A challenging EXAFS analysis problem

Bruce Ravel
Synchrotron Science Group
National Institute of Standards and Technology
&
Beamline for Materials Measurements
National Synchrotron Light Source II

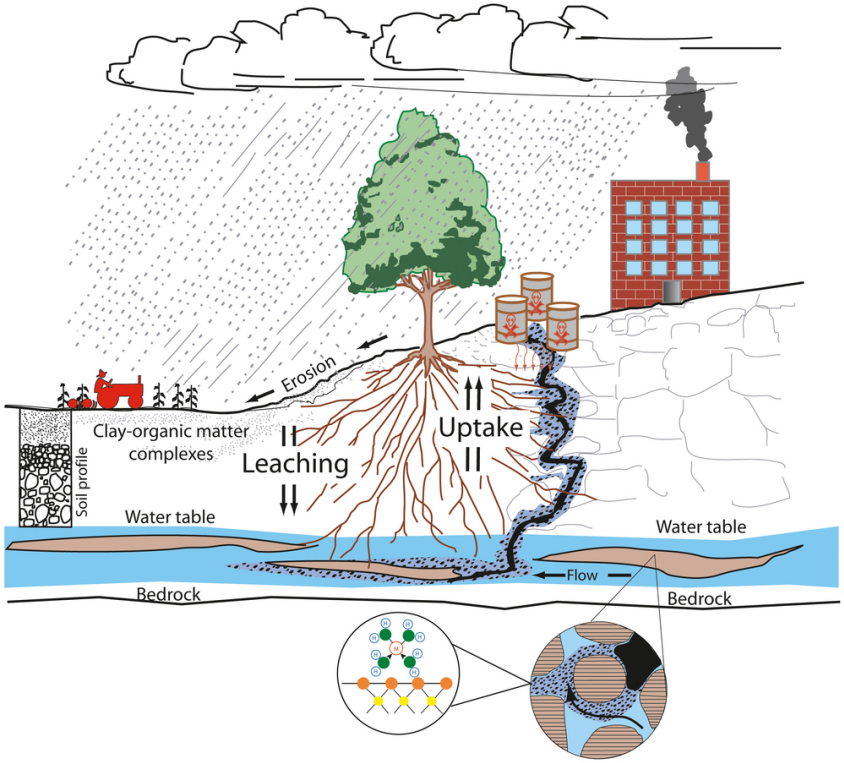
ASEAN Workshop on X-ray Absorption Spectroscopy
Synchrotron Light Research Institute

June 2–4, 2014

Copyright

Transport of metal contaminants in the environment

There are numerous natural and man-made point sources of toxic metals which find their way into water systems used for human and agricultural applications.



The safe use of water requires monitoring and eventual remediation of bioavailable metal species.

image from http://lightsources.org, Credit: Argonne National Laboratory

Real-time, field-ready sensors

Sophisticated laboratory and synchrotron methods exist to detect and speciate water contaminants at very low concentrations. The real-world task of environmental monitoring requires a fast, flexible, sensitive, selective method of detecting contaminants *in the field*.

Fast	Obtain results while still in the field
Flexible	Easy to carry and easy to use in the field
Sensitive	Detect contaminant concentrations below regulated human health hazard levels
Selective	Respond to the target metal but not to other metals

Real-time, field-ready sensors

Sophisticated laboratory and synchrotron methods exist to detect and speciate water contaminants at very low concentrations. The real-world task of environmental monitoring requires a fast, flexible, sensitive, selective method of detecting contaminants *in the field*.

We want Spock's tricorder!

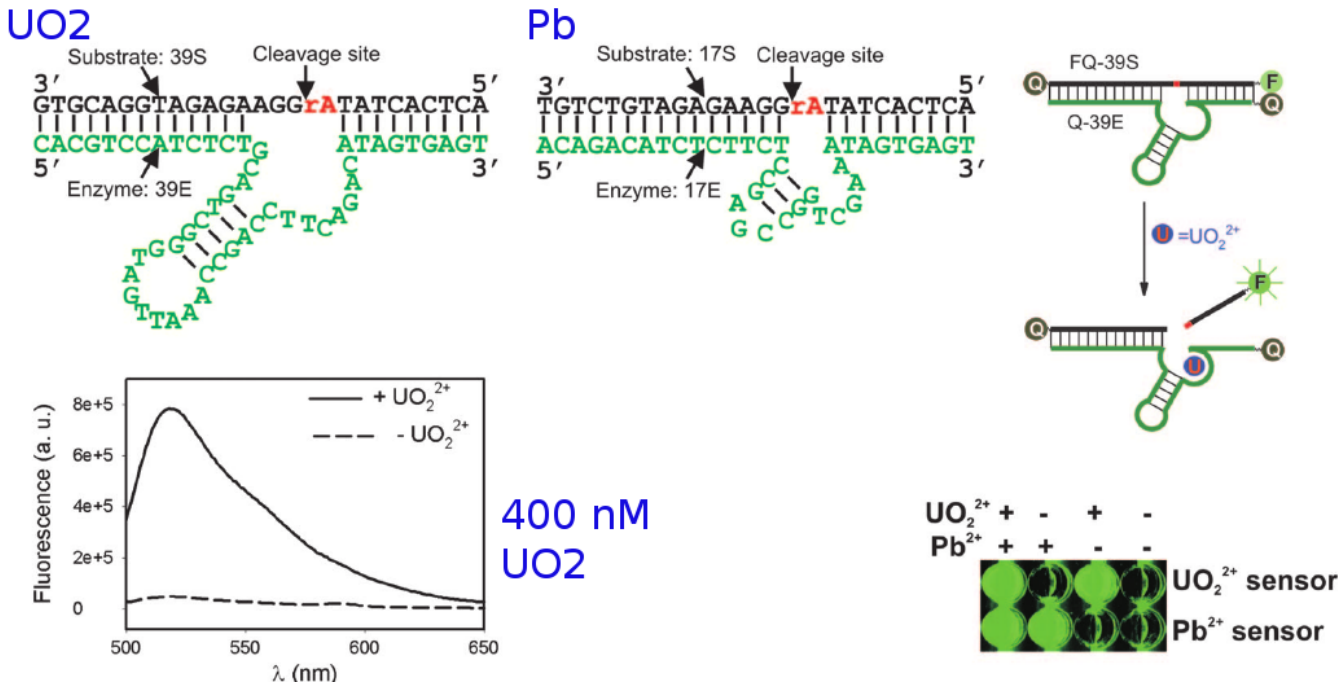
Fast Obtain results while still in the field

Flexible Easy to carry and easy to use in the field

Sensitive Detect contaminant concentrations below regulated human health hazard levels

Selective Respond to the target metal but not to other metals

Catalytic DNA-based sensors

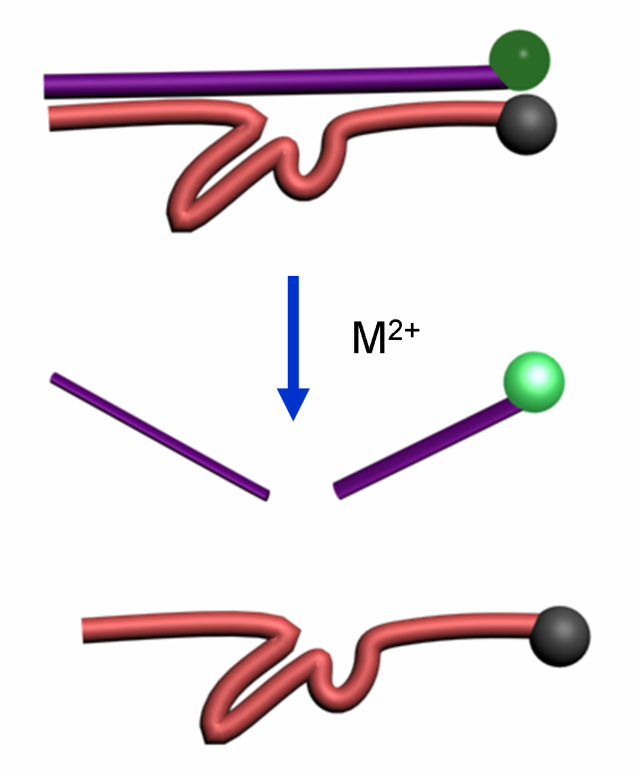


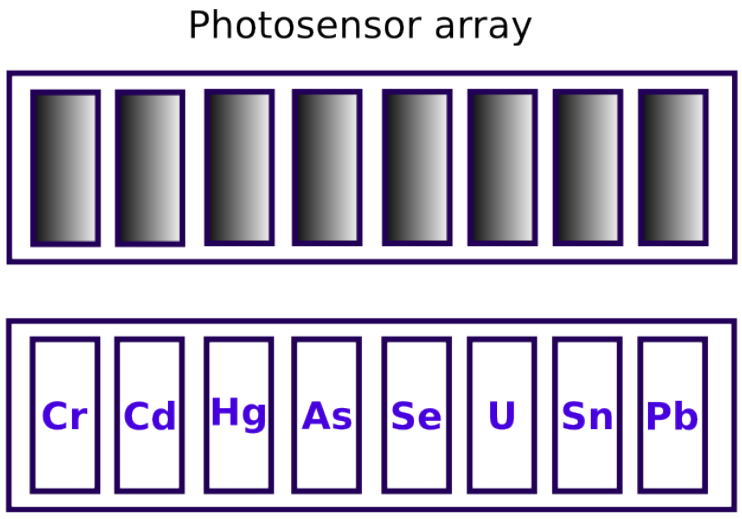
The sensor has a receptor, a cleavage site, and paired fluorophore and quencher.

J. Liu, et al. *A catalytic beacon sensor for uranium with parts-per-trillion sensitivity and millionfold selectivity* PNAS, 104:7 (2007) 2056-2061 DOI: 10.1073/pnas.0607875104

Building a sensor device

These DNA sensors can be incorporated into a hand-held device. Water is dropped onto an array of sensors and read using photodiodes.





DNA-based Hg sensor

U.S. EPA limit on Hg in water is 10 nM (2 ppb)

The DNA-based sensor for Hg has a detection limit of 2.4 nM

Questions:

- How and where does the metal bind?
- Under what conditions does the metal remain bound to the DNA?
- How many binding sites are there on a sensor?
- Do different metals behave differently?
- Can DNAzymes be designed more rationally?
- And, of course, what can XAS tell us about any of these questions (keeping in mind the very local nature of the XAS measurement)?

J. Liu and Y. Lu. *Rational Design of "Turn-On" Allosteric DNAzyme Catalytic Beacons for Aqueous Mercury Ions with Ultrahigh Sensitivity and Selectivity*, Angew. Chemie, **46**:40 (2007) 7587–7590
DOI: 10.1002/anie.200702006

XAS measurements

Solutions:

- 50 mM cacodylic acid as a buffer
- 100 mM $NaClO_4$ to maintain pH=6.10
- glycerol to promote glassification upon freezing

Samples:

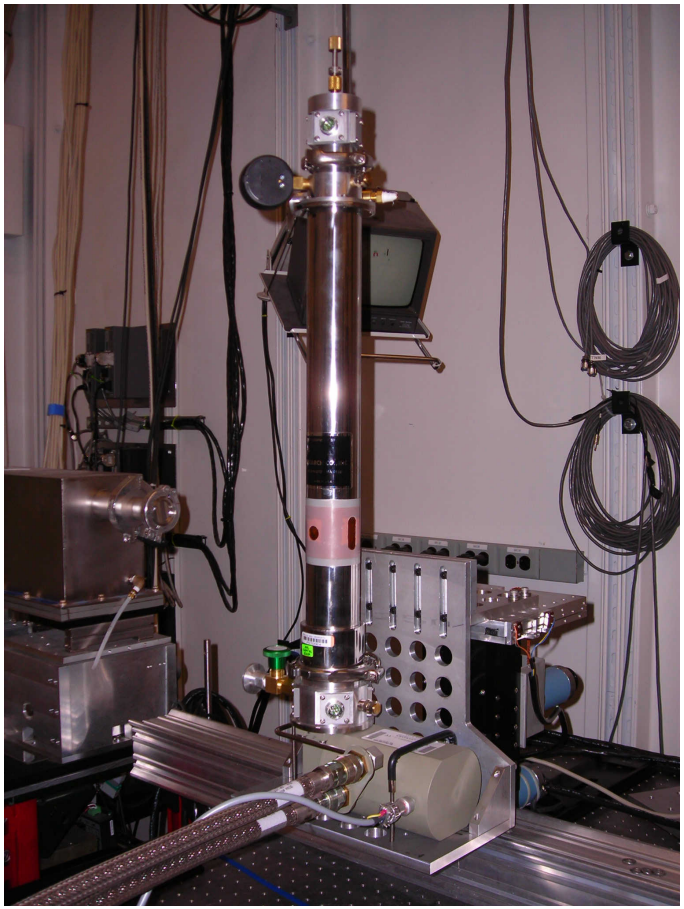
Control 15 mM Hg

Sample 3 mM Hg with 3 mM DNA

Sample with excess Hg 6 mM Hg with 3 mM DNA

Measure EXAFS at 10 K

Cryostat

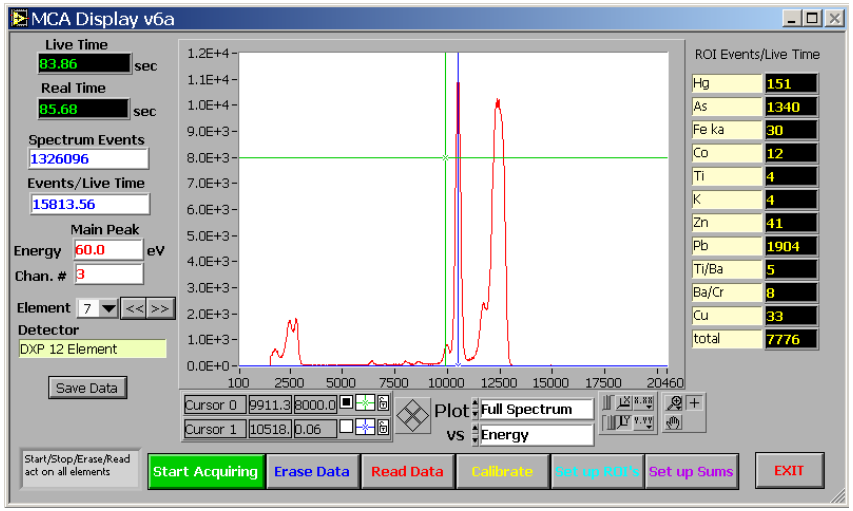


Displex cryostat at APS 20BM.

- He exchange gas
- 10 mm wide opening for beam
- ∼12 mm wide inner shroud
- Fluorescence measured through hole on side with a Ge detector
- At that time, 20BM did *not* have a focusing mirror

Unforced error #1

Here is the fluorescence spectrum:

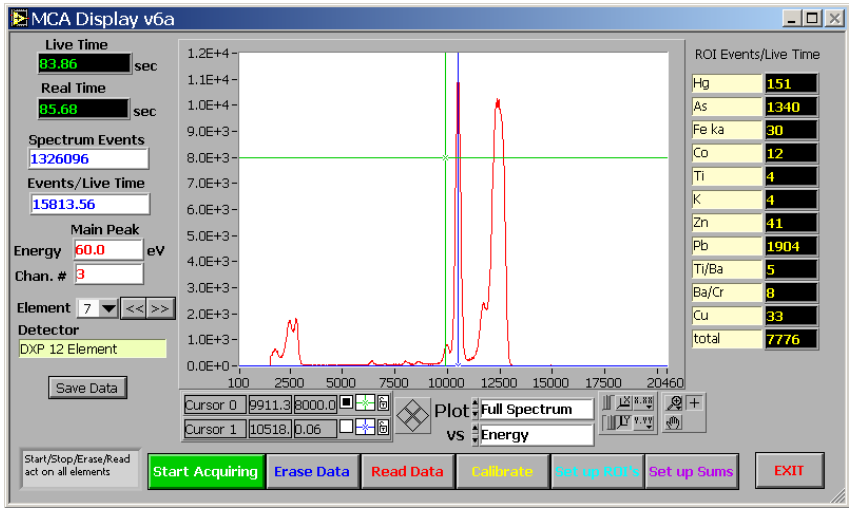


The Hg Lα peak is the tiny thing near the green line.

The neighboring peak is vastly larger!

Unforced error #1

Here is the fluorescence spectrum:



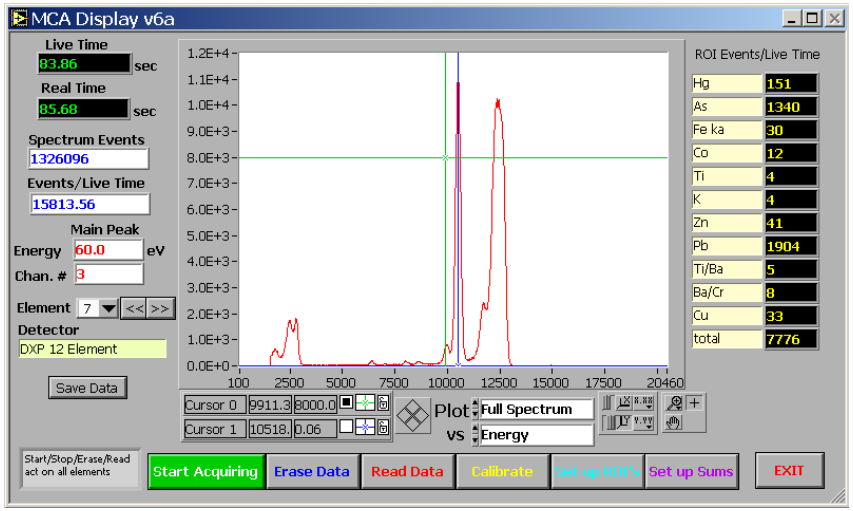
What's cacodylic acid?

The Hg Lα peak is the tiny thing near the green line.

The neighboring peak is vastly larger!

Unforced error #1

Here is the fluorescence spectrum:

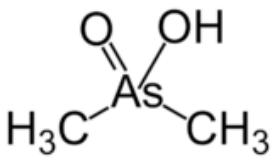


The Hg Lα peak is the tiny thing near the green line.

The neighboring peak is vastly larger!

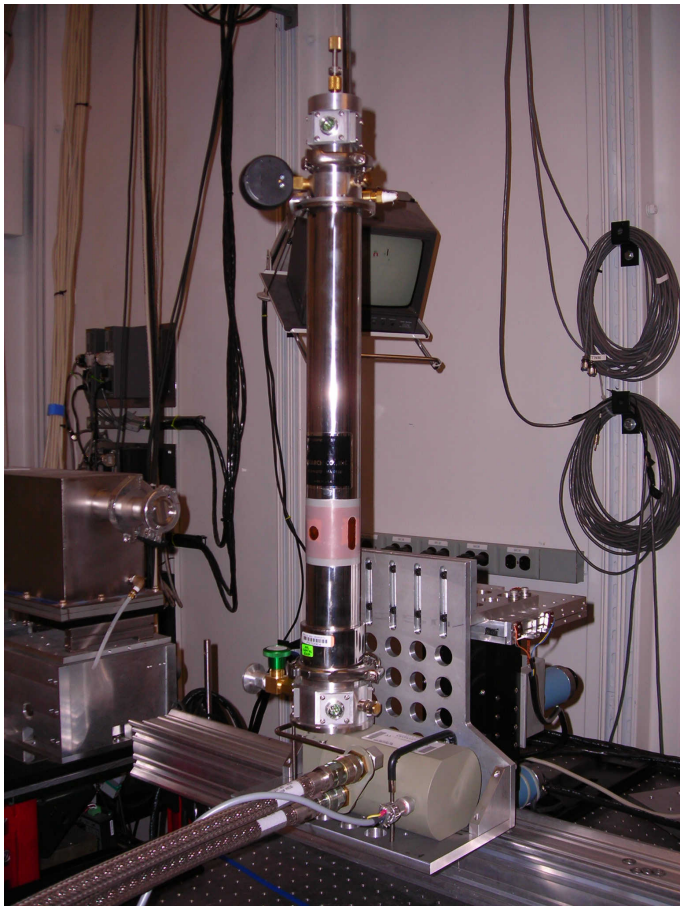
What's cacodylic acid?

Wikipedia tells me that cacodylic acid is:



The big peak is As Kα ($\sim$10.5 keV), our Hg Lα ($\sim$10 keV) peak is on its shoulder.

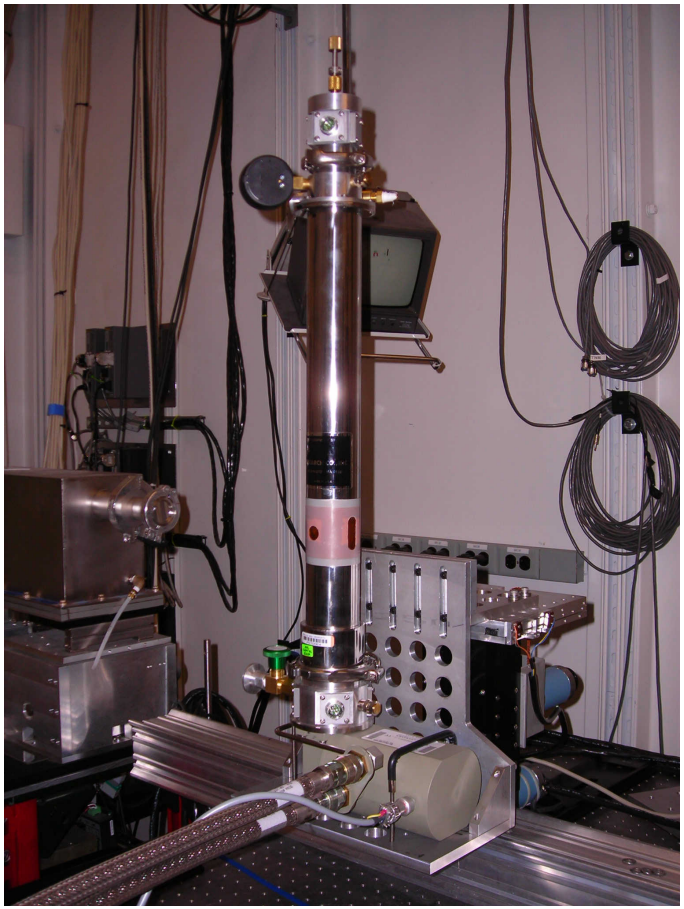
Unforced error #2



The samples were packaged back at the University of Illinois and were about 15 mm by 3 mm.

We had to put the samples in the cryostat upright and slit the beam down to ∼1 mm.

Unforced error #2



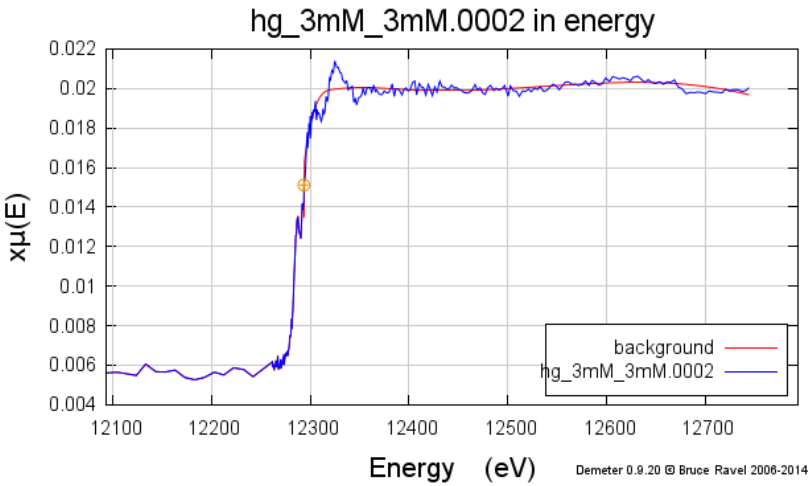
The samples were packaged back at the University of Illinois and were about 15 mm by 3 mm.

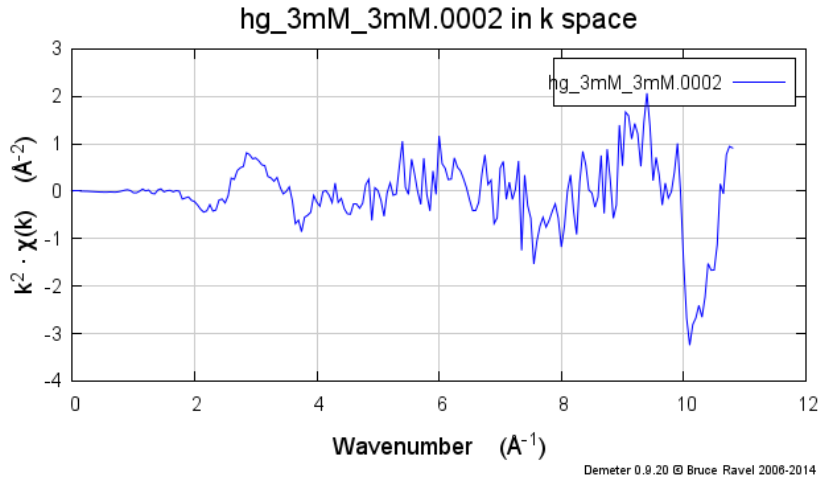
We had to put the samples in the cryostat upright and slit the beam down to ∼1 mm.

Plan ahead!

Forgetting about the details leads to much worse data!

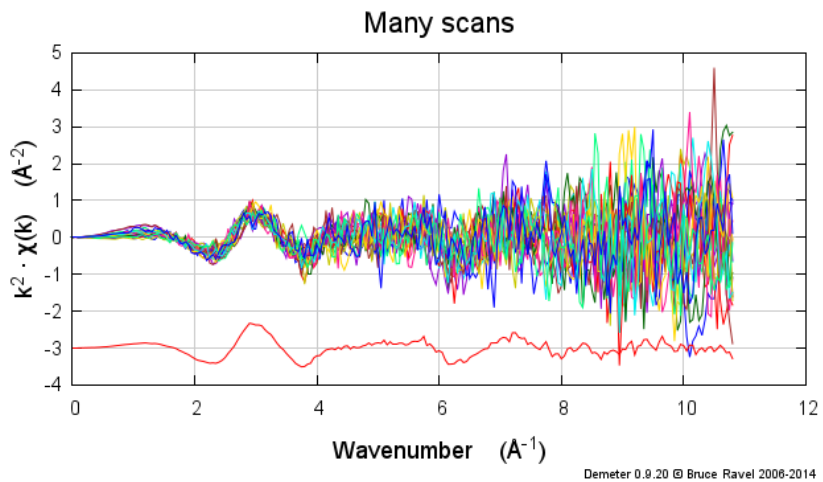
Our main sample



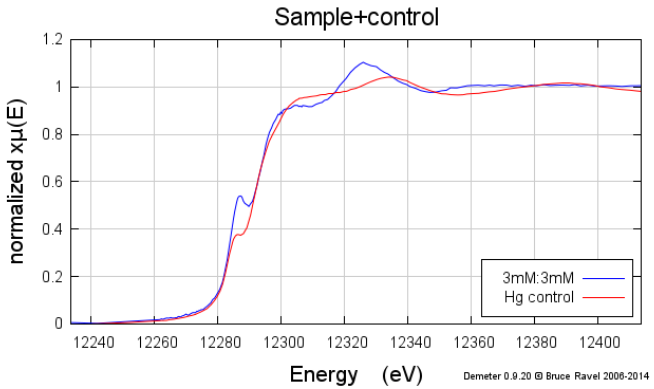


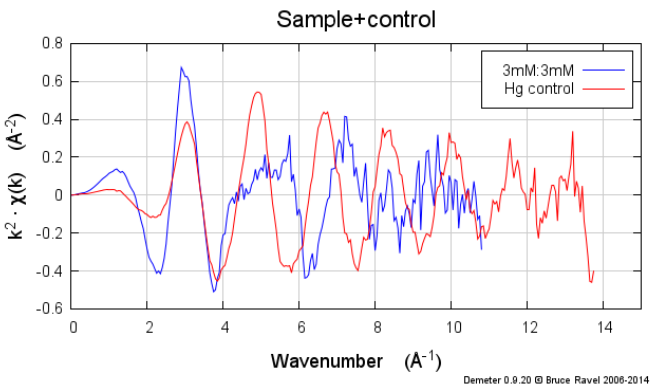
This poor data is due to low concentration, small beam, and large background from the As.

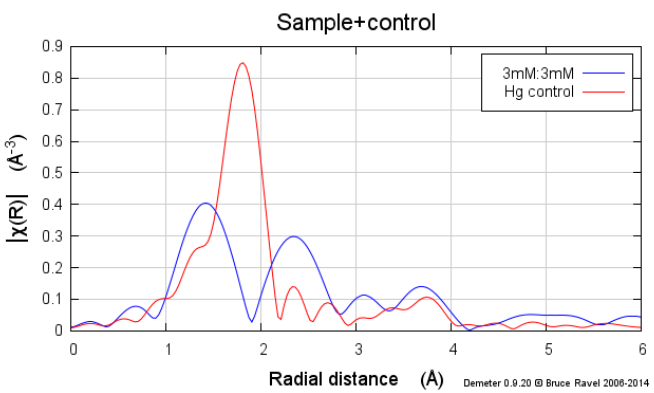
We measured 42 scans, taking about 22 hours.



Sample and control







Chemistry has certainly happened.

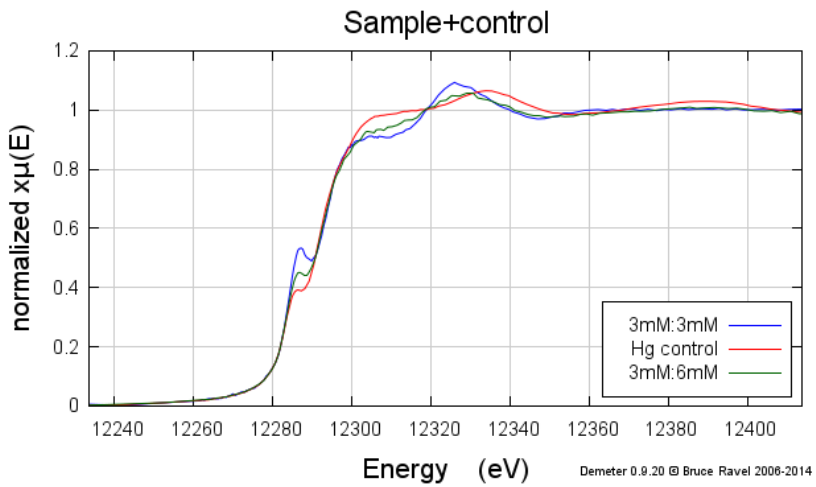
The control is clearly Hg in some kind of aqueous form.

The sample with DNA is clearly different from the control.

First question

Is all Hg taken up by the DNA?

To answer this, we measured a sample with excess Hg.

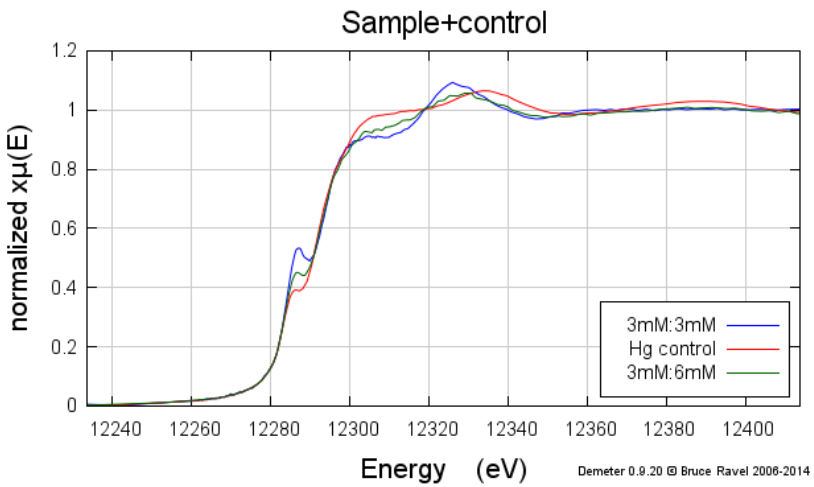


Let's go do some linear combination fitting. (Note the isosbestic points.)

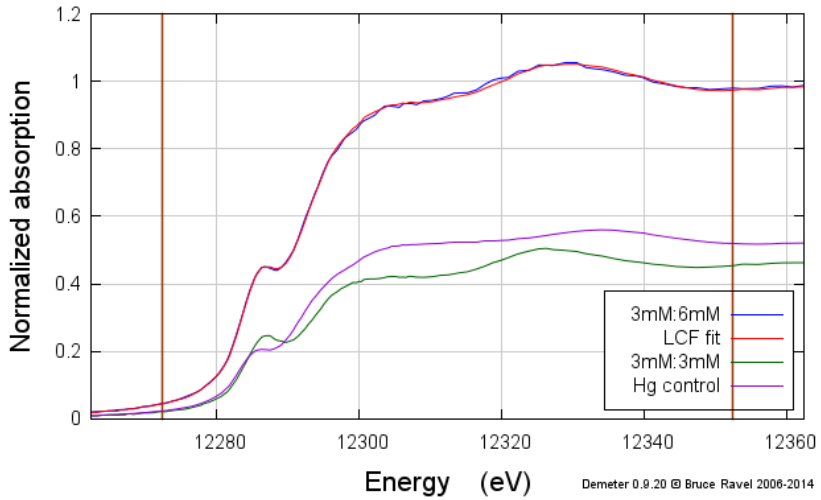
First question

Is all Hg taken up by the DNA?

To answer this, we measured a sample with excess Hg.

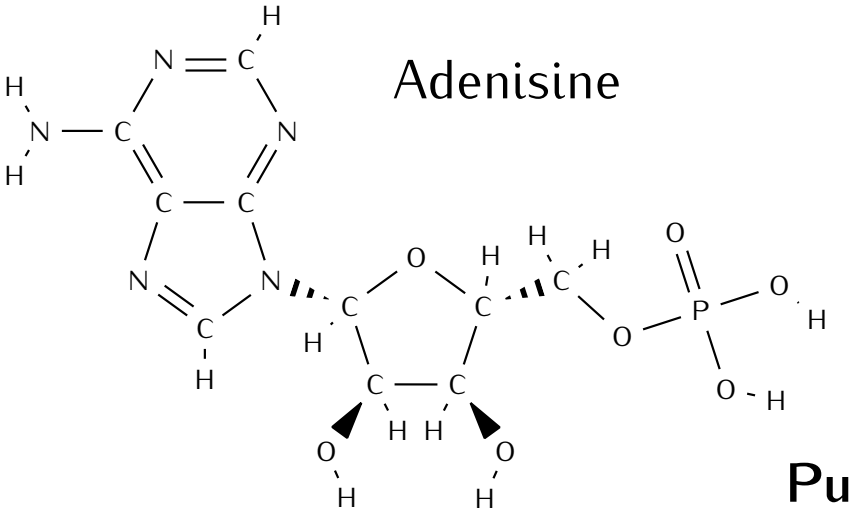


Let's go do some linear combination fitting. (Note the isosbestic points.)



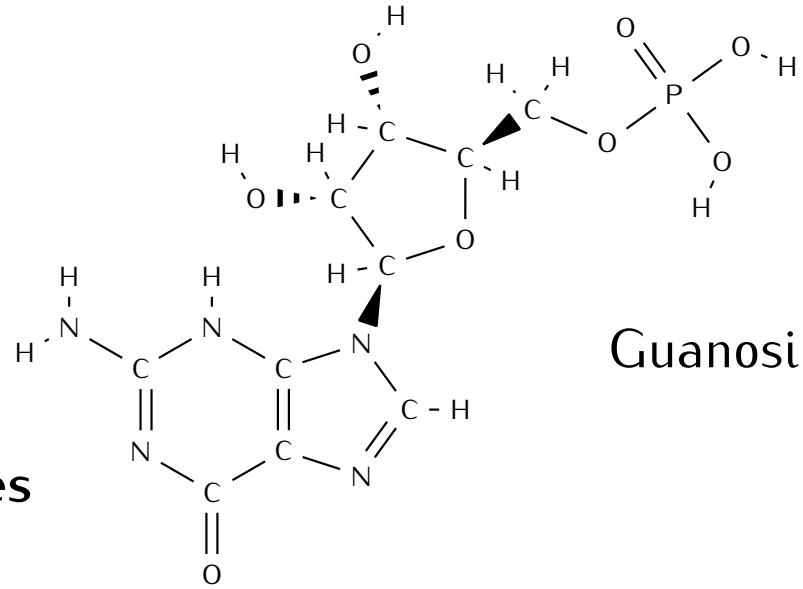
Yes, all the Hg is taken up by the DNA.
47(1)% sample + 53(1)% control

The nucleotides

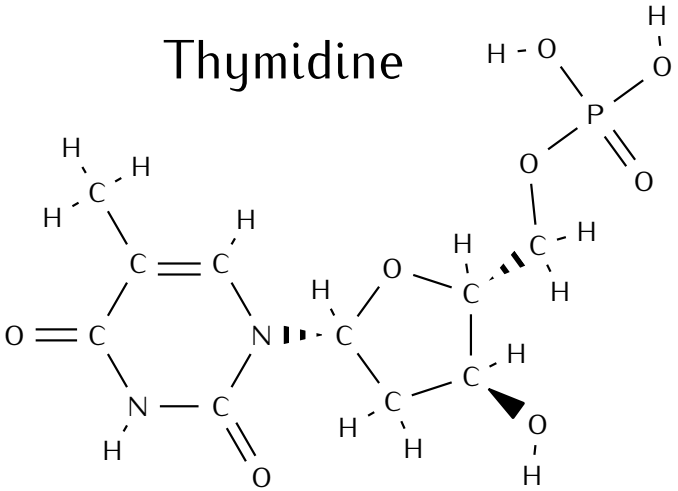


Adenisine

Purines

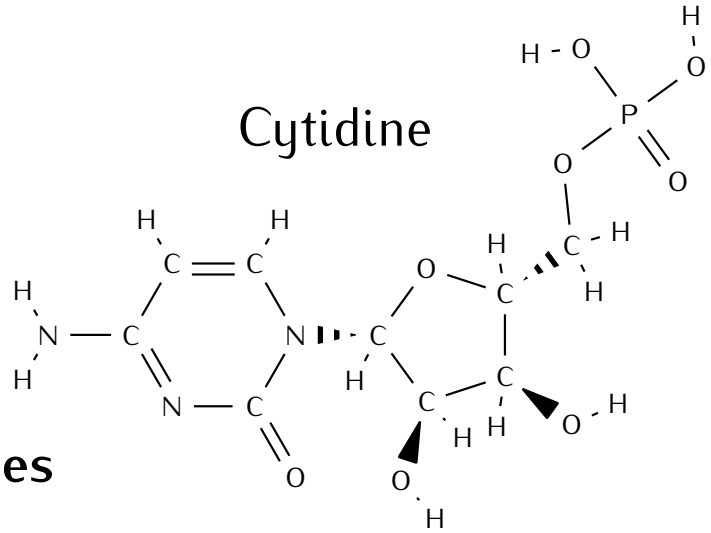


Guanosine



Thymidine

Pyridines



Cytidine

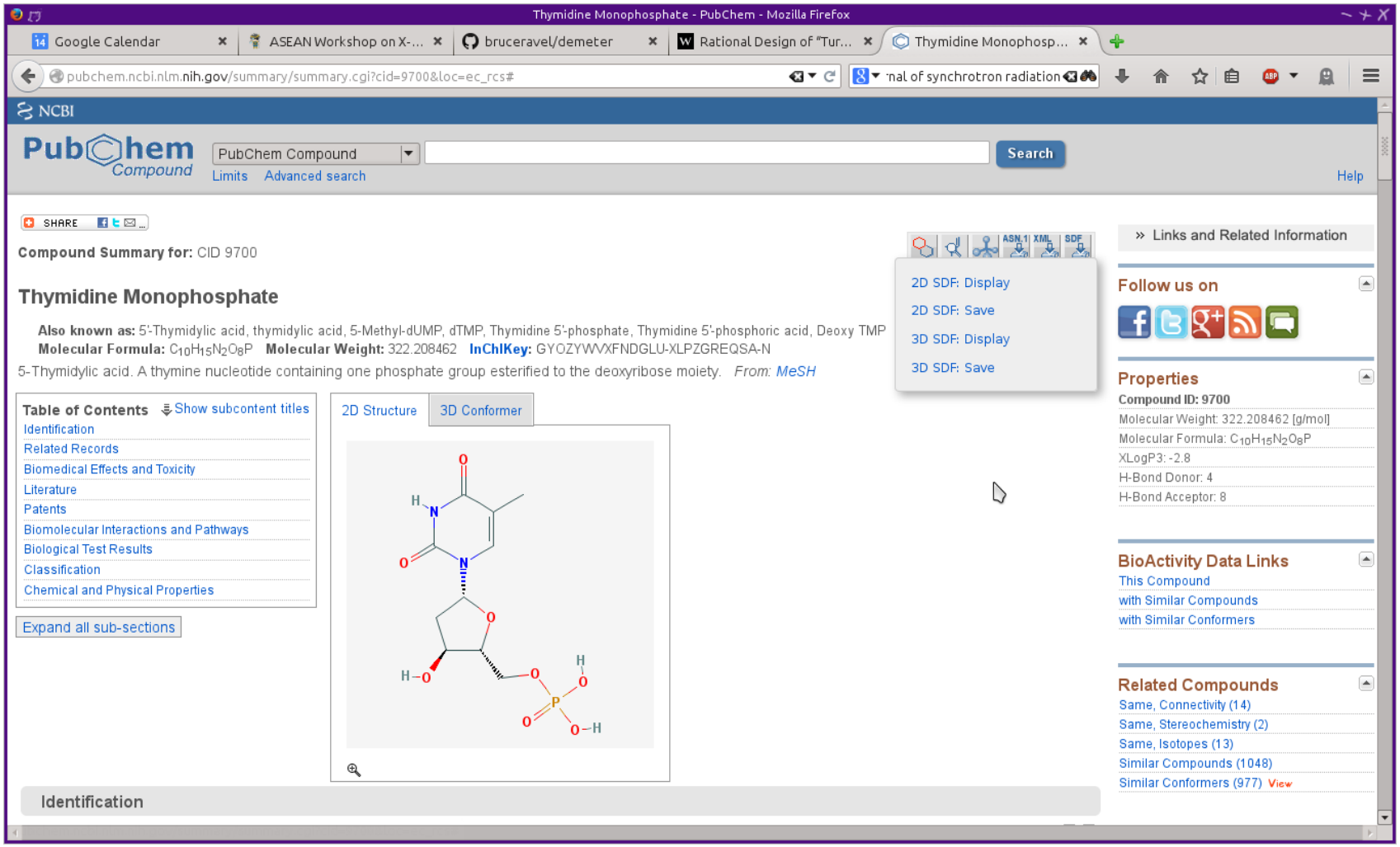
2D and 3D representations

The 2D figures on the previous page were generated from the canonical SMILES strings:

Adenisine	C1=NC2=C(C(=N1)N)N=CN2C3C(C(C(O3)COP(=O)(O)O)O)O
Thymidine	CC1=CN(C(=O)NC1=O)C2CC(C(O2)COP(=O)(O)O)O
Guanosine	C1=NC2=C(N1C3C(C(C(O3)COP(=O)(O)O)O)O)NC(=NC2=O)N
Cytidine	C1=CN(C(=O)N=C1N)C2C(C(C(O2)COP(=O)(O)O)O)O

Neat! But we need 3D structures to run FEFF...

Structure from PubChem



http://pubchem.ncbi.nlm.nih.gov/

Cartesian coordinates: 3D SDF file

```
9700
 -OEChem-05141416293D

36 37  0     1  0  0  0  0  0999 V2000
  -3.5515   -1.5175    0.1599 P   0  0  0  0  0  0  0  0  0  0  0  0
  -0.4389    1.3396    1.0202 O   0  0  0  0  0  0  0  0  0  0  0  0
  -0.9101    4.1569   -0.0812 O   0  0  0  0  0  0  0  0  0  0  0  0
  -2.7552   -0.1874    0.6247 O   0  0  0  0  0  0  0  0  0  0  0  0
   3.6173    1.7470    0.3907 O   0  0  0  0  0  0  0  0  0  0  0  0
   3.8378   -2.8022   -0.2452 O   0  0  0  0  0  0  0  0  0  0  0  0
  -2.5475   -2.2163   -0.8977 O   0  0  0  0  0  0  0  0  0  0  0  0
  -4.7267   -0.9241   -0.7790 O   0  0  0  0  0  0  0  0  0  0  0  0
  -4.0197   -2.4002    1.2798 O   0  0  0  0  0  0  0  0  0  0  0  0
   1.6113    0.5684    0.1973 N   0  0  0  0  0  0  0  0  0  0  0  0
   3.7127   -0.5224    0.0726 N   0  0  0  0  0  0  0  0  0  0  0  0
  -1.0101    2.8736   -0.6948 C   0  0  2  0  0  0  0  0  0  0  0  0
  -1.5699    1.8660    0.2995 C   0  0  1  0  0  0  0  0  0  0  0  0
   0.3733    2.3378   -0.9829 C   0  0  0  0  0  0  0  0  0  0  0  0
   0.7701    1.7196    0.3478 C   0  0  1  0  0  0  0  0  0  0  0  0
  -2.2796    0.6993   -0.3750 C   0  0  0  0  0  0  0  0  0  0  0  0
   1.0112   -0.6708    0.0146 C   0  0  0  0  0  0  0  0  0  0  0  0
   3.0176    0.6816    0.2323 C   0  0  0  0  0  0  0  0  0  0  0  0
   1.6792   -1.8209   -0.1381 C   0  0  0  0  0  0  0  0  0  0  0  0
   3.1656   -1.7831   -0.1119 C   0  0  0  0  0  0  0  0  0  0  0  0
   1.0130   -3.1449   -0.3336 C   0  0  0  0  0  0  0  0  0  0  0  0
  -1.6278    2.9841   -1.5911 H   0  0  0  0  0  0  0  0  0  0  0  0
  -2.2303    2.3332    1.0386 H   0  0  0  0  0  0  0  0  0  0  0  0

     (+ several more hydrogen atoms + bonding information)
```

Here is the "SDF" file for thymidine monophosphate from PubChem.

Along with lots of stuff not relevant to the EXAFS analysis, we find the Cartesian coordinates of all the atoms in thymidine monophosphate!

SDF = Structure data file

Cartesian coordinates: Feff input file

```
TITLE Hg decorating thymidine monophosphate
HOLE       4   1.0   *  Hg L3 edge (12284 eV), S0^2
*          mphase,mpath,mfeff,mchi
CONTROL    1      1     1     1
PRINT      1      0     0     0
RMAX       6.0

POTENTIALS
*    ipot   Z  element
       0   50   Hg
       1    8   O
       2    7   N
       3    6   C
       4   15   P
       5    1   H

ATOMS
*   x          y          z     ipot
  -3.5515   -1.5175    0.1599 4
  -0.4389    1.3396    1.0202 1
  -0.9101    4.1569   -0.0812 1
  -2.7552   -0.1874    0.6247 1
   3.6173    1.7470    0.3907 1
   3.8378   -2.8022   -0.2452 1
  -2.5475   -2.2163   -0.8977 1
  -4.7267   -0.9241   -0.7790 1
  -4.0197   -2.4002    1.2798 1
   1.6113    0.5684    0.1973 2
   3.7127   -0.5224    0.0726 2
  -1.0101    2.8736   -0.6948 3
  -1.5699    1.8660    0.2995 3
   0.3733    2.3378   -0.9829 3
*   (and so on...)
```

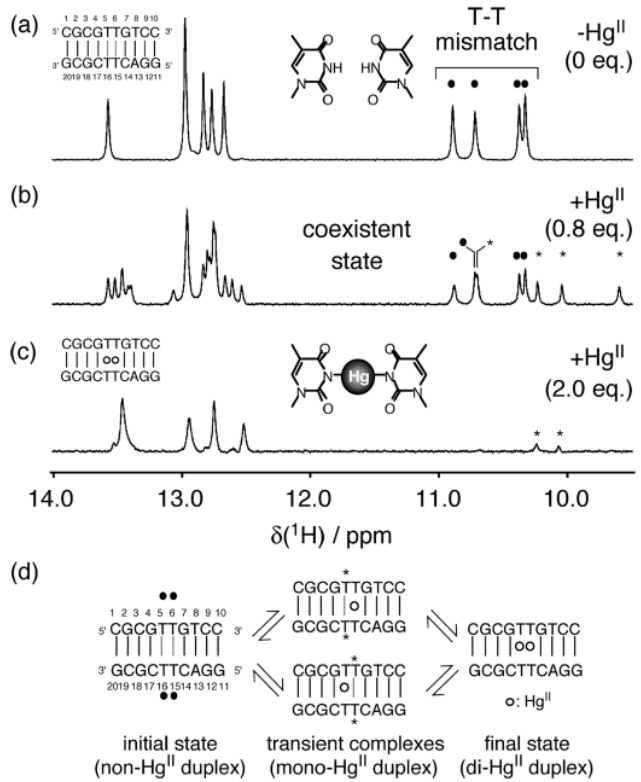
1. Do some cutting and pasting
2. Add some boilerplate for the header
3. Make a sensible `POTENTIALS` list

What about the Hg atom?

What is the likely location of the Hg atom?

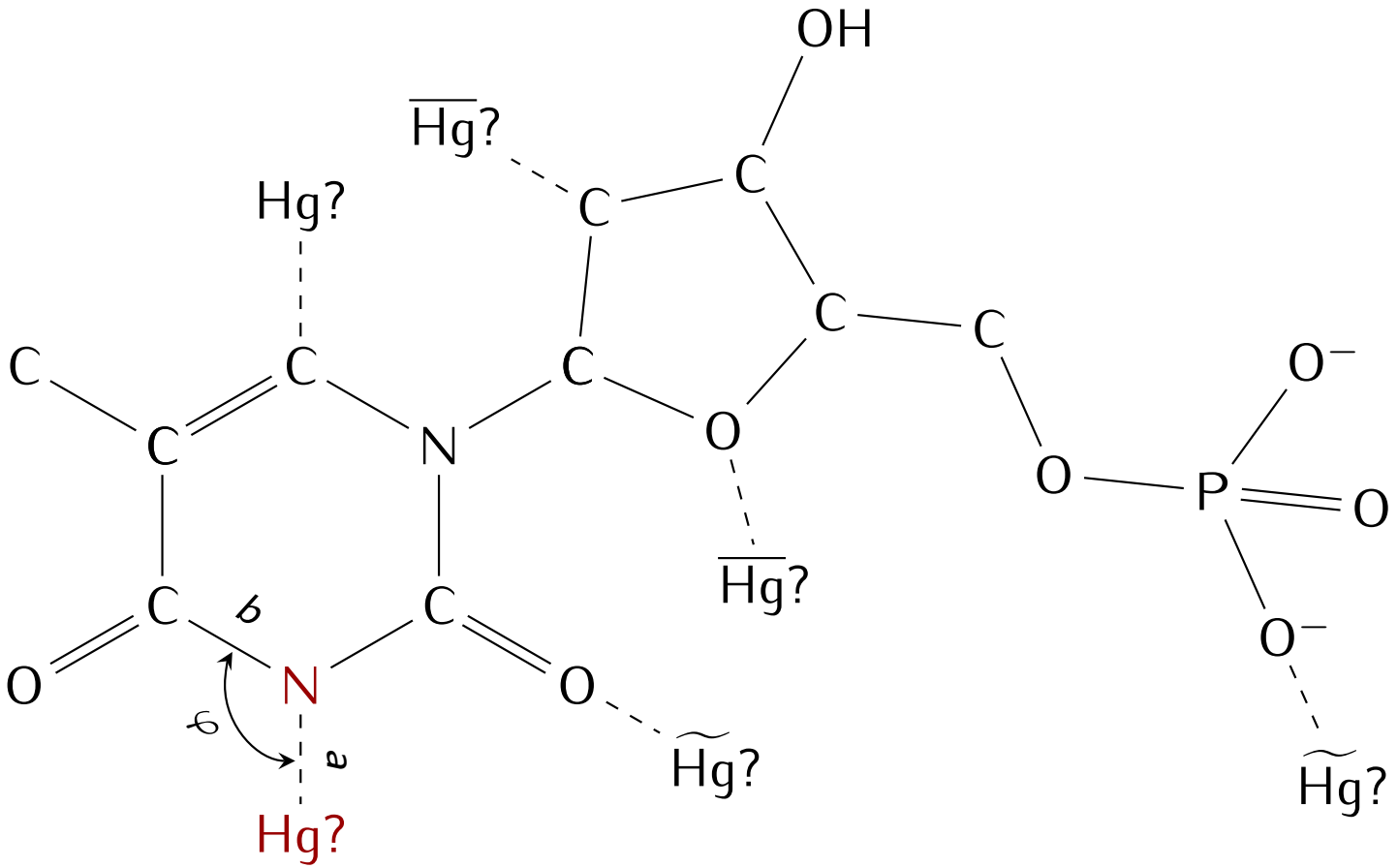
1. Thymine forms its hydrogen bond with adenisine via the N atom
2. The engineered DNA sensor is known to have a T-T mismatch
3. Earlier NMR work was interpreted at having the Hg bridging the T-T mismatch.

That said, I don't know much about this chemistry.



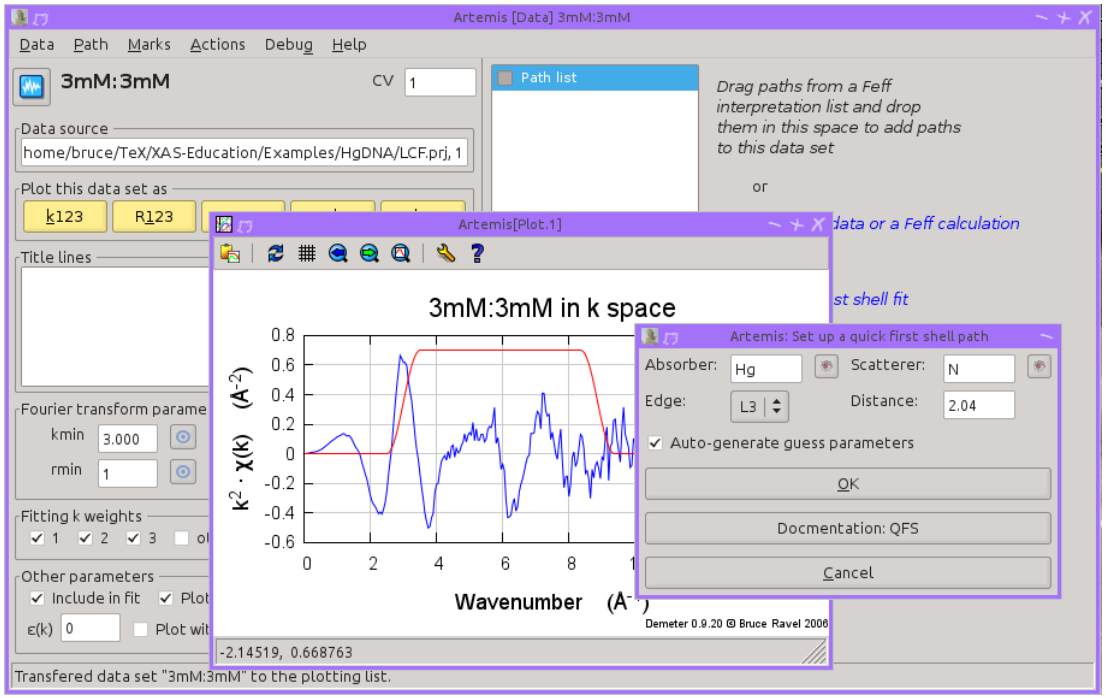
Y. Miyake, et al., *Mercury^{II}-Mediated Formation of Thymine-Hg^{II}-Thymine Base Pairs in DNA Duplexes*. J. Am. Chem. Soc. (2006) v.128, 2172-2173 DOI: 10.1021/ja056354d

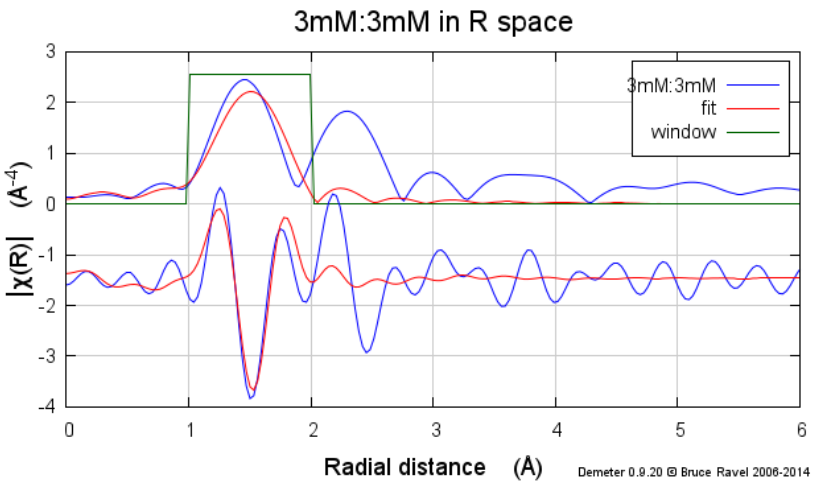
Decorating thymidine with Hg



Hg atom placement, 1

Do a quick first shell fit to determine the Hg 1st shell distance.





Not a great fit, but it tells us that the Hg atom is about 2.05 Å away from it's neighbor.

Hg atom placement, 2

Using the known nucleotide structures, I wrote a small program to solve some trigonometry:

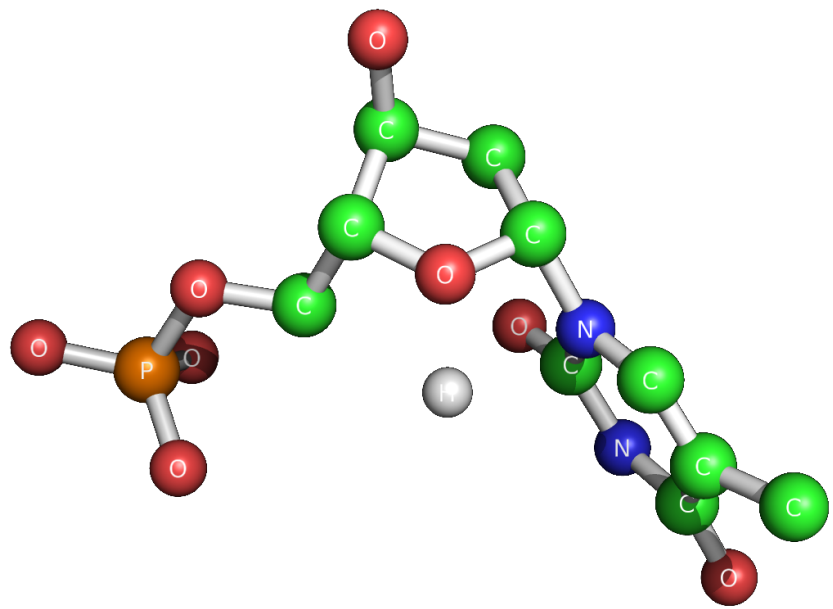
The Hg atom is ...

1. ... 2.05 Å away from its neighbor
2. ... in the same plane as the neighboring atoms
3. ... equidistant from the second neighbors (6- and 5-member ring options)
4. ... collinear with the 1^{st} and 2^{nd} neighbors (monodentate option)

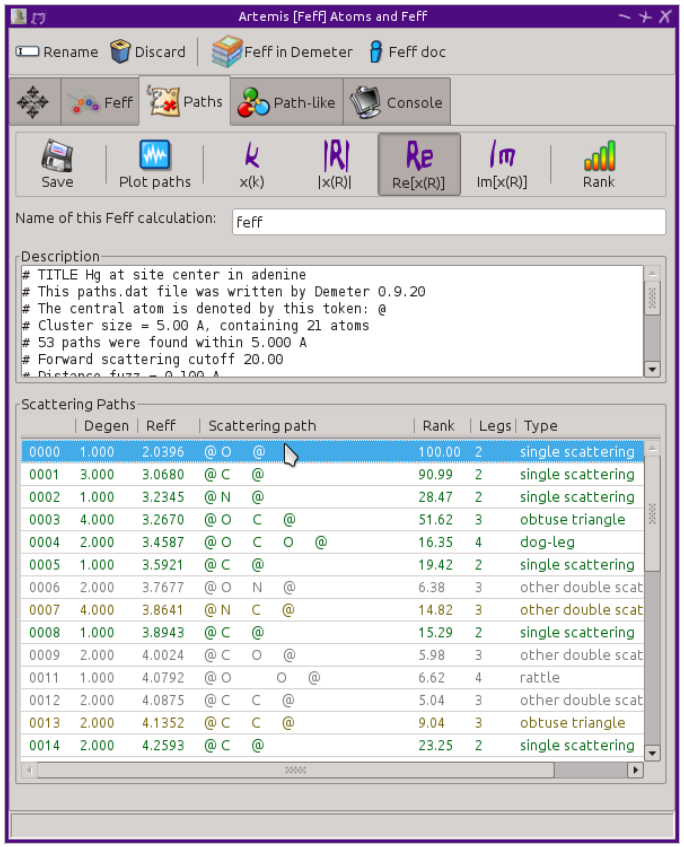
Finally, write out '`feff.inp`' files with Hg as the absorber.

5-member ring option: coordinates

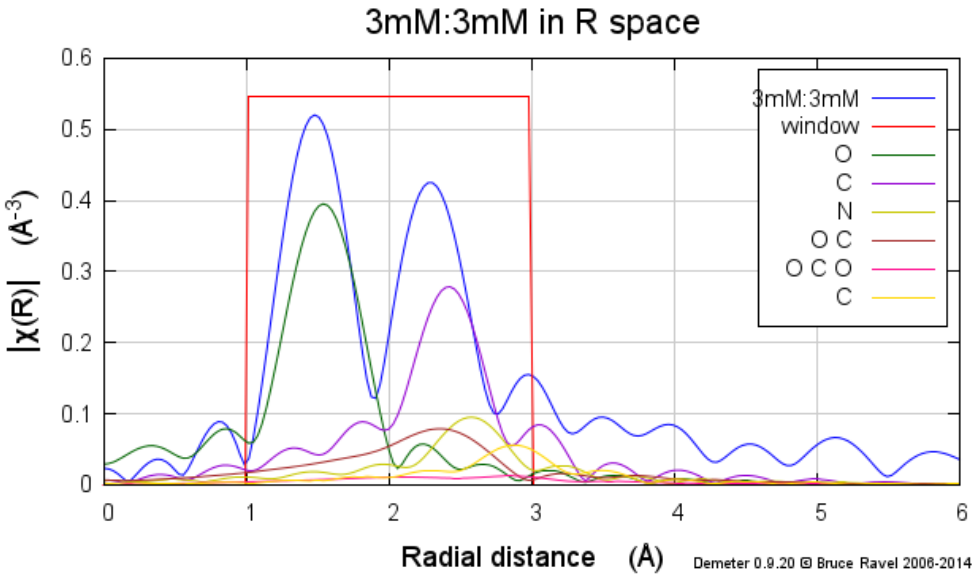
```
TITLE Hg decorating thymidine monophosphate
HOLE       4    1.0    *  Hg L3 edge (12284 eV), S0^2
CONTROL    1       1      1      1
PRINT      1       0      0      0
RMAX       6.0
POTENTIALS
*     ipot    Z  element
        0    50   Hg
        1     8   O
        2     7   N
        3     6   C
        4    15   P
ATOMS
*    x          y          z          ipot
      0.49977     0.63093     2.85314    0  Hg      0.00000
     -3.71800    -2.00000    -1.24900    1  O       6.44507
     -3.91000    -1.59800     0.29000    4  P       5.56632
     -5.35700    -1.58600     0.60000    1  O       6.65531
     -3.02000    -2.44600     1.11000    1  O       4.98947
     -3.35200    -0.12200     0.45900    1  O       4.59727
     -1.95500     0.01100     0.30500    3  C       3.59210
     -1.48700     1.40500     0.63700    3  C       3.07534
     -0.11000     1.31500     1.03100    1  O       2.05000
     -1.52300     2.38600    -0.51700    3  C       4.30462
     -1.77700     3.69900    -0.00600    1  O       4.77194
     -0.15400     2.19400    -1.16100    3  C       4.35705
      0.73400     1.75700     0.00100    3  C       3.07532
      1.68600     0.62300    -0.15600    2  N       3.23452
      1.54600    -0.35900    -1.10700    3  C       4.21393
      0.64900    -0.39000    -1.93000    1  O       4.89315
      2.52400    -1.32100    -1.06300    2  N       4.82117
      3.58700    -1.40200    -0.18100    3  C       4.78223
      4.40700    -2.31400    -0.22400    1  O       5.77995
      3.65500    -0.33500     0.78500    3  C       3.89431
      4.86400    -0.10800     1.62800    3  C       4.59276
      2.71300     0.59300     0.75000    3  C       3.05336
```



5-member ring option: paths



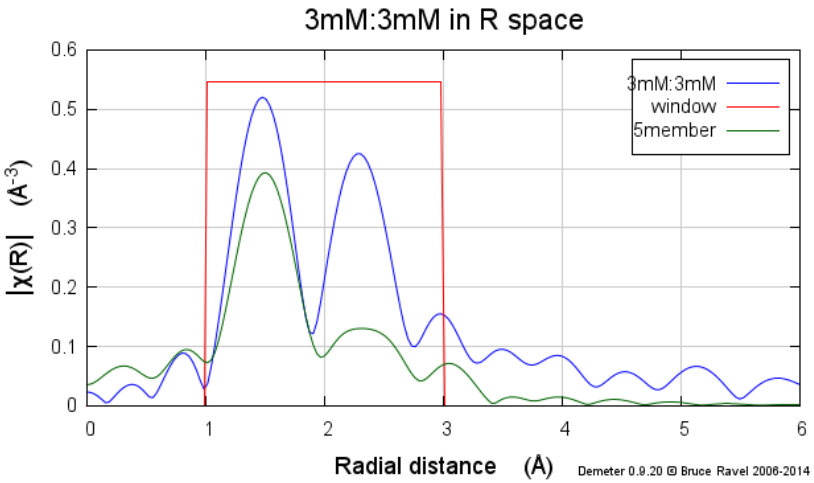
Run FEFF, drag-n-drop first 6 paths, transfer them to the plotting list, plot in R:

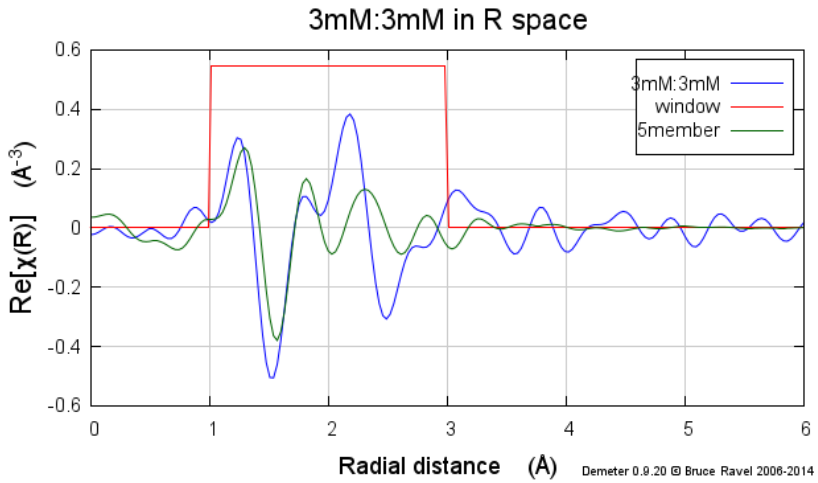


This looks sort of promising ... or does it?

5-member ring option: VPath

We fit a sum of paths to the data, so let's examine the sum of these paths. In ARTEMIS, this is called a "VPath."

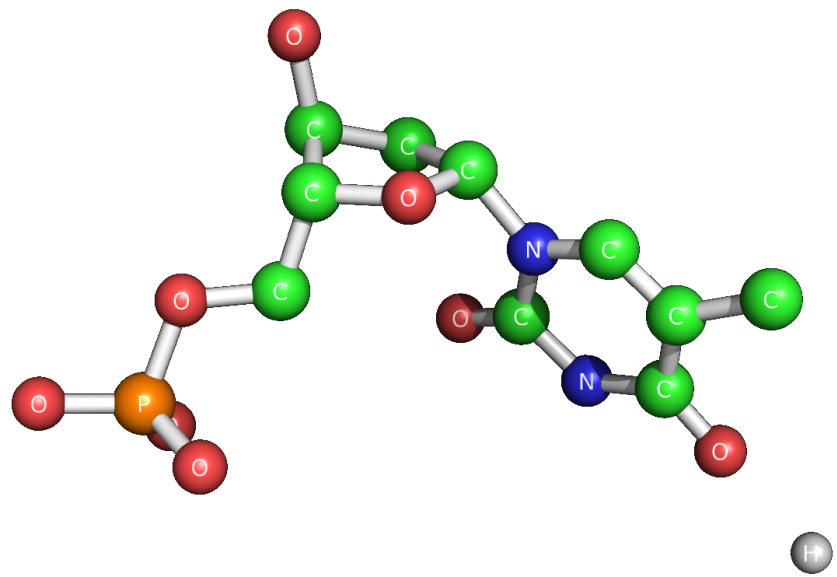




Not so promising, after all.

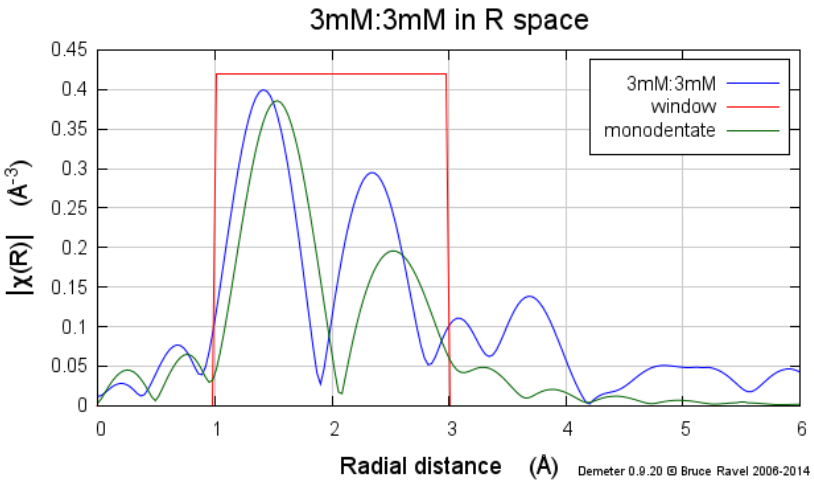
Monodentate option: coordinates

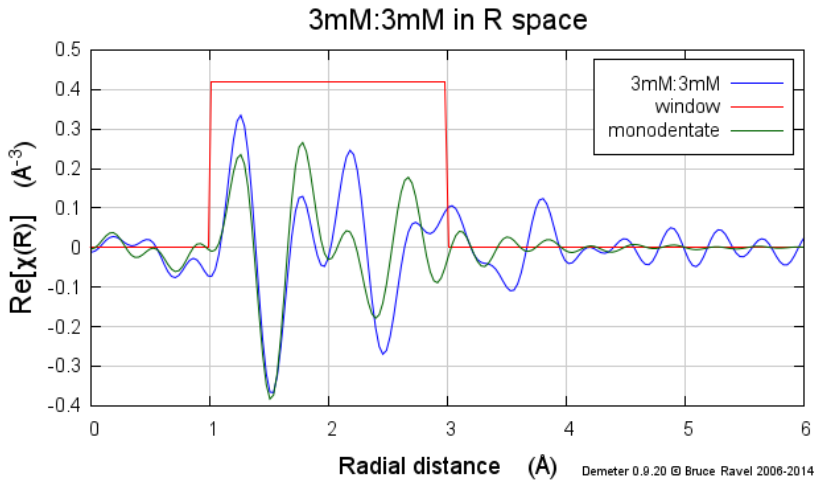
```
TITLE Hg decorating thymidine monophosphate
HOLE       4    1.0    *  Hg L3 edge (12284 eV), S0^2
CONTROL    1       1      1      1
PRINT      1       0      0      0
RMAX       6.0
POTENTIALS
*    ipot    Z  element
       0   50   Hg
       1    8   O
       2    7   N
       3    6   C
       4   15   P
ATOMS
*   x          y          z     ipot
     5.74339    -3.80032    -0.29408   0  Hg      0.00000
    -3.71800    -2.00000    -1.24900   1  O       9.67837
    -3.91000    -1.59800     0.29000   4  P       9.91863
    -5.35700    -1.58600     0.60000   1  O      11.35435
    -3.02000    -2.44600     1.11000   1  O       8.97789
    -3.35200    -0.12200     0.45900   1  O       9.83988
    -1.95500     0.01100     0.30500   3  C       8.61105
    -1.48700     1.40500     0.63700   3  C       8.95772
    -0.11000     1.31500     1.03100   1  O       7.88572
    -1.52300     2.38600    -0.51700   3  C       9.54572
    -1.77700     3.69900    -0.00600   1  O      10.62446
    -0.15400     2.19400    -1.16100   3  C       8.45356
     0.73400     1.75700     0.00100   3  C       7.48765
     1.68600     0.62300    -0.15600   2  N       6.00394
     1.54600    -0.35900    -1.10700   3  C       5.48832
     0.64900    -0.39000    -1.93000   1  O       6.34502
     2.52400    -1.32100    -1.06300   2  N       4.13555
     3.58700    -1.40200    -0.18100   3  C       3.22719
     4.40700    -2.31400    -0.22400   1  O       2.05000
     3.65500    -0.33500     0.78500   3  C       4.18739
     4.86400    -0.10800     1.62800   3  C       4.25452
     2.71300     0.59300     0.75000   3  C       5.43826
```



Monodentate option: VPath

Same exercise – run feff, drag-n-drop the first few paths, make a VPath, plot with the data.

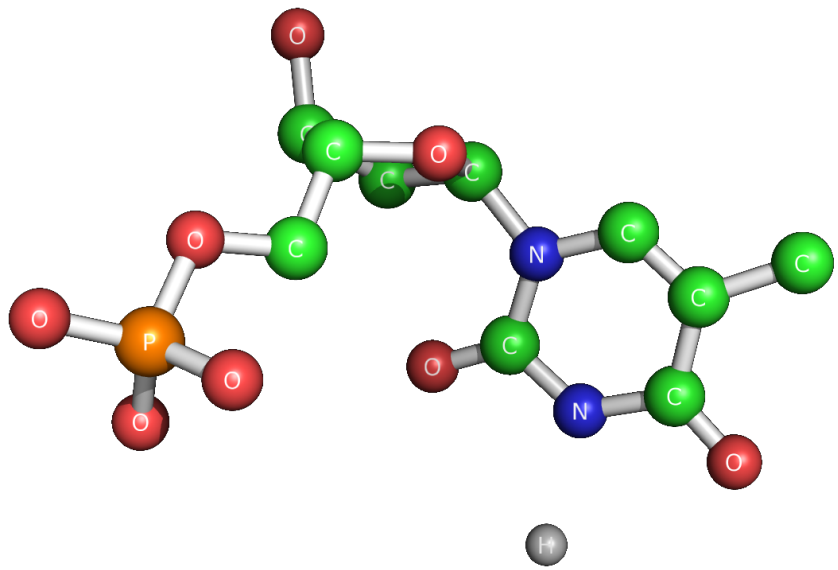




Better than the 5-member ring option, but still not so great.

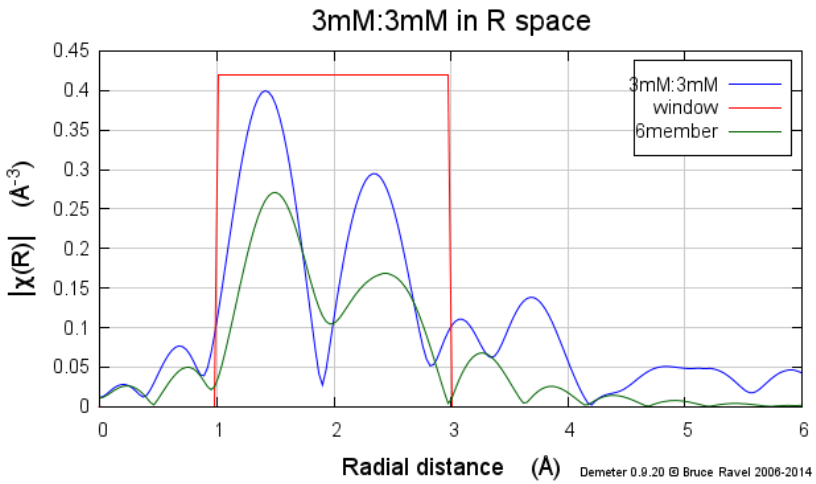
6-member ring option: coordinates

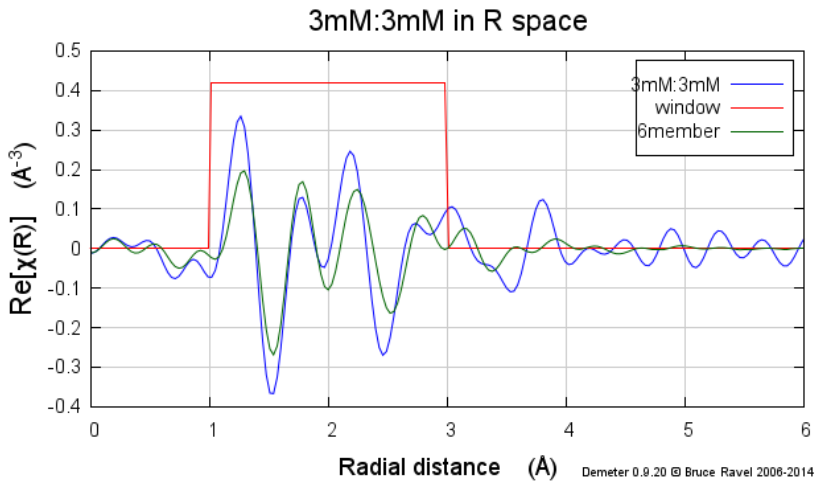
```
TITLE Hg decorating thymidine monophosphate
HOLE       4    1.0    *  Hg L3 edge (12284 eV), S0^2
CONTROL    1      1      1      1
PRINT      1      0      0      0
RMAX       6.0
POTENTIALS
*    ipot    Z  element
       0    50   Hg
       1     8   O
       2     7   N
       3     6   C
       4    15   P
ATOMS
*   x          y          z     ipot
     2.40463    -2.80748    -2.45560   0  Hg      0.00000
    -3.71800    -2.00000    -1.24900   1  O       6.29242
    -3.91000    -1.59800     0.29000   4  P       6.99112
    -5.35700    -1.58600     0.60000   1  O       8.43040
    -3.02000    -2.44600     1.11000   1  O       6.50160
    -3.35200    -0.12200     0.45900   1  O       6.98896
    -1.95500     0.01100     0.30500   3  C       5.87972
    -1.48700     1.40500     0.63700   3  C       6.51567
    -0.11000     1.31500     1.03100   1  O       5.95606
    -1.52300     2.38600    -0.51700   3  C       6.79387
    -1.77700     3.69900    -0.00600   1  O       8.11301
    -0.15400     2.19400    -1.16100   3  C       5.76519
     0.73400     1.75700     0.00100   3  C       5.44613
     1.68600     0.62300    -0.15600   2  N       4.19199
     1.54600    -0.35900    -1.10700   3  C       2.92421
     0.64900    -0.39000    -1.93000   1  O       3.03360
     2.52400    -1.32100    -1.06300   2  N       2.05000
     3.58700    -1.40200    -0.18100   3  C       2.92356
     4.40700    -2.31400    -0.22400   1  O       3.03859
     3.65500    -0.33500     0.78500   3  C       4.26357
     4.86400    -0.10800     1.62800   3  C       5.47827
     2.71300     0.59300     0.75000   3  C       4.68340
```



6-member ring option: VPath

Again – run FEFF, drag-n-drop the first few paths, make a VPath, plot with the data.





I actually like this one quite a bit! The amplitude is off by about a factor of 2, but the phase is quite close.

Parameterization

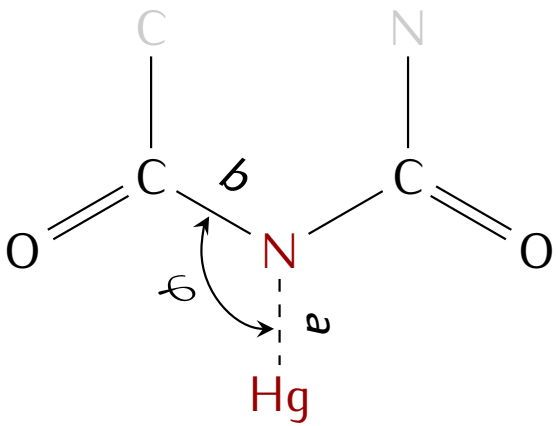
Number of independent points

k-range: $2\,\text{Å}^{-1}$ to $8.8\,\text{Å}^{-1}$ R-range: $1\,\text{Å}$ to $2.8\,\text{Å}$

$$N_{\text{idp}} = 2\Delta k \Delta R/\pi \approx 7.8$$

1. E_0 and `amp` are variables ..... (1,2)
2. Hg-N distance and σ^2 are variables ..... (3,4)
3. Hg-O distance and σ^2 are variables ..... (5,6)
4. Assume that the ring is completely rigid, this allows us to approximate the contributions of various single and multiple scattering paths without introducing any more variables.

Trigonometry



$$\varphi = 116.25^\circ$$

$$b = 1.378\,\text{Å}$$

a and $\sigma^2_{\text{Hg·N}}$ are variables of the fit.

Here's a formula for a triangle in a plane:

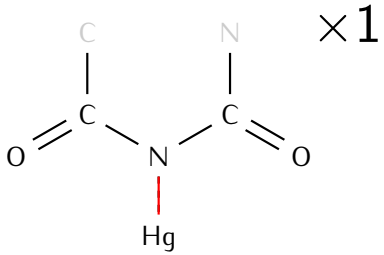
$$D(Hg - C) = \frac{a-b}{\cos(\theta)}\cos(\varphi/2)$$

$$\tan(\theta) = \frac{a+b}{a-b}\tan(\varphi/2)$$

Assuming the ring is **rigid**, then we approximate $\sigma^2_{\text{Hg·C}}$ (and others) by scaling geometrically from $\sigma^2_{\text{Hg·N}}$

Paths

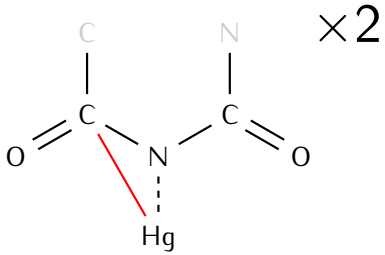
Path 1 (SS)



×1

ΔR_1 and σ_1^2 are variables

Path 2 (SS)

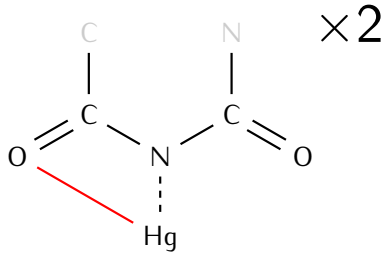


×2

ΔR_2 computed with trigonomtry

$\sigma_2^2 \propto \sigma_1^2$

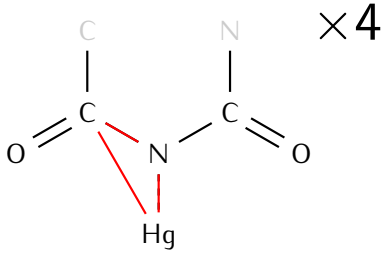
Path 3 (SS)



×2

ΔR_3 and σ_3^2 are variables

Path 4 (MS)

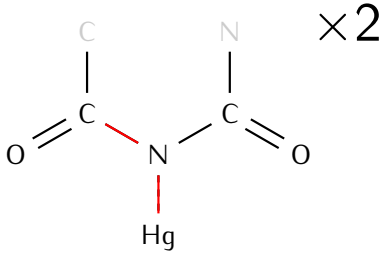


×4

ΔR_4 computed from paths 1 and 2

$\sigma_4^2 := \sigma_2^2$

Path 5 (MS)

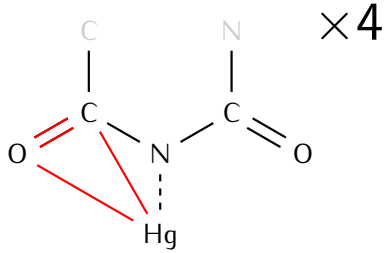


×2

ΔR_5 computed from path 1

$\sigma_5^2 := \sigma_2^2$

Path 6 (MS)

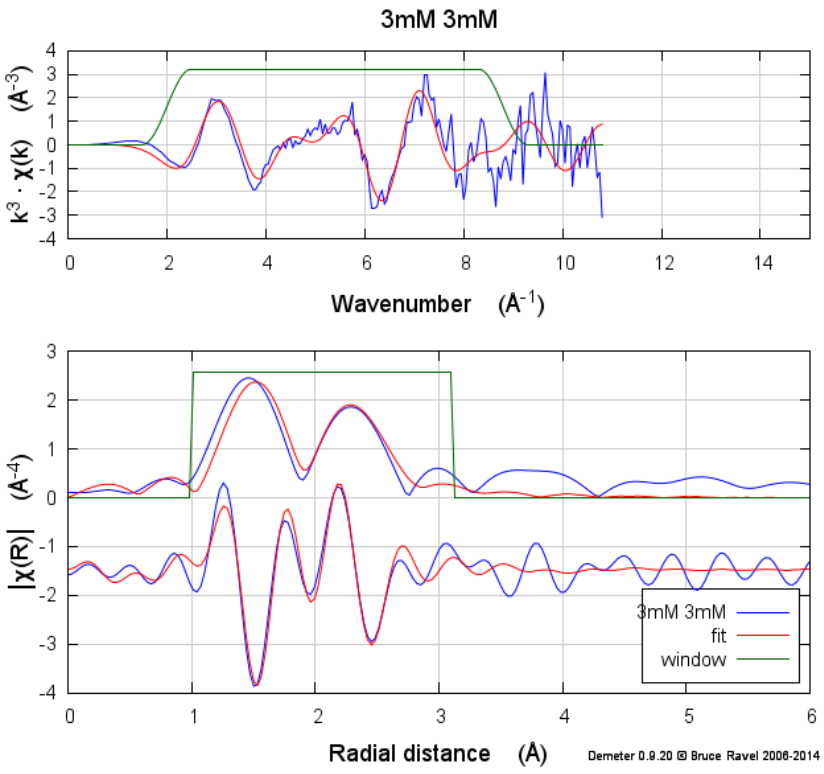


×4

ΔR_6 computed from paths 1 and 3

$\sigma_6^2 := \sigma_1^2 + \sigma_3^2$

Fitting result



amp	1.86 ± 0.44
E_0	1.41 ± 1.91
$\Delta R(N)$	0.006 ± 0.028
$\Delta R(O)$	-0.058 ± 0.063
$\sigma^2(N)$	0.0046 ± 0.0045
$\sigma^2(O)$	0.0096 ± 0.0081

Why is amp near 2?

The Hg atom bridges 2 thymines. Our FEFF model had Hg bound to 1 thymine. So S_0^2 is really 0.93(44)!

Uncertainties

- The data are short – i.e. little information content – and noisy
- The uncertainties are all quite large, although the best fit values all make sense
- S_0^2 came out right, although with large uncertainty
- The σ^2 approximations are sensible, but certainly not correct
- The assumption that the ring is rigid is sensible, but certainly not correct
- The assumption that the Hg atom sits in the plane of the ring is sensible, but certainly not correct

Our data are **consistent** with the Hg atom bound to the N atom in the 6-member nitrogenous base

What could we have done better?

- The As in the cacodylic acid hurt. Use a different buffer.
- The sample geometry hurt. Use better packaging or a focusing mirror.

Those two things could have increased efficiency by about an order of magnitude. Another couple inverse Ångstroms would have made a huge difference!

What have we learned?

- The science question required interpretation of both XANES and EXAFS
- Quick first shell fit to approximate the first shell distance
- Made input for FEFF from published structural data and a sensible guess for the location of the Hg atom
- Tried several possible coordination geometries, but only pursued the one that looked promising
- Dealt with limited information by applying interesting constraints
- We didn't exactly *solve* the structure, but we demonstrated that the EXAFS data are consistent with the assumption from NMR