

## What is the maximum yield in the solid state cinnamic acid dimerisation? A combinatorial mathematical approach

GAUTAM R DESIRAJU\* and V KANNAN†

School of Chemistry\* and School of Mathematics†, University of Hyderabad, Central University PO, Hyderabad 500134, India

MS received 22 October 1984; revised 18 November 1985

**Abstract.** Recurrence relations have been employed to obtain the maximum possible theoretical conversion in a topochemically controlled organic solid state reaction. Experimental deviations from this analytical value of  $\approx 82\%$  reflect the importance of defects, molecular relaxation and surface phenomena in real crystals.

**Keywords.** Configuration; recurrence relation; power series; partial fractions; topochemistry; cinnamic acid; photodimerisation; maximum yield.

### 1. Introduction

One of the oldest and best studied of organic solid state reactions is the photodimerisation of *trans*-cinnamic acid in UV light (Schmidt *et al* 1964). This is a good example of topochemical control in organic crystals because the three-dimensional structure of the starting material exclusively determines the chemical structure of the product.

In the  $\alpha$ -form, the nearest neighbours are related by a centre of symmetry and react to give an inversion symmetry truxillic dimer. In the  $\beta$ -form, however, the nearest neighbours are translationally related and the product is a mirror-symmetry truxinic dimer. The crystal symmetry of the monomer is thus preserved in the molecular symmetry of the dimer.

One of the interesting and less-studied aspects of this reaction is the fact that the  $\alpha$ - and the  $\beta$ -forms of the reactant differ in the maximum possible theoretical conversion to dimer (Schmidt *et al* 1964). In the  $\alpha$ -form each molecule has only one nearest neighbour. This neighbour is related to it by a crystallographic centre of symmetry and photoreaction is only possible between two such neighbours. Thus close to quantitative conversions are possible and are in fact realised in practice (95–100%, figure 1). However, in the  $\beta$ -form all the molecules are related by translation in a row to form the stack. Each molecule in the stack can react with either of its two equally close nearest neighbours. Therefore, as a result of independent events in the stack, a monomer molecule may find itself eventually isolated (figure 2). Therefore, if the product does not recrystallise as a separate phase, yields of photodimer can never be quantitative. In fact, maximum yields of dimer can be as low as 30% for the  $\beta$ -acids (Schmidt *et al* 1964).

\* To whom all correspondence should be addressed.

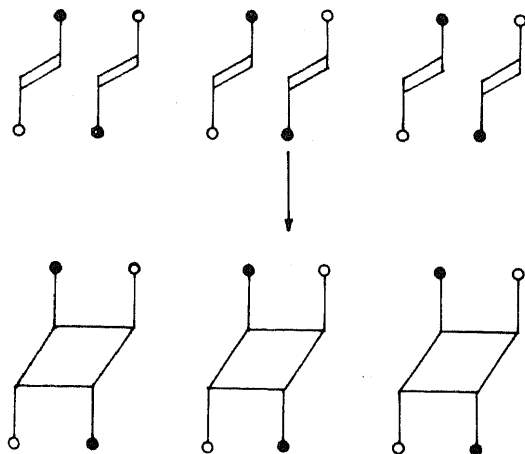


Figure 1. Photoreaction of crystalline  $\alpha$ -cinnamic acid (schematic).

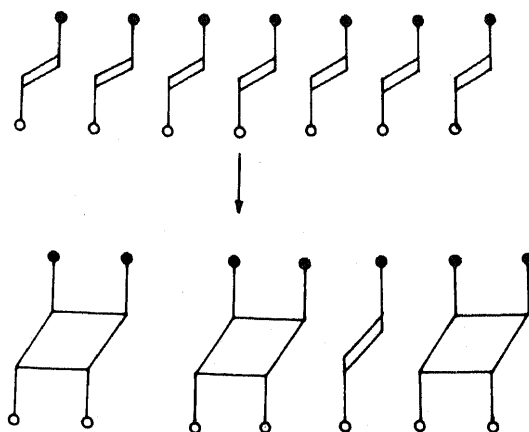


Figure 2. Photoreaction of crystalline  $\beta$ -cinnamic acid (schematic).

Can one, in fact, determine what the maximum possible theoretical yield is in a photoreaction of the  $\beta$ -crystals? The answer is yes, if a few simplifying assumptions are made. The result can be obtained analytically using recurrence relations.

### 2.1 Chemical formulation of the problem

The basic assumptions for the theoretical analysis of the photoreaction are as follows:

- Reaction is possible everywhere in the crystal at the outset and anywhere in the unreacted portions during the course of the reaction.
- Reaction at any molecule is with one of its two translationally related neighbours in the stack. If one neighbour has already reacted, reaction is only possible with the other. If both neighbours have reacted, the molecule cannot react.
- The reaction of molecules in one stack will not affect the molecules in any other

stack in the crystal. Within a stack, the probability that a reaction will occur at a particular site is not affected by a reaction having occurred nearby. In other words, each reaction is an independent event.

(d) The 'isolated' monomer molecules cannot crystallise out of the dimer matrix in the later stages of the reaction. This is a good assumption because quantitative yields are never obtained in the  $\beta$ -crystal photoreaction. In other words, the isolated monomers do not conglomerate and recrystallise during the later stages of the reaction as this would eventually lead to a quantitative yield of photo-product.

### 2.2 Mathematical formulation of the problem

The molecules can be abstractly represented by points in a line. A stack is such an arrangement of points. When two molecules react with each other, the corresponding points are joined to one another. Points that are not joined to any other point are called isolated points. The other points are called non-isolated points. Our chemical assumptions therefore mean that:

- (a) no two consecutive points can be isolated;
- (b) no point is joined to any point other than its predecessor or successor.
- (c) no point can be joined to more than one point. So, if  $p$  is joined to its successor  $q$ , then neither can  $p$  be joined to its predecessor, nor can  $q$  be joined to its successor.

Under these assumptions, we ask 'what is the probability that a point remains isolated?'

### 2.3 Theoretical analysis

We proceed as follows: first consider a finite sequence of  $n$  points and consider all possible configurations of these, called  $n$ -configurations, satisfying the above three requirements. Count the total number of isolated points in each of these configurations and compute the average number of isolated points in the  $n$ -configurations. Divide this number by  $n$  and take the limit as  $n$  tends to  $\infty$ . The answer gives the probability of the occurrence of isolated points in an infinite sequence.

2.3a *Some notations:* For each positive integer  $n$ , we let

$$G_n = \text{the number of } n\text{-configurations}$$

and

$$g_n = \text{the total number of isolated points in all these } n\text{-configurations put together.}$$

Then the precise problem on hand is to compute  $\lim_{n \rightarrow \infty} (1/n) \cdot (g_n/G_n)$  if it exists.

2.3b *Values of  $G_n$  and  $g_n$ :* For small values of  $n$ , the values of  $G_n$  and  $g_n$  can be directly computed. We list these values for our later use:

$$\left. \begin{array}{l} G_1 = 1 \quad g_1 = 1 \\ G_2 = 1 \quad g_2 = 0 \\ G_3 = 2 \quad g_3 = 2 \\ G_4 = 2 \quad g_4 = 2 \\ G_5 = 3 \quad g_5 = 3 \\ G_6 = 4 \quad g_6 = 6 \\ G_7 = 5 \quad g_7 = 7. \end{array} \right\} \quad (1)$$

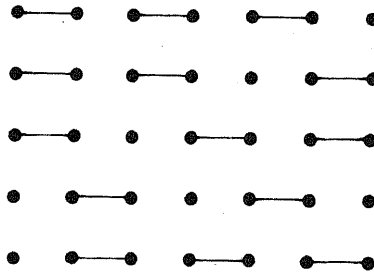


Figure 3. Five 7-configurations arranged in columns.

By way of explanation, we describe the details in one particular case, namely when  $n = 7$ . Figure 3 shows all possible 7-configurations.

Each column in figure 3 is a 7-configuration and there are five such. Thus  $G_7 = 5$ . If we count the total number of isolated points in all these, it is 7. Thus  $g_7 = 7$ .

2.3c *A recurrence relation for  $G_n$* : Consider the sequence of numbers  $G_1, G_2, \dots, G_n, \dots$ . If a formula for the  $n$ th term  $G_n$  in terms of the previous terms, holds for all values of  $n$ , it is called a recurrence relation satisfied by this sequence. Presently we obtain one such relation. Let us assume  $n \geq 3$ .

Among all  $n$ -configurations, there are two types. These are: (1) those that start with an isolated point, and (2) those that start with a non-isolated point.

If an  $n$ -configuration starts with an isolated point, the second and third points are necessarily joined to each other, and the remaining points form an arbitrary  $(n-3)$ -configuration. Therefore there are exactly as many  $n$ -configurations of the first type as there are  $(n-3)$ -configurations of both types. The latter number is  $G_{n-3}$ .

If an  $n$ -configuration starts with a nonisolated point, then the first two points are joined to each other and what remains is an arbitrary  $(n-2)$ -configuration. Thus there are exactly as many  $n$ -configurations of the second type as there are  $(n-2)$ -configurations of both types. This latter number is  $G_{n-2}$ .

Putting these together, we obtain the formula,

$$G_n = G_{n-2} + G_{n-3}, \text{ for all } n \geq 3. \quad (2)$$

2.3d *A recurrence relation for  $g_n$* : If an  $n$ -configuration is of the first type, then it has one isolated point more than its corresponding  $(n-3)$ -configuration. Such  $n$ -configurations contribute  $g_{n-3} + G_{n-3}$  isolated points, because, as we have already seen,  $G_{n-3}$  is the number of  $n$ -configurations of the first type.

If an  $n$ -configuration is of the second type, then it has exactly as many isolated points as its corresponding  $(n-2)$ -configuration. Therefore such  $n$ -configurations contribute  $g_{n-2}$  isolated points.

Taking these together, we obtain

$$g_n = g_{n-2} + g_{n-3} + G_{n-3}, \text{ for all } n \geq 3. \quad (3)$$

2.3e *Further values of  $G_n$  and  $g_n$* : Using these recurrence relations (2) and (3), not only can we check the validity of some values listed in (1), but can also compute the same more quickly for higher values of  $n$ . In fact, once it is given that  $G_1 = G_2 = g_1 = 1$  and  $g_2 = 0$ , the values of all other  $G_n$  and  $g_n$  are uniquely determined by (2) and (3). We require these values for  $n = 12$  and  $n = 13$  in our later proof. Since  $G_{13}$  and  $g_{13}$  can be

Table 1.

$n$	$G_n$	$g_n$	$g_n/nG_n$
5	3	3	1/5
6	4	6	1/4
7	5	7	1/5
8	7	12	3/14
9	9	17	17/81
10	12	24	1/5
11	16	36	9/44
12	21	50	50/273
13	28	72	9/49
14	37	102	34/185

computed only recursively, we have tabulated these values below upto the required stage.

Our problem is to find the limit of the sequence that appears in the last column of table 1. It is heartening to observe that each term here lies between 1/4 and 1/6. It will be shown that, eventually it lies between 1/5 and 1/6.

2.3f *An inequality for  $G_n$* : We use the recurrence relations (2) and (3) now to prove that

$$G_{n-1} \leq G_n \leq (3/2)G_{n-1} \quad (4)$$

for all  $n \geq 4$ . Suppose this is true for  $n = m - 1$ ,  $n = m - 2$  and  $n = m - 3$ . Then this is true for  $n = m$  also, because

$$\begin{aligned} G_m &= G_{m-2} + G_{m-3} \text{ by (2)} \\ &\geq G_{m-3} + G_{m-4} \text{ because (4) is true for } n = m - 2 \text{ and } n = m - 3 \\ &= G_{m-1} \text{ by (2),} \end{aligned}$$

and further,

$$\begin{aligned} G_m &= G_{m-2} + G_{m-3} \text{ by (2)} \\ &\leq (3/2)(G_{m-3} + G_{m-4}), \text{ for (4) is true for } n = m - 2 \text{ and } n = m - 3, \\ &= (3/2)G_{m-1} \text{ by (2).} \end{aligned}$$

A look at (1) and table 1 shows that (4) is true for  $n = 4, 5, 6$ . Therefore, by the principle of induction (4) is true for all  $n \geq 4$ .

2.3g *At most 1/5*: We claim that for large  $n$ , the value of  $g_n/(nG_n)$  is necessarily  $\leq 1/5$ . Suppose for some  $n$  we have,

$$(g_{n-2})/((n-2)G_{n-2}) \leq 1/5,$$

and

$$(g_{n-3})/((n-3)G_{n-3}) \leq 1/5. \quad (5)$$

Then

$$5g_{n-2} \leq (n-2)G_{n-2},$$

and

$$5g_{n-3} \leq (n-3) G_{n-3}. \quad (5')$$

Now

$$\begin{aligned} 5g_n &= 5(g_{n-2} + g_{n-3} + G_{n-3}), \text{ by (3),} \\ &\leq (n-2) G_{n-2} + (n-3) G_{n-3} + 5G_{n-3}, \text{ by (5'),} \\ &= n(G_{n-2} + G_{n-3}) + 2G_{n-3} - 2G_{n-2} \\ &\leq n(G_{n-2} + G_{n-3}), \text{ because } G_{n-3} \leq G_{n-2} \text{ by (4),} \\ &= nG_n \text{ by (2),} \end{aligned}$$

and therefore,

$$g_n/(nG_n) \leq 1/5. \quad (6)$$

Thus (5) implies (6) if  $n \geq 6$ . We check from table 1 that (6) holds for  $n = 12, 13$  and 14. It follows by an inductive argument that (6) holds for all  $n \geq 12$ .

2.3h *At least 1/6:* Similarly, if for some  $n$  we have

$$g_{n-2}/((n-2) G_{n-2}) \geq 1/6,$$

and

$$g_{n-3}/((n-3) G_{n-3}) \geq 1/6, \quad (7)$$

then

$$6g_{n-2} \geq (n-2) G_{n-2},$$

and

$$6g_{n-3} \geq (n-3) G_{n-3}. \quad (7')$$

and therefore

$$\begin{aligned} 6g_n &= 6(g_{n-2} + g_{n-3} + G_{n-3}), \text{ by (3),} \\ &\geq (n-2) G_{n-2} + (n-3) G_{n-3} + 6G_{n-3} \text{ by (7'),} \\ &= n(G_{n-2} + G_{n-3}) + 3G_{n-3} - 2G_{n-2} \\ &\geq n(G_{n-2} + G_{n-3}), \text{ because } 3G_{n-3} \geq 2G_{n-2} \text{ by (4),} \\ &= nG_n \text{ by (2),} \end{aligned}$$

and therefore,

$$g_n/(nG_n) \geq 1/6. \quad (8)$$

Observing that (8) holds for  $n = 5, 6, 7$ , we obtain from the above that (8) holds for all  $n \geq 5$ .

2.3i *Final result:* Thus for  $n \geq 12$ , we have

$$1/6 \leq g_n/(nG_n) \leq 1/5,$$

and therefore the required probability also lies between these two numbers. In fact,  $1/5$  and  $1/6$  are not sharp bounds for the above proof to work. One can prove that if for some  $n$  we have

$$\max(G_n/G_{n-1}, G_{n-1}/G_{n-2}, G_{n-2}/G_{n-3}) \leq (\alpha-3)/2$$

and

$$\min \{g_m/(mG_m): m = n-1, n-2, n-3\} \geq 1/\alpha$$

then

$$g_m/(mG_m) \geq 1/\alpha \text{ for all } m \geq n.$$

A dual result gives an upper bound for  $g_m/(mG_m)$ . An expanded tabulation upto  $n = 19$ , can now be used to prove that  $43/243 \leq (g_n/(nG_n)) \leq 34/185$  for all  $n \geq 19$ . This means that the required probability lies between 17.69 and 18.38%. Thus, when rounded off to the nearest integer, the answer is 18%.

The probability of occurrence of non-isolated points therefore lies between 81.72% and 82.31%.

#### 2.4 Proof that $\lim_{n \rightarrow \infty} (1/n)(g_n/G_n)$ exists and its value

The proof of the existence of the limit and its calculation involve heavier mathematics, which is outlined below.

Consider the power series

$$G(x) = 1 + G_1x + G_2x^2 + G_3x^3 + \dots + G_nx^n + \dots$$

Using the recurrence relation (2), one obtains

$$G(x)(1 - x^2 - x^3) = 1 + x,$$

$$G(x) = (1 + x)/(1 - x^2 - x^3).$$

Differentiating this power series term by term and then multiplying throughout by  $x$ , we get

$$G_1x + 2G_2x^2 + \dots + nG_nx^n + \dots = (x + 2x^2 + 4x^3 + 2x^4)/(1 - x^2 - x^3)^2.$$

In a similar way, one obtains from (3)

$$g(x) = g_1x + g_2x^2 + \dots + g_nx^n + \dots = x/(1 - x^2 - x^3)^2.$$

Let

$$1 - x^2 - x^3 = (1 - \alpha x)(1 - \beta x)(1 - \bar{\beta}x). \quad (9)$$

Then these two expressions can be resolved into partial fractions,

$$\left. \begin{aligned} x/(1 - x^2 - x^3)^2 &= A_1/(1 - \alpha x) + A_2/(1 - \alpha x)^2 + A_3/(1 - \beta x) + A_4/(1 - \beta x)^2 \\ &\quad + A_5/(1 - \bar{\beta}x) + A_6/(1 - \bar{\beta}x)^2, \\ (x + 2x^2 + 4x^3 + 2x^4)/(1 - x^2 - x^3)^2 &= B_1/(1 - \alpha x) + B_2/(1 - \alpha x)^2 \\ &\quad + B_3/(1 - \beta x) + B_4/(1 - \beta x)^2 \\ &\quad + B_5/(1 - \bar{\beta}x) + B_6/(1 - \bar{\beta}x)^2. \end{aligned} \right\} \quad (10)$$

Each term on the right can be expanded as a power series. The coefficient of  $x^n$  is given totally by

$$A_1 \alpha^n + A_2 (n+1) \alpha^n + A_3 \beta^n + A_4 (n+1) \beta^n + A_5 (\bar{\beta})^n + A_6 (n+1) (\bar{\beta})^n$$

in the first, and by a similar expression in the second. Therefore

$$\frac{g_n}{nG_n} = \frac{A_1 \alpha^n + A_2 (n+1) \alpha^n + A_3 \beta^n + A_4 (n+1) \beta^n + A_5 (\bar{\beta})^n + A_6 (n+1) (\bar{\beta})^n}{B_1 \alpha^n + B_2 (n+1) \alpha^n + B_3 \beta^n + B_4 (n+1) \beta^n + B_5 (\bar{\beta})^n + B_6 (n+1) (\bar{\beta})^n}.$$

Now  $1/\alpha$  is the unique real root of  $1 - x^2 - x^3$ . Since  $3/4$  is the approximate real root,  $\alpha$  is approximately  $4/3$ . By comparing the coefficients in (9),  $\alpha\beta\bar{\beta} = 1$ . Therefore  $|\beta| < 1$ . So  $|\beta|^n/|\alpha|^n$  is very small if  $n$  is large. When we divide both the numerator and the denominator of the above expression for  $g_n/(nG_n)$  by  $\alpha^n$ , we find therefore that the limit required is equal to the limit of  $[A_1 + A_2(n+1)]/[B_1 + B_2(n+1)]$  which is easily seen to be equal to  $A_2/B_2$ .

We next note from (10) that

$$A_2 = \lim_{x \rightarrow 1/\alpha} (1 - \alpha x)^2 x / (1 - x^2 - x^3)^2 \text{ and}$$

$$B_2 = \lim_{x \rightarrow 1/\alpha} (1 - \alpha x)^2 (x + 2x^2 + 4x^3 + 2x^4) / (1 - x^2 - x^3)^2.$$

Therefore

$$A_2/B_2 = \lim_{x \rightarrow 1/\alpha} x / (x + 2x^2 + 4x^3 + 2x^4)$$

$$= \alpha^3 / (\alpha^3 + 2\alpha^2 + 4\alpha + 2), \text{ and because } \alpha^3 = \alpha + 1,$$

$$A_2/B_2 = (\alpha + 1) / (2\alpha^2 + 5\alpha + 3) = (\alpha + 1) / (2\alpha + 3) (\alpha + 1)$$

$$= 1 / (2\alpha + 3).$$

Thus the actual value of the limit is  $1/(2\alpha + 3)$  where  $1/\alpha$  is the unique real root of  $1 - x^2 - x^3$  (since  $\alpha$  is approximately  $1.33$ , this limit is approximately equal to  $0.177$ , i.e.  $17.7\%$ ).

### 3. Experimental deviations from the theoretical value

#### 3.1 The role of thermal defects

Our results show that under the idealised conditions assumed by us, the maximum yield of  $\beta$ -truxinic dimer from solid state photoirradiated  $\beta$ -cinnamic acid is  $82\%$ . In practice, experimental yields vary from as low as  $60\%$  to as high as  $90\%$  for different  $\beta$ -acids.

Deviations from the idealised value occur because the assumptions in §2.1 need not strictly hold in all cases. While assumption (b) is self-evident in any non-defect controlled reaction such as the cinnamic acid dimerisation, assumptions (a), (c) and (d) may be contravened in several instances.

It has been observed that the crystallographic short axes for  $\beta$ -cinnamic acids vary between  $3.85$  and  $4.15$  Å. While any value in this range of translational distances is sufficient to permit initial photoreaction, any such reaction will be accompanied by a reduction in the distance between reacting molecules from  $\approx 4.0$  Å to  $\approx 1.6$  Å. This reduction must result in a concomitant relaxation of the neighbouring unreacted molecules so that the 'new' distance between incipient reactive molecules is now greater than the short axis. For crystals with short axes of  $4.0$  Å or less, such a relaxation probably does not affect photoreactivity. However, if the short axis is greater than  $4.0$  Å, molecular relaxation that follows initial reaction could place molecules beyond the photoreactive 'threshold' of  $4.2$  Å. Recently we have studied a case (Sarma and Desiraju 1984) where such a phenomenon, involving violation of assumption (a), probably occurs. We have observed that while 2,4-chlorocinnamic acid ( $\beta$ -structure, short axis



3.88 Å) has a maximum dimer yield of 90%, 6-chloro-3,4-methylenedioxcinnamic acid ( $\beta$ -structure, short axis 4.10 Å) will react to the extent of 70% dimer only.

Yields of dimer in  $\beta$ -crystals greater than 82% could arise due to autocatalysis, preferred surface reactions or to monomer crystallisation in the later stages of the reaction, in other words, in cases where assumptions 3 and 4 are not valid. However, this is not a situation often encountered and even Schmidt's definitive work in this area (Schmidt *et al* 1964) reports only three  $\beta$ -acids out of 14 where the maximum yields are significantly greater than 82%.

While it is generally true that maximum yields in  $\alpha$ -cinnamic acid solid state photodimerisations are very high, deviations from the theoretical value of 100% may reflect the presence of thermal defects in the crystal. Such defects have been invoked in a recent study on the kinetics of topochemical process (Hasegawa and Shiba 1982) where it has been commented that the rates of these reactions decrease above an optimal temperature ( $T_{opt}$ ). For  $\alpha$ -cinnamic acid,  $T_{opt}$  is 20°C and this means that the mismatch between potentially reactive double bonds owing to thermal fluctuations may inhibit solid state photoreactivity to some extent at temperatures above 20°C. Hasegawa and Shiba's study reports reaction rates as a function of temperature but an entirely analogous set of results would probably be obtained if one were to measure maximum dimer yields in an  $\alpha$ -crystal as a function of temperature. While the number of thermal defects could be obtained analytically, a corroborative number may probably be obtained experimentally from the maximum yield measurements.

Hasegawa and Shiba (1982) state that  $T_{opt}$  is rather high for hydrogen bonded crystals. It is possible that defects are better trapped in such crystals with a reduced probability of the crystal relaxing to an incipient reactive geometry. In complementary results (which demonstrate the reduced importance of defects for non-polar crystals) it has been observed that for  $\alpha$ -symmetry compounds such as 2-benzyl-5-benzylidenecyclopentanone or  $\alpha$ -benzylidene- $\gamma$ -butyrolactone, the conversion to dimer is quantitative (Jones *et al* 1980; Kearsley and Desiraju 1985). Interestingly, in the crystal of 3,4-dimethoxycinnamic acid, where only half the molecules are in a potentially reactive centrosymmetric ( $\alpha$ ) arrangement, the maximum dimer yield is 44% (Desiraju *et al* 1984). That this number is less than 50% is possibly because of the importance of thermal defects, and almost certainly means that dimer and monomer molecular phases do not separate with the progress of the reaction. This confirms the validity of assumption (d) in such reactions.

### 3.2 Mathematical modelling of defect reactions

The analytical treatment in §2.3 follows from assumptions concerning an equal probability of photoreaction throughout the crystal. While this is more-or-less true for defect-free regions in the bulk, the situation may be quite different in the crystal boundaries.

Let us now consider the situation where the points near the boundary of each stack behave differently from the points in the interior, as stated below: there is a number  $m$  such that  $m$  points in each of the two ends are sure to react, and cannot remain isolated.

Then for a given  $n$ , not all  $n$ -configurations are to be considered, but only those in which the above condition is met. Correspondingly, we have to compute new values for  $G_n$  and  $g_n$  lower than their original values. Let us denote these new values by  $\tilde{G}_n$  and  $\tilde{g}_n$ . Without any significant loss in generality, one may assume that  $m$  and  $n$  are even.

It is observed that an  $n$ -configuration is eligible in our new counting exactly when (i) none of the  $m$  points in the upper end and  $m$  points in the lower end remain isolated and (ii) the remaining  $n-2m$  points form a configuration that is eligible according to the earlier sense. Consequently, we have

$$\tilde{G}_n = G_{n-2m},$$

and

$$\tilde{g}_n = g_{n-2m},$$

whenever  $n \geq 2m$ . (These quantities become 0 when  $n < 2m$ , but in our problem,  $n$  far exceeds  $2m$ ). Therefore the number to be determined is  $(1/n) \cdot (\tilde{g}_n/\tilde{G}_n) = (1/n) \cdot (g_{n-2m}/G_{n-2m})$ . This lies between  $1/6$  and  $1/5$  whenever  $n \geq 12 + 2m$ . If  $m$  is a fixed quantity, then there is no change in the value of the limit.

$$\lim_{n \rightarrow \infty} (1/n) (\tilde{g}_n/\tilde{G}_n) = \lim_{n \rightarrow \infty} (1/n) (g_n/G_n).$$

If on the other hand  $m$  is not fixed, but only the ratio  $m/n$  is a fixed small quantity  $k$ , then because of the equality

$$(1/n) \cdot (\tilde{g}_n/\tilde{G}_n) = (1/n) \cdot (g_{n-2m}/G_{n-2m}) = (1-2k) (1/n-2m) (g_{n-2m}/G_{n-2m}),$$

the new value is  $(1-2k)$  times the old value. Since  $k$  is usually small, the value is almost the same.

Let us illustrate the same in a few particular cases: When  $n$  is approximately  $10^{20}$  and  $m$  is a fixed quantity (say, not more than  $10^5$ ), then the maximum yield is the same, viz, 82.3%. When  $n$  is approximately  $10^{20}$ , and  $m$  is  $10^{-10}n$ , then the maximum yield lies between 82.3% and a quantity slightly more than this. Because this quantity coincides with 82.3 in the first few significant decimal places, one concludes that here again the value is virtually unchanged.

### 3.3 Comparison with other related work

The problem of how to determine the number of isolated single points when adjacent pairs are selected at random from a line of points has been of interest in a number of seemingly unrelated chemical studies and much has been written on this subject. Flory (1939) has used this mathematical concept in obtaining the fraction of functional substituents of vinyl polymers which may undergo reaction in pairs while Jackson and Montroll (1958) discuss the average concentrations of trapped free radicals condensed from gaseous mixtures onto surfaces cooled to liquid helium temperatures. Flory's (1939) model has been commented upon (Cohen and Reiss 1963) and other extensions have appeared (Littlewood and Verrall 1973). A more detailed study using difference equations has been carried out (Page 1959). We wish to compare our study with these methods.

In the work of Flory (1939) and Page (1959), a mathematical modelling different from ours has been considered. These studies take into account the extra factor of 'the sequence in which the reactions among the molecules take place'. In our treatment, on the other hand, this order is immaterial. Since the present understanding of the factors which control the initiation and progress of organic solid state reactions is not well-developed, it is difficult to decide which of these two situations is the more appropriate.

In either case, the observed experimental values of the extent of reaction in different,  $\beta$ -cinnamic acids show considerable variation from the calculated values of 13% and 18% respectively of unreacted starting material.

Not only are the final answers different in the two cases, but the differences crop up for even small values of  $n$ . As a representative example, our value of  $g_5/G_5$  is 6/4, but the value of the corresponding quantity  $S_5$  is 16/15 according to Flory. There are two reasons for such a difference. First, as stated above, while counting the  $n$ -configurations, the sequence of reactions is taken into account. A configuration such as  $\text{---} \text{---} \text{---}$  would have been counted only once in  $G_5$  but twice in  $S_5$  since the initial relation could be either between the first and second points ( $\overset{1}{\cdot} \text{---} \overset{2}{\cdot} \text{---}$ ) or between the third and fourth points ( $\overset{2}{\cdot} \text{---} \overset{1}{\cdot} \text{---}$ ). So if the sequence of joining the points is important, there are more ways in which the configurations may be achieved. Secondly, Flory breaks up this aggregate of  $n$ -configurations into  $n$  parts, computes the average number of these isolated points separately in each of these parts, and finally takes the average of these averages to obtain  $S_n$ . Since the  $n$  parts are not of equal size, his  $S_n$  is different from the actual average of isolated points in these configurations.

Thus, because of a different assumption, and a different way of computing the averages, Flory's calculations are simpler than ours. In the paper by Page (1959) also, the first equation  $p_{i,n} = p_i p_{n-i}$  is incorrect, as one can actually verify that  $p_{3,5} = 1/6$  and  $p_2 = 0$ . For a similar reason as the one stated in the previous paragraph, his formula

$$p_n = (p_1 + \dots + p_{n-2})/(n-1)$$

is not correct. In fact, direct computations yield that

$$p_1 = 1; p_2 = 0; p_3 = 1/2; p_4 = 1/3; \text{ and } p_5 = 1/3,$$

where  $p_n$  is the probability that the left end point of an  $n$ -configuration is vacant. Here it is seen that  $p_5$  is not equal to  $(p_1 + p_2 + p_3)/4$ . These observations necessitate a thorough calculation, similar to the one we have carried out in §§2.3 and 2.4.

The work of Jackson and Montroll (1958) (which we became aware of at a late stage) is closer to ours because it starts with the same set of mathematical assumptions and arrives at the same result. But three facts deserve mention here. First, the two approaches agree only in the initial step of framing the recurrence relations; later they are entirely different. Secondly our arguments in §2.3 are conceptually very simple unlike those of Jackson and Montroll. Even without solving the recurrence relations we have been able to decisively conclude, merely by means of simple inequalities, that the required probability lies between 81.72 and 82.31%. Further our method shows easily that for all values of  $n \geq 19$ , the required probability deviates from the limiting value by less than 0.6%. Finally, even while calculating the exact limit in §2.4, the two methods are quite distinctive. Our answer is in terms of the unique real root of  $1 - x^2 - x^3$  whereas the result in the paper of Jackson and Montroll is in terms of the largest root of  $x^3 - x - 1$ .

#### 4. Conclusion

It has been shown that the maximum possible theoretical conversion of  $\beta$ -cinnamic acid crystals to  $\beta$ -truxinic acid through a solid state topochemical photodimerisation process is around 82%. The recurrence relations that have been used to obtain this

result may also be employed to describe the situations where preferential reaction occurs near the crystal surfaces.

#### **Acknowledgements**

We thank Prof. M Venkataraman for some results and one of the referees for helpful remarks. We thank Prof. G Wegner, Max-Planck Institute for Polymer Studies, Mainz, for drawing our attention to Flory's paper. One of us (GRD) thanks the DST (SERC) and the UGC, New Delhi, for financial support.

#### **References**

- Cohen E R and Reiss H 1963 *J. Chem. Phys.* **38** 680  
Desiraju G R, Kamala R, Kumari B H and Sarma J A R P 1984 *J. Chem. Soc. Perkin Trans. II* 181  
Flory P J 1939 *J. Amer. Chem. Soc.* **61** 1518  
Hasegawa M and Shiba S 1982 *J. Phys. Chem.* **86** 1490  
Jackson J L and Montroll E W 1958 *J. Chem. Phys.* **28** 1101  
Jones W, Nakanishi H, Theocharis C R and Thomas J M 1980 *J. Chem. Soc., Chem. Commun.* 610  
Kearsley S K and Desiraju G R 1985 *Proc. R. Soc. London* **A397** 157  
Littlewood B and Verrall J L 1973 *J. Chem. Phys.* **59** 1613  
Page E S 1959 *J. R. Stat. Soc.* **B21** 364  
Schmidt G M J, Cohen M D and Sonntag F I 1964 *J. Chem. Soc.* 1966  
Sarma J A R P and Desiraju G R 1984 *J. Chem. Soc., Chem. Commun.* 145