Biochem. J. (1976) **159**, 105–120
Printed in Great Britain

Subsite Mapping of Enzymes

DEPOLYMERASE COMPUTER MODELLING

By JIMMY D. ALLEN* and JOHN A. THOMA
Department of Chemistry, University of Arkansas, Fayetteville, AR 72701, U.S.A.

(*Received* 9 *March* 1976)

We have developed a depolymerase computer model that uses a minimization routine. The model is designed so that, given experimental bond-cleavage frequencies for oligomeric substrates and experimental Michaelis parameters as a function of substrate chain length, the optimum subsite map is generated. The minimized sum of the weighted-squared residuals of the experimental and calculated data is used as a criterion of the goodness-of-fit for the optimized subsite map. The application of the minimization procedure to subsite mapping is explored through the use of simulated data. A procedure is developed whereby the minimization model can be used to determine the number of subsites in the enzymic binding region and to locate the position of the catalytic amino acids among these subsites. The degree of propagation of experimental variance into the subsite-binding energies is estimated. The question of whether hydrolytic rate coefficients are constant or a function of the number of filled subsites is examined.

Several authors have invoked the subsite model to account for the enzymic properties of depolymerases such as proteinases (Schechter & Berger, 1967; Abramowitz *et al.*, 1967; Morihara & Oka, 1968; Morihara *et al.*, 1969; Atlas *et al.*, 1970; Thompson & Blout, 1970; Coggins *et al.*, 1974), nucleases (Cuatrecusas *et al.*, 1968; Lazarus *et al.*, 1968; Chou & Singer, 1970; Godefroy, 1970), and carbohydrases (Chipman & Sharon, 1969; Holler *et al.*, 1974; Robyt & French, 1963, 1970; Hiromi, 1970; Thoma *et al.*, 1970, 1971; Nitta *et al.*, 1971; Kato *et al.*, 1974; Iwasa *et al.*, 1974; Shibaoka *et al.*, 1974; Rexová-Benková, 1973; Thoma and Allen, 1976). The depolymerase subsite model depicts the substrate-binding region of the enzyme to be a tandem array of subsites (Fig. 1). Each subsite is complementary to, and interacts with, a substrate monomer unit. There are a number of different ways in which a substrate oligomer can interact with these subsites. For example, as depicted in Fig. 1, a tetramer can form eight positional isomers with a five-subsite enzyme. The population of each positional isomer is dictated by the energetics of interaction of the substrate monomer units with their respective subsites.

A substrate oligomer can bind non-productively so that a susceptible bond does not extend over the catalytic amino acids of the enzyme; alternatively the substrate can bind productively so that a susceptible bond does lie over the catalytic site, in which case the bond is cleaved. For the tetramer shown in Fig. 1 there are three productive positional isomers (binding-mode index IV,4; V,4; and VI,4); the remaining positional isomers are non-productive. When a substrate binds productively, the rate of bond hydrolysis is dictated by the hydrolytic rate coefficient, k_{+2}. (A complete listing of symbols is given in Table 1.)

The process of quantifying the subsite model is referred to as subsite mapping. To completely map the binding region of an enzyme, one must (1) determine the number of subsites, (2) locate the position of the catalytic amino acids within the subsites, (3) determine the subsite–substrate-monomer-unit binding energies and (4) determine the value of hydrolytic rate coefficients. The value of the subsite model will lie in the ability of a generated subsite map to quantitatively account for experimentally measurable parameters.

Two groups have attempted to place the depolymerase subsite model on a quantitative footing by using the amylase–maltodextrin system (Hiromi, 1970; Nitta *et al.*, 1971; Thoma *et al.*, 1970, 1971). Hiromi (1970) applied the subsite model to an exo-amylase, glucoamylase. Subsite mapping for exo-enzymes is simplified, because there is only one productive binding mode for each substrate. Hiromi's group (Nitta *et al.*, 1971) extended his development of the exo-subsite model to an endoamylase. These authors used qualitative estimates of the relative rates of bond cleavages in oligomeric substrates, i.e. bond-cleavage frequencies, together with the chain-length dependence of Michaelis parameters ($\tilde{K}_m$ and $\tilde{V}$) to predict a partial subsite map for *Aspergillus oryzae* α-amylase. However, as we have shown

* Present address: Department of Chemistry, University of California, Santa Barbara, CA 93106, U.S.A.

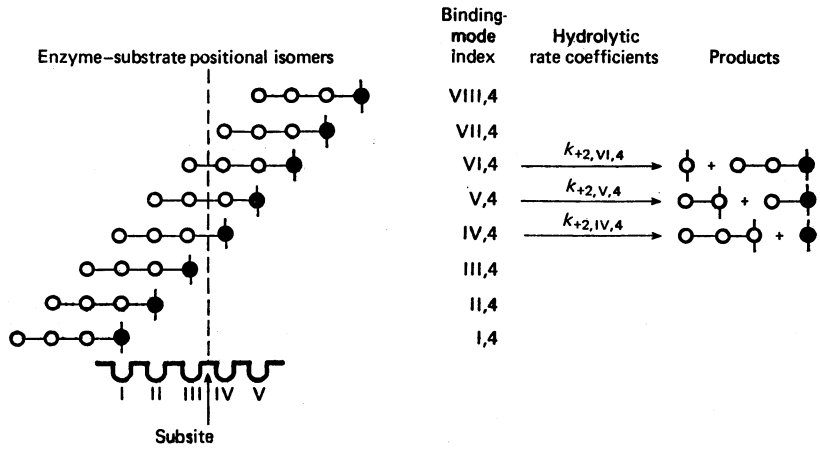


Fig. 1. *Positional isomers of maltotetraose on a hypothetical five-subsite enzyme*

⊔, Subsite on the enzyme; ↑, position of the catalytic site; ○, D-glucopyranoside unit; ⌀, reducing D-glucopyranoside unit; ◖, reducing radiolabelled D-glucopyranoside residue; –, α-(1→4) bond. The broken line is a visual aid to show the position of bond cleavage. The binding-mode index is a Roman number indicating the subsite holding the reducing unit followed by an Arabic number indicating the chain length of the substrate. When the substrate extends beyond the right of the binding region (VI,4; VII,4; VIII,4) virtual subsites are used to designate the binding mode index. The hydrolytic rate coefficients are of the form $k_{2,i,n}$ for chain length n in binding model i.

(Thoma & Allen, 1976; the following paper, Allen & Thoma, 1976) there are inherent difficulties in the approach used by these authors. We (Thoma *et al.*, 1970, 1971) have approached the endoamylase-subsite mapping problem differently, using quantitative bond-cleavage frequencies in conjunction with the chain-length dependence of $\bar{K}_m$ and $\bar{V}$. We have shown (Thoma & Allen, 1976) why this approach is preferred.

The present paper describes further improvements on the techniques used for quantitative subsite mapping. In our early work we (Thoma *et al.*, 1970) made a subjective judgement concerning subsite energy differences used to determine the number of subsites. Also, since subsite energies were not evaluated by a rigorous statistical analysis, disproportionate error was introduced into certain subsites; further, weighting factors were not used to take into account differences in precision of the experimentally determined parameters.

We report here the development of a computer simulation of the depolymerase-subsite model. By using a minimization routine, the computer model is able to predict a subsite map from experimental parameters. A statistical procedure is outlined to evaluate objectively the number of subsites and the position of the catalytic site. The experimental parameters can be weighted to take into account differences in experimental variance. Also, the complete subsite map is evaluated by using all of the experimentally determined parameters.

By using simulated data both with and without an approximation of experimental scatter, we have examined the effectiveness of the minimization algorithm in establishing a subsite map. The experimentally accessible parameters are tested for their ability to produce a unique subsite map. The propagation of experimental error into the predicted subsite energies and the acceleration factor is assessed and the possibility that the acceleration factor arises as an artifact of experimental error is discussed. Finally, a quantitative measure of the overall goodness-of-fit of a subsite map to experimental data is established.

In developing the subsite model, we (Thoma *et al.*, 1971) assumed that the hydrolytic coefficients for the various productive positional isomers were equal, so that the relative rate of bond cleavage was dependent only on the population of the respective productive positional isomers. However, the experimental chain-length dependence of $\bar{K}_{m,n}$ and $\bar{V}_n$ indicated that these hydrolytic coefficients were not a constant, but were a function of the number of filled subsites. When the approximation was made that filling a subsite contributed an average of 1.88 kJ/mol to the lowering of the activation-energy barrier, the experimental and computed parameters were brought into acceptable agreement. This energy has been referred to as 'strain' (Thoma *et al.*, 1971) or the acceleration

Table 1. *Symbols used*

A	Substrate.
b.c.f.	Bond-cleavage frequency.
E	Enzyme.
ΔG	Unitary free energy of binding.
ΔG_a	Acceleration factor.
i	Subsite index.
j	Maximum chain length substrate for which experimental data is available.
K	'Microscopic' dissociation constant.
K'	'Microscopic' association constant.
K_i	Inhibition constant.
K_m	Michaelis constant.
$K_{int.}$	Microscopic dissociation constant for a binding mode in which the entire binding region is occupied.
k_{+1}	'Microscopic' association rate constant.
k_{-1}	'Microscopic' dissociation rate constant.
k_{+2}	Hydrolytic rate coefficient.
l	Number of real subsites comprising the binding region of an enzyme.
m	Chain-length index for product.
min.	Minimum value.
n	Chain-length index for substrate.
P	Product.
Q	Normalized sum of the weighted-squared residuals.
R	Gas constant.
sim.	Simulated data computed from a subsite map.
T	Absolute temperature.
V	Maximum velocity.
W	Weighting factor.
ι	Binding-mode index specifying the real or virtual subsite occupied by the reducing-end glycosyl unit.
σ	Standard error or estimate of standard error.
[ ]	Concentration.
~	Measured or apparent value.
·	Time derivative.
○	α-D-Glucopyranoside unit.
⌀	Reducing α-D-glucopyranoside unit.
⧫	Reducing radioactively labelled α-D-glucopyranoside unit.

factor (Thoma & Allen, 1976). The possibility of the existence of this acceleration factor has led to considerable discussion in the literature (Iwasa *et al.*, 1974; Thoma & Allen, 1976), and will need to be critically re-examined to determine if it is an artifact of the model or corresponds to an actual physical phenomenon [see the following paper, Allen & Thoma (1976)].

Model

Derivations of the equations for subsite mapping and the underlying assumptions have been presented previously (Thoma *et al.*, 1970, 1971; Hiromi, 1970; Nitta *et al.*, 1971; Thoma & Allen, 1976). We will simply present a synopsis of the equations most relevant to the subsite-mapping procedures.

Synopsis of the subsite model

The subsite model can best be characterized by separation of the enzymic processes into 'microscopic' and 'macroscopic' events. The 'microscopic' constants describe the processes associated with one particular binding mode, i.e. positional isomer, of the n-mer substrate (see Fig. 1), whereas the 'macroscopic' constants describe the net interactions of the enzyme with the n-mer substrate. For example, each positional isomer for an n-mer is characterized by a 'microscopic' association constant; and, in addition, each productive positional isomer is characterized by a 'microscopic' hydrolytic coefficient. A measured 'macroscopic' association constant for the n-mer will be some combination of these 'microscopic' association constants, and the measured 'macroscopic' velocity for the n-mer will be some function of the 'microscopic' association constants and 'microscopic' hydrolytic coefficients. The first step in subsite mapping is to derive quantitative relationships for the 'microscopic' constants in terms of the subsite energies, and then to relate these 'microscopic' constants to measurable 'macroscopic' parameters. We will first examine how subsite interactions can be transformed into 'microscopic' constants.

Each subsite on the enzyme is characterized by a free energy of binding to a substrate monomer unit. The free energy of binding for a subsite is assumed to be unaffected by binding, or absence of binding, at other subsites. The subsite-binding energies govern the 'microscopic' association constant for a given positional isomer of n-mer by the relationship

$$\sum_{i=\iota-n+1}^{\iota} \Delta G_i + 10 = -RT \ln K'_{\iota,n} \qquad (1)$$

where ΔG_i is the unitary free energy of binding for subsite i, 10 kJ/mol is the cratic free energy of mixing (Gurney, 1953), and $K_{\iota,n}$ is the 'microscopic' association constant for a substrate of chain length n in binding mode ι. For example in Fig. 1 the microscopic association constant for the tetramer in binding mode V,4 ($K_{V,4}$) is $\exp[-(\Delta G_{II} + \Delta G_{III} + \Delta G_{IV} + G_V + 10)/RT]$.

The 'microscopic' binding constants can be converted to 'macroscopic' constants by consideration of the Michaelis scheme for binding and hydrolysis of a polymeric substrate

$$E + A_n \underset{k_{-1,\iota,n}}{\overset{k_{+1,\iota,n}}{\rightleftarrows}} EA_{\iota,n} \xrightarrow{k_{+2,\iota,n}} E + P_{n-m} + P_m \qquad (2)$$

where E is enzyme, A_n is a substrate of chain length n, $k_{+2,\iota,n}$ is the 'microscopic' hydrolytic coefficient for the positional isomer $EA_{\iota,n}$, and P_{n-m} and P_m are products of chain length $n-m$ and m from the non-reducing- and reducing-end of the substrate molecule respectively. The 'microscopic' Michaelis constant for the $EA_{\iota,n}$ complex in this scheme is $K_{\iota,n} = (k_{-1,\iota,n} + k_{+2,\iota,n})/k_{+1,\iota,n}$. When $k_{-1} \gg k_{+2}$ the inverse of the

Michaelis constant approximates an association constant. The 'microscopic' and 'macroscopic' association constants are related by the expression

$$1/\bar{K}_{m,n} = \sum_{\iota=1}^{\iota=l+n-1} 1/K_{\iota,n} = \sum_{\iota=1}^{\iota=l+n-1} K'_{\iota,n} \qquad (3)$$

where l is the number of subsites on the binding region of the enzyme. Therefore in Fig. 1, the measured $\bar{K}_{m,4}$ will be the inverse of the sum of the eight 'microscopic' association constants as determined from eqn. (1).

The measured maximum velocity for chain length n, $\bar{V}_n$, is determined not only by the 'microscopic' association constants, but also by the 'microscopic' hydrolytic coefficients. This relationship is expressed as

$$\frac{\bar{V}_n}{[E_0]} = \frac{\sum_{\iota=1}^{l+n-1} k_{+2,\iota,n}/K_{\iota,n}}{\sum_{\iota=1}^{l+n-1} 1/K_{\iota,n}} \qquad (4)$$

where the numerator of this fraction now contains only terms for productive positional isomers since $k_{+2,\iota,n} = 0$ for the non-productive binding. In Fig. 1, $\bar{V}$ for the tetramer will depend on the sum of the concentration of productive positional isomers (binding modes IV,4; V,4 and VI,4) times their respective hydrolytic coefficients divided by the sum of the concentration of all eight positional isomers. We see then that the 'macroscopic' parameters $\bar{K}_{m,n}$ and $\bar{V}_n$ are related to the 'microscopic' constants through eqns. (3) and (4); further, these microscopic constants are a function of the subsite energies of binding as described by eqn. (1).

As revealed by Fig. 1, there is another experimentally accessible parameter that will yield information about the subsites. The three productive binding modes for tetramer result in products that are characteristic of a particular positional isomer. The relative rate of formation of each product is proportional to the population of the positional isomer from which that product was formed, multiplied by the hydrolytic rate coefficient for that complex. These relative rates are called bond-cleavage frequencies.

If in eqn. (2) we label the end monomer unit of the oligomer so that we only observe product, P_m, formation, we can express the rate of formation of product from two adjacent binding modes ι and $\iota+1$ as

$$\frac{[\dot{P}_{\iota,m}]}{[\dot{P}_{\iota+1,m+1}]} = \frac{k_{+2,\iota,n}/K_{\iota,n}}{k_{+2,\iota+1,n}/K_{\iota+1,n}} \qquad (5)$$

where $[\dot{P}] = d[P]/dt$ and indicates the binding mode of the n-mer substrate which gives rise to that particular product. For example, in Fig. 1, $[\circ\dot{-}\bullet]/[\circ\text{-}\dot{\circ}\text{-}\bullet] = (k_{+2,V,4}/K_{V,4})/(k_{+2,VI,4}/K_{VI,4})$. A similar expression can be written for the formation of $\bullet$ from tetramer in binding mode IV,4. Since the $k_{2,\iota,n}$ term in the numerator of eqn. (5) goes to zero for non-productive complexes, bond-cleavage frequencies serve only as a probe of productive complexes.

Substrates that can span the entire binding region of the enzyme ($n \geqslant l$) provide a measure of the sum of the energies of interaction for all of the subsites. All binding modes in which the subsites are completely filled have the same dissociation constant, $K_{int.}$. We can write eqn. (3) in terms of $K_{int.}$ by partitioning it into

$$\frac{1}{\bar{K}_{m,n}} = \sum_{\iota=1}^{\iota=l-1} \frac{1}{K_{\iota,n}} + (n-l+1)\left(\frac{1}{\bar{K}_{int.}}\right) + \sum_{\iota=n+1}^{n+l-1} \frac{1}{K_{\iota,n}} \qquad (6)$$

The $K_{int.}$ term accounts for binding modes where the specificity region is completely filled. For example, with a five-subsite enzyme (Fig. 1) for the binding of a hexamer (not illustrated), the first term on the right hand of eqn. (6) accounts for binding modes I–IV in which subsites on the right side of the binding region are vacant; the middle term accounts for binding modes V and VI, in which all subsites are occupied, and the last term accounts for binding modes VII–X, where subsites on the left side of the binding region are vacant. Eqn. (6) has the form of a straight line and by plotting $(n-l+1)$ against $1/K_{m,n}$ for $n \geqslant l$ we obtain a slope of $1/K_{int.}$.

We have developed the relationship of four experimentally accessible parameters: $\bar{K}_m$, $\bar{V}$, bond-cleavage frequencies, and $\bar{K}_{int.}$ to the subsite free energies. Subsite mapping now requires the application of eqns. (1), (3), (4), (5), and (6) to: (1) determine the number of subsites; (2) determine the position of the catalytic site; (3) evaluate the free energies of the individual subsites; (4) determine the hydrolytic coefficients.

Acceleration factor

As we have shown previously (Thoma *et al.*, 1970; Thoma & Allen, 1976) bond-cleavage frequencies (eqn. 5) are the most reliable probe of the number of subsites and the position of the catalytic site. Bond-cleavage frequencies also yield information about subsite-binding energies and the hydrolytic coefficient. By solving eqn. (1) for $K_{\iota,n}$ and substitution into eqn. (5), on re-arrangement we obtain

$$RT \ln \frac{[\dot{P}_{\iota,m}]}{[\dot{P}_{\iota+1,m+1}]} = \Delta G_{\iota+1} - \Delta G_{\iota-n+1} + RT \ln \frac{k_{+2,\iota+1,n}}{k_{+2,\iota,n}} \qquad (7)$$

and hence, bond-cleavage-frequency ratios provide a measure of the difference between two subsite binding energies plus a function of the ratio of hydrolytic coefficients. Consider the two productive binding modes of tetramer (reducing-end labelled $\circ\text{-}\circ\text{-}\circ\text{-}\bullet$)

in Fig. 1, V,4 and VI,4, which yield reducing-end labelled dimer and trimer o–● and o–o–● respectively. In this case $\Delta G_{\iota+1}$ is ΔG_{VI}, and since site VI is a virtual subsite ΔG_{VI} has a zero free energy of binding. Therefore eqn. (7) simplifies to

$$RT\ln\frac{[o\text{–}●]}{[o\text{–}o\text{–}●]} = -\Delta G_{II} + RT\ln\frac{k_{+2,V,4}}{k_{+2,VI,4}} = -\Delta\tilde{G}_{II} \quad (8)$$

where $\Delta\tilde{G}$ is defined as an apparent free energy of binding and is a function of hydrolytic coefficients as well as a subsite-binding energy. By using bond-cleavage frequencies for a series of oligomeric substrates, it is possible to measure $\Delta\tilde{G}$ for every subsite on the enzyme-binding region, with the exception of the two subsites adjacent to the catalytic site (subsites III and IV in Fig. 1) which are occupied by every productive complex. For these apparent energies to be placed on an absolute scale, bond-cleavage frequencies for substrates which extend beyond the end of the binding region must be used.

If the hydrolytic coefficients are equal [e.g. in eqn. (7), $k_{+2,\iota,n} = k_{+2,\iota+1,n}$] then the apparent binding energies are equivalent to the true binding energies (i.e. $\Delta G_\iota = \Delta\tilde{G}_\iota$). When Thoma *et al.* (1971) made this approximation, and used the resulting binding energies in eqns. (1), (3) and (4) to predict $\tilde{K}_{m,n}$ and $\tilde{V}_n$, the fit between the calculated and experimental parameters was unacceptable. Further, the variance was not random (i.e. the residuals were not normally distributed) as if arising from experimental error, but was systematic, which is diagnostic of a poor model (Mannervik & Bártfai, 1973).

Thoma *et al.* (1971) found that the experimental and calculated values could be brought into good agreement if the approximation was made that binding of a substrate monomer unit lowers the activation-energy barrier by a constant amount. That is, that $k_{+2,\iota,n}$ is a function of the number of real filled subsites for binding mode ι. This observation is expressed mathematically as

$$k_{+2,\iota,n} \propto \exp\left(\sum_{i=\iota}^{\iota-n+1} \Delta G_{a,i}/RT\right) \quad (9)$$

where $\Delta G_{a,i}$ is the contribution of the ith real subsite to the acceleration of bond cleavage. Substitution of eqn. (9) into eqn. (8) gives

$$\Delta G_{II} = \Delta\tilde{G}_{II} + \Delta G_{a,\iota} \quad (10)$$

so that $\Delta\tilde{G}_\iota$ measured by the bond-cleavage frequencies must be increased by the factor $\Delta G_{a,II}$, the acceleration factor. By a minimization technique, $\Delta G_{a,\iota}$ was approximated as 1.88 kJ/mol. It is unlikely that $\Delta G_{a,\iota}$ is a constant for each filled subsite, and 1.88 kJ/mol is obviously an average value, which was found to account for the experimental data (Thoma *et al.*, 1971). We re-examine this approximation in the following paper (Allen & Thoma, 1976).

We have set forth the basic equations necessary for subsite mapping. In the next section we will show how these equations can be applied to experimental data through the use of a computer minimization model, and in the discussion we will develop the regime for subsite mapping and assess the errors of the technique.

Depolymerase computer model

A computer model of a depolymerase enzyme was developed by using eqns. (1), (3)–(6), and (11). A flow diagram of the program is given in Fig. 2 and a listing of the program with sample data is given by Allen (1975).* On the basis of our experience in subsite mapping, we have found that three different options are valuable. These options were incorporated into the computer model as MOD1, MOD2 and MOD3. 1. MOD1 accepts as input data the number of subsites, the position of the catalytic site and the subsite energies; the output gives calculated $K_{m,n}$, V_n, bond-cleavage frequencies and $K_{int.}$ for the input map. 2. MOD2 accepts the same input information as MOD1 in addition to any experimentally measured $\tilde{K}_{m,n}$, $\tilde{V}_n$, bond-cleavage frequencies, and $\tilde{K}_{int.}$; the output gives calculated $K_{m,n}$, V_n, bond-cleavage frequencies and $K_{int.}$ and a comparison of the computed and experimental parameters by using the criterion that will be established below. 3. MOD3 is the heart of the computer model. The input data are the experimentally determined parameters $\tilde{K}_{m,n}$, $\tilde{V}_n$, bond-cleavage frequencies and $\tilde{K}_{int.}$, in addition to the number of subsites, the position of the catalytic site and initialized subsite binding energies (see below). The output gives the optimum subsite energies consistent with the specified size and position of the catalytic site. This analysis is performed by using a minimization routine described below.

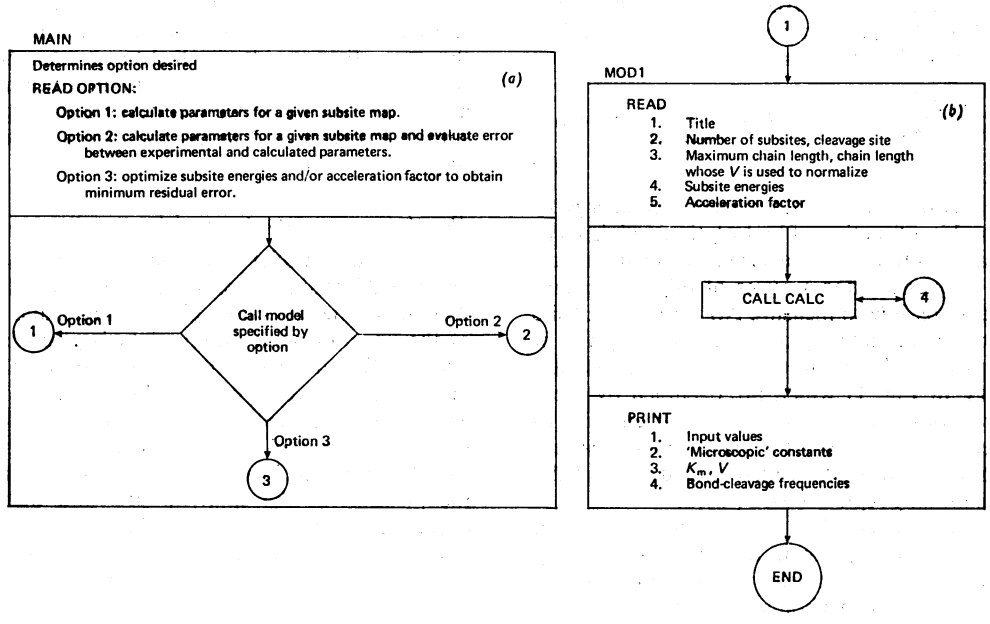
Minimization

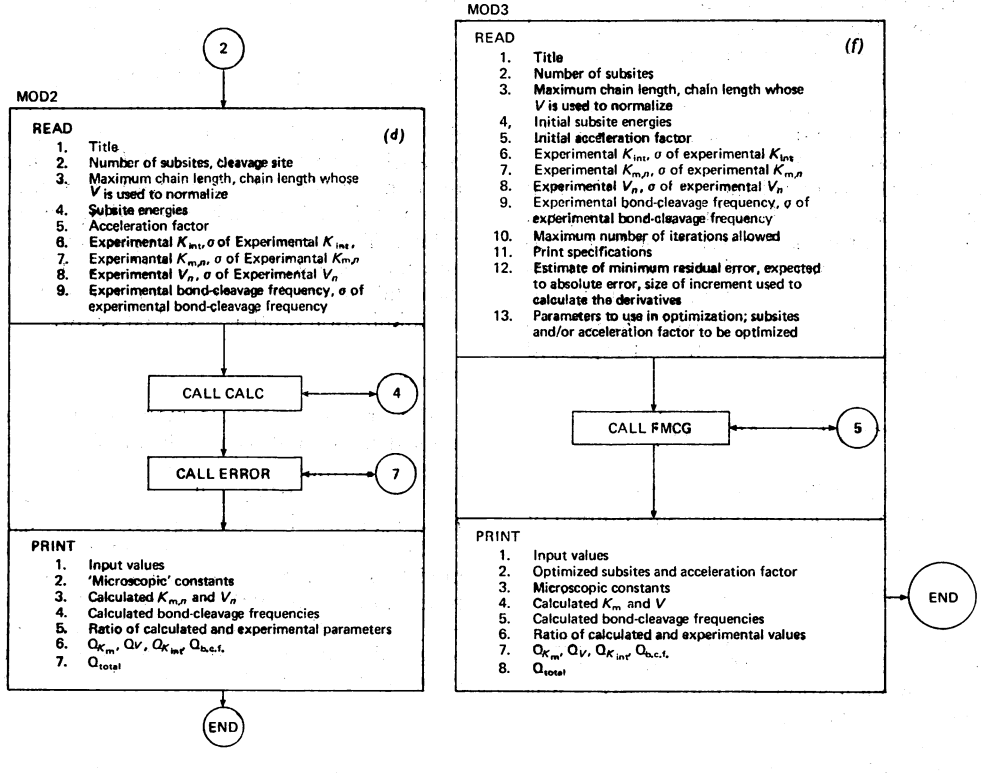
The process of subsite mapping, from a mathematical standpoint, is a problem of optimization. A subsite map must be generated that gives an optimum fit to the experimentally measured parameters: $\tilde{K}_{m,n}$, $\tilde{V}_n$, bond-cleavage frequencies and $\tilde{K}_{int.}$.

The first task is to establish a measure of fit of the model to the experimental data. The difference between an experimental and computed value ($X_{J,exptl.} - X_{J,calcd.}$) is the residual. We will use the sum of the weighted squared residuals as a measure of fit between experimental and calculated values or

$$Q = \sum_J W_J (X_{J,exptl.} - X_{J,calcd.})^2 \quad (11)$$

* Additional details and a copy of the program will be supplied by the author on request.


MAIN
Determines option desired
READ OPTION:
(a)
Option 1: calculate paramaters for a given subsite map.
Option 2: calculate parameters for a given subsite map and evaluate error between experimental and calculated parameters.
Option 3: optimize subsite energies and/or acceleration factor to obtain minimum residual error.
Call model specified by option
Option 1
Option 2
Option 3
1
2
3
MOD1
READ
(b)
1. Title
2. Number of subsites, cleavage site
3. Maximum chain length, chain length whose V is used to normalize
4. Subsite energies
5. Acceleration factor
CALL CALC
4
PRINT
1. Input values
2. 'Microscopic' constants
3. K_m, V
4. Bond-cleavage frequencies
END


2
MOD2
READ
(d)
1. Title
2. Number of subsites, cleavage site
3. Maximum chain length, chain length whose V is used to normalize
4. Subsite energies
5. Acceleration factor
6. Experimental K_{int}, σ of Experimental K_{int}
7. Experimantal $K_{m,n}$, σ of Experimantal $K_{m,n}$
8. Experimental V_n, σ of Experimental V_n
9. Experimental bond-cleavage frequency, σ of experimental bond-cleavage frequency
CALL CALC
4
CALL ERROR
7
PRINT
1. Input values
2. 'Microscopic' constants
3. Calculated $K_{m,n}$ and V_n
4. Calculated bond-cleavage frequencies
5. Ratio of calculated and experimental parameters
6. Q_{K_m}, Q_V, $Q_{K_{int}}$, $Q_{b.c.f.}$
7. Q_{total}
END
MOD3
READ
(f)
1. Title
2. Number of subsites
3. Maximum chain length, chain length whose V is used to normalize
4. Initial subsite energies
5. Initial acceleration factor
6. Experimental K_{int}, σ of experimental K_{int}
7. Experimental $K_{m,n}$, σ of experimental $K_{m,n}$
8. Experimental V_n, σ of experimental V_n
9. Experimental bond-cleavage frequency, σ of experimental bond-cleavage frequency
10. Maximum number of iterations allowed
11. Print specifications
12. Estimate of minimum residual error, expected to absolute error, size of increment used to calculate the derivatives
13. Parameters to use in optimization; subsites and/or acceleration factor to be optimized
CALL FMCG
5
PRINT
1. Input values
2. Optimized subsites and acceleration factor
3. Microscopic constants
4. Calculated K_m and V
5. Calculated bond-cleavage frequencies
6. Ratio of calculated and experimental values
7. Q_{K_m}, Q_V, $Q_{K_{int}}$, $Q_{b.c.f.}$
8. Q_{total}
END

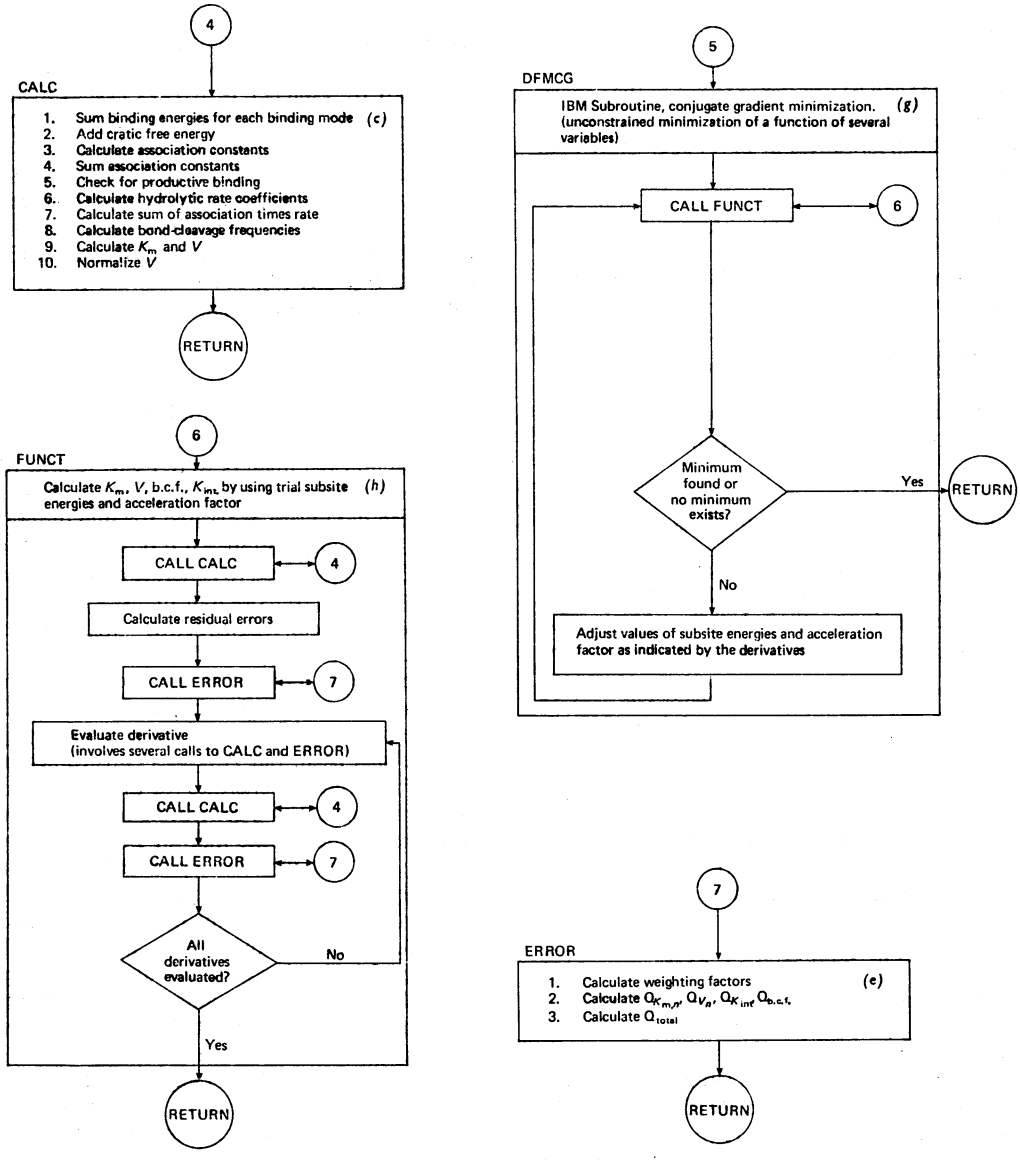


Fig. 2. *Depolymerase computer model: flow diagram*

The general logic of the depolymerase computer program is presented as a flow diagram; a complete Fortran listing is available with test input and output data (Allen, 1975). There are three available options: MOD1, MOD2 and MOD3. MOD1 is a subroutine that computes Michaelis parameters, K_m and V and bond-cleavage frequencies as a function of substrate-chain length. MOD1 accomplishes this computation by calling subroutine CALC. MOD2 is a subroutine that computes Michaelis parameters and bond cleavage frequencies as a function of substrate-chain length and compares these computed parameters with experimental values by calling subroutine ERROR. MOD3 is the minimization routine. Minimization of the sum of the squared residuals is accomplished by IBM subroutine DFMCG (IBM, 1970), which calls subroutine FUNCT to calculate Michaelis parameters, bond-cleavage frequencies, and to evaluate the derivative used in the search for the minimum.

where j is summed over the experimentally available data. W_j is the weighting factor, which will serve to give greater weight to the experimental parameters that are known with more certainty. The Q value can be normalized by dividing by j, the number of data points. Eqn. (11) can be written in terms of the experimentally measurable parameters to give four equations that establish goodness-of-fit for each individual model parameter as

$$Q_{K_m} = \sum_{n=1}^{j} W_{k_{m,n}} (\tilde{K}_{m,n,\text{exptl.}} - K_{m,n,\text{calcd.}})^2 \quad (12)$$

where j is the maximum chain length for which experimental $\tilde{K}_{m,n}$ values are available,

$$Q_V = \sum_{n=1}^{j} W_{v_n} (\tilde{V}_{n,\text{exptl.}} - V_{n,\text{calcd.}})^2 \quad (13)$$

where j is the maximum-chain-length substrate for which experimental $\tilde{V}_n$ values are available,

$$Q_{\text{b.c.f.}} = \sum_{n=3}^{j} \sum_{m=1}^{n-1} W_{\text{b.c.f.}_{n,m}} (\text{b.c.f.}_{n,m,\text{exptl.}} - \text{b.c.f.}_{n,m,\text{calcd.}})^2 \quad (14)$$

where $\text{b.c.f.}_{n,m}$ is the bond-cleavage frequency for chain length n, forming a product of chain length m, where j is the maximum-chain-length substrate for which bond-cleavage frequencies are available, and

$$Q_{K_{\text{int.}}} = W_{K_{\text{int.}}} (K_{\text{int.exptl.}} - K_{\text{int.calcd.}})^2 \quad (15)$$

The weighting factor is defined as the reciprocal of the experimental variance

$$W_j = 1/\sigma_j^2 \quad (16)$$

The goodness-of-fit of the model can be defined in terms of Q_{K_m} (eqn. 12), Q_V (eqn. 13), $Q_{\text{b.c.f.}}$ (eqn. 14), and $Q_{K_{\text{int.}}}$ (eqn. 15). All of the terms can be used to establish

$$Q_{\text{total}} = Q_{K_m} + Q_V + Q_{\text{b.c.f.}} + Q_{K_{\text{int.}}} \quad (17)$$

All of the Q values reported in the present paper have been normalized by division of Q by the number of data-point residuals constituting that Q value. In the present study the maximum chain length examined is twelve, so that Q_{K_m} (eqn. 12) is divided by 12, Q_V (eqn. 13) is divided by 10, $Q_{\text{b.c.f.}}$ (eqn. 14) is divided by 65, and Q_{total} is divided by 88. Note that the normalized Q_{total} does not necessarily equal the sum of the normalized Q values.

If Q is a measure of the goodness-of-fit of the model then the best fit is established when Q is a minimum, $Q_{\text{min.}}$, and the resulting map is the optimum map. Further, the value of $Q_{\text{min.}}$ for a subsite map is a measure of how well that particular map is able to account for the experimental data.

We tested several minimization procedures for their applicability to the subsite mapping problem. The various non-linear optimization techniques have been reviewed by Swann (1969). Two unconstrained gradient methods of minimization were examined. 1. A method originally proposed by Davidon (1959) and refined by Fletcher & Powell (1963), which is available in the IBM Scientific Subroutine Package as DFMFP (IBM, 1970). 2. A conjugate gradient method proposed by Fletcher & Reeves (1964) available in the Scientific Subroutine Package as DFMCG (IBM, 1970). Of these two gradient methods the conjugate gradient method (DFMCG) gave slightly more rapid convergence and was used throughout this work (Fig. 2). The derivatives were calculated by means of an interpolation formula that uses differences (Scarborough, 1962). All minimization calculations were done in double precision.

Simulation studies

In order to test the capabilities and the limitations of the computer minimization model, simulated experimental parameters were used. A flow diagram of the simulation studies is given in Fig. 3. The simulated data were generated for two different subsite maps using option MOD1 (discussed above). The subsite maps used to generate this simulated data will be referred to as parent maps. The subsite map for *Bacillus amyloliquefaciens* enzyme reported previously (Thoma *et al.*, 1971) was used as a basis for the two parent maps. In order to simulate a hypothetical enzyme that has a constant hydrolytic coefficient, the subsite map for *B. amyloliquefaciens* enzyme was used, but ΔG_a was constrained at zero. The subsite binding energies for this present map are given in Table 2. For a generating parent map with an acceleration factor, the map of *B. amyloliquefaciens* enzyme was used with the reported ΔG_a, 1.88 kJ/mol. The apparent binding energies (i.e. containing the acceleration factor) for this parent map are listed in Table 3. For both of these maps, $K_{m,n}$, V_n, and bonds cleavage frequencies were calculated for substrates up to chain length twelve and a $K_{\text{int.}}$ for the map was computed. These generated parameters will be referred to as simulated data without error (Fig. 3).

To examine the effect of experimental variance on the minimization procedure, an approximation of experimental scatter was incorporated into the simulated data. Program GAUSS in the IBM Scientific Subroutine Package (IBM, 1970) generates random numbers that conform to a normal distribution; so that, given a set of numbers and the expected experimental error in these numbers, the routine can incorporate experimental scatter into the numbers. To apply this program it was assumed that K_1 for glucose had 20% error and that the remaining Michaelis parameters ($n = 2$–12) had 10% error (Cleland, 1963). Each bond-cleavage frequency was assumed to have a constant standard error of 0.03. The resulting data are referred to as simulated data with error (Fig. 3).

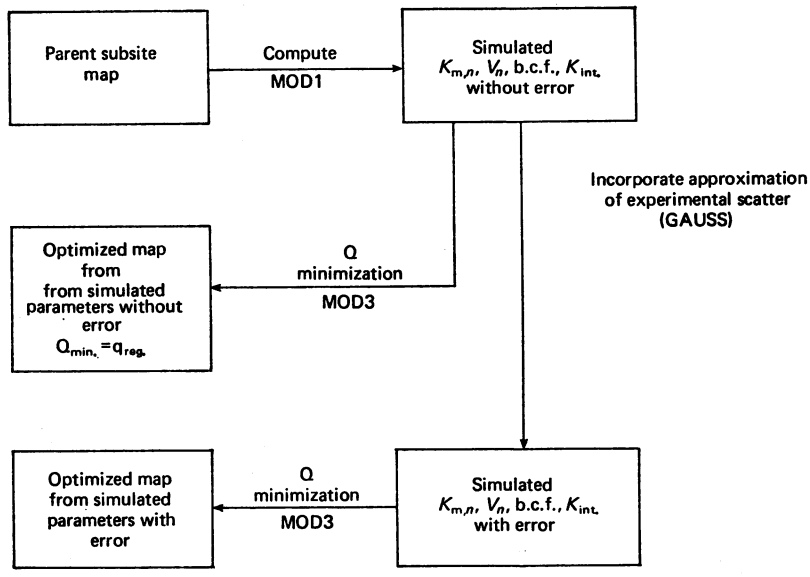


Fig. 3. *Flow diagram of the simulation regime followed in the simulation studies to test the effectiveness of the depolymerase computer model*

Two different parent maps were used. The parent subsite map in Table 2 was used as an enzyme with a constant hydrolytic rate coefficient ($\Delta G_a = 0$), and the parent subsite map in Table 3 was used as an enzyme with variable hydrolytic coefficients ($\Delta G_a > 0$). Both parent maps were used in MOD1 of the computer model to calculate Michaelis parameters, $K_{m,n}$ (eqn. 3), V_n (eqn. 4), bond-cleavage frequencies (b.c.f., eqn. 5) (for $n = 1$–12) and $K_{int.}$ (eqn. 6) to give simulated data without any experimental scatter. The data were made to approximate more closely experimentally measured parameters by incorporation of random scatter by using subroutine GAUSS. These simulated parameters were used in MOD3 of the computer model to regenerate a subsite map. The results are given in Tables 2 and 3.

Results and Discussion

Before the depolymerase computer model can be applied to experimental data, it must be examined for its reliability in establishing a subsite map. Various points must be considered. (1) Can the minimization model re-establish the parent subsite map with simulated data? (2) What effect does experimental variance have on the resulting subsite map? (3) What effects do the initialized values of subsite energies have on the minimum reached? (4) What is the most effective regime to follow in establishing a subsite map from experimental data? (5) Can the minimization model be used to establish the number of subsites and the position of the catalytic site? (6) What factors are reflected in the measure of goodness-to-fit, $Q_{min.}$? (7) Can experimental variance be responsible for the proposed acceleration factor, ΔG_a, or does ΔG_a reflect a real physical phenomenon? Below we will explore the application of the computer model to simulated data and in the following paper (Allen & Thoma, 1976) we show the application to experimental data.

Effectiveness of minimization

The simulated data offer a means of evaluating the effectiveness of the minimization model in establishing the absolute minimum and consequently the 'best' subsite map. As will be shown below, an error surface established by Q may have relative local minima in addition to the absolute minimum established where Q has reached its smallest possible value. In contrast with actual experimental data where the 'true' subsite map is not known, the subsite map predicted from the simulated data can be directly compared to the parent subsite binding energies that were used to compute the data. As outlined in the Model section and Fig. 3, the binding energies given in Table 2 (parent map) for a ten-subsite enzyme without an acceleration factor (i.e. $\Delta G_a = 0$) were used to generate $K_{m,n}$, V_n, bond-cleavage frequencies, and $K_{int.}$ for chain lengths up to twelve. A similar map with an acceleration factor (Table 3, parent map) was used to examine the effectiveness of the minimization model for an enzyme in which the hydrolytic rate coefficients are not constant, i.e. $\Delta G_a > 0$. The

Table 2. *Subsite map for a depolymerase with a constant hydrolytic coefficient*

The parent subsite map was adapted from the map proposed for *B. amyloliquefaciens* α-amylase (Thoma *et al.*, 1971) to test the computer model. ---- indicates the position of the catalytic site. The true binding energies ΔG_i, were used with the hydrolytic rate coefficients held constant (i.e. $\Delta G_a = 0$) as a parent map to generate simulated data with added experimental scatter as outlined in Fig. 3. The optimized map is the subsite binding energies predicted when a minimum was reached ($Q_{min.total}$) in the minimization routine (MOD3). Subsites I–V and VIII–X were initialized at a binding energy of zero and subsites VI and VII were initialized at 21 and −21 kJ/mol respectively. The $Q_{min.}$ values were calculated from eqns. (12)–(17).

		Binding energies (kJ/mol)	
Subsite no.		Parent map	Optimized map
I		−3.31	−3.22
II		−7.95	−8.41
III		0.59	1.21
IV		−2.51	−2.59
V		−7.11	−7.15
VI		12.55	12.55
VII		−11.05	−11.42
VIII		−5.43	−5.36
IX		−2.51	−2.34
X		6.69	6.74
ΔG_a (kJ/mol)		0	0
			Normalized sum of squared residual error
$Q_{min.}$	b.c.f.		0.9
	K_m		2.0
	V		0.9
	$K_{int.}$		0.1
	Total		1.0

minimization model requires as input data the number of subsites, the position of the catalytic site, and initialized subsite-binding energies where the search for a minimum will begin. The model then establishes the optimum subsite map within these established constraints. We show below how the number of subsites and the position of the catalytic site can be established from bond-cleavage frequencies; for the present we will assume that these values have been determined. The only remaining variable input is the initialized subsite-binding energies.

Subsite map with constant hydrolytic coefficients

We will first consider the case of a depolymerase where the hydrolytic coefficients are constant so that the acceleration factor has a value of zero. The parent subsite map is shown in Table 2.

Table 3. *Subsite map for a depolymerase with an acceleration factor*

The parent subsite map is the map proposed for *B. amyloliquefaciens* α-amylase with an acceleration factor of 1.88 kJ/mol (Thoma *et al.*, 1971). ---- shows the position of the catalytic site. Since the enzyme has an acceleration factor, apparent subsite binding energies are given that contain a rate coefficient term (eqn. 8). The optimized subsite maps were obtained by using the simulated parameters, with an approximation of experimental scatter, generated from the parent map. $Q_{min.}$ values were obtained by comparison of simulated parameters with parameters calculated from the optimized maps (see Fig. 3).

		Apparent binding energies (kJ/mol)		
			Optimized maps	
Subsite no.		Parent map	ΔG_a constrained at zero*	ΔG_a allowed to vary†
I		−5.19	−5.40	−5.10
II		−9.83	−10.17	−9.92
III		−1.30	−1.38	−1.00
IV		−4.39	−4.48	−4.18
V		−9.00	−8.83	−8.66
VI		10.67	22.51	11.59
VII		−12.93	−4.52	−12.89
VIII		−7.32	−7.41	−7.20
IX		−4.39	−4.69	−4.27
X		4.81	4.64	4.98
ΔG_a (kJ/mol)		1.88	0	1.67
			Normalized sum of squared residual error	
$Q_{min.}$	b.c.f.		0.9	1.0
	K_m		50.2	1.5
	V		15.9	0.6
	$K_{int.}$		0.9	0.5
	Total		9.3	1.0

* Subsites I–V and VIII–X were optimized by using $Q_{b.c.f.}$, and subsites VI and VII were optimized with $(Q_{K_m}+Q_V+Q_{K_{int.}})$; k_{+2} was forced to remain constant (i.e. $\Delta G_a = 0$).

† All of the subsite energies and ΔG_a were optimized to obtain $Q_{min.total}$.

Minimization with bond-cleavage frequencies. Bond-cleavage frequencies offer a means of assessing the apparent binding energies of the subsites with the exception of the two subsites adjacent to the catalytic site (subsites VI and VII in Table 2). The binding energies of these two subsites do not have any influence on bond-cleavage frequencies. With an enzyme where the hydrolytic coefficient is constant, these apparent binding energies are uncomplicated by the k_{+2} terms (eqn. 7) and are actual substrate–monomer-binding energies. When the energies of

Table 4. *Comparison of Michaelis parameters calculated from a local minimum*

The simulated Michaelis parameters were calculated from the parent map in Table 2 and an approximation of experimental scatter was added. The calculated Michaelis parameters used to compute the ratio under 'Local minimum map' were obtained from the optimized map where subsites VI and VII were established, through a local minimum, at −9.0 and 10.9 kJ/mol respectively. The calculated Michaelis parameters used to compute the ratio under 'Absolute minimum map' were obtained from the optimized subsite map when an absolute minimum was reached. The map is shown in Table 2 (optimized map).

	Simulated K_m/calculated K_m		Simulated V/calculated V	
Substrate chain length	Local minimum map	Absolute minimum map	Local minimum map	Absolute minimum map
1	0.9	1.6	—	—
2	0.5	0.8	0.4	1.0
3	0.8	1.2	0.4	1.1
4	1.4	0.9	0.7	0.8
5	4.9	0.6	3.8	0.8
6	18.8	0.9	10.4	0.9
7	10.5	1.1	5.6	1.0
8	3.4	1.0	2.3	1.3
9	1.8	0.9	1.4	1.2
10	1.9	1.0	1.3	1.2
11	1.8	1.0	1.3	1.2
12	1.6	0.9	1.0	1.0

subsites I–V and VII–X were initialized at 0 kJ/mol and optimized by using simulated bond-cleavage frequencies without error (eqn. 14), i.e. to reach $Q_{min.,b.c.f.}$, the optimized subsite energies were within ±0.013 kJ/mol of the parent subsite map. The minimized value of Q, $Q_{min.,b.c.f.}$ was 9.0×10^{-5}. The absolute minimum was always reached when optimizing with $Q_{b.c.f.}$ regardless of the initialized energies of the subsites. For example, when subsites I–V and VIII–X were initialized at +21 or −21 kJ/mol the same minimum was reached. Of course, the number of iterations necessary, and consequently the computation time, was decreased considerably when the initialized subsite-binding energies were close to the optimum subsite-binding energies. The observation that the absolute minimum is consistently obtained suggests that the error space created by $Q_{b.c.f.}$ is a smooth surface without relative local minima. Therefore there are no equivalent subsite maps that can account for the bond-cleavage frequencies; consequently bond-cleavage frequencies cannot predict a false map by becoming trapped in relative local minimum. We will show below that the error surface created by $(Q_{K_m} + Q_V)$ can contain a relative local minimum.

The simulated bond-cleavage frequencies with an approximation of experimental scatter were tested in a similar manner for convergence properties. As in the case for the perfect data, the final map was insensitive to the initial values of subsite binding energies. Of course, the experimental scatter introduced into the bond-cleavage frequencies will be propagated as error into the optimized subsite binding energies. In this case, with a constant standard error of ±0.03 in the simulated bond-cleavage frequencies approximating experimental precision, the optimized subsite energies were within ±0.059 kJ/mol of the parent subsite energies. The experimental scatter is also reflected in the minimized value of $Q_{b.c.f.}$; in this case $Q_{min.\ b.c.f.} = 0.9$.

Minimization with Michaelis parameters (establishment of a local minimum). The Michaelis parameters $K_{m,n}$ and V_n contain information about non-productive binding modes; consequently, they can be used to provide the binding energies of the two subsites adjacent to the catalytic site, subsites VI and VII, whose energies are inaccessible through bond-cleavage-frequency data. In addition, $K_{m,n}$ and V_n provide the information required to assess ΔG_a. Restraining subsites I–V and VIII–X at the energies predicted by bond-cleavage frequencies, subsites VI and VII can be optimized to accommodate the $K_{m,n}$ and V_n data. The sum of the binding energies for the entire span of subsites is set by $K_{int.}$. Thus subsites VI and VII are optimized as dictated by $(Q_{K_m} + Q_V + Q_{K_{int.}})$.

With subsites I–V and VIII–X set by bond-cleavage frequencies and subsites VI and VII initialized at 0 kJ/mol, subsites VI and VII were optimized to obtain $(Q_{min.,K_m} + Q_{min.,V} + Q_{min.,K_{int.}})$ by using simulated data without error from the parent map in Table 2. A relative local minimum was established with subsites VI and VII equal to −8.45 and 11.67 kJ/mol respectively and $Q_{min.,K_m} + Q_{min.,V} + Q_{min.,K_{int.}} = 52.6$. Hence, in contrast with the surface established by $Q_{b.c.f.}$, the error surface determined by Michaelis

parameters has at least one local minimum. The minimization routine has reached one of these relative minimums and cannot get out of the local valley to reach the absolute minimum. As pointed out by Fletcher & Reeves (1964), 'the best that can reasonably be expected is that the minimization process will lead as quickly as possible to the bottom of whatever valley it starts in'. Therefore greater care must be used when minimizing with Michaelis parameters rather than bond-cleavage frequencies.

The Michaelis parameters computed from the subsite energies of a local minimum map are compared in Table 4 with the simulated parameters and reveal a means of determining when a local minimum has been reached. One criterion of how well a model accounts for experimental data is the trends of the residuals (Mannervik & Bártfai, 1973). When the local minimum is reached, the ratios of simulated parameters to Michaelis parameters calculated from the relative local minimum map show definite trends, whereas the comparison of simulated parameters with the Michaelis parameters calculated from the subsite map that has reached an absolute minimum (Table 4) shows a more random distribution of error.

A relative local minimum can be avoided by using different initialized values for the subsite energies so that the valley established by the local minimum is not encountered in the minimization. In this case, when subsites VI and VII were initialized at +21 and −21 kJ/mol respectively, the absolute minimum was found. For the simulated data with experimental scatter, subsites VI and VII were predicted to be 12.26 and −11.88 kJ/mol respectively, which is within ±0.830 kJ/mol of the parent values.

Minimization of Q_{total}. A final refinement of the subsite map can be achieved by allowing all of the subsites to vary to establish the best fit as dictated by all of the experimentally accessible parameters, i.e. by minimization of Q_{total} (eqn. 17). This final optimized map is shown in Table 2 for the $\Delta G_a = 0$ model optimized from simulated data with error. Binding energies are predicted within ±0.620 kJ/mol of the parent binding energies.

To determine if experimental variance can erroneously introduce an acceleration factor, ΔG_a was allowed to vary by using simulated data, with error, of the parent map in Table 2, where $\Delta G_a = 0$. All of the subsite-binding energies and ΔG_a were optimized to establish $Q_{min.,total}$. When the minimum was reached, ΔG_a was computed to be −0.20 kJ/mol. A negative value for ΔG_a has no physical meaning (see the Model section) and is simply an artifact of the experimental scatter. The improvement in $Q_{min.,total}$ was minimal, from 1.0 when ΔG_a was forced to be zero to 0.9 when ΔG_a was optimized. Therefore an acceleration factor that is found to differ from zero by less than about 0.20 kJ/mol will be subject to suspicion.

Subsite map with an acceleration factor

The parent subsite map with $\Delta G_a = 1.88$ kJ/mol is shown in Table 3. The subsite energies are the apparent binding energies, $\Delta \tilde{G}_i$, and are transformed to actual substrate-monomer binding energies by addition of 1.88 kJ/mol to each apparent subsite-binding energy.

Minimization with bond-cleavage frequencies. When a depolymerase does not have constant hydrolytic rate coefficients then the binding energies predicted by bond-cleavage frequencies have a term containing hydrolytic coefficients (eqn. 7) and are apparent binding energies ($\Delta G_i - \Delta G_{a,i}$). As in the case of an enzyme without an acceleration factor, minimization of $Q_{b.c.f.}$ by using simulated data without error established the apparent subsite energies of subsites I–V and VIII–X of the parent map within ±0.013 kJ/mol. By using the data with experimental scatter, subsites I–V and VIII–X of the parent map were computed with an error of ±0.34 kJ/mol (Table 3, column 2). The established minimum was insensitive to the initialized values of apparent subsite-binding energies, again showing the absence of relative local minima in $Q_{b.c.f.}$.

Minimization with Michaelis parameters. To determine how well the data could be fit by forcing the hydrolytic coefficients to be constant ($\Delta G_a = 0$), subsites VI and VII were optimized (by using $Q_{K_m} + Q_V + Q_{K_{int}}$) with ΔG_a constrained at zero. Subsites I–V and VIII–X were set as established by the bond cleavage-frequency minimization. The optimized energies for these two sites are given in Table 3, column 2. The resulting $Q_{min,total} = 9.3$. The Michaelis parameters predicted by this map are compared in Table 5, columns 2 and 5 with the simulated Michaelis parameters from the parent subsite map (Table 3). It is apparent that the fit is poor; further, the residuals are obviously not randomly distributed, i.e. a trend is apparent in the ratio of the Michaelis parameters. This in fact is the same data trend which originally led to the proposal of an acceleration factor from the data of *Bacillus amyloliquefaciens* enzyme (Thoma *et al.*, 1971).

When the map was optimized by allowing ΔG_a to vary, the improvement in the fit was dramatic. All of the subsite-binding energies and ΔG_a were optimized by using Q_{total} to achieve the map in Table 2, column 3. $Q_{min,total}$ was decreased from 9.3 to 1.0, and the subsite energies are within ±0.34 kJ/mol of the parent energies, except subsite VI, which is 0.92 kJ/mol too high. The optimized accelerator factor is 1.67 kJ/mol, which is 0.21 kJ/mol lower than the parent ΔG_a. It is obvious from the Michaelis parameters predicted in Table 5, columns 3 and 4, that the fit has been drastically improved. We conclude that the subsite-binding energies can be predicted within about ±0.920 kJ/mol of the true value

Table 5. *Comparison of Michaelis parameters measured from a subsite map with $\Delta G_a = 1.88$ kJ/mol with those determined from optimized maps*

In order to approximate experimental data, simulated Michaelis parameters were computed from the parent subsite map of Table 3 where $\Delta G_a = 1.88$ kJ/mol and an approximation of experimental scatter was incorporated into the computed values. These simulated values were used in the minimization routine to obtain the optimum subsite map, where ΔG_a was constrained at zero and where ΔG_a was optimized.

	K_m or K_I (M)			V (normalized)*		
Substrate chain length	Simulated K_m (ΔG_a = 1.88 kJ/mol)	Simulated K_m/ calculated K_m (ΔG_a = 0 kJ/mol)†	Simulated K_m/ calculated K_m (ΔG_a = 1.67 kJ/mol)‡	Simulated V (ΔG_a = 1.88 kJ/mol)*	Simulated V/ calculated V (ΔG_a = 0 kJ/mol)†	Simulated V/ calculated V (ΔG_a = 1.67 kJ/mol)‡
1	5.03×10^{-1}	1.3	1.4		—	—
2	3.96×10^{-2}	0.8	0.8	3.69×10^{-6}	0.7	0.9
3	2.41×10^{-2}	1.4	1.2	9.00×10^{-5}	0.9	1.1
4	3.63×10^{-2}	18.8	1.1	2.15×10^{-3}	10.8	1.0
5	8.84×10^{-3}	31.4	0.8	6.50×10^{-3}	25.5	1.1
6	1.06×10^{-2}	37.7	1.1	3.16×10^{-2}	19.5	1.1
7	5.71×10^{-3}	20.3	1.1	1.71×10^{-1}	8.8	1.1
8	1.62×10^{-3}	5.9	1.0	5.53×10^{-1}	2.9	1.2
9	7.62×10^{-4}	3.0	1.0	1.04	1.4	1.2
10	7.86×10^{-4}	3.1	1.0	1.09	1.3	1.2
11	7.33×10^{-4}	3.0	1.0	1.16	1.3	1.2
12	6.59×10^{-4}	2.7	0.9	1.00	1.0	1.0

* Maximum velocities were normalized to the V of chain length 12.

† A comparison of the simulated Michaelis parameters in column (1) to those calculated by an optimization with ΔG_a constrained at zero (Table 3, column 2).

‡ A comparison of the simulated parameters in column 1 with those calculated by an optimization where ΔG_a was optimized in addition to the subsite energies. The optimum acceleration factor achieved was −1.67 kJ/mol (Table 3, column 3).

and the acceleration factor can be predicted within about ±0.21 kJ/mol.

Evaluation of the number of subsites and the position of the catalytic site. We have previously shown (Thoma & Allen, 1976) that the correct way to determine the number of subsites comprising the binding region and to locate the catalytic site is through the use of quantitative bond-cleavage frequencies for polymeric substrates that are long enough to span the entire binding region. Thoma *et al.* (1970) outlined a procedure for applying bond-cleavage frequencies to the task of measuring the number of subsites and locating the catalytic site.

Our procedure (Thoma *et al.*, 1970) used selected experimental bond-cleavage frequencies to determine apparent relative free energies of the subsites. The actual determination of the number of interacting subsites from these relative apparent free energies involved a visual inspection and a subjective judgement as to when differences in the binding energies became insignificant.

The depolymerase computer model can be used to determine the number of subsites objectively. In addition, all of the experimental bond-cleavage frequencies can be used, and weighting factors can be used to take into account differences in experimental precision. Since the value of $Q_{min.}$ is a criterion of the goodness-of-fit of a given subsite map, $Q_{min.,b.c.f.}$ offers a means to evaluate the correct number of subsites necessary to account for experimental bond-cleavage-frequency data. Table 6 shows the application of $Q_{min.,b.c.f.}$ to simulated bond-cleavage frequencies with added error.

In Table 6 selected subsites are optimized (indicated by X) and the remaining subsite energies are constrained to zero (i.e. no interaction with substrate). In each case the best fit is obtained, as measured by $Q_{min.,b.c.f.}$, for the allowed number of subsites. For example, in minimization number 1 when only subsites IV, V, VIII and IX are allowed to vary in the optimization, the best fit to the bond-cleavage frequencies has a $Q_{min.,b.c.f.} = 34$. Whenever a real subsite is added (minimizations 2–4, 6, 7), an improvement in the fit results, until the correct number of subsites is reached in minimization number 9. When the true number of subsites are optimized, $Q_{min.,b.c.f.} = 0.8$, and by adding additional subsites to either end of the binding region, no significant improvement in fit results. The number of subsites on the binding region of an enzyme will be indicated when the fit is no longer significantly improved by adding additional subsites. In the minimization

Table 6. *Evaluation of the number of substitutes on a depolymerase*

Simulated bond-cleavage frequencies were computed by using the parent map in Table 3 and an approximation of experimental scatter was introduced. In various minimizations certain subsites were optimized (indicated by X) to obtain $Q_{min.,b.c.f.}$ while the remaining subsites were constrained at a binding energy at zero. The subsite index numbers I–X are real subsites, while –I and XI–XIII are virtual subsites.

Minimization								Catalytic site								
no.	Subsite index no. . .	–I	I	II	III	IV	V	VI ↓ VII	VIII	IX	X	XI	XII	XIII	$Q_{min.,b.c.f.}$
1						X	X		X	X					34.0
2					X	X	X		X	X					33.0
3				X	X	X	X		X	X					20.0
4						X	X		X	X	X				12.0
5						X	X		X	X	X	X			12.0
6					X	X	X		X	X	X				12.0
7				X	X	X	X		X	X	X				4.0
8				X	X	X	X		X	X	X	X			4.0
9			X	X	X	X	X		X	X	X				0.8
10			X	X	X	X	X		X	X	X	X			0.8
11		X	X	X	X	X	X		X	X	X				0.8
12		X	X	X	X	X	X		X	X	X	X			0.8
13		X	X	X	X	X	X		X	X	X	X	X		0.8
14		X	X	X	X	X	X		X	X	X	X	X	X	0.8

where additional subsites were optimized, in addition to the indicated ten subsites, these virtual subsites were optimized to ±0.63 kJ/mol. The virtual subsite energies are due to experimental error and, in general, we can attribute binding energies for any end subsite in the range ±0.63 kJ/mol to experimental variance.

In the interest of conservation of computer time, a visual inspection of the data can often provide an estimate of the number of subsites that can then be objectively examined by use of the computer model. Consider a substrate of chain length n, which is capable of spanning the entire binding region of an l subsite enzyme (i.e. $n \geqslant l$). Such a substrate has $(n-l+1)$ binding modes, which result in all subsites being occupied with a substrate monomer unit. The dissociation constant for each of these positional isomers is $K_{int.}$ (eqn. 6). As established by eqn. (5) the bond-cleavage frequencies for two adjacent binding modes (indexed ι and $\iota+1$) in which all subsites are filled ($n>l$) is $(k_{+2,\iota,n}/K_{int.})/(k_{2,\iota+1}/K_{int.})$ which simplifies to $k_{+2,\iota}/k_{+2,\iota+1}$.

Since all of the real subsites are occupied in each binding mode, eqn. (9) predicts that $k_{+2,\iota}=k_{+2,\iota+1}$. Consequently, when $n>l$, the positional isomers in which all subsites are filled will result in equal bond-cleavage frequencies for the n-mer. So that, as will be seen in the companion paper (Allen & Thoma, 1976), often a visual examination of the bond-cleavage frequencies will reveal the approximate size and position of the cleavage point. If the bond-cleavage frequencies that result from completely filled subsites are small due to unfavoured binding, as in the case of *B. amyloliquefaciens* amylase (Thoma *et al.*, 1970), it may be necessary to resort to the analysis used by Thoma *et al.* (1970) to obtain an estimate of the size of the binding region and the position of the catalytic site. In any case, the subjective evaluation must be tested with the model to determine if questionable subsites are significant.

Evaluation of $Q_{min.,total}$. To determine the subsite binding energies of the subsites adjacent to the catalytic site and to evaluate ΔG_a for *B. amyloliquefaciens* amylase, Thoma *et al.* (1971) used a tabulation-minimization technique. The error surface of $Q_{K_m}+Q_V$ created by varying the subsites and ΔG_a was visually inspected. This procedure is limited to, at most, a two-variable system; and, as pointed out by Swann (1969) the tabulation minimization is inefficient. The conjugate gradient minimization model offers a more efficient technique and allows optimization of all subsites simultaneously to give the best overall fit. The bond-cleavage-frequency partial subsite map generated as outlined above offers a starting point for the subsequent optimizations. We therefore start with good initial values for all of the subsites with the exception of the two adjacent to the catalytic site. The information for optimization of these two subsites is contained in $\tilde{K}_{m,n}$ and $\tilde{V}_n$, with the additional restriction that the sum of all subsite free energies is constrained by $K_{int.}$.

The computation time for the minimization increases rapidly with the number of subsites. Therefore the best procedure is to use $(Q_{K_m}+Q_V+Q_{K_{int.}})$ to minimize the energies of the two subsites adjacent to the catalytic site, holding the remaining subsites at the energies established by the bond-cleavage-frequency analysis outlined above. Then

Table 7. *Evaluation of the contribution of experimental variance to $Q_{min.}$.*

The parent subsite maps of Tables 2 and 3 were used to compute Michaelis parameters, bond-cleavage frequencies, and $K_{int.}$. In order to approximate experimentally measured data, an estimate of experimental variance was incorporated into the computed parameters. Q values were calculated by comparison to the parent-map data with incorporated scatter with the parent-map data computed before scatter was added. The data with scatter were then used in the minimization routine to obtain the optimum subsite map as dictated by $Q_{min.,total}$.

	Table 2 map ($\Delta G_a = 0$ kJ/mol)		Table 3 map ($\Delta G_a = 1.88$ kJ/mol)	
Parameter	Q*	$Q_{min.}$†	Q*	$Q_{min.}$†
b.c.f.	0.9	0.9	0.9	1.0
K_m	2.1	2.0	2.1	1.5
V	1.4	0.9	1.4	0.6
$K_{int.}$	0.2	0.1	0.2	0.5
total	1.1	1.0	1.1	1.0

* The Q values reported were obtained by the comparison of the parent maps simulated parameters with experimental scatter with the parent map parameters with no added scatter.

† The $Q_{min.}$ values were obtained from the minimization routine where the parent map simulated parameters with scatter were used to obtain an optimized subsite map as determined by $Q_{min.,total}$.

when all subsites are set at good initial values; a final refinement of the map can be affected by optimizing all subsites by minimizing Q_{total} ($Q_{K_m}+Q_V+Q_{K_{int.}}$). After the best fit is established without an acceleration factor ($\Delta G_a = 0$), ΔG_a can be allowed to vary and the improvement in the resulting $Q_{min.,total}$ evaluated.

Evaluation of the contributions to $Q_{min.}$.

As pointed out by Mannervik & Bártfai (1973) the error in an optimized model will be

$$Q_{min.} = q_{reg.} + q_{var.} + q_{bias} \quad (18)$$

where $q_{reg.}$ is the contribution to the total error due to inadequacies of the minimization procedure (regression error), $q_{var.}$ is the contribution to the total error due to experimental error (variance error) and q_{bias} is the contribution to the error due to inadequacies of the model (model-bias error). As the mathematical depolymerase model approaches a true description of the physical system, q_{bias} approaches zero. If we are to assess the fit of a subsite map, we must sort out these contributing errors in $Q_{min.}$.

A measure of $q_{reg.}$ is provided by the application of the minimization routine to error-free simulated data (Fig. 3) from the subsite maps in Tables 2 and 3. With these data there is no experiment scatter, so that $q_{var.} = 0$; and the subsite model is a perfect description of the data, so that $q_{bias} = 0$. Hence, the $Q_{min.}$ reached will be equal to $q_{reg.}$, the error due to the inadequacies of the minimization algorithm in establishing the absolute minimum. In each case tested with error-free simulated data, the subsite-binding energies were predicted correctly within ±0.040 kJ/mol and ΔG_a was within ±0.004 kJ/mol of the parent values. $Q_{min.}$ ($q_{reg.}$) was always less than 0.01. We will see that $q_{reg.}$ is insignificant as compared with $q_{var.}$.

The simulated data with an incorporated estimate of experimental scatter provide a means of estimating $q_{var.}$. For the maps in Tables 2 and 3, experimental error was added to the simulated parameters and the values of Q were calculated from eqns. (12)–(17) (MOD2 option of the model); the values are given in Table 7. When these simulated parameters with experimental error were used to optimize the map, $Q_{min.,total}$ was improved from 1.1 to 1.0. These values are slightly improved over the parent subsite map Q values. That is, the experimental scatter has caused the optimum map to be slightly different from the true parent map. Table 7 shows that the upper limit of $q_{var.}$ for bond-cleavage frequencies is 1.0, for K_m is 2.1, for V is 1.4, and for $K_{int.}$ is 0.5. The upper limit for the $q_{var.,total}$ is 1.1. Since $q_{reg.}$ was always less than 0.01, the only significant factors contributing to $Q_{min.}$ are experimental error ($q_{var.}$) and the inadequacies of the model in approximating the physical system (q_{bias}). In the following paper (Allen & Thoma, 1976), by using experimental data, we will determine the q_{bias} value.

In light of these simulation studies, we conclude that the depolymerase minimization model is applicable to the subsite-mapping problem. The only problem encountered with the minimization routine was the establishment of relative local minima when optimizing only with Michaelis parameters. This difficulty can be overcome by using different initialization values. The computer model is able to evaluate objectively the number of subsites and the position of the catalytic amino acids. In addition the

final map can be optimized by using all of the available experimental parameters. The summed weighted squared residuals, Q, offer a means of objectively evaluating the fit of a generated subsite map to the experimental error.

In summary, we have established the best procedure for complete subsite mapping as follows. (1) Establish experimental conditions where complicating mechanisms such as bimolecular reactions and multiple attack are insignificant, or modify the model to account for these factors. (2) Use end-labelled substrates to determine quantitative bond-cleavage frequencies for chain lengths that are large enough to span the entire binding region. (3) Measure $\tilde{K}_m$ and $\tilde{V}$ as a function of substrate chain length. (4) Examine bond-cleavage frequencies to estimate the number of subsites and the position of the catalytic site, then apply the minimization model to bond-cleavage frequencies to test the estimate. (5) By using the binding energies of the subsites generated in step 4 for the subsites not adjacent to the binding site, use $\tilde{K}_{m,n}$ and $\tilde{V}_n$ to optimize the subsites adjacent to the catalytic site, then make a refinement of the map by optimizing all of the subsites using $\tilde{K}_{m,n}$, $\tilde{V}_n$ and bond-cleavage frequencies. (6) Allow the acceleration factor to vary, and optimize all of the subsites and ΔG_a. (7) Examine the improvement of the fit effected by allowing ΔG_a to vary, to determine if the hydrolytic rate coefficients are constant.

This work was supported by grants from the National Science Foundation and the University of Arkansas Computing Center.

References

Abramowitz, N., Schecter I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **29**, 862–867

Allen, J. D. (1975), Ph.D. Thesis, University of Arkansas

Allen, J. D. & Thoma, J. A. (1976) *Biochem. J.* **159**, 121–131

Atlas, D., Levit, S., Schechter, I. & Berger, A. (1970) *FEBS Lett.* **11**, 281–283

Chipman, D. M. & Sharon, N. (1969) *Science* **165**, 454–465

Chou, J. Y. & Singer, M. F. (1970) *J. Biol. Chem.* **245**, 1005–1011

Cleland, W. W. (1963) *Nature (London)* **178**, 463–465

Coggins, J. R., Kray, W. & Shaw, E. (1974) *Biochem. J.* **138**, 579–584

Cuatrecusas, P., Wilchek, M. & Anfinsen, C. B. (1968) *Science* **162**, 1491–1493

Davidon, W. C. (1959) *AEC Research and Development Report* no. ANL-5990 (Rev.)

Fletcher, R. & Powell, J. D. (1963) *Comput. J.* **6**, 163–168

Fletcher, R. & Reeves, C. M. (1964) *Comput. J.* **7**, 149–154

Godefroy, T. (1970) *Eur. J. Biochem.* **14**, 222–231

Gurney, R. W. (1953) *Ionic Processes in Solution*, pp. 80–112, McGraw-Hill, New York

Hiromi, K. (1970) *Biochem. Biophys. Res. Commun.* **40**, 1–6

Holler, E., Rupley, J. A. & Hess, G. P. (1974) *FEBS Lett.* **40**, 25–28

IBM (1970) *Programmer's Manual: System 360 Scientific Subroutine Package* (360*A-CM-03X*), version III, (IBM Publication no. GH 20-0205-3), pp. 211–229

Iwasa, S., Aoshima, H., Hiromi, K. & Hatano, H. (1974) *J. Biochem. (Tokyo)* **75**, 969–978

Kato, M., Hiromi, K. & Morita, Y. (1974) *J. Biochem. (Tokyo)* **75**, 563–576

Lazarus, H. M., Sparn, M. D. & Bradley, D. F. (1968) *Proc. Natl. Acad. Sci. U.S.A.* **60**, 1503–1510

Mannervik, B. & Bártfai, T. (1973) *Acta Biol. Med. Germ.* **31**, 203–215

Morihara, K. & Oka, T. (1968) *Biochem. Biophys. Res. Commun.* **30**, 625–630

Morihara, K., Oka, T. & Tsuzuki, H. (1969) *Biochem. Biophys. Res. Commun.* **35**, 210–214

Nitta, Y., Mizushima, M., Hiromi, K. & Ono, S. (1971) *J. Biochem. (Tokyo)* **69**, 567–576

Rexová-Benková, L. (1973) *Eur. J. Biochem.* **39**, 109–115

Robyt, J. & French, D. (1963) *Arch. Biochem. Biophys.* **100**, 451–467

Robyt, J. & French, D. (1970) *J. Biol. Chem.* **245**, 3917–3927

Scarborough, J. B. (1962) *Numerical Mathematical Analysis*, p. 133, Johns Hopkins Press, Baltimore

Schechter, I. & Berger, A. (1967) *Biochem. Biophys. Res. Commun.* **27**, 157–162

Shibaoka, T., Miyano, K. & Watanabe, T. (1974) *J. Biochem. (Tokyo)* **76**, 475–479

Swann, W. H. (1969) *FEBS Lett.* **2**, S39–S55

Thoma, J. A. & Allen, J. D. (1976) *Carbohydr. Res.* **48**, 105–124

Thoma, J. A., Brothers, C. & Spradlin, J. (1970) *Biochemistry* **9**, 1768–1775

Thoma, J. A., Rao, G. V. K., Brothers, C., Spradlin, J. & Li, H. L. (1971) *J. Biol. Chem.* **246**, 5621–5635

Thompson, R. C. & Blout, E. R. (1970) *Proc. Natl. Acad. Sci. U.S.A.* **67**, 1734–1740