

Martin Frank · Peter Gutbrod · Chokri Hassayoun ·
Claus-W. von der Lieth

Dynamic molecules: molecular dynamics for everyone. An internet-based access to molecular dynamic simulations: basic concepts

Received: 26 February 2003 / Accepted: 6 June 2003 / Published online: 8 August 2003
© Springer-Verlag 2003

Abstract Molecular dynamics is a rapidly developing field of science and has become an established tool for studying the dynamic behavior of biomolecules. Although several high quality programs for performing molecular dynamic simulations are freely available, only well-trained scientists are currently able to make use of the broad scientific potential that molecular dynamic simulations offer to gain insight into structural questions at an atomic level. The “Dynamic Molecules” approach is the first internet portal that provides an interactive access to set up, perform and analyze molecular dynamic simulations. It is completely based on standard web technologies and uses only publicly available software. The aim is to open molecular dynamics techniques to a broader range of users including undergraduate students, teachers and scientists outside the bioinformatics field. The time-limiting factors are the availability of free capacity on the computing server to run the simulations and the time required to transport the history file through the internet for the animation mode. The interactive access mode of the portal is acceptable for animations of molecules having up to about 500 atoms.

Keywords Molecular dynamics · Internet-based application · Conformational space

Introduction

The first molecular dynamics (MD) simulation of a small protein under vacuum conditions was reported at the end of the seventies. [1] During the last 25 years, MD has been a rapidly developing field of science and has become an established tool to study the dynamic behavior of biomolecules, complementary to experimental techniques. [2, 3] Simulations are performed on increasingly

larger molecular systems using more realistic simulation conditions and better sampling procedures. Like all computational branches of science, MD simulations benefit from tremendous improvements in computer hardware: simulations that a few years ago required supercomputers can now be carried out using standard personal computers. Due to longer sampling times and improvements in simulation techniques, nowadays many biomolecular properties can be computed that were not within reach 10 years ago. Simulations to investigate the transport of molecules through transmembrane channels have become feasible. [4] Recently, it has been demonstrated that MD simulations are adequate for exploring the time scales and energies of protein folding, an important unsolved problem in molecular biology. [5] Today, virtually all spatial structures of biomolecules obtained by X-ray crystallography or NMR spectroscopy are MD refined. Simulations aid in understanding biochemical processes and add a dynamic dimension to structural data. For example, the transformation of harmless prion protein into the disease-causing agent has recently been modeled. [6]

Several high quality programs for performing MD simulations are available for academic users. The traditional MD programs CHARMM, [7] AMBER [8] and GROMOS, [9] which were originally developed during the eighties, are constantly updated and intensively applied to a broad spectrum of scientific questions. TINKER, [10] NAMD [11] and GROMACS (GROningen MACHine for Chemical Simulations) [12] are freely available software for academic institutions. NAMD and GROMACS are based on parallel, object-oriented code designed for high-performance simulations of large biomolecular systems. NAMD scales to hundreds of processors on high-end parallel platforms and tens of processors on commodity clusters using switched fast Ethernet. GROMACS is extremely fast at calculating the non-bonded interactions that usually dominate simulations. The TINKER molecular modeling software is a complete and general package for molecular mechanics and dynamics, with some special features for biopoly-

M. Frank · P. Gutbrod · C. Hassayoun · C.-W. von der Lieth (✉)
German Cancer Research Center,
Central Spectroscopic Department B090,
Im Neuenheimer Feld 280, 69120 Heidelberg, Germany
e-mail: w.vonderlieth@dkfz.de

mers. TINKER has the ability to use any of several common parameter sets, such as AMBER94/96, CHARMM27, MM2(1991), MM3(2000), OPLS-AA and OPLS-UA.

The main focus of all the MD programs mentioned is to calculate molecular ensembles for a variety of simulation conditions with the possibility to apply constraints derived from experiments. Additionally, most MD programs provide tools to support the statistical analysis of various molecular properties. No or only modest graphical support is available. Although all programs provide a user manual and examples, it is often not easy, especially for the novice user, to set up the desired simulation protocol and to generate the molecular ensemble of interest. As a consequence, only a limited number of well-trained scientists working in the biocomputing or bioinformatics fields are currently able to make use of the broad scientific potential that MD simulations offer to gain insight into structural questions at an atomic level.

The aim of the "Dynamic Molecules" project is to open MD techniques to a broader range of users including undergraduate students, teachers and scientists outside the bioinformatics field. It is obvious that techniques of this complexity cannot be used without a basic understanding of the underlying physical principles as well as the scope and limitations of the methodology used. Therefore, different levels of information and training to support novice as well as advanced users are provided. Virtual workshops serve to introduce the underlying physical principles and demonstrate the type of application accessible by MD simulations. The "Simulation Manager" allows the user to start and analyze MD simulations for a characteristic selection of molecules. Specific workshops guide novice users on how to set up simulations and to analyze the trajectories generated.

Materials and methods

One of the fundamental principles of the project is to build the system on software modules that are available free of charge or at moderate expense. This approach guaranties that a broad community can access the service and that the portal can be transferred to other platforms easily. Therefore, the whole system was built on a Unix-based operating system and Apache as web server. PHP, a widely-used general-purpose script language that is especially suited for Web applications and dynamically generates HTML code, is used to access the underlying relational MySQL database. JavaScript is used to support a dynamic user interaction within the browser. We support Internet Explorer, Mozilla and Netscape Communicator as web browsers. The only required plugins, which must be downloaded to obtain full functionality, are Chime (www.mdli.com) and a viewer capable of reading the Scalable Vector Graphics (SVG) format to display scientific plots. Chime can be regarded as the standard plugin for web-based applications to visualize molecular structures. It offers an intuitive easy-to-use interface for novice users. Additionally, it provides many useful functions such as the animation of ensembles of structures generated by MD simulations. SVG is a new open standard of a vector graphics language recommended by the World Wide Web Consortium (W3C) (www.w3.org/Graphics/SVG/Overview.htm8). Written in XML, SVG drawings can be dynamic and interactive and are resolution-independent. We currently use the Adobe SVG

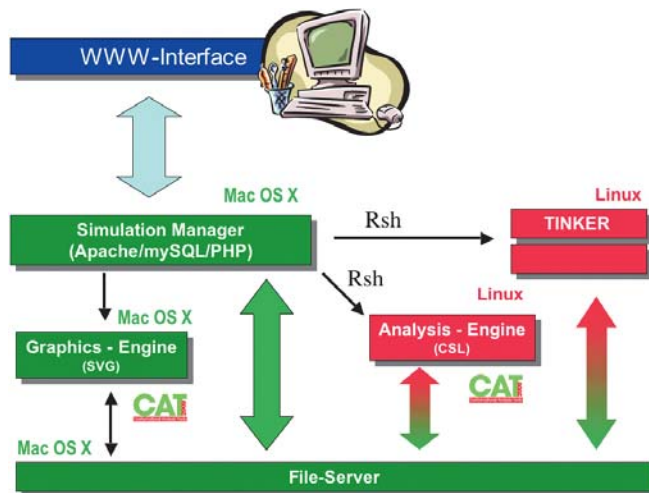


Fig. 1 Basic components and data flow

Viewer (www.adobe.com/svg/viewer/install/main.html) to display the scientific plots derived from the analysis of various trajectories.

The only browser-specific feature, which makes the input of internal coordinates to be analyzed more comfortable, is the LiveConnect feature of Chime available only for Netscape and Mozilla. LiveConnect creates a connection between HTML elements, Java, JavaScripts and plug-ins. The automatic analysis of ensembles of structures and generation of the SVG plots is accomplished with the program "Conformational Analysis Tools" (www.md-simulations.de/CAT/) which has been adopted to work smoothly together with the MD simulations web portal.

Figure 1 shows the basic elements of the "Dynamic Molecules" portal. The simulation manager is the central access point for users to start a new simulation, to monitor the current status of running jobs and to analyze ensembles of structures already generated. Molecules and simulation scripts are recalled from a pool of predefined entries and transferred to the Linux cluster, where the MD simulation is started. All files produced are copied to the file server. For each simulation a unique identification number is automatically generated. Bookkeeping of files is accomplished through a database, where job- and user-specific identifications for each file are stored. In case a specific simulation has already been performed, all files required are reloaded from the file server. The "Conformational Analysis Tools" are used to perform statistical analysis of an ensemble of structures and to generate the requested SVG plots. User interaction with the "Analysis Engine" is accomplished through a specific window of the simulation manager. The user has to indicate the atoms defining the internal coordinates (distances, angles, torsion angles) to be evaluated. The labeled atoms and the type of internal coordinates are transferred to a specific XML-based script language (CSL), which is subsequently executed by the "Conformational Analysis Tools software".

Hardware and software

Currently a PowerMac G4 running under Mac OSX is used as Web-Server. A Linux cluster with 18 processors (9x2 Athlon 2000+ dual processors) designed by Megaware, Chemnitz (www.megaware.de) carries out the time-consuming MD simulations. Linux RedHat 7.3 (www.redhat.com) is the operating system and the freely available portable batch system (www.openpbs.org) serves to organize an efficient distribution of the submitted jobs between the available processors. Currently, the

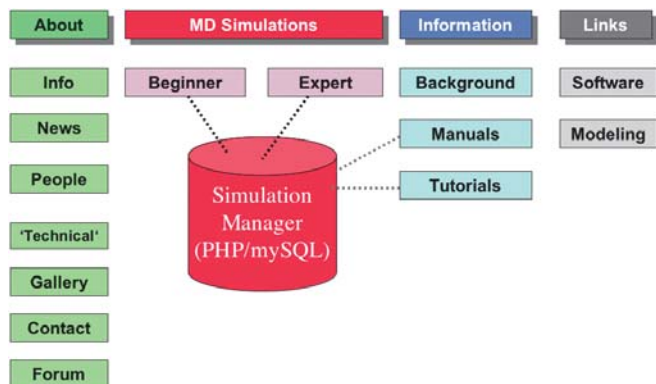


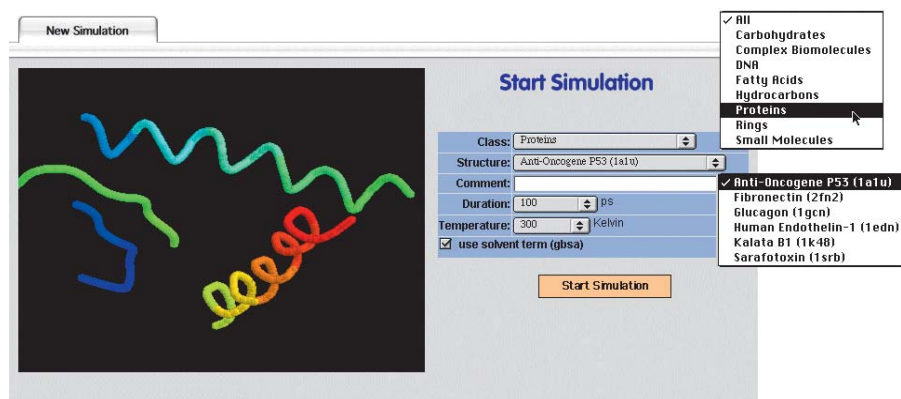
Fig. 2 “Dynamic Molecules” web portal: currently available features

dynamics module of the TINKER software suite is used to perform MD simulations based on the MM3(2000) force field for organic molecules and AMBER94 for proteins and DNA.

Results and discussion

The currently available features of the “Dynamic Molecules” homepage (see Fig. 2) are general and technical data describing the project and its progress (“About”), the interface which allows the user to set up, start and analyze MD simulations interactively (“Simulation Manager”), several levels to access scientific data (“Information”) and related web sites and software tools (“Links”). Additionally, several levels of education and user interactions are implemented to provide scientific introductions for beginners and more detailed technical support for advanced users. The main purpose of the workshops is to introduce the methods applied and to demonstrate the type of scientific questions that can be addressed using MD simulations. Manuals describe in detail how to set up specific simulations. A forum for discussions has been created to support the scientific communication between users.

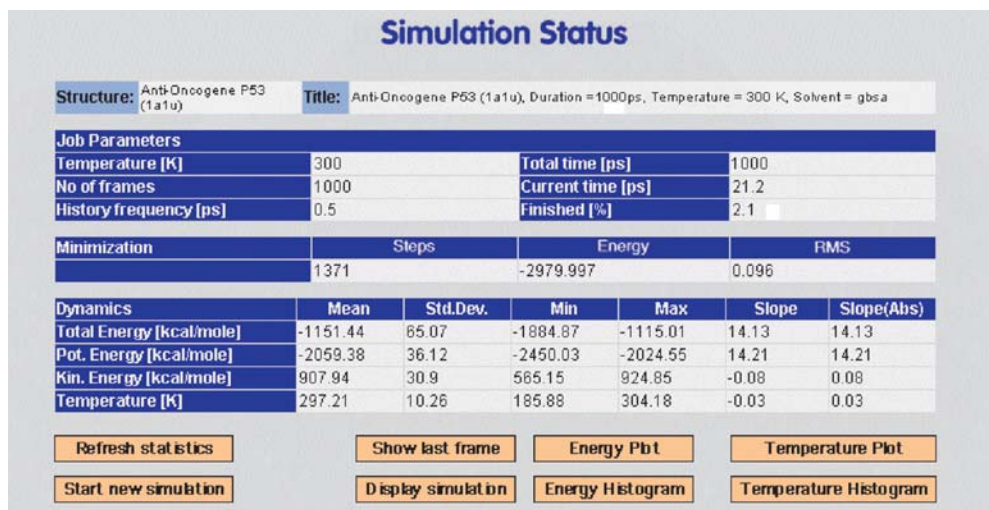
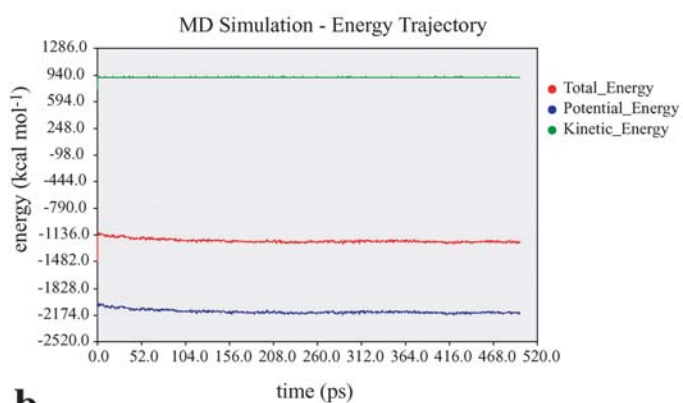
Fig. 3 Start simulation interface. A 3D structure is recalled from a predefined set of molecules. The simulation conditions have to be chosen by the user



The “Simulation Manager”, which is the central feature of the “Dynamic Molecules” portal, provides a so-called “Beginner Mode” focusing on introducing the methods for novice users and an “Expert Mode” which will enable the user to perform complete research projects. Since the basic concepts of the two modes are quite similar, we concentrate on describing the features implemented for the “Beginner Mode” in more detail.

The first step is to input the starting geometry of the molecule to be analyzed. The “Beginner Mode” provides access to the spatial structure of a representative set of predefined molecules. Currently, small organic compounds, DNA molecules in different conformations, peptides and small proteins as well as some biologically important sugars can be selected by indicating their chemical name (see Fig. 3). The second step is to define the simulation conditions such as temperature, duration of the simulation, description of the surroundings under which the molecules will be evaluated and to start the simulation. The simulation manager has access to a MySQL database where all the user interactions, simulation scripts and history files are stored. The “Beginner Mode” provides access to only a limited number of predefined molecules. It is checked whether the simulation in question has already been performed previously. If this is the case, the history files are recalled from the database. Otherwise the simulation is started on the Linux cluster. The user can control the status of the submitted simulations through the job status panel (see Fig. 4a). On demand, a plot of the development of the total energy (Fig. 4b) and temperature (Fig. 4c) of the active simulation can be displayed while the simulation is still running. Additionally, the 3D coordinates of the last frame can be displayed in a Chime window (Fig. 4d).

In a subsequent step, the molecular ensemble of structures generated must be analyzed and interpreted according to specific questions of scientific interest. Since it may happen that the applied simulation conditions are not adequately chosen for a certain molecular system and therefore rather meaningless structures are produced, it is necessary to check the validity of each simulation. A useful way is to start an animation (“Display Simulation”) (see Fig. 5) and to inspect visually if the structures

**a****b**

MD Simulation - Temperature

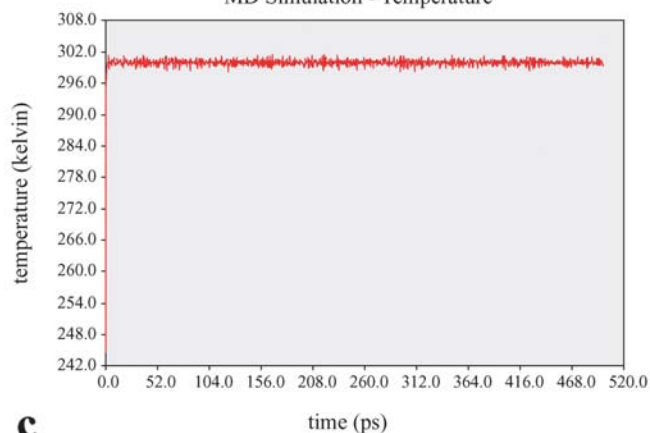
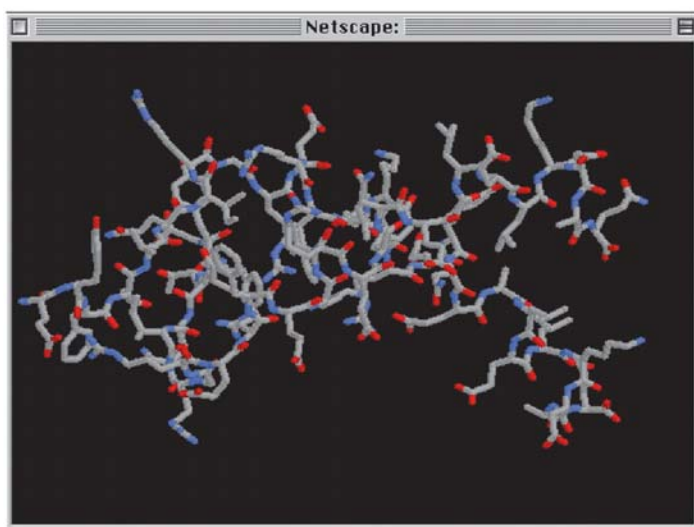
**c****d**

Fig. 4 a User control panel of a submitted MD simulation. Statistics for the temperature and the energy of the simulated

energy (b) and temperature (c) can be displayed. Additionally, the 3D coordinates of the last frame can be displayed in a Chime window (d)

Fig. 5 Animation of an MD simulation. To reduce the number of frames to be transferred through the Internet, the user can activate only certain parts of the total history file (here the frames between 500 and 700) or/and raise the increment step size. A toggle allows the user to stop the automatic animation and to scroll back and forth through the ensemble of loaded structures manually

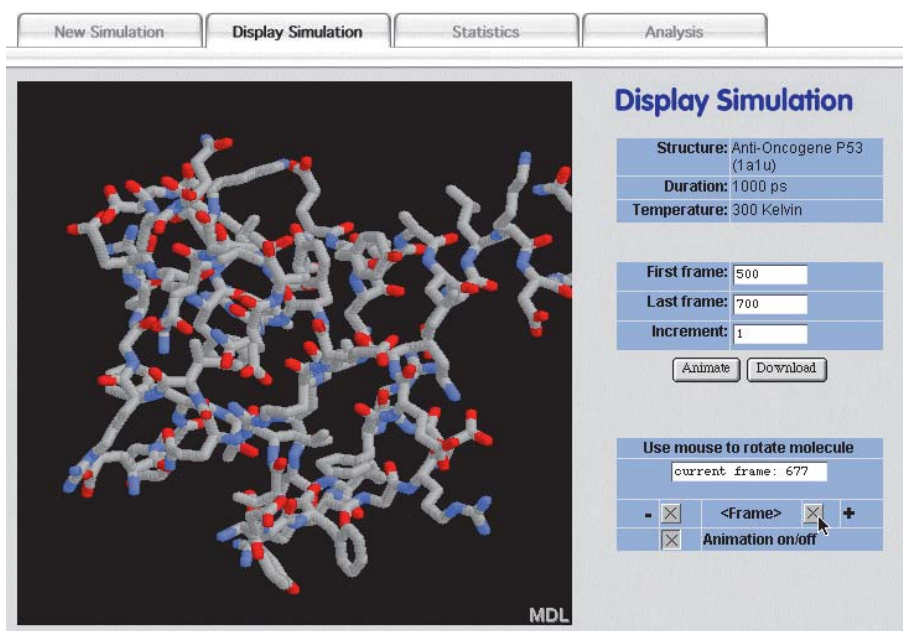
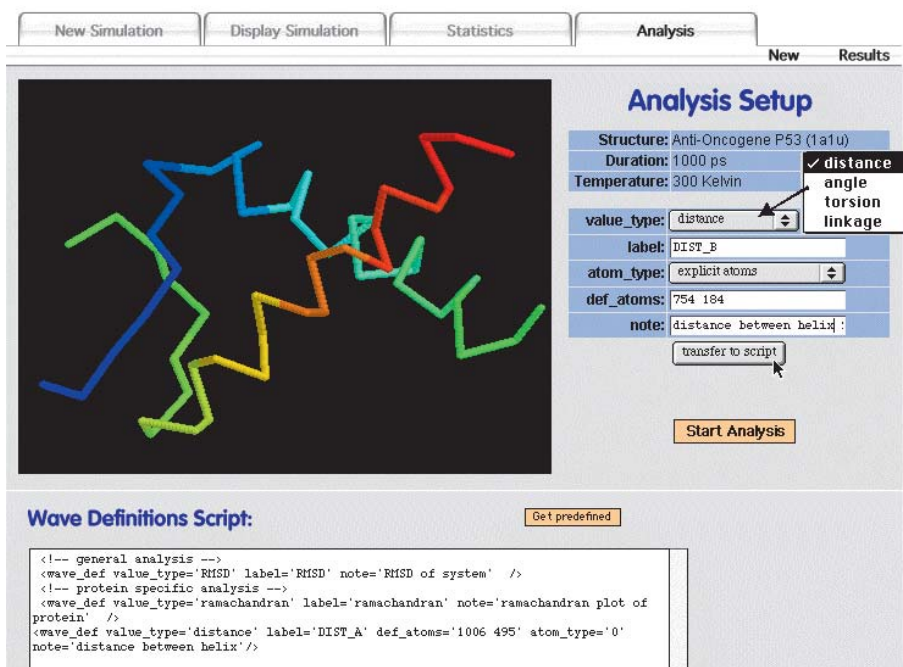


Fig. 6 Assignment of internal coordinates to be analyzed. First, the type of internal coordinate and a unique label has to be chosen. Second, the numbers of the atoms defining the internal coordinates have to be input. The “transfer to script” option adds commands (called wave definitions) into a script file, which is interpreted when the analysis is started



generated correspond to chemical common sense. Another mandatory check is to control the overall progress of temperature and total energy content for each simulation (“Statistics”).

The “Analysis” menu enables the user to evaluate the dynamic behavior of some specific internal coordinates such as distances, angles and dihedral angles. Their assignment is as follows (see Fig. 6). First, the type of internal coordinate and a unique label must be chosen. Second, the atom numbers defining the internal coordinates must be input. The atoms can be identified by

mouse-clicking onto the desired atom in the Chime window. Additionally, comments can be added that may help to document and organize series of simulations. The “transfer to script” option adds commands (called wave definitions) into a script file, which is read and interpreted by the “Conformational Analysis Tools” software when the analysis is started. For carbohydrates a list of predefined wave definitions is provided, which enables the user to evaluate all relevant internal coordinates with one command. The results of the analysis (see Fig. 7a) can be downloaded as tabulated data or as graphical

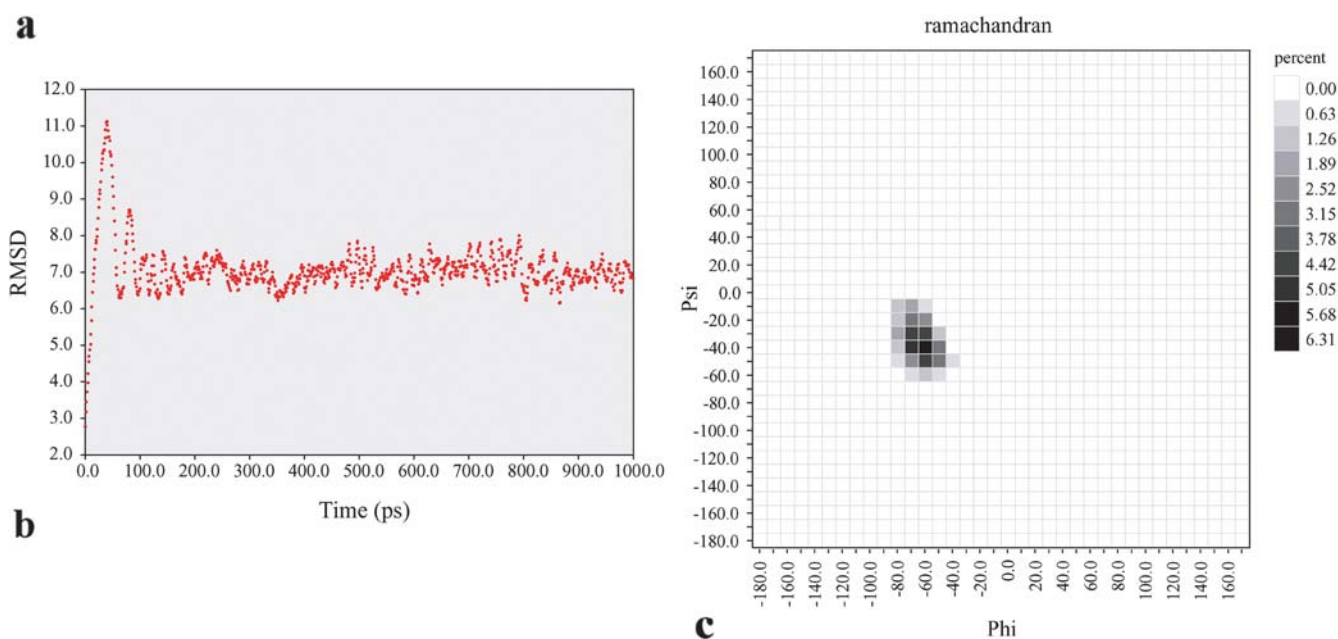
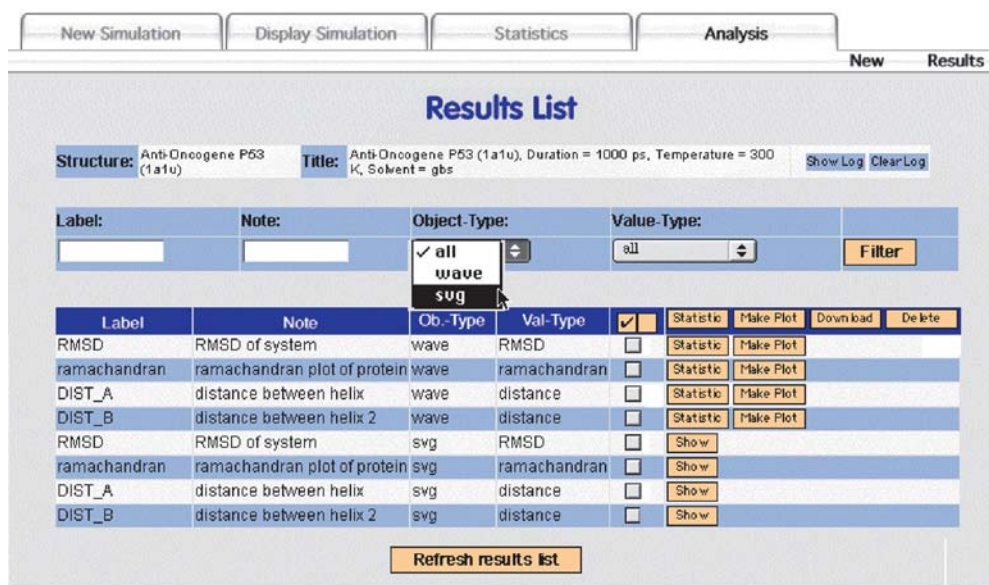


Fig. 7 **a** List of the calculated trajectories and plots generated for the anti-oncogene P53. The object type “svg” can be directly displayed. Files having the object type “wave” contain tabulated numerical data of a trajectory and can be combined with other waves to produce a new SVG plot. The plotted trajectory of the

RMSD of all atoms **b** indicates that a major conformational reorientation occurs at an early stage and that the resulting structure remains stable for the rest of the simulation. However, the Ramachandran plot of the Φ , Ψ backbone angles shows that the dominating α -helical structure for all residues is maintained (**c**)

representations, which can be displayed as SVG plots (see Fig. 7b–c). The “linkage” option (see Fig. 8a) produces a Ramachandran-like plot, which describes the population statistics of two successive torsion angles calculated from an MD simulation. Such maps are often used to visualize the conformational preferences of specific glycosidic linkages of carbohydrates (see Fig. 8b).

To perform MD simulations over the Internet, we are faced with the problem that the basic calculations may require rather long simulation times—hours, days, even

weeks of wall clock time—which are far beyond any interactive user interactions. On the other hand, scientists want to analyze the trajectories produced under different aspects according to their specific interests. It is often impossible to define beforehand all geometric parameters required to analyze a specific scientific question. For this purpose interactive user interactions are absolutely necessary. It is not sufficient to inform the user when the submitted simulation has been finished and transfer some standard evaluation results by email. Therefore, we have

Fig. 8 a For well-defined structural features which are routinely analyzed like the glycosidic angles of saccharides predefined wave definitions are provided, which can be activated through the “get predefined” button. **b** A conformational map (relative population density) of the glycosidic dihedral angle Φ, Ψ of lactose ($T=300$ K, duration 1,000 ps). The linkage option is used to produce a Ramachandran-like map. It is obvious that lactose populates two main conformations at Φ, Ψ of $20^\circ/-40^\circ$ and $50^\circ/0^\circ$

Wave Definitions Script: Get predefined

```

<!-- ----- Carbohydrate Linkages ----- -->
<wave_def value_type='linkage' label='GLB2_14L' def_atoms='25 23 24 16 17'
atom_type='0' note='1-4-linkage between GLB2 and GLC1' />

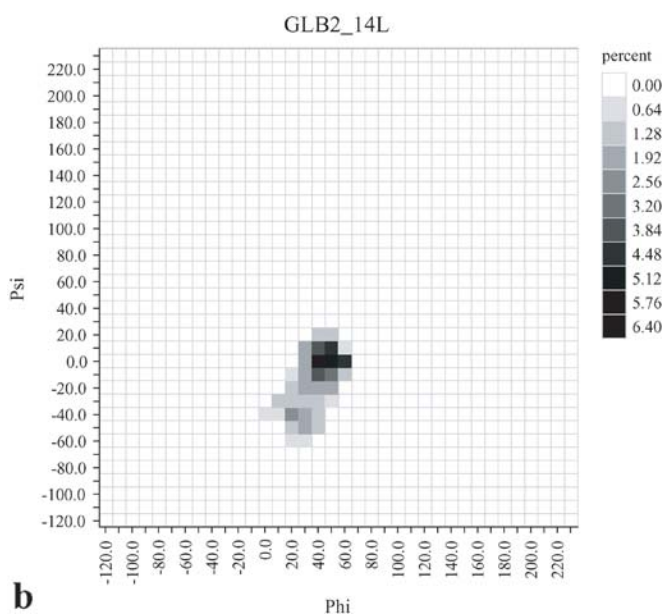
<!-- ----- phi/psi/omega Torsions ----- -->

<wave_def value_type='torsion' label='GLB2_14F' note='phi' atom_type='0' def_atoms=
'25 23 24 16' />
<wave_def value_type='torsion' label='GLB2_14Y' note='psi' atom_type='0' def_atoms=
'23 24 16 17' />
<wave_def value_type='torsion' label='GLB2_W' note='omega' atom_type='0' def_atoms=
'27 26 41 42' />
<wave_def value_type='torsion' label='GLC1_W' note='omega' atom_type='0' def_atoms=
'6 5 18 19' />

<!-- ----- OH Torsions ----- -->
<wave_def value_type='torsion' label='GLB2_20H' atom_type='0' note='OH-torsion'
def_atoms='32 29 30 31' />
<wave_def value_type='torsion' label='GLB2_30H' atom_type='0' note='OH-torsion'
def_atoms='36 33 34 35' />

```

a



included options that enable the user to separate the process to set up and start simulations from the process to analyze the data produced. Since the beginner mode is open to public access without any specific login procedure, users must be identified using automatic ways of identification. Practically, this is accomplished by setting hardware-specific cookies, which are recognized when the next call from the same hardware platform is identified. In such a way users can continue to analyze completed jobs at any time they log in. Thus, it is also guaranteed that each user will have access only to his list of results.

Summary and conclusion

“Dynamics Molecules” is the first internet portal that provides an interactive access to set up, perform and analyze molecular dynamic simulations. It is completely based on standard web technologies and uses only publicly available software. The “Beginner Mode”, which provides access to a characteristic selection of biologically relevant molecules, aims to demonstrate the size of molecules and the type of scientific questions that can be approached using the “Dynamic Molecules” portal. Undergraduate students, teachers and lecturers will be the primary users of this mode. It is the hope that the possibility to set up, run and analyze MD simulations using web-based interfaces and publicly available plugins will increase the knowledge about this computational

Table 1 Size of the uncompressed history files (1,000 frames in XYZ format) required to animate the simulated molecules

Molecule	Number of atoms	Size of XYZ file (MB)
Methanol	6	0.179
Butan-1-ol	15	0.432
Lactose	42	1.27
LewisX	71	2.00
GM1-Cer	228	6.39
Peptid	334	9.36
DNA-DS 10 bp	639	17.90
Protein	1,022	27.63

technique and thus encourage more people to apply it to their specific scientific problems.

The time-limiting factors are the availability of free capacity on the computing server to run the simulation and the time to transport the history file through the internet for the analysis of molecular ensembles. Since all history files generated for molecules available in the “Beginner Mode” are stored in the simulation database and will be recalled when the same simulation conditions are requested again, the necessary number of simulations to be performed will decrease with the progress of the project. The size of the history files to be download through the Internet depends on the number of atoms and the number of stored conformations—often called frames—during the simulation. For the “Beginner Mode” we decided always to store 1,000 frames independent of the simulation duration chosen by the user. Table 1 shows how the size of the uncompressed history files (in XYZ format as required by Chime for animations) increases with the number of atoms. Assuming a download speed of 50 kb/s, which is normally accessible by using DSL technology, 1,000 frames to animate molecules comprising between 50 and 100 atoms can be downloaded within 1 min. Standard compression procedures reduce the size of the history file to about one third. Therefore, a compressed format of the history file can be downloaded for larger molecules.

Biological molecules exhibit a wide range of time scales over which specific processes occur. Local motions (atomic fluctuations, sidechain motions) take place in the range of 10^{-15} to 10^{-9} s and are accessible with current standard hardware, whereas large-scale motions such as folding/unfolding of proteins proceed in the range of 10^{-7} to 10^4 s and are often not even approachable using state of

the art high performance computing facilities. Simulation periods in the range of 10^{-9} s with inclusion of explicit water molecules can now be regarded as standard for proteins. The “Dynamic Molecules” portal allows access to this time scale for isolated molecules applying a continuum model to mimic the influence of the aqueous environment.

We are currently working to set up the “expert mode”, which will allow us to work with user-definable molecules, to choose between a broad variety of commonly used scripts to perform MD simulations and many more and advanced functions to analyze trajectories. This mode is designed in such a way that trained people will be able to perform complete scientific projects based on Internet technologies.

Acknowledgement The “Dynamic Molecules” project is funded by a grant from the German Research Net (Deutsches Forschungsnetz: DFN) and the Federal Ministry of Education and Research in Germany (BMBF).

References

1. McCammon JA, Gelin BR, Karplus M (1977) *Nature* 267:585–590
2. Hansson T, Oostenbrink C, van Gunsteren W (2002) *Curr Opin Struct Biol* 12:190–196
3. Lee MR, Tsai J, Baker D, Kollman PA (2001) *J Mol Biol* 313:417–430
4. Tajkhorshid E, Nollert P, Jensen MO, Miercke LJ, O’Connell J, Stroud RM, Schulten K (2002) *Science* 296:525–530
5. Snow CD, Nguyen H, Pande VS, Gruebele M (2002) *Nature* 420:102–106
6. Billeter M, Wuthrich K (2000) *Arch Virol Suppl* 251–263
7. MacKerell Jr AD, Brooks B, Brooks III CL, Nilsson L, Roux LB, Won Y, Karplus M (1998) CHARMM: the energy function and its parameterization with an overview of the program. In: Schleyer PvR, Allinger NL, Clark T, Gasteiger J, Kollman PA, Schaefer III HF (eds) *The encyclopedia of computational chemistry*, 1st edn. Wiley, Chichester, pp 271–277
8. Cornell WD, Cieplak P, Bayly CI, Gould IR, Merz K, Ferguson DM, Spellmeyer DC, Fox T, Caldwell JW, Kollman PA (1995) *J Am Chem Soc* 117:5179–5197
9. Stocker U, van Gunsteren WF (2000) *Proteins* 40:145–153
10. Ren P, Ponder JW (2002) *J Comput Chem* 1497–1506
11. Kalé L, Skeel R, Bhandarkar M, Brunner R, Gursoy A, Krawetz N, Phillips J, Shinozaki A, Varadarajan K, Schulten K (1999) *J Comput Phys* 151:283–312
12. Lindahl D, Hess B, van der Spoel D (2001) *J Mol Model* 7:306–317