

The carbon–nutrient balance hypothesis: its rise and fall

Jason G. Hamilton,^{1*} Arthur R. Zangerl,² Evan H. DeLucia¹ and May R. Berenbaum²

¹Department of Plant Biology, University of Illinois, Urbana, IL 61801, U.S.A.

E-mail: jhamilt@life.uiuc.edu

²Department of Entomology, University of Illinois, Urbana, IL 61801, U.S.A.

Abstract

The idea that the concentration of secondary metabolites in plant tissues is controlled by the availability of carbon and nitrogen in the environment has been termed the carbon–nutrient balance hypothesis (CNB). This hypothesis has been invoked both for prediction and for *post hoc* explanation of the results of hundreds of studies. Although it successfully predicts outcomes in some cases, it fails to such an extent that it cannot any longer be considered useful as a predictive tool. As information from studies has accumulated, many attempts have been made to save CNB, but these have been largely unsuccessful and have managed only to limit its utility. The failure of CNB is rooted in assumptions that are now known to be incorrect and it is time to abandon CNB because continued use of the hypothesis is now hindering understanding of plant–consumer interactions. In its place we propose development of theory with a firm evolutionary basis that is mechanistically sophisticated in terms of plant and herbivore physiology and genetics.

Keywords

Carbon–nutrient balance, CNB, herbivory, heritability, optimal defence, plant–insect interactions, secondary metabolites.

Ecology Letters (2001) 4: 86–95

INTRODUCTION

The carbon–nutrient balance hypothesis (CNB) arose from attempts to explain differences in plant resistance to herbivory (Bryant *et al.* 1983). It postulates that the carbon–nutrient status of plants, as determined by resource availability, directly controls allocation to secondary metabolites. This allocation in turn affects palatability and resistance to herbivores. Appealing in its simplicity and the comparative ease with which it generates testable predictions, the CNB hypothesis has provided an important framework for hundreds of studies in plant–herbivore interactions over almost two decades. While CNB has successfully predicted the outcome of some experiments, it has also failed repeatedly (e.g. Herms & Mattson 1992; Karban & Baldwin 1997; Koricheva *et al.* 1998). Despite these failures, its use as an explanatory paradigm seems to continue unabated. We contend that continued progress in understanding patterns of production of secondary metabolites in plants and in understanding plant–herbivore interactions requires a shift to more complete models with a mechanistic basis in physiology and with an evolutionary underpinning. In this review, we argue

that continued use of CNB is no longer logically or philosophically defensible. We show why CNB cannot work as a general explanatory theory, and we suggest an alternative approach.

The carbon–nutrient balance hypothesis

CNB attempts to explain the concentrations of secondary metabolites in plant tissues as a function of the relative abundances of plant resources, particularly nitrogen and light. It rests on a few explicit assumptions. First, growth (construction of new cells) always takes priority over secondary metabolite production (Tuomi *et al.* 1991). That is, carbon and nitrogen are allocated to the production of secondary metabolites only after the requirements for growth are met. Second, nutrient limitations constrain growth more than rates of photosynthesis, and light limitations constrain photosynthesis more than growth (Bryant *et al.* 1983, 1988). Third, concentrations of precursor molecules are the most important determinant of the rate of secondary metabolite production (Reichardt *et al.* 1991). In contrast with other theories of allocation to plant secondary metabolite production, the CNB hypothesis does not explicitly

assume that the chemical structure or quantity of secondary metabolites produced reflects selection pressures exerted by herbivores (Tuomi *et al.* 1988).

The CNB hypothesis is really a variant of the much older “overflow metabolism” hypothesis that supposes that secondary metabolites are produced to reduce abnormal concentrations of cellular constituents (Lutz 1928; Haslam 1986; Berenbaum 1995). For example, when nitrogen limits plant growth, the CNB hypothesis predicts that carbohydrates will accumulate in plant tissues. This increased concentration of carbohydrates will lead, by simple mass-action, to an increased synthesis of carbon-based secondary metabolites such as phenolics and terpenes. Conversely, when light limits photosynthesis, all available carbohydrates will be shunted to growth, reducing the concentration of carbohydrates in plant tissues and lowering the carbon/nitrogen (C/N) ratio in plant tissues. This change in C/N ratio will lead to decreased synthesis of carbon-based secondary metabolites and, because of increased availability of nitrogen, to increased synthesis of nitrogen-containing secondary compounds such as alkaloids and cyanogenic glycosides.

An appealing aspect of the CNB hypothesis is that it makes specific predictions about patterns of allocation to plant secondary metabolite production. Nitrogen fertilisation and shading are both expected to lower C/N ratios and lead to a decrease in carbon-based defences. Although many experiments have yielded results that are consistent with this prediction, there are many others with results that are not consistent (e.g. Bryant *et al.* 1987; Herms & Mattson 1992; Lincoln *et al.* 1993; Lindroth *et al.* 1993; Gershenson 1994b; Ohnmeiss & Baldwin 1994). With more than 200 studies having direct bearing on the hypothesis, it is now clear that CNB fails to correctly predict outcomes in a substantial proportion of cases (Karban & Baldwin 1997; Koricheva *et al.* 1998). For example, Herms & Mattson (1992) cite 24 studies failing to confirm the predictions of CNB. More recently, Koricheva *et al.* (1998), in an explicit attempt to evaluate the generality of this hypothesis, conducted a meta-analysis of 147 studies testing the CNB hypothesis. They concluded that CNB cannot predict the quantity of any single carbon-based secondary compound (CBSC) but “can only make valid predictions concerning the total amount of carbon that can be allocated to CBSCs”. However, the appeal of the CNB hypothesis lay in its intended ability to predict concentrations of individual compounds (Bryant *et al.* 1987; Reichardt *et al.* 1991). Despite the wide-ranging shortcomings of CNB, the hypothesis continues to be invoked in the design and interpretation of experiments.

ATTEMPTING TO SAVE CNB

When the predictions of a hypothesis are shown to be incorrect, one possibility is to declare that the hypothesis is falsified and reject it (Popper 1959). However, this approach is too absolute in practice because there are few theories that cannot be falsified under certain circumstances. According to Kuhn (1996), a hypothesis can sometimes be salvaged by additional study. On other occasions, the scientific community can relegate the controversy to “a future generation” with more appropriate tools than are available in contemporary practice. Finally, the hypothesis can be replaced with a different hypothesis that will be subjected to further scrutiny.

There are three ways in which scientists have attempted to salvage CNB: (1) reformulating the predictions; (2) restricting the scope of the hypothesis and (3) modifying the hypothesis. We consider each of these efforts in turn.

Reformulating predictions of CNB

In a case where an experiment fails to support the predictions of a hypothesis, it is possible that the hypothesis was correct but misapplied. The CNB hypothesis was originally formulated to predict concentrations of secondary metabolites on the basis of changes in the production of secondary metabolites (Bryant *et al.* 1983, 1988). However, concentration has been shown to be a misleading measure of production of plant secondary metabolites because both numerator and denominator variables can change in response to environmental variation (Koricheva 1999). That is, the concentration of a secondary metabolite depends on synthesis of that compound and synthesis of all other plant constituents as well. For example, changes in secondary metabolism may not affect the concentration if these changes are accompanied by concomitant changes in specific leaf area. Chemical concentration may be an appropriate measure of quality of plant tissue for herbivores, but patterns of concentration alone cannot provide information on mechanisms (Koricheva 1999). This confusion has led different researchers to support or refute CNB based on similar results. For example, one study found increased concentrations of alkaloids under shade conditions, as predicted by CNB, but concluded that their results did not support CNB because the change in concentrations were due to reductions in amounts of nonstructural carbohydrates that significantly reduced leaf mass (Ralphs *et al.* 1998).

In another study, elevated CO₂ did not deplete leaf cyanogenic glycoside concentration as predicted by CNB (Gleadow *et al.* 1998). The authors suggest that the problem was not with the theory, but with the formulation of

the predictions. That is, instead of a decreased leaf nitrogen content due to increased carbohydrate concentration, elevated CO₂ may actually lead to increased nitrogen available for secondary metabolism because the efficiency of photosynthesis is increased. In other words, elevated CO₂ may allow a plant to reallocate nitrogen from photosynthesis to secondary metabolism.

Restricting the scope of CNB

Every hypothesis must have a stated domain (Pickett *et al.* 1994) – that is, application to a clearly delimited group of phenomena. One approach to reconciling contradictory results is to restrict the domain of the hypothesis to exclude the phenomena for which it is not predictive. Based on their observations of boreal woody species and graminoids, Bryant *et al.* (1983) proposed that carbon–nutrient balance affects secondary compound concentrations in all plants. In an early test of CNB, Bryant and coworkers determined that a slow-growing tree species adapted to a resource-limited environment did not respond as predicted (Bryant *et al.* 1987). Thus, exceptions were acknowledged early in the development of the theory. Over the past two decades, other authors have also concluded that CNB does not apply either to a particular group of secondary compounds or to a particular taxonomic group (Fig. 1). For example, Reichardt *et al.* (1991) concluded that CNB did not apply to “dynamic metabolites”, those metabolites that rapidly degrade, are mobile, or are metabolically labile. The result of restricting the domain of CNB, which was initially intended to predict concentrations of all secondary metabolites in all taxa, has limited its application to predicting only the total concentration of phenylpropanoids in woody species (Koricheva *et al.* 1998) (Fig. 1).

Restricting the domain of a paradigm is not, in general, a desirable course of action, because every restriction of the domain limits applicability. In the case of CNB the domain has been restricted to such an extent that it applies to only a handful of metabolites in a small number of taxa (Fig. 1).

Modifying CNB

Another possible explanation for the failure of a hypothesis to predict outcomes is that the hypothesis is incomplete – that is, its major assumptions are correct, but crucial elements are missing or need to be fine-tuned. A classic example of this approach to refining a paradigm involves early models of the orbits of heavenly bodies (Casper & Noer 1972). Greek and Roman astronomers had assumed that the earth is stationary at the centre of the universe and that all other heavenly bodies orbited the

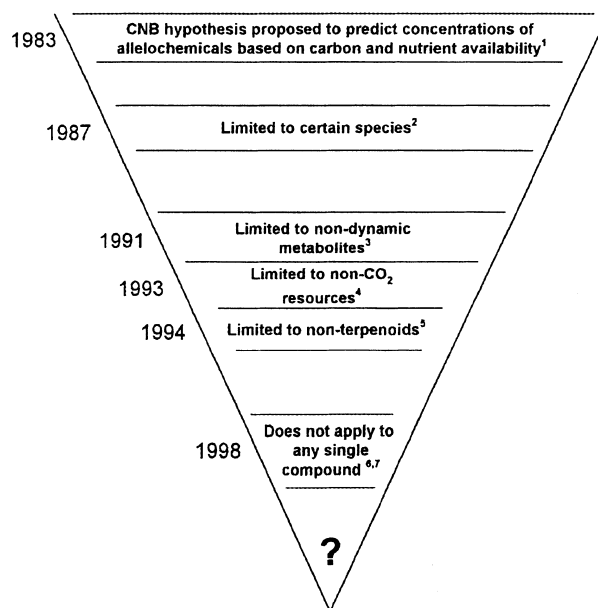


Figure 1 Illustration of the progressive narrowing of applicability of CNB from its inception in 1983 to present. ¹Bryant *et al.* 1983; ²Bryant *et al.* 1987; ³Reichardt *et al.* 1991; ⁴Lincoln *et al.* 1993; ⁵Gershenzon 1994b; ⁶Peñuelas & Estiarte (1998); ⁷Koricheva *et al.* 1998.

earth in circular motion. Observations inconsistent with this model prompted Copernicus to replace this model with a heliocentric model according to which planets orbit the sun in circular orbits. While correct in outline, this model still did not reproduce observations accurately. Johannes Kepler’s insight led him to realise that the basic model of Copernicus needed to be fine-tuned with the use of elliptical orbits instead of circular orbits.

Recently, modifications to CNB have been proposed that keep the fundamental assumptions of CNB but attempt to fine-tune it. For example, a genetically determined “basal allocation” has been suggested to mitigate environmentally controlled allocation (Hemming & Lindroth 1999; McDonald *et al.* 1999). A similar but reversed hierarchy suggests that environmentally based tendencies predicted by CNB are overridden by downstream biosynthetic pathways (Koricheva *et al.* 1998). These additions might improve the general model but do not address the weaknesses of the assumptions underpinning CNB.

THE FUNDAMENTAL ASSUMPTIONS OF CNB ARE NOW KNOWN TO BE INCORRECT

Irrespective of the successes of CNB, all of these efforts to salvage the hypothesis are not likely to succeed because it is now known that its fundamental assumptions were incorrect. One assumption is that the ability of a plant to

synthesise a defence compound can be predicted based on the atoms contained in the compound relative to the availability of these atoms in the plant. Thus, secondary compounds are classified into the functional groups of "carbon-based" or "nitrogen-based", even though these groups include biosynthetically unrelated molecules with different activities as well as different transport and storage requirements (Berenbaum 1995). It is now clear that the cost of a defence compound is not simply a function of the number and types of atoms it contains; the costs of biosynthetic machinery, storage, transport, and maintenance are also factors that make generalizations about N-containing and C-containing compounds impossible (Gershenzon 1994a). For example, alkaloids, nitrogen-containing compounds, are often more costly in terms of glucose than many phenolics that consist exclusively of C, H and O. In addition, C-based compounds may require more investment in N-containing enzymes for synthesis and storage than many N-based compounds (Gershenzon 1994a).

Another assumption is that growth always takes priority over secondary metabolism in allocation of resources (Tuomi *et al.* 1991). In other words, secondary metabolites can never have greater fitness value than growth. It is not integral to CNB that secondary metabolite concentrations sufficient to fend off herbivores and pathogens may be essential to a plant's survival. Arguing against this assumption is the vast body of evidence implicating many secondary compounds in defence (Rosenthal & Berenbaum 1991) and of specific regulatory elements in promoters of biosynthesis genes that upregulate the production of putative defence compounds in response to the presence of specific elicitors. These elicitors include microbial as well as arthropod products (Karban & Baldwin 1997). The very existence of herbivore- and pathogen-inducible secondary chemistry and the costs associated with such responses (Baldwin 1998; van Dam & Baldwin 1998) suggest that defence can override other aspects of plant function even when resources such as light and nutrients are severely restricted (Ohnmeiss & Baldwin 1994; Hemming & Lindroth 1999; Zangerl & Berenbaum 1994, 1995). Moreover, defence is not the only alternative to growth; reproduction, protection from such abiotic stresses as drought, ultraviolet light exposure, ozone and temperature extremes, and involvement in interspecific interactions other than herbivory, including competition and mutualisms, also make demands on resources (Seigler & Price 1976; Seigler 1977, 1998).

The assumption that simple mass-action drives rates of secondary metabolite production (Reichardt *et al.* 1991) implies that plants have little ability to control their chemical composition. On the contrary, secondary

metabolite production is finely regulated by enzymes physically, spatially, temporally and developmentally (Gershenzon 1994a). This assumption ignores the extremely well documented influence of genotype on chemical phenotype of a plant (e.g. Marquis 1992). A chemical phenotype is determined by three factors: genotype (G), environment (E) and interaction between genotype and environment ($G \times E$) (Falconer 1989):

$$P = G + E + G \times E$$

The separate contributions of genotype and environment to phenotypic variation are routinely quantified by applying standard quantitative genetic analyses. The result is a statistic, heritability, which is a ratio of the genetic variation to the sum of genetic and environmental variation. Heritability can range from 1 to 0, with 1 indicating that all variation among phenotypes is attributable to genetic variation and 0 indicating that all of the variation among phenotypes is attributable to environmental variation. A heritability of 0 does not necessarily mean that final concentration of secondary metabolites is independent of the action of genes; it merely means that all phenotypes in the population may share the same genes. Examination of heritabilities for a variety of secondary compounds reveals that, in most cases, between half and all phenotypic variation is genetic (Table 1). Indeed, many of the compounds have very high heritabilities, indicating that there is little if any environmental component. It might be argued that these heritabilities are inflated because they are based on data collected in uniform environments (common garden or greenhouse); however, a general survey of studies failed to detect any consistent bias in heritabilities between field and laboratory studies (Weigensberg & Roff 1996).

A different approach to elucidating the importance of environmental variation is to compare the degree to which it changes secondary metabolite concentrations relative to changes affected by variation in other factors. For example, the mean concentration of the furanocoumarin xanthotoxin, a potent toxin in the foliage of *Pastinaca sativa*, changes most dramatically as a function of damage (Zangerl & Berenbaum 1994, 1995), followed by ontogenetic changes (Zangerl 1986), and genotypic variation (Zangerl 1986), while light and nutrients have minimal effects (Zangerl & Berenbaum 1990). In *Plantago lanceolata*, far greater variation in the toxin aucubin is attributable to genotype (Fajer *et al.* 1992) than to either damage or nutrients (Fajer *et al.* 1992; Darrow & Bowers 1999). That the ordering of factors by importance differs among species is probably the rule, not the exception. For example, in *Betula pubescens* ssp. *tortuosa*, the factor with the greatest influence on total phenolics and condensed tannins is environment (fertilization and shade), followed

Table 1 Genetic variation in secondary compounds expressed as heritability (h^2), the proportion of phenotypic variation that is attributable to genetic effects. Unless otherwise noted, all heritabilities are narrow-sense (Lynch & Walsh 1998)

Secondary compound	h^2	Genus	Reference
Terpenes			
α -pinene	0.63	<i>Pinus</i>	Hanover (1966)
β -pinene	0.86	<i>Pinus</i>	Hanover (1966)
3-carene	0.94	<i>Pinus</i>	Hanover (1966)
camphene	0.38	<i>Pinus</i>	Hanover (1966)
limonene	0.87	<i>Pinus</i>	Hanover (1966)
cacalol	0.12–0.89 ^b	<i>Adenostyles</i>	Hägele & Rowell-Rahier (1999)
cacalol-trimer	0.48–0.63 ^b	<i>Adenostyles</i>	Hägele & Rowell-Rahier (1999)
Phenolics			
total phenolics	0.77–0.91 ^b	<i>Prunus</i>	Nachit & Feucht (1983)
tannins	0.66 ^a	<i>Lens</i>	Vaillancourt <i>et al.</i> (1986)
tannins	0.48–0.97 ^b	<i>Phaseolus</i>	Ma & Bliss (1978)
tannins	0.77	<i>Phaseolus</i>	Elia <i>et al.</i> (1997)
salicortin	0.20	<i>Salix</i>	Orians <i>et al.</i> (1996)
2'-cinnamoylsalicortin	0.59	<i>Salix</i>	Orians <i>et al.</i> (1996)
Alkaloids			
alkaloids	0.19–0.70 ^b	<i>Phalaris</i>	Ostrem (1987)
indole alkaloids	0.44–1.0	<i>Phalaris</i>	Barker & Hovin (1974)
pyrrolizidine alkaloids	0.52–1.0 ^b	<i>Senecio</i>	Vrieling (1990)
pyrrolizidine alkaloids	0.33–0.43 ^b	<i>Cynoglossum</i>	van Dam & Vrieling (1994)
gramine	0.72	<i>Phalaris</i>	Coulman <i>et al.</i> (1977)
gramine	0.70–0.77 ^b	<i>Hordeum</i>	Moharrampour <i>et al.</i> (1999)
hordenine	0.53	<i>Phalaris</i>	Coulman <i>et al.</i> (1977)
vicine	0.45	<i>Vicia</i>	Bjerg <i>et al.</i> (1985)
convicine	0.66	<i>Vicia</i>	Bjerg <i>et al.</i> (1985)
caffeine	0.80	<i>Coffea</i>	Montagnon <i>et al.</i> (1998)
morphine	0.76	<i>Papaver</i>	Lokostoth & Hezky (1994)
seneciphylline	0–0.86	<i>Adenostyles</i>	Hägele & Rowell-Rahier (1999)
senecionine	0.6–0.86	<i>Adenostyles</i>	Hägele & Rowell-Rahier (1999)
Furanocoumarins			
imperatorin	0.75–0.85	<i>Pastinaca</i>	Zangerl & Berenbaum (1990)
bergapten	0.64–0.85	<i>Pastinaca</i>	Zangerl & Berenbaum (1990)
xanthotoxin	0.43–0.78	<i>Pastinaca</i>	Zangerl & Berenbaum (1990)
sphondin	0.86–0.87	<i>Pastinaca</i>	Zangerl & Berenbaum (1990)
Glucosinolates			
total glucosinolates	0.87	<i>Brassica</i>	Rucker & Robbelen (1994)
sinigrin	0.72	<i>Brassica</i>	van Doorn <i>et al.</i> (1999)
progoitrin	1.0	<i>Brassica</i>	van Doorn <i>et al.</i> (1999)

^aInsufficient information to determine. ^bBroad-sense heritability.

distantly by genotype and damage (Ruohomäki *et al.* 1996). However, in *Betula pendula*, the most important factor seems to be genotype; of 36 different secondary metabolites or groups of metabolites examined, 100% were found to vary among genotypes, 61% were affected by shading, and 28% were affected by damage. Not only was genotype the most consistently influential factor, in most cases, it also had the greatest quantitative influence (Keinanen *et al.* 1999). Because secondary metabolite concentrations are often dictated by the joint influences of both environment and genetics, no hypothesis based exclusively on one or the

other of these two influences can be accurate all of the time. However, we maintain that there are substantive reasons to believe that hypotheses based exclusively on environmental effects will be prone to greater error than hypotheses with a genetic underpinning.

ENVIRONMENTAL VARIATIONS AND A THEORY OF DEFENCE

There are certainly instances where concentrations of secondary metabolites are affected by variation in the

environment (e.g. Herms & Mattson 1992; Marquis 1992; Koricheva *et al.* 1998). However, it cannot be assumed that an effect of environment is independent of the action of genes, as CNB assumes. There may be a direct effect of the environment on gene expression, suggesting induction or suppression. There may also be a genotype–environment interaction, suggesting that the degree of induction or suppression varies among genotypes. Differentiating among all of these possibilities is difficult and requires simultaneous manipulation of genotype and environmental factors in homogeneous settings (Karban 1992; Marquis 1992).

If there are effects of environment on secondary chemical concentrations that are not mediated by genes, such variation, while not heritable, may still affect the fitness of plant, herbivore or pathogen populations. However, despite all of the attention paid to CNB over the last two decades, few investigators have ascertained whether their experimental treatments fell within an ecologically relevant range, and even fewer determined whether environment-driven variations in defence affected herbivory or disease resistance. Environmental variation outside the ecological and evolutionary experience of a plant is for most purposes irrelevant, the exception being cases in which the response to novel environments is of interest (e.g. global change). The few studies that have investigated impacts on insect herbivores of environmentally induced variation in secondary metabolites have found that these changes do not affect insect herbivores in a predictable fashion (e.g. Lindroth *et al.* 1993; Kinney *et al.* 1997; Hemming & Lindroth 1999; Hunter & Forkner 1999; McDonald *et al.* 1999; Mutikainen *et al.* 2000).

PROSPECTS FOR A THEORY OF SECONDARY CHEMICAL CONCENTRATIONS

According to Kuhn (1996), “The decision to reject one paradigm is always simultaneously the decision to accept another, and the judgement leading to that decision involves the comparison of both paradigms with nature and with each other.” Thus, it is impractical to advocate outright rejection of the CNB hypothesis without at the same time suggesting an alternate testable hypothesis that is consistent with observed facts. To the extent that secondary metabolites function to reduce herbivory and disease and are genetically variable, treating them as defences makes them amenable to interpretation within the context of selection and evolution. For that reason, we propose that a theory that successfully predicts concentrations of defence-related secondary compounds is more likely to emerge from hypotheses that have an evolutionary underpinning that presupposes an adaptive value

to the manner in which defences are deployed. This approach involves identifying selective agents and quantifying the selective forces they exert (Berenbaum 1995). The adaptive value of any trait is determined by its fitness benefit relative to its fitness cost. If, on average, benefit exceeds cost, the trait is selectively advantageous, but, if cost exceeds benefit, the trait is disfavoured. As has been demonstrated by CNB, hypotheses based on nonadaptive processes are not likely to have broad applicability. While it cannot be assumed that all chemical traits are configured adaptively for defence (Marquis 1992), for those traits that do affect fitness, natural selection provides direction to their evolution as long as they are genetically variable (which, from inspection of Table 1, does not seem to be a limiting factor).

Like any other trait, the spatial and temporal allocation of chemical defences within a plant should be predictable if the selective factors shaping defence are properly understood. These factors, identified more than 25 years ago (McKey 1974), are the value of the plant tissue to the plant, the benefit of defence, and the probability of attack (Fig. 2). Of these three, the value of a plant tissue is most readily measured; it is the difference in fitness between intact plants and similar plants in which a known amount of the tissue is removed. Even though this value is relatively easy to measure, most researchers have relied on assumptions of value (Table 2). The benefit of defence is more difficult to measure because we typically cannot, short of genetically engineering knock-outs, eliminate the

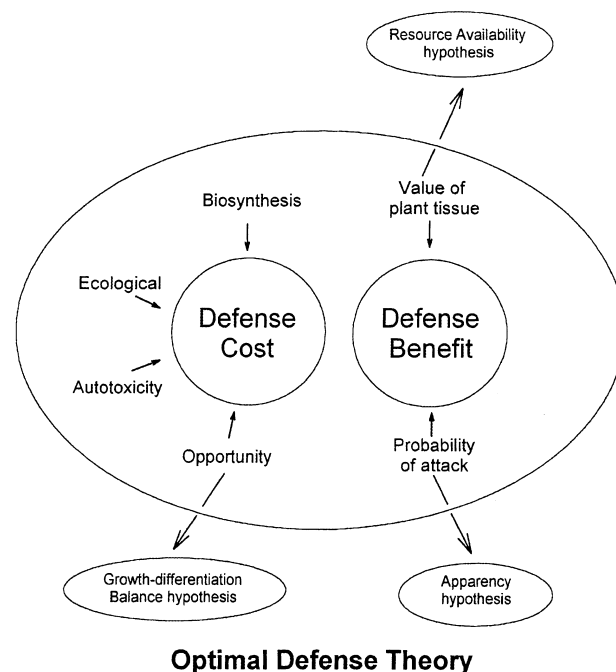


Figure 2 Relationships of existing hypotheses to an evolutionary theory of optimal defence.

Table 2 Tests of the optimal defence hypothesis

Plant part	Supports hypothesis?	Benefit of defence measured?	Cost of herbivory measured?	Likelihood of attack measured?	Reference
Leaves	no	no ^a	no ^b	no	Zangerl (1986)
Leaves	yes	no ^a	no ^b	yes	Van Dam <i>et al.</i> (1995)
Leaves	yes	no ^a	no	no	Ohnmeiss & Baldwin (1994)
Leaves	yes	no ^a	no ^b	no	Baldwin & Ohnmeiss (1994)
Reproductive parts	yes	no ^a	no	no	Baldwin & Karb (1995)
Reproductive parts	yes	no ^a	yes	no	Nitao & Zangerl (1987)
Roots, leaves, reproductive parts	yes	no ^a	no	yes	Zangerl & Rutledge (1996)

^aInvestigators assume defence is correlated with benefit. ^bInvestigators assume tissue value is correlated with photosynthesis.

defence of a plant and measure the associated loss in fitness due to herbivory (except some physical defences such as spines and thorns). For most chemical defences, we must rely on the existence of a correlation between fitness and defence of plants subjected to herbivores as evidence that the benefit of defence is proportional to the level of defence. In cases where this relationship does not hold, such as when multiple compounds act synergistically to enhance existing defences more than increasing concentrations of one compound (Berenbaum 1985), more detailed study is required.

Lastly, and perhaps most difficult of the three, is the estimation of the probability of attack, that is the likelihood that a herbivore will contact the plant at any given time. This factor determines when and for how long a plant or part of plant may benefit from defence. Likelihood of attack is probably not consistent from year to year and the pattern of defence probably reflects the average value of this variable over a period of years; as such, it is ideally measured over several seasons. A primary difficulty in measuring this variable is determining what constitutes “attack”. Incidence of physical damage is concrete evidence, but measures based on this criterion alone will not include visits by herbivores that rejected the plant without taking a bite. Thus, measurement of this factor must include studies of herbivore behaviour in addition to simply measuring plant damage. The use of plant damage as a measure of attack probability raises another problem. The incidence of damage is unlikely to be independent of secondary chemistry; it may vary inversely with defence, if the secondary chemicals are a deterrent, and positively with secondary chemicals if they serve as feeding stimulants or attractants, as it often does for specialist herbivores (Louda & Mole 1991). Identifying the ecological impact of a secondary compound also requires knowledge of specific herbivore feeding behaviour and physiology.

This evolution-based cost/benefit analysis, or “optimal-defence” theory (Rhoades 1979) makes clear predictions

that can be tested. For example, plants that are not subject to interaction with consumers (either herbivores or pathogens) have zero probability of attack and therefore derive no benefit from defence – they should not be defended if the defences have a cost. Plants subject to interactions with consumers should have defences as long as their benefit exceeds their cost. These principles are equally applicable to parts of plants as to whole plants (Zangerl & Rutledge 1996); if all plant parts are equally susceptible to loss, parts with the greatest fitness value are expected to be most heavily defended. Similarly, the timing of defence can be predicted; plant parts that are infrequently attacked or are subject to attack for limited periods should exhibit inducible defences, while those that are routinely attacked or attacked throughout their life spans should exhibit constitutive defences. Remarkably few investigators have attempted to compare actual patterns of defence to patterns predicted by optimal defence theory as outlined here. Among those who did, most found agreement (Table 2). However, strictly speaking, none of these investigators measured all of the factors essential to making predictions and, at best, measured directly only one of the three factors (Table 2).

Several other hypotheses formulated to predict optimal patterns of defence have been proposed over the years. Many of these are a step in the right direction but are incomplete because they address subsets of the factors important in the optimal defence hypothesis (Fig. 2). The resource availability hypothesis (Coley *et al.* 1985), for example, is predicated on the idea that plants living in resource-poor environments cannot readily replace resources lost to herbivores and therefore must more heavily defend the resources they have. In other words, the value of tissue to a plant in a resource-poor environment is higher than the value of a comparable amount of tissue to a plant growing in resource-rich environment. The resource availability hypothesis does not address the possibility that probability of attack may, in some cases, be the overriding factor governing the

benefit of defence. Similarly, the growth–differentiation balance hypothesis (Herms & Mattson 1992) addresses only opportunity costs, the costs to one function of diverting resources to another function, in this case the sacrifice of growth for the benefit of defence. However, its applicability is diminished when the probability of attack or value of tissue varies.

Because the hypothesis of optimal defence, as outlined here, is limited to “defences”, which by definition are traits that limit fitness losses in the presence of consumers, it could be argued that it is no more likely to be falsified than the theory of natural selection on which it is based. In a sense, this is true. The more profitable question to ask is whether the formulation of the hypothesis is sufficient to be universally predictive for the wide variety of defences encountered in plants. Already, there is some question as to whether the hypothesis applies to traits that have adverse effects on some herbivores but benefit others, either as cues or as sequestered defences against predation (Malcolm & Zalucki 1996; Siemsen & Mitchell-Olds 1996). Another difficulty is how the hypothesis applies to plants that appear to derive a benefit from consumption by herbivores (Marquis 1992; Paige 1999). In such cases of “overcompensation”, the plant apparently possesses tissues with negative fitness value; according to the optimal defence hypothesis, these tissues should be devoid of defences to encourage their removal by herbivores. Whether this phenomenon actually occurs is unknown. The solution to these problems will necessitate the elucidation of net costs and benefits of defence in the face of multiple herbivores.

CONCLUSIONS

No theory of chemical defence has ever been rejected; they all coexist by virtue of supportive evidence in one system or another (Berenbaum 1995). The argument could be made that for many of these hypotheses there are insufficient data to make definitive assessments. This is certainly not the case for CNB. This idea has captured the imagination of many investigators, and hundreds of studies have tested it, invoked it, or alluded to it. The culmination of all of this attention is that CNB fails as a predictive hypothesis and therefore must be rejected as a general phenomenon. For those species in which chemicals of a primary or secondary nature are affected by environmental variation, the implications of this variation for their interactions with herbivores are far from clear and require further investigation. We need to determine whether these variations have adaptive value and therefore are better addressed by other defence hypotheses.

In retrospect, the failure of CNB is understandable in that a “supply side” or “source-driven” formulation of

theory (Lerdau *et al.* 1994) ignores the homeostatic nature of organisms and denies an organism’s long history of evolution. For predictive value, theories grounded in evolution and adaptation are potentially more robust and deserve more attention. Perhaps the greatest contribution of CNB has been to spark interest in the secondary chemistry of plants, drawing new investigators into the field of chemical ecology. With this newly acquired interest and the skills necessary to pursue future studies, investigations of alternatives to the CNB hypothesis will flourish.

ACKNOWLEDGEMENTS

The authors thank Jennifer Hartz, Kate George, Jennie Tang and Maggie Prater for providing helpful comments on the manuscript. This work was supported in part by the University of Illinois Critical Research Initiatives Grant, the Department of Energy Office of Biological and Environmental Research grant DE-FG02-95ER62127 and the National Science Foundation grant DEB 9903867.

REFERENCES

- Baldwin, I.T. (1998) Jasmonate-induced responses are costly but benefit plants under attack in native populations. *Proc. Natl. Acad. Sci.*, 95, 8113–8118.
- Baldwin, I.T. & Karb, M.J. (1995) Plasticity in allocation of nicotine to reproductive parts in *Nicotiana attenuata*. *J. Chem. Ecol.*, 21, 897–909.
- Baldwin, I.T. & Ohnmeiss, T.E. (1994) Coordination of photosynthetic and alkaloidal responses to leaf damage in uninducible and inducible *Nicotiana sylvestris*. *Ecology*, 75, 1003–1014.
- Barker, R.E. & Hovin, A.W. (1974) Inheritance of indole alkaloids in reed canarygrass (*Phalaris arundinacea* L.) I. Heritability estimates for alkaloid concentration. *Crop Sci.*, 14, 50–53.
- Berenbaum, M.R. (1985) Brentown revisited: interactions among allelochemicals in plants. *Recent Adv. Phytochemistry*, 19, 139–169.
- Berenbaum, M.R. (1995) The chemistry of defense: theory and practice. *Proc. Natl. Acad. Sci.*, 92, 2–8.
- Bjerg, B., Nørgaard Knudsen, J.C., Olsen, O., Poulsen, M.H. & Sørensen, H. (1985) Quantitative analysis and inheritance of vicine and convicine content in seeds of *Vicia faba* L. *Z Pflanzenzuchtung*, 94, 135–148.
- Bryant, J.P., Chapin, F.S. III & Klein, D.R. (1983) Carbon/nutrient balance of boreal plants in relation to vertebrate herbivory. *Oikos*, 40, 357–368.
- Bryant, J.P., Chapin, F.S. III, Reichardt, P.B. & Clausen, T.P. (1987) Response of winter chemical defense in Alaska paper birch and green alder to manipulation of plant carbon/nutrient balance. *Oecologia*, 72, 510–514.
- Bryant, J.P., Tuomi, J. & Niemala, P. (1988) Environmental constraint of constitutive and long-term inducible defenses in woody plants. In: *Chemical Mediation of Coevolution* (ed. K.C. Spencer). Academic Press, San Diego, CA, U.S.A., pp. 367–389.

- Casper, B.M. & Noer, R.J. (1972). *Revolutions in Physics*. Norton, New York, NY, U.S.A.
- Coley, P.D., Bryant, J.P. & Chapin, F.S. III (1985) Resource availability and plant antiherbivore defense. *Science*, 230, 895–899.
- Coulman, B.E., Woods, D.L. & Clark, K.W. (1977) Distribution within the plant, variation with maturity, and heritability of gramine and hordenine in reed canary grass. *Can J. Plant Sci.*, 57, 771–777.
- van Dam, N.M. & Baldwin, I.T. (1998) Costs of jasmonate-induced responses in plants competing for limited resources. *Ecol. Lett.*, 1, 30–33.
- van Dam, N.M. & Vrieling, K. (1994) Genetic variation in constitutive and inducible pyrrolizidine alkaloid levels in *Cynoglossum officinale* L. *Oecologia*, 99, 374–378.
- van Dam, N.M., Vuister, L.W.M., Bergshoeff, C., De Vos, H. & Van der Meijden, E. (1995) The 'raison d'être' of pyrrolizidine alkaloids in *Cynoglossum officinale*. deterrent effects against generalist herbivores. *J. Chem. Ecol.*, 21, 507–523.
- Darrow, K. & Bowers, M. (1999) Effects of herbivore damage and nutrient level on induction of iridoid glycosides in *Plantago lanceolata*. *J. Chem. Ecol.*, 25, 1427–1440.
- van Doorn, J.E., van der Kruk, G.C., van Holst, G.J., Schoofs, M., Broer, J.B. & de Nijs, J.J.M. (1999) Quantitative inheritance of the progoitrin and sinigrin content in Brussels sprouts. *Euphytica*, 108, 41–52.
- Elia, F.M., Hosfield, G.L., Kelly, J.D. & Uebersax, M.A. (1997) Genetic analysis and interrelationships between traits for cooking time, water absorption, and protein and tannin content of Andean dry beans. *J. Am. Soc. for Hort Sci.*, 122, 512–518.
- Fajer, E.D., Bowers, M.D. & Bazzaz, F.A. (1992) The effect of nutrients and enriched CO₂ environments on production of carbon-based allelochemicals in *Plantago* – a test of the carbon nutrient balance hypothesis. *Am. Naturalist*, 140, 707–723.
- Falconer, D.S. (1989). *Introduction to Quantitative Genetics*. Longman, Essex, U.K.
- Gershenzon, J. (1994a) The cost of plant chemical defense against herbivory: a biochemical perspective. In: *Insect–Plant Interactions* (ed. E.A. Bernays). CRC Press, Boca Raton, FL, U.S.A., pp. 105–173.
- Gershenzon, J. (1994b) Metabolic costs of terpenoid accumulation in higher plants. *J. Chem. Ecol.*, 20, 1281–1328.
- Gleadow, R.M., Foley, W.J. & Woodrow, I.E. (1998) Enhanced CO₂ alters the relationship between photosynthesis and defense in cyanogenic *Eucalyptus cladocalyx* f. Muell. *Plant, Cell Environ.*, 21, 12–22.
- Hägele, B.F. & Rowell-Rahier, M. (1999) Genetic and environmental-based variability in secondary metabolite leaf content of *Adenostyles alliariae* and *A. alpina* (Asteraceae). A test of the resource availability hypothesis. *Oikos*, 85, 234–246.
- Hanover, J.W. (1966) Genetics of terpenes. I. Gene control of monoterpene levels in *Pinus monticola* DOUGL. *Heredity*, 21, 73–84.
- Haslam, E. (1986) Secondary metabolism – fact and fiction. *Natural Product Reports*, 3, 217–249.
- Hemming, J.D.C. & Lindroth, R.L. (1999) Effects of light and nutrient availability on aspen: growth, phytochemistry, and insect performance. *J. Chem. Ecol.*, 25, 1687–1714.
- Herms, D.A. & Mattson, W.J. (1992) The dilemma of plants: to grow or defend. *Q. Rev. Biology*, 67, 283–335.
- Hunter, M.D. & Forkner, R.E. (1999) Hurricane damage influences foliar polyphenolics and subsequent herbivory on surviving trees. *Ecology*, 80, 2676–2682.
- Karban, R. (1992) Plant variation: its effects on populations of herbivorous insects. In: *Plant Resistance to Herbivores and Pathogens: Ecology, Evolution, and Genetics* (eds R.S. Fritz & E.L. Simms). The University of Chicago Press, Chicago, IL, U.S.A., pp. 195–215.
- Karban, R. & Baldwin, I.T. (1997). *Induced Responses to Herbivory*. University of Chicago Press, Chicago, IL, U.S.A.
- Keinanen, M., Julkunen-Tiitto, R., Mutikainen, P., Walls, M., Ovaska, J. & Vapaavuori, E. (1999) Trade-offs in phenolic metabolism of silver birch: effects of fertilization, defoliation, and genotype. *Ecology*, 80, 1970–1986.
- Kinney, K.K., Lindroth, R.L., Jung, S.M. & Nordheim, E.V. (1997) Effects of CO₂- and NO₃-availability on deciduous trees: phytochemistry and insect performance. *Ecology*, 78, 215–230.
- Koricheva, J. (1999) Interpreting phenotypic variation in plant allelochemistry: problems with the use of concentrations. *Oecologia*, 119, 467–473.
- Koricheva, J., Larsson, S., Haukioja, E. & Keinanen, M. (1998) Regulation of woody plant secondary metabolism by resource availability: hypothesis testing by means of meta-analysis. *Oikos*, 83, 212–226.
- Kuhn, T.S. (1996). *The Structure of Scientific Revolutions*. 3rd edn. University of Chicago Press, Chicago, IL, U.S.A.
- Lerdau, M., Litvak, M. & Monson, R. (1994) Plant chemical defense: monoterpenes and the growth-differentiation balance hypothesis. *Trends Ecol. Evolution*, 9, 58–61.
- Lincoln, D.E., Fajer, E.D. & Johnson, R.H. (1993) Plant–insect herbivore interactions in elevated CO₂ environments. *Trends Ecol. Evolution*, 8, 64–68.
- Lindroth, R.L., Kinney, K.K. & Platz, C.L. (1993) Responses of deciduous trees to elevated atmospheric CO₂: productivity, phytochemistry, and insect performance. *Ecology*, 74, 763–777.
- Lokostoh, K. & Hezky, E.L. (1994) Inheritance of morphine content in poppies (*Papaver somniferum* L.). *Novenytermeles*, 43, 369–377.
- Louda, S. & Mole, S. (1991) Glucosinolates: chemistry and ecology. In: *Herbivores: Their Interactions with Secondary Plant Metabolites*, Vol. 1 (eds G.A. Rosenthal & M.R. Berenbaum). Academic Press, New York, NY, U.S.A., pp. 123–164.
- Lutz, L. (1928) Sur le rôle biologique du tannin dans la cellule végétale. *Bull. la Société Bot. France*, 75, 9–15.
- Lynch, M. & Walsh, B. (1998). *Genetics and Analysis of Quantitative Traits*. Sinauer, Sunderland, UK.
- Ma, Y. & Bliss, F.A. (1978) Tannin content and inheritance in common bean. *Crop Sci.*, 18, 201–204.
- Malcolm, S.B. & Zalucki, M.P. (1996) Milkweed latex and cardenolide induction may resolve the lethal plant defence paradox. *Entomologia Exp. Applicata*, 80, 193–196.
- Marquis, R.J. (1992) Selective impact of herbivores. In: *Plant Resistance to Herbivores and Pathogens: Ecology, Evolution, and Genetics* (eds R.S. Fritz & E.L. Simms). The University of Chicago Press, Chicago, IL, U.S.A., pp. 301–325.
- McDonald, E.P., Agrell, J. & Lindroth, R.L. (1999) CO₂ and light effects on deciduous trees: growth, foliar chemistry, and insect performance. *Oecologia*, 119, 389–399.
- McKey, D. (1974) Adaptive patterns in alkaloid physiology. *Am. Naturalist*, 108, 305–320.

- Moharrampour, S., Takeda, K., Sato, K., Yoshida, H. & Tsumuki, H. (1999) Inheritance of gramine content in barley. *Euphytica*, 106, 181–185.
- Montagnon, C., Guyot, B., Cilas, C. & Leroy, T. (1998) Genetic parameters of several biochemical compounds from green coffee, *Coffea canephora*. *Plant Breed*, 117, 576–578.
- Mutikainen, P., Walls, M., Ovaska, J. & Keinänen, M., Julkunen-Tiitto R. & Vapaavuori E. (2000) Herbivore resistance in *Betula pendula*: effect of fertilization, defoliation, and plant genotype. *Ecology*, 81, 49–65.
- Nachit, M. & Feucht, W. (1983) Inheritance of phenolic compounds, indoles and growth vigour in *Prunus* crosses (Cherries). *Z Pflanzenzüchtung*, 90, 166–171.
- Nitao, J.K. & Zangerl, A.R. (1987) Floral development and chemical defense allocation in wild parsnip (*Pastinaca sativa*). *Ecology*, 68, 521–529.
- Ohnmeiss, T.E. & Baldwin, I.T. (1994) The allometry of nitrogen allocation to growth and an inducible defense under nitrogen-limited growth. *Ecology*, 75, 995–1002.
- Orians, C.M., Roche, B.M. & Fritz, R.S. (1996) The genetic basis for variation in the concentration of phenolic glycosides in *Salix sericea*: an analysis of heritability. *Biochem. Syst. Ecol.*, 24, 719–724.
- Ostrem, L. (1987) Studies on genetic variation in reed canarygrass, *Phalaris arundinacea* L. I. Alkaloid type and concentration. *Hereditas*, 107, 235–248.
- Paige, K.N. (1999) Regrowth following ungulate herbivory in *Ipomopsis aggregata*: geographic evidence for overcompensation. *Oecologia*, 118, 316–323.
- Peñuelas, J. & Estiarte, M. (1998) Can elevated CO₂ affect secondary metabolism and ecosystem function? *Trends Ecol. Evolution*, 13, 20–24.
- Pickett, S.T.A., Kolasa, J. & Jones, C.G. (1994). *Ecological Understanding*. Academic Press, San Diego, CA, U.S.A.
- Popper, K.R. (1959). *The Logic of Scientific Discovery*. Basic Books, New York, NY, U.S.A.
- Ralphs, M.H., Manners, G.D. & Gardner, D.R. (1998) Influence of light and photosynthesis on alkaloid concentration in larkspur. *J. Chem Ecol.*, 24, 167–182.
- Reichardt, P.B., Chapin, F.S. III, Bryant, J.P., Mattes, B.R. & Clausen, T.P. (1991) Carbon/nutrient balance as a predictor of plant defense in Alaskan balsam poplar: potential importance of metabolite turnover. *Oecologia*, 88, 401–406.
- Rhoades, D.F. (1979) Evolution of plant chemical defense against herbivores. In: *Herbivores: Their Interaction with Secondary Plant Metabolites* (eds G.A. Rosenthal & D.H. Janzen). Academic Press, New York, NY, U.S.A., pp. 3–54.
- Rosenthal, G. & Berenbaum, M. (1991). *Herbivores: Their Interactions with Secondary Plant Metabolites*. 2nd edn. Academic Press, New York, NY, U.S.A.
- Rucker, B. & Robbelen, G. (1994) Inheritance of total and individual glucosinolate contents in seeds of winter oilseed rape (*Brassica napus* L.). *Plant Breed*, 113, 206–216.
- Ruohomäki, K., Chapin, F.S.I., Haukioja, E., Neuvonen, S. & Suomela, J. (1996) Delayed inducible resistance in mountain birch in response to fertilization and shade. *Ecology*, 77, 2302–2311.
- Seigler, D.S. (1977) Primary roles for secondary compounds. *Biochem. Syst Ecology*, 5, 195–199.
- Seigler, D. (1998). *Plant Secondary Metabolism*. Kluwer, Boston, MA, U.S.A.
- Seigler, D. & Price, P.W. (1976) Secondary compounds in plants: Primary functions. *Am. Naturalist*, 110, 101–105.
- Siemens, D.H. & Mitchell-Olds, T. (1996) Glucosinolates and herbivory by specialists (Coleoptera: Chrysomelidae, Lepidoptera: Plutellidae): consequences of concentration and induced resistance. *Environ. Entomol*, 25, 1344–1353.
- Tuomi, J., Fagerstrom, T. & Niemela, P. (1991) Carbon allocation, phenotypic plasticity, and induced defenses. In: *Phytochemical Induction by Herbivores* (eds D.W. Tallamy & M.J. Raupp). John Wiley & Sons, New York, NY, U.S.A., pp. 85–104.
- Tuomi, J., Niemela, P., Chapin, F.S. III, Bryant, J.P. & Siren, S. (1988) Defensive responses of trees in relation to their carbon/nutrient balance. In: *Mechanisms of Woody Plant Defenses Against Insects* (eds W.J. Mattson, J. Leveux & C. Bernard-Dagan). Springer-Verlag, New York, NY, U.S.A., pp. 57–72.
- Vaillancourt, R., Slinkard, A.E. & Reichert, R.D. (1986) The inheritance of condensed tannin concentration in lentil. *Can J. Plant Sci.*, 66, 241–246.
- Vrieling, K. (1990). *Costs and Benefits of Alkaloids of Senecio Jacobaea* L. Leiden University, Leiden.
- Weigensberg, I. & Roff, D.A. (1996) Natural heritabilities: can they be reliably estimated in the laboratory? *Evolution*, 50, 2149–2157.
- Zangerl, A.R. (1986) Leaf value and optimal defense in *Pastinaca sativa* L. (Umbelliferae). *Am. Midland Naturalist*, 116, 432–436.
- Zangerl, A.R. & Berenbaum, M.R. (1990) Furanocoumarin induction in wild parsnip: genetics and population variation. *Ecology*, 71, 1933–1940.
- Zangerl, A.R. & Berenbaum, M.R. (1994/1995) Spatial, temporal, and environmental limits on xanthotoxin induction in wild parsnip foliage. *Chemoecology*, 5/6, 37–42.
- Zangerl, A.R. & Rutledge, C.E. (1996) The probability of attack and patterns of constitutive and induced defense: a test of optimal defense theory. *Am. Naturalist*, 147, 599–608.

BIOSKETCH

Jason Hamilton is interested in examining the effects of insect herbivory on plant physiological processes and the effects of global change on plant–insect interactions.

Editor, J. P. Grover

Manuscript received 24 May 2000

First decision made 18 July 2000

Second decision made 19 September 2000

Manuscript accepted 27 September 2000