

Gruenwald Group

Interactions between gold nanoparticles and oleylamine ligands

Michael O. Chen

This document summarizes my work. Perhaps it could be of use to someone studying gold NPs and their superstructures in the future. All code, data files, output files, and other misc. supplementary material can be found in this¹ GitHub repository.

1 lammps simulation choices

Some key parts of the lammps MD simulation:

- Data was collected every 5000 timesteps no matter the length. This was done because it was thought that a constant resolution would be advantageous for certain statistical procedures.
- To calculate the coordination number a simple constant number was used, and if an atom was within this number distance of the NP, it was counted. This number was determined for each variable varied (ebind, lig num, s) by using only a few cases for each variable, not all of them by looking at g(r).
- A second shell was only clearly seen in the radial distribution function of the high vacuum solvent parameter ones, so this is why these were calculated only for s.

2 Effect of binding energy, number of ligands, and implicit solvent quality on coordination number(s)

2.1 Methodology

The basic process for measuring coordination number with respect to variation in each parameter was: find 2nd equilibrium point, then take an average over all measurements after and including this point.

The first equilibrium refers to the nve/limit period in which the randomly placed ligands organize themselves into a not unnatural position (which the random create_atoms may have placed them in). The second equilibrium is when the free energy of the system has been minimized, which is approximated here by the point at which the LJ potentials reach a minimum (and then oscillate there).

Finding the second equilibrium was broken down into 2 key parts: (1) determination of whether second equilibrium was reached in the simulation's time range and (2) determination of a specific timestep to use as the approximation for when second equilibrium starts. For part (2) to happen, part (1) must be satisfied. If either part is not satisfied, a longer simulation will be run.

2.1.1 Determination of whether second equilibrium was reached

The central guiding question was: after this time range, will PE increase or decrease without balance? To determine whether second eq. was reached, a 'tripod' test was used. Essentially, the idea is that there are three general ways in which second equilibrium has not been reached:

- Case 1: continuous monotonic change. Beyond the time range, the PE continues a monotonic increase or decrease clear in the time range.
- Case 2: continuous non-monotonic change. Beyond the time range, the PE changes in some way, which is a continuation of a non-monotonic trend in the time range. E.g., a spoon.
- Case 3: discontinuous change. In the time range, equilibrium has been established. Beyond the time range, the PE suddenly breaks from equilibrium and changes in some way.

Case 3 is considered very unlikely and is thus not considered. Case 1 and 2 are both considered plausible, and are accounted for. Case 1 is considered more likely than Case 2.

Each of the three 'legs' of the tripod test are checks for Case 1 or Case 2. In essence, if it is determined that it is plausible that Case 1 or Case 2 is happening during that leg's check, that leg has 'broken' and the tripod test has not been satisfied. Here are the legs:

- **Leg 1: Visual Inspection.** Visually inspect raw PE and 1/10 block PE for patterns *. If there is a clear Case 1 or Case 2, Leg 1 has broken (Case 2 can be said to be if there is 3 or less extrema).
- **Leg 2: Case 1 Brute-Force Check.** Correlation coefficient r is calculated for increasingly smaller blocks of the raw PE terminating at the end of the time range (figure is called 'leg2.png'). r will always oscillate wildly by the end of the time range, this section is ignored. Leg 2 breaks if both of the following are true: [1] r is always negative (or positive) and [2] r is close to 0.6 a lot of the time[†].
- **Leg 3: Case 2 sensing.** Blocks of 1/4 the time range are created over the entire time range. Within each block, the correlation coefficient r is calculated (figure: 'PE_fourth-block_r'). If the plot indicates 3 or less extrema, Case 2 is seen as plausible and Leg 3 breaks[‡].

If none of the tripod legs break, it has been determined that second equilibrium happens in this time range. Then, the specific point to use for 2nd eq. start must be determined.

* Found in the respective 'figures' folder as PE_raw.png and PE_tenth-block_avg.png

[†] This needs to be improved.

[‡] More needs to be done to quantify this. Under this system it is also redundant if Leg 1 contained large oscillations, it is really for when the plot is a bit smaller in amplitude, and what this means was not reasoned through. Maybe also looking at correlation at the end?

2.1.2 Determination of specific point for second equilibrium start

Two general scenarios are considered:

- If the equilibration appears to be monotonic: the first clear minimum within range of equilibrium should be used. If you argue it is later, I could just as well argue that is equilibrium fluctuation. It is assumed after the time range the oscillation will continue as it does.
- If the equilibration appears to not be monotonic: determine the slope of the final trend. If it is negative, find the first min. If it is positive. Find the first max[§]. **One-fifth rule:** because I see it as less likely that equilibration is not monotonic, the equilibration section must be at the very most one-fifth of the total time range. If it is more it is possible that after the time range the oscillations will adjust so that it is actually a monotonic equilibration. If equilibration is greater than one-fifth, a specific approximation cannot be chosen and a longer simulation must be run.

Other general priority considered: make sure nothing too high or too low is included. If something is clearly too high or low (it is not reached again), do not include it. Exclusion matters less, so go farther generally.

To determine the precise minimum or maximum, a zoomed in version of the raw PE graph was used, as well as the fat derivatives method (taking derivatives with large dx (and dy)). The fat derivatives is saved as 'PE_deriv.png'.

If either the tripod test has not been satisfied (2nd eq. was not reached) or part (2) cannot be satisfied (a specific approximation for 2nd eq. cannot be chosen), a longer simulation should be run.

Once the timestep to be used as the start of second equilibrium has been found, the average coordination number is calculated.

2.1.3 Calculation of average coordination number

In the 'out' folder for each run, 'coord_distr.dat' includes the coordination number measurement (column 1) for each timestep (column 0) after and including the chosen 2nd eq. point.

Next, 'out.txt' is a text file which includes the chosen 2nd eq. point as a timestep and as an index, the average PE and coordination number after 2nd equilibrium, and the standard deviation of the sample of coordination number measurements.

'coord_distr.dat' is the distribution of coordination number measurements after second equilibrium (a sample), and can be used to calculate the standard error of the mean, considering time correlation. This was not done because of time and my mathematical incompetence. Sample standard deviation was reported because it was easy to calculate, perhaps it can be used as looking at relative variation which may be skewed (?).

A note on organization: all results can be found in the 'results' folder. All of the files mentioned above (relating to individual sims)

can be found in the 'sims' folder. sim_analysis.ipynb is the notebook which does everything above and is in the github root.

'Consolidated' analysis, or analyzing the results from multiple sims is done in consolidated_analysis.ipynb, and the results saved in the 'consolidated' folder. The discussions below are all based on consolidated analysis.

2.2 Binding Energy

Below is shown the plot and table of how coordination number varied with binding energy, found using methodology described above. A fuller version of results and data can be found in results/sims/ebind. (The results below are contained in results/consolidated/, made using the consolidated_analysis.ipynb notebook.)

Table 1 # of closest neighbors[¶] for different binding energies.

Binding Energy (kcal/mol)	Coordination #	Sample st. dev.
2	18.931868574884096	3.903032618959191
4	100.19834041691965	6.117599658796531
6	204.17267206477732	5.61016256616313
8	281.2528565382988	4.832711211997653
10	331.5643117887052	3.8048110206378083
12	367.899221147038	2.124543356467281
14	393.42428160320304	2.0662491026325362
16	414.4988619119879	1.916586934479705
18	414.77272727272725	0.5554196624034861
20	—	— ^{**}

Figure 1 contains this data plotted.

Binding energy specific notes:

- Ebind = 2 is the only one in which PE increases towards eq, which makes sense since ligands binding is the least favorable
- Ebind = 10 may be slightly positive Case 1
- Ebind = 12 may be a min, max, dive Case 2. Oscillation in there makes me not increase.
- Ebind = 14 is NOT monotonic. It appears to rise after decreasing, then decrease after rising, then stay constant. Some clues point to another oscillation, but will assume not there.
- Ebind = 16 may very well be Case 2, but assumed not for now; it clearly breaks the one-fifth rule (and may have other issues). Another longer simulation is being run but won't make it into these plots.

2.3 Ligand Number

This section will follow the same pattern as the preceding section, but describe the variation of coordination number with *ligand number*. Generally these second 2 sections were not made with as much case as binding energy. Ligands had much more noise than binding energy.

Figure 2 contains this data plotted.

Ligand number specific notes:

- lig = 2 could very well be either Case 1 or Case 2. Assumed neither.

[§] Min/max are used as they indicate the possible ending of equilibration. More work was not able to be completed to fully justify this. Partly because I want to make sure the equilibration process is over, technically before could say no equilibrated...?

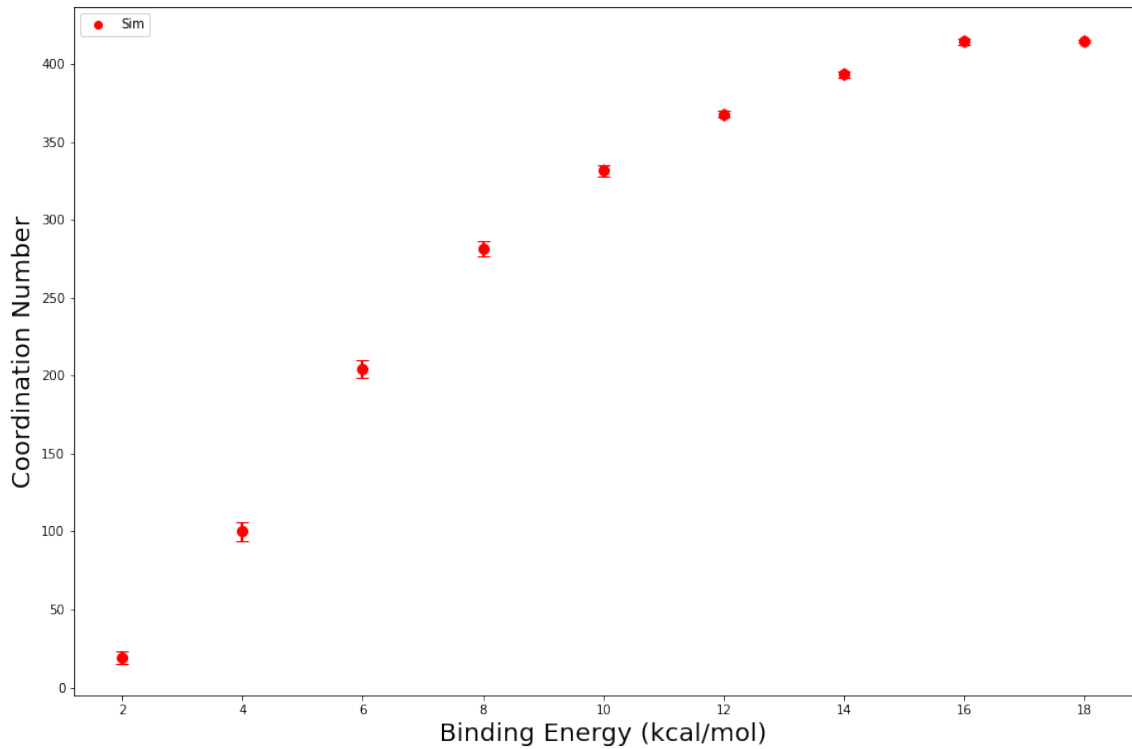


Fig. 1 Coordination number vs. binding energy.

Table 2 # of closest neighbors^{††} for different ligand numbers.

Binding Energy (kcal/mol)	Coordination #	Sample st. dev.
1	99.25751072961373	0.7812225848372742
2	190.82758620689654	3.0358963530836602
3	241.42566510172145	4.41219084406003
4	263.7913429522753	4.1358557457956433
5	274.2002781641168	4.157612065125467
6	281.6684266103485	4.2617684291139777
7	286.5438996579247	3.9707204444996456
8	291.5663390663391	4.213245337453836**
9	292.93969298245617	4.358481280376076
10	296.91769041769044	4.296153097439177
11	301.4266211604096	3.9033040534902783

- lig = 3 could be Case 1 or 2, or be breaking the one-fifth rule, there is a lot of ambiguity.
- lig = 5 assuming monotonic, could not be.
- lig = 6 has some serious issues.
- lig = 10 and 11 has some strong Case 1 appearance.

Longer runs should really be run for these above cases as well as 8-11.

2.4 Solvent parameter s

This is presented just like the above two sections, except everything is 5mil and possibly too short (25mils have been run on pre-emptable and if they run I will add them). The check for equilibrium was also done differently and less rigorously (so not very rigorous at all), but most of them have likely equilibrated (this was actually done a lot earlier before the code was refactored and slightly

modified, but it should be the same with the new code). Second coordination shell is actually calculated separately from everything using the file ‘results/sims/s/s_analysis.ipynb’. Figures 3 and 4 contains this data plotted.

3 Coordination structure of ligands in bad solution This was the only real qualitative thing studied, unfortunately I did not have enough time to do it really at all. I had the weird filament periodic boundary mess simulation, and these produced these images (Figures 5 and 6) which show perhaps one interesting fact: the second shell has ligands pointing out, indicating perhaps some order in these high vacuum environments of the ligands that could be elucidated given better simulation conditions. I believe these were created with 1100 ligands, 1.0 s, and 8.0 ebind.

I ran it with a larger box and it did run, and it appears to support the above (except the third shell has tails pointing out) but I did not have enough time to go more into depth. I recognize this work is essentially useless but wanted to include it just because I thought it might be interesting to look at.

Notes and references

- 1 <https://github.com/michaelchen78/AuNP-oleylamine>.

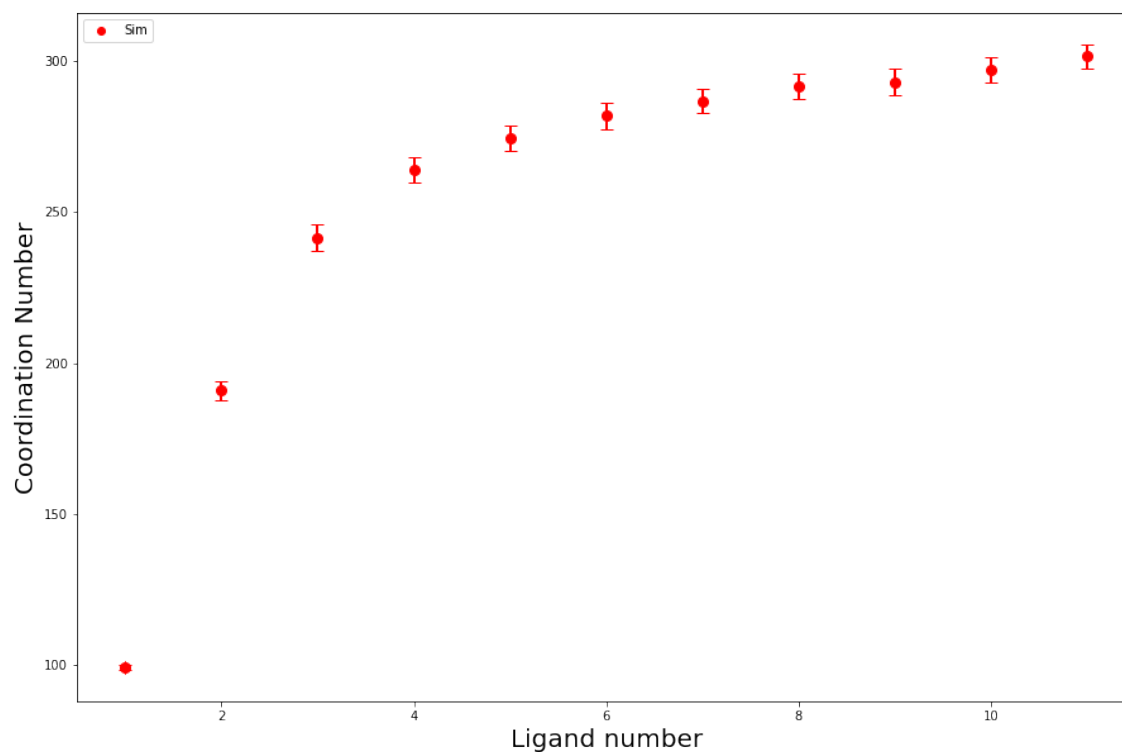


Fig. 2 Coordination number vs. ligand number.

Table 3 # of closest neighbors and second closest neighbors for different s values.

Binding Energy (kcal/mol)	Coordination #	Sample st. dev.	2nd Coordination #	2nd Sample st. dev.
0.1	254.5710928319624	0.28554641598119856	5.028910405161022	0.535352792878698
0.2	280.78537735849056	0.4492924528301887	4.615171669468029	0.64996555668592
0.3	314.8083900226757	1.0430839002267573	3.8884295190562024	1.0591811519232857
0.4	339.3578947368421	3.2221052631578946	3.5178778372541184	1.7607522429302371
0.5	354.7508813160987	8.898942420681552	3.1035438641149886	2.7523912701030615
0.6	362.86503067484665	15.024539877300613	3.0814888258229236	3.4595703072365214
0.7	367.18657937806876	19.734860883797054	2.81876904621537	3.8057436022229423
0.8	371.1621621621622	25.04054054054054	2.3560280009184624	4.053969212227784
0.9	373.6678787878788	28.008484848484848	2.422337089585397	3.945155042368548
1.0	376.36675126903555	32.380710659898476	2.259810992548742	4.3741453043562615

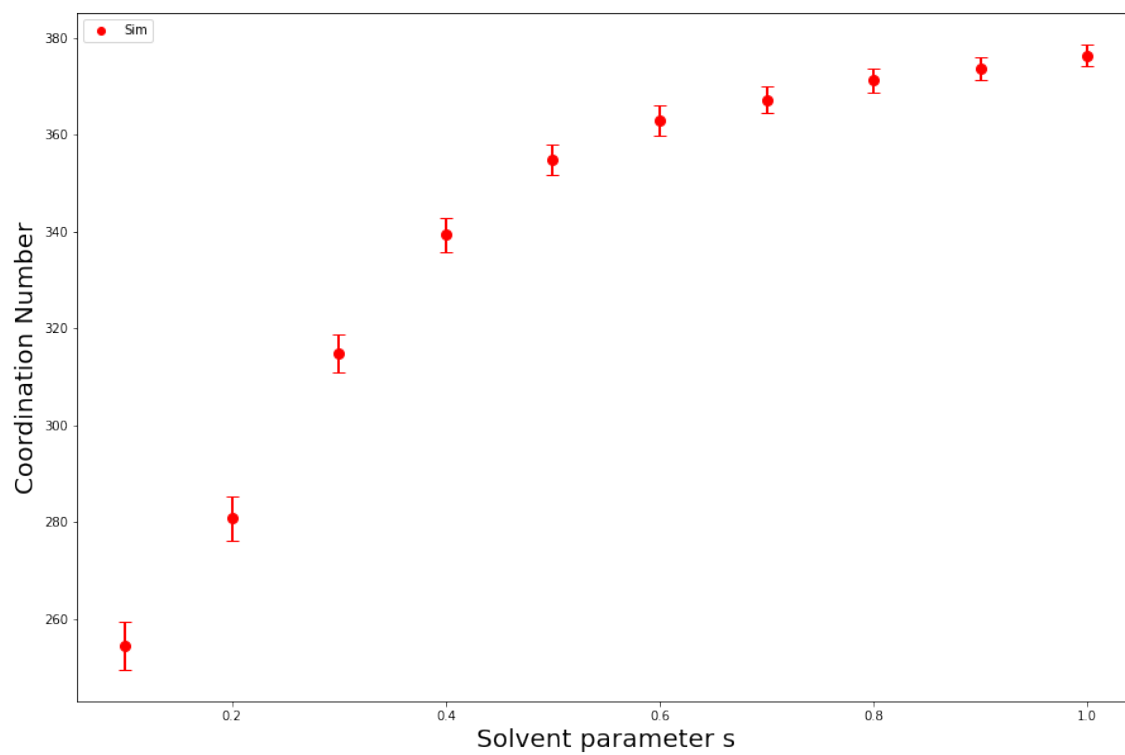


Fig. 3 Coordination number vs. solvent parameter s .

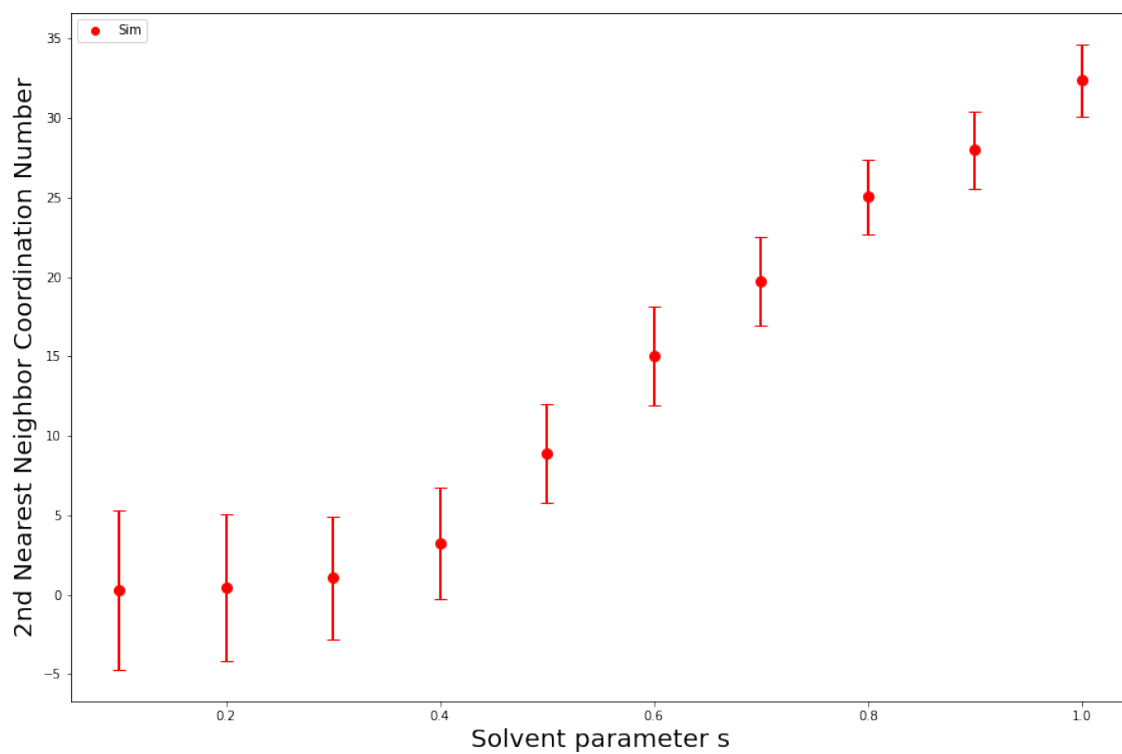


Fig. 4 2nd coordination number vs. solvent parameter s .

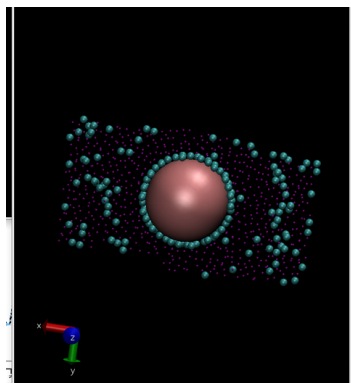


Fig. 5

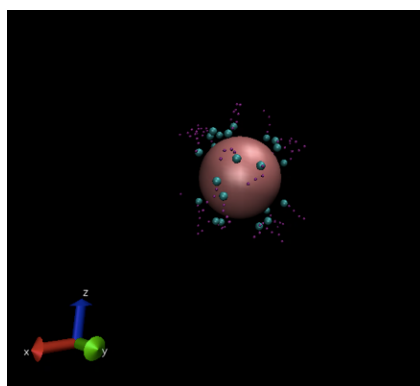


Fig. 6