Journal of Computer-Aided Molecular Design, 1 (1987) 171–176
ESCOM
J-CAMD 008

Thermodynamic cycle integration by computer simulation as a tool for obtaining free energy differences in molecular chemistry

W.F. van Gunsteren and H.J.C. Berendsen

Laboratory of Physical Chemistry, University of Groningen, Nijenborgh 16, 9747 AG Groningen, The Netherlands

Received 23 March 1987
Accepted 23 March 1987

Key words: Computer simulation; Free energy; Molecular dynamics; Thermodynamic integration; Statistical mechanics

SUMMARY

A new and promising development in the field of computer simulation of molecular systems is the so-called thermodynamic cycle integration technique, which combines well-known results from statistical thermodynamics with powerful computer simulation methods. The basic formulas, the development and the applications in the areas of drug design, protein engineering and conformational analysis of this elegant technique are discussed.

INTRODUCTION AND BACKGROUND

Molecular dynamics (MD) computer simulations have added considerably to our understanding, at the atomic level, of the properties of molecular systems, such as liquids or solutions of (bio)molecules during the past 15 years. A rather static picture of molecular conformation has been gradually transformed into a more dynamic one, where the molecular properties are dynamic averages over an ensemble of molecular conformations. The development of computer simulation techniques has been made possible by the continuous and rapid development of computer hardware. Every six to seven years the ratio of performance to price has increased by a factor of ten and due to the emerging parallel computing techniques the end of this increase is not yet in sight. At present small proteins in aqueous solution, involving many thousands of atoms, can be simulated over periods of about 10–100 ps. We refer to Refs. 1-4 for reviews on the subject.

FREE ENERGY VERSUS ENERGY

From a molecular dynamics trajectory, the statistical equilibrium averages can be obtained for any desired property of the molecular system for which a value can be computed at each point of the trajectory. Examples of such properties are the potential or kinetic energy of relevant parts of the system, structural properties and fluctuations, electric fields, diffusion constants, etc. A number of thermodynamic properties can be derived from such averages. However, two important thermodynamic quantities, the entropy and the (Gibbs) free energy can generally not be derived from a statistical average, They are global properties that depend on the extent of phase (or configuration) space accessible to the molecular system. Therefore, computation of the absolute free energy of a molecular system is virtually impossible. Yet, the most important chemical quantities like binding constants of donor-acceptor complexes or molecular solubilities are directly related to the free energy. However, over the past few years, several statistical mechanical procedures have evolved for evaluating *relative* free energy differences. They are rather demanding as far as computer time is concerned, but will open up a wide area of the most interesting applications in chemistry, e.g. in drug design and protein engineering.

FREE ENERGY DIFFERENCES BY THERMODYNAMIC INTEGRATION

There exist several methods for calculating relative free energy differences [5–8], of which we will discuss the two most important ones, viz. thermodynamic perturbation and integration methods. They make use of the fact that free energy changes related to small perturbations of a molecular system can be determined during a simulation. The free energy difference between two states A and B of a system can be determined from a MD simulation in which the potential energy function V is slowly changed such that the system slowly changes from state A to state B. In principle, the free energy is determined as the work necessary to change the system from A to B over a reversible path.

The method works as follows. Firstly, the Hamiltonian $H(\mathbf{p},\mathbf{q})$ (or only the potential term $V(\mathbf{q})$) is made a function of a parameter λ, such that $H(\mathbf{p},\mathbf{q},\lambda_A)$ characterizes state A of the system and $H(\mathbf{p},\mathbf{q},\lambda_B)$ state B. Then the Gibbs free energy of the system is also a function of λ:

$$G(\lambda) = -kT \ln\Delta(\lambda) \tag{1}$$

where k denotes Boltzmann's constant, T is the temperature and the isobaric partition function Δ is given by

$$\Delta(\lambda) = (h^{3N}N!)^{-1} \iiint \exp[-(H(\mathbf{p},\mathbf{q},\lambda) + PV)/kT]\, dV\, d\mathbf{p}\, d\mathbf{q} \tag{2}$$

Here P is the pressure and V the volume of the system and $\mathbf{q}$ and $\mathbf{p}$ are the generalized coordinates and momenta of the N particles. The free energy difference ΔG_{BA} then reads

$$\Delta G_{BA} = G(\lambda_B) - G(\lambda_A) = -kT \ln\{\Delta(\lambda_B)/\Delta(\lambda_A)\} \tag{3}$$

which can be expressed as an ensemble average

$$\Delta G_{BA} = -kT \ln\left\{\frac{\iiint \exp[-(H(\mathbf{p},\mathbf{q},\lambda_B)-H(\mathbf{p},\mathbf{q},\lambda_A))/kT]\exp[-(H(\mathbf{p},\mathbf{q},\lambda_A)+PV)/kT]dV d\mathbf{p}d\mathbf{q}}{\iiint \exp[-(H(p,q,\lambda_A)+PV)/kT]dV\, d\mathbf{p}\, d\mathbf{q}}\right\}$$

$$= -kT \ln\{<\exp[-(H(\mathbf{p},\mathbf{q},\lambda_B) - H(\mathbf{p},\mathbf{q},\lambda_A))/kT]>_{\lambda_A}\} \tag{4}$$

where the brackets $<...>_\lambda$ denote an ensemble average over $\mathbf{p},\mathbf{q}$ and V generated at a specific value of λ. Formula 4 is called the *perturbation formula*, since it will only yield accurate results when state B is close to state A. If this difference is large, the change from A to B must be split up in a number of steps between intermediate states that are close enough to allow for the use of Eq. 4 and then ΔG_{BA} is just the sum of the ΔG for all intermediate steps.

The thermodynamic *integration formula* is obtained by straightforward differentiation of (1) with respect to λ:

$$\frac{dG(\lambda)}{d\lambda} = <\frac{\partial H(\mathbf{p},\mathbf{q},\lambda)}{\partial\lambda}>_\lambda \tag{5}$$

In that case

$$\Delta G_{BA} = \int_{\lambda_A}^{\lambda_B} <\frac{\partial H(\mathbf{p},\mathbf{q},\lambda)}{\partial\lambda}>_\lambda d\lambda \tag{6}$$

If λ is being changed very slowly during a MD simulation, the integration (6) can be carried out during the MD run. Then ΔG_{BA} can be directly obtained for rather different states A and B. The continuous change in λ should be so slow that the system remains essentially in equilibrium for each intermediate value of λ.

Various parameterizations of the Hamiltonian are possible. One may change a covalent bond length b as

$$b(\lambda) = (1-\lambda)b_A + \lambda b_B \tag{7}$$

and do similarly for bond angles and other terms of the potential function. The choice should be made such that the change of $G(\lambda)$ as a function of λ is as smooth as possible.

THERMODYNAMIC CYCLES

The basis on which the thermodynamic cycle approach rests is the fact that the (Gibbs) free energy G is a thermodynamic state function. This means that as long as a system is changed in a reversible way the change in free energy ΔG will be independent of the path. Therefore, along a closed path or cycle one has $\Delta G = 0$. This result implies that there are two possibilities of obtaining ΔG for a specific process; one may calculate it directly using the techniques discussed above along a path corresponding to the process, or one may design a cycle of which the specific process

is only a part and calculate the ΔG of the remaining part of the cycle. The power of this thermodynamic cycle technique lies in the fact that on the computer also non-chemical processes such as the conversion of one type of atom into another type may be performed.

In order to visualize the method we consider the relative binding of two inhibitors I_A and I_B to an enzyme E. The appropriate thermodynamic cycle for obtaining the relative binding constant is

$$\begin{array}{ccc} & E + I_A \xrightarrow{1(exp)} (E{:}I_A) & \\ 3\,(sim.) \downarrow & & \downarrow 4\,(sim.) \\ & E + I_B \xrightarrow[2(exp)]{} (E{:}I_B) & \end{array} \qquad (8)$$

where the symbol: denotes complex formation.
The relative binding constant equals

$$K_2/K_1 = \exp\left[-(\Delta G_2 - \Delta G_1)/RT\right] \qquad (9)$$

However, simulation of processes 1 and 2 is virtually impossible since it would involve the removal of many solvent molecules from the active site of the enzyme to be substituted by the inhibitor. However, since (8) is a cycle we have

$$\Delta G_2 - \Delta G_1 = \Delta G_4 - \Delta G_3 \qquad (10)$$

and, if the composition of inhibitor I_B is not too different from that of I_A, the desired result can be obtained by simulating the non-chemical processes 3 and 4.

DEVELOPMENT AND APPLICATION OF THE THERMODYNAMIC CYCLE INTEGRATION TECHNIQUE

Different groups have contributed to the development of the thermodynamic cycle integration techniques [9–21]. The original idea is already old [9,10]. The perturbation approach (4) came first into use [11–15], followed by the integration formula (6) using discrete integration steps and simulations [16–18]. The latest development is the continuous integration approach [6,19–21].

It will be clear that the applications of the thermodynamic cycle integration technique (8) are manifold in chemistry. It can be used to study the relative free energy of solvation of species I_A and I_B. In that case the symbol E on the left hand side of (8) denotes the solvent and it should be omitted from the right hand side. So, processes 1 and 2 mimick the desolvation of I_A and I_B. Examples of this type of solution study are provided in Refs. 12,13,15,16,19–23. Relative binding constants for complex formation have also been calculated [20,24–26]. Other studies concern the effect of amino acid substitution upon inhibitor-enzyme binding [27,28]. Other applications that may be expected in the near future are the study of the stability of protein mutants, of DNA-repressor complexes, of different molecular conformations, etc.

APPROXIMATIONS, LIMITATIONS AND PERSPECTIVES

Up till now most studies concern changes in van der Waals parameters and atomic (dipolar) changes for relatively rigid molecules. In these cases the experimental free energies are generally reproduced within kT [14,19–28]. Difficulties arise when full atomic charges are created or annihilated. In that case, reaction field effects will contribute significantly to the free energy [20]. When the free energy change also depends on internal degrees of freedom, like torsion angles, the results may be less accurate [20].

The choice of the parameterization by λ of the Hamiltonian (or potential energy function) will influence the accuracy of the results [23,29]. Linear parameterization of the non-bonded terms is not always efficient [29]. Also for internal degrees of freedom use of a non-linear dependence of the torsional interaction on λ may be advisable [20].

When approximations are made in the molecular model, in the interaction function or by applying specific boundary conditions, it should be kept in mind that the same approximations must be made in both processes 3 and 4 of the thermodynamic cycle (8), in order to preserve a consistent cycle.

Accurate results are also dependent on correct sampling, that is, when the system is changed from state A to state B as a function of λ, the statistical sampling must be performed at the correct value of λ. In the continuous integration approach this means that the length of the MD simulation must be much longer than the relaxation time of the surroundings. When studying hydration this condition can be met, but conformational relaxation in proteins may require much longer than 10–100 ps simulations.

Finally, we may conclude that the method of thermodynamic cycle integration by computer simulation is a very promising and widely applicable tool in the study of molecular processes at the atomic level. It still requires a considerable amount of computing effort and its range and limitations are not yet fully explored. However, from available experience, it can be concluded that with the present state of the art instructive and rather accurate results can be obtained when the atomic interaction function is a reliable one, the change from state A to state B is not too drastic, and the parameterization ot the Hamiltonian is rightly chosen in order to allow for an accurate sampling of the free energy path.

REFERENCES

1 Van Gunsteren, W.F. and Berendsen, H.J.C., Biochem. Soc. Trans., 10 (1982) 301–305.

2 Hermans, J. (Ed) Molecular Dynamics and Protein Structure, Polycrystal Book Service, P.O. Box 27, Western Springs, IL 60558, 1985.

3 Ciccotti, G. and Hoover, W.G. (Eds.) Proceedings of the International School of Physics 'Enrico Fermi', course 97, on Molecular Dynamics Simulation of Statistical-Mechanical Systems, North-Holland, Amsterdam, 1986.

4 McCammon, J.A. and Harvey, S.C., Dynamics of Proteins and Nucleic Acids, Cambridge University Press, Cambridge, 1987.

5 Quirke, N. and Jacucci, G., Molec. Phys., 45 (1982) 823–838.

6 Berendsen, H.J.C., Postma, J.P.M. and van Gunsteren, W.F., In Hermans, J. (Ed) Molecular Dynamics and Protein Structure, Polycrystal Book Service, Western Springs, 1985, p. 43–46.

7 Frenkel, D., In Ciccotti, G. and Hoover, W.G. (Eds) Proceedings of the International School of Physics 'Enrico Fermi', course 97, North-Holland, Amsterdam, 1986, p. 151–188.

8 Mezei, M. and Beveridge, D.L., Ann. N.Y. Acad. Sci., 482 (1986) 1–23.

9 Kirkwood, J.G., J. Chem. Phys., 3 (1935) 300–313.
10 Zwanzig, R.W., J. Chem. Phys., 22 (1954) 1420–1426.
11 Torrie, G.M. and Valleau, J.P., Chem. Phys. Letters, 28 (1974) 578–581.
12 Okazaki, S., Nakanishi, K., Touhara, H. and Adachi, Y., J. Chem. Phys., 71 (1979) 2421–2429.
13 Postma, J.P.M., Berendsen, H.J.C. and Haak, J.R., Faraday Symp. Chem. Soc., 17 (1982) 55–67.
14 Tembe, B.L. and McCammon, J.A., Comput. Chem., 8 (1984) 281–283.
15 Jorgensen, W.L. and Ravimohan, C., J. Chem. Phys., 83 (1985) 3050–3054.
16 Mruzik, M.R., Abraham, F.F. and Pound, G.M., J. Chem. Phys., 64 (1976) 481–491.
17 Berens, P.H., Mackay, D.H.J., White, G.M. and Wilson, K.R., J. Chem. Phys., 79 (1983) 2375–2389.
18 Mezei, M., Swaminathan, S. and Beveridge, D.L., J. Am. Chem. Soc., 100 (1978) 3255–3256.
19 Postma, J.P.M., Molecular Dynamics of H_2O, thesis, University of Groningen, 1985.
20 Van Gunsteren, W.F. and Berendsen, H.J.C., In Stozowski, J. (Ed) Proceedings of the Symposium on Computational Methods in Chemical Design: Molecular Modelling and Graphics, Elmau, 1986, Oxford University Press, Oxford, 1987.
21 Straatsma, T.P., Berendsen, H.J.C. and Postma, J.P.M., J. Chem. Phys., 85 (1986) 6720–6727.
22 Lybrand, T.P., Gosh, I. and McCammon, J.A., J. Amer. Chem. Soc., 107 (1985) 7793–7794.
23 Bash, P.A., Singh, U.C., Langridge, R. and Kollman, P.A., Science (in press).
24 Lybrand, T.P., McCammon, J.A. and Wipff, G., Proc. Natl. Acad. Sci. U.S.A., 83 (1986) 833–835.
25 Wong, C.F. and McCammon, J.A., J. Amer. Chem. Soc. (in press).
26 Bash, P.A., Singh, U.C., Brown, F.K., Langridge, R. and Kollman, P.A., Science, 235 (1987) 574–576.
27 Wong, C.F. and McCammon, J.A., Isr. J. Chem. (in press).
28 Bash, P.A., Singh, U.C., Langridge, R. and Kollman, P.A., Science (in press).
29 Cross, A.J., Chem. Phys. Letters, 128 (1986) 198–202.