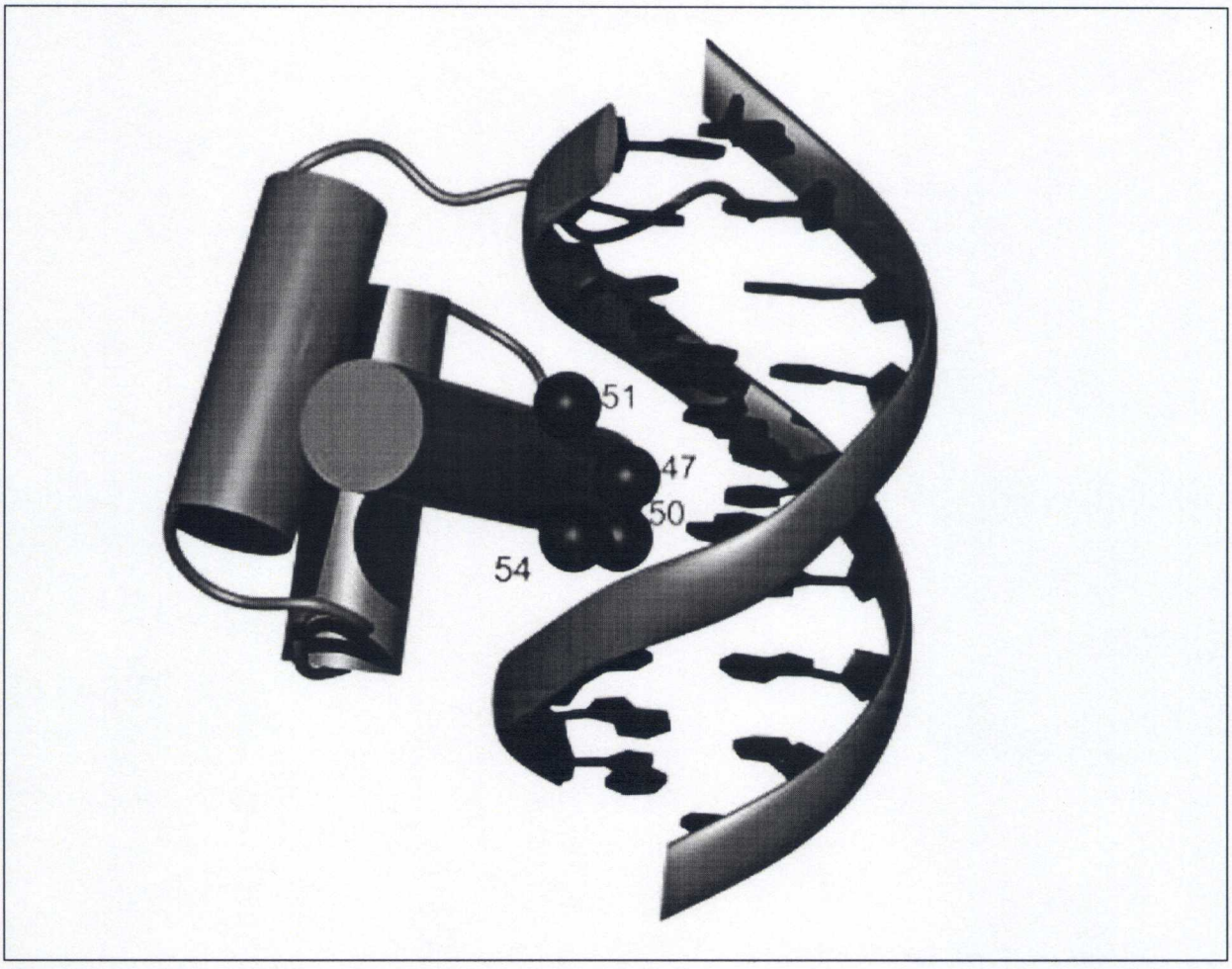
The newsletter of the **Centro Svizzero di Calcolo Scientifico / *Swiss Center for Scientific Computing***

August 1997
volume 6
number 3

CROSSCUTS



Homeodomain-DNA complex. See article p.19.

T A B L E O F C O N T E N T S

A User Project: Biological Macromolecules Studied by Molecular Dynamics Simulations

Martin Billeter, Peter Güntert, Peter Luginbühl, Reto Koradi and Kurt Wüthrich (Molecular Biology and Biophysics Institute, ETH Zurich)

Introduction

Metabolic processes in living cells are often dominated by macromolecules such as proteins or DNA. While proteins play an active role in almost all cellular events, the main purpose of DNA is the storage of genetic information. The functions of proteins and DNA are nicely illustrated by early events in the development of a single cell into a complex organism. Genetic information from the DNA must be read and expressed in a controlled manner, i.e., at the proper time and in the proper location of the growing organism. Homeodomains, which are part of larger proteins, are key actors in these control mechanisms. Their function is to recognize specific DNA fragments and, by binding to these, they identify starting points for the reading of genetic information.

Both proteins and DNA consist of many thousands of atoms and form linear chains with amino acids or nucleotides, respectively, as building blocks. Protein chains adopt unique three-dimensional folds, which are defined by the amino acid sequences. These three-dimensional (3D) structures determine their function; for example, the structure of a homeodomain is optimized for good interactions with DNA molecules. The nucleotide sequence in DNA chains represents the genetic information, and DNA chains are usually paired to form a double helix. In our laboratory we use nuclear magnetic resonance (NMR) experiments and various computer algorithms to determine the structure and to study the dynamics of macromolecules. One of our research projects focuses on the interaction of homeodomains with DNA. We have presented the first 3D structures at atomic resolution of a homeodomain in 1989 [1] and of a homeodomain-DNA complex in 1990 [2]. The refinement of the latter structure [3] and additional results [4] demonstrated the important role of water molecules penetrating the interface between homeodomain and DNA. These experimental data were then complemented by molecular dynamics (MD) simulations [5].

Computational Procedure

In classical molecular dynamics calculations the evolution of a molecular system is simulated by integration of the equations of motions. Each atom is represented by a mass point carrying a charge that interacts with other atoms in various ways. For example, a chemical bond between two atoms keeps their distance rigid, while Coulomb's law induces attractive or repulsive forces between any pair of atoms. The molecular system is given kinetic energy by attributing to each atom a velocity; the sum of these energies defines a temperature. In the simulation, the forces acting on each atom due to its neighbors are evaluated and the velocities are modified according to Newton's law. The system is then allowed to evolve linearly for a short time step, i.e., the atoms move with constant velocity. This combination of force evaluation, velocity updates and linear time steps is repeated many times. Thereby, the time steps must be sufficiently short to adequately sample the fastest motions in the system and avoid too close atomic contacts which could result in extremely strong repulsive forces. The size of a molecular system, about 10'000 atoms in the application below, and the choice of the time step between evaluations of the interatomic forces, typically a few femtoseconds (10^{-15} *s*; *fs*), strongly limit the time period that can be covered by MD simulations.

The original version of the MD simulation program OPAL [6] was designed for high efficiency on the NEC SX-3, with an average performance of 1.5 GFlops. Other features are its ease of use: The user interface provides a "programming language" which allows both interactive commands and the call of macros.

Command	Explanation
amber94	read parameters defining interatomic forces from library
read dg protein-DNA	read input structure from file *protein-DNA.dg*
build	build structure and initialize force field
inertia	transform coordinates to principal axes of inertia
velocity 293.0 force	set velocities based on forces at temperature 293 K
immerse water.asc	solvate structure in water bath from file *water.asc*
check	ensure that all atoms are within the surface boundary
pressure	report pressure
groupenergy	report intramolecular and intermolecular energies
temperature	report temperature
do i 12000	loop for 2000 MD runs of 1 ps each
md 400 100 0.0025	MD run: 400 steps of 2.5 fs, information every 100 steps
mdstat	report average values for energy, temperature
write asc snap$i(i3.3)	save snapshots in files numbered by loop variable *i*
end do	end of loop

Table 1: *Macro file for OPAL used to perform a standard MD simulation. Each line contains one OPAL command on the left, and an explanation of it on the right. These commands perform a MD simulation of 2 ns starting with a structure read from the file protein-DNA.dg, which is immersed into a water bath obtained from the file water.asc. Time steps are 2.5 femtoseconds, and every picosecond the current structure is saved.*

Table 1 lists all the commands required to perform a complete MD simulation of 2 nanoseconds (*ns*) length, with output of selected information (temperature...) every 250 *fs* and of intermediate structures every picosecond (*ps*). With these parameters, the application described subsequently (with nearly 10'000 atoms) required about 335 CPU hours for 800'000 time steps.

(Continued on page 20)

(Continued from page 19)

Results

A Two-Nanosecond Trajectory of the *Antennapedia* Homeodomain Bound to DNA

The system selected for the present investigation [5] consists of a homeodomain fragment with 54 amino acid residues and a DNA double helix with nine nucleotides on each chain (Figure 1).

This system contains the central elements of the *Antennapedia* homeodomain [fruit fly *Drosophila*] which controls antenna formation on the fruit fly *Drosophila*, and a target DNA [7]. The relative position of the two macromolecules in the initial configuration corresponds to the NMR structure of this complex [3]. This system was immersed in a water bath, resulting in a total number of atoms of 9721. Two nanoseconds were simulated in time steps of 2.5 *fs*. The evolution in time of the homeodomain-DNA-water system may be summarized as follows.

The homeodomain-DNA complex remained intact. The backbone of the protein changed very little, but larger motions were observed for its side chains. Of particular interest are side chains in the protein-DNA interface, some of which jumped between several DNA nucleotides over the course of the simulation. The interface part of the DNA remained rather rigid, offering a stable contact surface to the protein. During the simulation several water molecules exhibited different interaction behaviors with the macromolecules. Most water molecules resided in the bulk water surrounding the protein-DNA complex, with occasional short-lived contacts to the surfaces of the macromolecules. About 5% of the water molecules penetrated the protein-DNA interface and remained there for several hundred picoseconds.

Figure 2 illustrates the pathway of a water molecule that was initially located well outside the complex, then entered the macromolecular interface, and left it again after about 600 ps. The statistical significance of the run, i.e., the adequacy of the length of the simulation is demonstrated by the fact that the water molecules penetrating the protein-DNA interface during the simulation were initially distributed all over the system.

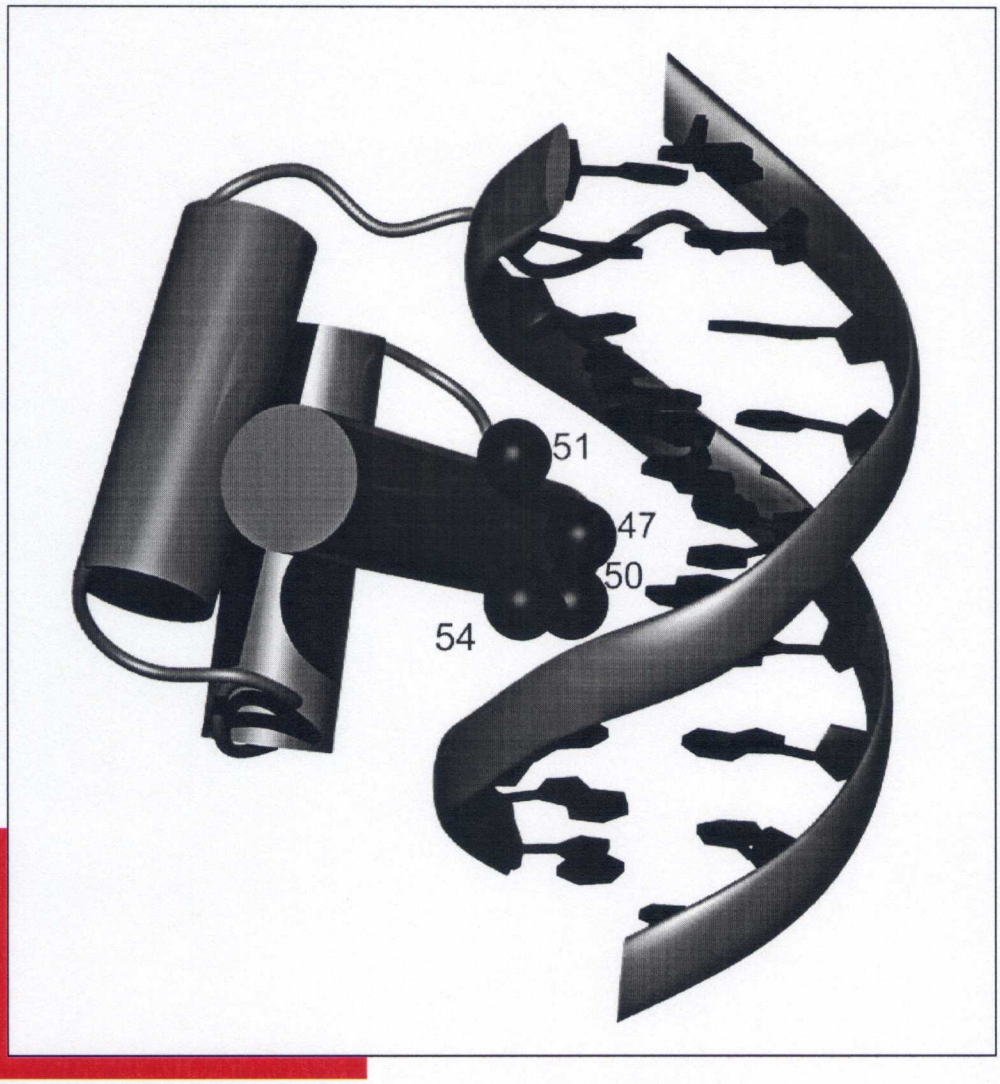


Figure 1: *View of the homeodomain-DNA complex studied in the present MD simulation. In this schematic drawing, the DNA is located on the right and represented by two ribbons for the backbone and plates for the sugar and base moieties. On the left, the backbone fold of the homeodomain is indicated by cylinders for helical structures and connecting bonds. Spheres show the positions of four side chains (with sequence numbers) located in the protein-DNA interface. (This figure was prepared with the program MOLMOL [9]).*

Direct and Water-Mediated Protein-DNA Contacts

Contacts between the homeodomain and the DNA that are responsible for the specific recognition of the DNA sequence are formed by the side chains of the amino acids at positions 47, 50, 51 and 54 (see Figure 1). These interactions can be grouped into hydrophobic contacts between carbon atoms (and their hydrogen atoms), and hydrogen bonds involving the polar character of oxygen and nitrogen atoms (and an intervening hydrogen).

Hydrophobic interactions are observed for the amino acids isoleucine 47 and methionine 54; they are highly populated but they open and close very frequently so that their lifetimes are short. In contrast, hydrogen bonds formed mostly by the amino acids glutamine 50 and asparagine 51 typically exist for more than 100 *ps*.

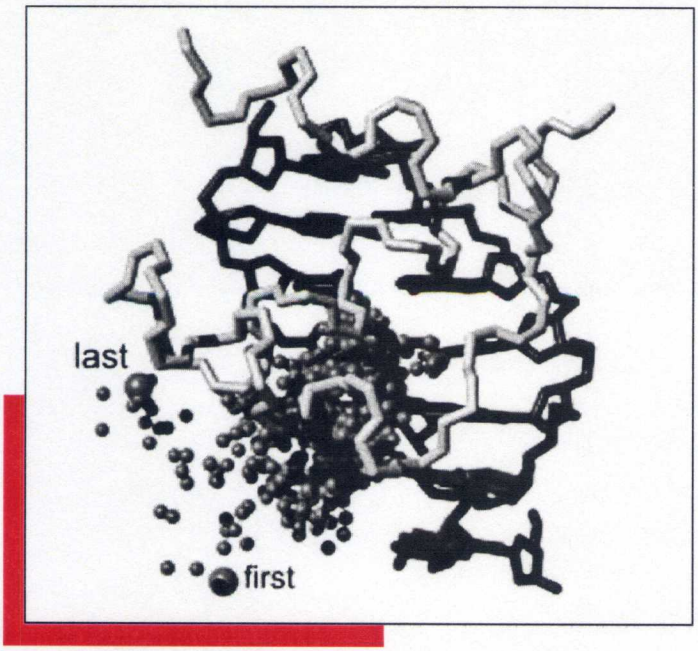


Figure 2: *Diffusion pathway of a water molecule in the interface of the homeodomain-DNA complex. The DNA is represented by dark bonds (bonds to hydrogen atoms are omitted for clarity), and the homeodomain backbone by bright bonds. The spheres indicate the various locations of a single water molecule during 700 ps of the MD simulation. The first and last spheres in this time period are identified. (This figure was prepared with the program MOLMOL [9]).*

These direct protein-DNA interactions are complemented by networks of hydrogen bonds mediated by water molecules in the macromolecular interface. For example, the water molecule presented in Figure 2 remains in the interface for a sufficiently long period to form hydrogen bonds to both the homeodomain and the DNA. In fact, for several hundred *ps* it simultaneously interacts with glutamine 50 and a DNA nucleotide, thus mediating an indirect protein-DNA contact. Such indirect interactions between the two macromolecules can have rather long lifetimes and be highly populated, although with varying water molecules.

Discussion

Specific interactions between different macromolecules, such as the recognition of specific DNA sequences by proteins, rely largely on the complementarity of the interacting surfaces. Complementarity of two molecules is not restricted to steric matching, but more generally includes favorable intermolecular interactions. The widely used assumptions of rigid interfaces in a complex may often not take proper account of entropy considerations. The increase in free energy that would result from complete immobilization of amino acid side chains on the protein surface upon interaction of this surface with another molecule can be a high prize to pay for ideal geometry of the intermolecular interactions. Our investigations using NMR spectroscopy and MD simulations combine atomic spatial resolution with temporal resolution over a wide range of frequencies.

The model of the *Antennapedia* homeodomain-DNA complex which emerges from the analysis of the presently described 2 *ns* MD trajectory is consistent with the NMR data. In addition, the trajectory provides a picture of possible pathways of water molecules in and out of the protein-DNA interface [5]. Further, it gives some leads for the interpretation of the range between the upper and lower bounds on the lifetimes of the hydration waters in the protein- DNA interface defined by the NMR measurements. The implication from the MD simulation is that the actual lifetimes of these waters are rather close to the lower limit of 1 *ns*. This high mobility implies constant interchange between a large number of different interaction networks, and we have to conclude that the observed specificity of the intermolecular recognition is in part the result of the ensemble of these rapidly interconverting non-bonding structures. The special role of the solvent must be largely based on the small size of the water molecule and its ability to form up to four hydrogen bonds. Water molecules appear to act not only as building blocks to improve the complementarity of the interaction surfaces of the protein and the DNA, but also as a lubricant to reduce entropic costs arising when a dense network of interactions is required for highly specific macromolecular recognition.

At present the continuation of our MD simulation project follows two lines. On the one hand we apply OPAL to biologically interesting molecules, such as the prion protein [8], which is associated with Creutzfeldt-Jakob disease (CJD) and bovine spongiform encephalopathy (BSE or "mad cow disease"), and various mutants of it. On the other hand we work on further development of OPAL, in particular also its implementation on parallel computers.

Acknowledgment

We thank the Swiss National Science Foundation for financial support (project 31.32033.91).

References

[1] Y.Q. Qian, M. Billeter, G. Otting, M. Müller, W.J. Gehring and K. Wüthrich, Cell 59, 573–580 (1989).

[2] G. Otting, Y.Q. Qian, M. Billeter, M. Müller, M. Affolter, W.J. Gehring and K. Wüthrich, EMBO J. 9, 3085–3092 (1990).

[3] M. Billeter, Y.Q. Qian, G. Otting, M. Müller, W. J. Gehring and K. Wüthrich, J. Mol. Biol. 234, 1084–1093 (1993).

[4] Y.Q. Qian, G. Otting and K.Wüthrich, J. Amer. Chem. Soc. 115, 1189–1190 (1993).

[5] M. Billeter, P. Güntert, P. Luginbühl and K. Wüthrich, Cell 85, 1057–1065 (1996).

[6] P. Luginbühl, P. Güntert, M. Billeter and K. Wüthrich, J. Biomol. NMR. 8, 136–146 (1996).

[7] W.J. Gehring, Y.Q. Qian, M. Billeter, K. Furukubo-Tokunaga, A.F. Schier, D. Resendez-Perez, M. Affolter, G. Otting and K. Wüthrich, Cell 78, 211–223 (1994).

[8] R. Riek, S. Hornemann, G. Wider, M. Billeter, R. Glockshuber and K. Wüthrich, Nature, 382, 180–182 (1996).

[9] R. Koradi, M. Billeter and K. Wüthrich, J. Mol. Graph. 14, 51–55, 29–32 (1996).