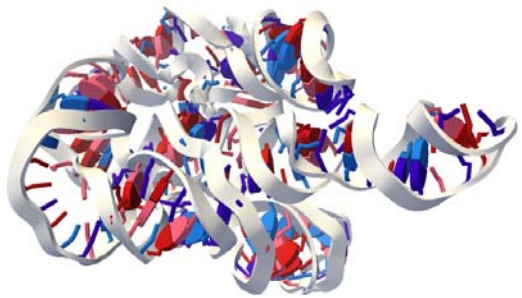
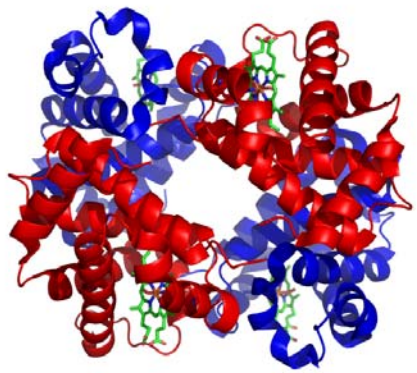
Mass Properties of Nucleic Acids and Proteins





BioE 215

May 4, 2007

Christopher Bruns

Coarse-grained representations in biology

Natural lump sizes:


atoms


residues


base pairs


duplexes, alpha helices, beta strands

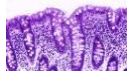

macromolecules

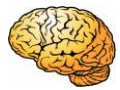

molecular assemblies


organelles

(picometer to micron scale)


cells


tissues


organs


limbs


organisms


societies

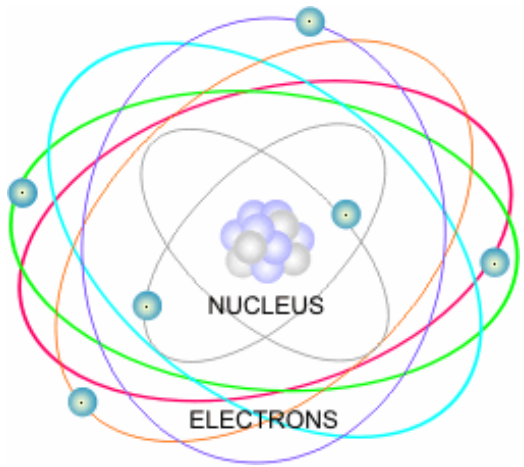

ecosystems


biospheres


intergalactic federations

(micron to mega-light-year scale)

Part I: Atoms

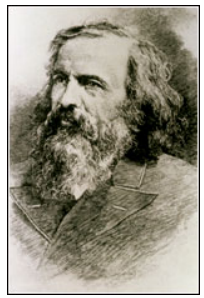


Atomic Theory

John Dalton
(1766-1844)

- Chemical elements are made of tiny particles called atoms
- All atoms of a given element are identical
- The atoms of a given element are different from those of any other element
- Atoms of one element can combine with atoms of other elements to form compounds. A given compound always has the same relative numbers of types of atoms.
- Atoms cannot be created, divided into smaller particles, nor destroyed in the chemical process. A chemical reaction simply changes the way atoms are grouped together.

Periodic table of the elements



Dmitri Mendeleev
(1834-1907)

hydrogen 1 **H** 1.0079																	helium 2 **He** 4.0026
lithium 3 **Li** 6.941	beryllium 4 **Be** 9.0122											boron 5 **B** 10.811	carbon 6 **C** 12.011	nitrogen 7 **N** 14.007	oxygen 8 **O** 15.999	fluorine 9 **F** 18.998	neon 10 **Ne** 20.180
sodium 11 **Na** 22.990	magnesium 12 **Mg** 24.305											aluminium 13 **Al** 26.982	silicon 14 **Si** 28.086	phosphorus 15 **P** 30.974	sulfur 16 **S** 32.065	chlorine 17 **Cl** 35.453	argon 18 **Ar** 39.948
potassium 19 **K** 39.098	calcium 20 **Ca** 40.078		scandium 21 **Sc** 44.956	titanium 22 **Ti** 47.867	vanadium 23 **V** 50.942	chromium 24 **Cr** 51.996	manganese 25 **Mn** 54.938	iron 26 **Fe** 55.845	cobalt 27 **Co** 58.933	nickel 28 **Ni** 58.693	copper 29 **Cu** 63.546	zinc 30 **Zn** 65.39	gallium 31 **Ga** 69.723	germanium 32 **Ge** 72.61	arsenic 33 **As** 74.922	selenium 34 **Se** 78.96	bromine 35 **Br** 79.904	krypton 36 **Kr** 83.80
rubidium 37 **Rb** 85.468	strontium 38 **Sr** 87.62		yttrium 39 **Y** 88.906	zirconium 40 **Zr** 91.224	niobium 41 **Nb** 92.906	molybdenum 42 **Mo** 95.94	technetium 43 **Tc** [98]	ruthenium 44 **Ru** 101.07	rhodium 45 **Rh** 102.91	palladium 46 **Pd** 106.42	silver 47 **Ag** 107.87	cadmium 48 **Cd** 112.41	indium 49 **In** 114.82	tin 50 **Sn** 118.71	antimony 51 **Sb** 121.76	tellurium 52 **Te** 127.60	iodine 53 **I** 126.90	xenon 54 **Xe** 131.29
caesium 55 **Cs** 132.91	barium 56 **Ba** 137.33	57-70 *	lutetium 71 **Lu** 174.97	hafnium 72 **Hf** 178.49	tantalum 73 **Ta** 180.95	tungsten 74 **W** 183.84	rhenium 75 **Re** 186.21	osmium 76 **Os** 190.23	iridium 77 **Ir** 192.22	platinum 78 **Pt** 195.08	gold 79 **Au** 196.97	mercury 80 **Hg** 200.59	thallium 81 **Tl** 204.38	lead 82 **Pb** 207.2	bismuth 83 **Bi** 208.98	polonium 84 **Po** [209]	astatine 85 **At** [210]	radon 86 **Rn** [222]
francium 87 **Fr** [223]	radium 88 **Ra** [226]	89-102 **	lawrencium 103 **Lr** [262]	rutherfordium 104 **Rf** [261]	dubnium 105 **Db** [262]	seaborgium 106 **Sg** [266]	bohrium 107 **Bh** [264]	hassium 108 **Hs** [269]	meitnerium 109 **Mt** [268]	ununnilium 110 **Uun** [271]	unununium 111 **Uuu** [272]	ununbium 112 **Uub** [277]		ununquadium 114 **Uuq** [289]				

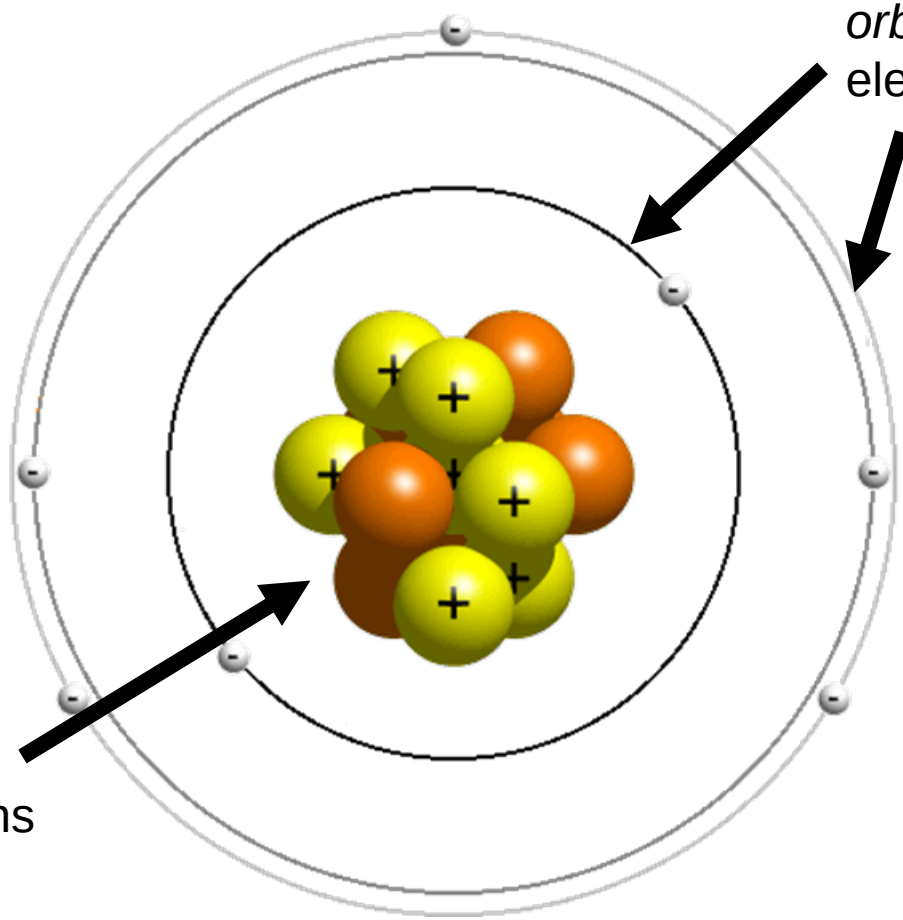
Key:
element name
atomic number
symbol
atomic weight (mean relative mass)

*lanthanoids	lanthanum 57 **La** 138.91	cerium 58 **Ce** 140.12	praseodymium 59 **Pr** 140.91	neodymium 60 **Nd** 144.24	promethium 61 **Pm** [145]	samarium 62 **Sm** 150.36	europium 63 **Eu** 151.96	gadolinium 64 **Gd** 157.25	terbium 65 **Tb** 158.93	dysprosium 66 **Dy** 162.50	holmium 67 **Ho** 164.93	erbium 68 **Er** 167.26	thulium 69 **Tm** 168.93	ytterbium 70 **Yb** 173.04
**actinoids	actinium 89 **Ac** [227]	thorium 90 **Th** 232.04	protactinium 91 **Pa** 231.04	uranium 92 **U** 238.03	neptunium 93 **Np** [237]	plutonium 94 **Pu** [244]	americium 95 **Am** [243]	curium 96 **Cm** [247]	berkelium 97 **Bk** [247]	californium 98 **Cf** [251]	einsteinium 99 **Es** [252]	fermium 100 **Fm** [257]	mendelevium 101 **Md** [258]	nobelium 102 **No** [259]

http://fermat.csci.unt.edu/~mikler/Courses/biocomputing/resources/periodic_table.gif

Bohr model of the atom

(not to scale)



orbitals, containing electrons

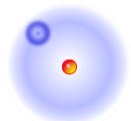
nucleus, containing protons and neutrons

Niels Bohr
(1885-1962)

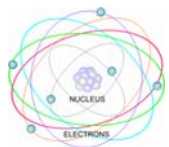
Atomic length scale

- Macroscopic length units
  - meter
- Atomic length units
  - nanometer = 10^{-9} m
  - Ångstrom = 10^{-10} m
  - picometer = 10^{-12} m

- Diameter of hydrogen atom = 2.4 Ångstroms

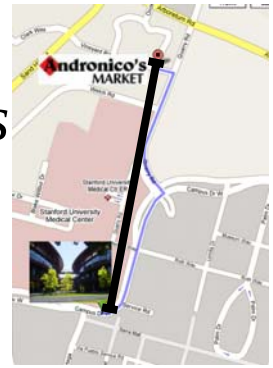


- Diameter of carbon atom = 3.4 Ångstroms

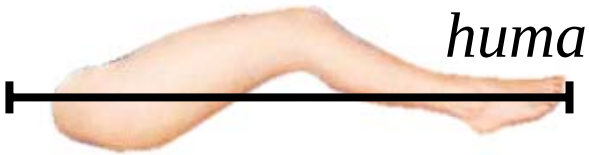


10^{3} kilometer

Clark center to Andronico's

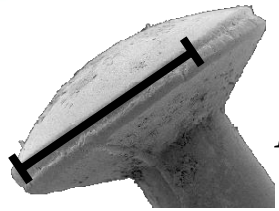


1 meter



human leg

10^{-3} millimeter

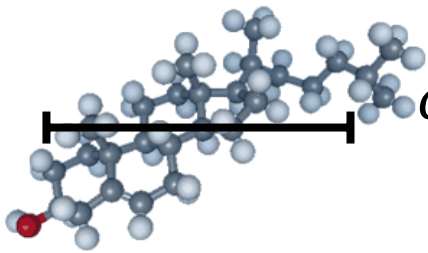


pin head

10^{-6} micrometer (micron)

sperm neck

10^{-9} nanometer



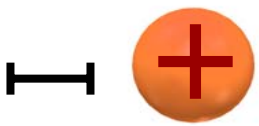
cholesterol molecule

10^{-10} Ångstrom

hydrogen atom

10^{-12} picometer

10^{-15} femtometer

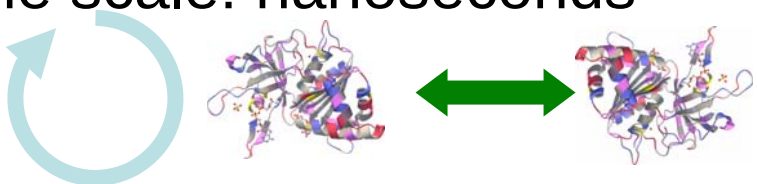


proton

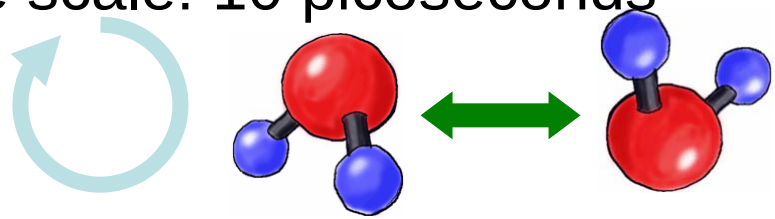
Atomic time scale

- Macroscopic time unit: second
- Atomic simulation time units
  - picosecond = 10^{-12} s
  - femtosecond =10^{-15} s
- Protein folding time scale: microseconds to hours

- Protein tumbling time scale: nanoseconds



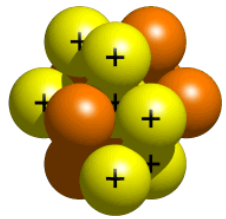
- Water tumbling time scale: 10 picoseconds



$$F = ma$$

Four fundamental forces of physics

1. Gravity: unimportant at atomic scale
2. Electromagnetic forces: everything else
3. Strong nuclear force
4. Weak nuclear force



Quantum vs. classical mechanics

Quantum mechanics

- Schrödinger equation

  $H\psi = E\psi$
- Expensive computation
- Especially useful at small size scales
- More precise
- Orbitals, wave functions, quantized spin
- Explains bonding and collisions between atoms

Classical mechanics

- Newton's second law

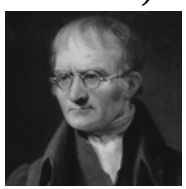
  $F = ma$
- Easier computation
- Especially useful at large size scales
- Velocities, accelerations, continuously differentiable angular velocity

$$F = ma$$

Atomic mass scale

Amedeo Avogadro
(1776-1856)

Johann Josef
Loschmidt
(1821-1895)



John Dalton
(1766-1844)

- Avogadro constant
  - $N_A = (6.022\,141\,79 \pm 0.000\,000\,30) \times 10^{23}$ mole^{-1}
  - The number of carbon-12 (^{12}C) atoms in 12 grams (0.012 kg) of unbound carbon-12 in its rest-energy electronic state
  - Basis for atomic mass units
  - Terminology: N_A atoms = 1 *mole* of atoms
- Macroscopic mass units: kilogram, gram
- Atomic mass units

  1 Dalton = 1 atomic mass unit (amu) =
  $1/N_A$ grams = 1.66×10^{-27} kg

  mass of hydrogen atom: 1.008 Daltons
  mass of uranium atom: 238.03 Daltons

Atomic energy scale

(not nuclear reaction energy scale…)

Force is the derivative of energy

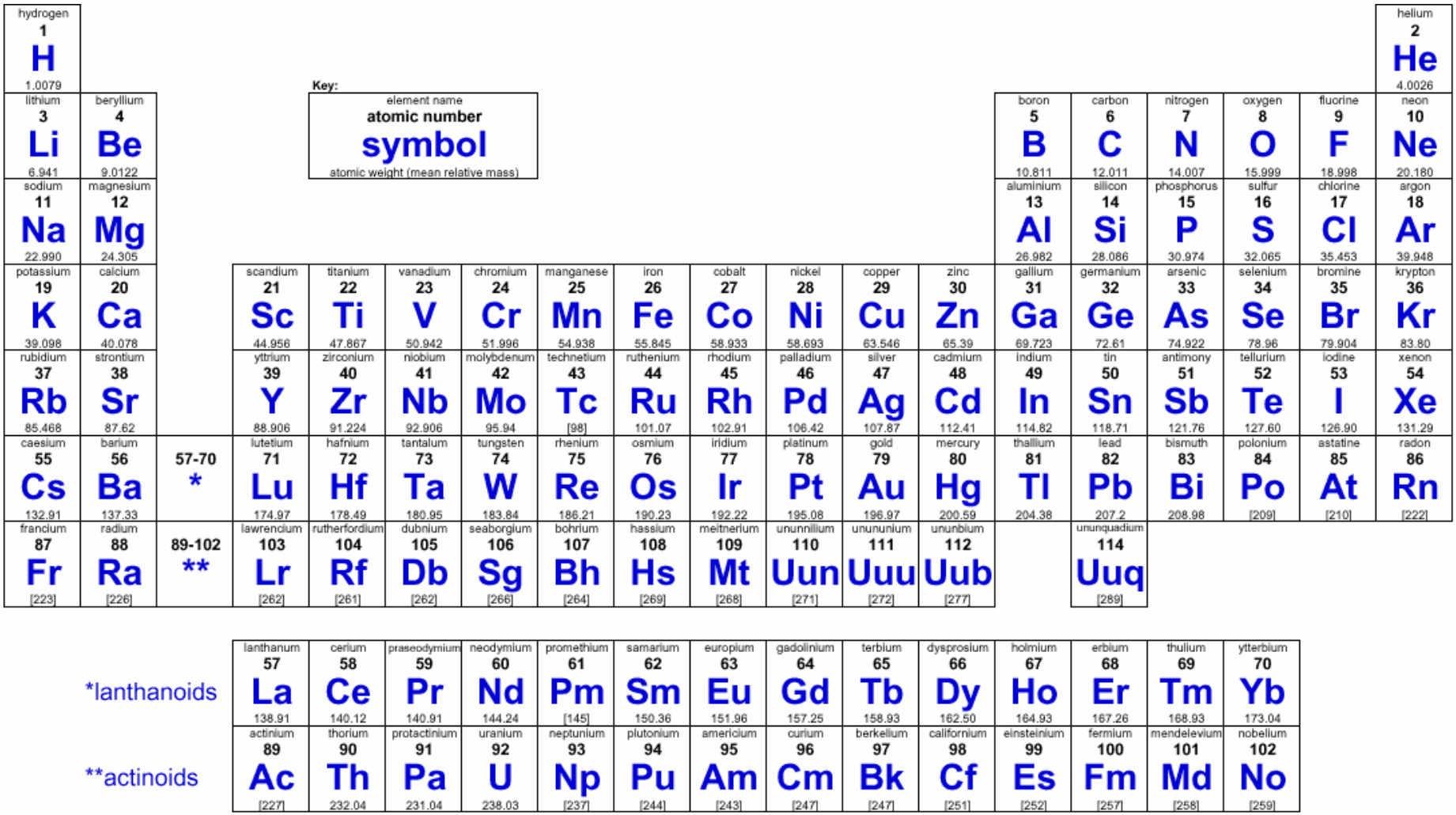
Macroscopic energy units

- joule = 1 kg m^2 / s^2

Atomic energy units

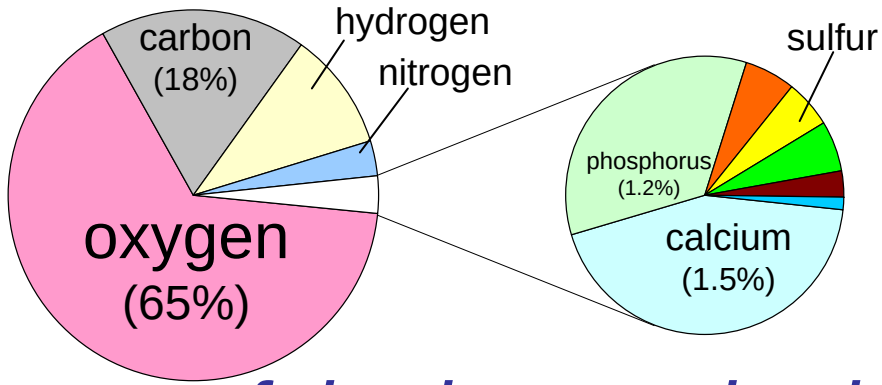
- kilojoule per mole = $1000/N_A$ joules
- kilocalorie per mole = $4184/N_A$ joules

Periodic table of the elements



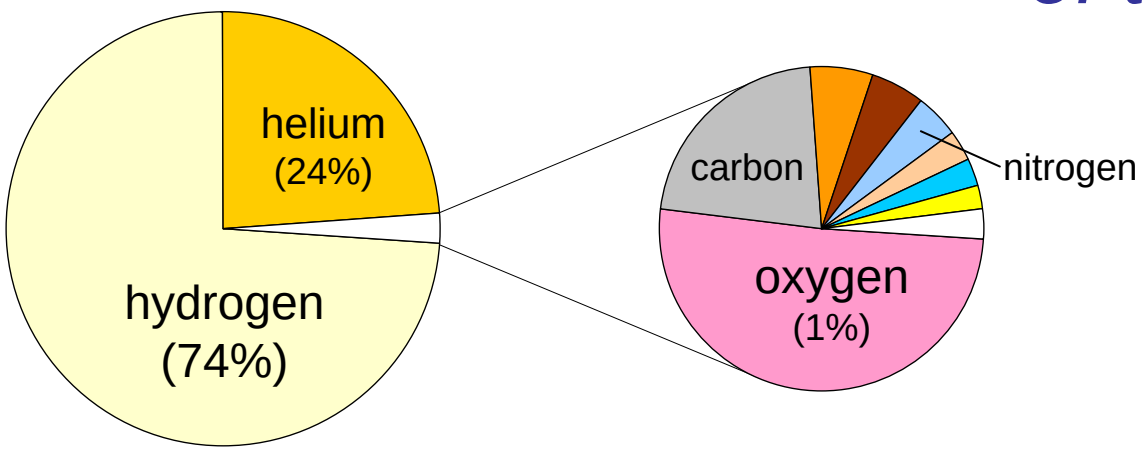
http://fermat.csci.unt.edu/~mikler/Courses/biocomputing/resources/periodic_table.gif

Elemental composition (by mass)



of the human body

of the universe



Determining atomic mass

- Determined experimentally
- Atomic mass can be looked up in a Periodic Table, in the CRC Handbook of Chemistry and Physics, or any of many other reference sources
- ChemicalElement.cpp
- Atomic mass is NOT exactly the sum of the masses of the electrons, protons, and neutrons in the atom
  - "$E = mc^2$" has a significant effect in atom formation.
  - Those special nuclear forces do affect the final mass.

Isotopes

- Same element, but differs in number of neutrons in the nucleus
- Changes mass, but does not affect chemical properties (much)
- Mass found in periodic table is usually the natural abundance weighted average of isotope masses
- That's why although carbon-12 is *defined* as the mass standard, the periodic table shows the mass of carbon as 12.011, rather than 12.0000000…
- Why use weighted average?
  - You probably don't know which isotopes are in your molecule
  - Easier to measure experimentally

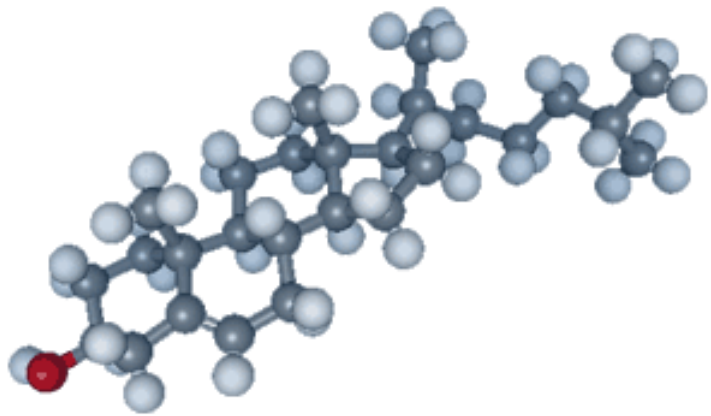
An atom “is” a point mass

- For Clintonesque definition of “is”
- Mass is concentrated in the center
  - Nucleus holds 4999/5000ths of the atomic mass
  - Nucleus occupies one billionth of the atomic volume
- Internal angular velocity and angular momentum don’t apply
  - Quantum spin values are not continuous

Atoms are *not* point charges

- Electron cloud occupies most of the volume, and possesses all of the negative charge
- But classical mechanics usually treats atoms as if they were point charges

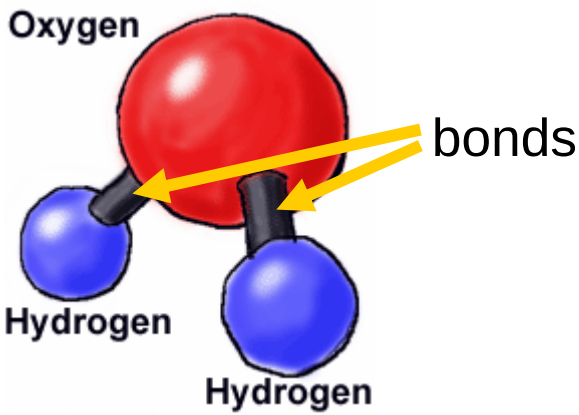
Part II: Molecules



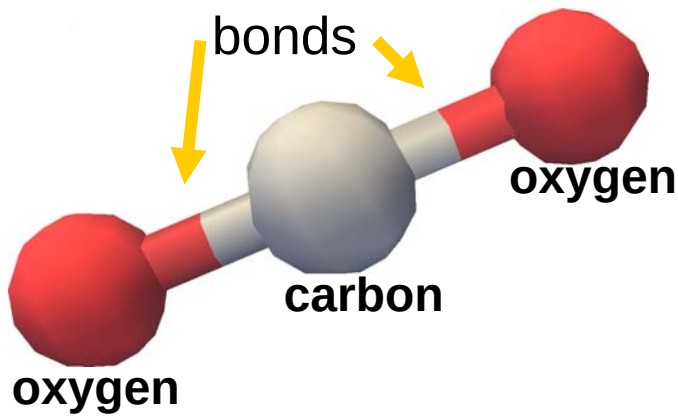
What is a molecule?

- a group of *atoms* connected by *covalent bonds (contrasted with ionic bonds and hydrogen bonds)*

water molecule (H_2O)



carbon dioxide (CO_2)



Combining masses

Molecular mass *is* the sum of the masses of the atoms in the molecule

- Effect of energy of molecule formation upon mass is negligible.

(unlike atomic mass)

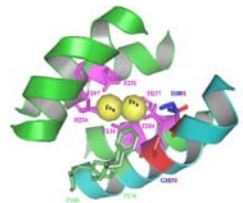
Molecule mass properties are SO simple

- Compared to, say, musculoskeletal mass properties
- Because atoms “are” point masses
- Good thing, because we can neither see, feel, touch nor savor molecule mass properties
- CAVEAT: Molecules are NOT rigid bodies, but it may be useful to model them as rigid bodies.
- Molecule mass: $$m^{\text{mol}} = \sum_{\text{atoms}} m^{\text{atom}_i}$$
- Molecule center of mass: $$\vec{r}^{\,\text{mol}_{cm}/\vec{O}} = \frac{1}{m^{\text{mol}}} \sum_{\text{atoms}} m^{\text{atom}_i} * \vec{r}^{\,\text{atom}_i/\vec{O}}$$
- Molecule inertia dyadic:
$$I^{\text{mol}/\vec{O}} = \sum_{\text{atoms}} m^{\text{atom}_i} * [\underline{1} * (\vec{r}^{\,\text{atom}_i/\vec{O}} \cdot \vec{r}^{\,\text{atom}_i/\vec{O}}) - \vec{r}^{\,\text{atom}_i/\vec{O}} * \vec{r}^{\,\text{atom}_i/\vec{O}}]$$

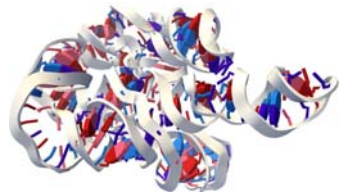
Viewing Molecule Models

- RasMol
  - Roger Sayle (1996)
  - http://www.umass.edu/microbio/rasmol/index2.htm


- Pymol
  - Warren Delano
  - http://pymol.sourceforge.net/

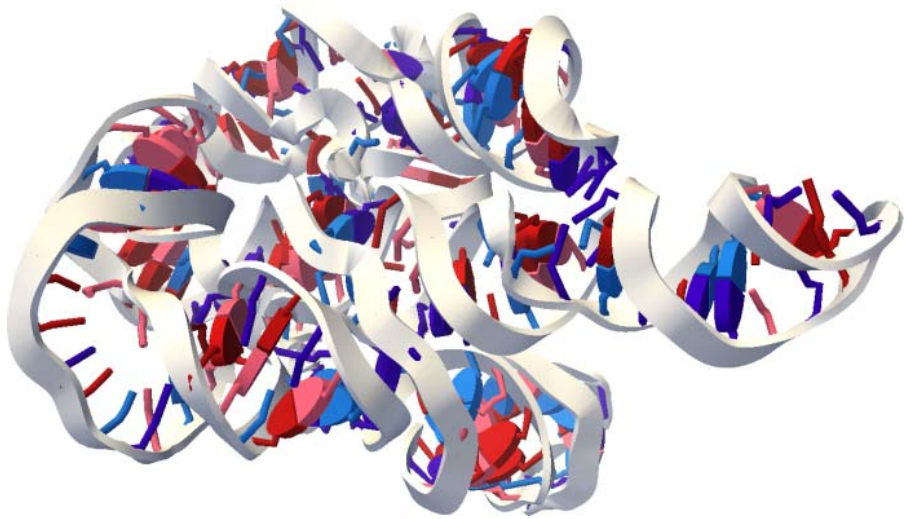



- Tornado
  - Christopher Bruns
  - https://simtk.org/home/rna-viz-proto

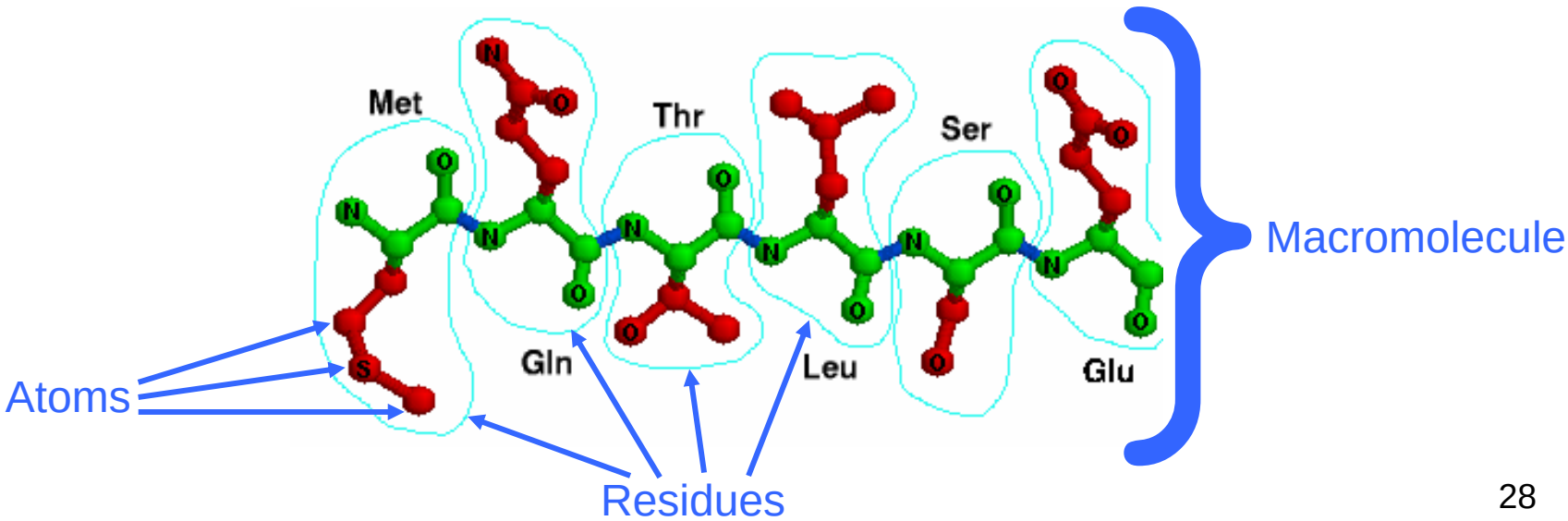
- Chimera, Chime, VMD, many others
  - not necessarily free

Part III: Nucleic Acids and Proteins

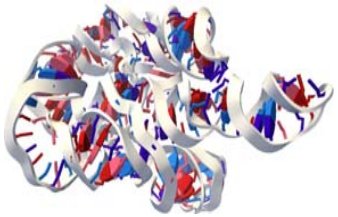


Macromolecules, residues, and atoms

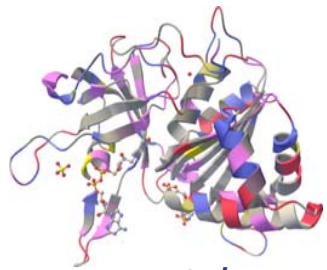
- RNA, DNA, and protein are large molecules called *macromolecules*
- Macromolecules consist of chains of *residues*



Macromolecules, residues and atoms



RNA



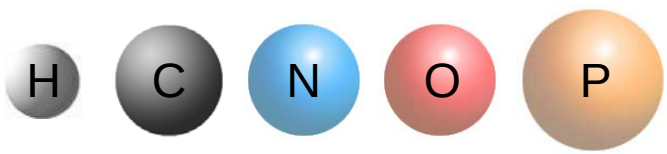
protein

Nucleic Acids (RNA and DNA)

adenylate

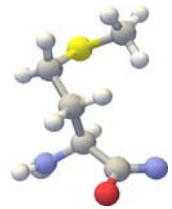
5 nucleotide residues
- adenylate (A)
- cytidylate (C)
- guanylate (G)
- thymidylate (T)
- uridylate (U)

- 5 chemical elements
  - hydrogen (H)
  - carbon (C)
  - nitrogen (N)
  - oxygen (O)
  - phosphorus (P)



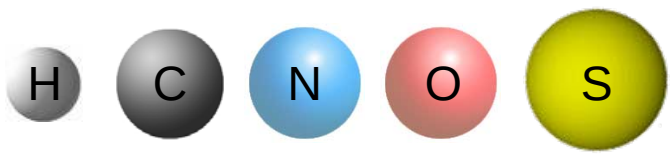
Proteins

- 20 amino acid Residues
  - alanine, arginine, aspartate, asparagine, cysteine, glutamate, glutamine, glycine, histidine, isoleucine, leucine, lysine, methionine, phenylalanine, proline, serine, threonine, tryptophan, tyrosine, valine

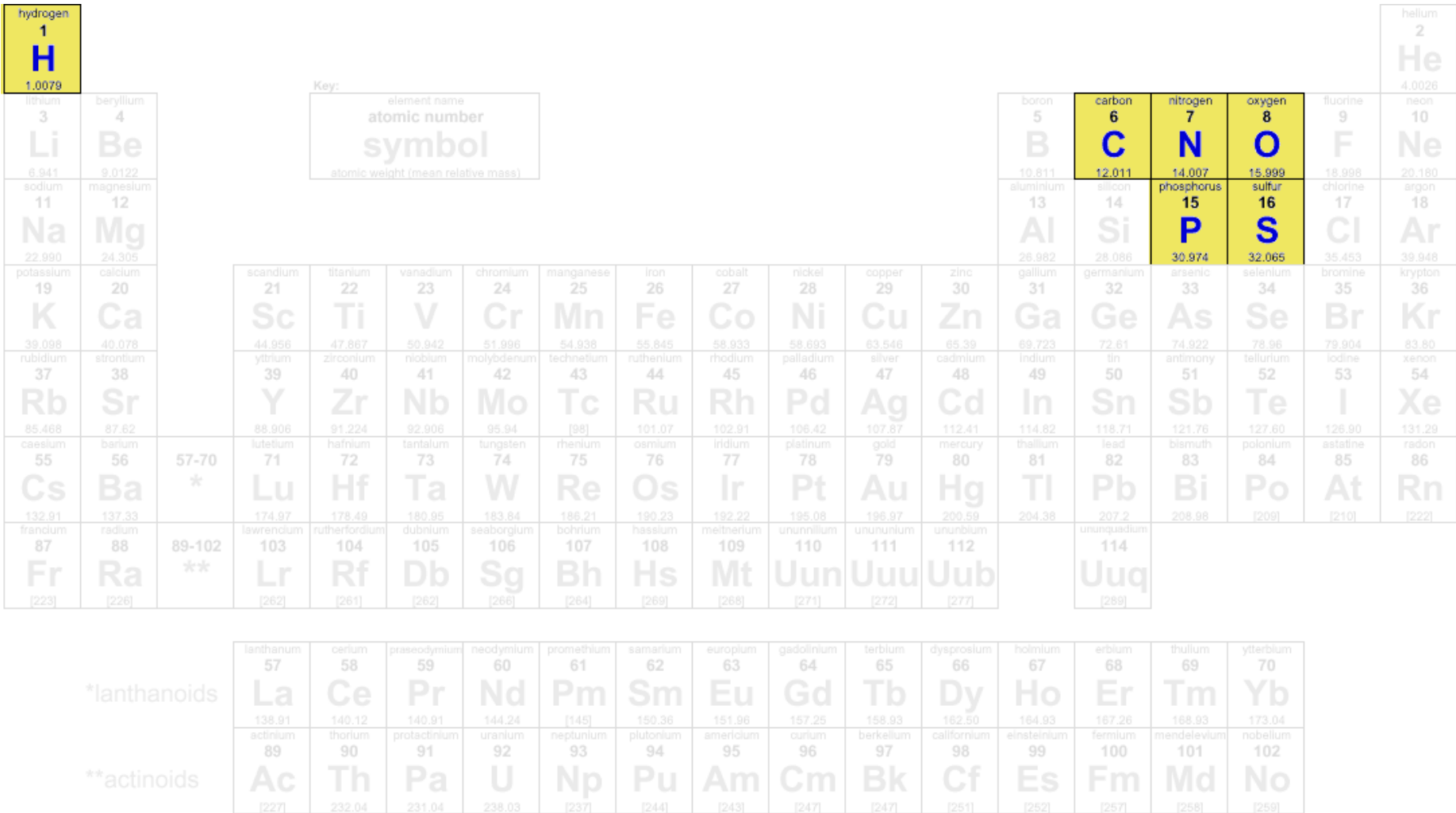


methionine

- 5 chemical elements
  - hydrogen (H)
  - carbon (C)
  - nitrogen (N)
  - oxygen (O)
  - sulfur (S)

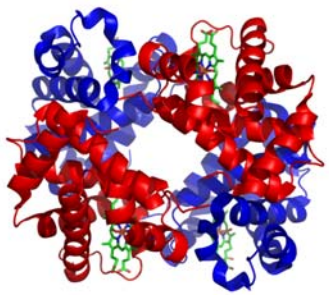


Six elements of DNA, RNA and proteins



Example protein: Hemoglobin

- Red blood cell is 99% hemoglobin (dry weight)
- Hemoglobin mass = 62000 Da



Hemoglobin residues

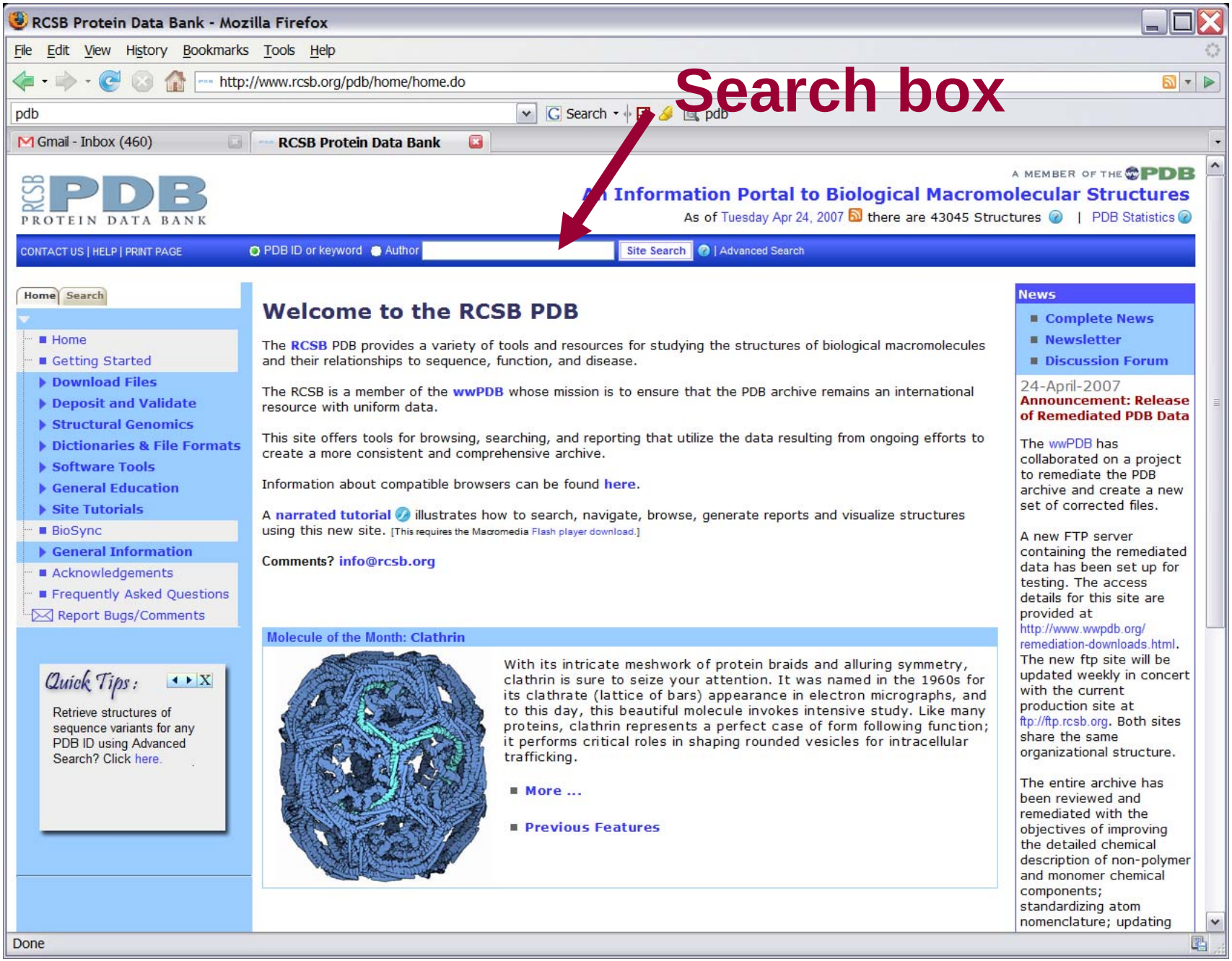
A	Ala	alanine	72
R	Arg	arginine	12
N	Asn	asparagine	20
D	Asp	aspartate	30
Q	Gln	glutamine	8
E	Glu	glutamate	24
G	Gly	glycine	40
H	His	histidine	38
I	Ile	isoleucine	0
L	Leu	leucine	72
K	Lys	lysine	44
M	Met	methionine	6
F	Phe	phenylalanine	30
P	Pro	proline	28
S	Ser	serine	32
T	Thr	threonine	32
Y	Tyr	tyrosine	12
V	Val	valine	62
W	Trp	tryptophan	6
C	Cys	cysteine	6

Hemoglobin Atoms

carbon	2818
nitrogen	764
oxygen	788
sulfur	12
hydrogen	4376
iron	4

PDB: The Protein Data Bank

- http://www.rcsb.org/pdb/
- Repository of DNA, RNA, and protein atomic structures
- Contains experimental results, not perfect models for simulation
- Entries identified by 4 character ID, e.g. “1MRP”, “1GRZ”
- Use “advanced search” at PDB web site to find structures.



RCSB PDB : Structure Explorer - Mozilla Firefox

File Edit View History Bookmarks Tools Help

http://www.rcsb.org/pdb/explore/explore.do?structureId=1MRP

pdb

Search pdb

Gmail - Inbox (461) | RCSB PDB : Structure Expl...

RCSB PDB PROTEIN DATA BANK

4-character ID

A MEMBER OF THE PDB

An Information Portal to Biological Macromolecular Structures

As of Tuesday Apr 24, 2007 there are 43045 Structures | PDB Statistics

CONTACT US | HELP | PRINT PAGE | PDB ID | Keyword | Author | Site Search | Advanced Search | RSS Feed for the Latest Released Structures

Home | Search | Structure | Queries

Help | Structure Summary | Biology & Chemistry | Materials & Methods | Sequence Details | Geometry

- 1MRP
- Download Files
  - PDB File
  - PDB gz
  - PDB File (Header)
  - mmCIF File
  - mmCIF gz
  - mmCIF File (Header)
  - PDBML/XML File
  - PDBML/XML gz
  - PDBML/XML File (Header)
  - Structure Factors File
  - Structure Factors gz
  - Biological Unit Coordinates
- FASTA Sequence
- Display Files
- Display Molecule
- Structural Reports
- External Links
- Structure Analysis
- Help

Quick Tips :

To view **sequence details** of this structure click on the *Sequence Details* tab above the summary page.

1MRP DOI 10.2210/pdb1mrp/pdb

Red - Derived Information

Title	FERRIC-BINDING PROTEIN FROM HAEMOPHILUS INFLUENZAE
Authors	Bruns, C.M., Nowalk, A.J., Arvai, A.S., Mctigue, M.A., Vaughan, K.G., Mietzner, T.A., Mcree, D.E.
Primary Citation	**Bruns, C.M., Nowalk, A.J., Arvai, A.S., McTigue, M.A., Vaughan, K.G., Mietzner, T.A., McRee, D.E.** Structure of Haemophilus influenzae Fe(+3)-binding protein reveals convergent evolution within a superfamily. *Nat.Struct.Biol. v4 pp.919-924 , 1997* [Abstract]
History	Deposition 1997-05-14 Release 1998-01-28
Experimental Method	Type X-RAY DIFFRACTION Data [EDS]

Parameters	Resolution[Å]	R-Value	R-Free	Space Group
	1.60	0.179 (obs.)	0.240	P 2_1 2_1 2

Unit Cell	Length [Å]	a	132.19	b	52.45	c	41.26
	Angles [°]	alpha	90.00	beta	90.00	gamma	90.00

Molecular Description Asymmetric Unit Polymer: 1 Molecule: FERRIC IRON BINDING PROTEIN Chains: _

Classification **Iron Transport**

Source Polymer: 1 Scientific Name: **Haemophilus influenzae** Expression system: **Escherichia coli**

Chemical Component

Identifier	Name	Formula	Drug Similarity	Hapten Similarity	Ligand Structure	Ligand Interaction
PO4	PHOSPHATE ION	O_4P^{3-}			[View]	[View]

Images and Visualization

<< Asymmetric Unit >>

Display Options

KiNG
Jmol
WebMol
MBT SimpleViewer
MBT Protein Workshop
QuickPDB
All Images

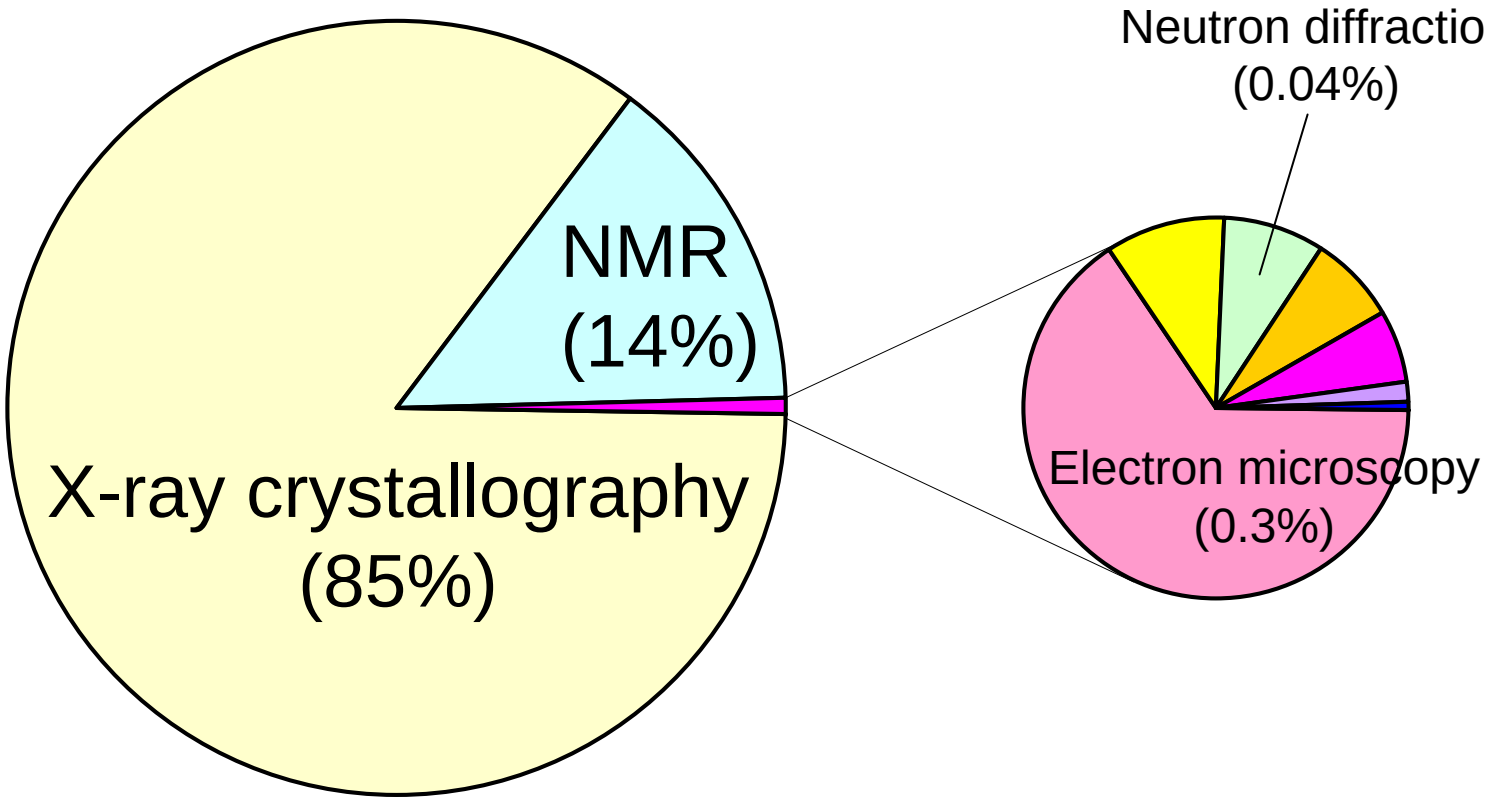
http://www.rcsb.org/pdb/rss/LastLoad

Where do PDB files come from?

Experimental technique	Number	Percent
X-ray diffraction (crystal)	36421	84.995%
Nuclear Magnetic Resonance (NMR)	6208	14.487%
Electron microscopy	145	0.338%
Fiber diffraction (X-ray)	22	0.051%
Neutron diffraction	19	0.044%
Powder diffraction (X-ray)	17	0.040%
Electron diffraction	14	0.033%
Electron tomography	4	0.009%
Fluorescence transfer	1	0.002%
Total	42851	100.000%

(at 90% sequence identity April 23, 2007)

Where do PDB files come from?

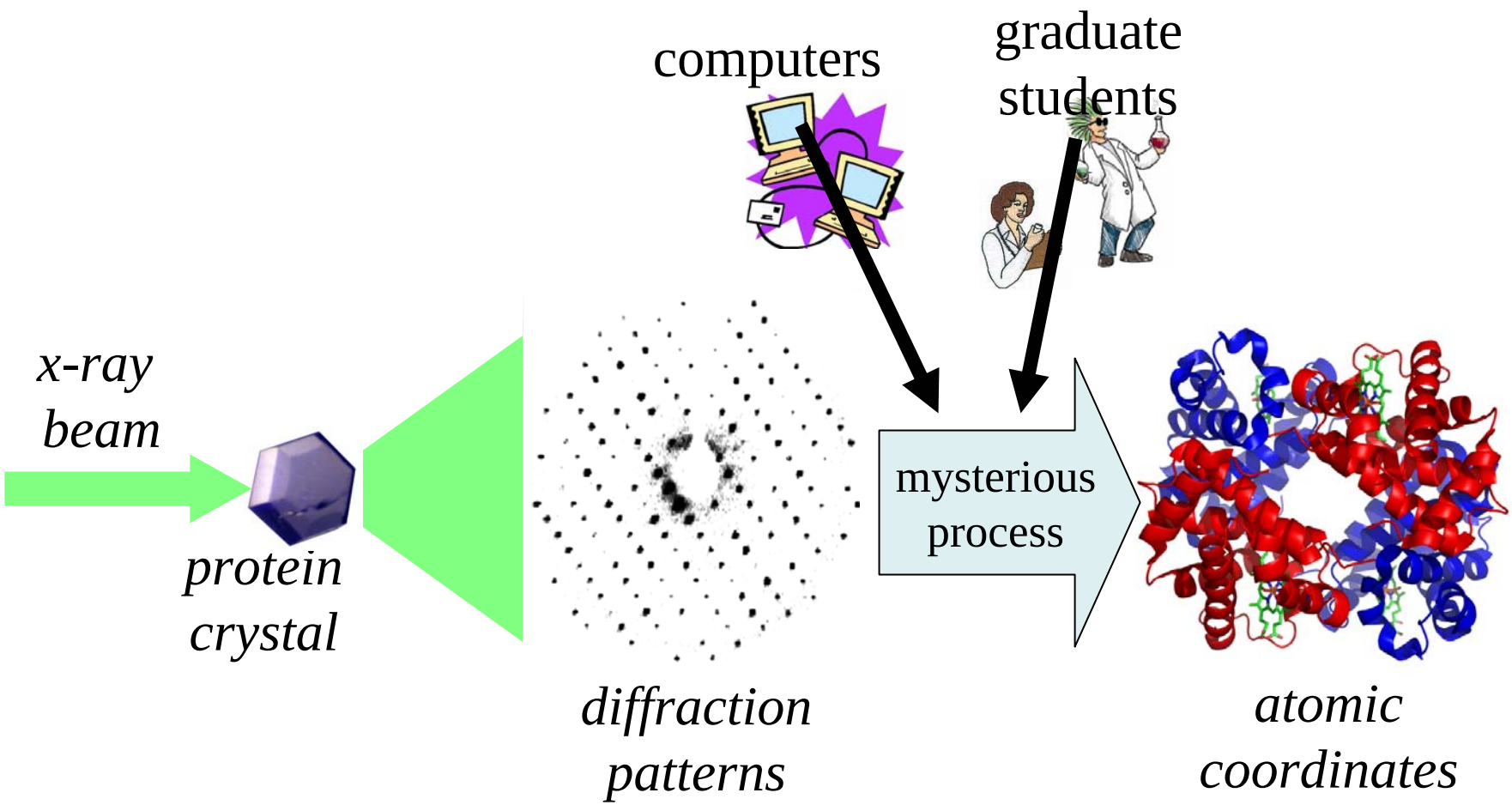


(at 90% sequence identity April 23, 2007)

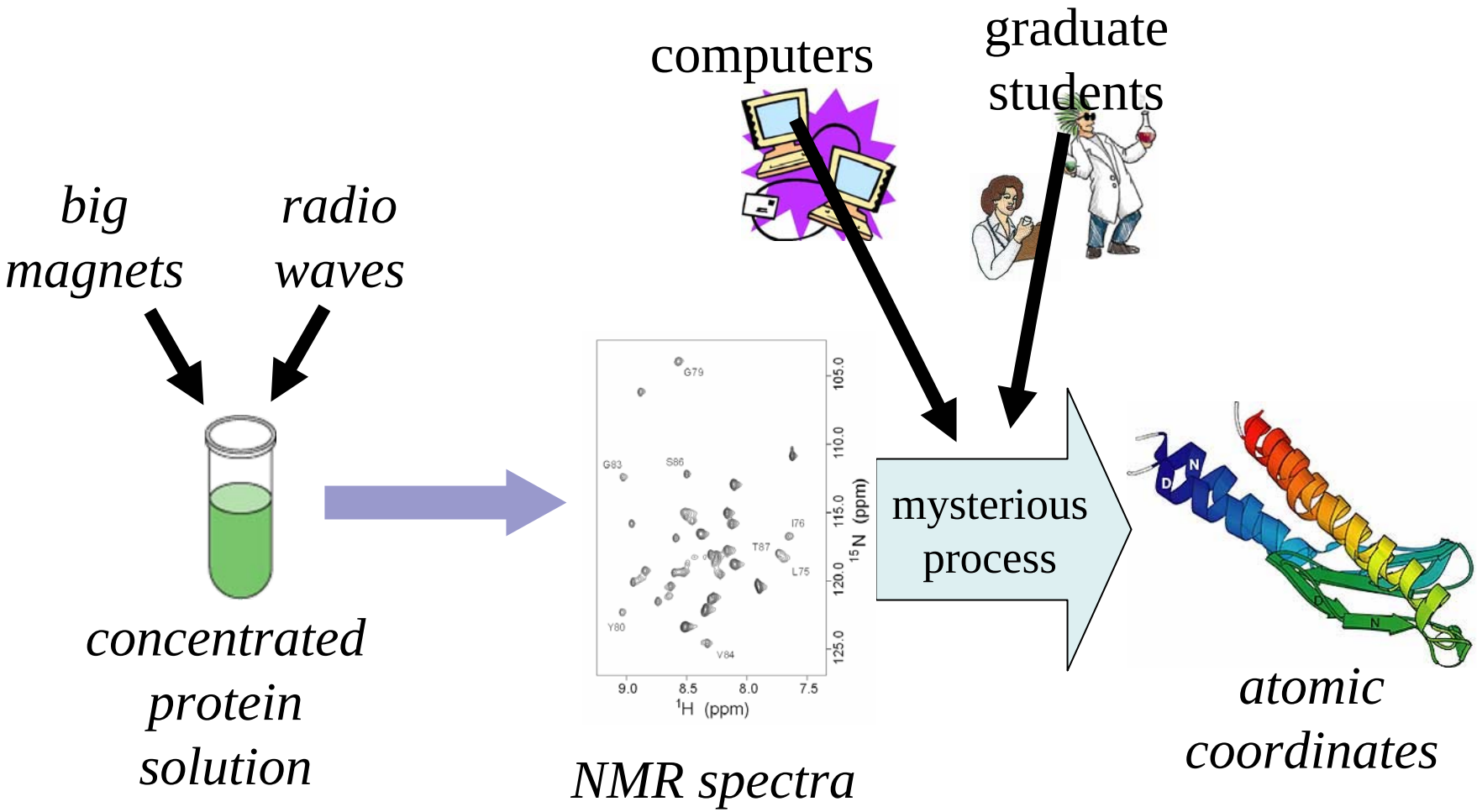
Crystallography vs. NMR

- X-ray crystallography
  - Usually lack hydrogen atom positions
  - Structural time scale is hours
- NMR – Nuclear Magnetic Resonance
  - Molecular dynamics simulations are part of the structure determination process
  - Produces multiple related versions of a molecular structure
  - nanosecond to microsecond structure time scale

X-ray crystallography



NMR spectroscopy



Haemophilus influenzae

apo (iron free)

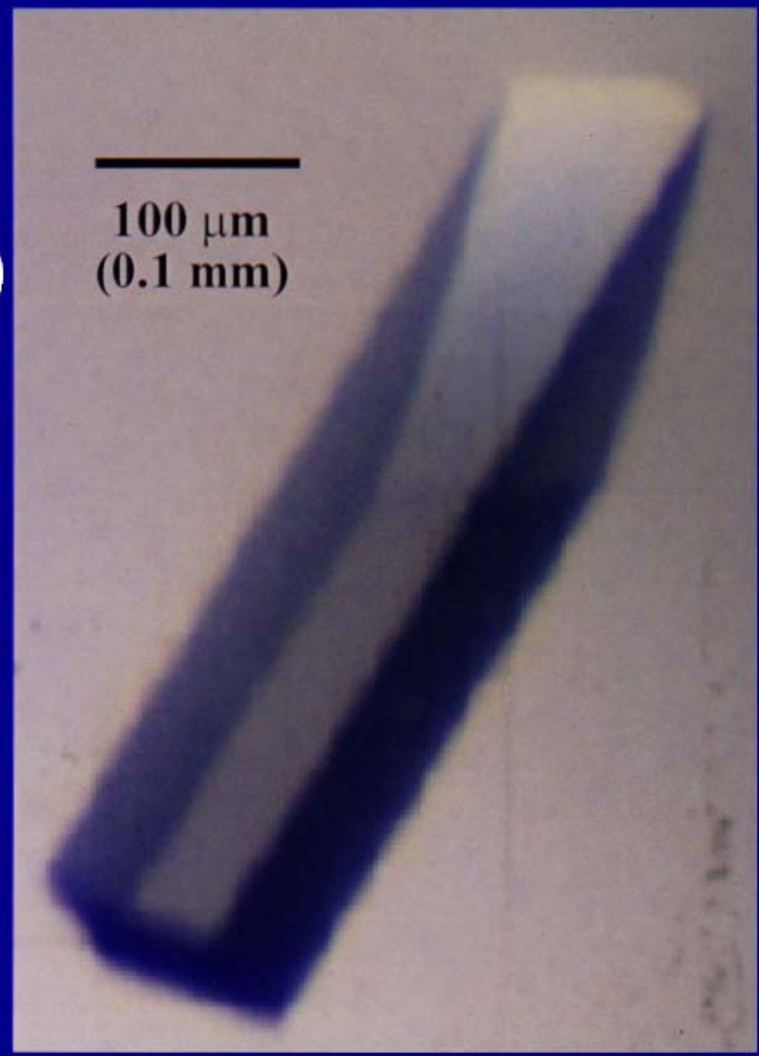
Fe^{+3} binding protein (apo-hFBP)

orthorhombic space group $P2_12_12$

a = 106.7 Å b = 77.4 Å c = 34.2 Å

$\alpha = \beta = \gamma = 90^\circ$

resolution = 1.75 Å

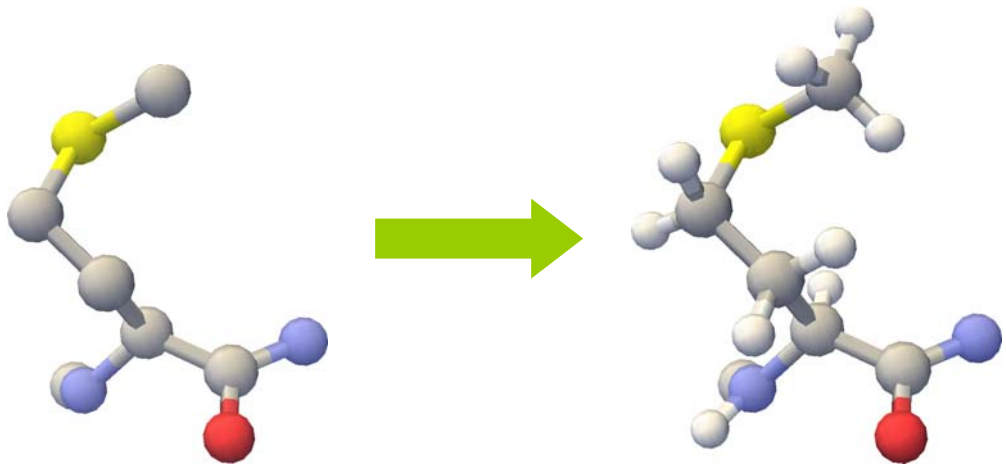


Resolution vs. accuracy

- “Resolution” in an x-ray crystal structure is a technical term relating to the diffraction data. It has units of length. Resolution is typically in the range of 1 (better) to 3 (worse) Angstroms.
- The accuracy of a crystal structure is typically 1/10 of the resolution for well determined parts of the structure. Typically 0.2 Angstroms.
- Occasionally you may hear ill-informed computational chemists state that a “three Angstrom RMSD” is “close to the accuracy of a crystal structure”. Such people are confused.

Missing hydrogen atoms

- X-ray crystallography produces essentially no electron density information for hydrogen atoms
  - Most x-ray structures in the PDB lack hydrogen atom positions
- Neutron diffraction does locate hydrogen atoms
- There exist computer programs that automatically add hydrogen atom positions
  - H++ http://biophysics.cs.vt.edu/H++/



Missing segments

- Mobile loops can be invisible in electron density maps.
- Thus PDB models can lack atomic coordinates for parts of a molecule

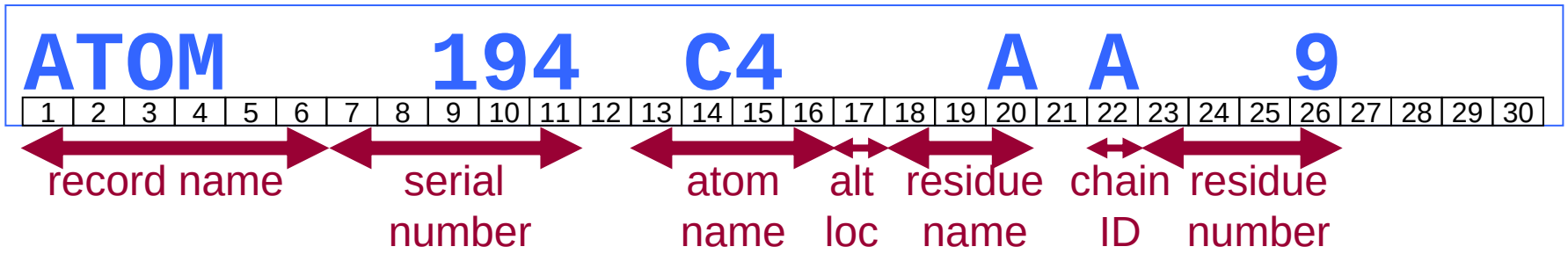
PDB File format

- Text format – humans can read it too
- Most of the information is atomic coordinates
- PDB files usually contain multiple molecules, including water molecules

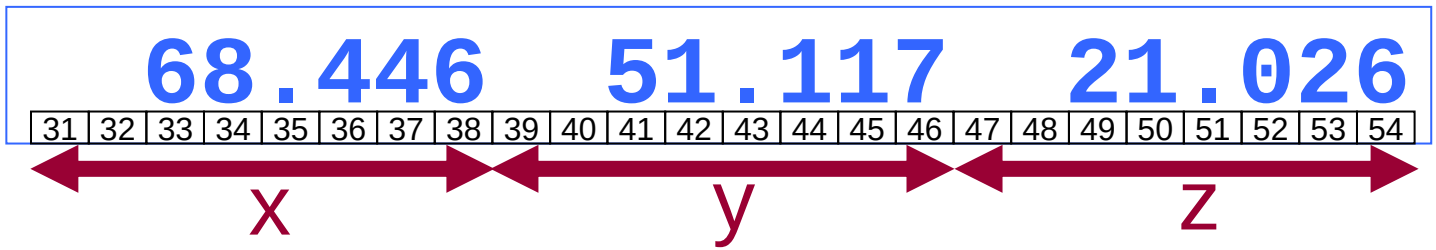
```
…
ATOM   1751  N   GLY C 250      32.286   1.882  43.206  1.00 22.00
ATOM   1752  CA  GLY C 250      32.365   1.086  41.969  1.00 21.39
ATOM   1753  C   GLY C 250      31.538   1.735  40.864  1.00 20.79
ATOM   1754  O   GLY C 250      30.621   2.527  41.152  1.00 21.58
…
```

Anatomy of an ATOM record

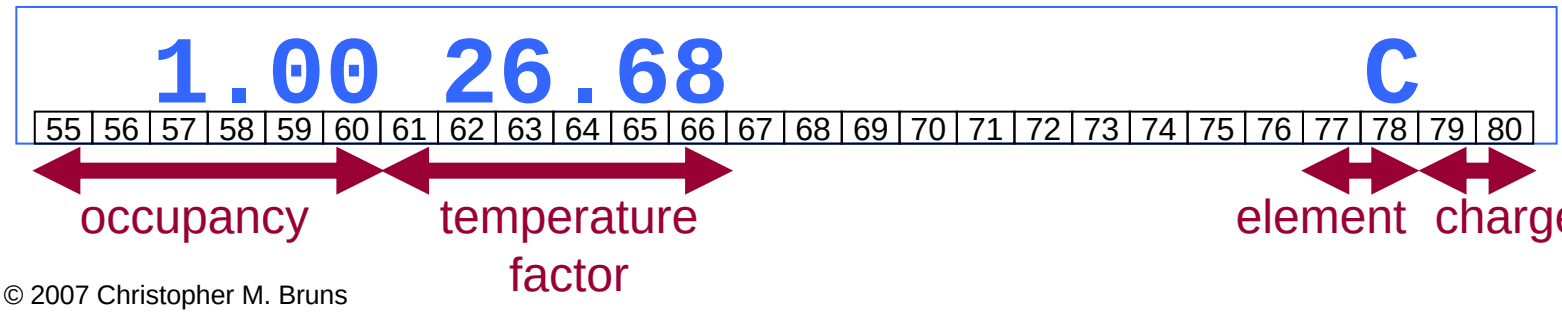
columns 1 to 30



columns 31 to 54: atomic coordinates



columns 55 to 80



PDB ATOM record format

COLUMNS	DATA TYPE	FIELD	DEFINITION
1 - 6	Record name	"ATOM "	
7 - 11	Integer	serial	Atom serial number.
13 - 16	Atom	name	Atom name.
17	Character	altLoc	Alternate location indicator.
18 - 20	Residue name	resName	Residue name.
22	Character	chainID	Chain identifier.
23 - 26	Integer	resSeq	Residue sequence number.
27	AChar	iCode	Code for insertion of residues.
31 - 38	Real(8.3)	x	Orthogonal coordinates for X in Angstroms
39 - 46	Real(8.3)	y	Orthogonal coordinates for Y in Angstroms
47 - 54	Real(8.3)	z	Orthogonal coordinates for Z in Angstroms
55 - 60	Real(6.2)	occupancy	Occupancy.
61 - 66	Real(6.2)	tempFactor	Temperature factor.
77 - 78	LString(2)	element	Element symbol, right-justified.
79 - 80	LString(2)	charge	Charge on the atom.

http://www.wwpdb.org/documentation/format23/v2.3.html

ATOM and HETATM records

- ATOM records – atoms in “canonical” amino acid and nucleotide residues
- HETATM records – all other atoms
  - includes water molecules and small ions
  - includes unusual or modified nucleotides and amino acids, so it is not safe to ignore HETATM records when you want an entire macromolecule
- ATOM and HETATM lines have the same format, except for the record name

```
ATOM    194  C4    A A   9      68.446  51.117  21.026  1.00 26.68           C
HETATM  195  P   2MG A  10      61.504  51.328  21.232  1.00 44.21           P
```

PDB atomic coordinates

- x, y, z coordinates in units of Ångstroms (10^{-10} meters)
- Right-handed orthogonal coordinate system with origin established from details of crystal geometry

```
                                        X       Y       Z
ATOM    194  C4    A A   9      68.446  51.117  21.026  1.00 26.68           C
HETATM  195  P   2MG A  10      61.504  51.328  21.232  1.00 44.21           P
```

Coarse-grained representations in biology

Natural lump sizes:

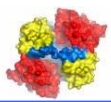

atoms


residues


base pairs

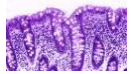

duplexes, alpha helices, beta strands

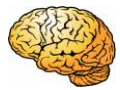

macromolecules


molecular assemblies


organelles


cells


tissues


organs


limbs


organisms


societies


ecosystems


biospheres


intergalactic federations