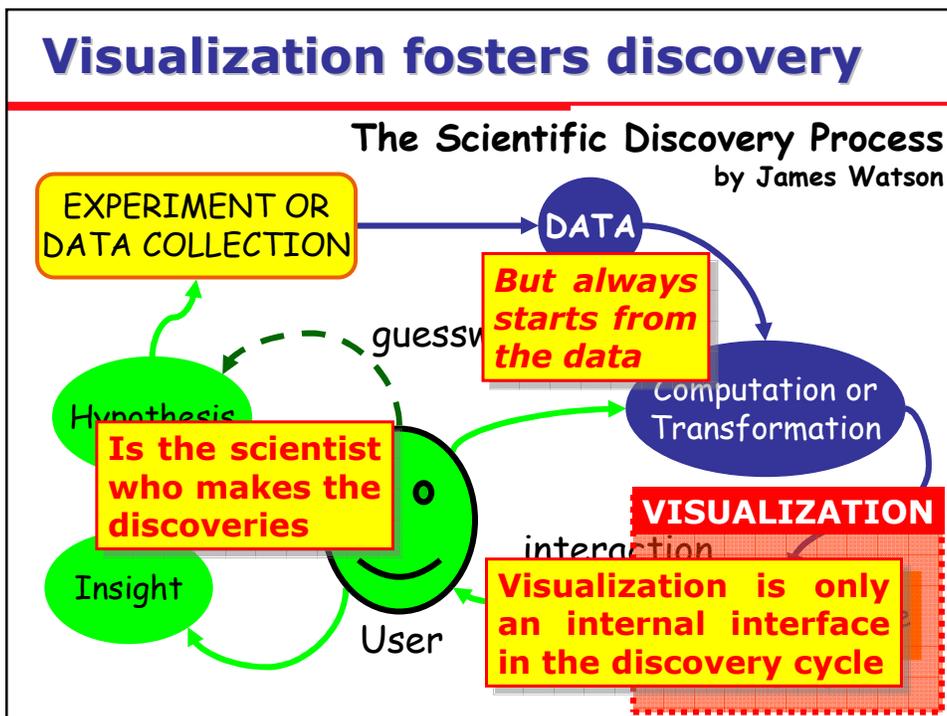


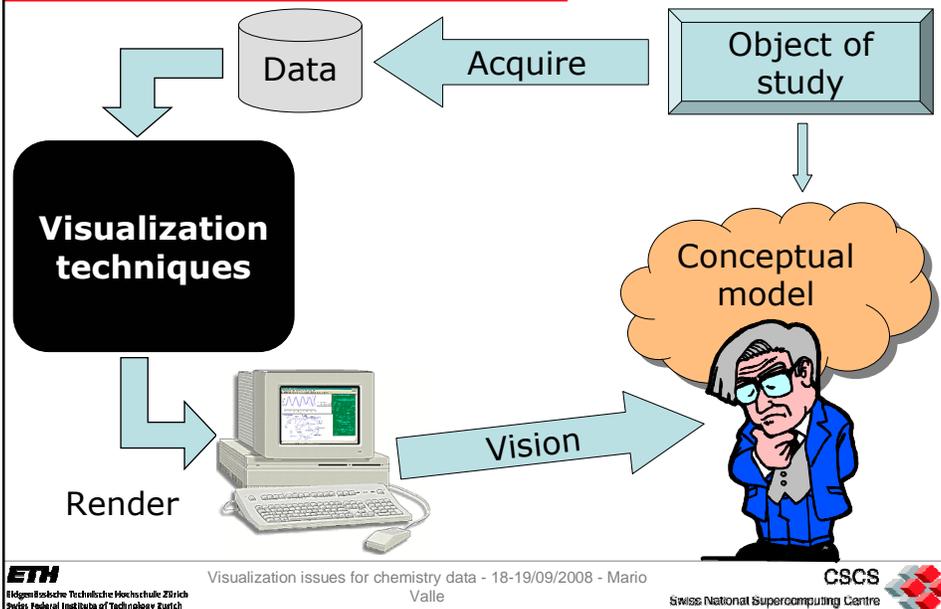
TITLE REF. Number: 188872 REF. Code: rdd2 M. Space Group: P1  
 CELL 0.71069 30.0658 21.9373 4.0980 90.000  
 ZERR 1 0.0000 0.0000 0.0000 0.0000  
 LATT -4  
 SYMM - Swiss National Supercomputing Centre  
 SYMM 0.25000 - X, 0.25000 + Y, 0.25000 + Z  
 SYMM 0.75000 + X, 0.75000 - Y, 0.25000 + Z  
 SFAC C 2.31000 20.84392 1.02000 10.20751 1.58860 0.56873 0.86550  
 51.65125 0.21560 0.00330 0.00160 11.500 0.68000 12.31100  
 SFAC N 12.21261 0.00570 3.13220 9.89331 2.01250 28.99750 14.16600  
 0.58260 -11.52901 0.00610 0.00330 19.600 0.68000 14.16600  
 SFAC O 3.04850 13.27711 2.28680 5.70111 1.54630 0.32390 0.00000  
 32.90894 0.25080 0.01060 0.00600 32.500  
 UNIT 104 32 32  
 FVAR 1.00000  
 O1 3 -0.01002 0.05177 1.00141 11.0000 0.0500  
 O2 3 -0.01671 0.14669 0.69403 11.0000 0.0500  
 N1 2 0.02997 0.07429 0.87290 11.0000 0.0500  
 N2 2 0.02073 0.12343 0.71921 11.0000 0.0500  
 C1 1 -0.00089 0.00001 1.19609 10.5000 0.0500  
 C2 1 0.05909 0.15339 0.56535 11.0000 0.0500  
 C3 1 0.05240 0.21065 0.44370 11.0000 0.0500  
 C4 1 0.08785 0.24028 0.29736 11.0000 0.0500  
 C5 1 0.00000 0.00000 0.00000 11.0000 0.0500  
 C6 1 0.00000 0.00000 0.00000 11.0000 0.0500  
 C7 1 0.09968 0.12458 0.54459 11.0000 0.0500  
 H1 1 1.00000 0.00000 0.00000 11.0000 0.00500  
 H2 1 2.00000 0.00000 0.00000 11.0000 -1.00000  
 REM SPGR Fdd2  
 END

**CSCS**  
 - Swiss National Supercomputing Centre  
**Mario Valle**  
**CECAM - 18-19/09/2008**

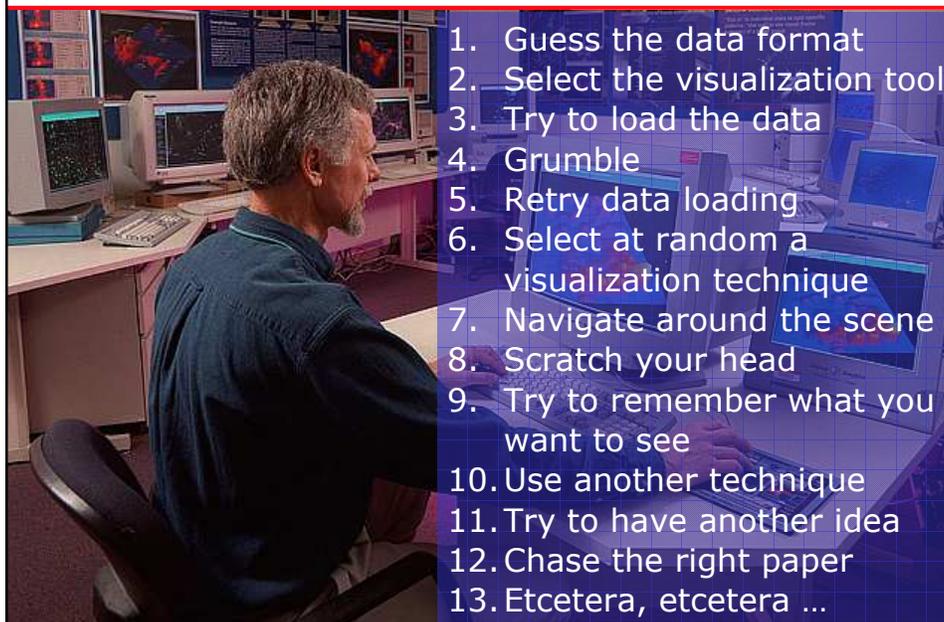
# Visualization issues for chemistry data



## The visualization process



## The real visualization cycle

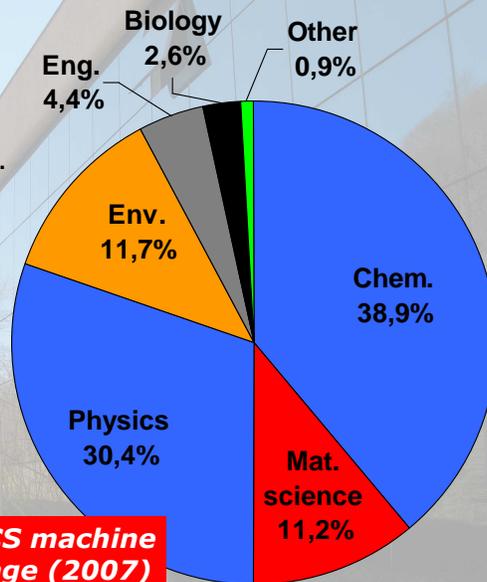


## Chemistry support at CSCS

In 2007 almost **exactly half** of the CSCS machines time has been assigned to chemistry related projects.

The visualization group works in **strict cooperation with the users.**

Not only for day-to-day support, but acting also as a **catalyst** for introducing innovative techniques and methods that help their work.



## The archetypal problem...

```
1 bernina
2 Entering Gaussian System, Link 0=g98
3 Initial command:
4 /apps/gaussian/g98-rev-A.11.3/g98/11.exe /scratch/gauss.7453/Gau-7456.inp -scdir=/scratch/gauss.7453/
5 Entering Link 1 = /apps/gaussian/g98-rev-A.11.3/g98/11.exe FID= 7457.
6
7 Copyright (c) 1988,1990,1992,1993,1995,1998 Gaussian, Inc.
8 All Rights Reserved.
9
10 This is part of the Gaussian(R) 98 program. It is based on
11 the Gaussian 94(TM) system (copyright 1995 Gaussian, Inc.),
12 the Gaussian 92(TM) system (copyright 1992 Gaussian, Inc.),
13 the Gaussian 90(TM) system (copyright 1990 Gaussian, Inc.),
14 the Gaussian 88(TM) system (copyright 1988 Gaussian, Inc.),
15 the Gaussian 86(TM) system (copyright 1986 Carnegie Mellon
16 University), and the Gaussian 82(TM) system (copyright 1983
17 Carnegie Mellon University). Gaussian is a federally registered
18 trademark of Gaussian, Inc.
19
20 This software contains proprietary and confidential information,
21 including trade secrets, belonging to Gaussian, Inc.
22
23 This software
24 used, copied
25 written lic
26
27 The followi
28 contracts u
29
30
31
32 Use, duplication or disclosure by the US Government is subject
```

**The Gaussian log file  
Made for humans, not for visualization tools**

# A puzzle for visualization tools

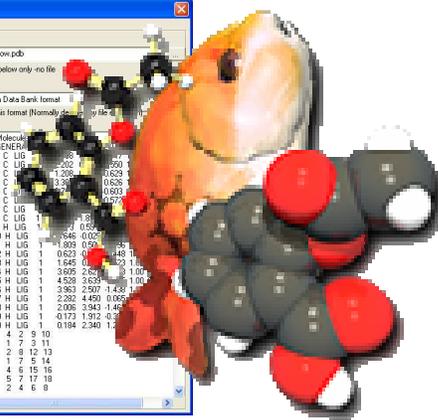
```

HEADER      1 Configuration at time step      4800.00
ATOM       1  N   ASP      1      14.927  -1.261  -21.608  -0.52  0.00
ATOM       2  HN  ASP      1      14.887  -2.262  -21.481  0.29  0.00
ATOM       3  CA  ASP      1      14.856  -0.745  -22.986  0.04  0.00
ATOM       4  HA  ASP      1      15.901  -0.688  -23.290  0.09  0.00
ATOM       5  CB  ASP      1      14.029  -1.678  -23.900  -0.03  0.00
ATOM       6  HB1 ASP      1      13.045  -1.215  -23.986  -0.01  0.00
ATOM       7  HB2 ASP      1      14.469  -1.602  -24.894  -0.01  0.00
ATOM       8  CG  ASP      1      13.834  -3.180  -23.648  0.80  0.00
ATOM       9  OD1 ASP      1      13.554  -3.809  -24.700  -0.80  0.00
ATOM      10  OD2 ASP      1      13.887  -3.809  -22.529  -0.80  0.00
ATOM      11  C   ASE      1      14.276  0.676  -23.013  0.54  0.00
ATOM      12  O   ASP      1      14.394  1.340  -24.055  -0.58  0.00
ATOM      13  N   ILE      2      13.776  1.127  -21.895  -0.41  0.00
ATOM      14  HN  ILE      2      13.822  0.443  -21.154  0.27  0.00
ATOM      15  CA  ILE      2      13.807  2.541  -21.468  -0.06  0.00
ATOM      16  HA  ILE      2      14.643  3.012  -21.984  0.09  0.00
ATOM      17  CB  ILE      2      12.558  3.270  -21.979  0.13  0.00
ATOM      18  HB  ILE      2      12.273  2.957  -22.983  0.02  0.00
ATOM      19  CG2 ILE      2      11.342  3.013  -21.077  -0.32  0.00
ATOM      20  HG21 ILE      2      11.191  1.934  -21.102  0.09  0.00
ATOM      21  HG22 ILE      2      10.416  3.334  -21.554  0.09  0.00
ATOM      22  HG23 ILE      2      11.478  3.383  -20.061  0.09  0.00
ATOM      23  CG1 ILE      2      12.735  4.862  -22.021  -0.04  0.00
ATOM      24  HG11 ILE      2      13.725  5.027  -22.447  0.02  0.00
ATOM      25  HG12 ILE      2      12.796  5.238  -21.000  0.02  0.00
ATOM      26  CD1 ILE      2      11.683  5.605  -22.779  -0.07  0.00
ATOM      27  HD1 ILE      2      11.179  4.915  -23.456  0.02  0.00
ATOM      28  HD2 ILE      2      12.118  6.412  -23.369  0.02  0.00
ATOM      29  HD3 ILE      2      10.989  6.004  -22.040  0.02  0.00
    
```

- Why should I guess the atom type from the atom name?
- Why should I force one unrelated field to contain a user defined value?
- Why a trajectory should be stored in an *ad-hoc* manner?

Offered with compliments to you by your friendly PDB file format

# Universal converter/reader?



**Open Babel** is a project designed to interconvert between many chemistry file formats

**Unfortunately it fall short in various areas**

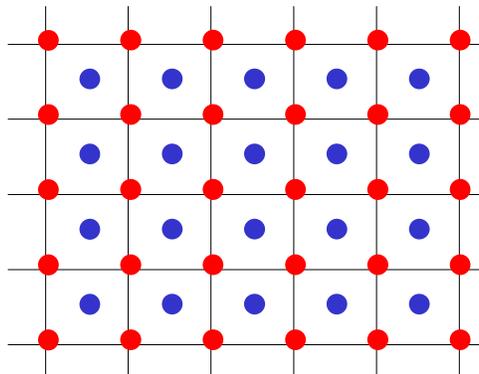
## 1<sup>st</sup> humble request

Provide us with a file format:

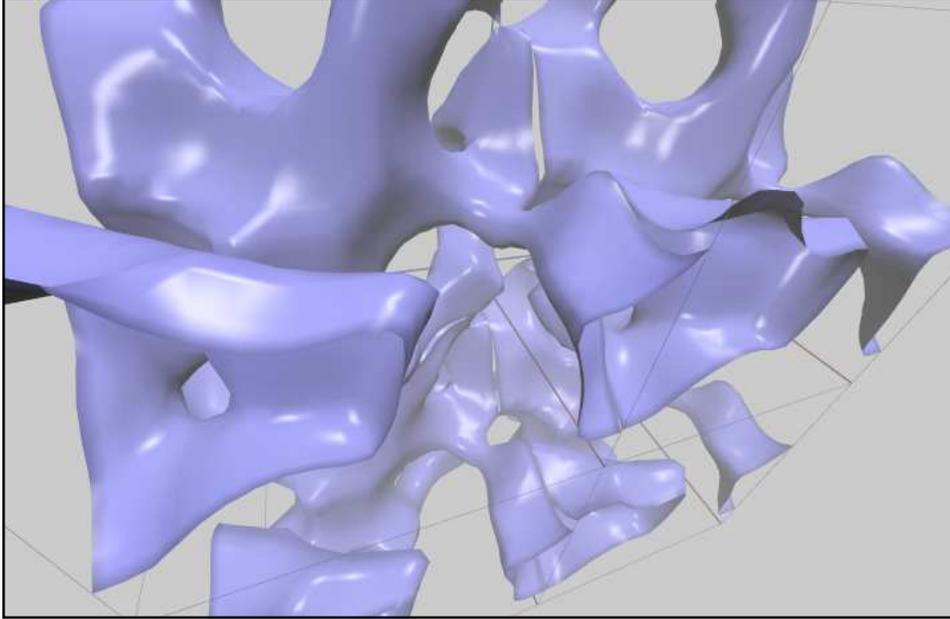
- Easily parseable
- That supports all the information the users need
- That supports all the information the user do not know yet they need

## Where are the values defined?

Gaussian Cube, PLT, CHGCAR  
are unclear on this point



## Where are the values defined?

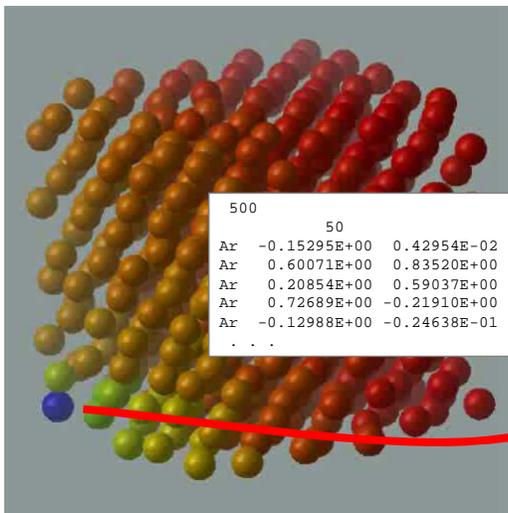


## 2<sup>nd</sup> humble request

Provide us with a file format:

- With a well defined semantic
- That does not play dirty tricks to spare a few bytes
- That accepts user contributions to clarify corner cases

## What is wrong with this data?



What sounds strange (or interesting...) with this data?

```
500
      50
Ar -0.15295E+00  0.42954E-02  0.67474E-01  0.30441E+00  0.78083E+00
Ar  0.60071E+00  0.83520E+00 -0.81022E-01  0.41648E+00
Ar  0.20854E+00  0.59037E+00  0.81560E+00  0.33224E+00
Ar  0.72689E+00 -0.21910E+00  0.53502E+00  0.28482E+00
Ar -0.12988E+00 -0.24638E-01  0.16533E+01  0.30124E+00
. . .
```

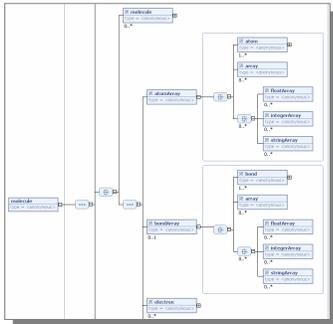
A bug in my XYZ data reader...

## 3<sup>rd</sup> humble request

Provide us with methods to write the files from custom (one-off) code:

- Simple to use (at the level of a bunch of "printf")
- Flexible
- Simple to use (already said that?)

## But there are well built formats



CML

```

data_global
  _chemical_name 'Calcite'
loop
  _publ_author_name
  'Graf D L'
  _journal_name_full "American Mineralogist"
  _journal_volume 46
  _journal_year 1961
  _journal_page_first 1283
  _journal_page_last 1316
  _publ_section_title
  ;
  Crystallographic tables for the rhombohedral carbonates
  ;
  _chemical_formula_sum 'Ca C O3'
  _cell_length_a 4.9900
  _cell_length_b 4.9900
  _cell_length_c 17.0615
  _cell_angle_alpha 90
  _cell_angle_beta 90
  _cell_angle_gamma 120
  _cell_volume 367.916
  _symmetry_space_group_name_H-M 'R -3 c'
loop
  _symmetry_equiv_pos_as_xyz
  'x, y, z'
  '2/3+x, 1/3+y, 1/3+z'
  '1/3+x, 2/3+y, 2/3+z'
  'x, x-y, 1/2+z'
  '2/3+x, 1/3+x-y, 5/6+z'
  '1/3+x, 2/3+x-y, 1/6+z'
  
```

CIF

## Similar experience with CFD data



1. Proprietary
2. Difficult to read
3. Format not documented



1. Standard
2. Covers everything
3. Readers/Writers available



## Similar experience with CFD data



1. Proprietary
2. Difficult to read
3. Format not documented



1. Need post conversion
2. Designed by committee
3. Always something missing
4. Low quality implementations

## 4<sup>th</sup> humble request

Give us a file format that provides a read/write library:

- Simple to use
- That supports all the information the user (do not know yet) needs
- That can be easily extended

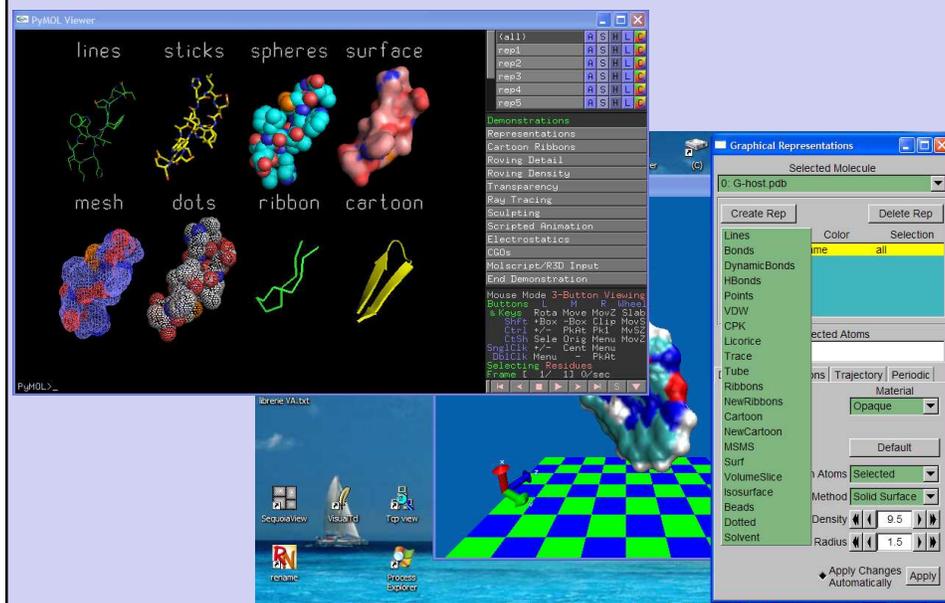
# Representational conformism



# CONFORMITY

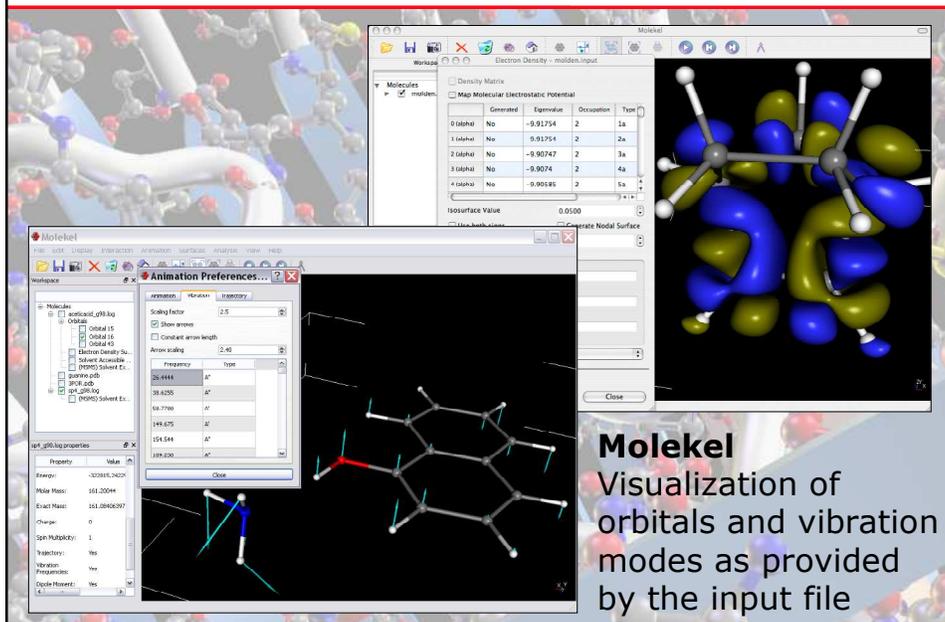
WHEN PEOPLE ARE FREE TO DO AS THEY PLEASE,  
THEY USUALLY IMITATE EACH OTHER.

# Lots of tools, same representations



The screenshot displays the PyMOL Viewer interface. On the left, a menu lists various representation styles: lines, sticks, spheres, surface, mesh, dots, ribbon, and cartoon. The main window shows a 3D molecular model of a protein structure. On the right, the 'Graphical Representations' panel is open, showing settings for the selected molecule 'G-host.pdb'. The panel includes options for 'Create Rep', 'Delete Rep', 'Color', and 'Selection'. The 'Color' dropdown is set to 'name' and the 'Selection' dropdown is set to 'all'. Other settings include 'VolumeSlice', 'Atoms', 'Method' (Solid Surface), 'Density' (9.5), and 'Radius' (1.5). The 'Apply Changes Automatically' checkbox is checked.

## Show what the data provide



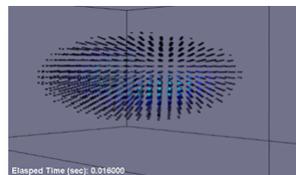
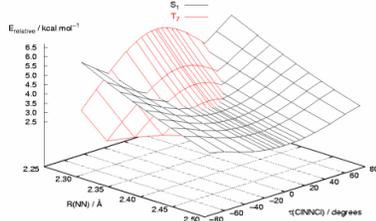
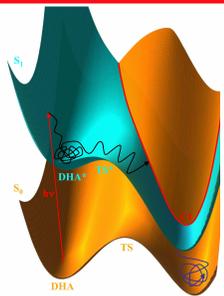
## Extend the visual language!

The limits of my language  
are the limits of my mind.

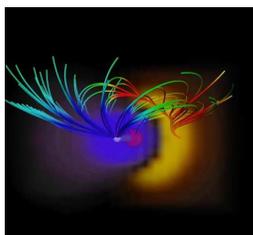
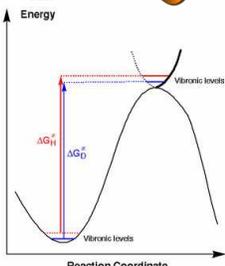
All I know  
is what I have words for.

*Ludwig Wittgenstein,  
Philosophical Investigations, 1953*

# Quantum chemistry viz support



Dance of the Probability Fluid  
Robert Wyatt (U. Texas)



**ETH**  
Eidgenössische Technische Hochschule Zürich  
Swiss Federal Institute of Technology Zurich

Visualization issues for chemistry data - 18-19/09/2008 - Mario Valle

**CSCS**  
Swiss National Supercomputing Centre

# Multidimensional data support

**CI**  
Chemistry International  
Vol. 21, No. 2,  
March 1999

[1999, Vol. 21](#)  
[No. 2](#)  
[\(March\)](#)  
[News from IUPAC](#)  
[Other Societies](#)  
[Reports from IUPAC Bodies](#)  
[New Books](#)  
[Reports from Commissions](#)  
[Prizes and Awards](#)  
[Conferences](#)  
[Announcements](#)  
[Conferences](#)

[CI Homepage](#)

**Chemistry International**  
Vol. 21, No. 2

**March 1999**

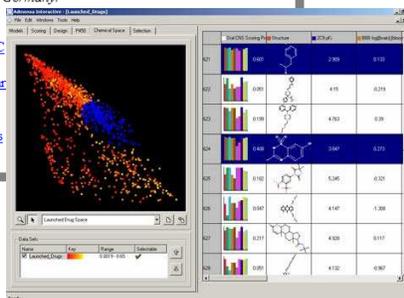
**News from IUPAC**

A New NMR Data Standard for the Exchange and Archiving for Multidimensional Data Sets

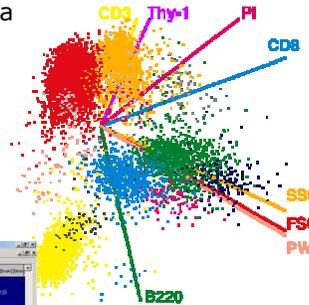
*ISAS, Institut für Spektrochemie und Angewandte Spektroskopie, Bunsen-Kirchhoff-Str.11, Postfach 10 13 52, 44013 Dortmund, Germany.*

[Introduction](#)  
[Proposal to IUPAC](#)  
[Work Plan](#)  
[New Data Dictionary](#)  
[Geography](#)  
[Conclusion](#)  
[Acknowledgements](#)  
[References](#)

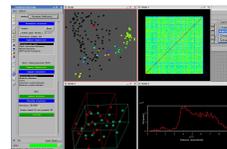
ADMensa  
Interactive



Biplot representation of blood chemistry data



Crystal fingerprinting





# PROBLEMS

NO MATTER HOW GREAT AND DESTRUCTIVE YOUR PROBLEMS MAY SEEM NOW,  
REMEMBER, YOU'VE PROBABLY ONLY SEEN THE TIP OF THEM.

## Summarizing our requests

When design a new data format, remember, it will be used (also) for visualization; and therefore:

- The users will outsmart you
- The visualization people will populate your nightmares
- The code developers will try not to use your libraries

