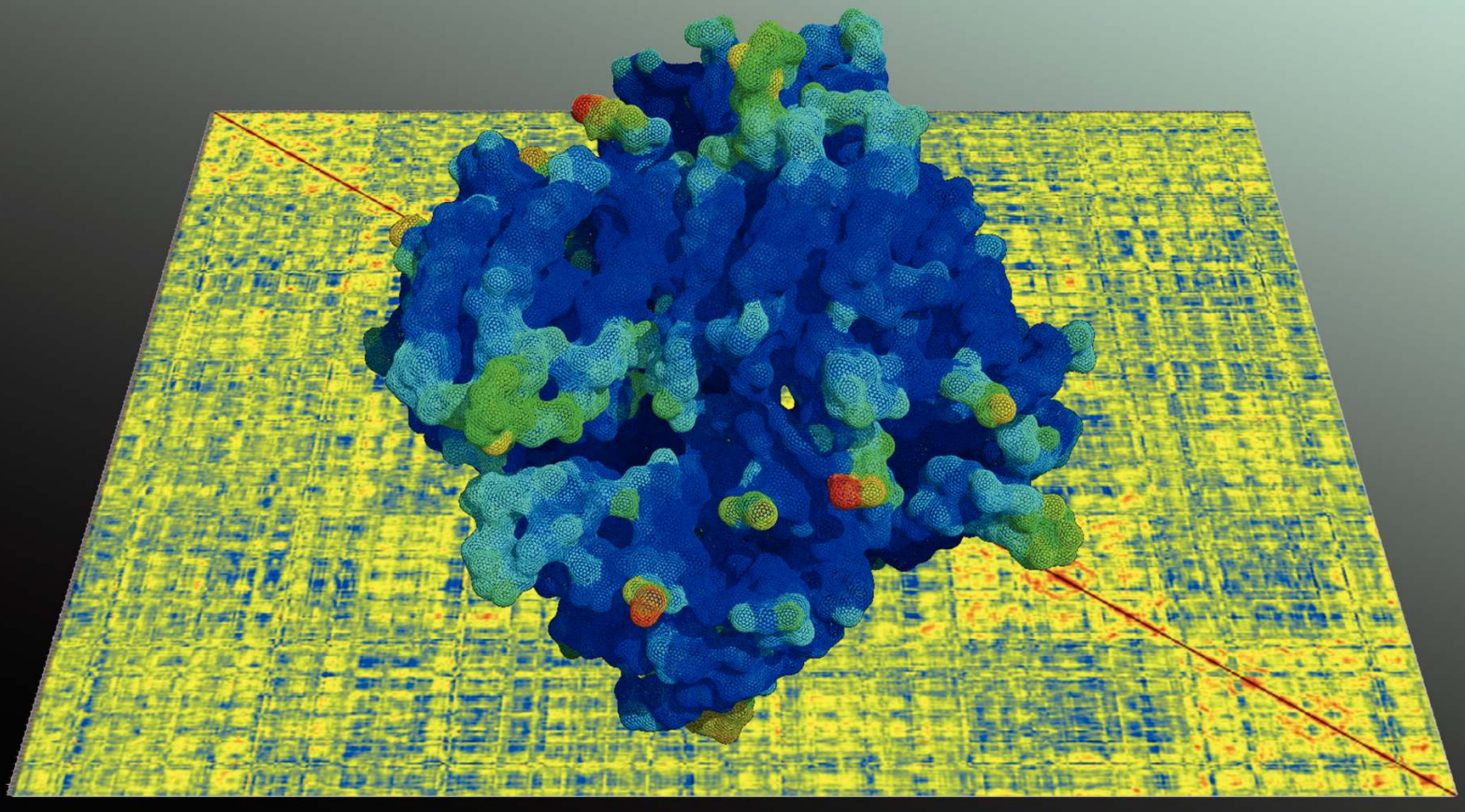
Volume 34 | Issues 25–26 | 2013
Included in this print edition:
Issue 25 (September 30, 2013)
Issue 26 (October 5, 2013)

Journal of COMPUTATIONAL CHEMISTRY

Organic • Inorganic • Physical
Biological • Materials

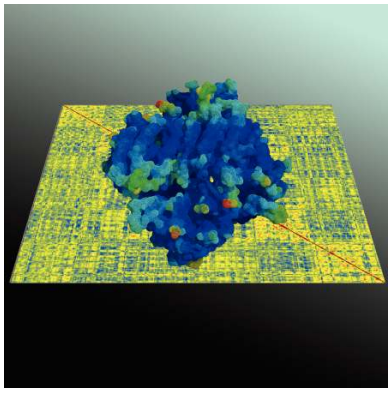
www.c-chem.org



Editors:
Charles L. Brooks III • Masahiro Ehara • Gernot Frenking • Peter R. Schreiner

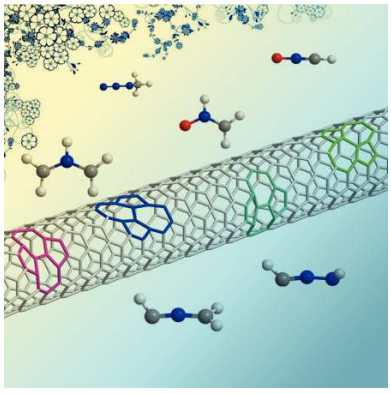
WILEY

Journal of COMPUTATIONAL CHEMISTRY
Organic • Inorganic • Physical
Biological • Materials



Principal Component Analysis

The cover shows an artistic rendering of some of the results obtained through the application of the program grcarma by Panagiotis Koukos and Nicholas Glykos on page 2310. The background surface is a color representation of the variance–covariance matrix of LmbE, a homohexameric protein. The molecular model in the foreground shows a dimer of the same protein colored according to its atomic root mean squared fluctuations.



Cycloaddition

The cover depicts a Stone–Wales (SW) defective single-walled carbon nanotube (SWNT) surrounded with several 1,3-dipoles such as azomethine ylide and nitrone. Using a two-layered ONIOM method by Tao Yang, Xiang Zhao, and Shigeru Nagase on page 2223, 1,3-dipolar cycloadditions of several 1,3-dipoles upon SW defective and defect-free SWNTs are systematically studied. Compared with defect-free SWNTs, the SW defective SWNTs exhibit the higher chemical reactivity. The larger the diameter of the SWNT, the lower chemical reactivity the SW defective SWNT shows.

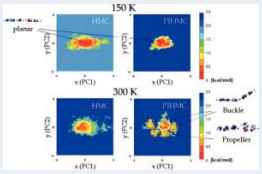
Coming Soon

Look for these important papers in upcoming issues

Nuclear quantum effect and temperature dependency on the hydrogen-bonded structure of base pairs

Masashi Daido et al.

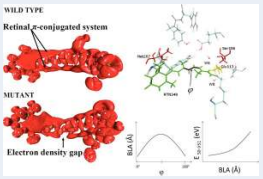
The ratio of buckle mode (PC1) for Watson–Crick-type guanine–cytosine pairs with quantum simulation decreases compared to that with classical simulation, whereas that of propeller mode (PC2) with quantum simulation increases. Nonplanar structures of base pairs become stable due to the nuclear quantum effect. DOI: 10.1002/jcc.23399



Electron density deformations provide new insights into the spectral shift of rhodopsins

Erix Wiliam Hernández-Rodríguez et al.

The method of deformed atoms in molecules (DAM) is used to study the electron distribution in 11-*cis*-retinal in bovine and human wild-type rhodopsins and two mutations associated with retinitis pigmentosa disease: M207R and S186W. This study provides new insights into mechanisms that regulate spectral shifts in rhodopsin and its mutants. DOI: 10.1002/jcc.23414



© 2013 Wiley Periodicals, Inc.

Grcarma: A Fully Automated Task-Oriented Interface for the Analysis of Molecular Dynamics Trajectories

Panagiotis I. Koukos and Nicholas M. Glykos*

We report the availability of grcarma, a program encoding for a fully automated set of tasks aiming to simplify the analysis of molecular dynamics trajectories of biological macromolecules. It is a cross-platform, Perl/Tk-based front-end to the program carma and is designed to facilitate the needs of the novice as well as those of the expert user, while at the same time maintaining a user-friendly and intuitive design. Particular emphasis was given to the automation of several tedious tasks, such as extraction of clusters of structures based on dihedral and Cartesian principal component analysis, secondary structure analysis, calculation and display of root-mean-square deviation (RMSD) matrices, calculation of entropy, calculation and analysis of variance–covariance matrices, calculation of the fraction of native contacts, etc. The program is free-open source software available immediately for download.
© 2013 Wiley Periodicals, Inc.

DOI: 10.1002/jcc.23381

Introduction

Performing biomolecular simulations of relatively simple macromolecular systems (for example a protein without any modifications plus surrounding solvent and ions using periodic boundary conditions and with full electrostatics) has in recent years become so straightforward, that even groups without extensive experience on the subject can—and do—use molecular dynamics simulations as a more-or-less black box tool in an attempt to answer biologically relevant questions [1–3]. This trend of popularizing molecular dynamics simulations immediately creates a need for highly automated analysis tools that can hide the technical details away from the final user while maintaining a formal correctness for the underlying analyses [4–6]. One way of combining user friendliness with formal correctness of the underlying analyses is through the application of fully automated pipelines which encode for accepted—in the field—analysis procedures with as little as possible user intervention [7–9]. In this spirit, we have developed grcarma, a set of pipelines built around the program carma [10] which encode for some of the most popular techniques for the analysis of biomolecular simulations. For an example of the automation afforded by the program, we note that only two clicks with the cursor in grcarma's graphical user interface suffice for the user to (a) perform a dihedral principal component analysis of the whole trajectory, (b) perform a cluster analysis in the space defined by the top three principal components, (c) isolate the frames corresponding to each cluster, (d) for each cluster perform least-squares fitting on the backbone atoms, and finally, (e) isolate—on a per cluster basis—average structures, representative structures as well as a superposition of 500 equally spaced structures from each cluster (with the structures colored according to the corresponding atomic root mean squared fluctuations (RMSF), see Fig. 2d for an example). Although automation was the guiding principle in designing grcarma, we believe that the simplicity and leanness of its interface can make it attractive to experienced users as well.

Program Description

grcarma is written in Perl and makes use of the Tk module for graphics. It is a lightweight program with a small memory footprint and negligible CPU requirements. It is tuned for the DCD/PSF world and is immediately compatible with all molecular dynamics programs that produce such trajectory files, for example CHARMM [11] or NAMD [12], without any need for trajectory format conversions. The user interface is minimalistic and consists of a single main window containing the tasks available (Fig. 1a) plus a text box for informational messages (see Fig. 1, underneath the window marked as b). The main way of interacting with grcarma's main window is through the list of buttons on the left panel of the main window (Fig. 1a). Each of those buttons represents a particular task that can be performed and requires a set of flags along with their arguments to be passed to carma (see Fig. 1b for an example based on Cartesian principal component analysis [PCA]).

The list of procedures encoded in grcarma are outlined in Table 1 (compare also with Fig. 1a) and will not be discussed in detail. Most of these procedures are fully automated in the sense that suitable default values are available for most (if not all) of the required parameters. Some specific examples will be presented below. The dihedral PCA and Cartesian PCA tasks allow the automatic isolation of a user specified number of clusters, the creation of separate DCD files containing the frames belonging to the various clusters, fitting of the DCD files for the removal of global rotation–translations, and finally the creation of three PDB files for each cluster. The first PDB file contains the average structure, the second is the cluster's representative structure, while the third is a superposition of 500 equally spaced structures from the given

P. I. Koukos, N. M. Glykos
Department of Molecular Biology and Genetics, Democritus University of Thrace, University Campus, 68100 Alexandroupolis, Greece
Fax: (+30) 25510 30620 E-mail: glykos@mbg.duth.gr
© 2013 Wiley Periodicals, Inc.

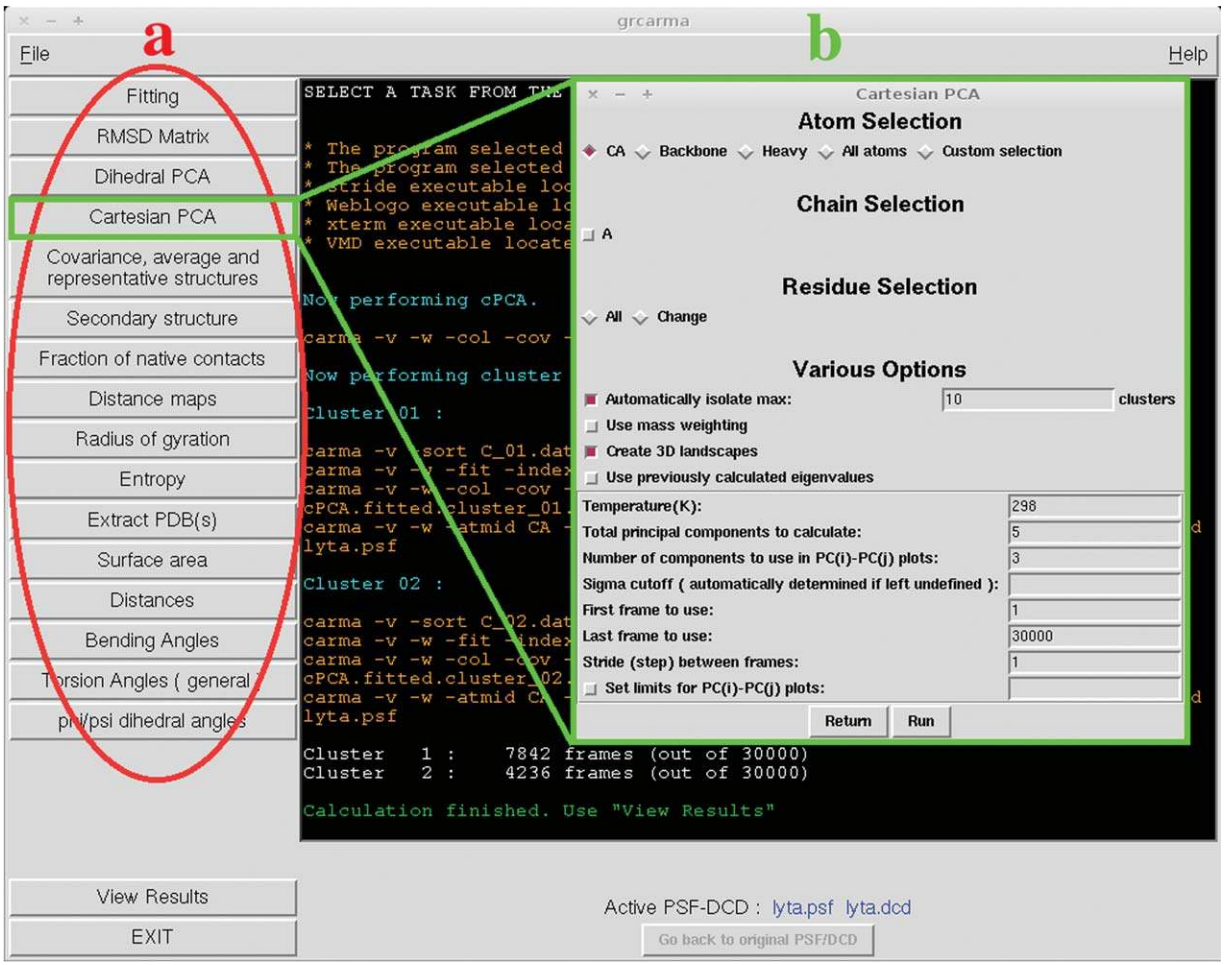


Figure 1. The main window of grcarma. On the left, the task selection panel marked as (a) is clearly visible. The window marked as (b) is a typical task window, in this case for the Cartesian PCA analysis. Various elements are visible including (from top to bottom) the atom, chain and residue selection options, task-specific elements (such as the number of principal components to calculate), followed by specifications for the first and last frames to include in the analysis as well as the step (stride) between the frames. Underneath the window marked as (b), the text box with informational messages is visible.

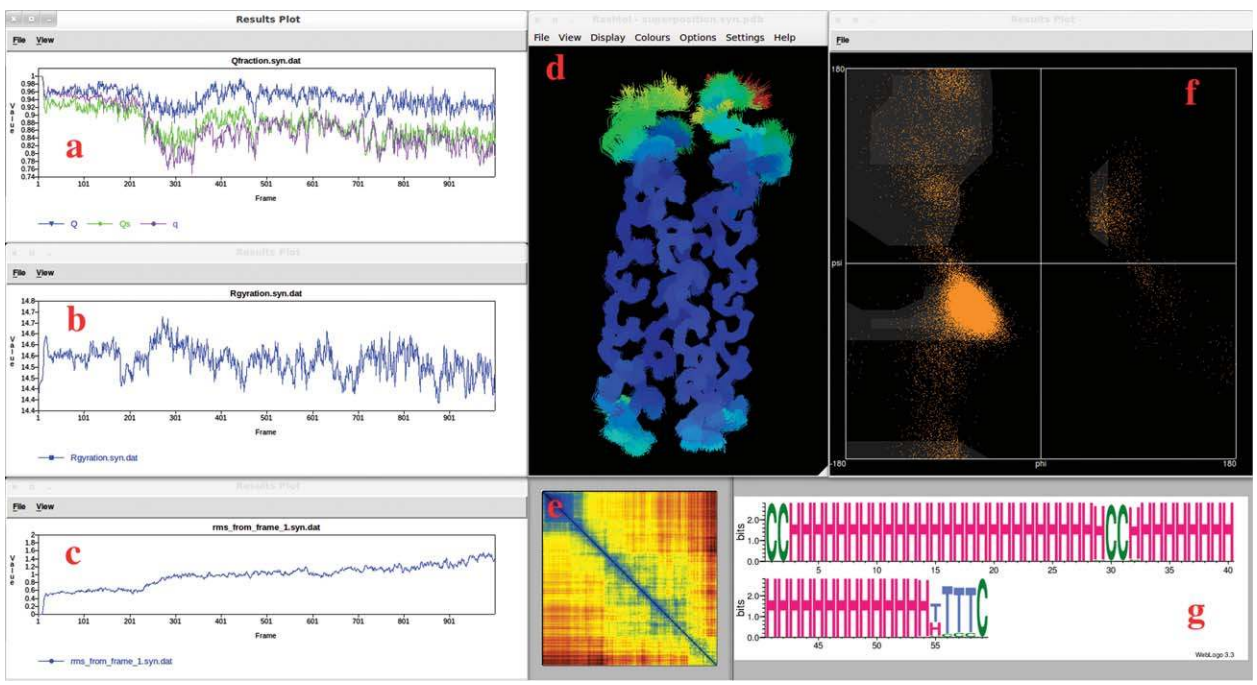


Figure 2. A collage of results produced through the application of grcarma. (a) Evolution (versus frame number) of the fraction of native contacts using three similarity measures. (b) Evolution of the radius of gyration. (c) Evolution of the rms deviation from the first frame of the trajectory. (d) Representation of the backbone atoms of a cluster (produced through Cartesian PCA) colored according to atomic RMSFs. (e) Color representation of the frame-to-frame RMSD matrix. (f) Ramachandran plot of selected residues of the molecule. (g) Weblogo [14,15] representation of the secondary structure assignments produced by STRIDE [13].

Table 1. Analysis pipelines implemented in grcarma.

Task	Description
Fitting	Removal of global rotations–translations and superposition of every frame on the first (or a user-selected frame) using any combination of atom-residue-chain selections.
RMSD matrix	Calculation and color representation of frame-to-frame RMSD matrices using user-selected number and type of atoms and residues
Dihedral/Cartesian PCA	Dihedral/Cartesian principal component analysis of the trajectory including automatic isolation and characterization of structurally prominent clusters.
Covariance, average, and representative structures	Calculation of the variance–covariance matrix, and production of PDB files depicting the average structure, the representative structure as well as a superposition of 500 structures from the trajectory.
Secondary structure	Secondary structure assignments (with STRIDE [13]) and color representations of the results as a function of simulation time and with a weblogo [14,15] representation.
Fraction of native contacts	Calculation of fraction of native contacts in the form of three similarity measures: Q, Qs, and q [16].
Distance maps	Calculation and rendering of distance maps together with their corresponding RMS deviations for selected atomic species.
Radius of gyration	Calculation of the mass-weighted radius of gyration for selected atomic species.
Entropy	Recursive calculation of the entropy using a progressively greater number of frames.
Extract PDB(s)	Extraction of PDB files from the trajectory.
Surface area	Calculation of a metric related to the total surface area of selected atomic species.
Distances	Calculation of distances between two specified atoms.
Bending angles	Calculation of the bending angles between three specified atoms.
Torsion angles (general)	Calculation of the torsion angle between four specified atoms.
Phi/psi dihedral angles	Calculation of the phi/psi dihedral angles of a molecule.

cluster. For another example, the systematic calculation of the solute's entropy is another task that has been completely automated with grcarma. The task repeatedly calculates the solute entropy using two well established procedures based on PCA [17,18]. Entropy is calculated several times, every time including more frames in the analysis than the previous, until the calculation has been performed for all frames contained in the trajectory. The task is so heavily automated that the only required parameter is the temperature at which the simulation was performed.

To give an idea of the wealth of information that grcarma can produce automatically, Figure 2 shows a collage of a small selection of results obtained through its application. These range from graphs showing the evolution (versus simulation time) of the fraction of native contacts (Fig. 2a), to calculation of the frame-to-frame RMSD matrix (Fig. 2e), to calculation of superimposed structures belonging to a cluster derived from the trajectory and colored according to the RMSFs of the respective atoms (Fig. 2d).

In summary, we believe that the automation of the tasks included in grcarma, together with its simple and intuitive design, will make this program a useful addition to the armory of analysis tools available to the macromolecular simulation community.

Program Availability

grcarma is free and open source software, published under a Simplified BSD License. grcarma is cross-platform supporting the majority of recent versions of Windows, GNU/Linux, and MacOSX operating systems. The source code of the program together with ready-made executables for all major architectures are available via both the github (https://github.com/pkoukos/grcarma) and sourceforge (http://sourceforge.net/projects/grcarma/) repositories.

Keywords: molecular dynamics • principal component analysis • graphical user interface

How to cite this article: P. I. Koukos, N. M. Glykos. *J. Comput. Chem.* **2013**, *34*, 2310–2312. DOI: 10.1002/jcc.23381

[1] S. A. Adcock, J. A. McCammon, *Chem. Rev.* **2006**, *106*, 1589.
[2] G. G. Dodson, D. P. Lane, C. S. Verma, *EMBO Rep.* **2008**, *9*, 144.
[3] J. D. Durrant, J. A. McCammon, *BMC Biol.* **2011**, *9*, 71.
[4] W. Humphrey, A. Dalke, K. Schulten, *J. Mol. Graphics* **1996**, *14*, 33.
[5] D. Sellis, D. Vlachakis, M. Vlassi, *Bioinformatics Biol. Insights* **2009**, *3*, 99.
[6] P. Kota, *In Silico Biol.* **2006**, *7*, 0008.
[7] A. Hospital, P. Andrio, C. Fenollosa, D. Cicin-Sain, M. Orozco, J. L. Gelpi, *Bioinformatics* **2012**, *28*, 1278.
[8] A. V. Popov, Y. N. Vorobjev, D. O. Zharkov, *J. Comput. Chem.* **2013**, *34*, 319.
[9] S. Jo, W. Jiang, H. S. Lee, B. Roux, W. Im, *J. Chem. Inf. Model.* **2013**, *53*, 267.
[10] N. M. Glykos, *J. Comput. Chem.* **2006**, *27*, 1765.
[11] B. R. Brooks, C. L. Brooks III, A. D. Mackerell, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, S. Boresch A. Caflisch, L. Caves, Q. Cui, A. R. Dinner, M. Feig, S. Fischer, J. Gao, M. Hodoscek, W. Im, K. Kuczera, T. Lazaridis, J. Ma, V. Ovchinnikov, E. Paci, R. W. Pastor, C. B. Post, J. Z. Pu, M. Schaefer, B. Tidor, R. M. Venable, H. L. Woodcock, X. Wu, W. Yang, D. M. York, M. Karplus, *J. Comput. Chem.* **2009**, *30*, 1545.
[12] J. C. Phillips, R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kalé, K. Schulten, *J. Comput. Chem.* **2005**, *26*, 1781.
[13] D. Frishman, P. Argos, *Proteins: Struct. Funct. Genet.* **1995**, *23*, 566.
[14] T. D. Schneider, R. M. Stephens, *Nucleic Acids Res.* **1990**, *18*, 6097.
[15] G. E. Crooks, G. Hon, J. M. Chandonia, S. E. Brenner, *Genome Res.* **2004**, *14*, 1188.
[16] S. S. Cho, Y. Levy, P. G. Wolynes, *Proc. Natl. Acad. Sci. USA.* **2006**, *103*, 586.
[17] J. Schlitter, *Chem. Phys. Lett.* **1993**, *215*, 6.
[18] I. Andricioaei, M. Karplus, *J. Chem. Phys.* **2001**, *115*, 6289.

Received: 25 April 2013
Accepted: 25 June 2013
Published online on 15 July 2013