attention times required to carry out condensations for each group. Also included are attention times for protection of mononucleotides and oligonucleotides. The condensations which utilize trityl-protecting groups involve only protection of the 3′-OH and observation of the condensation reaction. Cyanoethyl-protected condensations require that the cyanoethyl group be added prior to each condensation since it is removed during post-condensation treatment.

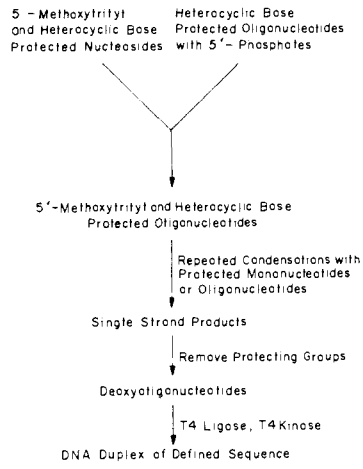


Figure 3. Overall synthesis strategy.

The yield of each reaction is important in determining the time required to carry out the reaction. If the yield of a given reaction is low, it is necessary to carry out the reaction with larger amounts of raw material. This may or may not require more attention time depending on the initial scale of the reaction. Data were gathered on the effect of scale of reaction on the attention time required for mononucleotide protection, single-strand condensation reactions, and duplex joining. The data are adequately described by the simple power law model given in eq 3, where T is the attention time in hours required to react a mixture containing Q grams of reactants. T_0 is the time in hours to process Q_0 grams of reactant. For example, if the size of a reaction step is doubled, the average attention time is $(2/1)^{0.3}$ = 1.23 times greater. The increase in time is due to the fact that in some cases, it is more convenient to prepare several small batches of material rather than scaling the reaction to larger columns, etc.

$$T = T_0(Q/Q_0)^{0.3} \quad (3)$$

The scale of reactions in a reaction path depends on the desired amount of material and the yield of each reaction in the path. Hence, for the evaluation of general reaction paths, it is necessary to be able to predict the yield of any reaction which might occur in the paths. For general classes of reactions, this is a well-known and yet unsolved problem. Several approaches have been advocated for predicting the yields of reactions.[12] If the reaction proceeds to equilibrium, and all the species in the system are known, it is possible to compute the equilibrium composition and yield from the free energies of the individual components. If selectivity in the reaction is due to different kinetics for competing reactions, an approach based on linear free-energy relations may be used. This approach has proved useful for several families of compounds.[14a]

For the DNA synthesis reactions, we have taken an empirical approach based on fitting simple functions to yield data presently available. These functions are then used to predict the yield of reactions which have not yet been performed. Data are available for 212 single-strand DNA

Table II. Time Requirements[a] for the Protection of Mononucleotides and the Condensation of Trityl- and Cyanoethyl-Protected Oligonucleotides

Preparation of Mononucleotides (Protection)	
A	20
C	20
G	20
T	1
Trityl A	40
Trityl C	40
Trityl G	40
Trityl T	20
Condensation Reactions (Single Strands)	
Protection of 3′-OH	12
Reaction observation	4
Total for trityl protected reaction	16
Cyanoethyl protection	20
Protection of 3′-OH	12
Reaction observation	4
Total for cyanoethyl protected reaction	36

[a] In hours.

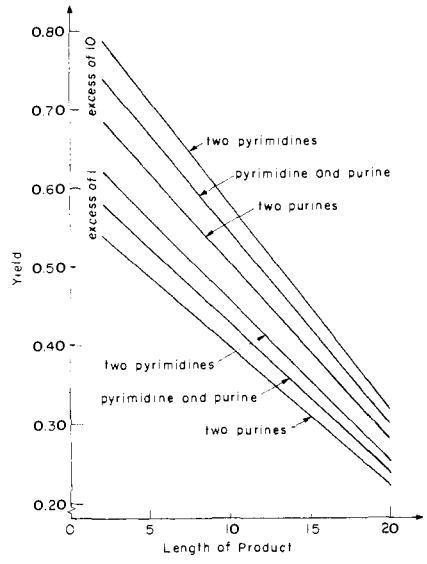


Figure 4. Overall molar yield for the condensation reactions used in the chemical synthesis of oligonucleotides.

reactions: 39 dinucleotide formation reactions and 173 reactions which produce oligonucleotides.[14b] In each of the following models, the yield (Y) is defined as the molar quantity of purified, recovered product divided by the molar quantity of the reactant with the free 3′-hydroxyl groups, expressed as a fraction. This definition is used regardless of which reactant is present in excess. The reactants are taken to be fully protected, and the product has a free 3′-hydroxyl group and a free 5′-phosphomonoester or 5′-trityl protection, depending on the initial condition of that nucleoside. Hence the reported yield is for a complete reaction *and separation* cycle in the reaction path. It was not possible to determine separately the yield loss due to separation from the published data.

Excess (X) is defined as the ratio of the molar quantity of the component with the free 5′-phosphomonoester to the molar quantity of the component with the free 3′-hydroxyl group. Thus, X can be less then one, although it very seldom is. Length (L) refers to the number of nucleoside residues in the product. Type refers to the bases which are attached to the 3′ and 5′ ends which are being joined. The yield model is given in eq 4 and plotted in Figure 4. The amount of condensing agent used had no significant effect on the yield for the data studied. The standard error (square root of variance) is 6 percentage points of yield.[12]

$$Y = [0.80 - 0.025(L_p)](\alpha)[1 - 0.25\exp(-0.3X)] \quad (4)$$

$$\alpha = \begin{cases} 1.07 \text{ if two pyrimidines} \\ 1.00 \text{ if pyrimidine and purine} \\ 0.93 \text{ if two purines} \end{cases}$$

Separation Evaluation for Chemical Reactions. Following reaction work-up, the reaction mixture is separated to recover purified products and, when possible, purified unreacted starting materials. The separation commonly is performed by ion-exchange chromatography in a buffered water-ethanol solution with a salt gradient used to control elution of the species in the mixture.[15] Other separation procedures such as liquid-liquid extraction with trityl-containing compounds[16] and gel filtration[15r] have also been used. The synthesis of di- and trinucleotide blocks using liquid-liquid extraction has been attempted.[17]

Model of Ion-Exchange Chromatography. The attention time required to separate a reaction mixture depends on the difference in charge between the product, by-products, and reactants which make up the reaction mixture. To a smaller extent, the separation depends on the nucleotides which make up each chain. If the charge difference is large, the separation is relatively easy. For this case, only the basic attention time associated with preparing the column, loading the mixture, monitoring the separation, and collecting the samples is required. When the charge difference is small, a greater amount of attention time is required. The column must be run more slowly, and multiple passes through the ion-exchange column may be required. The model described below is based on attention-time data collected for over 30 separations performed on single-strand condensation reaction mixtures. The major by-product is the pyrophosphate of the mononucleotide or oligonucleotide block containing the 5′-phosphomonoester. The charge on the pyrophosphate can sometimes be as great or greater than the charge on the product. Therefore, separation techniques employing gradients of salt and alcohol have been employed to achieve the desired purification of product.[15o-v]

Table III summarizes the model. The correction factor for the base content of each single chain is used to compute the chain's equivalent length. The correction is due to the fact that the bases bind differently to the packing in the ion-exchange column. Hence, they do not move through the column at a rate that is directly proportional to their charge. For small products of length four nucleotides or less, the separation attention time is 16 hr. If the condensation involves trityl-protected products, the attention time is given in Table III. The attention time depends on the relative size of the product, L_p, the reactants, L_3 and L_5, and the by-product pyrophosphate. Attention time T_1 represents the attention time required to separate the product from the free 5′-phosphomonoester reagent. The time T_2 is for the separation of the product from the pyrophosphate by-product.

If the reaction involves a cyanoethyl-protected oligonucleotide, the separation of the reaction mixture will differ from the trityl-protected reaction mixture because of the presence of one additional charge on the cyanoethyl-protected strand. (The cyanoethyl group is removed from the phosphomonoester prior to separation.) Table III gives the form of the model. Pyrophosphates are formed in the reaction and have a charge of $2C_5 - 2$ when C_5 is the charge of the oligonucleotide carrying the free 5′-phosphomonoester.

Table III. Time Requirements for the Separation of Condensation Reaction Mixtures

Length Correction Factors
A = 0.00, C = 0.15
G = 0.35, T = 1.00

Effective Length
$L' = L$ + (sum of correction factors)/L

Effective Charge
Trityl products: $C_p = L'_p$, $C_3 = L'_3$, $C_5 = L'_5$, $C_{pp} = 2C_5 - 2$
Cyanoethyl products: $C_p = L'_p + 1$, $C_3 = L'_3 + 1$, $C_5 = L'_5 + 1$, $C_{pp} = 2C_5 - 2$

Separation Times
Time = 16 hr for all separations where the product has a length less than or equal to 4

Trityl products:
$T_1 = (C_p + C_5)/(C_p - C_5)$
$T_2 = 0$ if $C_p/(2C_5 - 2) \geq 1$
$= 20$ if $C_p/(2C_5 - 2) \leq {}^2/_3$
$= 60\{1 - [C_p/(2C_5 - 2)]\}$ if ${}^2/_3 \leq [C_p/(2C_5 - 2)] \leq 1$
Total time = $12 + T_1 + T_2$

Cyanoethyl products:
D_1 = largest of C_p, C_3, C_5, and C_{pp}
D_2 = second largest of C_p, C_3, C_5, and C_{pp}
D_3 = third largest of C_p, C_3, C_5, and C_{pp}
D_4 = lowest of C_p, C_3, C_5, and C_{pp}

$T_1 = (D_1 + D_2)/(D_1 - D_2)$
$T_2 = (D_2 + D_3)/(D_2 - D_3)$
$T_3 = (D_3 + D_4)/(D_3 - D_4)$

Total time = 10 + highest of T_1, T_2, and T_3 + one-half the sum of the other two

The attention times T_1, T_2, and T_3 represent the three separations between the four main species in the reaction mixture.

Analysis Models for Chemical Reactions. The analysis of reaction mixtures fractionated on DEAE cellulose is by ultraviolet spectroscopy. The ratio of absorbance at different wavelengths is the desired analytical result. For most syntheses where the base composition of product is not the same as starting material, the absorbance ratios will be different. For all syntheses, isolated products must be further analyzed to ensure the correct base composition. Methods include paper chromatography in several solvent systems after complete removal of protecting groups and after removal of only base-labile protecting groups. Completely deprotected samples are also treated with enzymes snake venom phosphodisterase, spleen phosphodiesterase, and alkaline phosphatase, and the base composition measured. The average analysis attention time is 16 hr for cyanoethyl products and 20 hr for trityl products. A small correction for an unequal distribution of A, T, G, and C is applied. If A, T, G, and C are unequally distributed, more time is required to ensure that the desired base content has been incorporated in the product. Table IV illustrates the attention-time model for single-strand analysis.

Time Value of Oligonucleotides. The value of a reaction product is equal to the time necessary to prepare the reagents required plus the time necessary to carry out the reaction, separation, and analysis steps which yield the product. With this definition and the models described above, it is possible to determine the time value of any given oligonucleotide, if the reaction path leading to the product is known. If the yield of a given reaction is low, the time required to carry out the reaction, separation, and analysis steps to produce the desired amount of product will be increased as discussed above. A power law model has been fitted to the attention times required to carry out the reaction,

Table IV. Time Requirements for the Analysis of Single-Strand Reaction Products

Cyanoethyl Products
Time = 16 hr

Trityl Products
Ultraviolet peak factors
$H_A = 0.64$(number of A's)
$H_C = 1.60$(number of C's)
$H_G = 0.50$(number of G's)
$H_T = 1.00$(number of T's)
H_{max} = highest of H_A, H_C, H_G, and H_T
$U' = H_{max}/H_A + H_{max}/H_C + H_{max}/H_G + H_{max}/H_T -$ (number of H's > 0)
$U = U'$ if $U' < 5$
$U = 5$ if $U' \geq 5$
Time = $16 + 4(0.90 + 0.02U)$

separation, and analysis steps for a single reaction. As with the reaction step alone, a power law model with an exponent of 0.3 describes the data. The model is given by eq 5 where

$$T_{R,S,A} = T_0(Q/Q_0)^{0.3} \quad (5)$$

$T_{R,S,A}$ is the time required to carry out a single sequence of reaction, separation, and analysis. Q is the quantity in gram moles of the reactants, and T_0 and Q_0 are the basis time and amounts, respectively.

When the yield of a reaction is low, it is necessary to increase the attention time for both that reaction, separation, and analysis sequence as well as for the time required to produce the reactants. The expression for the time value of a product is given in eq 6 where TVP is the time value of

$$\mathrm{TVP} = T_{R,S,A} + \left(\frac{\mathrm{TV5'}}{\mathrm{yield}}\right)^{0.3} + \left(\frac{\mathrm{TV3'(excess)}}{\mathrm{yield}}\right)^{0.3} \quad (6)$$

the product; $T_{R,S,A}$ is the attention time for the reaction, separation, and analysis; TV5′ is the time value of the reagent with the free 5′-phophomonoester; TV3′ is the time value of the reagent with the free 3′-hydroxyl; excess is the molar excess of the reagent with the free 3′-hydroxyl; and yield is the molar yield of the reaction and separation steps (computed from eq 4).

Evaluation of the Enzymatic Joining of Chemically Synthesized Deoxyoligonucleotides. Duplex-joining reactions utilize the tendency of complementary strands of nucleotides to base-pair by hydrogen bonding. Suppose that two deprotected single strands of nucleotides are allowed to base pair on an appropriate template so that their 3′ and 5′ ends are adjacent. It would then be possible to use the enzyme T4 ligase to join the two single strands.[18] By forming single strands which overlap each other, it is possible to have the complementary strand in the uncompleted duplex serve as the template. Equations 7, 8, and 9 illustrate several possible duplex joining reactions.

The amount of experience with the ligase joining reactions is not as great as with the single-strand condensations. A review of the previous work with these reactions indicates that approximately 500 attention hr are required to successfully carry out such a reaction.[18] The steps involved are: (1) deblocking the single strands; (2) carrying out extensive model studies on small amounts of material to determine reaction behavior; and (3) final reaction.

Duplex Evaluation Models. The same areas, reaction, separation, and analysis, that are important in single-strand synthesis are also important in duplex joining reactions. Insufficient data are available for an analysis of ligase-catalyzed joining reactions. Therefore, a simple model is used. The reaction model assumes: (1) all deoxyoligonucleotides

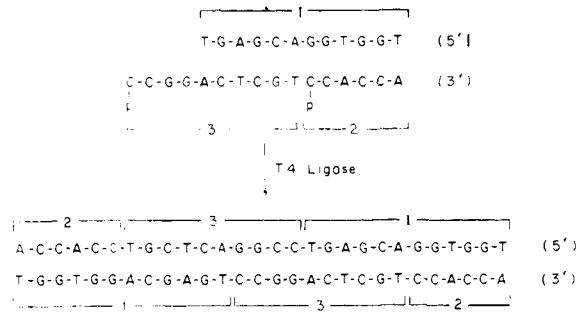
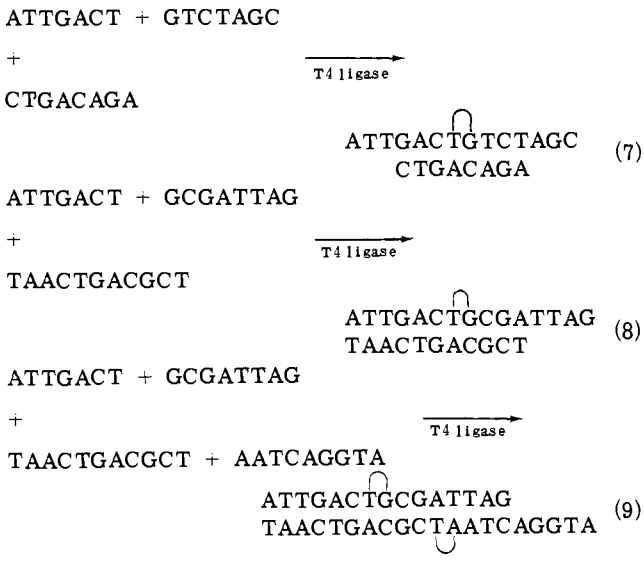


Figure 5. Interfering complementary groups in the formulation of DNA duplexes.

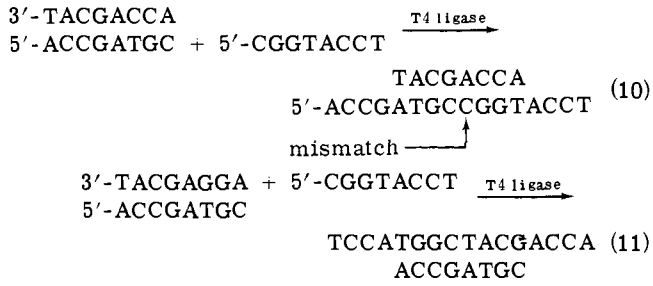


are present in equimolar quantities; (2) the yield is 45%; (3) the reaction attention time (which includes a model study of the reaction) is fixed at 500 hr; and (4) a detailed search for interfering complementary groups within sequences is carried out.

As in the single-strand evaluations, the time value of a product is based on the deoxyoligonucleotide time values plus the reaction, separation, and analysis time. A 0.3 power law model is also used.

The search for potentially complementary groups within the deoxyoligonucleotides used to form the product is an important feature of duplex-reaction evaluation. The basis of the duplex reaction is the formation of hydrogen bonds between overlapping segments of the reagents. The hydrogen bonding is brought about by heating the deoxyoligonucleotides in solution and then slowly cooling the mixture to allow the overlapping ends to achieve the most thermodynamically stable pairing. The goal is to have the hydrogen bonding yield the nucleotide sequence in the target molecule. Unfortunately, it is possible that hydrogen bonds can form between nucleotide sequences which are not the desired sequence. Figure 5 illustrates an example from synthesis of the yeast alanine tRNA gene.[19] The product expected from the T4 ligase reaction was the duplex of chemically synthesized segments 1, 2, and3. Segment 2 was expected to be joined to segment 3 by the T4 ligase. Instead, the dimer shown was formed quantitatively. Therefore, complementary single-stranded ends must be excluded from any synthetic reaction sequence.

The hydrogen bonding between oligonucleotide strands need not be exact (A with T, G with C) for the joining to occur. For example, T4 ligase catalyzed joining reactions have been observed where A is opposite C,[19,20] and C is opposite T^{21} in the DNA duplex. A hypothetical example is eq 10. In this example, an A to C base pair is tolerated, and



the wrong ligase joining reaction occurs. The desired reaction is shown in eq 11. In order to evaluate whether two sequences are similar enough to cause formation of the wrong product in a ligase mixture, the bond strengths in Table V were assigned. The duplex-reaction evaluation model is summarized in Table VI. If the hydrogen bonding is not exact in a potentially wrong joining reaction, and the average hydrogen bond strength for the overlapping segments is 7 or greater, an additional 500 attention hr are required. The 500 hr are required to develop reaction conditions which favor the desired hydrogen bonding. If the average bond strength is less than 7, it is assumed that the reaction will not proceed to undesirable product. If the undesirable matches cannot be prevented (an arbitrary average bond strength greater than 8), the program is constructed so as to reject these potential sequences and to resynthesize the sequences with break points (for enzymatic joining) at different sites. The break point for sequences shown in eq 10 would therefore be rejected (average bond strength is 8). The current model assigns 10,000 attention hr to such rejected sequences. Therefore, the program selects against the presence of these groups in troublesome sequences.

Separation of the reaction products from starting materials and by-products is performed on Biogel-A (0.5 m).[18] The duplex separation model for attention time is based on the difference in size that exists between the desired product and the by-product which is closest in size to the product. The by-product which is closest in size to the product is the duplex in which the smallest end deoxyoligonucleotide (oligonucleotide on the end of the duplex product) is missing. (This assumes no complementary group problems.) If an oligonucleotide is missing from the "middle of the du-

Table V. Relative Bond Strengths between Base Pairs (on a 1–10 scale used in calculating complementary behavior)

A-A	1	C-G	10
A-C	5	C-T	1
A-G	1	G-G	1
A-T	7	G-T	7
C-C	1	T-T	5

Table VI. Time Requirements for Duplex-Forming Reactions

Reaction Attention Time (TRA)
TRA = 500 hr (includes model study)

Reagent Cost (TRC)

$$\text{TRC} = \left(\frac{\text{reagent cost}}{\text{yield}}\right)^{0.3} \quad (\text{yield} = 45\%)$$

Additional Study Required due to Complementary Problems (TCOMP)
10,000 if all base pairs are A–T and C–G
500 if some mismatches but average bond strength ≥ 7
0 if average bond strength < 7

Time = TRA + TRC + TCOMP

Table VII. End Group Labels for Unambiguous Duplex Analysis

Making Certain All Strands Are in Product $\frac{A\cdots B}{C\cdots D}$

(1) End groups
  (a) 3′ ends of B and C must not be the same as each other or any internal strand 3′ end
  (b) 5′ ends of A and D must not be the same as each other or any internal strand 5′ end (different labels may be used to distinguish identical nucleotides)
(2) Internal groups: no two different subgroups of reagents may have the same number and type of both 3′ ends and labeled 5′ ends. The following are violations of this rule:
  G^+—T and G^+—T
  G^-—T, C*—A and G^-—A, C*—T

Making Certain All Strands Are in the Right Location
(1) After 3′ degradation, no 3′ labeled group may be the same
(2) The 5′ ends of strands with duplicate 3′ nucleotides must have different labels
(3) If only two labels, P^{32} and P^{33}, are used, and only internal phosphates are present, the P^{33} must be placed as follows: (strand position)

T1 T2 T3 T4 ···
B1 B2 B3 B4 ···

No. of strands	Molecule configuration	P^{33} label placement
3	Any	None required, optional in all locations
4	T1,T2,T3,B1	None required, optional in all locations
4	T1,T2,B1,B2	T1 or B2
5	T1,T2,T3,B1,B2	T1 or B2
6	T1,T2,T3,T4,B1,B2	T2 and B2 or T1 and T3
6	T1,T2,T3,B1,B2,B3	T1 and T2; T1 and B2; T2 and B3; or B2 and B3
7	T1,T2,B1,B2,B3,B4,B5	T1 and B4 or B2 and B5
7	T1,T2,T3,B1,B2,B3,B4	T1,T2 and B3 or T2,B3 and B4
8	T1,T2,T3,B1,B2,B3,B4,B5	T1,T2 and B3; T1,B2 and B4; T2,B3 and B5; B2,B3 and B5; or B2,B4 and B5
8	T1,T2,T3,T4,B1,B2,B3,B4	T1,T2 and B2; T1,T3 and B3; T2,B2 and B4; or T3,B3 and B4

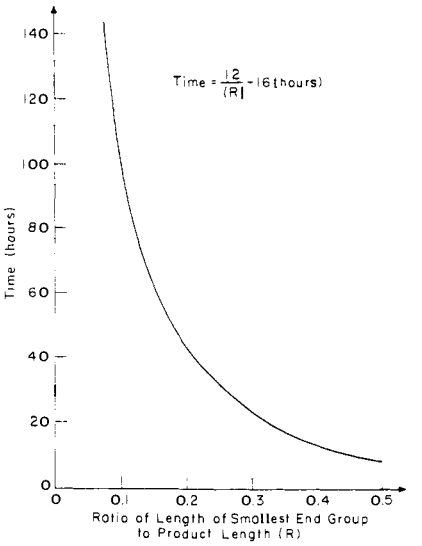


Figure 6. Time requirements for the separation of duplex-forming reaction mixtures.

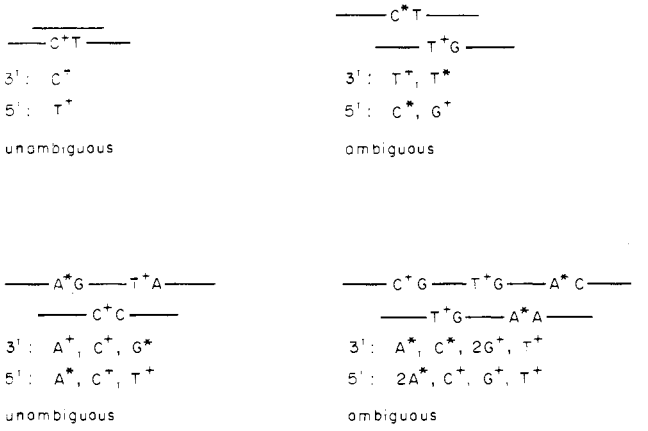
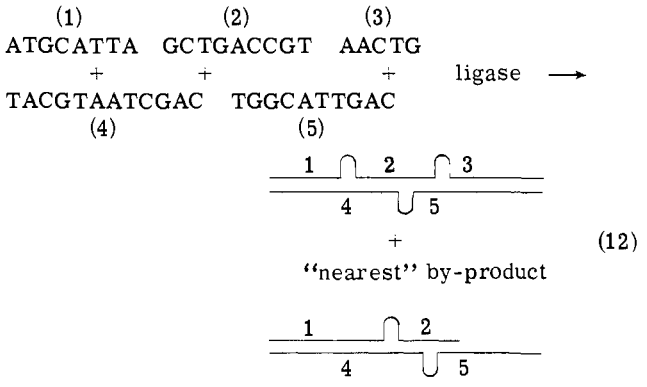


Figure 7. Examples of end group labels for the analysis of duplex products. A P^{32} label is denoted by + and a P^{33} label by *.

plex," the duplex will not be joined, and two by-products much smaller than the product will be formed. Equation 12 illustrates this situation. The duplex separation model is summarized in Figure 6.



(12)

Duplex Analysis. Three different classes of information are used to ensure that the correct duplex has been formed during reaction. (1) The fractions which result from separation confirm that the appropriate size has been obtained for the product. (2) Resistance to phosphatase indicates that the free phosphate ends have reacted. (3) Nearest neighbor and 5′-mononucleotide analyses of radioactive phosphate are used to determine that all the reagent strands have been incorporated in the product, and that the strands are in the appropriate positions relative to each other. The attention time for analysis by separation is taken as a simple fraction (½) of the separation attention time. The resistance to phosphatase requires an average of 50 attention hr.

Analysis for location of radioactive phosphate can be quite complicated. The common approach is to label the 5′-phosphate end of certain of the single strands with radioactive phosphorus (either P^{32} or P^{33}). The reaction product is then degraded in two different ways. Degradation to 3′-phosphate nucleotides is possible using micrococcal and spleen phosphodiesterase.[19] The use of pancreatic deoxyribonuclease and venom phosphodiesterase gives hydrolysis to 5′-phosphate nucleotides. Hence, the radioactive phosphate can be made to associate with the 3′- or 5′-nucleotides adjacent to the ligase joining point. Several examples of unambiguous and ambiguous end group analyses are given in Figure 7. The problem of where labels must be introduced to give unambiguous joining information was solved for combinations of single strands to form duplexes. The solution is summarized in Table VII. Several other options exist which could resolve ambiguous analyses. The bihelical strands may be separated into two single strands and each strand analyzed individually. It is also possible to use different activity levels of radioactive P^{33} or P^{32} to uniquely identify the end groups.[18]

Because of the complexity of the end group analysis

Table VIII. Time Requirements for Duplex Analysis

Separation Analysis (T)
Half the separation score
Resistance to Phosphatase (TRP)
50 hr
End Group Analysis (TEGA)
Three reagents: 50 hr
Four or more reagents
All 3′ ends different: 100 hr
3′ ends appearing a maximum of two times and all 5′ ends different: 150 hr
All others: 200 hr
T = TA + TRP + TEGA

problem, the simple evaluation function given in Table VIII was used.

In summary, models have been developed which predict the attention time required to carry out any given sequence of reaction, separation, and analysis steps. The time requirements depend on the type of reaction used (single-strand condensation or enzymatic duplex joining using T4 ligase), the yield of the reaction, the difference in charge or size between the product, by-products, and reactants, and the ease of analysis of the product.

Applying the Synthetic Method to DNA Synthesis

A computer program DINASYN (*D*eoxyr*I*bo*N*ucleic *A*cid *SYN*thesizer) was written by Jones[22] for determining the optimal reaction path to any general DNA molecule. The program is based on a dynamic programming approach to the search for optimal synthesis paths.

Working the synthesis problem forward (dynamic programming approach) involves two major steps. The first is to generate all possible subgroups in the target molecule from length one to the final molecule size. Then one must evaluate the cost of making each subgroup, starting with the cost of the raw materials (the four protected mononucleotides).

Initially, all the dinucleotides are considered. Optimum reaction conditions, separation time, and analysis time are determined for each of these groups. Next, the times for the trinucleotide groups are calculated, optimizing over the joining location (which two subgroups are joined to form the target group). The joining point which gives the lowest time is selected as optimum, and this time is used in the next level of optimization. The evaluation of the subgroups continues until the evaluation of the target molecule (the last subgroup) is completed.

It should be noted that the optimization carried out for all groups of three or more nucleotides is an optimization over the total cost of the reaction sequence. This is possible since the subgroups used in these reactions have already had their optimum synthesis path determined. These minimum subgroup reaction path times are added to the time of the last reaction to form the total time, which must be minimized to give the minimum cost joining point and optimal reaction path for each group. This step-by-step build-up of optimal reaction subpaths is continued until eventually the optimal synthesis path is generated.

An example of the subgroups and the reaction evaluations necessary for each subgroup optimization in the sequence TCAGGA is shown in Table IX. Note that only 34 evaluations are required compared with 210 (42 paths times 5 reaction evaluations per path) for an exhaustive search procedure. Computational savings greatly increase as the length of the target strand increases.

Table IX. An Example of the Dynamic Programming Search Procedure for Selection of the Optimal Reaction Path to the Hexanucleotide TACGAC[a]

TACGAC			
T + ACGAC			
TA + CGAC			
TAC + GAC			
TACG + AC			
TACGA + C			
TACGA	ACGAC		
T + ACGA	A + CGAC		
TA + CGA	AC + GAC		
TAC + GA	ACG + AC		
TACG + A	ACGA + C		
TACG	ACGA	CGAC	
T + ACG	A + CGA	C + GAC	
TA + CG	AC + GA	CG + AC	
TAC + G	ACG + A	CGA + C	
TAC	ACG	CGA	GAC
T + AC	A + CG	C + GA	G + AC
TA + C	AC + G	CG + A	GA + C
TA	AC	CG	GA
T + A	A + C	C + G	G + A
T	A	C	G
START	START	START	START

[a] 34 reaction evaluations are necessary.

Multiple Groups and Information Flow Recycle

In a large DNA molecule, one would suspect that several segments within the molecule would be identical. This is indeed the case. In fact, more multiple groups have been found in 36 bihelical DNA's than would be expected if the nucleotide sequences were random.[22] The presence of a multiple group can have a large effect on time required for the synthesis of a DNA. If a large oligonucleotide is found to appear twice in the target molecule, it may be possible to devise a synthetic plan which uses this group twice. If this is the case, it would only be necessary to make a larger batch of the material and use it twice in the synthesis. Since the time required for larger batches follows a 0.3 power law, considerable savings can be realized if multiple appearing groups can be utilized in the reaction path. A computer program has been written to find all the multiple groups in any given bihelical DNA. The synthesis strategy is then modified to indicate that potential savings could occur if these groups occur in the reaction path. While the multiple groups offer considerable potential for savings, it is necessary to modify the dynamic programming search procedure to incorporate them.

The basic assumption on which the principle of optimality rests is that the information flow in the decision making process is strictly serial. This means that decisions made at a given stage must not depend on any information other than that which directly preceeds it. For DNA synthesis, as previously discussed, this assumption was true. However, the inclusion of multiple group considerations changes the information flow to nonserial. The problem is that when one starts working through a synthesis path, attention times must be assigned. These times are used to decide which path to select at each optimization level. If a group has the possibility of being used twice, should it be assumed that it will be used twice and assigned a lower value? The trouble is that when the multiple group is first encountered, it is not known whether it will be used again later in the synthesis path. Therefore, a decision must be made on information not known, and later decisions depend on the immediate decision. This results in a feedback or recycle of information which nullifies the principle of optimality. A bounding pro-

cedure is currently in use to allow the consideration of multiple groups while still maintaining the dynamic programming approach. The heuristic is invoked that the largest multiple groups that do not interfere with each other should be used. These groups are assigned a lower time value depending on how often they appear in the desired molecule. That is, if a group appears three times, its time value is approximately one-third of the time value if it was used only once. The optimal path is then selected using the dynamic programming approach. If the multiples are used in the path selected by the dynamic programming approach, no changes are necessary. The path is not necessarily optimal however.[22] If a path is selected which used the multiple appearing group only once, a dilemma exists. Is the correct path one which uses all the multiple appearing groups in the synthesis path or one which gives no multiple credit for use of the subgroup? If the first situation is true, the synthesis path must be changed, and if the second is true, increasing the cost of the components used may make another path more attractive. A bounding technique, in which a lower bound on multiple appearing group times is repeatedly used, has also been developed, but because of its complexity and relatively small effect on the paths, it was not used.[22]

For the synthesis of single strands of oligonucleotides, the dynamic programming approach is rapid and very useful. However, the synthesis of complete duplexes from single strands and combinations of single strands and duplexes is too large a problem to attack directly with the dynamic programming approach. Because of the size of the problem, a decomposition approach is used.

Decomposition at the Duplex Synthesis Level

The overall synthesis problem is separated into two parts. The first part involves identifying a set of single strands in the target molecule duplex which is synthesized in the optimal manner using the dynamic programming approach. These single strands are then used as starting materials, and the dynamic programming approach is used to determine the optimal sequence for joining the single strands into the desired duplex.

The identification of the critical set of single strands is currently done *via* interaction between the synthetic chemist and the DINASYN program. The sets of multiple groups which occur in the desired DNA are displayed for the synthetic chemist's information. He then selects "cut points" in the desired duplex which preserve the larger number of long multiple groups. The program then performs a preliminary check of the feasibility of each cut point. The potential for self-complementary groups, the length of single strand, and the overlap between chains in alternate strands are all determined. If an infeasibility is detected, the program operator has the option to select another cut point to avoid the problem. The set of single strands which results from the cut-point selection is then checked for remaining multiple groups. If multiple groups do remain, they are "uncoupled" by an exhaustive search method to decide which multiples should be saved if conflicts occur. This section uses the heuristic that larger multiple groups should be saved before smaller ones, and that the greatest number of multiples should be saved. Examples of conflicting multiple groups and their resolution are shown in Figure 8. The single strand synthesis is then performed for each single strand using the multiple credits determined in the above section.

Once the times for each single strand are determined, it is necessary to assemble the single strands in the optimal way. This problem is conceptually different from the previous problem in that more than two subgroups can be joined in one reaction step. These subgroups can either be single strands or duplexes. All duplexes used in reactions must

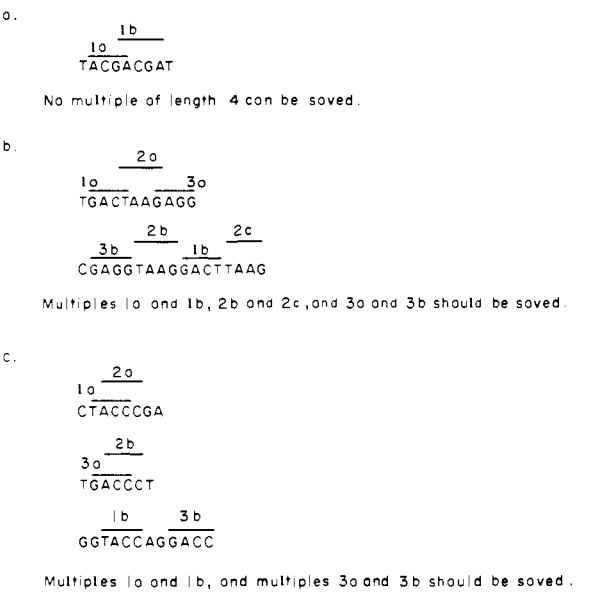


Figure 8. Examples of the sorting of interacting multiple groups to maximize multiple group retention.

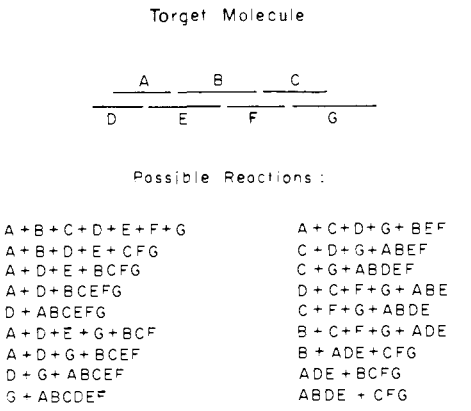


Figure 9. Duplex-forming reactions for a duplex containing seven single strands.

contain three or more groups. A duplex made from two single strands is only hydrogen bonded and is the same as two single strands since no ligase reaction has occurred. Figure 9 illustrates the different reactions which might be used to make a duplex containing seven single strands. The single strand time values being given, the synthesis is searched in the forward direction optimizing over the subgroups used in the joining reaction until the target DNA sequence has been encountered. The only heuristics used are that no subgroups larger than two-thirds of the final molecule size are used (because of separation problems), and that each reaction may involve no more than eight starting materials (either single strands or duplexes). For more than eight groups, the search time becomes prohibitive. The reaction, separation, and analysis models for duplexes are used in the optimization. Following the selection of the duplex-reaction sequence, the program prints out the reaction pathway and the associated evaluation information (time values, excesses of reagents, yields, etc.).

Example of the DINASYN Program

The DINASYN program has been used to synthesize reaction paths to a number of DNA's.[22] The reaction paths have indicated a reduction in attention time of from 20 to 50% over the plans developed by synthetic chemists using an heuristic approach. In addition, the complete synthesis plan can be generated in much less time using the DINASYN pro-

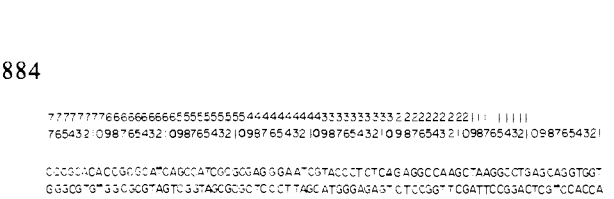


Figure 10. Schematic representation of the nucleotide sequence for the structural gene for an alanine transfer ribonucleic acid from yeast.

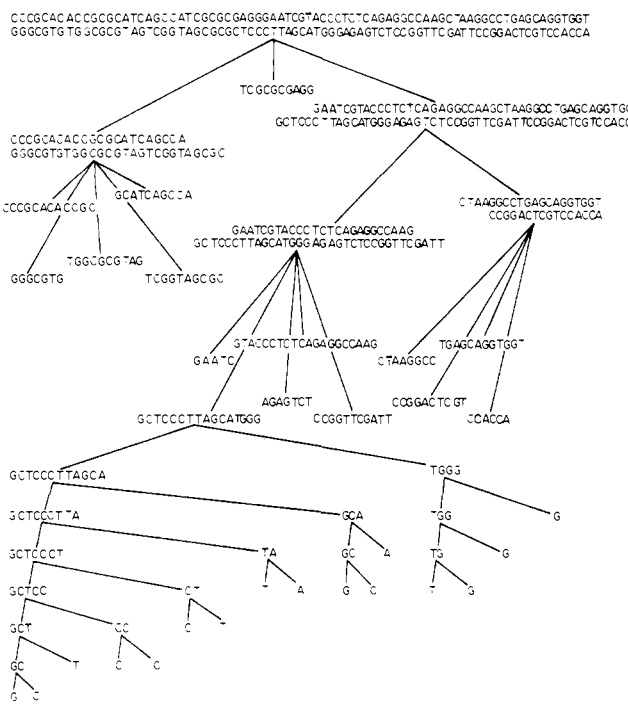


Figure 11. Part of the synthesis path used in the synthesis of a gene for alanine tRNA from yeast.

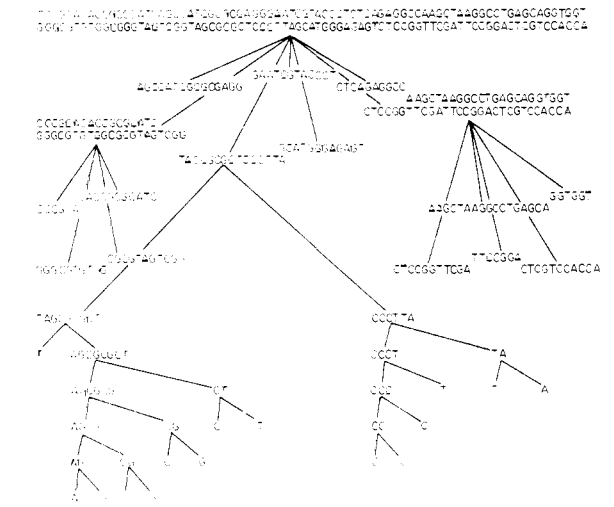


Figure 12. Part of the optimal synthesis plan for alanine tRNA from yeast. This plan was developed using the DINASYN program.

gram. The ease of reaction path synthesis allows rapid investigation of many alterations in the synthetic plan. As an example of the potential savings which could accrue for a specific DNA, consider the gene shown in Figure 10. Part of the synthesis tree used to synthesize this gene[18] is shown in Figure 11. The execution of this synthetic plan required slightly over 20 man years of attention time. The evaluation of this path using the models in the DINASYN program gave 22 man years of attention time. Part of the optimal path generated using the program is illustrated in Figure 12. The attention time required for this path is about 11 man years or a savings of 50% over the path actually executed. The savings in time are due to the extensive use of multiple groups and the parallel nature of the reaction path. The reduction in the "depth" of the path considerably reduces the yield losses. The combination of seven components in the last duplex joining reaction is interesting. The reaction has the proper overlap of chains, no mismatch or self complementary pairings, is easily separated, and has relatively few analysis problems. The complete tree for the synthesis with the proper conditions at each reaction step is the output from the program.[22]

Acknowledgments. Part of this work was performed while the first three authors were in the Department of Chemical Engineering at the Massachusetts Institute of Technology. This work was partially supported by NIH Grant Number 1 R01 GM 22041-01.

References and Notes

(1) Department of Chemical Engineering, University of California, Berkeley, Calif.
(2) Distrigas, Inc., Boston, Mass.
(3) Department of Chemistry, University of Colorado, Boulder, Colo.
(4) Faculty of Medicine, University of Calgary, Calgary, Alberta, Canada.
(5) H. G. Khorana, K. L. Agarwal, H. Büchi, M. H. Caruthers, N. K. Gupta, K. Kleppe, A. Kumar, E., Ohtsuka, U. J. Raj Bhandary, J. H. van de Sande, V. Sgaramella, T. Terao, H. Weber, and T. Yamada. *J. Mol. Biol.*, **72,** 209 (1972).
(6) M. H. Caruthers, K. Agarwal, P. Besmer, P. Cashion, M. Fridkin, E. Jay, A. Kumar, P. Loewen, R. Miller, K. Minamoto, B. RamaMoorthy, A. Panet, J. H. van de Sande, T. Sekiya, N. Siderova, and H. G. Korana, Abstracts, Ninth International Congress of Biochemistry, Stockholm, No. 3.10.
(7) P. C. Loewen, T. Sekiya, and H. G. Khorana, *J. Biol. Chem.*, **249,** 217 (1974).
(8) M. H. Caruthers, D. Goeddel, and D. Yansura, unpublished research.
(9) C. L. Harvey, R. Wright, A. F. Cook, D. T. Marchuk, and A. L. Nussbaum, *Biochemistry*, **12,** 208 (1973).
(10) E. J. Corey and W. T. Wipke, *Science*, **166,** 179 (1969).
(11) H. G. Khorana, *Pure Appl. Chem.*, **17,** 349 (1968).
(12) G. J. Powers and R. Jones, *AIChE J.*, **6,** 1204 (1973).
(13) R. Bellman, "Dynamic Programming," Princeton University Press, Princeton, N.J., 1957.
(14) (a) P. R. Wells, "Linear Free Energy Relationships," Academic Press, London, 1968. (b) The yield data are from the series of papers, "Studies on Polynucleotides," by Khorana and his coworkers; see ref 15.
(15) "Studies on Polynucleotides": (a) XXX, T. M. Jacob and H. G. Khorana, *J. Amer. Chem. Soc.*, **86,** 1630 (1964); (b) XLIII, E. Ohtsuka, M. W. Moon, and H. G. Khorana, *ibid.*, **87,** 2956 (1965); (c) XLIV, T. M. Jacob and H. G. Khorana, *ibid.*, **87,** 2971 (1965); (d) XLV, S. A. Narang and H. G. Khorana, *ibid.*, **87,** 2981 (1965); (e) XLVI, S. A. Narang, T. M. Jacob, and H. G. Khorana, *ibid.*, **87,** 2988 (1965); (f) LX, H. Kossel, M. W. Moon, and H. G. Khorana, *ibid.*, **89,** 2148 (1967); (g) LXII, S. A. Narang, T. M. Jacob, and H. G. Khorana, *ibid.*, **89,** 2158 (1967); (h) LXIV, T. M. Jacob, S. A. Narang, and H. G. Khorana, *ibid.*, **89,** 2177 (1967); (i) LXV, H. Kossel, H. Büchi, and H. G. Khorana, *ibid.*, **89,** 2185 (1967); (j) LXVI, E. Ohtsuka and H. G. Khorana, *ibid.*, **89,** 2195 (1967); (k) LXXXVII, N. K. Gupta, E. Ohtsuka, H. Weber, S. H. Chang, and H. G. Khorana, *Proc. Nat. Acad. Sci. U.S.*, **60,** 285 (1968); (l) LXXXVIII, N. K. Gupta, E. Ohtsuka, V. Sgaramella, H. Büchi, A. Kumar, H. Weber, and H. G. Khorana, *ibid.*, **60,** 1338 (1968); (m) XCIII, J. Hachman and G. H. Khorana, *J. Amer. Chem. Soc.*, **91,** 2749 (1969); (n) CIII, H. G. Khorana, *et al.*, *ibid.*, **94,** 1120 (1972); (o) CIV, H. Weber and H. G. Khorana, *J. Mol. Biol.*, **72,** 219 (1972); (p) CV, H. Büchi and H. G. Khorana, *ibid.*, **72,** 251 (1972); (q) CVI, A Kumar, E. Ohtsuka, and H. G. Khorana, *ibid.*, **72,** 289 (1972); (r) CVII, E. Ohtsuka, A. Kumar, and H. G. Khorana, *ibid.*, **72,** 309 (1972); (s) CVIII, A. Kumar and H. G. Khorana, *ibid.*, **72,** 329 (1972); (t) CIX, K. L. Agarwal, A. Kumar, and H. G. Khorana, *ibid.*, **72,** 351 (1972); (u) CX, M. H. Caruthers, J. H. van de Sande, and H. G. Khorana, *ibid.*, **72,** (1972); (v) CXI, M. H. Caruthers and H. G. Khorana, *ibid.*, **72,** 407 (1972).
(16) M. H. Caruthers and J. H. van de Sande, unpublished research.
(17) K. L. Agarwal, A. Yamazaki, and H. G. Khorana, *J. Amer. Chem. Soc.*, **93,** 2754 (1971).
(18) M. H. Caruthers, K. Kleppe, J. H. van de Sande, V. Sgaramella, K. L. Agarwal, H. Büchi, N. K. Gupta, A. Kumar, E. Ohtsuka, U. L. Raj Bandary, T. Terao, H. Weber, T. Yamada, and H. G. Khorana, *J. Mol. Biol.*, **72,** 475 (1972).
(19) V. Sgaramella and H. G. Khorana, *J. Mol. Biol.*, **72,** 427 (1972).
(20) G. M. Tslapolis and S. A. Narang, *Biochem. Biophys. Res. Commun.*, **39,** 631 (1970).
(21) P. Loewen, P. Bessmer, R. C. Miller, and H. G. Khorana, unpublished research.
(22) R. Jones, "Computer-Aided Synthesis of DNA," Master's Thesis, Department of Chemical Engineering, Massachusetts Institute of Technology, 1972.