Steve Stein, NIST

Mass Spectrometry Data Center

Biomolecular Measurement Division

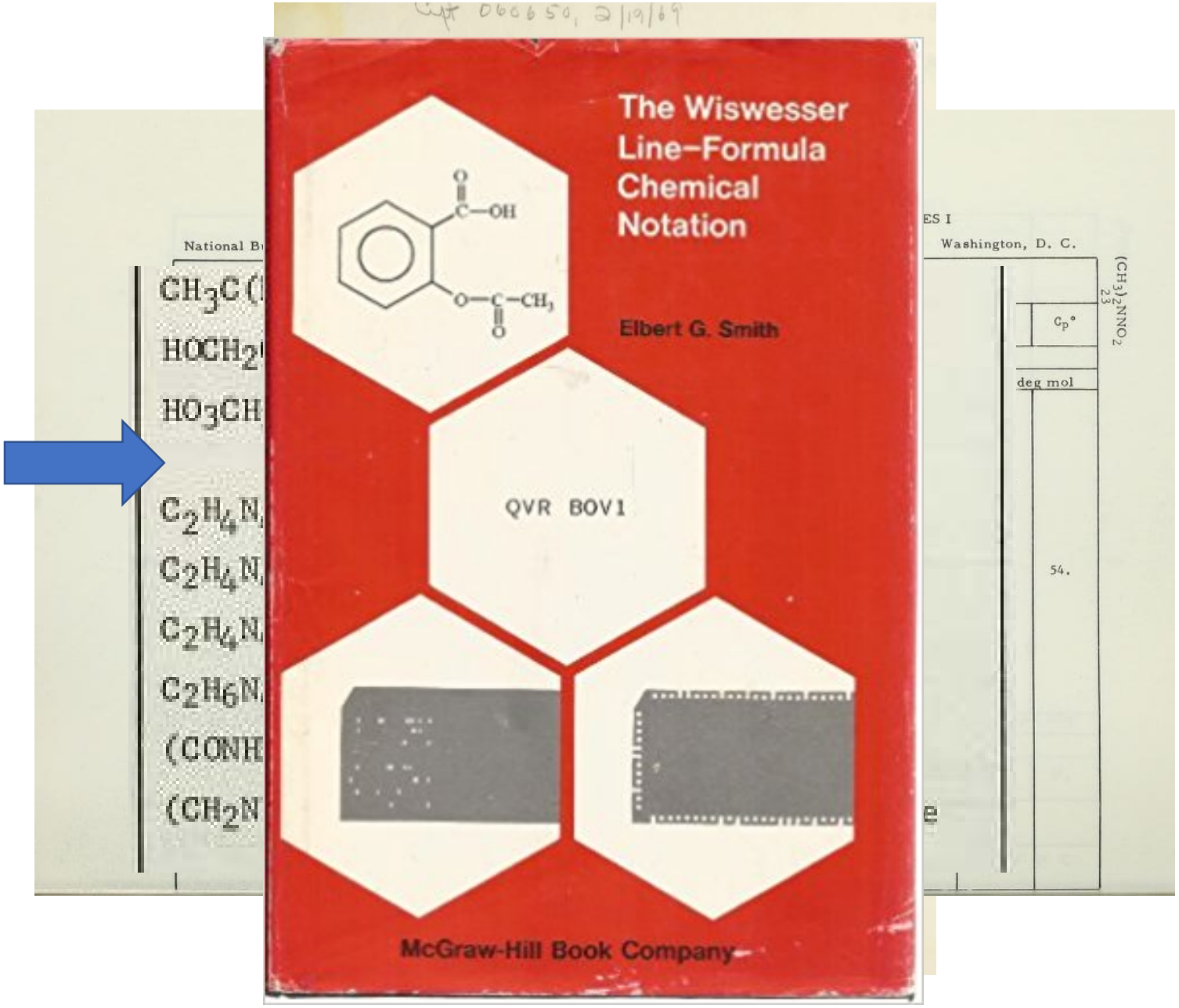
The Origin of the ‘n’ In InChI

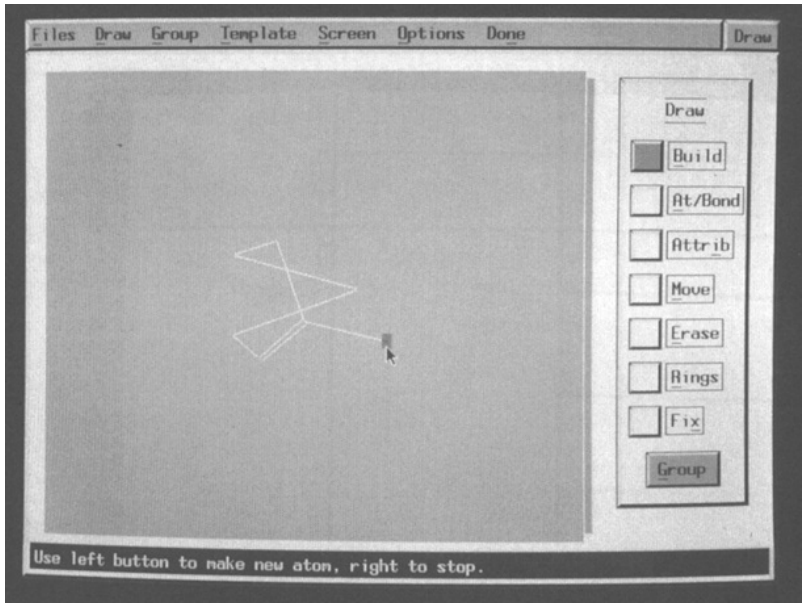
NIST

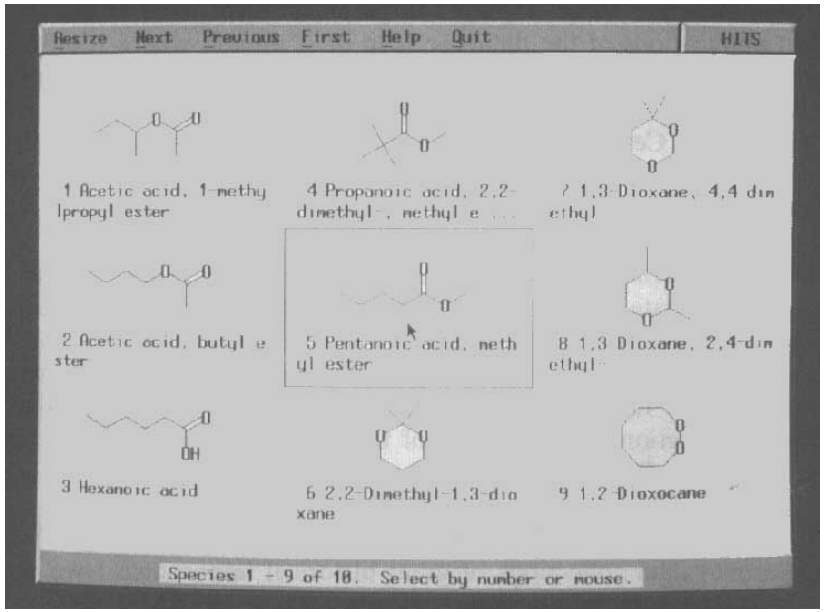
National Institute of Standards and Technology

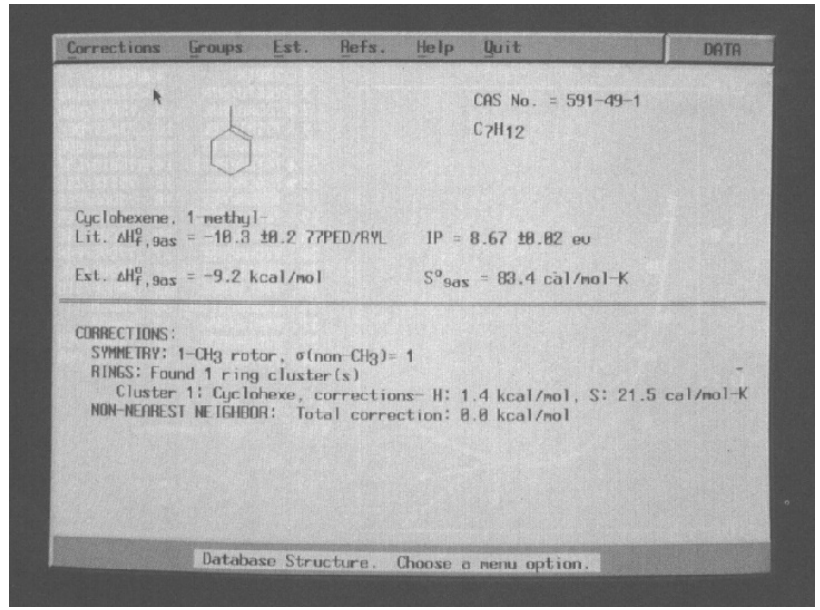
U.S. Department of Commerce

Frederick D. Rossini
NBS 1928-1950









NIST National Institute of Standards and Technology U.S. Department of Commerce **NIST Chemistry WebBook, SRD 69**

Search ▾ | NIST Data ▾ | About ▾

The NIST Structures and Properties Group Additivity Model

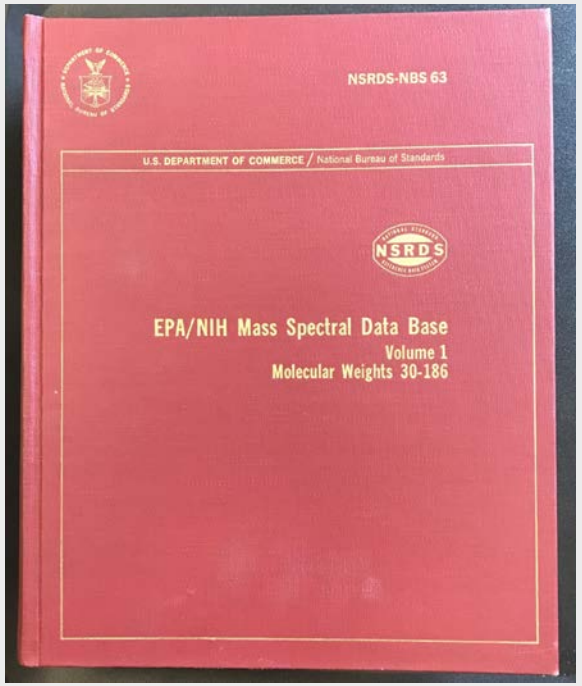
The Structures and Properties software fully implements Benson's Group Additivity method[1,2, 3] for estimating gas phase enthalpies of formation, entropies and heat capacities. The implementation includes a variety of corrections, including ring strain, symmetry and non-nearest neighbor effects such as cis, gauche and ortho interactions. Benson's group values are used except when more recent data call for significant changes or where it was necessary to compensate for differences in computed symmetry numbers (discussed below). Some entropy and heat capacity group values missing from Benson's tables are taken from the DIPPR Data Evaluation Handbook[4].

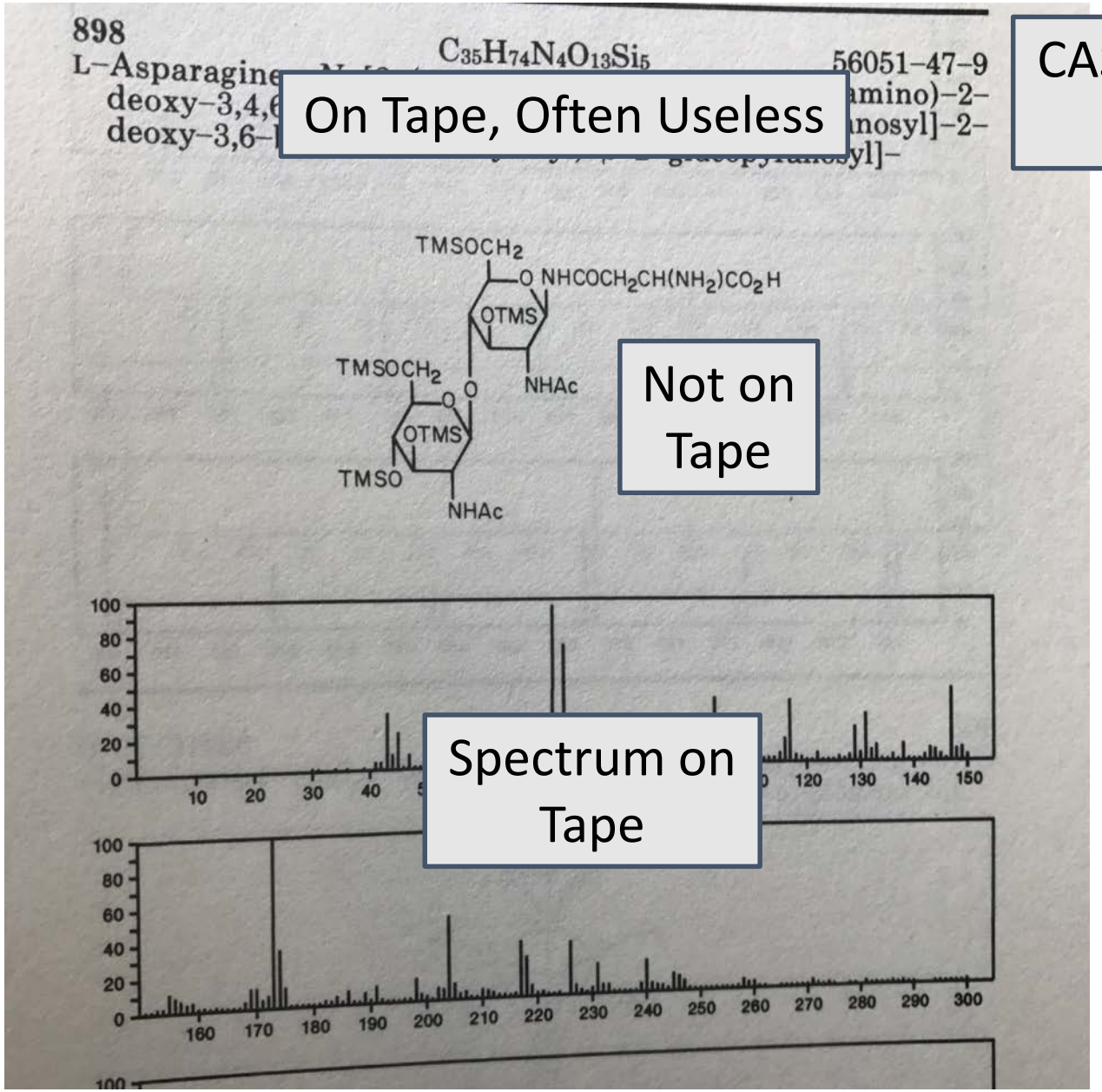
Based primarily on more recently reported data, about 30 new groups have been assigned enthalpy of formation values and nearly 40 new ring corrections were added[5]. The principal source of experimental data for the new groups and ring corrections was "Thermochemical Data of Organic Compounds"[6]. The treatment of symmetry and equivalent isomers is generally the same as Benson's, but differs in some respects, as discussed later in this section.

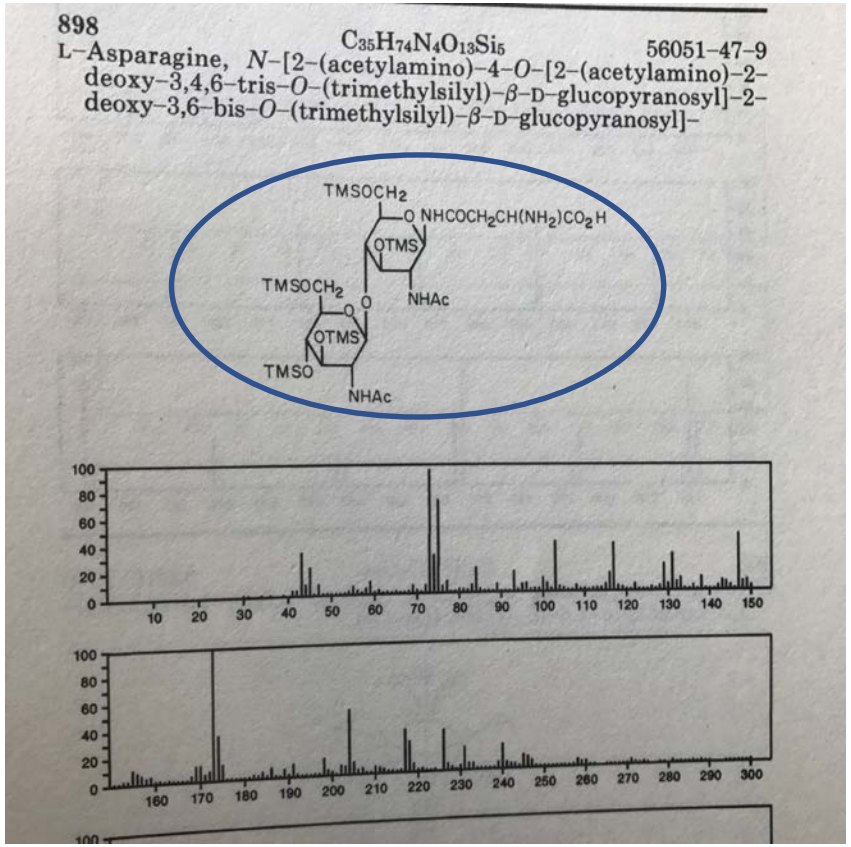
For flexible ring systems and trivalent nitrogen compounds the program occasionally gives values different from those recommended by Benson. This is done intentionally to predict more reliably entropies of ring systems for which no data exists, although agreement for a limited number of known compounds (some substituted cyclohexanes) may be slightly degraded. The rationale for our choice of symmetry numbers is given below.

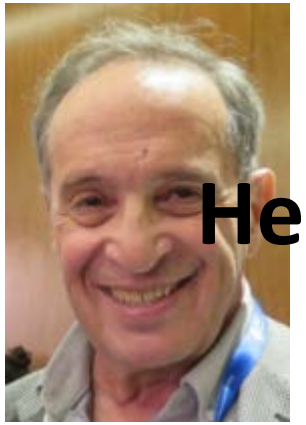
Benson's "whole molecule" corrections to the entropy for symmetry (σ) and for numbers of equivalent optical isomers (n) require an understanding of the molecular conformation, which often is not available. Therefore, such corrections can be difficult to apply, especially for

EPA/NIH Mass Spectral Data Base Transferred to NBS in 1986




898
$C_{35}H_{74}N_4O_{13}Si_5$
56051–47–9
L–Asparagine
deoxy–3,4,6
deoxy–3,6–
amino)–2–
nosyl]–2–
yl]–
On Tape, Often Useless
CAS rn/Formula on Tape
TMSOCH2
O NHCOCH2CH(NH2)CO2H
OTMS
TMSOCH2
O
O
NHAc
OTMS
TMSO
NHAc
Not on Tape
Spectrum on Tape





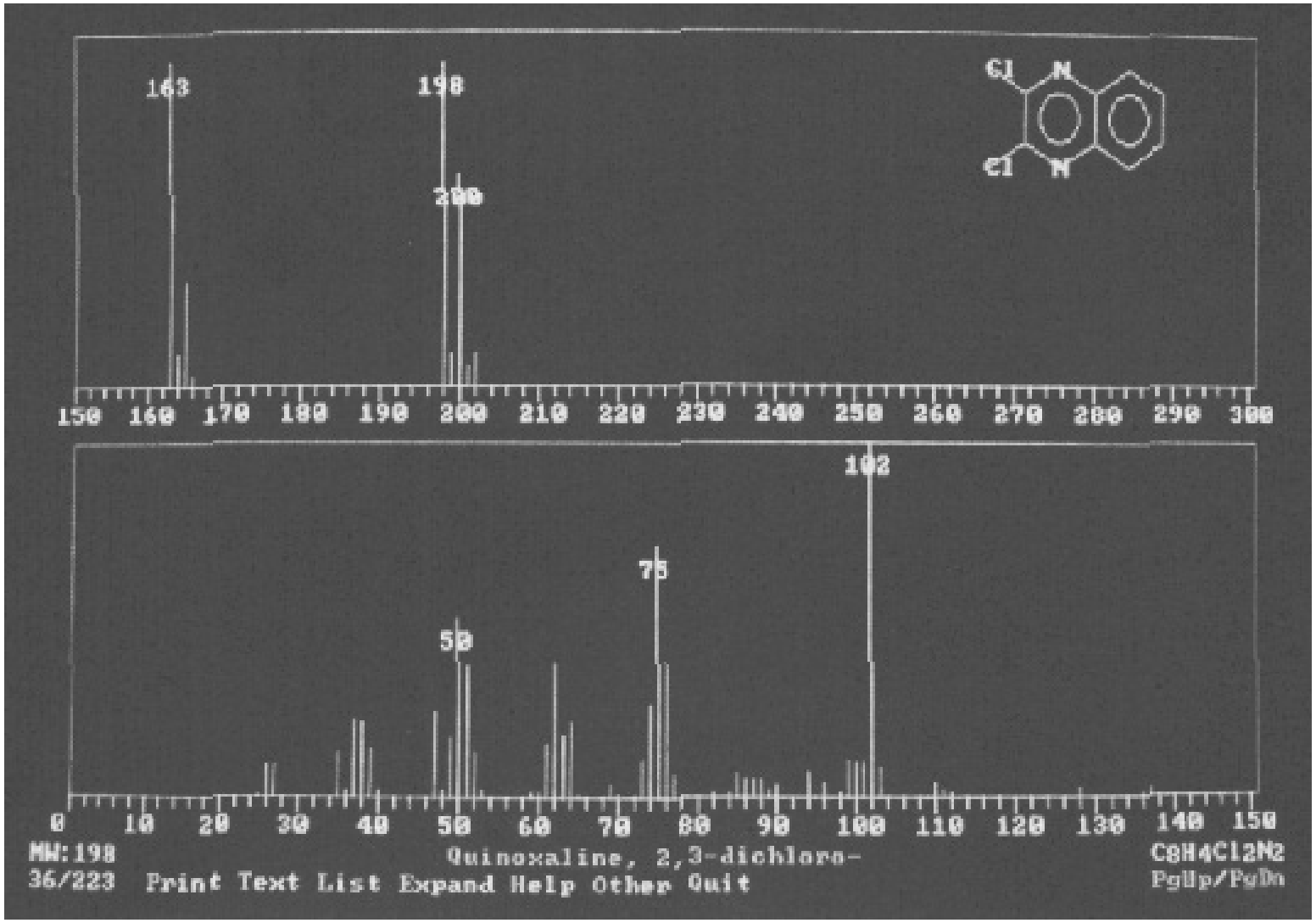
Help !!!

Connection Tables
from On-Line Database

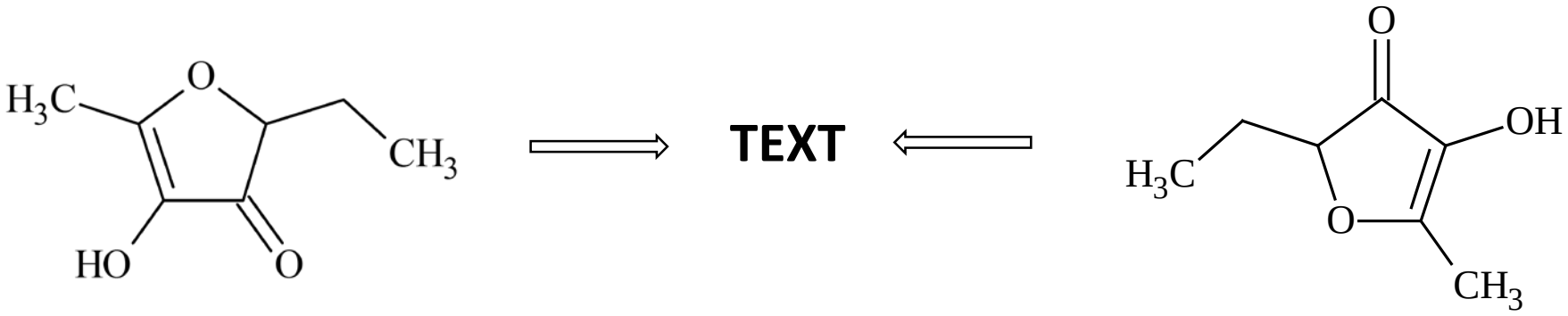
Software CT to 2D
Craig Shelly, Kodak

NIST/EPA/NIH Mass Spectral Database 1988

36 - 360K Floppies



Need to Find Replicate Compounds



Developed path searching algorithm

Chemists synthesize a single naming system

David Adam, London

An international team of chemists
ing on something that chemist
lacks — a consistent and comp
way of labelling all chemical comp

The new technique will apply
algorithms to molecular structu
ing a unique string of letters for any
compound. The new
to replace common chemical nam
allow easier linking to compounds
chemical databases and journals.

"The hope is that all organizat
handle information on chemicals w
to use a single format to say what the
is," says Alan McNaught, general m
the production division at the Roy
of Chemistry in Cambridge, who co
the project for the International
Pure and Applied Chemistry (IUPAC).

Right now there is no single international standard for identifying chemicals. The IUPAC and the American Chemical Society use different rules. Some drug companies, as well as different branches of chemistry, have their own chemical-naming systems. Even simple structures can cause confusion. For example, the formal name for acetic acid, the main ingredient in vinegar, is ethanoic acid.

IUPAC believes that its new system — which would be freely available to all — could unify the different approaches. Tentatively

IUPAC Chemical Identifier

InChI

Ichi Ocha

(2-amino-4-
nedioic acid.

TIM EVANS/NPI

chemists test the idea. It labels each atom in a compound in a way that does not depend on how the structure is drawn, and converts the label to a string of characters. The format has not been finalized, but at present ethane is 'C3C 2-1 or e le, and acetone is
process is reversible, so molecular structures can be generated from the identifiers.

The next step is to extend the system to include more complex organic compounds,

as polymers, and ultimately to tackle ganic compounds. By adding it to soft- packages commonly used to draw ical structures, the NIST team hopes ChI will enter into widespread use.

n effect, the IChI number will provide each chemical molecule with a digital object identifier (DOI) — a concept increasingly being applied to everything from scientific papers to individual genes. Jonathan Goodman, a chemist at the University of Cambridge, says chemistry suits this approach well. "Molecules are a wonderful unit of information to treat in this way," he says. "They are complex enough to have lots of interesting features and difficulties but simple enough to represent quite a small subset." ■

www.iupac.org/projects/2000/2000-025-1-800.html

Nature, May 23, 2002

Lucifer's Lawyer

https://www.pinterest.com/pin/194147433910548679/

For Breakfast

Milk Toast

at least once a week in place of a cereal. It is wholesome, nourishing and delicious.

Slices toasted bread	1 tablespoon flour
1 quart boiling milk	½ cup cold milk
1 tablespoon butter	1 teaspoon salt

Add butter to boiling milk; mix flour with cold milk and stir to hot milk with salt. Pour over toast and serve hot. Some like milk toast sprinkled with sugar, others pr salt and pepper; or a little grated nutmeg may be sprinked over the top.

IUPAC InChI

IUPAC International Chemical Identifier

http://www.mrbreakfast.com/superdisplay.asp?recipeid=148

InChIKey Plays a Central Role in our MS Library Program

- Compare Spectra Across Libraries
  - Essential evaluation tool
  - First, non-stereo block used
  - Recent paper in JASMS
- Spectrum Dissemination
  - Spectra from NIST Webbook
  - Work in progress: Metabolomics
- As a Compound Accession Number
  - Underway

InChIKey as an MS Accession Number

- Useful, but insufficient
  - Must deal with uncertain structural features
- Multiple isomers
  - Meso – DL (multiple stereocenters)
  - Distinguish by retention time
- Derivatization
  - GC/MS often requires chemical adduction at uncertain positions

Uncertain Pairs of Isomers

- Meso/DL
  - Naturally occur together
- Double Bond (sp2) Z/E isomerism
  - Mostly older data
- Spectra usually very similar
- InChIKey (no stereo) + Elution Order
  - Use uncertain bond type

Derivatives for GC/MS

- Trimethylsilyl (TMS) most common
- Replace H at OH, NH2, NH, SH on precursor
  - Position not always predictable
  - Enol possibility for keto-enol tautomers
  - Steric effects
- Analyst concerned only with precursor

- Will use 2 representations
  - InChIKey for ‘Probable’ Structure
  - InChIKey (precursor) + Order # + Text Descriptor (e.g., TMS)