CSCS

Swiss National Supercomputing Centre

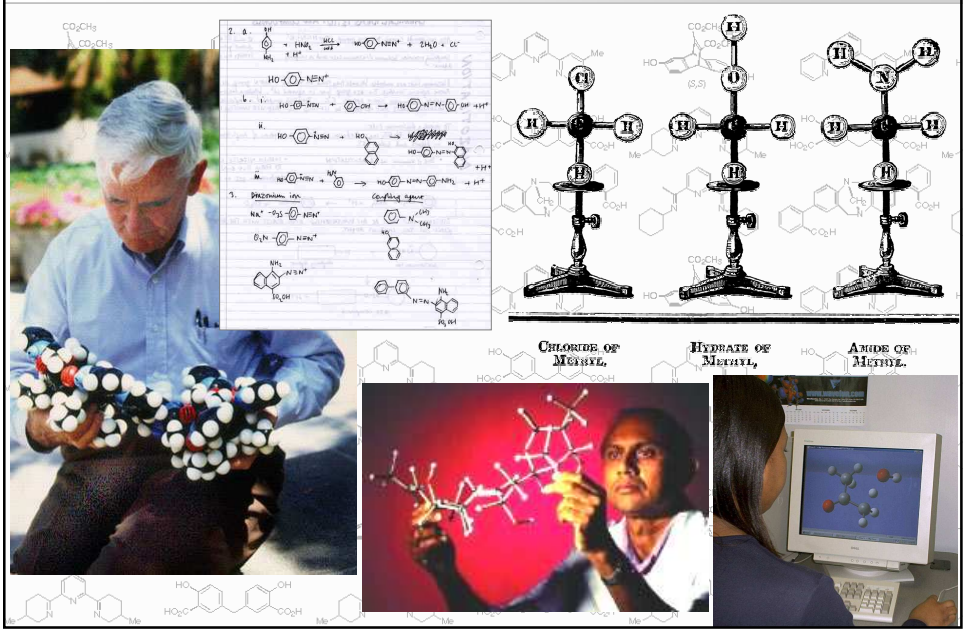
Ing. Mario Valle

EGEE'07 – 01/10/2007

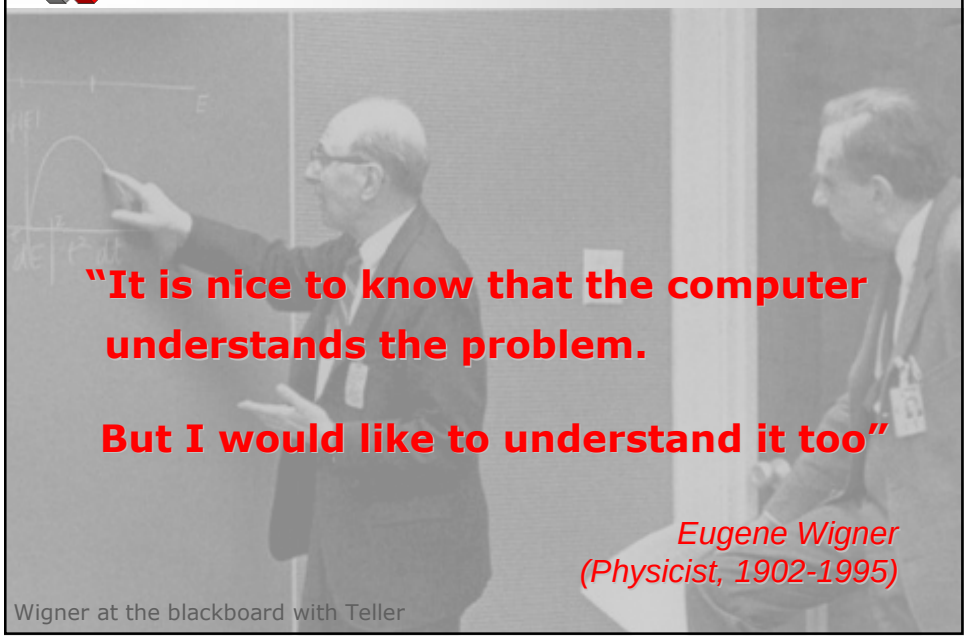
Chemistry Visualization Tools in an Integrated Discovery Cycle

Chemists are visual people

Chemistry relies on graphics



Chemistry relies on computation



"It is nice to know that the computer understands the problem.

But I would like to understand it too"

Eugene Wigner
(Physicist, 1902-1995)

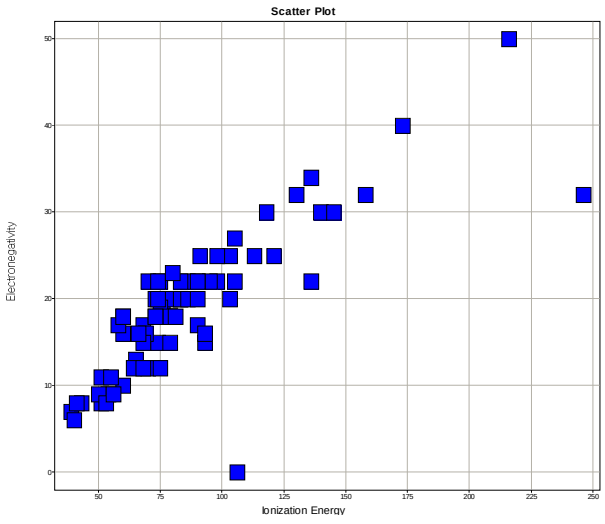
Wigner at the blackboard with Teller

Do you see anything interesting?

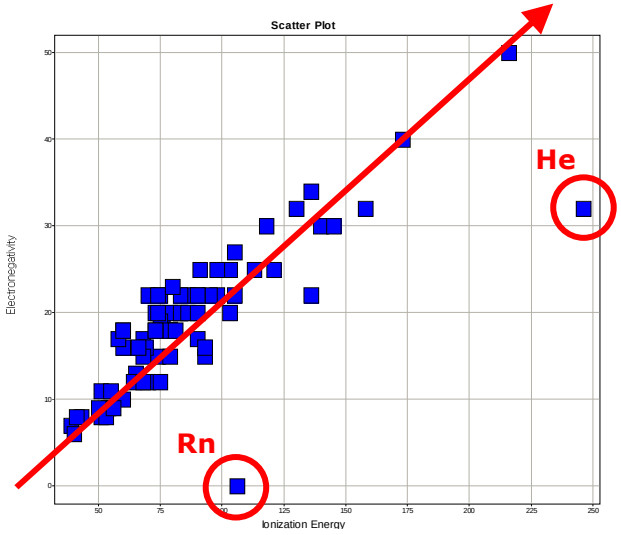
	A	B	C	D	E	F	G	H	I	J	K
1	Element	*P1	*P2	Atomic Num	Atomic Mas	Atomic Radi	Ionic Radius	Ionization E	Electronega	*C1	*C2
2	Ac	140	0	89	227	200	126	51	11	62	56
3	Ag	630	80	47	107	144	129	75	18	124	40
4	Al	750	160	13	27	143	67	60	16	28	25
5	Ar	1050	160	18	39	98	154	158	32	176	51
6	As	870	120	33	75	120	72	98	22	115	33
7	At	990	40	85	210	140	76	95	22	119	22
8	Au	630	40	79	197	144	99	91	25	131	22
9	B	750	200	5	10	85	41	83	20	101	8
10	Ba	80	40	56	137	222	149	51	8	46	56
11	Be	80	200	4	9	112	59	93	15	82	15
12	Bi	870	40	83	209	150	117	73	20	140	27
13	Br	990	120	35	79	114	182	118	30	161	44
14	C	810	200	6	12	77	30	113	25	82	1
15	Ca	80	120	20	40	197	114	60	10	70	51
16	Cd	690	80	48	112	151	109	90	17	113	43
17	Cl	990	160	17	35	100	167	130	32	173	47
18	Co	500	120	27	59	125	83	79	18	120	30
19	Cr	320	120	24	52	128	75	68	17	91	28
20	Cs	20	40	55	132	265	181	39	7	7	56
21	Cu	630	120	29	63	128	87	76	19	118	32
22	F	990	200	9	19	72	119	173	40	39	1
23	Fe	440	120	26	55	126	83	79	18	115	32
24	Fr	20	0	87	223	269	194	40	6	1	56
25	Ga	750	120	31	69	135	76	60	18	89	31
26	Ge	810	120	32	72	122	87	79	20	118	33
27	H	20	240	1	1	32	0	136	22	40	1
28	He	1050	240	2	4	31	93	246	32	1	1
29	Hf	200	40	72	178	159	85	70	12	95	44
30	Hg	690	40	80	200	151	116	103	20	147	27
31	I	990	80	53	126	133	206	105	27	153	44
32	In	750	80	49	114	167	94	58	17	93	42
33	Ir	500	40	77	192	136	82	90	22	116	25
34	K	20	120	19	39	227	152	43	8	37	56
35	Kr	1050	120	36	83	112	169	140	30	163	47

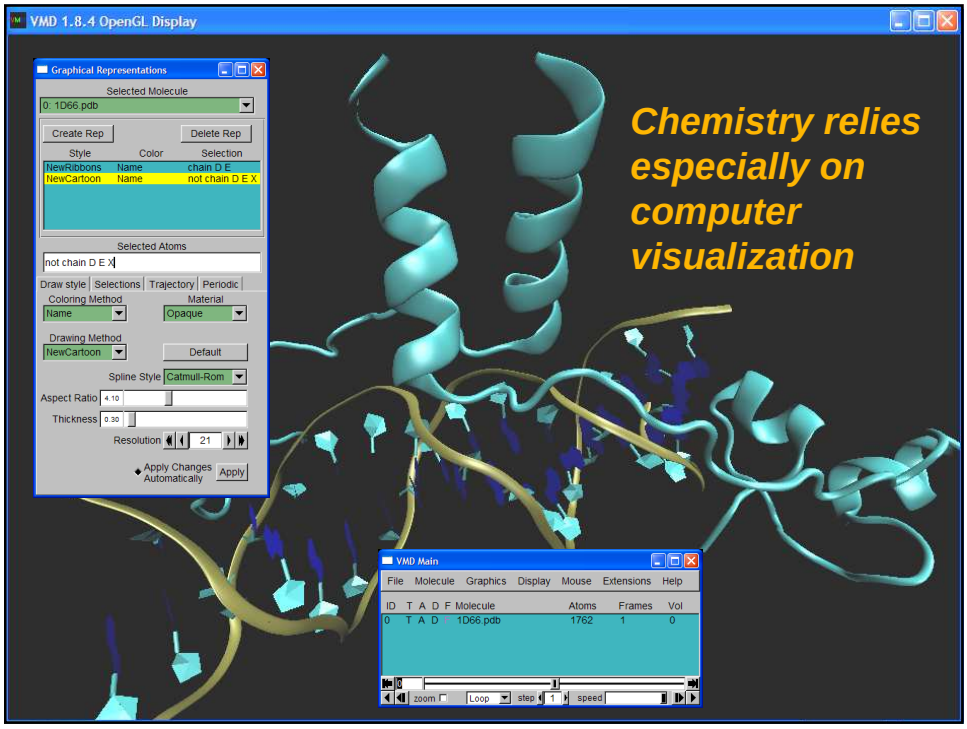
periodic

From data to information

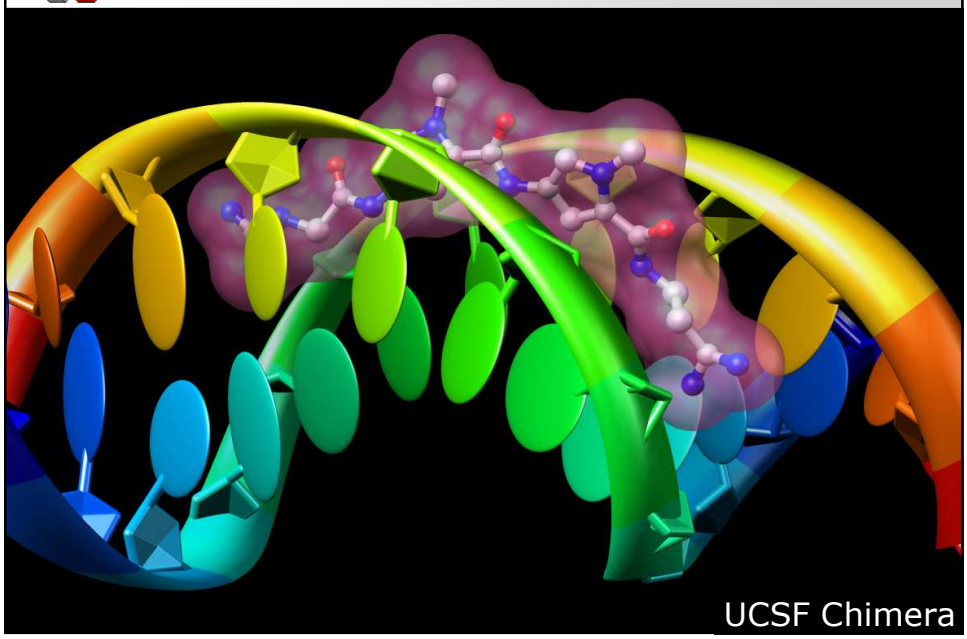


From data to information

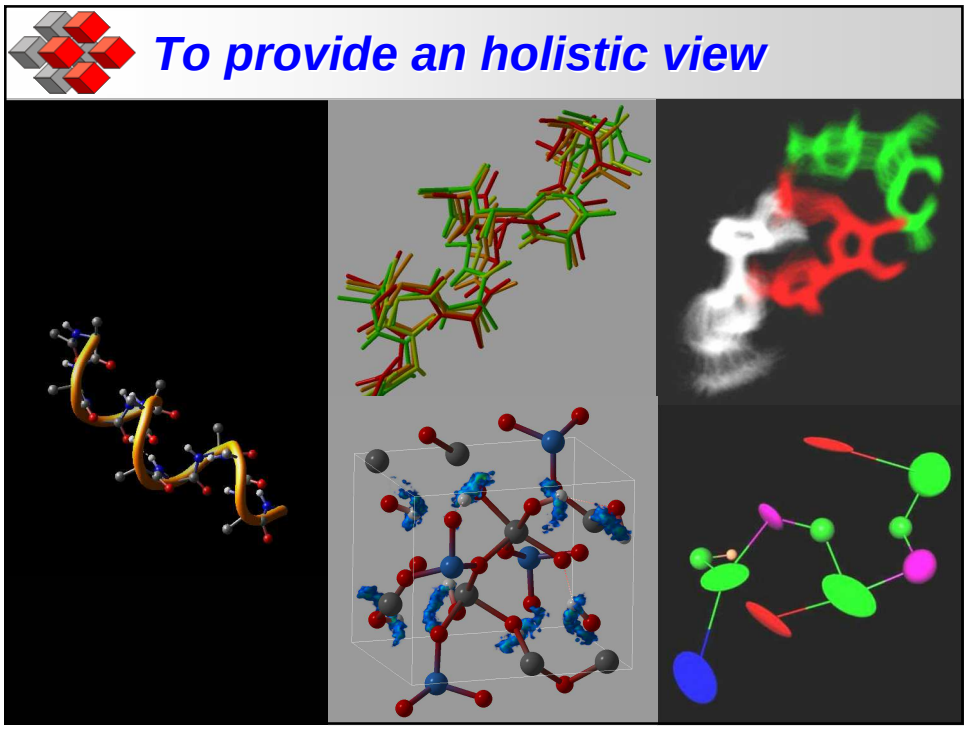


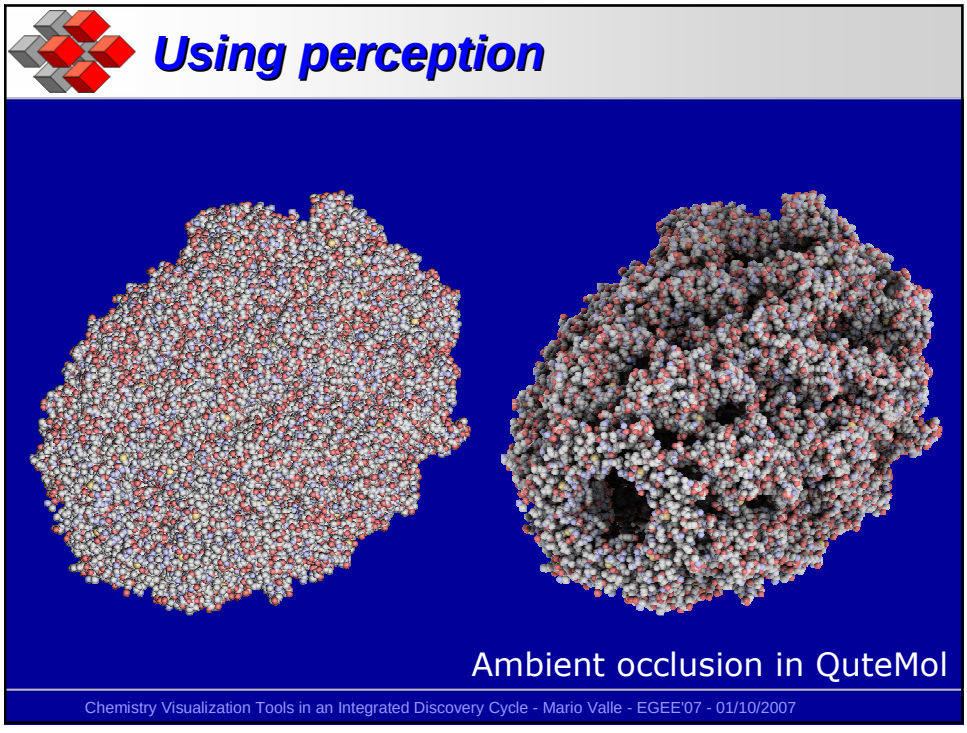


To remove unneeded details

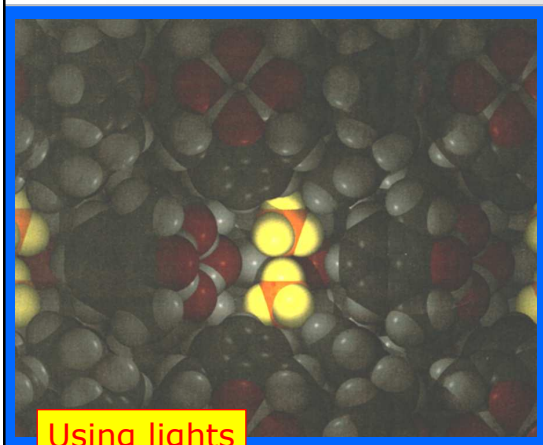


To see correlations

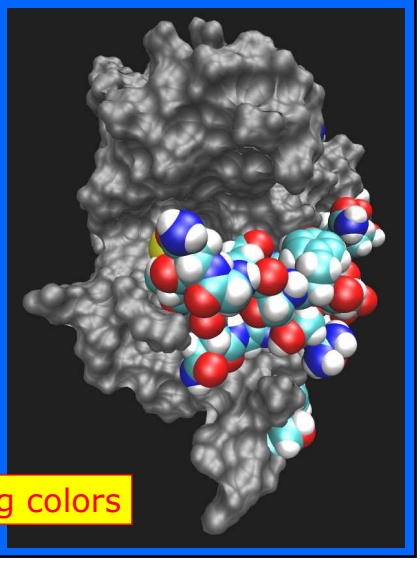

To provide an holistic view


Using perception
Ambient occlusion in QuteMol
Chemistry Visualization Tools in an Integrated Discovery Cycle - Mario Valle - EGEE'07 - 01/10/2007

And focusing attention

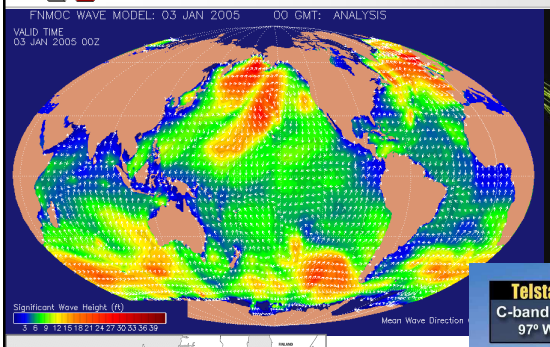


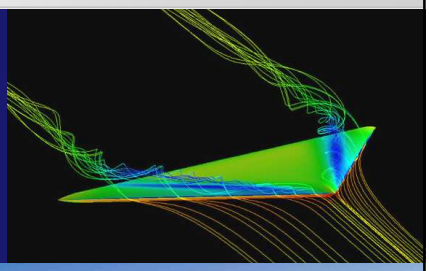
Using lights

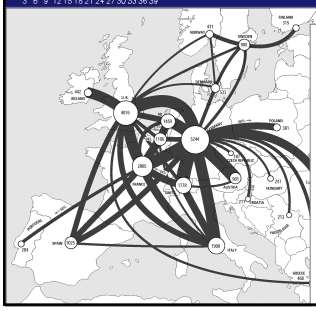


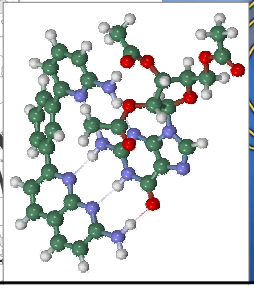
Using colors

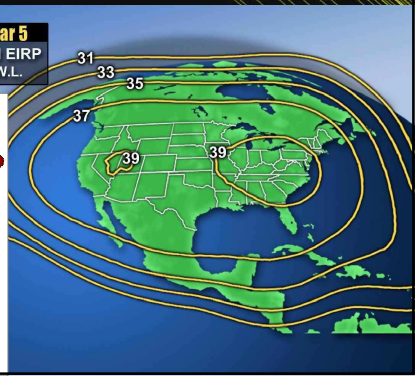
The goal: to see the unseen



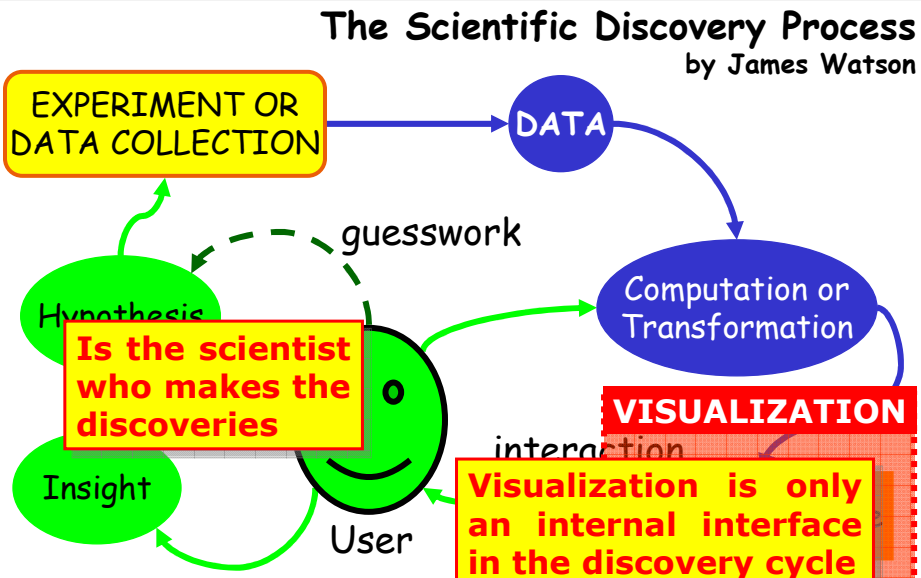




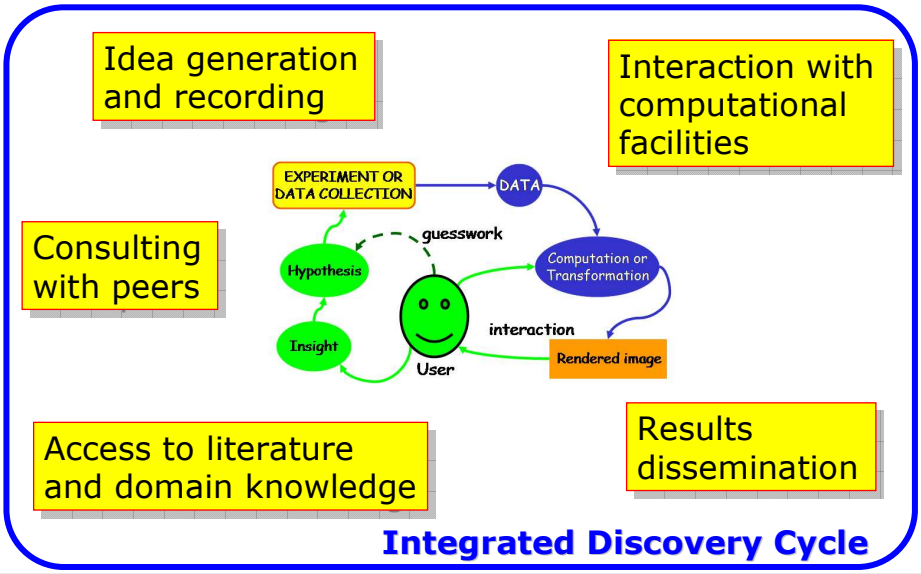




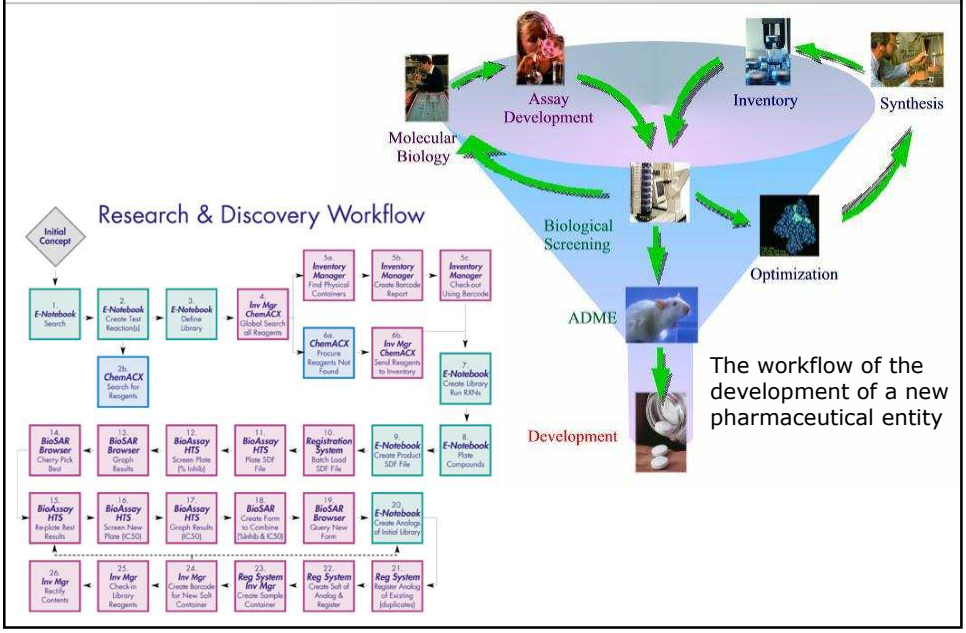
Visualization central & active role



More around the discovery cycle

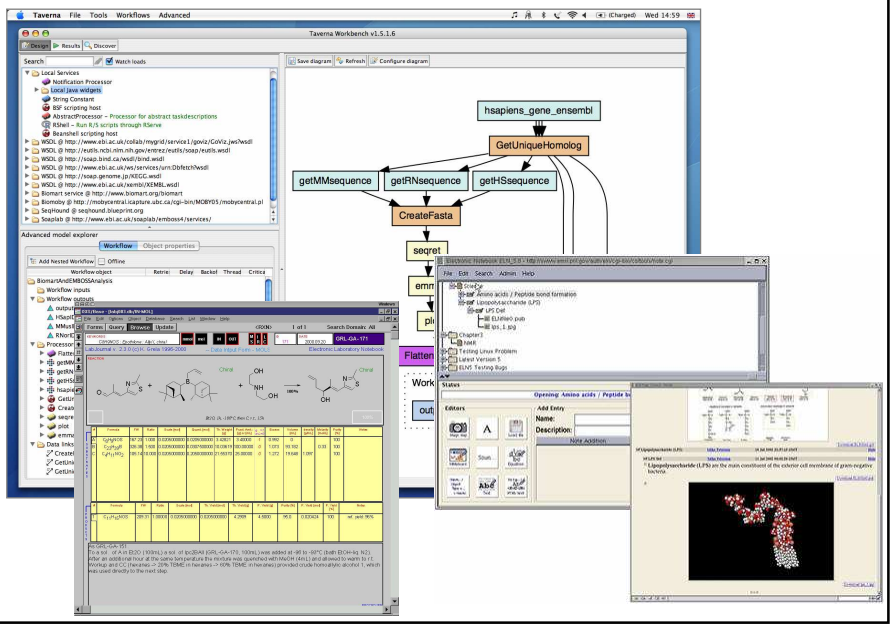


Integrated (industrial) cycles

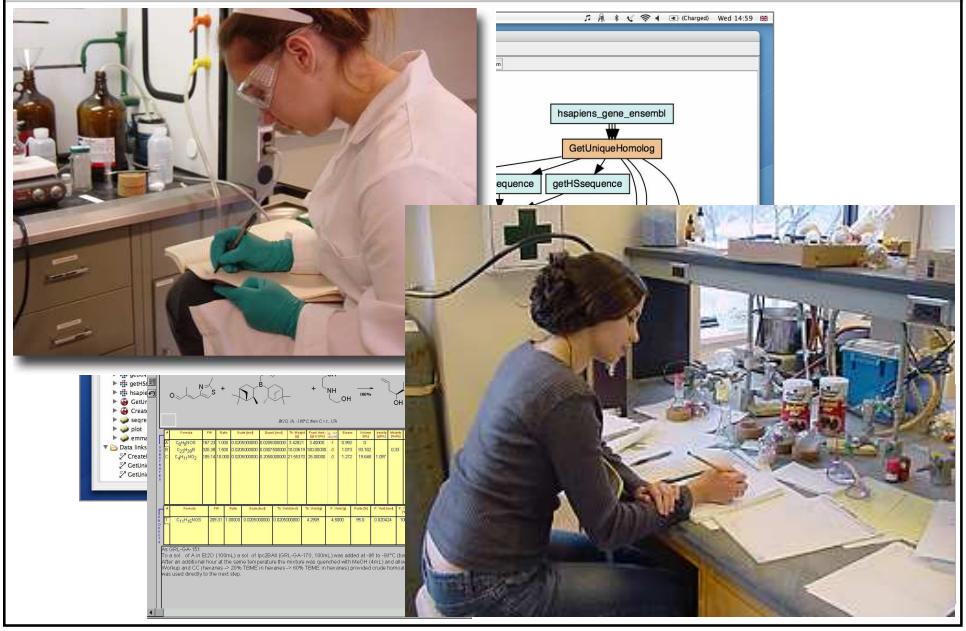


The workflow of the development of a new pharmaceutical entity

Integrated cycle tools support



But in academic research…

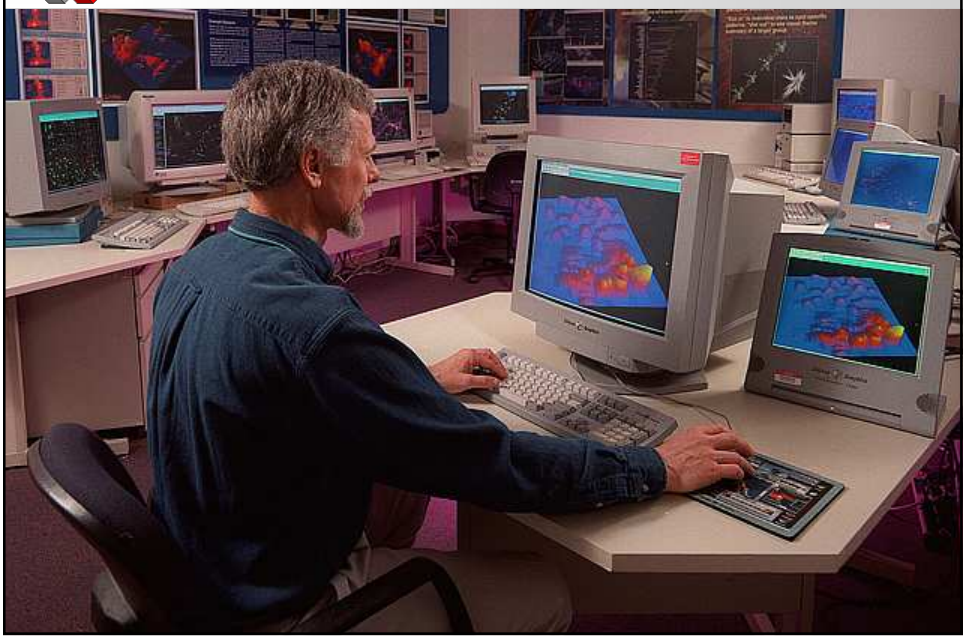


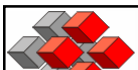
The real integrated environment

The real visualization cycle

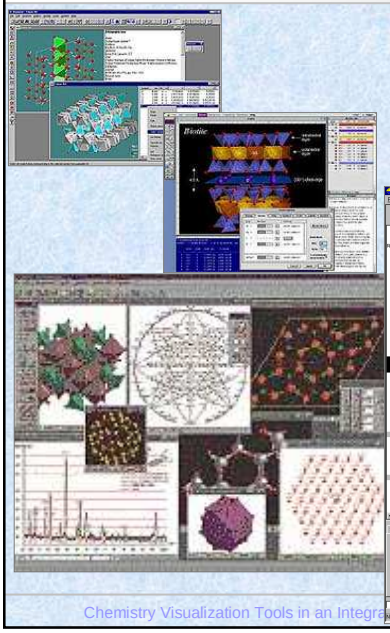
1. Guess the data format
2. Select the visualization tool
3. Try to load the data
4. Grumble
5. Retry data loading
6. Select at random a visualization technique
7. Navigate around the scene
8. Scratch your head
9. Try to remember what you want to see
10. Use another technique
11. Try to have another idea
12. Chase the right paper
13. Etcetera, etcetera …

Focus on the visualization tools

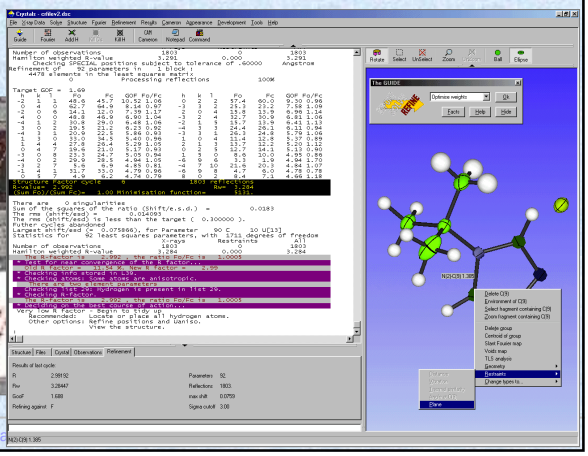




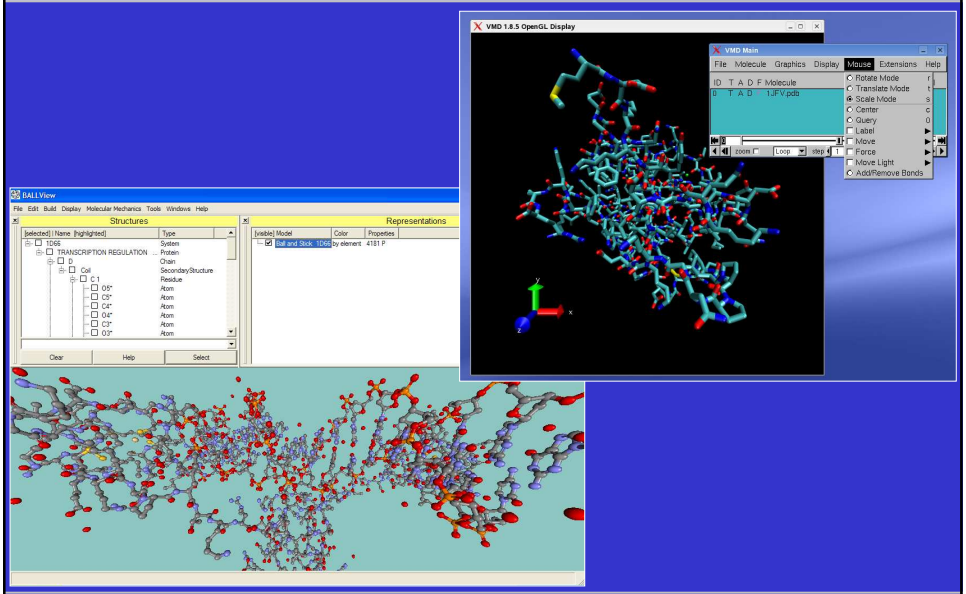
First difficulty: tools inflexibility

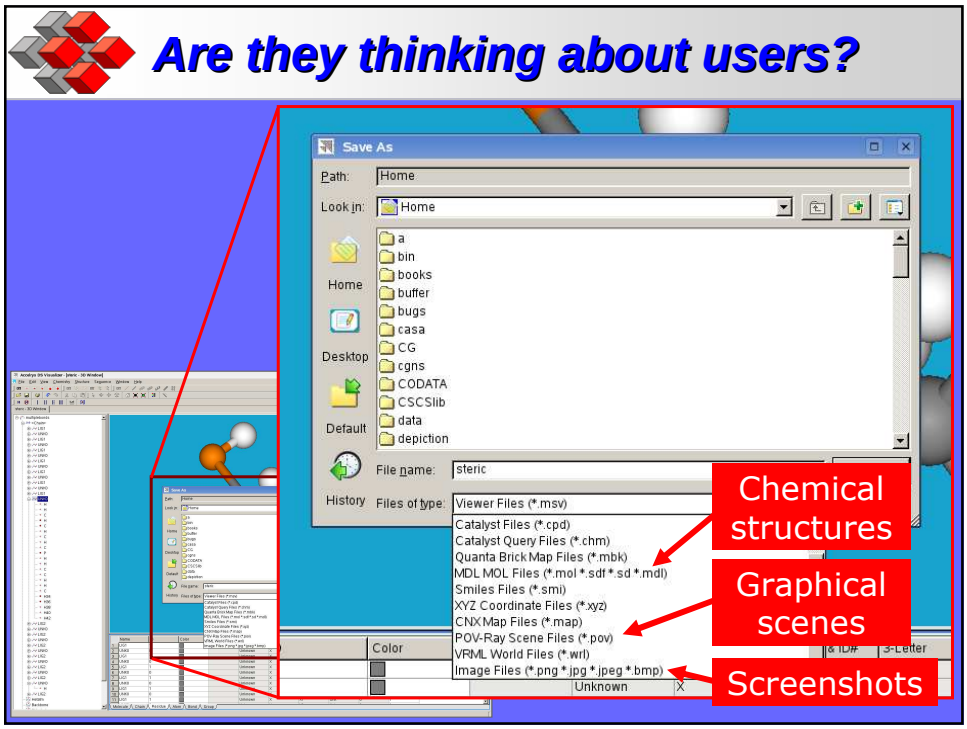


For example there are nice crystallography programs that do not support dynamic data and do not allow extension



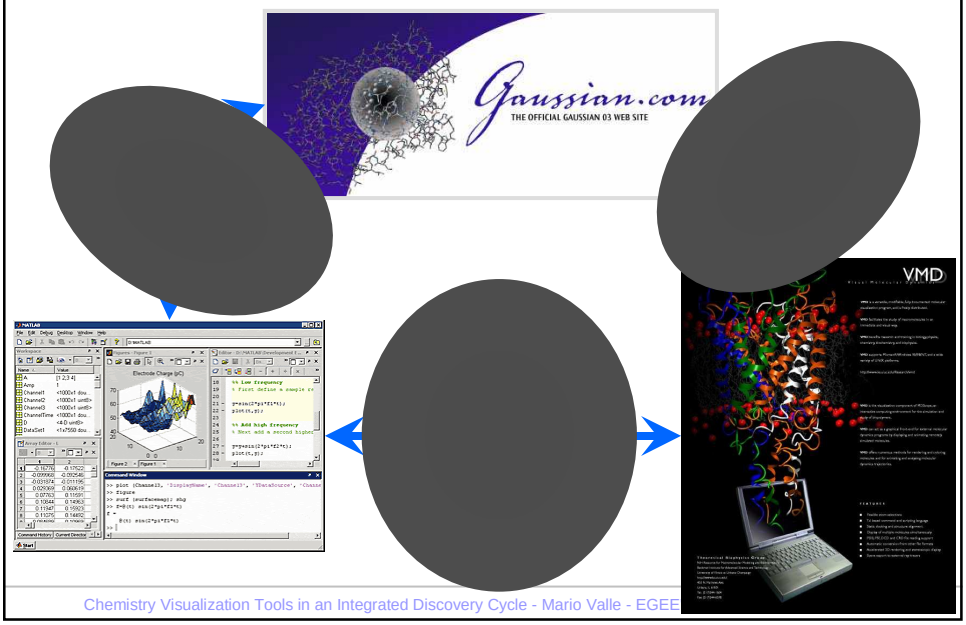
Usability inconveniences



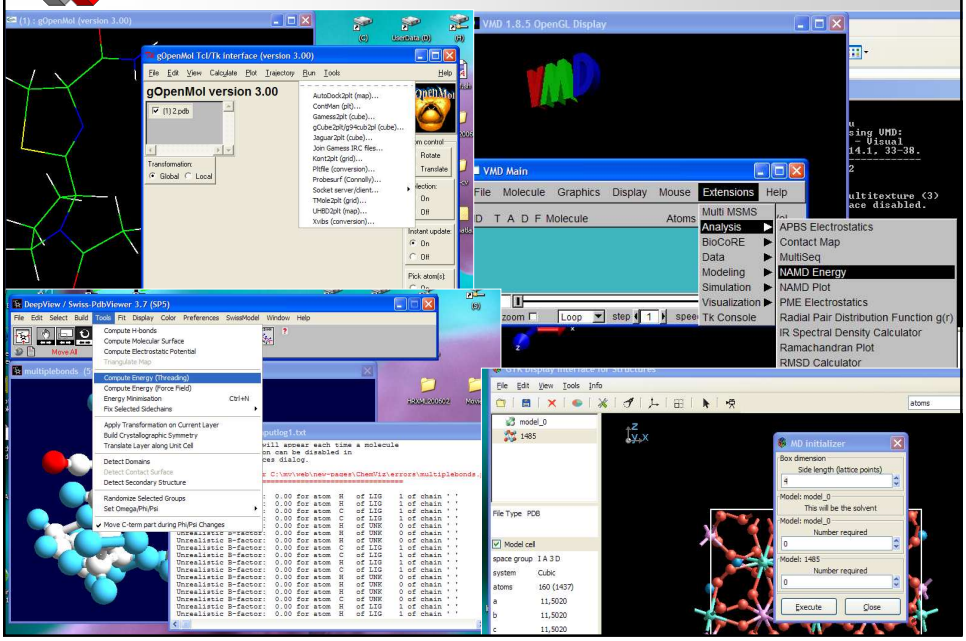

Are they thinking about users?
Save As
Path:
Home
Look in:
Home
a
bin
books
buffer
bugs
casa
CG
cgns
CODATA
CSCSlib
data
depiction
Home
Desktop
Default
History
File name:
steric
Files of type:
Viewer Files (*.msv)
Catalyst Files (*.cpd)
Catalyst Query Files (*.chm)
Quanta Brick Map Files (*.mbk)
MDL MOL Files (*.mol *.sdf *.sd *.mdl)
Smiles Files (*.smi)
XYZ Coordinate Files (*.xyz)
CNX Map Files (*.map)
POV-Ray Scene Files (*.pov)
VRML World Files (*.wrl)
Image Files (*.png *.jpg *.jpeg *.bmp)
Color
Unknown
Chemical structures
Graphical scenes
Screenshots


Games put focus on the action
File Edit Simulation Analyze View Effects Options Window Help
ATOM PROPERTIES
OBJECT PROPERTIES
60 Hz Mem 93 Sys L
YASARA & WHAT IF
Obj all SimOff 9780 Atoms

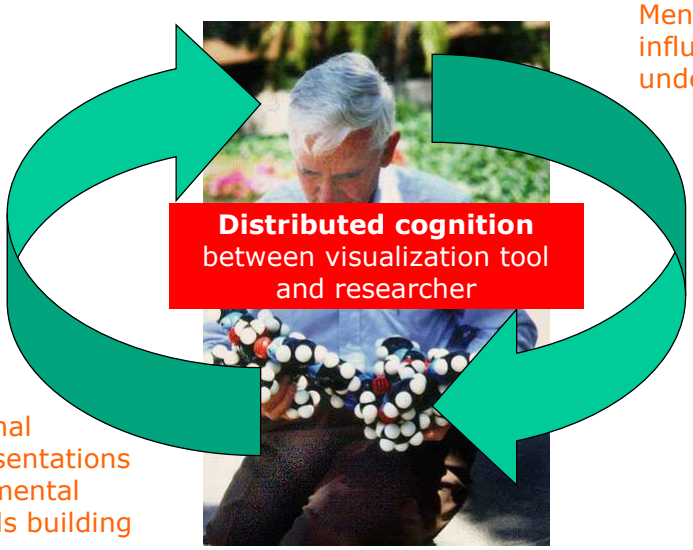
Difficult fit in the discovery loop



Menu-based integration only



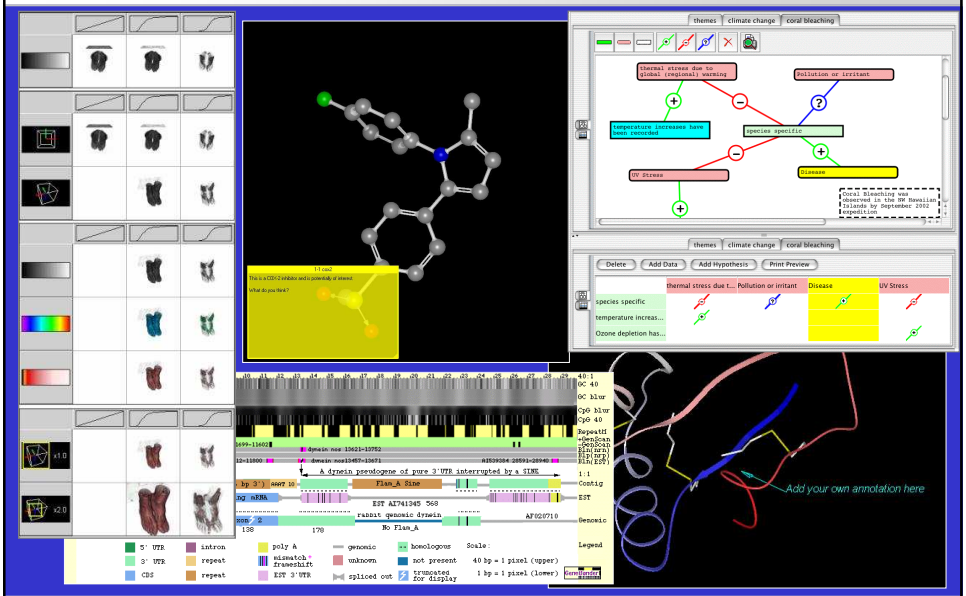
Integration in the thinking loop



Mental models influence understanding

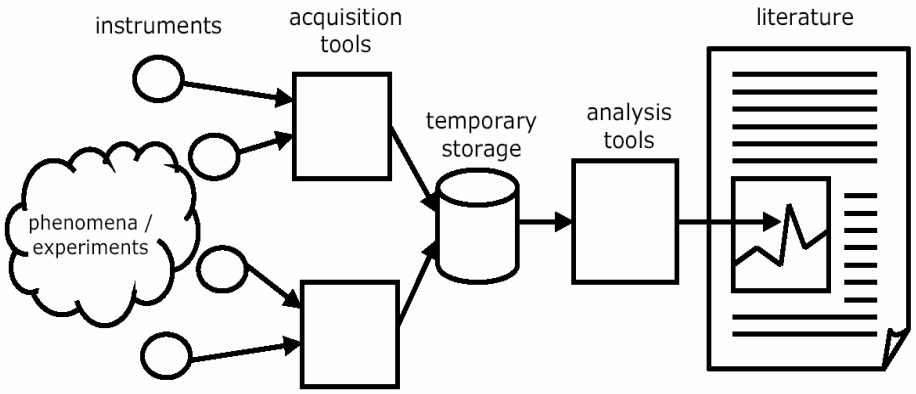
External representations help mental models building

Thinking process support



Integration with the data

Non existent.
It is often publish and forget …

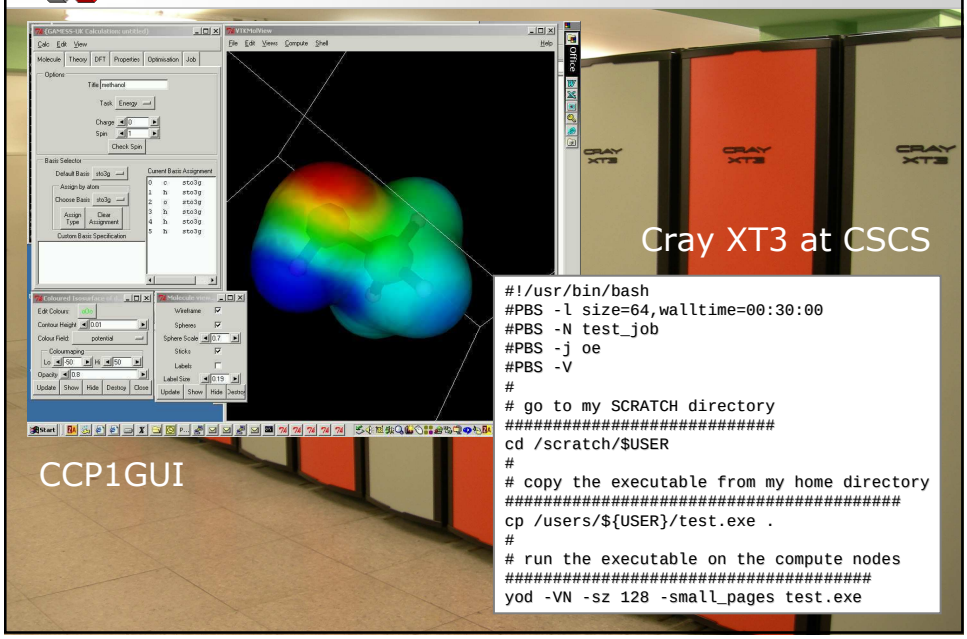


Data management issues

```
coco_jupiter2.pdb:HEADER    PROTEIN
coco_jupiter2.pdb:COMPND    coco_jupiter2.pdb
coco_jupiter2.pdb:AUTHOR    GENERATED BY BABEL 1.6

n20001.pdb:HEADER     n20001.pdb
n20001.pdb:TITLE      Nafion: N2_23-9 Polimeri con 4 Monomer
n20001.pdb:CMPND      time=  600.000               file.pdb
n20001.pdb:AUTHOR     generato dal programma hist_to_pdb

paolo.pdb:REMARK    GENERATED BY TRJCONV
paolo.pdb:HEADER    Benzene_1 t=   0.00000
paolo.pdb:REMARK    THIS IS A SIMULATION BOX

SPINDEN_OLD.pdb:PDBFIL pdb file created from cpmd Wannier file: SPINDEN

to_alpha2.pdb:REMARK File: to_alpha.pdb
to_alpha2.pdb:REMARK Atoms in molecule: 256
to_alpha2.pdb:REMARK This file created by progra
to_alpha2.pdb:REMARK --------------------------

TruduNew.pdb:REMARK    GENERATED BY ME
TruduNew.pdb:HEADER    Argon solid-liquid

why-no-hbonds.pdb:REMARK    GENERATED BY TRANS
why-no-hbonds.pdb:HEADER    360 H2O (TIP5P,1bar,300K) t=  10.00000
```

A quick survey of my chemistry file collection reveals that metadata are *"terra incognita"*

Job submission



Challenges to integration

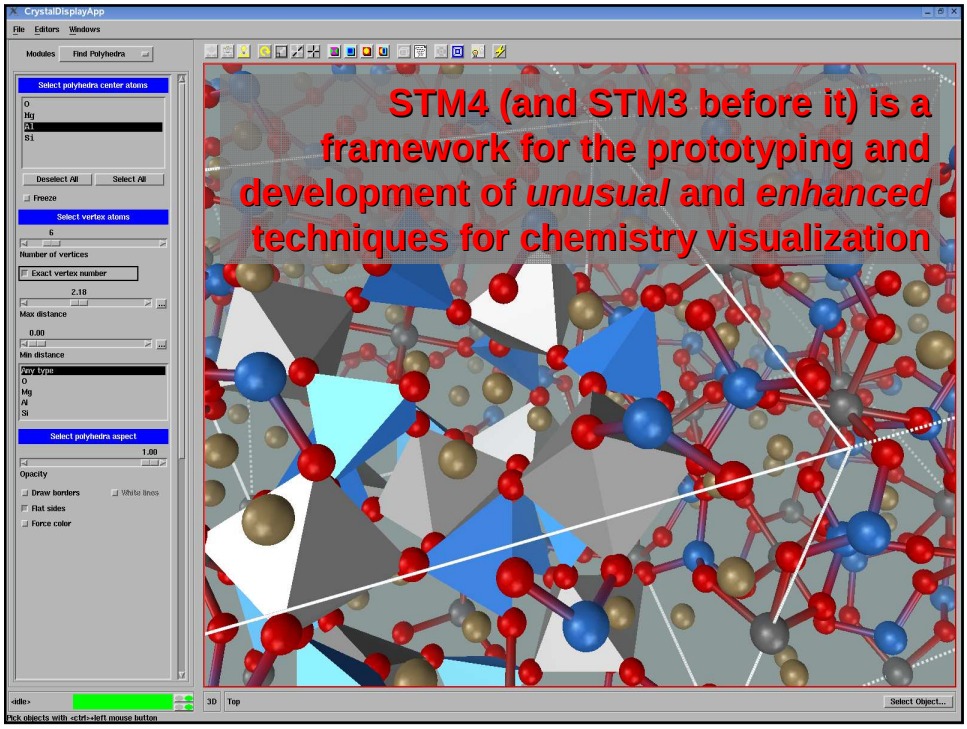
- Visualization is a proven, valuable tool to foster understanding
  - But it is only a part of the whole process
- It is not a technology problem
  - There are plenty of interesting tools, but we must make them usable
- Ideal solutions cannot be force-feed to the mean researcher
  - They want results from their research
- Researchers abandon their habits (and tools) only if really convinced it is worth the hassle
  - They invest time in their research not in learning new tools

Learn real integration needs

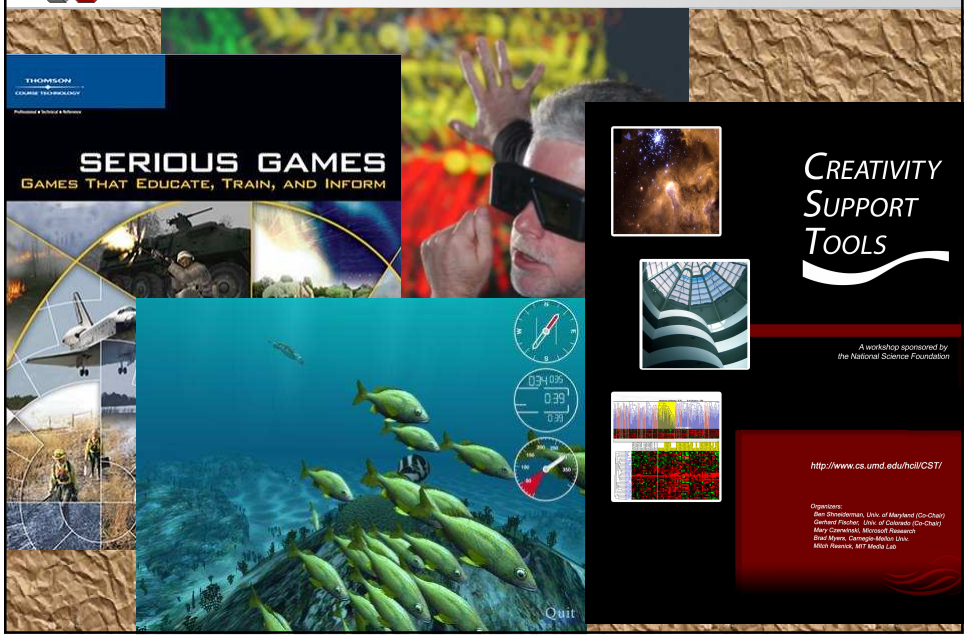
The first step is to increase dialog between visualization experts and chemistry researchers

- To stimulate chemistry researchers to ask for new functionalities in their visualization tools
- To make visualization experts learn what really helps the researchers' work
- To avoid solving the wrong problem

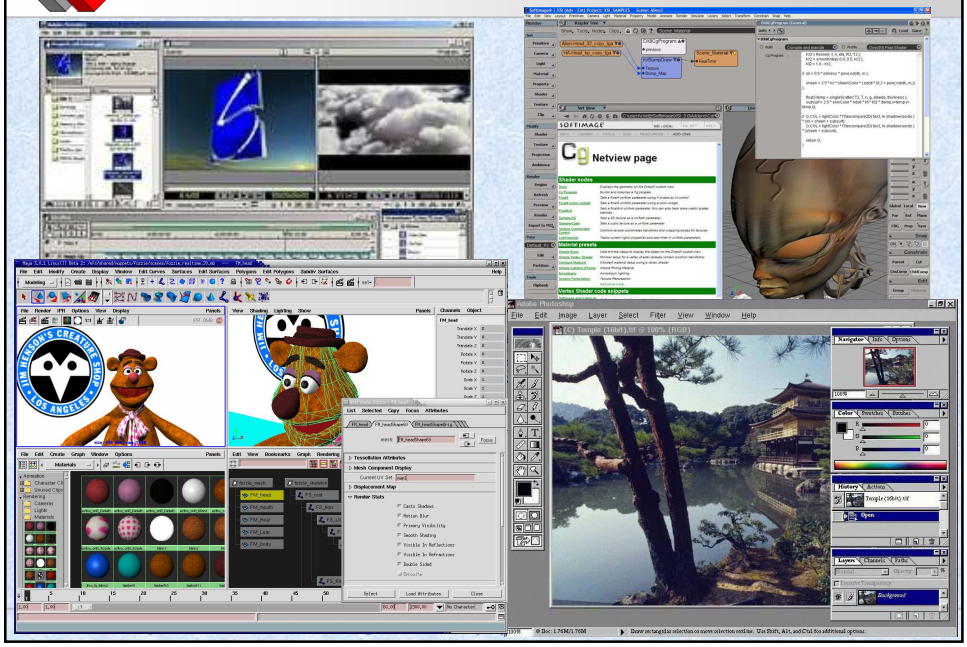

Prof. Michele Parrinello



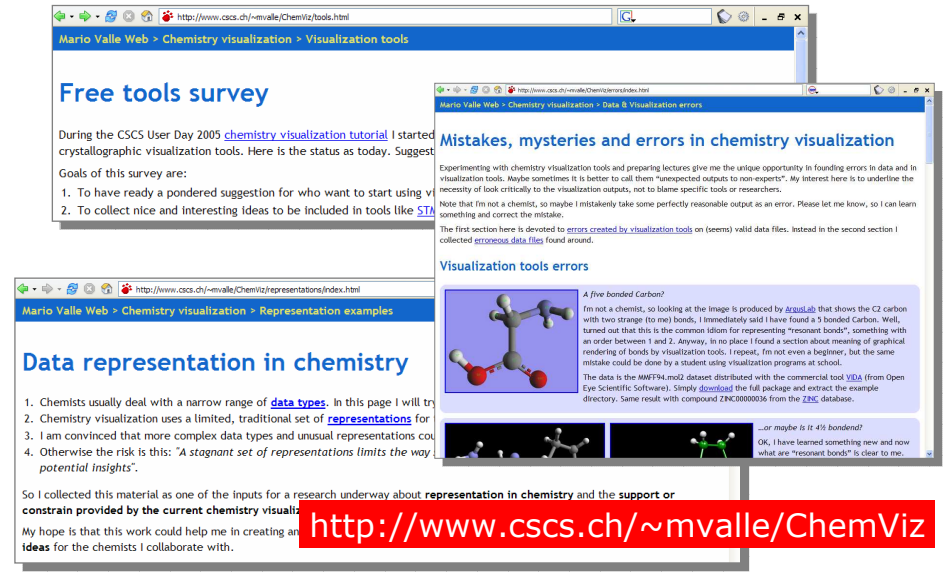
Collect ideas from other fields



Look at the "Perfect tools" around



Sharing results



The journey just started