View metadata, citation and similar papers at core.ac.uk

brought to you by CORE

provided by NASA Technical Reports Server

COMPUTER SIMULATION OF PROTEIN SYSTEMS

D. J. OSGUTHORPE,
P. DAUBER-OSGUTHORPE,
J. WOLFF,
D. H. KITSON
AND
A. T. HAGLER

THE AGOURON INSTITUTE

LA JOLLA, CALIFORNIA

Computer Simulation of Protein Systems

D. J. Osguthorpe, P. Dauber-Osguthorpe, J. Wolff, D. H. Kitson and A. T. Hagler
The Agouron Institute, 505 Coast Blvd. South, La Jolla, California 92037.

Introduction. Significant advances are being made in the theoretical treatment of the conformation and dynamics of biological molecules. Several recent convergent developments are responsible for opening up new fields of investigation. They include:

1. The development and application of powerful theoretical techniques taken from statistical physics such as Monte Carlo and molecular dynamics simulations to biological systems.
2. The development of powerful computational hardware such as the Cyber 205.
3. The development of interactive graphics systems.
4. The increasing availability of experimental structural and dynamic data such as the ever-growing data base of protein crystal structures, small peptide crystal structures and the structural and dynamic properties of these same molecules in solution.

These developments enabled us to undertake the project of studying ligand binding to dihydrofolate reductase (DHFR). This is an extremely important enzyme, as it is the target of several drugs (inhibitors) which are used clinically as antibacterials, antiprotozoals and in cancer chemotherapy.[1,2] DHFR catalyzes the NADPH (reduced nicotinamide adenine dinucleotide phosphate) dependent reduction of dihydrofolate to tetrahydrofolate, which is used in several pathways of purine and pyrimidine biosynthesis, including that of thymidylate.[3] Since DNA synthesis is dependent on a continuing supply of thymidylate, a blockade of DHFR resulting in a depletion of thymidylate can lead to the cessation of growth of a rapidly proliferating cell line.

DHFR exhibits a significant species to species variability in its sensitivity to various inhibitors. For example, trimethoprim, an inhibitor of DHFR, binds to bacterial DHFR's 5 orders of magnitude greater than to vertebrate DHFR's.[4,5] We were interested in studying the structural mechanics, dynamics and energetics of a family of dihydrofolate reductases to rationalise the basis for the inhibition of these enzymes and to understand the molecular basis of the difference in the binding constants between the species. This involves investigating the conformational changes induced in the protein on binding the ligand, the internal strain imposed by the enzyme on the ligand, the restriction of fluctuations in atom positions due to binding and the consequent change in entropy. X-ray crystallographic structures of DHFR from a few species, in complex with various ligands, are known,[6-8] as well as partial data about the structures in solution.[9-11] The availability of the structure, in the form of atomic coordinates for the enzyme system, is a prerequisite for performing any kind of energy calculations. In addition, due to the size of these systems as discussed below, only the availability of supercomputers such as the Cyber 205 make this project feasible.

Computational Techniques. The techniques we use to investigate the DHFR system all require the calculation of the potential energy of the molecular system. This potential energy is expressed in terms of an analytical representation of all internal degrees of freedom and interatomic distances, as in eqn. (1).

$$\begin{aligned} V = {} & \sum\{D_b[1 - e^{-\alpha(b-b_0)}]^2 - D_b\} + 1/2 \sum H_\theta(\theta - \theta_0)^2 \\ & + 1/2 \sum H_\phi(1 + s\cos n\phi) + 1/2 \sum H_\chi \chi^2 \\ & + \sum\sum F_{bb'}(b - b_0)(b' - b_0') \\ & + \sum\sum F_{\theta\theta'}(\theta - \theta_0)(\theta' - \theta_0') + \sum\sum F_{b\theta}(b - b_0)(\theta - \theta_0) \\ & + \sum F_{\phi\theta\theta'}\cos\phi(\theta - \theta_0)(\theta' - \theta_0') + \sum\sum F_{\chi\chi'}\chi\chi' \\ & + \sum\epsilon[2(r^*/r)^9 - 3(r^*/r)^6] + \sum q_i q_j / r \end{aligned} \tag{1}$$

This type of representation of the potential energy in terms of the internal (valence) degrees of freedom is called a Valence Force Field. Such valence force fields have long been used in vibrational spectroscopy in order to carry out normal mode analysis.[12] Basically the terms in equation (1) express

the energies required to deform each internal coordinate from some unperturbed "standard" value denoted by the subscript "0". The first term is a Morse potential which describes the energy required to stretch each bond from its relaxed value, b_0. The second term represents the energy stored in each valence angle when it is bent from its "standard" value, θ_0. The third term represents the intrinsic energy required to twist the molecule about a bond by a torsion angle, ϕ. The fourth term represents the energy required to distort intrinsically planar systems by χ from their planar conformation, i.e. the out of plane term. The next terms represent various couplings between internal coordinates, which are known to be necessary from studies of vibrational spectra.[13] They are the bond-bond, angle-angle, bond-angle, angle-angle-torsion and out of plane cross-term respectively. The last 3 terms describe the exchange repulsion, dispersion and coulombic interactions that occur between non-bonded atoms.

The parameters D_b , H_θ , H_ϕ , H_χ , and F_{ij} are the force constants for the corresponding intramolecular deformation, r^* and ϵ characterize the size of the atoms and the strength of the van der Waals interaction between them, while the q_i are the partial charges carried by each atom. The parameters for the functions were derived from fitting a wide range of experimental data including crystal structure, unit cell vectors and the orientation of the asymmetric unit, sublimation energies, molecular dipole moments, molecular structure, vibrational spectra and strain energies of small organic compounds.[14-19] Ab-initio molecular orbital calculations have also been used in conjunction with the experimental data to give information on charge distributions, energy barriers and coupling terms, both to supplement and confirm the results obtained from the experimental data.[20, 21]

Minimisation. Given the analytical representation of the potential energy in eqn. (1), we can minimize this energy with respect to all internal degrees of freedom, i.e. solve the equation

$$\partial E/\partial x_i = 0 \qquad i = 1, 3n \tag{2}$$

where the x_i are the cartesian coordinates of the molecule.

The minimisation results in the "minimum energy structure" of the system. Analysis of the minimum energy structure reveals the basic structural features of the system along with the interatomic forces underlying this minimum energy conformation. At the minimum, we can take second derivatives of the energy and construct the mass weighted second derivative matrix. From the eigenvalues of this matrix the vibrational frequencies may be obtained and the normal modes from the eigenvectors.[22] The conformational entropy of the system can now be calculated from the vibrational frequencies using the Einstein relations.[23] The conformational entropy of a system plays an important role in both conformational equilibria and binding.[24]

Molecular dynamics. Molecular dynamics is the numerical integration of Newtons classical equations of motion. Having specified the potential, we define the initial conditions of the system, the coordinates of the protein, inhibitor, solvent and a set of initial velocities. Once the initial conditions are given, Newtons equations of motion

$$-\delta V(\bar{r}_i \cdots \bar{r}_n)/\delta \bar{r}_i = \bar{F}(\bar{r}_i \cdots \bar{r}_n) = m_i d^2\bar{r}_i/dt^2 \tag{3}$$

are integrated forward in time, in order to compute the atomic trajectories $\bar{r}_i$ (t)...$\bar{r}_n$ (t) as functions of time. The forces are calculated from the energy expression in eqn. (1) by taking analytical derivatives. We then take a small time step, Δt, of $\approx 10^{-15}$ sec. and applying the acceleration as calculated from Newtons law (eqn. 3), we update the velocity and position of each atom, to a new velocity and position using a Gear[25] predictor-corrector algorithm or a Verlet algorithm.[26] The forces and acceleration at the new positions are then calculated and we repeat the procedure, thus tracing the trajectories of the atoms.

Calculations on the Cyber. One of the systems we are studying, the E. coli DHFR-Trimethoprim complex, is the system we have been using to develop the programs on the Cyber 205. Table I lists the no. of atoms, internal coordinates and non-bond interactions for this system, to demonstrate the

magnitude of the calculation involved.

Table I

E. coli Dihydrofolate Reductase System

	atoms
E. coli Dihydrofolate Reductase	2490
Trimethoprim	40
155 Waters	465
	2995

Internal Coordinates	
Bonds	2875
Valence Angles	4785
Torsion Angles	6784
Bond-Bond cross-terms	4785
Bond-Angle cross-terms	9570
Angle-Angle cross-terms	7584
Angle-Angle-Torsion cross-terms	6784
Non-bond pairs	$\approx$1,600,000

Minimisation and molecular dynamics both require computing the energy using eqn. (1), changing the coordinates and repeating this process many times. Note that each energy calculation involves evaluating the appropriate terms in eqn. (1) for each of the internals listed in table I. Thus the last three terms in eqn. (1) need to be evaluated for each of the 1,600,000 non-bonded pairs. As the time required to compute the change in the coordinates once the energy has been calculated is small, the time required to calculate the energy determines the time to perform the minimisation, or how many steps of dynamics can be done. For a minimisation the number of iterations depends on how close to zero we require the derivatives, for a conjugate gradient minimiser previous experience indicates that about 3 times the number of atoms iterations are required to get derivatives to less than 0.05 kcal/molÅ, which is about 10,000 iterations for the protein. In molecular dynamics we would like to simulate at least 100 picoseconds, preferably a nanosecond, as this is still a very short time compared to molecular events such as binding. This requires 100,000 iterations at a 1 femtosecond timestep. Thus the speed with which the energy calculation is carried out is crucial.

Non-bond interaction calculation. Table II shows the timings of the energy routines used to compute eqn. (1) on the VAX 11/780 and the Cyber 205 for the Dihydrofolate Reductase system. The non-bond part of the calculation takes by far the major portion of the CPU time, 78% of the iteration time on the VAX, so this was vectorised first. The routine computes the non-bond energy, see eqn. (1), by calculating the interaction between all pairs of atoms, except for bonded atoms and 1-3 interactions. For a 10Å cutoff this is $\approx 1.6\times 10^6$ pairs, which is the reason this is the major time consuming portion of the energy calculation. This was implemented on the VAX by a residue neighbour list in

Table II

Comparison of the Timing of Energy Calculation routines for 1 Iteration

Routine	VAX 11/780	CYBER Vectorised Large Pages
Bonds	2.42	0.055
Valence Angles	9.06	0.13
Torsion Angles[2]	30.69	0.55
Bond-bond	5.25	0.14
Bond-Angle	11.9	0.25
Angle-Angle	16.55	0.17
Out of Plane	2.35	0.10[1]
Non-Bond	448.98	1.23
Iteration Timing[2]	573.58	2.7

1. The out of plane routine is not vectorised.
2. The iteration timing is slightly larger than the sum of all the individual routine timings as it includes the time for the minimisation routine itself.

which for each residue a list of all the residues it interacts with is stored. This neighbour list is set up prior to the non-bond calculation and has to be recalculated every so often if a cutoff is used. In the non-bond calculation a loop is performed over all the residues and for each residue the interactions of all atoms in it with all atoms of the residues in the neighbour list of this residue are computed. This routine was vectorised by calculating the interaction of 1 atom with all its neighbouring atoms as vector operations. This gives vector lengths of up to 1000 for a 10Å cutoff. A bit vector with the length of the number of atoms in the molecule is set up for each atom which indicates whether an atom interacts with this atom or not. This is a large array, $N^2/2$, where N is the number of atoms, but because of the bit addressing capability of the Cyber 205 this only takes up 70,000 words in memory. The performance improvement of this routine after vectorisation is 365 over the VAX, which includes the intrinsic scalar speed of the Cyber 205, some 14 times faster than the VAX. The vectorisation of the non-bond routine took approximately 1 month.

Valence energy calculation. The valence energy and cross-term routines take $\approx$20% of the iteration time on the VAX. These routines were vectorised next, starting with the torsion angle routine which is the next major time consuming routine, 6% of the iteration time on the VAX. The bond, valence angle and torsion angle routines already used a list of the internals in the VAX version. These were all vectorised by creating vectors for the bonds, valence angles and torsion angles, which gives vector lengths from 3000 to 9000 for the dihydrofolate reductase system, see table I. These vectorisations resulted in performance improvements of 37 to 90 over the VAX in these routines.

To date we have achieved a net gain in speed over the VAX 11/780 of 212 for the enzyme simulation study described above.

References

1. J.R. Bertino, *Arthritis and Rheumatism*, **79**, 16 (1973).
2. J. Bertino and D. Johns, in *Cancer Chemotherapy*, ed. I. Brodsky, vol. 2, p. 9, Grune and Stratton. New York (1972).
3. M. Osborn and F.M. Huennekens, *J. Biol. Chem.*, **969**, 233 (1958).
4. J. Burchall and G.H. Hitchings, *Mol. Pharmacol.*, **1**, 126 (1965).
5. B.R. Baker, in *Medicinal Chemistry 3rd ed.*, ed. A. Burger, Wiley-Interscience, New York (1970).
6. D.A. Matthews, R. Alden , J.T. Bolin, S.T. Freer, R. Hamlin, N. Xuong, J. Kraut, M. Poe, M. Williams, and K. Hoogstein, *Science*, **197**, 452 (1977).
7. D.A. Matthews, R.A. Alden, J.T. Bolin, D.J. Filman, S.T. Freer, R. Hamlin, W.G.J. Hol, R.L. Kisliuk, E.J. Pastore, L.T. Plante, N. Xuong, and J. Kraut, *J. Biol. Chem.*, **253**, 6946 (1978).
8. D.A. Matthews, R.A. Alden, S.T. Freer , N. Xuong, and J. Kraut, *J. Biol. Chem.*, **254**, 4144 (1979).
9. G. C. K. Roberts, J. Feeney, A. S. V. Burgen, V. Yuferov, J. G. Dann, and R. Bjur, *Biochemistry*, **13**, 5351 (1974).
10. J. Feeney, B. Birdsall, J. P. Albrand, G. C. K. Roberts, A. S. V. Burgen, P. A. Charlton, and D. W. Young, *Biochemistry*, **20**, 1837 (1981).
11. A. Gronenborn, B. Birdsall, E. Hyde, G. Roberts, J. Feeney, and A. Burgen, *Mol. Pharmacol.*, **20**, 145 (1981).
12. O. Ermer, *Structure and Bonding*, **27**, 161, Berlin (1976).
13. S. Califano, *Pure Appl. Chem.*, **18**, 353 (1969).
14. A.T. Hagler, E. Huler, and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5319 (1974).
15. A.T. Hagler and S. Lifson, *J. Am. Chem. Soc.*, **96**, 5327 (1974).
16. A.T. Hagler, S. Lifson, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5122 (1979).
17. A.T. Hagler, P. Dauber, and S. Lifson, *J. Am. Chem. Soc.*, **101**, 5131 (1979).
18. P. Dauber and A.T. Hagler, *Accts. of Chem Res.*, **13**, 105 (1980).
19. P. Dauber-Osguthorpe, J. Wolff, and A. T. Hagler. work in progress
20. S. Lifson, A.T. Hagler, and P. Dauber, *J. Am. Chem. Soc.*, **101**, 5111 (1979).
21. A.T. Hagler and A. Lapiccirella, *Biopolymers*, **15**, 1167 (1976).
22. E.B. Wilson, J.C. Decius, and P.C. Cross, in *Molecular Vibrations*, McGraw Hill, New York (1955).
23. T.L. Hill, in *An Introduction to Statistical Thermodynamics*, Addison-Wesley, Reading, Mass. (1960).
24. P. Dauber , M. Goodman, A.T. Hagler, D.J. Osguthorpe, R. Sharon, and P.S. Stern, *Proc. of the ACS Symposium on Supercomputers in Chemistry*, **173**, 161 (1981).
25. C.W. Gear, *Numerical Initial Value Problems in Ordinary Differential Equations*, Prentice Hall, Englwood Cliffs, N.J.
26. L. Verlet, *Phys. Rev.*, **159**, 98 (1967).