Available online at www.sciencedirect.com
SCIENCE DIRECT®
European Journal of Medicinal Chemistry 39 (2004) 107–111
EUROPEAN JOURNAL OF MEDICINAL CHEMISTRY
www.elsevier.com/locate/ejmech

Short communication

Analysis of datasets showing which compounds kill which organisms: inferring two systems

T.P. Hutchinson

Department of Psychology, Macquarie University, Sydney, NSW 2109, Australia

Received 18 August 2003; received in revised form 25 November 2003; accepted 27 November 2003

Abstract

Experiments are sometimes conducted to show which of several compounds is successful at inactivating which of several microorganisms. This paper proposes a method of analysing the datasets obtained. The example refers to 33 compounds (2,4-dihydroxythiobenzanilides) and seven microorganisms (dermatophytes) [data from Eur. J. Med. Chem. 35 (2000) 393]. The conclusion is that two systems or modes of action are needed to explain the data.

Keywords: 2,4-Dihydroxythiobenzanilides; Antifungal activity; Dermatophytes; Data condensation

1. Introduction

Matysiak et al. [1] reported on the fungistatic activity (against seven dermatophytes) of 33 compounds (2,4-dihydroxythiobenzanilides). I wish to point out an interesting feature of the patterns of which compounds were successful in inactivating which dermatophytes, namely, that two systems or modes of action (not one, and not three) are needed to explain the data. The method of data analysis that leads to this conclusion is potentially useful with other datasets of the same type.

The background to the study by Matysiak et al. [1] is (on the one hand) the known biological activity of thiobenzanilides, and (on the other) the difficulty of synthesising many members of this group of compounds. Having developed a new route of synthesis, Matysiak et al. were able to obtain 33 2,4-dihydroxythiobenzanilides, that differed in substituents, and to study their fungistatic activity. Other publications from the same team (e.g. [2,3]) have also reported on the antifungal activities of these compounds. The chemical structure is Ar_1–(C:S)–(NH)–Ar_2, where the aryl group Ar_1 has hydroxyl groups in positions 2 and 4, and Ar_2 has various substituents in positions 2–5.

The aim of the present paper is to describe a method of processing datasets that show which compounds kill or inactivate which organisms, and to give the results when applied to the data of Matysiak et al. Section 2 draws attention to the importance of differences between two compounds as to which organisms they inactivate, and identifies alternative strategies for proceeding with data analysis if that is found. Section 3 will introduce the reader to the notation that will be used, showing how a single-system theory would be presented. The key contribution of this paper is the two-system theory of Section 4. Section 5 concludes the paper with discussion.

2. Alternative strategies for understanding data

If a compound inactivates one organism but not another, we are not surprised, but simply say that the first organism is less tolerant (or more sensitive) than the second. If an organism is inactivated by one compound but not by another, we are not surprised, but simply say that the first compound is in some sense stronger than the second. However, in Table 1, we see the following: compounds 17 and 24 inactivate dermatophyte B but not E, whereas compounds 8 and 23 inactivate E but not B. Such a pattern cannot be explained by the compounds differing in strength, or by the dermatophytes differing in tolerance. At this point, we are faced with a choice of strategies of data analysis.

E-mail address: phutchin@psy.mq.edu.au (T.P. Hutchinson).

doi:10.1016/j.ejmech.2003.11.006

Table 1
Inactivation (coded 1) or not (0) of seven dermatophytes (A–G) by 33 compounds at a concentration of 15.7 μg/ml. Eleven patterns were observed; code numbers are given in the first column. The total number of dermatophytes inactivated in each pattern is given. The final column shows which compounds gave rise to the pattern

Pattern	A	B	C	D	E	F	G	Total	Compounds
0	0	0	0	0	0	0	0	0	25, 26
1	0	1	0	0	0	0	0	1	17, 24
2	0	0	0	0	1	0	0	1	8, 23
3	1	1	0	0	0	0	0	2	6
4	1	1	0	1	0	0	0	3	9, 33
5	1	1	0	1	1	0	0	4	32
6	1	1	0	0	1	1	0	4	4
7	0	1	0	0	1	1	1	4	2
8	1	1	0	1	1	1	0	5	28
9	1	1	0	1	1	1	1	6	1, 7, 10, 12, 15, 16, 20, 22, 29, 30
10	1	1	1	1	1	1	1	7	3, 5, 11, 13, 14, 18, 19, 21, 27, 31

Codes 1–33 are used here for the compounds, though Roman numerals I–XXXIII were used by Matysiak et al. [1]. A–G refer to *Epidermophyton floccosum I, Epidermophyton floccosum II, Microsporum gypseum, Trichophyton galline, Trichophyton interdigitale, Trichophyton mentagrophytes*, and *Trichophyton rubrum*.

1. This is beyond what we are trying to explain. The reason might be measurement error. All we can do is to write an equation that captures the idea of random error, as in Section 3 below.
2. Alternatively, we might consider it as evidence that there are at least two systems in these dermatophytes that are vulnerable to these compounds, in other words, two modes of action of this set of compounds on this set of dermatophytes.
   The second alternative will be pursued in the present paper. (The patterns where each of the dermatophytes was inactivated or each was not inactivated, which will be labelled 10 and 0 in Table 1, are uninteresting in this context, and will be omitted from further consideration.)

In the study of what brain-injured patients can and cannot do, such a finding—that one patient succeeds in one task but not the other, whilst another patient succeeds in the second task but not the first—is termed "double dissociation", and is regarded as good evidence of the existence of different mental modules, one being damaged in one patient and the other in the second patient (for example, Ref. [4]).

3. Notation, and a theory having a single system

The format of data taken to be the starting point will be classifications of each compound as either inactivating or not inactivating each of the dermatophytes. Thus in principle the table of data has seven columns and 33 rows, with each entry being a 1 or a 0. Only 11 of the 128 possible response patterns were in fact observed, and thus the data can be presented more concisely, see Table 1. The data are from Table 2 of [1], and refer to a concentration of 15.7 μg/ml. (It is clear from Ref. [1: p. 399, 401] that this is an appropriate concentration to consider.) It will be convenient to refer to compounds by the numbers 0–10 given in Table 1, e.g. pattern 2 will be shorthand for compounds 8 and 23. (In some datasets, there is more detailed information, e.g. minimal inhibition concentrations may be given, as indeed they were in Table 2 of [1]. How to take advantage of this is a question for the future.)

Starting from the idea of a "tolerance" of an organism and a "strength" of a compound, we might proceed as follows. Let a_i be the strength of compound i and t_j be the tolerance of dermatophyte j. Then the probability P_{ij} of inactivation of dermatophyte j by compound i is found via $\ln[P_{ij}/(1 - P_{ij})] = a_i - t_j$ (where ln means natural logarithm). This model has found wide application. In educational testing, it is termed the Rasch model (for example, Ref. [5]) for predicting the probability of examinee i successfully answering item j, from the ability of the examinee and the difficulty of the item. However, it seems too simple in the present context: it presupposes that strength and tolerance are single numbers, there are not different types of strength or tolerance.

Section 4 below will develop an alternative. Each compound will have two strengths, not one, and each dermatophyte will correspondingly have two tolerances, not one. This is adapted from a proposal for analysing datasets showing which (brain-injured) patients succeed in which task [6]. But first, to introduce the notation that will be used, results for a single-system model will now be presented.

All we know about the system is the order of the sensitivities of the dermatophytes to interference with the system. In effect, a system is defined by an ordered list of dermatophytes. The convention that will be followed is that the most sensitive (least tolerant) dermatophytes are listed on the left, and the least sensitive (most tolerant) on the right. If one dermatophyte is inactivated, so are all those to its left in the list, because they are more sensitive. Naturally, the most successful single-system theories arrange dermatophytes roughly in order of the number of compounds that inactivate them. Equally best is this order:

B A E D F G C

This theory predicts that increasing strength of attack will lead to inactivation of first B, then both B and A, then

successively E, D, F, G, and C will be added to the list of dermatophytes inactivated. Are any wrong predictions made? Yes—for example, inactivation of E alone is observed (pattern 2, i.e. compounds 8 and 23), but this theory says that if E is inactivated, so also are B and A.

The compounds can be included in the same list. It is convenient to identify them by the pattern number in Table 1. (There are 11 of these, as compared with 33 compounds.) A compound is stronger than a dermatophyte (and thus inactivates it) if the compound is to the right of the dermatophyte. Naturally, the compounds appear roughly in order of the number of dermatophytes that they inactivate:

2 B 1 A 3,4 E 6 D 5 F 8 G 9 C

To identify which dermatophytes are inactivated by any particular compound, we simply find which are to the left of it in the list.

How many wrong predictions are made by this theory that has only one type of strength and one type of tolerance? In the case of three patterns (five compounds), there is a single wrong prediction: for example, pattern 6 (i.e. compound 4) is predicted to inactivate A, B, and E, whereas in fact it inactivates A, B, E, and F; and in the case of one compound, there are two wrong predictions.

4. Theories having two systems

4.1. Method of deriving the theory from the data

The systems are deduced one dermatophyte at a time.

(1) Patterns having exactly one dermatophyte inactivated. In pattern 1 (Table 1), only B is inactivated, and in pattern 2, only E. Therefore, at least two systems are required, B being most sensitive for one and E being most sensitive for the other.

(2) Patterns having exactly two dermatophytes inactivated. A and B was the only pattern of this kind observed, suggesting that the theory so far should be:

AB

E

For example, if the second system is not damaged, first B alone and then both A and B are inactivated by damaging the first system.

(3) Patterns having exactly three dermatophytes inactivated. A, B, D was the only pattern of this kind observed, suggesting that the theory so far should be

BAD

E

(Actually, at this point, the second system might be E B or even E B D, E A B D, etc. Most of these possibilities, but not E B, are eliminated at step 4.)

(4) Patterns having exactly four dermatophytes inactivated. There were three of these. A, B, D, E requires no further elaboration of the theory, being predicted by what we already have. A, B, E, F requires F to be added after E; and B, E, F, G requires G to be added after F. Thus the theory so far is:

BAD

EFG

(5) Patterns having exactly five or exactly six dermatophytes inactivated. Patterns A, B, D, E, F and A, B, D, E, F, G were the only ones observed, and they require no further elaboration of the theory, as they are consistent with what we already have.

This procedure is satisfactory with a small dataset. But, as far as I know, exhaustive consideration of all possibilities (by computer) is the only way of guaranteeing that the best theory is found.

4.2. Results for data of Matysiak et al.

It is now possible to make a succinct statement of the theory that has been obtained. There are two vulnerable systems, which are:

- B A D (C E F G)
- E F G (A B C D)
- Both systems are essential.

(Brackets indicate that we do not know whether these dermatophytes need to be listed, or in what order.)

According to this theory, increasing strength of attack on the first system will lead to inactivation of first B, then successively A and D become inactivated also; increasing strength of attack on the second system will lead to inactivation of first E, then successively F and G also; attack on both systems will lead to such patterns as inactivation of B and E, of B and E and F, and so on. Patterns such as inactivation of A alone, or inactivation of D and F and no others, will never be observed.

Ordered lists of strengths of the patterns (i.e. the 33 compounds) accompany the ordered lists of dermatophytes, and may be shown as below:

- 2 B 1,7 A 3,6 D 4,5,8,9 ...
- 1,3,4 E 2,5 F 6,8 G 7,9 ...
- Both systems are essential.

To identify which dermatophytes are inactivated by any particular compound, we find which are to the left of it in *either* list. For example, compounds within pattern 6 inactivate A, B, and (through attack on a different system) E and F. On comparison with the data (Table 1), all the predictions are found to be correct.

The following minor variation of the theory is also compatible with the data:

- B A D ...
- E B F G ...
- Both systems are essential.

(There is no objection to dermatophytes appearing in both lists, as B does here.)

4.3. Comments on theory

This style of theory has several distinctive features, some being advantages and some being disadvantages.

1. The theory tries to follow meekly, wherever the pattern of 0s and 1s leads. There is no input from preconceptions about which dermatophytes resemble which others most closely, or which compounds resemble which others most closely. Indeed, because of the passive nature of the treatment of data, some people might say this is not development of theory, but is description or condensation of data.
2. Any attempt to understand the orderings of dermatophytes and the orderings of compounds will come later. Because nothing we know about the dermatophytes has yet been used, it would be impressive if their orderings made sense in terms of their characteristics; and similarly for the compounds.
3. There is no random element, no means of explaining any divergence of data from theory. And as there is no modelling of randomness, there cannot be any statistical testing.
4. Binary data (i.e. inactivated vs. not inactivated) is the starting point. It is not clear how to take advantage if greater level of detail is available. For example, Ref. [1] gives the minimal inhibition concentration, not merely whether or not inactivation occurred at a single concentration.
5. A theory is judged successful because the observed patterns of inactivation are consistent with it. But what of the non-observation of patterns that are predicted to be possible (for example, inactivation of B and E only)? This is not a failure of the theory, because 33 is sufficiently few compounds and it is expected that many possible patterns will not be observed.

4.4. What if either system is sufficient?

So far, it has been assumed that both systems are needed in order for the organism to remain active, i.e. failure of either system causes inactivation. But what if either system is sufficient for the organism to avoid inactivation, i.e. only failure of both systems leads to inactivation? One possible theory of this type is the following:

- (E F G) B A D C
- (A B C D) E F G
- Either system is sufficient.

(Brackets indicate that the order of the dermatophytes is not known.) The dermatophytes inactivated are those that are to the left of the relevant strengths of attack in *both* lists. In other words, a dermatophyte is not inactivated if it is to the right of the relevant strengths of attack in either list. Thus inactivation of E alone is possible—when the first system survives in B, A, D, and C, and the second system survives in F and G. Similarly, inactivation of B alone is possible—when the first system survives in A, D, and C, and the second system survives in E, F, and G. And, indeed, all the observed patterns of inactivation (Table 1) are compatible with this theory.

The method of derivation of theories of this type is by working from the right hand end, and focussing attention on dermatophytes that are not inactivated, rather than those that are. There is only one pattern that has only a single dermatophyte not inactivated, namely, C. Therefore, C must appear in the rightmost position for at least one system. There is only one pattern that has exactly two dermatophytes not inactivated, namely, C and G. Therefore, G and C are in the two rightmost positions for one system, or respectively, in the rightmost position for one system and the other. And so on.

5. General discussion

Experimental data showing how effective a number of compounds are against a number of microorganisms are often presented in tabular form. When examining such tables, several features are noticed successively: the average or typical level of effectiveness of this class of compounds against this list of microorganisms; which compounds are most effective and which are least effective; which microorganisms are most tolerant and which are least tolerant; and whether some compounds are especially effective against one set of microorganisms, and other compounds against a different set of microorganisms. It is the last of these that is the subject of this paper. Information was given at the end of Section 3 about the extent to which a single-system theory fails to fit the data. With the present dataset, a two-system theory gives an error-free fit. Whether this is a worthwhile improvement, whether it is plausible that there are two systems or modes of action, and whether a "both systems are essential" or an "either system is sufficient" theory is the more plausible—these are matters for debate by experts on these particular compounds and dermatophytes.

The chief concern in this paper has been to obtain two ordered lists of dermatophytes as a summary or condensation of the observed patterns of inactivation. What these lists actually are is another matter for experts—is it appropriate to call them two systems, or two processes, or two mechanisms of action? Taking an empirical approach to this, we might ask why some compounds are stronger than others are. Is there something special about compounds 8 and 23 (i.e. those that have been referred to as pattern 2) that make them ineffective against the first system, and about compounds 17, 24, 6, 9, 33 (i.e. patterns 1, 3, 4) that make them ineffective against the second system? We might examine if the order of the compounds in the first list is correlated with various chemical or physical properties, and similarly if the order of the compounds in the second list is correlated with different properties. For example, it would be very interesting if one but not the other correlated with lipophilicity of the compound. Actually, for the present dataset, that is not the case; instead, the lipophilicity parameter R_{Mw} (given in Table 2 of [1]) is

moderately correlated with the rank positions of the compound in the first system and in the second: the correlations are 0.40 and 0.57.

It is possible that the results obtained here are an over-interpretation of the data, that the distinction between dermatophytes B, A, D (on the one hand) and E, F, G (on the other) is illusory—the appearance of a distinction may have resulted from random measurement error or from differences between the dermatophytes that are of no wider importance. Whatever might be said about the results with this specific dataset, I hope the method of analysis may prove useful with other data. Many papers having data of this form could summarise it in the form of ordered lists of organisms and compounds, as in Section 4.2. (Such datasets are reasonably common—of the 13 papers in the September and October 2003 issues of *European Journal of Medicinal Chemistry*, six have data showing the effect of several compounds on several organisms.) For the data in Table 1, it has been possible to infer a good deal about what the systems might be, but not everything: the first three dermatophytes in each system were identified, but the order of the less sensitive ones is not known. This has been practicable without resort to a computer. Other datasets might require sophisticated software or much computing power in order to find the best-fitting systems.

There has been discussion in Section 4.3 of the limitations of this style of theory. An intriguing question for the future is whether the model can be extended in such a way as to incorporate (firstly) the additional information given by minimal inhibition concentrations (additional beyond the binary classification as inactivated or not, that is), and (secondly) random or uninteresting variation, thus opening up the possibility of statistical testing.

References

[1] J. Matysiak, A. Niewiadomy, G. Macik-Niewiadomy, T. Kornillowicz, European Journal of Medicinal Chemistry 35 (2000) 393–404.
[2] A. Niewiadomy, J. Matysiak, A. Zabinzka, J.K. Rozylo, B. Senczyna, K. Jozwiak, Journal of Chromatography A 828 (1998) 431–438.
[3] J. Matysiak, A. Niewiadomy, G. Macik-Niewiadomy, European Journal of Pharmaceutical Sciences 10 (2000) 119–123.
[4] M. Coltheart, in: B. Rapp (Ed.), Handbook of Cognitive Neuropsychology, Psychology Press, Hove, 2001, pp. 3–21.
[5] W.M. Yen, in: M.C. Alkin (Ed.), Encyclopedia of Educational Research, sixth ed., vol. 2, Macmillan, New York, 1992, pp. 657–667.
[6] T.P. Hutchinson, Language and Cognitive Processes 18 (2003) 165–174.