Journal of Computer-Aided Molecular Design, 7 (1993) 535–555
ESCOM

J-CAMD 206

Computer-aided drug design: A free energy perturbation study on the binding of methyl-substituted pterins and N5-deazapterins to dihydrofolate reductase

Peter L. Cummins and Jill E. Gready*

Department of Biochemistry, University of Sydney, Sydney, N.S.W. 2006, Australia

Received 16 October 1992
Accepted 17 February 1993

Key words: Molecular dynamics; Thermodynamics; Hydration; Ligand–protein interactions

SUMMARY

Molecular dynamics simulation and free energy perturbation techniques have been used to study the relative binding free energies of 8-methylpterins and 8-methyl-N5-deazapterins to dihydrofolate reductase (DHFR). Methyl-substitution at the 5, 6 and 7 positions in the N-heterocyclic ring gives rise to a variety of ring substituent patterns and biological activity: several of these methyl derivatives of the 8-methyl parent compounds (8-methylpterin and 8-methyl-N5-deazapterin) have been identified as substrates or inhibitors of vertebrate DHFR in previous work. The calculated free energy differences reveal that the methyl-substituted compounds are thermodynamically more stable than the primary compounds (8-methylpterin and 8-methyl-N5-deazapterin) when bound to the enzyme, due largely to hydrophobic hydration phenomena. Methyl substitution at the 5 and/or 7 positions in the 6-methyl-substituted compounds has only a small effect on the stability of ligand binding. Furthermore, repulsive interactions between the 6-methyl substituent and DHFR are minimal, suggesting that the 6-methyl position is optimal for binding. The results also show that similarly substituted 8-methylpterins and 8-methyl-N5-deazapterins have very similar affinities for binding to DHFR. The computer simulation predictions are in broad agreement with experimental data obtained from kinetic studies, i.e. 6,8-dimethylpterin is a more efficient substrate than 8-methylpterin and 6,8-dimethyl-N5-deazapterin is a better inhibitor than 8-methyl-N5-deazapterin.

INTRODUCTION

Dihydrofolate reductase (DHFR) catalyses the nicotinamide adenine dinucleotide phosphate (NADPH)-dependent reduction of folate to dihydrofolate and tetrahydrofolate, and is a target for various antineoplastic and antibacterial drugs. DHFR is also a relatively small and stable enzyme, which makes it suitable for high-resolution crystallographic studies. These factors have made DHFR an extensively studied enzyme, in terms of both its structure and function [1,2]. We

*To whom correspondence should be addressed.

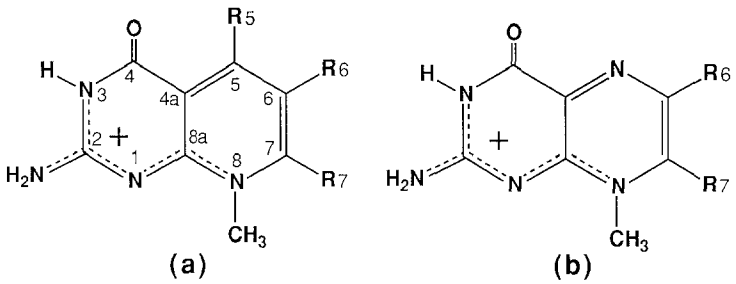

Fig. 1. Structures for cations (i.e. N3 protonated forms) of the 8-methyl substituted N5-deazapterins (a) and pterins (b): R5, R6 and R7 may be either H or CH_3.

are currently engaged in developing several classes of new compounds as possible substrates or inhibitors of DHFR [3–9], with the aim of obtaining improved lead compounds for further development as cytotoxic drugs. The N3 protonated forms of two such classes of compounds, based on 8-methyl substituted pterin and N5-deazapterin derivatives, are shown in Fig. 1. The details of the design rationale and physico-chemical properties for these pterin substrates have been discussed at length elsewhere [3–7,10–15]. We simply note here that the 8-R substituent is central to the design rationale, as this results in higher basicities for formation of the cations (shown in Fig. 1) and hence the pterins becoming active substrates. Related design rationales based on the results of ab initio calculations [16] and experimental evidence for N5-deazafolate [17,18] also suggested that 8-R-N5-deazapterins [8,9] might well be inhibitors.

While 8-R-substitution is a necessary condition for the substrate activity [3–6], a variety of ring-substituent patterns is also possible, at least in principle, by the addition of methyl groups at the 5, 6 and 7 positions of the heterocyclic ring in the case of 8-methyl-N5-deazapterin and the 6 and 7 positions in 8-methylpterin (see Fig. 1). The primary compounds, 8-methyl-N5-deazapterin (8d) and 8-methylpterin (8p), and their various methyl derivatives are listed in Table 1, together with the abbreviations that will be used in the following text and tables. The pterins exhibit a range of substrate and binding activity depending on pH, enzyme source (avian and human) and

TABLE 1
LIGANDS AND ABBREVIATIONS USED IN THE PRESENT WORK

Ligand name	Abbreviation
5,6,7,8-Tetramethyl-N5-deazapterin	5678d
6,7,8-Trimethyl-N5-deazapterin	678d
5,7,8-Trimethyl-N5-deazapterin	578d
5,6,8-Trimethyl-N5-deazapterin	568d
7,8-Dimethyl-N5-deazapterin	78d
6,8-Dimethyl-N5-deazapterin	68d
5,8-Dimethyl-N5-deazapterin	58d
8-Methyl-N5-deazapterin	8d
6,7,8-Trimethylpterin	678p
7,8-Dimethylpterin	78p
6,8-Dimethylpterin	68p
8-Methylpterin	8p

ring-substituent pattern [3–7]. The primary compound of the pterin series (8p) and the 6-substituted compound (68p) were found to be good substrates of vertebrate DHFRs, while pterins with a 7-methyl group (78p and 678p) displayed minimal substrate activity. As anticipated, the deazapterins 8-methyl-N5-deazapterin and 6,8-dimethyl-N5-deazapterin have been found to inhibit the reduction of substrate by DHFR and display strong binding in ternary complexes with NADPH cofactor [8,19].

As part of our overall strategy for the rational design of biologically active molecules, we are examining the utility of computer-based approaches such as free energy perturbation (FEP) methods [20,21] for studying the thermodynamics of substrate and inhibitor binding to DHFR. In the FEP method, the free energy differences are obtained by a 'chemical transformation' or a 'mutating' of the potential energy parameters of a ligand A to those of a second ligand B during a molecular dynamics (MD) simulation. The relative thermodynamic stabilities of A and B are given by the free energy difference

$$\Delta\Delta F_{bind} = \Delta F_{bind} - \Delta F_{solv} \tag{1}$$

where ΔF_{solv} is from the mutation A $\rightarrow$ B carried out for the free ligand in aqueous solution and ΔF_{bind} is from the mutation A $\rightarrow$ B carried out for the ligand bound to the protein in an aqueous environment.

Several FEP calculations on the binding of known inhibitors of DHFR, such as the drugs methotrexate and trimethoprim, have been reported in the literature [22–27]. In this paper, we present the results of FEP calculations of $\Delta\Delta F_{bind}$ for the N3 protonated forms of the various methyl derivatives of 8-methyl-N5-deazapterin (8d) and 8-methylpterin (8p) (Table 1) bound to clDHFR·NADPH (DHFR from chicken liver) [28–30]. Errors in the calculation of free energy differences via MD simulations may arise from several sources, including the approximate nature of simulation force-fields, difficulties in sampling configuration space, the truncation of long-range (i.e. electrostatic) interactions, and the imposition of boundaries and choice of ensemble [31–35]. It is beyond the scope of this paper to examine all such possible sources of error: in this work we confine our attention to estimating the magnitudes of errors that may arise from incomplete sampling of configuration space during the MD simulations. An analysis of the mean structures and atomic fluctuations (root-mean-square vibrational amplitudes) obtained from MD simulation is also discussed with reference to the efficiency of configuration-space sampling.

The ultimate goal of rational drug design is to optimize the structural characteristics of a molecule for a particular function, in this case, for strong binding to the active site of DHFR. The primary aim of this work is therefore to test whether the standard FEP methodology can furnish useful information for guiding the design process. Since many of the derivatives listed in Table 1 have not yet been synthesized or assayed, an obvious design question is whether some would give rise to significantly improved binding over those derivatives which have already tested positive for biological activity (e.g. 68p, 8p, 68d and 8d). Accordingly, the thermodynamic quantities ΔF_{bind}, ΔF_{solv} and $\Delta\Delta F_{bind}$ (Eq. 1) are given relative to a common reference, i.e. the known inhibitor 8-methyl-N5-deazapterin (8d), enabling an estimation of the differences in ligand-binding to DHFR between any pair of compounds. In order to check the validity of the computer simulation predictions, the results of the free energy calculations are correlated with the available experimental data [3,4,7,8,19] for the ligands bound to DHFR.

METHODS

Free energy perturbation methods

The potential energy function used in the MD simulations has the general form

$$V = V_{bad} + V_{ele} + V_{vdw} + V_{hb} \quad (2)$$

where V_{bad} represents the intramolecular part of the potential energy, which includes all bond, angle and dihedral terms, V_{ele} is the electrostatic term arising from nonbonded interactions between atomic charges, V_{vdw} is the nonbonded van der Waals (vdW) interaction term, and V_{hb} is a hydrogen bond interaction term. The all-atom and united-atom protein force-field parameters of Weiner et al. [36,37] were adopted for the DHFR molecule. The more accurate all-atom model [36] was used for residues within 8 Å of the ligand centre of mass (c.m.), while the united-atom model (in which hydrogens attached to carbon are not represented explicitly) [37] was used for the remainder of the protein. The NADPH cofactor force-field parameters used are described in an earlier work [38]. Water molecules were assigned the TIP3P force-field parameters of Jorgensen et al. [39]. Structural parameters for the ligands were obtained at the semiempirical AM1 level [40]. Atomic partial charges for the ligands were also computed at the AM1 level from the molecular electrostatic potential (MEP) using an efficient interaction energy method [41], while standard values were used for vdW parameters as described in previous work [38]. For hydrogens H(N2) and H(N3), which are involved in direct H-bonding with the enzyme active site (via Glu^-) [4,6], we have used a modified form for the vdW interaction which correctly takes into account the effects of H-bonding [42].

The theory behind free energy perturbation calculations can be found elsewhere [21,32,43]. In the present simulations, which were carried out using the AMBER programs [44], the transformation between initial and final Hamiltonian (i.e. Eq. 2) states, A and B, was achieved via λ-coupling [21,32], where the entire ligand molecule was taken to be the perturbed group of atoms. The coupling parameter λ was divided into discrete values, λ_i, to yield 'windows' of uniform width $\Delta\lambda_i$. For each window an equilibration simulation was performed followed by a simulation period in which all data were collected for the calculation of the perturbation free energy as $\Delta F(+\Delta\lambda_i)$ and $\Delta F(-\Delta\lambda_i)$. The perturbation $\Delta\lambda_i$ must be chosen small enough to ensure that $\Delta F(+\Delta\lambda_i) \approx \Delta F(-\Delta\lambda_i)$. Thus two estimates of the total free energy change were obtained by summing over windows for the $\pm\,\Delta\lambda$ perturbations

$$\Delta F(\pm\,\Delta\lambda) = \sum_i \Delta F(\pm\,\Delta\lambda_i) \quad (3)$$

Mutation pathways

While there are 66, i.e. $n(n-1)/2$ for n compounds, distinct mutations which directly connect any pair of the 12 compounds listed in Table 1, it is unnecessary to consider them all since the free energy is a path-independent quantity. In theory, the minimum number of FEP simulations that are required to obtain all free energy differences is therefore only 11, i.e. $n-1$. Nevertheless, since in practice configuration space is never completely sampled during the time scale of a typical simulation, the carrying out of some additional simulations may provide some indication of the errors due to sampling. Thus free energy changes between two ligands can be estimated by

following various pathways that each involve a differing sequence of mutations. For example, to obtain the free energy difference between 678d and 8d one would normally think of performing the mutation 678d → 8d directly. However, the desired free energy can also be obtained by recognizing that the change should not be dependent on the path taken and performing the sequence of mutations 678d → 68d → 8d where the total free energy is simply the sum of free energies for the two mutations in the sequence. We have chosen a subset of 19 (given in the Results section) of the 66 possible mutations: these involve $CH_3 \rightarrow H$ mutations of a single methyl substituent and mutations between similarly substituted pterins and N5-deazapterins.

Simulation conditions for calculation of ΔF_{solv}

The procedures used to calculate the relative solvation free energies in this work follow those used by Rao and Singh [45] in their FEP study of hydrophobic hydration. There are two main reasons for using this approach unmodified. First, the types of mutation being carried out here are very similar to those considered by Rao and Singh. Second, Rao and Singh obtained satisfactory agreement with experiment for a range of large to small molecular ion mutations and apolar to polar molecule mutations, suggesting that the parameters governing the behaviour of the simulations, such as the extent of solvation, nonbonded interaction cutoffs, number of windows in FEP calculations, etc., are capable of adequately reproducing the general features of the hydration phenomena, including free energy differences. The essential details of these simulation conditions are given below.

The ligands were first solvated with boxes of Monte Carlo water using the EDIT module of AMBER. Only those water molecules within 12 Å (in the x, y and z directions) of any solute atom were retained to form a rectangular simulation box. The simulations were carried out for the closed, isothermal, isobaric (NTP) ensemble at a temperature of 300 K and pressure of 1 atm. Periodic boundary conditions were applied and a cutoff radius of 8 Å was used for all nonbonded interactions between solute–solvent and solvent–solvent molecules (with constant dielectric $\varepsilon = 1$).

As shown in Fig. 2, the calculations were divided into four stages. Energy minimization (BORN module) was followed by MD equilibration for 7 ps before the mutations were started (GIBBS module). To reduce the likelihood of possible path-dependency errors that may arise in the calculation of ΔF_{solv}, the mutation A → B was carried out in two stages via an intermediate state A′ with coordinate coupling [45]. The intermediate state A′ takes on the electrostatic force field of the final state B, but has the same intramolecular and vdW force field as the initial state A. Thus the mutation A → A′ gives the 'electrostatic' contribution ΔF_{ele}. The transformation is completed by the mutation A′ → B to give the 'vdW' contribution ΔF_{vdw}, where the remaining potential energy terms are changed to the final state B. Coordinate coupling (CC) was applied in the calculation of ΔF_{vdw}, and the SHAKE algorithm [46] was used to constrain all bond lengths to their specified equilibrium values.

Simulation conditions for calculation of ΔF_{bind}

The starting coordinates for the clDHFR·NADPH·ligand complexes were modelled on the clDHFR·$NADP^+$·biopterin X-ray structure [30]. The clDHFR·NADPH·ligand complexes were neutralized by adding counterions and the region within a radius of 16 Å of the ligand centre was solvated using EDIT. Note that in the present simulations, the active-site residue Glu^{30} is ionized

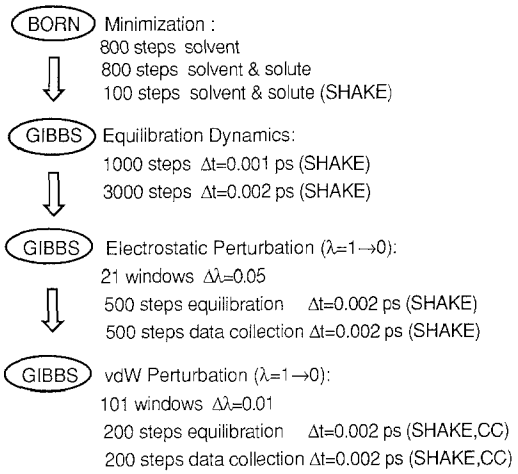

Fig. 2. Simulation conditions for the calculation of the solvation free energy differences ΔF_{solv}: Δt = integration time step, CC = coordinate coupling, $\Delta\lambda$ = increment in the coupling parameter.

and neutralized by the protonated form of the ligand (refer to Fig. 2a of Ref. 6). Due to the prohibitive computational cost of solvating the entire protein in a periodic simulation box, we have used an efficient active-site dynamics approach [47] for the calculation of ΔF_{bind}. In this approach, the system is partitioned into 'non-dynamic' atoms whose positions are frozen at the

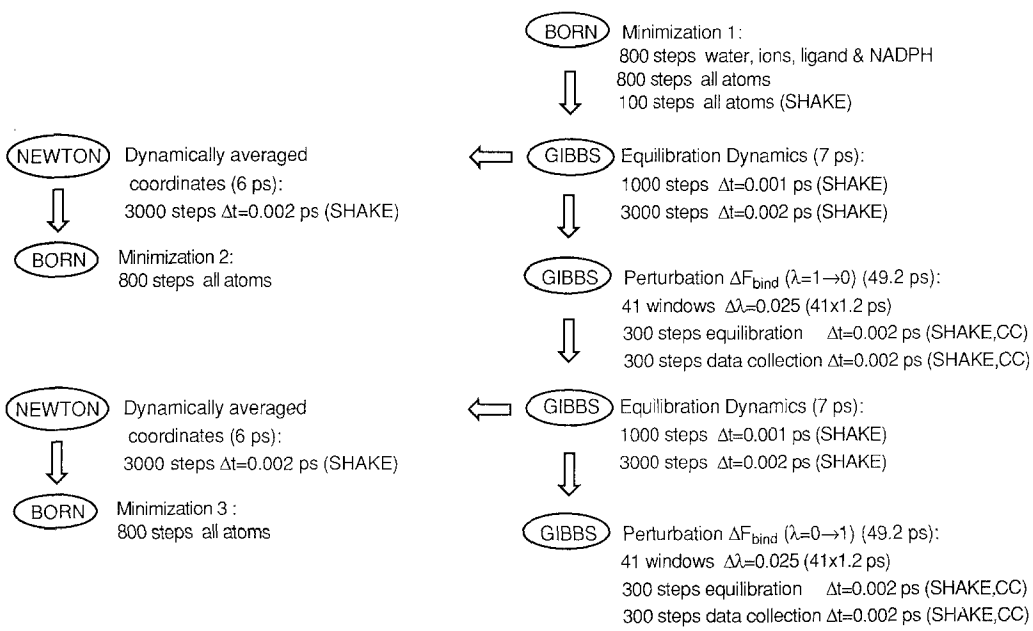

Fig. 3. Simulation conditions for the calculation of the binding free energy differences ΔF_{bind}: Δt = integration time step, CC = coordinate coupling, $\Delta\lambda$ = increment in the coupling parameter.

initial X-ray coordinate values and 'dynamic' atoms which are the only ones allowed to move during the various stages of the calculations. The dynamic atoms were defined by including all protein residues, counterions and water within a radius of 16 Å from the ligand centre of mass. Using these truncation conditions all of the important active-site protein residues are included and typically 50–60 water molecules were involved in the dynamics. Water molecules and counter ions were restrained from leaving the dynamics region by applying a small harmonic (0.6 kcal/mol/Å^2) radial restoring force [47]. A dynamics zone of 16 Å radius also appears to be a reasonable choice on the basis of previous work where this technique has been used to study ligand-protein binding in DHFR [22–27]. The simulations were carried out at a constant temperature of 300 K and using an 8 Å cutoff for all nonbonded interactions (constant dielectric $\varepsilon = 1$), consistent with the calculation of the relative solvation free energies, ΔF_{solv}.

The procedure for calculating ΔF_{bind} and performing the coordinate averaging is summarized in Fig. 3. Energy minimization (BORN module) was followed by a series of equilibration and FEP simulations with coordinate coupling (CC) [45] and the application of SHAKE [46] to constrain bond lengths to equilibrium values. In the calculation of ΔF_{bind}, a simulation was carried out for both the forward mutation (λ: 1→0) and the reverse mutation (λ: 0→1). The final coordinates for the forward mutation (λ: 1→0) were used to start the simulation for the reverse mutation (λ: 0→1). The final coordinates from the 7 ps equilibration simulations at both $\lambda = 1$

TABLE 2
DIMENSIONS OF THE PERIODIC SIMULATION BOXES USED IN THE SOLVATION FREE ENERGY CALCULATIONS

Simulation	Lengths (Å) of periodic box[a]			Number of solvent molecules
	x	y	z	
5678d → 678d	33.19	30.64	26.05	851
5678d → 578d	33.19	30.64	26.05	851
5678d → 568d	33.19	30.64	26.05	851
678d → 78d	32.96	30.09	26.07	833
678d → 68d	32.96	30.09	26.07	833
578d → 78d	32.86	30.78	26.02	835
578d → 58d	32.86	30.78	26.02	835
568d → 68d	32.91	30.67	25.84	834
568d → 58d	32.91	30.67	25.84	834
68d → 8d	33.14	30.43	25.97	857
58d → 8d	32.02	30.70	25.84	831
78d → 8d	32.78	29.91	26.02	831
678d → 678p	32.96	30.09	26.07	833
68d → 68p	33.14	30.43	25.97	857
78d → 78p	32.78	29.91	26.02	831
8d → 8p	31.73	29.93	25.84	804
678p → 78p	32.92	30.07	26.10	830
678p → 68p	32.92	30.07	26.10	830
68p → 8p	33.07	30.39	25.97	847

[a] z-axis is perpendicular to plane of the heterocyclic ring (Fig. 1).

and $\lambda = 0$ were subject to a further 6 ps of MD over which the atomic coordinates were averaged and rms vibrational amplitudes computed (NEWTON module) [48]. The mean structures obtained from the simulations using NEWTON were then energy minimized for the purposes of graphical analysis using the MidasPlus program [49] and least-squares superposition.

RESULTS

Computed free energies

The dimensions of the periodic box and the numbers of water molecules for the 19 mutations carried out in solvent according to the procedure outlined in Fig. 2 are given in Table 2. The corresponding solvation free energy changes ΔF_{solv} are given in Table 3. The arrows, e.g. in 68d $\rightarrow$ 8d, indicate the direction of the forward (λ: 1$\rightarrow$0) mutation. The binding free energies ΔF_{bind} for the same 19 mutations computed according to the procedure in Fig. 3 are given in Table 4. The ΔF_{bind} value is the mean and standard error of the four free energy changes calculated using the procedure in Fig. 3, i.e. $\Delta F(+\Delta\lambda)$ and $\Delta F(-\Delta\lambda)$ for both the forward (λ: 1$\rightarrow$0) and reverse (λ: 0$\rightarrow$1) mutations.

The effects on ΔF_{bind} of increases in simulation times (from ca. 50 ps to 100 ps) during the perturbation calculations were assessed by performing simulations for the mutations 78d $\rightarrow$ 8d and 678d $\rightarrow$ 678p using doubled equilibration and data collection times, i.e. using 1200 steps per

TABLE 3
DIFFERENCES IN FREE ENERGIES OF SOLVATION ΔF_{solv} (KCAL/MOL), CALCULATED ACCORDING TO THE PROCEDURE OUTLINED IN FIG. 2

Simulation	ΔF_{ele} [a]	ΔF_{vdw} [a]	ΔF_{solv} [b]
5678d $\rightarrow$ 678d	−0.93 ± 0.00	−0.85 ± 0.01	−1.78 ± 0.01
5678d $\rightarrow$ 578d	−0.34 ± 0.01	−0.36 ± 0.02	−0.70 ± 0.03
5678d $\rightarrow$ 568d	−0.20 ± 0.00	−0.22 ± 0.00	−0.44 ± 0.00
678d $\rightarrow$ 78d	−0.33 ± 0.01	−0.64 ± 0.03	−0.97 ± 0.04
678d $\rightarrow$ 68d	−0.15 ± 0.03	−0.98 ± 0.06	−1.13 ± 0.09
578d $\rightarrow$ 78d	−1.10 ± 0.02	−0.04 ± 0.01	−1.14 ± 0.03
578d $\rightarrow$ 58d	−0.23 ± 0.01	−0.29 ± 0.01	−0.52 ± 0.02
568d $\rightarrow$ 68d	−0.96 ± 0.01	−0.54 ± 0.04	−1.50 ± 0.05
568d $\rightarrow$ 58d	−0.29 ± 0.00	−0.84 ± 0.03	−1.13 ± 0.03
68d $\rightarrow$ 8d	−0.61 ± 0.01	−0.88 ± 0.05	−1.49 ± 0.06
58d $\rightarrow$ 8d	−1.06 ± 0.00	−1.23 ± 0.04	−2.29 ± 0.04
78d $\rightarrow$ 8d	−0.30 ± 0.01	−0.77 ± 0.02	−1.07 ± 0.03
678d $\rightarrow$ 678p	−3.40 ± 0.01	−0.72 ± 0.02	−4.12 ± 0.03
68d $\rightarrow$ 68p	−3.34 ± 0.02	−0.44 ± 0.01	−3.78 ± 0.03
78d $\rightarrow$ 78p	−3.88 ± 0.04	−0.47 ± 0.02	−4.35 ± 0.06
8d $\rightarrow$ 8p	−3.74 ± 0.02	−0.29 ± 0.01	−4.03 ± 0.03
678p $\rightarrow$ 78p	−0.87 ± 0.01	−0.10 ± 0.07	−0.97 ± 0.08
678p $\rightarrow$ 68p	−0.15 ± 0.03	−0.12 ± 0.02	−0.27 ± 0.05
68p $\rightarrow$ 8p	−0.85 ± 0.01	−1.22 ± 0.07	−2.07 ± 0.08

[a] Mean and standard error of the two free energies, $\Delta F(+\Delta\lambda)$ and $\Delta F(-\Delta\lambda)$, obtained by double-wide sampling (see text).
[b] The total free energy change (ΔF_{solv}) is the sum of the electrostatic (ΔF_{ele}) and vdW (ΔF_{vdw}) components.

window instead of the 600 steps indicated in the Fig. 3 procedure. Only the free energies $\Delta F(+\Delta\lambda)$ and $\Delta F(-\Delta\lambda)$ for the forward (λ: 1→0) mutation were calculated. The result for the 78d → 8d mutation was -1.23 ± 0.06 kcal/mol, which compares favourably with the value of -1.09 ± 0.02 kcal/mol calculated from the results for the shorter (50 ps) time-scale simulation (Table 4). The 100-ps simulation carried out for the 678d → 678p mutation gave -4.53 ± 0.04 kcal/mol compared with -4.94 ± 0.02 kcal/mol for the 50 ps simulation time.

In order to compare the ligands according to their affinity for binding to DHFR, we have estimated ΔF_{solv} and ΔF_{bind} relative to the inhibitor 8-methyl-N5-deazapterin (8d) in the following manner. The data in Tables 3 and 4 were used to calculate the free energy changes, ΔF_{solv} and ΔF_{bind}, via several pathways starting from 8d. The results of this calculation are given in Table 5. Then the final estimates of the free energy changes, ΔF_{solv} and ΔF_{bind} relative to 8d, were obtained by averaging the free energies over the different pathways. These final results are listed in Table 6. The final errors in ΔF_{solv} and ΔF_{bind} have been arrived at in one of two ways. Firstly, errors were computed by ignoring the errors for the individual pathways quoted in Table 5 and taking the standard deviation from the mean obtained by averaging the free energies over pathways. Alter-

TABLE 4
DIFFERENCES IN FREE ENERGIES OF BINDING ΔF_{bind} (KCAL/MOL), CALCULATED ACCORDING TO THE PROCEDURE OUTLINED IN FIG. 3

Simulation	ΔF [λ: 1→0][a]		ΔF [λ: 0→1][a]		ΔF_{bind}[b]
	$+\Delta\lambda$	$-\Delta\lambda$	$+\Delta\lambda$	$-\Delta\lambda$	
5678d → 678d[c]	−1.52	1.55	0.91	−1.01	−1.20 ± 0.30
5678d → 578d[c]	0.25	−0.32	0.30	−0.27	0.00 ± 0.33
5678d → 568d	−0.55	0.49	−0.21	0.28	−0.14 ± 0.44
678d → 78d	−0.26	0.12	0.24	−0.37	−0.25 ± 0.10
678d → 68d	−0.55	0.56	1.11	−1.10	−0.83 ± 0.32
578d → 78d	−0.98	1.02	1.08	−1.09	−1.04 ± 0.05
578d → 58d	0.23	−0.22	−0.67	0.71	0.46 ± 0.27
568d → 68d	−1.15	1.07	1.56	−1.72	−1.38 ± 0.31
568d → 58d	−0.16	0.10	−0.45	0.38	0.14 ± 0.32
68d → 8d	0.15	−0.05	0.12	−0.11	−0.01 ± 0.13
58d → 8d	−1.58	1.49	1.34	−1.32	−1.43 ± 0.12
78d → 8d[c]	−1.06	1.11	2.79	−2.84	−1.95 ± 1.00
678d → 678p	−4.92	4.96	5.95	−5.91	−5.44 ± 0.57
68d → 68p	−3.95	3.97	4.26	−4.28	−4.12 ± 0.18
78d → 78p	−4.46	4.50	4.82	−4.83	−4.65 ± 0.20
8d → 8p[c]	−3.95	3.96	3.79	−3.62	−3.83 ± 0.16
678p → 78p	0.16	−0.18	−0.02	0.08	0.11 ± 0.07
678p → 68p	0.13	−0.12	0.57	−0.61	−0.23 ± 0.41
68p → 8p[d]	0.44	−0.45	−0.25	0.35	0.37 ± 0.09

[a] ΔF values are given for $+\Delta\lambda$ and $-\Delta\lambda$ perturbations from the forward (λ: 1→0) and reverse (λ: 0→1) mutations.
[b] Estimate of ΔF_{bind} given by the mean values and standard errors.
[c] Change in the ligand-binding geometry occurred during the $\lambda = 1 \rightarrow 0$ mutation (see text and Figs. 9 and 10 for 8d → 8p and 78d → 8d).
[d] Change in the ligand-binding geometry occurred during the $\lambda = 0 \rightarrow 1$ mutation (see text).

TABLE 5
SOLVATION AND BINDING FREE ENERGIES, ΔF_{solv} AND ΔF_{bind} (KCAL/MOL), RELATIVE TO 8-METHYL-N5-DEAZAPTERIN (8d) FOR VARIOUS MUTATION PATHWAYS

Ligand	Pathway	ΔF_{solv}[a]	ΔF_{bind}[a]
5678d	← 568d ← 68d ← 8d	3.41 ± 0.06	1.53 ± 0.51
5678d	← 568d ← 58d ← 8d	3.84 ± 0.04	1.43 ± 0.51
5678d	← 568d ← 68d ← 678d ← 78d ← 578d ← 58d ← 8d	3.43 ± 0.10	0.87 ± 0.61
678d	← 5678d ← 568d ← 68d ← 8d	1.63 ± 0.06	0.33 ± 0.59
678d	← 68d ← 8d	2.62 ± 0.11	0.84 ± 0.32
678d	← 78d ← 578d ← 58d ← 8d	2.64 ± 0.07	0.18 ± 0.27
678d	← 78d ← 78p ← 678 p ← 678d ← 68d ← 8d	2.85 ± 0.15	−0.41 ± 0.55
678d	← 678p ← 78p ← 78d ← 678d ← 68d ← 8d	2.39 ± 0.15	1.09 ± 0.57
578d	← 5678d ← 568d ← 68d ← 8d	2.71 ± 0.07	1.53 ± 0.61
578d	← 58d ← 8d	2.81 ± 0.04	0.97 ± 0.28
578d	← 58d ← 568d ← 68d ← 8d	2.38 ± 0.08	1.07 ± 0.52
568d	← 68d ← 8d	2.99 ± 0.08	1.39 ± 0.31
568d	← 58d ← 8d	3.42 ± 0.05	1.29 ± 0.31
568d	← 68d ← 678d ← 78d ← 578d ← 58d ← 8d	3.01 ± 0.11	0.73 ± 0.48
78d	← 8d	1.07 ± 0.03	1.95 ± 1.00
78d	← 678d ← 68d ← 8d	1.65 ± 0.11	0.59 ± 0.32
78d	← 578d ← 58d ← 8d	1.64 ± 0.05	−0.07 ± 0.25
78d	← 78p ← 678p ← 678d ← 68d ← 8d	1.88 ± 0.14	−0.66 ± 0.56
68d	← 8d	1.49 ± 0.06	0.01 ± 0.13
68d	← 678d ← 78d ← 578d ← 58d ← 8d	1.51 ± 0.10	−0.65 ± 0.38
68d	← 68p ← 678p ← 678d ← 68d ← 8d	2.01 ± 0.12	−0.71 ± 0.72
68d	← 678d ← 678p ← 78p ← 78d ← 678d ← 68d ← 8d	1.26 ± 0.17	−1.29 ± 0.63
58d	← 8d	2.29 ± 0.04	1.43 ± 0.12
58d	← 568d ← 68d ← 8d	1.86 ± 0.08	1.53 ± 0.44
58d	← 568d ← 68d ← 678d ← 78d ← 578d ← 58d ← 8d	1.88 ± 0.11	0.54 ± 0.56
678p	← 678d ← 68d ← 8d	−1.50 ± 0.10	−4.60 ± 0.59
678p	← 78p ← 78d ← 678d ← 68d ← 8d	−1.73 ± 0.15	−3.35 ± 0.50
678p	← 68p ← 68d ← 8d	−2.02 ± 0.08	−3.88 ± 0.42
68p	← 68d ← 8d	−2.29 ± 0.06	−4.11 ± 0.22
68p	← 678p ← 678d ← 68d ← 8d	−1.77 ± 0.12	−4.83 ± 0.72
68p	← 68d ← 678d ← 78d ← 578d ← 58d ← 8d	−2.27 ± 0.10	−3.47 ± 0.42
78p	← 78d ← 8d	−3.28 ± 0.04	−2.70 ± 0.82
78p	← 78d ← 678d ← 68d ← 8d	−2.70 ± 0.13	−3.24 ± 0.35
78p	← 678p ← 678d ← 68d ← 8d	−2.47 ± 0.10	−4.49 ± 0.55
8p	← 8d	−4.03 ± 0.03	−3.83 ± 0.16
8p	← 68p ← 68d ← 8d	−4.36 ± 0.10	−3.74 ± 0.23
8p	← 68p ← 678p ← 678d ← 68d ← 8d	−3.84 ± 0.14	−4.46 ± 0.68

[a] The free energies have been estimated for various pathways using the results of the FEP calculations (ΔF_{solv} and ΔF_{bind}) listed in Tables 3 and 4. Errors have been estimated as $n^{-1/2} \Sigma\sigma_i$ where σ_i are the standard errors quoted in Tables 3 and 4 (n = number of mutations in pathway).

natively, the errors were calculated simply by taking the average of the errors listed in Table 5. In Table 6, we give the larger of these two error estimates. The resulting estimates of $\Delta\Delta F_{bind}$ relative to 8-methyl-N5-deazapterin (8d) calculated using Eq. 1 are also listed in Table 6. The errors in $\Delta\Delta F_{bind}$ were obtained by adding the errors of the composite free energies in accordance with the usual rules for adding standard deviations.

Structures of ligand–protein complexes obtained by MD simulation

The procedure for calculating ΔF_{bind} in Fig. 3 includes two starting structures for the MD analysis stage of the calculations (NEWTON module), thus yielding a set of 38 individual simulations involving clDHFR·NADPH·ligand complexes for analysis. We note first that the potential energy function (Eq. 2) reproduces the essential features of the starting X-ray structure, since the initial energy minimized coordinates (minimization 1 in Fig. 3) were only slightly changed from the starting X-ray structure; the differences from least-squares fit of the dynamic heavy atom positions being less than 0.4 Å with maximum deviations not exceeding 1.0 Å. There were also no significant structural changes in the mean MD coordinates obtained from the simulations using NEWTON after they had been energy minimized (minimizations 2 and 3 in Fig. 3). Figures 4 to 7 illustrate various aspects of the protein-backbone structure obtained from selected MD simulations (minimizations 2 and 3 in Fig. 3) and of the average over all 38 simulations. Note that the low flat regions (Figs. 4–6) correspond to residues composed of nondynamic atoms, which obviously have zero vibrational amplitude (Fig. 7). The plots in Fig. 4 give the distances between the protein-backbone alpha-carbon (C^{α}) positions after a least-squares superposition of the C^{α} coordinates of two representative MD structures. The segments involved in secondary protein structure (β-sheet and α-helix) [2,28,29] are indicated. The broken line traces distances between clDHFR structures where the bound ligand is 68d in each of the two complexes involved in the

TABLE 6
FREE ENERGIES (KCAL/MOL) RELATIVE TO 8-METHYL-N5-DEAZAPTERIN AND AVERAGED OVER THE VARIOUS MUTATION PATHWAYS

Ligand	ΔF_{solv}[a]	ΔF_{bind}[a]	$\Delta\Delta F_{bind}$[b]
5678d	3.56 ± 0.24	1.28 ± 0.36	−2.28 ± 0.60
678d	2.43 ± 0.47	0.43 ± 0.68	−2.00 ± 1.15
578d	2.63 ± 0.23	1.02 ± 0.57	−1.61 ± 0.80
568d	3.14 ± 0.24	1.14 ± 0.36	−2.00 ± 0.60
78d	1.56 ± 0.34	−0.05 ± 0.63	−1.61 ± 0.97
68d	1.57 ± 0.32	−0.65 ± 0.55	−2.23 ± 0.87
58d	2.01 ± 0.24	1.18 ± 0.52	−0.83 ± 0.76
678p	−1.75 ± 0.26	−3.94 ± 0.63	−2.19 ± 0.89
78p	−2.82 ± 0.42	−3.87 ± 0.88	−1.05 ± 1.30
68p	−2.11 ± 0.29	−4.14 ± 0.68	−2.03 ± 0.97
8p	−4.08 ± 0.26	−4.01 ± 0.39	0.07 ± 0.65

[a] Mean of the free energies obtained for the various mutation pathways listed in Table 5 (see text for explanation of procedure for error estimation).

[b] $\Delta\Delta F_{bind} = \Delta F_{bind} - \Delta F_{solv}$ (Eq. 1). Errors are sums of the errors for ΔF_{bind} and ΔF_{solv}.

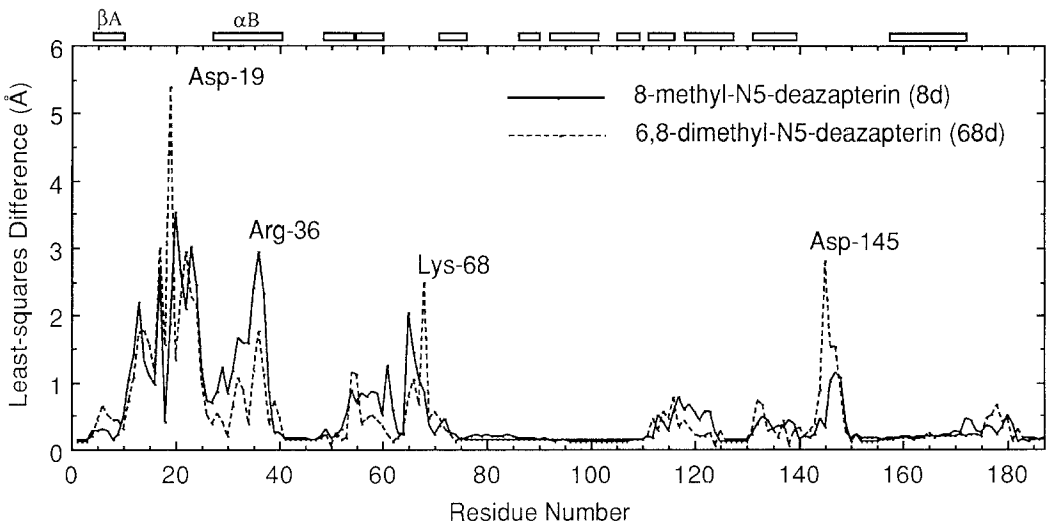

Fig. 4. Least-squares superposition of the DHFR C^{α} coordinates of two structures obtained from the MD simulations. The broken line traces the deviations (Å) between C^{α} positions for a case where the ligand bound to DHFR is 6,8-dimethyl-N5-deazapterin (68d) in both sets of coordinates involved in the fit (i.e. the coordinates from minimization 3 for the 568d → 68d calculation and the coordinates from minimization 2 for 68d → 8d). Similarly, the unbroken line is for a case where the ligand bound to DHFR is 8-methyl-N5-deazapterin (8d) in both sets of coordinates involved in the fit (i.e. the coordinates from minimization 3 for the 68d → 8d and 58d → 8d calculations). Bars marked βA, αB, etc. indicate portions of secondary protein structure (β-sheet and α-helix).

least-squares fit (from the 568d → 68d and 68d → 8d mutations), while the unbroken line is for two complexes where the bound ligand is 8d (from the 68d → 8d and 58d → 8d mutations).

Figures 5 and 6 give the distances between the C^{α} positions after least-squares superposition of MD structures and the initial X-ray clDHFR structure. In Fig. 5, the broken line traces the distances between the MD simulation and the X-ray clDHFR coordinates (i.e. coordinates from the clDHFR · NADPH · biopterin complex) where the bound ligand in the MD simulation is 68d, while the unbroken line traces the distances where the ligand in the MD simulation is 8d (both MD structures taken from the 68d → 8d mutation simulations). In Fig. 6, the coordinates of all 38 MD structures (minimizations 2 and 3 in Fig. 3) have been averaged and the resulting structure least-squares fitted to the C^{α} coordinates of the X-ray structure. The maximum and mean rms vibrational amplitudes of the C^{α} atoms for the set of 38 simulations are given as a function of residue number in Fig. 7, together with the rms deviations in the amplitudes. The active site of an 8-methyl-pterin (678p)-bound complex obtained after MD simulation (minimization 2 in Fig. 3) is shown in Fig. 8. The H-bonded ion-pair formed between the carboxylate group of Glu^{30} and extended-guanidinium nitrogens (N3 and N2) represents the only specific interaction between ligand and enzyme. The carbonyl oxygen, O4, of the ligand, the indole nitrogen of Trp^{24} and carboxylate oxygen of Glu^{30} participate in a weakly H-bonded network through two water molecules which were present in the original X-ray crystal structure [30]. Hydrophobic side chains of Leu^{22} make van der Waals (vdW) contacts at the N5 position of the pterin ring, while similar steric interactions between 7-methyl or 8-methyl substituents and the Val^{115} side chain are also possible. During the course of an FEP mutation, a small proportion (ca. 10%) of the MD

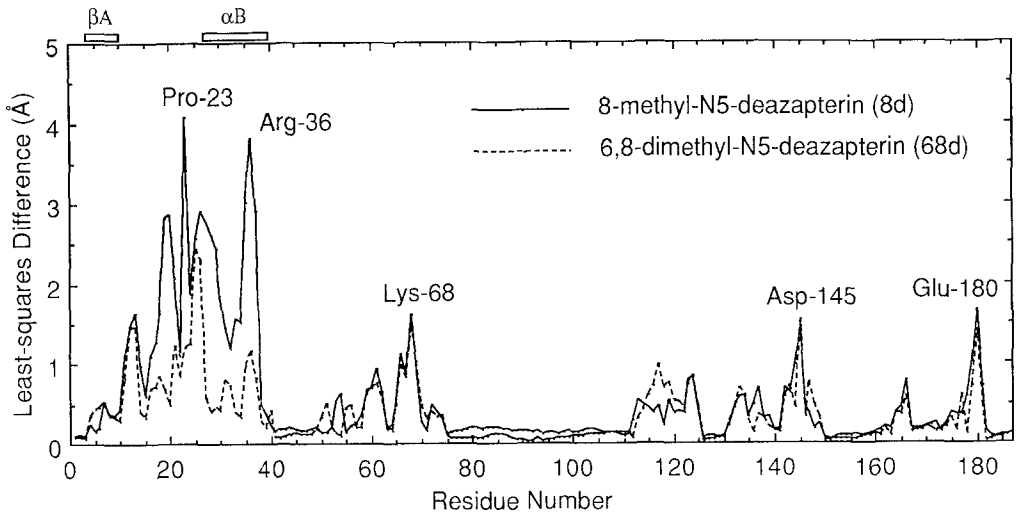

Fig. 5. Results of a least-squares superposition analysis showing the deviations (Å) between the DHFR C^{α} coordinates obtained from a typical MD simulation and X-ray crystallography. The ligands bound to DHFR in the MD simulation are 6,8-dimethyl-N5-deazapterin (broken line) and 8-methyl-N5-deazapterin (unbroken line), both from the 68d → 8d simulation (i.e. the coordinates obtained from minimizations 2 and 3 in Fig. 3). X-ray coordinates are from the clDHFR · $NADP^+$ · biopterin complex [30]. Bars marked βA, αB indicate portions of secondary protein structure (β-sheet and α-helix).

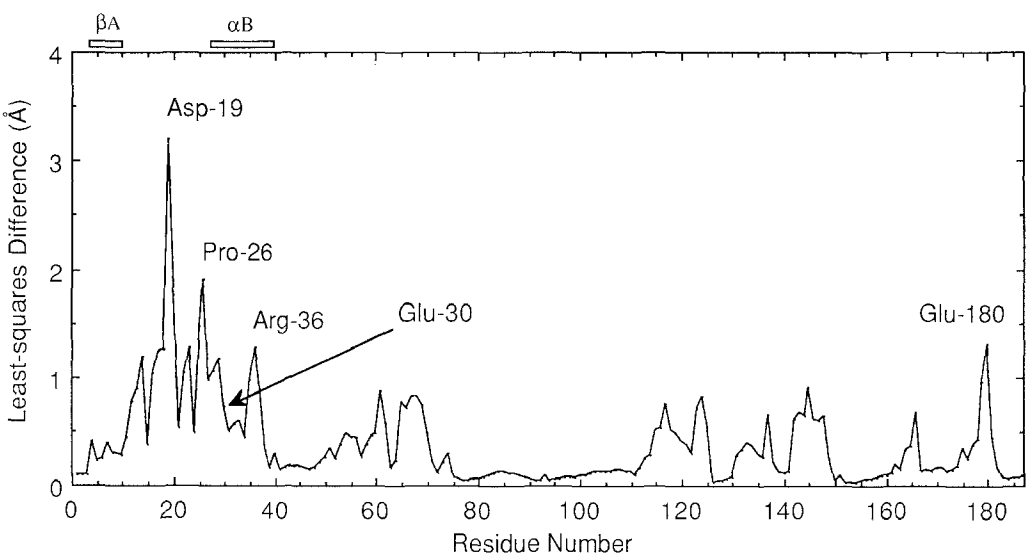

Fig. 6. Results from a least-squares superposition of DHFR C^{α} coordinates showing the deviations (Å) between the average of all 38 MD simulation structures (i.e. the average of the coordinates obtained from minimizations 2 and 3 for the 19 calculations performed according to the scheme in Fig. 3) and the X-ray crystal structure coordinates (i.e. coordinates from the clDHFR · $NADP^+$ · biopterin complex [30]). Bars marked βA, αB indicate portions of secondary protein structure (β-sheet and α-helix).

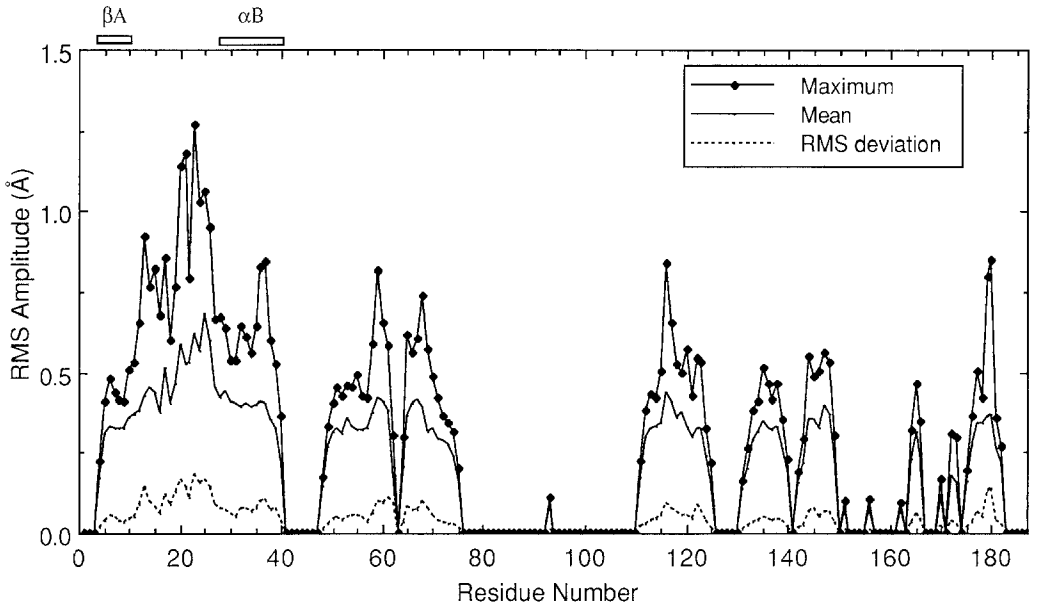

Fig. 7. Root-mean-square (rms) vibrational amplitudes obtained from the 38 MD simulations (NEWTON, Fig. 3) of clDHFR·NADPH·ligand complexes (the maximum, mean and standard deviation of amplitudes are given – see text). Bars marked βA, αB indicate portions of secondary protein structure (β-sheet and α-helix).

structures (in the 5678d → 678d, 5678d → 578d, 78d → 8d, 8d → 8p and 68p → 8p calculations) underwent significant nonreversible changes in the mode of ligand binding to DHFR. These configurational changes usually involved a rotational reorientation of the ligand and Phe^{34} side chain as shown in Fig. 9 for the 8p complex from the 8d → 8p mutation. Also, a Phe^{34} side chain rotation occurred during the 5678d → 678d mutation, but without any change in the ring orienta-

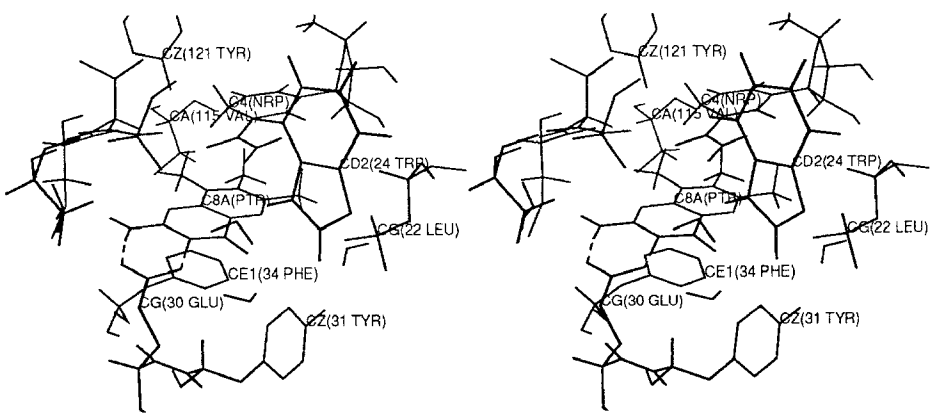

Fig. 8. Active-site structure of the clDHFR·NADPH·6,7,8-trimethylpterin complex given by the minimization 2 coordinate set (Fig. 3) from the 678p → 78p calculation. The H bond between the carboxylate side chain of Glu^{30} and the extended-guanidinium group of the ligand (PTR) is indicated by broken lines. Hydrogens on DHFR and cofactor (NRP) have been omitted for clarity.

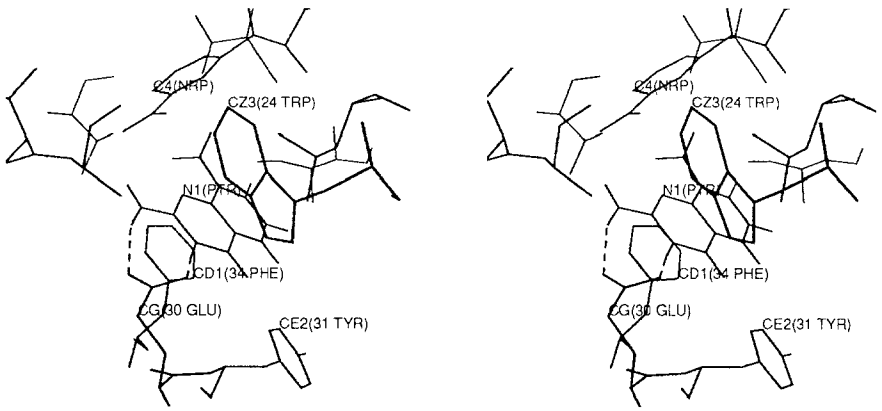

Fig. 9. Active-site structure of the clDHFR · NADPH · 8-methylpterin complex given by the coordinates of minimization 3 (Fig. 3) from the 8d → 8p calculation. Note the orientations of the pterin (PRT) and phenylalanine rings as compared with the normal orientation as shown in Fig. 8. Hydrogens on DHFR and cofactor (NRP) have been omitted for clarity.

tion of the ligand. In the 78d → 8d mutation the orientations are normal (i.e. like in Fig. 8), however, the normal H-bonding pattern between ligand and Glu^{30} has been disrupted and is mediated by H-bonding with water molecules, as shown for the 8d complex in Fig. 10.

DISCUSSION

Structures of ligand–protein complexes obtained by MD simulation

Before discussing results for the free energies computed by the FEP method, we shall refer to Figs. 4–10 and focus on some of the key structural features of the clDHFR · NADPH · ligand

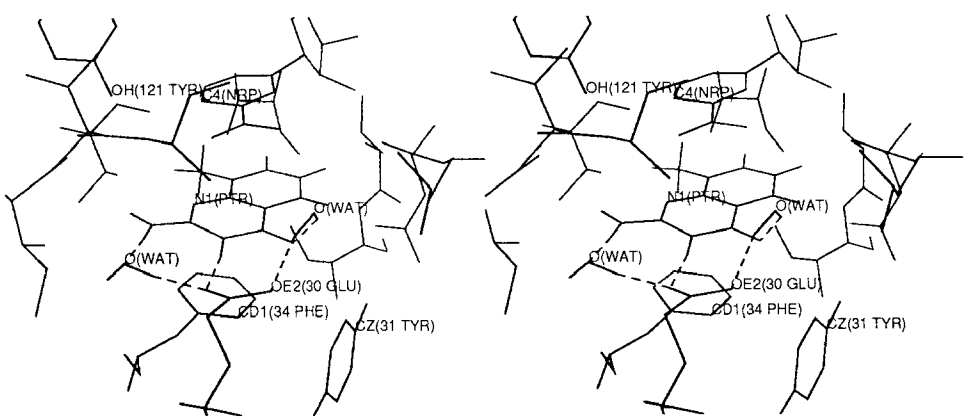

Fig. 10. Active-site structure of the clDHFR · NADPH · 8-methyl-N5-deazapterin complex given by the coordinates of minimization 3 (Fig. 3) from the 78d → 8d calculation. Dashed lines indicate a H-bonding pattern between ligand (PTR), protein and water which is different from the normal configuration shown in Fig. 8. Hydrogens on DHFR and cofactor (NRP) have been omitted for clarity.

complexes obtained from the MD simulations and the possible implications for the computed free energy changes.

The two examples shown in Fig. 4 illustrate structural variations in the protein backbone that arise when the simulations are carried out under slightly different starting conditions. For two simulations where the bound ligand is 68d, the backbones differ by as much as 5 Å (Asp^{19}). The largest differences generally occur in the solvent-exposed loops (residues 11–27, which are analogous to the mobile Met^{20} loop in the *E. coli* structure), whereas regions of the protein involved in secondary structures (βA and αB) tend to show much smaller differences. However, there are still quite large differences (e.g. Arg^{36} in 8d complexes) in the αB substructure. Note from Fig. 8, that the side chains of two of the loop residues, Leu^{22} and Trp^{24}, interact with the ligand directly via vdW contacts (Leu^{22}), or indirectly via H-bonding with crystallographic water molecules (Trp^{24}).

The largest differences between the MD and X-ray backbones (Fig. 5) occur in the loop structure comprising residues 11–27: for the simulation where the ligand bound to DHFR is 8d, the maximum deviation is as large as 4 Å (Pro^{23} and Arg^{36}). Generally, deviations in the positions of residues 11–40 are smaller for the simulation where the ligand is 68d. For both 8d and 68d, smaller deviations from the X-ray structure are observed for residues higher in the sequence (above 40) which becomes punctuated more frequently with nondynamic residues. Comparing Figs. 5 and 6, it is apparent that the overall deviation in the structure obtained by averaging over all 38 simulations (Fig. 6) is considerably less than that for a structure obtained from an individual simulation where the averages have been taken over 6 ps of dynamics: the only very significant deviations remain in Asp^{19} (3 Å) and Pro^{26} (2 Å). This reduction in the overall deviation from X-ray coordinates suggests that the MD structures are scattered more or less randomly about the X-ray structure. Consequently the coordinate averages taken over all simulations may give a better representation of the ensemble-average structure than the mean coordinates from a single MD simulation.

Information on the dynamics of the clDHFR molecule may be gleaned from the atomic rms vibrational amplitudes, i.e. the dynamic rms fluctuations from the mean configurations sampled by NEWTON over 6 ps (Fig. 7). The maximum amplitudes from the set of 38 simulations show considerable variation, particularly for the loop residues 11–27. Note that residues 11–27 which show the largest displacement from the X-ray structure (Figs. 5 and 6) also exhibit the largest rms amplitudes. On averaging the rms amplitudes for all 38 MD structures, a considerable smoothing of the dynamic fluctuations is obtained. The α-helix residues (28–40) have a uniform averaged rms fluctuation of about 0.4 Å, while the loop residues (11–27) have a peak rms fluctuation of about 0.7 Å.

The observation that considerable variations in mean geometries result from slightly different initial conditions (i.e. ligand binding and/or starting coordinates) would indicate that not all of the available configuration space is being sampled during the time-scale (6 ps) of a NEWTON simulation and hence also the perturbation simulations (Fig. 3). Thus any real differences between the various clDHFR·NADPH·ligand complexes would be unresolvable in the MD structures due to incomplete sampling. Note also that the accuracy of the computed free energy depends on the efficiency of configuration-space sampling. Another potential problem in the free energy calculations arises from changes in ligand-binding geometry. Figures 9 and 10 depict geometries which clearly differ from the normal or crystallographic-like geometry typified by the structure shown in Fig. 8. These changes appear only during the course of a mutation and are

therefore path dependent, i.e. artifacts of the FEP calculations; evidently the perturbation is sufficiently large to cause such changes. Clearly such path-dependencies may pose significant problems for the prediction of free energies, due to the large changes in the configuration space which lead to incorrect or unrealistic final configurational states. However, it is worth noting that by performing calculations for alternative mutations which do not lead to changes in binding geometries, it is possible to identify and avoid these particular sampling difficulties which could give rise to large errors in the free energies.

Computed free energies

The changes in the free energies of solvation (ΔF_{solv}) for the simulations (Fig. 2) were given in Table 3, together with the electrostatic (ΔF_{ele}) and vdW (ΔF_{vdw}) components. The vdW components for simulations involving the mutation of CH_3 to H are all negative, which is a predictable outcome on the basis of known hydration phenomena [45]. What cannot be predicted for the $CH_3 \rightarrow H$ mutation is the direction of change for the electrostatic component which in this case is also negative, thereby leading to negative total free energies of solvation. In contrast, the relatively large and negative ΔF_{ele} that is computed for mutations between the pterin and N5- deazapterin classes is expected on the basis of the change in polarity that results from mutating carbon (C5) to nitrogen (N5) in the heterocyclic ring (Fig. 1).

The proximity of the methyl substituent to the carbonyl oxygen (O4) appears to have some bearing on the pattern of ΔF_{solv}, since the free energy changes follow the order 58d < 68d < 78d (where the final ligand is 8d) and 678d < 578d < 568d (where the initial ligand is 5678d). This trend may be expected for the vdW component due simply to hydration effects involving the water structure surrounding the carbonyl group, i.e. an *R*-substituent would tend to partially shield the carbonyl from its polar interactions with the solvent. However, note that the purely electrostatic component (ΔF_{ele}) of the free energy also follows much the same trend.

The results for calculations of ΔF_{bind} given in Table 4 indicate small free energy changes (< 0.5 kcal/mol) for the mutation of a methyl group to hydrogen at the C6 position. On average, slightly larger free energy changes are obtained for mutations at the C7 position, while the largest changes are obtained for mutations involving a methyl group positioned at C5. From inspection of Fig. 8, most of the change at C5 probably involves steric (i.e. vdW) interactions between CH_3 of the ligand and the alkyl side chain of Leu^{22}. Note also that the free energy changes involving the mutation of a methyl group to hydrogen are generally smaller in magnitude than for the corresponding change in solution (Table 3). In contrast, the changes in ΔF_{bind} that result from a mutation from a pterin to N5-deazapterin are somewhat larger in magnitude than the corresponding ΔF_{solv} values, indicating that the active site is quite polar. The changes in ligand-binding geometry that were observed for some mutations (i.e. 5678d → 678d, 5678d → 578d, 78d → 8d, 8d → 8p, 68p → 8p) do not in general cause significant variations (less than about 0.5 kcal/mol) in the free energy changes, except for the reverse 78d → 8d mutation where water molecules have disrupted the H-bonding between ligand and Glu^{30} (Fig. 10) and resulted in a 1.7-kcal/mol free energy deviation from the forward mutation results.

Comparing the results obtained for different pathways (Table 5), the variation in both ΔF_{solv} and ΔF_{bind} for a given ligand is, in many instances, greater than 1.0 kcal/mol. These free energy variations may be largely the result of the variations in configuration-space sampling discussed above. Increasing the simulation times over which the free energies are computed may change free

energy differences by ca. 0.5 kcal/mol (678d → 678p mutation, see Results section). It is not clear, however, that increasing simulation times by this amount would have a significant effect on reducing the overall pathway variations in Table 5. Sampling appears to be somewhat more complete in the solvation simulations: the variations due to configuration-space sampling in the calculation of ΔF_{solv} are usually not more than about 0.5 kcal/mol, i.e. less than the corresponding variations in ΔF_{bind}.

Relative thermodynamic stabilities

Note that since the averaged structures appear to be more representative of the ensemble as given by the initial X-ray structure (compare Fig. 5 with Fig. 6), the free energies averaged over different pathways may also yield more reasonable a priori estimates of the ensemble average free energy, despite the relatively large statistical uncertainty indicated by the results in Table 6. It should be noted, however, that the average of thermodynamic properties such as the free energy, which depend on a series expansion of ensemble-averaged moments of the configurational energy, will most likely be much slower to converge; properties such as average structures would converge more rapidly [43].

The solution thermodynamics of these compounds has not been studied experimentally and consequently there are no experimental values for ΔF_{solv}. However, the general trend in the calculated ΔF_{solv} values indicates an expected hydrophobic hydration pattern, i.e. that the more methyl-substituted compounds are thermodynamically less stable in solution [45]. The results in Table 6 show that 5678d is the least stable in water, followed by 678d, 578d and 568d which are in turn less stable than 78d, 68d, and 58d. A similar trend is observed when comparing the stability of the methyl-substituted pterins relative to the parent compound 8p. Note that in solution the pterins are considerably more stable (about 3 kcal/mol) relative to the N5-deazapterins due to increased polarity of the N5-heterocyclic ring in the latter. The differences in $\Delta\Delta F_{bind}$ between similarly substituted pterins and deazapterins, e.g. 68d and 68p, are found to be quite small, usually much less than 1 kcal/mol. The results of kinetic experiments on the enzymic reduction of pterin substrates show that 68p has a higher substrate affinity (by about one order of magnitude or 1.5 kcal/mol in the K_m) than 8p, and that 68d is a better inhibitor than 8d [3,4,8,19]. These results are consistent with the higher affinities predicted by the FEP calculations for the binding of 68p and 68d compared with 8p or 8d. However, although negligible substrate activity has been observed for 678p, the computed free energy differences predict that it is comparable with the substrate 68p in its affinity for binding to clDHFR, i.e. the simulations clearly indicate that the equilibrium complexes involving 67p and 678p are thermodynamically stable. This is an interesting result, since significant overlap with the nicotinamide cofactor may be expected on the basis of a recently proposed transition-state model for folate reduction [50]. Thus the inactivity of the 7-methyl-substituted compounds may be due to steric factors in the step leading to hydride-ion transfer rather than being due to weak binding of the ligand.

CONCLUSION

Molecular dynamics (MD) and free energy perturbation (FEP) methods have been used to estimate the relative affinities for the binding of protonated 8-methyl-substituted pterin and N5-deazapterin ligands to dihydrofolate reductase (from chicken liver X-ray coordinates). The

combined effect of sampling errors in ΔF_{solv} and ΔF_{bind} (Eq. 1) amounts to uncertainties that may be larger than 1 kcal/mol (i.e. a factor of 10) in the relative thermodynamic stabilities of ligand binding ($\Delta\Delta F_{bind}$ in Eq. 1). Such errors make the accurate quantitative prediction of free energies difficult, particularly when differences of only one, two or three kcal/mol are involved. It seems clear from the results of this study that questions relating to the sampling of configuration space should be addressed. In the present calculations, the muations in solution were performed over about 120 ps, while a 40-ps simulation time was used for the ligand-binding mutations. Although run over a much shorter time than the solvation simulations, the binding simulations were also run for the reverse mutation so that the likely effects of incomplete sampling have been incorporated into our error estimates. Note that by evaluating free energy differences over many different pathways we have applied what should be an even more stringent test of sampling. It appears from the error estimates that much longer overall simulation times may be required to accurately resolve small free energy differences, as indicated also directly by the 80 ps mutation (678d $\rightarrow$ 678p) which was carried out in order to gauge the effect of increasing simulation time. However, it should be stressed that increasing simulation time alone may not adequately resolve free energy changes and some form of forced sampling technique [51] may be required.

Given constraints on computer time, the accuracy to which free energies can be calculated routinely is another issue that needs to be addressed. Due to the long computer time required to run a simulation for a fully solvated protein, it has become common practise [22–27] to perform only the mutations in solvent under NTP conditions (with periodic boundary conditions), while using the active-site approach for protein simulations. Boundary effects would be practically eliminated by solvating the protein and constructing the NTP (or NTV) ensembles consistently for each of the simulations. It may be necessary to perform calculations on more fully solvated protein systems, at additional computational expense, or to explore the use of more computationally efficient models which take into account the effects of solvent on electrostatic interactions and protein dynamics [52,53]. There are also other possible sources of error, such as the force fields, and the assumed binding geometry. In many instances the electrostatic contribution to $\Delta\Delta F_{bind}$ is large and may be affected by the charge distribution of the ligand. The ligands' atomic charges were computed from the AM1 electrostatic potential and it is not certain how the free energy differences would change if ab initio potentials were used. In addition, although there is certainly strong indirect evidence [3–8] for the binding geometries we have used, these have not yet been verified by X-ray crystallography. All of these factors could have a significant bearing on the accuracy of the results, and deserve careful scrutiny in future applications of the FEP methodology.

Despite the uncertainties outlined above, we may tentatively draw the following conclusions from the present study. Within the error limits of the computed free energy differences, the general trends indicate that, relative to the parent 8-methyl-N5-deazapterin (8d) or 8-methylpterin (8p), the compounds methyl substituted in the 5, 6 or 7 positions are thermodynamically more stable when bound to DHFR largely by virtue of a hydrophobic effect, i.e. methyl substitution reduces the affinity of the ligand for the solvent more than it reduces affinity for the DHFR active-site. The stability of ligand-binding ($\Delta\Delta F_{bind}$) to DHFR appears to be optimal with a 6-methyl substituent: additional 5-methyl and/or 7-methyl substitution has little effect on the strength of binding. This saturating effect is most likely due to the fact that the methyl substituents are in close proximity to one another. Thus the first hydrophobic substituent effectively produces a solvent cage in its vicinity [45] which allows adjacent methyl groups to be accommo-

dated with a much smaller free energy change. Also, with regard to the stability of ligand binding, the FEP calculations indicate only minor differences between similarly substituted pterins and N5-deazapterins, e.g. the DHFR·ligand binding constants of 6,8-dimethylpterin (68p) and 6,8-dimethyl-N5-deazapterin (68d) are predicted to be of the same order of magnitude.

The results are consistent with what is known experimentally for the substrates 8p and 68p and for the inhibitors 8d and 68d, in that the 6-substituted compound has been found to be the more tightly bound substrate or inhibitor. It is encouraging that we have found these trends to be in qualitative agreement with the available experimental data. Such trends, if they can be predicted reliably by computer-based methods, are clearly of considerable practical benefit in drug design, both in terms of screening derivatives of compounds for later synthesis and for gaining a better understanding of the binding processes and desolvation effects which influence the thermodynamic stabilities of ligand–protein complexes.

ACKNOWLEDGEMENTS

Financial support from the National Health and Medical Research Council (NH&MRC), and equipment grants from the Clive and Vera Ramaciotti and Utah Foundations are gratefully acknowledged. The NH&MRC provided funds for a laboratory-based IBM RS/6000 320 model workstation and we also acknowledge the University of Sydney Computing Service for providing time on their IBM RS/6000 530 and 540 machines. We thank Professor J. Kraut for the DHFR X-ray coordinates and Dr. U.C. Singh for the AMBER code.

REFERENCES

1 Blakley, R.L., In Blakley, R.L. and Benkovic, S.J. (Eds.) Folates and Pterins – Chemistry and Biochemistry of Folates, Vol. 1, Wiley, New York, 1984, pp. 191–253.
2 Kraut, J. and Matthews, D.A., In Jurnak, F.A. and McPherson, A. (Eds.) Biological Macromolecules and Assemblies, Vol. 3, Wiley, New York, 1987, pp. 1–72.
3 Thibault, V., Koen, M.J. and Gready, J.E., Biochemistry 28 (1989) 6042.
4 Gready, J.E., In Curtius, H.-Ch., Ghisla, S. and Blau, N. (Eds.) Chemistry and Biology of Pteridines 1989, de Gruyter, Berlin, 1990, pp. 23–30.
5 Koen, M.J., Haynes, R.K., Gready, J.E. and Pilling, P.A., In Curtius, H.-Ch., Ghisla, S. and Blau, N. (Eds.) Chemistry and Biology of Pteridines 1989, de Gruyter, Berlin, 1990, pp. 94–97.
6 Cummins, P.L. and Gready, J.E., Proteins: Struct. Funct. Genet., in press.
7 Jeong, S.S., Wormell, P. and Gready, J.E., In Blau, N., Curtius, H.-Ch., Levine, R. and Yim, J. (Eds.) Pteridines and Related Biogenic Amines and Folates, Hanrim, Seoul, 1992, pp. 277–292.
8 Gready, J.E., Ivery, M.T.G., Koen, M.J. and Yang, H.J., In Blau, N., Curtius, H.-Ch., Levine, R. and Yim, J. (Eds.) Pteridines and Related Biogenic Amines and Folates, Hanrim, Seoul, 1992, pp. 265–276.
9 Koen, M.J. and Gready, J.E., J. Org. Chem., 58 (1993) 1104.
10 Gready, J.E., J. Comput. Chem., 6 (1985) 377.
11 Gready, J.E., Biochemistry 24 (1985) 4761.
12 Williams, M.L. and Gready, J.E., J. Comput. Chem., 10 (1989) 35.
13 Jordan, M.J. and Gready, J.E., J. Comput. Chem., 10 (1989) 186.
14 Brown, D.J. and Jacobsen, N.W., J. Chem. Soc., (1961) 4413.
15 Pfleiderer, W., In Blakley, R.L. and Benkovic, S.J. (Eds.) Folates and Pterins – Chemistry and Biochemistry of Pterins, Vol. 2, Wiley, New York, 1985, pp. 43–114.
16 Gready, J.E., to be published.

17 Stone, S.R., Montgomery, J.A. and Morrison, J.F., Biochem. Pharmacol., 33 (1984) 175.
18 Davies, J.F., Delcamp, T.J., Prendergast, N.J., Ashford, V.A., Freisheim, J.H. and Kraut, J., Biochemistry, 29 (1990) 9467.
19 Ivery, M.T.G. and Gready, J.E., In Ayling, J.E., Nair, M.G. and Baugh, C.M. (Eds.) Chemistry and Biology of Pteridines and Folates, Plenum Press, New York, 1993, in press.
20 Tembe, B.L. and McCammon, J.A., Comput. Chem., 8 (1984) 281.
21 Singh, U.C., Brown, F.K., Bash, P.A. and Kollman, P.A., J. Am. Chem. Soc., 109 (1987) 1607.
22 Singh, U.C., Proc. Natl. Acad. Sci. USA, 85 (1988) 4280.
23 Singh, U.C. and Benkovic, S.J., Proc. Natl. Acad. Sci. USA, 85 (1988) 9519.
24 Brooks, C.L., Int. J. Quantum Chem. Quantum Biol. Symp., 15 (1988) 221.
25 Brooks, C.L, and Fleischman, S.H., J. Am. Chem. Soc., 112 (1990) 3307.
26 Fleischman, S.H. and Brooks, C.L., Proteins: Struct. Funct. Genet., 7 (1990) 52.
27 McDonald, J.J. and Brooks, C.L., J. Am. Chem. Soc., 113 (1991) 2295.
28 Volz, K.W., Matthews, D.A., Alden, R.A., Freer, S.T., Hansch, C., Kaufman, B.T. and Kraut, J., J. Biol. Chem., 257 (1982) 2528.
29 Matthews, D.A., Bolin, T.J., Burridge, J.M., Filman, D.J., Volz, K.N., Kaufman, B.T., Beddell, C.R., Champness, J.N., Stammers, D.K. and Kraut, J., J. Biol. Chem., 260 (1985) 381.
30 a. Matthews, D.A., Oatley, S.J. and Kraut, J., unpublished results for clDHFR · NADP+ · biopterin complex.
30 b. McTigue, M.A., Davies II, J.F., Kaufman, B.T. and Kraut, J., Biochemistry, 31 (1992) 7264.
31 Van Gunsteren, W.F., Prot. Eng., 2 (1988) 5.
32 Van Gunsteren, W.F., In van Gunsteren, W.F. and Weiner, P.K. (Eds.) Computer Simulation of Biomolecular Systems, ESCOM, Leiden, 1989, pp. 27–59.
33 Pettitt, B.M., In van Gunsteren, W.F. and Weiner, P.K. (Eds.) Computer Simulation of Biomolecular Systems, ESCOM, Leiden, 1989, pp. 94–100.
34 Pearlman, D.A. and Kollman, P.A., In van Gunsteren, W.F. and Weiner, P.K. (Eds.) Computer Simulation of Biomolecular Systems, ESCOM, Leiden, 1989, pp. 101–119.
35 Pearlman, D.A. and Kollman, P.A., J. Am. Chem. Soc., 113 (1991) 7167.
36 Weiner, S.J., Kollman, P.A., Nguyen, D.T. and Case, D.A., J. Comput. Chem., 7 (1986) 230.
37 Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P.K., J. Am. Chem. Soc., 106 (1984) 765.
38 Cummins, P.L., Ramnarayan, K., Singh, U.C. and Gready, J.E., J. Am. Chem. Soc., 113 (1991) 8247.
39 Jorgensen, W.L., Chandrasekhar, J., Madura, J.D., Impey, R.W. and Klein, M.L., J. Chem. Phys., 79 (1983) 926.
40 Dewar, M.J.S., Zoebisch, E.G., Healy, E.F. and Stewart, J.J.P., J. Am. Chem. Soc., 107 (1985) 3902.
41 Cummins, P.L. and Gready, J.E., Chem. Phys. Lett., 174 (1990) 355.
42 Hagler, A.T., Maple, J.R., Thacher, T.S., Fitzgerald, G.B. and Dinur, U., In van Gunsteren, W.F. and Weiner, P.K. (Eds.) Computer Simulation of Biomolecular Systems, ESCOM, Leiden, 1989, pp. 149–167.
43 Mezei, M. and Beveridge, D.L., Ann. N.Y. Acad. Sci., 482 (1986) 1.
44 Singh, U.C., Weiner, P.K., Caldwell, J.W. and Kollman, P.A., AMBER (Version 3.2), Department of Pharmaceutical Chemistry, University of California, San Francisco, CA, 1988.
45 Rao, B.G. and Singh, U.C., J. Am. Chem. Soc., 111 (1989) 3125.
46 Van Gunsteren, W.F. and Berendsen, H.J.C., Mol. Phys., 34 (1977) 1311.
47 Brooks, C.L., Brunger, A. and Karplus, M., Biopolymers, 24 (1985) 843.
48 This version of NEWTON has been locally modified to perform direct averaging of coordinates.
49 Ferrin, T., Huang, C., Pettersen, E.F., Couch, G.S. and Jarvis, L., MidasPlus program, Computer Graphics Laboratory, University of California, San Francisco, CA, 1989.
50 Bystroff, C., Oatley, S.J. and Kraut, J., Biochemistry, 29 (1990) 3263.
51 Mezei, M., J. Comput. Phys., 68 (1987) 237.
52 Solmajer, T. and Mehler, E.L., Prot. Eng., 4 (1991) 911.
53 Solmajer, T. and Mehler, E.L., Int. J. Quantum Chem., 44 (1992) 291.