Phytochemistry, 1970, Vol. 9, pp. 1049 to 1063. Pergamon Press. Printed in England.

QUALITATIVE AND QUANTITATIVE CO-OCCURRENCE OF TERPENOIDS AS A TOOL FOR ELUCIDATION OF THEIR BIOSYNTHESES*

EUGENE ZAVARIN†

University of California Forest Products Laboratory, Richmond, California, U.S.A.

(*Received* 22 *July* 1969, *in revised form* 11 *October* 1969)

Abstract—A mathematical and conceptual framework has been developed for the utilization of qualitative and quantitative co-occurrence of natural compounds in establishing biosynthetic hypotheses. The methods were applied to a series of cases of co-occurrence of monoterpenoids, and the results were utilized for substantiation and clarification of our ideas on the biosyntheses of α- and β-pinene, β-phellandrene, camphene, tricyclene, bornyl acetate, γ-terpinene, α-thujene, sabinene, terpinolene, 3-carene, limonene and myrcene.

INTRODUCTION

IN OUR work on xylem and cortex volatile oils from *Pinus* and *Abies*, we often encountered regularities in the occurrence of compounds constituting these oils. Some regularities were qualitative, i.e. two or more compounds had a tendency to occur together, while some were quantitative, i.e. compounds tended to occur in quantities having a mathematical relationship to each other. Although, as discussed later, other factors could be involved in some instances, the linkage of biosynthetic pathways is probably primarily responsible for these regularities.

The biosynthesis of monoterpenes has been reviewed several times.[1] The generally accepted idea of biosynthesis involves transformation of mevalonate ion into isopentenyl- and dimethylallyl pyrophosphates, with the two combining to form a molecule of geranyl- (or neryl-) pyrophosphate. Not much reliable information is available on transformations leading from geranyl- or nerylpyrophosphate to individual monoterpenes, and one has to be satisfied with a model (Fig. 1) which presupposes formation of a carbonium ion intermediate followed by appropriate transformations of this entity. This model, advanced by Ruzicka and later expanded by others,[1,2] has been found very useful for predicting possible chemical structures for monoterpenoids, although in many respects it represents an oversimplification. Thus, the processes in question are most likely concerted in nature. Furthermore, the model ignores the enzymatic involvement, which should constitute a powerful factor in terpene differentiation processes.[3,4]

* Presented in part at the Phytochemical Society of North America Meeting, Tucson, Arizona, 6–8 June, 1968.

† Forest Products Chemist.

[1] J. H. RICHARDS and J. B. HENDRICKSON, *The Biosynthesis of Steroids, Terpenes and Acetogenins*, p. 207, W. A. Benjamin, New York (1964). J. R. HANSON, *Perf. and Essen. Oil. Rec.* **58**, 787 (1967). G. WEISSMANN, "The distribution of terpenoids", in *Comparative Phytochemistry* (edited by T. SWAIN), p. 97, Academic Press, New York (1966).

[2] L. RUZICKA, *Experientia* **9**, 357 (1953).

[3] R. M. GASCOIGNE, *J. Chem. Soc.* 879 (1958).

[4] R. ROBINSON, in *Biosynthesis of Terpenes and Steroids* (edited by G. E. W. WOLSTENHOLME and M. O'CONNOR) p. 229, Little, Brown & Co., Boston (1959).

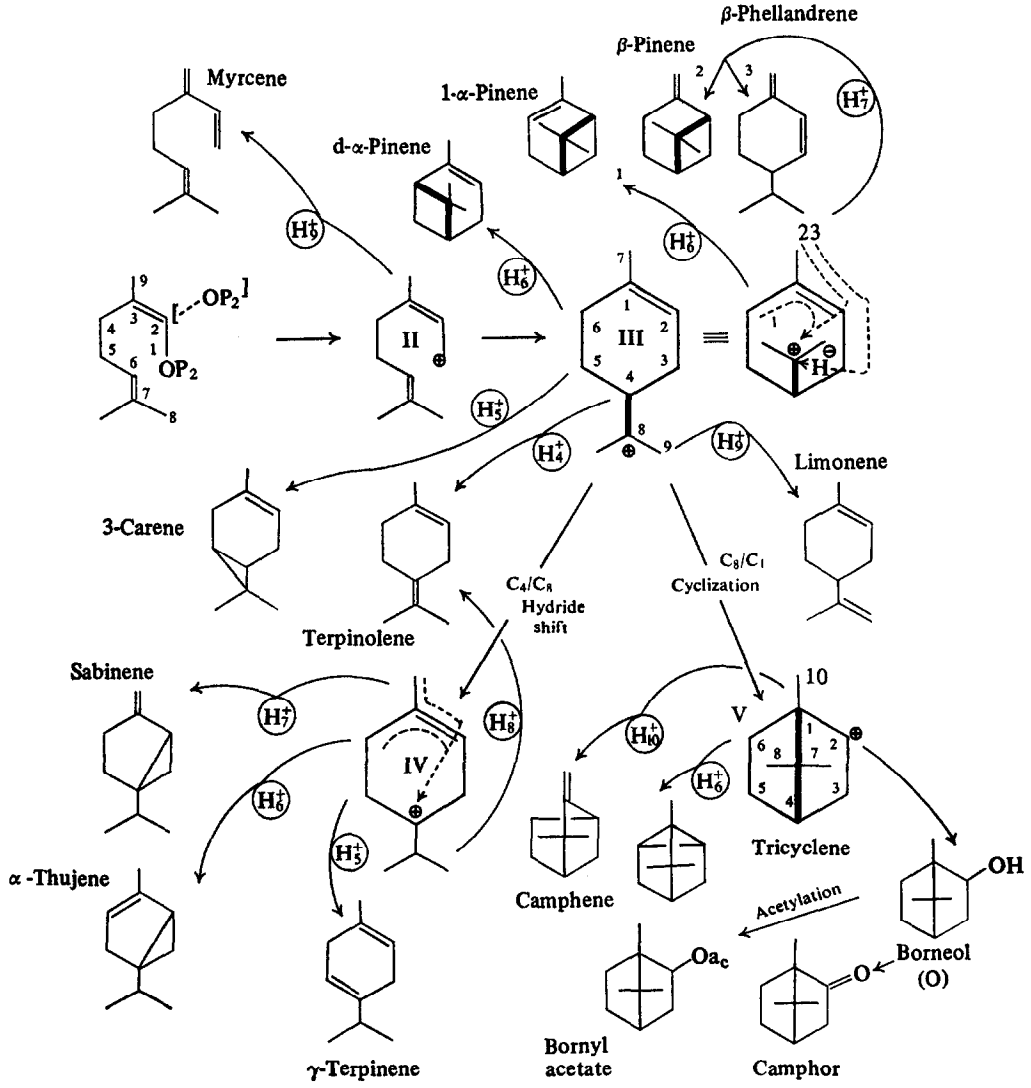

FIG. 1. PROBABLE PATHWAYS OF MONOTERPENOID BIOSYNTHESIS.

However, even within the framework of the carbonium ion model, we still need experimental data associating individual monoterpenes with specific biogenetic routes, since most of the tracer studies available are concerned with merely demonstrating the incorporation of mevalonate or acetate in monoterpenes. Where more detailed experimental results are on hand, like those of Sanderman,[5,6] Banthorpe and their co-workers,[7,8] the data are occasionally contradictory and additional efforts are desirable. In a few cases, radioactive tracer

[5] W. SANDERMANN, *Holzforschung* **16**, 45 (1962).
[6] W. SANDERMANN and W. SCHWEERS, *Tetrahedron Letters* 257, 259 (1962).
[7] D. V. BANTHORPE and K. W. TURNBULL, *Chem. Commun.* 177 (1966).
[8] D. V. BANTHORPE and D. BAXENDALE, *Chem. Commun.* 153 (1965).

experiments can yield only ambiguous results and, therefore, basically different approaches are required.

In the present paper, we report on cases of qualitative and quantitative co-occurrence of monoterpenoids, and develop a conceptual and mathematical framework for utilizing these and similar results in elucidation of biosynthetic mechanisms. Although the conclusions arrived at are mostly hypothetical, the usefulness of this inferential methodology seems justified by the inapplicability (in some cases) of more direct methods as well as by providing "the basis for direct experiment as well as affording evidence for the *generality* of a scheme that may be thoroughly tested in only a few species", as argued by Richards and Hendrickson for the inferences based on the comparison of structures.[9]

Monoterpenoids seem a particularly appropriate subject for a study of the principles of co-occurrence, as the compounds are relatively few, their transformations are simpler and better understood than those of higher terpenoids, and they are more amenable to quantitative analysis by GLC. The methodology arrived at can be later applied to other classes of compounds.

RESULTS AND DISCUSSION

Qualitative Co-occurrence

The qualitative co-occurrence, usually designated simply as "co-occurrence", has been used in a large number of instances to intimate a close biosynthetic relationship between two or more natural compounds. Thus, co-occurrence of caryophyllene and humulene in *Eugenia caryophyllata* has been cited as evidence for their biogenesis from the same cation by Richards and Hendrickson.[10] Erdtman and Norin use the argument of co-occurrence to imply possible formation of thujopsane, widdrane, cuparane, and cedrane type sesquiterpenoids from the same *cis*-farnesyl type precursor.[11] Parker, Roberts and Ramage use co-occurrence for hypothesizing about the biosynthesis of pseudoivalin.[12] Nayak and Dev assumed from the co-occurrence of longifolene and longicyclene that they were formed from the same carbonium ion.[13] Many other examples could be quoted.

The argument is statistical in nature, and rests upon the assumption that if several compounds are found together much more often than expected on the basis of pure chance they are likely to be biogenetically particularly closely related. The argument seems reasonable because with an increase in the number of biogenetic steps separating compounds in question, the chance of genetic and evolutionary change should increase and result in a greater chemical differentiation. It follows that full utilization of the co-occurrence phenomenon should include calculation of the probability, p, for two or more compounds to be found together on the basis of pure chance. Small values for the probability would imply that co-occurrence cannot be satisfactorily explained by chance alone and possibility of biosynthetic linkage would be indicated. The necessary mathematical tools have been worked out by probability mathematicians and the solution is given in so-called hypergeometric distribution functions

[9] J. H. RICHARDS and J. B. HENDRICKSON, *The Biosynthesis of Steroids, Terpenes and Acetogenins*, p. 10, W. A. Benjamin, New York (1964).
[10] J. H. RICHARDS and J. B. HENDRICKSON, *The Biosynthesis of Steroids, Terpenes and Acetogenins*, p. 230, W. A. Benjamin, New York (1964).
[11] H. ERDTMAN and T. NORIN, "The chemistry of the Order Cupressales", in *Fortschritte der Chemie Org. Naturstoffe* (edited by L. ZECHMEISTER), p. 245, Springer-Verlag, New York (1966).
[12] W. PARKER, J. S. ROBERTS and R. RAMAGE, *Quart. Rev. Chem. Soc. (London)* **21**, 350 (1967).
[13] U. R. NAYAK and S. DEV, *Tetrahedron Letters* 243 (1963).

presented in the experimental part of this study.[14] The calculations require knowledge of the total number of sources examined (S), the number of sources where compounds co-occur ($S_{a,b...n}$), and the numbers of sources where each individual compound has been identified (S_a, S_b, . . . S_n).

Probability calculations have not been used much in literature in regard to co-occurrence; most authors, while noting the existence of two or more compounds together in certain sources, tend at the same time to ignore how often they were found singly. Often this cannot be avoided as in many cases the necessary data are not available in sufficient quantity although it obviously involves the danger of a bias favoring or disfavoring biosynthetic linkage.

However, even with calculations above some bias seems unavoidable, and critical utilization of experimental data as well as acceptance of rather low p values only is indicated. Some of the bias arises through the assumption that each compound has in all analyses the same chance of being missed. This is probably true for cases where compounds of interest are present in relatively large amounts, which are difficult to overlook in an analysis; it is much less true, however, for so-called trace constituents particularly if literature data are used. In this case quantitative variability of mixture composition, variability in sensitivity of analytical methods (e.g. lumping analyses by GLC and by classical, preparative methodology) and personal factors could introduce a bias into probability values by preferential identification of several components together. In some cases, particularly if "traces" are separated by a considerable quantitative gap, it seems advantageous to leave out trace constituents and base the calculations on occurrence in major quantities only.

Another source of bias can represent consideration of samples of unequal taxonomic relation to each other. Obviously, the more closely related two plants are the more close their chemistry is likely to be. Thus, a bias would be introduced by lumping together in calculations samples from a series of populations belonging to the same species and samples obtained from single populations of diverse species. This type of bias can be minimized, but not removed, by considering differences between samples of equal taxonomic rank only (e.g. species belonging to the same genus or individuals belonging to the same population); since plant taxa are never equally related to each other, this is only a compromise solution.

In our work on *Abies* cortical oleoresins,[15] association of large values for camphene (10 per cent or higher) with tricyclene (1 per cent or higher) was noted in four firs, *Abies concolor*, *A. grandis*, *A. veitchii* and *A. koreana*, while fourteen firs either lacked these constituents or contained them in lower amounts.* Calculation of the probability of chance co-occurrence gave $p = 3{\cdot}3 \times 10^{-4}$, which is sufficiently small to assume biosynthetic closeness of these compounds.

Another case involved co-occurrence of sabinene, γ-terpinene and α-thujene in xylem oleoresin of *Pinus*. All three were detected in appreciable quantities in *Pinus muricata* (southern variant)—sabinene up to 73·5 per cent (54·5 per cent average), α-thujene up to 11 per cent (2·5 per cent average) and γ-terpinene up to 3·5 per cent (1·0 per cent average).[16,17] Later, sabinene was discovered in amounts larger than 0·5 per cent in *P. monticola*, *P. lamber-*

* Only GLC analyses were considered, which were enlarged by inclusion of *Abies firma* and *A. homolepis* —both containing lower amounts of either compound. Oleoresin of *A. koreana* was re-analyzed and this increased tricyclene value to 2·0%.

[14] W. FELLER, *An Introduction to the Probability Theory and Its Applications*, Vol. I, p. 41, John Wiley and Sons, New York (1958).
[15] E. ZAVARIN and K. SNAJBERK, *Phytochem.* 4, 141 (1965).
[16] M. B. FORDE and M. M. BLIGHT, *New Zealand J. Bot.* 2, 44 (1964).
[17] N. T. MIROV, E. ZAVARIN, K. SNAJBERK and K. COSTELLO, *Phytochem.* 5, 343 (1966).

tiana, P. monophylla,[18] *P. flexilis,*[18,19] *P. strobiformis,*[19] *P. coulteri, P. contorta,*[20,21] *P. stankewiczii*[22] and *P. roxburghii,*[23] and γ-terpinene in *P. monticola*[18] and *P. roxburghii.*[23] Taking as 94 the number of pine species investigated, the calculations give $p_{ts} = 9{\cdot}0 \times 10^{-4}$ for the co-occurrence of sabinene and γ-terpinene, $p_{tt} = 3{\cdot}2 \times 10^{-2}$ for the co-occurrence of α-thujene and γ-terpinene, and $p_{stt} = 2{\cdot}9 \times 10^{-5}$ for the co-occurrence of all three compounds. These low values for probability of co-occurrence by chance alone suggest that a case of close linkage of biosynthetic mechanisms is likely to have been encountered again. This deduction is strengthened by co-occurrence of α-thujene, sabinene and γ-terpinene in *Juniperus* foliage.[24–26] The same compounds were also noticed together in *Cupressus* leaf volatile oils.[27]

The analytical data of Mirov on *Pinus* oleoresins[28] allow further computations. However, of the pairs of terpenes tested, only 3-carene/terpinolene pair had a chance probability value low enough to be of any importance—$p_{ct} = 1{\cdot}8 \times 10^{-4}$. The connection of this and previous cases of qualitative co-occurrence to specific biosynthetic sequences will be discussed in connection with quantitative co-occurrence.

Quantitative Co-occurrence

Theoretical. Much less attention has been given to quantitative co-occurrence of natural products, i.e. to mathematical relationships between the amounts in which compounds occur in natural sources. Several cases of positive and negative correlations have been noted by Smith,[29] Peloquin,[30] and Hanover[31] working with oleoresins of *Pinus* and *Pseudotsuga.*

In our work on *Pinus* and *Abies* many quantitative correlations between the amounts of individual monoterpenoids were encountered[32] and some are reproduced in Figs. 3–5. The regression lines obtained were largely linear, fitting the general equation

$$a \pm bx = y \tag{1}$$

Three special cases are of particular significance for our treatment to follow.

(1) $b = 0$, i.e., y is independent of variations in x (Fig. 2, line C, Fig. 9 of Ref. 32).
(2) $a = 0$, i.e., changes in y are proportional to changes in x (Fig. 3 and Fig. 2 line A).
(3) $b/a \leqslant -1$, i.e. one variable can be completely substituted by the other (Fig. 2 line B, and Fig. 13 of Ref. 32).

In some cases (Fig. 3) the clustering of the points was definite, in other cases the deviations were greater and use of regression analysis was desirable. In this case, recourse was made to

[18] E. ZAVARIN, "Evolution of volatile terpenoids", in *Evolutionary Biology* (edited by T. DOBZHANSKY, M. K. HECHT and W. C. STEERE), Vol. IV, Appleton-Century-Crofts, New York; to appear in 1970.
[19] R. H. SMITH, *U.S. For. Serv. Res. Note PSW*-135 (1967).
[20] R. H. SMITH, *Phytochem.* 3, 259 (1969).
[21] R. H. SMITH, *Forest Sci.* 13, 246 (1967).
[22] N. T. MIROV, E. ZAVARIN and K. SNAJBERK, *Phytochem.* 5, 97 (1966).
[23] S. JUVONEN, Papers presented at the 10th Apothecary and Pharmacist Congress, Helsinki, 1964, pp. 302 (1965).
[24] F. M. COUCHMAN and E. VON RUDLOFF, *Can. J. Chem.* 43, 1017 (1965).
[25] A. R. VINUTHA and E. VON RUDLOFF, *Can. J. Chem.* 46, 3743 (1968).
[26] S. B. TEPPEEV and M. I. GORYAEV, *Uch. Zap. Kabardino-Balkar. Univ. Ser. S. Khoz. Khim. Biol.* 29, 371 (1966).
[27] E. ZAVARIN and L. LAWRENCE, to be published.
[28] N. T. MIROV, Composition of Gum Turpentines of Pines, *U.S. Dep. Ag. Tech. Bull.* No. 1239.
[29] R. H. SMITH, *U.S. Forest Service Research Paper, PSW*-15 (1964).
[30] R. L. PELOQUIN, M.S. Thesis, Stanford University, Palo Alto, California (1964).
[31] J. W. HANOVER, *Heredity* 21, 73 (1966).
[32] E. ZAVARIN, K. SNAJBERK, T. REICHERT and E. TSIEN, *Phytochem.* 9, 377 (1970).

computer methodology allowing easy calculation of regression coefficients –*b*– (slopes), regression constants –*a*–, correlation coefficients –*r*–, standard deviations of the fit, and other statistics.

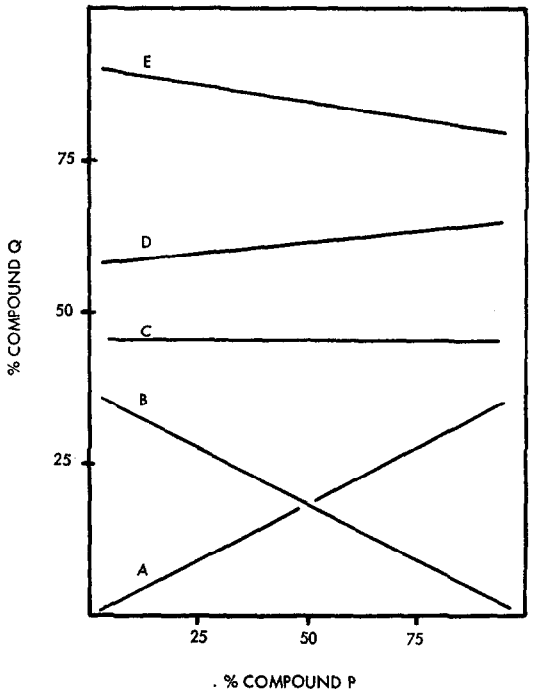

FIG. 2. SEVERAL SPECIAL CASES OF LINEAR REGRESSIONS, CORRELATING QUANTITIES OF MONOTERPENOIDS.

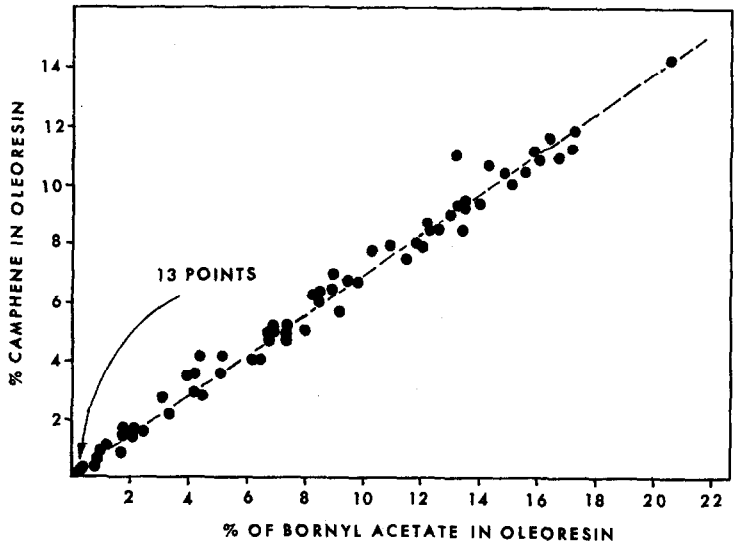

FIG. 3. % CAMPHENE VS. % BORNYL ACETATE PLOT FOR SEVERAL *Abies* SPECIES.
The % camphene vs. % tricyclene plot was practically identical to this.

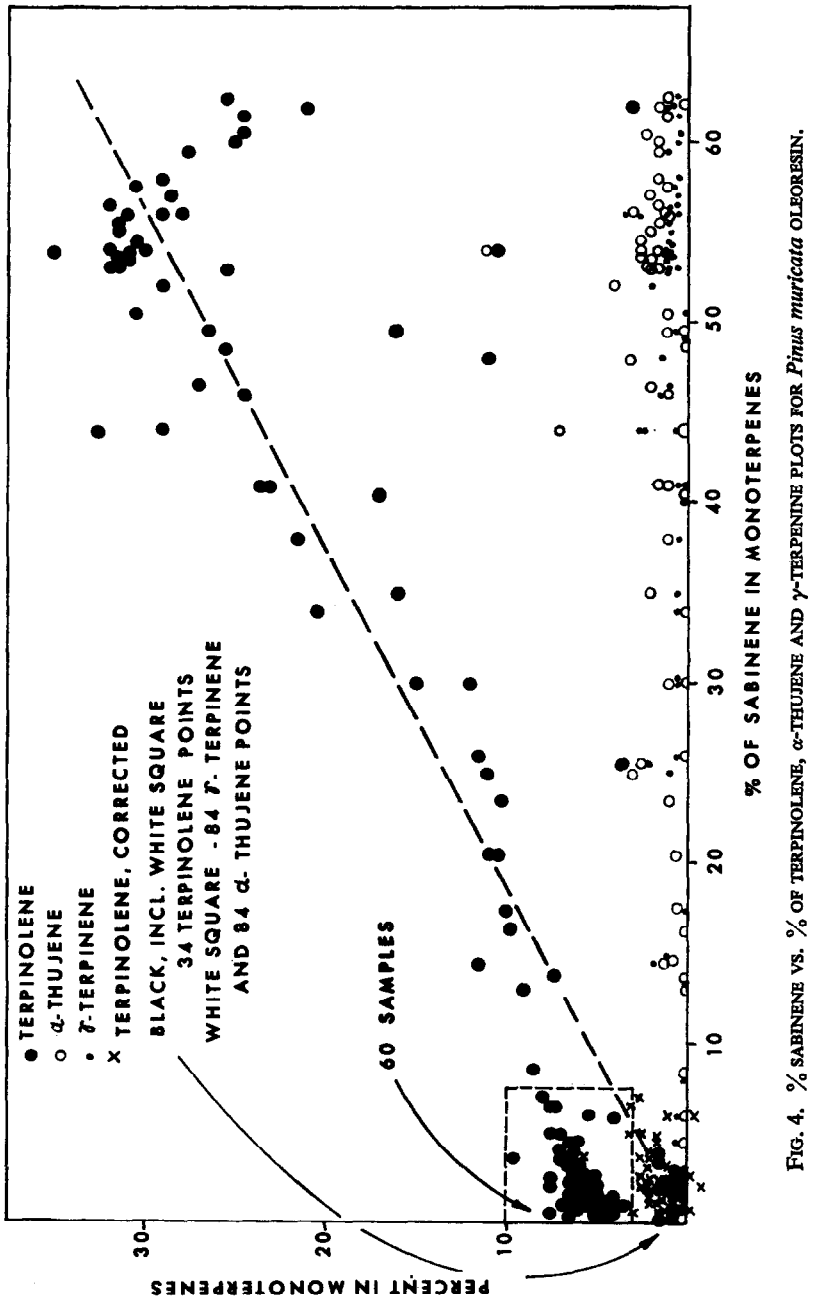

FIG. 4. % SABINENE VS. % OF TERPINOLENE, α-THUJENE AND γ-TERPENINE PLOTS FOR *Pinus muricata* OLEORESIN.

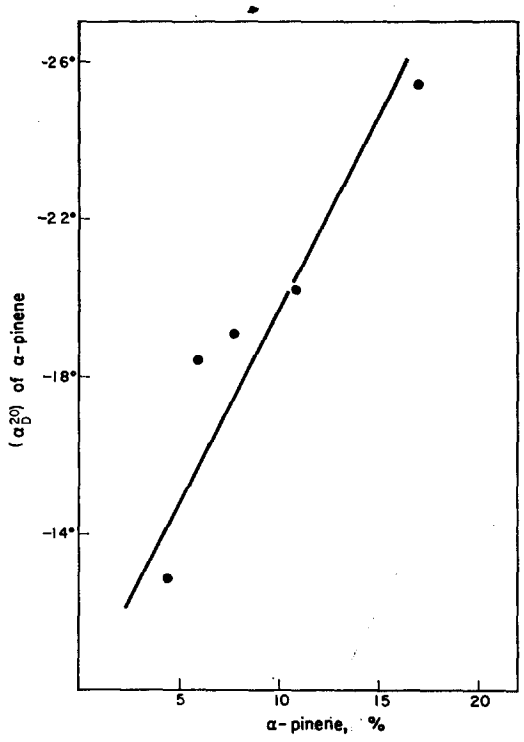

FIG. 5. % α-PINENE VS. (α_d^{20}) PLOT FOR *Pinus ponderosa* OLEORESIN.

Several reasons could be advanced for the existence of these correlations,* namely: (1) linkage of genes; (2) spurious linkage, arising from mathematical methods employed; and (3) linkage of biosynthetic reaction sequences.

Gene linkage is not likely to be involved in correlations between individual terpenoids in work based on random sampling in the field and extending over a wide geographic range, and where absence of strong selective pressures on terpenoid composition can be assumed. With gene exchange spreading over enormous time intervals, crossing-over and related phenomena are likely to result in breaking of existing gene linkages and the randomization of information within the gene pool. These conditions might not be met, however, when sampling in highly localized, isolated areas, or in experiments with artificial hybrids obtained by controlled pollination using only a few parental genotypes. Still another consideration is a terpenoid evolution favoring the linked chromosome structures; in the case of oleoresin, however, there is no evidence for this happening, and we are assuming it to be a minor factor in the following treatment.

Terpene composition is usually expressed in per cent. In the case of oleoresins, the weight of terpenoids or of total oleoresin represent the bases commonly used, while in the case of leaf oils, total oil weight, green or dry leaf weights are employed. This normalization procedure can result in appearance of spurious correlations. The most obvious case is that of a two-

* In the discussion to follow it is assumed that the variability due to environmental and physiological factors is small or has been made small by standardization of the collection procedures. In case of samples collected in several distant locations, geographic variations, too (e.g. chemical clines), could change or obscure the existing relationships. This can be easily demonstrated and the effect made negligible by reduction of the total geographic sampling area.

component mixture, where any variability must be of necessity connected with $a = 100$, $b = -1{\cdot}0$ and $r = -1{\cdot}0$, the correlation being meaningless. Here, it is uncertain whether change in one of the components is being naturally compensated by a change in the second component, or whether the amount of the first component is changing independently and the correlation obtained is purely artificial. In multicomponent systems, this spurious element becomes less important. Any independent increase (or decrease) in the first of the two components becomes compensated by the corresponding decrease (or increase) in all other normalized constituents and affects the second component proportionally less. This suggests a way to test the correlations encountered for the presence of an artificial element. Thus, spurious contribution can probably be regarded as negligible if little change in slope, $-b-$, of the regression lines is obtained after enlarging the percentage basis to include much larger numbers and amounts of components. In the case of oleoresin, total terpene weight and oleoresin weight represent the two convenient test bases. Calculations made using cortical terpene data of *A. lasiocarpa*[32] indicated, in all cases examined, negligible changes of the regression slopes (constant "b"), following the enlargement of per cent base from total terpene to total oleoresin. Although further testing is still desirable, it thus seems that spurious contribution is small and the correlations are real, i.e. that monoterpenoids represent a closely knit family of compounds, with the amounts of most materials varying at the expense of each other within this family.

The probable connection of the above quantitative correlations to biosynthesis has been assumed or hinted by several authors.[30,31] On one occasion[31] an attempt was made to explain them through mechanistic speculations, based solely on the known *in vitro* rearrangements of monoterpenes with inorganic agents. Unfortunately, no attention was given to the biosynthetic ideas of Ruzicka and others, as well as to the isotopic tracer experimental information.

Although it is probable that linkage of biosynthetic mechanisms is by far the most important contributing factor to the existence of these correlations, it is not easy to utilize these correlations for clarification of biosynthetic questions. Thus, positive as well as negative correlations can be connected with either parallel or consecutive arrangement of biosynthetic steps, and associating certain types of correlations with particular step arrangements would yield ambiguous results only. Furthermore, as the carbonium ion picture of terpene biosynthesis is likely to be only a model, this could result in additional difficulties. For these reasons any proposed methods of interpretation must be considered as rules only, whose validity must be carefully ascertained by the agreement of the results of their application with the biosynthetic ideas derived from experiments based on the isotopic labeling, intermediate isolation, structural correlations and other similar independent evidence.

The experimental data available suggests that monoterpenoids represent a group of compounds which are varying in quantity largely at each other's expense. Furthermore, material exchange does not need to affect all terpenoids in the same proportion. Thus, as exemplified for di- and monoterpenes of *A. lasiocarpa* (p. 1061), the quantities of some compounds are not affected by variations in others, i.e. biosynthetic mechanisms are not intimately linked and enzymic systems *independently* control the quantities of the two groups of compounds. In other instances (Fig. 13 of Ref. 32), an organism is able to increase the quantity of one constituent only at the expense of other components (*substitution*)—i.e. the enzymic system is partially restricted in its freedom. In still other cases (Figs. 3–5) the organism is able to produce two or more compounds with quantities only in a *fixed ratio*, i.e. the enzymic system has little control over the relative amounts produced, and the biosynthetic paths are very

closely linked. In this latter case, presupposition of a close common intermediate becomes an attractive hypothesis. Thus, we are faced with a gradation in ability of the enzymic systems to independently control the production of individual terpenoids. If we now assume that this gradation is related to the similarity of biosynthetic mechanisms operating, two rules can be postulated.

(I) Compounds in a substitutional or proportional relationship (Fig. 2, lines A and B) should be biosynthetically closer to each other than to a compound to which they stand in an independent relationship (Fig. 2, line C).

(II) Compounds in a proportional relationship (Fig. 2, line A), should be biosynthetically closer to each other than to compounds to which they stand in substitutional or independent relationship (Fig. 2, lines B and C).

These rules should apply mainly to the three special cases indicated (Fig. 2, lines A–C), although the correlations are often describable in intermediate terms only where the rules are less likely to apply. Thus, with *a* large and *b* small (Fig. 2, lines D and E), it would be difficult to say whether we are closer to independence or to the other two relationships. The concept of "biosynthetic similarity", used above, represents another difficulty. It is a qualitative and relative concept corresponding to the one employed with qualitative co-occurrence. In the field of monoterpenoids it involves the use of terms such as close common intermediates, formation through the same hydride shifts, parallelism in electron flows, and the like.

Application

Camphene, tricyclene, bornyl acetate. All three compounds are often encountered in nature. They are rarer in *Pinus* xylem oleoresins[28] and more common in *Abies* cortical oleoresin[15] and *Picea* leaf oils.[33,34] Often, they are accompanied by the chemically closely related isoborneol, free borneol, santene and camphor; although low in *Abies*, the latter becomes one of the quantitatively more important constituents of the leaf oil of certain *Picea* species.

As proposed by Ruzicka[2] on the basis of chemical evidence, the biosynthetic differentiation leading to individual members of the group is connected with stabilization of the 2-bornane carbonium ion (V) (Fig. 1), which arises by cyclization of the 1-p-menthene-8-carbonium ion (III). The intermediate V as a common precursor thus ties the camphene-related terpenoids into a particularly closely knit unit, separating them from other terpenoids. This hypothesis was strengthened later by von Rudloff,[35] who noted the qualitative co-occurrence of most of these compounds in volatile oils of *Picea* and few other genera, and by Banthorpe and Baxendale who did tracer experiments with camphor.[8]

In our work on the cortical oleoresin from *A. grandis*, *A. concolor* var. *lowiana* and intergrading populations, a strong proportionality relationship between percentages of camphene, tricyclene and bornyl acetate (Fig. 3) was observed. With camphene as a dependent variable, the regression slopes had values of $b = +9{\cdot}5$ for tricyclene, and $b = +0{\cdot}7$ for bornyl acetate, with regression constants *a* close to zero. The other compounds present in more than few per cent included α-pinene (3·5–24·0 per cent), β-pinene (5·0–84·0 per cent), limonene (0·2–32·5 per cent), and β-phellandrene (1·5–62·5 per cent), and correlated negatively to the first three. Other camphene-rich firs, such as *A. veitchii* and *A. concolor* var. *concolor*, as well as few

[33] M. VON SCHANTZ, *Planta Medica* **13**, 369 (1965).
[34] M. VON SCHANTZ and S. JUVONEN, *Acta Bot. Fennica* **73**, 1 (1966).
[35] E. VON RUDLOFF, *Phytochem.* **5**, 331 (1966).

individuals of *A. magnifica* and *A. procera* where camphene content rose to measurable values, exhibited the same behavior. The same apparently holds for *A. sibirica*, with 18·6 per cent and 26·1 per cent for camphene and bornyl acetate[36] fitting well the value for the slope.

Application of rule II requires that camphene, tricyclene and bornyl acetate be closer related to each other than to α-pinene, limonene and other monoterpenes occurring in the same sources. This is corroborated by the low *p* value for the chance co-occurrence of these compounds (as discussed earlier), and is thus in complete accord with and lends strong support to the ideas above on biosynthesis of these materials.

Sabinene, α-thujene, α-terpinene and *terpinolene*. Although the first three terpenoids are not uncommon in the plant kingdom, they appear to be rare in Pinaceae, being found so far in larger than trace amounts only in ten *Pinus* xylem oleoresins and in trace to small amounts in a few *Abies* cortical and *Picea* leaf terpenes.[18] They are ubiquitous in Cupressaceae, however, commonly occuring in *Juniperus* and *Cupressus* leaf oils. Terpinolene seems somewhat more widespread but occurs usually in small amounts (below 5 per cent) in many coniferous oleoresins and leaf oils.

The formation of sabinene, α-thujene, α- and γ-terpinene was postulated to involve[1,2] stabilization of 1-p-menthene-4-carbonium ion (IV) arising by 4–8 hydride shift from 1-p-menthene-8-carbonium ion (III) (Fig. 1). It is thus this hydride shift which is responsible for consolidating these compounds into one family and separating them from other terpenes. Biogenesis of sabinene, α-thujene and some close oxygenated compounds was studied using radioactive tracers by Sandermann and his co-workers[5,6] and by Banthorpe and Turnbull.[7] Although the results were contradictory in some features, they were compatible with the hydride shift hypothesis.

In our work on the xylem oleoresin of *Pinus muricata*[17] a proportionality relationship between sabinene, α-thujene and α-terpinene has been observed (Fig. 4). Other compounds present in the same oleoresin and correlating negatively with sabinene included α-pinene, β-pinene, limonene, 3-carene and β-phellandrene. According to rule II and in agreement with the low values for chance probability of co-occurrence, the three first compounds should be biosynthetically closer to each other than to the other compounds mentioned, which agrees with biogenetic ideas outlined above.

More interesting, however, is a proportionality relationship between sabinene and terpinolene, encountered in the same source (Fig. 4). Terpinolene can be visualized as forming by proton loss either from C_8 of carbonium ion IV, or from C_4 of carbonium ion III. The proportionality relationship encountered should indicate, thus, a particularly close biosynthetic tie between sabinene and terpinolene and strongly suggests IV as the common precursor for the two compounds.

3-Carene and terpinolene. 3-Carene represents a terpene commonly encountered in plants, including the Pinaceae, the biosynthesis of which has not yet been examined. Its formation most likely involves the attack of C_8 of 1-p-menthene-8-carbonium ion (III) on C_5, with loss of the corresponding proton (Fig. 1).

In our work on *A. amabilis* cortical oleoresin, a strong proportionality was observed between 3-carene (2·2–57·5 per cent) and terpinolene (0·0–4·0 per cent) with correlation coefficient of $r = +0{\cdot}84$, regression slope $b = +0{\cdot}055$, and regression constant close to zero (terpinolene as dependent variable). Other compounds present in substantial quantities were

[36] A. P. PENTEGOV, V. A. PENTEGOVA and M. A. CHIRKOVA, *Akad., Nauk SSSR; Sibir. Otd; Trudy Chim. Metallurg, Inst.* **13**, 5 (1959).

α-pinene (5·5–40·0 per cent), β-pinene (2·5–24·0 per cent), limonene (0·0–35·5 per cent) and β-phellandrene (3·0–63·0 per cent). Positive correlation between 3-carene and terpinolene was noted also in case of *A. lasiocarpa* cortical oleoresin[32]. Hanover quoted a positive correlation between 3-carene and an "unknown" with GLC retention times closely corresponding to terpinolene in the cortical oleoresin of *P. monticola* ($r = +0{\cdot}625$)[31] and in the xylem oleoresin of *Pseudotsuga menziessii* ($r = +0{\cdot}557$).[37] The low *p* figure for the co-occurrence of terpinolene and 3-carene calculated from Mirov's data, and mentioned before, is also in agreement with the above.

As can be seen in Fig. 4, with *Pinus muricata* xylem oleoresin a large number of sabinene-terpinolene points at the sabinene values of 0–5 per cent cluster well above the regression line. All these points correspond to the oleoresin obtained from "intermediate" chemical variant of this pine containing 69·0 to 94·5 per cent 3-carene and trace–8·5 per cent terpinolene with further constituents including α-pinene (1–20 per cent), sabinene (trace–7 per cent) and myrcene (0·5–3·0 per cent); other variants of this pine lack 3-carene for all practical purposes. Using data obtained with *P. muricata* population from near Monterey where sabinene is present in amounts less than 1 per cent, the 3-carene/terpinolene ratio calculated as 1:0·055, in perfect agreement with slope of *A. amabilis* regression line.

Since it is difficult to imagine how 3-carene can be formed through 1-p-menthene-4-carbonium ion (IV), the strong linkage of terpinolene to 3-carene suggests the existence of an alternate biogenetic route for terpinolene—namely, loss of the C_4 proton in 1-p-menthene-8-carbonium ion (III). In conformity with that, correcting terpinolene values for the amount synthesized through the carbonium ion III path (using the *A. amabilis* value of 0·055 for the regression slope) placed the deviating sabinene-terpinolene points near the theoretically expected values (Fig. 4).

This example demonstrates the potential power of the method, as it does not seem possible to devise an isotopic tracer experiment capable of distinguishing the two routes to terpinolene, and estimating their individual contribution in a given case.

β-Pinene and β-phellandrene. These compounds represent rather common plant terpenes. Formation of β-pinene has been proposed to take place by[1,2] the attack by the C_8 of the C_2 of the double bond, in the 1-p-menthene-8-carbonium ion (III) with the loss of the proton at C_7 (Fig. 1). An alternative non-carbonium ion mechanism which involves internal cycloaddition of myrcene has been also proposed by Ruzicka.[2,38] Biosynthesis of β-phellandrene has not been well treated in literature, although clearly formation by cycloaddition or a Diels–Alder reaction does not apply.

In our work on *A. lasiocarpa* cortical terpenoids,[32] a number of correlations were observed. These included striking proportionality relation between β-pinene and β-phellandrene ($r = +0{\cdot}95$, $b = +0{\cdot}406$ and *a* close to zero, β-phellandrene as independent variable), and a substitutional relation between either of these two compounds and limonene (for sum of β-pinene and β-phellandrene, $r = -0{\cdot}96$, $b = -0{\cdot}98$ and $a = 78$—limonene as independent variable). Low negative correlations were furthermore encountered between the pairs α-pinene/limonene, 3-carene/β-phellandrene and 3-carene/β-pinene.

The strong proportional relation between β-pinene and β-phellandrene according to rule II suggests a particular close mechanistic link between these two compounds, closer

[37] J. W. HANOVER and M. M. FURNISS, *Joint Proc. 2nd Genetics Workshop of the Soc. of Am. Foresters and the 7th Lake States Tree Improv. Conf.* (1965). U.S. Forest Serv. Res. Paper NC-6, 23 (1966).

[38] J. H. RICHARDS and J. B. HENDRICKSON, *The Biosynthesis of Steroids, Terpenes and Acetogenins*, p. 218, W. A. Benjamin, Inc., New York (1964).

than the link between either of them and limonene or 3-carene. Since for limonene and 3-carene, as well as for β-phellandrene, carbonium ion mechanisms are favoured[1,2] the internal cycloaddition mechanisms proposed for the formation of β-pinene can be ruled out.

Although an explanation of the particularly close biosynthetic link between β-pinene and β-phellandrene must probably await the time when the processes of terpene biosynthesis are known in more detail, the model depicted in Fig. 1 seems fairly satisfactory. In both cases it involves a loss of proton at C_7, with the electron flow through C_1 and C_2 and differentiation in β-pinene and β-phellandrene connected with the C_2—C_8 bond formation in the case of β-pinene, and C_3—C_8 hydride shift in the case of β-phellandrene.

l-α-Pinene and d-α-pinene. The appearance of α-pinene in two enantiomeric forms in *Pinus* was ascribed to two different biosynthetic paths—only one of them involving the interaction of positively charged C_8 of carbonium ion III with the double bond at C_1.[39] This is corroborated by the results of tracer experiments with *P. nigra austriaca* by Sandermann *et al.*[6] Of the two mechanisms, the one involving a double bond and leading to the l-compound is clearly closer to β-pinene, the difference residing essentially in loss of the proton at C_6 instead of at C_7, while direct attack at C_6 has few features in common with β-pinene biosynthesis. It would, thus, be expected that variability in β-pinene content would be more closely tied to variability of l-α-pinene than to variability of d-α-pinene contents (Fig. 1).

In *Pinus* oleoresins, both l- and d-α-pinenes represent rather common compounds, while β-pinene is always in l-form.[28] In our work on *P. ponderosa* xylem oleoresin[40] we observed a positive correlation, although not a complete proportionality, between α-pinene (3·5–20 per cent in turpentine) and β-pinene (14·0–67·5 per cent in turpentine) ($r = +0{\cdot}65^*$), $b = +0{\cdot}184$, $\alpha = 1{\cdot}30$, β-pinene as independent variable. Both α-pinene and β-pinene correlated negatively to 3-carene (0·0 to 60·5 per cent in turpentine; $r = -0{\cdot}53$ and $-0{\cdot}76$ respectively); correlations to limonene (0·0–34·0 per cent), myrcene (0·0–23·0 per cent) and trace constituents were not significant. The specific rotation of the isolated α-pinene amounted on the average to $[\alpha_d^{20}] = -19°$, i.e. it represented largely the l-α-pinene. Still, with specific rotation of pure enantiomers amounting to about 55°,[28] the presence of substantial amounts of the d-compound in the α-pinene was indicated. Determination of the optical rotation of α-pinene at the representative points of the regression curve indicated the α-pinene rotation becoming more negative with increase in β-pinene or α-pinene content of the turpentine (Fig. 5). This stronger proportional relationship between β-pinene and l-α-pinene than between β-pinene and d-α-pinene affords further support for the biosynthetic considerations above.

Monoterpenoids and diterpenoids. Diterpenoids (in particular the resin acids) represent, together with monoterpenoids, the large part of most xylem and cortical oleoresins. Biosynthesis of monoterpenoids is thought to involve dephosphorylation of geranyl- or neryl-pyrophosphates in combination with cyclizations and other transformations of the resulting carbonium ions. The formation of diterpenoids is thought to proceed[1,2] through reaction of geranylpyrophosphate with two molecules of isopropenylpyrophosphate to geranylgeranyl-pyrophosphate, which through cyclizations and oxidations leads to resin acids and related materials. The synthesis of these two terpenoid classes is thus well separated and no significant correlation between variability of monoterpenoid composition and diterpenoids would be expected.

* Correlation coefficient rose to $r = +0{\cdot}94$, using two straight lines joined at an angle.

[39] L. Westfelt, *Svensk Kemisk Tidskrift* 79, 456 (1967).

[40] E. Zavarin and F. W. Cobb, Jr., to be published.

In our work on *A. lasiocarpa* cortical oleoresin,[32] in spite of the exceptionally strong compositional fluctuations within monoterpenoids, the quantitative changes in amount of each individual compound were largely independent of the monoterpene (100-diterpene) content of the total oleoresin in good agreement with independence rule I. Thus, with total monoterpene content as the dependent variable, the correlation and regression coefficients for limonene (range 0·0–91·0 per cent) and β-phellandrene (range 0·0 to 66·5 per cent, both total terpene basis) amounted to b = –0·015, r = –0·11 and b = +0·035, r = 0·19 only, while the regression constants were a = 29·8 and a = 28·2 respectively.

At the present time, we are continuing work on the quantitative co-occurrence by extending it into the field of sesquiterpenoids, where several biosynthetic problems exist which are not soluble or only difficultly soluble by other available methods.

EXPERIMENTAL

Oleoresin Collection and Analysis

The methods used in collecting xylem and cortical oleoresin were previously described.[15] Cortical oleoresin of *Abies concolor* var. *lowiana*, and *A. grandis*, pure and intergrading populations, 120 samples, were collected in northern California, Oregon, Washington, Idaho, and Montana. *A. amabilis* samples were obtained from 37 trees in northern California, Oregon, Washington and British Columbia. Samples of *A. balsamea* (252 trees), *A. fraseri* (55 trees), *A. magnifica* (127 trees), and *A. procera* (98 trees) cortical oleoresin were obtained from populations spanning entire ranges of these species. *A. bracteata* analyses were based on several samples collected in Santa Lucia mountains while cortical oleoresins of Japanese species, *A. veitchii* (11 trees), *A. mayriana* (16 trees), *A. mariessii* (10 trees), *A. firma* (18 trees) and *A. homolepis* (8 trees) stemmed from one or two populations of each in Japan. Analysis of xylem oleoresin of *Pinus ponderosa* was based on 134 individual samples collected in the University of California Blodget Experimental Forest in the central Sierra Nevada of California.

The methods of monoterpene hydrocarbon analysis were described before.[15] The ratio of bornyl acetate to camphene was determined, using 10% Carbowax 20 M on Chromosorb P, 60/80, $\frac{1}{16}$ in. i.d. × 5 ft copper column at 168°, and using a response factor of 1·33 determined in a separate experiment.

α-Pinene was isolated from the turpentine samples obtained by distillation of the *P. ponderosa* oleoresin using a Varian Autoprep Model A-200 instrument, equipped with thermal conductivity detector and $\frac{3}{8}$ in. o.d. × 8 ft copper column filled with 10%, 1,2,3-tris-cyanoethoxypropane on Chromosorb P, 60/80 acid-washed; helium flow rate was 200 ml/min. The purity of the isolated α-pinene fractions were better than 99% as determined by analytical GLC. The optical rotation of α-pinene was determined in toluene solution using 0·1 ml microcells.

Probability Calculations

The calculations are based on the following problem of probability mathematics. In a population of n elements, n_1 are red and $n_2 = n - n_1$ are black. A group of r elements is chosen at random. We seek the probability p_k that the group so chosen will contain k red elements.

Given S = number of separate sources examined, with subsets S_a, S_b = number of source containing compound A or compound B (separately or together), and S_{ab} = number of sources containing both compounds, use of hypergeometric distribution functions[16] gives the following chance for the two compound AB, co-occurrence:

$$p_{ab} = \frac{S_a!\,S_b!(S - S_b)!(S - S_a)!}{S!\,S_{ab}!(S_a - S_{ab})!(S_b - S_{ab})!(S + S_{ab} - S_a - S_b)!} \tag{1}$$

with: $n = S$, $n_1 = S_a$, $r = S_b$, $k = S_{ab}$.

For the special cases of $S_b = S_{ab}$ and $S_a = S_b = S_{ab}$, (1) reduces to:

$$p_{ab} = \frac{S_a!(S - S_b)!}{S!(S_a - S_b)!} \tag{2}$$

and

$$p_{ab} = \frac{S_a!(S - S_a)!}{S!} \tag{3}$$

respectively.

Expression (1) can be extended also to the co-occurrence of more than two compounds. For three compounds, probability for co-occurrence for the compound C and compounds A and B combination, $p_{(ab)c}$, can

be calculated from (1) by setting $n_1 = S_{ab}$, r = Sc and R = S_{abc}—the latter two defined in conformity to the former. The final probability for the ABC co-occurrence is then given by:

$$p_{abc} = p_{ab} \cdot p_{(ab)c} \tag{4}$$

Acknowledgements—We thank Dr. Thomas Reichert, Mr. Karel Snajberk, Miss Elaine Tsien, Mrs. Wilhelmina Hathway, and Mrs. Joanne Chin of the Forest Products Laboratory (Berkeley) for analytical work. Thanks also go to Dr. William B. Critchfield and Dr. Stanley Krugman of the Pacific Southwest Forest and Range Experiment Station (Berkeley), to Professor William B. Libby, School of Forestry and Conservation of the University of California, Berkeley, to Professor Richard Kepner, Chemistry Department, University of California, Davis, and to Dr. James D. Cumming, University of California Forest Products Laboratory, for many discussions and suggestions, and to the National Science Foundation for supporting this work in part by Grant No. GB3954.