MOTIFS AS ELEMENTS IN PROTEIN STRUCTURE AND FUNCTION: A PROPOSAL

by

Carlos Corredor*

Resumen

Corredor, C.: Motifs as elements in protein structure and function: A proposal. Rev. Acad. Colomb. Cienc. **22** (85): 519-526. 1998. ISSN 0370-3908.

En este artículo pretendemos mostrar que los motivos en las proteínas se comportan como unidades individuales cargadas colocadas en forma muy precisa en un espacio acuoso de tal manera que para cualquier sistema proteína - agua de solvatación exista un máximo de atracciones y un mínimo de repulsiones entre los motivos que constituyen la proteína. Cualquier región en la superficie de una proteína formada por las porciones hidrofílicas de diferentes motivos genera un campo electromagnético no homogéneo que fija moléculas de agua. Un sustrato es cualquier molécula que genera un campo electromagnético complementario a la región activa de la proteína. La interacción entre una proteína y su sustrato se puede concebir como la neutralización de dos campos electromagnéticos que al reaccionar producen un nuevo sistema con una energía interna mayor que la de la proteína o de su sustrato. Al tratar de encontrar un nuevo mínimo de energía interna y, por consiguiente, un sistema estable tiene que ocurrir un cambio conformacional que simplemente implica una reacomodación de los motivos para obtener un máximo de atracciones y un mínimo de repulsiones entre motivos. Es este cambio conformacional lo que causa la actividad de las proteínas.

Palabras claves: Proteínas, estructura, función.

Summary

In this article we propose that protein motifs behave as single charged entities precisely located in a water space in such a fashion that for a whole protein – solvating water system, there is a maximum of attractions and a minimum of repulsions between the motifs making up the protein. Any region of the surface of a protein made up of hydrophylic portions of different motifs generates an uneven electromagnetic field which fixes water molecules to it. A substrate

* Miembro de Número, Academia Colombiana de Ciencias Exactas, Físicas y Naturales. Dirección actual: Decanato, Facultad de Ciencias, Universidad Javeriana, Bogotá. E-mail: ccorredo@javercol.javeriana.edu.co

* Current Address: Departamento de Nutrición y Bioquímica, Facultad de Ciencias, Universidad Javeriana, Bogotá, Colombia.

is a molecule that generates a complementary field to the active region of the protein. Protein – substrate interaction can be thought of as two electromagnetic fields neutralizing each other upon reaction to make a new system with a larger internal energy than either that of the protein or that of the substrate. In order to find a new minimum internal energy and therefore a stable system, there must be a conformational change which simply means that the motifs will rearrange in such a way that a new condition of maximum attractions and minimum repulsions is obtained. It is this conformational change that causes known protein actions.

Key words: Proteins, structure, function.

Secondary structure has long been accepted to be due mainly to two features of the primary or back-bone structure: a) the tetrahedral angles around the a-carbon of the aminoacids which will change the direction of the axis of the chain in a precise fashion and b) the hydrogen bonds that are established between a carbonyl oxygen of a given peptide bond and a conveniently placed amino group of another peptide bond, thus located because of a). In this manner, two types of structures are possible: alfa chains and beta pleated sheets. It could be argued that once a given type of secondary structure starts to form it could theoretically encompass the whole of a protein. This, however, has not been found in nature. Even the first protein whose tertiary structure was elucidated, myoglobin, which is made up of alfa helices, is not a *single* a-helix.

With the elucidation of more and more tertiary structures of proteins, it is now apparent that most native proteins are made up of successive a helices, b-strands or mixtures of a-helixes and b-strands in a precisely determined sequence and three dimensional structure called motifs. It has also been recognized that motifs can group in larger structures, domains, having an organization, and some times a function, of their own.

There are now several recognized motifs and we now know how they are located in several proteins. In fact, in this paper we will argue that the actual building blocks from which functional proteins are made up are motifs, rather than aminoacids, even though the motifs are themselves made up of aminoacids.

The proposal that motifs are the basic building blocks is more than just a refinement of our knowledge of protein structure. We will argue that this notion can actually explain protein function in a rather simple and elegant way.

Let us consider the simplest of motifs: an a-helix and a b-strand in a water milieu. Established knowledge allows us to predict that both the helix and the strand will maintain their structure in water at temperatures between just above 0°C and less than 60°C, that is, at temperatures normally compatible with life. This means that, even though any hydrogen bond between a given carbonyl oxygen in a peptide bond and the amino nitrogen of another one breaks and reforms at a given rate at any given temperature, no new bonds will be established after each breaking instance with different peptide oxygens or nitrogens or with water in this temperature range. Therefore, aminoacids linked in an a-helix or a b-strand will behave as a single structure, with little or no individual aminoacid mobility away from the secondary structure's axis, apart from limited rotation of the aminoacid residues and stretching of interatomic bonds.

Let us now remember a second concept: neither the a-helix nor the b-strand are simple successions of peptide bonds. In fact, the a-carbon that connects them is linked to a so-called aminoacid residue. And any one of the 20 residues which make up normal proteins occupy a given space determined by the van der Waals radii of the atoms that conform it, space that must project from the axis of the secondary structure.

A rapid inspection of space filling models of different amino acid residues shows that, even though generally speaking, the space which most of them occupy is roughly similar, each amino acid surface is different from that of other amino acids. Furthermore, there are a few which are quite bulky, while others have long hydrocarbon chains. If we now inspect a space filling model of a peptide, we can ascertain that its surface will be determined by the outermost atoms of the amino acid residues which make it up and will, for the most part, hide the backbone.

The van der Vaals radii of atoms indicate the ground state outer orbits of the outermost electrons, i.e. the limits between individual atoms which can not be occupied by any other atom at normal temperatures. This limit can only be overcome when cooling a substance to a fraction of a degree above 0°K to give us a Bose-Einstein plasma, which, of course, is not the case with real life proteins. The van der Vaals radii, then, allow us to construct a model of the topological surface of a peptide.

However, it is well known that both basic and acid amino acids carry charges in water soluble proteins at physiological pH´s. This means that for a better approach to the understanding of protein interactions, two more elements must be taken into account: a) the protein´s solvent and b) the electro magnetic field radiating from the topological surface of the atoms. For this discussion, we will only consider water as a solvent, but similar arguments can be used *mutis mutandi* when lipid is the solvent.

Let us now give our attention to the electromagnetic field radiating from the topological surface of the peptide. We are accustomed to think only of whole negative or positive charges such as the ones in acid or basic amino acids. However, it is well accepted that bonds between atoms of different electronegativity such as OH, SH, NH, CO are permanently polarized and carry a fractional negative charge, d^-, on the more electronegative atom and fractional positive charge, d^+, on the other one. This polarization will also generate an electromagnetic field, albeit smaller than the one originating in whole charges, which will also radiate from the topological surface of the peptide.

Quantum mechanical considerations will lead us to suppose that in any given bond between two atoms with a similar electronegativity the greatest probability of finding the binding electrons is at the center of the axis between the two atoms. However, there is a small, but finite, probability that the binding electrons may in a given instant be closer to one than to the other atom, thus producing a transient polarization. These covalent bond polarizations will, on the whole, cancel each other out, unless two chains are so close together that these very minute electromagnetic fields can actually interact and produce attractions. This is what happens in the case of hydrocarbons with chains above 6 carbon atoms which are liquid and, if long enough, solid at room temperature.

With these considerations in mind, we can then argue that, superimposed above the topological surface of any given peptide, there is an electromagnetic field generated by unit charges from the acid or basic amino acids, by the fractional charges due to polar amino acids and by the fractional, reversible, minute charges due to bond polarizations occurring near the surface. Such an electromagnetic field can be visualized as a map of contour force lines.

In order to draw such a map we must consider what force would a test charge moving parallel to the protein surface feel. According to Coulomb´s law, the force between unit charges is directly proportional to the product of the charges and inversely proportional to the square of the distance. In the case of polar bonds the force is inversely proportional to the cube of the distances and in the case of reversible dipoles (London forces), it is inversely proportional to the fourth or even the fifth power of the distance. In figure 1 we show how forces felt by a test charge would decline with the distance from the protein surface, that is, from the van der Waals radii of its outermost atoms. For calculations we arbitrarily assumed charges to be 1 electron equivalent charge units (e.e.c.u.) for unit charges, 0.3 e.e.c.u. for polar bonds and 0.1 e.e.c.u. for reversible dipoles.

As can be observed, at a given distance only unit charges exert a significant force. However, as the distance decreases, the force due to dipoles and even that due to reversible dipoles become significant. What this means in terms of the protein surface considered is that there are attractive or repulsive forces of different magnitudes on that surface which will be able to affect any near-by molecules in different forms depending on the distance between the interacting surfaces.

The closest molecules to any given protein will, of course, be the solvent molecules. As initially indicated, we will only consider water molecules in this discussion. Water is a dipole with d^+ and d^- fractional charges on both hydrogens and paired unshared electrons respectively. This means that water molecules will solvate both the unit charges of acid and basic amino acids and the OH or SH dipoles of polar ones. On the other hand, the force between water molecules is stronger than the one that could be exerted by the reversible dipoles of neutral amino acids. This means that water will not "wet" the non polar surfaces, and in fact, will tend to actually confine them closer to the peptide backbone.

We must now remember that the different types of amino acids are not evenly distributed along the primary chain. This means that, in any given motif, there is a preponderance of polar and charged aminoacids on the surface closer to the water solvent which will be solvated, while most of the non polar ones will be located on the other surface which will interact preferentially with a similar surface of another motif. In this manner the protein tends to adopt a close packed tertiary structure with a hydrophobic core and a water solvated surface.

Now we must recall that the protein surface is actually made up of the amino acids which are in contact with the solvent, independently of the position a given amino acid occupies in the peptide chain sequence. This

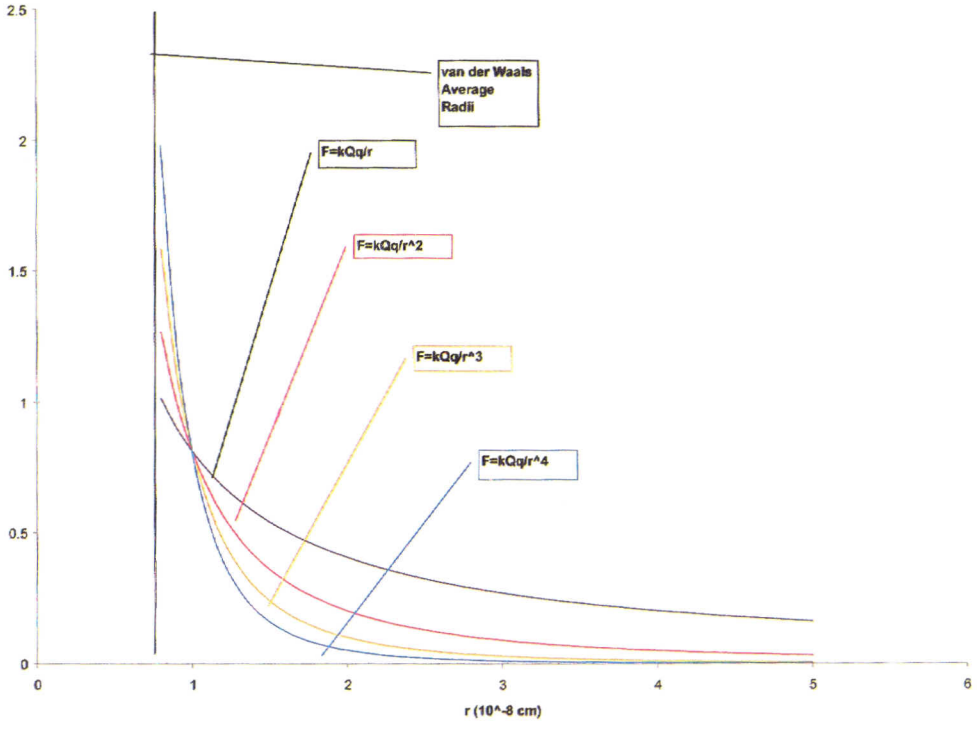

Figure 1. The force in electron-volts felt by a test charge at any distance, r, from a given point at the protein surface was calculated by the appropriate equation where k = electrostatic constant; Q and q = charge in electron charge equivalents and distance, Å.

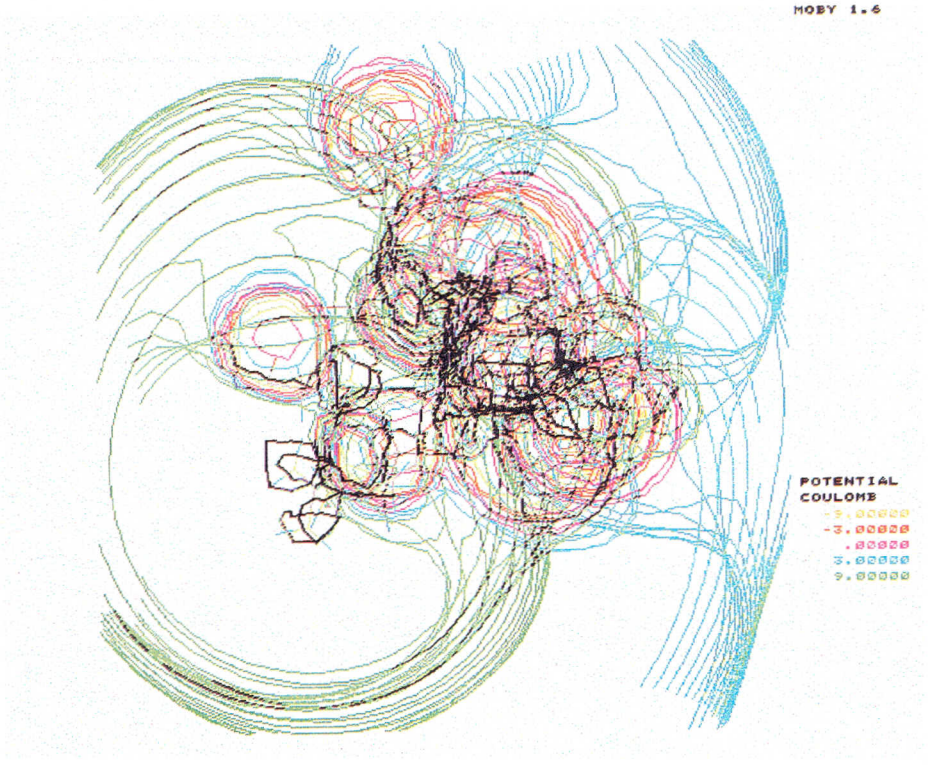

Figure 2. Isopotential map of the force field generated by the peptide GVPINVSCT in water solution using MOBY graphics program with AMBER potentials.

has two implications: 1) that amino acids which are adjacent on the surface may be quite apart in the backbone sequence and 2) that a change in protein conformation may result in a change of the amino acids occupying any given region of the surface. The latter also means that a new surface electromagnetic field will be generated at those sites at the surface where changes have occurred. We must emphasize that if, as we propose, the protein building blocks are motifs, the amino acids which will appear at or disappear from the surface pursuant to a conformational change will depend on the way motifs will rearrange themselves as a new internal energy minimum is sought. This conception also means that it is possible that any given chain of motifs might have more than one energy minimum and, therefore, spontaneously adopt more than one conformation.

Due to the nature of the water dipole, the orientation of the axis of the water molecules will change depending on whether it is interacting with a positive or a negative charge or a given fractional charge on the amino acid residues. This actually means that there is an uneven layer of fixed water that, due to the orientation of its molecules, grossly reproduces the electromagnetic field generated by the surface amino acids some distance into the body of undisturbed water.

We can now consider a functional protein made up of individual motifs connected by small unstructured amino acid chains that can be thought of as connecting pins. Let us further think of each motif as a rigid structure made up of a lot of minute magnets of different strengths attached around the axis of the motif. It can then be easily imagined that the motifs will try to accommodate in such a manner that most opposite pole magnets in different motifs will face each other. Let us change this rather naive picture for one in which instead of minute magnets we have a fairly rigid motif structure on whose surface there are whole or fractional charges of different strengths, each generating a radiating field. Again, it can be argued that motifs will so arrange themselves in space, that the system **protein-adsorbed water** will adopt a particular and unique conformation in which the alignment of positive in front of negative whole or fractional charges will be such that attractions will be maximized while same sign charges will tend to be as far apart as possible and therefore, the repulsive forces will be minimized.

We could, then, write that the internal energy of the system will be minimal when the summatory of all attractive forces is larger than the summatory of all repulsive forces.

$$E_{min} = \frac{\Sigma \text{ attractive forces}_{aa\text{-}aa+aa\text{-}water}}{\Sigma \text{ repulsive forces}_{aa\text{-}aa+aa\text{-}water}} >>>$$

where aa-aa and aa-water indicate aminoacid-aminoacid and aminoacid-water interactions.

Let us now consider the interaction between a protein and its natural substrate. Koshland argued that in the case of enzymes, the substrate and the active site fit each other like a lock and key, and that there must be at least three points of contact to account for differential union with estereoisomers. This concept has been extended to the proposal that rather than a lock and key fit, the interaction is more like that of a hand and glove, thus introducing the idea of fit between corresponding topological surfaces rather than between individual charge points.

We would now like to further extend the concept by suggesting that the interaction -and fit - occurs between complementary electromagnetic fields which are generated by the atoms at the topological surfaces of the protein and of its substrate. In this fashion, the interaction will not be limited to a few correctly placed aminoacids, but rather to a whole patch of electromagnetic attractive forces on a specific area of the surfaces of both the protein and the molecule, large or small, with which it will interact.

To better visualize this concept, let us consider a small arbitrary peptide. The electric field generated by the aminoacids can be calculated as shown in the appendix. We can then draw lines joining points of equal electric force to give us a map such as the one depicted in Fig 2. For all practical purposes, the force at any one point in any line drawn would be what a test charge located at that point away from the peptide would feel. As can be appreciated, rather than peaks of single charges, what we now have is a field of varying intensity over the whole surface of the peptide.

Let us now locate an appropriate motif with its own electromagnetic field near the depicted peptide and observe the manner in which the two fields interact. If the electromagnetic contours fit, that is if, for each positive field in one surface, there is a negative one on the other one and viceversa, the two surfaces will attract one another and the two molecules will get close enough at those surfaces, and only at those particular electromagnetically complementary surfaces. Of course, the fit can range from absolute to rather loose, and this will be reflected in different affinity constants of the reactants, from very large for the former to rather small for the latter.

Let us now consider a water system in which only those two types of molecules are present. We can argue that thermal motion will cause collisions between molecules and that if the collision occurs in such a fashion that it will involve the electromagnetically complementary surfaces, it will be a productive collision and the two molecules will now form a binary system whose behaviour we will explore below. On the other hand, if there is no electromagnetic complementarity between the fields above the reactant surfaces, there will be an elastic collision between the two and they will move apart.

If a binary system protein-substrate is formed, then we can expect neutralization of the interacting electromagnetic fields. Such neutralization will alter the internal energy balance in both the protein and the substrate. This will necessarily determine that a new minimal energy state will have to be achieved. To do so, the different motifs making up the protein will have to rearrange themselves in a new way and in so doing they will drag with them the substrate. We will then have a binary system in which the conformation of the parts will have suffered modification. Such modification will then explain the particular protein activity: catalysis, transport, recognition, movement, etc.

To explore this concept further, let us think of a protein in a physiological solution with its substrate at a given distance. Let us now assume that in their random thermal motion they approach each other. As they get closer, the larger charges on both protein and substrate orient and modify the direction of the approach. As the molecules get closer together, the electromagnetic fields due to the dipoles acquire importance and there is a rectification of the direction of approach until both complementary electromagnetic fields are aligned and the two molecules are joined by the neutralization of the complementary fields.

However, previous to actual interaction, intervening water molecules must get out of the way. This might occur if, at the same time when the complementary electromagnetic fields of protein and substrate begin to neutralize each other, the attraction of the protein surfaces for the water molecules diminishes, thus allowing them to incorporate themselves in the water structure and out of the way of the approaching interacting molecules. During the approach, both the protein and the substrate interacting surfaces maintain their relative topologies for a time, but as they get closer together there is an induced displacement of the fields that allows for a better fit between both surfaces

One important consequence of this conception is that it does not matter what amino acids make up the given motifs participating at the protein joining surface as long as the electromagnetic field they exert as a whole is approximately the same in all cases, thus conserving for the most part its complementarity with the appropriate substrate. For this to hold, it is of paramount importance that the internal energy of the system remains approximately constant when there are amino acid substitutions at any given point. Otherwise, a new energy minimum will be sought through a conformational change and the protein will exhibit a different or no activity.

Amino acid substitutions that do not alter significantly the internal energy of the protein or the surface electromagnetic field would account for what are known as mute mutations in which the affinity constant of the protein for the substrate will change slightly, but the reaction is otherwise unaffected.

The consideration of the protein-substrate interaction as the neutralization of two complementary electromagnetic fields can also explain why some proteins are rather unespecific in their interaction with certain related substrates. Such is the case with hexokinases which, even though they have a higher affinity for a given hexose, can actually ligate and catalyze the phosphorilation of most hexoses whose dipoles are similar but their location in space is different giving a similar electromagnetic field but with varying intensities as the dipoles change their disposition in space.

An interesting example of the meaning of this proposal can be seen in the way that neutrophiles phagocyte different invading microorganisms but do not attack self cells. It can be assumed that neutrophiles cannot have the possibility to recognize any eventual invader. However, we can also assume that they have receptors for an "activated" antibody, that is, one that has been attached to an antigenic determinant in, for instance, the surface of an invading bacteria. In this manner, the neutrophile will recognize and join a single electromagnetic field located in all "activated" antibodies and will recognize and phagocyte it - and anything attached to it - only when the antibody has been activated.

Let us illustrate this phenomenon by considering a bacterial infection by a non-encapsulated organism. No matter the strain of invading bacteria, there will be a particular clone of antibodies which will recognize the electromagnetic field generated by one or more of the bacterial surface antigenic determinants. The site of recognition of the antigenic determinant is the variable region

of the inmunoglobulin which is composed of portions of both the light and heavy chains. This region generates an electromagnetic field which is exquisitely complementary to that of the antigenic determinant so that it will interact with it neutralizing the whole site of contact. As explained above, this will cause a change in the internal energy of the system which will now seek a minimum level through a change in conformation, that is, in the relations between individual motifs making up the antibody molecule. This change in conformation will result in bringing up to the surface a region previously buried in the Fc fragment, that is in the constant region of the inmunoglobulin. This region, in turn, will generate a unique electromagnetic field which is complementary to a receptor site in the surface of the neutrophile. When these two fields interact neutralizing each other, the system will seek a new energy minimum through a conformational change that will result in pulling of the surface receptor protein of the neutrophile from the inside to form a phagosome, drawing with it the bacteria attached to the antibody bound to the surface receptor site.

If one proposes that all Fc fragments in any given class of antibodies are the same, any antibody-antigenic determinant electromagnetic fields neutralization will result in the same conformational change at the Fc part thus allowing neutrophiles to phagocyte any type of foreign body, as long as there is an antibody against it. Similar considerations can be made for hormone-receptor, transporter-transportee, enzyme-substrate or any other kind of interaction between a protein and its substrate leading to a particular action.

In conclusion, we propose that motifs are the individual secondary structure elements which behave as single charged entities precisely located in a water space in such a fashion that for the whole protein – solvating water system there is a maximum of attractions and a minimum of repulsions. Any region of the surface of the protein generates an electromagnetic field which fix water molecules to it and extends through them the field into the body of surrounding water. A substrate is a molecule that generates a complementary field to the active region of the protein. Protein – substrate interaction can then be thought of as two electromagnetic fields neutralizing each other upon reaction to make a new system with a larger internal energy than either that of the protein or that of the substrate. In order to find a new minimum internal energy and therefore a stable system, there must be a conformational change which simply means that the motifs will rearrange in such a way that a new condition of maximum attractions and minimum repulsions is obtained. It is this conformational change which causes any protein effect when in contact with is substrate.

APPENDIX

*Leonardo Lareo**

In order to calculate the electromagnetic contours of a motif as described in the present article it is to be noted that the complete mathematical description of a molecule includes relativistic quantum mechanics considerations. However, there is not a quantum mechanical theory for this particular purpose. Therefore, we will start with a non relativistic description of Schrödinger equation:

$$H\Psi(R,r) = E\Psi(R,r)$$

where H is the system hamiltonian and Y is the wave function which depends on the nucleus (R) and electrons (r) coordinates. This equation is far too complex for practical use, so one must make a series of approximations. The Born-Oppenheimer approximation, first proposed in 1927, separates the electrons movement from that of the nucleus through two independent equations. The first one describes the movement of the electron:

$$H\Psi(R;r) = E\Psi(R;r)$$

which depends only parametrically on the nucleus position.

This equation defines $E(R)$ as a function of the nucleus coordinates. This energy is usually referred to as *surface potential energy.* The second equation describing the nucleus movement above the aforementioned potential energy, $E(R)$, is

$$H\Phi(R) = E\Phi(R)$$

This is the fundamental equation of quantum dynamics, but since the nuclei are relatively heavy objects, the quantum mechanical effects are quite small and the equation can be replaced by Newton's classical mechanics equation.

$$dV/dR = md^2R/dt^2$$

Solutions to this equations comprise what is now referred to as molecular dynamics.

Potential V is an expression of the surface energy potential and is called *force field* which can be empirically described. The coordinates of the actual atoms in a molecule combined with the force field are an expression of the molecule's energy.

Energy values and graphs of the equipotential surface of a protein can be calculated from different models for the force field. Differences between different models are due to the number of factors taken into account. However, generally speaking, they consider energies due to oscillation of bond angles, bond length, dihedric angle oscillations, hydrogen bonds, van der Waals interactions, electrostatic coulombic interactions, etc. In this article we used the AMBER model of force field and the graph was made under MOBY.