Proteolytic Fragmentation of Fibrinogen. II. Kinetic Modeling of the Digestion of Human and Bovine Fibrinogen by Plasmin or Trypsin†

Richard I. Shrager, Elemer Mihalyi,* and David W. Towne

ABSTRACT: The kinetic data presented in the previous paper (Mihalyi, E., et al. (1976), *Biochemistry 15,* preceding paper in this issue), with respect to the fragmentation of human and bovine fibrinogen by either plasmin or trypsin, were compared with several chemical kinetic models. The models were derived mathematically on the basis of the three-nodular structure of fibrinogen (Hall, C. E., and Slayter, H. S. (1959), *J. Biophys. Biochem. Cytol. 5,* 11) and the asymmetrical cleavage sequence first proposed by Marder, V. J., et. al. ((1969) *J. Biol. Chem. 244,* 2111). The parameters were determined by nonlinear curve fitting. The whole process could be described accurately by only two rate constants. Several variant models were tested and, although a clear cut choice cannot be made, one of these, the protected three-bonds model, appears to give the best fit in most cases. This model assumes that the chain segment that distinguishes F from X protects certain other chains (the bonds) from proteolytic cleavage.

One may begin with the usual caveat about models. Any set of data can be fit by many models. A good model is one that fits a broad set of data without needless complexity, and, preferably, with model parameters that suggest some physical meaning. A note on symbols is also in order. The symbols F, D_L, etc. will represent chemical species in reaction diagrams, and concentrations of those species in mathematical equations will be represented by the symbols F, D_L, etc., respectively. This convention is simple and sufficiently unambiguous for present purposes.

Methods

The model proposed here for the digestion of fibrinogen is one involving three subunits D_L, D_R, and E, with six connecting chains or bonds. The subunits D_L and D_R will be collectively referred to as D. Two other chain portions, which may or may not protect the first six, are also present. These are severed from the molecule at such a fast rate compared with the other cleavages that this process can be approximated here as a single step, governed by a single rate constant. For convenience, these chain segments will be designated as chain 7, bearing in mind that this represents actually two chains. The complete molecule, called F, is pictured in Figure 1, along with its transient products X and Y, and its final products D and E. First, it may be assumed that chains 1 through 6 are all available for digestion in any order, though possibly at differing rates. One exception is that chain 7 may protect the others until it is severed. Further, it is assumed that free, active enzyme concentration is constant throughout the experiment, so the decays of the bonds are first-order reactions with rate constants denoted k_1 to k_6 for chain 1 through 6 and k_7 for chain 7.

Figure 2 shows the possible paths the reaction might follow for a particular molecule. The first subscript on X and the Y's in the right column are concerned with chains 1, 2, and 3. The second subscript on X and the Y's in the bottom row are concerned with chains 4, 5, and 6. If chain 7 does not protect the others, then F may be disregarded in the diagram, since F is merely a heavy form of X.

A system of differential equations that describes the entire reaction with each subspecies would be quite large. There are nine configurations called $X_{2,1}$, if each chain behaves differently, and so on through the combinations, yielding 49 species of X and 14 species of Y as shown in the superscripts of Figure 2. But only the aggregates F, X, Y, D, and E can be observed. Therefore the equations will be much simpler.

When the various models are matched to the data later in the paper, a symmetric molecule will always be assumed, i.e., the fragments D_L and D_R are identical, and bonds 1, 2, and 3 are identical with bonds 4, 5, and 6, respectively. However, the models will be derived without requiring symmetry, because it is possible that, at some future time, data distinguishing the two sides will be available.

In the first model, it is assumed that chain 7 offers no protection to the other six chains. It is possible that some of the bonds are severed simultaneously, in which case they appear kinetically as one bond. The treatment of fewer than six effective bonds is considered later in this paper. Assuming first-order independent decay of bonds (chains) with rates k_1, k_2, . . ., k_6, the probability that the ith bond is severed by time t is given by:

$$p_i(t) = 1 - \exp(-k_i t)$$

The severing of an individual bond is really a second-order process, relying on the concentration of enzyme as well as the concentration of intact bonds. However, if the fraction of bound enzyme is small so that a constant amount of free enzyme may be assumed, then the first-order treatment may be retained, remembering that the decay rates now contain free enzyme concentration as a factor.

If the bonds decay independently, the probability that any combination of bonds is severed is the product of the individual probabilities. For large numbers of molecules, the concentration of a given species is practically proportional to its probability. The time course of the reaction may then be described in the following way. The formulas for the concentrations D_L, D_R, and E all derive from the product rule for the probability of several independent events (the breaking of the six bonds). D is defined as the sum of D_L and D_R. The formula for Y fol-

† From the National Institutes of Health, Division of Computer Research and Technology, Laboratory of Statistical and Mathematical Methodology, Bethesda, Maryland 20014 (R.I.S.), and the National Heart and Lung Institute, Laboratory of Cell Biology, Bethesda, Maryland 20014 (E.M. and D.W.T.). *Received August 15, 1975.*

lows from the rules for the release of D: there must be one molecule of D for every molecule of Y plus two molecules of D for every molecule of E, or $D = Y + 2E$. Finally, X is related to Y and E by conservation of matter. In normalized (dimensionless) concentrations, the formulas for the time course are:

$$D_L = (1 - \exp(-k_1 t)) \times (1 - \exp(-k_2 t))(1 - \exp(-k_3 t)) \quad (1)$$

$$D_R = (1 - \exp(-k_4 t))(1 - \exp(-k_5 t))(1 - \exp(-k_6 t))$$

$$D = D_L + D_R,\ E = D_L D_R,\ Y = D - 2E;\ X = 1 - Y - E$$

In the second model, it is assumed that chain 7 protects the others until it is severed, and the reasoning is as follows: since the X, Y, D, E system can be described by a linear system of differential equations with constant coefficients (decay rate constants), and if one knows the impulse response (the result when everything starts as $X_{3,3}$) which is given in eq 1, and further that the reaction F → X is first order with rate constant k_7, then (1) $F = \exp(-k_7 t)$; (2) the input rate from F into the X, Y, D, E system is $-dF/dt$ or $+ k_7 \exp(-k_7 t)$; (3) if an impulse response $P(t)$ and the input rate function $I(t)$ of a linear system with constant coefficients are known, then the resulting response $R(t)$ is given by the convolution of P and I (see any elementary treatment of control theory on this point):

$$R(t) = \int_0^t P(u) I(t - u) du$$

(Note: u is a variable of integration which disappears when the definite integral is evaluated.) In particular, $D_L(t)$ is given by

$$D_L(t) = \int_0^t (1 - \exp(-k_1 u))(1 - \exp(-k_2 u)) \times (1 - \exp(-k_3 u)) k_7 \exp(-k_7 (t - u)) du$$

with a similar expression for $D_R(t)$. Performing the integration, one gets

$$\begin{aligned} D_L = 1 + k_7[&-\exp(-k_1 t)/(k_7 - k_1) \\ &- \exp(-k_2 t)/(k_7 - k_2) - \exp(-k_3 t)/(k_7 - k_3) \\ &+ \exp(-(k_1 + k_2)t)/(k_7 - k_1 - k_2) \\ &+ \exp(-(k_1 + k_3)t)/(k_7 - k_1 - k_3) \\ &+ \exp(-(k_2 + k_3)t)/(k_7 - k_2 - k_3) \\ &- \exp(-(k_1 + k_2 + k_3)t)/(k_7 - k_1 - k_2 - k_3)] \\ &- \left[\frac{1}{k_7} - \frac{1}{k_7 - k_1} - \frac{1}{k_7 - k_2} - \frac{1}{k_7 - k_3} + \frac{1}{k_7 - k_1 - k_2} \right. \\ &\left. + \frac{1}{k_7 - k_1 - k_3} + \frac{1}{k_7 - k_2 - k_3} - \frac{1}{k_7 - k_1 - k_2 - k_3}\right] \\ &\times k_7 \exp(-k_7 t) \quad (2) \end{aligned}$$

with a similar expression for $D_R(t)$ substituting k_4, k_5, and k_6 for k_1, k_2, and k_3, respectively. The remainder of the system is given by:

$$D = D_L + D_R,\ E = D_L D_R,\ Y = D - 2E,\ F = \exp(-k_7 t),\ X = 1 - F - Y - E$$

(A computational note: only the first three exponentials in the second term of eq 2 need be computed. The others are products of the first three.)

With either eq 1 or eq 2, a model of relative concentration is generated in which the maximum possible concentrations

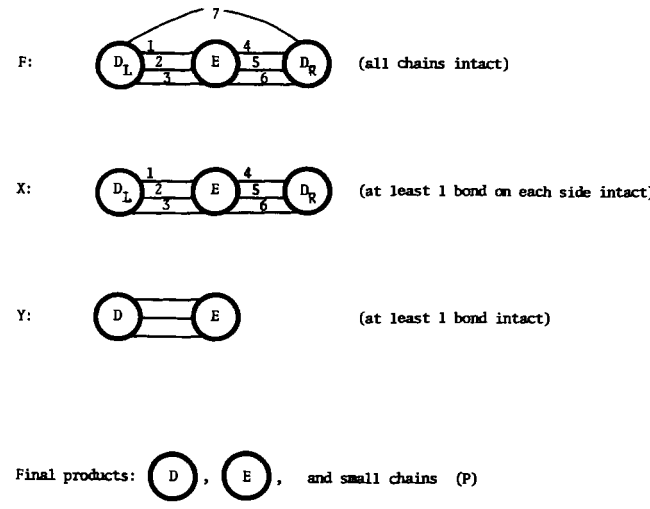

FIGURE 1: Species present during digestion.

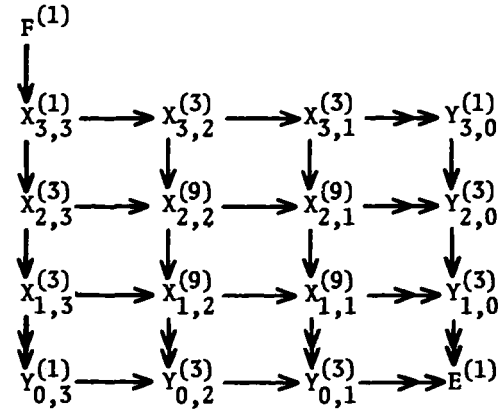

FIGURE 2: Proposed mechanism. $X_{i,j}$ is species X with i bonds on side 1 intact and j bonds on side 2 intact. Likewise, $Y_{0,i}$ and $Y_{i,0}$ are species Y with i bonds intact. Each molecule follows one of the paths from F (or $X_{3,3}$) to E. A molecule of D is released whenever a double arrow (→→) is transversed. Decay rates cannot be assigned to the arrows unless several decay rates are assumed equal. The superscripts are the number of subspecies if each bond has a distinct decay rate.

of F, X, Y, D_L, D_R, and E are all 1. Maximum possible concentration of D is 2 since D is defined as the sum of D_L and D_R. The observed data, however, are in units of optical density, so the following procedure was used to convert the model-generated numbers and data to relative optical density (ROD):[1]

1. The last point in each time course experiment was regarded as equilibrium. From this point the normalized absorption coefficients (NAC) of D, E, and P (the sum of the small fragments $P_1 + P_2$) were estimated since they were the only products present:

$$NAC(D) = \frac{1}{2} OD(D_\infty)/[OD(D_\infty) + OD(E_\infty) + OD(P_\infty)]$$

$$NAC(E) = OD(E_\infty)/[OD(D_\infty) + OD(E_\infty) + OD(P_\infty)]$$

$$NAC(P) = OD(P_\infty)/[OD(D_\infty) + OD(E_\infty) + OD(P_\infty)]$$

NAC(D), NAC(E), and NAC(P) were further refined by including them as parameters in the curve fits presented later in this paper.

[1] Abbreviations used: ROD, relative optical density; NAC, normalized absorption coefficients.

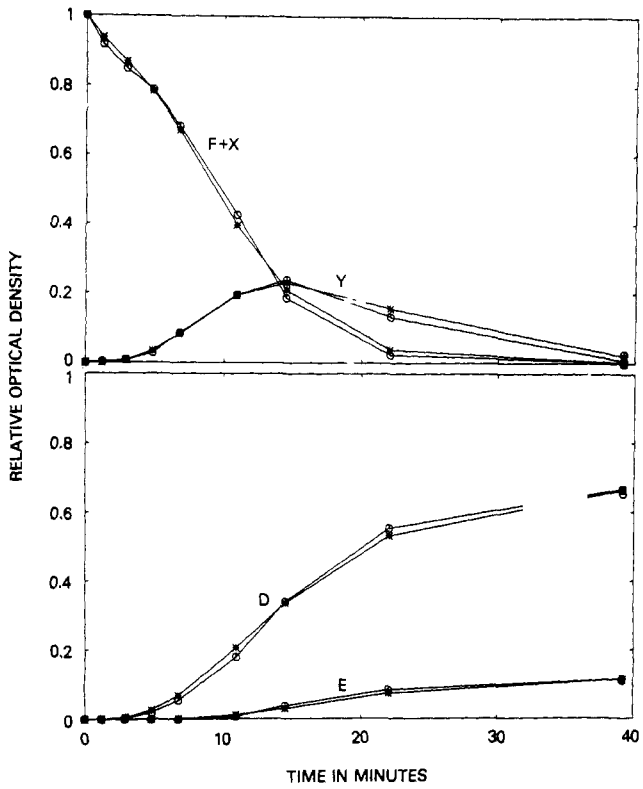

FIGURE 3: Bovine fibrinogen with trypsin as observed (curve labeled with 0's) and as computed by the protected three-bond model (curve labeled with *'s). This fit is a typical result for all experiments.

2. The remaining NAC's were assumed to be sums of their components:

$$NAC(F) = 2NAC(D) + NAC(E) + NAC(P) = 1$$

$$NAC(X) = 2NAC(D) + NAC(E)$$

$$NAC(Y) = NAC(D) + NAC(E)$$

3. For each time point, the computed relative optical densities were given by:

$$ROD_C(F) = F * NAC(F) = F$$

$$ROD_C(X) = X * NAC(X)$$

$$ROD_C(Y) = Y * NAC(Y)$$

$$ROD_C(D) = D * NAC(D)$$

and

$$ROD_C(E) = E * NAC(E)$$

The data, in OD, were converted to ROD by dividing each OD by total OD for that time point. For example:

$$ROD_{obsd}(X) = OD(X)/(OD(F) + OD(X) + OD(Y) + OD(D) + OD(E) + OD(P))$$

A curve-fitting procedure (Shrager, 1970, 1972) was called upon to adjust the rates in eq 1 and 2 so that the sum of squares over all time and species:

$$S = \sum w_i(ROD_{obsd} - ROD_C)^2$$

was minimized. The weights w were set at 2500, because the errors in ROD were close to 0.02, and the appropriate weight for the Chi-square goodness of fit statistic is:

$$w = 1/(\text{standard error})^2$$

for independent errors. For a simple statistical treatment, normal independent errors in ROD and no error in time were assumed.

The curve-fitting procedure also provided asymptotic standard errors for the fitted parameters and the ability to constrain parameters to linear equalities or inequalities. The constraints chosen were that all rate constants must be equal.

$$k_1 = k_2 = k_3 = k_4 = k_5 = k_6$$

for the models with three bonds on a side, and

$$k_1 = k_2 = k_4 = k_5 \text{ with } k_3 = k_6 = 10$$

for the models with two bonds on a side and

$$k_1 = k_4 \text{ with } k_2 = k_3 = k_5 = k_6 = 10$$

for the models with one bond on a side.

The reasons for setting such constraints were threefold:

1. Individual rates for each bond could not be well resolved by the program since individual subspecies of X and Y could not be observed.

2. Among slowly decaying bonds, a rapidly decaying bond (say $k = 10$) cannot be distinguished from a nonexistent bond.

3. It is possible to resolve differing rates for the decay of X and the decay of Y since these are observable. However, the data do not exhibit any systematic disagreement with a symmetric model; i.e., some fits would be slightly improved with X decaying slower than Y, while other fits would be slightly improved by the reverse. Therefore, the data give no convincing evidence of assymmetry or cooperativity in the decay rates. Therefore, the only thing resolved will be an average rate for the slowest decaying bonds, with the more rapid decaying bonds seeming to disappear. Thus, one or two bonds per side can be modeled using a three-bond-per-side model.

Results

There are, finally, six models to consider: one, two, or three bonds per side, with or without the protection of chain 7. The four sets of data to wihch the models were compared are described in the first paper of this series and a typical result is shown in Figure 3. The results are presented in Table I. For the unprotected models, one and two bonds per side gave poor fits to all sets of data, so they are not shown. Table I shows only the two rate constants k_1 and k_7 because the others were constrained to either $k_i = k_1$ or $k_i = 10$ depending on the model. The NAC's of D, E, and P were also parameters, but they were linear, well-determined, and practically independent of k_1 and k_7. For statistical reasons, they were included in determining degrees of freedom and standard errors, but the fitted NAC's, being much like those observed, are not included in Table I. A measure of the goodness of fit is the Chi-squared "probability" (p), which is typically 0.5, and the higher it is (maximum 1.0) the better the fit. Assuming normally distributed data, correct weights, and no error in time, then p is the true probability that the sum of squares would be this bad or worse. The calculated "probabilities" are not so precise; they are, rather, similarity measures between model and data, but hopefully they are informative enough to allow reasonable model selection. Since all degrees of freedom were greater than 30, the normal approximation to the Chi-squared distribution was used (Hodgman, 1954).

The bovine fibrinogen with plasmin data yielded a surprisingly poor fit to the most favored model in the other data. The major difference seems to lie in one data point: the first ob-

TABLE I: Numerical Results.[a]

Data	Params	Models: 3 Bonds Unprotected	3 Bonds Protected	2 Bonds Protected	1 Bond Protected
Bovine plasmin	k_1	0.158 ± 0.030	0.192 ± 0.040	0.160 ± 0.026	0.166 ± 0.036
	k_7	0.346 ± 0.083	0.600 ± 0.120	0.446 ± 0.079	0.187 ± 0.033
	$p(S)$	0.1949	0.0001	0.1446	0.6950
Bovine plasmin[b]	k_1	0.155 ± 0.03	0.155 ± 0.03		
	k_7	10 (fixed)	10 (fixed)		
	$p(S)$	0.3409	0.3409		
Bovine trypsin	k_1	0.105 ± 0.015	0.140 ± 0.030	0.130 ± 0.027	0.112 ± 0.029
	k_7	0.251 ± 0.069	0.309 ± 0.067	0.192 ± 0.042	0.115 ± 0.026
	$p(S)$	0.0618	0.9994	0.3974	0.0000
Human plasmin	k_1	0.197 ± 0.035	0.206 ± 0.035	0.161 ± 0.024	0.147 ± 0.016
	k_7	3.23 ± 1.5	3.18 ± 1.4	1.92 ± 0.51	0.351 ± 0.040
	$p(S)$	0.7357	0.7486	0.1314	0.0000
Human trypsin	k_1	0.0593 ± 0.0077	0.0064 ± 0.0080	0.0608 ± 0.0076	0.0586 ± 0.0084
	k_7	0.871 ± 0.220	0.410 ± 0.067	0.176 ± 0.021	0.0785 ± 0.0089
	$p(S)$	0.0000	0.0375	0.0000	0.0000

[a] All these rate constants pertain to 9.3 units of trypsin/ml and to 2.19 CTA units of plasmin/ml. In molar concentrations these correspond to 1.48 μmol/ml trypsin and 0.83 μmol/ml plasmin. [b] Earliest X observation ignored.

servation of $F + X$, which appears much higher than one would expect from the other runs. On the chance that it was an outlier, that one point was removed, whereupon k_7 became indeterminate (subject to great fluctuations with little effect on the fit) for the three-bond models. With k_7 fixed at 10, the probability of those two models rose to a reasonable value as shown in Table I. Another experiment was run to check the value of $F + X$ at $t = 1.32$, and the new value was 0.87 instead of 0.93, still not low enough to favor the six-bond protected model, but enough to illustrate that the value in question is not well-resolved. A final verdict must await better data at early times.

Note that when k_7 is fast (say $k_7 > 2$), the three-bond models become practically indistinguishable because there is little difference between rapidly decaying protection and no protection. Also, when k_7 becomes fast (as in the 3-bond models for human fibrinogen with plasmin) or when early data is sparse, k_7 becomes poorly determined. This last fact again indicates that denser early data should be sought, especially for plasmin. (Although F and X are not separated in the data, the loss of optical density in $F + X$ can be seen to stem from two causes: (1) X becoming Y, which means that as ROD(F + X) decreases, ROD(Y) must increase, and (2) loss of mass in $F + X$ which we interpret as F becoming X, and which by contrast does not result in an increase in ROD(Y). From the involvement of Y, one can distinguish mathematically, and almost by eye, which loss $F + X$ is incurring at which time, and at what rate, provided there are sufficient data at early times.)

The low probabilities for human fibrinogen with trypsin were due to a higher noise level than anticipated. If smaller uniform weights were used, all probabilities would rise, the parameters would not change, and the protected three-bond model would still be best by far.

In conclusion, the most reasonable choice of model at this time is the protected three-bond-per-side model, where plasmin digests chain 7 more rapidly than trypsin does. With somewhat less conviction, one could claim that chain 7 protects only against trypsin.

Other models also promise to fit the data well, if and when

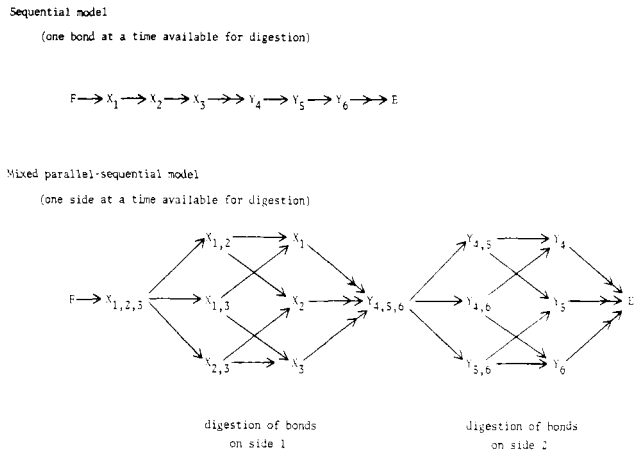

FIGURE 4: Alternate mechanisms. Subscripts tell which bonds are intact and available for digestion. A molecule of D is released, whenever a double arrow (→→) is traversed.

some evidence arises to justify their use. Three models in particular come to mind: (1) the strict sequential model in which the chains are digested in a fixed order; (2) the mixed sequential-parallel model in which bonds 1, 2, and 3 are digested in parallel followed by bonds 4, 5, and 6 in parallel, and (3) the mixed sequential-parallel model in which bonds 1, 2, and 3 are digested in sequence while bonds 4, 5, and 6 are also digested in sequence. The first two of these models are diagrammed in Figure 4.

As an example of further evidence, the work of Furlan et al. (1975) suggests that fragments of α chain appear first, followed by β and γ chains in either order. This could indicate that the α chain decay rate constants k_1 and k_4 are considerably faster than k_2, k_3, k_5, and k_6, but this model would behave like a two-bond-per-side model which does not give good results on data presented here. An alternate model has the α chain protecting the β and γ chains until it is severed. The form of the model is derived using the same reasoning preceding eq 2,

except that now, the α chains are the protectors. Without details the result is:

$$D_L = 1 + k_1\left[-\frac{\exp(-k_2t)}{k_1-k_2} - \frac{\exp(-k_3t)}{k_1-k_3} + \frac{\exp(-(k_2+k_3)t)}{k_1-k_2-k_3}\right] + k_1\exp(-k_1t)\left[-\frac{1}{k_1} + \frac{1}{k_1-k_2} + \frac{1}{k_1-k_3} - \frac{1}{k_1-k_2-k_3}\right] \quad (3)$$

with a similar expression for D_R in k_4, k_5, and k_6. The remainder of the system is:

$$D = D_L + D_R,\ E = D_L D_R,$$
$$Y = D - 2E, \text{ and } X = 1 - Y - E$$

The above model, when run against the time course data, yielded the following probabilities:

bovine − plasmin: 0.40 better than preferred model

bovine − trypsin: 0.25, worse than preferred model

human − plasmin: 0.71, same as preferred model

human − trypsin: 0.0008, worse than preferred model

The results are not impressive, especially considering that there was one more parameter here, namely the α-chain decay rate (k_1 and k_4) which was permitted to differ from the other k's. Possibly, the early α-chain fragments observed by Furlan et al. (1975) are not in the segment connecting D to E. On the other hand, none of the above results gave an unreasonable looking fit, so this model is not ruled out.

Discussion

The kinetic model (2) was constructed on the basis of the Hall and Slayter (1959) trimodular model which appears most consistent with the data presented here. There is further chemical evidence in the literature for a fast release of the C-terminal portions of the Aα chains (Furlan and Beck, 1972; Gaffney and Dobos, 1971; Mills and Karpatkin, 1972; Mosseson et al., 1973; Pizzo et al., 1972). All intermediates in question have been observed. Their molecular weights are consistent with the model ($MW_Y \approx MW_D + MW_E$, and so on) as is their order of appearance (Y and D appear as X disappears, while E is somewhat delayed). Even with a perhaps overly simple model (initial dimers ignored, six rates assumed equal, first-order kinetics), it is possible to get good qualitative agreement with the data. Likewise, good agreement is observed for a model (3) based on the data of Furlan et al. (1975). However, the observed order of appearance seems inconsistent with models that permit E to appear along with initial D or Y (Mosseson et al., 1973; Koppel, 1967). Likewise, models that have one bond, or one rate-limiting bond (with other bonds decaying rapidly), per side with no protection (Takagi and Doolittle, 1975) seem to be eliminated because such models require the almost immediate appearance of Y (i.e., maximum slope for Y at or very near $t = 0$), whereas data presented here exhibit a definite delay in Y information. In summary, models with three influential bonds per side, and involving some form of protection, seem favored over the other models under investigation.

Data from the pH-stat (previous paper in this series) can be resolved well as sums of two exponentials. At first, this fact seems consistent with the two rate constants k_1 and k_7 from the large fragment data presented here, but the rates from the two sets of data do not agree very well, nor is there any consistent factor that relates them from case to case. This is not surprising, because the large fragment data are concerned with about 8 bonds per structure molecule, whereas pH-stat data are concerned with about 80 bonds per molecule or piece thereof. The pH-stat data will be dominated by those 90 percent of the bonds not accounted for by k_1 and k_7.

References

Furlan, M., and Beck, E. A. (1972), *Biochim. Biophys. Acta 263,* 631.

Furlan, M., Seelich, T., and Beck, e. A. (1975), *Biochim. Biophys. Acta 400,* 112.

Gaffney, P. J., and Dobos, P. (1971), *FEBS Lett. 15,* 13.

Hall, C. E., and Slayter, H. S. (1959), *J. Biophys. Biochem. Cytol. 5,* 11.

Hodgman, C. (1954), Mathematical Tables from Handbook of Chemistry and Physics, Cleveland, Ohio, CRC Press.

Koppel, G. (1967), *Z. Zellforsch. 77,* 443.

Marder, V. J., Shulman, N. R., and Carroll, W. R. (1969), *J. Biol. Chem. 244,* 2111.

Mills, D., and Karpatkin, S. (1972), *Biochim. Biophys. Acta 271,* 163.

Mosseson, M. W., Finlayson, J. S., and Galanakis, D. K. (1973), *J. Biol. Chem. 248,* 7913.

Pizzo, S. V., Schwartz, M. L., Hill, R. L., and McKee, P. A. (1972), *J. Biol. Chem. 247,* 636.

Shrager, R. (1970), *J. Assoc. Comput. Mach. 17,* 446.

Shrager, R. (1972), *Commun. ACM 15,* 41.

Takagi, T., and Doolittle, R. F. (1975), *Biochemistry 14,* 940.