Hindawi Publishing Corporation
ISRN Computational Biology
Volume 2013, Article ID 640125, 5 pages
http://dx.doi.org/10.1155/2013/640125

*Research Article*

# Saving Significant Amount of Time in MD Simulations by Using an Implicit Solvent Model and Elevated Temperatures

**Ifat Shub,**[1,2] **Ehud Schreiber,**[1] **and Yossef Kliger**[1]

[1] *Compugen Ltd., 69512 Tel Aviv, Israel*
[2] *The Mina and Everard Goodman Faculty of Life Sciences, Bar Ilan University, 52900 Ramat-Gan, Israel*

Correspondence should be addressed to Yossef Kliger; kliger@cgen.com

Received 31 January 2013; Accepted 25 February 2013

Academic Editors: F. Barbault and F. Fanelli

Copyright © 2013 Ifat Shub et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Molecular dynamic simulations are used for investigating various aspects of biological processes. Such simulations often require intensive computer power; therefore several solutions were developed to minimize the computer power needed, including the usage of elevated temperatures. Yet, such simulations are still not commonly used by the wide scientific community of chemists and biochemists. For about two years now, the molecular simulations suite GROMACS enables conducting simulations using implicit solvent models to further decrease runtimes. In order to quantify the saving in computer power, and to confirm the validity of the models, we followed the simple dissolution process of a single NaCl molecule. The results reveal approximately 350-fold decrease in real-world runtime when using an implicit solvent model and an elevated temperature, compared to using explicit water molecules and simulating at room temperature. In addition, in a wide range of temperatures, the dissolution times of NaCl are distributed, as expected, exponentially, both in explicit and in implicit solvent models, hence confirming the validity of the simulation approach. Hopefully, our findings will encourage many scientists to take advantage of the recent progress in the molecular dynamics field for various applications.

## 1. Introduction

Molecular dynamics (MD) is thecomputer simulation of the physical movements of molecules. The molecules' atoms are allowed to interact for a period of time, and the trajectories of the motion of the system are followed. In order for an MD simulation to be meaningful, the run has to be long enough for the system to visit all the energetically relevant conformations; that is, the system has to be ergodic in the timescale of the simulation [1]. An important benefit of the ergodicity of the system is that it allows MD to follow only one copy of the molecule instead of looking at a snapshot of many copies of the molecule. On the other hand, very long runtimes are required in order to achieve ergodicity, since the time step of the simulation has to be short enough to allow following the fastest motion of the system, namely, bond stretching and bond bending. These short time steps impose working in femtosecond resolution. However, "interesting" biochemical processes typically take at least microseconds and sometimes even milliseconds. Therefore, MD simulations of biochemical processes require intensive calculations and thus depend on expensive and hard-to-achieve hardware. To resolve this, various approaches were developed ranging from special-purpose computers [2], through algorithms developed to enhance MD simulation like those that enable the usage of implicit water [3] and those that allow "alchemical" transformation [4], to Monte Carlo simulations [5].

Since the launch of GROMACS [6] version 4.5 in 2010, it is possible to perform MD simulations using implicit solvent modeling. This, naturally, allows decrease in runtimes. To calculate the efficiency of the implicit modeling in reducing simulation runtimes, we modeled the dissolution of NaCl. The results reveal a significant decrease in real-world runtime when using an implicit solvent model, rather than using an explicit solvent one, that is, explicitly incorporating water molecules. Furthermore, simulating at elevated temperatures allowed further decrease in runtime.

The appropriateness of implicit water, high temperature simulations can be questioned, as it certainly involves extra approximations and assumptions compared to explicit water, room temperature simulations. To address these concerns, we focus on the dissolution times of the NaCl system. We perform multiple simulations at several temperatures and verify that the expedited simulation scheme still obeys the following two theoretical expectations (to be elaborated later).

(1) The average dissolution time has the expected dependence upon the temperature.

(2) The dissolution times have the expected (exponential) distribution, parameterized only upon the average.

## 2. Methods

Molecular Dynamics simulations were performed using the GROMACS 4.5.3 software suite and the OPLS-AA force field [7]. For the explicit water runs, a dodecahedron box was used. Generalized-Born (GB) potentials were used for the implicit solvent model. Exact commands used in the simulations are described in the Supplementary Material available online at http://dx.doi.org/10.1155/2013/640125.

When performing MD simulations with explicit water molecules, one of the parameters that should be taken into account is the system box size. Wrong choice of the box size may affect the simulations output (Figure 1). For the dodecahedron box, the dimensions are set to the diameter of the system (largest distance between atoms) plus twice the specified distance. When the specified distance was 1 nm, we observed that after their separation the $Na^+$ and $Cl^-$ ions reapproached each other many times during the simulation. This indicates that the box (with periodic boundary conditions) is too small. When using a specified distance of 3 nm or 6 nm, the two ions continued to be separated for hundreds of picoseconds (Figure 1). We decided to use a specified distance of 3 nm, as this seemed to be large enough for our system, yet with about 8.5 times shorter runtime compared to the 6 nm box.

Interactions between $Na^+$ and $Cl^-$ ions were determined with the *g_mindist* command, using the option of calculating the number of contacts between two defined groups.

The NaCl "salt crystal" (actually pair of ions in our simulations) is electrically bound inside the water medium. Thermal excitations may, however, separate the ions far enough so that they may diffuse from one another and the crystal may dissolve. Such dissolution, at least for low temperatures, is a rare event, and the ions fluctuate many times in the bound state before they succeed in overcoming the potential barrier and dissolving. This has the following implications.

(1) The ions fluctuate randomly in an approximate harmonic oscillator potential near their minimum energy displacement, hitting the potential barrier with a frequency approximately constant with temperature. The probability of the thermal excitation to pass above that constant barrier is, however, given by a Boltzmann factor which is exponential in nature. The exponent is negative and is proportional to the barrier height and inversely proportional to the temperature. Therefore, the logarithm of the average dissolution time is expected to depend in a linear fashion on the inverse temperature.

(2) The dissolution process is supposed to be memory-free; that is, the bound state is oblivious of the elapsed time. In particular, the rate of the process, given that it has not yet occurred, is constant. In such a case, just like with the case of nuclear decay, the distribution of the dissolution time is exponential. Performing many simulations (at a given temperature), we therefore may employ the Lilliefors statistical test of the null hypothesis that the sample comes from an exponential distribution against the alternative that it does not come from this distribution. It is a 2-sided goodness-of-fit test and is suitable when a fully specified null distribution is unknown, as its scale parameter (the average time) must be estimated from the samples. As a check, we have also employed the corresponding Lilliefors tests assuming the normal and extreme value distributions (where both scale and location parameters must be estimated). The statistical analysis was performed using the Matlab command *lillietest*.

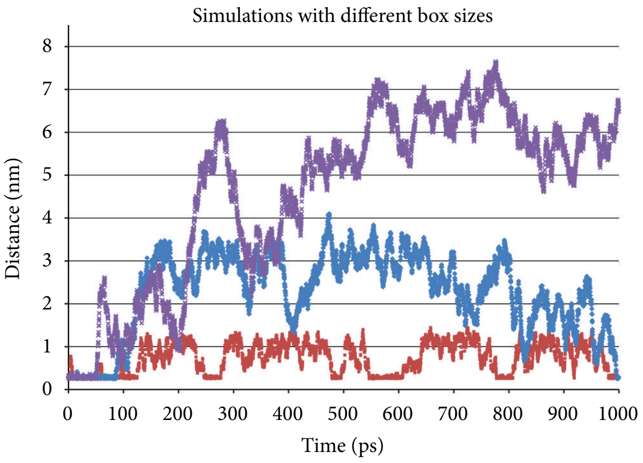


Figure 1: Simulations with different box sizes. Three typical simulations with different box sizes (dimensions are set to the diameter of the system (largest distance between atoms) plus twice the specified distance of 1 nm, 3 nm, or 6 nm) for one NaCl molecule with explicit water in 450°K. The distance (nm) between the $Na^+$ and the $Cl^-$ ions is measured over time (ps). In the simulation of the 1 nm box (red) the distance returned to the baseline value (~0.26 nm) very quickly after the ions were separated. This artificial behavior does not happen with the 6 nm box (purple); however the simulation is much slower. This artifact occurs when using the 3 nm box (blue), but only after about 1000 ps, a time period long enough for our purposes.

Table 1: Lilliefors test for the different test sets. The table includes the results of the 4 sets of runs; for each set the median of the 50 runs is calculated. In addition the $P$ values of the Lilliefors statistical test are presented for each of the 3 distributions: normal, extreme value, and exponential. The null hypothesis is that the sample comes from the given distribution. We see that the normal and extreme value distributions can be rejected (with a 5% $P$ value), while the exponential distribution, which is the expected one, cannot.

| Water modeling approach | Explicit | Explicit | Implicit | Implicit |
|---|---|---|---|---|
| Temperature (°K) | 298 | 450 | 298 | 450 |
| Median simulation dissolution time (ps) | 86.4 | 15.6 | >1000 | 234.853 |
| Median real-world dissolution simulation time | 57.2 min | 13.4 min | >0.73 min | 0.16 min |
| $P$ value to reject H0: exponential | **0.2808** | **0.1902** | | **0.0906** |
| $P$ value to reject H0: normal | 0.001 | 0.001 | | 0.0125 |
| $P$ value to reject H0: extreme value | 0.001 | 0.001 | | 0.001 |

Table 2: NaCl dissolution simulations using implicit solvent model at different temperatures. For each temperature, 50 repetitions were performed. $P$ values of the Lilliefors statistical test are presented for each of the 3 distributions: normal, extreme value, and exponential. The null hypothesis is that the sample comes from the given distribution. An ideal range of temperatures was found (in bold), in which the dissolution times of the 50 repetitions are distributed exponentially. At extremely elevated temperatures the null hypothesis, according to which the distribution is exponential, was rejected (in italic).

| | Dissolution time | | | | $P$ value | |
|---|---|---|---|---|---|---|
| Temperature (°K) | 1/Temp | Average (ps) | Log Avg | Exponential | Normal | Extreme value |
| 450 | 0.002222 | 269.344546 | 2.430308 | 0.0906 | 0.0125 | 0.001 |
| 500 | 0.002 | 135.70892 | 2.132608 | 0.1037 | 0.0022 | 0.001 |
| 550 | 0.001818 | 102.92944 | 2.01254 | **0.5** | **0.0028** | **0.001** |
| 600 | 0.001667 | 62.97156 | 1.799144 | **0.5** | **0.001** | **0.001** |
| 650 | 0.001538 | 36.57108 | 1.563138 | **0.5** | **0.001** | **0.001** |
| 700 | 0.001429 | 23.38584 | 1.368953 | **0.5** | **0.001** | **0.001** |
| 750 | 0.001333 | 13.9286 | 1.143907 | **0.5** | **0.001** | **0.001** |
| 800 | 0.00125 | 13.80488 | 1.140033 | **0.5** | **0.001** | **0.001** |
| 850 | 0.001176 | 10.72328 | 1.030328 | **0.3643** | **0.001** | **0.001** |
| 900 | 0.001111 | 7.34044 | 0.865722 | **0.5** | **0.001** | **0.001** |
| 1050 | 0.000952 | 4.7986 | 0.681115 | **0.5** | **0.0135** | **0.001** |
| 1200 | 0.000833 | 2.20284 | 0.342983 | 0.0884 | 0.0129 | 0.001 |
| 2000 | 0.0005 | 0.77276 | −0.11196 | 0.0803 | 0.001 | 0.001 |
| 3000 | 0.000333 | 0.46648 | −0.33117 | 0.0815 | 0.001 | 0.001 |
| 4000 | 0.00025 | 0.3098 | −0.50892 | *0.0038* | 0.001 | 0.001 |
| 6000 | 0.000167 | 0.14592 | −0.83589 | *0.001* | 0.001 | 0.001 |
| 10000 | 0.0001 | 0.084 | −1.07572 | *0.001* | 0.001 | 0.001 |

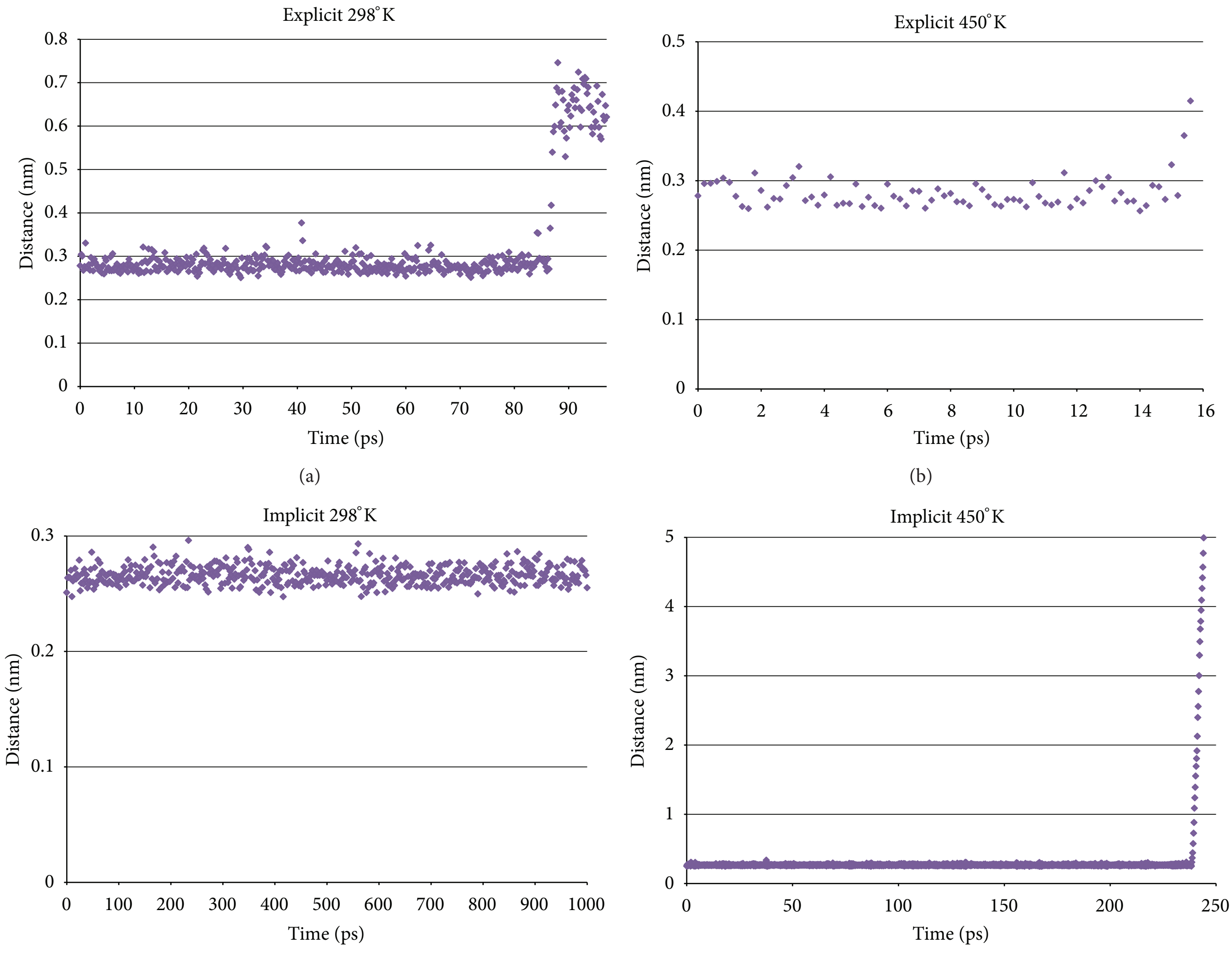
## 3. Results and Discussion

*3.1. Explicit Solvent versus Implicit Solvent Simulations in Two Different Temperatures.* We used a very simple process—the dissolution of one "molecule" of NaCl. We started with a conservative approach in which explicit water and "physiological temperature" were used (Figure 2(a)). We repeated the simulation 50 times, using different random seeds, and confirmed that, as expected, the NaCl dissolution times are distributed exponentially (Table 1). This supports the validity of the simulation.

Next, to accelerate the process we used implicit water instead of explicit water. Further acceleration of run times was achieved by using elevated temperatures. The results reveal that the two changes did not harm the validity of the simulations (Table 1 and Figure 2).

Even though we used "nonrealistic" temperatures, we were able to show that qualitatively it is possible to model a chemical process of NaCl dissolution. This is consistent with a previous study of MD simulations with explicit solvent and the OPLS-AA force field, showing that the actual temperature values are not accurately reproduced, due to the imperfection of the force fields, but that the qualitative differences between temperatures are reproduced correctly [8].

*3.2. Finding the Range of Temperatures for Running the Simulation.* In the simple case of NaCl, we could actually wait for dissolution, and one clear cutoff, that is, the distance



(a)

(b)

(c)

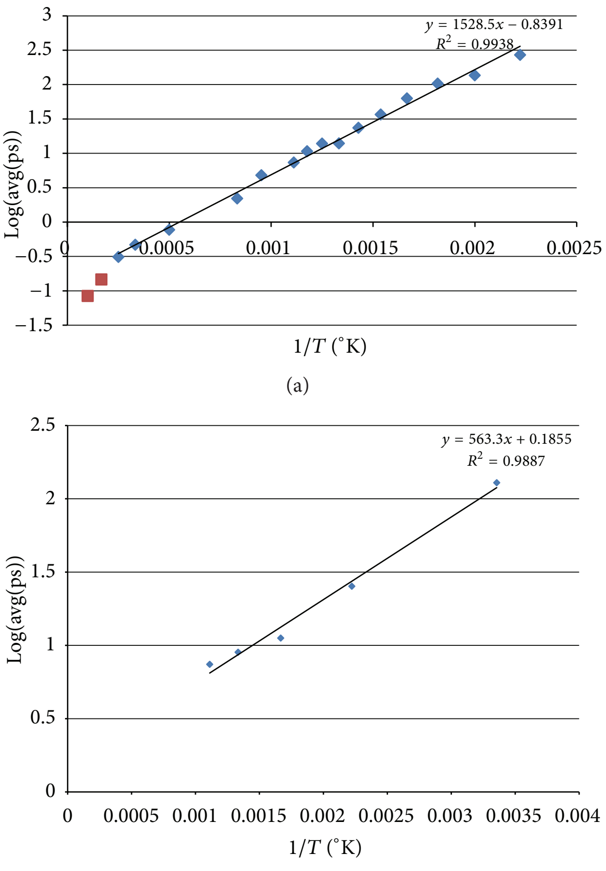
(d)

Figure 2: Typical simulations of NaCl. Typical runs (we chose runs in which the dissolution time between the Na and the Cl is close to the median of the distribution of the dissolution times) are shown. (a and b) show that elevated temperatures shorten the dissolution time in explicit solvent simulations (at 298°K the NaCl dissolves in about 85 ps, while at 450°K it does so after only about 16 ps). (c and d) show the same in implicit solvent simulations (at 298°K the NaCl does not dissolve in 1000 ps, whereas at 450°K it dissolves in about 240 ps). Although the dissolution time using implicit water is longer than in the explicit water case, the real-world runtime is much shorter (see Table 1). The real-world runtime of the 450°K simulation using implicit water is 84 times shorter than 450°K simulation using explicit water.

between $Na^+$ and $Cl^-$, is enough to determine the dissolution time. Again, at each temperature, we ran 50 independent MD simulations using explicit or implicit solvent. We used the Lilliefors test to determine each temperature's data set distribution (Table 2). As expected, a linear correlation between log (average dissolution time) and 1/temperature was found both when using the explicit and implicit models (Figure 3).

## 4. Conclusions

We were able to show that it is possible to shorten run times of MD simulations with implicit solvent and elevated temperatures. This requires much less intensive computational power and enables performing significantly longer simulations in shorter runtimes. We have shown that the dissolution times of a model simple complex (bound pair of NaCl ions) can be reliably extracted in such expedited molecular simulations, as they display both the expected average dependence on temperatures and the expected distribution. Note that the off-rate of a binary complexation reaction is given just as the inverse of the average dissociation time, so we have shown that in principle it is amenable to determination by simulation. More complex cases, for example, protein-protein interactions, require much longer real-world simulation runtime, which will be achievable by using implicit solvent models and elevated temperatures. These findings will hopefully inspire modeling of more complex systems.



(a)

(b)

Figure 3: NaCl runs in different temperatures using implicit/explicit solvent. (a) shows data of simulations with an implicit solvent model, while (b) corresponds to an explicit solvent model; that is, it involves explicit water molecules. A linear correlation between log (average dissolution time (ps)) and $1/T$ (1/°K) is manifested. The two points marked in red are outside the ideal temperature range (the sample is not distributed exponentially (see Table 2)).

## Conflict of Interests

The three authors hereby declare that they do not have any conflict of interests relevant to this paper. Specifically, GROMACS is a free molecular simulation engine that they used in this study, but they have no conflict of interests regarding this software.

## References

[1] A. Barducci, M. Bonomi, and M. Parrinello, "Metadynamics. Wiley interdisciplinary reviews," *Computational Molecular Science*, vol. 1, no. 5, pp. 826–843, 2011.

[2] D. E. Shaw, P. Maragakis, K. Lindorff-Larsen et al., "Atomic-level characterization of the structural dynamics of proteins," *Science*, vol. 330, no. 6002, pp. 341–346, 2010.

[3] D. Qiu, P. S. Shenkin, F. P. Hollinger, and W. C. Still, "The GB/SA continuum model for solvation: a fast analytical method for the calculation of approximate Born radii," *Journal of Physical Chemistry A*, vol. 101, no. 16, pp. 3005–3014, 1997.

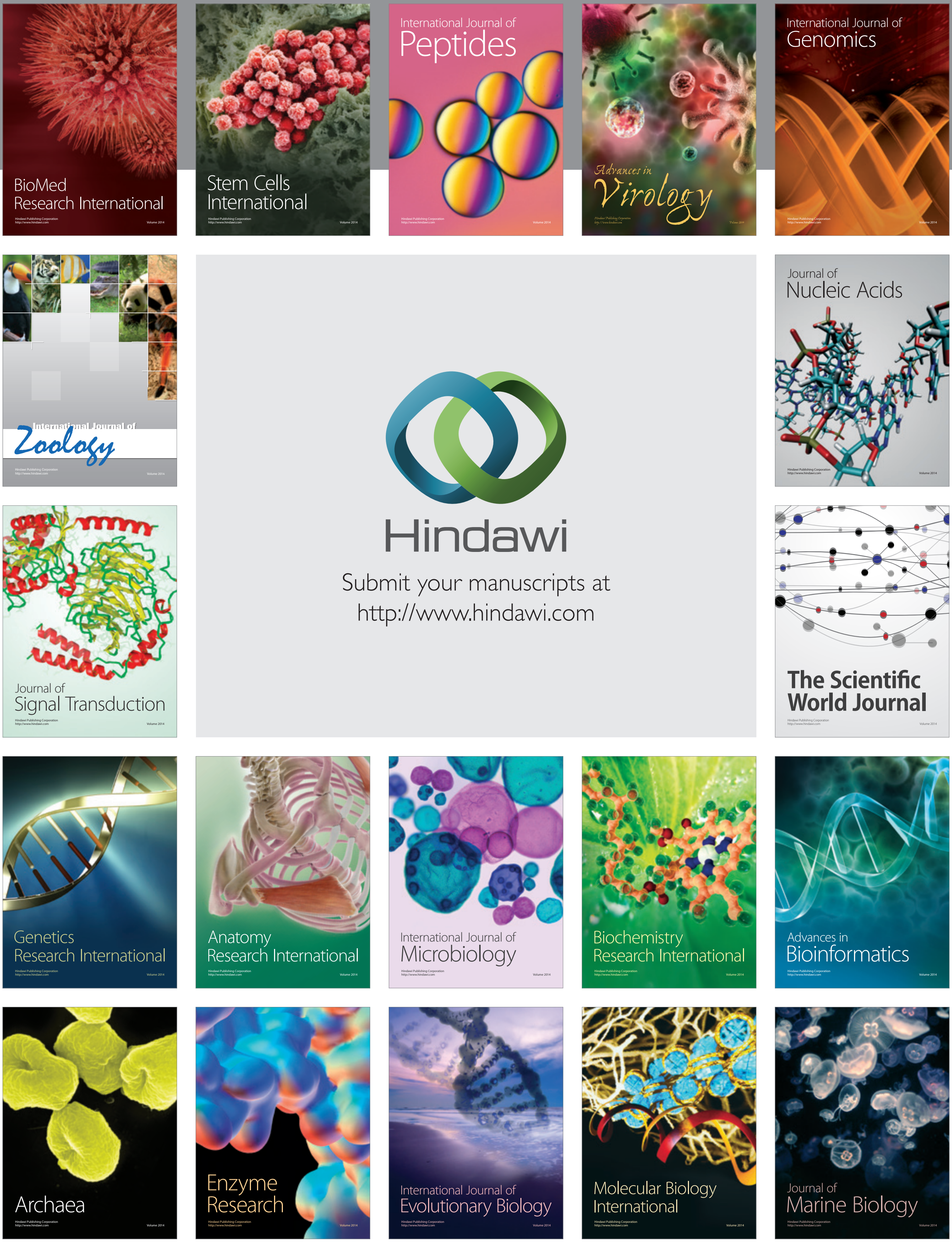
[4] M. R. Shirts, D. L. Mobley, and J. D. Chodera, "Alchemical free energy calculations: ready for prime time?" *Annual Reports in Computational Chemistry*, vol. 3, pp. 41–59, 2007.

[5] B. L. de Groot, R. M. Scheek, A. Amadei, G. Vriend, and H. J. C. Berendsen, "Prediction of protein conformational freedom from distance constraints," *Proteins*, vol. 29, no. 2, pp. 240–251, 1997.

[6] D. Van Der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen, "GROMACS: fast, flexible, and free," *Journal of Computational Chemistry*, vol. 26, no. 16, pp. 1701–1718, 2005.

[7] G. A. Kaminski, R. A. Friesner, J. Tirado-Rives, and W. L. Jorgensen, "Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides," *Journal of Physical Chemistry B*, vol. 105, no. 28, pp. 6474–6487, 2001.

[8] R. Zhou, "Trp-cage: folding free energy landscape in explicit water," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 23, pp. 13280–13285, 2003.


BioMed Research International
Stem Cells International
International Journal of Peptides
Advances in Virology
International Journal of Genomics
International Journal of Zoology
Journal of Nucleic Acids
Journal of Signal Transduction
The Scientific World Journal
Genetics Research International
Anatomy Research International
International Journal of Microbiology
Biochemistry Research International
Advances in Bioinformatics
Archaea
Enzyme Research
International Journal of Evolutionary Biology
Molecular Biology International
Journal of Marine Biology

Hindawi

Submit your manuscripts at
http://www.hindawi.com