numerous thanks
Randy

::::::::::::::
227
::::::::::::::
Return-Path: <BUCHANAN@SUMEX-AIM.ARPA>
Received: from rockefeller.ARPA (rocky4.ARPA) by rocky3 (4.12/4.7)
        id AA03094; Thu, 8 Aug 85 01:53:14 edt
Received: from SUMEX-AIM.ARPA (sumex-aim.arpa.ARPA) by rockefeller.ARPA (4.12
        id AA04046; Thu, 8 Aug 85 01:53:06 edt
Date: Wed 7 Aug 85 22:51:18-PDT
From:BruceBuchanan<BUCHANAN@SUMEX-AIM.ARPA>
Subject: Re: dendral performance
To: DAVIS%MIT-OZ@MIT-MC.ARPA
Cc: lederberg@SUMEX-AIM.ARPA, bgb@SUMEX-AIM.ARPA, waleson@SUMEX-AIM.ARPA
In-Reply-To: Message from "Randall Davis <DAVIS%MIT-OZ@MIT-MC.ARPA>" of Wed 7
Message-Id: <12133398993.20.BUCHANAN@SUMEX-AIM.ARPA>

Randy,
  I suspect the Dreyfus brothers will try to argue that a narrow scope of expertise (as Dendral or Mycin have) is not expertise at all. So you have to be prepared to argue that depth is at least as important. Does one have to know most of everything about most all subjects in an area to be an expert? How narrowly can one define the area?

  Dendral got to be pretty good at interpreting the mass spectra of some sub-families of steroids: estrogenic steroids and the keto-androstanes in particular. [I think it could have been given much more breadth if we had been more diligent.] I guess I believe it was doing as well as Djerassi's post-docs, who were experts-in-training in mass spectroscopy.

  There was one experiment [again, which I think we could have repeated for many other sub-families if we had tried] in which DENDRAL was able to provide correct structures for the compounds in a mixture, when presented with a single mass spectrum for the whole mixture. This kind of systematic analysis is extremely tedious and almost never done by hand. A collaborator in Europe sent us the spectrum without telling us the components, except to say that they were all estrogenic steroids. We ran DENDRAL on each spectrum -- there were about five as I recall -- and then compared DENDRAL's answers with the collaborator's. DENDRAL got them all, and even suggested an extra component of one or two of the mixtrues that the collaborator acknowledged could have been residues in the instrument from the previous sample. This is written up in Jnl of Am.Chem.Soc. 1973, pp.6078-6084.

  More generally, the structure generator in DENDRAL performs a task that almost no chemist can perform on any but very small problems: complete and non-redundant generation of ALL isomers of a given empirical formula. Moreover, it can be constrained to produce a complete and non-redundant list of structures that are consistent with an arbitrary set of constraints (e.g., as inferred from mass spectra). Very small empirical formulas will allow many hundreds or a few thousand structures.

Ray Carhart, in hand-checking CONGEN, manually listed 3000 structures for some C6 or C7 formula -- not only did he miss some but he included a few duplicates that he failed to see because of the symmetries. I think only Josh at that time had more experience with thinking about systematic generation of chemical structures.

In an experiment with published articles in the literature, we showed that chemists often failed to note all the structures that DENDRAL found to be consistent with the published constraints. Whether they just overlooked them (as we think) or were merely taking some shortcuts in writing up their work (as some claimed), we don't know. In either case they were less careful than they would have been if they had been using CONGEN.

Ed Feigenbaum is fond of saying that Dendral is better than the world's best chemists. In some senses it can be said to be. But the Drefus brothers are not incorrect in assuming that there are some things a chemist can do that Dendral cannot.

Hope this helps. Josh, please add or correct as needed.

regards,
bgb
