

## Has Nomenclature Gone to Pot?

R. SCHOENFELD

Editorial and Publications Service, CSIRO, East Melbourne, Victoria 3002, Australia

Received November 26, 1979

The chaotic state of chemical nomenclature, as printed in the current literature, is exemplified in the case of cannabinoid products. A plea is made for stricter enforcement of the official rules by editors of primary journals. Flaws in the prevailing cannabinoid nomenclature are explained. Remedies are suggested, both in terms of systematic nomenclature and in terms of the rules on the naming of natural products recently promulgated by IUPAC.

If traffic in chemical nomenclature is in a chaotic state, this is not due to a lack of signposts. The International Union of Pure and Applied Chemistry and the International Union of Biochemistry have entrusted several committees with the task of unifying and codifying the laws whereby we chemists are expected to manipulate our huge vocabulary without ambiguity. These committees have distinguished members, who take their task very seriously; the rules they publish are models of clarity and commonsense. But the success of every legislature depends not only on the collective wisdom of the legislators; what is still needed is the willing cooperation of the average citizen and the efficiency of the cop on the beat.

### MEANWHILE, DOWN AT THE PRECINCT...

I am such a cop. Together with my colleagues on the editorial staff of the *Australian Journal of Chemistry* I am entrusted with the task of arresting and deporting every illegal chemical name that tries to take up residence in our precinct. Cops are seldom popular, and the reader may well assume that we are subject to much abuse from the sponsors of such illicit immigrants. Not so; we enjoy the support of our community because, instead of sternly presenting each offender with a long ticket listing all his infringements, we discuss the law with him and suggest remedial action. We find that citizens of our precinct are often ignorant of the law, but willing to learn and eager to obey it, even if it means having a few pages of their manuscript retyped.

Most of the really difficult crime we have to deal with is, I am sorry to say, imported from other precincts. Very few chemical journals, these days, send their editors out to police the nomenclature beat. Tracking down incorrect names and substituting lawful ones can be a time-consuming business, and many (indeed most) scientific publishing houses feel that they cannot afford a large enough police force. They thus allow incorrect names to appear on their pages, whence they flow into the body of *Chemical Abstracts*, and console themselves with the thought that the crime will not go forever unpunished. In fact, when the ill-named compound reaches the *Chemical Abstracts* Subject Index, the team of nomenclatural supercops headed by Dr. K. Loening will descend on it, strip off the criminal name, and insert the lawful name instead.

But not all chemists consult (although all should) the Subject Index in order to perfect their nomenclature; also, this Index appears, perforce, after a certain delay. In the meantime the literature is polluted with much counterfeit nomenclature, accepted in good faith by unwitting citizens. It is for this reason that we feel we ought to patrol our own precinct. We have not found the additional cost to be excessive. It is true that, during the first three years or so, the patrol was only kept going by the frantic oxidation of midnight oil. But, as our nomenclatural ignorance diminished and the skill of our authors grew, our fuel bills came down. These days, the pursuit

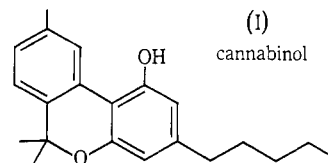
of illicit nomenclature takes up a manageable 5% of editorial time. Now, editorial time amounts to, at most, 30% of a publisher's budget; and the cost of publishing a paper is, perhaps, 1-2% of the cost of doing the research for it. Hence, is there really any economic excuse for willingly allowing nonsense names to appear in chemical publications?

### MAKING A HASH OF "HASH"

This is as much preamble as a writer of detective stories may permit himself. Having introduced myself as a cop, I owe my readers a fast-paced yarn, and there is no better background for a cops-and-robbers story than the drug scene. So let me tell you of my recent encounter with that potent principle (or rather, with its name) of hashish and marijuana, the compound tetrahydrocannabinol. Whether marijuana, in spite of its wide acceptance, is a social evil is the subject of much debate; but I hope to demonstrate beyond debate that the present nomenclature of its constituents, in spite of its wide acceptance, is totally nefarious.

Tetrahydrocannabinol, and its abbreviation THC, is nowadays a household word. It may appear in the daily press and in news broadcasts; announcers, with amusing frequency, pronounce "cannabinol" as "cannibal". In the scientific literature, this topical compound is named  $(-)-\Delta^1$ -tetrahydrocannabinol,  $(-)-\Delta^1$ -THC,  $(-)-\Delta^9$ -tetrahydrocannabinol, or  $(-)-\Delta^9$ -THC. All four names refer to the same compound, and the number of names is raised to eight by those chemists who, believing the stereochemistry of the molecule to be implicit in the term tetrahydrocannabinol, omit the initial minus sign.

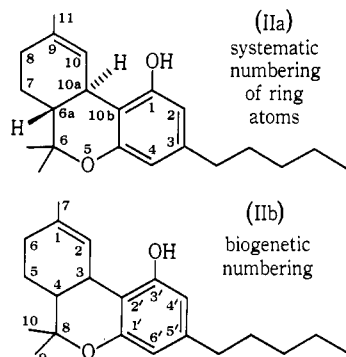
Eight names, and all are wrong. To arrive at the source of the present confusion, we must go back to the compound which serves as base for the nomenclature of all constituents of *Cannabis sativa*. This is cannabinol (I), abbreviated by some, horror of horrors, to CBN.



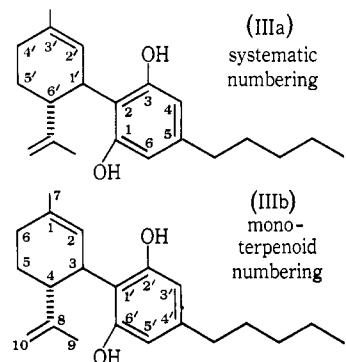
Cannabinol owes its predominant nomenclatural status mostly to historical accident. The compound is physiologically inactive, and may not be present in fresh *Cannabis* extracts at all. But early investigators did all their work on resins obtained, after long delay, from police or customs officials, and one of their principal techniques of isolation was distillation under moderate vacuum. It is thus not hard to see why cannabinol received attention sooner than its tetrahydrogenated cousin.

Most of the structure of cannabinol was established in the thirties (like a good crime writer, I shall mention the person

principally responsible only at the end of this article), and thereafter the names of all other tricyclic constituents were referred to "cannabinol". This was done much in the same way in which early investigators of steroids referred all names to "cholesterol". However, the steroid chemists soon saw the error of their ways and evolved a rational nomenclature whose base compounds were a group of hydrocarbons such as  $5\alpha$ -cholestane. Cannabis chemists, however, persisted with their names, and soon confusion was further confounded when their ranks were split by a schism as far as numbering the atoms of the molecular skeleton was concerned. It is time to call the villain, tetrahydrocannabinol, onstage. It, and the two systems of numbering, are depicted in structures IIa and IIb.



The ring atoms (*only* the ring atoms) of structure IIa are numbered according to sane systematic nomenclature (IUPAC Rules A-22 and B-3.4). The numbers are those shown in "The Ring Index" (and its successor, the "Parent Compounds Handbook") for 6*H*-dibenzo[*b,d*]pyran. The numbering of IIb follows biogenetic principles; tetrahydrocannabinol may be considered as having arisen from a monoterpene and a substituted resorcinol, and indeed it is accompanied in the plant by compounds such as "cannabidiol" (III), in which the pyran ring is either "not yet formed" or "split open". Monoterpene-plus-phenolic numbering was used for cannabidiol, and this numbering was then transferred to tetrahydrocannabinol. This had the advantage that, say, in biogenetic experiments with isotopes, carbon 1 in structure IIb corresponded to carbon 1 in structure IIIb.



Thus both groups of schismatics, those favoring IIa and those addicted to IIb, had logical reasons for their preference, but the overall result of the schism, which divides cannabis chemists about evenly, has been totally disastrous. The reader will now understand why the same compound is sometimes prefixed by  $\Delta^1$  and sometimes by  $\Delta^9$ . He will not be amused to learn that in a recent, otherwise excellent, monograph on cannabis half the chapters were written by members of the IIa orthodoxy and half by IIb dissenters, so that the one compound scored two berths in the index, quite without cross-reference. Perhaps the reader will grieve with me when he learns that a journal of high prestige, edited by colleagues of mine whose scholarship is impeccable, recently found itself

publishing a "IIa" and a "IIb" article side by side!

### WAS HE PUSHED?

Our cop story has now progressed to the point where a corpse has definitely been discovered. Next comes the question, "Was it accident or crime?"

Here there can be no doubt. Even a rookie cop can spot without trouble the following evidence of nomenclatural unlawfulness:

(i) Any system of trivial nomenclature must be based on the name of a nonfunctional structure, not on that of a phenol. Nonobservance of this rule has recently led to abominations such as "oxotetrahydrocannabinol", which contravene the fundamental IUPAC rule of senior functions (the keto group, being senior to the hydroxyl, should be named as suffix).

(ii) Suppose we forget about (i) for a moment, and admit cannabinol as our base name. This molecule is achiral. The addition of four hydrogens in going from I to II creates two chiral centers, whose geometry is not expressed by the prefix "tetrahydro". Some authors have meticulously referred to the natural, levorotary, compound by means of the descriptor (-). Others, as we have seen, have cheerfully assumed that tetrahydrocannabinol always designates the "natural" drug. Thus there is now much uncertainty whether work recently described in the literature has been performed on the racemic compound or the stereoisomer.

(iii) The symbol  $\Delta$  is no longer used in modern nomenclature to express unsaturation, except in special circumstances (e.g., IUPAC Rule C-55) which have nothing to do with the present case.

(iv) Let us forget about point iii for a moment and allow  $\Delta$  to function as a symbol of unsaturation. That brings an even more hideous crime to light. The fact that unsaturation exists between carbons 9 and 10 (in the numbering of IIa) or between 1 and 2 of IIb has already been expressed by our definition of cannabinol. By using  $\Delta$  signs, the cannabis chemist is thus expressing one feature of a molecule twice. To the serious nomenclator, that is a crime as dreadful as seeing  $\text{PhCH}_2\text{Cl}$  called *chlorobenzyl chloride*.

[Naturally, there is a reason why the cannabis chemist has tried to recall  $\Delta$  from retirement. Without it, structure IIa would have had to be called 6a,7,8,10a-tetrahydrocannabinol, and the four locant numerals take up much space. But that is the cannabis chemist's own fault: he should have, as we shall see later, started from a different base name.]

### THE CRYPTIC CLUES

A case has thus been made out against the name tetrahydrocannabinol. Before discussing remedial action, I should also like to present my case against the barbarous abbreviations THC and CBN, and their bastard offspring. I have already vituperated elsewhere<sup>1</sup> against the use of capitalized acronyms to express chemical structures. I should like to add the following specific charges:

(i) THC as part of a chemical formula should mean only one thing: a methylene group substituted by tritium. Admittedly most chemists, at the time I write, do not seem much interested in  $\text{CH}^3\text{H}$ . But by the time this article appears, tritium nuclear magnetic resonance may already be in fairly general use. Cannabis chemists will say that they could never err in recognizing the symbol THC. Would they still be so sure if they saw, out of context, the symbol  $\text{THC}^+$ ?

(ii) The acronymic notation has given rise to two contradictory bastard notations, which often inhabit the same article. One is spelled-out prefix plus THC, the other atomic symbols plus THC. Thus recently, in one paragraph, have I seen (the IIa numbering is used) 11-OH- $\Delta^9$ -THC, 8 $\alpha$ ,11-di-OH- $\Delta^9$ -THC, and 11-oxo- $\Delta^9$ -THC.

Suppose someone were to prepare, from the recently described 11-Br- $\Delta^9$ -THC, the acid 11-CO<sub>2</sub>H- $\Delta^9$ -THC. It would take a mind reader to distinguish this name from that of the lower homolog, the known  $\Delta^9$ -THC-11-oic acid. Moreover, in such a name, the letter C would first mean "carbon" and then "cannabinol". Is this a system of communication worthy of scientists?

### HOW TO REFORM THC PUSHERS

I believe I have proved that a crime has been committed. Being a modern, socially conscious cop, I have not pointed an accusing finger at the numerous criminals; this article will be noteworthy for the scarcity of its references. I am much more interested in remedy than in punishment. How do we best get out of this mess?

It seems to me that, first of all, we must familiarize ourselves again with systematic nomenclature. The molecule we are considering is not particularly complex by modern standards. The merest beginner could pick up a Ring Index and find the ring system 6H-dibenzo[b,d]pyran. The rest is just a matter of routine substitution, and addition of hydrogens. The correct systematic name is listed by *Chemical Abstracts* under the parent heading 6H-dibenzo[b,d]pyran-1-ol and, transcribed into its "uninverted" form, reads as follows: (6a*R*-*trans*)-6a,7,8,10a-tetrahydro-6,6,9-trimethyl-3-pentyl-6H-dibenzo[b,d]pyran-1-ol.

It is obvious that, for anyone having access to a Ring Index or Parent Compounds Handbook, deriving this name is no trouble;<sup>4</sup> the only feature of the name worth explaining is the stereochemical descriptor. For asymmetric compounds whose absolute stereochemistry is known, *Chemical Abstracts* designates the first-cited chiral center (i.e., the one on the lowest numbered atom) as *R* or *S* according to the Cahn-Ingold-Prelog rules. All other centers are then named relative to the first. Under IUPAC rules the stereochemistry could, of course, also be expressed as 6a*R*,10a*R*.

Users of the illicit trivial nomenclature will object that, simple as the systematic name is, it is yet too long and cumbersome to be inserted regularly into running text. This is, of course, true, but the same problem is encountered in all fields of preparative and structural chemistry, and authors cope with this difficulty by talking of "the phenol (I)" or "the hydroxy derivative".

Thus the case for creating a trivial nomenclature of cannabinoids is not particularly strong, and all attempts at nomenclatural reform should begin with the realization that the systematic nomenclature exists, is not too forbiddingly difficult, and may not be ignored. At the *Australian Journal of Chemistry* we have adopted the policy of being, temporarily at least, tolerant of the present nomenclature in the Results and Discussion sections; but we insist that correct systematic names be given, either in parentheses or as footnotes, in the Experimental Part. If this practice were to become widespread, it is possible that the present unsound nomenclature might just fade away.

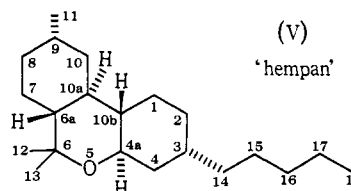
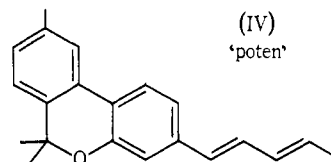
### LOOPHOLES IN THE LAW

But let us now assume that the craving of cannabis chemists for a trivial nomenclature of their own cannot be denied. If the patient cannot be cured, can we not at least substitute some less harmful drug?

Fortunately we can. Help for the suffering addict arrives in the form of Section F of the IUPAC rules for nomenclature of organic chemistry.<sup>2</sup> This has been specially devised for the benefit of natural-products chemists who may have to describe a series of closely related compounds of sufficient complexity to make systematic nomenclature cumbersome. Section F generalizes the technique that has worked so well in the field

of steroid chemistry. Each new "semisystematic" nomenclature is based on a "parent structure" which may not contain functional groups (thus cholestane becomes a parent structure rather than the hydroxyl-containing cholesterol) but incorporates as many as possible of the features (i.e., the bond skeleton and the geometry) that the family of natural products has in common. This parent structure should express "either the lowest or the highest possible state of hydrogenation". Saturated structures take the ending -ane when they are hydrocarbons (as in the case of steroids) and -an when they contain a heteroatom (as in the case of cannabinoids). The unsaturated parent compounds analogously end in -ene or -en.

A quick investigation of the possible options rules out the use of the "most unsaturated" form for the parent structure of cannabinoids. Such a parent structure would, of course, express the unsaturation of the phenolic ring and would yield a simple name for what is now cannabinol (I). But it does not express the geometry of the two chiral centers of II, and thus descriptors would have to be added to the name. Moreover, "maximum unsaturation" would have to extend into the pentyl side chain. Thus this would have to be either lopped off the parent structure (an action leading to the subsequent introduction of the prefix "pentyl") or changed to pentadienyl (so that four additional "hydro" atoms would have to be named). Thus we are unlikely ever to encounter in the literature a term such as my hypothetical "poten" (IV) (from *pot*, a popular term for a psychomimetic drug)—which may be just as well, because it could have been mistaken for a similarly spelled Irish beverage of remarkable mind-blowing potency.



A "most saturated" structure that could serve very well as parent for a new cannabinoid nomenclature is V, which I have named "hempan".<sup>5</sup> This expresses the geometry of the common cannabinoid molecules; the only nuisance is that the saturation of the phenolic ring creates four additional chiral centers. These have been arbitrarily designated in structure V; for all cannabinoid compounds known this designation is of no importance because in them the four centers have lost their asymmetry through the introduction of double bonds. If more saturated compounds were to be isolated in the future, and their symmetry were to differ from that of "hempan", this could easily be indicated by descriptors such as 4a $\beta$ H (IUPAC Rule F-6.3).

A semisystematic nomenclature based on structure V would be free of the errors in logic that taint present usage, and the names would still be acceptably simple. Thus the villain of our story, (-)- $\Delta^1$ -tetrahydrocannabinol, would become 1,3,4a(10b),9-hempatetraen-1-ol. The greater the change to the basic structure, the more would the new system outperform the old. Thus the semisystematic nomenclature could cope with such modifications as migration or disappearance of the phenol group, against which the "THC" system is helpless.

Like a benevolent cop who has preached on road safety to some offending motorist, I leave the cannabis chemists to ponder their misdeeds, and to think of remedies. Before closing the subject, however, a few words should still be said about the numbering of the ring system. I think nomenclatural reform should begin right away with the abandonment of the "biogenetic" numbering shown in IIb. This was instituted for the best of motives, to show the close link between cannabinol (I) and cannabidiol (III). But, unfortunately, the numbering of cannabidiol (shown as IIIb) that was transferred to cannabinol was itself wrong. As the "characteristic function" of that molecule (the phenolic hydroxyl) is on the aromatic ring, this ring becomes the parent into which all other parts of the molecule are substituted, and is thus entitled to unprimed numerals. The systematic name of cannabidiol is thus (1'*R*-*trans*)-2-[3-methyl-6-(1-methylethenyl)-2-cyclohexenyl]-5-pentyl-1,3-benzenediol. The systematic numbering is shown in structure IIIa.

Thus, the numbering of IIb, with all due respect to the eminent men still using it, is a compounding of errors and the sooner it disappears from the literature the better. It took a great deal of effort to compile first "The Ring Index" and then the "Parent Compound Handbook"; now that these admirable works are available, their names and numberings should be accepted.

A word of caution is required concerning the carbons of the side chains. In any semisystematic nomenclature, they may be given unprimed numbers running on from the last number of the ring atoms, as shown in the "hempan" structure. In systematic nomenclature, this is not permitted. Thus an "11-hydroxy" compound in the semisystematic nomenclature would become a "9-(hydroxymethyl)" compound in a paper giving systematic names.

The reader may now wonder whether my cop story has a happy or unhappy ending. The crime has been uncovered, and the miscreant duly cautioned. But will he heed the warning? Perhaps, among the powers possessed by cannabinoids to modify human behavior, there is also the power to make re-

search workers who stare at cannabinoid crystals impervious to the demands of sane nomenclature. Is being blown out of one's nomenclatural mind a professional risk?

Fortunately, I am able to reassure the reader. A very interesting case history is available in the files of the precinct station. In the early thirties, an English investigator performed work of particular distinction in determining the structure of cannabinol. Later, this man abandoned the work bench for an editor's desk and became a prominent member of the IUPAC Commission on Organic Nomenclature; he also wrote the best introductory text on the subject.<sup>3</sup> His name is R. S. Cahn, and he is well known to students as the author, with Ingold and Prelog, of the Sequence Rules without which the geometry of molecules could not be accurately described.

Here we have undisputable proof that work on *Cannabis* does not disqualify a chemist from thinking logically about nomenclature. May I, as a harassed cop, appeal to leading workers in the field for help?

## REFERENCES AND NOTES

- (1) R. Schoenfeld, "The Chemist's English. Part XVI", *Proc. R. Aust. Chem. Inst.*, **43**, 222 (1976).
- (2) IUPAC Commission on Nomenclature of Organic Chemistry, "Natural Products and Related Compounds", *Eur. J. Biochem.*, **86**, 1 (1978).
- (3) R. S. Cahn and O. C. Dermer, "Introduction to Chemical Nomenclature", 5th ed, Butterworths, London, 1979.
- (4) Here some nomenclatural dirty linen must be washed. IUPAC, in a fit of indecision, permitted "hydro" to be used, at the nomenclator's option, as a "detachable" prefix (i.e., it is classed among true substitution prefixes such as methyl and pentyl, and cited in IUPAC's alphabetical order) or as "nondetachable" prefix (i.e., it follows after all substitution prefixes have been listed). *Chemical Abstracts* follows the former system; the *Australian Journal of Chemistry*, for a multitude of logical and practical reasons, follows the latter. Thus my colleagues and I would print . . . 6,6,9-trimethyl-3-pentyl-6a,7,8,10a-tetrahydro-6*H*. . . . The reader will observe that this arrangement brings the "hydro" atoms together with the "indicated hydrogen" 6*H*. A recent IUPAC Information Bulletin (No. 54) contains the welcome news that the nondetachable listing of hydro will become mandatory at the next revision.
- (5) Names such as *poten* and *hempan* have only been coined to illustrate the argument. It would be presumptuous of me to make genuine suggestions; these should come from leading workers in the field.

## Cambridge Crystallographic Data Centre. V. An Integrated System of Printed Indexes

FRANK H. ALLEN

Crystallographic Data Centre, University Chemical Laboratory, Cambridge CB2 1EW, England

Received March 19, 1979

Computer programs have been developed for the generation of an integrated set of six printed indexes from the Bibliographic File of the Cambridge Crystallographic Data Centre. The most important information elements, i.e., compound name, molecular formula, authors names, journal reference, are chosen for specific inversion; this set of inversions is cross-linked by a multipoint index of variable information content. Traditional techniques are augmented by keyword-in-context and element-in-context layouts. The system has three main roles: as an in-house aid for file maintenance, as an adjunct to computerized bibliographic searches, and as a search tool in its own right.

## INTRODUCTION

The Cambridge Crystallographic Data Centre<sup>1</sup> maintains computer-based files of bibliographic,<sup>2</sup> chemical connectivity,<sup>1</sup> and numeric structural data<sup>3</sup> for organics, organometallics, and metal complexes studied by X-ray and neutron diffraction. The Centre is also responsible for dissemination of the data base: in machine-readable form, together with software for

search retrieval and display;<sup>1,4</sup> via traditional printed publications in the reference-book series "Molecular Structures and Dimensions" (MSD);<sup>5-8</sup> via a current awareness service.

At its inception in 1965 the Centre was faced with problems of file definition, system organization, and software development, while simultaneously assimilating both current and backlog input. The first priority was the establishment of a