AMoS: Agent-based Molecular Simulations

Ioanna K. Mprouza, Fotis E. Psomopoulos, and Pericles A. Mitkas

Department of Electrical and Computer Engineering
Aristotle University of Thessaloniki
Thessaloniki, 54 124, Greece
{ioanna_brouza@hotmail.com, fpsom@danae.ee.auth.gr, mitkas@eng.auth.gr}

Abstract

Molecular Dynamics (MD) is a form of computer simulation wherein atoms and molecules are allowed to interact for a period of time, utilizing theories from mathematics, physics and chemistry. At the core of any MD simulation lies the potential function (or force field), which describes the interactions between the particles of the simulation.

In this paper a new framework for MD simulations is presented, which utilizes software agents. Every agent in our multi-agent system corresponds to a single particle and probes its environment for candidate agent-particles with which an interaction is possible.

The framework is applied on protein structural data (PDB files) using an implicit solvent environment and a time step of 5 femtoseconds. Although the system is fully parameterized, the experiments were based on a specific force field and set of parameters, known as ENCAD.

Keywords: Molecular Dynamics, Force Field Equations, Multi-Agent Systems, Protein Prediction Structure, Simulation, Protein Data Bank

1. Introduction

Biologists have always tried to understand the many ways in which nature "constructs" life and how organisms operate and survive. In recent years, however, the complexity of the problems in biology on one hand, and the emergence of computers as a powerful analytical tool on the other hand, have led biologists into a close cooperation with computer scientists. A new science was born due to this cooperation, known as Bioinformatics. Computer science provides the algorithmic and theoretical background for biological problems, while concepts from biology may lead to different approaches in informatics issues.

One of the fundamental and most interesting problems in bioinformatics, is the study of proteins. Proteins are the macromolecules responsible for most of the operations and functions in a living organism. They are large organic compounds consisting of aminoacids (which may number from 50 to 500), where aminoacids are molecules with a certain structure. There are exactly twenty different aminoacids that participate

in the composition of proteins. There is currently a lot of ongoing research regarding the study of protein structure and its correspondence with the functional behaviour of the protein.

This paper describes AMoS (Agent-based Molecular Simulations), a system that we have developed in order to estimate the motion of macromolecules consisted of aminoacids. AMoS is a multi-agent system that imitates the macromolecule in a real world environment. The main goal is to represent every molecular atom with an intelligent software agent. In computer science, a software agent is a piece of software that acts on behalf of a user or other program. Such "action on behalf of" implies the ability to decide when (and if) an action is appropriate. The idea is that agents are not strictly invoked for a task, but activate themselves. Intelligent agents are characterised by autonomy, mobility, communication and interaction with each other. The user describes the system s/he desires to simulate and AMoS returns as output the estimated motion of the input macromolecule.

AMoS is based on the deterministic and multidisciplinary method of simulation known as Molecular Dynamics (MD). MD is used in different kinds of systems with varying level of detail, ranging from quantum mechanics to molecular mechanics and gives the motion of macromolecules utilising known energy functions-force fields [Lifson (1968), Warshel (1970)].

2. Theoretical background

Molecular Dynamics refers to the computational method that provides the analysis of a system in terms of time and is the method that led to the solution of N-body problem [Rahman (1964)]. The target of this problem is to find the motion of N bodies based on classical mechanics, having as input their initial coordinates, their masses and their velocities. The difficulty rises from the fact that the number of particles might be so large that it may be impossible to analytically calculate the properties of the system. The MD methods circumvent this difficulty by using numerical approaches based on theories of mathematics, chemistry and physics combined with computer algorithms. At the core of any MD simulation lies the potential function (or force field), which describes the terms of interaction between the particles of the simulation [Rapaport (2004)].

Most of the force fields that are used in proteins simulations are based on the function of potential energy U_{Total} that is given in Eq. 1, where all parameters are related to the types of energy they describe [Ponder (2003)].

$$U_{Total} = \sum_{AllBonds,b} k_b \cdot (b - b_0)^2 + \sum_{\substack{AllBond \\ Angles,\theta}} k_\theta \cdot (\theta - \theta_0)^2 + \\ + \sum_{\substack{AllTorsion \\ Angles,\varphi}} k_\varphi \cdot [\cos(n \cdot (\varphi + \delta)) + 1] + \\ + \sum_{\substack{NonBonded \\ Pairs}} \left[\frac{q_i \cdot q_j}{r_{ij}} + \frac{A_{ij}}{r_{ij}^{12}} - \frac{C_{ij}}{r_{ij}^{6}} \right] \quad \textbf{(1)}$$

Although AMoS is fully parameterized and allows the use of any force field, our experiments depend on a particular force field, which is called ENCAD. The potential function that was used for our simulations is a semi-empirical analytical function in Cartesian coordinates (Eq. 2). Further details on the ENCAD force field, including its parameters, are presented in [Levitt (1995)].

$$U_{Total} = \sum_{AllBonds,\ b} k_b^i \cdot (b_i - b_0^i)^2 + \\ + \sum_{\substack{AllBond \\ Angles,\ \theta}} k_\theta^i \cdot (\theta_i - \theta_0^i)^2 + \\ + \sum_{\substack{AllTorsion \\ Angles,\ \varphi}} k_\varphi^i \cdot [1 - \cos(n^i \cdot (\varphi_i - \varphi_0^i))] + \quad \textbf{(2)} \\ + \sum_{\substack{AllNonBonded \\ Distances,\ r<r_c}} A_{sc} \cdot \varepsilon^{ij} \cdot \left[\left(\frac{r_0^{ij}}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{r_0^{ij}}{r_{ij}} \right)^{6} - S_{vdw}^{A}(r_{ij}) \right] + \\ + \sum_{\substack{AllNonBonded \\ Distances,\ r<r_c}} \left[\frac{q^i \cdot q^j}{r_{ij}} - S_{els}^{A}(r_{ij}) \right]$$

An understanding of the five terms of Eq. 2 was essential for the development of the multi-agent system, in order to provide all the needed characteristics for the agents.

The first term ($\sum_{AllBonds,\ b} k_b^i \cdot (b_i - b_0^i)^2$) describes the potential energy between two atoms due to the covalent bond that connects them. At this point it must be noted that every bond is elastic. Bond length stretching energy is described by a quadratic function, which has a minimum when the bond with length b_i reaches its state of

equilibrium b_0^i (Fig. 1). The parameters b_0^i and k_b^i depend on the type of atoms that participate in the covalent bond.

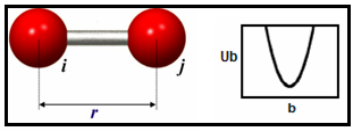

Fig. 1. *Two atoms which are connected with a covalent bond (left). The diagram of the bond length energy function is shown on the right.*

The second term in Eq. 2 ($\sum_{\substack{AllBond\\ Angles,\ \theta}} k_\theta^i \cdot (\theta_i - \theta_0^i)^2$) refers to the potential energy that is developed due to the bending of the angle created by three atoms. Two of these three atoms are connected with the third with a covalent bond. The bond angle bending energy is described by a quadratic equation, like the bond length stretching energy, and it exhibits a minimum (Fig. 2) when the angle θ_i reaches its state of equilibrium (θ_0). The parameters θ_0 and k_θ^i are based on the type of the three atoms and the way they form the angle.

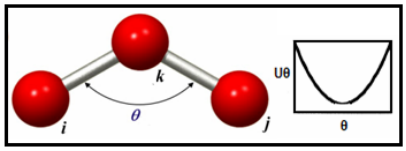

Fig. 2. *The formation of an angle between 3 atoms and two covalent bonds (left). The diagram of the bond angle bending energy θ (right).*

The third term of Eq. 2 ($\sum_{\substack{AllTorsion\\ Angles,\ \varphi}} k_\varphi^i \cdot [1 - \cos(n^i \cdot (\varphi_i - \varphi_0^i))]$) is more complicated as it concerns two structures of four molecules. The first structure is represented in the first image of Fig. 3 and the energy is known as true dihedral or torsion angle twisting energy. The second structure is the potential energy that is provoked due to an out-of-plane dihedral or torsion angle (middle image of Fig. 3). The potential energy of the third term is one or more cosine functions, that take values in the range $[-k_\varphi^i, k_\varphi^i]$, with periodicity n and value of equilibrium φ_0^i (Fig. 3). Parameters k_φ^i, φ_0^i και n are defined by the type of the four atoms and the way they are connected.

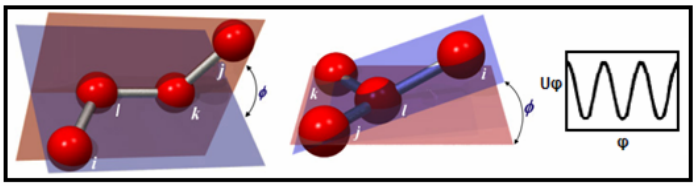

Fig. 3. *The two images in the left show the two categories of torsion angle that is formed of four atoms and three covalent bonds. The third image represents the diagram of the the torsion angle twisting energy.*

The fourth term ($\sum_{\substack{AllNonBonded \\ Distances,\ r<r_c}} A_{sc} \cdot \varepsilon^{ij} \cdot \left[\left(\frac{r_0^{ij}}{r_{ij}} \right)^{12} - 2 \cdot \left(\frac{r_0^{ij}}{r_{ij}} \right)^{6} - S_{vdw}^{A}(r_{ij}) \right]$) refers to the potential energy known as Van der Waals. The parameter r_{ij} is the distance between atoms i and j. A_{sc} is a scale to reduce Van der Waals repulsion to compensate for truncation and $S_{vdw}^{A}(r)$ is a shifting function that ensures smooth truncation. The pair of interacting atoms i and j are chosen to include all atoms closer than the cutoff distance r_c (Figure 4), but exclude atoms that are connected by three or fewer bonds. The values of ε and r_0 used for a particular atom pair are the geometric means of the tabulated values. The energy arising from this interaction varies with the separation distance between the two atoms. This energy is zero at infinite distance, but as the separation is reduced the energy decreases, passing through a minimum and from there on increasing rapidly.

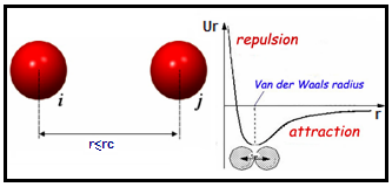

Figure 4. *Two atoms that are subjected to Van der Waals interactions (left). The diagram of Van der Waals energy (right).*

Finally, we have the term that gives the electrostatic energy $\frac{q^i \cdot q^j}{r_{ij}} - S_{els}^{A}(r_{ij})$, where the atomic charges (q^i and q^j) refer to units of electrons, r_{ij} is the distance between atoms i and j, and $S_{els}^{A}(r_{ij})$ is a shifting function that ensures smooth truncation at the cutoff distance, r_c but exclude atoms that are connected by three or

fewer bonds. Fig. 5 shows the structure of the molecular atoms and the diagram of the electrostatic energy.

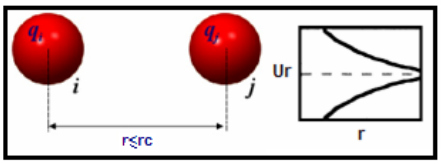

Fig. 5. *Two atoms that are subjected to electrostatic interactions (left). The diagram of electrostatic energy (right).*

It is necessary to say that the electrostatic and Van der Waals terms of the potential energy function are really important for the simulations because of their complexity, while the three first terms that are based on interactions caused by one, two or three covalent bonds usually have values near the value of the equilibrium state.

After the evaluation of the potential energy, the next step is to calculate the force that is exerted on every atom. For this reason the force equation $F_i = -\frac{\partial U}{\partial x_i}$ (Eq. 3) is used. Based on the second law of Newton $F_i = m_i \cdot \alpha_i$ (Eq. 4), we calculate the acceleration that every atom has ($\alpha_i = \frac{F_i}{m_i}$). The atom's new position and velocity are given by Eqs. 5 and 6 which were developed according to classic mechanics (Beeman method presented in [Beeman (1976)]).

$$u_i(t+\Delta t) = u_i(t) + \left[2 \cdot \alpha(t+\Delta t) + 5 \cdot a_i(t) - \alpha_i(t-\Delta t)\right] \cdot \frac{\Delta t}{6} \quad \textbf{(5)}$$

$$x_i(t+\Delta t) = x_i(t) + u_i(t) \cdot \Delta t + \left[4 \cdot \alpha_i(t) - \alpha_i(t-\Delta t)\right] \cdot \frac{\Delta t^2}{6} \quad \textbf{(6)}$$

The pointer i that exists in variables F_i, x_i, α_i and u_i has an integer value of 1 to 3 and refers to the three components in the three-dimensional space.

3. System architecture

The main goal of our approach was to provide a software platform for molecular simulations, with emphasis on aminoacid macromolecules i.e. protein chains. In this system, intelligent software agents represent atoms in three-dimensional space.

The platform can be described as a Multi-Agent System (MAS). MAS comprise agents that are scattered around the environment and act either autonomously, or in collaboration. On one hand, there are cooperative agents working towards a common

goal, who interact in order to solve problems that are beyond the individual capacities or knowledge of each problem solver. On the other hand we have selfish agents who pursue their own interests. In the case of the AMoS MAS, the agents cooperate in order to model the motion of the protein in three dimensional space. Each atom of every aminoacid participating in the protein chain is represented with an agent.

The logical diagram of the system is presented in Fig. 6. In order to illustrate the way AMoS works, the input and output of the system will be described in detail, as well as the creation and the role of the participating agents.

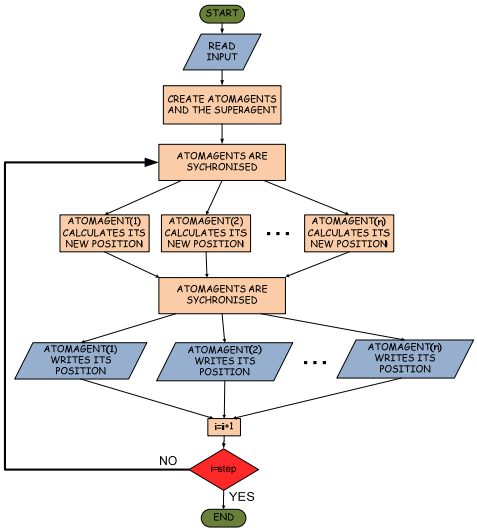

Fig. 6. *AMoS logical diagram*

The main user input file is called *InputParameters.properties*, and defines the simulation system. There are two major parameters, *protein* and *stop*. Parameter *stop* refers to the steps of the simulation. Parameter *protein* is a string in the form "$i,AMIN_1,AMIN_2,\ldots,AMIN_i$", where i is an integer representing the number of aminoacids that comprise the macromolecule in question. Parameter $AMIN_i$ is a string of three capital letters and it is unique for every aminoacid.

According to the parameter *protein* and especially its aminoacids, the system starts the simulation by reading PDB files. These files are 20, one for each aminoacid, and provide the structure of the atoms that constitute the aminoacid. In every PDB file, there are details about the coordinates and the way the atoms are connected in the three-dimensional space.

During the initialization phase of the platform, the coordinates of all aminoacids are transformed into a global system according to their connection with each other, and a software agent is created for every atom involved. This agent has all the characteristics of the atom, such as the coordinates, the mass, the code and the

symbol, which are needed in order to calculate the energy and its new position. The agents that represent the atoms are called AtomAgents. Apart from the AtomAgents, there is one more agent, the SuperAgent, which has the role of the coordinator in that it is responsible for the synchronization of the AtomAgents and the termination of the simulation.

During the actual simulation, every AtomAgent imitates the operations of the real atom it represents. First of all, it finds all the atoms that are closer than four covalent bonds and asks for their coordinates. In this way, it calculates the three potential energies that are related to interactions where covalent bonds participate (Eq. 2). Then it chooses those of the remaining atoms that are closer than r_c and calculates the electrostatic and Van der Waals interactions (Eq. 2). The sum of these five energies gives the total potential energy of the atom, from which the force the AtomAgent is subjected to can be calculated. Finally, Eqs. 5 and 6 are used to give the new coordinates of the atom.

During the simulation the system produces a number of text files, one each atom. Every line of the file identifies the position of the atom in the *n*-th step of the simulation with three double precision numbers corresponding to the *x*, *y* and *z* coordinates. In order to produce a visual result, Matlab was used and a function was developed which creates a video of the protein motion based on the .txt files. There are two modes of video reproduction. In the first mode, the individual atoms of the aminoacids are represented with spheres of different shape and colour according to their type (**H**, **O**, **N**, **C**, **S**), while in the second mode the whole aminoacids are rendered with equal spheres of different color.

4. Experiments

In order to evaluate the AMoS platform, several experiments were performed using different input files. As discussed in the previous section, all input parameters exist in the *"InputParameters.properties"* file (*protein, stop, dt, rc, stepZeroVel* and *stepZeroAccel)*. The main parameter is *protein* which gives the number and the sequence of the aminoacids that are about to be simulated. Parameter *stop* defines the number steps in the simulation. The four parameters *dt,* r_c*, stepZeroVel και stepZeroAccel* are essential for the potential energy and the equations of motion. Electrostatic and Van der Waals forces are exerted to two molecular atoms whose distance is less than r_c. Parameter *dt* is the time step which is used in the equation of motion (Equation 5 and 6). At this point, it is important to note that all the measures are normalized [Rapaport (2004)].

It is known that a macromolecule consisting only of Alanines creates a helix-like structure [Bortolussi (2005)]. In order to evaluate the platform, several experiments were performed using poly-alanine chains. The main parameters and execution time

of the most characteristic experiments are presented in Table 1. In order to illustrate the results, experiments 5 and 8 are described in detail.

Table 1. *Experiments that were made with the use of AMoS.*

No	Protein	stop	dt	agents	time length (hh:mm:ss)
1	3,ALA,ALA,ALA	1000	10^{-5}	34	00:02:25
2	3,ALA,ALA,ALA	10000	10^{-5}	34	00:24:23
3	3,ALA,ALA,ALA	20000	10^{-2}	34	00:51:10
4	3,ALA,ALA,ALA	60000	10^{-2}	34	02:24:23
5	3,ALA,ALA,ALA	100000	10^{-2}	34	06:04:46
6	6,ALA,ALA,ALA,ALA,ALA,ALA	1000	10^{-2}	64	00:13:44
7	6,ALA,ALA,ALA,ALA,ALA,ALA	10000	10^{-2}	64	02:08:36
8	6,ALA,ALA,ALA,ALA,ALA,ALA	60000	10^{-2}	64	21:41:09
9	9,ALA,ALA,ALA,ALA,ALA,ALA, ALA,ALA,ALA	1000	10^{-2}	94	00:36:30
10	12,ALA,ALA,ALA,ALA,ALA,ALA, ALA,ALA,ALA,ALA,ALA,ALA	1000	10^{-2}	124	01:30:07
11	15,ALA,ALA,ALA,ALA,ALA,ALA, ALA,ALA,ALA,ALA,ALA,ALA, ALA,ALA,ALA	1000	10^{-2}	154	02:36:24
12	16,ALA,LEU,MET,GLU,ALA, GLN,HIS,ALA,LEU,LYS,MET, GLU,ALA,HIS,LEU,ALA	1000	10^{-2}	245	18:24:35

Experiment 5 simulates a macromolecule of three Alanines, where the main parameters have values of *stop=100000* and $dt=10^{-2}$. During the video rendering of the simulation, every aminoacid was represented by a sphere. Fig. 7 shows 12 frames of this video, and although it is obvious that there is a motion where the atoms conserve their relative positions, the emerging structure is not readily distinguishable.

Taking this under consideration, experiment 8 was performed using six Alanines with *stop=60000*. This time, the video was rendered in both modes (using atoms and aminoacid representation, respectively). Fig. 8 shows the separate atom rendering while Fig. 9 presents the same video frames with the aminoacid representation. This time, the expected helix-like structure is evident in both cases.

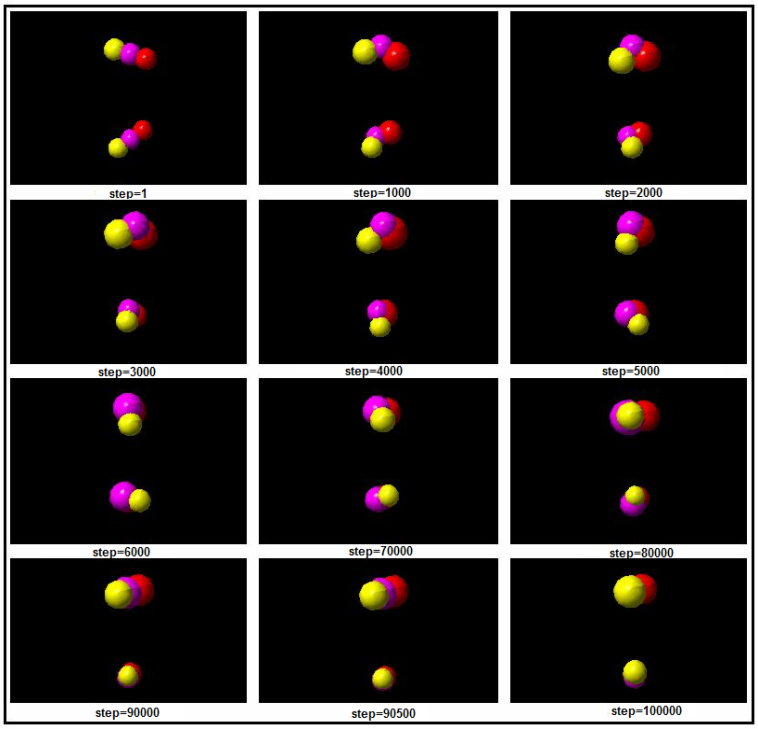

Fig. 7. *Simulation of a macromolecule with three alanines for 100000 steps and* $dt=10^{-2}$

5. Conclusions and Future Work

AMoS is a molecular simulations platform, which creates a MAS representing the actual system. Thanks to software agents there is an almost perfect correspondence between reality and the simulation, where every agent represents a particular molecular atom. The agents operate in the same way as atoms, by interacting with each other and continually moving without breaking down the macromolecule.

In this paper, we utilized a particular empirical force field known as ENCAD and through a series of experiments we ascertain the correct usage of this potential function. Many experiments were preformed with macromolecules of Alanines, where we observed their helix-like structure.

The use of software agents allowed for a detailed molecular simulation, overcoming the problem of the large number of calculations. Every agent calculates the total force to which the corresponding atom is subjected. This was accomplished by using communication among the agents, either through message exchange or directly. The

result is that each agent quickly finds the agents with which it interacts and all agents are synchronized.

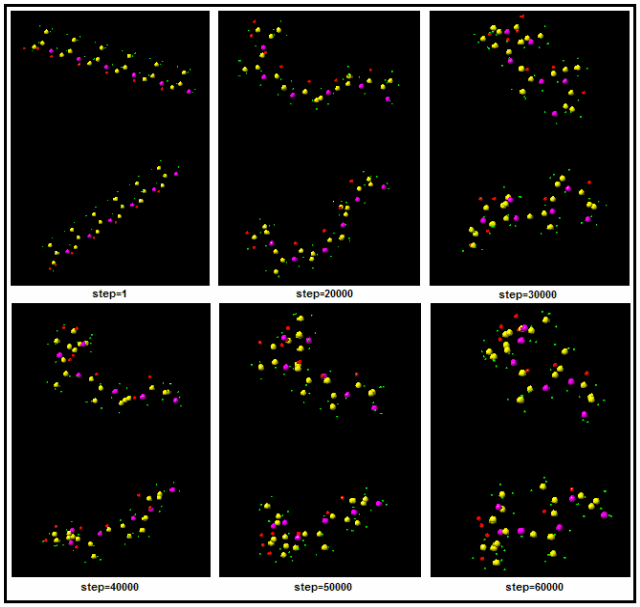

Fig. 8. *Simulation of a macromolecule constituted by six alanines for 60000 steps and* $dt=10^{-2}$

An important framework has been created, allowing for some interesting future extensions. AMoS can accept the usage of any force field, with minor changes in the code. Moreover, several advantages of the agent technology are yet to be utilized. For example, an AtomAgent can be further designed to support decisions, such as if it is preferable to move in a new position than stay in its old one, according to the total energy change of the system.

Moreover, the advantages of software agents make it possible to study the properties and the "behaviour" of molecular complexes by a series of experiments. In particular, it is possible to create categories with sections of atoms that will have a certain motion according to the results of the experiments. In this way, agents will be able to detect such sections and predict their position without the usage of a potential function.

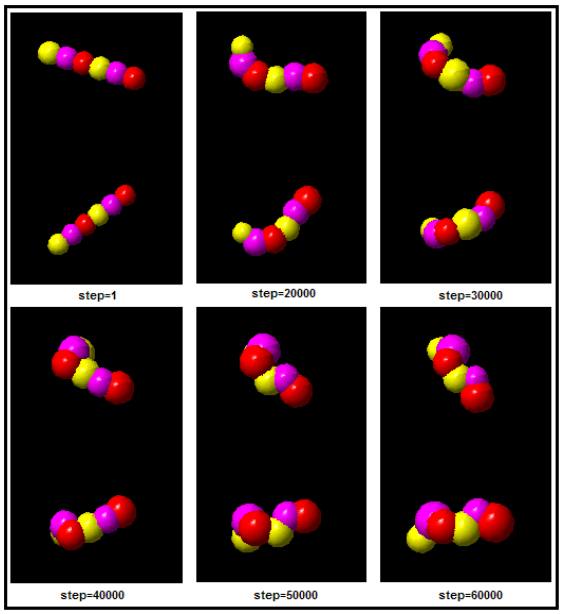

Fig. 9. *Simulation of a macromolecule with six alanines for 60000 steps and dt=10^{-2} (representation with aminoacids)*

References

1. D. Beeman (1976), *Some Multistep Methods for Use in Molecular Dynamics Calculations,* Journal of Chemical Physics, pp. 20.
2. L. Bortolussi et al, (2005), Protein Folding Simulation in CCP
3. M. Levitt et al, (1995), *Potential energy function and parameters for simulations of the molecular dynamics of proteins and nucleic acids in solution*, Computer Physics Communications 91.
4. S. Lifson and A. Warshel (1968), *A Consistent Force Field for Calculation of Conformations, Vibrational Spectra and Enthalpies of Cycloalkane and n-Alkane Molecules*, Journal of Chemical Physics, pp 49.
5. J.W. Ponder, D.A. Case (2003), *Force Fields for Protein Simulation,* The Scripps Research Institute, La Jolla, California.
6. A. Rahman (1964), Correlation of motion of atoms in liquid argon, Phys. Rev., pp.136A.
7. D.C. Rapaport (2004), The Art of Molecular Dynamics Simulation, Cambridge Un. Press.
8. A. Warshel and S. Lifson (1970), *Consistent Force Field Calculations. II Crystal Structures, Sublimation Energies, Molecular and Lattice Vibrations, Molecular Conformations and Enthalpies of Alkanes*, Journal of Chemical Physics, pp. 53.