$F(\ )$ = perturbed feasible region resulting from the perturbation in the parenthesis
F^0 = unperturbed feasible region
F^* = permanently feasible region
$I(\ )$ = set of indices of the constraints tight at ( )
J_t = complement of ϕ_t
N = number of primal variables
M = number of constraints
P = set of permissible perturbations in the model coefficients
R = set of indices of rows which have a dependency in the coefficients of that row
T = number of terminal sets
α_{ij} = bounds on the perturbations in the a_{ij}
β_i = bounds on the perturbations in the b_i
μ_{ij} = coefficients of the row dependency equations
η_{ij} = coefficients of the column dependency equations
ϕ_t = set of indices of those constraints of terminal set t which are perturbed to their maximum values
$\rho(\ )$ = radius of the flexibility set
$\mathcal{M}$ = set of integers between $-M$ and $+M$
$\mathcal{F}$ = set of all perturbed feasible regions
Superscripts = elements in a sequence
Subscripts = component of a matrix
Prefix δ = a perturbation of the quantity following it

Mathematical Symbols

ϵ = a member of
$\neq$ = not equal to
$\emptyset$ = empty set
$\cap$ = set intersection
$\cup$ = set union
$\subset$ = set inclusion
: = such that
$||\ ||$ = Euclidean norm

LITERATURE CITED

Beightler, C. S. and D. J. Wilde, "Sensitivity Analysis gives better insight into Linear Programming," *Hydrocarbon Proc. Petrol. Ref.*, **44**, 111 (Feb., 1965).

Dantzig, G. B., "Large Scale Linear Programming," in *Mathematics of the Decision Sciences*, Part I, G. B. Dantzig and A. F. Veinott, Jr. (eds.), American Math. Soc. (1968).

Friedman, Y., and G. V. Reklaitis, "Flexible Solutions to Linear Programs under Uncertainty: Inequality Constraints," *AIChE J.*,

Rockafeller, R. T., *Convex Analysis*, Princeton University Press, New Jersey (1970).

Manuscript received June 4, 1974; revision received and accepted September 20, 1974.

Synthesis of Fault Tolerant Reaction Paths

The problem of synthesizing an optimal low risk reaction network from unreliable reactions is formulated. An expected cost criterion function for decision-making is derived both for the case of a continuously operating network of reactions and for the case of reaction paths that involve batch reactions. It is shown that in general the decision space has a nonserial structure and the search for the optimal path will involve use of network search methods. An example is given where the optimal reaction path is synthesized for deoxyribonucleic acid. This example demonstrates the decision-making strategy for a class of batch reaction systems.

GARY J. POWERS
DANIEL L. CUMMINGS
and
RAM N. S. RATHORE

Department of Chemical Engineering
Carnegie-Mellon University
Pittsburgh, Pennsylvania 15213

SCOPE

In planning the synthesis of even the simplest of molecules, the possibility of failure at each step must be carefully considered. In general, the best synthetic plan involves the conversion of the most readily available and cheapest raw materials into the desired product in the fewest number of routine steps and in the highest possible yield. The objective function which is used to measure the success of a synthetic plan will depend on the context under which the synthesis is executed. In large-scale industrial syntheses, the costs of raw materials and operations and the safety of the processing system are of major importance. In syntheses performed in the laboratory the speed and ease with which a compound can be obtained is of more interest.

In either case the reliability of the synthetic plan must be considered. In industrial syntheses a given synthetic plan may be abandoned because raw material costs have changed, because a catalyst is poisoned by an impurity, because an unforeseen impurity is difficult to separate, or because mechanical equipment required to handle the reaction mixture and products cannot be reliably designed or maintained. In the laboratory an intermediate may be too unstable to store or too insoluble to allow use of convenient laboratory apparatus. In addition, there may be no analytical technique suitable for characterization of the reaction products. Problems of this sort are difficult to anticipate and sometimes economically impossible to overcome. For this reason a synthetic plan which contains reliable reactions and alternate routes or convenient detours will be wiser and safer than one whose success is wholly dependent on one critical reaction. In this paper a strategy is presented for finding synthesis pathways which have a low risk and still maintain the desirable features

of high yield, few steps, and low cost reagents. With the method described in the paper it is necessary to know, or be able to predict, the probability of failure and cost of any reaction which might be used in the reaction path.

CONCLUSIONS AND SIGNIFICANCE

This work is significant in that the reliability of reaction steps is incorporated in the synthetic planning of complex reaction networks. Modern concepts of decision theory have been applied to define the decision space and the expected cost of the reaction paths. The expected cost of a raction path is shown to depend on the cost and reliability of the reactions involved, the probable costs of the reagents used in the reactions, and the probable cost of being stranded at intermediate products.

In many cases the decision space can be represented in the form of a tree in which the number of paths leading to the target molecules from the raw materials is very large. This makes exhaustive search for the optimal reaction path infeasible, and it is necessary to use partial enumeration methods such as branch and bound algorithms or backtrack programming. In addition to these methods, one important class of reaction network problems has a serial structure which allows the use of dynamic programming, a more efficient search strategy. An example of the synthesis of deoxyribonucleic acid illustrates how dynamic programming can be applied to establish the optimal low risk reaction path. This example also shows that the optimal reaction path can change when the probability of failure of reactions is considered.

In planning the synthesis of a given target molecule, the probability that certain steps in the reaction path might fail is an important consideration. The failure of steps in the reaction path can have two effects:

1. In a batch process, the steps may need to be repeated several times to achieve a successful operation. The increased number of trials increases the cost of that operation and hence increases the cost of that reaction path. In a continuous process it may be necessary to blend off-specification material with higher-quality product, to sell the material at a lower price, to dispose of the material as a waste, or in extreme cases, to shut the process down and rebuild parts of the plant.

2. It is possible that a particular operation in the reaction path may prove infeasible. In this case it is necessary to seek an alternative reaction path which does not include that operation. For a batch process, if no alternatives exist from that point in the path, it is necessary to start the synthesis again from starting materials. The effort required to make the precursors to the failed operation is then lost. In a continuous process it will be necessary to shut down the process and modify it so that the reaction becomes feasible or so that an alternate reaction path can be used. Hence the selection process for the optimal reaction paths must consider both deterministic costs, such as the cost of raw materials and the effort required to carry out the reaction, separation and analysis steps in the path, as well as probabilistic factors, such as the probability of failure of each operation and the probability that the cost of making a precursor could be lost.

In this paper we develop the expected cost criterion function for the general reaction synthesis problem. This function accounts for the failure of reaction steps and illustrates the decision-making strategies for establishing the optimal reaction system.

Problem Formulation

Consider the problem of finding the optimum reaction path to a molecule represented by ABCDE. The symbols represent the structure and functionality of the molecule. Systematic means for representation of organic molecules have been developed and are discussed by Hendrickson (1971) and Powers and Jones (1973). Raw materials A, B, C, D, and E are available for the synthesis of ABCDE. Reactions are known which cause the raw materials to combine to form larger molecules. The cost of carrying out each reaction ($RC_{j,k}$) which combines compound j with compound k is assumed known. In addition, the probability that the reaction (j, k) will fail on a given trial is available from prior data as $P_{j,k}$. In the case of continuous reactions $P_{j,k}$ represents the number of failures per unit time. The cost of raw material i is known to be C_i per mole.

The objective is to produce the target molecule t at the lowest possible cost. The minimization problem is given in Equation (1) and is read: minimize the sum of the reaction costs and raw material costs over the set of reaction paths which convert available raw materials into the target molecule.

$$\underset{(t)}{\text{Min}} \{\Sigma RC_{j,k} + \Sigma C_i\} \qquad (1)$$

Figure 1 illustrates the graph of precursor molecules and the reactions which could be involved in the synthesis of

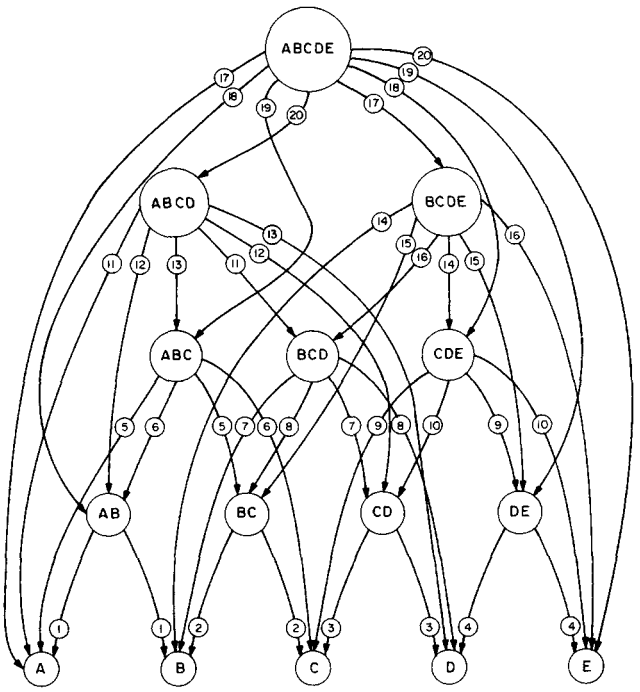


Fig. 1. The generation of precursors from the molecule ABCDE. Two arcs bearing the same number represent one reaction operator (Rathore et al., 1974).

the target molecule ABCDE. In preparation of Figure 1 only binary reactions (two small molecules react to produce a larger third molecule) are allowed and no precursor larger than the target molecule is considered. The nodes in the graph represent the precursors (or intermediate products) and arcs represent the reactions which may constitute a batch experiment in the laboratory or a continuous chemical process of a large petrochemical complex. It can be shown that there are 14 different reaction paths for the synthesis of molecule ABCDE. All these paths are imbedded in the graph of Figure 1. Powers and Jones (1973) show that the number of possible reaction paths increases very rapidly as the number of functional groups on the target molecules increases.

The optimization involves determination of optimal reaction conditions and the reaction path to the target molecule. For example, there is only one reaction path to the compound AB. It involves the reaction A + B → AB. The optimum conditions for carrying out the reaction (temperature, pressure, concentration, separation method, etc.) could be determined and the cost of product AB determined. The cost is given by Equation (2) where * indicates optimum costs.

$$C^*_{AB} = RC^*_{A,B} + C^*_A + C^*_B \tag{2}$$

The optimum cost to produce the molecule ABC depends on the reaction path used to produce ABC. The two possible reaction paths involve*

1. A + B → AB
   AB + C → ABC
2. B + C → BC
   BC + A → ABC

The costs for producing ABC are given in Equation (3)

$$C^*_{ABC} = \text{Min} \begin{cases} \text{Min } (RC_{AB,C} + RC_{A,B} + C_A + C_B + C_C) \\ \text{Min } (RC_{A,BC} + RC_{B,C} + C_A + C_B + C_C) \end{cases} \tag{3}$$

where the first minimization is over the sum of the costs along each reaction path and the second minimization is the selection between the paths.

The minimization along the complete reaction path is necessary as interactions can occur between reactions on a path. For example, unconverted raw materials or byproducts from one reactor might be passed to the input of a following reactor. In addition, the energy released in one reaction might be used to supply energy required in another reaction. In still another case the byproducts from one reaction might be used as a reagent for another reaction. All of these interactions indicate the need for a search technique which considers coupling between the reactions along a path. Several algorithms have been developed in the operations research field for finding minimum cost paths in graphs of the type shown in Figure 1. Branch and bound search (Mitten, 1970), backtrack programming (Golomb and Baumert, 1965) and nonserial dynamic programming (Bertelé and Brioschi, 1972) are all applicable to the problem.

There is, however, a large class of reaction networks where the interactions between reactions along a path are small, for example, in batch reactions where the products are completely separated prior to the next reaction, and in continuous reaction networks if no recycle of material or energy occurs between the reactions.* When the reactions are independent the following cost function applies:

$$C^*_{ABC} = \text{Min} \begin{cases} RC^*_{AB,C} + RC^*_{A,B} + C^*_A + C^*_B + C^*_C \\ RC^*_{A,BC} + RC^*_{B,C} + C^*_A + C^*_B + C^*_C \end{cases} \tag{4}$$

where the first level of minimization is performed for each reaction independent of the other reactions. This type of problem has a serial information flow and can be solved by dynamic programming. The general recursion formula is

$$C^*_{j,k} = \text{Min } \{RC^*_{j,k} + C^*_j + C^*_k\} \tag{5}$$

The following formulation assumes that the reaction path selection is a serial problem. If the problem is not serial, the more general graph search techniques mentioned above should be used.

Dynamic programming has been used by Powers and Jones (1973) and Powers et al. (1974) for planning the synthesis of bihelical deoxyribonucleic acids. The pathways found using this technique have commonly been 40% less costly than those generated using heuristic selection methods. On a number of occasions the reaction path selected by the dynamic programming method contained reactions which were failure prone. The question was then raised: How should the probability of failure of a reaction step influence the selection of the optimal reaction path? One approach to this problem is given in the next section.

Expected Cost of Precursors

The procedure for selecting the optimal reaction path to a given target molecule must consider the probability of failure of reactions on the reaction paths, the probability of being stranded at intermediate products, and the desired probability of achieving the target molecule.

A successful reaction step involves reaction, separation, and analysis of the reaction products so that the next step in the reaction path can be executed. The reaction step fails if any one part of it fails. For example, if a reaction gives the wrong products, or a low yield of the desired product it could be considered a failure. The exact definition of the reaction failure will depend on the environment in which the synthesis is performed. In the laboratory the definition of failure of a reaction may be very different from that used in a large-scale plant. Similarly the definition of failure of the separation and analysis parts of overall reaction step depends on the environment. In the chemistry laboratory, failure to separate a reaction mixture is primarily dependent on the differences in physical and chemical properties that exist between the species in the mixture. A wide range of techniques including chromatography, extraction, crystallization, etc. could be used. The major source of failures in laboratory separations is the species having property differences sufficiently small so that no technique can give the required purity products for the next reaction step. Isomers and large molecules of slightly different molecular characteristics are most troublesome.

Continuous industrial reactions may fail because of equipment failures or deactivation of catalysts. Fires, explosions, plugging of pipes, valves and pumps, etc. may occur in chemical plants and can cause significant business, property and personnel damage. Whitaker (1973) and Browning (1969a, 1969b) describe the range of the probability of failure in chemical plants and estimate the failure consequences. A failure of a unit in a chemical plant may lead to deterioration of product quality, or reduction of production, or even shutdown of the entire chemical plant. A proper definition of failure must account for all these effects.

* The reaction AC + B → ABC is not allowed.
* The recycle of material around a reactor when the conversion of reagents is low does not add interactions between reactions.

The prediction of the probability of failure of the reaction depends on the ability to predict failures in the reaction, separation, and analysis operations. If a particular class of reaction steps has been carried out a number of times it may be possible to use these data to predict the future probabilities of failure reactions. If probability data do not exist it may be necessary to resort to more subjective probabilities based on expert opinion.

The probability of being stranded at an intermediate product depends on the number or reactions that can convert the intermediate into other intermediates or products and the probability of failure of these reactions. The major concern is that the reaction path selected be flexible enough to allow detours if a particular reaction step proves infeasible.*

The desired probability of success for the complete reaction path can be determined in several ways. In an industrial environment the overall risk allowed in the reaction path is determined by the potential market value of the product, the capital position of the firm, and whether government regulations are involved. In the laboratory the situation is not as clear. What probability of failure is acceptable to a research chemist when he or she is starting a large synthetic project? Is it different for a researcher just beginning a career than for an established, well recognized worker? Several studies by sociologists of science have indicated that the normal practice of science involves a very low level of risk taking (Kuhn, 1967).

Regardless of the source of probability data and the amount of risk taking, it is necessary to develop the proper decision-making framework in which to select the best reaction path. In the previous sections the use of dynamic programming to select the optimum reaction path was described. In that method the deterministic costs associated with consuming reagents and carrying out reaction steps were used. What is needed is a way to combine this deterministic approach with the probabilisitc considerations discussed above. The expected value criterion supplies one means for combining both the deterministic costs with the probability of failure. The theory on which this approach is based was developed by von Neumann and Morganstern (1944) and has been presented more recently by Raiffa (1968). The goal is to generate a value for each precursor molecule which correctly reflects the measurable rewards, the probability of success of each part of the reaction step, and a measure of the synthesis planner's regard for risk. In the following analysis the expected value of an event is defined as the sum of the deterministic costs of the outcomes multiplied by the probability of their occurrence. Decision-makers, in general, are averse to risk and consider the value of an event to be less than its expected value. The value which an event has to the decision-maker is often termed his utility. A number of different utility functions have been described and used in formal decision analysis (Simon, 1966). The extent of risk aversion is not a constant personal or organizational characteristic. For illustrative purposes, we have chosen to use the expected value as the measure of utility for this problem.

In the previous sections the value of a precursor was defined as the cost of the reaction step (including separation and analysis) which produced it plus the cost of the reagents required for the reaction. The consideration of failure of reaction steps requires several changes in this definition. First it is necessary to define the expected cost of a precursor in terms of:

1. The expected cost of carrying out the reaction step,
2. The expected cost of the reagents, and
3. The expected cost of being stranded at the current position.

The recurrence formula using expected costs is given by Equation (6).

$$\bar{C}^*_{jk} = \underset{[j,k]}{\text{Min}} \left[\overline{RC}^*_{j,k} + \bar{C}^*_j + \bar{C}^*_k \right] + \underset{[jk,l]}{\Pi} P_{m,n} \left[\underset{[j,k]}{\text{Min}} \left\{ \overline{RC}^*_{j,k} + \bar{C}^*_j + \bar{C}^*_k \right\} \right] \quad (6)$$

A bar over a variable indicates expected cost. The first expression in Equation (6) is the same as for the deterministic case with the exception that expected values are used. The optimization is over all reaction paths leading to compound jk. The second term is the contribution to the expected cost due to the probability of being stranded at species jk. The pi product is over all the reactions (jk, l) leaving species jk. The value of the event is this joint probability times the value of the molecule jk.

Computing $RC^*_{j,k}$ for Continuous Reactions. In a continuous industrial process the expected cost of the reaction will depend on the frequency of failures, average down time associated with the major failings, and the deterministic cost of running a process during the available time. Browning (1969a) has developed relationships for computing maximum probable loss (PML) for failures in chemical plants. The losses include loss of fuel, material, facilities, and process equipment. In addition, the losses due to interruption of business and personnel injury are included. Let us assume that the annual probability of a failure f_i is p_i; then

$$\overline{RC}^*_{j,k} = d(RC^*_{j,k}) + \Sigma\, p_i(\text{PML}) \quad (7)$$

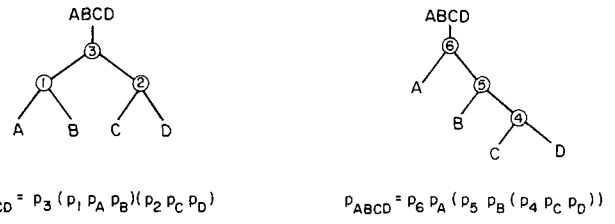
where d is the availability of the plant in days/year, and $RC^*_{j,k}$ (\$/day) is the optimum cost of combining molecules j and k. The value of $\overline{RC}^*_{j,k}$ given from Equation (7) can be used in Equation (6) to compute the expected cost of the precursor molecules in the network of continuous reactions.

Computing $\overline{RC}^*_{j,k}$ for Batch Reactions. The expected cost of a batch reaction step depends on the cost of a single trial of the reaction and the probability of failure of the step. A single reaction step with a probability of failure $P_{m,n}$ defines a Bernoulli trial (Drake, 1967). If a single trial of a reaction step (reaction, separation, and analysis) is not successful it can be repeated. In the following discussion, it is assumed that in case of failure of a reaction all the reagents can be recovered during the separation phase of the reaction step. This corresponds to the case where the reaction does not go and no byproducts are produced. With slight modification the case for loss of reagents can also be considered.

If the probability of failure does not change from trial to trial, a Bernoulli process describes the overall probability of success. If the trials are not independent, a Markov process describes the probabilities. For example, if one learns something about a reaction step when it fails that may change the probability of failure for the next trial a Markov process is required. In the following discussion a simple Bernoulli process is used. The probability of k_0 successes in n independent Bernoulli trials is given by Equation (8).

$$p_{k0} = \binom{n}{k_0} P_s^{k_0} (1 - P_s)^{n-k_0} \quad (8)$$

* This practice of planning for detours is common in the chemical industry. For example, if a large ethylene plant is destroyed by explosion and fire a number of other sources of ethylene or other feed stocks must be found to keep the rest of the reactions which depend on ethylene in operation.



$P_{ABCD} = P_3 (P_1 P_A P_B)(P_2 P_C P_D)$

$P_{ABCD} = P_6 P_A (P_5 P_B (P_4 P_C P_D))$

Fig. 2. Two reaction paths to the molecule ABCDE. The probability of failure of the reaction system is given below each path.

where P_s is the probability of success in a single trial and

$$\binom{n}{k_0} = \frac{n!}{(n-k_0)!\,k_0!}$$

The probability of exactly zero successes in n trials is given by

$$p_0 = (1 - P_s)^n \tag{9}$$

Using probability of failure $P_F = 1 - P_s$, the number of trials to achieve a process probability of zero successes p_0 is given by

$$n = \ln p_0 / \ln P_F \tag{10}$$

Equation (10) defines the number of trials to the first success with probability of success of $(1 - p_0)$. With this definition of the reaction step it is possible to achieve probabilities of failure of the reaction step between P_F and zero by varying the number of trials used. In other situations it may be necessary to have more than one success.* Equation (8) can be used to predict the required trials for these cases.

If the desired overall probability of failure is fixed for each reaction step, n can be computed from Equation (10). The optimum expected cost of the reaction step $\overline{RC}^*$ equals nRC^* where RC^* is the optimal cost for carrying out a single trial of the reaction steps.

The probability of failure of the reaction path depends on the probability of failure of each reaction step. If a probability of failure of the complete pathway leading to the target molecule is specified, recycle of information will exist in the reaction path. A number of combinations of the values of probability of failure of each reaction step in the path can lead to the same overall probability of failure of the reaction path. In other words, it will be difficult to establish values of n and p_0 at each step. This problem cannot be solved by simple dynamic programming. A direct search strategy could be used in which the number of trials for each reaction subproblem is a decision variable. In the following example the probability of failure of each and every reaction step is fixed, by varying the number of trials n, at an acceptably low value so that the effect on the overall path probability of failure is absent. Figure 2 illustrates two reaction paths to the molecule ABCD. If the probability of failure of the reactions steps is equal, the probability of failure for both paths is equal. The probability of failure to make a target molecule from a single binary reaction is

$$P = p_x p_y p_z \tag{11}$$

where p_x is the probability of failure of the reaction step and p_y and p_z are the probabilities of failure to have the proper reagents.

* In the present discussion it is assumed that all the reagents are committed to each trial. Another strategy could involve dividing the reagents and carrying out the reactions sequentially. In these situations more than one success may be required.

Synthesis of Deoxyribonucleic Acids

A previous paper (Powers and Jones, 1973) describes the use of dynamic programming to select optimal reaction paths to bihelical deoxyribonucleic acids. A part of the synthetic plan to a DNA commonly involves the preparation of single strands of DNA up to 20 nucleotides in length. The reaction, separation, and analysis steps associated with the formation of these polymers have been modeled. The models predict the attention time required to carry out any combination of reaction steps. These models are described in detail by Powers and Jones (1973), Powers et al. (1974), Jones (1973), and Randall (1973).

A small target molecule which is required for a larger synthesis plan is AGAGTCT. The letters stand for the nucleotides adenosine, cytosine, guanosine, and thymidine. A phosphodiester linkage joins the monomers. The problem is to make the target molecule given starting materials A, C, G, and T. There are 132 direct reaction paths to this molecule. The best ten of these 132 paths are shown in Figure 3 along with their deterministic attention time. Reaction path 1 is the one selected by a dynamic programming scheme using the recursion formula given in Equation (5).

A simple model for the probability of failure of a single trial of a reaction step is given in Table 1. The model is based on the size of one of the groups taking part in the reaction. If a large attacking group is used the probability of failure of the reaction step is higher. The attacking group is the group which ends up on the right hand side in the notation used here. For example, the reaction AGA + G → AGAG has an attacking group of length one (G) and a probability of failure of 0.05. The reaction AG + AGTCT → AGAGTCT has an attacking group of length five (AGTCT) and a probability of failure of 0.4.

Each reaction step is required to have a probability of failure of 0.01. The number of Bernoulli trials required to give this probability for various single trial probabilities of

1.	AGAGTCT AGAG+TCT AG+AG TC+T A+G A+G T+C	695.33	2.	AGAGTCT AGAG+TCT AG+AG T+CT A+G A+G C+T	703.28
3.	AGAGTCT AGAG+TCT AGA+G TC+T AG+A T+C A+G	704.57	4.	AGAGTCT AGA+GTCT AG+A GT+CT A+G G+T C+T	709.43
5.	AGAGTCT AGAGT+CT AGA+GT C+T AG+A G+T A+G	711.49	6.	AGAGTCT AGAG+TCT AGA+G T+CT AG+A C+T A+G	712.51
7.	AGAGTCT AGAGT+CT AGAG+T C+T AG+AG A+G A+G	713.29	8.	AGAGTCT AGAGTC+T AGA+GTC AG+A GT+C A+G G+T	724.35
9.	AGAGTCT AGA+GTCT AG+A G+TCT A+G TC+T T+C	725.36	10.	AGAGTCT AGA+GTCT A+GA GT+CT G+A G+T C+T	729.20

Fig. 3. The ten best reaction paths to the DNA AGAGTCT when probability of failure is not considered. The attention time in hours is given next to each path.

failure is also given in Table 1.

Pathway number 6 in Figure 3 was selected by heuristic methods and carried out by Dr. H. G. Khorana and coworkers as part of the synthesis of the gene for the alanine transfer ribonucleic acid in yeast (Khorana et al., 1972).

When the probabilities given in Table 1 are used in the recursion formula given by Equation (6), the pathways give in Figure 4 are found to be best. Figure 5 shows how the paths have shifted when the probability of failure is considered.

The expected cost of being stranded at an intermediate molecule is a small fraction of the total cost. This is due to the low probability of failure assumed for each reaction step. For higher probabilities of failure the reaction intermediates which have alternate reaction paths leading to the target molecule will be more favored.

The differences in the cost of these pathways are small. For the complete synthesis of a large DNA the difference between a satisfactory reaction path selected by heuristic means and the optimal path is often 50% or larger (Powers and Jones, 1973). Note that consideration of probabilities of failure shifts the best path closer to that selected by heuristic means.

TABLE 1. THE PROBABILITY OF FAILURE OF A DNA CONDENSATION REACTION AS A FUNCTION OF LENGTH OF ATTACKING GROUP

Length of attacking group	Probability of failure	Number of trials to achieve $p_f = 0.01$
1	0.05	1.53
2	0.10	2.00
3	0.20	2.86
4	0.25	3.32
5 and greater	0.40	5.02

1.	AGAGTCT AGAG+TCT AGA+G TC+T AG+A T+C A+G	922.79	
2.	AGAGTCT AGAG+TCT AGA+G T+CT AG+A C+T A+G	930.56	
3.	AGAGTCT AGAGT+CT AGAG+T C+T AGA+G AG+A A+G	943.32	
4.	AGAGTCT AGAGTC+T AGA+GTC AG+A GT+C A+G G+T	954.45	
5.	AGAGTCT AGAGT+CT AGA+GT C+T AG+A G+T A+G	968.69	
6.	AGAGTCT AGAGTC+T AGAG+TC AGA+G T+C AG+A A+G	970.85	
7.	AGAGTCT AGA+GTCT AG+A GTC+T A+G GT+C G+T	979.38	
8	AGAGTCT AGAGTC+T AGAGT+C AGAG+T AGA+G AG+A A+G	983.07	
9.	AGAGTCT AGA+GTCT A+GA GTC+T G+A GT+C G+T	999.15	
10.	AGAGTCT AGAG+TCT AG+AG TC+T A+G A+G T+C	999.18	

Fig 4. The ten best reaction paths to the DNA AGAGTCT when probability of failure is considered.

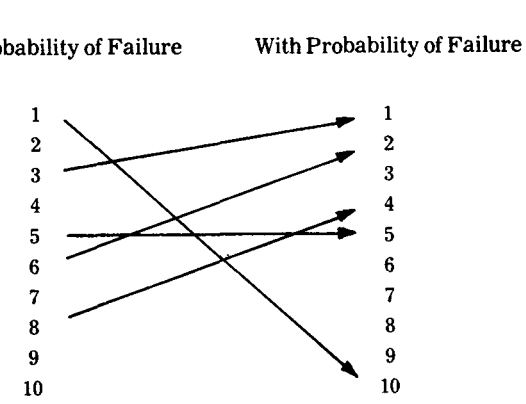


Fig. 5. The change in rank of reaction paths when probability of failure is considered.

ACKNOWLEDGMENT

Part of this work was carried out when G. J. Powers and D. C. Cummings were at Massachusetts Institute of Technology and R. N. S. Rathore was at Tufts University.

LITERATURE CITED

Bertelé, V., and F. Brioschi, *Nonserial Dynamic Programming*, Academic Press, New York (1972).

Browning, R. L., "Calculating Loss Exposures," *Chem. Eng.*, 239 (Nov. 17, 1969a).

———., "Estimating Loss Probabilities," *ibid.*, 135 (Dec. 15, 1969b).

Drake, A. W., *Fundamentals of Applied Probability Theory*, McGraw-Hill, New York (1967).

Golomb, A. O., and L. Baumert, "Backtrack Programming," *J. ACM*, **12**(4), 516 (1965).

Hendrickson, J. G., "A Systematic Characterization of Structures and Reactions for Use in Organic Synthesis," *J. Am. Chem. Soc.*, **93**, 6847 (1971).

Jones, R., "Computer-Aided Synthesis of DNA," Masters thesis, Mass. Inst. Technol., Cambridge (1972).

Khorana, H. G., L. L. Agarwal, H. Buchi, M. H. Caruthers, N. Gupta, K. Kleppe, A. Kumar, E. Ohtsuka, U. L. Rajbhandary, J. H. van de Sande, V. Sgaramella, R. Rerao, H. Weber, and T. Yamada, *J. Am. Chem. Soc.*, **94**, 1120 (1972).

Kuhn, T., *The Structure of Scientific Revolutions*, 2nd edit., The University of Chicago Press (1962).

Mitten, L. S., "Branch and Bound Methods: General Formulation and Properties," *Operations Research*, **18**(1), 24 (1970).

Powers, G., and R. Jones, "Reaction Path Synthesis Strategies," *AIChE J.*, **19**, 1204 (1973).

———., G. Randall, J. H. van de Sande, M. Caruthers, and H. Khorana, *J. Am. Chem. Soc.*, in press.

Raiffa, H., "Decision Analysis, Introductory Lectures on Choices under Uncertainty," Addison-Wesley, Reading, Mss. (1968).

Randall, G., "Reaction Models for DNA Synthesis," Masters thesis, Mass. Inst. Technol., Cambridge (1972).

Rathore, R. N. S., K. A. Van Wormer, and G. J. Powers "Synthesis Strategies for Multicomponent Separation Systems with Energy Integration," *AIChE J.*, **20**, 491 (1974).

Simon, H. A., *Theories of Decision-Making in Economics and Behavioral Science*, Survey IX, "Surveys of Economic Theory," Vol. III, p. 39. St. Martin's Press, New York (1966).

von Neumann, J., and O. Morganstern, *Theory of Games and Economic Behavior* p. 121, Wiley, New York (1944).

Whitaker, G. D., "Statistical Reliability Models for Chemical Process Plant," *The Chem. Eng.* (London), 471, (Oct., 1973).

Manuscript received February 8, 1974; revision received November 1 and accepted November 4, 1974.