



## **Lunch and Learn Workshop:**

# **Small Molecule Geometry Optimizations and CCS Calculations**

Michael D. Bartberger  
Molecular Engineering  
Therapeutic Discovery  
Amgen Inc.

michael.bartberger@amgen.com

Carlos Larriba-Andaluz  
School of Engineering and Technology  
IUPUI

clarriba@iupui.edu

# Goal of this workshop

**To facilitate the basic use of CCS-prediction algorithms by non-computational specialists (i.e. mass spectrometry experimentalists) in a practical fashion, optionally at no cost**

- I. Describe a few of the (numerous) program systems / methods available to input a molecular structure of interest and perform geometry optimizations and atomic charge determinations on a QM energy surface
- II. Provide a method to easily convert the output of (I) to input suitable for the popular MOBCAL and IMoS CCS prediction software

*What this workshop is not:*

- *Advanced (though Carlos will present background on CCS methodology)*
- *Comprehensive*
- *An endorsement of one software package over another*
- *An endorsement of one CCS method over another (e.g. PA vs. TM)*
- *“The only way to do it”*

# Example used in this workshop

analytical  
chemistry

Article

pubs.acs.org/ac

## Structural Resolution of 4-Substituted Proline Diastereomers with Ion Mobility Spectrometry via Alkali Metal Ion Cationization

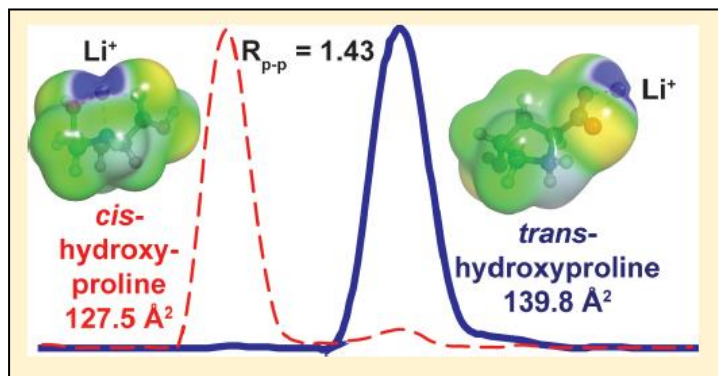
Tawnya G. Flick,<sup>\*,†</sup> Iain D. G. Campuzano,<sup>‡</sup> and Michael D. Bartberger<sup>‡</sup>

<sup>†</sup>Department of Oral Attribute Sciences, Amgen, Inc., Thousand Oaks, California 91320, United States

<sup>‡</sup>Department of Molecular Structure and Characterization, Amgen, Inc., Thousand Oaks, California 91320, United States

*Analyt. Chem.* **2015**, 87, 3300.

(B3LYP/6-31++G(d,p))



charge carrier	<i>cis</i> -hydroxyproline		<i>trans</i> -hydroxyproline	
	exp.	theor.	exp.	theor.
H	126.5	$120.7 \pm 1.4$	125.9	$120.3 \pm 1.5$
Li	127.5	$120.0 \pm 1.3$	139.8	$128.0 \pm 1.4$
Na	129.5	$122.8 \pm 1.0$	137.3	$129.0 \pm 1.6$
K	130.6	$124.6 \pm 1.1$	134.1	$129.6 \pm 1.5$
Cs	133.4	$129.4 \pm 1.4$	136.0	$129.5 \pm 1.5$

- I. Set up (sketch) lithiated *cis*-hydroxyproline; convert to QM input file using provided template
- II. Set up QM run (rather than wait for result, we will utilize pregenerated results in next step)
- III. Convert output into a format suitable for a MOBCAL and IMoS TM CCS evaluation
- IV. Launch CCS calculations ; compare (pregenerated) results to published value

# Example used in this workshop

analytical  
chemistry

Article

pubs.acs.org/ac

## Structural Resolution of 4-Substituted Proline Diastereomers with Ion Mobility Spectrometry via Alkali Metal Ion Cationization

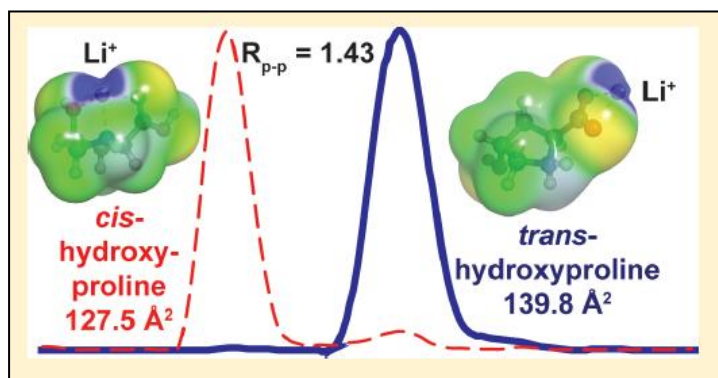
Tawnya G. Flick,<sup>\*,†</sup> Iain D. G. Campuzano,<sup>‡</sup> and Michael D. Bartberger<sup>‡</sup>

<sup>†</sup>Department of Oral Attribute Sciences, Amgen, Inc., Thousand Oaks, California 91320, United States

<sup>‡</sup>Department of Molecular Structure and Characterization, Amgen, Inc., Thousand Oaks, California 91320, United States

*Analyt. Chem.* **2015**, 87, 3300.

(B3LYP/6-31++G(d,p))



charge carrier	cis-hydroxyproline		trans-hydroxyproline	
	exp.	theor.	exp.	theor.
H	126.5	120.7 ± 1.4	125.9	120.3 ± 1.5
Li	127.5	120.0 ± 1.3	139.8	128.0 ± 1.4
Na	129.5	122.8 ± 1.0	137.3	129.0 ± 1.6
K	130.6	124.6 ± 1.1	134.1	129.6 ± 1.5
Cs	133.4	129.4 ± 1.4	136.0	129.5 ± 1.5

- I. Set up (sketch) lithiated cis-hydroxyproline; convert to QM input file using provided template
- II. Set up QM run (rather than wait for result, we will utilize pregenerated results in next step)
- III. Convert output into a format suitable for a MOBCAL and IMoS TM CCS evaluation
- IV. Launch CCS calculations ; compare (pregenerated) results to published value

# Software packages used in this workshop

I. Avogadro: Cross-platform molecular editor (sketcher) with interactive (MM-based) geometry optimizer for rough cleanup of sketched structure and generation of input for Gaussian, GAMESS (and other) electronic structure programs. Free of charge. (<https://avogadro.cc/>)

II. Gaussian: Extremely feature-rich, fast, robust electronic structure package for all manner of QM calculations. Commercial software. (<http://www.gaussian.com>)

III. Firefly: (A. Granovsky, MSU) Originally based on GAMESS QM package (Gordon et al., Iowa State). Current version a complete rewrite of integral routines in a manner optimized for Intel CPU architecture. Fast SCF / DFT / MPn / CASSCF, though currently only numerical DFT frequencies (GAMESS has analytical DFT frequencies but code overall slower). Runs on Windows / Mac / Linux. (<http://classic.chem.msu.su>) or (<http://www.msg.ameslab.gov/gamess>)

IV. logtomfj.py: (A.B. Birkholz, Amgen) Python 2 utility to convert Gaussian, Firefly or GAMESS output containing geometric information and ESP-derived atomic charges to MOBCAL input (also readable by IMoS). (Also some small scripts / wrappers written by MDB).

V. OpenBabel: File format conversion utility. Free of charge. ([http://openbabel.org/wiki/Main\\_Page](http://openbabel.org/wiki/Main_Page))

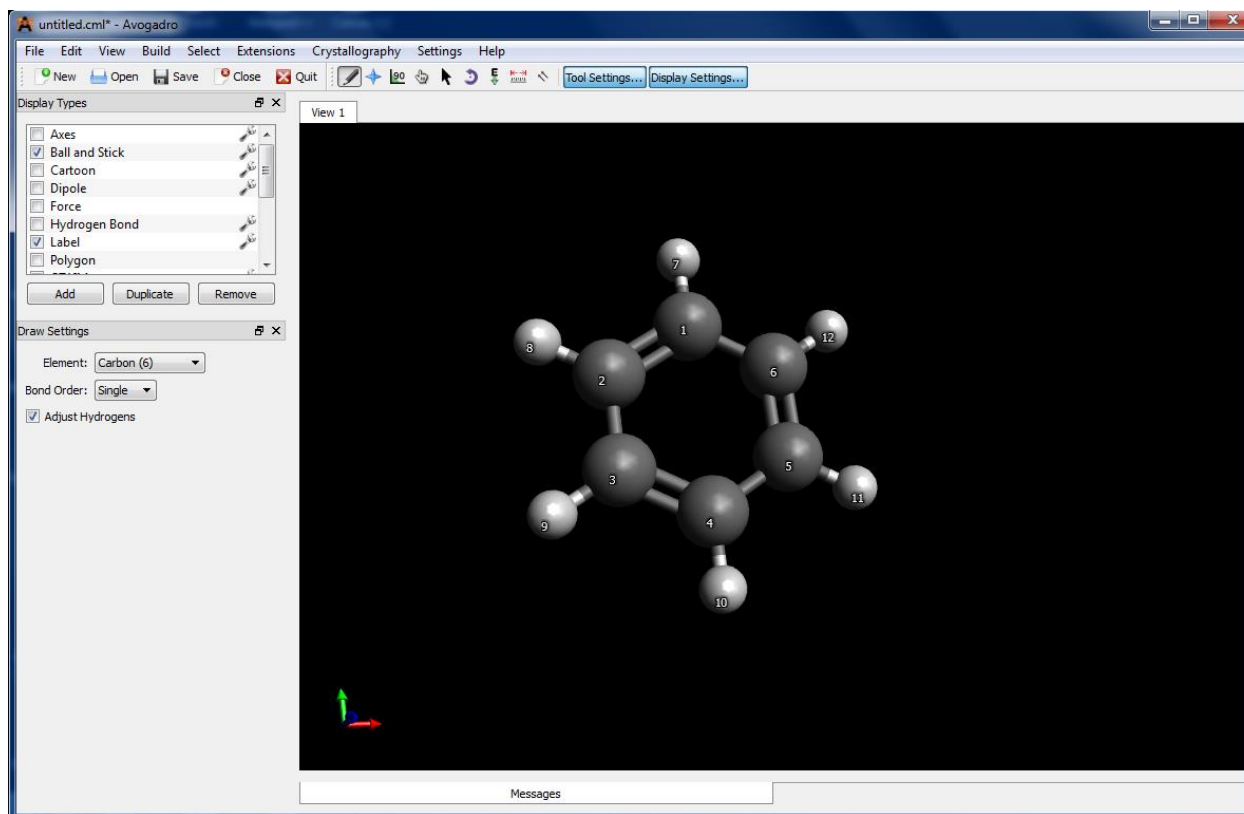
VI. MOBCAL and IMoS: MOBCAL (M.F. Jarrold; <https://www.indiana.edu/~nano/software>) with atomic LJ parameters as modified by Campuzano. IMoS (C. Larriba-Andaluz; <http://www.imospedia.com/>) as discussed by Carlos in second half of workshop

*Computing setup used in workshop: Intel-based laptop, Linux OS (Ubuntu 16.04 LTS but any reasonable modern distro with Fortran compiler (for MOBCAL) and Python 2 and (t)csh interpreters (for scripts) will do). (Suitably equipped Windows computer should work as well).*

# Using Avogadro

Create (skeleton) input deck  
for some QM programs

Draw, Rotate, Minimize

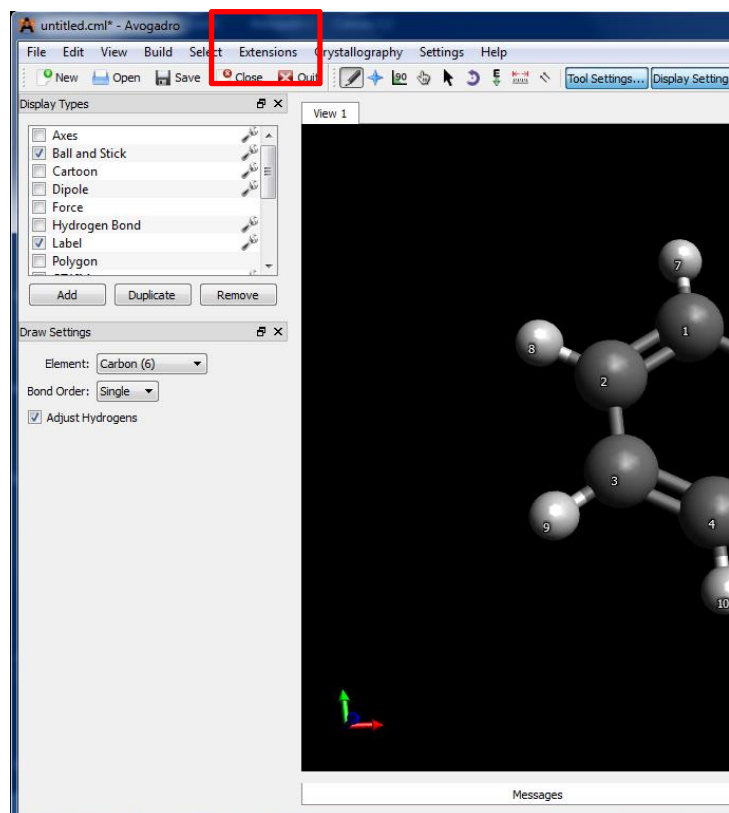


Sketch /  
Minimize  
options

# Using Avogadro

Create (skeleton) input deck  
for some QM programs

Draw, Rotate, Minimize



The Gaussian Input dialog box is shown with the following settings:

- Title: Title
- Calculation: Geometry Optimization
- Theory: B3LYP
- Charge: 0
- Output: Standard
- Format: Cartesian
- Processors: 1
- Basis: 6-31G(d)
- Multiplicity: 1
- Checkpoint: ☐

Buttons: Hide Preview, Reset, Use Form, Generate..., Close

Preview of the input file:

```
#n B3LYP/6-31G(d) Opt

Title

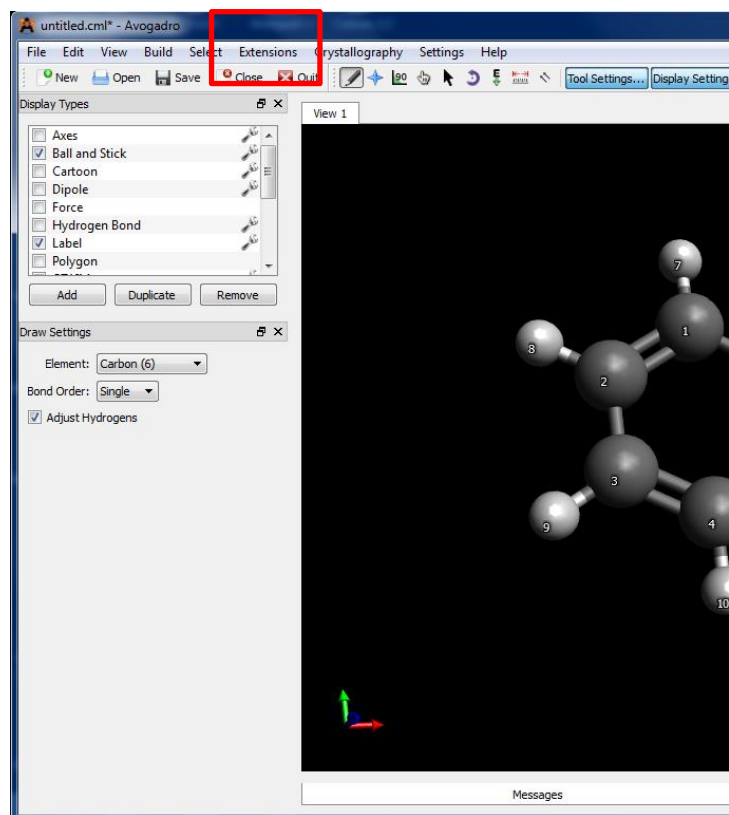
0 1
C   -1.58557    2.06660    0.00000
C   -2.76382    1.31254    0.00000
C   -2.69927   -0.08604   -0.00000
C   -1.45578   -0.72955   -0.00000
C   -0.27610    0.02409   -0.00000
C   -0.34106    1.42419   -0.00000
H   -1.63967    3.14896    0.00000
H   -3.72595    1.81021    0.00000
H   -3.61289   -0.66740   -0.00000
```

Sketch /  
Minimize  
options

# Using Avogadro

Create (skeleton) input deck  
for some QM programs

Draw, Rotate, Minimize



```
#n B3LYP/6-31G(d) Opt

Title

0 1
C   -1.58557    2.06660    0.00000
C   -2.76382    1.31254    0.00000
C   -2.69927   -0.08604   -0.00000
C   -1.45578   -0.72955   -0.00000
C   -0.27610    0.02409   -0.00000
C   -0.34106    1.42419   -0.00000
H   -1.63967    3.14896    0.00000
H   -3.72595    1.81021    0.00000
H   -3.61289   -0.66740   -0.00000
```

Absolute beginners: use coordinates from confs  
drawn w/ Avogadro + my supplied templates

Advanced users: automated conformational  
sampling via commercial apps plus workflows (e.g. “for” loops) to perform QM on many conformations  
(check out OpenBabel to convert from e.g. multi .MOL2s to many Gaussian .coms)



# Anatomy of Gaussian input file suitable for CCS studies

% ("Link0") lines :

Machine (hardware)-level  
Directives

# (Route) lines :

specifies type of calculation  
(always terminate w/ blank line)

Comment line(s) can be as  
long as desired ; terminate  
comments w/ blank line

Total molecular charge (+1 in  
this case) and unpaired  
spin multiplicity (1 in this  
case ; singlet)

A monoradical is a doublet ; so  
e.g. a radical cation would  
read '1 2' )

Geometry specification  
immediately follows

**Don't forget to end with  
a blank line!**

%mem=500MB

%chk=cis\_li\_hydroxyproline\_ready

#b3lyp/6-31++g\*\* opt freq pop(mk,dipole)

cis\_li\_hydroxyproline conf 1

1 1

N	-3.77017	0.23383	-0.18328
C	-3.31535	-0.51534	-1.35390
C	-4.48791	-1.45314	-1.63971
H	-2.37386	-1.06968	-1.15509
H	-3.15423	0.18462	-2.20714
C	-5.64007	-0.42183	-1.51420
O	-4.55464	-2.44241	-0.62820
H	-4.43196	-1.95784	-2.62897
C	-5.17726	0.59919	-0.42636
H	-6.67033	-0.82438	-1.41260
H	-5.61882	0.13663	-2.47873
C	-5.96082	0.54201	0.90029
H	-5.29469	1.62504	-0.85129
O	-5.89957	-0.48169	1.55963
O	-6.70482	1.58843	1.31521
H	-7.20041	1.55046	2.14548
H	-5.42034	-2.92335	-0.68635
Li	-4.78142	-1.17913	0.51378
H	-3.31833	1.17203	-0.07288



# (Most) important parts of a Gaussian output file

*Did my geometry optimization complete?*

Item	Value	Threshold	Converged?
Maximum Force	0.000078	0.000450	YES
RMS Force	0.000009	0.000300	YES
Maximum Displacement	0.001620	0.001800	YES
RMS Displacement	0.000248	0.001200	YES

Predicted change in Energy=-7.799983D-08

Optimization completed.

**-- Stationary point found.**

*Am I indeed at a minimum? (Not an "accidental" TS)*

Harmonic frequencies (cm\*\*<sup>-1</sup>), IR intensities (KM/Mole), Raman scattering activities (A\*\*4/AMU), depolarization ratios for plane and unpolarized incident light, reduced masses (AMU), force constants (mDyne/A), and normal coordinates:

*Look at the first few harmonic frequencies (all positive for min.)*

	1	2	3
	A	A	A
Frequencies --	79.8678	145.0224	216.7994
Red. masses --	5.6770	5.1449	3.5657
Frc consts --	0.0213	0.0638	0.0987
IR Inten --	0.1933	6.6097	0.1218

Charges from ESP fit, RMS= 0.00181 RRMS= 0.01412:

*#pop(mk,dipole)*

*requests atomic charges*

*based on a fit to the total*

*electrostatic potential & constrained*

*to reproduce dipole moment – I like Merz/Kollman*

ESP charges:

	1
1 N	-0.748752
2 C	0.069471
3 C	0.201446

.

.

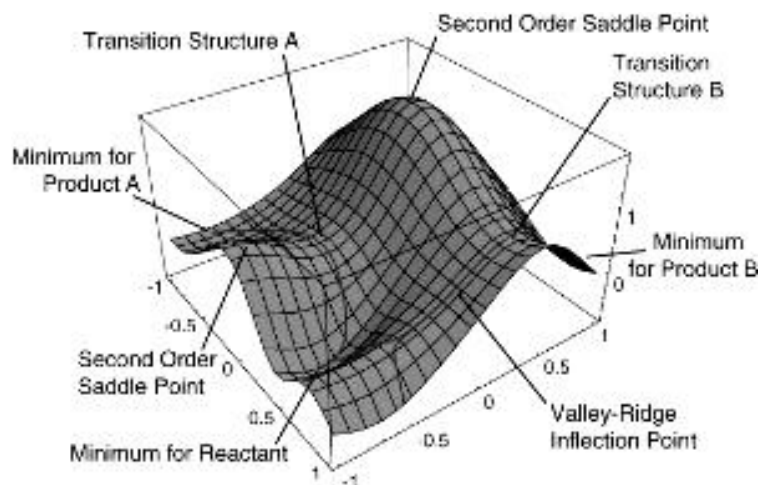
**Strongly recommended over the (default) Mulliken charges in most programs!**

18	Li	0.828794
19	H	0.366854

# (Most) important parts of a Gaussian output file

Vibrational frequency analysis (“Hessian”: 2<sup>nd</sup> – derivative matrix) characterizes stationary points on PES as minima or TS and provides entropy (for DG, Dgrel determination) – perhaps can skip this step / keyword in QM input file if you are only interested in a single conformation

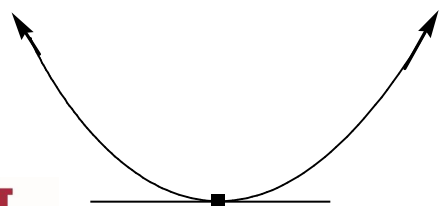
can be compared  
to experimental  
vibrational (IR)  
spectra



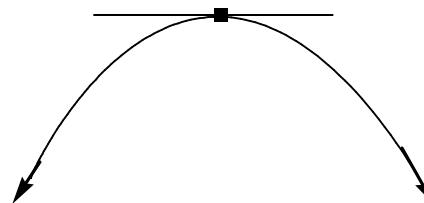
All positive force constants (vibrational frequencies) for minima (as it should be!)

Exactly one “negative” (actually imaginary) frequency for TS, corresponding to motion along reaction coordinate

III. Akin to concavity (2<sup>nd</sup> derivative) test of a parabolic function



“concave up”  
(minima)  
1<sup>st</sup> deriv: zero  
2<sup>nd</sup> deriv: **positive**



“concave down”  
(minima)  
1<sup>st</sup> deriv: zero  
2<sup>nd</sup> deriv: **negative**

# Anatomy of a functionally equivalent Firefly / GAMESS input

! is used for  
comments

- various  
keywords belong  
in their respective  
“groups” (\$)

- read Firefly /  
GAMESS docs  
to understand  
keywords

- I prefer Firefly  
over GAMESS  
(when not using  
Gaussian!) but  
up to you  
(change B3LYP1  
to B3LYP5 if  
using GAMESS)

- Symmetry use  
is hard for  
newbies in FF /  
GAMESS (but  
easy in  
Gaussian!) -  
leave as C1 here  
if using  
FF/GAMESS

```
$CONTRL SCFTYP=RFH DFTTYP=B3LYP1 RUNTYP=OPTIMIZE $END
$CONTRL MAXIT=200 NZVAR=1 UNITS=ANGS NPRINT=-5 $END
!
! Remember to properly adjust charge and multiplicity!
!
$CONTRL ICHARG=1 MULT=1 $END
$ZMAT DLC=.TRUE. AUTO=.TRUE. $END
$SYSTEM TIMLIM=999999999 MEMORY=500000000 $END
$BASIS GBASIS=N31 NGAUSS=6 NDFUNC=1 NPFUNC=1 DIFFS=.TRUE. DIFFSP=.TRUE. $END
$STATPT NSTEP=999 $END
!
! Delete or comment the following line if frequencies not desired
!
$STATPT HSEND=.TRUE. $END
$SCF DIRSCF=.TRUE. $END
$ELDENS IEDEN=1 $END
$ELPOT IEPOT=1 WHERE=PDC $END
$PDC PTSEL=CONNOLLY CONSTR=DIPOLE $END
$ELMOM IEMOM=2 WHERE=POINTS $END
$POINTS
ANGS 1
0.0 0.0 0.0
$END
$DATA
cis-hydroxyproline lithiated conf 1
C1
N      7.0      -3.77017      0.23383      -0.18328
C      6.0      -3.31535     -0.51534     -1.35390
.
.
$END
```

# (Most) important parts of a Firefly / GAMESS output file

*Did my geometry  
optimization complete?*

```
MAXIMUM GRADIENT = 0.0000715    RMS GRADIENT = 0.0000321
***** EQUILIBRIUM GEOMETRY LOCATED *****
```

*Am I indeed at a  
minimum? (Not  
an “accidental”  
TS)*

```
FREQUENCIES IN CM**-1, IR INTENSITIES IN DEBYE**2/AMU-ANGSTROM**2
```

*Look at the first  
few harmonic  
frequencies (all  
positive for min.)*

```

              1
FREQUENCY:    53.77
REDUCED MASS: 3.02486
IR INTENSITY: 0.33668
```

*atomic charges  
based on a fit to the total  
electrostatic potential & constrained  
to reproduce dipole moment –*

ELECTROSTATIC POTENTIAL

***Strongly recommended over the  
(default) Mulliken charges in most programs!***

ATOM	CHARGE	E.S.D.
N	-0.8396	0.0201
C	0.0040	0.0299
.		
.		
LI	0.8580	0.0059
H	0.4208	0.0078

note that charges do differ between programs  
(differences in implementation, grids, etc). – stick with  
one program (but note nearly identical final CCSs)



IUPUI

AMGEN

# Anatomy of MOBCAL input (.mfj) & control (mobcal.in) files

Description

Number of coordinate sets to be read from input stream (for conf. averaging if desired)

Number of atoms

Units

“calc” / “equal” / “none” –  
calc = charge distribution will be read in and determined from supplied atomic charges in input stream

Atom radius scale factor

X / Y / Z / atomic # / charge  
(if “calc” chosen)

To run:

1) “log2mfj.py” – will build this for you from a “good” Gaussian / FF / GAMESS output containing ESP charges

Input=cis\_li\_hydroxyproline\_ready

1

19

ang

calc

1.0000

0.48650	-0.36658	-1.24918	14	-0.748752
1.87127	-0.61232	-0.77145	12	0.069471
1.73529	-0.13479	0.66687	12	0.201446
2.59249	-0.03128	-1.35065	1	0.096494
2.15979	-1.67117	-0.79838	1	0.080306
0.43789	-0.82686	1.11788	12	-0.274480
1.53880	1.30726	0.57469	16	-0.749622
2.59984	-0.35470	1.29597	1	0.097248
-0.43286	-0.86477	-0.18961	12	0.292049
-0.04922	-0.30856	1.94750	1	0.127036
0.65357	-1.84625	1.44783	1	0.117665
-1.62239	0.08015	-0.07809	12	0.601711
-0.80095	-1.87781	-0.37822	1	0.079321
-1.55592	1.27415	-0.36157	16	-0.596599
-2.71232	-0.49896	0.39640	16	-0.593546
-3.42788	0.15762	0.49119	1	0.515928
1.73285	1.73856	1.41775	1	0.488675
0.24542	1.68395	-0.84293	7	0.828794
0.29803	-0.80018	-2.14967	1	0.366854

2) “runmobcal\_n2.exe myjob.mfj” – builds mobcal.in and runs jobs in a temporary directory under working dir (look for .tmp.\* dir - results will be copied and temp dir deleted when job is done (hours for N<sub>2</sub>!))

“mobcal.in” : small control file expected by code containing name of input, name of output, and random seed

```
cis_li_ohpro_gaus
cis_li_ohpro_gaus.mob
28
```

# MOBCAL's trajectory method (TM) considered most suitable (at least for small / druglike molecules)

look at very end of MOBCAL  
output file

*From Firefly's  
B3LYP / 6-31++G(d,p)  
optimized geometry and  
ESP ("CONNOLLY" ~=  
"MK") atomic charges*

*From Gaussian's  
B3LYP / 6-31++G(d,p)  
Optimized geometry and  
ESP ("MK" ~=  
"CONNOLLY") atomic  
charges*

```
[mdb@laptop]$ tail *.mob  
==> cis_li_ohpro_ff.mob <==
```

```
number of complete cycles (itn) =    10  
number of velocity points (inp) =    40  
number of random points (imp) =   1000  
total number of points = 400000
```

```
inverse average (second order) TM mobility = 5.4130E+03  
average TM cross section = 1.1967E+02  
standard deviation (percent) = 1.0376E+00  
number of failed trajectories =    0
```

```
==> cis_li_ohpro_gaus.mob <==
```

```
number of complete cycles (itn) =    10  
number of velocity points (inp) =    40  
number of random points (imp) =   1000  
total number of points = 400000
```

```
inverse average (second order) TM mobility = 5.4148E+03  
average TM cross section = 1.1971E+02  
standard deviation (percent) = 5.3229E-01  
number of failed trajectories =    0
```

# Need speed? Smaller basis sets for are OK for simple systems (if geometries reasonable)

look at very end of MOBCAL output file

*From Firefly's  
B3LYP / 6-31G(d)  
optimized geometry and  
ESP ("CONNOLLY" ~=  
"MK") atomic charges:*

```
[mdb@laptop]$ tail *.mob  
==> cis_li_ohpro_ff.mob <==
```

```
number of complete cycles (itn) =    10  
number of velocity points (inp) =    40  
number of random points (imp) =   1000  
total number of points = 400000
```

```
inverse average (second order) TM mobility = 5.4048E+03  
average TM cross section = 1.1949E+02  
standard deviation (percent) = 1.0213E+00  
number of failed trajectories =    0
```

```
==> cis_li_ohpro_ready.mob <==
```

```
number of complete cycles (itn) =    10  
number of velocity points (inp) =    40  
number of random points (imp) =   1000  
total number of points = 400000
```

```
inverse average (second order) TM mobility = 5.4177E+03  
average TM cross section = 1.1977E+02  
standard deviation (percent) = 1.1270E+00  
number of failed trajectories =    0
```

*So....what basis set for QM part?*

*e.g. 6-31G(d) for cations ;  
6-31+G(d) for anions is an OK  
start (compare final TM CCS  
here (using 6-31G(d)) to those  
on previous slide (using  
6-31++G(d,p))*

*caveat – large basis and/or  
dispersion might be desirable /  
necessary for good  
geometries for larger  
systems*