Algorithms for Protein Modelling and Engineering
Professor. Pralay Mitra
Department of Computer Science and Engineering
Indian Institute of Technology, Kharagpur
Lecture No. 14
Molecular Surface

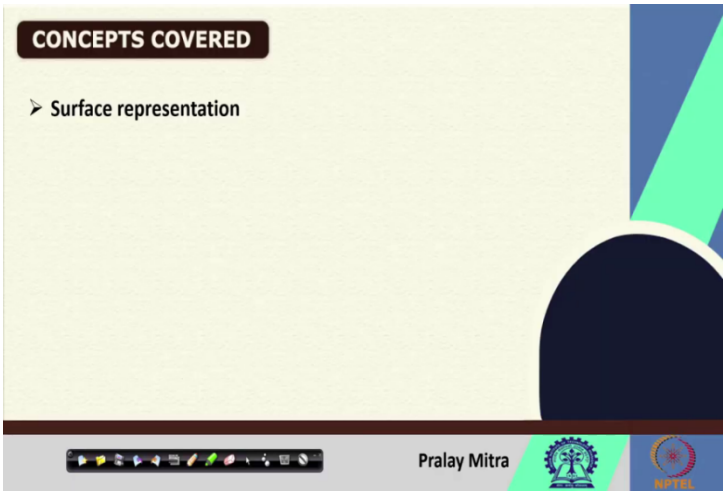
Welcome back, in this lecture we will discuss the molecular surface. We discussed this molecular surface in the context of surface complementarity, molecular recognition etcetera. There are different algorithms for determining surface complementarity, a few of them we discussed, and we will discuss some others also as the course will proceed.

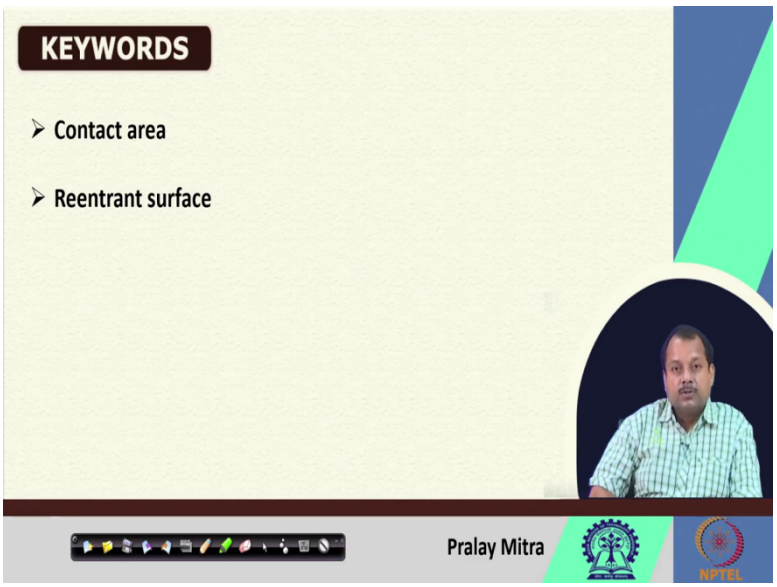
Now, it is time to discuss a little more in detail regarding the molecular surface. Whatever we have done is, digitize the protein molecule by putting it in a grid. We decided on the grid step and grid size because grid size will be determined by the size of the protein molecule.

After determining the grid step size, we decide whether one protein molecule's atom belongs to a grid cell or not. As you understand that deciding on the grid step size tells a lot. If the grid step size is very small then we are going for finer detailing, if it is large, then the problem is that we have to go for some thresholding. And because of the thresholding, you remember that it may be possible that one atom belongs to more than 2 or more 3 cells.

Today we are going to discuss two other techniques for determining or defining the molecular surface which is most widely used in this protein modelling and protein engineering. One is the Connolly surface and the other is the solvent-accessible surface area (SASA) or accessible surface area (ASA) as computed by the NACCESS. We shall discuss both, and in this context grid-based surface representation also.

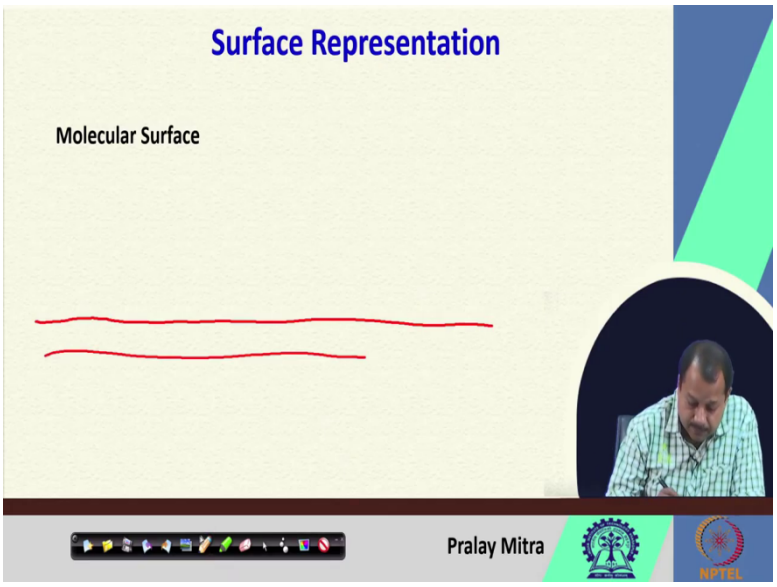
(Refer to Slide Time: 2:35)





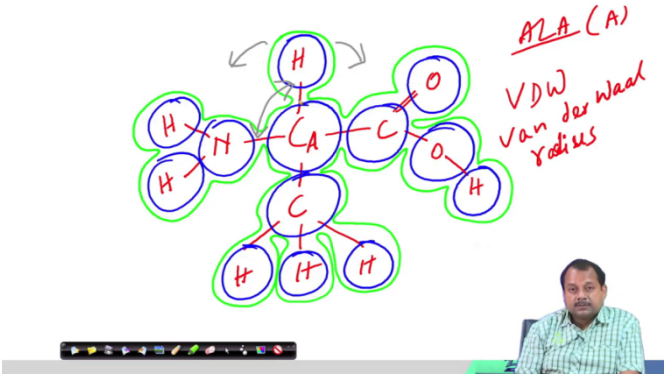
Today, we are going to cover the surface representation, contact area, and re-entrant surface.

(Refer to Slide Time: 2:54)



In the surface representation, let us first start with the molecular surface.

(Refer to Slide Time: 3:07)



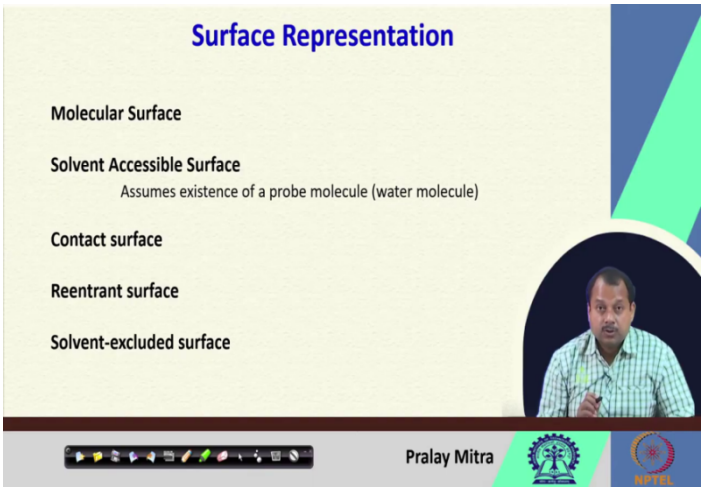
Let us assume that one protein molecule is there and that protein molecule contains only one amino acid – alanine (A or ALA). In this case, all the solid lines that you can see are the covalent bonds and whenever there is no such solid line that means there is no covalent bond. Usually, we assume that each atom is represented by its van der Waal radius or in short VDW.

If this VDW radius means, that this carbon is something like this, this carbon is something like this, this carbon is something like this oxygen is like this. Then I am putting in one circle but it will be a separate circle and the reason is that this is my O and then actually if I go this is my H now, it will be good surely, for this it will be for this, this, this, this, this so something approximately like this.

Here the blue colour circle indicates the atom represented by its VDW radii. Now, if it is represented by VDW radii then when I say the molecular surface if I use say another colour say green, it will be indicating something like this. Also when I am drawing this then one thing you possibly have noted is this is a bond. A bond is a bit stretchable which is not much of a problem.

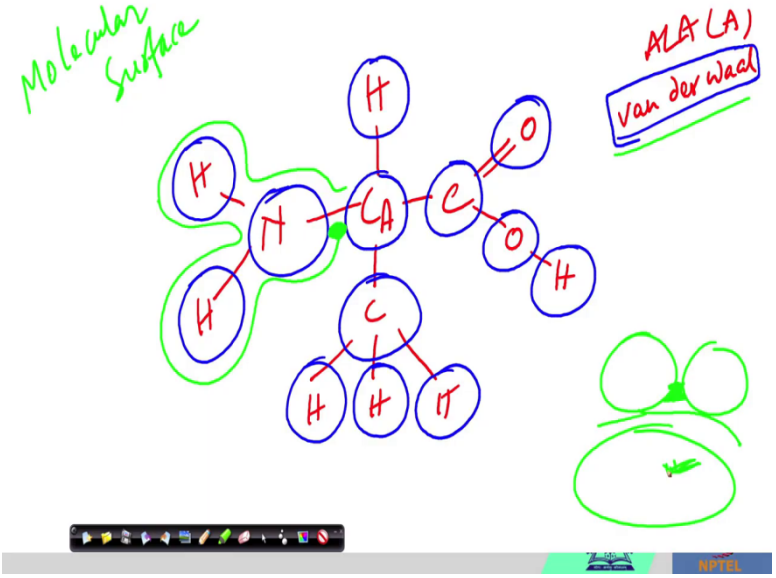
But, the angle may change because there may be some rotation, okay and because of the rotation what may happen is that, this bond angle may change. If this changes, then it will go this way or it will go this way or in two dimensional in three dimensional it can come towards me or go away from me and for all these changes actually, the surface to some extent will keep on changing. This is about the molecular surface. Now, this molecular surface is very difficult to compute and also we perhaps are not much interested in this molecular surface.

(Refer to Slide Time: 6:50)



So, what we are possibly interested in is the solvent-accessible surface, which assumes the existence of a probe molecule (here water molecule) and that water molecule when accesses the part of the alanine molecule, probably is of our interest not anything else. I shall come to that, but let me show you other possibilities - contact surface, re-entrant surface, solvent excluded surface, so these are the surfaces that we are going to discuss.

(Refer to Slide Time: 7:51)



Now, I have to draw it again, not a problem. If I start with a red colour as CA then I had C then double bond O, H, C, H, H, H, N, H, H, H, then here it is H. This is alanine and if I consider that van der Waal radius again using the colour blue this is this is, this is, this is, this is, and this is ok! This is my VDW radius. Now, I mentioned the molecular surface.

The molecular surface is this blue colour surface. If I can compute in some way then I can consider that as a van der Waal surface because this blue colour is for van der Waal radius. In

reality, they are not hard sphere ball-and-stick models. They are kind of a cloud, and because of that something like these will be there in between. That way what I can get is the molecular surface. Fine, that is as per the definition.

But, the question is that when I say the molecular surface or the surface, then what is my intention, why do I need to calculate the surface that is the question and if I can perhaps answer that question then accordingly I can define the surface also.

So, first of all, if I say this is my surface this green colour detail, if this is my surface so, is that of any use, perhaps not. What as of now, we have done is that we calculated the surface to know two things, whether there is an overlap of that surface area with another say molecule in general. I can take whether there is an overlap so, that we can calculate whether in that region they are interacting or they are placing side by side or not that is one point.

If say it is inside the water molecule, then whether the water molecule can access that particular atom or the surface of that particular atom or not. These two things - whether a biomolecule can interact with the surface of that molecule, and whether a water molecule can access the surface of that particular atom or the molecule. If it is, then that is of our interest otherwise say these green regions that I am marking here. This big green region perhaps is not accessible to any biomolecule because whenever say one biomolecule irrespective of whether it is a protein, DNA, RNA, or any small molecule ligand, the small peptide will come and they will try to interact. Perhaps they will not be able to reach this region. It is the same for the water molecule also because that water molecule or the biomolecule also has some radii. If I assume that those water molecules are biomolecules then they also have some VDW radii.

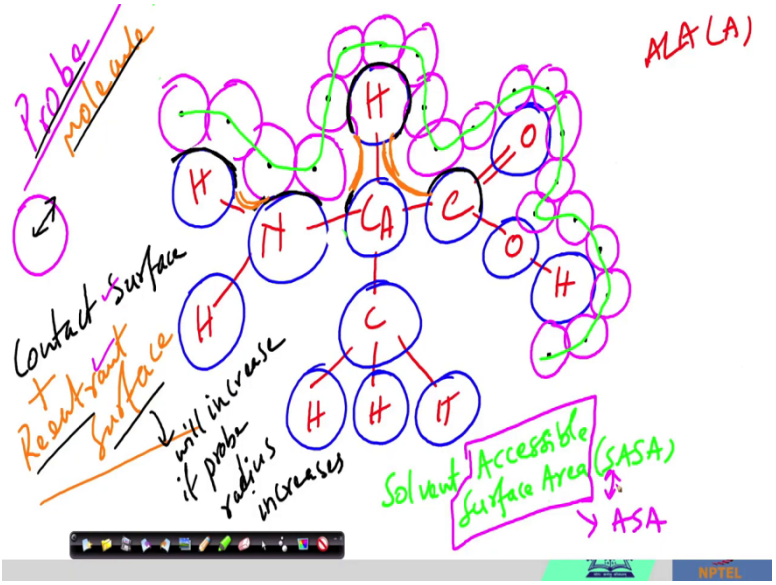
Now, let us assume two circles are given and another large circle is there, so they can go and touch up to these, but this region will not be accessible by this large molecule, if this region is not accessible by this large molecule, then that is not of our interest and I do not bother about that one.

That way we are not interested in this molecular surface, but this is the definition that you should know, even if we are not interested in this van der Waal surface (that blue colour surface). We are interested in the surface of a molecule which will have some overlap or contact with the other biomolecules or with the water molecules that are of our interest.

With that definition, we are going to define the surface which will be useful for us. So for that let me erase this green colour which is not actually of our interest. In this one I am erasing this green colour, molecular surface as I mentioned is not our interest erasing but you should remember the definition of the molecular surface, you should remember the definition of the van der Waal surface but the van der Waal radii is of interest. Since that suggests that you cannot penetrate inside one atom, so a few more things I have to erase from here. That Leonard-Jones potential that you remember or 6-12 potential, where there is a reciprocal part of r^6 and r^{12}, where one is the attractive force and the other is the repulsive force. Because of the repulsive force, I cannot go inside or penetrate any atom. OK!

Let me select the pink colour. Let us assume that there is one water molecule and that water molecule is rolling on this particular alanine molecule.

(Refer to Slide Time: 15:03)



This probe, let us consider as one sphere in 3D, in 2D it will be looking like a circle now, this sphere will keep on rolling on to this, so we need to roll, so it will roll like this. Although the radius of all my pink colour circles is not looking similar, please consider that they are of equal radius. These pink circles, that way it will go. When it will go then you can see two different things. Number 1 is that region where this pink colour circle or pink colour sphere in 3D is touching the blue colour circle or blue colour sphere in 3D which is called the contact surface because there is a contact. You can also see some of the regions where it cannot go inside. If I consider yellow or golden colour, there are some regions like this one, so this region or if I consider this region, in this region this probe molecule cannot enter.

Hence, this CA cannot be touched by this probe molecule because of the existence of this hydrogen and carbon which are very close, that situation may arise.

In this case, since hydrogen is of small size then perhaps it can touch a little bit, but may not be based upon the orientation of the H, there may be some situations where it cannot able to touch. That is called the re-entrant surface. One is the contact surface, where is probe molecule is touching the actual atom of the molecule and another situation is that when it is not able to touch the surface of any atom, but I am considering for the completeness of my surface calculation, why completeness?

If I draw this one then this is my contact area, partly this may be contact area this is my contact area completely this black colour and then this black colour, now see between these two black colours here, how do I connect, how the line will be continuous. If the line is not continuous, then I cannot define a surface with that one so continuity will be considered by this as if the probe molecule is there, then it will reach up to this one but not beyond that one.

That way, I can get this re-entrant surface. This contact surface plus this re-entrant surface will give you one surface, sorry I missed it with the golden colour. This contact surface and the re-entrant surface in combination will give one surface. That surface is the smallest surface region on the protein molecule within which a probe molecule cannot go or penetrate.

Now, you see that when I am defining in this way, then the re-entrant surface of this part, the re-entrant surface, or the portion of the re-entrant surface will increase if the radius of this probe will increase. This re-entrant surface will increase if the probe radius increases and vice-versa, if it will decrease, then what will happen? This pink circle is of the smaller size it can go deeper inside then the portion of the re-entrant surface will reduce and it will go to the part of the contact surface, so the contact surface increases and the re-entrant surface reduces.

Also, if the probe radius is infinitesimally small then it will be nothing but the molecular surface, small portion of the re-entrant surface will of course be there if it is not a dot but, that will be very small. If you want a better fitting for the molecular surface then you need to reduce the diameter or the radius of the probe molecule for the better fitment. But, if you go for the coarse level fitting, then you increase the probe molecule then the question is, what is the correct probe radius?

The answer is not known, it depends upon your application and of course, it depends upon your implementation, it depends upon the programme through which you are implementing

the probe molecule calculation. So, when you are implementing this probe molecule when you are implementing this surface calculation using this probe molecule, then definitely there will be a limit on the size of the probe molecule. It cannot go infinitesimally small. That is all about the contact and re-entrant surface.

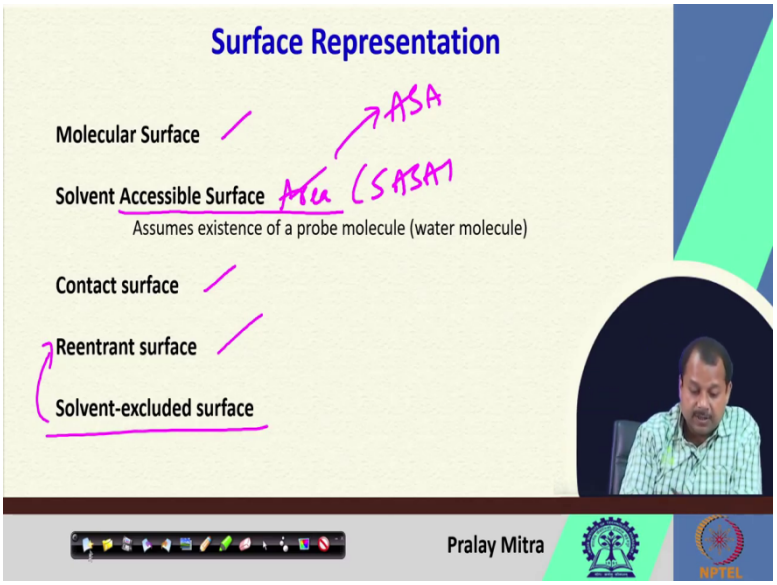
However, some algorithms suggest, if that touches and this is my touching area even if that is not of my interest, so what I will do instead of taking the sum of the contact surface and re-entrant surface, I will roll the probe on the protein molecule the same way and during that rolling I will look for the trajectory of the centre of this probe molecule. The trajectory of the centre of this probe molecule - let me take green because green is not there. If I take that then I will get one surface something like this, so the surface area is covered by this green line. I consider this green line, then I will get another surface area that usually is called the solvent accessible surface area SAS or sorry not SAS solvent accessible surface SASA sometimes it is also only called the ASA, which means they consider this then to be ASA solvent accessible surface area.

Solvent accessible surface area means the surface area calculated by the trajectory of the centre of the probe which rolls over there. This is all about contact area or contact surface re-entrant surface solvent accessible surface area or accessible surface area both are used interchangeably.

Then the final question remains if the probe molecule size reduces then it will be better fitting and if the probe molecule size increases, it will be less fitting and definitely because of that fact, the surface area computation will change. Also never it is a good point to mention that the accessible surface area and the surface area computed by the contact surface plus the re-entrant surface will be the different accessible surface area that is always higher compared to the contact surface and the re-entrant surface.

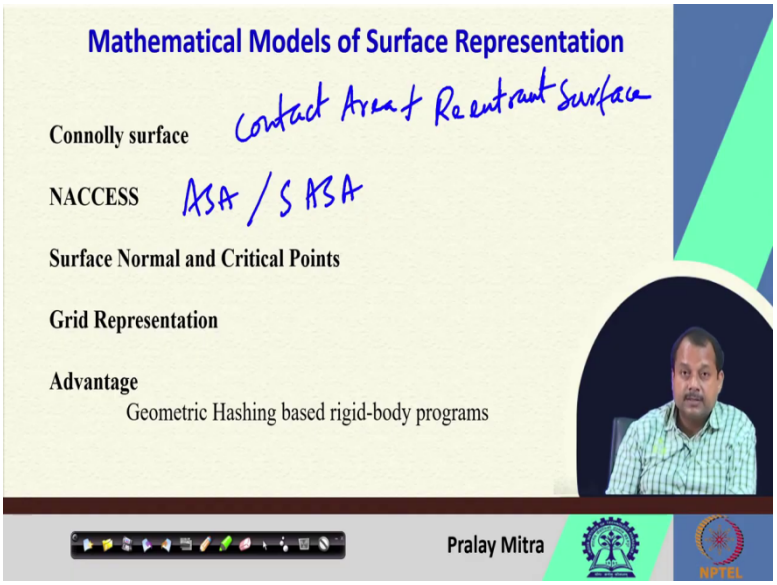
But, even whether it is a contact surface or the re-entrant surface or accessible surface area depending upon the probe radius the value will change, so what is the good number usually by default what is used is 1.4Å. This is used assuming that the probe radius for the water molecule is 1.4Å. You can always reduce that one or increase that one. If you increase or decrease there will be a limit based upon the algorithm or the implementation you are using, so be careful about that.

(Refer to Slide Time: 27:25)



We discussed the molecular surface then the contact surface, the re-entrant surface, and the solvent-accessible surface area (SASA) or only the accessible surface area (ASA) both are used synonymously. The solvent-excluded surface is the re-entrant surface because it excludes the solvent, the solvent cannot enter between two spheres then this region and this region is not able to touch. So, that is the solvent excluded surface area.

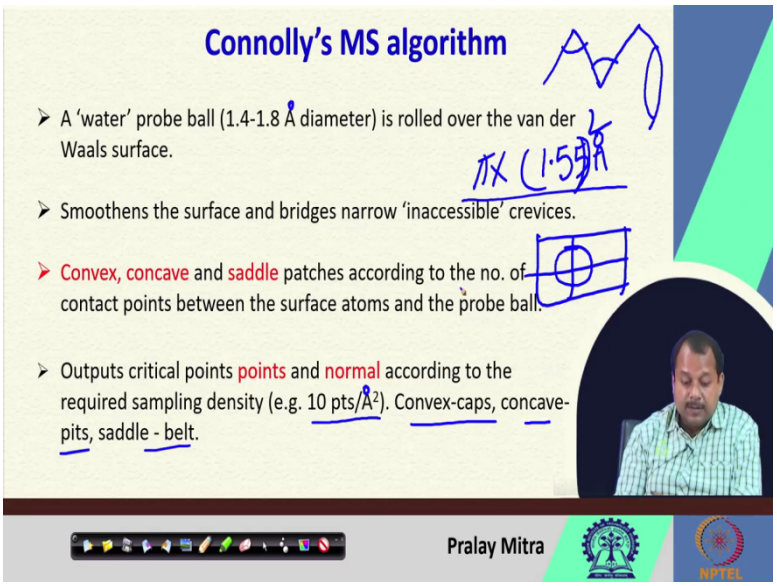
(Refer to Slide Time: 28:12)



Connolly developed one algorithm based on the solid angle and normal point calculation on the surface and he computed the contact area plus the re-entrant surface. NACCESS is another programme. Connolly surface is a little bit computation intensive whereas NACCESS is very fast and NACCESS computes the accessible surface area or ASA. This computes ASA or SASA - both are the same.

Apart from surface normal, critical points are also being computed. These surface normals and critical points are similar to the Connolly surface grid representation that we have discussed. We also see that using the grid representation we can compute the surface complementarity or the surface matching. Surface complementarity is a separate issue that will come later and we also see the advantage of this grid representation in geometric hashing based rigid body docking as well as the brute force method where we exploited the fast Fourier transform to speed up our computation.

(Refer to Slide Time: 29:58)



In Connolly's MS algorithm, a water probe is a ball of 1.4Å to 1.8Å diameter. Here there is a small typo, it will be Å diameter. When it is rolled over the van der Waal surface then it smoothens the surface and bridges a narrow inaccessible service crevices.

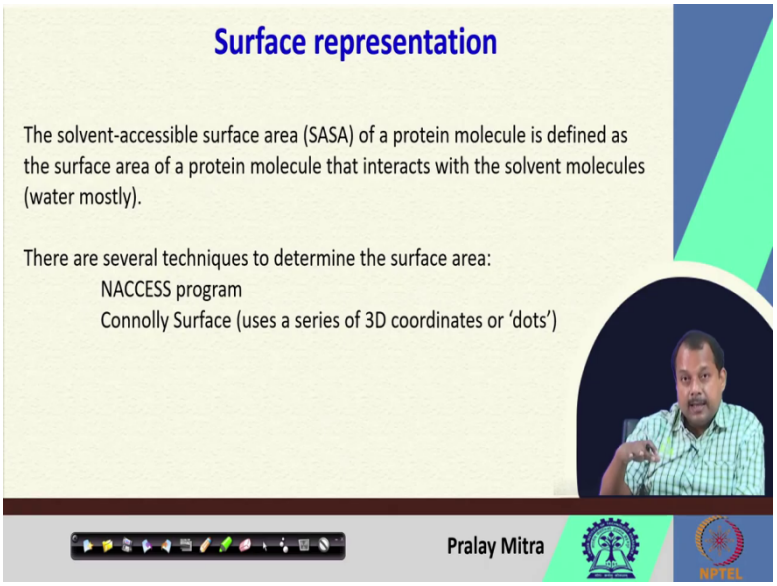
Then, convex, concave and saddle point patches according to the number of contact points between the surface atom and the probe ball is being identified and it outputs critical points and normal according to the required sampling density so that you can mention. Usually 10 points per Å. Again here it will be Å², 10 points per Å² it is being computed these convex gaps

concave pits and saddle belt so that you can think of this, so this point, this point and this point.

Based upon that one the saddle belt or convex gap or concave pits, so that may vary but usually 10 points per angstrom square is there. Now, from here also understand that it is now not one particular atom because if I consider one particular atom and say for hydrogen or say Sulphur or something, if I just assume that it is 1.55Å, then the total surface area will be $\Pi \times r^2$.

Now, this is the total surface area of 10 points divided by this $Å^2$ which means corresponding to one atom it is not only the one point that we have considered. Initially, for the brute force technique or grid representation it will be at most 4 in 2D that we consider that a grid cell is like this and the atom is placed here it is not also like that, so it is much more that one in Connolly's MS algorithm. Also, you have to remember based on that one that computation time will increase.

(Refer to Slide Time: 32:20)



Anyway, the solvent-accessible surface area of a protein molecule is defined as the surface area of a protein molecule that interacts with the solvent molecules - water mostly. That is the definition of the SASA or ASA in general. There are several techniques, but the most widely used for ASA calculation is the NACCESS because of its speed, but the Connolly MS algorithm or Connolly surface algorithm is also there and it provides more direct information regarding the surface normal and specifically when you are interested to work with the

surface normal and you wish to go in more detail or dense point representation then Connolly surface is perhaps the right one for your purpose.

That’s it regarding this molecular surface representation and now one more algorithm we will discuss next that will exploit the knowledge of this molecular surface. Thank you very much.