Twenty Five Years of Progress in Cheminformatics

Dr. Wendy A. Warr

http://www.warr.com
Our Authors

Wendy Warr & Associates
<table>
<thead>
<tr>
<th>Journal Name Analysis</th>
<th>Count</th>
</tr>
</thead>
<tbody>
<tr>
<td>Tetrahedron Latt.</td>
<td>9</td>
</tr>
<tr>
<td>Angew. Chem.</td>
<td>8</td>
</tr>
<tr>
<td>J. Am. Chem. Soc.</td>
<td>7</td>
</tr>
<tr>
<td>Tetrahedron</td>
<td>7</td>
</tr>
<tr>
<td>Chem. Ber.</td>
<td>6</td>
</tr>
<tr>
<td>Angew. Chem., Int. Ed. Engl.</td>
<td>5</td>
</tr>
<tr>
<td>References not containing information for this analysis</td>
<td>5</td>
</tr>
<tr>
<td>AIP Conf. Proc.</td>
<td>4</td>
</tr>
<tr>
<td>J. Chem. Soc., Perkin Trans. 2</td>
<td>4</td>
</tr>
<tr>
<td>Nachr. Chem., Tech. Lab.</td>
<td>4</td>
</tr>
</tbody>
</table>

Histogram Entries 1-18 of 110
ACS Award for Computers in Chemical and Pharmaceutical Research

1986 Raymond E. Dessy
1987 W. Todd Wipke
1988 W.A. Goddard III
1989 Christie G. Enke
1990 Peter C. Jurs
1991 John A. Pople
1992 Ernest R. Davidson
1993 W. Clark Still
1994 Michael J.S. Dewar
1995 Peter A. Kollman
1996 Norman L. Allinger
1997 Harold A. Scheraga
1998 William L. Jorgensen
1999 Corwin H. Hansch
2000 Donald G. Truhlar
2001 Martin Karplus
2002 Irwin D. Kuntz
2003 Kendall N. Houk
2004 W. Graham Richards
2005 Peter Willett
2006 Johann Gasteiger
ACS Division of Chemical Information
Herman Skolnik Award

Herman Skolnik (1976)          Peter Willett (1993)

Wendy Warr & Associates
ACS Division of Chemical Information
Herman Skolnik Award

- Ernst Meyer (1990)
- Reiner Luckenbach (1995)
- Clemens Jochum (1995)
- Johann Gasteiger (1997)

- Guenter Grethe (2001)
- Hugo Kubinyi (2006)
CROSSBOW

Computerized Retrieval of Structure Based On Wiswesser

Structure Representation


Structure Representation

Substructure Searching

- NIH/EPA Chemical Information System
- DARC
- CAS ONLINE
- MACCS/ISIS
Meyer’s Formula Reading Machine
Imlac Terminal

Wendy Warr & Associates
VT640 Terminal
1984 VAX 11/750

- Clock speed 6 MHz
- 2 Mb memory
- 134 Mb fixed disk
- Two 67 Mb exchangeable disk drives
- Shared peripherals
- £100,000 (1984 price)
Windows
Icons
Mice
Pointers
Microcomputers

Wendy Warr & Associates
Ivar Ugi
1930 - 2005
CAOS

Reaction Programs

• Reaction retrieval
• Synthetic analysis
  – Synthesis design
  – Reaction prediction
  – Mechanism elucidation
Reaction Retrieval

- **REACCS/ISIS**
- **ORAC**
- **CASREACT**
Synthetic Analysis

• Logic-oriented
  – IGOR
  – EROS
  – SYNGEN
  – CAMEO
  – WODCA

• Information-oriented
  – LHASA
  – SECS/CASP
Reaction Classification

InfoChem Classification Algorithm

CLASSIFY
Cambridge Structural Database


3D Structures

Gasteiger Most Cited

3D Searching


3D Searching


Molecules Twist

Wendy Warr & Associates
Molecular Diversity
Selection

$10^{180}$ possible drugs...$10^{18}$ likely drugs...
$10^7$ known compounds...$10^6$ commercially available...$10^6$ in corporate databases...
$10^4$ in drug databases...$10^3$ commercial drugs...
$10^2$ profitable drugs
Library Design

- Random screening is too expensive
- Its hit rate is low
- False positives may be a problem
- HTS consumes expensive compounds
- There are too many possible compounds
- How to choose those most likely to be hits?
Selecting Diverse Subsets

- Clustering
- Dissimilarity-based selection
- Partitioning/cell-based approaches
- Optimization-based methods
Measuring Diversity

Diversity


Product-based or Reagent-based Design

Leads and Drugs


Leads and Drugs


Methods for vHTS

- Identifying drug-like structures
- 2D similarity
- 3D pharmacophores
- Docking
Docking


Docking


GOLD docking result for PDB 1FKG
Most Cited. Top Twenty

Most Cited. Top Twenty


Most Cited


Most Cited


Most Cited


Most Cited


Most Cited


Acknowledgments

• Chemical Abstracts Service
• Eric Shively, CAS
• Michael Lynch
The Next 25 Years?
Fearless predictions…

“Where a calculator on the Eniac is equipped with 18,000 vacuum tubes and weighs 30 tons, computers in the future may have only 1,000 vacuum tubes and perhaps weigh 1.5 tons.”

Popular Mechanics, March 1949
Basic Principle of Forecasting

“Give them a number or give them a date, but never both.”

- Edgar Fiedler
Technology Adoption Cycle

- Peak of inflated expectations
- Slope of enlightenment
- Trough of disillusionment
- Plateau of productivity

Wendy Warr & Associates
Information Overload

- The Lord’s Prayer uses 56 words
- The Ten Commandments, 297 words
- The American Declaration of Independence, 300 words
- The EEC Directive on the import of caramel and caramel products uses 26,911 words
A Mobile CAS Service
New Infrastructures, Middleware

- e-Science
- Cyberinfrastructure
- eSciDoc
“Here we sit side by side with those on whose shoulders we stand”

Michael Lynch, 2002