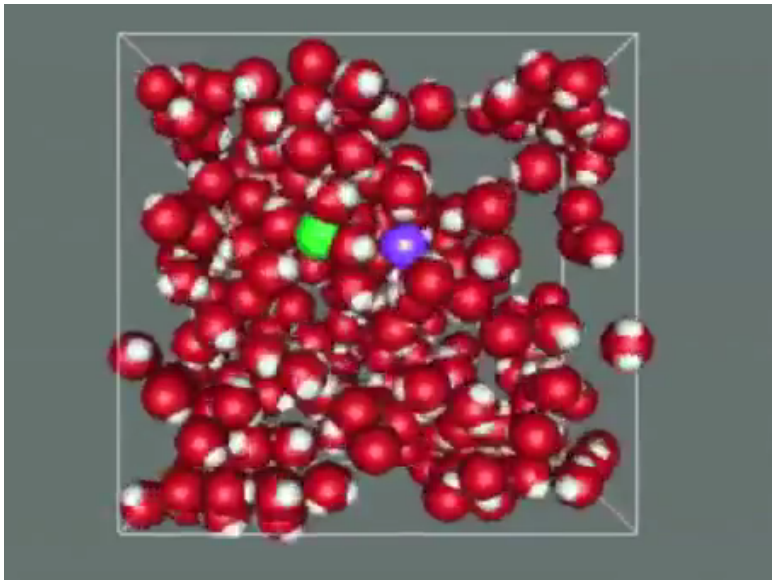
Computational Chemistry & Classical Molecular Dynamics
Prof. B. L. Tembe
Department of Chemistry
Indian Institute of Technology- Bombay

Lecture – 36
Molecular Dynamics using Gromacs-4: Liquid Water: Input Files

Hello, we will continue our discussion on classical molecular dynamics simulations, in the last lecture, you were introduced to several types of files, some of those files you can edit using your vi editor, some of the files you cannot edit because they are binary files, the examples of binary files are your tpr file, trr file and xtc file, now why we use binary files instead of regular ASCII files is because binary files occupy much less space on the hard disk than your text files.

Text files are ASCII files, they take a lot of space and you know space on the computer becomes a very primary issue because your trajectories are long, lacks of steps are carried out and you cannot save all that information if it is in the text format, okay, so before we continue our discussion on simulation on water today, we will show you some videos on molecular dynamics simulations, can you start the first video please.

(Refer Slide Time: 01:16)

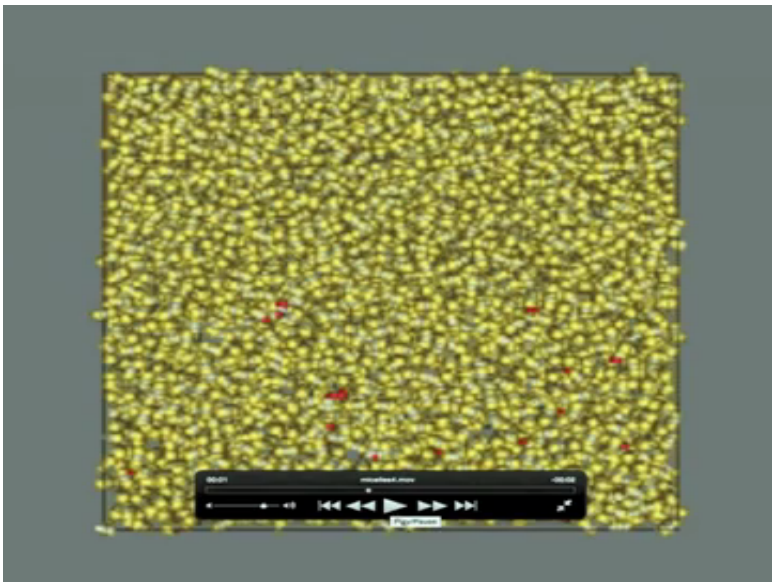


Yeah, this is the video of NaCl in pure water, the green one is chloride and the purple one is Na+, this is the short video, okay, so then you can show some other video, yeah sure, so you will see that in classical molecular dynamics, all the molecules are represented as some sites;

interacting sites. If it is a water simulation, the water is represented by 3 sides that is one side for oxygen and 2 sides for water, okay.

So, now this is for a super critical fluid, yeah, this is for a super critical fluid and this is pure water, this is that is the simulation for pure water, yeah, you will see that the white objects are hydrogens and the red object is the oxygen. Now, why do not we see a simulation on a more complex system?

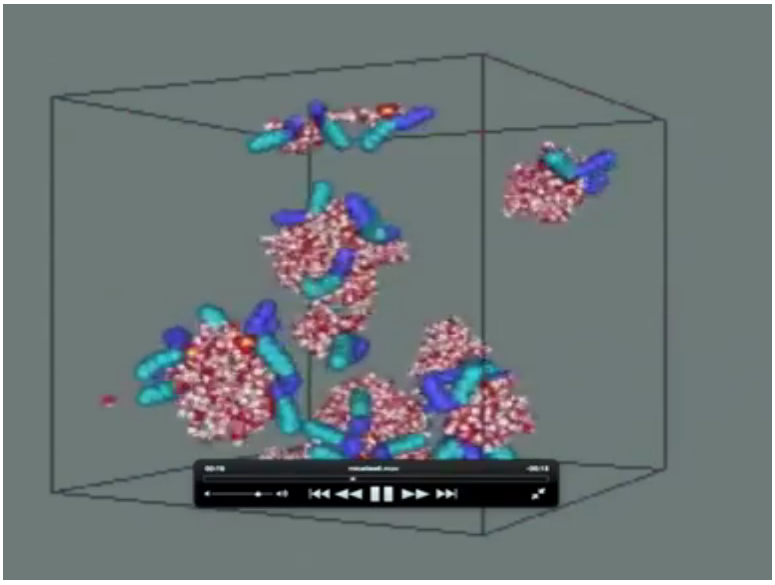
(Refer Slide Time: 02:09)



Yeah we can show them representation of a micellar system, so we can call it classic molecular dynamics because all the equations of motions are all classical that is based on the Newtonian equation of motion, whereas in a quantum mechanical simulation, it will be based on Schrodinger equation or Dirac equation or more complex versions of that so, since you have seen the video, you can see that whatever simulations are done you can actually physically see on the screen.

And of course, remember that these are all simulations; simulations are not 100% reality because all simulations are based on some model potentials, models of algorithms, so it is not a 100% reality but it gives us a fairly good knowledge of what reality is about.

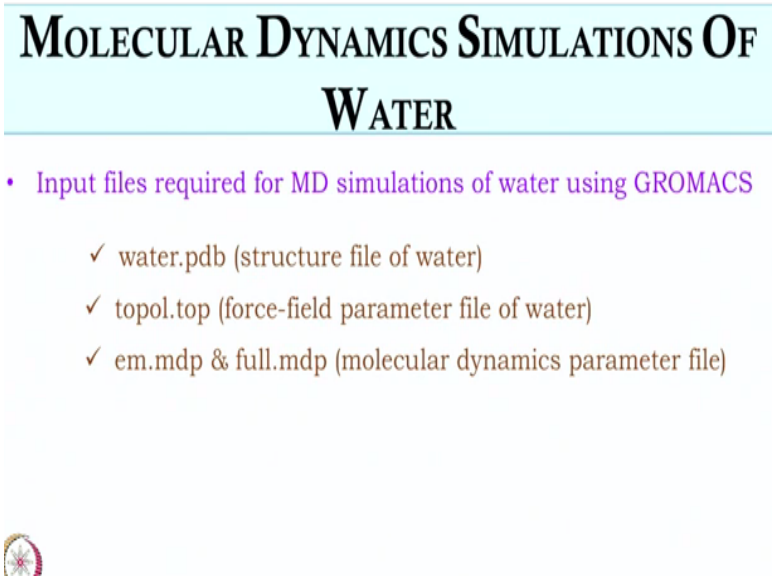
(Refer Slide Time: 02:49)



So, this is simulation of a; this is for a micellar system, it has been aggregated in the solution, okay, this is more complex system than water or methanol or argon, okay. So, now what we will do; once we have seen that now, we will start go back rather to see what are the files that you need for water simulations, okay. So, today we will be discussing how to create input files for water, so that you can do molecular dynamics of water, okay.

So, we again begin with your pdp files, topology files, so let us start with simulations for water.

(Refer Slide Time: 03:34)



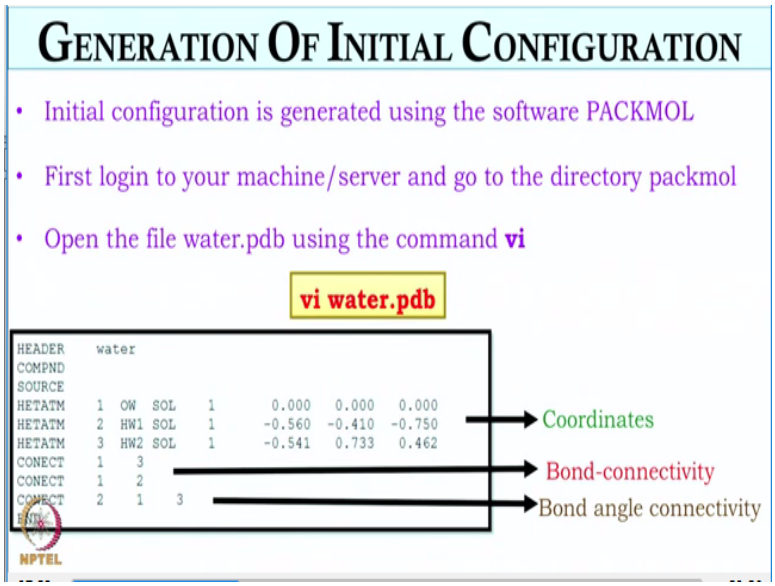
So, for molecular dynamics simulations of water, we need the input file which are required is the structure file of water that is water.pdb, then we require topology file that is where in the force field parameter of water is included, then we need mdp file which is a molecular

dynamics parameter file, we need 2 mdp file; one is the energy minimisation mdp file and one is for the full molecular dynamics.

So, can you comment on where the these files are available? These files are available in the Gromacs site itself, okay and also in the manual Gromacs; in the manual of Gromacs, the format is given, okay, like should be the format of a topology file of a molecular dynamics parameter file and pdp file is you can create a pdb file using Gaussview, Avogadro or it is available free of cost from automated topology builder or RSB protein data bank side.

Or if I know the structure of the file, I can create it manually also, yeah but you can create it manually for simple molecule but when it becomes more complicated then, it is easier for us to downloaded from somewhere or to create it and then optimise the geometry and then use it, okay, yeah and to generate; but you did not say about the imp file, yeah imp file will come later when we discuss about the generation of initial configuration.

(Refer Slide Time: 04:54)



So, to generate initial configuration, we will use the help of the software Packmol which is a free software and first, we have to login to our machine or server and then we have to go the directory packmol using the command cd and will have the file water. pdb which we can view using the command vi, so vi water.pdb, this is the pdb file of water, okay, let us see this very carefully because it gives information about all the details.

Yeah, so the last 3 columns represent the coordinate of x, y and z axis of 3 site of water since water is 3 site molecule; one is oxygen, 2 is hydrogen, so OW HW1 and HW2 are the site of

water molecule and the last 3 column are the coordinate of x, y and z axis and the first 3 lines that is header, water compound source that is common for all pdb files, and earlier some files had a blank line, is that allowed or no? That is allowed, blank line does not mean anything like, okay, you can neither write something or you cannot write something.

Because you can see now the Gromacs structure is so build that it is very, very flexible as well as data is concerned, yeah, whereas when we wrote our own programs earlier, you see that even if there is a blank line, it will make a big mess because the blank line will be neither 0 or nothing, so if program; in the program that you write unless such flexibility is there, you have to be very, very careful about the data structure, exactly

Because one dot or one single digit, it can mess up the whole thing, yes, okay, so it is nice that they have worked on a fairly good structure and I would like to comment on those x, y, z coordinates, so before that what is that 1, 2, 3 on the left; 1, 2, 3 is the atom number, okay, like there are 3 atoms in water, so it is 1, 2, 3, okay and SOL is the residue name for water, okay and so instead of SOL, can I call it POL.

So but then POL has to be same throughout, okay like if we use POL in water.pdb, so you have to use the same residue name in say topology file or in the itp file, everywhere, everywhere you have to use the same okay, residue name; residue name is user defined, whatever the user define but whenever other people are making pdb file, they are using SOL as a common feature, yeah SOL is the common feature, okay.

And what are the next lines, you have 1, 2, 3, yeah 1 connect 1, 3 means that the bond; oxygen and hydrogen is connected, right, okay 1 and 3 and 1 and 2 so that is the bond connectivity which is represented and the last line where connect 2, 1, 3 it is a bond angle connectivity, okay, so that means you do not have to give the actual bond angle, the computer knows how to calculate the angle since you say 1, 2, 3 have an angle.

The actual bond angle and bond length everything you have to specify in the topology file or in the itp file, no, in a pdb file, you do not have to actually write the bond angle and bond length but I would say that information is already contained in the coordinates, yeah, you can calculate the bond angle as well; because suppose, we will leave it as an exercise, now you know that 000 is the coordinate of oxygen atom, yes.

And the second one -.560 – .410 and -.750, these are the x, y, z coordinates of the first hydrogen, yes, okay. So, if you treat it as a vector, now treat this -.50 as an x axis, next one as y, third one as z, it is like xi + yj + zk, now you can calculate the length of this, the length will be x square + y square + z square; square root, now since the oxygen is at the centre, this length will be just the OH bond length, yes, okay.

So, similarity the second one, you see the second one is slightly different from the first that is the bond length between oxygen and hydrogen, yes, so I would urge you to calculate these bond lengths, now suppose you want the distance between 2 hydrogen atoms, you can calculate from here because you have x1, y1, z1, x2, y2, z2, so if these are the 2 coordinates, the distance between them is the x2 – x1 square + y2 – y1 square + z2 – z1 square, square root over that, so that will give you the bond length between 2 hydrogens.

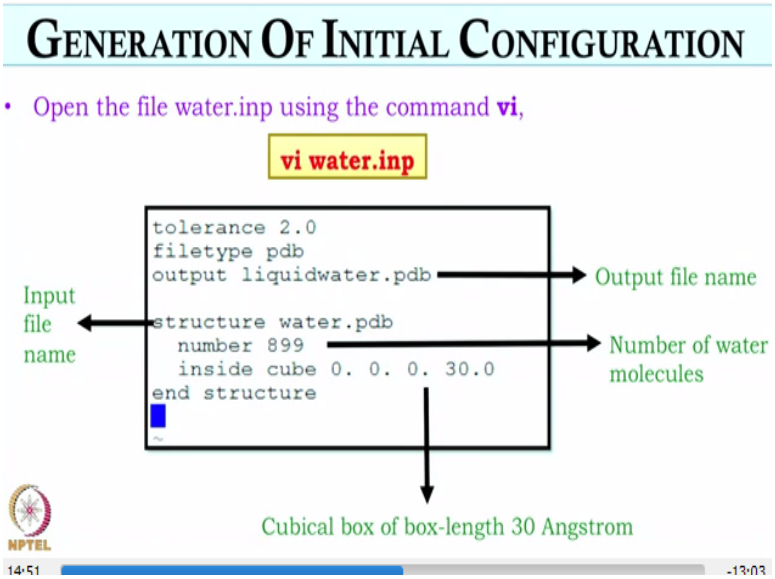
And if you want a bond angle, you have the first vector for OH and second vector for OH2, OH1, OH2 just take a dot product that is will give you a cos theta and cos inverse of that will give you the bond lengths, so and make sure you have these unit vectors because if it is not a unit vector, you will have a problem, so what yo do; OH1 OH2 take unit vectors in those directions, then take the dot product cos inverse then you will get the bond angle, okay.

So, therefore whenever you see this coordinates, I urge you to make it a habit that you check these things because ultimately, the result you get you are responsible for it so you have to make sure that the coordinates are correct and this is the basic input file, so in case something is wrong here, then everything will be wrong, yes, so in fact, you will be using several molecules know, maybe you have a mixture of alcohol and water, yeah alcohol, you can use many, many solvent molecules.

So, make sure that whatever your pdb file is there, you check the bond lengths and the bond angles at least 1 or 2, so that you know that it is a reasonable thing and also now these which kind of a model for water is, is it spc; this is spc and furthermore, there are many, many models of water with different bond lengths, different bond angles, so we will be doing with spc water.

So, in your home assignments, may be we will ask them to do some other model of water, yeah, there is TIP3P, TIP4P, 4 site or 5 site maybe continue with, so there are different models with different sites, so as you go along you practice with other models in your results, yeah, you can, any other questions on this now, things are okay, no, thanks sir, okay, yeah. So, now we have the water.pdb file and now we have to put in a box.

(Refer Slide Time: 10:51)



So, we have to define the box length and how many number of water molecules are to be put in what is the name of output file, so all these are contained in the imp file, okay, imp file is an input in packmol okay, so you can open the file water. imp using the command vi, so vi water.imp will give you a file like this, so but this one is it not created by us? You can create it or you can download it also.

Like initially, if someone is the beginner, they do not know what is the format of an imp file, so they can go to the site of packmol okay, where there are input examples for water urea mixture, for bilayer, for single water molecules, so they can download the imp file and pdb file from there as a beginner and it is free of cost but now water is there only one imp file or there are many?

The imp file format for water or for any molecule is the same, you have to have a tolerance you have to define the minimum distance between 2 atom or molecule, you have to define the file type, you have to give an output file name, so what is the meaning of tolerance 2 here, tolerance 2 is the distance between 2 molecule is 2 < or > 2 angstrom or minimum 2 angstrom that is user defined, you can give 2, 3.

So, I would consider that 2 is very small between 2 water molecules, you can give 3 as well, I would like to give something like 2.8 yeah for example that is allowed to give, okay, so I can change that 2 to 2.5, yes, yes, definitely but I cannot change the next 2 lines no, next 2 line file type has to be pdb because packmol; in packmol, the input file format is the pdb file format and so now we have a; and the third line output that is dependent on user whatever the file name they want to give.

So, in this file, the fourth one is the blank line yes, so can I remove the blank line, yeah you can remove the blank line or can I give 5 blank lines, yeah you can give 5 blank line also does not matter and the one thing that people can change now in this imp file, you have 899 water molecules, so they can change it and 30 box length, yeah, so I would ask you as an assignment take 20 as the box length that is an angstrom and calculate how many water molecules are there.

Yeah, so may be when they do the execution, you can change that 30 to 20 at 899 to some other number, yeah, so before we come to the next lecture please calculate how many molecules are there in the box of 20 by 20 by 20, so again now pdb files are all in angstrom; pdb file are all in angstrom unit and whenever you we have to use it in Gromacs, we have to have in a gro format, correct and the unit has to be in nanometre.

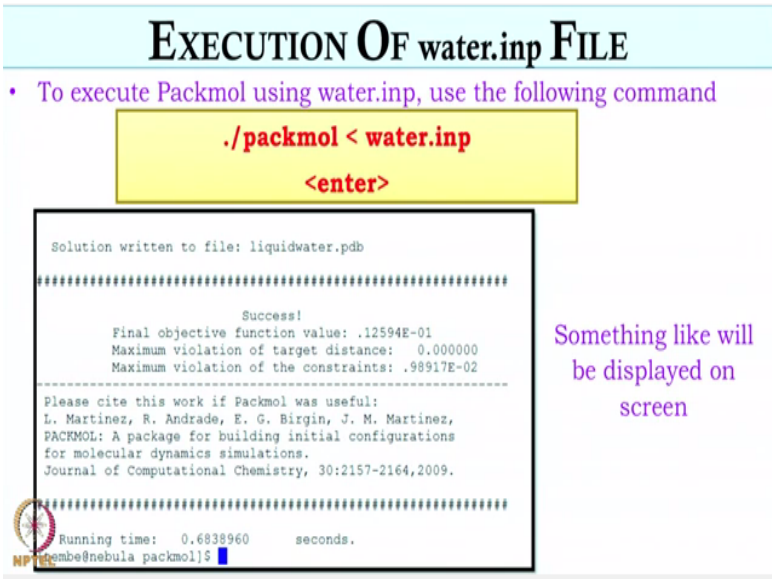
So that we will do when we go along, so what is the next file now after this pdb, so next is structure where we have to give the pdb file of water or methanol or any other molecules that we are simulating, number is user defined, we have to first calculate the number according to the box length and density of the molecule but that is already done here know, yeah that is already done here.

And then inside cube, it means that the it is a cubical box, the box length is 30 angstrom, you can have a different box as well okay, I have taken here cubical box and then you have to end the structure now, okay with structure and end structure, yes. Now, that between inside the cube there is 1 or 2 space, yes, can I make it 3 spaces, no because inside a cube like inside is the separate word and cube is the separate word so there is one space in between.

So, I cannot change the space here, no what about that 000 now, yeah that space you can change, so instead of whatever space given on the screen, one space is there, 0. space, 0. space, 0. space, so can I make it 0. space space space 0. space space space, yeah, you can do that it does not mind, it should be on that line, it cannot be on the next line, it should be on that line itself, yeah, so and in case some of you get an error you just go back and make sure things are in this order, okay.

We have already seen that is the output file name, this is the input file, this is the number of water molecules and this is the cubical box of box length 30 angstrom, okay.

(Refer Slide Time: 14:56)



So, now we have to execute packmol using this water.imp file, so to do that we will use the following command, ./packmol space < water.imp, so that means water.imp is an input file yeah, packmol is an executable command, yes and the result will be storing some other file name, okay, so the result will be stored in as it is written their solution, written to file liquid water. pdb that is the output file name okay.

And after successful execution it was so; show like this in the screen success and then the time that it has taken to run the file, so what is this violation of the constraints? See, you have success, yeah then final objective function value then maximum violation of target distance like whatever the tolerance we are giving, okay so what is the violation that they have done so that is what is written there okay, maximum violation of target distance.

But if the tolerance is < 2 or if it does not satisfy the thing then it would not write the pdb file, it will say something like the structure is forced to be written in that box length, so now these constraints or these bond constraints or some other constraints; bond constraint because everything is dependent on the poly atomic molecule is dependent on the bond length and bond angle and distance between molecule.

No but my other question is what this packmol is doing, is really packing several molecules in a box, yeah, correct and we are giving the distance like if there is 1 water molecule first we are placing it in centre let us say, correct and the other water molecule it is placing 2 angstrom far away from the first one correct, then again when it is putting the third one it is placing it 2 angstrom far away from the second one, correct, so this is the way it is packing the molecules in the system.

So but the way we used to write our own programs, we just take a molecule and put it somewhere there is no change of bond length or bond angle, yeah bond length and bond angle would not change here, correct but then still there is a; so what is that constraint then; constraint is the distance between the molecule because that is the only constraint we are giving that there you see when we were doing argon simulation, we have used tolerance of 3 angstrom, correct.

So that is the distance between 2 argon atom should be minimum 3 angstrom, it cannot be <3 but so here it is allowing some flexibility, yeah, right because if you do not allow for this flexibility, it is sometimes very difficult to put all the molecules in the box, okay, so now you see that although there are some constraints here that is may be some lengths are < expected when you do energy minimisation, automatically, this will be pushed away, yeah, okay.

And what is the last line; running time is 0.6seconds, it took 0.6 seconds to execute the packmol, so now since the last line says some line is there that is the working directory is not it, yeah that is packmol directory, so we are still in packmol directory. Now, we have to copy the file liquid wter.pdb in the directory where we want to do simulation of water, so for that first we have to create a directory mkdir water, okay outside the packmol directory.

And then we will copy the file liquid water.pdb to that directory. So, now there is also is there not a Gromacs directory where all the things are uploaded; uploaded but you still there is a

Gromacs directory because unless you have a Gromacs directory, you cannot run the command of Gromacs, right but though there is a Gromacs directory, you have to create a directory for water of for argon.

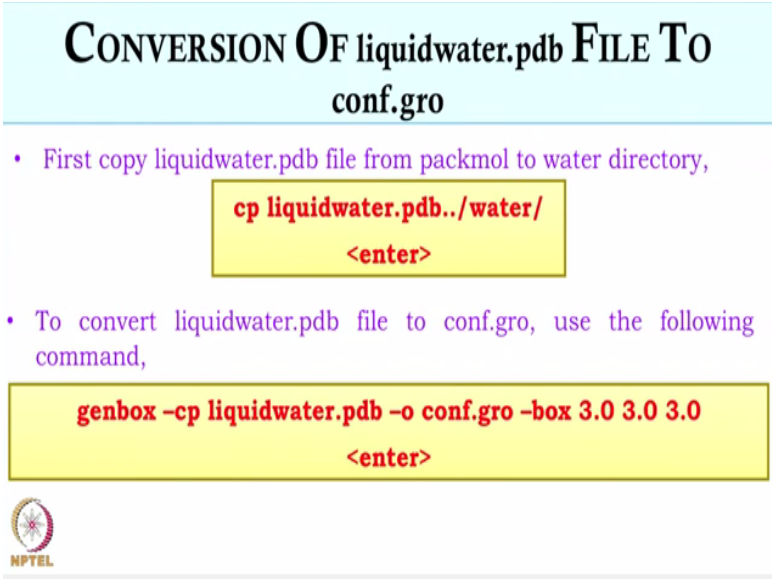
So, all these commands like g_rdf all that there are all in Gromacs directory; they all are in groamcs directory because if they are not there in the directory, you cannot run, nothing will happen exactly, so when, now you are putting all these files in water directory, you create 1, yeah when you run the final md rum, yeah do you run it from water directory or do you run; we run it from water directory.

Because water directory is anyways in the Gromacs directory okay so you have created that directory, so these water directory is the sub directory of Gromacs, yeah, it is the sub directory of Gromacs, so then that is how you make sure that everything that is there in Gromacs will be known in a sub directory of Gromacs, yes or suppose you have downloaded Gromacs in your documents or downloads directly, so it is not required at you have to have water directory in the Gromacs directory.

Since Gromacs is installed in downloads, correct, so you can create a directory in download folder, okay as well or let us say I create my directory in some other drive, I can always give a path wherever the Gromacs files are, yeah, so for that you need to have a script file correct, so which means the it will be more and more complicated if your files are kept all over the place, yeah, so it is best to have all sub directories inside the Gromacs, yeah.

Because that will be easier for the user to handle it, correct, correct, okay, so now once we have the water.pdb file, so we will convert into a gro file okay, to use it in Gromacs, okay.

(Refer Slide Time: 19:56)



And to do that first, we will copy liquid water.pdb from packmol to water directory and then will give the command to copy this cp liquid water.pdb to water and then okay but this means the water directory has already been created yeah, we did not have that command here, no for creating water directory, we just to have use mkdir command, right and now to convert liquid water.pdb to .gro file, we will us the following command.

In that command water; liquid water.pdb.. yeah should there will be a place after pdb, no otherwise it will; okay, cp space liquid water. pdb right, then you have to give I do not think there has to be a space because if there is no space it will the whole thing will be like path know, you are copying from one place to another place, so I think there should be a space, okay, so it is a good challenge to users whatever we give on our screen, when you practice that it is always a; they say know that is always a slip between the cup and the lip.

So, you make sure that whenever there is a problem you can always send an email but by and large you will be able to since you have no practice of almost 10 weeks, you will be able to correct these things, okay, yeah, so you have copied know to water directory and now will convert the pdb file to a gro file using the command genbox, so genbox space – cp then we have to give the input file that is liquid water.pdb then space –o which is the output file space output file name that is configuration.gro or conf.gro then space –box.

Then we have to give the box length, so suppose that –box 333 is not there, then what will happen, it would not take, it will just create a gro file and then the box lengths space which we have in gro file it will be 000 since we are not specify any box length, so if do not give the

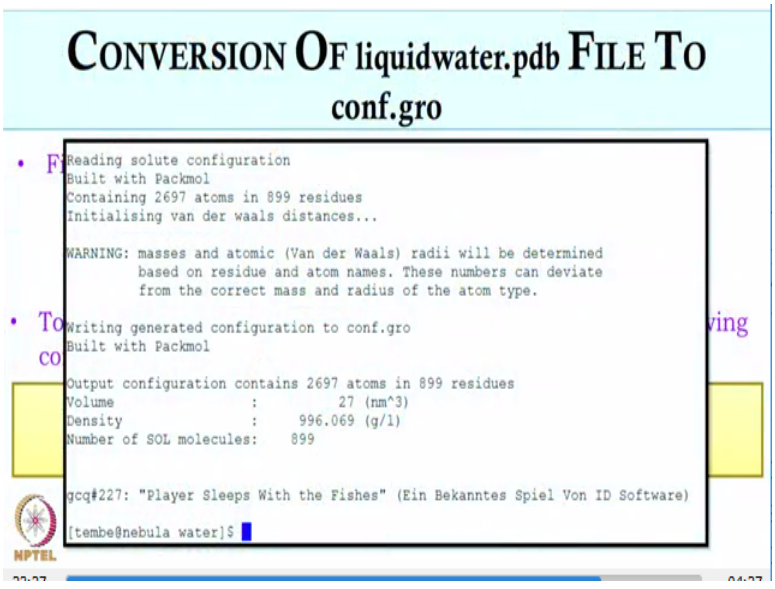
last line, I have to go back to the gro file and then you have to edit it and give that probably in the last line, yeah in the last line, okay.

So, now but best is to write the box length here itself but now this is still now, it has already gone into nanometres, yeah it has gone into nanometre because it is a gro file and in Gromacs the distance unit has to be in nanometre, now what is that-cp stand for in gen box? –cp is the input file, it is the flag for specifying the pdb file, okay and –o is the flag for the output file, so it has to be –cp not just -c, no not just –c is an input file for gro file, okay.

When we will used as conf.gro as an input for say, energy minimisation or full molecular dynamics simulation, then we will use –c space conf.gro, okay –c is the flag for as an input for gro file, okay. But here which means they have to be careful what options are given, yeah, if they just type gen box space –h; h is the help command, they will know what is the flag required for an input file or for an output file and all the flags will be listed there, yeah all the flags.

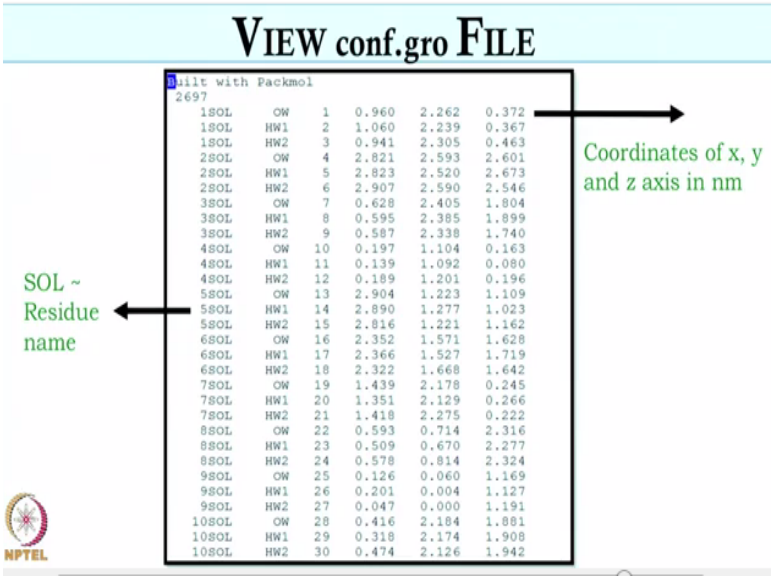
And there it will be written at this is an input file option, this is an output file option, so far which command which is the input file, yeah all of them will be written there. But –o is always for output, -o is always for output, okay. So, conf.gro this is the file again it is the text file which can be edited yeah it is the text file it can be edited, okay.

(Refer Slide Time: 23:27)



So, something like this will occur after we type the gen box command, so you have basically, edited the gro file now to show this or it just the output on the screen, it is the output on the screen.

(Refer Slide Time: 23:33)



Now I will show the gro file, no, go back to the output, let us see, so what are all these things now, so it is like, it is reading the solute configuration which is built with packmol, right, which contain this many number of atom, okay because there are 899 water molecules, so 899 *3, so there will be that much yeah, so they are initialising the Vander Waals distance and the total volume is 27 nanometre cube since the box length is 3 nanometre, right.

So, this is the density 996 gram per litre and total number of solvents molecule is 899, so this is the thing that will be displayed. Basically, all the input parameters which we are specifying okay and this the gro file so, SOL; it is the residue name then OW, HW1, HW2 this is the atom name and the last 3 line are the coordinate of x, y, z axis in nanometre, okay, so now you have created coordinates of 2697 molecules; atoms, yes, now that means 899 molecules yes.

So and your first column will give the site number yes that is if you go from 1 to 2697 yes it will grow from; no, it will first column is the solvent residue name, so first you see there is one sol, one sol, one sol, so first molecule, one solvent is first molecule, right, so it has 3 sides, so there is 3 one solvent, one solvent, one solvent, so 123 will correspond to one sol, one sol, one sol, those are OW HW1 HW2.

Those are the index number of the first water molecules correct, okay then 4, 5, 6 is the index number of the second water molecule, so this goes this way, so which means the left column one will be written 899 times, sorry, sorry, no, yeah one will be; 2899, the third column which is 1, 2, 3, 4 it will come 2697 time because third column is the atom or site number, it is here and the first column is the molecule number yes it is a residue number.

So by molecule you mean the residue, yeah, okay, so but do they always call all molecules are as residues, yeah because in Gromacs molecules are called as residue or residue number because show me that residue looks like some precipitate and all that no, because in Gromacs, this is the general practice to use it, okay and the extreme 3 columns on the right are the coordinates, yeah they are the coordinates of x, y, z axis in nanometre for each site for each atom.

So, now since when we started the first coordinate was 000 oxygen and 2 hydrogen, yeah, now these all have changed, yeah so which means once you run packmol, nothing may remain in the original site, no, okay, so for users you may just take any water molecule and calculate the length between oxygen and hydrogen yeah one and oxygen and hydrogen 2 make sure those bond distance is are the same as the original one.

Yeah I have already given it as an assignment to calculate the OH bond distance between first and fourth molecule or like that and if you have lot of time, on some evening, you can go to this coordinate file and calculate these distances make sure your water molecules are the correct geometry, it will be a large wok to do because there are 899 water molecules but they can also, no, now you see what you can do you have now leant Fortran, yeah, you can read this as an input file.

Yeah and then it can write; then calculate you can write a program to calculate the distances between all OH bonds, yeah, you see how many bonds are; 1 angstrom exactly and what is the average bond length in your water and standard deviation, lots of; see this data you can use as an input file for lot of calculation using your earlier programs, if necessary I will put that on websites, we will conclude this lecture right now.

And what we will do next time, next we will see the topology file of water and we will through the energy minimisation, so in the next lecture we will do the topology file and the

remaining step and finally go to the final calculation on water, yeah, okay, thank you we will conclude this lecture now and continue with this in the next class.