

Too Broad Generic Disclosures: A Problem for All

JAMES. F. SIBLEY

Patents, Licensing and Trade Marks Department, Shell International Petroleum Co. Ltd., Shell Centre,
London SE1 7NA, England

Received September 14, 1990

The use of generic, or Markush, structures in patents is reviewed, and some examples of excessively broad Markush specifications are presented. The advantages and disadvantages of such claims to the patentee, the patent examiner, the database producer, and others in the scientific community are described.

It has been possible for some time to carry out chemical structure searches in the Chemical Abstracts Registry File for specific compounds described in both the open literature and patent specifications provided characterizing data were given. This database does not cover generic structures or "paper" chemicals. Now databases are being developed which use as input the full generic disclosures found in patent specifications. A typical generic structure is shown in Figure 1.

This structure is taken from claim 1 of GB 1599915. Varying values were given for all the different R groups, for X, Y, and *m*. As is fairly normal in such structures, Y could be absent altogether or there could be 1, 2, or 3 Y groups attached at any of the free positions on the phenyl ring. If there were more than one Y group, then they could be the same or different.

These generic formulas are often referred to as Markush structures or Markush formulas but this is, in fact, incorrect. In 1924, in US 1506316, Eugene Markush claimed a process for the manufacture of dyes which comprised coupling with a halogen-substituted pyrazolone, a diazotized unsulfonated material selected from the group consisting of aniline, homologues of aniline, and halogen substitution products of aniline.

The Markush-type claim, in which, in the absence of a suitable commonly accepted generic expression commensurate in scope with the monopoly desired, the applicant enumerates the different species that are to be grouped together into a genus, is largely a phenomenon of the United States patent system.

Generic formulas were used in patent specifications before 1924 not only in Europe but also in the U.S. Figure 2 shows claim 3 of US 1150580 issued in 1915 and assigned to Eli Lilly.

X₁ and X₂ were each alkyl or hydrogen, Y was an acid radical, and Z was a monovalent group capable of being introduced by means of the Grignard reaction.

In general, there can be no argument against the use of generic structures in patent specifications, and in most cases, they will not cause problems to those who have to process the structures for input to a generic database such as Markush Darc, produced by Derwent Publications in the U.K., or Marpat, produced by Chemical Abstracts in the U.S. The difficulties occur when the applicant makes excessively broad disclosures in an extravagant or convoluted way which is not readily comprehensible to the reader.

Klaus Goehring and I wrote a paper¹ drawing attention to these difficulties and Dr. Jenny, a European patent attorney working for Ciba-Geigy, has recently replied² defending the status quo, and the present talk is based on these papers. The matter was also discussed at a conference in Montreaux last year.

Jenny says that the number of patents involved is insignificant, but he bases this only on the small number of patent applications that the EPO and WIPO have been unable to search. Currently, Derwent is processing about six such specifications each week; specifications which are referred to by Derwent as "nasties" and "supernasties", that is about 300 specifications each year, and the number is rising. Of course, this is still a fairly small fraction of the total applications published in the chemical area, but that it happens at all degrades the retrieval systems used by both the public and the patent offices. It is interesting to note that, unlike Jenny, the President of the European Patent Office considers the problem sufficiently important to have set up a committee to study the matter.

The breadth of the disclosure is not usually the problem, more often it is the complexity and lack of coherence in the disclosure that causes the difficulties. Homological generics, while covering an infinite number of individual compounds are very easily accommodated.

For example, while Dr. Christ was concerned that the generic formula



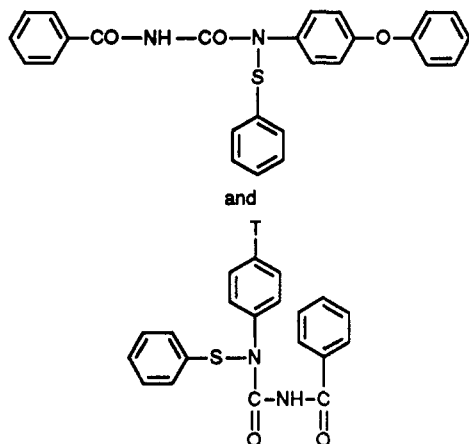
wherein "alk" covered groups containing up to 20 carbon atoms embraced 10⁴⁰ individual compounds³ such a disclosure would be very simple indeed to put into a generic database.

Jenny also states that "The simple fact of the matter is that these problems would not be solved even if—to the detriment of the inventor—only patents of quite narrow scope were granted". Patents issued in Japan, the USSR, and the countries of Eastern Europe have always had claims of narrow scope, and it is no coincidence that these specifications cause no difficulties to the documentalist. It is somewhat disappointing that there are signs that Japan is moving toward the broader claims seen in specifications published in the United States and Western Europe.

The computer-based structure systems have made a very great difference to the retrieval of chemical information, and experienced users will find amusing Jenny's assertion that the human brain is better than the computer at detecting generic/specific and generic/subgeneric connections.

The human brain far outstrips the computer in the area of serendipity, which plays a major part in patent searching, but when it comes to chemical structures, the human brain is often conditioned by the way in which the query was phrased or drawn.

This can be illustrated, albeit rather embarrassingly, by the failure of an experienced searcher to pick up the relationship between compounds containing the structures



wherein T could be, inter alia, a phenoxy group.

The computer on the other hand works only on connection tables, it does not process drawn structures, and it is quite incapable of being influenced by tiredness or boredom while involved in the search. The only mistakes it can make are those induced by incorrect input possibly due to the problems we are to discuss today.

There are two quotations which I would like you to bear in mind during the discussions this afternoon. They are as follow:

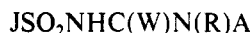
"Claims must define the boundaries of the invention precisely and clearly so that the public may know what they are prohibited from doing during the existence of the monopoly, and what they are to have at the end of the term as a consideration for the grant."

Secondly from Mr Sieders of AKZO N.V., who said "Any applicant should be granted a monopoly when he has inventively contributed to the state of the art and only in so far as he has actually done so. That is what he is entitled to and nothing more or less".

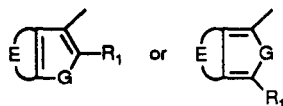
Mr. Sieders also made the point that there should be a reasonable relationship between the contribution to the art and the monopoly claimed although he accepted that it was not always easy to establish what a reasonable relationship would be.

We will now look at three of the specifications which have caused problems starting with US 4737184. We will concentrate only on claims 1 and 2.

Claim 1 is directed to a group of sulfonyleureas which have the general formula



This is a well-known formula which has appeared in very many specifications. In this case we will look first at the group J, a ring structure described in claim 1 in a way which I believe the human brain would find quite hard to decipher. The relevant part of claim 1 is shown below.



E is a bridge of 3 or 4 atoms, which may be substituted or unsubstituted, containing 0-2 heteroatoms selected from the group consisting of oxygen, sulfur, and nitrogen and also containing 1-4 atoms of carbon, said bridge together with two carbon attachment sites forming a partially saturated 5-6-membered carboxylic or heterocyclic ring; or E is a bridge of 3 or 4 atoms which may be substituted or unsubstituted containing at least 1 heteroatom selected from 0-1 oxygen or

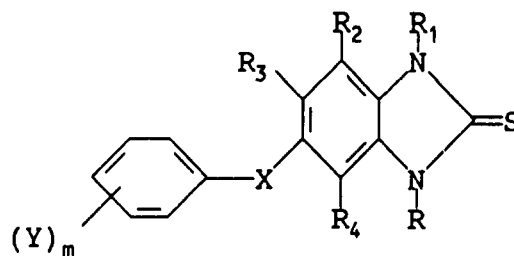


Figure 1.

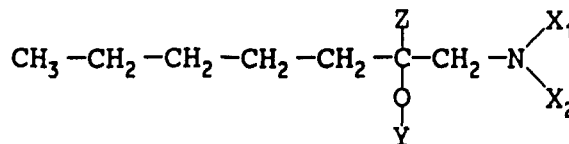


Figure 2.

sulfur or 0-2 nitrogen and 1-3 atoms of carbon, said bridge together with two carbon attachment sites forming an unsaturated 5-6-membered heterocyclic ring, with the proviso that when E contains two atoms of oxygen or sulfur, then they must be separated by at least one atom of carbon and that oxygen and sulfur are only linked to each other if the sulfur is in the form of SO or SO₂; in the bridging group E nitrogen may take the form of N or N-O, sulfur may take the form of S, SO, or SO₂, and one of the atoms of carbon may be a carbonyl, thiocarbonyl, or the cyclic 5- and 6-membered ketals thereof.

It contains two alternative definitions for "E", an essential part of ring J, which will broadly determine which rings are covered by J in the claim. Clearly, the first definition allows for carbocyclic rings or heterocyclic rings containing up to two heteroatoms and the ring must be partially saturated which will exclude aromatic rings. In the second definition for "E", carbocyclic rings are excluded for there must be at least one and at most three heteroatoms in the ring which must be unsaturated, but presumably not necessarily aromatic, and the selection of heteroatoms is restricted, the maximum being two nitrogen atoms and one oxygen or sulfur atom. Then there is a proviso for E which can only apply to the first type of ring because it deals first with the case when there are two atoms of oxygen or sulfur and secondly with case where both sulfur and oxygen are present, conditions which are excluded from the second type of ring. The claim contains other criteria regarding substituents and optional additional rings, and I would contend that the range of fused ring heterocyclic compounds covered by this formula is not easily derived without the expenditure of a considerable amount of time.

Clarification may be sought from claim 2 in which the applicant sets out some 166 ring formulas for J which he considers are covered by the general description in claim 1. These 166 formulas embrace no fewer than 612 different fused heterocyclic rings which will need to be processed for input to the database. Of course, the person doing the input must check that by using only the structures from claim 2, the full scope of claim 1 has been covered. In fact, despite the number of rings set out in claim 2, it does not, for example, cover the cases where E forms a 6-membered ring containing, in addition to the two bridge atoms, one carbon atom, two nitrogen atoms, and one oxygen or sulfur atom as provided for in the second definition. The input must also cover the possible ketal rings.

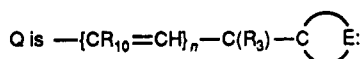
Now attention must be turned to the variables and substituents set out in both claim 1 and claim 2. The substituents can cascade one within another, thus R₆ can be, for example, -R₈, -SR₈, or -OR₈ where R₈ can, in some cases, be substituted once or twice with R₁₅, and R₁₅ in turn can be for example -OR₁₀ where R₁₀ can be a cycloalkyl or phenyl group substituted with R₁₄.

Then the person doing the input must check the provisos given in both claim 1 and claim 2 to see if they materially affect the input. In this case, the choice of ring A is restricted when the compound is a sulfonylthiourea.

Finally, attention must be given to the specification itself, does it disclose a wider group of compounds than the claims and the answer is yes. The possibility that ring A was 1,3,5-triazine was included in the specification but dropped during prosecution, and this must be included if the full scope of the disclosure is to be covered.

Let us turn to another patent specification also concerned with sulfonylureas. This time it is a European application No. 209232, and again we will concentrate on claims 1 and 2.

The general formula is the same as that given above in which ring J may be phenyl, thiophene, pyrazole, or pyridine all of which carry a substituent Q, which is defined as follows:

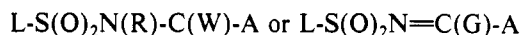


E is a bridge of 4 or 5 atoms, which may be substituted or unsubstituted, containing 0-2 heteroatoms selected from the group consisting of oxygen, sulfur, and nitrogen, and also containing 2-5 carbon atoms, said bridge together with one carbon attachment site forming a partially saturated or a fully unsaturated, nonaromatic 5- or 6-membered carbocyclic or heterocyclic ring, with the proviso that when E contains two atoms of oxygen or sulfur, they must be separated by at least one atom of carbon, and that oxygen and sulfur are only linked to each other if the sulfur is in the form SO or SO₂; in the bridging group E, sulfur may take the form of S, SO, or SO₂ and one or two of the carbon atoms may be a carbonyl, thiocarbonyl, or the cyclic 5- and 6-membered ketals thereof.

The definition of E is different but again some clarification is available from claim 2. Here are shown 80 ring structures which embrace 161 different individual rings. This time we can see easily that claim 2 covers only part of the totality of ring E for there are very few examples of rings with adjacent heteroatoms and even these are limited to those in which one of the heteroatoms is sulfur in the form of SO₂.

The database producer will need to be especially vigilant in this case for this time there are provisos which restrict the choice of ring A in the case of sulfonylthioureas, the types of ring when E contains a thiocarbonyl group and position of Q with certain rings J.

A sulfonamide case is our third example. This time it is US 4838925 in which claim 1 is directed to compounds having the structure:



where A can be any one of 19 different structures and L can be chosen from any one of 25 structures of which 12 are substituted by R₁. R₁ can be any one of 21 different groups, most of which can be further substituted. We are going to look only at the possibilities that R₁ can be -(CH₂)_nQ or -(CH₂)_nQ₁, for Q can be chosen from 27 different rings and Q₁ from 87 different rings.

We have seen three different variations of the proliferating ring syndrome. In the first, the variations were part of the immediate basic structure, in the second they were an essential substituent to the basic structure, while in the third they are an optional substituent to some, but by no means all, of the rings forming the basic structure. They will all need a great deal of time to analyze, compress into more generic formulas, and check before they can be input to the database. The manpower costs and the possibility of error are enormous.

These patents each claim a monopoly over many millions of individual compounds. Probably each one will cover more than 10²⁵ compounds. Now, no one objects to the applicant being able to obtain a monopoly more extensive than that

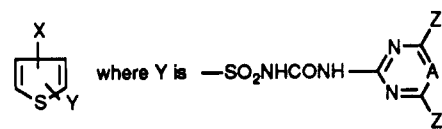
actually exemplified. But what would be an acceptable extrapolation?

In the first specification we looked at, the applicant has illustrated in claim 2, 612 possibilities for the ring J but only three have been made and tested. The generic formula in claim 1 allows the substituent R₁ to be selected from some 71 million chemically different groups of which just one, hydrogen, has been used. Similarly, both the X and Y attached to the pyrimidine ring may be chosen from over 370 000 different groups but only three, methyl, methoxy, and chlorine, have been made and tested. The situation is not better in the other two specifications we have looked at this afternoon. So, is there a reasonable relationship between the monopoly and the contribution to the art?

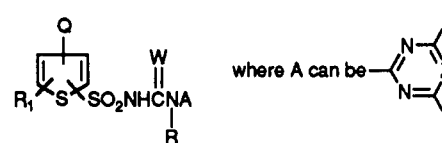
Another question might be...has there been any inventive contribution to the state of the art at all?

These are just three of a great number of patent specifications claiming herbicidal sulfonylureas and sulfonamides, each and every one of which has indicated that the presence of either of these groups attached to a nitrogen-containing heterocyclic group provides activity regardless of the other groups present. Each and every one of these specifications has covered many thousands if not many thousands of millions of individual compounds; so can we really be surprised when another thousand or another million very similar compounds have the same activity. Surely, these can at best be discoveries, an invention would be a nonherbicidal compound of the same structure.

The cascading of substituents poses problems for those who are studying infringement, for it is often difficult to decide when the compounds described in one specification fall within the scope of an earlier patent. For example, US 4822402 claims compounds of the formula



and wherein X is a halogenoalkoxyalkyl group. The claims are very modest, and not unexpectedly, the applicant is Japanese. I suggested to a colleague that such a claim falls within the scope of a slightly earlier sulfonylurea case EP 207609 which claimed



At first he could not see where the overlap occurred, but if in the European specification R₁ is hydrogen and Q is a C₁₋₄ alkyl substituted by R₈ and R₈ is OR₉ where R₉ is haloalkyl, then there must be at least some overlap.

I will turn now to the absolute classic generic specification, PCT Application No. 8704321 which is based on two U.S. priority applications. The specification is directed toward a process for reducing moisture loss from plants by the application of certain chemicals. The specification is 789 pages long and has 110 claims, of which claim 105 is directed toward novel compounds. The specification was, as far as I am aware, the first on which an international search authority, in this case the European Patent Office, refused to establish a search report.

We are going to look only at a part of claim 105. This claim is itself 53 pages long, and it embraces compounds which can have any one of 33 different generic formulas. We will concentrate only on structure 32. This claims as novel, compounds of the general formula

R_1 -X- R_{36}

R_1 is defined in two parts. First we are told

R_1 is a substituted or unsubstituted, carbocyclic or heterocyclic ring system selected from a monocyclic aromatic or nonaromatic ring system, a bicyclic aromatic or nonaromatic ring system, a polycyclic aromatic or nonaromatic ring system, and a bridged ring system which may be saturated or unsaturated

and then

or R_1 is a substituted heteroatom or substituted carbon atom, or a substituted or unsubstituted, branched or straight chain containing two or more carbon atoms or heteroatoms in any combination.

Each part of the definition is followed by a list of the permitted substituents, but we will look at these in a moment. It is enough at this stage to realize that R_1 can be any cyclic or acyclic organic grouping and probably any inorganic grouping.

The linking group X is defined as a covalent single bond or double bond, a substituted or unsubstituted heteroatom or substituted carbon atom, or a substituted or unsubstituted, branched or straight chain containing two or more carbon atoms or heteroatoms in any combination.

In other words, X can vary from a single covalent bond to a chain of assorted atoms stretching from Washington to Los Angeles with branches at every highway intersection in between, the branches spreading in all directions until they reach an ocean. So, after looking at the definitions of R_1 and X, we are unable to restrict in any way the compounds covered by this formula.

That leaves just R_{36} . Now R_{36} is a substituted or unsubstituted, asymmetrical heterocyclic ring system having at least three nitrogen atoms that is selected from a monocyclic aromatic or nonaromatic ring system, a bicyclic aromatic or nonaromatic ring system, a polycyclic aromatic or nonaromatic ring system, and a bridged ring system which may be saturated or unsaturated.

We can easily establish that any 5-membered ring containing three, four, or five nitrogen atoms only must be symmetrical so that they fall outside the definition of R_{36} . Of the 6-membered rings, only 1,2,4-triazine is asymmetrical and falls within the definition of R_{36} . But what about 5- and 6-membered rings with three nitrogen atoms and another heteroatom, such as 1,2,3,4-thiadiazole, 1,2,3,5-thiadiazole, 1,2,3,4-oxadiazole, 1,2,3,5-oxadiazole, and 1,2,3,6-oxadiazole; these are all asymmetrical so they are in. Once one gets into fused ring systems, suitable asymmetric groups abound, for example, 2H-pyrazolo[3,4-b]pyridine, 3H-imidazo[4,5-b]pyridine, 1H-pyrrolo[2,3-d]pyridazine, pyrazolo[3,4-b][1,4]oxazine, and so on. So formula 32 covers any substituted suitable heterocyclic ring such as those already mentioned in which the ring, any linking group present, and the substituent can be further substituted by any one or more of the substituents listed.

We ought to look at the list of substituents, for having gotten this on their computer, the applicants have used it not only 71 times in this specification but also in at least three other PCT applications.

We need to look only at the substituents listed for R_{36} because the other listings are identical. Each listing covers some two pages and contains a number of specific substituents such as hydrogen, hydroxy, amino, nitro, and cyano and a number of generic groupings such as alkyl, alkoxy, halogen, and acyloxy. In addition, the substituent or substituents may be -X, =X, -X= R_3 , =X- R_3 , or -X- R_3 .

We already know that X may be a single or double covalent bond, or an optionally substituted heteroatom or substituted

Table I

	-X =X	-X- R_3 =X- R_3
-Cl	-Cl	-Cl
-OH	-OH	-OH
-NH ₂	-NH ₂	-NH ₂
-OCH ₃	-OCH ₃	-OCH ₃
-SCH ₃	-SCH ₃	-SCH ₃
-SPh	-SPh	-SPh
-CH ₃	-CH ₃	-CH ₃
-C ₂ H ₅	-C ₂ H ₅	-C ₂ H ₅
-CN	-CN	-CN
-NCO	-NCO	-NCO
-NO ₂	-NO ₂	-NO ₂
-C(O)NH ₂	-C(O)NH ₂	-C(O)NH ₂
-C(O)CH ₃	-C(O)CH ₃	-C(O)CH ₃
-S(O) ₂ CH	-S(O) ₂ CH ₃	-S(O) ₂ CH ₃
=N-OH	=N-OH	=N-OH
=O	=O	

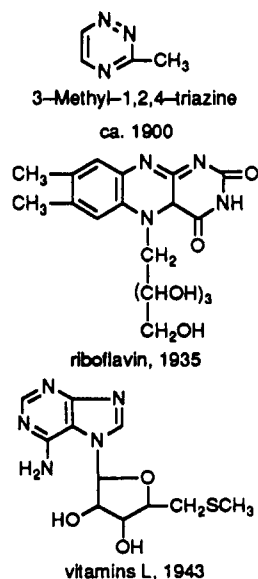
carbon atom, or an unrestricted chain of atoms, and we can easily determine that R_3 is the same as R_1 . So we can set out possible substituents in a table. (See Table I.)

First we have a selection of the groups actually set out in the listing, for example chloro, hydroxy, amino, methoxy, thiophenyl, methyl, nitrile, carboxyamino, hydroxyimino, and oxo. Then we know that the substituent can be -X or =X and from our knowledge of X we can see that we have a new set of substituents which includes chloro, hydroxy, amino, etc. Finally we must look at the possibility that the substituent can have the form -X- R_3 and =X- R_3 . We know what R_3 can be, and bearing in mind that X can be a single or double bond and therefore disappear, we have a third set of substituents including chloro, hydroxy, amino, etc.

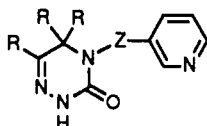
As I understand it, the U.S. Patent Office is unwilling to accept the phrase "optionally substituted" because it is open ended and because if a group is truly optional, it cannot be part of the invention. If this is the case, perhaps someone will spend a little time explaining how the travesty we have just looked at is more restricted since it covers not once but at least twice every substituent group known to man and then some. The only substituent listed once is hydrogen.

But, since there are no provisos concerning the nature of the substituents, they can all be hydrogen, and so the substitution is "optional".

Among the compounds falling within this part of claim 105 that the applicant considered novel in 1986 are the following:



But what if it were granted in its present form? What subsequent specifications would fall within the scope of this small part of claim 105. I glanced through the abstracts in a single class of the Derwent Manual Code and found some 40-odd patents filed after the publication date of PCT 8704321 that would infringe. One, for example, is EP 314615 which claims compounds of the formula



where Z is an aminomethylene or methanimino group. It is interesting that the EPO did not cite the PCT specification when carrying out the search on this European application, for there are some quite interesting pyridyl-substituted 1,2,4-triazines set out in tables in the text.

There are probably just two reasons why specifications

appear with broad complex claims such as we have discussed this afternoon. The first is that the technical area is subject to much research and that the applicant must draft the claims carefully to prevent overlap. The second, is greed.

What must the database producers do with such specifications? Do they try to process them, or do they just throw them into a garbage bin? What are patent offices to do? Do they discount them and hope that they will go away, do they spend time processing them, or just give in and grant them? What do patent attorneys do? Do they carry on pretending that the problems caused by such specifications have nothing to do with them and that they must retain the freedom to continue in the same way? Lastly what are the searchers to do?

REFERENCES AND NOTES

- (1) Goehring, K.; Sibley, J. F. *World Pat. Inf.* **1989**, 11 (1), 50.
- (2) Jenny, F. A. *World Pat. Inf.* **1990**, 12 (2), 71-75.
- (3) Christ, H. *Mitt. Dtsch. Patentanwalte* **1987**, 78, 121-135.

Very Broad Markush Claims;¹ A Solution or a Problem? Proceedings of a Round-Table Discussion Held on August 29, 1990[†]

G. W. A. MILNE

National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892

Received September 30, 1990

INTRODUCTION

On August 29, 1990, during the National Meeting of the American Chemical Society in Washington, DC, a symposium was held to consider very broad Markush claims. Two formal papers were presented, dealing with the advantages and disadvantages of broad generic claims, and these were followed by an open, round-table discussion. The two papers, by Sibley² and by Brown³ are published immediately prior to this article, and there follows a transcript of the round-table discussion. The transcript has been edited but not materially changed or abridged.

Gerry Vander Stouw: Could we take places and come to order please? On behalf of the Divisions of Chemical Information and Chemistry and the Law, I would like to welcome you to our final technical program to be held in this room this week, not, however, our final technical program because we do have a cosponsored session on risk assessment tomorrow in room four in this hotel. That will be an all-day session on risk assessment. At this time, I would like to turn the podium over to Mike Dixon, who will be chairing this program on very broad generic structures and Mike, I think, will introduce the rest of the program and also perhaps explain to you how it is all going to work. Mike.

Dixon: Gerry, thank you. Good afternoon and welcome to this Symposium entitled Very Broad Markush Claims; A Solution or a Problem? This morning, those of you who were at the sessions organized by Mike Feider of Dow Chemical, if you are the same crowd, my compliments to you on your perseverance, you heard what Markush claims were from the point of view of search and retrieval. What we are going to hear this afternoon is, is it all worthwhile? There are inherent

problems depending upon which side of the table one is sitting upon with regard to Markush formulas.¹ There are those, dare I suggest, from industry who wish to hide in the Markush claim all sorts of valuable information and put a sort of smoke screen around the particular compound for which they wish to obtain a patent. On the other side of the table, there is the competition, the chemists and the patent examiners in the Patent Offices who really need to be able to test validity and to find the needle in the proverbial haystack.

In between, there are information scientists who like to play with these compounds and invent all sorts of mystical problems around them and call them things like "nasties". We will hear about those, I am sure, this afternoon. Our format will be slightly different from that of the usual program. We will begin with two formal papers, and then we will have an open discussion which will be lead by various people, panelists here, who I will introduce subsequently.

Before going further, I have first of all, to thank the gentleman who has just walked in with this great pile of paper, Bill Milne, editor of the *Journal of Chemical Information and Computer Sciences*, who is really the organizer of this symposium, for his efforts in getting it all together. Secondly, I have a request to make of each and everyone of you. That is, if you have the temerity to ask a question or make a point, would you please state your name, your affiliation, and speak clearly. We may find it necessary to repeat the questions. If we repeat them incorrectly, please make us repeat it until we get it right. But, if you can speak up, it would help those at the back of the room and the lady on my right, far far right by the wall, who I am told will complain if we don't speak loudly enough.

So, now we are going to introduce our first participant, Jim Sibley. Jim has a B.Sc. and a Ph.D. from the University of London. That is one of the better universities of course. That is London, England, not London, Ontario. Jim has been in

[†] Financial support for this Symposium was provided by the ACS Division of Chemical Information and by the law firm of Spensley, Horn, Jubas and Lubitz, of Los Angeles.