

# Coordinates to Structure

Atomistic Modeling of Small-Angle Scattering  
Data Using SASSIE-web

September 21-23, 2016

Advanced Photon Source

Argonne National Laboratory, Argonne, IL

# MD

A technique for computing equilibrium and transport properties of a classical many-body system.

## Experiment:

sample in instrument, measure over time ... measure longer until data converge.

## MD:

model system with  $N$  particles, solve  $\mathbf{F} = m\mathbf{a}$  until properties do not change (Equilibrate) then you “measure” (i.e. average a property) until data converge.

# MD

They can both suffer from the similar issues

experiment	md
sample not prepared correctly	incorrect starting model structure
measurement too short	simulation too short
system undergoes irreversible change (aggregate etc.)	structure gets stuck in non-ergodic hole
didn't quite measure what we thought	ahem, a bug in your analysis code

# Pseudo-code

```
read_input_parameters(temp, dt, number_of_timesteps, pressure,  
approximation parameters)
```

```
initialize(coordinates, velocities)
```

Loop over timesteps:

```
    calculate_forces(r)
```

```
    integrate_equation_of_motion(r,p) --> propagates r & p
```

```
    t = t + dt
```

# $F = - \text{del } U \rightarrow$ Verlet Integrator

$$U(r) = 4\epsilon \left[ \left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^6 \right]$$

$$f_x(r) = -\frac{\partial u(r)}{\partial x} = -\left(\frac{x}{r}\right) \left(\frac{\partial u(r)}{\partial r}\right)$$

$$f_x(r) = \frac{48x}{r^2} \left[ \left(\frac{1}{r}\right)^{12} - \left(\frac{0.5}{r}\right)^6 \right]$$

$F = - \text{del } U \rightarrow \text{Verlet Integrator}$

**Many numerical integrators have been evaluated**

**Verlet**

**Leapfrog Verlet**

**Velocity Verlet**

**4d Verlet**

**Velocity Verlet 2**

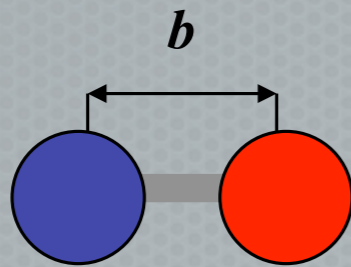
**...**

**Key  $\rightarrow$  energy conservation ( $\sim 10^{-6}$ )**

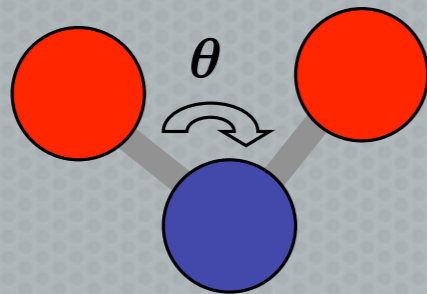
# Bonded Systems

## Potential

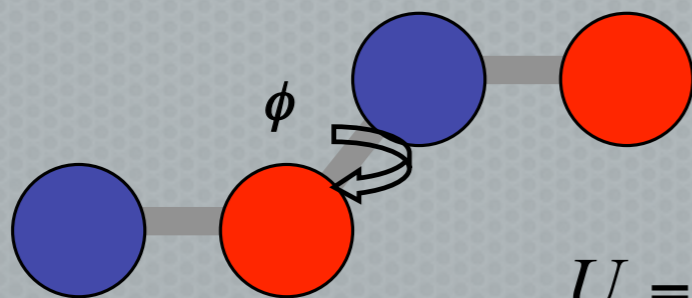
### *Strong bonded interactions*



$$U = K(b - b_0)^2$$

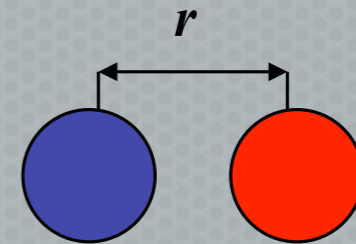


$$U = K(\theta - \theta_0)^2$$



$$U = K(1 - \cos(n\phi))$$

### *Non bonded interactions*



$$U = \left(\frac{A}{r}\right)^{12} - \left(\frac{B}{r}\right)^6$$



$$U = \frac{q_1 q_2}{r}$$

# Bonded Systems

$$\begin{aligned} V = & \sum_{\text{bonds}} k_b (b - b_0)^2 + \sum_{\text{angles}} k_\theta (\theta - \theta_0)^2 + \sum_{\text{dihedrals}} k_\phi [1 + \cos(n\phi - \delta)] \\ & + \sum_{\text{impropers}} k_\omega (\omega - \omega_0)^2 + \sum_{\text{Urey-Bradley}} k_u (u - u_0)^2 \\ & + \sum_{\text{nonbonded}} \epsilon \left[ \left( \frac{R_{\min_{ij}}}{r_{ij}} \right)^{12} - \left( \frac{R_{\min_{ij}}}{r_{ij}} \right)^6 \right] + \frac{q_i q_j}{\epsilon r_{ij}} \end{aligned}$$

# Approximations

MD calculations take too long...

Many schemes are used to save cpu cycles. The tricks have no physical significance . . .

They should not affect the results of the calculation in any way.

# Approximations

Force (summed over pairs of atoms)

$$N(N-1)/2 \rightarrow O(N^2)$$

tricks to get  $O(N)$ :

neighbor lists (verlet, cell, etc.)

PME :  $N \ln(N)$

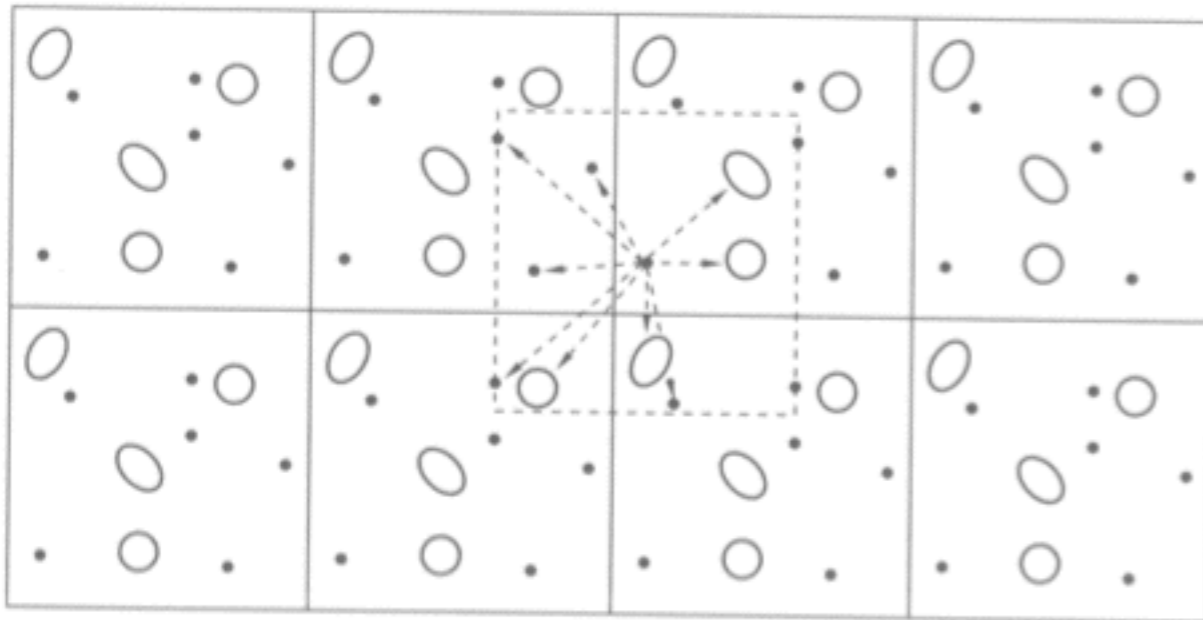
Parallelization Methods

Multiple time steps

built-in and mature

# PBC

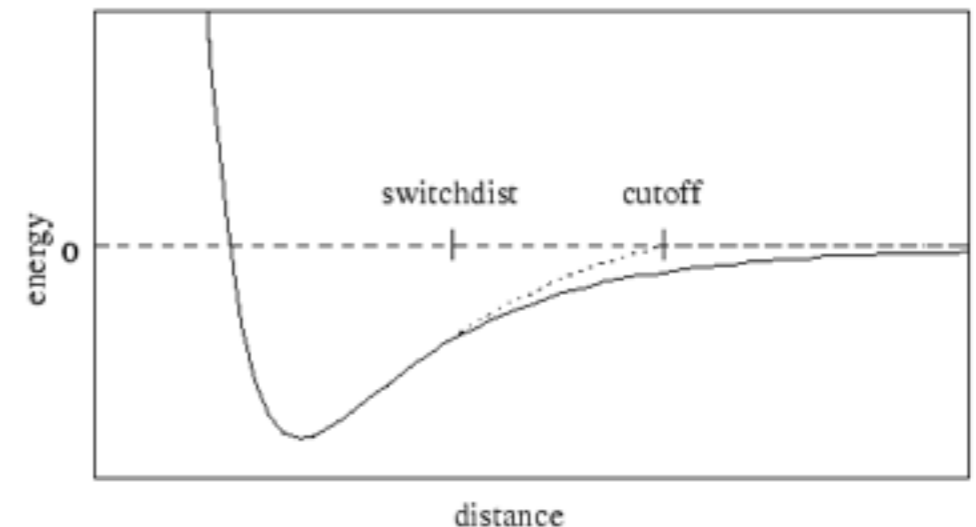
3D : Fraction @ surface  $\rightarrow N^{-1/3}$   
1000 atoms  $\rightarrow$  49% at surface  
 $10^6$  atoms  $\rightarrow$  6% at surface



But or  $O(N^2)$  pair  
distance calculation just  
went to infinity

Need cutoffs [ $r_{\text{cut}} < L/2$ ]

MD (truncated &  
smoothed)



**Figure 1:** Graph of van der Waals potential with and without the application of the switching function. With the switching function active, the potential is smoothly reduced to 0 at the cutoff distance. Without the switching function, there is a discontinuity where the potential is truncated.

# Summary

Provide the MD “program”

Coordinates

Topology (connectivity)

Force Field Parameters

Temperature

Pressure

Parameters for approximations

MTS, NNL, PME, PBC, cutoffs, restraints

Output information (filenames)

**Equilibrate** (convergence) ... then “measure”

# Barriers: BUILD; EQUILIBRATE; PROPAGATE; ANALYZE

What software package(s) and force-fields do I use?

**Starting structure?**

**How do I clean up the structure?**

**How do I set up a trajectory** (time or space)?

How do I calculate scattering observables correctly?

# Overview

## **PDB ---> PSF: from coordinates to bonding & structure**

Anatomy of PDB file

Binary files

FF topology file

Psfgen (PDB + topology) --> PSF

FF parameter file

## **PDB + PSF + INPUT FILE : run a NAMD minimization**

Minimization (in vacuo)

Adding water & ions

PBC & dynamics parameters

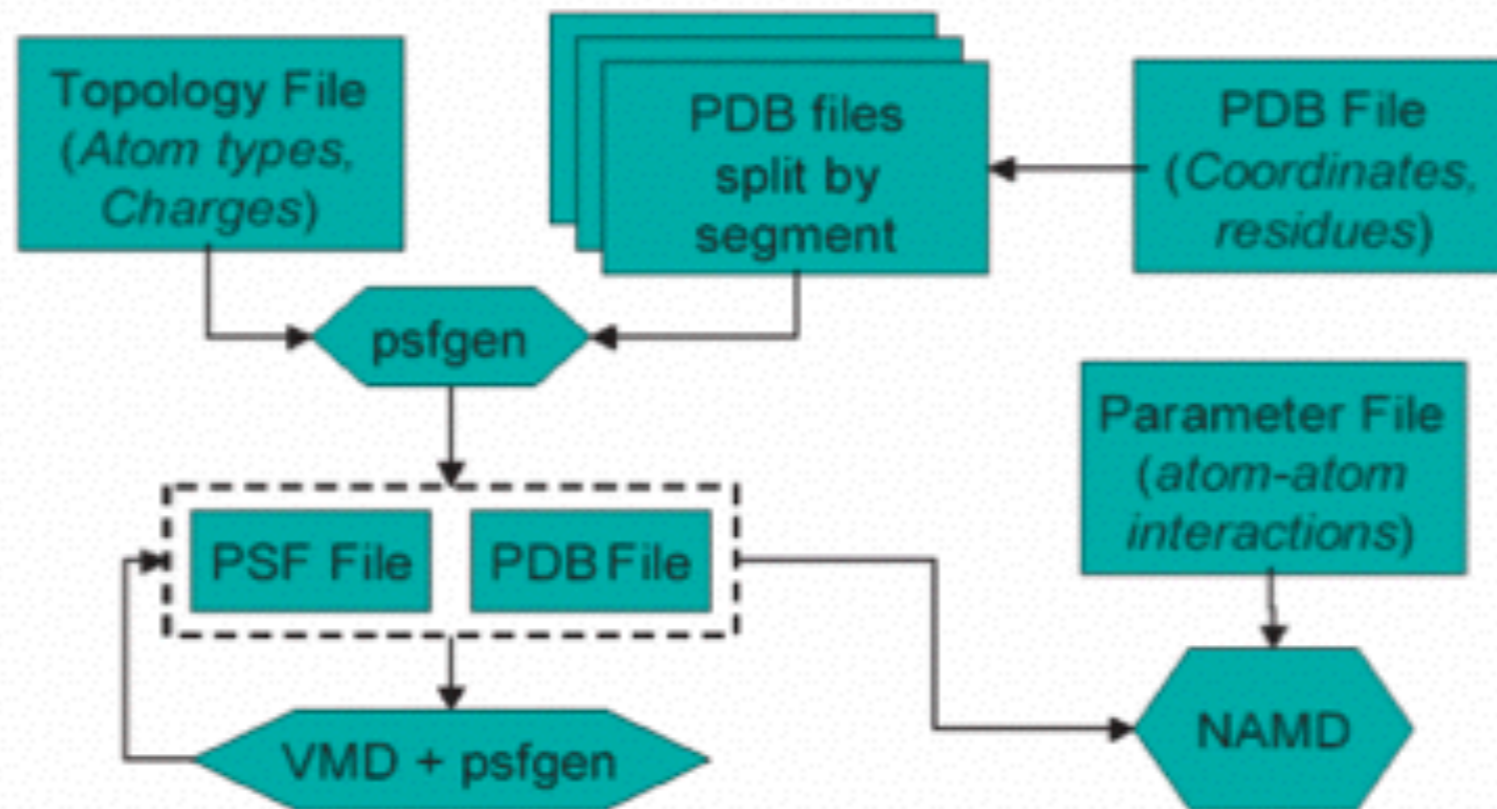
Restarting a simulation (continuation)

Fixed atom simulations & IMD

Output files

Lab I & II preview

# The Process



# Coord. Files

## Protein Data Bank (PDB)

<http://www.pdb.org>

[ftp://ftp.wwpdb.org/pub/pdb/doc/format\\_descriptions/Format\\_v32\\_letter.pdf](ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v32_letter.pdf)

## Open Babel

[http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page)

> 110 chemical file formats

- Common cheminformatics formats
  - Canonical SMILES format (can)
  - Chemical Markup Language (cml, mrv)
  - InChI format (inchi)
  - MDL MOL format (mol, mdl, sdf, sd)
  - **Protein Data Bank format (pdb, ent)**
  - SMILES format (smi, smiles)
  - Sybyl Mol2 format (ml2, sy2, mol2)
- Utility formats
  - Compare molecules using InChI (k)
  - Copy raw text (copy)
  - Fastsearch format (fs)
  - Fingerprint format (fpt)
  - General XML format (xml)
  - Generic Output file format (dat, output, out, log)
  - MolPrint2D format (mpd)
  - Multilevel Neighborhoods of Atoms (MNA) (mna)
  - Open Babel molecule report (molreport)
  - Open Babel report format (report)
  - Outputs nothing (nul)
  - Reads raw text (text)
  - Title format (txt)
  - **XYZ cartesian coordinates format (xyz)**
- Other cheminformatics formats
  - Accelrys/MSI Biosym/Insight II CAR format (arc, car)
  - Accelrys/MSI Cerius II MSI format (msi)
  - Accelrys/MSI Quanta CSR format (csr)
  - MCDL format (mcdl)
  - MSI BGF format (bgf)
  - PubChem format (pc)
- Computational chemistry formats
  - ADF cartesian input format (adf)
  - ADF output format (adfout)
  - CAChe MolStruct format (cache, cac)

# Coord. Files

## Protein Data Bank (PDB)

<http://www.pdb.org>

[ftp://ftp.wwpdb.org/pub/pdb/doc/format\\_descriptions/Format\\_v32\\_letter.pdf](ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v32_letter.pdf)

## Open Babel

[http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page)

> 110 chemical file formats

- [Chemical Resource Kit 3D format \(crk3d\)](#)
- [Chemical format \(gpr\)](#)
- [Molden format \(molden, mold, molf\)](#)
- [PCModel Format \(pcm\)](#)
- [UniChem XYZ format \(unxyz\)](#)
- [ViewMol format \(vmol\)](#)
- [YASARA.org YOB format \(yob\)](#)
- [Kinetics and Thermodynamics formats](#)
  - [ChemKin format \(ck\)](#)
  - [Thermo format \(tdd, therm\)](#)
- [Molecular dynamics and docking formats](#)
  - [Amber Prep format \(prep\)](#)
  - [Autodock Protein Data Bank \(pdbqt\)](#)
  - [DL-POLY CONFIG \(CONFIG\)](#)
  - [DL-POLY HISTORY \(HISTORY\)](#)
  - [Dock 3.5 Box format \(box\)](#)
  - [GROMOS96 format \(gr96\)](#)
  - [MacroModel format \(mmod, mmd\)](#)
  - [Tinker MM2 format \(txyz\)](#)
  - [XTC format \(xtc\)](#)
- [Volume data formats](#)
  - [ADF TAPE41 format \(t41\)](#)
  - [Gaussian cube format \(cube, cub\)](#)
  - [OpenDX cube format for APBS \(dx\)](#)
- [Miscellaneous formats](#)
  - [M.F. Sanner's MSMS input format \(msms\)](#)
- [Biological data formats](#)
  - [FASTA format \(fasta, fa, fsa\)](#)
  - [PQR format \(pqr\)](#)
- [Obscure formats](#)
  - [Alchemy format \(alc\)](#)
  - [CCC format \(ccc\)](#)
  - [Feature format \(feat\)](#)
  - [SMILES FIX format \(fix\)](#)
  - [XED format \(xed\)](#)

# Anatomy of a PDB file I

HEADER  
OBSLTE  
TITLE  
CAVEAT  
COMPND  
SOURCE  
KEYWDS  
EXPDTA  
NUMMDL  
MDLTYP  
AUTHOR  
REVDAT  
SPRSDE  
JRNL  
REMARK (1-999)  
REMARK 290 Cryst. Sym.  
REMARK 300 Biomol.  
**REMARK 465 Missing AA**  
**REMARK 470 Missing atom(s)**  
DBREF  
SEQADV  
**SEQRES**  
MODRES  
HET  
HETNAM

HETSYN  
FORMUL  
HELIX  
SHEET  
**SSBOND**  
LINK  
CISPEP  
SITE  
CRYSTL1  
OIGXn  
SCALEn  
MTRIXn  
MODEL (starts a model)  
**ATOM**  
ANISOU  
TER (ends a chain)  
HETATM  
ENDMDL (ends a model)  
CONNECT  
MASTER  
END

## FOCUS:

**name of compound**  
**species & tissue**  
**journal citation**  
**sequence**  
**symmetry**  
**missing atoms**  
**modified atoms/bonds**  
**heteroatoms**  
**coords**  
**multiple models**

PDB files are not necessarily written with  
your goals in mind!!!

# Anatomy of a PDB file II

```

      1           2           3           4           5           6           7
1234567890123456789012345678901234567890123456789012345678901234567890
ATOM      1  N      GLN  A      1           9.921    8.724    2.612    1.00    0.00           PCLN  N
ATOM      2  HT1   GLN  A      1           9.866    9.721    2.563    0.00    0.00           PCLN  H
ATOM      3  HT2   GLN  A      1           9.309    8.315    1.935    0.00    0.00           PCLN  H
    
```

Column(s)	Data Type	Definition	Example
1 - 6	Record Name		"ATOM "
7 - 11	Integer	Atom serial number	" 1"
13 - 16	Atom	Atom name	" N "
17	Character	"altloc"	" "
18 - 20	resName	Residue name	"GLN"
22	chainID	Chain indentifier	"A"
23 - 26	Integer	Residue sequence #	" 1"
27	Character	"iCode"	" "
31 - 38	Real (8.3)	X coord. (ang.)	" 9.921"
39 - 46	Real (8.3)	Y coord. (ang.)	" 8.724"
47 - 54	Real (8.3)	Z coord. (ang.)	" 2.612"
55 - 60	Real (6.2)	Occupancy	" 1.00"
61 - 66	Real (6.2)	Temperature factor	" 0.00"
77 - 78	String	Element symbol (RJ)	" N"
79 - 80	String	Charge	" "
(73 - 76)	String	Segment name	"PCLN"

# Anatomy of a PDB file II

```

      1           2           3           4           5           6           7
1234567890123456789012345678901234567890123456789012345678901234567890
ATOM      1  N      GLN  A      1           9.921    8.724    2.612    1.00    0.00           PCLN  N
ATOM      2  HT1   GLN  A      1           9.866    9.721    2.563    0.00    0.00           PCLN  H
ATOM      3  HT2   GLN  A      1           9.309    8.315    1.935    0.00    0.00           PCLN  H
  
```

Column(s)	Data Type	Definition	Example
1 - 6	Record Name		"ATOM "
7 - 11	Integer	Atom serial number	" 1"
13 - 16	Atom	Atom name	" N "
17	Character	"altloc"	" "
18 - 20	resName	Residue name	"GLN"
22	chainID	Chain indentifier	"A" --> " "
23 - 26	Integer	Residue sequence #	" 1"
27	Character	"iCode"	" "
31 - 38	Real (8.3)	X coord. (ang.)	" 9.921"
39 - 46	Real (8.3)	Y coord. (ang.)	" 8.724"
47 - 54	Real (8.3)	Z coord. (ang.)	" 2.612"
55 - 60	Real (6.2)	Occupancy	" 1.00"
61 - 66	Real (6.2)	Temperature factor	" 0.00"
77 - 78	String	Element symbol (RJ)	" N"
79 - 80	String	Charge	" "
(73 - 76)	String	Segment name	"PCLN"

[ftp://ftp.wwpdb.org/pub/pdb/doc/format\\_descriptions/Format\\_v32\\_letter.pdf](ftp://ftp.wwpdb.org/pub/pdb/doc/format_descriptions/Format_v32_letter.pdf)

pdbx : 2017

# PDB Notes

Proteins: N to C terminus

Nucleic acids: 5' - 3'

No ordering for polysaccharides

Chains terminated by "TER"

MODEL --> ENDMDL

AltLoc --> means there is more than one structure reported (duplicate)

Model can have no more than 99,999 atoms, otherwise split into separate PDB ENTRIES  
--> in practice we can have any number of atoms in our files (99999) & (9999)->SEG

Terminal oxygen atoms: OXT (proteins) & O5' or OP3 (nucleic acids)

**MISSING ATOMS/RESIDUES (hydrogens!)**

**DUPLICATE RESIDUES / ATOMS (AltLoc)**

**WRONG PROTEIN (e.g. you only want part of a complex)**

**POST-TRANSLATION MODS (HEME, phosphoserine, etc.)**

**SEQUENCE (HIS tags, cloning artifacts, etc.)**

**RS-SR'**

**N-TERMINAL MODIFICATIONS**

**X-tal / NMR WATERS / IONS**

# Binary Coordinate Files

## DCD (CHARMM/NAMD/XPLOR)

TRJ or MD CRD (Amber)

NETCDF (Amber, MMTK)

TRJ (Gromacs)

etc.

“naming” information is not in a binary file, just general system info. and x,y,z coordinates: need PDB or PSF + DCD

**Compared to concatenated PDB files (separated by END statements) DCD files are ~ 7 - 10 X smaller**

**Allows one to optimize file I/O code to speed up saving & reading data**

**Non-trivial to bootstrap your own code (endian byte ordering, 32/64 bit, . . .)**

# Simple Example

## PDB of a atom (oh2.pdb)

```
ATOM      1  OH2  TIP3  A      1      0.000  0.000  0.000  1.00  0.00  WAT1  O
```

## XYZ of a atom (oh2.xyz)

1

```
OH2      2.000000  2.000000  2.000000
```

## PDB of a atom moving (oh2\_move.pdb)

```
ATOM      1  OH2  TIP3  A      1      0.000  0.000  0.000  1.00  0.00  WAT1  O
END
ATOM      1  OH2  TIP3  A      1      1.000  0.000  0.000  1.00  0.00  WAT1  O
END
ATOM      1  OH2  TIP3  A      1      2.000  0.000  0.000  1.00  0.00  WAT1  O
END
ATOM      1  OH2  TIP3  A      1      3.000  0.000  0.000  1.00  0.00  WAT1  O
END
```

# Simple Example

PDB of a atom (oh2 pdb)

ATOM 1

.00 WAT1 O

XYZ of a

1

OH2

PDB of a

ATOM 1

0.00 WAT1 O

END

ATOM 1

0.00 WAT1 O

END

ATOM 1

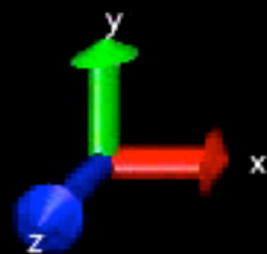
0.00 WAT1 O

END

ATOM 1

0.00 WAT1 O

END



# Onward

Okay, we have coordinates but how do we begin to deal with bonding and forces . . . MD?

PDB  
(a file)

Dictionary of building blocks  
(a file)  
“topology for all systems”

1. PSFGEN  
2. CHARMM  
(programs)

PSF  
(a file)

# Topology File

[http://mackerell.umaryland.edu/CHARMM\\_ff\\_params.html](http://mackerell.umaryland.edu/CHARMM_ff_params.html)

Specifically:

**top\_all27\_prot\_na.inp**

## KEYWORDS

- MASS** : define atom type names and masses
- RESI** : define building blocks (amino acids, etc.)
- PRES** : patches to correct / change structure  
(NTER, CTER, DISU, etc.)
- DECL** : to deal with joining blocks  
proteins: -CA, -C, -O, +N, +HN, +CA  
nucl. acids: +P, +O1P, +O2P, +O5', -O3'
- DEFA** : default patches to apply (NTER, CTER)
- AUTO ANGLES DIHE** : tells "program" to enumerate angles  
and dihedrals when blocks are put  
together

Text file (read top to bottom)

Keyword driven

Comments --> ! or \*

**Topology file is still only  
a piece of the puzzle . . .**

**Need a "parameter" file  
at some point.**

# Anatomy of a Topology File I

```
MASS      1  H      1.00800  H ! polar H
MASS      2  HC     1.00800  H ! N-ter H
MASS      3  HA     1.00800  H ! nonpolar H
MASS      4  HT     1.00800  H ! TIPS3P WATER HYDROGEN
MASS      5  HP     1.00800  H ! aromatic H
MASS      6  HB     1.00800  H ! backbone H
MASS      7  HR1    1.00800  H ! his he1, (+) his HG,HD2
MASS      8  HR2    1.00800  H ! (+) his HE1
MASS      9  HR3    1.00800  H ! neutral his HG, HD2
MASS     10  HS     1.00800  H ! thiol hydrogen
MASS     11  HE1    1.00800  H ! for alkene; RHC=CR
MASS     12  HE2    1.00800  H ! for alkene; H2C=CR
MASS     20  C      12.01100  C ! carbonyl C, peptide backbone
MASS     21  CA     12.01100  C ! aromatic C
MASS     22  CT1    12.01100  C ! aliphatic sp3 C for CH
MASS     23  CT2    12.01100  C ! aliphatic sp3 C for CH2
MASS     24  CT3    12.01100  C ! aliphatic sp3 C for CH3
MASS     25  CPH1   12.01100  C ! his CG and CD2 carbons
MASS     26  CPH2   12.01100  C ! his CE1 carbon
```

. . . etc. . . .

**“Atom Types” != “PDB Atom Name”**

# Anatomy of a Topology File II

. . .

```
RESI CLA      -1.00 ! Chloride Ion
GROUP
ATOM CLA CLA -1.00
PATCHING FIRST NONE LAST NONE
```

**CLA** --> expected name in PDB file

**CLA** --> "atom type" in topology file (MASS)

```
RESI SOD      1.00 ! Sodium Ion
GROUP
ATOM SOD SOD 1.00
PATCHING FIRST NONE LAST NONE
```

```
RESI CAL      2.00 ! Calcium Ion
GROUP
ATOM CAL CAL 2.00
PATCHING FIRST NONE LAST NONE
```

MINIMAL FORMAT FOR A "RESIDUE":

```
RESI "NAME"          TOTAL CHARGE
GROUP
ATOM "PDBNAME" "ATOM TYPE NAME" CHARGE
"DEFAULT PATCH STATEMENT"
```

a residue is defined between the initial RESI statement until another "keyword" is found (RESI, MASS, etc.) or EOF . . .

. . .

Generally, you don't edit these files, you USE them . . .

# Anatomy of a Topology File III

## RESI:

DEFA FIRS NTER LAST CTER  
AUTO ANGLES DIHE

```
RESI ALA          0.00
GROUP
ATOM N    NH1    -0.47  !      |
ATOM HN   H      0.31  !  HN-N
ATOM CA   CT1    0.07  !      |      HB1
ATOM HA   HB     0.09  !      |      /
GROUP      !  HA-CA--CB-HB2
ATOM CB   CT3   -0.27  !      |      \
ATOM HB1  HA     0.09  !      |      HB3
ATOM HB2  HA     0.09  !  O=C
ATOM HB3  HA     0.09  !      |
GROUP      !
ATOM C    C      0.51
ATOM O    O     -0.51
BOND CB CA N HN N CA
BOND C CA C +N CA HA CB HB1 CB HB2 CB HB3
DOUBLE O C
IMPR N -C CA HN C CA +N O
DONOR HN N
ACCEPTOR O C
IC -C CA *N HN 1.3551 126.4900 180.0000 115.4200 0.9996
IC -C N CA C 1.3551 126.4900 180.0000 114.4400 1.5390
IC N CA C +N 1.4592 114.4400 180.0000 116.8400 1.3558
IC +N CA *C O 1.3558 116.8400 180.0000 122.5200 1.2297
IC CA C +N +CA 1.5390 116.8400 180.0000 126.7700 1.4613
IC N C *CA CB 1.4592 114.4400 123.2300 111.0900 1.5461
IC N C *CA HA 1.4592 114.4400 -120.4500 106.3900 1.0840
IC C CA CB HB1 1.5390 111.0900 177.2500 109.6000 1.1109
IC HB1 CA *CB HB2 1.1109 109.6000 119.1300 111.0500 1.1119
IC HB1 CA *CB HB3 1.1109 109.6000 -119.5800 111.6100 1.1114
```

**N** --> expected name in PDB file

**NH1** --> "atom type" in topology file (MASS)

A residue may then be subdivided into GROUPs, which contain several atoms whose total charge is neutral or a unit charge. This subdivision is used in the calculation of nonbonded interactions using the group keyword and in the extended electrostatics options.

RESI: all "internal" residues

# Anatomy of a Topology File IV

## Internal Coordinates:

“optional”

one IC line for each atom in RESI

have information about N-1 & N+1 RESI

(1)	(2)	(3)	(4)	(1-2) bond length	(1-2-3) bond angle	(1-2-3-4) dihedral angle	(2-3-4) bond angle	(3-4) bond length
IC -C	CA	*N	HN	1.3551	126.4900	180.0000	115.4200	0.9996
IC -C	N	CA	C	1.3551	126.4900	180.0000	114.4400	1.5390
IC N	CA	C	+N	1.4592	114.4400	180.0000	116.8400	1.3558
IC +N	CA	*C	O	1.3558	116.8400	180.0000	122.5200	1.2297
IC CA	C	+N	+CA	1.5390	116.8400	180.0000	126.7700	1.4613
IC N	C	*CA	CB	1.4592	114.4400	123.2300	111.0900	1.5461
IC N	C	*CA	HA	1.4592	114.4400	-120.4500	106.3900	1.0840
IC C	CA	CB	HB1	1.5390	111.0900	177.2500	109.6000	1.1109
IC HB1	CA	*CB	HB2	1.1109	109.6000	119.1300	111.0500	1.1119
IC HB1	CA	*CB	HB3	1.1109	109.6000	-119.5800	111.6100	1.1114

```

ATOM N      NH1      -0.47  !      |
ATOM HN     H        0.31  !  HN-N
ATOM CA     CT1      0.07  !      |      HB1
ATOM HA     HB       0.09  !      |      /
GROUP      !  HA-CA--CB-HB2
ATOM CB     CT3     -0.27  !      |      \
ATOM HB1    HA       0.09  !      |      HB3
ATOM HB2    HA       0.09  !  O=C
ATOM HB3    HA       0.09  !      |
GROUP      !
ATOM C      C        0.51
ATOM O      O       -0.51
    
```

position of (1) determined from positions of 2-4

position of (4) determined from positions of 1-3

**Thus, given a 3 atom seed all coordinates of RESI can be generated (using the IC definitions)**

\* --> on 3rd atom delineates an improper IC (“advanced”)

# Anatomy of a Topology File V

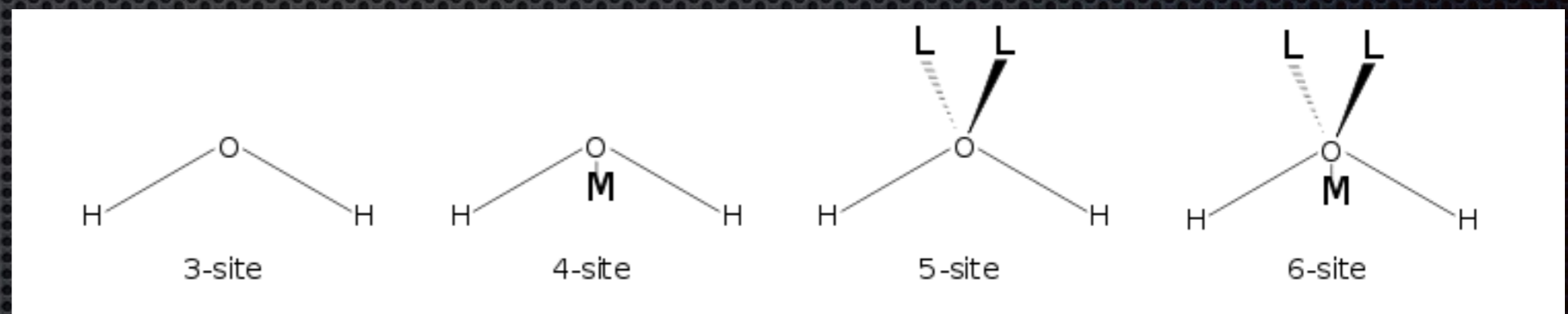
## Water

...

```
RESI TIP3                0.000 ! tip3p water model, generate using noangle nodihedral
GROUP
ATOM OH2  OT          -0.834
ATOM H1   HT           0.417
ATOM H2   HT           0.417
BOND OH2  H1 OH2  H2  H1 H2    ! the last bond is needed
ANGLE H1  OH2  H2             ! required
ACCEPTOR OH2
PATCHING FIRS NONE LAST NONE
```

...

	TIPS <sup>[4]</sup>	SPC <sup>[5]</sup>	TIP3P <sup>[6]</sup>	SPC/E <sup>[7]</sup>
$r(\text{OH}), \text{\AA}$	0.9572	1.0	0.9572	1.0
HOH, deg	104.52	109.47	104.52	109.47
$A \times 10^{-3}, \text{kcal } \text{\AA}^{12}/\text{mol}$	580.0	629.4	582.0	629.4
$B, \text{kcal } \text{\AA}^6/\text{mol}$	525.0	625.5	595.0	625.5
$q(\text{O})$	-0.80	-0.82	-0.834	-0.8476
$q(\text{H})$	+0.40	+0.41	+0.417	+0.4238



[http://en.wikipedia.org/wiki/Water\\_model](http://en.wikipedia.org/wiki/Water_model)

# Anatomy of a Topology File VI

## Patches: PRES

```
PRES NTER          1.00 ! standard N-terminus
GROUP              ! use in generate statement
ATOM N            NH3  -0.30 !
ATOM HT1          HC   0.33 !           HT1
ATOM HT2          HC   0.33 !           (+)/
ATOM HT3          HC   0.33 !  --CA--N--HT2
ATOM CA           CT1   0.21 !           |   \
ATOM HA           HB   0.10 !           HA   HT3
DELETE ATOM HN
BOND HT1 N HT2 N HT3 N
DONOR HT1 N
DONOR HT2 N
DONOR HT3 N
IC HT1  N    CA    C    0.0000  0.0000  180.0000  0.0000  0.0000
IC HT2  CA   *N   HT1  0.0000  0.0000  120.0000  0.0000  0.0000
IC HT3  CA   *N   HT2  0.0000  0.0000  120.0000  0.0000  0.0000
```

ATOM statements here:

- (1) **ADD new atoms (HT1, HT2, HT3)**
- (2) **MODIFY & CHANGE the type and charge (N, CA, HA)**

Patches are used to:

**Alter protonation state  
(ASPP, GLUP, HS2, ...)**

**Create RS-SR' bonds  
(DISU)**

**Attach HEME groups  
(PHEM)**

**Alter terminii (NTER,  
CTER, GLYP, PROP)**

...

# Anatomy of a Topology File VII

## Patches (cont.): more examples . . .

```
PRES GLYP          1.00 ! Glycine N-terminus
GROUP              ! use in generate statement
ATOM N      NH3    -0.30 !
ATOM HT1    HC     0.33 !  HA1    HT1
ATOM HT2    HC     0.33 !  | (+)/
ATOM HT3    HC     0.33 !  --CA--N--HT2
ATOM CA     CT2    0.13 !  | \
ATOM HA1    HB     0.09 !  HA2    HT3
ATOM HA2    HB     0.09 !
DELETE ATOM HN
BOND HT1 N HT2 N HT3 N
DONOR HT1 N
DONOR HT2 N
DONOR HT3 N
IC HT1  N   CA   C   0.0000  0.0000  180.0000  0.0000  0.0000
IC HT2  CA  *N   HT1 0.0000  0.0000  120.0000  0.0000  0.0000
IC HT3  CA  *N   HT2 0.0000  0.0000  120.0000  0.0000  0.0000
```

```
PRES PROP         1.00 ! Proline N-Terminal
GROUP              ! use in generate statement
ATOM N      NP     -0.07 !  HA
ATOM HN1    HC     0.24 !  |
ATOM HN2    HC     0.24 !  -CA   HN1
ATOM CD     CP3    0.16 !  / \ /
ATOM HD1    HA     0.09 !  N(+)
ATOM HD2    HA     0.09 !  / \
ATOM CA     CP1    0.16 !  -CD   HN2
ATOM HA     HB     0.09 !  | \
BOND HN1 N HN2 N   !  HD1 HD2
DONOR HN1 N
DONOR HN2 N
IC HN1  CA  *N   CD   0.0000  0.0000  120.0000  0.0000  0.0000
IC HN2  CA  *N   HN1 0.0000  0.0000  120.0000  0.0000  0.0000
```

```
PRES ACE          0.00 ! acetylated N-terminus
GROUP              ! use in generate statement
ATOM CAY    CT3   -0.27 !
ATOM HY1    HA     0.09 !  HY1  HY2  HY3
ATOM HY2    HA     0.09 !  \ | /
ATOM HY3    HA     0.09 !  CAY
GROUP              !  |
ATOM CY     C      0.51 !  CY=OY
ATOM OY     O     -0.51 !  |
!
BOND CY CAY CY N CAY HY1 CAY HY2 CAY HY3
DOUBLE OY CY
IMPR CY CAY N OY
IMPR N CY CA HN
ACCEPTOR OY CY
IC CY     N   CA   C   0.0000  0.0000 -60.0000  0.0000  0.0000
IC CY     CA  *N   HN   0.0000  0.0000  180.0000  0.0000  0.0000
IC CAY    CY  N   CA   0.0000  0.0000  180.0000  0.0000  0.0000
IC N      CAY *CY  OY   0.0000  0.0000  180.0000  0.0000  0.0000
IC OY     CY  CAY  HY1  0.0000  0.0000  180.0000  0.0000  0.0000
IC OY     CY  CAY  HY2  0.0000  0.0000   60.0000  0.0000  0.0000
IC OY     CY  CAY  HY3  0.0000  0.0000 -60.0000  0.0000  0.0000
```

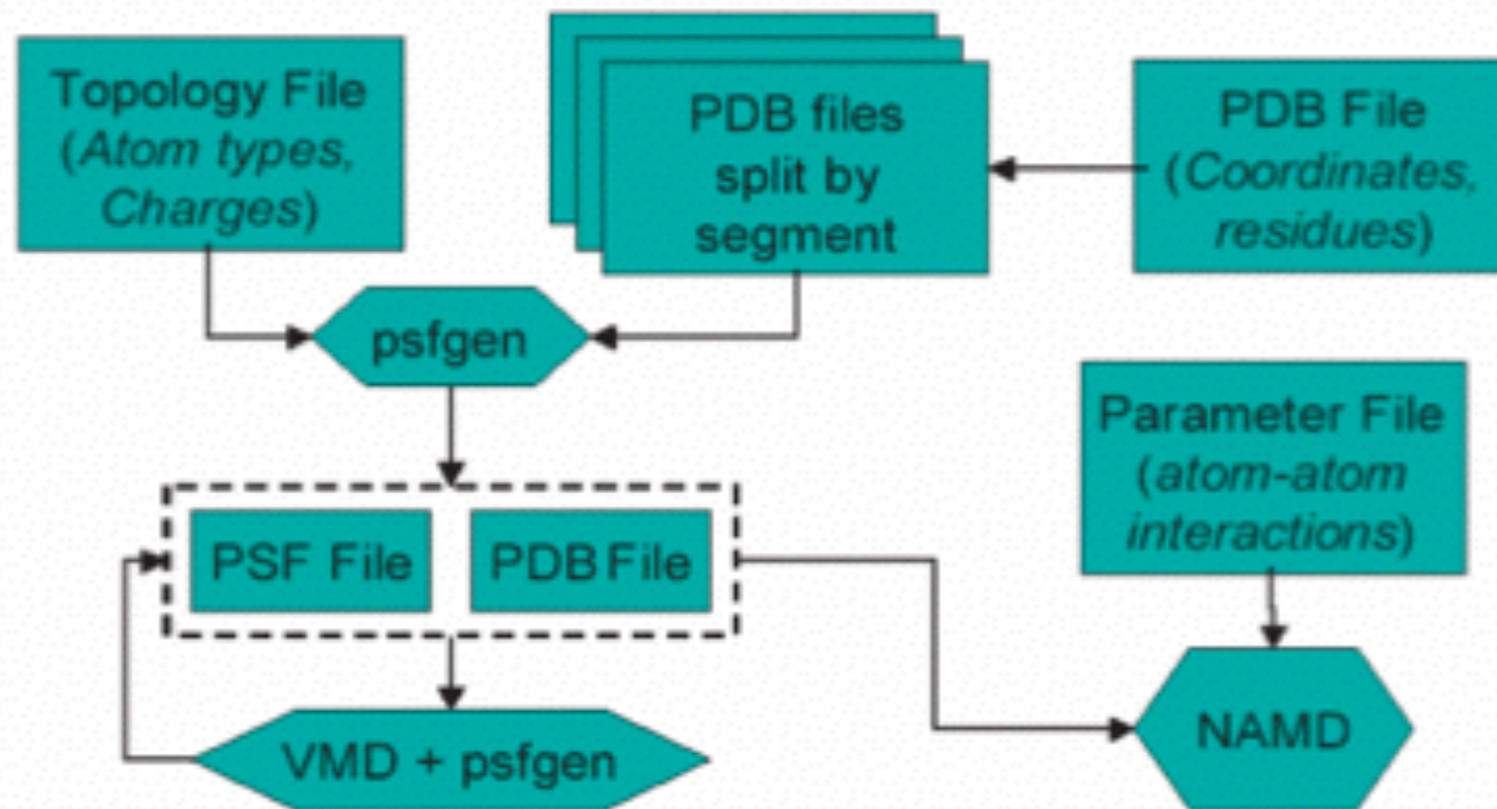
# Anatomy of a Topology File VIII

```
. . .  
PRES 5TER          0.00 ! 5'-terminal HYDROXYL patch, from MeOH  
                    ! use in generate statement  
  
GROUP  
ATOM H5T  HN5    0.43  
ATOM O5'  ON5   -0.66  
ATOM C5'  CN8B   0.05  
ATOM H5'  HN8    0.09  
ATOM H5'' HN8    0.09  
!  
DELETE ATOM P  
DELETE ATOM O1P  
DELETE ATOM O2P  
!  
BOND H5T   O5'  
DONO H5T   O5'  
BILD H5T  O5'  C5'  C4'  0.0000  0.00 180.00  0.00 0.0000
```

```
PRES 3TER          0.00 ! 3'terminal HYDROXYL patch, from MeOH  
                    ! use in generate statement  
  
GROUP  
ATOM C3'  CN7    0.14  
ATOM H3'  HN7    0.09  
ATOM O3'  ON5   -0.66  
ATOM H3T  HN5    0.43  
BOND O3'  H3T  
DONO H3T  O3'  
BILD H3T  O3'  C3'  C4'  0.9600 114.97 148.63 111.92 1.5284
```

. . .

# The Process



# PSFGEN

A program to read your structure files (separate PDB files for each segment) and use the topology file (“dictionary”) to build the complete system.

The program needs an input file to tell it what to do

**USAGE (unix prompt):**

**>psfgen build\_system.psfgen >& build.out &**

Sample input file (here named: build\_system.psfgen):

```
topology /home/mdschool/toppar/top_all27_prot_na.inp
```

```
segment TIP3 {  
  auto none
```

```
  pdb output_building/oh2.pdb
```

```
}  
coordpdb output_building/oh2.pdb TIP3
```

```
guesscoord
```

```
writesf output_building/one_water.psf
```

```
writepdb output_building/one_water.pdb
```

← path to topology file

← define a segment  
← pre-build command

← path to pdb file

← path coordinate file (pdb)

← use IC to add H etc.

**OUTPUT FILES**

# PSFGEN II (OUTPUT)

```
ATOM      1  OH2 TIP3 A      1      0.000  0.000  0.000  1.00  0.00      WAT1 O
```

```
(jc@ocho)psfgen/single_water% cat build.out
```

```
...
```

```
Created by CHARMM version 27 1
duplicate residue key CAL will be ignored
duplicate residue key DUM will be ignored
building segment TIP3
disabling angle autogeneration
disabling dihedral autogeneration
reading residues from pdb file output_building/oh2.pdb
extracted 1 residues from pdb file
Info: generating structure...
Info: segment complete.
reading coordinates from pdb file output_building/oh2.pdb for segment TIP3
Info: guessing coordinates for 2 atoms (0 non-hydrogen)
Warning: poorly guessed coordinates for 2 atoms (0 non-hydrogen):
Warning: poorly guessed coordinate for atom H1      TIP3:A      TIP3
Warning: poorly guessed coordinate for atom H2      TIP3:A      TIP3
Info: writing psf file output_building/one_water.psf
total of 3 atoms
total of 3 bonds
total of 1 angles
total of 0 dihedrals
total of 0 impropers
total of 0 cross-terms
Info: psf file complete.
Info: writing pdb file output_building/one_water.pdb
Info: Atoms with guessed coordinates will have occupancy of 0.0.
Info: pdb file complete.
```

## HINTS:

**“poorly guessed ...”**

search PDB for 0.000

search PDB for 99999.999

**new PDB file:**

**important! (H, patches, ...  
etc.)**

**NEVER ASSUME that since  
the script ran that it did what  
you expected**

inspect output

inspect output files

inspect structure (VMD)

# PSFGEN II (OUTPUT)

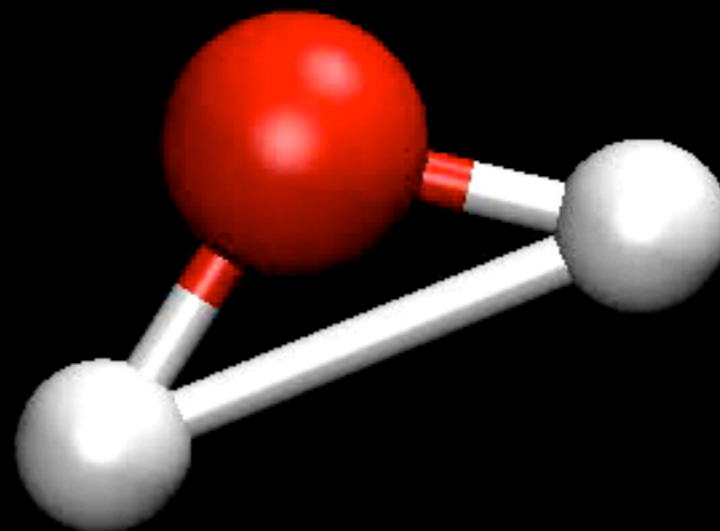
```
ATOM      1  OH2  TIP3  A      1      0.000      0.000      0.000      1.00      0.00      WAT1  O
```

```
REMARK original generated coordinate pdb file
ATOM      1  OH2  TIP3      0      0.000      0.000      0.000      1.00      0.00      TIP3  O
ATOM      2  H1   TIP3      0      1.000      0.000      0.000      0.00      0.00      TIP3  H
ATOM      3  H2   TIP3      0     -0.326      0.946      0.000      0.00      0.00      TIP3  H
END
```

# PSFGEN II (OUTPUT)

```
ATOM      1  OH2 TIP3 A      1      0.000      0.000      0.000      1.00      0.00      WAT1 O
```

```
REMARK original  
ATOM      1  OH  
ATOM      2  H1  
ATOM      3  H2  
END
```



```
0.00      TIP3 O  
0.00      TIP3 H  
0.00      TIP3 H
```

# PSFGEN III

## protein

Sample input file (here named: build\_system\_protein.psfgen):

```
>psfgen build_system_protein.psfgen >& build_protein.out &
```

```
topology /home/mdschool/toppar/top_all27_prot_na.inp
alias residue HIS HSE
segment PCLN {
  first NTER
  last CTER
  pdb output_building/icln.pdb
}
coordpdb output_building/icln.pdb PCLN
guesscoord
writepsf output_building/new_icln.psf
writepdb output_building/new_icln.pdb
```

← path to topology file

← correct naming issues

← pre-build commands

← path to pdb file (sequence)

← path coordinate file (pdb)

← use IC to add H etc.

← **OUTPUT FILES**

all terminal patches go in segment area

# PSFGEN IV (OUTPUT)

```
(jc@ocho)psfgen/protein% cat build_protein.out
```

...

```
Created by CHARMM version 27 1
duplicate residue key CAL will be ignored
duplicate residue key DUM will be ignored
aliasing residue HIS to HSE
building segment PCLN
setting patch for first residue to NTER
setting patch for last residue to CTER
reading residues from pdb file output_building/icln.pdb
extracted 116 residues from pdb file
Info: generating structure...
Info: skipping improper N-C-CA-HN at beginning of segment.
Info: skipping conformation C-N-CA-C at beginning of segment.
Info: skipping conformation C-CA-N-HN at beginning of segment.
Info: skipping bond C-N at end of segment.
Info: skipping improper C-CA-N-O at end of segment.
Info: skipping conformation CA-C-N-CA at end of segment.
Info: skipping conformation N-CA-C-O at end of segment.
Info: skipping conformation N-CA-C-N at end of segment.
Info: segment complete.
reading coordinates from pdb file output_building/icln.pdb for segment PCLN
Warning: failed to set coordinate for atom H   GLN:27  PCLN
Warning: failed to set coordinate for atom HB3  GLN:27  PCLN
Warning: failed to set coordinate for atom HG3  GLN:27  PCLN
Warning: failed to set coordinate for atom H   GLN:28  PCLN
Warning: failed to set coordinate for atom HB3  GLN:28  PCLN
Warning: failed to set coordinate for atom HG3  GLN:28  PCLN
Warning: failed to set coordinate for atom H   GLN:29  PCLN
```

...

```
Info: guessing coordinates for 287 atoms (5 non-hydrogen)
Warning: poorly guessed coordinates for 127 atoms (2 non-hydrogen):
poorly guessed coordinate for atom HT1  GLN:27  PCLN
poorly guessed coordinate for atom HT2  GLN:27  PCLN
poorly guessed coordinate for atom HT3  GLN:27  PCLN
```

## HINTS:

**“Warning: failed to set coor. ...”**

pdb <-> topology naming  
issues (use more alias  
commands or fix pdb file)

## new PDB file:

important! (H, patches, ... etc.)

**NEVER ASSUME that since the  
script ran that it did what you  
expected**

inspect output  
inspect output files  
inspect structure (VMD)

Info: skipping bond C-N at end of segment.  
Info: skipping improper C-CA-N-O at end of segment.  
Info: skipping conformation CA-C-N-CA at end of segment.  
Info: skipping conformation N-CA-C-O at end of segment.  
Info: skipping conformation N-CA-C-N at end of segment.  
Info: segment complete.  
reading coordinates from pdb file output\_building/icln.pdb for segment PCLN  
Warning: failed to set coordinate for atom H GLN:27 PCLN  
Warning: failed to set coordinate for atom HB3 GLN:27 PCLN  
Warning: failed to set coordinate for atom HG3 GLN:27 PCLN  
Warning: failed to set coordinate for atom H GLN:28 PCLN  
Warning: failed to set coordinate for atom HB3 GLN:28 PCLN  
Warning: failed to set coordinate for atom HG3 GLN:28 PCLN  
Warning: failed to set coordinate for atom H GLN:29 PCLN  
...  
Info: guessing coordinates for 287 atoms (5 non-hydrogen)  
Warning: poorly guessed coordinates for 127 atoms (2 non-hydrogen):  
Warning: poorly guessed coordinate for atom HT1 GLN:27 PCLN  
Warning: poorly guessed coordinate for atom HT2 GLN:27 PCLN  
Warning: poorly guessed coordinate for atom HT3 GLN:27 PCLN  
Warning: poorly guessed coordinate for atom HB1 GLN:27 PCLN  
Warning: poorly guessed coordinate for atom HG1 GLN:27 PCLN  
Warning: poorly guessed coordinate for atom HB1 GLN:28 PCLN  
Warning: poorly guessed coordinate for atom HG1 GLN:28 PCLN  
Warning: poorly guessed coordinate for atom HB1 GLN:29 PCLN  
...  
Info: writing psf file output\_building/new\_icln.psf  
total of 1745 atoms  
total of 1762 bonds  
total of 3177 angles  
total of 4651 dihedrals  
total of 297 impropers  
total of 0 cross-terms  
Info: psf file complete.  
Info: writing pdb file output\_building/new\_icln.pdb  
Info: Atoms with guessed coordinates will have occupancy of 0.0.  
Info: pdb file complete.  
PCLN:27  
PCLN:28  
PCLN:29

## HINTS:

**“Warning: failed to set coor. ...”**

pdb <-> topology naming  
issues (use more alias  
commands or fix pdb file)

**new PDB file:**

important! (H, patches, ... etc.)

**NEVER ASSUME that since the  
script ran that it did what you  
expected**

inspect output

inspect output files

inspect structure (VMD)

## Original pdb file (icln.pdb)

## Topology for RESI GLN

```

ATOM      1  N   GLN A  27
ATOM      2  CA  GLN A  27
ATOM      3  C   GLN A  27
ATOM      4  O   GLN A  27
ATOM      5  CB  GLN A  27
ATOM      6  CG  GLN A  27
ATOM      7  CD  GLN A  27
ATOM      8  OE1 GLN A  27
ATOM      9  NE2 GLN A  27
ATOM     10  H   GLN A  27
ATOM     11  HA  GLN A  27
ATOM     12  HB2 GLN A  27
ATOM     13  HB3 GLN A  27
ATOM     14  HG2 GLN A  27
ATOM     15  HG3 GLN A  27
ATOM     16  HE21 GLN A  27
ATOM     17  HE22 GLN A  27

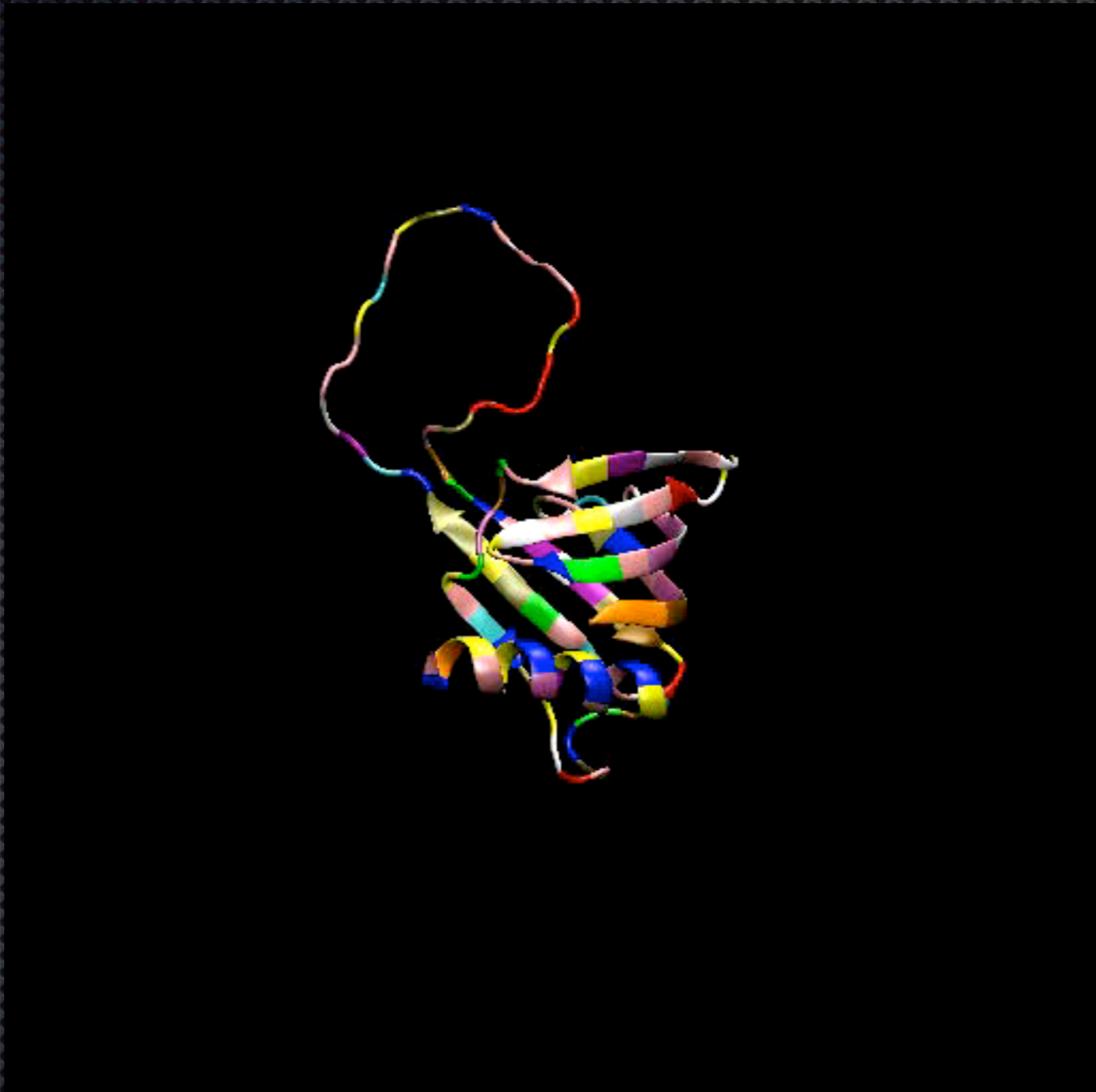
```

. . .

```

RESI GLN      0.00
GROUP
ATOM N        NH1    -0.47  !      |
ATOM HN       H       0.31  !  HN-N
ATOM CA       CT1     0.07  !      |      HB1 HG1 OE1   HE21 (cis to OE1)
ATOM HA       HB      0.09  !      |      |      |      ||     /
GROUP        !  HA-CA--CB--CG--CD--NE2
ATOM CB       CT2    -0.18  !      |      |      |      \
ATOM HB1      HA      0.09  !      |      HB2 HG2      HE22 (trans to OE1)
ATOM HB2      HA      0.09  !  O=C
GROUP        !      |
ATOM CG       CT2    -0.18
ATOM HG1      HA      0.09
ATOM HG2      HA      0.09
GROUP
ATOM CD       CC      0.55
ATOM OE1      O      -0.55
GROUP
ATOM NE2      NH2    -0.62
ATOM HE21     H       0.32
ATOM HE22     H       0.30
GROUP
ATOM C        C       0.51
ATOM O        O      -0.51

```



# PSFGEN V

## protein II (with patches)

Sample input file (here named: build\_system\_protein\_II.psfgen):

>psfgen build\_system\_protein\_II.psfgen >& build\_protein\_II.out &

```
topology /home/mdschool/toppar/top_all127_prot_na.inp
```

← path to topology file

```
alias residue HIS HSE
```

```
alias atom ILE CD1 CD
```

← correct naming issues

```
segment HYDR {
```

```
  first NTER
```

```
  last CTER
```

```
  pdb output_building/hydrolase.pdb
```

```
}
```

← pre-build commands

← path to pdb file (sequence)

```
coordpdb output_building/hydrolase.pdb HYDR
```

← path coordinate file (pdb)

```
patch DISU HYDR:22 HYDR:157
```

```
patch DISU HYDR:42 HYDR:58
```

```
patch DISU HYDR:128 HYDR:232
```

```
patch DISU HYDR:136 HYDR:201
```

```
patch DISU HYDR:168 HYDR:182
```

```
patch DISU HYDR:191 HYDR:220
```

← post-build commands

```
guesscoord
```

← use IC to add H etc.

```
writesf output_building/new_hydrolase.psf
```

← **OUTPUT FILES**

```
writpdb output_building/new_hydrolase.pdb
```

# PSFGEN VI (OUTPUT)

```
(jc@ocho)psfgen/protein% cat build_protein_II.out
```

...

```
Created by CHARMM version 27 1
duplicate residue key CAL will be ignored
duplicate residue key DUM will be ignored
aliasing residue HIS to HSE
aliasing residue ILE atom CD1 to CD
building segment HYDR
setting patch for first residue to NTER
setting patch for last residue to CTER
reading residues from pdb file output_building/hydrolase.pdb
extracted 220 residues from pdb file
Info: generating structure...
Info: skipping improper N-C-CA-HN at beginning of segment.
Info: skipping conformation C-N-CA-C at beginning of segment.
Info: skipping conformation C-CA-N-HN at beginning of segment.
Info: skipping bond C-N at end of segment.
Info: skipping improper C-CA-N-O at end of segment.
Info: skipping conformation CA-C-N-CA at end of segment.
Info: skipping conformation N-CA-C-O at end of segment.
Info: skipping conformation N-CA-C-N at end of segment.
Info: segment complete.
reading coordinates from pdb file output_building/hydrolase.pdb for segment HYDR
Warning: failed to set coordinate for atom O ASN:245 HYDR
Warning: failed to set coordinate for atom OXT ASN:245 HYDR
applying patch DISU to 2 residues
applying patch DISU to 2 residues
applying patch DISU to 2 residues
applying patch DISU to 2 residues
applying patch DISU to 2 residues
Info: guessing coordinates for 1563 atoms (2 non-hydrogen)
Warning: poorly guessed coordinates for 19 atoms (2 non-hydrogen):
```

```
Warning: poorly guessed coordinate for atom HT1 ILE:16 HYDR
```

```
Warning: poorly guessed coordinate for atom HT2 ILE:16 HYDR
```

```
Warning: poorly guessed coordinate for atom HT3 ILE:16 HYDR
```

```
Warning: poorly guessed coordinate for atom HC ILE:22 HYDR
```

## HINTS:

**“Warning: failed to set coor. ...”**

pdb <-> topology naming  
issues (use more alias  
commands or fix pdb file)

**new PDB file:**

important! (H, patches, ... etc.)

**NEVER ASSUME that since the  
script ran that it did what you  
expected**

inspect output

inspect output files

inspect structure (VMD)

applying patch DISU to 2 residues  
applying patch DISU to 2 residues  
applying patch DISU to 2 residues  
applying patch DISU to 2 residues

Info: guessing coordinates for 1563 atoms (2 non-hydrogen)

Warning: poorly guessed coordinates for 19 atoms (2 non-hydrogen):

Warning: poorly guessed coordinate for atom HT1 ILE:16 HYDR  
Warning: poorly guessed coordinate for atom HT2 ILE:16 HYDR  
Warning: poorly guessed coordinate for atom HT3 ILE:16 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:33 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:46 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:67 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:99 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:105 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:108 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:114 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:123 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:137 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:155 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:158 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:163 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:185 HYDR  
Warning: poorly guessed coordinate for atom HG LEU:209 HYDR  
Warning: poorly guessed coordinate for atom OT1 ASN:245 HYDR  
Warning: poorly guessed coordinate for atom OT2 ASN:245 HYDR

Info: writing psf file output\_building/new\_hydrolase.psf

total of 3160 atoms  
total of 3196 bonds  
total of 5767 angles  
total of 8444 dihedrals  
total of 569 impropers  
total of 0 cross-terms

Info: psf file complete.

Info: writing pdb file output\_building/new\_hydrolase.pdb

Info: Atoms with guessed coordinates will have occupancy of 0.0.

HYDR:245

HYDR:22 HYDR:157

HYDR:42 HYDR:58

HYDR:128 HYDR:232

HYDR:136 HYDR:201

HYDR:168 HYDR:182

HYDR:191 HYDR:220

## HINTS:

**“Warning: failed to set coor. ...”**

pdb <-> topology naming  
issues (use more alias  
commands or fix pdb file)

**new PDB file:**

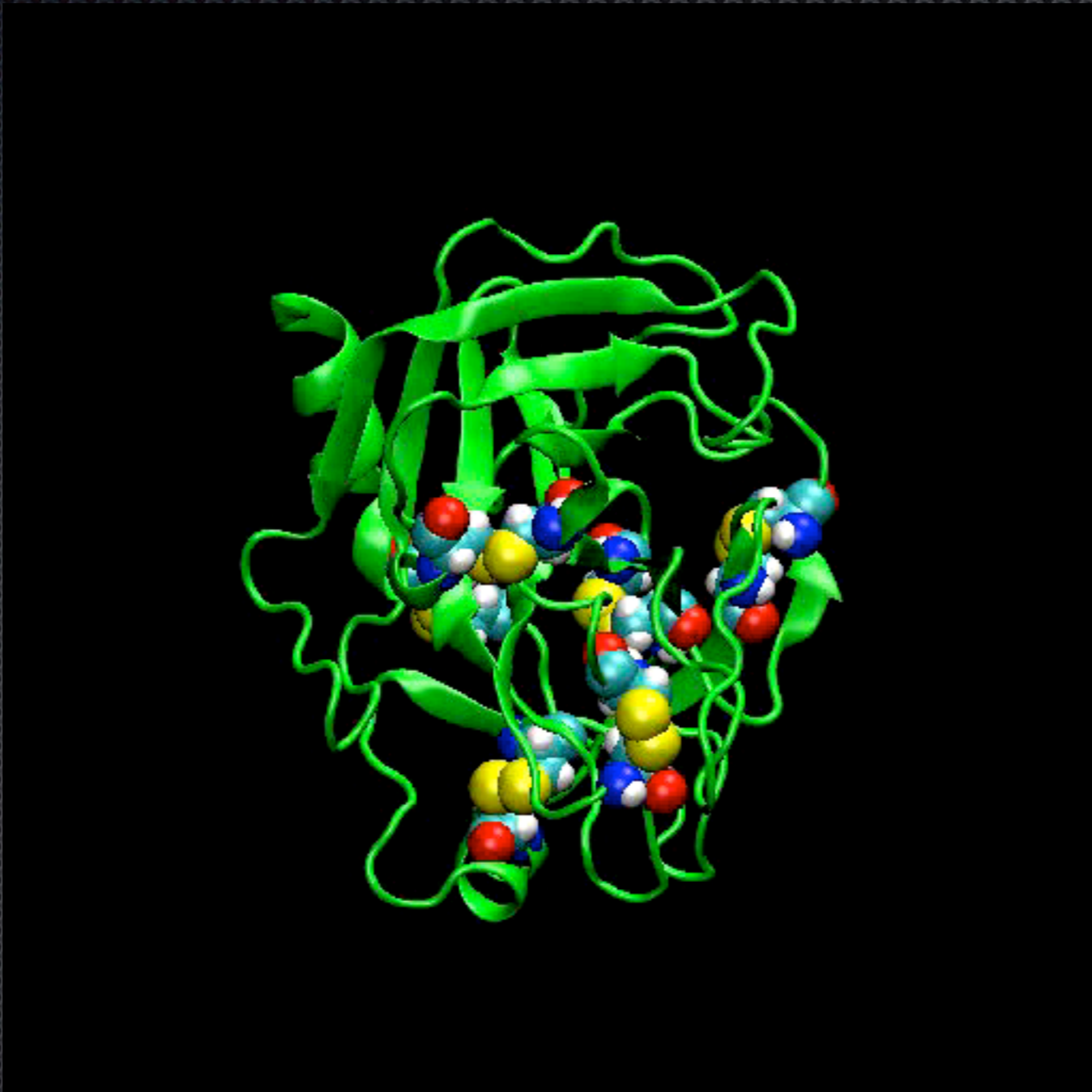
important! (H, patches, ... etc.)

**NEVER ASSUME that since the  
script ran that it did what you  
expected**

inspect output

inspect output files

inspect structure (VMD)



# PSFGEN VII

## protein III (multiple segments with patches)

Sample input file (here named: build\_system\_protein\_III.psfgen):

>psfgen build\_system\_protein\_III.psfgen >& build\_protein\_III.out &

```
topology /home/mdschool/toppar/top_all127_prot_na.inp
```

← path to topology file

```
alias residue HIS HSE  
alias atom ILE CD1 CD
```

← correct naming issues

```
segment FIXA {  
  first NTER  
  last CTER
```

← pre-build commands

```
  pdb output_building/chain_a.pdb
```

← path to pdb file (sequence)

```
}
```

```
coordpdb output_building/chain_a.pdb FIXA
```

← path coordinate file I (pdb)

```
segment FIXB {
```

```
  first NTER
```

← pre-build commands

```
  last CTER
```

```
  pdb output_building/chain_b.pdb
```

← path to pdb file II (sequence)

```
}
```

```
coordpdb output_building/chain_b.pdb FIXB
```

← path coordinate file II (pdb)

```
patch DISU FIXA:42 FIXA:58
```

```
patch DISU FIXA:122 FIXB:132
```

← post-build commands

```
patch DISU FIXA:168 FIXA:182
```

```
patch DISU FIXA:191 FIXA:220
```

```
patch DISU FIXB:88 FIXB:99
```

```
patch DISU FIXB:95 FIXB:109
```

```
patch DISU FIXB:111 FIXB:124
```

```
guesscoord
```

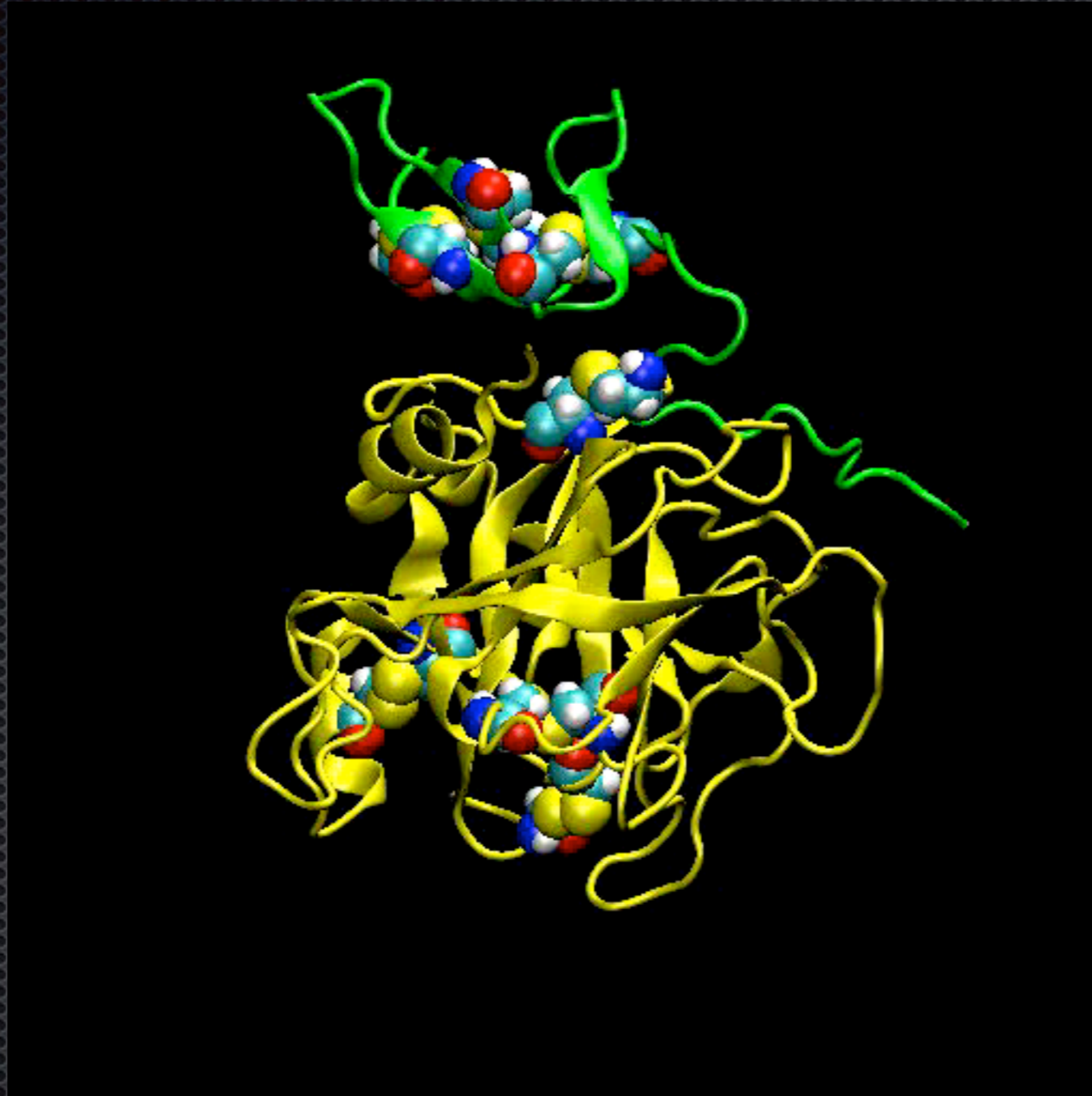
← use IC to add H etc.

```
writesf output_building/new_factorIXA.psf
```

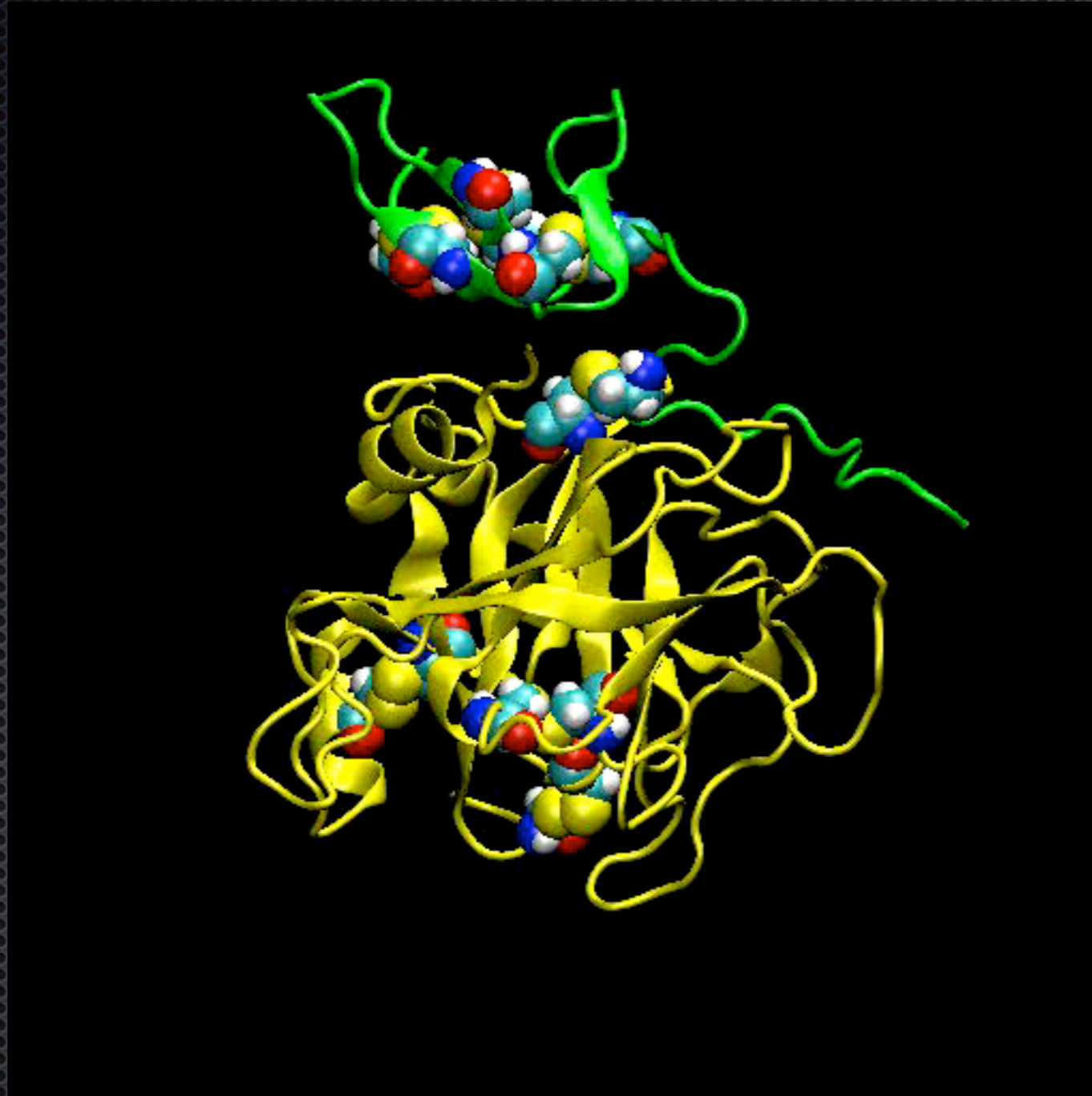
← OUTPUT FILES

```
writepdb output_building/new_factorIXA.pdb
```

# PSFGEN VII (OUTPUT)



# PSFGEN VII (OUTPUT)



# PSFGEN IX

## protein/dna complex

Sample input file (here named: build\_system\_protein\_dna\_complex.psfgen):

```
>psfgen build_system_protein_dna_complex.psfgen >& build_protein_dna_complex.out &
```

```
topology /home/mdschool/toppar/top_all127_prot_na.inp
```

```
alias residue HIS HSE  
alias atom ILE CD1 CD  
alias atom SER HG HG1  
alias atom CYS HG HG1  
alias atom THY C5A C5M
```

```
segment INT1 {  
  first NTER  
  last CTER  
  pdb output_building/fixed_int1.pdb  
}
```

```
coordpdb output_building/fixed_int1.pdb INT1
```

```
segment INT2 {  
  first NTER  
  last CTER  
  pdb output_building/fixed_int2.pdb  
}
```

```
coordpdb output_building/fixed_int2.pdb INT2
```

```
segment INT3 {  
  first NTER  
  last CTER  
  pdb output_building/fixed_int3.pdb  
}
```

```
coordpdb output_building/fixed_int3.pdb INT3
```

```
segment INT4 {  
  first NTER  
  last CTER
```

System:

4 Integrase proteins

2 LEDGF proteins

3 pieces of DNA (6 chains)\*

```

first 5TER
last 3TER
pdb output_building/dna1.pdb
}
coordpdb output_building/dna1.pdb DNA1

segment DNA2 {
first 5TER
last 3TER
pdb output_building/dna2.pdb
}
coordpdb output_building/dna2.pdb DNA2

segment DNA3 {
first 5TER
last 3TER
pdb output_building/dna3.pdb
}
coordpdb output_building/dna3.pdb DNA3

segment DNA4 {
first 5TER
last 3TER
pdb output_building/dna4.pdb
}
coordpdb output_building/dna4.pdb DNA4

segment DNA5 {
first 5TER
last 3TER
pdb output_building/dna5.pdb
}
coordpdb output_building/dna5.pdb DNA5

segment DNA6 {
first 5TER
last 3TER
pdb output_building/dna6.pdb
}
coordpdb output_building/dna6.pdb DNA6

guesscoord

writepsf output_building/new_complex.psf
writepdb output_building/new_complex.pdb

```

**System:**

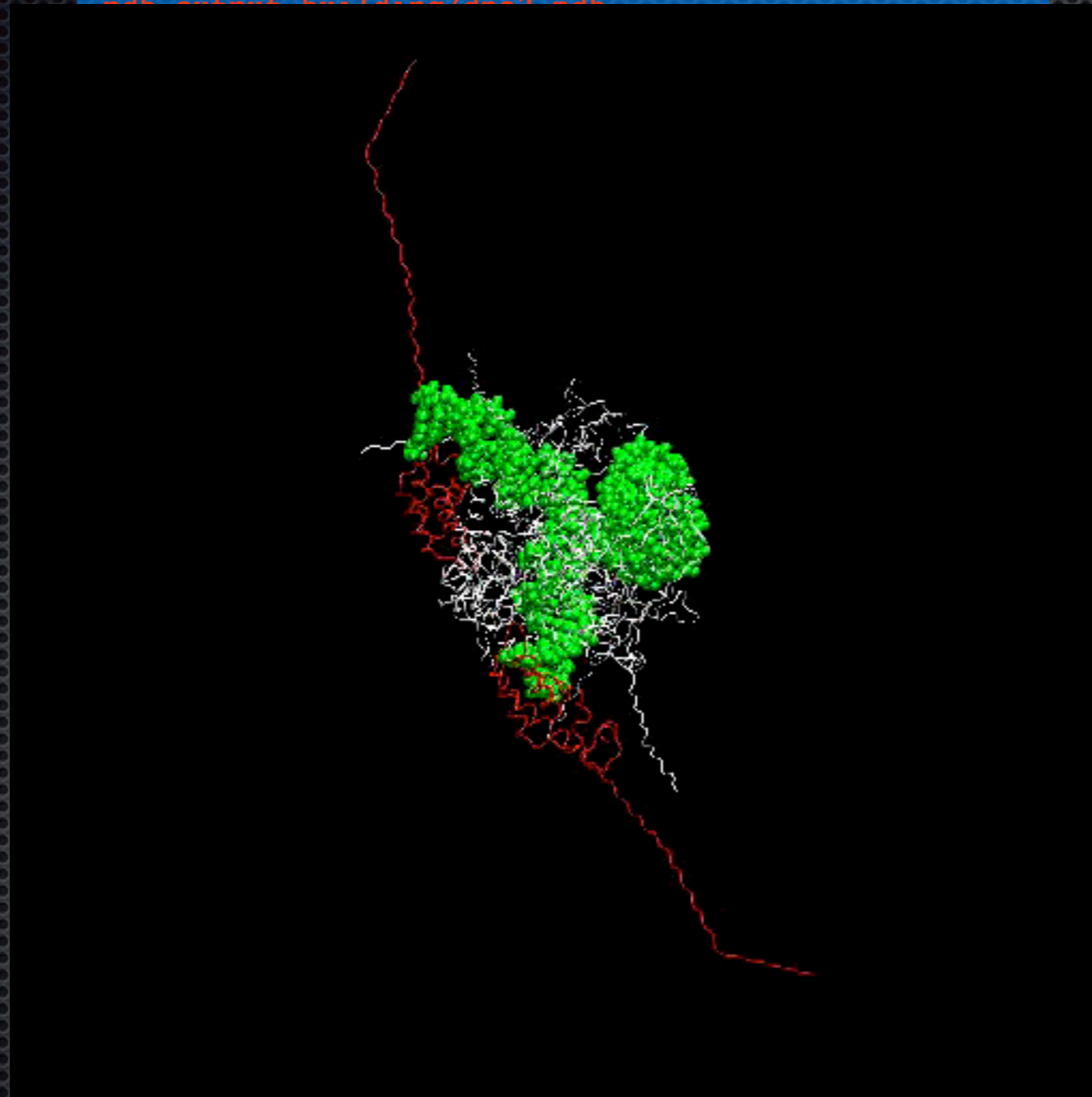
**4 Integrase proteins**

**2 LEDGF proteins**

**3 pieces of DNA (6 chains)\***

```
first 5TER
last 3TER
pdb output_building/dna1.pdb
}
coordpdb output_building/dna1.pdb DNA1

segment DNA2 {
  first 5TER
  last 3TER
  pdb output_building/dna2.pdb
}
```



**System:**

**4 Integrase proteins**

**2 LEDGF proteins**

**3 pieces of DNA (6 chains)\***

```
guesscoord
```

```
writepsf output_building/new_complex.psf
writepdb output_building/new_complex.pdb
```

**Actually, for DNA you need to apply two patches (one specific patch for each residue type for every residue individually):**

**DEO1 : PYRIMIDINES**

**DEO2: PURINES**

**LAB\_2: (today)**

# PSF File of Water

PSF

```
3 !NTITLE
REMARKS original generated structure x-plor psf file
REMARKS topology /Users/curtisj/research/toppar/top_all27_prot_na.inp
REMARKS segment TIP3 { first NONE; last NONE; auto none }

3 !NATOM
1 TIP3 A TIP3 OH2 OT -0.834000 15.9994 0
2 TIP3 A TIP3 H1 HT 0.417000 1.0080 0
3 TIP3 A TIP3 H2 HT 0.417000 1.0080 0

3 !NBOND: bonds
1 2 1 3 2 3

1 !NTHETA: angles
2 1 3

0 !NPHI: dihedrals

0 !NIMPHI: impropers

. . . etc. . . .
```

**“Atom Types” != “PDB Atom Name”**

# PSF File of Protein I

PSF

6 !NTITLE

REMARKS original generated structure x-plor psf file

REMARKS 2 patches were applied to the molecule.

REMARKS topology /Users/curtisj/research/toppar/top\_all27\_prot\_na.inp

REMARKS segment PCLN { first NTER; last CTER; auto angles dihedrals }

REMARKS patch NTER PCLN:27

REMARKS patch CTER PCLN:142

1745 !NATOM

1	PCLN	27	GLN	N	NH3	-0.300000	14.0070	0
2	PCLN	27	GLN	HT1	HC	0.330000	1.0080	0
3	PCLN	27	GLN	HT2	HC	0.330000	1.0080	0
4	PCLN	27	GLN	HT3	HC	0.330000	1.0080	0
5	PCLN	27	GLN	CA	CT1	0.210000	12.0110	0
6	PCLN	27	GLN	HA	HB	0.100000	1.0080	0
7	PCLN	27	GLN	CB	CT2	-0.180000	12.0110	0
8	PCLN	27	GLN	HB1	HA	0.090000	1.0080	0
9	PCLN	27	GLN	HB2	HA	0.090000	1.0080	0
10	PCLN	27	GLN	CG	CT2	-0.180000	12.0110	0
11	PCLN	27	GLN	HG1	HA	0.090000	1.0080	0
12	PCLN	27	GLN	HG2	HA	0.090000	1.0080	0

. . .

1744	PCLN	142	LEU	HD22	HA	0.090000	1.0080	0
1745	PCLN	142	LEU	HD23	HA	0.090000	1.0080	0

# PSF File of Protein II

. . .

1762 !NBOND: bonds

1	5	2	1	3	1	4	1
5	6	7	5	7	8	7	9
10	7	10	11	10	12	13	10
13	14	15	13	15	16	15	17
18	5	18	20	19	18	20	21
20	22	22	23	24	22	24	25
24	26	27	24	27	28	27	29

. . .

1738	1740	1738	1741	1742	1736	1742	1743
1742	1744	1742	1745				

. . .

# PSF File of Protein III

...

3177 !NTHETA: angles

1	5	6	1	5	18	2	1	5
2	1	4	2	1	3	3	1	5
3	1	4	4	1	5	5	18	19
5	18	20	5	7	9	5	7	8
5	7	10	7	10	12	7	10	11
7	10	13	7	5	6	7	5	18

...

1739	1738	1740	1740	1738	1741	1742	1736	1737
1743	1742	1745	1743	1742	1744	1744	1742	1745

...

# PSF File of Protein IV

. . .

4651 !NPHI: dihedrals

1	5	7	10	1	5	7	8
1	5	7	9	1	5	18	20
1	5	18	19	2	1	5	7
2	1	5	18	2	1	5	6
3	1	5	7	3	1	5	18
3	1	5	6	4	1	5	7

. . .

1739	1738	1736	1742	1740	1738	1736	1742
1741	1738	1736	1742				

. . .

# PSF File of Protein V

. . .

297 !NIMPFI: impropers

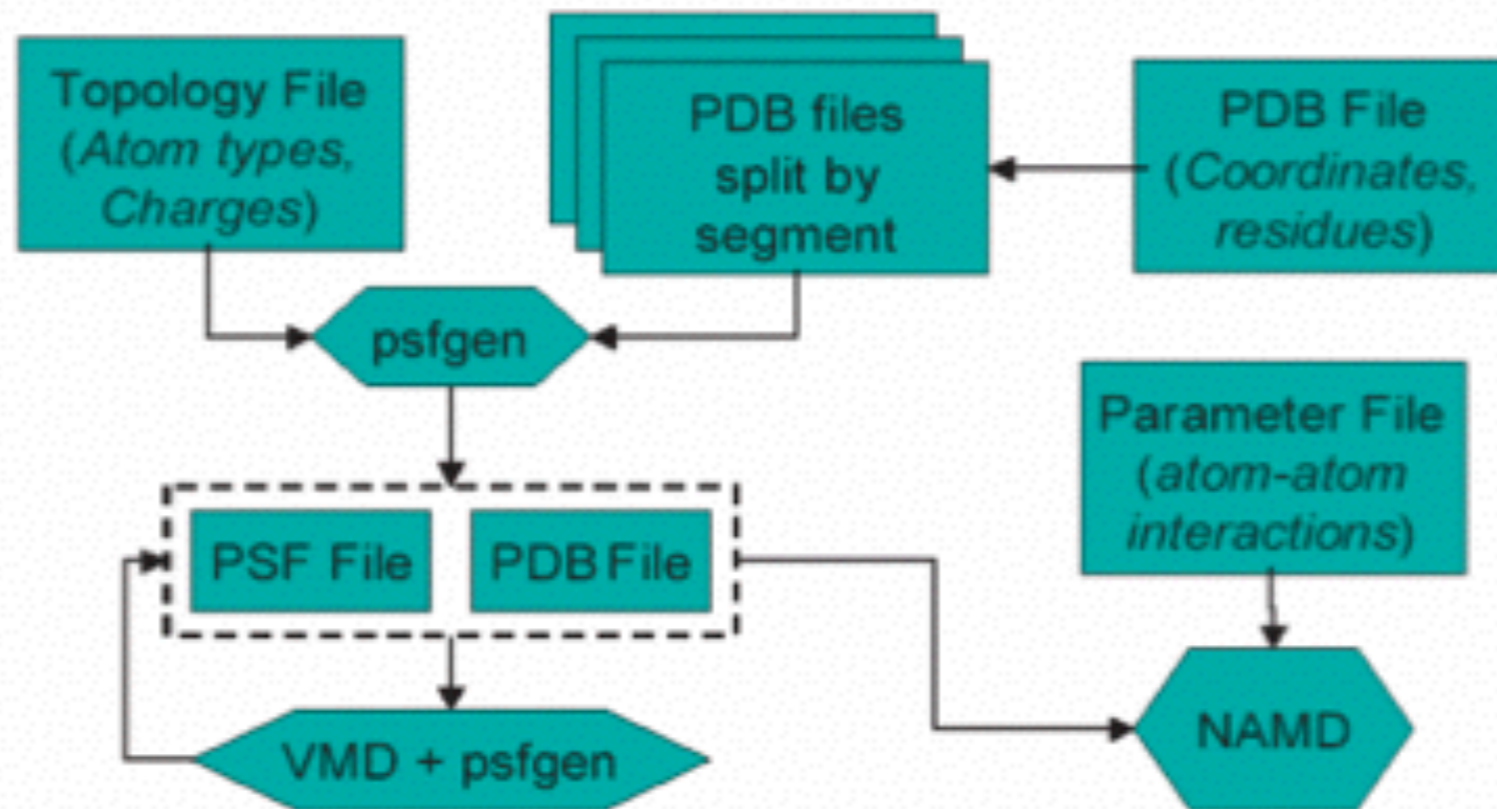
13	15	10	14	13	10	15	14
15	13	16	17	15	13	17	16
18	5	20	19	20	18	22	21
30	32	27	31	30	27	32	31
32	30	33	34	32	30	34	33
35	22	37	36	37	35	39	38

. . .

1724	1718	1729	1725	1726	1731	1728	1727
1729	1724	1731	1730				

. . .

# The Process



# Parameter File I

## BONDS

```
!  
!V(bond) = Kb(b - b0)**2  
!  
!Kb: kcal/mole/A**2  
!b0: A  
!  
!atom type Kb          b0  
!  
!Carbon Dioxide  
CST  OST   937.96      1.1600 ! JES  
!Heme to Sulfate (PSUL) link  
SS   FE    250.0       2.3200 !force constant chosen so  
      !equilibrium bond length optimized to reproduce  
      !CSD survey values of  
      !2.341pm0.01 (mean, standard error)  
      !adm jr., 7/01  
C    C     600.000     1.3350 ! ALLOW ARO HEM  
      ! Heme vinyl substituent (KK, from propene (JCS))  
CA   CA    305.000     1.3750 ! ALLOW  ARO  
      ! benzene, JES 8/25/89  
CE1  CE1   440.000     1.3400  !  
      ! for butene; from propene, yin/adm jr., 12/95  
CE1  CE2   500.000     1.3420  !  
      ! for propene, yin/adm jr., 12/95  
CE1  CT2   365.000     1.5020  !  
      ! for butene; from propene, yin/adm jr., 12/95  
.  
.  
.
```

# Parameter File II

## ANGLES

```
!  
!V(angle) = Ktheta(Theta - Theta0)**2  
!  
!V(Urey-Bradley) = Kub(S - S0)**2  
!  
!Ktheta: kcal/mole/rad**2  
!Theta0: degrees  
!Kub: kcal/mole/A**2 (Urey-Bradley)  
!S0: A  
!  
!atom types      Ktheta      Theta0      Kub      S0  
!  
!Carbon Dioxide, JES  
OST  CST  OST      3000.00  180.0000  ! CO2, JES  
  
CA   CA   CA      40.000   120.00    35.00    2.41620  ! ALLOW  ARO  
      ! JES 8/25/89  
CE1  CE1  CT2      48.00    123.50    !  
      ! for 2-butene, yin/adm jr., 12/95  
CE1  CE1  CT3      48.00    123.50    !  
      ! for 2-butene, yin/adm jr., 12/95  
  
. . .
```

# Parameter File II

## DIHEDRALS

```
!  
!V(dihedral) = Kchi(1 + cos(n(chi) - delta))  
!  
!Kchi: kcal/mole  
!n: multiplicity  
!delta: degrees  
!  
!atom types          Kchi      n      delta  
!  
C      CT1  NH1  C      0.2000  1      180.00 ! ALLOW PEP  
          ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c  
C      CT2  NH1  C      0.2000  1      180.00 ! ALLOW PEP  
          ! ala dipeptide update for new C VDW Rmin, adm jr., 3/3/93c  
C      N    CP1  C      0.8000  3      0.00  ! ALLOW PRO PEP  
          ! 6-31g* AcProNH2, ProNH2, 6-31g*//3-21g AcProNHCH3 RLD  
4/23/93  
CA     CA   CA   CA     3.1000  2      180.00 ! ALLOW   ARO  
          ! JES 8/25/89  
CA     CPT  CPT  CA     3.1000  2      180.00 ! ALLOW   ARO  
          ! JWK 05/14/91 fit to indole
```

. . .

# Parameter File IV

## IMPROPER

```
!  
!V(improper) = Kpsi(psi - psi0)**2  
!  
!Kpsi: kcal/mole/rad**2  
!psi0: degrees  
!note that the second column of numbers (0) is ignored  
!  
!atom types          Kpsi          psi0  
!  
CPB  CPA  NPH  CPA  20.8000      0      0.0000 ! ALLOW HEM  
      ! Heme (6-liganded): porphyrin macrocycle (KK 05/13/91)  
CPB  X    X    CE1  90.0000      0      0.0000 ! ALLOW HEM  
      ! Heme (6-liganded): substituents (KK 05/13/91)  
CT2  X    X    CPB  90.0000      0      0.0000 ! ALLOW HEM  
      ! Heme (6-liganded): substituents (KK 05/13/91)  
CT3  X    X    CPB  90.0000      0      0.0000 ! ALLOW HEM  
      ! Heme (6-liganded): substituents (KK 05/13/91)  
!HA  C    C    HA   20.0000      0      0.0000 ! ALLOW  PEP POL ARO  
      ! Heme vinyl substituent (KK, from propene (JCS))
```

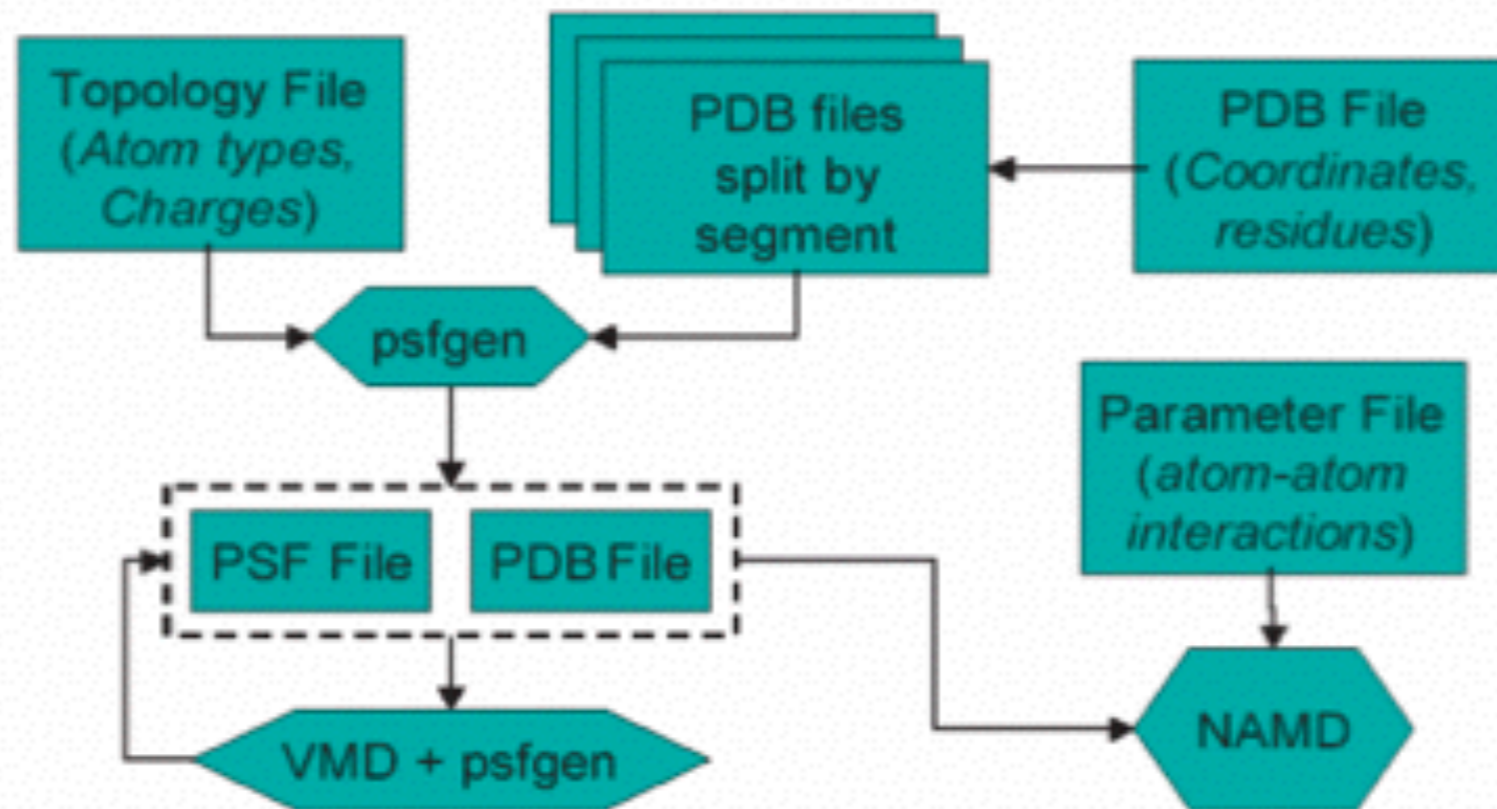
. . .

# Parameter File V

```
NONBONDED nbxmod 5 atom cdie1 shift vatom vdistance vswitch -
cutnb 14.0 ctofnb 12.0 ctonnb 10.0 eps 1.0 e14fac 1.0 wmin 1.5
!adm jr., 5/08/91, suggested cutoff scheme
!
!V(Lennard-Jones) = Eps,i,j[(Rmin,i,j/ri,j)**12 - 2(Rmin,i,j/ri,j)**6]
!
!epsilon: kcal/mole, Eps,i,j = sqrt(eps,i * eps,j)
!Rmin/2: A, Rmin,i,j = Rmin/2,i + Rmin/2,j
!
!atom ignored epsilon Rmin/2 ignored eps,1-4 Rmin/2,1-4
!
!carbons
C 0.000000 -0.110000 2.000000 ! ALLOW PEP POL ARO
! NMA pure solvent, adm jr., 3/3/93
CA 0.000000 -0.070000 1.992400 ! ALLOW ARO
! benzene (JES)
CC 0.000000 -0.070000 2.000000 ! ALLOW PEP POL ARO
! adm jr. 3/3/92, acetic acid heat of solvation
CD 0.000000 -0.070000 2.000000 ! ALLOW POL
! adm jr. 3/19/92, acetate a.i. and dH of solvation
```

. . .

# The Process



# NAMD

**UIUC:**

<http://www.ks.uiuc.edu/Research/namd/>

**Open source** (binaries and code) . . . no re-distribution

**Many hardware platforms** (desktop, GPU, Clusters, etc.)

**Superior parallel performance** (Charm++, infiniband, myranet, etc.)

**Simple input file syntax** and expandable via TCL scripting

**Charmm / Amber / Coarse Grain / etc. FF**

**Excellent tutorials and support**

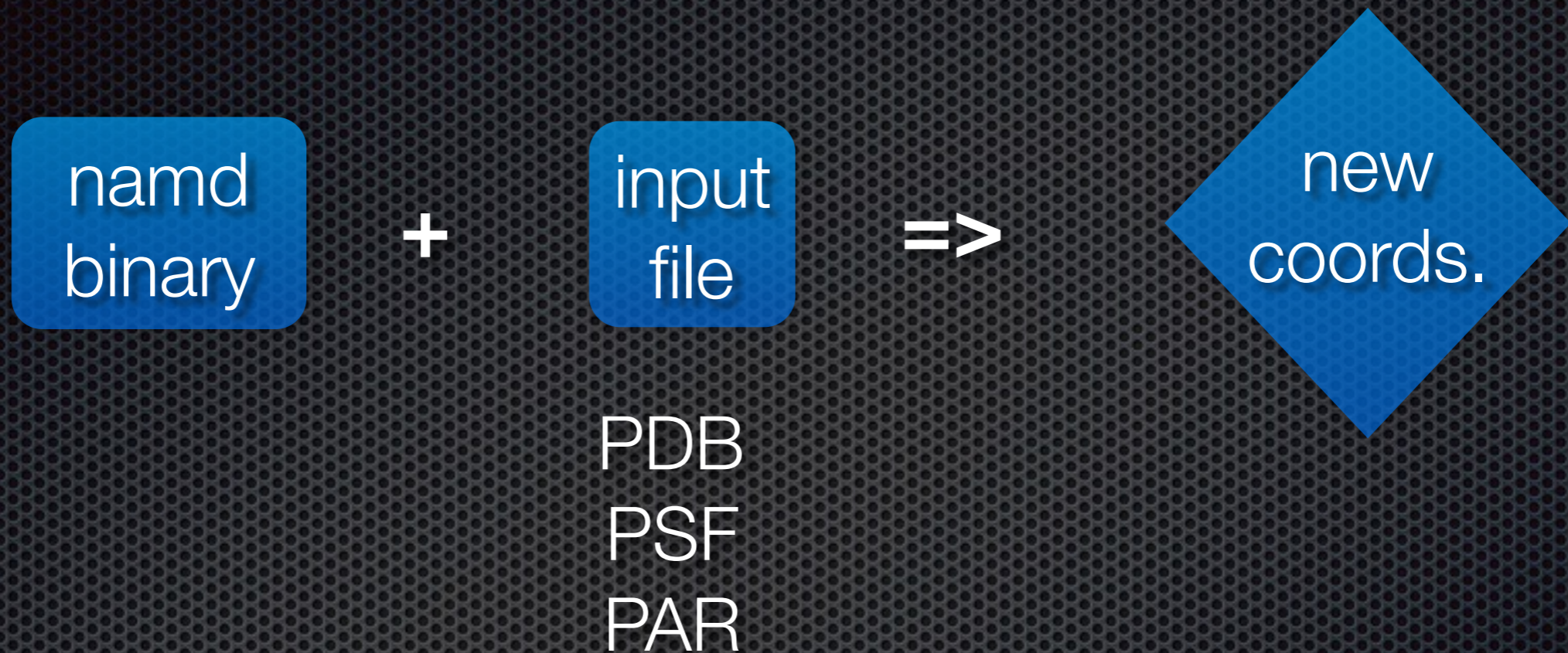
**Linked w/ VMD (visualization & control)**

**Runs many types of MD simulations**

NVE, NVT, NPT  
Constant area  
Steered MD  
Interactive MD (IMD)  
MD Flexible Fitting  
Free-Energy  
etc.

- [AIX-POWER-lapi](#) (IBM POWER clusters)
- [AIX-POWER-multicore](#) (IBM POWER single node)
- [Linux-x86](#) (32-bit Intel/AMD with ethernet)
- [Linux-x86-TCP](#) (TCP may be better on gigabit)
- [Linux-x86\\_64-multicore](#) (64-bit Intel/AMD single node)
- [Linux-x86\\_64](#) (64-bit Intel/AMD with ethernet)
- [Linux-x86\\_64-TCP](#) (TCP may be better on gigabit)
- [Linux-x86\\_64-ibverbs](#) (InfiniBand via OpenFabrics OFED, no MPI needed)
- [Linux-x86\\_64-ibverbs-smp](#) (InfiniBand plus shared memory, no MPI needed)
- [Linux-x86\\_64-CUDA](#) (NVIDIA CUDA acceleration)
- [Linux-x86\\_64-ibverbs-CUDA](#) (NVIDIA CUDA with InfiniBand)
- [MacOSX-x86](#) (Mac OS X for Intel processors)
- [MacOSX-PPC](#) (Mac OS X for PowerPC)
- [Solaris-x86\\_64](#)
- [Win32](#) (Windows XP, etc.)
- [Win64-MPI](#) (Windows HPC Server)

# NAMD II



MD parameters

1. Minimization
2. MD (time)
3. Directed MD

# NAMD INPUT FILE

```
# sample NAMD configuration file for Minimization
```

```
# molecular system
```

```
coordinates      output/new_icln.pdb
structure        output/new_icln.psf
temperature      300
```

PDB

PSF

```
# force field
```

```
paratypecharmm  on
parameters      /home/mdschool/toppar/par_all27_prot_na.inp
```

PAR

```
# approximations
```

```
exclude         scaled1-4
1-4scaling      1.0
switching       on
switchdist      10
cutoff          12
```

MD

parameters

```
# output
```

```
outputname      output/min0
binaryoutput    no
```

```
# run control
```

```
minimize        1000
```

# RUNNING NAMD

## USAGE (unix prompt):

### Start a run:

```
>namd2 min0 >& min0.out &
```

### Check status:

```
>tail -90f min0.out
```

### Multi-processor run:

```
>charmrun ++local ++p 2 /usr/local/bin/namd/namd2 min0 >& min0.out &
```

# CHARMM INPUT FILE

```
* title
*

open unit 1 read form name ~/toppar/top_all127_prot_na.inp
read rtf card unit 1
close unit 1

open unit 1 read form name ~/toppar/par_all127_prot_na.inp
read param card unit 1
close unit 1

open unit 1 read card name traI_308_to_380.pdb
read sequence pdb unit 1
close unit 1

generate TRAA first glyp last cter setup

open unit 1 read card name linker1_phyre_1.pdb
read unit 1 coor pdb
close unit 1

!ic seed
ic fill preserve
ic param
ic build

open unit 1 write form name new_linker1_phyre_1_plus_tail.pdb
write coor pdb unit 1
* fixed TRAA
* by jc 05/24/11
```

```
close unit 1
```

```
!ic seed  
ic fill preserve  
ic param  
ic build
```

```
open unit 1 write form name new_linker1_phyre_1_plus_tail.pdb  
write coor pdb unit 1  
* fixed TRAA  
* by jc 05/24/11
```

```
close unit 1
```

```
OPEN UNIT 1 WRITE CARD NAME new_linker1_phyre_1_plus_tail.psf  
WRITE PSF CARD UNIT 1  
*PSF File  
*
```

```
close unit 1
```

```
! Perform the minimisation.  
Minimise Abnr Nstep 5000 Nprint 50 Tolgrd 0.1 Inbfrq 0 Imgfrq 0
```

```
! Test the first derivatives.  
Test First Step 0.000001 Tol 0.000001
```

```
! Save the optimised coordinates.  
Open write card unit 17 name new_min_fixed_full10.pdb  
Write coor pdb unit 17  
* minimized full PARP  
* by jc 9/27/10  
*  
close unit 17
```

# AMBER INPUT FILE

```
#!/bin/csh -f
```

```
cat > mdin <<EOF
```

```
MD run, Temp=300K, start
```

```
#06-07-00#
```

```
11 A water box, const eps=1.0,
```

```
nscm=0, npscal=1, frameon=1 - pointing LP's as in Jim's code
```

```
const pressure MD, seperate temp coupling
```

```
&cntrl
```

```
irest=1, ibelly=0, imin=0,
```

```
ipol=1,
```

```
ntx=5, ntxo=1, ig=71277, tempi=0.00,
```

```
ntb=2,
```

```
ntt=0, temp0=300.0, tautp=0.2,
```

```
ntp=1, pres0=1.0, comp=44.6, taup=0.2,
```

```
nscm=0,
```

```
nstlim=10, t=1170.0, dt=0.001,
```

```
ntc=2, tol=0.0000005,
```

```
ntf=2, nsnb=5,
```

```
cut=9.0, dielc=1.0,
```

```
ntpr=1, ntwx=5, ntwv=5, ntwe=5,
```

```
ioutfm=0,
```

```
ntr=0, ntave=100, ntrx=1
```

```
/
```

```
&ewald
```

```
ew_type = 0, verbose=0,
```

```
vdwmeth=1, maxiter=10, indmeth = 3, diptau = 1.0,
```

```
frameon=1, irstdip=1, scaldip=0,
```

```
/
```

```
EOF
```

```
set output = mdout ubiquitin
```

```
ipol=1,
ntx=5, ntxo=1, ig=71277, tempi=0.00,
ntb=2,
ntt=0, temp0=300.0, tautp=0.2,
ntp=1, pres0=1.0, comp=44.6, taup=0.2,
nscm=0,
nstlim=10, t=1170.0, dt=0.001,
ntc=2, tol=0.0000005,
ntf=2, nsnb=5,
cut=9.0, dielc=1.0,
ntpr=1, ntwx=5, ntwv=5, ntwe=5,
ioutfm=0,
ntr=0, ntave=100, ntrx=1
/
&ewald
  ew_type = 0, verbose=0,
  vdwmeth=1, maxiter=10, indmeth = 3, diptau = 1.0,
  frameon=1, irstdip=1, scaldip=0,
/
EOF

set output = mdout.ubiquitin

$DO_PARALLEL $TESTsander -O -i mdin -p prmtop -c inpcrd -o $output || goto
error

../dacadif $output.save $output
# ../dacadif -t 2 rstcip.save rstcip
# ../dacadif -t 2 restrt.save restrt
../dacadif -t 2 mden.save mden
../dacadif mdcrd.save mdcrd
../dacadif -t 2 mdvel.save mdvel
/bin/rm -f mdin mdinfo
exit(0)
```

# CPMD INPUT FILE

```
&CPMD
MOLECULAR DYNAMICS
RESTART WAVEFUNCTION COORDINATES LATEST
MAXSTEP
  100
STORE
  10
EMASS
  800.0
TEMPERATURE
  10.0
MOVIE
STRUCTURE BONDS ANGLES
&END
```

```
&SYSTEM
SYMMETRY
  0
POISSON SOLVER TUCKERMAN
CELL ABSOLUTE DEGREE
  32.0 32.0 32.0 90.0 90.0 90.0
CUTOFF
  70.
CHARGE
  -2.0
&END
```

```
&DFT
FUNCTIONAL BLYP
&END
```

```
&CPMD
MOLECULAR DYNAMICS
RESTART WAVEFUNCTION COORDINATES LATEST
MAXSTEP
 100
STORE
 10
EMASS
 800.0
TEMPERATURE
 10.0
MOVIE
STRUCTURE BONDS ANGLES
&END
```

```
&SYSTEM
SYMMETRY
 0
POISSON SOLVER TUCKERMAN
CELL ABSOLUTE DEGREE
 32.0 32.0 32.0 90.0 90.0 90.0
CUTOFF
 70.
CHARGE
 -2.0
&END
```

```
&DFT
FUNCTIONAL BLYP
&END
```

SYMMETRY

0

POISSON SOLVER TUCKERMAN

CELL ABSOLUTE DEGREE

32.0 32.0 32.0 90.0 90.0 90.0

CUTOFF

70.

CHARGE

-2.0

&END

&DFT

**FUNCTIONAL BLYP**

&END

**&ATOMS**

**\*S-q6**

**GOEDECKER**

0 0 0

1

2.108937 -1.441863 -1.341704

**\*O\_MT\_BLYP**

**KLEINMAN-BYLANDER**

**LMAX=P LOC=P**

17

0.000000 0.000000 0.000000

4.512669 0.120945 -1.283126

1.441863 -2.114603 -3.998664

2.570030 -3.953307 0.132280

-2.619161 -2.553024 3.970312

5.867605 -7.723315 -5.092820

3.061359 -.893838 3.688743

-.166295 -8.443304 -3.548911

7.819691 -5.767457 .211645

# SUMMARY

**WHEW!!! So many details!!!**

What have we accomplished?

You have now seen the way that coordinates are handled and how FF are integrated to describe a molecular (atomic) system. **This is a general method!**

## Break

LAB 0 --> Software installation & intro to SASSIE-web

## Lunch

LAB I --> Overview of VMD

## Break

LAB II --> YOU generate systems (PDB->PSF-> NAMD)