Molecular Dynamics Simulations of Biomolecular Systems

Tadeusz Kuliński

Institute of Bioorganic Chemistry, Polish Academy of Sciences
Noskowskiego 12-14, Poznan, Poland

E-mail: tadkul@ibch.poznan.pl

Introduction

During the last decade the main goal of molecular biology became to understand biological function in terms of structure, interactions and processes at the microscopic level. Some experimental techniques such as X-ray crystallography, NMR spectroscopy, time-resolved fluorescence spectroscopy are routinely used to provide an atomic picture of the structure and mobility of biomolecules like proteins, peptides, nucleic acids and their fragments and components. More flexible molecules are less accessible to structure determination by these methods. The most of the spectroscopic techniques yield the information about dynamics limited only to the groups of atoms, fragments of the molecules, but not for all biomolecule. Energetic information may be obtained experimentally for whole systems by differential scanning microcalorimetry, but cannot be measured at the atomic level. Due to the limitations of experimental techniques, the characterisation of a biomolecular system at the atomic level is incomplete. This incomplete molecular picture makes difficult to understand the link between the biological function of the molecular system and its molecular structure flexibility, conformational dynamics and interactions.

A new alternative way to study biomolecular systems at the atomic level is the simulation on a computer. Despite the extremely fast progress of the computing power available such calculations have to be based on a compromise between the complexity of the description of the molecular system, the number of atoms included in the system and relative numbers of computational cycles required for reliable description of the processes investigated.

We briefly present the physical basis and strategies of computer approaches utilized for the calculations of molecular properties. Then we present two examples of the different complexity to show the general methodology and procedures used for the simulations of biomolecular systems. One of them concerns the predictions of the enzymatic properties of genetically engineered protein, second one describes the analysis of different properties of ribonucleic and deoxyribonucleic acids on the level of its constituents.

Theory and strategies of computer molecular simulations

The mathematical description of a molecule including quantum mechanical and relativistic effects is very complex. Molecular mechanics and dynamics of complex molecule have to be based on empirical data that implicitly incorporate all the relativistic and quantum

effects, neglecting its explicit description. The description of the molecular system is based on the nonrelativistic Schroedinger equation:

$$H\Psi(R,r) = E\Psi(R,r) \qquad 1$$

where H is the Hamiltonian of the system is the wavefunction. R and r are the positions of the nuclei and electrons, respectively, and E is the energy of the system.

In the Born-Oppenheimer approximation the motion of the electrons can be decoupled from the nuclei, giving two separate equations. The first equation describes the electronic motion which depends only parametrically on the position of the nuclei R:

$$H\Psi(r;[R]) = E\Psi(r;[R]) \qquad 2$$

This equation defines an energy E[R], which is a function of only the coordinates of the nuclei, usually called the potential energy surface. The second equation describes the motion of the nuclei on this potential energy surface E[R|:

$$H\Psi(R) = E\Psi(R) \qquad 3$$

The direct solution of the eq. 2 is the subject of ab initio quantum calculations. In semiempirical approach many of the integrals needed for the solution of the eq. 2 are approximated with empirically fitted functions. Both approaches solve for the electronic wavefunction and energy as a function of nuclear coordinates.

Eq. 3 describe the evolution of the structure of the molecule. Simultaneous calculations for the solution of the eq. 2 and 3 for the complete calculation of the molecular system is extremely time and computer power consuming. That is why usually empirical fit to the potential energy surface is used to solve eq. 3. Next approximation is when the quantum mechanical effects are neglected as insignificant to the whole molecule description, and eq. 3 reduces to Newton equation of motion:

$$-\frac{dE}{dR} = m\frac{d^2R}{dt^2} \qquad 4$$

This approach for the calculation of molecular structure time behavior is called molecular dynamics. When the time evolution of the structure of the system is ignored and the calculations are focused on finding particular static geometries and their energies, it is called molecular mechanics.

The empirical fit to the potential energy surface E[R], called forcefield. describes the bond part of the potential energy and interatomic interactions including van der Waals and electrostatic interactions in different functional forms, depending on the type of the atoms involved. The forcefield describes entire classes of molecules with compromised accuracy, as the extrapolation from the empirical data of the representative set of molecules. The elaboration of the best forcefields is a task of its own and different forcefields are developed and constantly updated by the comparison of the computational results with experimental data for selected model systems by specialised scientific groups.

The coordinates of considered molecule combined with the forcefield data create the energy expression for the molecule, e.g. potential energy of this molecule as a function of its coordinates and is an input to the computer simulation programs.

Molecular Modeling Study of the Thermodynamics on Enzyme-Substrate interaction

The creation of new drugs is so far mainly a costly trial and error procedure of the synthesis of new chemicals with unknown binding properties. One of the developing application of computer molecular modelling is the rational drug design. It involves calculation of binding constants between target biomolecule and new drug molecule by exploring 3D structural information known from x-ray crystallography.

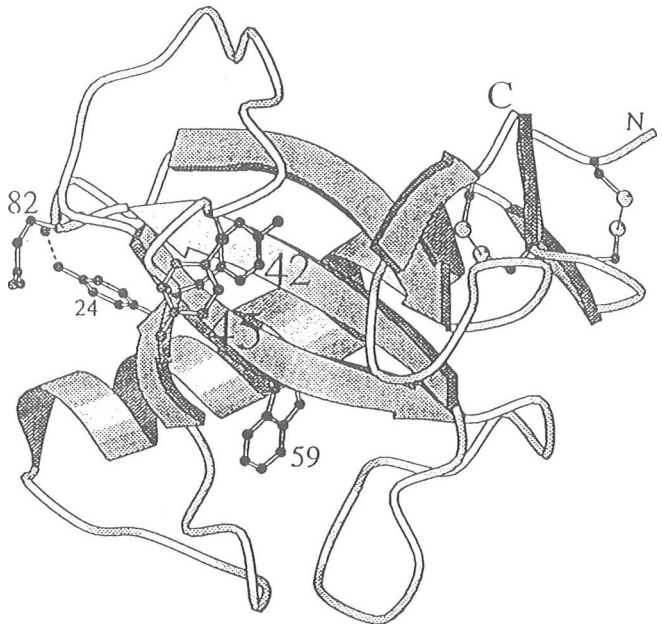

Fig 1. Symbolic presentation of ribonuclease T1 structure. Only few out from total 105 ammoacids are presented explicitly.

Ribonuclease T1 (RNT1). enzyme extracted from the fungus Aspergillus oryzae is an ideal system for the study of protein activity. It is a small protein, consisting of 105 aminoacids with well described crystal structure (more than 11 different crystal structures known), and clearly defined active site. RNT1 specifically cleaves single stranded RNA on the 3' side of guanylic residue. The active site has two functional parts, a recognition site that binds the guanosine ring and a catalytic center. The recognition site is highly specific for guanosine against adenosine and other nucleosides. The guanosine base appears to be recognised by specific hydrogen bond with amino acids residues Asn-43. Asn-44, Glu-46. and Asn-98. The base is also bound by stacking between Tvr-45. acting as a lid of the recognition site and Tyr 42 at the bottom (Fig. 1). Oxygen 6 of the guanosine ring is hydrogen bonded to the backbone

nitrogen (Fig. 2). These hydrogen bonds will be lost for other potential substrate, modified base 2 aminopurine (2AP).

The calculation started from the cry stal structure of RNT1:2'-GMP inhibitor complex determined by Saenger with 19 A resolution [4], To create the initial structure for wild type enzy me - substrate complex, the phosphate group was moved from the 2' to the 3' position on the sugar ring and located avoiding any bad van der Vaals contacts with the protein. Then an 18 A radius sphere of water molecules was placed around Tyr-42 . enclosing the active site and the w hole ligand. This system, that consists of about 2700 atoms, was simulated for 180 ps with a time step of 0.002ps to dynamically reach the optimal conformation in the solvent after introduced changes. Then the structure averaged over the last 100 ps of the simulation was energy minimised with 100 steps of Adopted Basis Newton-Raphson minimisation to relax strains and applied as the start for free energy simulations. CHARMM programme [5] with parameter set PARAM19 was used for the simulations.

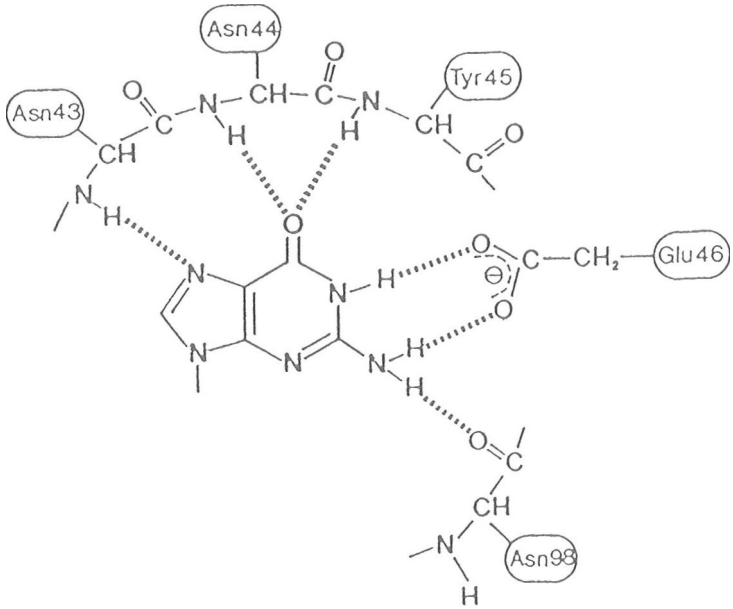

Fig 2. Scheme of the hydrogen bondings net between guanosine and the protein in the recognition site of the RNT1.

Site directed mutagenesis of cloned RNT1 gene make possible now to express and overproduce in bacteria variants of this enzyme characterised by modified enzymatic properties [1]. The modification of binding properties may be studied beforehand by free energy perturbation method, in which the relative free energy difference between different ligands binding to altered protein may be calculated, in order to reduce costs and time of biochemical experiments.

The effect of the point mutation of the enzyme by the substitution of Glu46 by Gln46 on the ligand binding was studied by computer calculation with free energy perturbation methods [2]. The theory of free energy simulations frequently used for biological systems has been extensively described by several autors (for reviews see [3]).

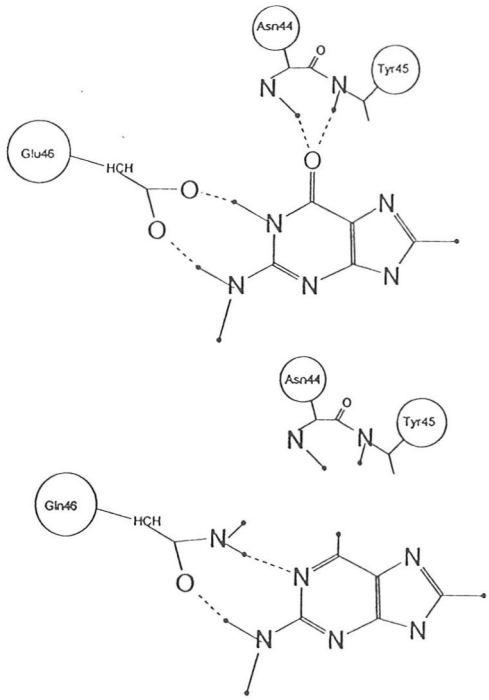

Fig. 3. Changes in the substrate - enzyme interaction scheme :
a). part of the natural recognition site (RNT1-Glu46 and 3'GMP)
b). postulated hydrogen bonding interactions for mutated enzyme RNT1-Gln46 and new substrate 2AP.

The free energy simulation of the substrate modification and enzyme modification included several steps, which will be described briefly below (for more details see [2]).

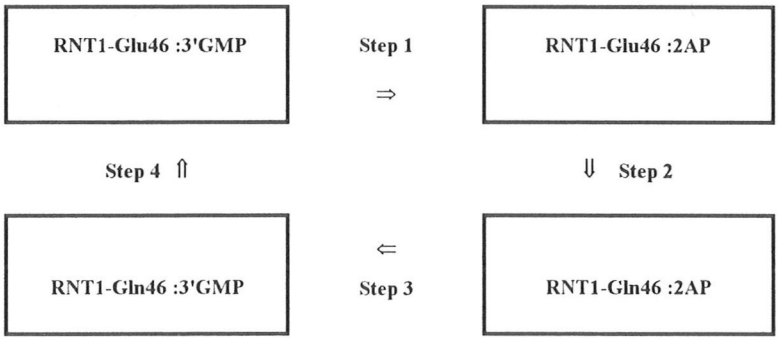

Fig. 4. Scheme of the perturbations steps used for the mutants of Ribonuclease T1 binding evaluation.

To simulate 2AMP bound to RNT1 instead of 3'-GMP , the 06, C6, N1 and H1 atoms in 3'-GMP were changed to the N1 and C6 atoms of 2AP in the previous structure. Free energy runs were performed for such system for 20 ps. where the first 5 ps simulation was used to relax modified structure. and another 15 ps of simulation were used for calculating the free energies. The next simulation was performed for the 3'-GMP in the same configuration as in the active site placed in the water sphere of 14 A for 180 ps to calculate the free energy of unbound solvated substrate. The mutation from Glu-46 to Gln-46 requires changes in the enzyme structure of all atoms starting from C_β of the glutamine into Gln sidechain atoms. This system had to be simulated lor 100 ps tor each free energy simulation run to reach equilibration of this quite large changes in the protein structure. Finally 180 ps normal MD simulations were performed for each of the four complexes with the two different ligands and the two enzyme structures. The last part of these simulations was used for studies of the structure and dynamics of these complexes.

The MD structures of the active site of the four complexes compared with the starting structure shows how the binding mode may be changed for altered enzyme or substrate. The structure of the active site of the complex RNT1-Glu-46 with natural substrate 3'GMP is similar to the X-ray structure of the complex with inhibitor. 2'-GMP.

When the substrate was replace by 2AP. the binding mode of the wild-type protein changed dramatically. The lid of the binding pocket, consisting of Tvr-45 and Asn-43 opened the binding centre, pointing out into the water. 2AP ring was displaced by twisting and shifting out from the enzyme. There were no close contacts between the recognising residues and the 2AP ring within natural enzyme.

For the modified enzyme, when Glu-46 was changed into Gln-46. the binding site structure with the 2AP substrate became again similar to the natural ligand-enzyme complex New substrate was better localised within recognition site than within the wild type enzyme. Tyr-45 was in the closed lid position, and the hydrogen bond between N1 atom of 2AP and hydrogen from Gln-146 amino group was formed, as expected. The other hydrogen bond between amino group of 2AP and the oxygen of Gln-46 was counteracted by a repulsive force between two amino groups.

Table I:
Free energy changes DDG calculated for binding of different substrates to natural and mutated RNT1.

Step	Time window [ps]	AAG kcal/mol
1	20	+2.8 +-0.7
2	100	-3.5 +-0.8
3	20	+2.3+-0.5
4	100	-1.0 0.5

The binding of the guanosine to modified RNT1-Gln-46 resembled to that Of 2AP. as well as to the binding of 3'-GMP to wild type of the enzyme. Tyr-45 is closing recognition site in a similar way as for 2AP. The oxygen group of Gln-46 formed hydrogen bonds to guanosine N1 and N2 hydrogens, but the amine group of Gln-46 and guanosine ring were pushing themselves away. Oxygen 06 of guanine formed two hydrogen bonds to the backbone NH group of Gln-46 and OH group of Tyr-42 side chain.

This picture may explain the microscopic details of substrate binding and recognition at the level of particular atoms within this complex structure. The free energy of binding is the physical parameter describing globally the binding affinity of a ligand to an enzyme, including all entropic contributions and potential energies of interactions. From these simulations we see that the free energy for the binding of 2AP to wild type RNT1 is higher then for 3'-GMP, which means that 2AP binds much weaker then the natural substrate. The other free energy simulations show that the mutant protein RNT1-Gln-46 binds 2AP better then the wild type protein, and prefers it against 3'-GMP. The small free energy difference for binding of 3'-GMP to the natural enzyme and to mutated form shows the same comparable binding affinity of both structures for normal substrate.

This studies show that complex biological systems can be studied by free energy methods and properties of modified molecules may be predicted.

Molecular Dynamics Computer Simulation of Nucleic Acids Hydration

The function of nucleic acids depends on their very specific three-dimensional structures which result from a balance between various intramolecular noncovalent forces (electrostatic forces due to the negative charged phosphates, stacking and hydrogen bonding interactions between bases, conformational energy of sugar-phosphate backbone) and interactions with an aqueous medium. Since the first studies on DNA fibres, it has been known that water activity can affect the nucleic acid conformations. Although it is well establish now that the DNA polymorphism can be induced by a change in the solvent conditions, the mechanism how these transitions occur on the molecular level is not fully clarified. Despite DNA and RNA differ only slightly in the chemical structure, there are fundamental differences in the structural and functional behaviour between both nucleic acid molecules. In contrast to the DNA polymorphism, the RNA structure seems to be quite insensitive to variations in the degree of hydration. There have been a number of proposals trying to explain the different properties in terms of influence of the 02'H hydroxyl present in RNA but not in DNA.

Molecular Dynamics (MD) computer simulations have been undertaken to compare the hydration pattern of the ribo and deoxyribo nucleosides in the aqueous solutions at the level of microscopic interactions and elucidate the role of the hydration water molecules on conformational dynamics of nucleic acids and in mediation of intermolecular interactions [6],

Cytidine (Cyd) and 2'-deoxycytidine (dCyd)) (Fig. 5), the monomeric units of RNA and DNA respectively, differ in the chemical structure of the furanose ring by the only one oxygen atom in the 2' position. This difference leads to the important and sometimes counterintuitive differences in the physico-chemical properties of these compounds and their derivatives observed experimentally. Although Cyd possesses one more hydrophilic group, it is six times less soluble in water than dCyd. The different association abilities of Cyd and dCyd

in the aqueous solutions have been found by Fourier transform infrared spectroscopy (FTIR) [7].

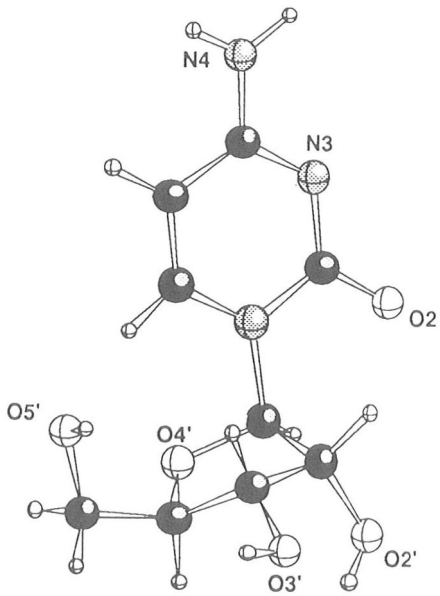

Fig. 5. The molecular structure of cytidine nucleoside. 2'[Deoxy-cytidine has the hydrogen atom instead of hydroxyl group in the 2' position of the furanose ring.

The simulation in each case was started by placing nucleosides randomly in a simulation cell filled with water molecules. Both the nucleosides and the water molecules were kept rigid in the first equilibrium period at 50 ps In the next equilibration period of 50 ps the nucleosides and the salts, if present, were allowed to become fully flexible. Finally, a complete MD simulation was carried out during 100 ps to study hydration structure and dynamics. The water molecules were kepi rigid even in the production period. All the simulations were performed with standard NVE^{MD} MD method using a modified version of McMOLDYN [8] computer simulation package.

Crystallographic coordinates for the nucleosides were taken from the Cambridge Crystallographic Data Base. Amber force field was used for the non-bonded and bonded interactions [9]. Water molecules were furnished with SPC parameters [10],

For the analysis of the data produced during computer simulation of this very dynamic system the atom-atom radial distribution functions (RDF) have been used to examine the structure of simulated aqueous solutions around selected centers of nucleosides. The RDF g(ij. r) describe the spatial organisation of atoms j around a central atom i and give the probability of finding a pair of atoms $i\,j$ a distance r apart, relative to the probability expected for a completely random distribution at the same density i.e. bulk density. r_O [11]. The integral z(ij. r) of g(ij. r)

$$z(ij,r) = 4\pi\rho_{0}\int_{0}^{r} g(ij,r)r^{2}dr$$

gives the number of atoms *j* around an atom *i* as a function of their distance r and is called the running integral of the RDF. From the combination of these functions one can extract the information about the nature of interactions and stability of possible hydrogen bonds when formed.

Table II: The average numbers of the hydrogen bonds formed by the Cyd and dCyd sugar hydrophilic centres with water molecules in the 0.5M and 5M simulated aqueous solutions.

		Cyd		dCyd	
		0.5 M	**5.0 M**	**0.5 M**	**5.0 M**
O...H_wO_w	**02'**	1.3	0.11	-	-
o. H_u • 1 xA	**03'**	1.7	0.09	1.5	0.10
O...O_w>= 2.8A	**05'**	1.6	0.09	1.6	0.09
	04'	0.6	0.08	0.8	0.07
OH...O_w	**02'**	1.0	0.02	-	-
<H...O_w>= 2.3A	**03'**	0.9	0.12	0.5	0.06
<O...O_w>=3.lA	**05'**	0.5	0.12	1.0	0.06

MD simulations have indicated that the hydration of Cyd and dCyd base moieties in the aqueous solution has a hydrophobic character. The radial distribution of water molecules around heterocyclic bases indicates that none of the potential Watson-Crick base pair centers form stable hydrogen bonds with the water molecules (Fig. 6). Similar distances between the O_w and H_w atoms of water and those from base suggest an apolar character of the hydration and the formation of something akin to a clathrate cage structure of water around base moieties of nucleosides.

The hydration of Cyd and dCyd sugar moieties shows a hydrophilic character and the three-dimensional networks of hydrogen bonded water molecules to all sugar hydrophilic centers are formed differently around the ribose and 2'-deoxyribose. The sugar hydroxyl groups participate in the hydrogen bonding with water molecules acting both as H-donor and as H-acceptor, see Table II and Table III. There are significant differences between Cyd and dCyd in an H-donor capacity of the sugar hydroxyl groups. The 02'H group exhibits the strongest H-donor ability in the 0.5M Cyd solution, while the 05'H group in 0.5M dCyd solution. The increase in concentration brings about more pronounced changes in H-donor and H-acceptor activity of sugar hydroxyl groups for Cyd than for dCyd.

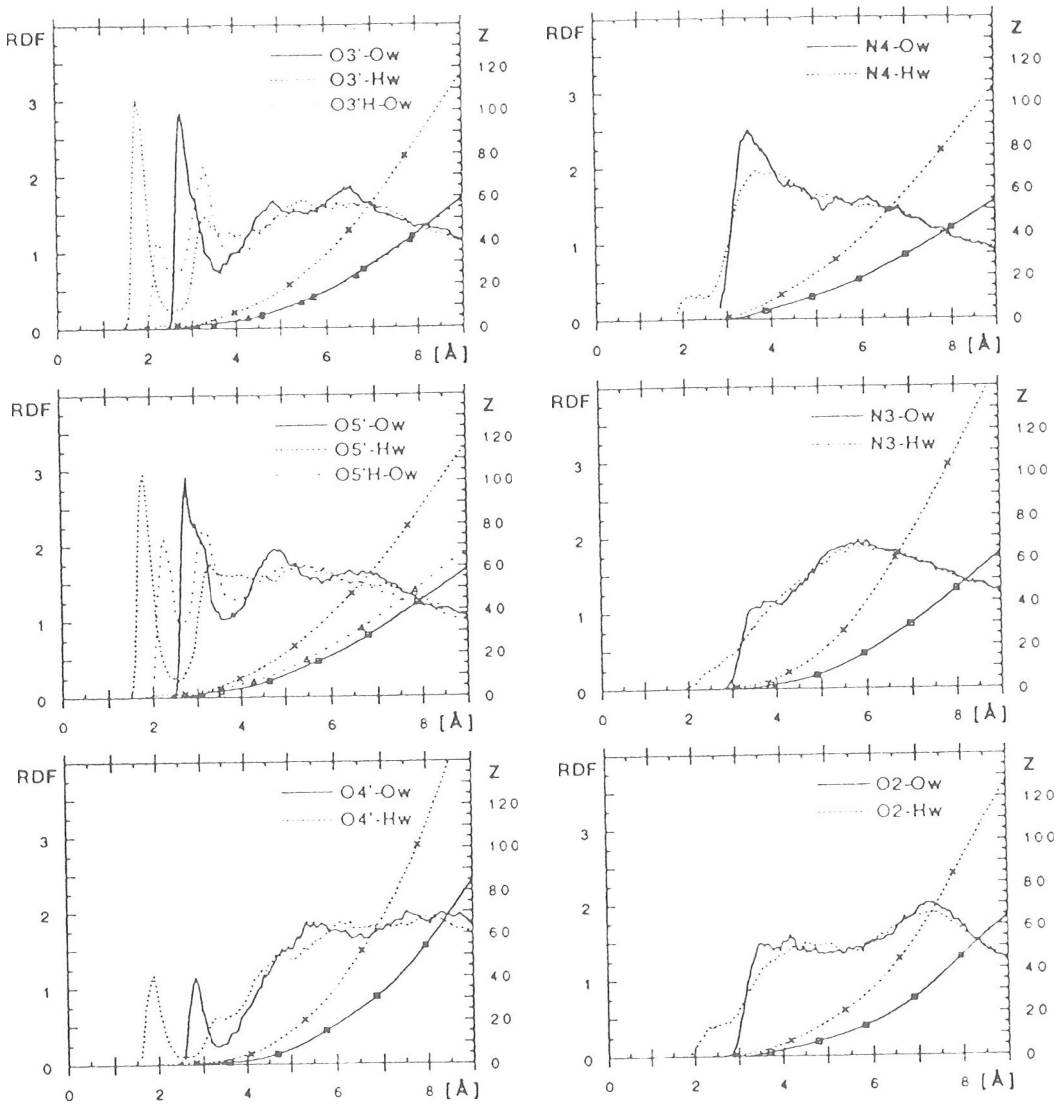

Fig. 6. The radial distribution of water molecules around all potential hydrophilic centres of Cytidine nucleoside in the simulated 0.5 M aqueous solutions:
a) - the ribose moiety, b) - the base moiety.

The analysis of interaction energies between water and nucleoside molecules has revealed that hydration interactions of Cyd are stronger than that of dCyd by an energy corresponding to one hydrogen bond which comes mainly from the different ribose and 2'-deoxyribose hydration.

Table III: Average distribution of water molecules in the first hydration shell around the sugar hydrophilic centres of Cyd and dCyd in the 0.5M and 5M simulated aqueous solutions.
R - the positions of the first minima for $g(0\text{-}0_w)$,
z(R) - the numbers of water molecules in the first hydration shell.

		Cyd		dCyd	
		0.5M	**5.0M**	**0.5M**	**5.0M**
02'	**R**	3.8 A	4.0 A	-	-
	z(R)	3.1	0.26	-	-
03'	**R**	3.7 A	3.8 A	3.7 A	3.6 A
	z(R)	3.1	0.26	2.7	0.19
05'	**R**	3.6 A	3.8 A	3.7 A	3.6 A
	z(R)	2.2	0.26	3.2	0.18
CM'	**R**	3.4 A	3.5 A	3.4 A	3.3 A
	z(R)	0.5	0,08	0.8	0.06

The data presented here indicate that MD computer simulations may be an extremely useful tool to study the hydration phenomena and to elucidate some atomic details of the hydration pattern in aqueous solution inaccessible to experimental methods.

References:

[1] Grunert, H P. Landt, O. Zirpel-Giesebrecht, M. Backmann, J. Heinemann, U. Saenger, W. Hahn, U. Trp59 to Tyr substitution enhances the catalytic activity of RNase T1 and of the Tyr to Trp variants in positions 24, 42 and 45. Protein Eng. 1993 (6(7)) pp 739-744
[2] Elofsson, A. Kulinski, T. Rigler, R. Nilsson, L. Site specific point mutation changes specificity: a molecular modelling study by free energy simulations and enzyme kinetics of the thermodynamics in ribonuclease T1 substrate interactions. Proteins 1993 (17(2)) pp 161-175

[3] Beveridge, D L. DiCapua. F M. Free energy via molecular simulation: applications to chemical and biomolecular systems. [Review]. Annual Review of Biophysics & Biophysical Chemistry 1989 (18) pp 431-492

[4] Heinemann.U. & Saenger ,W.(1982): Specific protein-nucleic acid recognition in ribonuclease T1-2',-guanylic acid complex: an X-ray studv Nature,299,27.

[5] Brooks. B.R.. Bruccoleri. R E., Olafsen. B.D., States. D.J., Swaminathan. S., Karplus. M., CHARMM: A program for macromolecular energy minimization and dynamics calculations. J. Comp. Chem. 4, 187-217, 1983.

[6] K.Kulinska and A.Laaksonen. "Hydration of cytidine, 2'-deoxyc\tidine and their phosphate salts in the aqueous solutions. A Molecular Dynamics computer simulations study ", J. Biomolecular Struct Dynamics 11, 1307-1325, 1994.

[7] Kulmska. K .,Sarzynska, J. and Wiewiorowski, M.(1991) *Vibrational Spectroscopy* 1: 277

[8] Laaksonen, A.(1986) *Comput Phys Comm* 42: 271.

[9] Weiner. S.J., Kollman, P.A., Nguyen. D.T., Case. 1). A. (1986) *J Comp Client,* v.7, 2:230

[10] Berendsen. H.J.C., Postma. J.P.M., van Gunsteren. W.F. and Hermans. J.(1981). in Intennolecular Forces, edited bv Pullman. B., p. 331.

[11] Allen. M.P. and Tildesley. D.J., *Computer Simulations of Liquids,* Oxford Science Publications. (1987).