ACTA CHEMICA SCANDINAVICA 11 (1957) 439—460

# Insulin Crystals

## V. The Nucleation and Growth of Insulin Crystals

JØRGEN SCHLICHTKRULL

*Novo Terapeutisk Laboratorium, Copenhagen, Denmark*

Beef-insulin is crystallized from a buffered solution of sodium chloride as geometrically similar rhombohedrons. During the process of crystallization running determinations of the concentration, $c_s$, of dissolved insulin are made. Simultaneously, samples are taken in which the crystallization is immediately arrested by diluting with a solution of zinc chloride. The concentrations of crystalline insulin, $C$, are determined by analyses of the stabilized samples. The total volumes and surfaces of the crystals, as well as the cumulative distributions of sizes are also determined in these samples by a special counting technique.

The total number ($N$), length ($L$), surface ($S$) and volume ($V$) of crystals are found to be mutually proportional during crystallization, except for the short initial stage.

The rate of linear crystal growth, $\dot{g}$, is found to be independent of crystal size and proportional to $(c_s-c)^2$, where $c$ is the final concentration of dissolved insulin.

The rate of nucleation is found to be proportional to the product of $g$ and $N$, $L$, $S$ and $V$, respectively, except for the short initial stage of crystallization.

The experimental results are predicted from the simple concept that nuclei are mainly formed on the surface of the crystals. This self-reproduction of crystals was not observed when the insulin was crystallized in the twin-like distorted shape.

The outer surface of the system, exposed to glass and air, is of prime importance to the process of nucleation. Nucleation is diminished when the crystallization is carried out in a paraffin enclosure or when Tween 80 is added, and it is considerably enhanced by foam or other increase in surface.

The purpose of the present paper is to investigate the rate of nucleation and growth of insulin crystals as functions of the concentration of insulin in the crystallization mixture. In earlier papers [1,2] a crystallization method was described which yielded perfectly shaped rhombohedral zinc-insulin crystals irrespective of the species of insulin. This method is suitable for a study of the kinetics of insulin crystallization for several reasons:

1) During the process of crystallization, which is not too rapid for study, samples can be taken in which the process of crystallization can be immediately arrested with zinc ions [2].

2) The fraction of crystalline insulin in the stabilized samples can be determined accurately [2].

3) The dimensions of a crystal are completely defined by one parameter, *e. g.* the largest diagonal, $l$, in the crystal projection on a plane through the microscope stage [2].

4) The distributions of crystal sizes in the stabilized suspensions can be determined by a counting procedure [2].

## SYMBOLS*

$c_s$ = The concentration of dissolved insulin in per cent of the total insulin concentration which is 1.7 % or 2.42 mg/ml insulin nitrogen.

$C$ = The concentration of crystalline insulin in per cent of the total insulin concentration.

$t$ = The time in min from the moment buffer is added to the acid insulin solution.

$t_n$ = A moment of nucleation.

$t_*$ = The time when crystallization is finished.

$l$ = The size of crystals, *i. e.* the largest diagonal in the crystal projection on a plane through one of the faces. As the crystals lie on their faces, the plane is parallel to the slide on the microscope stage. $l$ is measured in $\mu$.

$\mathfrak{P}(l,t)$ = The cumulative distribution of crystal sizes at the moment $t$, *i. e.* the number per ml of suspension of crystals which are greater than $l$ at the moment $t$.

$\mathfrak{p}(l,t) = -\frac{\delta \mathfrak{P}(l,t)}{\delta l}$, *i. e.* the distribution of sizes at the moment $t$.

$\dot{g} = \frac{\delta l(\mathfrak{P},t)}{\delta t}$, *i. e.* the rate of linear crystal growth which is supposed to be independent of the crystal size.

$g = \int_0^t \dot{g}\,dt$ is called the growth function.

$l_{max}(t) = g(t)$ as it is supposed that nucleation occurs already at $t = 0$.

$\mathfrak{N} = N = \int_0^g \mathfrak{p}(l,t)dl$ is the number of crystals per ml of suspension.

$\mathfrak{L} = L = \int_0^g l \cdot \mathfrak{p}(l,t)dl$ is the total length of crystals per ml of suspension.

$\mathfrak{S} = S = \int_0^g l^2 \cdot \mathfrak{p}(l,t)\,dl$ is the total surface of crystals per ml of suspension, the unit of area being the area of a rhombohedron with size $l = 1\ \mu$.

* $P$, $p$, $h$, $N$, $L$, $S$ and $V$ are typed in gothics when the independent variable is $t$ and in italics when it is $g$.

$\mathfrak{V} = V = \int_0^g l^3 \cdot \mathfrak{p}\,(l, t)\,\mathrm{d}l$ is the total volume of crystals per ml of suspension, the unit of volume being that of the rhombohedron with size $l = 1\ \mu$.

$\bar{l}$ = $L/N$, *i. e.* the average crystal size.

$m_i$ = $l_{i+1} + l_i$, where $l_i$ are class limits in the counting scheme.

$d_i$ = $l_{i+1} - l_i$.

$n_i'$ = The number of crystals in 1 small square in the counting chamber belonging to the size interval $l_i < l < l_{i+1}$.

$c$ = The volume in ml of crystallization mixture corresponding to 1 small square in the chamber.

$S_i$ = $f_{si} \cdot n_i'$ is the surface per ml of suspension of crystals belonging to the size range $l_i < l < l_{i+1}$, the unit of area being the area of a rhombohedron with size $l = 1\ \mu$.

$\mathfrak{h}$ = $\frac{\mathrm{d}\mathfrak{N}}{\mathrm{d}t}$ is the rate of nucleation, as it is assumed that nuclei are immediately transformed into crystals.

$h_s$ = $\mathfrak{h}/\dot{g}$ is called the specific rate of nucleation.

$r$ and $u$ are positive constants in the eqn (40), $S = rV^u$.

$m$ is the concentration of a crystallizing substance present as monomer dissolved particles.

$p$ and $q$ are positive powers.

$a$, $a_1$, $a_2$, $a_3$, $a_4$ = constants.

$b_1$, $b_2$, $b_3$, $b_4$ = constants.

$c_1$, $c_2$, $c_3$, $c_4$, $c_5$, $c_6$, $c_7$ = constants.

$\alpha$ = const. $\times\ g$.

$\beta = \frac{\sqrt{3}}{2} \cdot \alpha$.

## METHODS

1. *Crystallization.* The method of crystallization has been described recently [1,2]. The volume of crystallization mixture used was 4 litres, and the concentration of insulin was 1.7 %, *i. e.* 2.42 mg insulin-nitrogen per ml. Any foam that formed was removed.

2. *Determination of the insulin in solution,* $c_s$. 80—90 ml of the crystallizing suspension are filtered through filter paper covered with kieselguhr. Pressure is applied in order to keep down the filtration time to within 1 min. The filtrate is acidified with 1 N HCl in order to avoid crystallization on the walls of the flask. The concentration of insulin in solution, $c_s$, is determined by the semi-micro Kjeldahl analysis and expressed in per cent of the total concentration of insulin in the suspensions.

3. *Determination of the crystalline fraction, C.* A 10 ml sample is stabilized with zinc, the amorphous precipitate is dissolved, and the amount of crystalline insulin determined by semi-micro Kjeldahl analysis. The procedure has been described earlier [2]. The concentration of insulin in the crystalline state, $C$, is expressed in per cent of the total concentration of insulin in the suspension.

4. *Determination of the total volume ($\mathfrak{V}$) and surface ($\mathfrak{S}$) of crystals, and the cumulative distribution of crystal sizes ($\mathfrak{P}$).**

The cumulative distributions of crystal sizes are $\mathfrak{P}(l, t)$. At the time $t$, $\mathfrak{P}$ crystals per ml are bigger than $l$. Hence, the distribution of crystal sizes is

$$\mathfrak{p}(l,t) = -\frac{\delta \mathfrak{P}(l,t)}{\delta l} \tag{1}$$

The total number, length, surface and volume of crystals per ml of suspension are defined by

$$\mathfrak{N}(t) = \int_0^{l_{max}} \mathfrak{p}\ (l,t)\ \mathrm{d}l \tag{2}$$

$$\mathfrak{L}(t) = \int_0^{l_{max}} l\ \mathfrak{p}\ (l,t)\ \mathrm{d}l \tag{3}$$

$$\mathfrak{S}(t) = \int_0^{l_{max}} l^2\ \mathfrak{p}\ (l,t)\ \mathrm{d}l \tag{4}$$

$$\mathfrak{V}(t) = \int_0^{l_{max}} l^3\ \mathfrak{p}\ (l,t)\ \mathrm{d}l \tag{5}$$

Thus, an insulin rhombohedron with size $l = 1\ \mu$ has unit size, unit surface and unit volume.

The $\mathfrak{N}(t)$ values are not estimated directly from countings, as it is difficult to count very small crystals. The surface areas and the volumes are calculated from the $\mathfrak{P}(l, t)$ estimates obtained by the counting technique described earlier [2]. In analogy with the previous [2] formula for $V_i$

$$S_i = f_{si} \cdot n_i' \tag{6}$$

where

$$f_{si} = \frac{3m_i^2 + d_i^2}{12\ c} \tag{7}$$

## THEORY

### 1. The growth of crystals.

The time, $t$, is counted in min from the moment buffer is added to the insulin solution. It is assumed that crystals are formed so early that

$$\dot{\mathfrak{N}}\ (0) > 0 \tag{8}$$

where the dot symbolizes differentiation with respect to time. The number of crystals per ml of suspension increases with time as shown by the fictitious curve in Fig. 1.

* These symbols are typed in gothics when the independent variable is $t$ and in italics when it is $g$.

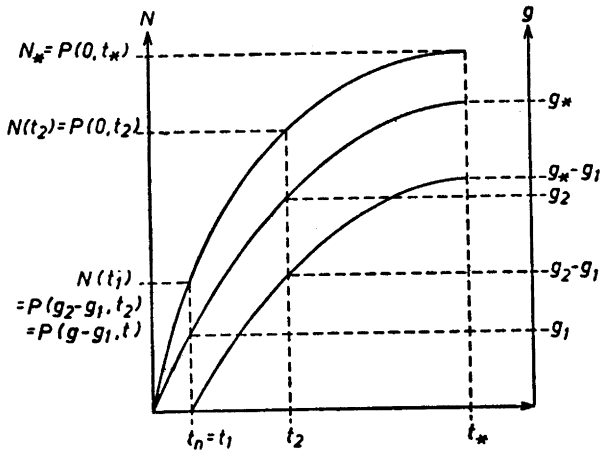


*Fig. 1.*

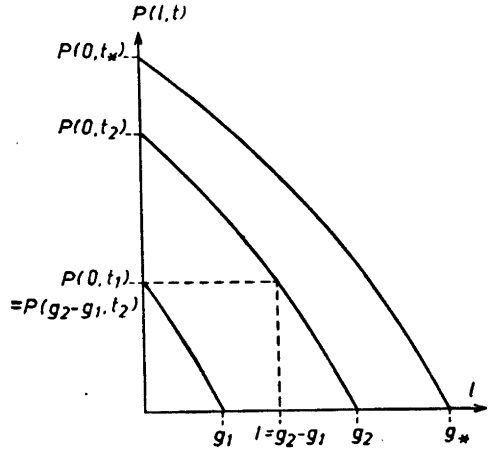


*Fig. 2.*

The crystallization is finished at the time $t_*$ defined by

$$\dot{C}(t_*) = 0 \tag{9}$$

At that time the crystals are not further increasing in number, *i. e.*

$$\dot{\mathfrak{N}}(t_*) = 0 \tag{10}$$

Strictly speaking $t_* = \infty$, but in an actual experiment any $t$ value above a certain limit may be symbolized through $t_*$.

The growth of a crystal proceeds as insulin is deposited on the faces which thereby move outwards in directions perpendicular to the faces. The rhombohedral shape is maintained, and the three faces meeting in an obtuse vertex move with identical rates. It is of course possible that the three opposite faces move at another rate, or even not at all. It is now assumed that the rate of movement of a face does not at any time vary between crystals having the same or different sizes. As the dimensions of a crystal are completely defined by the size, $l$, it follows from this assumption that at the time $t$ all crystals grow with the same linear rate of growth, $\dot{l}(t)$. In order to distinguish between size and growth, $\dot{l}$ is denoted by the symbol $\dot{g}$. It is seen that $\dot{g}$ is a function of one parameter only, *e. g.* the time.

The maximum crystal size

$$l_{\max}\ (t) = \int_0^t \dot{g}(t)\mathrm{d}t = g(t) \tag{11}$$

is shown by the middle fictitious curve in Fig. 1. While this $g(t)$ curve indicates the size of crystals formed at $t = 0$, the curve below indicates the size of crystals formed at the arbitrary moment $t = t_n$. The moment $t_n$ is called the moment of nucleation of these crystals. As all crystals grow with identical rates of linear growth, it follows that the size of these crystals is always $g(t)$—$g(t_n)$.

The cumulative distributions, $\mathfrak{P}(l,\ t)$, can be expressed in terms of $\mathfrak{N}(t)$ and $g(t)$. The $\mathfrak{P}$ function is considered at the arbitrary moment $t = t_2$ (Fig. 2). The total number of crystals is

$$\mathfrak{P}(0,\ t_2) = \mathfrak{N}(t_2) \tag{12}$$

At the moment $t_2$ no crystals are bigger than the maximum size, $g_2 = g(t_2)$, and consequently

$$\mathfrak{P}(g_2,\ t_2) = 0 \tag{13}$$

The crystals of the size $l$ were formed at the moment $t$ when the biggest crystals had the size $g_2$—$l$. At that moment $\mathfrak{N}(t)$ crystals were present. These are consequently the crystals which at the time $t_2$ are bigger than $l$. Hence,

$$\mathfrak{P}(l,\ t_2) = \mathfrak{N}(t) \tag{14}$$

where $t$ is derived from

$$g(t) = g_2 - l \tag{15}$$

Inversely,

$$\mathfrak{N}(t) = \mathfrak{P}(g_2—g(t),\ t_2) \tag{16}$$

or with $g$ as sole parameter,

$$N(g) = P(g_2—g,\ g_2) = P(g_3—g,\ g_3),\ \textit{etc.} \tag{17}$$

Thus, the number of crystals can be calculated from determinations of $P$ and $g$.

As shown graphically in Fig. 2, eqn. (17) reveals the relationship between the $P$-functions at different $g$ (or $t$) values. The horizontal distance between two $P(l)$ curves is constant, since it follows from eqn. (17) that

$$P(l,\ g_1) = P(l+g_2—g_1,\ g_2) \tag{18}$$

The same horizontal displacement of the distributions, eqn. (1), is observed by differentiation of eqn. (18) which thereby becomes

$$p(l,\ g_1) = p(l+g_2—g_1,\ g_2) \tag{19}$$

## 2. The nucleation

It is assumed that nuclei are immediately transformed into crystals. Hence, the rate of nucleation is

$$\mathfrak{h}(t) = \dot{\mathfrak{N}}(t) \tag{20}$$

It follows from this definition and eqn. (16) that

$$\mathfrak{h}(t) = p(g_2—g(t),\ g_2) \cdot \dot{g}(t) \tag{21}$$

Thus, the rate of nucleation at the moment $t$ is the product of a value in the distributions of crystal sizes and the growth rate. The physical meaning of this equation can be made clear from Fig. 2 by moving one of the curves the distance $\mathrm{d}g$ to the right in the time $\mathrm{d}t$. The point of interception between curve and $P$-axis thereby moves $\mathrm{d}N$ upwards.

The derivative of $N$ with respect to $g$

$$h_s(g) = p(g_2—g(t),\ g_2) \tag{22}$$

is obtained from eqn. (21). It is called the specific rate of nucleation and is simply a value in the distributions of crystal sizes. When the function $h_s(g)$ is known, it follows from eqn. (22) that the crystals are completely defined for every value ($g_2$) of the growth function. Hence, not only the size distributions but also the total volume, $V$, surface, $S$, and length, $L$, are defined by $h_s(g)$ and the $g$-value. Inversely, as will be demonstrated, inferences about $h_s(g)$ can be drawn from $V(g)$, $S(g)$ and $L(g)$.

The volume is given by eqn. (5). According to eqn. (22)

$$p(l,\ g) = h_s(x) \tag{23}$$

where

$$l = g — x \tag{24}$$

from which

$$V(g) = \int_0^g (g—x)^3 h_s(x)\mathrm{d}x \tag{25}$$

or

$$V(g) = g^3N(g) — 3g^2\int_0^g x h_s(x)\mathrm{d}x + 3g\int_0^g x^2 h_s(x)\mathrm{d}x — \int_0^g x^3 h_s(x)\mathrm{d}x \tag{26}$$

The surface, $S$, and the length, $L$, may be written in a quite analogous way, *viz.*

$$S(g) = g^2N(g) — 2g\int_0^g x h_s(x)\mathrm{d}x + \int_0^g x^2 h_s(x)\mathrm{d}x \tag{27}$$

and

$$L(g) = gN(g) - \int_0^g x h_s(x)\mathrm{d}x \tag{28}$$

By differentiation of the eqns. (26), (27) and (28) with respect to $g$ the following is found:

$$\begin{array}{llllll} V' = 3S & S' = 2L & L' = N & N' = h_s(g) \\ V'' = 6L & S'' = 2N & L'' = N' & \\ V''' = 6N & S''' = 2N' & & \\ V'''' = 6N' & & & \end{array} \tag{29}$$

from which can be read that

$$h_s(g) = N'(g) = L''(g) = \frac{1}{2}S'''(g) = \frac{1}{6}V''''(g) \tag{30}$$

## 3. Volume and surface relations

From determinations of $S$ and $V$ in the course of crystallization, inferences can be drawn about the rates of growth and nucleation. The surface

$$S = f(V) \text{ or } \mathcal{S} = f(\mathcal{V}) \tag{31}$$

is expressed as a function of the volume. As according to eqn. (29)

$$\dot{\mathcal{V}} = 3\mathcal{S}\dot{g} \tag{32}$$

it follows that

$$\dot{g} = \dot{\mathcal{V}}/3f(\mathcal{V}) \tag{33}$$

The specific rate of nucleation is obtained from eqn. (30). The eqn. (31) is differentiated

$$S' = f'(V) \cdot V' \tag{34}$$

with respect to $g$. As according to eqn. (29)

$$V' = 3S \tag{35}$$

it follows that

$$S' = 3f'(V) \cdot S \tag{36}$$

and by inserting eqn. (31)

$$S' = 3f' \cdot f \tag{37}$$

When differentiation is repeated twice more, it is found that

$$h_s(g) = \frac{27}{2}\,(f^3\, f''' + 4f^2 f' f'' + f(f')^3) \tag{38}$$

and when eqn. (33) is inserted

$$h(t) = \frac{27}{6}\,\dot{\mathcal{V}}(f^2 f''' + 4ff' f'' + (f')^3) \tag{39}$$

As an example, let it be possible to express $S$ by

$$S = r \cdot V^u \qquad g(t_1) < g < (t_2) \tag{40}$$

where $r$ and $u$ are positive constants. The eqn. (38) then becomes

$$h_s(g) = 81r^4u\left(u-\frac{1}{2}\right)\left(u-\frac{2}{3}\right)V^{4u-3} \qquad g(t_1) < g < g(t_2) \tag{41}$$

A few special cases of eqn. (40) will now be considered.

*Case 1, $u = 2/3$.* From eqn. (41) it is seen that

$$h_s(g) = 0 \qquad g(t_1) < g < g(t_2) \tag{42}$$

which means that crystallization proceeds without nucleation. The inverse inference — that equation (40) is valid with $u = 2/3$ when $h_s = 0$ — is generally false as can be understood from the eqns. (26) and (27). It is, however, true when the crystal suspension is known to be monodisperse at $t = t_1$ as may happen, *e. g.*, from a previous seeding.

*Case 2, $u = 3/4$.* This case is compatible with the observations made during the initial stage of the insulin crystallization. The specific rate of nucleation is calculated from eqn. (41) to be

$$h_s(g) = \frac{3^4}{2^6}\,\mathrm{r}^4 \qquad g(t_1) < g < g(t_2) \tag{43}$$

Thus, the rate of nucleation is proportional to the rate of linear growth. As the distributions of crystal sizes are identical with $h_s(g)$, all sizes of crystals formed in the time interval occur with equal frequency, *i. e.*

$$\mathrm{p}(l,\ t) = \frac{3^4}{2^6}\,\mathrm{r}^4 \qquad (t_1 < t < t_2) \tag{44}$$

when

$$l < g(t) - g(t_1)$$

*Case 3, $u = 1$.* This is the case actually observed except for the initial period of crystallization. The eqn. (41) shows that

$$h_s(g) = \frac{27}{2}\,r^4V \qquad g(t_1) < g < g(t_2) \tag{45}$$

which by means of eqn. (40) is changed into

$$h_s(g) = \frac{27}{2}\,r^3S \qquad g(t_1) < g < g(t_2) \tag{46}$$

From the eqns. (29) and (40) it follows further that

$$h_s(g) = 9r^2L \qquad g(t_1) < g < g(t_2) \tag{47}$$

and

$$h_s(g) = 3rN \qquad g(t_1) < g < g(t_2) \tag{48}$$

Thus, in the time interval, $t_1 < t < t_2$, the specific rate of nucleation is proportional to the number of crystals present, their total length, surface or volume. The constant mutual ratios are

$$N = 3rL = \frac{9}{2} r^2 S = \frac{9}{2} r^3 V \qquad g(t_1) < g < g(t_2) \tag{49}$$

This equation shows further that the average crystal size, $L/N$, is constantly

$$\bar{l} = 1/3r \qquad (t_1 < t < t_2) \tag{50}$$

The distribution of crystal sizes is obtained from eqns. (16), (22) and (48) which give

$$\mathfrak{p}(l,\ t) = 3r\mathfrak{P}(l,\ t) \qquad (t_1 < t < t_2) \tag{51}$$

when

$$l < g(t) - g(t_1)$$

From eqns. (1) and (51) it follows that

$$\mathfrak{p}(l,\ t) = f(t)\ \mathrm{e}^{-3rl} \qquad (t_1 < t < t_2 \text{ and } l < g(t)—g(t_1)) \tag{52}$$

where $f(t)$ is a function of time only.

The cumulative distribution of crystal sizes is obtained by means of eqn. (51) from which

$$\mathrm{d}\ln\mathfrak{P}(l,\ t) = —3r\mathrm{d}l \tag{53}$$

and

$$\ln\mathfrak{P}(l,\ t) = —3rl + \ln\mathfrak{N}(t) \tag{54}$$

in which

$$t_1 < t < t_2 \qquad \text{and} \qquad l < g(t) - g(t_1) \tag{55}$$

Thus in the interval (55) the logarithms of the cumulative distributions are graphically represented by the straight lines, eqn. (54). As the slopes of these lines are identical, the lines may be moved in a graph to cover each other. At $t = t_2$, eqn. (54) gives

$$\ln\mathfrak{P}(g(t_2) - g(t),\ t_2) = — 3r(g(t_2) — g(t)) + \ln\mathfrak{N}(t_2) \tag{56}$$

Using eqn. (16), the addition of the eqns. (54) and (56) shows that

$$\ln\mathfrak{P}(l,\ t) = — 3rx + \ln\mathfrak{N}(t_2) \tag{57}$$

where

$$x = l + g(t_2) — g(t) \tag{58}$$

Thus, all $\mathfrak{P}$-values observed in the time interval $t_1 < t < t_2$ will lie on the straight line, eqn. (57), provided the $l$-values are smaller than $g(t)—g(t_1)$.

## 4. The origin of nuclei

A fundamental theory of nucleation and crystal growth has been developed by Christiansen [3]. According to this theory

$$\dot{\mathfrak{N}} = \text{const. } m^p \tag{59}$$

where $m$ is the concentration of substance present as monomer dissolved particles and $p$ is a positive integer. This relation was demonstrated by the same author to be valid for the precipitation of barium sulphate ($p = 8$). The rate of linear growth of barium sulphate was found to be

$$\dot{g} = \text{const. } m^q \tag{60}$$

where $q$ is positive ($q = 4$). The power $q$ was also explained by the theory.

When the eqns. (59) and (60) are applied to the present case of insulin crystallization, the specific rate of nucleation is found to be

$$h_s(g) = \text{const. } \dot{g}^{(\frac{p}{q}-1)} \tag{61}$$

This equation is compatible with the observations in the initial stage of insulin crystallization, although as already mentioned,

$$h_s(g) = \text{const. } (V,\ S,\ L \text{ or } N) \tag{62}$$

in the main part of crystallization. Hence, the hypothesis is set up that the insulin nuclei have different origins. It is assumed that some of the insulin nuclei are formed spontaneously according to eqn. (61), and further that other insulin nuclei are formed by interaction of the solution with the faces, edges or vertices of the crystals. Thus, in the initial stage of crystallization, where crystals are few and interaction therefore insignificant, nucleation may proceed according to eqn. (61). Later the interaction becomes of overwhelming significance and thereby masks the process, eqn. (61), so much that the specific rate of nucleation follows eqn. (62).

## ANALYSIS OF DATA FROM THE CRYSTALLIZATION EXPERIMENT

### 1. Changes with time in the concentrations of dissolved and crystalline insulin

At $t = 0$ the insulin is almost altogether (94 %) precipitated as amorphous particles. In the course of the following 50 min, 26 % of the insulin is gradually dissolved. At $t = 50$, the crystalline fraction appears in chemically measurable quantities. It increases first rapidly and later more slowly as seen in Fig. 3 and Table 1. At the same time the concentration of dissolved insulin decreases. The drop in $c_s$ indicates that crystals grow from dissolved insulin.

### 2. The linear growth rate

As the crystals apparently grow from dissolved insulin being deposited on the faces, it is of prime interest to determine the functional relationship between the linear growth rate and the concentration of dissolved insulin. The $C(t)$ and $c_s(t)$ data will be used first in connection with eqn. (33). In the next paragraph it will be demonstrated that

$$G = rv \qquad (t > 60) \tag{63}$$

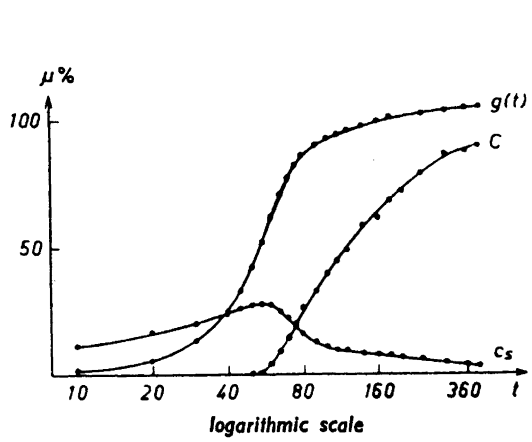


*Fig. 3.*

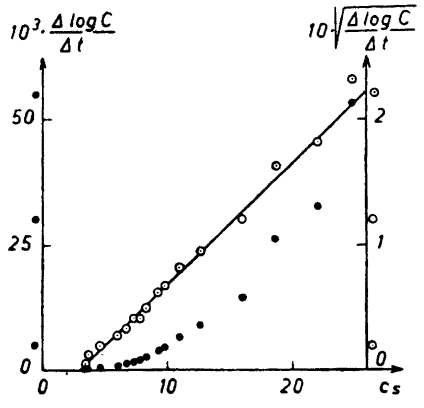


*Fig. 4.*

where $r$ is a constant. Hence,

$$\dot{g} = \frac{(\dot{\ln v})}{3r} \qquad (t > 60) \tag{64}$$

and

$$\dot{g} = \frac{(\dot{\log C})}{3rM} \qquad (t > 60),\ (M = 0.4343) \tag{65}$$

Estimates of the growth rate can be obtained from

$$\dot{g}_i \approx \frac{1}{3rM} \cdot \frac{\log C_{i+1} - \log C_{i-1}}{t_{i+1} - t_{i-1}} \qquad (t > 60) \tag{66}$$

in which $i$ is the sample number. The fractions on the right side of the equation are calculated and plotted against $c_s$ in Fig. 4.

The configuration of the points in the graph suggests a power relation, and when the square roots of the fractions are tentatively plotted against $c_s$, it appears that

$$\dot{g}(t) = c_7(c_s(t) - 3)^2 \qquad (t > 60) \tag{67}$$

where $c_7$ is a constant. Assuming this equation to be valid for $t < 60$ also, it follows that

$$g(t) = c_7 \cdot \int_0^t (c_s - 3)^2 \, dt \tag{68}$$

The value of the integral corresponding to each sample is calculated by the approximative formula

$$\int_0^{t_x} (c_s - 3)^2 \, dt \approx \frac{1}{2} \sum_{i=1}^{i=x} (t_i - t_{i-1}) \left[(c_{s_i} - 3)^2 + (c_{s_{i-1}} - 3)^2\right] \tag{69}$$

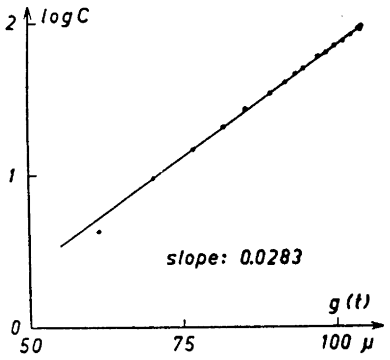


*Fig. 5.*

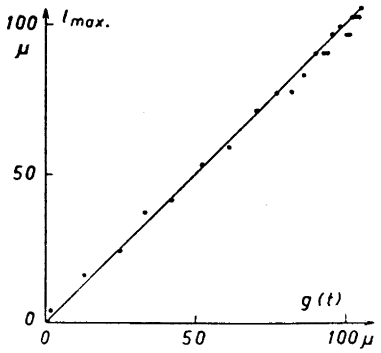


*Fig. 6.*

where i = 1 when $t = 0$. When crystallization was finished, the biggest crystals found were $g_* = 105\ \mu$. The value of the integral at the same time, $t_*$, is calculated as $3.20 \times 10^4$. Hence,

$$c_7 = \frac{105}{3.20} \times 10^{-4} = 3.28 \times 10^{-3} \tag{70}$$

which is used in eqn. (68) to calculate a $g(t)$ estimate for each sample. These estimates are now plotted against $\log C$ in Fig. 5.

The slope of the straight line is 0.0283. As from eqn. (65)

$$g(t) = \frac{\log C(t)}{3rM} + \text{const.} \quad (t > 60) \tag{71}$$

it follows that

$$r = \frac{0.0283}{3M} = 0.022 \tag{72}$$

In order to compare the $g(t)$ estimates with the size of crystals actually found, $g(t)$ is plotted (Fig. 6) against the maximum size, $l_{max}$, of crystals observed in all samples including those from $t < 60$. The calculated and observed maximum sizes are in agreement, also for $t < 60$.

Finally the $g(t)$ values are plotted against different quantiles in the cumulative distributions of crystal sizes (Fig. 7). The quantiles are derived from the countings by picking out of the counting tables the crystal sizes corresponding to the $\varrho$-value in question. The theoretical lines inserted in the figure were calculated from eqns. (68) and (88). It appears from Fig. 7 that crystals of different sizes grow with identical rates of growth and further that this rate of growth is $\dot{g}$ as calculated by means of eqn. (67). Thus, the linear growth rate is proportional to the square of $(c_s - 3)$, which may be the supersaturation but presumably is not. When the crystal suspension is left for several days, the $c_s$ value drops below 2 %. Further, when the crystals are filtered off and resuspended in a fresh insulin-free medium of the same small molecular

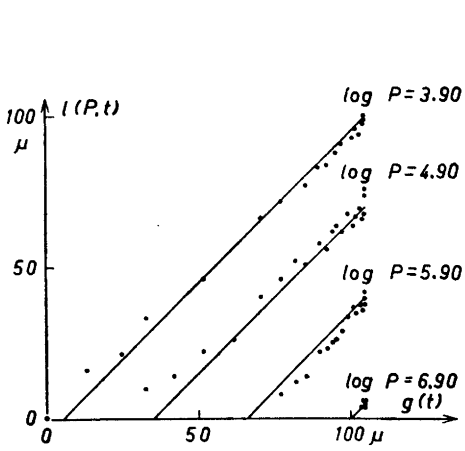


*Fig. 7.*

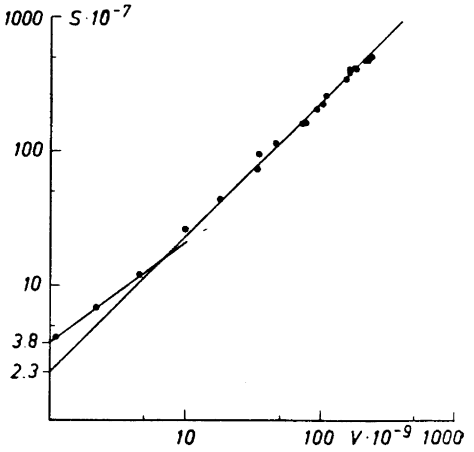


*Fig. 8.*

composition, $c_s$ is found to be only 0.5 % after equilibration for two days. These supplementary findings must be considered in the chemical interpretation of eqn. (67).

### 3. The rate of nucleation

In the present case of crystallization the functional relationship between the total volume and surface of crystals is so simple that inferences can be drawn about the specific rate of nucleation. The $V$ and $S$ means from two, and sometimes more countings are plotted against each other on double logarithmic paper in Fig. 8.

The crystallization is divided into two periods, A and B, by the two straight lines with slopes 3/4 and 1. In the first initial stage

$$\text{(A)} \qquad S \approx 3.8 \times 10^{\frac{1}{4}} \times V^{\frac{3}{4}} \qquad (10^9 < V < 10^{10}) \tag{73}$$

and in the second main part of the crystallization

$$\text{(B)} \qquad S \approx 0.023 \times V^1 \qquad (10^{10} < V < V_*) \tag{74}$$

The value (0.023) of the constant in the last equation is in agreement with the former value (0.022) from the preceding paragraph.

From eqns. (73) and (43) it is calculated that

$$\text{(A)} \qquad h_s \approx \frac{3^4}{2^6} \times 3.8^4 \times 10^1 = 2 \times 10^3 \tag{75}$$

and from eqns. (74) and (48) it follows that

$$\text{(B)} \qquad h_s \approx 0.07 \times N \tag{76}$$

The eqn. (75) is a special case of eqn. (61), *viz.* the case where $p = q$. The assumption is now made, that during the whole period of crystallization nucleation proceeds according to eqn. (61) in which $p = q$. There is no incompatibility between this assumption and the eqn. (76) when it is further assumed that a second nucleation process takes place on the crystals. When $N$ exceeds $10^5$ this second process becomes dominating and especially as $N$ increases to $10^7$, the final value.

It follows from these assumptions that

$$h_s = a + b_1N + b_2L + b_3S \tag{77}$$

in which $a$ is the constant specific rate of spontaneous nucleation, $b_1N$ is the specific rate of nucleation from the vertices, $b_2L$ at the edges and $b_3S$ on the faces. As $N$, $L$, $S$ and $V$ become mutually proportional when about 1 % of the final crystals are formed, it is not possible to infer from the data whether or not the surface nucleation occurs on the faces or rather at the edges or vertices. However, the agreement of the data with

$$h_s = a_1 + b_1N \quad \text{(vertices)} \tag{78}$$
$$h_s = a_2 + b_2L \quad \text{(edges)} \tag{79}$$
$$h_s = a_3 + b_3S \quad \text{(faces)} \tag{80}$$
$$h_s = a_4 + b_4V \quad \text{(eqn. 45)} \tag{81}$$

will be investigated. When these four equations are differentiated by means of eqn. (29) they become ordinary differential equations

$$N'(g) = a_1 + b_1N(g) \tag{82}$$
$$N''(g) = b_2N(g) \tag{83}$$
$$N'''(g) = 2b_3N(g) \tag{84}$$
$$N''''(g) = 6b_4N(g) \tag{85}$$

having the solutions

$$N(g) = c_1\ e^{\alpha}\ (1—e^{-\alpha});\ c_1 = \frac{a_1}{b_1}\ ;\ \alpha = b_1g \tag{86}$$

$$N(g) = c_2\ \sinh\alpha;\ c_2 = \frac{a_2}{\sqrt{b_2}}\ ;\ \alpha = \sqrt{b_2}\cdot g \tag{87}$$

$$N(g) = c_3\ e^{\alpha}[1 + e^{-\frac{3}{2}\alpha}\ (\sqrt{3}\cdot\sin\beta - \cos\beta)];\ c_3 = \frac{a_3}{3c_4}\ ;\ c_4 = \sqrt[3]{2b_3}\ ; \tag{88}$$

$$\alpha = c_4g;\ \beta = \frac{\sqrt{3}}{2}\ \alpha$$

$$N(g) = c_5\ (\sinh\alpha + \sin\alpha);\ c_5 = \frac{a_4}{2c_6}\ ;\ c_6 = \sqrt[4]{6b_4}\ ;\ \alpha = c_6\ g \tag{89}$$

It follows from eqn. (48) that

$$\frac{d\ln N}{dg} = 3r \tag{90}$$

in the second period of crystallization. Hence,

$$a = 3rg = \frac{0.0283}{0.4343}\, g = 0.0652\, g \tag{91}$$

or

$$b_1 = \sqrt{b_2} = \sqrt[3]{2b_3} = \sqrt[4]{6b_4} = 0.0652 \tag{92}$$

When crystallization is finished the growth function has reached the value $g_* = 105\ \mu$ and, as will be explained later, the number of crystals is $1.02 \times 10^7$ per ml. Hence,

$$c_1 = \frac{1}{2}\, c_2 = c_3 = \frac{1}{2}\, c_5 = 1.02 \times 10^7/\text{exp.}\ (0.0652 \times 105) = 1.1 \times 10^4 \tag{93}$$

The numerical values of the constants, eqns. (92) and (93), are introduced into the eqns. (86)—(89), which are then used to calculate sets of $N(g)$ values to be compared with $P$-values from the observed cumulative distributions of crystal sizes.

The cumulative distribution of sizes when crystallization is finished (asterisk) can be expressed by these equations, as from eqns. (14) and (15)

$$\mathrm{P}_*(l) = N(g) \tag{94}$$

where

$$g = g_* - l \tag{95}$$

The curves in Fig. 9 represent the logarithms of $N(g)$ from eqns. (86), (87), (88), (89), (92) and (93) against $l$, where $l$ is connected to $g$ by eqn. (95). The dots represent the almost coinciding observation data gained from the three last samples. The data fit best to the curves calculated according to eqns. (87) and (88) corresponding to nucleation at edges and faces, respectively.

In order to utilize the information gathered from all the countings, their data are compiled in one graph (Fig. 10). Let the crystal sizes at $t = t_*$ be denoted by $x$, then it follows from eqn. (18) that

$$\mathrm{P}_*(x) = \mathrm{P}(l,\ t) \tag{96}$$

where

$$x = l + g_* - g(t) \tag{97}$$

The substitution, eqn. (97), is used in the presentation of all the $\mathrm{P}(l, t)$ values observed (Fig. 10). The scattering of the dots increases towards the lower region of the graph where the P-values are based upon counts of a few crystals only. The trend of the dots is in agreement with Fig. 9 and with the theoretical curves although too many points are below the curve in the interval $20 < x < 40$. The eqns. (87) and (88) fit best to the observations depicted in this figure also.

When the method of least squares is used (Fig. 10) on $\log \mathrm{P}_*(x)$ where $\log \mathrm{P}_*(x) > 4.4$, the slope is found to be —0.0283. A straight line with this slope was already expected in accordance with eqn. (57) and Fig. 5. The line intercepts the $\log \mathrm{P}_*(x)$ axis at 7.01. Hence, the final number of crystals is estimated as $1.02 \times 10^7$ per ml.

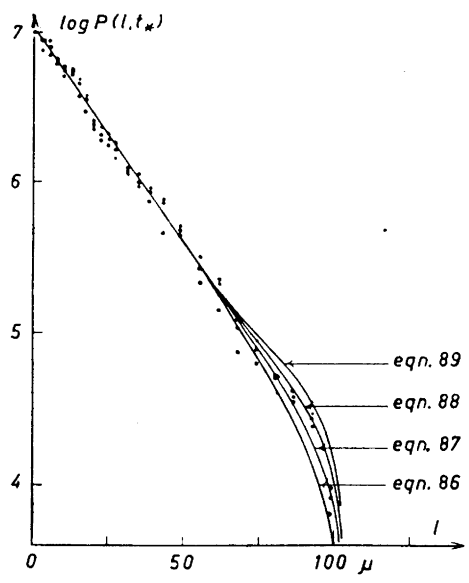


*Fig. 9.*

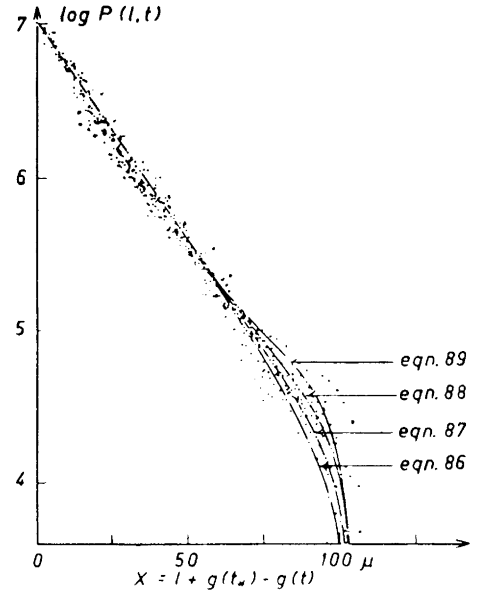


*Fig. 10.*

The total volumes and surfaces expected in accordance with eqns. (68), (88), (92) and (93) have been calculated and presented in Table 1 and as a curve in Fig. 11 in comparison with the observed values. The eqn. (80) is written

$$h_s = a_3 + \frac{b_3}{3} \cdot V' \tag{98}$$

which by integration becomes

$$V(g) = \frac{3}{b_3} [N(g) - a_3 g] \tag{99}$$

from which $V(g)$ can be calculated by inserting $g$ from eqn. (68) and $N(g)$ from eqn. (88). It follows from eqn. (99) that the number of insulin crystals is proportional to the weight or volume of crystallized insulin, provided $a_3$ is relatively small. The surface was calculated from eqn. (80)

$$S = \frac{1}{b_3} [N'(g) - a_3] \tag{100}$$

in which

$$N'(g) = \frac{1}{3} a_3 \, (e^{a} + 2e^{-\frac{1}{2}a} \cos\beta) \tag{101}$$

is obtained by differentiating eqn. (88). There is a reasonable agreement between the calculated and observed values.

*Table 1. Data relating to the crystallization experiment.*

| (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) | (11) | (12) |
|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 2 | 6.1 | 0.0 | 0.0 | — | — | — | — | 3 | 2 | 0.1 |
| 2 | 10 | 11.0 | 0.4 | 1.3 | — | — | — | — | 21 | 2 | 0.4 |
| 3 | 20 | 16.7 | 0.0 | 5.2 | — | — | — | — | 62 | 2 | 1.3 |
| 4 | 30 | 19.8 | 0.0 | 13.1 | 0.1 | 0.0 | 0.5 | 0.2 | 93 | 2 | 2 |
| 5 | 40 | 24.1 | 0.0 | 24.9 | 0.3 | 0.2 | 1.4 | 1.1 | 146 | 4 | 5 |
| 6 | 45 | 25.9 | 0.0 | 32.8 | 1.1 | 0.7 | 4.1 | 2.7 | 172 | 6 | 10 |
| 7 | 50 | 27.3 | 0.4 | 42.0 | 2.2 | 1.9 | 6.8 | 6.2 | 178 | 11 | 19 |
| 8 | 55 | 27.4 | 0.8 | 51.5 | 4.6 | 4.5 | 12 | 13 | 180 | 20 | 36 |
| 9 | 60 | 27.3 | 4.2 | 61.3 | 10 | 10 | 26 | 26 | 178 | 38 | 68 |
| 10 | 65 | 24.9 | 9.4 | 70.2 | 18 | 20 | 43 | 47 | 157 | 68 | 107 |
| 11 | 70 | 22.1 | 14.3 | 77.1 | 34 | 33 | 73 | 76 | 120 | 106 | 127 |
| 12 | 75 | 18.8 | 20.0 | 82.0 | 35 | 47 | 96 | 105 | 82 | 147 | 120 |
| 13 | 80 | 16.0 | 26.2 | 85.6 | 47 | 59 | 115 | 132 | 55 | 185 | 102 |
| 14 | 90 | 12.6 | 32.9 | 89.9 | 78 | 80 | 168 | 175 | 30 | 245 | 74 |
| 15 | 100 | 11.0 | 39.8 | 92.5 | 72 | 95 | 162 | 208 | 21 | 291 | 61 |
| 16 | 110 | 9.8 | 44.6 | 94.1 | 94 | 107 | 210 | 232 | 15 | 323 | 49 |
| 17 | 120 | 9.3 | 48.8 | 95.4 | 104 | 116 | 231 | 252 | 13 | 351 | 46 |
| 18 | 140 | 8.3 | 58.4 | 97.7 | 109 | 135 | 261 | 293 | 9 | 408 | 38 |
| 19 | 160 | 7.9 | 61.2 | 99.4 | 155 | 150 | 355 | 327 | 8 | 456 | 36 |
| 20 | 180 | 7.3 | 68.1 | 101.0 | 164 | 165 | 406 | 365 | 6 | 506 | 31 |
| 21 | 200 | 6.7 | 72.0 | 102.0 | 165 | 179 | 388 | 388 | 5 | 540 | 24 |
| 22 | 240 | 6.0 | 78.8 | 103.3 | 181 | 196 | 418 | 425 | 3 | 588 | 18 |
| 23 | 300 | 4.6 | 86.5 | 104.6 | 177 | 212 | 426 | 461 | 1 | 640 | 5 |
| 24 | 360 | 3.7 | 87.6 | 105.0 | 239 | 219 | 517 | 474 | 0 | 657 | 1 |
| 25 | 400 | 3.4 | 89.6 | 105.0 | 214 | 219 | 483 | 474 | 0 | 657 | 0.7 |
| 26 | 1 400 | 2.3 | 92.7 | 105.0 | 222 | 219 | 485 | 474 | 0 | 657 | 0.0 |

Legend to table:
(1) $i$, sample number.
(2) $t$, time in minutes from the moment of mixing.
(3) $c_S$, the percentage of insulin in solution.
(4) $C$, the percentage of insulin in the crystalline state.
(5) $g(t) = 3.28 \times 10^{-3} \int_0^t (c_S - 3.0)^2 \, dt$.
(6) $V \times 10^{-9}$, calculated from the distributions of crystal sizes observed. See eqns. (1) and (5).
(7) $V \times 10^{-9}$, predicted values from eqn. (99) in which $a_3 = 2\,094$, $b_3 = 1.385 \times 10^{-4}$ and $N(g)$ is calculated from eqn. (88) where $g = g(t)$ from column (5).
(8) $S \times 10^{-7}$, calculated from the distributions of crystal sizes observed. See eqns. (1), (4), (6) and (7).
(9) $S \times 10^{-7}$, predicted values from eqns. (100) and (101) in which $a_3 = 2\,094$, $b_3 = 1.385 \times 10^{-4}$ and $g = g(t)$ from column (5).
(10) $\dot{g} \times 10^2 = 0.328 \, (c_S - 3.0)^2$
(11) $h_S \times 10^{-3} = 10^{-3} \, N'(g)$. $N'(g)$ is calculated from eqn. (101), $g = g(t)$ from column (5).
(12) $\dot{h} \times 10^{-3} = 10^{-3} \times \dot{g} \times h_S$.

## 4. The microscopic appearance

The photomicrograph (Fig. 12) shows the appearance of the crystals upon completion of crystallization. The crystals are apparently not distributed in space at random. The number of crystals smaller than 4 $\mu$ (175 × magnified) was counted in the photomicrograph and estimated as 41. Out of these, 15 were free, 10 occurred in pairs, 14 were found at the edges or vertices of

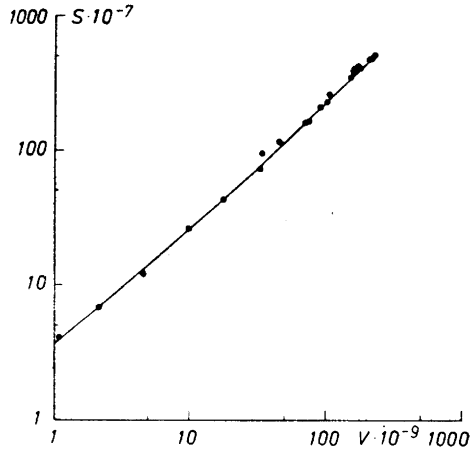


*Fig. 11.*

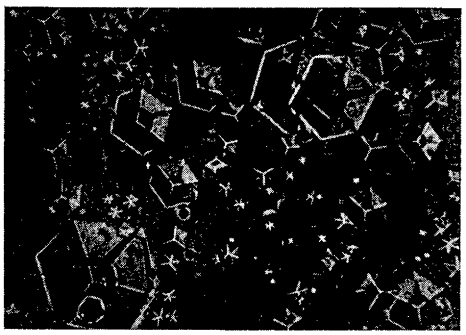

*Fig. 12.*

bigger crystals, and 2 on the crystal faces. It is concluded that either small new crystals wander about and become attracted and held fast by other crystals, or they originate on them.

## OTHER EXPERIMENTS

1. *Crystallization of pig insulin from a clear solution.* The eqns. (78), (79), (80) and (81) which were formed on the hypotheses set up to explain the $\varrho(l, t)$ function contain some remarkable consequences. The specific rate of nucleation, being a function of $g$, is seen to be independent of $\dot{g}$ and of the concentration of dissolved insulin. However, in a single experiment $\dot{g}$ is certainly a function of $g$ and it is therefore possible to express $h_s$ and $\mathfrak{h}$ by means of $\dot{g}$. It turns out that $h_s(\dot{g})$ and $\mathfrak{h}(\dot{g})$ are complicated functions which seem to have no underlying pattern from the point of view of chemical interpretation. Thus, up to the present it has not been possible to find any explanation except for the simple hypothesis of nucleation by crystal self-reproduction. In order to test experimentally whether $h_s$ is chemically independent of $\dot{g}$, crystallizations were made in which the concentration of dissolved insulin, and hence $\dot{g}$, varies with $g$ in a different pattern, *i. e.* by crystallizing from a clear solution.

The composition was: 1 % pig insulin, 0.04 % $Zn^{++}$, 0.05 M sodium citrate, 15 % (v/v) acetone, pH = 6.2. Volume 1 l. The crystallizations were carried out in ordinary glass beakers equipped with a mechanical glass stirrer. The acetone prevented any foam formation. In one of the two experiments the glass was covered with a paraffin layer and the surface was covered with paraffin oil. The resulting cumulative distributions of crystal sizes are shown in Fig. 13.

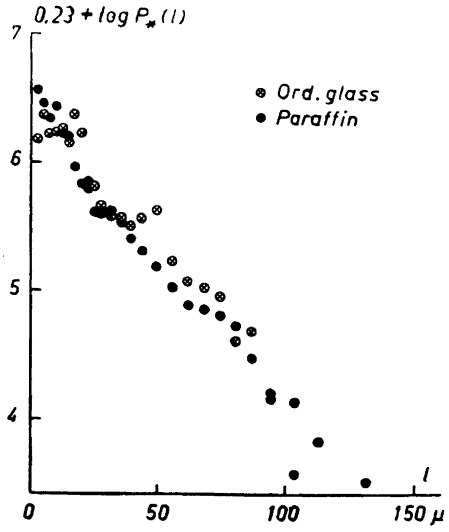


*Fig. 13.*

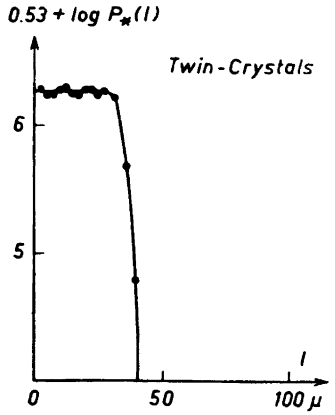


*Fig. 14.*

To all log $\rho_*$ values are added log $\frac{1.7}{1} = 0.23$ so as to render direct comparison with previous figures possible. It is seen that the change of the $\dot{g}(g)$ function together with the elimination of the amorphous phase did not abolish the straight line relationship expected in accordance with the hypothesis of crystal reproduction.

2. *Crystallization of beef insulin from a clear solution.* An experiment was made using the same composition as in the case of pig insulin crystallization except for the insulin concentration which was 0.5 % and pH which was 6.7. The crystals were as expected more or less distorted apparently by twin formation [1]. The resulting cumulative distribution of crystal sizes is shown in Fig. 14. The shape of the curve is inconsistent with the eqns. (86)—(89) derived from the hypothesis of crystal reproduction. Furthermore the distribution of sizes shows a maximum at about 35 $\mu$ corresponding to a rather monodisperse suspension. It means that $h_s$ drops sharply when the concentration of dissolved insulin decreases. Assuming that the nucleation occurs spontaneously within the solution and further that eqn. (61) can be employed in the present case, it follows that $p > q$ as in the case of barium sulphate. Thus, the mechanism of nucleation in the citrate-acetone buffer appears to depend on the species of insulin. In the sodium chloride medium, however, there is no such dependence as illustrated by the 4-litre experiment for which beef insulin was used.

It is concluded that the observed pattern of nucleation is in accordance with the fundamental concepts of spontaneous nucleation, eqn. (61), when the crystals are of the usual distorted twin-like type. The data of crystallization of plain single rhombohedrons are also in agreement with the fundamental con-

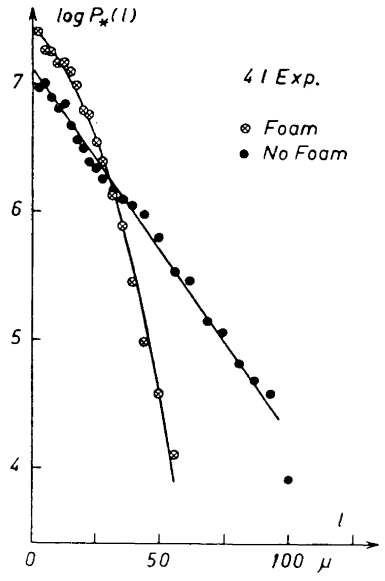


*Fig. 15.*

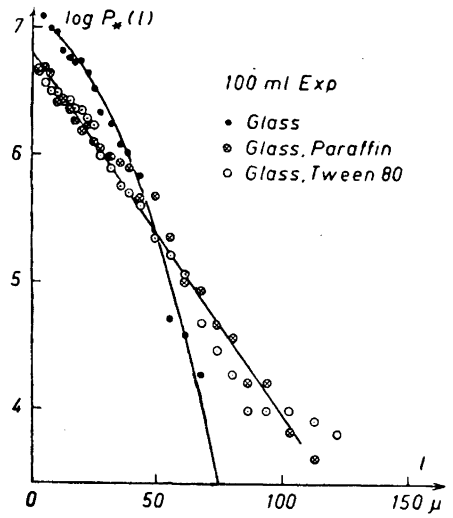


*Fig. 16.*

cepts when it is accepted that a second population of nuclei originates on the existing crystals. In this latter case it was estimated that $p = q$ in eqn. (61), but the masking effect of the apparent self-reproduction makes this estimate somewhat uncertain.

3. *The outer surface.* The investigation of the nucleation process is complicated by an apparent nucleation process taking place in the outer surface of the suspension.

When insulin crystallizes slowly from a calm solution, it is observed that crystals are primarily formed in the glass surface, in the surface exposed to air and on foreign bodies (*e. g.* a hair fiber) if such are suspended in the solution. In the 4-litre experiment some foam is formed on the surface. If the foam is not quantitatively removed, the crystallization yields a larger number of crystals, the maximum size becomes smaller and the graphical picture of $\log \varrho_*(l)$ becomes curved as shown in Fig. 15.

In another experiment the surface was increased (without perceptible foam) by reducing the volume of the crystallizing suspension to 100 ml. The resulting cumulative distribution of sizes (Fig. 16) resembles that of the 4-litre experiment with foam. Thus, it seems that nucleation may occur in the surface to such an extent as to cause a significant change in the distribution of sizes.

Inversely, it has been demonstrated by experiment, that the apparent excess nucleation from the surface can be reduced or eliminated. The 4-litre experiment was repeated on a 100 ml scale in 3 experiments carried out in identical apparatus. The one experiment served as the control for the case with relatively large surface. Number two beaker was covered with paraffin

and the surface of the suspension was covered with paraffin oil. The third suspension contained 0.05 % of the surface active agent Tween 80. It is seen from Fig. 16 that the paraffin cover as well as the Tween 80 causes a bigger maximum size of crystals, a smaller crystal number, and the reappearance of the straight line relationship characteristic of the 4-litre experiment when foam is absent.

## DISCUSSION

Nucleation may occur within the solution, in the interfaces between solution and glass or air and on the surface of the crystals. The great changes in nucleation produced by foam or other increases in the surface, by the paraffin coating, and by Tween 80 indicate that the rate of nucleation increases towards the surface. The hypothesis of surface nucleation on the crystals provides a simple explanation of the size distributions observed. It has not been possible to determine whether this nucleation is located on the faces or rather at the edges or vertices. It is interesting that this type of nucleation is apparently wholly or partially absent when the insulin crystallizes in the twin-like distorted shape. One can imagine that the crystal reproduction may be the repeated development and disengagement of twins.

## REFERENCES

1. Schlichtkrull, J. *Acta Chem. Scand.* **11** (1956) 1459.
2. Schlichtkrull, J. *Acta Chem. Scand.* **11** (1957) 291.
3. Christiansen, J. A. *Fysisk Tidsskrift* **2** (1955) 200.

Received October 24, 1956.