Journal of Computer-Aided Molecular Design, 7 (1993) 305–323
ESCOM

J-CAMD 203

An approximate but efficient method to calculate free energy trends by computer simulation: Application to dihydrofolate reductase–inhibitor complexes

Paul R. Gerber[a], Alan E. Mark[b] and Wilfred F. van Gunsteren[b,*]

[a]*Pharmaceutical Research and Development, F. Hoffmann-La Roche AG, CH-4002 Basle, Switzerland*
[b]*Department of Physical Chemistry, Swiss Federal Institute of Technology Zürich, ETH-Zentrum, CH-8092 Zürich, Switzerland*

Received 5 October 1992
Accepted 29 January 1993

Key words: Molecular dynamics; Free energy differences; Linear approximation; DHFR–inhibitor complexes

SUMMARY

Derivatives of free energy differences have been calculated by molecular dynamics techniques. The systems under study were ternary complexes of Trimethoprim (TMP) with dihydrofolate reductases of *E. coli* and chicken liver, containing the cofactor NADPH. Derivatives are taken with respect to modification of TMP, with emphasis on altering the 3-, 4- and 5-substituents of the phenyl ring. A linear approximation allows the encompassing of a whole set of modifications in a single simulation, as opposed to a full perturbation calculation, which requires a separate simulation for each modification. In the case considered here, the proposed technique requires a factor of 1000 less computing effort than a full free energy perturbation calculation. For the linear approximation to yield a significant result, one has to find ways of choosing the perturbation evolution, such that the initial trend mirrors the full calculation. The generation of new atoms requires a careful treatment of the singular terms in the non-bonded interaction. The result can be represented by maps of the changed molecule, which indicate whether complex formation is favoured under movement of partial charges and change in atom polarizabilities. Comparison with experimental measurements of inhibition constants reveals fair agreement in the range of values covered. However, detailed comparison fails to show a significant correlation. Possible reasons for the most pronounced deviations are given.

INTRODUCTION

One of the most promising aspects of molecular dynamics simulations is the possibility to calculate changes in free energy that originate from a modification in the Hamiltonian of the system under consideration. The general theory for such calculations is well established [1] and a variety of test systems have been studied in great detail. Reliable estimates for changes in equilib-

*To whom correspondence should be addressed.

rium constants of complex formation upon modification of a given ligand would be of great practical interest. In particular the pharmaceutical industry has a special interest in protein–inhibitor systems. Such systems are, however, computationally very demanding due to their size, especially when treated including solvent. A thorough treatment of a single inhibitor modification is currently still a matter of days or even weeks of computing time on the most powerful modern computers. Thus, such calculations can hardly be considered suitable for preliminary exploration or experiment planning. In terms of practical drug design the prediction of the relative free energy of binding of a single modified inhibitor will be of little use. It will, in almost all cases, be more efficient to synthetize and test the compound directly. The prediction of free energy trends for a large number of possible changes of the inhibitor would, in contrast, be of considerable help in guiding synthetic choices. This would be true as long as a range of modifications could be treated in a single calculation and the calculation yielded reasonable predictive power. The present work is an investigation of a possible method of this type. The method is related to work by Wong [2] on the analysis of the sensitivity of free energy calculations to changes in force field parameters. The method exploits the fact that the derivative of the total free energy of a system with respect to each term of the interaction potential can be calculated without the need to explicitly perform a specific perturbation, or to actually modify the Hamiltonian. We then explore the validity of using a linear approximation to extrapolate from these derivatives to the total change in free energy associated with a given modification of the Hamiltonian. The advantage of this approach is that because all of the derivatives are independent, they can be accumulated simultaneously during a single molecular dynamics simulation. In addition, all linear combinations of the calculated derivatives that lead to realistic final systems are included within the scope of the approximation. Clearly, if the results of such a treatment correlate to some extent with experimental findings such a method would form the basis of a very useful tool in areas such as drug design.

The methodology is tested by application to the extensively studied system of dihydrofolate reductase (DHFR)–inhibitor complexes.

THEORY

The relative free energy between two states, A and B, of a system can be determined using molecular dynamics simulation techniques if the potential energy function of the system, V, is expressed in terms of a coupling parameter, λ, such that when $\lambda = 0$ the system corresponds to state A ($V_{effective} = V_A$) and when $\lambda = 1$ the system corresponds to state B ($V_{effective} = V_B$). In this case the difference in Gibbs free energy between states A and B, $\Delta G_{A \to B}$, can be shown to be given by the so-called free energy thermodynamic integration formula (1)

$$\Delta G = \int_0^1 \left\langle \frac{\partial V}{\partial \lambda} \right\rangle_\lambda d\lambda \tag{1}$$

where the brackets $\langle \cdots \rangle_\lambda$ represent an ensemble average at a given value of λ. To determine the change in free energy by using Eq. 1, the parameter λ can either be slowly changed from 0 to 1 in a reversible manner during the course of the simulation or else the ensemble average can be determined at a series of specific values of λ and the integration performed numerically.

Other formulations can be used to determine changes in free energy from computer simulations. The advantage of the integration formula is that, because the potential, V, is normally

expressed as a linear combination of terms, it is possible using Eq. 1 to separate the various contributions to the total free energy. The derivative with respect to λ of each specific interaction in the system can be determined and integrated separately [3]. It is this principle that forms the basis of the approach we are proposing. We consider the system only in the initial state ($\lambda = 0$). Thus, during the simulation, there is no mixing of the potentials of the initial and final states and the system evolves as if it were unperturbed. An arbitrary final state can then be defined in which the charge and Lennard–Jones parameters of all atoms of interest are perturbed simultaneously. Partial derivatives with respect to each of these force-field parameter changes can then be calculated analytically and averaged for each atom. The derivatives will indicate the sign and magnitude in the free energy terms for a given change in charge or Lennard–Jones parameters of a specific atom for the ensemble of states sampled. The derivatives will, however, only be strictly valid for small changes of the force field parameters of a single atom. To estimate the change in free energy for a physically feasible modification, the derivatives with respect to different parameters must be combined and weighted appropriately. In the present study we investigate the possibility that the derivatives can be combined and weighted linearly.

Figure 1 illustrates the essence of the linear approximation. It assumes that the initial slope (dashed line) in a graph of the free energy versus perturbation parameter points towards a final value of the free energy which lies near the exact one (solid line). The validity of this assumption will largely depend on two factors: (i) on the degree of correspondence of states sampled when $\lambda = 0$ with the final state ($\lambda = 1$) of the perturbed system, and (ii) on the choice of the functional form used to calculate the derivatives. Obviously, one can imagine cases in which a linear assumption is not justified due to either of these factors. In this study we intend, however, to consider only enzyme–inhibitor systems. To a first approximation the protein can be thought of as providing a relatively rigid environment. In such cases the ensemble of states sampled when $\lambda = 0$ may well approximate those at $\lambda = 1$ as the response of the protein to a given perturbation is limited. The choice of the functional form of the derivatives remains, however, a critical issue.

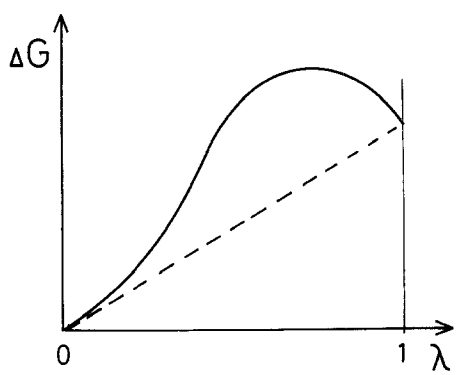


Fig. 1. Schematic representation of a possible free energy versus λ graph. The dashed line indicates the extrapolation of the initial linear approximation to the final value of the perturbation parameter. Clearly, the actual curve can distinctly deviate from linearity, but a suitable choice of the functional form of the development of the perturbation near $\lambda = 0$ can help to improve the linearity of the graph.

In standard free energy simulations the dependence of the potential energy function, V, on the coupling parameter, λ, is frequently simply a linear combination of the potential energy function in the initial state, V_A, and in the final state, V_B. The effective potential energy function is thus given by

$$V_{effective} = (1-\lambda)V_A + \lambda V_B \tag{2}$$

The derivative with respect to λ in this case is simply given by the difference between the potential in the initial and final states

$$\frac{\partial V}{\partial \lambda} = V_B - V_A \tag{3}$$

Coulombic interactions

For coulombic interactions the potential between two atoms, i and j, with charges, q_i and q_j, separated by a distance, r_{ij}, is given by

$$V^{C_{ij}} = \frac{q_i q_j}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \tag{4}$$

where ε_r is the relative dielectric constant. In a conventional simulation a change in q_i or q_j or both q_i and q_j would yield a single derivative for the pairwise interaction using Eq. 3. In the current work we treat the change of each parameter independently. This is equivalent to defining a perturbation parameter, λ_k, for each change and determining the partial derivatives. For example, when considering changes in q_i and q_j we would have

$$V^{C_{ij}}(\lambda_{q_i},\lambda_{q_j}) = \frac{[(1-\lambda_{q_i})q_i^A + \lambda_{q_i}q_i^B][(1-\lambda_{q_j})q_j^A + \lambda_{q_j}q_j^B]}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \tag{5}$$

If both charges q_i and q_j are altered, there will be two derivative contributions

$$\left.\frac{\partial V^{C_{ij}}}{\partial \lambda_{q_i}}\right|_{\lambda_{q_j}=0} = \frac{(q_j^B - q_j^A)q_i^A}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \tag{6}$$

and

$$\left.\frac{\partial V^{C_{ij}}}{\partial \lambda_{q_j}}\right|_{\lambda_{q_i}=0} = \frac{(q_j^B - q_j^A)q_i^A}{4\pi\varepsilon_0\varepsilon_r r_{ij}} \tag{7}$$

The first arises from changing q_i while keeping q_j at its original value, and the second from changing q_j while keeping q_i constant.

Lennard–Jones interactions

A similar argument applies for Lennard–Jones (LJ) interactions. The LJ potential between atoms i and j is given by

$$V^{LJ_{ij}} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \tag{8}$$

The derivative with respect to a change in the LJ parameters of atom i is given by

$$\left.\frac{\partial V^{LJ_{ij}}}{\partial \lambda_{LJ_i}}\right|_{\lambda_{LJ_j}=0} = 4\varepsilon_{ij}^{BA}\left[\left(\frac{\sigma_{ij}^{BA}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}^{BA}}{r_{ij}}\right)^{6}\right] - 4\varepsilon_{ij}^{AA}\left[\left(\frac{\sigma_{ij}^{AA}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}^{AA}}{r_{ij}}\right)^{6}\right] \tag{9}$$

In terms of the linear approximation we propose it should be noted that, while $V_{effective}$ is linearly dependent on λ, the derivatives given by Eqs. 6, 7 and 9 are strongly dependent on the distance separating the two atoms. The derivatives rapidly approach infinity as r_{ij} approaches zero. Normally, atoms are prevented from approaching each other very closely by the repulsive part of the LJ interaction and the response of the system to a change in charge or LJ parameters for well separated atoms will be approximately linear. A difficulty will be encountered, however, when new atoms are created. Such atoms experience no interactions in the initial state and may pass very close to their neighbours. Because of the singular nature of both the coulombic and LJ interactions this will give rise to extremely large contributions to the derivatives from a very small region of phase space. The problem associated with growing atoms is particularly severe with regard to LJ interactions [4]. Coulombic interactions grow more gently with decreasing distance and the singularity could be effectively avoided by placing nascent atoms at a distance of 0.1 nm from their bonded partners. This is well within the van der Waals shell of the adjacent atoms and precludes very close contacts. In the case of the LJ interactions, however, we considered it necessary to take additional measures to avoid the effects of close encounters, which completely dominate the value of the derivative.

Growing atoms in a gentle way is a recurring difficulty in free energy calculations. The problem is to ensure that the process of growing is distributed evenly over the entire interval in which the interaction develops to its full strength. Using a simple linear combination of the initial and final states as in Eq. 3 leads to initial values for the derivative which are far from representative for the whole interval. In cases where a full perturbation is performed these initial values will contribute in only a minor way to the total integral and the problem can be effectively handled by sampling more frequently as λ approaches zero. In our case the biasing effects of the derivative values at $\lambda = 0$ are severe. To avoid this problem we have chosen the following form for a developing LJ interaction between atoms i and j

$$V^{LJ_{ij}} = 4\varepsilon_{ij}\left[\frac{1}{[\alpha(\lambda) + (r_{ij}/\sigma_{ij})^6]^2} - \frac{1}{\alpha(\lambda) + (r_{ij}/\sigma_{ij})^6}\right] \tag{10}$$

in which the λ-dependent cutoff parameter, α, disappears as the interaction grows to its full size. The derivative at the beginning of the perturbation interval has the form

$$\left.\frac{\partial V^{LJ_{ij}}}{\partial \lambda_{LJ_i}}\right|_{\lambda_{LJ_i}=0,\ \lambda_{LJ_j}=0} = 4\varepsilon_{ij}^{BA}\left[\frac{1}{[\alpha(0) + (r_{ij}/\sigma_{ij}^{BA})^6]^2} - \frac{1}{\alpha(0) + (r_{ij}/\sigma_{ij}^{BA})^6}\right] \tag{11}$$

At the end of the interval the derivative has the same value as for the regular form (Eq. 9), provided the cutoff parameter, α, disappears faster than linear at $\lambda = 1$. The initial value of α is of minor importance as long as it ensures that the interaction at zero distance is of the same order of magnitude as the absolute value of the minimum of the interaction function. For this work we have chosen a value of $\alpha(0) = 0.618$. This leads to a value $V^{LJ} = 4\varepsilon$ when $r_{ij} = 0$. Figure 2 gives an impression how the interaction Eq. 10 develops to its full LJ form with decreasing values of α. It also illustrates that for distances larger than the eventual position of the minimum, which contribute the largest phase-space volume, the form very closely approaches that of a normal LJ potential. For this reason we considered Eq. 11 well suited for a linear approximation. To use Eq. 10 when growing atoms in full perturbation calculations a functional form of $\alpha(\lambda) = \alpha(0)\,(1-\lambda)^2$ may be appropriate.

Scaling the derivatives

The magnitude of the calculated derivatives using Eqs. 6, 7, and 9 will of course depend on the coulombic and LJ parameters chosen for atoms i and j in states A and B. In order to relate the calculated values to a chemically feasible modification the derivatives must be scaled appropriately. This is a simple matter in the case of the coulombic interactions as the derivative depends linearly on the difference in charge between states A and B. For example, if the initial charge on atom i in state A was q_x and the charge on atom i in state B in the simulation was taken to be q_y

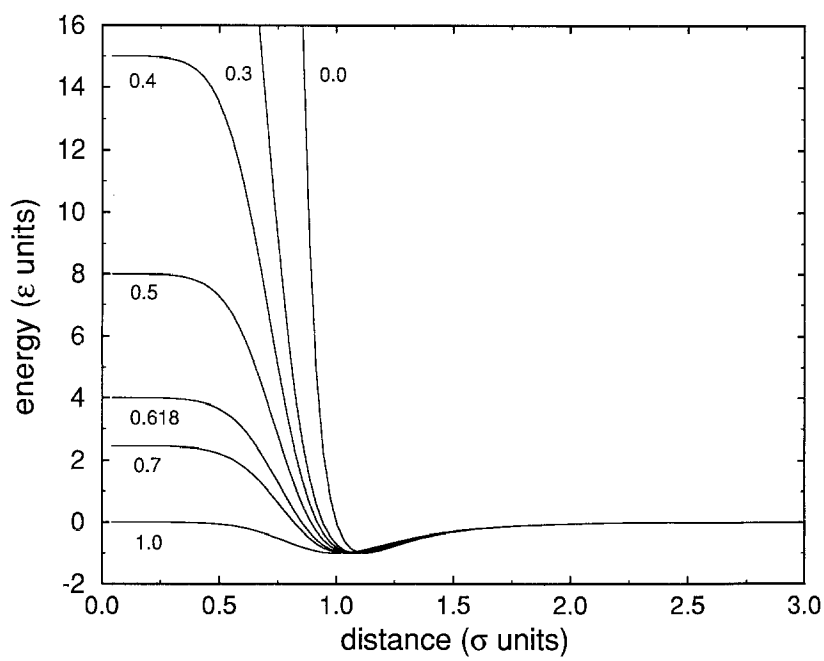


Fig. 2. Lennard–Jones functions (Eq. 10) for different values of the soft core parameter α. In our situation only two limiting cases are of importance, namely the fully grown potential ($\alpha = 0$) and the value of $\alpha = 0.618$ chosen for the start of a growing atom ($\lambda = 0$).

the scaling factor to obtain the appropriate derivative for atom i having a charge of q_z in state B would be

$$scale\ factor = \frac{q_x - q_z}{q_x - q_y} \tag{12}$$

The derivatives of the LJ interactions depend in a non-linear fashion on two parameters, ε and σ, for each pair of atoms. An exact treatment would require the determination of two derivatives for each atom and a functional form in which ε and σ were both directly dependent on λ. For simplicity we have calculated in Eq. 11 the LJ derivatives by taking the difference between the interaction of the perturbed atom in its final and initial states. To scale the LJ derivatives we proceed in an analogous manner to that used to scale the derivatives of the coulombic interactions. We assume that the derivative will be linearly dependent on the difference in an effective LJ strength, S^{LJ}, of a given atom in states A and B. While this is certainly an approximation it should be remembered that, primarily, we are not interested in exact values for the mutation of one atom into another. Rather we wish to observe general trends in the free energy. Such trends should not depend critically on the precise choice of S^{LJ}. For values of the relative LJ strengths of different atom types we used the relative volumes of the attractive part of the LJ potential. The volumes could be estimated as follows. Based on normal combination rules the energy of LJ terms is proportional to the square root of the product of the ε parameters for each of the two atoms. The range of the LJ interaction is given by the average of the σ parameters for each of the two atoms. Therefore, a plausible value for $S^{LJ}{}_i$ of an atom type i should be proportional to the square root of ε_i times the cube of the sum of its radius with an average radius of its partners.

$$S_i^{LJ} = \frac{K\sqrt{\varepsilon_i}}{(\sigma_i + \sigma_{av})^3} \tag{13}$$

The proportionality constant, K, was assumed to be the same for all atom types. For the average van der Waals radius, $\sigma_{av}/2$, we took a value of 0.18 nm. The appropriate scaling factors were then obtained by substituting values of $S^{LJ}{}_i$ for q_i in Eq. 12.

It is clear that in obtaining estimates for the changes in relative free energy in this way we are ignoring a number of potentially important factors. In particular the linear combination of the partial derivatives does not account for higher order effects, such as mutual interaction between two nascent charges or response of the system to a new charge. Such effects would require the evaluation of higher order derivatives of free energy differences with respect to the λ_k and is not practical in the current context. In addition, the effects of bonded interactions, bonds, angles and dihedral angle terms are ignored.

SYSTEMS UNDER STUDY

In order to test the method we have considered the binding of analogues of the inhibitor trimethoprim (TMP) to bacterial and vertebrate forms of dihydrofolate reductase (DHFR). This system was chosen as it is of considerable clinical interest and has been extensively studied from both an experimental and theoretical standpoint, e.g. see Ref. 5. X-ray structures have been

published for such complexes [6] and binding data are available for a wide variety of 3,4,5-phenyl substituted (2,4-diaminobenzyl)pyrimidines [7,8] of which TMP is the trimethoxy case. Derivatives were determined with respect to every atom of the inhibitor (TMP). In addition, each oxygen of the methoxy groups carried two dummy atoms, and the terminal methyl groups three. This enabled us to consider substituents up to three bonds long with simple branchings. Assuming just 10 different substituents could replace each of the three methoxy groups of TMP we could in this study estimate the free energy change for over a 1000 different compounds.

The calculations were performed using ternary complexes of chicken liver (cl) and *Escherichia coli* (ec) DHFR containing the cofactor NADPH. An experimental structure of the cl–DHFR complex was available [6]. In the case ec–DHFR the appropriate complex was modelled by replacing folate by TMP in an analogous experimental structure [9]. For the replacement we were guided by a model ternary complex proposed by Matthews et al. [6].

To relate the calculated free energy values to changes in binding energy it is necessary to consider the free energy associated with a given modification of TMP both bound to the protein and free in solution. The reference system for isolated TMP was taken initially in the conformation as found in the complex with cl–DHFR. The calculated free energies could then be related to the experimental values using the following thermodynamic cycle.

$$
\begin{array}{ccccc}
[\mathrm{DHFR(cl):I}] & \leftarrow \mathrm{exp} \rightarrow & \mathrm{I} & \leftarrow \mathrm{exp} \rightarrow & [\mathrm{DHFR(ec):I}] \\
\uparrow & & \uparrow & & \uparrow \\
\mathrm{sim} & & \mathrm{sim} & & \mathrm{sim} \\
\downarrow & & \downarrow & & \downarrow \\
[\mathrm{DHFR(cl):J}] & \leftarrow \mathrm{exp} \rightarrow & \mathrm{J} & \leftarrow \mathrm{exp} \rightarrow & [\mathrm{DHFR(ec):J}]
\end{array}
$$

The free inhibitor (I) is shown in the middle. On both sides the complexes with DHFR of either species are represented. The second line gives the same cases but for a modified inhibitor (J). Experimentally obtained differences in free energies correspond to the horizontal paths. The vertical changes correspond to the nonphysical process of mutating one ligand into another in the course of the simulation. As the total change in free energy for any cycle must sum to zero, the differences in the vertical paths should be equal to the differences in the horizontal paths.

Simulations and force field

All simulations were performed using a modified version of the GROMOS package of programs in conjunction with the standard GROMOS force field for solvent simulations [10]. In this force field apolar hydrogens are treated as united atoms together with the heavy atoms to which they are attached. Polar hydrogens which are treated explicitly experience coulombic but no LJ interactions. Partial atomic charges are adjusted to yield a proper description of hydrogen bonding effects. This results in a somewhat overemphasized polarity of bonds to explicitly treated hydrogens. During the simulations bond distances were constrained to their reference values using the procedure SHAKE [11]. The systems were treated using periodic boundary conditions

of either orthorhombic (rectangular box) or cubic symmetry (truncated octahedron) (see Table 1). Nonbonded interactions were treated using a twin range method in which short-range (< 0.8 nm) LJ and coulombic interactions were updated every time step (2 fs) while longer range (< 1.5 nm) coulomb interactions were updated at the same time as the pair list, every 10 fs [12]. The dimensions of the periodic box were chosen such that every pair of atoms could only experience interaction if both atoms were in the same molecule. The dimensions of the box were adjusted continuously during a simulation to have the pressure coupled to a reference value of one atmosphere with a time constant of 0.5 ps [13]. Similarly, the velocities were rescaled during simulation to keep the temperature near the reference value of 300 K with a time constant of 0.1 ps [13]. Space not occupied by the atoms of the solute molecule was filled by simple point charge (SPC) water molecules [14].

The three systems, two ternary complexes and free TMP were first equilibrated for 50 ps. Each trajectory was then split into n (seven) separate trajectories by choosing new random initial velocities. Upon continuing the simulations these sets diverged mutually with increasing root mean square deviations (rmsd). The mutual rmsd values stabilized at a value around 0.1 nm for C^{α} atoms after some 10 ps. This divergence was carried on for even longer times, depending on the size of the system. The advantage of this kind of parallel development is that running n well-separated trajectories for a given time is most likely to cover phase-space more evenly than a single trajectory of n-fold duration. In addition, this procedure allows for efficient data accumulation on multiprocessor machines. Table 1 contains various quantities that characterize the simulation and data acquisition.

Perturbations

In order to be able to cover a wide variety of 3,4,5-phenyl substituted (2,4-diaminobenzyl)pyrimidines [7,8], we allowed the perturbation to modify every atom of the inhibitor. In addition, each oxygen of the methoxy groups carried two dummy atoms and the terminal methyl

TABLE 1
DATA DESCRIBING THE SIMULATIONS AND TYPICAL CONFORMATIONS[a]

	System (DHFR complex)		
	E. coli	Chicken liver	Free TMP
Box shape	octahedron	octahedron	rectangular
Initial box size (nm)	6.46	7.02	3.45, 3.42, 3.40
Number of water molecules	3 704	4 802	1 338
Total number of atoms	12 738	16 369	4 055
Equilibration time (ps)	50	50	50
Divergence time (ps)	>10	>10	>10
Acquisition time (ps)	70	56	210
Pyrimidine dihedral angle (°)	5 (27)	68 (14)	
Phenyl dihedral angle (°)	79 (17)	81 (8)	

[a] The dihedral angles (see Fig. 3) are averages over the final conformations of the seven diverged trajectories. Root mean square deviations (given in brackets) confirm the significance of the difference in the pyrimidine dihedral.

groups carried three dummy atoms. Dummy atoms do not experience nonbonded interactions with other atoms. They effectively have no influence on the equilibrium properties of the system and are only used to determine the derivatives of the free energy corresponding to an atom being created at that location. The molecule with the attached dummy atoms is displayed in Fig. 3. Choosing dummy atoms in these positions enabled us to consider substituents up to three bonds long with simple branchings. To keep the dummy atoms at the methyl carbons in well-defined orientations a torsional potential with a single minimum and a transition barrier of 42 kJ/mol was applied. Equilibrium torsional angles were 180° for the first dummy, +60° for the second and −60° for the third one. The final value of the charge of an originally neutral atom was taken to be a positive elementary charge, while charged atoms evolved to neutrality. These changes are only used to define the derivatives and have no effect on the dynamics of the system. Similarly, LJ-interacting atoms developed to noninteracting ones, while dummy atoms, initially not affected by the LJ interaction had a final interaction strength corresponding to a bare carbon. Each of the atoms of interest were perturbed independently during the simulation and the derivatives with respect to the LJ and the electrostatic potential energy parameters were stored separately. The values of the derivatives do not depend in any way on whether surrounding atoms were also perturbed.

RESULTS AND DISCUSSION

It has been previously shown that the bound conformation of TMP in the two ternary complexes is different [6]. This was confirmed by the simulations and is illustrated by a comparison of typical values for the dihedral angles of the two bonds connecting the pyrimidine and phenyl rings given in Table 1. It should be noted that although TMP exhibits two-fold rotational symmetry at the phenyl ring, this symmetry is not born out in the MD simulations of the complexes. The time scale for the rotation of the phenyl ring within the binding pocket is much

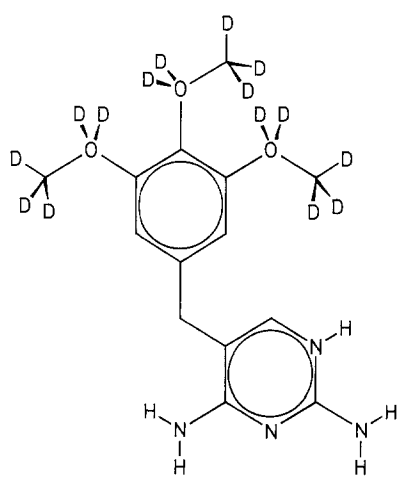
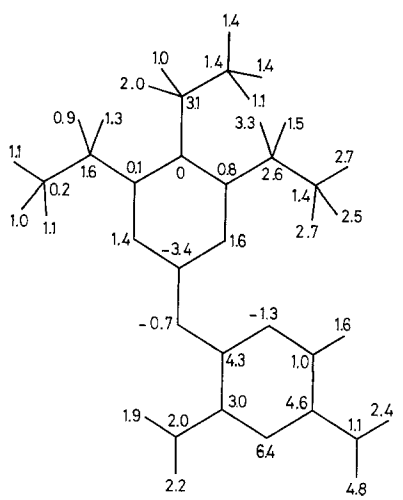


Fig. 3. Representation of trimethoprim, including all dummy atoms (labelled D) that were used in the simulations. In all the following figures containing this structure, atom labels and indications of steric arrangement will be omitted for clarity, but are understood to be identical to the ones shown in this figure. The 3-position of the phenyl substitution points to the right.

greater than that accessible during the simulation. For the isolated molecule in solution, however, the motion about the two dihedral angles was sufficiently free to make a distinction between the two symmetry-related conformers not possible. Consequently all results for the case of the isolated inhibitor have been symmetrized. This was achieved by averaging the values associated with equivalent atoms of the substituents at positions 3 and 5 on the phenyl ring and atoms 2 and 6, and atoms 3 and 5 on the phenyl ring itself. Where nonsymmetrized results for the two complexes are presented, they have been related to the average conformations given in terms of the phenyl dihedral angles given in Table 1. The values given there are to be understood as dihedral angles with respect to carbon 6 of the phenyl ring. Thus, in the complex with the bacterial DHFR the 3-position of the phenyl ring is taken to be oriented towards the cofactor and Ala[19], while in the chicken liver DHFR the 3-position of the phenyl points towards the Pro[26] residue. The binding of TMP to the *E. coli* and chicken liver enzymes is well illustrated in Figs. 4 and 5 of Matthews et al. [6] and we would refer the reader to the associated detailed description of both 3D structures contained therein.

The most direct way of representing the results of the derivative calculations is to use maps of the structure of TMP including the dummy atoms of the perturbation. These maps illustrate differences between derivatives for a complex and the corresponding values in the free inhibitor. Figures 4 and 5 show two such maps for LJ interactions. The number associated with each atom indicates the amount by which the complex would lower its free energy with respect to free TMP, if the LJ strength of the atom was increased by the value of a bare carbon, each atom having been perturbed independently. The values lie in the range of −3.4 to 7.4 kJ/mol. The values for the bacterial and mammalian enzymes are very similar and indicate that there is only a limited potential to develop more selective inhibitors by modifying LJ strengths or atomic polarizabilities. A conspicuous fact is, nevertheless, that the values are positive with very few exceptions. This



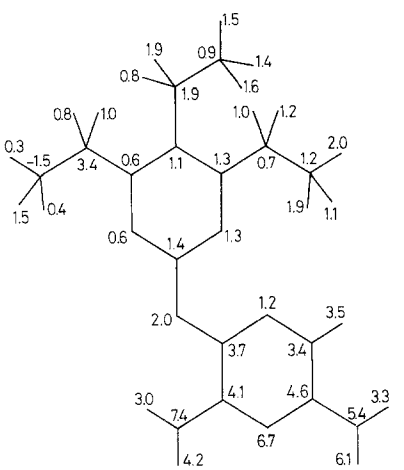
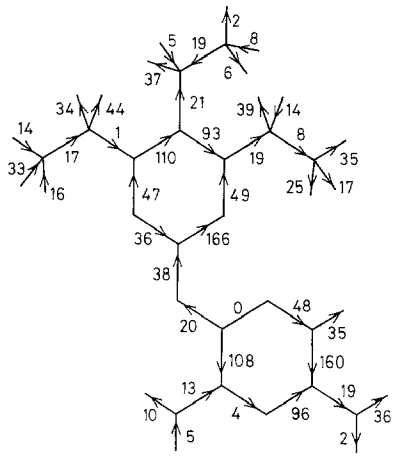


Fig. 4 (Left). Change in free energy upon increasing the LJ strengths of atoms by the strength corresponding to a bare carbon. Values are in kJ/mol. This map shows the difference between the complex of trimethoprim (TMP) with *E. coli* DHFR and NADP+ and the free trimethoprim in water. The free TMP data have been symmetrized with respect to exchanging 3- and 5-substituents.

Fig. 5 (Right). Same as Fig. 4 but for the complex with chicken liver DHFR.



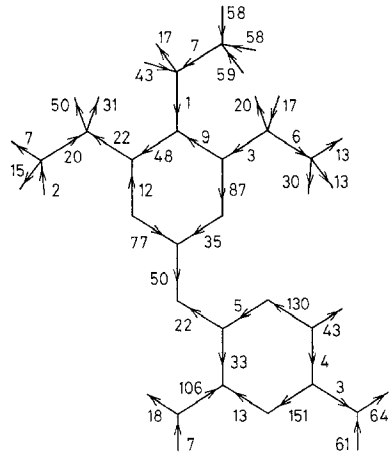


Fig. 6 (Left). Free electric fields along bonds, i.e. the change in free energy upon moving a positive elementary charge along a bond in the direction indicated by the arrow. Values are in kJ/mol. This map shows the difference between the complex of trimethoprim with *E. coli* DHFR and NADP+ and the free trimethoprim in water. The free TMP data have been symmetrized with respect to exchanging 3- and 5-substituents.
Fig. 7 (Right). Same as Fig. 6 but for the complex with chicken liver DHFR.

might indicate that water provides a relatively loose van der Waals contact when compared to a protein environment. However, this finding may be related to the fact that in the SPC water model hydrogens have no LJ interactions.

Much more pronounced results are found for the coulombic interaction as is illustrated by the maps of Figs. 6 and 7. Derivatives of the electrostatic interactions were determined for the modification of the charge on individual atoms. In fact this would correspond to the creation-isolated charges. We are, however, primarily interested in modifications of the inhibitor that result in no change in net charge. For this reason the derivatives have been transformed to

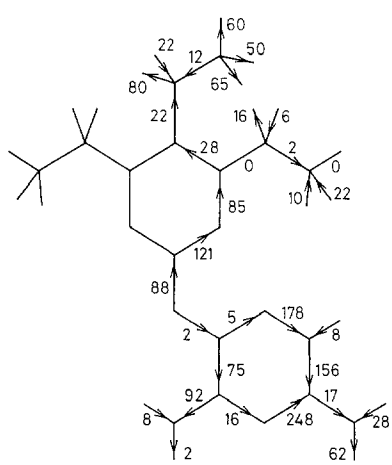


Fig. 8. Difference map between Fig. 6 and Fig. 7. Since there is no obvious correspondence of phenyl ring orientation in the two complexes, this representation has been symmetrized with respect to exchanging 3- and 5-substituents.

TABLE 2
PARTIAL ATOMIC CHARGES AT SUBSTITUENTS AS OBTAINED FROM ADJUSTING CHARGE-DEPENDENT ELECTRONEGATIVITIES [15][a]

Substituent	Phenyl-C	1. atom	2. atom	3. atom	4. atom	5. atom
OMet	17	−43	26			
	18	−36	18			
OH	17	−42	25			
	15	−55	40			
OEt	17	−43	21	5		
OCHO	22	−28	30	16	−40	
	18	−36	36	20	−38	
OSO_2Met	18	−33	19	−11	−11	18
	18	−36	36	−18	−18	18
SMet	5	−19	14			
	10	−20	10			
Cl	9	−9				
	14	−14				
NH_2	2	−32	15	15		
	0	−82	41	41		
$NMet_2$	3	−33	15	15		
	0	−60	30	30		
NO_2	29	23	−26	−26		
NHCOMet	9	−18	18	21	−44	14
Met	−4	4				
CF_3	8	40	−16	−16	−16	
Et	−12	3	9			
MetOH	−1	19	−37	19		
	0	15	−55	40		
MetOMet	0	18	−37	19		
CMet=C	0	−4	4	0		
OCF_3	20	−33	52	−13	−13	−13

[a]Second rows give established values of the GROMOS force field [10]. Units are in percent of an elementary charge. Hydrogens participate in a united atom, unless explicitly mentioned.

correspond to the amount by which the free energy of the complex would be lowered, compared to that of the free inhibitor, if a positive elementary charge would be moved along the bond from one atom to the next in the direction indicated by the arrow. This number is obtained by combining the effects of lowering the partial charge of one atom and increasing the partial charge on the other. The values range from 0 to 166 kJ/mol. Changing the polarization of a bond by shifting even a small fraction of an elementary charge may easily change the free energy of binding by several kT. In terms of the selectivity between the bacterial and mammalian enzymes the effect is even more pronounced. Figure 8 shows the difference map, ec–cl. It should be noted that the binding environments of the two enzymes are different. It is thus, not obvious how to assign a proper correspondence of symmetry-related conformations, e.g. the designation of a particular substituent as being at position 3 or at position 5 being in effect arbitrary. Rather than separate consideration of both possible combinations the values presented in Fig. 8 have, for simplicity, been symmetrized. Nevertheless, dramatic differences are observed in several places.

TABLE 3
DIFFERENCES OF FREE ENERGY DERIVATIVES FOR CHANGING A METHOXY SUBSTITUENT OF TRIMETHOPRIM IN VARIOUS POSITIONS AT THE PHENYL RING [a]

	Charges from electronegativity			Charges from standard GROMOS		
	Substitution at position					
Substituent	3	4	5	3	4	5
–	4.85	−1.84	−1.61	4.05	−0.52	−0.24
OH	0.71	−2.48	−7.71	−1.32	−3.53	−10.72
OEt	−6.32	−0.64	0.66	−6.32	−0.64	0.66
OCHO	9.32	−1.51	−5.33	6.13	−0.04	−1.98
OSO_2Met	−11.32	−6.89	−6.40	−7.97	−6.34	−8.86
SMet	−3.87	−5.56	−1.66	−2.63	−5.29	−2.38
Cl	1.41	−3.76	−2.95	2.33	−2.71	−2.98
NH_2	−2.06	−5.16	−8.26	−8.16	−7.77	−18.74
$NMet_2$	−7.85	−7.95	−7.66	−11.45	−8.10	−10.90
NO_2	6.07	−3.25	1.34			
NHCOMet	−4.24	−13.17	−23.30			
Met	−1.00	−6.47	−2.85			
CF_3	2.04	−4.31	5.39			
Et	−5.17	−8.33	−1.49			
MetOH	−5.67	−10.89	−2.09	−11.10	−10.08	2.88
MetOMet	−11.61	−13.32	−5.31			
CMet=C	−4.85	−7.44	−3.23			
OCF_3	3.72	−3.50	−12.26			

[a]Difference between values for the complex with *E. coli* DHFR and the ones for the free inhibitor. Values given are in kJ/mol.

In particular, shifts within the two ring systems and differences in the dummy atoms of the 4-substituent would be predicted to have major effects on the selectivity of the inhibitor. The large differences in the pyrimidine ring region were surprising. One might have expected little difference between the two species, since the pyrimidine ring binds within the active site of the enzyme and the local environment of both enzymes is arranged to perform the same catalytic function. Upon examination of the neighbourhood of the pyrimidine ring in the two complexes, it became clear that during the simulations considerably more water molecules lie in close proximity to the pyrimidine ring in the *E. coli* complex than in the chicken liver complex.

From the maps one can progress towards envisaging specifically altered structures. We restricted this part to modifying the phenyl substituents in order to be able to make the connection to the experimental values quoted by Selassie et al. [7,8]. To this end we calculated the effect of changing single substituents in any of the three positions for the two complexes as differences to the free inhibitor. This calculation requires, as a central ingredient, partial atomic charges for all substituents in question. Table 2 shows the charges that were utilized. They originate from a simple calculation of adjusting charge-dependent electronegativities in the spirit of Gasteiger and Marsili [15]. Since the GROMOS force field makes use of enhanced partial charge values near polar hydrogen atoms, we also included the corresponding values for some of the substituents. The

TABLE 4
DIFFERENCES OF FREE ENERGY DERIVATIVES FOR CHANGING A METHOXY SUBSTITUENT OF TRIMETHOPRIM IN VARIOUS POSITIONS AT THE PHENYL RING[a]

	Charges from electronegativity			Charges from standard GROMOS		
	Substitution at position					
Substituent	3	4	5	3	4	5
–	5.18	2.51	−6.59	4.69	3.09	−5.17
OH	2.43	2.96	−11.36	2.07	4.61	−14.81
OEt	−4.04	1.12	−0.26	−4.04	1.12	−0.26
OCHO	0.50	−16.48	−1.03	−1.27	−12.11	0.14
OSO_2Met	−7.04	−9.66	−3.01	−4.59	−17.60	−1.43
SMet	0.21	−1.49	−5.92	0.30	−1.81	−5.65
Cl	3.25	−0.13	−8.79	3.09	−0.16	−7.70
NH_2	2.92	4.01	−14.89	1.50	9.70	−25.80
$NMet_2$	−1.09	0.00	−10.66	−2.12	3.01	−14.31
NO_2	1.61	−7.86	3.07			
NHCOMet	4.91	−20.62	−14.93			
Met	3.67	−0.05	−11.62			
CF_3	1.80	−7.70	2.61			
Et	1.46	−0.67	−8.75			
MetOH	−0.49	8.61	−13.48	−4.67	21.05	−14.10
MetOMet	−2.81	6.10	−13.80			
CMet=C	0.13	−2.12	−6.64			
OCF_3	3.44	−26.95	−1.07			

[a] Difference between values for the complex with chicken liver DHFR and the ones for the free inhibitor. Values given are in kJ/mol.

values displayed in Figs. 6 and 7, together with the partial charges of Table 2, then yield the derivatives of the free energy under substituent replacement. The values quoted in Tables 3 and 4 also include the LJ contribution. In calculating these values an average over the different possibilities to achieve the mutation of the substituent was taken. This may be illustrated for the case where a methoxy group is mutated to a hydroxy one. This can either be achieved by changing the terminal methyl into a hydrogen, or alternatively by deleting the methyl and letting either of the dummy atoms at the oxygen develop into a hydrogen.

In these tables a few points are particularly noteworthy. For the *E. coli* complex it appears that replacing the methoxy substituent at position 4 by any of the groups tested should increase the stability of the complex. Substitutions at position 5 are also predominantly favourable. At the 3-position positive and negative values occur. A similar trend is observed in the case of the chicken liver complex, with substitutions at positions 4 and 5 being predominantly favourable. A few ligands show conspicuously large values. In general, these can be traced back to modifications involving the introduction of large dipole moments along bonds to dummy atoms. The most extreme effects occur at the 4-substituent in the chicken liver complex, where already moderately polar substituents, such as trifluormethoxyl or *n*-acetyl acquire large values for the calculated free energy change. The case of the 4-substituent of the chicken liver complex was investigated in more

TABLE 5
COMPARISON OF EXPERIMENTAL INHIBITION DATA WITH DIFFERENCES OF FREE ENERGY DERIVATIVES[a]

Phenyl substituents			DHFR from *E. coli*				DHFR from chicken liver			
3	4	5	Exp.	MD	MDG	QSAR	Exp.	MD	MDG	QSAR
Et	Et	Et	−44.8	−51.0	−51.7	−46.7	−30.1	−27.2	−27.7	−26.3
OMet	OMet	OMet	−46.3	−36.0	−36.6	−45.3	−22.8	−19.3	−19.7	−22.3
OMet	NMet2	OMet	−44.2	−44.0	−44.7	−46.6	−23.8	−19.3	−16.7	−24.4
OMet	Br	OMet	−46.9	−39.7	−39.4	−45.5	−26.0	−19.4	−19.9	−25.0
OMet	SMet	OMet	−46.3	−41.6	−41.9	−46.4	−24.6	−20.8	−21.6	−23.7
OMet	CMet=C	OMet	−46.6	−43.5	−44.1	−46.6	−23.9	−21.4	−21.9	−24.4
MetOH	–	MetOH	−36.2	−45.6	−45.4	−36.3	−18.5	−30.8	−35.4	−18.9
OMet	–	OMet	−44.2	−37.9	−37.2	−42.9	−23.6	−16.8	−16.6	−23.7
OEt	–	OEt	−44.1	−43.5	−42.8	−43.7	−23.7	−21.1	−20.9	−23.7
Met	–	Met	−40.4	−41.7	−41.1	−41.5	−26.4	−24.8	−24.6	−26.2
OH	OH	–	−37.0	−41.3	−46.9	−37.4	−20.6	−22.5	−25.2	−23.1
OCHO	OCHO	–	−40.9	−38.0	−34.6	−38.9	−26.8	−41.9	−38.3	−25.1
OMet	OMet	–	−44.3	−37.6	−36.9	−41.8	−25.6	−25.9	−24.9	−24.5
OH	OMet	–	−39.2	−38.9	−43.3	−39.1	−22.1	−25.5	−29.9	−23.9
OMet	OSO_2Met	–	−45.5	−44.5	−43.2	−43.1	−26.3	−35.6	−42.5	−25.2
OMet	OH	–	−43.2	−40.1	−40.4	−40.1	−24.7	−22.9	−20.3	−23.5
OSO_2Met	OMet	–	−44.7	−43.9	−43.2	−41.8	−26.0	−29.9	−29.0	−24.4
CF_3	OMet	–	−44.1	−35.6	−34.9	−40.9	−28.6	−24.1	−23.1	−28.4
MetOH	–	–	−36.0	−45.1	−48.5	−38.9	−24.7	−25.1	−26.5	−23.8
OSO_2Met	–	–	−39.7	−45.8	−43.8	−39.3	−24.8	−27.4	−25.9	−25.1
MetOMet	–	–	−37.8	−51.1	−49.0	−39.3	−25.1	−26.2	−25.8	−24.4
OH	–	–	−37.1	−40.7	−43.9	−36.6	−22.2	−23.0	−26.8	−25.7
OMet	–	–	−39.7	−39.5	−37.4	−39.2	−25.5	−23.4	−21.8	−26.2
Met	–	–	−38.4	−40.5	−38.4	−38.9	−27.1	−23.2	−23.6	−27.1
Cl	–	–	−38.1	−38.1	−36.1	−39.1	−28.7	−20.4	−19.7	−28.5
CF_3	–	–	−40.3	−37.4	−35.4	−38.5	−28.2	−21.6	−20.0	−29.1
–	NH_2	–	−36.1	−37.9	−40.6	−37.3	−21.4	−16.7	−10.5	−21.3
–	NHCOMet	–	−39.5	−45.9	−46.0	−39.2	−24.4	−41.3	−40.8	−24.1
–	OSO_2Met	–	−37.9	−39.7	−39.2	−39.4	−24.7	−30.4	−37.8	−24.9
–	OH	–	−37.0	−35.3	−36.4	−36.4	−23.9	−17.8	−15.6	−24.3
–	NO_2	–	−35.6	−36.0	−36.1	−37.9	−25.1	−28.6	−28.1	−28.1
–	OMet	–	−39.1	−32.8	−32.9	−38.0	−24.6	−20.7	−20.2	−25.1
–	$NMet_2$	–	−38.9	−40.7	−40.9	−39.3	−23.0	−20.7	−17.2	−24.4
–	Met	–	−37.2	−39.2	−39.3	−37.3	−26.2	−20.8	−20.2	−26.6
–	Cl	–	−37.0	−36.5	−35.6	−37.4	−27.7	−20.9	−20.4	−27.7
–	OCF_3	–	−37.7	−36.3	−36.4	−38.0	−24.8	−47.7	−47.1	−28.4
–	–	–	−35.4	−34.6	−33.4	−35.6	−27.0	−18.2	−17.1	−26.1

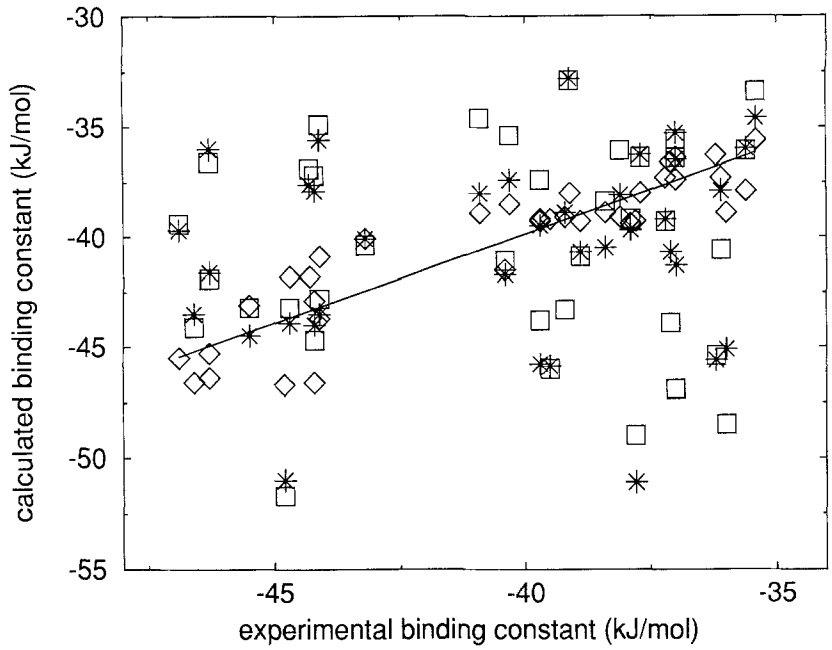


Fig. 9. Comparison of logarithms of experimental inverse inhibition constants and calculated values for the difference of free energy derivatives. The experimental values have been appropriately scaled, using a temperature of 300 K, to yield free energy values in kJ/mol. The calculated values have been shifted to reproduce the mean of the experimental ones. The graph shows data for the complex of trimethoprim with *E. coli* DHFR and $NADP^+$ and for free trimethoprim in water. ✳, MD simulation, using partial charges from electron negativities (Table 2); □, MDG simulation, using GROMOS partial charges (Table 2); ◇, QSAR values from Refs. 7 and 8. The solid line is a linear least-squares fit of the QSAR data.

detail to attempt to trace the origin of the large electric fields along the bonds to the dummy atoms. Inspection of the trajectories revealed that water molecules were frequently present near this substituent. Furthermore, these water molecules show a high degree of orientation, with their respective hydrogen atoms pointing towards the substituent. Such an orientation is favoured by the close presence of charged and polar groups, such as Arg^{70}, Lys^{68}, Gln^{35} and Asn^{64}. Water molecules are highly polar and carry large partial charges. As the hydrogen atoms of the waters experience no LJ repulsion, it is possible that very close encounters (<0.1 nm) between water hydrogens and dummy atoms can occur. Although, the coulombic potential has a comparatively mild singularity as compared to the LJ repulsion, such encounters, though probably rare, may still bias the calculated derivatives, which in the linear approximation used here determines the resulting free energy difference.

←

[a] Also shown are values from QSAR fitting [7,8]. The two sets of MD data differ by the partial charges used at the substituent atoms (MDG used GROMOS standard charges, where available). Values are in kJ/mol. The experimental values are derived from measured inhibition constants using a temperature of 300 K. Because the calculation yields only differences, the given values have been placed on the energy scale such as to reproduce the average of the experimental values.

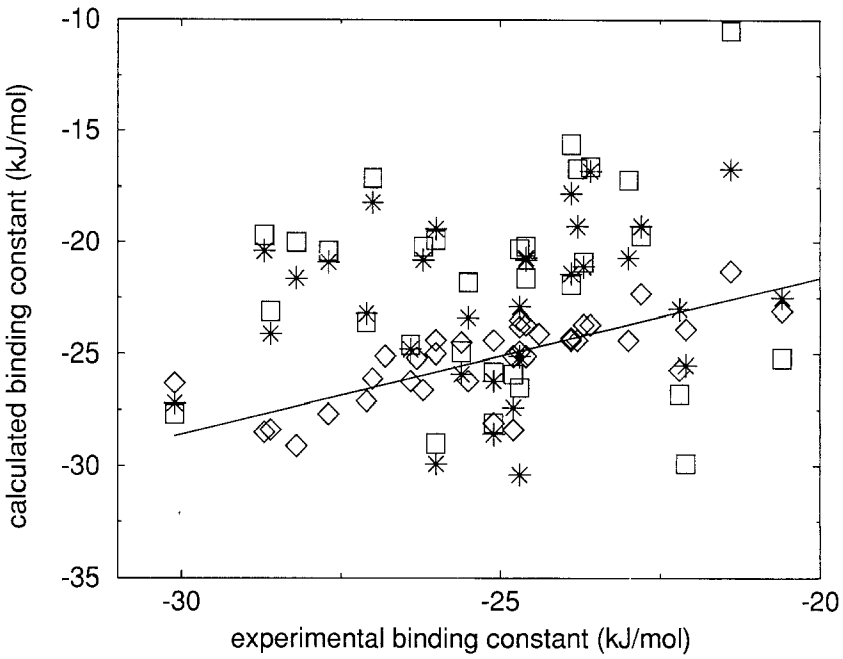


Fig. 10. Same as Fig. 9 but for the complex with chicken liver DHFR.

Finally, for comparison with the experiment, Table 5 contains a list of those compounds, that differ from trimethoprim by modified phenyl substitutions, and which are within reach of our perturbation. For the cases of asymmetric substitution (unequal 3- and 5-phenyl substituents) the orientation with the lower free energy was taken. Table 5 shows the experimental values together with the results from the simulation for the two sets of charges used, as well as a fit to the experiment using QSAR techniques [7,8]. The correlation between the experimental and calculated values is shown in Figs. 9 and 10. Apart from problems associated with polar substitutions at dummy atoms as discussed above which lead to several eccentric values, the spread of the results from the simulation is comparable to that of the experimental values. Details are, however, poorly reproduced. This is illustrated by the low correlation between calculated and experimental values. The values of the correlation coefficients are below 0.2, marginally better for the set of charges that omit increased hydrogen polarity. This may be viewed as a further indication that the derivatives of the coulombic interaction are biased.

Possibly the most disappointing aspect of this study was that the conspicuous role of 3,5-methoxy, trisubstituted benzylpyrimidines in specifically inhibiting *E. coli* DHFR, as compared to vertebrate DHFR, is hardly born out in the calculation. Clearly in this regard and in comparison to QSAR results, the method we have proposed does poorly with this system. It should be remembered, however, that the QSAR method depends on empirical relationships based on a large set of experimental data. The present method is in contrast essentially ab initio in character. It could in principle be used in cases where no experimental information in regard to alternative ligands is available.

CONCLUSIONS

Given the poor correlation between experimental inhibition constants and calculated derivatives, it is premature to propose the method as a useful tool for the prediction of inhibition trends. Nevertheless, maps of the type shown in Figs. 4 to 8 may serve as indicators of possible novel modifications of the inhibitors. We feel that with refinements, methods such as this have the potential to be useful in situations where little experimental information is available, or where a completely new class of inhibitor modification is envisaged. Obviously, however, further work is needed to analyse the validity of the various assumptions that have been made in the current study. One hope is that this may be done by tracing the sources of the most conspicuous discrepancies between experiment and calculation. There is a need to investigate whether a more gentle growth of the coulombic interaction of newly generated charges would result in better agreement with experimental results. The importance of cross terms and bonded terms also need to be considered as does the effect of scaling the LJ potential in the manner we have chosen.

In summary, this work has pointed toward a novel way of using a single molecular dynamics situation to investigate the possible effects of a wide range of modifications of a single ligand. When compared to full free energy perturbation calculations, the linear approximation technique reported here is easily a factor of 1000 more efficient at the expense of a loss of accuracy due to the neglect of the non-linear dependency of the free energy upon the force field parameters. A new method to grow atoms during free energy calculations has also been introduced. While it is clear that the simplistic approach we have taken in this study is inadequate, we feel that similar approximate methods do have potential in the field of drug design.

REFERENCES

1 Beveridge, D.L. and DiCapua, F.M., Annu. Rev. Biophys. Biophys. Chem., 18 (1989) 431.
2 Wong, C.F., J. Am. Chem. Soc., 113 (1991) 3208.
3 Gao, J., Kuczera, A., Tidor, B. and Karplus, M., Science, 244 (1989) 1069.
4 Cross, A.J., Chem. Phys. Lett., 128 (1986) 198.
5 McDonald, J.J. and Brooks, III, C.L., J. Am. Chem. Soc., 114 (1992) 2062.
6 Matthews, D.A., Bolin, J.T., Burridge, J.M., Filman, D.J., Volz, K.W., Kaufman, B.T., Bedell, C.R., Champness, J.N., Stammers, D.K. and Kraut J., J. Biol. Chem., 260 (1985) 381.
7 Selassie, C.D., Fang, Z-X., Li, R-L., Hansch, C., Debnath, G., Klein, T.E., Langridge, R. and Kaufman, B.T., J. Med. Chem., 32 (1989) 1895.
8 Selassie, C.D., Li, R-L., Poe, M. and Hansch, C., J. Med. Chem., 34 (1991) 46.
9 Bystroff, C., Oatley, S.J. and Kraut, J., Biochemistry, 29 (1990) 3263.
10 van Gunsteren, W.F. and Berendsen, H.J.C., Groningen Molecular Similation (GROMOS) Library Manual, Biomos, Nijenborgh 16, 9747 AG Groningen, The Netherlands, 1987.
11 Ryckaert, J.-P., Ciccotti, G. and Berendsen, H.J.C., J. Comput. Phys., 23 (1977) 327.
12 van Gunsteren, W.F. and Berendsen, H.J.C., Angew. Chem. Int. Ed. Engl., 29 (1990) 992.
13 Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. and Haak, J.R., J. Chem. Phys., 92 (1984) 3684.
14 Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F. and Hermans, J., In Pullman, B., (Ed.) Intermolecular Forces, Reidel, Dordrecht, 1981, pp. 331–342.
15 Gasteiger, J. and Marsili, M., Tetrahedron, 36 (1980) 3219.