Separation Science and Technology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t713708471

Microheterogeneous Collagen Model on the Basis of a Chromatographic Study on Hydroxyapatite Columns

Tsutomu Kawasaki[a]
[a] LABORATOIRE DE GENETIQUE MOLECULAIRE INSTITUT DE RECHERCHE EN BIOLOGIE MOLECULAIRE FACULTE DES SCIENCES, PARIS, FRANCE

To cite this Article Kawasaki, Tsutomu(1982) 'Microheterogeneous Collagen Model on the Basis of a Chromatographic Study on Hydroxyapatite Columns', Separation Science and Technology, 17: 3, 407 — 452
To link to this Article: DOI: 10.1080/01496398208068550
URL: http://dx.doi.org/10.1080/01496398208068550

SEPARATION SCIENCE AND TECHNOLOGY, 17(3), pp. 407–452, 1982

Microheterogeneous Collagen Model on the Basis of a Chromatographic Study on Hydroxyapatite Columns

TSUTOMU KAWASAKI

LABORATOIRE DE GENETIQUE MOLECULAIRE
INSTITUT DE RECHERCHE EN BIOLOGIE MOLECULAIRE
FACULTE DES SCIENCES
PARIS 5, FRANCE

Abstract

Earlier, on the basis of experimental chromatograms of collagen obtained on hydroxyapatite columns, a microheterogeneous collagen model with fluctuating primary structures around a mean structure was proposed. In this paper this model is examined in detail on the basis of a new chromatographic theory. Good fits were obtained between theory and experiment. It was estimated that the probability, π, that aspartic and glutamic residues exist at correct positions on the three α collagen peptide chains is about 0.8. The following conclusions were derived at the same time: (a) the energy of adsorption, ε, for a carboxyl group of collagen on to a crystal site of hydroxyapatite is 0.7 kcal/mol. (b) The replusive interaction energy, $\tilde{E}$, for one of collagen molecules with another on the crystal surface is 3 kcal/mol when these make maximum contact. (c) The square-root, h', of the area on the collegen surface where a carboxyl group can move freely (around the mean position) is 2.5–3.5 Å. (d) It is likely that a considerable adsorption of phosphate ions from the buffer occurs on the collagen surface adsorbed on the hydroxyapatite surface. The experimental values obtained in (a)–(c) are all reasonable ones; this strongly supports the microheterogeneous model itself. It is suggested that the ambiguous primary structure arises at a translational level of biosynthesis.

I. INTRODUCTION

Chromatographies of calf-skin collagen (in its native state) were carried out on hydroxyapatite (HA) columns by applying several different column lengths and linear molarity gradients of phosphate buffer (pH 6.8) (*1*). Collagen is heterogeneous. In agreement with the prediction from the general theory of HA chromatography (*2–6*), when the slope of the phosphate

gradient is extremely small, the resolution of the column increases and multipeak chromatograms are obtained (*1*). Several typical results are reproduced in Figs. A1–A4 in Appendix I. In parallel with the behaviors of synthetic poly-L-Asp and poly-L-Glu (with adsorption carboxyl groups) and DNA (with adsorption phosphate groups), the elution of collagen from the HA column is achieved almost independently of the molarities of KCl or NaCl co-existing in the phosphate buffer. The elution is carried out only by phosphate ions (Ref. 7, Appendix I). It can be deduced (Ref. 7, Appendix I) (*8*) that collagen is adsorbed onto one of the two types of crystal sites (called C sites) by using carboxyl groups. The desorption of collagen from the crystal surfaces is carried out by competition with phosphate ions from the buffer that are also adsorbed onto C sites. (For the arrangement of C sites and the orientation of collagen molecules on the HA surface, see Discussion Section B). This would mean that each peak in a multipeak chromatogram (Figs. A3 and A4) consists of molecules that are adsorbed by using (almost) the same number of carboxyl groups.

It is confirmed that rechromatography of the different parts of the chromatogram gives patterns with narrow widths eluted at phosphate molarities almost equal to those where the fractions appeared in the first chromatogram. The sum of the independent chromatograms of the fractions in the original chromatogram is almost identical with the original chromatogram (*1*). Amino acid analysis, measurements of the intrinsic viscosity, optical rotatory dispersion, and electron-microscopical observations of SLS fibers (*9*) for different chromatographic fractions were performed. No differences, however, have been detected so far (*7*). On the other hand, the "determination" of all or most of the primary structure of collagen (for the sample which is not yet subjected to HA chromatography) has been done successfully by several authors (Appendix III).

On the basis of these data and the shape of the multipeak chromatograms obtained on HA columns (Figs. A3 and A4), a microheterogeneous collagen model was proposed (*7*, *10*). This model states that the primary structure of collagen is ambiguous, fluctuating around a mean structure that can be "determined" by the organochemical method. The probability that aspartic and glutamic residues (with carboxyl groups) exist at "correct" positions on the three α peptide chains (Appendix III) is less than unity. Collagen, therefore, is a statistical assembly of a large number of slightly different components. The reasoning behind this model is mentioned below. First, from the rechromatography experiment showing that there are no mutual transitions among different collagen components, and the chemical and physical identities among them (within the limits of experimental errors; see above) being compatible with the well-established fact that collagen is homogeneous in both dimensions and shape (Appendix III), it can be

deduced that the differences in different collagen components reside in their primary structures. The possibilities that these are due to heterogeneous deaminations among molecules and/or posttranslational modifications of the molecular structure can almost be excluded (see Discussion Section F). Second, provided each peak in a multipeak chromatogram (Figs. A3 and A4) consists of a single molecular species with a proper primary structure, then "the organochemical determination of the primary structure for the mixture of several molecular species that correspond to the number of the peaks" would be impossible. Third, the chromatograms (Figs. A3 and A4) are characterized by properties (a) that the intervals between the peaks are roughly constant and (b) that the heights of peaks decrease gradually on the right-hand side of the chromatogram. From the analogy with the binomial distribution, the existence of a statistical assembly of a large number of molecular components can be suggested behind this type of chromatogram (cf. Fig. A6). It can be assumed that, due to the ambiguity in the primary structure, the carboxyl groups of the acidic amino residues are distributed statistically around the mean arrangement on the surface of each collagen component; the discrete distribution in peaks in a multipeak chromatogram is a result of the stepwise variation in number of carboxyl groups per molecule that are available for the reaction with crystal C sites. Fourth, from the intervals between the peaks in the multipeak chromatogram, the differences in adsorption energies between molecules involved in the neighboring peaks can be estimated independently of the microheterogeneous model. If this model is introduced, these energy differences should represent the adsorption energy, ε, for a carboxyl group onto a C site (see above). In Ref. *3* a value of about 0.5 kcal/mol was obtained for these energy differences by using Fig. A3(d) (identical with Fig. 9c in that paper). On the other hand, according to the microheterogeneous model, the total shape of a multipeak chromatogram should depend upon the ε value since the geometrical configurations of each molecule on the crystal surface should follow a Boltzmann distribution (cf. Theoretical Section B). Thus it can be seen in Figs. A1–A4 that some components are concentrated at the extreme left-hand parts of the chromatograms. The degree of this concentration should depend upon the ε value. In Ref. *10* the ε value of the order of 0.5 kcal/mol was obtained by using this method (cf. Theoretical Section B). The fact that the same ε value was obtained by using two independent methods strongly supports the microheterogeneous model. In the present paper a best ε value of 0.7 kcal/mol will be obtained (Theoretical Section C). Lastly, from the analysis of the chromatogram for a mixture of nucleoside phosphates with different phosphate chain lengths, it can be estimated that the adsorption energy (onto a C site) for a univalent phosphate group is 0.9–1 kcal/mol (*8*). This is close to the value for the adsorption energy of a carboxyl group estimated above.

We present below two other experimental data that would support indirectly the microheterogeneous collagen model. Thus parts (a), (b), and (c) of Fig. A5 in Appendix I illustrate three chromatograms (with small sample loads) of lysozyme, a typical globular enzymatic protein, obtained under almost the same experimental conditions as those attained in Figs. A1, A2(b), and A3(d) and A4(b), respectively. Lysozyme is adsorbed by using basic groups onto the other crystal sites (P sites) existing on different HA surfaces from those where C site exist (*3*, *11*, *12*) (see Discussion Section E). It can be seen in Fig. A5 that, in contrast to collagen, lysozyme is a highly homogeneous protein. The width of the chromatographic peak of lysozyme in Fig. A5(c) is of the same order of magnitude as the widths of the peaks in the multipeak chromatograms in both Figs. A3(d) and A4(b). The increase in the resolution of the column occurring in the order of Figs. A1, A2(b), and A3(d) and A4(b) can be explained in terms of the decrease in the width of the lysozyme peak occurring in order of parts (a), (b), and (c) of Fig. A5 (*3*). This demonstrates that a peak in the multipeak collagen chromatogram, in fact, consists of the molecules that are chromatographically homogeneous being adsorbed by using (almost) the same number of carboxyl groups (see above).

Figure A6 in Appendix I illustrates a chromatogram of a synthetic poly-L-lysine·HBr sample with molecular weights ranging between 1500 and 8000 daltons (degrees of polymerization 7–38). This can a priori be assumed to be an assembly of a number ($\approx$22) of components with a statistical distribution for chain lengths. Poly-L-lysine is adsorbed onto P sites by using ε-amino groups of the side chains (except for the case of extremely small chain lengths) (*12*)). It can be seen in Fig. A6 that, in parallel with collagen (Figs. A3 and A4), the chromatogram consists of several peaks, the intervals between the peaks are roughly constant, and the heights of peaks decrease gradually on the right-hand side of the chromatogram. The discrete distribution of the peaks (the number of which is much smaller than the total number, about 22, of the components) should be a result of the stepwise variation in number of functional ε-amino groups per molecule that are available for the reaction with P sites (*12*). The chromatogram of poly-L-lysine (Fig. A6) would justify indirectly the interpretations given above to multipeak collagen chromatograms (Figs. A3 and A4). From the intervals of the neighboring peaks in the poly-L-lysine chromatogram (Fig. A6), the value of 2–2.2 kcal/mol for the adsorption energy of an ε-amino group onto a P site can be obained (*12*).

In the period when the microheterogeneous collagen model was first proposed (*7*, *10*), the general theory of HA chromatography was in an early stage of development, and the model was criticized only roughly. In Ref. *6* a method for a much more precise calculation of the theoretical chromatogram was developed for mixtures of molecules with the same dimensions and the same shape. Account was taken of both the mutual repulsive interactions

among molecules occurring on the crystal surfaces of HA and the longitudinal diffusion in the column. It can be assumed that collagen molecules with a rodlike shape (Appendix III) are adsorbed in parallel with one another on the crystal surface, avoiding the mutual superposition. The molecular interactions should occur mainly through the sides of the rods (*6*). In the present paper the microheterogeneous collagen model is reexamined on the basis of the new theory (*6*). Satisfactory fits are obtained between the theoretical and experimental results. From the width in the distribution of the peaks in the multipeak chromatogram, it can be estimated that the mean probability, π, that aspartic and glutamic residues (with carboxyl groups) exist at "correct" positions on the three α collagen peptide chains is about 0.8. It can be suggested, however, that the corresponding probabilities are much closer to unity with both glycyl and imino residues (see Appendix III). The following conclusions are derived at the same time. (a) The energy of adsorption, ε, for a carboxyl group of collagen onto a C site of HA is about 0.7 kcal/mol. (b) The repulsive interaction energy, $\tilde{E}$, for one of collagen molecules with others on the crystal surface is about 3 kcal/mol, provided one of the two sides of the rod is completely brought into contact with (or, more precisely, keeps the minimum distance from) other molecules that are adsorbed side-by-side. (c) The square root, h', of the area on the molecular surface where a carboxyl group can move freely without considerable loss of energy (around the mean position) is 2.5–3.5 Å. And (d) it is likely that a considerable adsorption of phosphate ions from the buffer occurs onto the molecular surface of collagen adsorbed on the crystal surface. The experimental values obtained in points (a)–(c) above are all reasonable ones. This, arguing in the other direction, strongly supports the microheterogeneous model itself.

A brief introduction of collagen is made in Appendix III. It can be suggested that ambiguous primary structures of the three α chains arise at a translational level of biosynthesis (see Appendix III).

In the arguments below, the numbers of equations bearing an asterik represent the equations that appear in Ref. *6*.

II. THEORETICAL

A. Distribution $D(\psi)$ of the Parameter $\psi \equiv x + (kT/\varepsilon) \ln \tau$ in the Microheterogeneous Collagen Mixture: The Relationship with the Chromatogram

Provided there are no mutual molecular interactions nor longitudinal diffusion in the column, the elution phosphate molarity, m, of any collagen

component should be represented by Eq. (10*). The factor, $q_{(\rho')}$, involved in Eq. (10*) (cf. Eq. 5*) can be written as

$$q_{(\rho')} = \beta e^{\psi\varepsilon/kT} \tag{1}$$

where

$$\psi \left(\equiv -\frac{Q_{(\rho')}}{\varepsilon} \right) = x + \frac{kT}{\varepsilon} \ln \tau \tag{2}$$

In Eq. (2), $-\varepsilon(\varepsilon > 0)$ represents the adsorption energy for a carboxyl group of collagen onto one of the C sites of HA; x, the average number, in the equilibrium state, of carboxyl groups per molecule that react with C sites; and τ, the number of effective geometrical configuration(s) of a molecule on the HA surface in the equilibrium state. Since collagen has a rodlike shape (Appendix III), it can be assumed that the molecule is adsorbed by using the lateral surface. The geometrical molecular configuration on the crystal surface is limited within the rotation of the rod around its main axis (cf. Section B). By taking into account the Boltzmann distribution for the molecular configuration, x and $\ln \tau$ can be represented by Eqs. (A2*) and (A3*) (where x and τ are written as $x_{(\rho')}$ and $\tau_{(\rho')}$), respectively. It can be considered that $-\psi\varepsilon$ or $Q_{(\rho')}$ represents a type of adsorption free energy per molecule (Ref. 6, Appendix). Therefore, the elution molarity, m (Eq. 10*), is governed by ψ. Moreover, m increases almost linearly with an increase of ψ, because Eq. (10*) can be written approximately as

$$m \approx \frac{1}{x'\varphi'} \ln (x'\varphi' s\beta) + \frac{\varepsilon}{x'\varphi' kT} \psi \tag{3}$$

The approximate linearity between m *and* ψ can numerically be confirmed (cf. Figs. 4b and 5); this means that the distribution, $D(\psi)$, approximately represents the chromatogram itself.

It can be assumed that the mean elution molarity of the multipeak collagen chromatogram in Fig. A3(d) is almost equal to the molarity that should occur with an infinitesimal sample load or in the absence of molecular interactions. Thus, in Fig. 1, the two unfilled circles show plots of elution phosphate molarities at the centers of gravity of the experimental chromatograms in Fig. A3(d) (left-hand side circle) and Fig. A4(b) (right-hand side circle) versus sample loads. In both Figs. A3(d) and A4(b), the slope, $g'_{(P)}$, of the phosphate gradient is the same (3.75×10^{-5} M/mL), and the length, L, of the column also is almost identical [281 cm (Fig. A3d) and 285 cm (Fig. A4b)]. This would mean that the extrapolated value of these two experimental molarities to sample load zero gives the elution molarity at sample

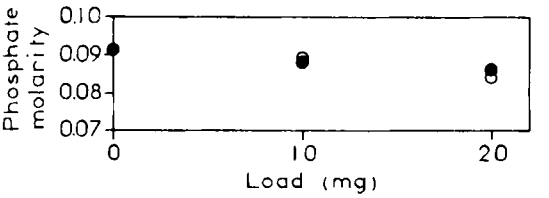

FIG . 1.Open circles: Plots of elution phosphate molarities at the centers of gravity of experimental chromatograms in Figs. A3(d) (left-hand-side circle) and Fig. A4(b) (right-hand-side circle) in Appendix I versus sample loads. (Reproduced from Fig. 5 in Ref. *13*.) Filled circles: Corresponding plots for theoretical chromatograms in Fig. 5(d) (left-hand-side circle), Fig. 8(d) (middle circle), and Fig. 9(b) (right-hand-side circle).

load zero that should be realized under the same experimental condition as that in Fig. A3(d). It can be seen in Fig. 1 that the extrapolated molarity is almost equal to the molarity for the left-hand side unfilled circle. In Section B the calculation of $D(\psi)$ will be performed in such a way that the elution molarity at the center of gravity of the theoretical multipeak chromatogram obtained by using Eq. (10*) be almost equal to the molarity at the center of gravity of the experimental chromatogram in Fig. A3(d).

For the parameters involved in Eq. (10*) (or Eq. 3), the following values will be used: $\varphi' = 6.7\ M^{-1}$ (*3, 8*), $x' = 40$ (*3* see below), $\beta = 2 \times 10^{-3}$ (see below), $T = 4°C$, $m_{in} = 0.001\ M$ (see the legend of Fig. A1), and $s = 2.36 \times 10^{-5}$ M/cm $\times$ 281 cm $= 6.36 \times 10^{-3}\ M$, where 2.36×10^{-5} is the slope $g'_{(P)}$ ($= 3.75 \times 10^{5}$ M/mL of the phosphate gradient expressed in unit of M/cm (Ref. *3*; see the legend of Fig. A3). As collagen is homogeneous in both dimensions and shape (Appendix III), the assumption of a constant x' value among different components is reasonable (but see Discussion Section E). The value of β was estimated on the basis of both the β value for P crystal sites (*12*) and some assumptions concerning the ratio between P and C crystal surfaces. Actually, however, the chromatogram depends only slightly upon the β value because, in Eq. (3), β appears in the logarithmic term, ε and ψ taking values of the order of 0.5 (kcal/mol) and 20, respectively (see Section B).

B. Calculation of $D(\psi)$

Earlier (*7*), it was mentioned that to consider the geometrical configuration of a collagen molecule on the crystal surface of HA is equivalent to replacing the actual lateral molecular surface by an abstact surface that consists of a number of elementary surfaces with an area of h^2 Å^2, respectively, and to partitioning carboxyl groups among these elementary surfaces. This consideration is improved below.

Thus it can in general be assumed that collagen is adsorbed onto one or more than one array of C sites on the crystal surface (cf. Discussion Section B). Provided the molecule is adsorbed onto only one array of C sites, let us consider a region with a finite width h_1 over the array. It can be assumed that only carboxyl groups on the molecular surface facing this region can react with C sites. If the molecule is adsorbed onto more than one array of C sites, carboxyl groups existing over the regions for any arrays under consideration would react with C sites. However, in general it is possible that the crystal regions that belong to different arrays of C sites are overlapped with one another. The same carboxyl group can react with more than one of the C sites that belong to different arrays. Even if the regions belonging to different arrays of crystal sites are separated from one another, the widths of different regions can be different. Therefore, in any instance when the molecule is adsorbed onto more than one array of sites, let us define h_1 as the quantity obtained by dividing the total width of the crystal regions where the adsorption occurs by the number of the regions, i.e., the number of the arrays of C sites under consideration. In the case of overlapping, the overlapped parts are subtracted from the total width. Now, the parameter α_1 defined as

$$\alpha_1 = nh_1/l \tag{4}$$

where n is the number of the arrays of C sites under consideration and l is the distance around the main axis of the rodlike molecule on the molecular surface, gives the probability that, in a given molecular configuration, a carboxyl group exists on a part or parts of the lateral molecular surface facing the n crystal region(s). The total width of the part(s) of the lateral molecular surface facing the n crystal region(s) should be almost equal to nh_1.

Let us define another width h_2 over each crystal site. This width is measured in the direction parallel to the array of C sites under consideration, in contrast with h_1 that was measured perpendicular to the array. We assume that, of carboxyl groups that exist on the lateral molecular surface facing the n crystal region(s) (over the n array(s) of C sites), only those existing within the widths h_2 (concerning respective C sites), can react with the C sites. If the widths h_2 belonging to different C sites are overlapped, we redifine h_2 to be equal to the interdistance between neighboring C sites on the array under consideration. Now, the parameter α_2 defined as

$$\alpha_2 = h_2/d \tag{5}$$

where d is the interdistance between the neighboring C sites on the array on which collagen is adsorbed, gives the probability that a carboxyl group that exists on the lateral molecular surface facing the n crystal region(s) reacts with a C site.

Therefore, the parameter ϕ defined as

$$\phi = \alpha_1 \alpha_2 = h^2/ld \tag{6}$$

where

$$h = \sqrt{n}h' \tag{7}$$

and

$$h' = \sqrt{h_1 h_2} \tag{8}$$

gives the probability that, in a given molecular configuration on the crystal surface, a given carboxyl group existing on the total lateral molecular surface reacts with a C site.

It would be reasonable to assume that the carboxyl group (attached at the top of the side chain of aspartic or glutamic residue) can move almost isotactically around the mean position on the molecular surface. Therefore, if both h_1 and h_2 are smaller than, or equal to, the minimal interdistance between the neighboring C sites measured in any direction on the crystal surface, then both h_1 and h_2 should be almost equal to h'. h'^2 should measure the area on the lateral molecular surface in which a carboxyl group can move freely without considerable loss of energy (in order for it to react with a C site). If h_1 or h_2 is greater than the minimal interdistance between the C sites, a modified interpretation should be given to h'.

Let us introduce an integer ν that has the closest value to $1/\phi$:

$$\nu \approx 1/\phi \tag{9}$$

ν is related to h by the relationship

$$\nu \approx ld/h^2 \tag{10}$$

(see Eq. 6). Replacing $1/\phi$ with an integer ν means replacing the actual lateral collagen surface with an abstract surface with (almost) the same total area as the original surface. The abstract surface consists of ν elementary domains, and each elementary domain consists of a number of elementary surfaces with an area of h^2 Å^2, respectively. In the arguments below, the terminologies "elementary domain" and "elementary surface" should definitively be distinguished. ϕ now represents the probability that, in a given molecular configuration on the crystal surface, a given carboxyl group belongs in a particular one of the ν elementary domains. Only in that elementary domain is the reaction with the C crystal site possible. Further, calling X the total number of carboxyl groups per molecule, it is possible to

define the probability that X carboxyl groups are partitioned among ν elementary domains on the abstract molecular surface with the proportion $x_1{:}x_2{:}\cdots x_\nu$ where $x_1, x_2, \ldots, x_\nu$ fulfill the relationship

$$\sum_{j=1}^{\nu} x_j = X \tag{11}$$

and each x_j can be

$$x_j = 0, 1, 2, \ldots, X \tag{12}$$

Since the total number of elementary surfaces is finite, the maximum possible value of x_j can be less than X. However, as the total number of amino residues per molecule (about 3000) is much greater than the number X of acidic residues with carboxyl group (about 230; see Appendix III), the total number of elementary surfaces should also be much greater than X (cf. Discussion Section D). The approximation of Eq. (12) practically would not influence the final result of the calculation.

In Eq. (11), j and ν have physical meanings of the jth-type geometrical configuration of the molecule on the crystal surface and the total number of possible geometrical configurations, respectively. x_j, therefore, represents the number of carboxyl groups per molecule that react with C sites provided the jth-type configuration is being realized. $x_1, x_2, \ldots, x_\nu$ are related to x and $\ln \tau$ (see Eq. 2) through Eqs. (A2*)–(A4*) (where x and τ are written as $x_{(\rho')}$ and $\tau_{(\rho')}$, respectively).

According to the microheterogeneous collagen model (Introduction Section), acidic amino residues exist at "correct" positions on the three α peptide chains (Appendix III) with a mean probability, π, that is less than unity. The "correct" positions are those that can be "determined" by the organochemical method. It can almost be confirmed (Appendix III) that the total number (per peptide chain) of acidic amino residues estimated on the basis of the organochemically "determined" primary structure (i.e., the total number of "correct" positions) is almost equal to the number estimated from the amino acid analysis (i.e., the total number itself of acidic amino residues). Further, as X (= 230) is large, the X value would almost be constant among different collagen components even in the presence of the ambiguous primary structure (cf. Appendix III). This would mean that X carboxyl groups are partitioned semistatistically among elementary surfaces of each collagen component. With a mean probability, π, each of the X carboxyl groups are localized on each of the X particular elementary surfaces that correspond to the X "correct" positions of the acidic amino residues on the α chains.

It is now necessary to know the distribution law of the X particular elemental surfaces (called hereafter "correct" elementary surfaces) on the

abstract molecular surface, or the partition law of the X "correct" elementary surfaces among the ν elementary domains. This partition law should be common among different collagen components. Earlier (7), a reasonable assumption for the partition law was proposed. Thus we rearrange the value of j, which indicates the ν respective elementary domains, in decreasing order following the magnitude of the number of partitioned "correct" elementary surfaces. Writing $\zeta_1, \zeta_2, \ldots, \zeta_\nu$ for these numbers, we therefore have

$$\zeta_1 \geq \cdots \geq \zeta_j \geq \zeta_{j-1} \geq \cdots \geq \zeta_\nu \tag{13}$$

where

$$\sum_{j=1}^{\nu} \zeta_j = X \tag{14}$$

and each ζ_j can be

$$\zeta_j = 0, 1, 2, \ldots, X \tag{15}$$

Some comment on Eq. (12) (see above) can also be applied to Eq. (15). We now make M attempts to partition at random the X "correct" elementary surfaces among the ν elementary domains, and define the average value, $\bar{\zeta}_j$, of ζ_j as

$$\bar{\zeta}_j = \lim_{M \to \infty} \frac{1}{M} \sum_{m=1}^{M} (\zeta_j)_m \tag{16}$$

where $j = 1, 2, \ldots, \nu$, and m indicates the mth trial. We finally assume that the number, $\tilde{\zeta}_j$, of "correct" elementary surfaces partitioned to the jth elementary domain is

$$\tilde{\zeta}_j = [\bar{\zeta}_j + 0.5] \tag{17}$$

where $j = 1, 2, \ldots, \nu$, and the square brackets indicate that the decimal values are discounted in order for $\tilde{\zeta}_j$ to be an integer. According to Eq. (17), the sum of $\tilde{\zeta}_j$ for j can, in general, be slightly different from X. Fortunately, however, the sum in our calculation (Fig. 2) is exactly equal to X. If $\pi = 1$, then $x_j = \tilde{\zeta}_j$ for any value of j and for any collagen component. If $\pi < 1$, the values of x_j for different components are distributed statistically around the mean value $\tilde{\zeta}_j$. It should be emphasized that what is important for our purpose is the distribution of the values of x_j around $\tilde{\zeta}_j$ for each j, and not the exact value of $\tilde{\zeta}_j$.

Figure 2 illustrates a type of calculation of $\tilde{\zeta}_j$ as a function of j when $X = 230$ and $\nu = 18$ (for the assumption of the ν value, see below). Thus the

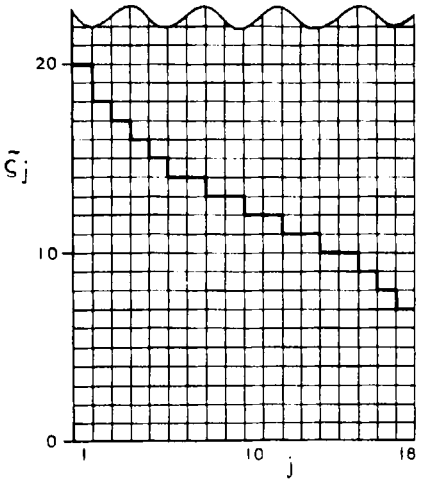

FIG. 2. Abstract surface of collagen. Squares represent elementary surfaces with an area of h^2 Å^2, respectively. A vertical column of squares piled up above a particular value of j represents the jth elementary domain. The ordinate value of the thick folded line gives the value of $\tilde{\zeta}_j$ for each j. The domain under the thick line is composed of X (= 230) "correct" elementary surfaces in which X carboxyl groups should exist provided that $\pi = 1$. When $\pi < 1$, the number, x_j, of carboxyl groups that exist within the jth elementary domain is fluctuating among different collagen components around the mean value $\tilde{\zeta}_j$.

plane consisting of a number of squares represents the total abstract collagen surface. The squares represent the elementary surfaces (with an area of h^2 Å^2, respectively). A vertical column of squares piled up above a particular value of j on the abscissa shows the jth-type elementary domain. The ordinate value of the thick folded line gives the value of $\tilde{\zeta}_j$ for each j. Therefore, the total domain under the thick folded line is composed of X(= 230) "correct" elementary surfaces in which X carboxyl groups should exist provided that $\pi = 1$.

Now, if a set $(x_1, \ldots, x_j, \ldots, x_\nu)$ is assigned to each model molecule in the mixture, the value of ψ for each molecule can be calculated from Eqs. (A2*)–(A4*) and Eq. (2). If the data are accumulated for a large number of molecules with a constant π value, the distribution $D(\psi)$ can be obtained. In an extreme case when $\pi = 0$, the partition of the X carboxyl groups among the ν elementary domains occurs completly at random. In this instance the frequency of the occurrency of a set $(x_1, \ldots, x_j, \ldots x_\nu)$ can be represented analytically (see Appendix II). Here, however, the calculations of $D(\psi)$ for any π values are made by using the Monte Carlo method with 5000 trials. (For details, see the legend of Fig. 3.)

Figure 3 illustrates several types of calculations of $D(\psi)$ for random (part a) and semirandom (parts b–d) distributions of 230 carboxyl groups on the abstract surface of each collagen component. Several different π values are assumed. It also is assumed that $\varepsilon = 0.7$ (kcal/mol) (see Eq. 2) and that $\nu = 18$ (see below). The calculations for the diagrams in parts (b)–(c) of Fig. 3 have been made by using Fig. 2. The value, 0.7 kcal/mol, of ε is that which finally is concluded to be best (see Section C). The distribution $D(\psi)$ depends upon the ν value. If $\varepsilon = 0.7$ kcal/mole, then, for any value of π, it is when $\nu = 18$ that the phosphate molarity at the center of gravity of the theoretical chromatogram calculated by using Eq. (10*) coincides best with the corresponding molarity for the experimental chromatogram in Fig. A3(d) (see Section A). This is the reason why the value of 18 has been chosen for ν. $D(\psi)$ should approximately represent the chromatogram (Section A). By comparing Fig. 3 with both Figs. A3 and A4, it can be found that the distribution $D(\psi)$ is very similar to the experimental chromatograms. Thus, first, $D(\psi)$ consists of several discrete peaks with (almost) the same intervals between them as in experimental chromatograms. Second, the heights of the peaks decrease gradually on the right-hand side of the distribution. And, lastly, some components are concentrated in the extreme left-hand part of the distribution. Especially, it is when $\pi = 0.8$ (Fig. 3c) that a best fit with the experiment is obtained concerning the width in the distribution of the peaks or the total number of peaks in the chromatogram. This will be shown more explicitly if the theoretical chromatogram is calculated by taking into account both the mutual molecular interactions and the longitudinal diffusion in the column (Section C).

Part (a) of Fig. 4 illustrates the distribution, D_M, of the maximum value, x_M, of $x_1, \ldots, x_j, \ldots, x_\nu$, i.e.,

$$x_M = \max(x_1, \ldots, x_j, \ldots, x_\nu) \tag{18}$$

for the case when $\pi = 0.8$, or for model molecules that are identical with those used for the calculation of the diagram in Fig. 3(c). This diagram is reproduced in part (b) of Fig. 4. As x_M takes only integral values, $D_M(x_M)$ represents a completely discrete distribution. The values of this distribution are shown on the right-hand side ordinate axis in Fig. 4(a). However, in order to compare this distribution with both distributions $D(\psi)$ (Fig. 4(b)) and $D'(x)$ (Fig. 4c; see below). In Fig. 4(a), each peak of the distribution is represented as a triangle with a mean width of 0.2. This is equal to the width of elements in the abscissa in both parts (b) and (c) of Fig. 4 that has been introduced for convenience' sake (see the legend of Fig. 3). On the left-hand ordinate axis in Fig. 4(a) are shown the D_M values that are normalized in order for the total areas under all the triangular peaks to be equal to unity.

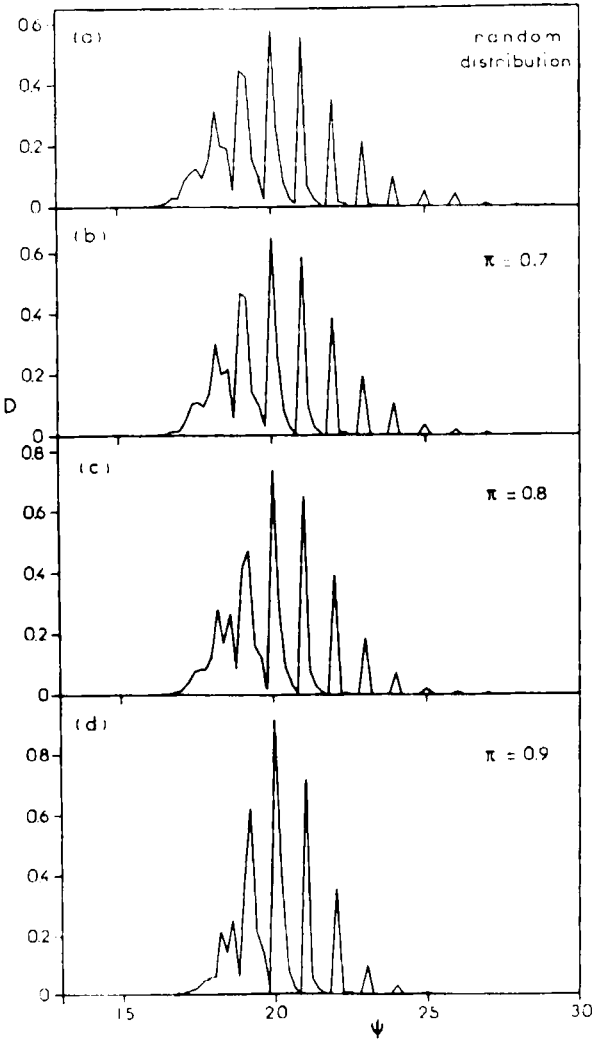

FIG. 3. Functions $D(\psi)$ for random (a) and semirandom (b)–(d) distributions of X (= 230) carboxyl groups on the abstract surface of each collagen component that consists of ν (= 18) elementary domains. It is assumed that $\varepsilon = 0.7$ kcal/mol. Calculations of the diagrams were performed by using the Monte Carlo method with 5000 trials; the current ψ value was divided into elements of 0.2 in. size, and the probabilities of occurrences of ψ values between $\psi' - 0.1$ and $\psi' + 0.1$ were estimated, respectively, where $\psi' = I, I + 0.2, I + 0.4, \ldots$, and I represents integers. The curves $D(\psi)$ are drawn by connecting the probabilities occurring in respective elements. If elements smaller than 0.2 in. size are used, substructures appear in the curves. Due to longitudinal molecular diffusion in the column, however, these substructures do not appear in the final chromatogram. For the calculations of the diagrams in (b)–(d), the set $(x_1, \ldots,$

Part (c) of Fig. 4 illustrates the distribution D' of the x values for the same model molecules as those for both parts (a) and (b). Now, if ε approaches infinity, then ψ tends to x (see Eq. 2), and x tends to x_M (compare Eqs. A2* and A4* with Eq. 18). This means that both $D(\psi)$ and $D'(x)$ tend to $D_M(x_M)$. In this instance, any molecule in the mixture is adsorbed on the HA surface in the energetically most stable configuration(s) by using the maximum possible number of carboxyl groups. Actually, due to the finite value of ε, however, the number of adsorption groups per molecule reacting with crystal sites makes a Boltzmann distribution (Eq. A4*). Figure 4(c) shows that the average value, x, of the reacting adsorption groups still makes essentially a discrete distribution when $\varepsilon = 0.7$ (kcal/mol). However, some components are more concentrated in the left-hand part of the distribution (Fig. 4c) than with the energetically most stable adsorption (Fig. 4a). With distribution $D(\psi)$, the concentration of the components becomes still eminent (Fig. 4b). This is because the smaller the x value, the more important is the contribution of the entropy term, $\ln \tau$, to the parameter ψ (see Eq. 2). It can be shown that the concentration of some components in the left-hand part of the distribution increases with decrease in the ε value in both $D'(x)$ and $D(\psi)$. Therefore, by comparing the degree of this concentration with that seen in the experimental chromatogram, the value of ε can roughly be estimated. It is by using this method that earlier (*10*) a value of the order of 0.5 kcal/mol was obtained for ε (see Introduction Section). In this estimation, however, the distribution $D'(x)$ was used as an approximation of $D(\psi)$, where the approximation of $x' = \infty$ is also involved.

C. Theoretical Chromatograms

Figure 5 illustrates several theoretical chromatograms obtained in the absence of mutual molecular interactions. It is assumed that $\varepsilon = 0.7$ kcal/mol and that $\pi = 0.8$, the corresponding function $D(\psi)$ being depicted in both Figs. 3(c) and 4(b). The calculation of Fig. 5 was made by using Eqs.

$x_j, \ldots, x_\nu$) was assigned to each molecular component by using Fig. 2. The method of this assignment is slightly different from that proposed earlier (*7*). Thus, by using a uniform series of real random numbers comprized between 0 and 1 produced by the computer, a real number, p, is assigned to every elementary surface under the thick folded line $\tilde{\zeta}_j$ in Fig. 2. If $p \leq \pi$, a carboxyl group is assumed to enter this elementary surface. If $p > \pi$, the real number p is again produced. This number is transformed into an integer j' comprized between 1 and ν. It is assumed that the carboxyl group enters an elementary surface that belongs in both the j'th elementary domain and the domain above the thick folded line in Fig. 2. After having finished the trials for all X elementary surfaces existing under the thick folded line $\tilde{\zeta}_j$, the total number of carboxyl groups is estimated for every elementary domain or every vertical column of the elementary surfaces. This represents the value of x_j.

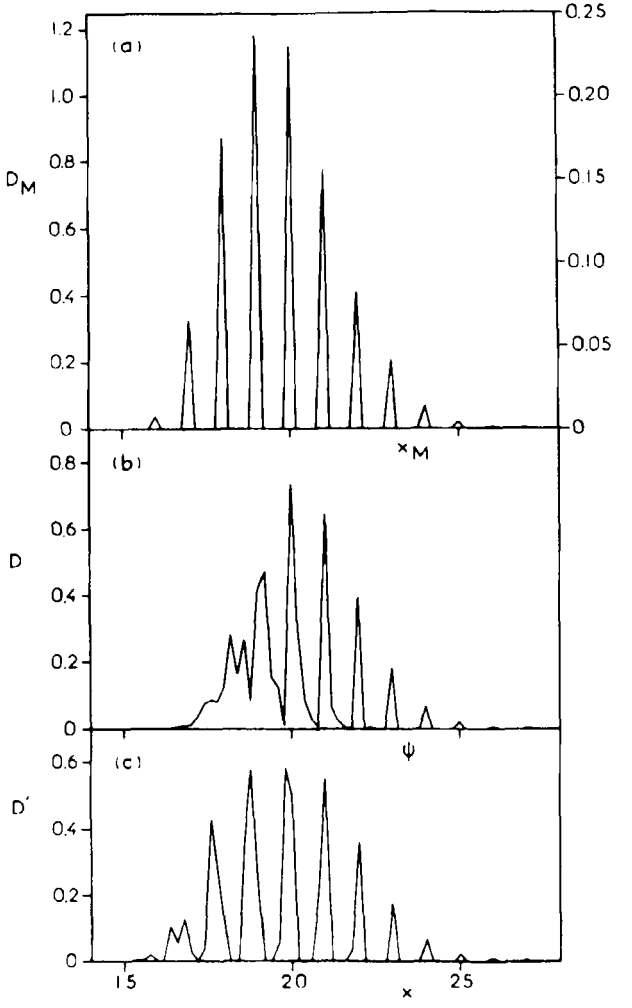

FIG. 4. (a) Distribution, D_M, of the maximum value, x_M, of $x_1, x_2, \ldots, x_\nu$. (b) Distribution, D, of ψ. (c) Distribution, D', of the average value, x, of $x_1, x_2, \ldots, x_\nu$. The same 5000 model molecules are used as those used for the calculation of Fig. 3(c). The diagram in (b) is identical with that in Fig. 3(c). Thus it is assumed $X = 230$, $\nu = 18$, and $\pi = 0.8$, and, in (b) and (c), it also is assumed $\varepsilon = 0.7$ kcal/mol. As x_M takes only integral values, $D_M(x_M)$ (a) makes a completely discrete distribution. Each peak in the distribution $D_M(x_M)$ is represented as a triangle with a mean width of 0.2, however (see text). The method of the representation of the distribution $D'(x)$ (c) is the same as that of $D(\psi)$ (b) which is explained in detail in the legend of Fig. 3. $D_M(x_M)$ represents the limiting distributions of both $D(\psi)$ and $D'(x)$ when ε approaches infinity.

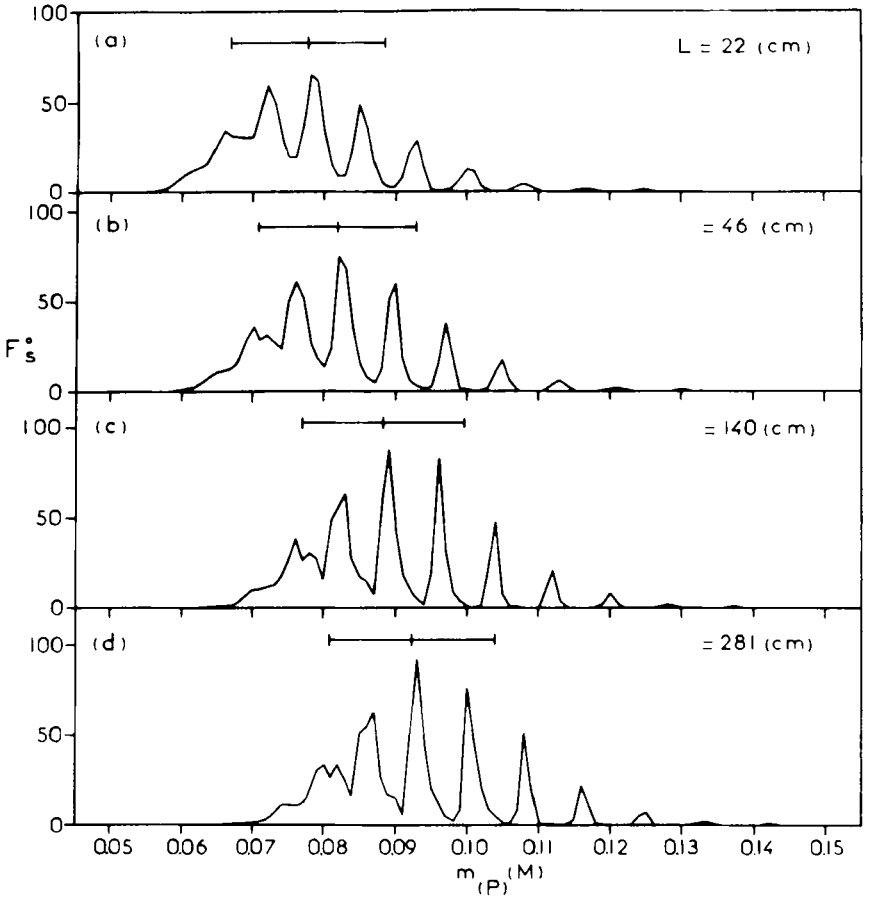

FIG. 5. Theoretical chromatograms in the absence of mutual molecular interactions, obtained for the microheterogeneous collagen model with $\varepsilon = 0.7$ kcal/mol and $\pi = 0.8$. Both centers of gravity and the standard deviations of the total chromatograms are shown. The corresponding function $D(\psi)$ is depicted in both Figs. 3(c) and 4(b). Account is taken of the longitudinal molecular diffusion in the column; the value 0.3 cm is used for the diffusion parameter θ_0. The experimental conditions in (a)–(d) are identical with those in the experimental chromatograms shown in the corresponding parts of Fig. A3 in Appendix I.

(11*)–(17*), taking into account the longitudinal diffusion in the column. The value 0.3 cm was used for the diffusion parameter θ_0 (eq. 11*) which had been estimated earlier (*3*). The experimental conditions in parts (a)–(d) of Fig. 5 are identical with those in the experimental chromatograms shown in the corresponding parts of Fig. A3. It can be seen in Fig. 5 that, when the column is long, the width in the total theoretical chromatogram is similar to the width in the experimental chromatogram. When the column is short, however, the width in the experimental chromatogram becomes larger than that in the theoretical chromatogram. This is due to the fact that, when the ratio of the sample load to the column length is high, the elution positions of the components that are eluted at relatively small phosphate molarities are displaced to smaller molarities in the presence of the components with high elution molarities. The displacement effect can also be seen in both Figs. A4(a) and (b) in comparison with Figs. A3(b) and (d), respectively. The displacement effect and the deformation of the total shape of the chromatogram occurring associated with the former effect can both be explained only by assuming the existence of replusive interactions among sample molecules adsorbed on the HA surfaces (*6*).

Figures 6–9 depict theoretical chromatograms calculated by taking into account both the mutual repulsive molecular interactions on the crystal surfaces and the longitudinal diffusion in the column. The molecular model is the same as that used in Fig. 5 with both $\varepsilon = 0.7$ kcal/mol and $\pi = 0.8$. For the calculation, Eqs. (1*), (3*)–(8*), and (24*) were used. The value 3 kcal/mol was assumed for $\tilde{E}$ (Eq. 8*), the repulsive interaction energy for one of collagen molecules with another when these make a maximum contact. The experimental conditions in Figs. 6–9 are identical with those in the experimental chromatograms shown in the corresponding parts of Figs. A1–A4, respectively. Some details for the method of calculations of Figs. 6–9 are shown in the legend of Fig. 6. It can be seen in Figs. 6–9 and Figs. A1–A4 that satisfactory fits are obtained between the theoretical and experimental results. Both the shapes and the widths of the corresponding chromatograms obtained under any experimental conditions resemble each other. The fact that the resolution of the column increases, in general, with a decrease in the slope of the phosphate gradient (in order of Fig. A1, Fig. A2, and Figs. A3 and A4) is explained satisfactorily by the theory. Especially under high resolution conditions, good fits are obtained concerning the total number of peaks. Experimental chromatograms, in general, begin more abruptly at their left-hand parts than theoretical chromatograms, however. A possible explanation for this difference will be given in Discussion Section A.

The intermediate and right-hand side filled circles in Fig. 1 show plots of elution molarities at the centers of gravity of theoretical chromatograms in Figs. 8(d) and 9(b) vs sample loads, respectively; these correspond to the

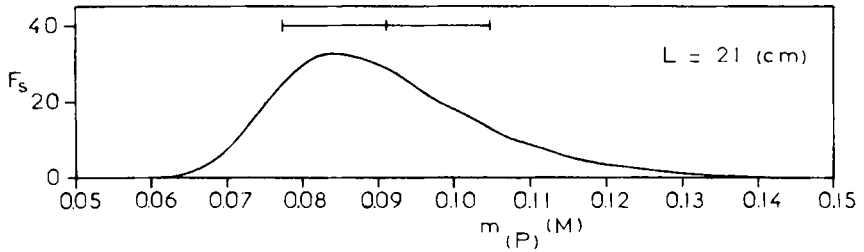

FIG. 6. Theoretical chromatogram for the microheterogeneous collagen model (the same as that used in Fig. 5) with $\varepsilon = 0.7$ kcal/mol and $\pi = 0.8$, the corresponding function $D(\psi)$ being depicted in both Figs. 3(c) and 4(b). The experimental condition is identical with that in the experimental chromatogram shown in Fig. A1 in Appendix I. Account is taken of both the mutual repulsive interactions among molecules occurring on the crystal surfaces of HA and the longitudinal diffusion in the column. For mutual interactions it is assumed that $\tilde{E} = 3$ kcal/mol. Concerning the longitudinal diffusion, the value 0.3 cm is used for the parameter θ_0; it also is assumed that the value of $\bar{\sigma}_x$ (eq. 24*) should be equal to the standard deviation of the chromatographic peak (with infinitesimal sample load) for a single collagen component that is involved just at the center of gravity of the distribution $D(\psi)$. For the calculation of the standard deviation of the peak of the single collagen component (see above), eqs. (11*)–(14*) were used. (For the principle of this method for the calculation of $\bar{\sigma}_s$, see Ref. 6) Instead of $\chi^*_{(\rho'')}$ (eqs. 3* and 7*) for molecular component ρ'' (where $\rho'' = 1, 2, \ldots, \rho$), the amount of components existing within the elementary widths between $\psi' - 0.1$ and $\psi' + 0.1$ of the variable ψ (eq. 2) has been considered (where $\psi' = I, I + 0.2, I + 0.4, \ldots$, and I represents integers; cf. the legend of Fig. 3). This amount should be represented in unit such that it is equal to unity provided the whole column is saturated with molecules. For this representation it is necessary to know both the reduced amount α (to column diameter = 1 cm) of the sample load and the effective surface area of HA per unit volme of packed crystals in the column. The former value is shown in the legend of Fig. A1 in Appendix I; the latter value was estimated to be 1.6×10^4 cm^2/mL, both referring to the adsorption experiment of collagen (*11*, Appendix I) and on the basis of the assumption of the compact adsorption of collagen on the HA surface (Section III).

experimental chromatograms in Figs. A3(d) and A4(b), respectively. The corresponding plots for the latter two chromatograms are shown by unfilled circles in Fig. 1 (see Section A). It can be seen that the experimental plots coincide satisfactorily with the theoretical ones. The extreme left-hand side filled circle in Fig. 1 shows the plot for the theoretical chromatogram with infinitesimal sample load in Fig. 5(d). The experimental condition (except sample load) for Fig. 5(d) is identical with those in both Figs. 8(d) and A3(d), and almost identical with those in both Figs. 9(b) and A4(b). It can be seen in Fig. 1 that the elution molarity for the theoretical chromatogram in Fig. 8(d) (intermediate filled circle) and that for the corresponding experimental chromatogram in Fig. A3(d) (left-hand side unfilled circle) are both almost equal to the molarity for the theoretical chromatogram with infinitesimal sample load in Fig. 5(d) (left-hand side filled circle). This result is self-

consistent with a fundamental assumption for the calculation of the function $D(\psi)$ (the second paragraph in Section A).

The values 0.7 kcal/mol, 3 kcal/mol, and 0.8, for the parameters, ε, $\tilde{E}$ and π, respectively, which have been used for the calculations of Figs. 6–9, can be considered to be best values. Earlier (*3*) the value of about 0.5 kcal/mol was obtained for ε from the measurements of the intervals between neighboring peaks in the multipeak chromatogram in Fig. A3(d) (see Introduction Section). In order to calculate, by using this value of ε, theoretical chromatograms which coincide best with the experimental chromatograms in both Figs. A3 and A4 (concerning the intervals between the neighboring peaks), it is necessary to assume the value of about 5 kcal/mol for $\tilde{E}$. Under this assumption, however, the width of each peak in the total chromatogram becomes too large in comparison with the interval between the peaks. Especially when the column is short, all peaks become continuous, and the chromatogram that apparently consists of almost only one large peak is obtained. This is the reason why a slightly larger value, 0.7 kcal/mol, of ε and a slightly smaller value, 3 kcal/mol, of $\tilde{E}$ have been chosen.

Figures 10(a) and (b) depict two theoretical chromatograms obtained under the same experimental condition as that realized in both Figs. 8(d) and A3(d), assuming that the distribution of carboxyl groups on the lateral collagen surface is completely random and that $\pi = 0.9$, respectively. It also is assumed that $\varepsilon = 0.7$ kcal/mol and that $\tilde{E} = 3$ kcal/mol. The corresponding functions $D(\psi)$ for Figs. 10(a) and (b) are shown in Figs. 3(a) and (d),

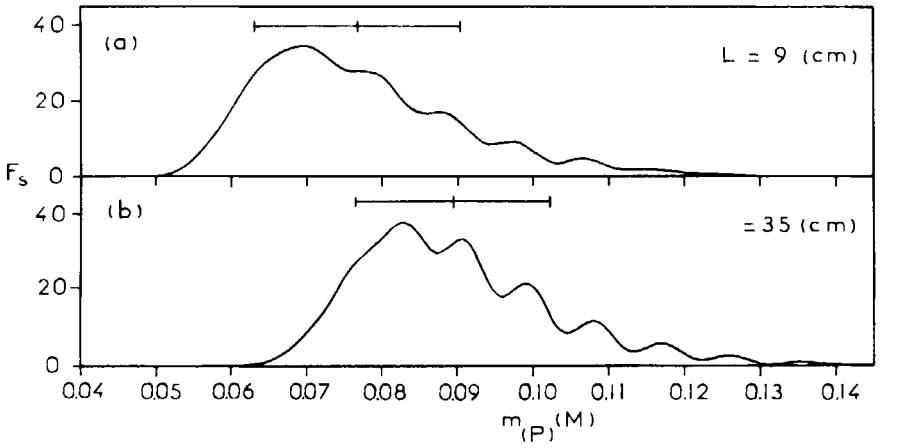

FIG. 7. As Fig. 6 but for experimental conditions identical with those in the corresponding parts of Fig. A2 in Appendix I.

respectively. By comparing Fig. 10 with Fig. A3, it can be understood that, provided the distribution of carboxyl groups is random, both the total width and the number of peaks in the theoretical chromatogram are too large in comparison with the experimental diagram. If $\pi = 0.9$, a reverse relationship occurs between the theoretical and experimental chromatograms.

Figure 11 illustrates theoretical chromatograms that correspond to those in Fig. 8. The value 0.7 is assumed for π instead of the value 0.8 in Fig. 8. The function $D(\psi)$ for Fig. 11 is shown in Fig. 3(b). The widths of the total chromatograms in Fig. 11 are slightly larger than those in the corresponding parts in Fig. 8. It appears that it is Fig. 8, rather than Fig. 11, that coincides best with Fig. A3, the experimental chromatograms.

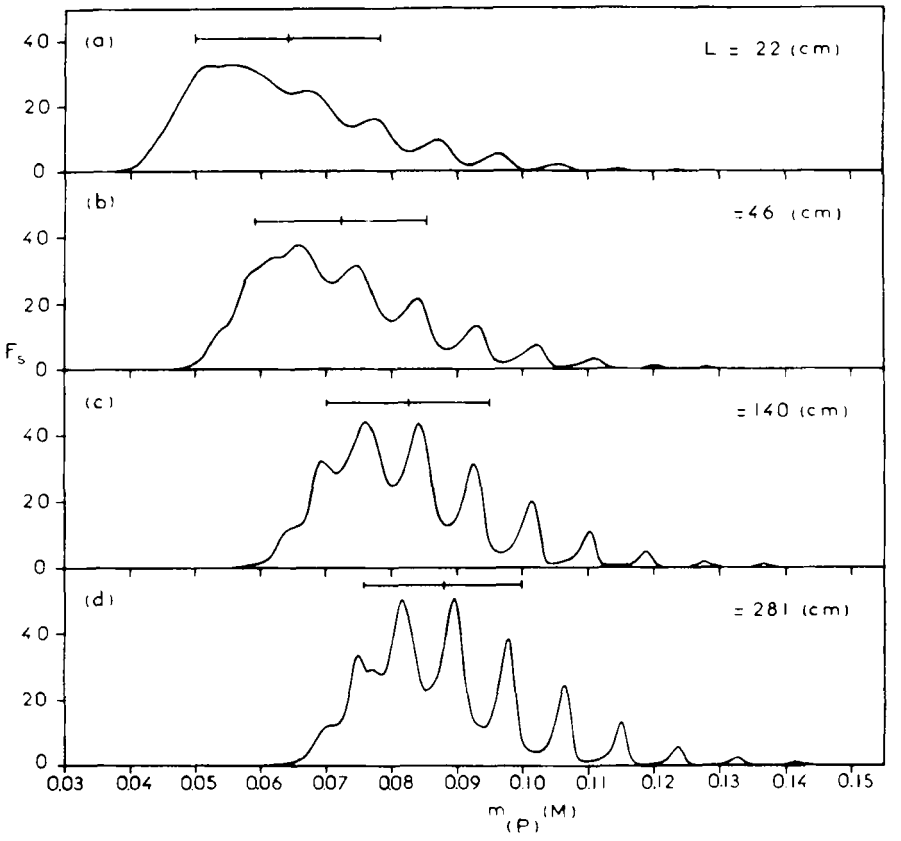

FIG. 8. As Fig. 6 but for experimental conditions identical with those in the corresponding parts of Fig. A3 in Appendix I.

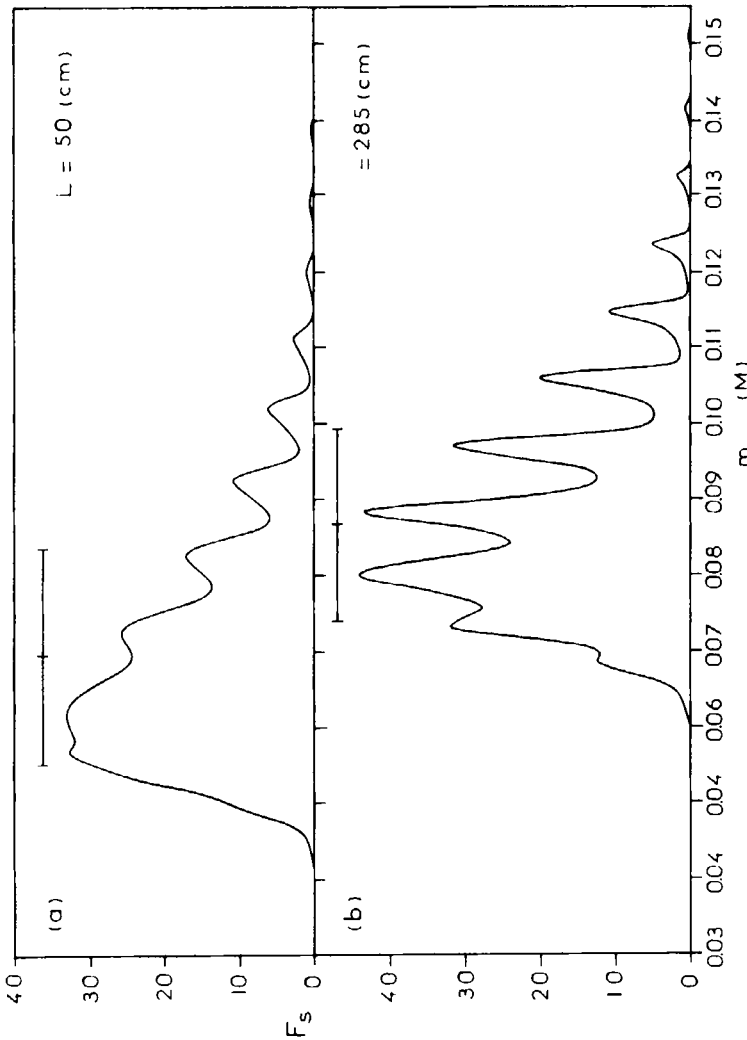

FIG. 9. As Fig. 6 but for experimental conditions identical with those in the corresponding parts of Fig. A4 in Appendix I.

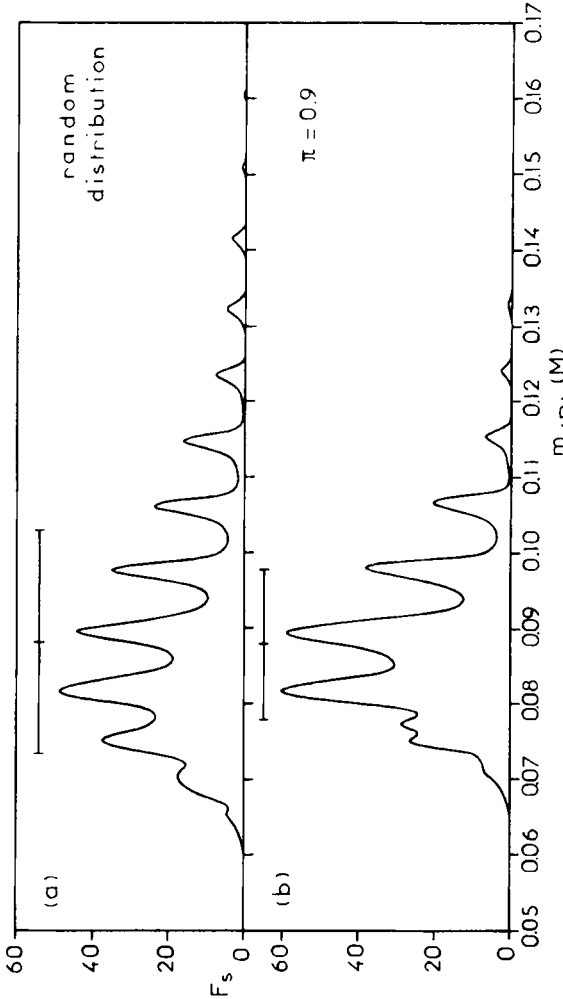

FIG. 10. (a) As Fig. 8(d) but for the case when the distribution of carboxyl groups on the abstract collagen surface is random. The corresponding function $D(\psi)$ is shown in Fig. 3(a). (b) As Fig. 8(d) but for the case where $\pi = 0.9$. The corresponding function $D(\psi)$ is shown in Fig. 3(d).

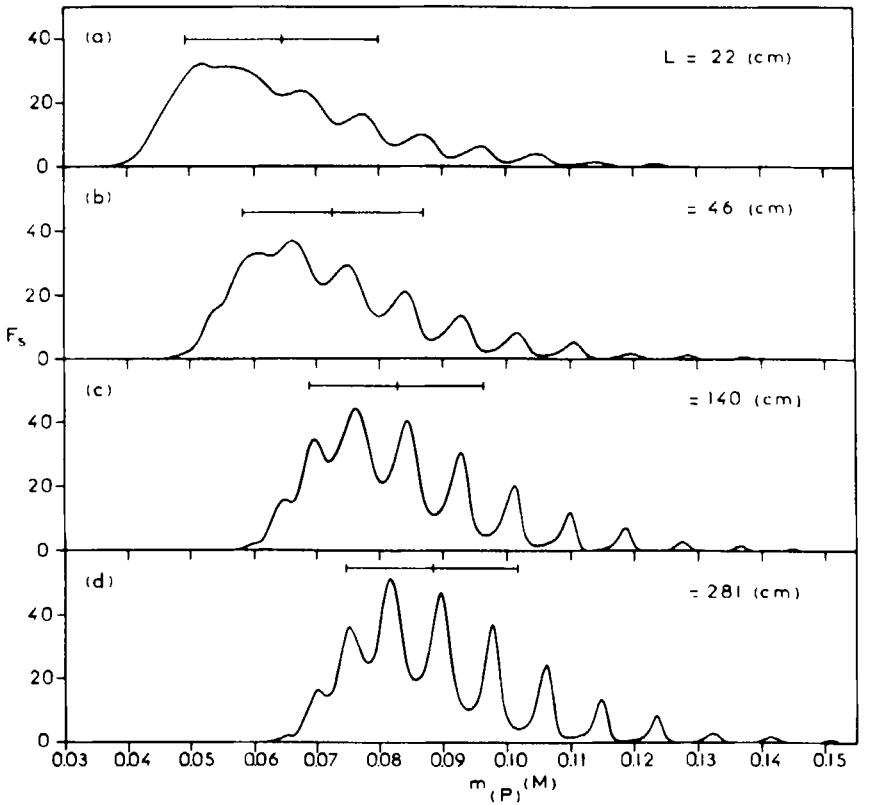

FIG. 11. As Fig. 8 but for the case where $\pi = 0.7$. The corresponding function $D(\psi)$ is shown in Fig. 3(b).

III. DISCUSSION

A. Why Do Experimental Chromatograms Begin More Abruptly at Their Left-Hand Parts than the Theoretical Chromatograms?

Compare Figs. A1–A4 with Figs. 6–9. It can be suggested that this discrepancy arises from the approximations involved in the theory (a) that the distribution of molecules on the crystal surfaces follows a Bragg-Williams approximation, (b) that only the short-range molecular interactions are of importance, and (c) that the probability that any part of a side of a rodlike molecule is brought into contact with other molecules is proportional to the square-root of the molecular density on the crystal surface (*6*). In contrast to

these approximations, let us examine another extreme approximation such that, when the molecular density on the crystal surface is high, there can be two adsorption states of molecules. In the first state, due to the importance of the entropy factor, the molecules are assumed to be distributed at random on the crystal surface while keeping mutually parallel orientations (called as a random phase). From the analogy with the usual solution, if the molecular density is low, it would be the random phase that actually is realized. In approximations (a)–(c) above, the random phase is assumed independent of the molecular density. However, when the molecular density is high but not high enough for the crystal surface of HA to be completely saturated with molecules, then it might be possible that both the collagen and the solvent make a bidimensional crystalline phase on the crystal surface. Thus, due to the importance of the repulsive interactions among collagen molecules (for repulsive interactions, see Introduction Section in Ref. *6*), the free energy of the system would be at a minimum when the direct contact between any pair of collagen molecules is avoided or minimized, making a crystalline phase. In chromatography, the crystalline phase would, therefore, be realized immediately after the phosphate gradient has been applied or even before the application of the gradient. Since, in the crystalline phase, the apparent adsorption energy per molecule [i.e., the sum of the adsorption energy (defined here as positive) and the repulsive interaction energy with other molecules (defined here as negative)] is larger than in the random phase with the same molecular density, the desorption of molecules from the HA surface should occur in a higher phosphate molarity in the crystalline phase. However, once the desorption occurs in the crystalline phase and the molecular density on the HA surface is reduced, then it would be the random phase that takes place (see above). Since, in the random phase, the apparent adsorption energy per molecule is smaller, the desorption of molecules would still continue until the molecular density attains a smaller value where a new equilibrium is realized. As a result, the chromatogram would begin abruptly at its left-hand part.

Actually, the crystalline and the random phases might not be distinguished definitively. Nevertheless, it is probable that, because of the influence of the repulsive molecular interactions upon the distribution of molecules on the HA surface, the experimental chromatogram begins abruptly at its left-hand part in comparison with the theoretical prediction made on the basis of assumptions (a)–(c) above.

B. Orientation of Collagen on the Hydroxyapatite Surface

It can be deduced (*8*) that crystal C sites on which collagen adsorption occurs exist on side faces of the HA crystals with thin blade-like shapes (see

Figs. A1 and A2 in appendix II of Ref. 7 where microphotographs of HA crystals used for chromatography are shown). It can also be deduced (*8*) that these side crystal faces correspond to (**a**, **c**) [or (**b**, **c**)] crystal planes on which **a** (or **b**) and **c** vectors lie parallel and perpendicular to the two flat crystal surfaces, respectively. On the side crystal faces, C sites can be considered to be arranged with a minimal interval of | **a**| (= | **b**|) i.e., 9.42 Å in the **a** (or **b**) direction and a minimal interval of | **c**| /2, i.e., 3.44 Å in the **c** direction (see Fig. 2 in Ref. *8* where the arrangement of C sites on the crystal surface is schematically represented). In Appendix I of Ref. *11*, it was estimated under some assumptions that the thickness of the blade-like crystal would be about 700 Å. This means that, if collagen (with a length of 3000 Å; see Appendix III) is adsorbed in parallel with the **c** direction, it should be adsorbed by using only a quarter of the total length. If the adsorption occurs in the **a** (or **b**) direction, the total molecular length presumably is used for the adsorption (but see below).

The diameter of the collagen molecule is about 15 Å (Appendix III). This would mean that, if the molecule is orienting in the **c** direction, the adsorption is done on only one or two arrays of C sites that are arranged in the **c** direction, since the interval between the neighboring two arrays of C sites is 9.42 Å (see above). If the molecule is orienting in the **a** (or **b**) direction, two to four arrays of C sites may be used. Thus the interval between the neighboring two arrays of C sites now is 3.44 Å (see above), so that if two, three, four, and five arrays are used, the distance between the two exterior arrays is 3.44, 6.88, 10,32, and 13.76 Å, respectively. Therefore, assuming that collagen is an idealized cylinder with a diameter of 15 Å, the distances from the molecular surface to crystal sites on the two exterior arrays of them can be estimated to be about 0.2, 0.9, 2 and 4.5 Å in respective cases. From this consideration it seems unlikely that more than four arrays of C sites are used.

It can now be concluded that the total number of crystal sites existing under an adsorbed collagen molecule (in such a way that they can react with the molecule) generally is much larger when the adsorption is made in the **a** (or **b**) direction by using the total molecular length than when the adsorption occurs in the **c** direction by using only a quarter of the length. Therefore, it is presumable that the former orientation is energetically much more stable than the latter; it should be the former orientation that actually is realized. The possibility that collagen is adsorbed in an intermediate orientation between the two orientations examined above cannot be excluded completely. However, no stable adsorption seems to be achievable in the intermediate orientation. It should be added that, from a combination of crystallographic, stereochemical, and chromatographic studies, poly-L-lysine

with a highly stretched conformation (cf. Fig. A6 in Appendix I) is deduced to be adsorbed on the (**a**, **b**) crystal surfaces (i.e., the two flat surfaces of the blade-like crystal; see above) in parallel with the **a** (or **b**) direction (Ref. *12*, Appendix III).

The possibility that some amounts of crystals with lengths in the **a** (or **b**) direction that are less than 3000 Å co-exist in the HA preparation cannot be excluded. Therefore, when the molecular density on the crystal surfaces is very high or just at the beginning of chromatography, some molecules might be adsorbed on the crystal surfaces by using some part of the molecular structure. However, when the migration of molecules occurs on the column, the molecular density on the crystal surfaces should not be so high; it should be this state that mainly determines the shape of the chromatogram. In this state, virtually all molecules would be adsorbed by using the total molecular structure in order to obtain energetic stabilities. Even assuming that the adsorption of molecules is made by using part of the molecular structure in any process of chromatography, the occurrence of the multipeak chromatogram cannot be explained in terms of this mechanism since chromatography is virtually a quasi-static process (*2–6*).

C. Is It Themodynamically Possible that Collagen Molecules Are Adsorbed Compactly on the Hydroxyapatite Surfaces at the Beginning of Chromatography in Spite of Repulsive Mutual Interactions?

By using the value 0.7 kcal/mol of ε (Theoretical Section C) and the mean value, about 20, of x (Fig. 4c), the energy of adsorption per collagen molecule can be estimated to be about 14 kcal/mol on the average. Therefore, even provided molecules are adsorbed compactly on the HA surface, the apparent adsorption energy per molecule (see Section A) is positive with a value of 11 kcal/mol, since $\tilde{E} = 3$ kcal/mol (Theoretical Section C). This means that even the compact adsorption is thermodynamically possible.

In Appendix I of Ref. *11* it was mentioned that the adsorption capacity of HA for DNA is much smaller than that for collagen. DNA is also adsorbed on to C sites (by using phosphate groups) (*2*). It was suggested (Ref. *11*, Appendix I) that collagen is adsorbed onto not only C sites but also P sites just at the beginning of chromatography. However, it seems more likely that the small adsorption capacity for DNA arises from the fact that there are strong repulsive interactions among DNA molecules on the HA surface and/or that the conformations of DNA (with high molecular weights) are closer to random coils than to rods.

D. Values of h and h'

See Eqs. (6)–(8). On the basis of a reasonable assumption that collagen is adsorbed in the **a** (or **b**) direction on the side faces of the HA crystal (Section B), the value of d (Eq. 5) can be estimated to be 9.42 Å. The value of l (Eq. 4) can be estimated to be 47.1 Å by using the cylindrical collagen model with a diameter of 15 Å (Appendix III). Since $\nu = 18$ (Theoretical Section B), from Eq. (10) we have $h^2 = 24.6$ Å^2 or $h = 5.0$ Å. In this manner of adsorption, two to four arrays of C sites [arranged in the **a** (or **b**) direction] should be used for the adsorption of a molecule (Section B). This means that $n = 2$–4 (Eq. 7). Hence, from Eq. (7) the value of h' can be estimated to be 3.5, 2.9, and 2.5 Å assuming $n = 2$, 3, and 4, respectively. All these values are smaller than, or almost equal to, the minimal interdistance 3.44 Å between the neighboring C sites arranged in the **c** direction. This is also the minimal value among all possible interdistances between C sites. It can, therefore, be assumed (Theoretical Section B) that both h_1 (Eq. 4) and h_2 (Eq. 5) are essentially equal to h' (Eq. 8). h'^2 should measure the area on the lateral molecular surface in which a carboxyl group (attached at the top of the side chain of aspartic or glutamic residue) can move freely without considerable loss of energy. The values 2.5–3.5 Å of h' obtained above are reasonable ones that are self-consistent with the physical interpretation of h'.

By using the cylindrical collagen model of 3000 × 15 Å (Appendix III), the total area of the lateral molecular surface can be estimated to be 1.41×10^5 Å^2, while we have a value 24.6 Å^2 of h^2 (see above). This latter represents the area of an elementary surface on the abstract collagen surface, the total area of which is almost equal to that of the actual lateral molecular surface (Theoretical Section B). The total number of elementary surfaces on the abstract collagen surface can now be estimated to be about 5700. As, on the other hand, $n = 2$–4 (see above), the total number of the surfaces with an area of h'^2 can be estimated to be 11,000–23,000 (Eq. 7). This is 4–8 times as large as the total number, about 3000, of amino residues constituting an intact collagen molecule (see Appendix III). This means that the area assigned to an amino residue on the cylindrical collagen surface is 4–8 times as large as h'^2. This result of the simple calculation is also consistent with the physical meaning of h' (see above).

E. Adsorption of Phosphate Ions onto the Collagen Surface Adsorbed on the Hydroxyapatite Surface

The total number of C sites that are involved within two to four arrays of the sites [arranged in the **a** (or **b**) direction] existing under an adsorbed collagen molecule with a length of 3000 Å can be estimated to be 640 to

1270. These values are much larger than both the value 40 of x' (Theoretical Section A) and the values 15–25 of x (Fig. 4(c). The x' value has been estimated by comparing chromatographic data (with small sample loads) obtained under several different s values with Eq. (10*) (*3*). Even assuming that the molecule is adsorbed in the **c** direction by using only a quarter of the total molecular length, the total number of C sites existing under an adsorbed molecule is much larger than both x' and x values (see Section B).

It appears that the small value of x' can be explained by assuming that the majority of crystal sites existing under an adsorbed collagen molecule can react with phosphate ions from the buffer. The phosphate ions can freely enter the interspaces between the adsorbed molecule and the crystal surface. Collagen involves glycyl and imino residues occupying 1/3 and 1/5–1/4 of the toal residues, respectively (Appendix III). As these residues have no protruding side chains, some interspaces might, in fact, be realized between the adsorbed molecule and the crystal surface. Since, however, the diameter of phosphate ion (about 5 Å) occupies one-third of the diameter of the collagen rod (about 15 Å), it seems to be difficult to explain the extremely small x' value on the basis of this hypothesis only.

Further, assuming that the value 40 of x' (which is only twice as large as the mean value 20 of x) literally represents the number of C sites on which the adsorption of phosphate ions is impossible due to the presence of an adsorbed collagen molecule (see Theoretical Section A), a co-relation can be expected to occur between the x and x' values for different collagen components. The larger the x value, the larger should be the x' value, because x reacting sites are involved in x' sites. In order for x' to be constant, x' should be much larger than x. On the basis of the microheterogeneous collagen model, the shape of the experimental chromatogram can be explained only by assuming that x' is constant. In fact, the elution molarity of the molecule should roughly be determined by the factor ψ/x' in the second term on the right-hand side of Eq. (3), since the value of this term actually is much larger than the value of the term $(\ln x')/(x'\varphi')$ involved in the first term on the right-hand side of Eq. (3). Further, ψ/x' is about equal to x/x' (Figs. 4b and c). If there is a strong co-relation between the x and x' values, the ratio x/x' should roughly be constant. This means that any collagen component should be eluted at almost the same phosphate molarity.

The requirement that the apparent x' value should be both small and constant can be fulfilled only by assuming that the true x' value is very large but that a considerable adsorption of competing phosphate ions occurs onto the collagen surface adsorbed on the HA surface. Thus in our chromatographic model (*2–6*) the energetical interaction between the sample molecule and the competing ion is not taken into consideration. However, it is reasonable to assume that the number x' of crystal sites on which the

adsorption of competing ions is impossible due to the presence of an adsorbed molecule apparently decreases if the adsorption of the ions occurs on to the surface of the molecule. Further, if the total number of competing ions adsorbed on a molecule is large, the ratio of the variation in this number to the average number among different collagen components should be small; the apparent x' value would almost be constant. It is reasonable to assume that the phosphate adsorption is done mainly on the basic amino residues with a positive charge (lysine, hydroxylysine, arginine, and histidine). The total number of these residues per intact collagen molecule, about 300, is large (Appendix III).

Both the much smaller x' value for collagen than that expected from the molecular dimensions and the considerable adsorption of phosphate ions onto proteins in general can be suggested from the following experimental data apart from the microheterogeneous collagen model. The elution phosphate molarities of native nuclear DNA with molecular weights extending in a considerably wide range involving the molecular weight equal to that for collagen (about 3×10^5 daltons; see Appendix III) are almost constant, being equal to about 0.22 M (*14, 15*). This datum provides an experimental proof demonstrating that the elution molarity of a given molecule is governed by an intensive factor x/x' ($\approx \psi/x'$) (see above). Now, this elution molarity is only three times as high as the elution molarity, about 0.08 M, for collagen (Figs. A1–A4 in Appendix I). DNA is adsorbed onto C sites by using phosphate groups (see Section C). Since the density of phosphate groups on the DNA surface is much higher than the density of carboxyl groups on the collagen surface, the number of adsorption groups per unit molecular weight that are used for the reaction with crystal sites would be much larger with DNA than with collagen, while the elution phosphate molarity should roughly be determined by the factor x/x' (see above). This would mean that the apparent x' value per unit molecular weight of collagen is much smaller than the corresponding value for DNA.

Second, with basic proteins, it can be deduced that molecules are adsorbed (by using basic groups) onto P sites that are arranged on (**a**, **b**) crystal surfaces different from the surfaces where C sites exist (see Section B). These molecules compete with cations from the buffer for P sites (*2, 8, 12, 16*). By applying molarity gradients of potassium ions, the chromatographic behaviors of several different basic proteins were compared between the potassium phosphate and potassium chloride systems (with the same pH 6.8). It was found, however, that for any protein the elution potassium molarity in the phosphate system is about 30% as low as the molarity in the chloride system (*16*). This suggests that, due to adsorption of phosphate ions on the protein surface, the total volume of the molecule apparently has increased, and that, due to this increase, the value of x' (concerning P sites)

has also increased about 30% in the phosphate system. For this increase a considerable amount of phosphate ions on the protein surface would be necessary. This suggests that, provided basic proteins are adsorbed onto C sites in the presence of phosphate ions, the apparent x' value (concerning C sites) would drastically decrease in comparison with the case of no phosphate adsorption on the molecular surface.

It should be added that, for basic proteins (lysozyme and cytochrome *c*), the x' values (concerning P sites) estimated from molecular dimensions are essentially identical with those estimated from the chromatographic data (*11*) (Ref. *12*, Appendix IV). Effective potassium adsorption on the protein surface does not occur.

F. Is It Not Possible to Explain the Experimental Collagen Chromatograms on the Basis of Other Models Than the Microheterogeneous Model?

1. *Deaminations*. With a number of synthetic peptides, spontaneous deaminations of asparginyl and glutaminyl residues do occur under physiological solvent conditions (*17*). The deamination half-times vary between 6 d to 9 years depending markedly upon the sequence in amino residues in peptides. The *in vivo* behaviors of cytochrome *c*, aldolase, and lysozyme are consistent with the deamination half-times of model peptides with the same amide sequences as these peptides (*17*). The deamination experiments for peptide fragments with the same amide sequences as those involved in collagen are not reported.

It can be deduced, however, that deaminations do not occur or that they occur only very slightly with collagen. Thus in Table 1 are shown amide contents (obtained by amino acid analysis) of collagen from several different sources and their subunit components, expressed as numbers per 1000 amino residues (for details, see the legend of Table 1). It can be seen in Table 1 that the amide contents are highly constant among collagen peptides originating from different sources. Theoretically, the amide contents of β_{11} (dimer of α_1) and β_{22} (dimer of α_2) should be equal to those of α_1 and α_2, respectively. The amide content of β_{12} (dimer of α_1 and α_2) should be intermediate of those of α_1 and α_2. These points, in fact, are well reflected in amino acid analysis (Table 1). This would mean that the precision in the experimental data in Table 1 is high. Further, from the primary structure of the α_1 chain obtained by combining parts of the structures of the α_1 chains from calf- and rat-skin collagens that have been "determined" by the organochemical method (see Appendix III), the amide content per 1000 amino residues of the α_1 chain can be estimated to be 39.8 (where a residue that is not yet determined to be either glutamic or glutaminyl has been assumed to be glutamic). This value is

TABLE I

Amide Contents (obtained by amino acid analysis) of Collagens from Several Different Sources and Their subunit Components, Expressed as Numbers per 1000 Amino Residues (reproduced from Ref *18*)[a].

Source of collagen	Total	α_1	β_{11}	β_{12}	α_2
Calf-skin	44.2		$\alpha = 42.9$,	$\beta = 42.2$	
Rat-skin	41	42	41	42	43
Rat-tail tendon	40	39	40	41	41
Carp swim bladder	40	37	39	43	46
Cod-skin	46	$\alpha_1 = 43$,	$\alpha_2 = 46$,	$\alpha_3 = 54$	
Dog fish shark skin	38	39	39	36	
Human skin	36.9	37.9	37.2	44	45
			$\beta_{22} = 45$		

[a] For α_1 and α_2 components, see Appendix III. Cod-skin collagen is composed of α_1, α_2, and α_3 components in contrast with other collagens composed of two α_1 and one α_2. β_{11}, β_{12}, and β_{22} represent dimers: $\alpha_1 - \alpha_1$, $\alpha_1 - \alpha_2$, and $\alpha_2 - \alpha_2$, respectively. α and β (for calf-skin collagen) show mixtures before fractionations of α_1 and α_2, and β_{11} and β_{12}, respectively.

close to the values obtained from amino acid analysis (Table 1). If spontaneous deamination occurred, essentially constant amide contents among different collagens seen in Table 1 should be highly fortuitous. It seems unlikely that such a high fortuity actually occurs.

The rechromatography experiment of collagen (Introduction Section) would exclude the possibility that deaminations occur on the HA column during the chromatographic process or generally when the collagen is dissolved in the buffer used for chromatography (see the legend of Fig. A1 in Appendix I). This would also exclude the possibility that the molecular structure is destroyed anyhow.

2. *Posttranslational Modifications of the Molecular Structure.* These modifications involve (a) conversion of proline to hydroxyproline, (b) conversion of lysine to hydorxylysine, (c) glycosylation of hydroxylysine, and (d) formation of intramolecular cross-links at the terminal part(s) of the molecular structure (see Appendix III).

It is difficult to attribute the collagen heterogeneity detected on HA columns (Figs. A1–A4 in Appendix I) to posttranslational modifications. Thus, if the terminal part(s) of the molecular structure where the cross-links are formed are eliminated by pepsin treatment, essentially the same chromatogram as that occurring before pepsin treatment is obtained (*1*). This excludes the possibility that the chromatographic collagen heterogeneity arises from modification (d). It can be assumed that modification (c) scarcely changes the elution phosphate molarity from HA column because the

carbohydrate content of a collagen molecule is extremely small. Only some hydroxylysyl residues (out of about 3000 total amino and imino residues) are attached to carbohydrates in the form of galactosylhydroxylysine and glucosylgalactosylhydroxylysine (see Appendix III).

It is reported (*19*) that a slight heterogeneity in an α chain (Appendix III) can be detected on a DEAE-cellulose column. This is mainly concerned with the variation in hydroxylations of lysyl residues. Even assuming that hydroxylations of both prolyl and lysyl residues (modifications a and b) occur heterogeneously among different molecules, it is difficult to explain the heterogeneous HA chromatograms (Figs. A1–A4 in Appendix I) in terms of this heterogeneity. First, if the heterogeneous hydroxylations do occur, this should not influence the distribution of functional carboxyl groups on the collagen surface. From a titration experiment it can be assumed that all carboxyl groups are exposed on the molecular surface (Appendix III). Second, no synthetic polypeptides without charge are retained on the HA column (*20*). This would mean that the hydroxyl group without charge does not virtually react with the HA surface. If it reacts, the reaction energy should be much smaller than the adsorption energy, 0.7 kcal/mol, for a carboxyl group. The value 0.7 kcal/mol is a reasonable one judging apart from the microheterogeneous collagen model since the adsorption energy for a univalent phosphate group is 0.9–1 kcal/mol (Introduction Section). Further, the intervals between the neighboring peaks in the multipeak collagen chromatograms (Figs. A3 and A4 in Appendix I) should reflect the energy differences of about 0.7 kcal/mol, and this statement holds independently of the microheterogeneous model (Introduction Section; see Ref. *3*). Therefore, it is difficult to assume that, due to the heterogeneity in hydroxylations, the heterogeneous chromatograms (Figs. A3 and A4 in Appendix I) occur.

APPENDIX I

See Figs. A1–A6.

APPENDIX II

When the partition $\{x_1, \ldots, x_j, \ldots, x_\nu\}$ of X carboxyl groups occurs at random among ν elementary domains on the abstract collagen surface, then, representing by λ a set $(x_1, \ldots, x_j, \ldots, x_\nu)$, the probability of the occurrence of the λth-type set, $\Phi(\lambda)$, can be written as

$$\Phi(\lambda) = \frac{1}{\nu^X} P_\lambda Q_\lambda \qquad \text{(A1)}$$

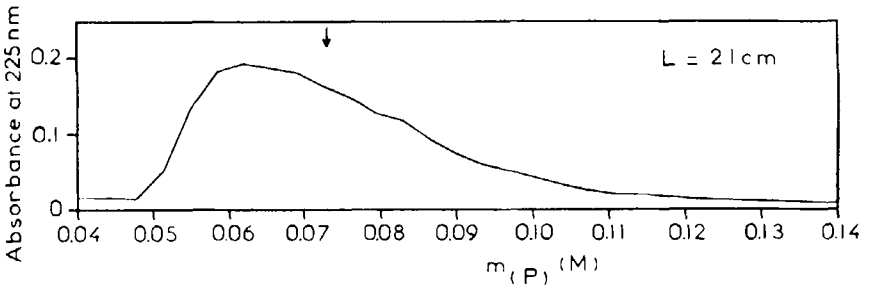

FIG. A1. Experimental chromatogram of calf-skin native collagen at 4°C represented as a function of phosphate molarity, $m_{(P)}$, in sodium phosphate buffer (pH 6.8). The linear molarity gradient of the buffer is applied. Both 0.15 *M* NaCl and 1 *M* urea exist in the buffer; this is necessary for collagen to be dissolved in the buffer. The concentration of urea, 1 *M*, is far below that causing denaturation (*21*), which was verified on HA columns (*1*, *22*). The arrow shows the position of the center of gravity of the total chromatogram. The experimental conditions are as follow: column length $L = 21$ (cm), reduced slope of the phosphate gradient (to column diameter 1 cm) $g'_{(P)} = 1.25 \times 10^{-3}$ M/mL, initial phosphate molarity $m_{in(P)} = 0.001$ *M*, and sample load (reduced value to column diameter = 1 cm) $\alpha = 2$ mg. (Reproduced with modifications from Fig. 1 in Ref. *1*.)

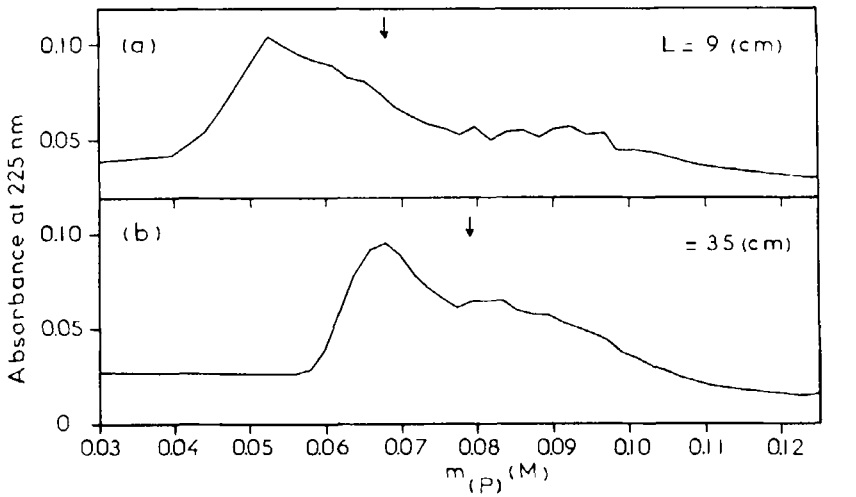

FIG. A2. As Fig. A1 but column lengths are different and $g'_{(P)} = 4.5 \times 10^{-4}$ M/mL. (Reproduced with modifications from Fig. 2 in Ref. *1*.)

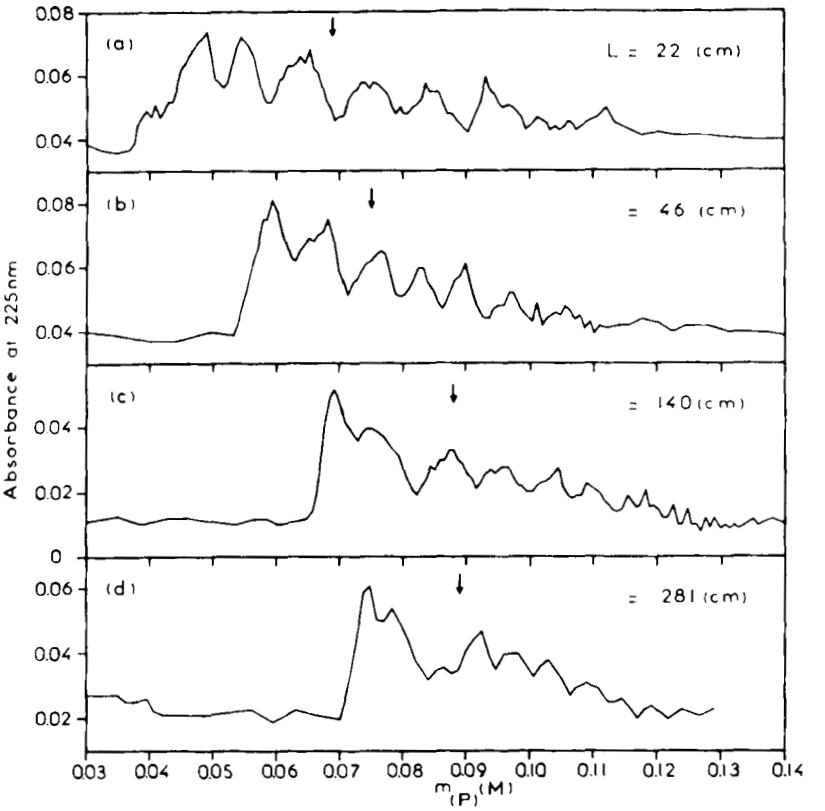

FIG. A3. Fig. A1 but column lengths are different, $g'_{(P)} = 3.75 \times 10^{+5}$ M/mL, and $\alpha = 10$ mg. (Reproduced with modifications from Fig. 6 in Ref. *1*.)

where P_λ and Q_λ are weights. These can be represented as

$$P_\lambda = \frac{\nu!}{\prod_{u=0}^{X} J_\lambda(u)!} \tag{A2}$$

and

$$Q_\lambda = \frac{X!}{\left[\prod_{j=1}^{\nu} x_j!\right]_\lambda} = \frac{X!}{\prod_{u=0}^{X} (u!)^{J_\lambda(u)}} \tag{A3}$$

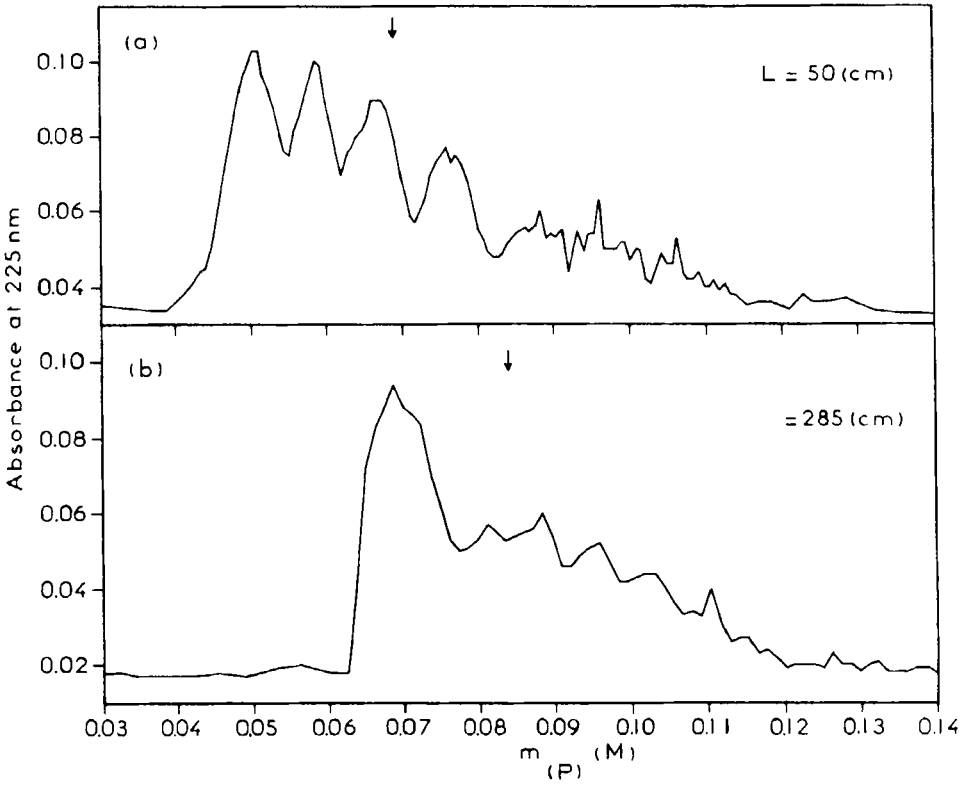

FIG. A4. As Fig. A3 but column lengths are different and $\alpha = 20$ mg. (Reproduced with modifications from Fig. 7 in Ref. *1*.)

where

$$J_\lambda(u) = \left[\sum_{j}^{\nu} \delta_{x_j,u} \right]_\lambda \tag{A4}$$

$\delta_{x_j,u}$ being the Kronecker's delta. Both relationships

$$\sum_{u=0}^{X} J_\lambda(u) = \nu \tag{A5}$$

and

$$\sum_{u=0}^{X} u J_\lambda(u) = \sum_{j=1}^{\nu} x_j = X \tag{A6}$$

are fulfilled. Equations (A2)–(A4) were presented earlier (*7*); where some misprints are involved, however.

The function $\Phi(\lambda)$ (occurring when $\pi = 0$) is identical with that obtained provided the distribution of the X carboxyl groups on the total abstract molecular surface is random. The total number of elementary surfaces is much greater than the number of "correct" elementary surfaces (see text). The following argument should be avoided: When the distribution of X carboxyl groups on the total abstract surface is random, then, with a finite probability, the carboxyl groups should enter the X "correct" elementary surfaces. This means that the π value is greater than zero even when the distribution of the carboxyl groups is random. This argument is self-inconsistent. The statement that the carboxyl groups enter the X "correct"

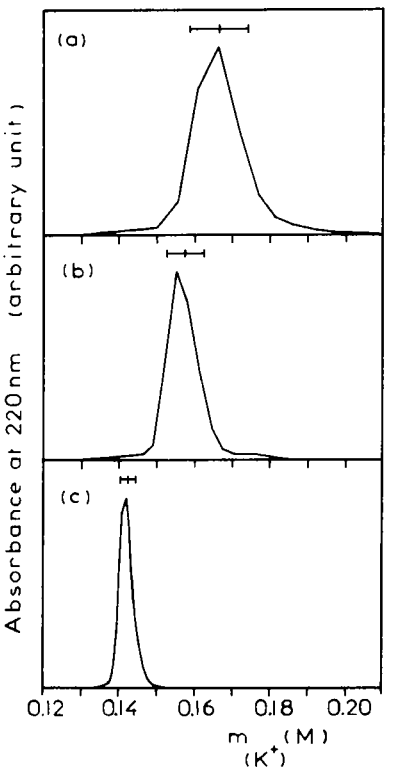

FIG. A5. Experimental chromatograms of egg-white lysozyme with small sample loads at about 25°C represented as functions of potassium molarity, $m_{(K^+)}$, 1.5 times as high as $m_{(P)}$, in potassium phosphate buffer (pH 6.8) used to make the linear gradient. The experimental conditions in (a), (b), and (c) are almost identical with those in Fig. A1, Fig. A2(b), and Figs. A3(d) and A4(b), respectively. Thus, in (a), $L = 21$ cm and $g'_{(P)} = 1.25 \times 10^{-3}$ M/mL; in (b), $L = 55$ cm and $g'_{(P)} = 4.5 \times 10^{-4}$ M/mL; and in (c), $L = 276$ cm and $g'_{(P)} = 3.75 \times 10^{-5}$ M/mL. (Reproduced from Fig. 6 in Ref. *3* or, with modifications, from Fig. 8 in Ref. *13*.)

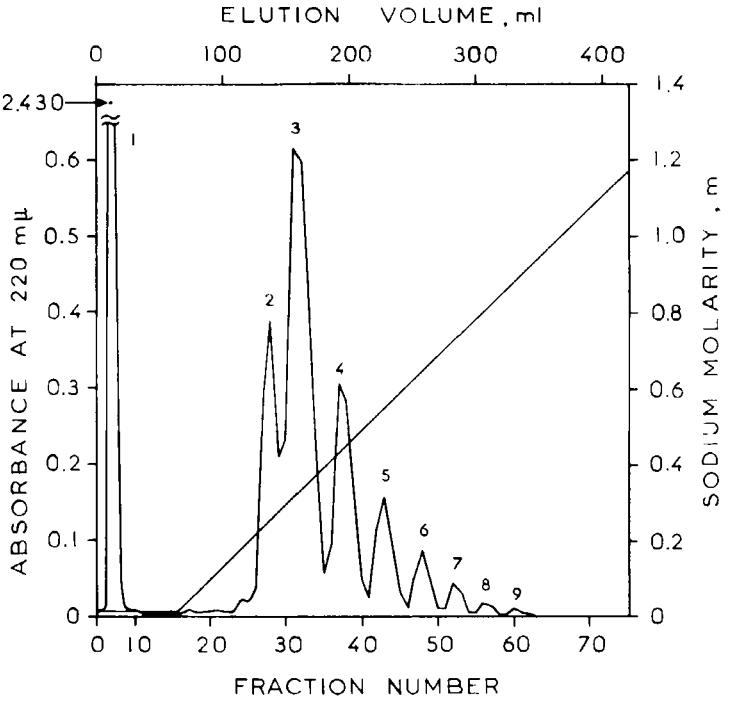

FIG. A6. Experimental chromatogram at room temperature of a synthetic heterogeneous poly-L-lysine·HBr sample with molecular weights ranging between 1500 and 8000 daltons (degrees of polymerization 7–38). Sodium phosphate buffer (pH 6.8) was used to make the linear gradient. The experimental conditions are $L = 22$ cm and $g'_{(P)} = 2.2 \times 10^{-3}$ M/mL. (Reproduced from Fig. 1 in Ref. *12* or, with modifications, from Fig. 5a in Ref. *20*.)

elementary surfaces with a finite probability is incompatible with the statement that the distribution of the carboxyl groups is random, because the distribution of the X "correct" elementary surfaces on the total abstract surface is given.

APPENDIX III

Collagen is the major structural protein involved in a wide range of vertebrate and invertebrate species (*23*). Calf-skin collagen investigated in the present paper belongs in type I out of, at least, several types of collagens that can be distinguished genetically in higher animals (*9*). This molecule, with a molecular weight slightly less than 3×10^5 daltons, can be represented by a rod of 3000×15 Å (*24*). This is made up of three polypeptide chains

(called α chains) consisting of slightly more than 1000 amino (and imino) residues, respectively. Two of the α chains (α_1 chains) are identical and the other (α_2 chain) has a slightly different amino acid composition (and sequence) (*9, 18, 25, 26*). The three α chains together form a triple helix and a rodlike external shape (see above), the N and C terminals appearing at the same ends of the rod (*23*). Collagen partially involves molecules in which α chains are cross-linked by a covalent bond through lysyl and hydroxylysyl residues existing at the terminal part(s) of the chains (*9, 23, 24, 26*). Cross-links are formed between not only α chains belonging in a single collagen molecule but also those belonging in different molecules. In the latter case, collagen polymers are produced. These are hardly extractable under the condition where monomer molecules usually are extracted (*27, 28*); a slight amount of polymers that are contaminating the monomer preparation can be eliminated by treatment with protease at the same time with the elimination of the intramolecular cross-links (*27, 28*) (cf. Discussion Section F-2). The collagen sample used in the experiments in Figs. A1–A4 in Appendix I was prepared by the method of Ref. *27* (*1*). To only some of hydroxylysyl residues (out of 3000 total residues per molecule) are attached carbohydrates in the form of galactosylhydroxylysine and glucosylgalactosylhydroxylysine (*9, 23, 26*). All or most of the amino acid sequences in both α_1 and α_2 chains have successfully been "determined" by using the organochemical method (*9, 26, 29*). Recently, slight modifications were added (*30, 31*). Thus the greater parts of both α_1 and α_2 chains consist of repetitions of triplet structures of Gly-X-Y where X and Y can be amino and imino residues. Glycine occupies almost one-third of the total residues of an α chain; proline and hydroxyproline occupy one-fifth to a quarter. Therefore, X and Y positions are often occupied by imino residues. Both aspartic and glutamic residues with carboxyl groups occupy about 7% of the total residues. About 230 acidic residues are involved in a intact collagen molecule. The carboxyl groups are all free and titratable (*32*).

The collagen biosynthesis can be characterized by at least two prominent features: (1) the protein is first synthesized as a precursor form, procollagen. This is made up of two pro-α_1 and one pro-α_2 peptide chains being longer at the N and C terminals by several hundred amino acids (about 40% larger in size than the α chains; for biosyntheses of the pro-α chains, see below). (2) The biosynthesis involves several unusual post-translational modifications which occur after assembly of amino acids into the three polypeptide chains of the molecule (*9, 33, 34*). These modifications involve (a) conversion of proline to hydroxyproline, (b) conversion of lysine to hydroxylysine, (c) glycosylation of hydroxylysine (see above), and (d) oxidative deamination of lysine and hydroxylysine to yield the cross-link-precursor aldehydes allysine and hydroxyallysine at the terminal part(s) of the α peptides (see above) (*9*).

Pro-α_1 and pro-α_2 chains are synthesized separatedly as single chains by sequential amino acid additions from the amino terminal ends (like syntheses of polypeptides of other proteins) (*33, 35, 36*). The estimation that polysomes participating in collagen synthesis are larger than a 23-ribosomal aggregate but smaller than a 50–60-ribosomal aggregate is consistent with a monocistronic message coding for a pro-α chain (*33, 34, 37*).

The microheterogeneous collagen model with the π value of 0.8 can be considered to be compatible with the fact that the primary structure of the α chains can be "determined" by the organochemical method. The process of this "determination" is reviewed in detail by Fietzek and Kühn (*26, 29*). Thus the determination of the amino acid sequence is carried out for relatively small peptide fragments of the α chains, mainly with the aid of the Beckman Sequenator (Sequencer), by repeating successively the Edman-type degradation cycles. In a cycle, N terminal residue of the peptide fragment is released and converted finally to a phenylthiohydratoinine (PTH) derivative, which is extracted and identified. The residual peptide which is now one amino acid shorter is subjected to the next degradation cycle. For the indentification of the PTH-amino acid, thin layer or gas-liquid chromatography is used. It is reported (*29*), however, that, as the peptide degradation proceeds, the detection of the liberated amino residue becomes more and more hampered by overlaps and a general rise of background. This is a reason why the amino acid sequence determination can be performed only for small peptide fragments (see above). Yield relative to the first step of the degradation declines to 50% after only 20 degradation steps, and to 5% at the 45th step. The decrease of the yield may be accounted for by the cumulative effect of only a slight blocking of the reaction at each step. A concomitant increase of the background would be due (at least partially) to unspecific cleavage of the peptide chain (*29*).

According to the microheterogeneous collagen model, the recovery of the chromatographic spot or the peak for a "correct" amino acid appearing in thin layer or gas–liquid chromatography (see above) should be equal to π or 80% with aspartic or glutamic acid, provided that the experiment is ideal and that there is no unspecific cleavage of the peptide chain or blocking of the degradation reaction (see above). The spots or the peaks for "incorrect" amino acids would be distributed around the main one (for the "correct" amino acid) in diluted states. It would be practically impossible to observe "incorrect" spots or peaks, and/or it would be difficult to distinguish them from the artificial backgrounds.

We suggest below that the ambiguous primary structures of the α chains arise at a translational level of biosynthesis. This speculation is based on the fact that the amino acid compositions of the α chains estimated from the amino acid analysis (written as $P_1, P_2, \ldots, P_n$ where $\sum_{i=1}^{n} P_i = 1$, i

representing the type of genetically coded amino acid, i.e., the type classified without discriminating hydroxyproline from proline, and hydroxylysine from lysine; the actual value of n is 18) is similar to the compositions calculated from the primary structures that have been "determined" by the organochemical method (written as $P'_1, P'_2, \ldots, P'_n$ where $\sum_{j=1}^{n} P'_j = 1$). In fact, the total primary structure of the α_1 chain can be obtained by combining the partial structures "determined" for calf- and rat-skin collagens. Concerning the α_2 chain, the data from which the complete primary structure can be constructed are not published. The differences in amino acid sequences between the common parts of calf- and rat-skin peptide chains are only very slight for both α_2 and α_2 chains (*9, 26*). It is presumable that, even with the α_2 chain, the amino acid compositions estimated from the two methods are similar. (For the amide compositions for the α_1 chain estimated by using the two methods, see Discussion Section F-1). Further, since the amino acid compositions are not so different between the α_1 and α_2 chains (*18*) (for amide compositions, see Table 1), the arguments below, for simplicity, will be made without distinguishing these peptide chains. Now, the following relationship should, in general, be fulfilled:

$$\begin{bmatrix} P_1 \\ P_2 \\ \cdot \\ \cdot \\ \cdot \\ P_n \end{bmatrix} = \begin{bmatrix} \pi_{11} & \pi_{12} & \cdots & \pi_{1n} \\ \pi_{21} & \pi_{22} & \cdots & \pi_{2n} \\ \cdot & & & \cdot \\ \cdot & & \cdots & \cdot \\ \cdot & & & \cdot \\ \pi_{n1} & \pi_{n2} & \cdots & \pi_{nn} \end{bmatrix} \begin{bmatrix} P_1 \\ P'_2 \\ \cdot \\ \cdot \\ \cdot \\ P'_n \end{bmatrix} \tag{A7}$$

where π_{ij}, fulfilling $\sum_{i=1}^{n} \pi_{ij} = 1$, represents the probability that a "correct" position j for amino acid j on the α chains is occupied by amino acid i. Therefore, provided the primary structures of the α chains are completely unique, then $\pi_{ij} \stackrel{\checkmark}{=} \delta_{ij}$ where δ_{ij} is the Kronecker's delta. On the other hand, we have an experimental relationship:

$$\begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_n \end{bmatrix} \approx \begin{bmatrix} P'_1 \\ P'_2 \\ \vdots \\ P'_n \end{bmatrix} \tag{A8}$$

(see above).

It can be assumed that the probability of the fortuitous occurrence of a matrix (π_{ij}) fulfilling both Eq. (A8) and the condition such that the mean value, π, of π_{ii} for aspartic and glutamic acid be of the order of 0.8 is low. It

can, therefore, be suggested that, in order for the matrix (π_{ij}) to fulfill actually both Eq. (A8) and the small π value, a feedback mechanism exists that is provoked by collagen itself. It does not seem likely now that the value, 0.8, of π is caused by the heterogeneous cistron structure of DNA because, first, we have no experimental data suggesting the existence of such a high heterogeneity in the ciston structure as can explain this π value. There are only several types of collagens that can be distinguished genetically (see above). Second, even assuming the existence of the highly heterogeneous cistron structure, it seems to be difficult to assume any feedback mechanism controlling the distribution law among different structural genes. It also seems to be difficult to assume the feedback mechanism regulating the transcription process from DNA to *m*-RNA being compatible with both Eq. (A8) and the small π value.

Let us assume that the ambiguous primary structures of the α chains arise at a translational level. For this to occur, two possibilities exist: (a) binding between "incorrect" amino acid and *t*-RNA, and (b) binding between "incorrect" codon and anticodon. In order for point (a) to occur, the activation energies for bindings of "correct" pairs of molecules provided by aminoacyl-*t*-RNA synthetases (*38*) would not be so different from the energies for "incorrect" pairs. In order for point (b) to occur, the binding energies for "correct" pairs of codon and anticodon would not perhaps be so different from those for "incorrect" pairs. In any instance, however, it would be reasonable to assume that, for entropical reason, the probability π_{ij} depends upon the effective concentration (denoted by P_i°) of amino acid i in the amino acid pool. It would, therefore, be possible that, due to a feedback mechanism provoked by collagen, the concentrations $P_1^\circ, P_2^\circ, \ldots, P_n^\circ$ are regulated in order for Eq. (A8) to be realized.

Both the calculations of theoretical chromatograms and the estimation of the π value have been made on the basis of the assumption that the total number of acidic amino residues per collagen molecule is constant, being equal to 230 (Theoretical Section B). Provided the collagen biosynthesis is carried out in such a way that amino acids are chosen completely at random from the amino acid pool [presumably with a composition similar to that occurring within collagen itself, not being so different from the amino acid composition for procollagen (*33*)], then the probability that m_i amino acid i are incorporated into a collagen molecule (consisting of 3000 amino residues) should be $\binom{3000}{m_i} P_i^{\circ m_i}(1 - P_i^\circ)^{3000-m_i}$; this is a binomial distribution with a form similar to that of a Gaussian distribution with a standard deviation equal to $[3000\ P_i^\circ(1 - P_i^\circ)]^{1/2}$. When the amino acid i is incorporated into the "correct" positions on the α chains with a probability π_{ii} (see above), it could roughly be approximated such that, beside the "correct" incorporation, the amino acid i is incorporated at random into the

α chains from the amino acid pool where the apparent probability of the occurrence of amino acid i is $P_i^\circ(1 - \pi_{ii})$ instead of P_i°. By using this approximation, the standard deviation in number of amino acid i in a collagen molecule can be represented as $(3000\, P_i^\circ(1 - \pi_{ii})[1 - P_i^\circ(1 - \pi_{ii})])^{1/2}$. Now, by assigning to i the acidic amino acids and by using the values 230 and 0.8 for $3000\, P_i^\circ$ and π_{ii} $(= \pi)$, respectively, the standard deviation for the acidic amino residues can be estimated to be 6.7. This value, in fact, is much smaller than the total number, 230, in a intact collagen.

For globular proteins (ovalbumin and hemoglobin), Loftfield (*39*, *40*) estimates that the frequency of replacement of a "correct" amino acid with a sterically similar "incorrect" amino acid on a polypeptide chain should be at most of the order of 3×10^{-4} (the case of substitution of valine for isoluecine). On the other hand, Pauling (*41*) estimates, on a physicochemical basis, that the probability of the occurrence of the Michaelis-Menten complex of "incorrect" amino acid and aminoacyl-*t*-RNA synthetase should be at least 5% in some instances. In this estimation an effort is made to decrease as much as possible the possibility of the entrance of the "incorrect" amino acid into the template of the enzyme. Numerous "incorrect" aminoacylations occurring due to misrecognition of the "correct" *t*-RNA by the valyl-*t*-RNA synthetase are reported in heterologous systems (*42*). For instance, by using valyl-*t*-RNA synthetase from *Bacillus stearothermophilus*, $t\text{-RNA}^{\text{Ile}}$ and $t\text{-RNA}_f^{\text{Met}}$ from *Escherichia coli* are acylated to the extents of 78 and 20%, respectively (*42*). In both bacterial and mammalian cells, extreme starvation for certain amino acids results in translational errors that can be easily detected by two-dimensional polyacrylamide gel electrophoresis (*43*). These errors can be predicted from misreading of pyrimidines for purines at the third position of the codon (*43*). In order to explain high precisions for occurrences of "correct" amino residues on a polypeptide chain that can be seen at least with globular proteins (see above), some proofreading mechanisms are proposed (*44–48*).

In fact, a highly unique primary structure would teleologically be necessary with globular proteins. The situation is different with collagen, however. Thus the greater parts of both α_1 and α_2 collagen peptides consist of the triplet structures of Gly-X-Y where X and Y can be any residues (see above). Nevertheless, the tertiary triple-helical structure (see above) is shown to be represented by the conformation that $(\text{Gly-Pro-Pro})_n$ takes (*23*, *49*). Further, it is demonstrated experimentally (*50*) that the triple-helical collagen structure (naturally consisting of two α_1 and one α_2 chains; see above) can be formed from both three α_1 and three α_2 chains. This again shows that precise amino acid sequences in the α chains are not required for the formation of the collagen structure, since the amino acid compositions and sequences are different between α_1 and α_2 chains (see above). However,

the existence of glycyl residues in glycyl positions in the $(Gly\text{-}X\text{-}Y)_n$ structure is necessary for the formation of the collagen structure (*23, 49*). This suggests that the π_{ii} value for glycine is close to unity. A high π_{ii} value for imino residues can also be suggested because the stability of the collagen structure is sensitive to the imino acid content (*51*) (see below). Concerning amino acids other than glycine and imino acids, we find no reasons why they should exist at "correct" peptide positions with high probabilities.

Finally, we present some experimental data showing high homogeneities (in a certain sense) in both α_1 and α_2 peptide chains. Thus sharp peaks or bands of both α_1 and α_2 components with random coil conformations can be obtained both chromatographically [on CM-cellulose columns (*24, 25*)] and electrophoretically. The electrophoreses are carried out on polyacrylamide gels in both the absence (*24, 50, 52, 53*) and the presence (*54*) of sodium dodecylsulfate (SDS), and on starch gels in the absence of SDS (*24, 55, 56*). With polyacrylamide gels in the presence of SDS, it can be suggested (*54*) that the separation of the α_1 and α_2 components (with essentially the same molecular weights) arises from the difference in rigidities in molecular structures depending upon the imino acid content, and not from the difference in charges. It can be suggested, that, with starch gel electrophoresis in the absence of SDS also, the separation of the two components is not due to the charge difference. Thus a good resolution can be achieved only in a small range of pH 4–5 (*56*). In this pH range the α_2 peptide [with the higher and lower contents of basic and acidic amino residues, respectively (*18*)] migrates more rapidly than the α_1 peptide. When pH is more acid (3.58), however, it is not the basic α_2 but rather the acidic α_1 component that migrates most rapidly (*56*). This datum suggests that the molecular conformations and the change in conformations occurring with change in pH are both fundamentally different between the α_1 and α_2 chains. CM-cellulose is a cation-exchanger (*57*). Since, however, the pH (4.8; see Ref. *25*) that is applied to the CM-cellulose chromatography for the separation of the two collagen peptides (see above) is just within the small pH range (4–5) only where a good electrophoretic separation can be achieved on starch gels (see above). It is very probably that the chromatographic separation is virtually due to the fundamental difference in molecular conformations. The effective number (or density) of positively charged groups existing on the "surface" of a molecule with a "random coil" conformation would depend markedly upon the type of "random coil" conformations.

It would be necessary to see what happens if the α_1 and α_2 peptides are subjected to isoelectric focusing electrophoresis. This might detect the slight distributions in total charges within both α_1 and α_2 chains which can be expected from the microheterogeneous collagen model (see above).

For further informations on collagen research, the reader is recommended to see a recent review written by Eyre (*58*).

Acknowledgments

The author is grateful to Dr G. Bernardi for the interest in this work. He also wishes to thank Dr J. Ninio for some useful biological discussions. Calculations were performed on the CDC 6600 computer of the Faculty of Sciences, University of Paris.

REFERENCES

1. T. Kawasaki and G. Bernardi, *Biopolymers, 9*, 269 (1970).
2. T. Kawasaki, *Sep. Sci. Technol., 16*, 325 (1981).
3. T. Kawasaki, *Ibid., 16*, 439 (1981).
4. T. Kawasaki, *Ibid., 16*, 817 (1981).
5. T. Kawasaki, *Ibid., 17*, 885 (1981).
6. T. Kawasaki, *Ibid.*, In Press.
7. T. Kawasaki, *J. Chromatogr., 82*, 219 (1973).
8. T. Kawasaki, *Ibid., 151*, 95 (1978).
9. K. A. Piez, in *Biochemistry of Collagen* (G. N. Ramachandran and A. H. Raddi, eds.), Plenum, New York, 1976, p. 1.
10. T. Kawasaki, *J. Chromatogr., 82*, 237 (1973).
11. T. Kawasaki, *Ibid., 93*, 313 (1974).
12. T. Kawasaki, *Ibid., 157*, 7 (1978).
13. T. Kawasaki and G. Bernardi, *Biopolymers, 9*, 257 (1970).
14. D. A. Wilson and C. A. Thomas, Jr., *Biochim. Biopys. Acta, 331*, 333 (1973).
15. G. Bernardi, *Methods Enzymol., 21*, 95 (1971).
16. G. Bernardi, *Ibid., 27*, 471 (1973).
17. A. B. Robinson, *Proc. Natl. Acad. Sci. U. S. A., 71*, 885 (1974).
18. J. E. Eastoe, in *Treatise on Collagen*, Vol. 1 (G. N. Ramachandran, ed.), Academic, New York, 1967, p. 1.
19. R. A. Trelstad, A. H. Kang, B. P. Toole, and J. Gross, *J. Biol. Chem., 247*, 6469 (1972).
20. G. Bernardi, *Methods Enzymol., 22*, 325 (1971).
21. J. Gross and D. Kirk, *J. Biol. Chem., 233*, 235 (1958).
22. G. Bernardi and T. Kawasaki, *Biochim. Biophys. Acta, 160*, 301 (1968).
23. W. Traub and K. A. Piez, *Adv. Protein Chem., 25*, 243 (1971).
24. K. A. Piez, in *Treatise on Collagen*, Vol. 1 (G. N. Ramachandran, ed.), Academic, New York, 1967, p. 267.
25. K. A. Piez, E. Eigner, and M. S. Lewis, *Biochemistry, 2*, 58 (1963).
26. P. P. Fietzek and K. Kühn, *Int. Rev. Connect. Tissue Res., 7*, 1 (1976).
27. A. L. Rubin, M. P. Drake, P. F. Davidson, D. Pfahl, P. T. Speakman, and F. O. Schmitt, *Biochemistry, 4*, 181 (1965).
28. M. P. Drake, P. F. Davidson, and F. O. Schmitt, *Ibid., 5*, 301 (1966).
29. P. P. Fietzek and K. Kühn, *Top. Curr. chem., 29*, 1 (1972).
30. K. A. Piez and B. L. Trus, *J. Mol. Biol., 122*, 419 (1978).
31. H. Hofman, P. P. Fietzek, and K. Kühn, *Ibid., 125*, 137 (1978).
32. A. Veis, in *Treatise on Collagen*, Vol. 1, (G. N. Ramachandran, ed.), Academic, New York, 1967, p. 367.
33. D. J. Prockop, R. A. Berg, K. I. Kivirikko, and J. Uitto, in *Biochemistry of Collagen* (G. N. Ramachandran and A. H. Raddi, eds.), Plenum, New York, 1976, p. 163.
34. P. Bornstein, *Annu. Rev. Biochem., 43*, 567 (1974).
35. J. Vuust and K. A. Piez, *J. Biol. Chem., 245*, 6201 (1970).

36. J. Vuust and K. A. Piez, *Ibid., 247,* 856 (1972).
37. E. Lazarides and L. N. Lukens, *Nature (New Biol.), 232,* 37 (1971).
38. L. L. Kisselev and O. O. Favorova, *Adv. Enzymol., 40,* 141 (1974).
39. R. B. Loftfield, *Biochem. J., 89,* 82 (1963).
40. R. B. Loftfield, *Ibid., 128,* 1353 (1972).
41. L. Pauling, *Festschrift Arthur Stoll,* Birkhauser, Basel, 1957, p. 597.
42. R. Giegé, D. Kern, J. P. Ebel, H. Grosjean, S. DeHenau, and H. Chantrenne, *Eur. J. Biochem., 45,* 351 (1974).
43. J. Parker, J. W. Pollard, J. D. Friesen, and C. P. Stanners, *Proc. Natl. Acad. Sci. U.S.A., 75,* 1091 (1978).
44. J. J. Hopfield, *Ibid., 71,* 4135 (1974).
45. J. Ninio, *Biochimie, 57,* 587 (1975).
46. A. R. Fersht and M. M. Kaethner, *Biochemistry, 15,* 3342 (1976).
47. F. von der Haar and F. Cramer, *Ibid., 15,* 4131 (1976).
48. T. Yamane and J. J. Hopfield, *Proc. Natl. Acad. Sci. U.S.A., 74,* 2246 (1977).
49. G. N. Ramachandran and C. Ramakrishnan, in *Biochemistry of Collagen* (G. N. Ramachandran and A. H. Raddi, eds.), Plenum, New York, 1976, p. 45.
50. C. Trocz and K. Kühn, *Eur. J. Biochem., 7,* 454 (1969).
51. P. H. von Hippel, in *Treatise on Collagen,* Vol. 1 (G. N. Ramachandran, ed.), Academic, New York, 1967, p. 253.
52. M. L. Tanzer, D. Monroe, and J. Gross, *Biochemistry, 5,* 1919 (1966).
53. M. Stark and K. Kühn, *Eur. J. Biochem., 6,* 534 (1968).
54. H. Furthmayer and R. Timpl, *Anal. Biochem., 41,* 510 (1971).
55. V. Näntö, J. Maatela, and E. Kulonen, *Acta. Chem. Scand., 17,* 1604 (1963).
56. V. Näntö, J. Pikkarainen, and E. Kulonen, *J. Am. Leather Chem. Assoc., 60,* 63 (1965).
57. E. Peterson and H. A. Sober, *J. Am. Chem. Soc., 78,* 751 (1956).
58. D. R. Eyre, *Science, 207,* 1315 (1980).

Received by editor January 23, 1981