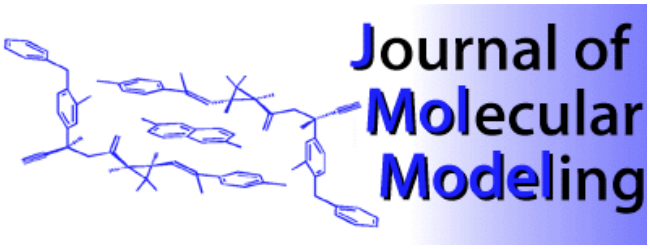
*J. Mol. Model.* **1999**, *5*, 153 – 168




**FULL PAPER**



# Molecular Dynamics Simulations of L-Lactate Dehydrogenase: Conformation of a Mobile Loop and Influence of the Tetrameric Protein Environment

**Rebecca K. Schmidt and Jill E. Gready**

Division of Biochemistry and Molecular Biology, John Curtin School of Medical Research, Australian National University, GPO Box 334, Canberra ACT 2601 Australia. Tel: 61 2 6279 8304; Fax: 61 2 6249 0415; E-mail: Jill.Gready@anu.edu.au

*Received: 15 March 1999/ Accepted: 17 June 1999/ Published: 9 September 1999*

**Abstract** L-lactate dehydrogenase (LDH) catalyses the interconversion of pyruvate and L-lactate in the presence of the coenzyme NADH. Molecular dynamics (MD) simulations have been performed for LDH complexed with NADH and the pyruvate-analogue inhibitor oxamate with the aim of characterising important influences on maintaining the geometry and hydrogen bond network of the active site. Two features in particular were found to dominate. First, the tetrameric protein environment is found to play a significant role in maintaining the active-site geometry. Simulations of the monomer alone reproduce the crystallographic structure poorly, and at least part of the neighboring subunit is necessary to prevent water penetration into the active site and to provide rigidity to the α1G-α2G helix immediately adjacent to the active site. These results offer one explanation for the observation that the monomer is not biologically active. Second, the conformation of Arg109 (part of the mobile loop which closes over the active site) is shown to play a key role in maintaining the active-site geometry. In some simulations, a torsional rotation in the side chain of Arg109 results in the breaking of crystallographic hydrogen bonds which are important for polarising the carbonyl bond of the substrate. This conformational change appears to be a trigger for the opening of the mobile loop. Long-range nonbonded interactions are found to be influential in maintaining the proper crystallographic conformation of Arg109. Thus, we conclude that to adequately model LDH, at least part of the neighboring subunit must be included in the MD simulations and nonbonded interactions must be properly represented to ensure that Arg109 remains in the crystallographic conformation. Out of a set of simulation protocols tested here, one meets both these criteria and will be used for the generation of starting structures for future reaction-mechanism calculations.

**Keywords** L-lactate dehydrogenase, Molecular dynamics, Loop, Hydrogen bond, Protein electrostatics

*Correspondence to:* J.E. Gready

## Introduction

L-Lactate dehydrogenase (LDH, EC 1.1.1.27) reversibly interconverts pyruvate and L-lactate in the presence of the coenzyme NADH. Typical kinetic constants for the human enzyme [1] (kcat=350 $s^{-1}$ and $K_M$=0.08 mM at pH=6) indicate a rate acceleration of $10^{14}$ compared with the analogous reaction in solution [2]. The enzymic reaction is achieved by the transfer of a proton from His195 and a hydride ion from NADH (residue numbering as defined by Eventoff et al. [3]). Interestingly, these chemical steps are not rate-limiting in the overall reaction, but instead the slowest step is a conformational change that must occur before the reaction proceeds. This rate-limiting conformational change has been deduced from experiment by the lack of an isotope effect [1] as well as the decreased reaction rate observed in more viscous solutions [4]. Crystallographic comparisons of the apo and ternary complexes [5] reveal that the conformational change is the motion of a loop (residues 98-113, as shown in Figure 1) which closes over the active site. From tryptophan fluorescence experiments [6], this loop motion has been shown to occur on the same time scale as the overall enzymic rate. Both computational [7] and crystallographic [5] studies show that the motion of the loop can be described as a displacement of a rigid flap (residues 103-107) via flexible "hinge" residues on either side of the flap.

After the loop closes over the bound substrate and coenzyme, the active-site configuration required for catalysis is completely formed (Figure 2). The loop contributes two residues (Gln102 and Arg109) which interact directly with the substrate. Gln102 hydrogen bonds to the inhibitor oxamate in the crystal structure. (Note that this interaction is not possible for the true substrates pyruvate or L-lactate, in which the amino group is replaced by a methyl group.) This amino acid has been postulated to differentiate among substrate side chains, and has been the target for site-directed mutagenesis studies which aimed to vary the specificity of the enzyme [8,9]. Arg109 on the mobile loop also forms hydrogen bonds with the carbonyl oxygen (O1) and one carboxyl oxygen (O3) of the substrate. This interaction has been postulated to polarise the carbonyl bond and, thus, speed the reaction. Spectroscopic measurements show that upon binding to the enzyme, the substrate C=O vibrational frequency decreases, corresponding to an increase in the single bond character of this carbonyl bond, and, thus, its polarisation [10]. This polarisation increases the positive charge on the carbon and the negative charge on the oxygen ($C^+$-$O^-$), facilitating hydride and proton transfer, respectively.

Other residues in the active site which are not part of the loop also interact with the substrate. Arg171 anchors the substrate via a strong interaction with the carboxyl group [11,12], positioning the substrate to accept a proton and hydride ion from His195 and NADH, respectively. Thr246 also forms a

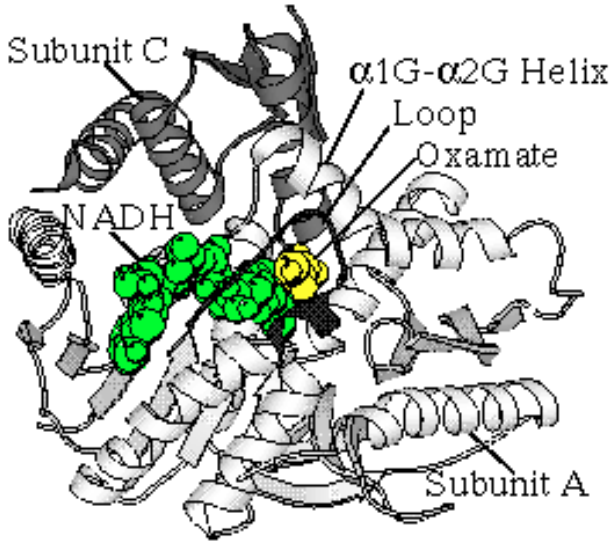



**Figure 1** *Molscript diagram of the crystallographic structure of the ternary complex of LDH from dogfish muscle (PDB entry 1LDM). Subunit A (excluding the 18 residue N-terminus) is shown in light grey with the flexible loop (residues 98-113), shown in black, closing over the active site bound with the coenzyme NADH and the inhibitor oxamate (in space filling representations). Also shown is the portion of Subunit C (medium grey) which is included in the spherical calculations (see Methods)*

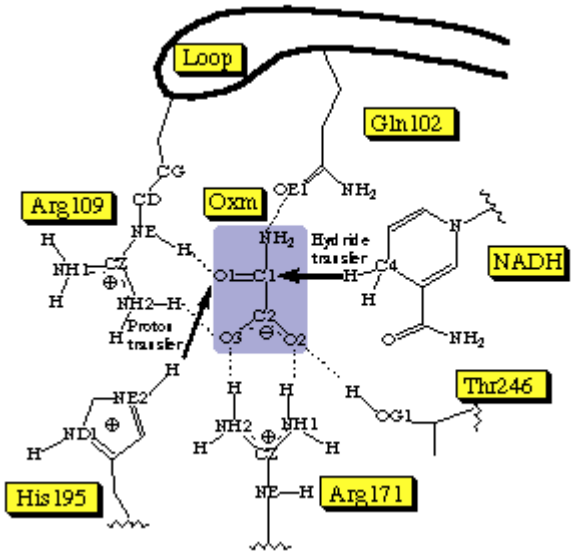



**Figure 2** *Schematic diagram of the active site of LDH showing important amino acid residues and the inhibitor oxamate (Oxm), with significant atoms labelled. Hydrogen bonds found in the crystallographic structure are indicated by dashed lines, while the two arrows indicate the path of the proton transfer from His195 and the hydride-ion transfer from NADH. Note that Arg109 and Gln102 are part of the flexible loop which closes over the active site.*

hydrogen bond to the substrate, although site-directed mutagenesis studies indicate that this hydrogen bond is less crucial than the activation of the dihydronicotinamide ring of the cofactor by the Thr246 hydroxyl group [13]. Many ordered waters are also observed crystallographically in the active site near Glu194, Arg109, Asp168, and Asn140. In the crystallographic structure of the apo enzyme (i.e. without substrate or NADH), the mobile loop is open and fewer ordered waters are observed in the active site, implying that one of the functions of the closed loop is to exclude mobile water from the active site.

While crystallographic and mutagenesis studies have provided a wealth of information on the catalytic mechanism and loop behaviour of this enzyme [14,15], theoretical approaches can provide supplementary information at an atomic level of detail that is not accessible via experiments. A range of theoretical techniques have been used to study LDH, including quantum mechanical calculations of isolated substrates [16,17] or fragments of key species in the active site [18-21], as well as empirical calculations of the electrostatic energy [22]. Several molecular dynamics (MD) studies have appeared [23-25], with one particularly interesting high-temperature study which characterised the motion of the loop [7].

The ultimate goal of this current work is to continue previous characterisation of the enzymic reaction using hybrid quantum mechanical/molecular mechanical (QM/MM) techniques [26]. Before performing QM/MM simulations, however, it is necessary to perform MD simulations of the enzymic system for several reasons. First, these simulations are required to generate relaxed starting coordinates for QM/MM studies. Second, MD simulations are less computationally costly than QM/MM calculations, and so the simulation protocol for the portion of the system which will be treated using molecular mechanics in the QM/MM studies can be optimised economically. Key to optimising the protocol for the MD simulations is characterising important influences on maintaining the geometry of the active site and selecting simulation conditions which adequately represent these influences.

Of central importance in the MM region is determining what proportion of the enzyme tetramer is required to model the system accurately. Normally, the enzyme is active only as a tetramer, although in *Bacillus stearothermophilus* the enzyme is in equilibrium between the dimer and tetramer. Historically, MD studies have tended to focus on smaller monomeric proteins in order to minimise computational time, and for this same reason some previous studies of LDH have included only one monomer of the full tetramer [25]. While MD simulations of the full tetramer of LDH might be feasible computationally, applying QM/MM methods to such a large model may not be necessary and, thus, would be an inefficient use of computer time. As each active site is independent, it may be possible to model the enzyme using a small portion of the tetramer, but the effects of such truncation are unknown. Here, simulations of one monomer are compared with those using either the dimer, tetramer, or a sphere of protein residues within 24 Å of the substrate. Monomeric simulations are found to reproduce the crystallographic structure poorly, and the problems with these simulations offer insight into the lack of biological activity for the monomer.

A second feature of the enzyme which is crucial to reproduce in the simulations is the conformation of the mobile loop which closes over the active site. In particular, one residue from the mobile loop (Arg109) must be properly oriented to polarise the substrate. Not all simulations, however, maintain the crystallographic orientation of this loop due to conformational changes in Arg109. This conformational change, and the hydrogen bond exchange which results, is similar to that found in a high-temperature MD simulation in which the loop starts to open [7]. By varying the simulation conditions (such as the dielectric constant, size of the water cap, and the use of counterions), we show that the conformation of the flexible loop is strongly affected by long-range nonbonded interactions, and only when they are adequately balanced does the proper closed conformation result.

## Computational details

A series of MD simulations was first performed for one monomer of the normally tetrameric LDH bound with the inhibitor oxamate and the coenzyme NADH. In oxamate, an amino group replaces the methyl group found in the biological substrate pyruvate. The inhibitor was used so that direct comparisons with the crystal structure could be made, without any ambiguities arising from differences in protein structure resulting from the binding of pyruvate or L-lactate instead of oxamate. The starting coordinates for the monomer simulations were obtained from the 2.1 Å crystal structure of the dogfish muscle enzyme [27] (Brookhaven Protein Data Bank code 1LDM). As the first 18 residues of the N-terminus of subunit A contact only subunit D (Figure 3), these residues were not included in the monomer calculations. Based on evidence that the reaction occurs optimally at pH=6 [1,28] and assuming typical $pK_a$ values for amino acids in proteins [29], Asp and Glu were not protonated while His, Lys, and Arg were protonated.

Two oxamate molecules are observed in each monomer in the crystallographic structure; only the one found in the active site was included in the monomer calculations. Each monomer of the crystal structure also includes 245 ordered water molecules. Of these crystallographic waters, those found in the interior of the protein were included in the computational model, comprising sixteen ordered waters in the active site as well as three far from the active site, i.e., a total of 19 out of the 245 crystallographic waters. The remaining crystal waters are on the surface of the protein and were not included except for those which fall under the spherical water cap which is overlaid on the protein to solvate the active site. This solvent cap has a radius of either 24 or 28 Å from oxamate, the smaller of which is sufficient to cover the exterior of the mobile loop with several layers of water. Waters in the cap

**Table 1** *Key to Simulation Conditions*

| | Protein [a] | Water cap radius (Å) [b] | Counter ions [c] | Dielectric constant [d] |
|---|---|---|---|---|
| Mono24_1 | Monomer | 24 | | 1.0 |
| Mono24_c1 | Monomer | 24 | * | 1.0 |
| Mono24_1r | Monomer | 24 | | $1.0*r_{ij}$ |
| Mono28_1 | Monomer | 28 | | 1.0 |
| Mono28_c1 | Monomer | 28 | * | 1.0 |
| Mono28_1r | Monomer | 28 | | $1.0*r_{ij}$ |
| Sph24_1 | Sphere | 24 | | 1.0 |
| Dim24_1 | Dimer | 24 | | 1.0 |
| Tetra24_1 | Tetramer | 24 | | 1.0 |
| Sph24_1r | Sphere | 24 | | $1.0*r_{ij}$ |
| Dim24_1r | Dimer | 24 | | $1.0*r_{ij}$ |
| Tetra24_1r | Tetramer | 24 | | $1.0*r_{ij}$ |

*[a] The extent of the normally tetrameric protein included in the calculation. Monomer includes only subunit A; Sphere includes subunit A plus portions of subunit C to result in a sphere of protein residues of approximately 24 Å radius around the oxamate (see Figure 1); Dimer includes subunits A and C; and Tetramer includes all four subunits. The N-terminal tails are included only in the tetramer.*

*[b] The radius of the solvating water cap centred on the oxamate.*

*[c] * indicates the presence of counterions on the exterior of the protein not covered by the water cap.*

*[d] A linear dielectric constant is indicated by 1.0 while a distance-dependent dielectric constant is indicated by $1.0*r_{ij}$.*

are prevented from evaporating via a harmonic restraining potential (k=1.5 kcal/mol·Å$^2$).

All simulations were performed using the AMBER 4.1 program [30] using the Cornell all-atom protein force field [31] and TIP3P water [32]. The parameters for the adenine nucleotide portion of NADH were adapted from D-adenosine in the Cornell force field. Parameters and charges for the nicotinamide portion of NADH have been described previously [33]. The oxamate charges were determined at the SCF/6-31+G* basis set level, with bond, angle, and torsion parameters adapted from the standard protein force field.

Six simulations were performed for the monomer system using a variety of protocols (Table 1). The effect of the size of the water cap was explored by running simulations with a cap radius of either 24 or 28 Å. The effect of the dielectric constant used in the calculation of the electrostatic energy $E_{elec}$ was determined by running simulations with either a linear dielectric constant ($\varepsilon$=1) or a distance-dependent dielectric constant ($\varepsilon=1*r_{ij}$) in the equation

$$E_{elec} = \sum_{i<j} \frac{q_i q_j}{\varepsilon r_{ij}} \quad (1)$$

where $r_{ij}$ is the distance between two charges $q_i$ and $q_j$. (While the use of $\varepsilon=1*r_{ij}$ is not formally correct with the Cornell force field, these simulations are useful in gauging the influence of long-range interactions, and in particular the possible adverse effect of the exterior of the protein which is not shielded by solvent in these calculations.) In two simulations, counterions were added to the exterior of the protein not enclosed by the water cap using the CION utility in AMBER 4.1. Counterions were placed wherever the absolute value of the electrostatic potential (calculated using a constant dielectric $\varepsilon$=1) was greater than 0.5e. This procedure resulted in the addition of twelve chloride ions to the protein so that the final total charge of the monomer was +2.

In addition to the monomer calculations, two series of simulations for larger systems were performed. One series used a dielectric constant of $\varepsilon$=1, and the second used $\varepsilon=1*r_{ij}$. In each series, increasing proportions of the tetramer are included. In the calculations abbreviated as “sphere”, all protein residues within roughly 24 Å from the oxamate are included. Thus, in addition to subunit A, residues 22-81 and 241-263 of subunit C were included. The same crystallographic and cap waters as in the monomer simulations were included (except for those waters which overlapped with the additional protein), as well as the oxamate and NADH in the one active site. The dimer calculations included both subunits A and C, with the exception of the 18 residue N-termini, which are more closely associated with subunits D and B, respectively (Figure 3). Both subunits in the dimer calculations also included: the active-site oxamate and NADH; all interior crystallographic waters included in the monomer calculations; crystallographic waters found in the interface between the A and C subunits; and a 24 Å water cap centred on the oxamate of subunit A. Finally, the tetramer calculation included all four subunits in their entirety, including all four N-termini as well as two oxamates and one NADH for each

subunit. All interior crystallographic waters used in the monomer calculation were included for each subunit, as well as any crystallographic waters in the interfaces between the subunits and a 24 Å water cap surrounding the active-site oxamate of subunit A.

All simulations were equilibrated in the same manner. This minimisation sequence was designed to first minimise the water coordinates while holding protein, oxamate(s) and coenzyme(s) fixed, so that strong forces due to badly placed waters do not distort the protein structure. First, the positions of the water hydrogens were minimised (while the oxygen atoms remained fixed) for 100 steps using the steepest descent method followed by the conjugate gradient method until the the norm of the gradient of the energy was below 0.5 kcal/mol·Å. The positions of all water atoms were then minimised to a gradient of 0.1 kcal/mol·Å. Finally, 10 ps of MD were performed with only the water allowed to move.

Once the water structure was optimised around the crystallographic protein structure, the entire system was slowly minimised to prevent distortions due to the release of possible strains in the crystallographic structure. Initially, the positions of hydrogens alone were minimised to a gradient of 0.5 kcal/mol·Å. Next, the protein backbone was held fixed while the remainder of the system was minimised to 0.5 kcal/mol·Å while under the influence of a strong restraining potential to the crystal structure (k=100 kcal/mol·Å$^2$). The backbone was then allowed to move, and a series of minimisations with decreasing restraints to the crystal structure (force constants: k=100, 50, 15, and 2 kcal/mol·Å$^2$) were performed to a gradient of 0.1 kcal/mol Å. Finally, all restraints were released, and the system was fully minimised to 0.1 kcal/mol·Å.

The minimised structures were then used as starting coordinates for MD simulations. Only those residues which were within approximately 17 Å of the oxamate were allowed to move. For the monomer simulations, this mobile belly consisted of residues 36-30, 52-55, 94-120, 135-149, 160-178, 189-203, 225-263, 269-275, and 285-291 which includes the mobile loop which covers the active site. In the multimer simulations (sphere, dimer or tetramer), the mobile belly also included residues 55-72 of the C subunit. The simulations were run at a constant temperature of 300°K with the time constant for the coupling of the heat bath to both the solvent and the remainder of the system equal to 0.2 ps. All bonds to hydrogens were fixed using the SHAKE algorithm, which allows a timestep of 0.001 ps. Each simulation was run for 310 ps with coordinates saved every 25 timesteps. The cutoff for nonbonded interactions was 14 Å and the nonbonded cutoff list was updated every 25 timesteps.

RMS deviations from crystallographic coordinates were calculated using the AMBER 4.1 utility CARNAL after first removing all rotations and translations of the fixed portion of the protein (i.e. those residues *not* in the mobile belly) with respect to the crystallographic structure. A typical time history of the root mean square (RMS) deviation of the heavy atoms of the mobile belly for a monomer simulation is shown in Figure 4. The RMS deviation initially rises until a plateau is reached at approximately 110 ps, and thereafter small fluctuations (0.13 Å) are observed around a mean of 1.13 Å. As all simulations reached such a plateau within 110 ps, in all cases this first portion was discarded as the equilibration period and the remaining 200 ps was analysed as the data collection period.

CARNAL was also used to calculate selected distances, torsions, and hydrogen bond occupancies. Hydrogen bond occupancy is defined as the percentage of the simulation in which two species are hydrogen bonded according to conventional geometric criteria (donor-acceptor distance < 3.5 Å and donor-hydrogen-acceptor angle > 120°). Interaction energies between different parts of the system were calculated using the AMBER 4.1 utility ANAL employing the same force field parameters and variables as for the simulations.

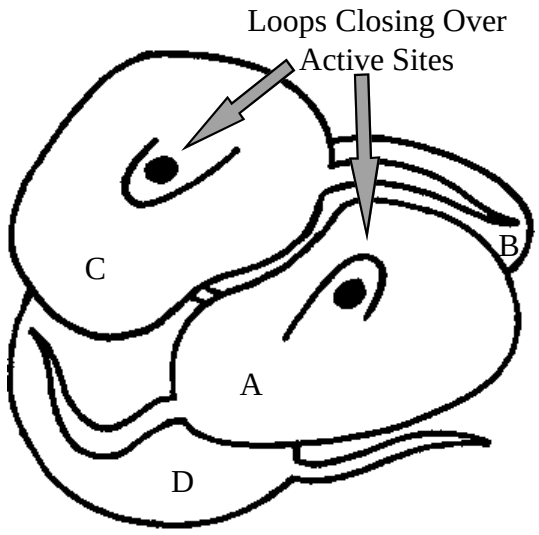



**Figure 3** *Schematic diagram of the entire tetrameric complex, in the same orientation as for Figure 1. The active sites and loops for subunits A and C are shown, as well as the N-terminal tails of subunits A, C, and D, which interact with subunits D, B, and A, respectively, to stabilise the tetramer. The N-terminal tail of subunit B is not visible.*

## Results

### *Monomer calculations*

**General observations** The average RMS deviations for the data collection period for each of the monomer simulations are shown in Table 2. The heavy atoms of the mobile belly in the protein reproduce the arrangement in the crystal structure quite well, with small deviations ranging from 0.73 to 1.18 Å. The RMS deviations for NADH heavy atoms are all

**Table 2** *RMS Deviations (Å) [a] in the Monomer Simulations*

| | Protein mobile belly | Loop flap backbone |
|---|---|---|
| Mono24_1 | 1.13 (0.13) | 1.60 (0.23) |
| Mono24_c1 | 1.18 (0.12) | 1.74 (0.28) |
| Mono24_1r | 0.81 (0.05) | 1.08 (0.09) |
| Mono28_1 | 1.12 (0.12) | 1.59 (0.25) |
| Mono28_c1 | 1.05 (0.11) | 1.19 (0.19) |
| Mono28_1r | 0.73 (0.05) | 1.49 (0.11) |

*[a] RMS deviations of the heavy atoms of the protein mobile belly and the backbone atoms of the loop flap (residues 103-107) are calculated after first removing rotations and translations with respect to the fixed non-belly portion of the protein. The RMS deviation is then averaged over every saved timestep of the collection period of the dynamics. RMS fluctuations are shown in parentheses.*

approximately 0.68 Å, while for oxamate the range is wider, from 0.43 to 1.19 Å (data not shown). In general, these RMS deviations are quite small, a result which might stem partially from the fact that only those atoms within 17 Å of the oxamate were allowed to move.

RMS deviations, however, are only gross indications of the quality of a simulation because local deviations in structure and hydrogen bonding can be masked in an average sense. It is more informative to examine hydrogen bond occupancy as shown in Table 3. Two of the most important interactions are with the proton donor His195 and the hydride-ion donor NADH. With the exception of Mono28_1, oxamate hydrogen bonds to His195 nearly 100% of the time for all simulations. Likewise, the NADH C4 to oxamate C1 distance of 3.5± 0.2 Å in all simulations (data not shown) is close to the experimental distance of 3.20 Å. Therefore, structures from all simulations except Mono28_1 are poised to allow the chemical reaction to occur with respect to the geometry of the proton and hydride-ion donors relative to the substrate.

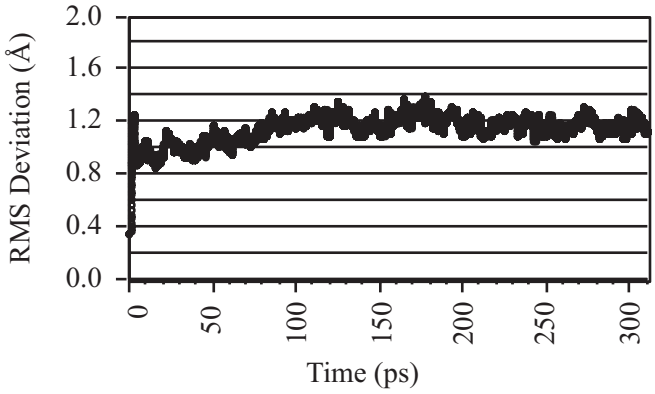



**Figure 4** *Time history of the RMS deviation of the heavy atoms of the mobile belly in Mono24_1 from the X-ray structure. Note that the RMS deviations are measured after rotating snapshots (taken every 0.025 ps) so that the heavy atoms of the protein which are not in the mobile belly have the best RMS overlap with the crystallographic structure.*

Hydrogen bonds between Arg171 and oxamate display a high degree of occupancy throughout all simulations. Other hydrogen bonds observed crystallographically, however, are not maintained in all of the monomer calculations. For example, while a high occupancy for the Thr246-oxamate hydrogen bond is observed for some of the trajectories, in others (Mono24_c1, Mono28_1, and Mono28_c1) this hydrogen bond is broken and exchanged for one to water. While it is common in all monomer simulations for Thr246 OG1 to be an acceptor in a hydrogen bond to a water bound in the active site, when the Thr246 hydrogen bond to oxamate is broken, the hydroxyl hydrogen is also hydrogen bonded to a second water. This second water has penetrated the active site via a pathway accessible to bulk solvent but which is usually blocked by the adjacent subunit C in the tetramer. In Mono24_1, additional water has penetrated the active site as well, and although the Thr246-oxamate hydrogen bond is not exchanged in this case, the additional water shifts Arg171 so that its hydrogen bonds to oxamate are not as linear (Figure 5a).

Likewise, the hydrogen bond occupancy for Gln102 and Arg109 varies among the monomer calculations. These residues are part of the active-site loop which closes over the substrate pocket. This loop, which is intrinsically more flexible than other parts of the protein, is expected to be more sensitive to the protein and water environment, as well as any deficiencies in the force field or protocol. As may be seen in Table 2, the loop backbone atoms always have a higher RMS deviation than the heavy atoms of the rest of the protein belly, and this difference increases even more if the RMS deviations of all loop heavy atoms are considered.

As a result, in two of the monomer calculations (Mono24_1 and Mono28_1) Arg109 is not positioned in its correct orientation. In these cases, the Arg109 CG-CD-NE-CZ torsion moves from its crystallographic value of 83° to about 160°, effectively rotating the plane of the guanidinium group so that the two usual hydrogen bonds to oxamate are not possible (Figure 5a). The distance between oxamate O1 and Arg109 NE increases only slightly (from 2.9Å to about 3.2Å) but the angle precludes hydrogen bond formation. An alternate hydrogen bond between the donor Arg109 NH1 and acceptor His195 O is then formed, replacing a crystallographi-cally observed hydrogen bond between His195 O and a water.

This shift in Arg109 is also accompanied by a shift in the hydrogen bonding pattern of Gln102. In the crystallographic structure, Gln102 OE1 is hydrogen bonded to the amino group of the inhibitor oxamate (even though such a hydrogen bond is impossible with either pyruvate or L-lactate). If Arg109 is in the crystallographic conformation (Figure 5c), the Gln102-oxamate hydrogen bond is typically not very highly populated in the monomer simulations and is exchanged for hy-

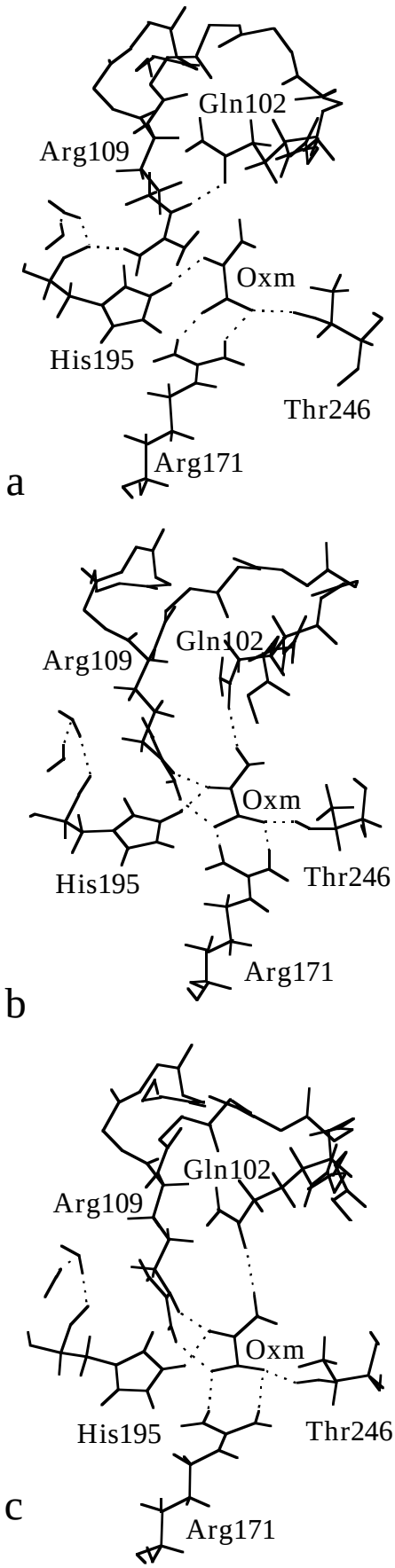



**Figure 5** *Structures of the last timestep for (a) Mono24_1 and (b) Sph24_1 compared with (c) the crystallographic structure. The loop backbone is shown as well as other critical active site residues and two waters which surround the substrate-analogue oxamate (Oxm). Hydrogen bonds are indicated by dashed lines.*

drogen bonds to water. In the alternative configuration (Figure 5a), in which the guanidinium group of Arg109 rotates, Gln102 OE1 acts as an acceptor in a hydrogen bond with donor Arg109 NE. Interestingly, in a high temperature MD simulation of LDH [7], the formation of this same hydrogen bond was observed to trigger loop opening. While in the current simulations the loop does not fully open and the hinge torsions remain near crystallographic values, it is not clear whether the loop would eventually open if the simulations were continued for long enough or at a higher temperature to speed this very slow (350 $s^{-1}$) conformational change.

A time history of the hydrogen bond exchange that occurs when Arg109 shifts is shown in Figure 6 for the first 20 ps of the simulation employing the protocol Mono24_1. The crystallographic structure (Figure 5c) is maintained for the first 6 ps, with hydrogen bonds between oxamate and Arg109 NE and NH2 present. After 6 ps, however, Arg109 rotates and these hydrogen bonds are exchanged for non-crystallographic Arg109-Gln102 and Arg109-His195 hydrogen bonds. Note, however, that the crystallographic hydrogen bond between Gln102 and oxamate is observed with a high occupancy in the alternative conformation. The crystallographic hydrogen bond between Arg109 NE and oxamate O1 reforms intermittently after 6 ps, due to the atoms' close proximity even after Arg109 rotates, but the crystallographic hydrogen bond between Arg109 NH2 and oxamate O3 is broken for the duration of the 310 ps simulation. Even though a distinct conformational change has occurred after 6 ps, all loop-loop hydrogen bonds (as defined previously [5]) persist, as shown by the top five hydrogen bonds in Figure 6. This fixed intra-loop hydrogen bonding pattern is consistent with the model of the loop as a rigid flap moving only in the flexible hinges.

**Effect of simulation conditions on simulation behaviour** Given these general observations about the behaviour of the monomer simulations, we can now examine whether any correlation exists between the observed behaviour and a given protocol, and use this analysis to choose the protocol which produces the most physically realistic results. The simulations should reproduce the crystal structure accurately in terms of RMS deviations, as well as maintain the hydrogen bond geometry of the active site. As discussed above, important specific geometric features the simulations must maintain are the orientations of the proton and hydride-ion donors with respect to the substrate, and the proper placement of the loop residue Arg109.

In the evaluation of the quality of a simulation, it is difficult to separate effects which are direct consequences of the protocol, from random fluctuations which might be rare events unique to one trajectory and independent of the protocol. This

**Table 3** *Percentage of Collection Period of Dynamics in the Monomer Simulations in which Oxamate is Hydrogen Bonded to Given Active-Site Residues (See Figure 2)*

| Donor Acceptor | Arg109 NE Oxm O1 | Arg109 NH2 Oxm O3 | Arg171 NH1 Oxm O2 | Arg171 NH2 Oxm O3 | His195 NE2 Oxm O1 | Thr246 OG1 Oxm O2 | Oxm N Gln102 OE1 |
|---|---|---|---|---|---|---|---|
| Mono24_1 | 1 | 0 | 99 | 100 | 98 | 100 | 68 |
| Mono24_c1 | 99 | 100 | 98 | 100 | 99 | 0 | 0 |
| Mono24_1r | 100 | 100 | 100 | 100 | 99 | 100 | 14 |
| Mono28_1 | 17 | 0 | 100 | 99 | 24 | 0 | 8 |
| Mono28_c1 | 96 | 78 | 100 | 100 | 99 | 47 | 14 |
| Mono28_1r | 100 | 100 | 95 | 100 | 99 | 100 | 3 |

event might trap the system in an unusual higher energy conformation which persists on the timescale of the simulations, resulting in the poor conformational sampling commonly observed in MD simulations [34]. For instance, in the physical world, the mobile loop does open and close, and, thus, one would expect a "good" MD simulation to model this opening. However, the rate of this opening is 350 $s^{-1}$, so that the loop should open once in a room temperature MD simulation of duration $10^9$ ps. While there is certainly a small chance that in these short 310 ps simulations the loop would open (or at least start to), clearly the probability is quite low, and if multiple trajectories with the same protocol display this loop opening, it is more likely that the motion is not physically realistic but instead is due to shortcomings of the protocol. Therefore, before different protocols are compared, trajectories run with the same protocol will first be compared.

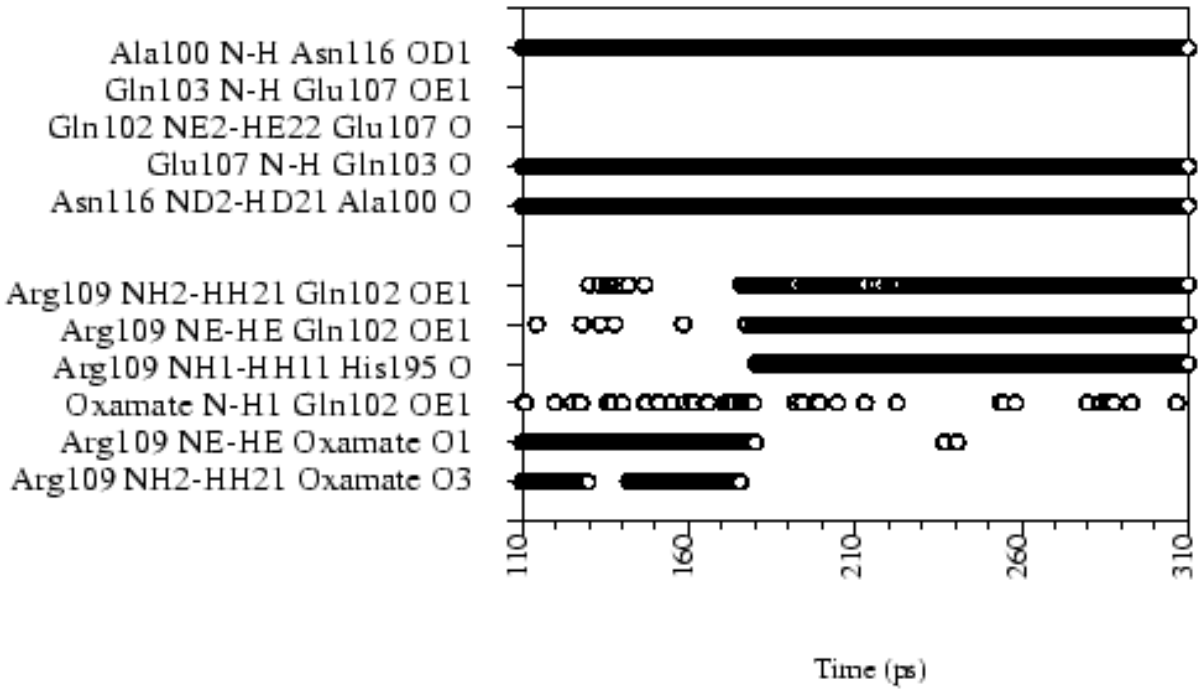



**Figure 6** *Time history of hydrogen bonds during the first 20 ps of the Mono24_1 simulation. Each row corresponds to a given hydrogen bond, as labelled on the vertical axis, and ❍ indicates that a hydrogen bond (defined by the geometric criteria given in Methods) is present for a given timestep. Starting from the top of the graph, the first five hydrogen bonds are transverse loop-loop hydrogen bonds observed crystallographically. The next three are non-crystallographic hydrogen bonds which are formed in the alternate Arg109 conformation (Figure 5a). The final three hydrogen bonds are crystallographically observed hydrogen bonds between oxamate and Gln102 or Arg109 (Figure 5c).*

**Table 4** *RMS Deviations (Å) [a] in the Multimer Simulations with Results from Mono24_1 and Mono24_1r Shown for Comparison*

| | Protein mobile belly | Loop flap backbone |
|---|---|---|
| Mono24_1 | 1.13 (0.13) | 1.60 (0.23) |
| Sph24_1 | 0.61 (0.40) | 0.28 (0.04) |
| Dim24_1 | 1.00 (0.08) | 1.82 (0.33) |
| Tetra24_1 | 1.05 (0.08) | 2.00 (0.27) |
| Mono24_1r | 0.81 (0.05) | 1.08 (0.09) |
| Sph24_1r | 0.59 (0.04) | 0.33 (0.05) |
| Dim24_1r | 0.89 (0.06) | 1.63 (0.21) |
| Tetra24_1r | 0.81 (0.05) | 1.47 (0.08) |

*[a] RMS deviations of the heavy atoms of the protein mobile belly and the backbone atoms of the loop flap (residues 103-107) are calculated after first removing rotations and translations with respect to the fixed non-belly portion of the protein. The RMS deviation is then averaged over every saved timestep of the collection period of the dynamics. RMS fluctuations are shown in parentheses.*

If the motion of the loop and Arg109 is a rare event, it is not expected to recur in multiple simulations, but if it occurs more frequently, then it is more likely to be due to deficiencies in the protocol or other errors in the model.

**Comparison of trajectories using the same protocol** Two additional simulations were run for three different protocols using the same simulation conditions and starting coordinates, but with different random numbers used to assign the initial velocities. For the protocol Mono24_1, the average RMS deviations for the belly of the protein are similar for the three simulations (1.13, 1.14, and 1.35 Å) and hydrogen bonding patterns are also similar, with the crystallographic Arg109-oxamate hydrogen bonds broken in all three cases. Results for the three trajectories employing protocol Mono24_1r are similarly consistent. The RMS deviations of the protein belly are nearly identical (0.81, 0.81, and 0.80). Very similar hydrogen bonding behaviour is observed for the three trajectories, with mostly 100% occupancies for all hydrogen bonds except for Gln102-oxamate, which is low for all three simulations. Repeats of Mono28_1 simulations, however, exhibit larger variability. The RMS deviations of the protein belly compared with the crystal structure are 1.12, 1.19, and 1.01 Å for the three trajectories. However, significant differences in hydrogen bonding patterns belie the similar RMS deviations, highlighting the risks of relying on such a gross analysis of the geometry of the active site. Two out of the three simulations have low occupancy of the hydrogen bond between oxamate and His195, which is critical for the chemical reaction. Two out of the three simulations have poorly placed Arg109 and all have low hydrogen bond occupancy for Gln102-oxamate. None of the three has both well positioned His195 and Arg109.

In summary, the multiple Mono24_1 and Mono24_1r trajectories are more reproducible than the Mono28_1 trajectories. Even though different behaviour was seen in the three trajectories of Mono28_1, all three displayed significant shifts in hydrogen bonding which would reduce the ability of the model to emulate physical behaviour. Therefore, it appears that trends in the stability of the hydrogen bond patterns and the RMS deviations from the crystallographic structure should be good guides to the quality of the protocol.

**Effect of size of water cap** By comparing the pairs Mono24_1 / Mono28_1, Mono24_c1 / Mono28_c1, and Mono24_1r / Mono28_1r, the effect of increasing the radius of the water cap from 24 to 28 Å can be gauged. The 24 Å cap is large enough so that several layers of water are present on the exterior of the loop. However, if this cap size is inadequate, imbalances in interactions on either side of the loop might cause the loop (and hence Arg109) to shift. For all pairs, the RMS deviation of the belly in the 28Å cap simulations is slightly better than that observed with the smaller water cap (Table 2). The behaviour of the loop, though, appears to be independent of the size of the water cap, with all pairs exhibiting the same behaviour (either remaining in the crystallographic conformation or shifting to the alternative conformation). Therefore, the anomalous movement of the loop does not appear to be due to insufficient solvation on the exterior. The larger, and more computationally expensive, 28 Å cap is thus not warranted, and indeed in some cases may even make the simulation worse, as the simulation with the lowest occupancy of the crucial His195-oxamate hydrogen bond used a 28 Å cap (Mono28_1).

**Effect of counterions** The influence of counterions can be gauged by comparing the pairs Mono24_1 / Mono24_c1 and Mono28_1 / Mono28_c1. The RMS deviations are very similar within each pair. In both Mono24_1 and Mono28_1, Arg109 is in the alternative conformation (Figure 5a), but upon adding counterions to these simulations (Mono24_c1 and Mono28_c1), the crystallographic Arg109-oxamate hydrogen bonds are maintained. Therefore, counterions appear to improve the simulations, and in particular the Arg109 orientation.

**Effect of dielectric constant** By comparing the pairs Mono24_1 / Mono24_1r and Mono28_1 / Mono28_1r, it is clear that changing the dielectric constant has a large effect on the trajectories, and indeed influences the trajectories most dramatically among all variables tested in the monomer simulations. Using a distance-dependent dielectric ($\varepsilon=1*r_{ij}$) always results in a lower RMS deviation in each pair, and Mono28_1r has the lowest deviation from the crystal structure of all monomer simulations. This low RMS deviation is coupled in Mono24_1r with nearly 100% occupancies for all crystallographic hydrogen bonds except for Gln102-oxamate. Such high occupancy is generally the case for simulations

**Table 5** *Percentage of Collection Period of Dynamics in the Multimer Simulations in which Oxamate Is Hydrogen Bonded to Given Active-Site Residues with Results from Mono24_1 and Mono24_1r Shown for Comparison*

| Donor Acceptor | Arg109 NE Oxm O1 | Arg109 NH2 Oxm O3 | Arg171 NH1 Oxm O2 | Arg171 NH2 Oxm O3 | His195 NE2 Oxm O1 | Thr246 OG1 Oxm O2 | Gln102 OE1 Oxm N |
|---|---|---|---|---|---|---|---|
| Mono24_1 | 1 | 0 | 99 | 100 | 98 | 100 | 68 |
| Sph24_1 | 100 | 99 | 100 | 100 | 99 | 97 | 93 |
| Dim24_1 | 1 | 0 | 100 | 100 | 99 | 100 | 3 |
| Tetra24_1 | 25 | 1 | 55 | 100 | 92 | 100 | 81 |
| Mono24_1r | 100 | 100 | 100 | 100 | 99 | 100 | 14 |
| Sph24_1r | 100 | 100 | 100 | 100 | 100 | 100 | 86 |
| Dim24_1r | 100 | 93 | 84 | 100 | 100 | 100 | 93 |
| Tetra24_1r | 55 | 21 | 94 | 100 | 97 | 100 | 28 |

with $\varepsilon=1*r_{ij}$, and neither monomer simulation employing a distance-dependent dielectric has a poorly positioned Arg109. Similarly good results have been observed when using a distance-dependent dielectric with both the Cornell [35] and the Weiner [36] AMBER force fields.

However, as argued in the Discussion, despite these excellent results, use of the distance-dependent dielectric is not formally correct in these simulations which already include explicit solvation, and might be making the system unphysically rigid by decreasing long-range interactions and altering both the water and protein force fields. Increased rigidity, for example, can be observed in the α1G-α2G helix in simulations employing a distance-dependent dielectric (Figure 7a). In contrast, for the monomer calculations using a constant dielectric, in which the loop and Arg109 shift, significant movement is observed for this helix (Figure 7b). In a crystallographic comparison of the apo and ternary complexes, [5] this helix was considered to be a minor mover, so that the distance-dependent dielectric simulations appear to be modelling the physical system better in this respect. In the monomer simulations, the α1G-α2G helix is in contact with the flap of the loop on one side and the water cap on the other. In the biological tetrameric molecule, however, in addition to contacting the flap of the loop, the end of the α1G-α2G helix interacts with the neighbouring subunit C instead of solvent water. This adjacent rigid scaffolding might be expected to be important in maintaining the active-site geometry.

*Multimeric systems*

The question then arises: if more of the tetramer is included in the simulations, will the movement of the α1G-α2G helix (and the loop and Arg109, in turn) be reduced, thus avoiding the use of the questionable distance-dependent dielectric? To answer this question, a series of simulations was performed using increasing proportions of the tetrameric molecule: (1) a sphere of all protein within ~24 Å of the oxamate; (2) a dimer; and (3) a tetramer. These additional simulations were performed with either constant or distance-dependent dielectrics. To reduce computational time, the smaller water cap (radius=24Å) was used, and the effect of counterions was not studied.

**Effect of including more protein** In both series of calculations, simulations using a sphere of protein residues (Sph24_1 and Sph24_1r) produce the lowest RMS deviations from the experimental structure (0.61 and 0.59 Å, Table 4) which are even lower than for the best monomer calculation (Mono28_1r, 0.73 Å). The monomer, dimer, and tetramer calculations result in similar RMS deviations for the protein in the constant dielectric (~1.0 Å) and distance-dependent series (~0.8 Å).

In both the Sph24_1 and Sph24_1r simulations, the backbone of the α1G-α2G helix remains near the crystallographic conformation (Figure 7c). In turn, the loop does not shift, with the backbone of the loop flap displaying very low deviations from the crystal structure (0.28 and 0.33 Å, Table 4), and Arg109 and oxamate forming the two crystallographically observed hydrogen bonds for nearly 100% of both simulations (Table 5). Surprisingly, however, including the neighbouring subunit does not always ensure that Arg109 is properly placed. In Dim24_1 and Tetra24_1, and even Tetra24_1r (the only distance-dependent dielectric simulation with an incorrect Arg109 position), Arg109 has rotated to its alternative conformation (Figure 5a) and the loop has shifted (Figure 7d). Note, however, that in Tetra24_1 the α1G-α2G helix does not move as much as in Mono24_1 (Figure 7b), and the movement at the end of the helix which directly contacts subunit C is markedly decreased. Therefore, even though the presence of the neighbouring subunit C decreases the flexibility of the α1G-α2G helix, the loop and Arg109 move due to some other flaw in the force field or simulation protocol.

While not all of the multimer simulations result in the proper positioning of Arg109 and loop, adding more of the tetramer to the computational model certainly improves the occupancy of the Thr246-oxamate hydrogen bonds. All of the multimer simulations exhibit nearly 100% occupancy

**Table 6** *Interaction Energy of the Loop with the Remainder of the System [a]*

| | Oxamate | NADH | Mobile belly | Charged non-belly | Neutral non-belly |
|---|---|---|---|---|---|
| Mono24_1 | -74.5 | -103.2 | -237.1 | 2.8 | -4.1 |
| Sph24_1 | -73.9 | -102.6 | -239.7 | 24.8 | -2.6 |
| Dim24_1 | -73.8 | -103.4 | -217.8 | 47.2 | -3.4 |
| Tetra24_1 | -73.6 | -101.1 | -238.0 | 66.8 | -4.9 |
| Mono24_1r | -41.0 | -56.0 | -165.7 | 0.2 | -4.6 |
| Sph24_1r | -40.2 | -54.9 | -166.0 | 1.7 | -4.5 |
| Dim24_1r | -40.4 | -55.8 | -163.5 | 3.2 | -4.8 |
| Tetra24_1r | -39.8 | -55.7 | -166.3 | 5.0 | -5.1 |

*[a] The total interaction energy (in kcal/mol) was calculated between all atoms of the loop (residues 98-113) and the active-site oxamate; NADH; mobile belly; and the charged or neutral non-belly (all charged or neutral amino acids in the fixed protein). The coordinates were taken from the minimisation sequence after the water had been relaxed and the protein hydrogens were minimised, so that the heavy atoms of the protein, cofactor, and substrates were still in the crystallographic conformation.*

(Table 5). No additional waters penetrated the active site via the pathway near Thr246, as was observed in the monomer simulations, and Arg171 maintains linear hydrogen bonds with oxamate. Thus, the neighbouring subunit which blocks this "hole" is important for excluding water from the active site.

Most other interactions with amino acids in the multimer calculations are maintained properly. All simulations maintain hydrogen bonds between oxamate and His195 and Arg171 for most of the trajectory (Table 5), and all display an average NADH C4 to oxamate O1 distance of about 3.5± 0.2 Å (data not shown). Hydrogen bonds between oxamate and Gln102 persist for both Sph24_1 and Sph24_1r, with an occupancy of 93% and 86%, respectively. The occupancy of this hydrogen bond is variable for the dimer and tetramer simulations, as in the monomer simulations, probably due to the increased RMS deviations of the loop. Overall, then, among both monomer and multimer simulations, Sph24_1 and Sph24_1r are clearly the best, with low RMS deviations and accurate replication of the crystallographic hydrogen bonding pattern in the active site.

**Effect of dielectric constant** As in the monomer simulations, when comparing pairs with and without distance-dependent dielectric (Sph24_1 / Sph24_1r, Dim24_1 / Dim24_1r, and Tetra24_1 / Tetra24_1r), simulations employing $\varepsilon=1*r_{ij}$ always exhibit a lower RMS deviation. But in contrast to the monomer simulations, Sph24_1 and Sph24_1r exhibit nearly identical hydrogen bonding patterns, so that no additional benefit is gained from using a distance-dependent dielectric. For the dimer, using $\varepsilon=1*r_{ij}$ improves the hydrogen bonding pattern, while for the tetramer both Tetra24_1 and Tetra24_1r reproduce the crystallographic structure poorly, with movement observed in both the loop and Arg109.

## Discussion

The aim of this study was to examine the physical factors which influence the stability of the active-site geometry for LDH. This analysis acts as a guide in optimising the simulation protocol which will be used in the MM region of future reaction mechanism calculations using the hybrid QM/MM method. This optimum simulation protocol will also be used in MD simulations to generate physically realistic starting coordinates for the reaction-mechanism calculations. Here, a series of simulations was evaluated according to their ability to reproduce the hydrogen bonding pattern of the active site and to exhibit a low average RMS deviation from the crystallographic structure. Several issues appear to be crucial to modelling this system accurately: the conformation of Arg109 and the mobile loop of which it is part, and the influence of the surrounding tetramer on each active site.

Previous studies [23] have also concluded that the conformations of Arg109 and the mobile loop are critical discriminating features among simulations. This region of the protein is intrinsically mobile, as the loop must open and close to allow substrate and cofactor binding and release, and, hence, the average temperature factor of the flap of the loop (residues 103-107) is larger than the average for the whole protein (25.2 compared with 18.5 $Å^2$/atom [27]). Movement in this region, however, is observed on a shorter timescale (310 ps) in our simulations than is expected. The loop appears to be shifting due to deficiencies or imbalances in the protocol or force field, resulting in one common alternative configuration in which Arg109 does not hydrogen bond to the oxamate, but instead is hydrogen bonded to His195 and Gln102 (Figure 5a). This new hydrogen bond is not observed in the best simulations as the loop remains closed as expected

for this timescale. This newly formed hydrogen bond between Arg109 NH2 and Gln102 OE1 was identified in a previous high temperature MD simulation of LDH to be the trigger for the opening of the loop [7]. While in the current room temperature simulations the loop does not open significantly, perhaps if the simulations were continued for a longer time or at higher temperatures the formation of this new hydrogen bond might initiate loop opening. Given that different programs and force fields have been used in the two studies, the similarity between the two results is encouraging. Equally encouraging is the observation in the current simulations that during the Arg109 conformational change, loop-loop hydrogen bonding remains constant (Figure 6), which is consistent with previous studies [5,7] which describe the loop as a rigid flap moving only in the flexible hinges.

The crystallographic hydrogen bonds between substrate and Arg109 are postulated to polarise the substrate carbonyl bond, facilitating both proton and hydride-ion transfer [10]. In experiments where Arg109 is replaced by glutamine [37], the resulting mutant enzyme is still active, but $k_{cat}$ is dramatically decreased from 250 to 0.6 $s^{-1}$. Thus, while Arg109 is not essential for the reaction to occur, the enzymic rate acceleration is assisted greatly by its binding and polarisation effects. The shift in the Arg109 hydrogen bonding arrangement observed in some simulations could, therefore, be expected to have major implications for future QM/MM simulations of the chemical reaction.

Initially, we suspected that the reason Arg109 shifts in the monomer simulations is the absence of a neighbouring subunit C. This neighbour contacts the end of the α1G-α2G helix of subunit A, and in its absence, excessive movement might be expected in the helix, and the mobile loop, in turn. Indeed, upon adding portions of the immediately adjacent subunit in the protein sphere simulations (Sph24_1 and Sph24_1r), reduced movement was observed in both the α1G-α2G helix and the loop containing Arg109. However, this trend did not continue in the dimer and tetramer simulations, with Dim24_1, Tetra24_1 and Tetra24_1r all displaying loop movement which disrupted the Arg109-oxamate hydrogen bonds. As significant movement of the α1G-α2G helix is not observed in these simulations, the sole cause of the loop shift is not excessive movement of this helix.

Presumably, however, maintaining the rigidity of this helix is still important even if it does not necessarily prevent loop motion. Other residues may move from their crystallographic positions if this helix shifts, so adding additional protein in the multimer systems (sphere, dimer and tetramer) appears necessary to model this system accurately. The surrounding protein environment is also necessary to prevent penetration of bulk water into the active site via a pathway near Thr246. The tetrameric active site is normally shielded from bulk water after closure of the mobile loop, with only a few well-ordered waters in the active site, but allowing more water to enter the active site could affect the reaction markedly. In addition to potentially influencing loop opening, this water penetration was shown to alter the active-site geometry by shifting Arg171 in some monomer simulations (Figure 5a) so that its hydrogen bonds to oxamate are not as linear. Additional water in the active site also alters the electrostatic environment. Therefore, the rest of the tetramer is important for maintaining the active-site geometry, and must be represented adequately in the simulations. This analysis suggests why only tetramers or dimers (in the case of *Bacil-*

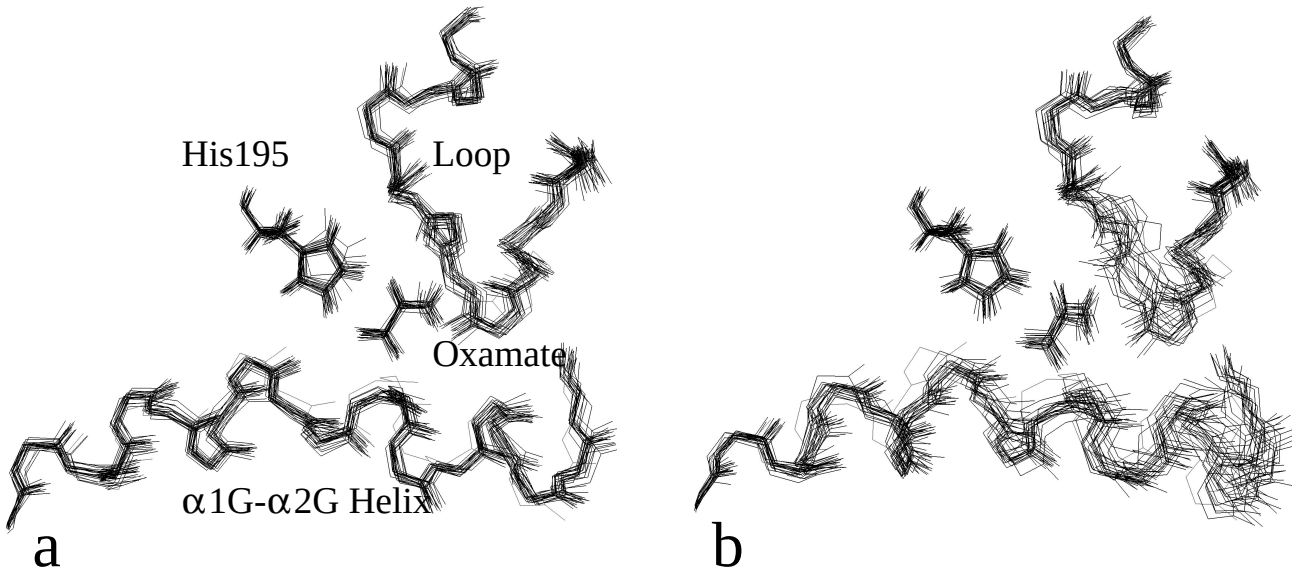



**Figure 7** *Superposition of twenty snapshots taken at 10 ps intervals from the 200 ps data collection period of simulations (a) Mono24_1r and (b) Mono24_1. Rotations and translations with respect to the fixed non-belly portion of the protein are first removed. The backbone atoms of both the loop and the a1G-a2G helix are shown, as well as all atoms of oxamate and the side chain of the catalytic His195.*

*lus stearothermophilus*), and not monomers, are active biologically.

Given the importance of including part of the neighbouring subunit, it is surprising that the results of the dimer and tetramer simulations are worse than those employing the protein sphere, as in theory the larger systems should represent the environmental influence of the surrounding protein more accurately. One possible explanation for the loop movement in the dimer and tetramer simulations (in addition to some monomer simulations) is improper representation or balance of the long-range electrostatic interactions. Long-range interactions may well stabilise overall protein structure, as has been observed in simulations employing Ewald sums [35] or dual cutoffs [38]. When a nonbonded cutoff is used, the neglect of interactions beyond the cutoff may be problematic [39].

The suggestion that loop movement is still observed in the dimer and tetramer because of improperly represented long-range electrostatic interactions is supported by several observations. First, most simulations employing a distance-dependent dielectric maintain the crystallographic orientation of Arg109 and the loop. A distance-dependent dielectric reduces the contribution of long-range electrostatic interactions relative to local interactions, and so the influence of charges far from the active site is minimised. Second, the addition of counterions (which effectively neutralise charges at long ranges) improves simulations in the monomer simulations. Third, the surface area of the artificial protein-vacuum interface is larger for the dimer and tetramer than for the sphere and monomer, with a larger number of charged amino acids on the surface. Finally, the sphere model is spherically symmetrical with a constant radius of protein or water surrounding the mobile belly, which may prevent an imbalance of nonbonded interactions from a given point.

Imbalances in long-range electrostatics might be removed by submerging the entire tetramer in a box of water, subjecting the system to periodic boundary conditions, and using Ewald sums to calculate the nonbonded interactions. While the MD simulation behaviour would most likely be improved, this approach is impractical as the eventual goal is to perform QM/MM calculations, which would be prohibitively expensive for a model with so many atoms. Alternatively, the current simulations could be run without nonbonded cutoffs, which would increase CPU time dramatically as well. Using a nonbonded cutoff has the advantage of decreasing computational time, and although not ideal, in some of the protocols the nonbonded interactions do appear to be well balanced despite this approximation. This may be the result of using a longer cutoff (14 Å) than is typically used (8 Å) in an effort to minimise the magnitude of the forces at the cutoff.

The artificial protein-vacuum interface might well be responsible for the imbalance of the long-range electrostatic interactions. Ideally, the monomer, sphere, dimer, or tetramer models in the MD simulations should be surrounded by either protein or water which provide salt bridges, hydrogen bonds, and electrostatic shielding. These effects of the environment are modelled for the most important portion of the surface by adding a water cap over the active-site region. While an artificial boundary between the vacuum and the outside of the water cap still exists, it appears a water cap with radius 24 Å is sufficiently large to prevent anomalies. However, the remainder of the protein surface not under the

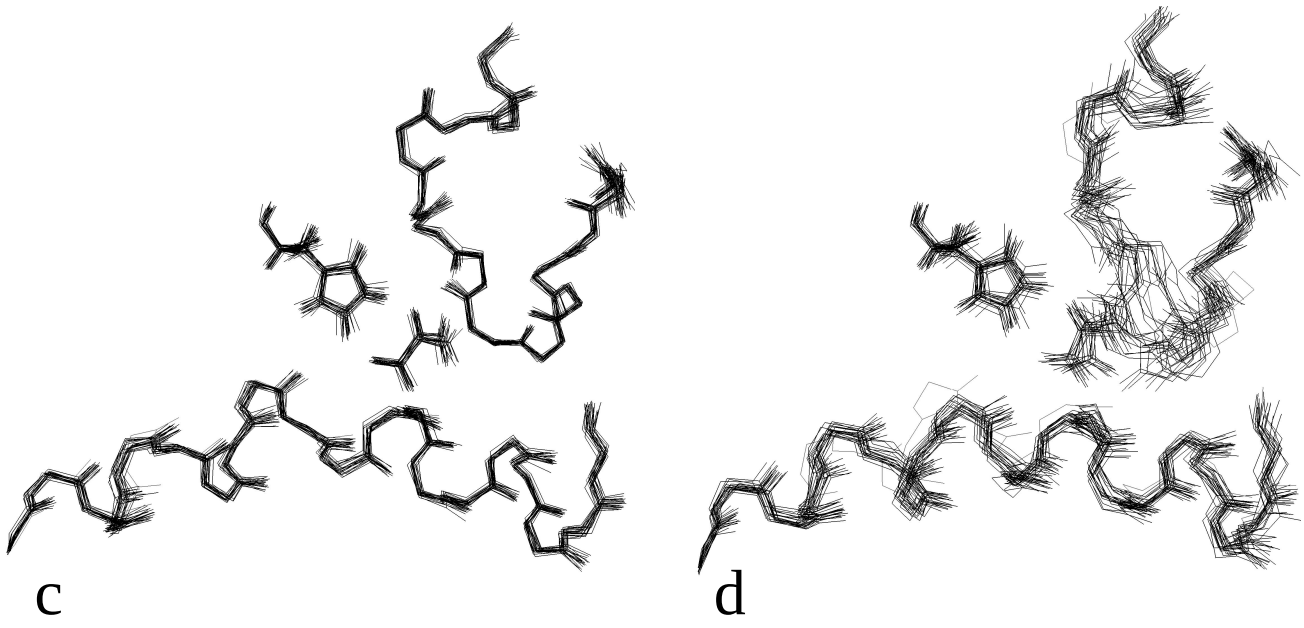



**Figure 7 (continued)** *Superposition of twenty snapshots taken at 10 ps intervals from the 200 ps data collection period of simulations (c) Sph24_1 and (d) Tetra24_1. Rotations and translations with respect to the fixed non-belly portion of the protein are first removed. The backbone atoms of both the loop and the a1G-a2G helix are shown, as well as all atoms of oxamate and the side chain of the catalytic His195.*

water cap still encounters an artificial vacuum. Charged residues in this region are particularly problematic as electrostatic interactions are long-range, and physically unrealistic charges might affect protein dynamics in the active site. Although these effects were presumed originally to be small, as the exterior of the protein is fixed and a nonbonded cutoff of 14 Å excludes most long-range interactions, it is possible that the surface can indirectly influence the active site in an unpredictable fashion via interactions with the outer regions of the mobile belly as well as the loop. Indeed, the interaction energy between the loop and the fixed protein outside of the mobile belly increases as more of the tetramer is added to the model (Table 6). Most of this interaction with the non-belly is due to charged amino acids, while the interaction with neutral amino acids in the non-belly is much smaller. In contrast, the interaction of the loop with oxamate, NADH, and the mobile belly is comparable among the monomer, protein sphere, dimer, and tetramer simulations. The magnitude of all interactions is reduced when a distance-dependent dielectric is used, and, in particular, interactions between the loop and the non-belly, explaining why the loop shift is much less likely in these simulations. Thus, the unfavourable long-range influence of the charged amino acids on the surface of the protein, which are not shielded by the normal protein or water environment, appears be the source of the imbalance in electrostatics which causes the loop to shift.

Attempting to reduce the influence of these charged surface residues by adding counterions improved the monomer simulations in that Arg109 and the loop remained in the crystallographic orientation. However, counterions were used only with the monomer simulations, and other problems were apparent resulting from the absence of a neighbouring subunit (water penetration and increased movement of the α1G-α2G helix). Further studies would be required to evaluate whether adding counterions can be beneficial for the larger systems. Two protocols, however, (Sph24_1 and Sph24_1r) generated good trajectories without use of counterions.

In both of the simulations which employ a protein sphere, the long-range electrostatics appear to be properly represented, so that Arg109 and the loop remain fixed. This inclusion of a small portion of the neighbouring subunit results in a low RMS deviation from the crystallographic structure. The choice between using a constant (Sph24_1) or a distance-dependent dielectric (Sph24_1r) must then be made. Simulations using a distance-dependent dielectric are generally very good, even when only one monomer of the tetramer is included. However, it is uncertain whether using this distance-dependent dielectric might produce the right answer for the wrong reason. Other studies have shown that atomic fluctuations in simulations with a distance-dependent dielectric are too small when compared with that predicted by crystallographically measured B factors, so that the structure remains very close to experiment but with excessively damped internal motions [35]. Here, simulations using $\varepsilon=1*r_{ij}$ underestimate the effects of long-range electrostatics which appear to cause conformational changes in the active site. This effect is fortuitous, in that the crystallographic structure is maintained accurately, but it is difficult to predict other subtle effects on the dynamics and solvation which might not reflect physical reality so well. Distance-dependent dielectrics were designed to mimic solvation implicitly by providing the dampening effect water exerts on long-range electrostatic interactions due to its high dielectric constant. When used with explicit water molecules, however, the consequences are uncertain, as dielectric screening is already accounted for and, thus, is doubly counted. The TIP3P model will not be correct using $\varepsilon=1*r_{ij}$ as it was developed and optimised using $\varepsilon=1$ [32], while the Cornell protein force field was designed for use with explicit water and $\varepsilon=1$ [31]. Therefore, the magnitude and balance of water-water and water-protein interactions will be altered in an unpredictable fashion. In particular, the stronger TIP3P water-water interaction may be the reason the loop and Arg109 remained fixed. Fortunately, the potential pitfalls of using a distance-dependent dielectric can be avoided by using the Sph24_1 protocol. Crystallographic conformations of the loop and Arg109 are observed, even when using a constant dielectric $\varepsilon=1$; water penetration is prevented by the inclusion of a small portion of neighbouring subunit; and nonbonded interactions appear to be well-balanced. This protocol is the method of choice for simulating this protein system.

This conclusion has been arrived at by testing protocols using only a single trajectory of limited duration (310 ps), which is quite short on the biological timescale. It can be problematic to evaluate the quality of a protocol based on such short simulations, because of the difficulty in distinguishing between random conformational transitions and those arising more directly and reproducibly from the simulation protocol. Results may not necessarily be improved by simply extending the simulation time to nanoseconds, as MD simulations can often have a persistent dependence on the initial conformation and velocities, resulting in inadequate conformational sampling. Because of potential problems with conformational sampling, a recent paper has recommended that many shorter trajectories are preferable to one long trajectory when testing protocols [34]. This concern has been addressed here to a limited degree by repeating several trajectories for selected protocols (Mono24_1, Mono24_1r, and Mono28_1). The three trajectories for Mono24_1 and Mono24_1r are very similar, while for Mono28_1 they are similar in that all have a serious flaw (either the shifting of Arg109 or the lack of a His195-oxamate hydrogen bond). While it would certainly be desirable to perform more and longer trajectories for each protocol, the trends in physical behaviour seem clear enough when only a single trajectory is performed. This reproducibility might be due to the decreased conformational freedom for this system in which the exterior of the protein is fixed, in contrast to simulations of proteins where all atoms are allowed to move.

## Conclusion

A series of MD simulations has been analysed to identify the key influences on maintaining the geometry of the LDH ac-

tive site. Two main conclusions stem from this work. First, the conformations of the mobile loop and Arg109 (which is important for polarising the substrate) are sensitive to long-range nonbonded interactions. Second, the monomer is not biologically active because the active-site geometry is adversely affected by the absence of the remaining tetramer. Without the neighbouring protein subunits, water penetrates the active site, disrupting some hydrogen bonds and distorting others, and the electrostatic environment of the reaction is altered. The neighbouring subunit is also important for providing a rigid scaffolding so that the α1G-α2G helix does not shift excessively.

MD simulations must, therefore, represent nonbonded interactions accurately and include at least part of the remaining tetramer in order to model this system properly. One protocol (Sph24_1) satisfies both these requirements in a computationally efficient manner by including all protein residues within approximately 24 Å of the substrate. In simulations using this protocol, the loop and, hence, Arg109 are properly positioned; bulk water does not penetrate the active site; and the α1G-α2G helix does not fluctuate far from its crystallographic orientation. This simulation protocol will be used for generating starting LDH-NADH-pyruvate and LDH-$NAD^+$-lactate structures for hybrid QM/MM calculations, and for the MM region in these calculations.

**Acknowledgements** Funding for RKS from an Australian Research Council Postdoctoral Fellowship is gratefully acknowledged, as well as generous computing grants and support from the Australian National University Supercomputing Facility. The ANU Strategic Development Fund provided additional funding.

## References

1. Cortes, A.; Emery, D. C.; Halsall, D. J.; Jackson, R. M.; Clarke, A. R.; Holbrook, J. J. *Prot. Sci.* **1992**, *1*, 892.
2. Burgner, J. W. I.; Ray, W. J. J. *Biochemistry* **1984**, *23*, 3636.
3. Eventoff, W.; Rossman, M. G.; Taylor, S. S.; Torff, H.-J.; Meyer, H.; Keil, W.; Kiltz, H.-H. *Proc. Natl Acad. Sci. USA* **1977**, *74*, 2677.
4. Demchenko, A. P.; Rusyn, O. I.; Saburova, E. A. *Biochim. Biophys. Acta* **1989**, *998*, 196.
5. Gerstein, M.; Chothia, C. *J. Mol. Biol.* **1991**, *220*, 133.
6. Waldman, A. D. B.; Hart, K. W.; Clarke, A. R.; Wigley, D. B.; Barstow, D. A.; Atkinson, T.; Chia, W. N.; Holbrook, J. J. *Biochem. Biophys. Res. Commun.* **1988**, *150*, 752.
7. Philippopoulos, M.; Xiang, Y.; Lim, C. *Prot. Eng.* **1995**, *8*, 565.
8. Wilks, H.; Hart, K.; Feeney, R.; Dunn, K.; Muirhead, H.; Chia, W.; Barstow, D.; Atkinson, T.; Clarke, A.; Holbrook, J. *Science* **1988**, *242*, 1541.
9. Wilks, H. M.; Halsall, D. J.; Atkinson, T.; Chia, W. N.; Clarke, A. R.; Holbrook, J. J. *Biochemistry* **1990**, *29*, 8587.
10. Deng, H.; Zheng, J.; Clarke, A.; Holbrook, J. J.; Callender, R.; Burgner, J. W. I. *Biochemistry* **1994**, *33*, 2297.
11. Hart, K. W.; Clarke, A. R.; Wigley, D. B.; Chia, W. N.; Barstow, D. A.; Atkinson, T.; Holbrook, J. J. *Biochem. Biophys. Res. Commun.* **1987**, *146*, 346.
12. Hart, K. W.; Clarke, A. W.; Wigley, D. B.; Waldman, A. D. B.; Chia, W. N.; Barstow, D. A.; Atkinson, T.; Jones, J. B.; Holbrook, J. J. *Biochim. Biophys. Acta* **1987**, *914*, 294.
13. Sakowicz, R.; Kallwass, H.; Parris, W.; Gold, M.; Jones, J. B. *Biochem. Biophys. Res. Commun.* **1992**, *182*, 1309.
14. Clarke, A. R.; Atkinson, T.; Holbrook, J. J. *TIBS* **1989**, *14*, 101.
15. Clarke, A. R.; Atkinson, T.; Holbrook, J. J. *TIBS* **1989**, *14*, 145.
16. Norris, K. E.; Bacskay, G. B.; Gready, J. E. *J. Comput. Chem.* **1993**, *14*, 699.
17. Norris, K. E.; Gready, J. E. *J. Mol. Struct. (Theochem)* **1993**, *279*, 99.
18. Andres, J.; Moliner, V.; Krechl, J.; Silla, E. *Bioorg. Chem.* **1993**, *21*, 260.
19. Wilkie, J.; Williams, I. H. *J. Am. Chem. Soc.* **1992**, *114*, 5423.
20. Ranganathan, S.; Gready, J. E. *J. Chem. Soc. Faraday Trans.* **1994**, *90*, 2047.
21. Andres, J.; Moliner, V.; Krechl, J.; Silla, E. *J. Chem. Soc. Perkin Trans. 2* **1995**, 1551.
22. Gelpi, J. L.; Jackson, R. M.; Holbrook, J. J. *J. Chem. Soc. Faraday Trans.* **1993**, *89*, 2707.
23. Dafforn, T. R.; Badcoe, I. G.; Sessions, R. B.; El Hawrani, A. S.; Holbrook, J. J. *Proteins: Struct. Funct. Genet.* **1997**, *29*, 228.
24. van der Spoel, D.; Vogel, H. J.; Berendsen, H. J. C. *Proteins: Struct. Funct. Genet.* **1996**, *24*, 450.
25. van Beek, J.; Callender, R.; Gunner, M. R. *Biophys. J.* **1997**, *72*, 619.
26. Ranganathan, S.; Gready, J. E. *J. Phys. Chem. B* **1997**, *101*, 5614.
27. White, J. L.; Hackert, M. L.; Buehner, M.; Adams, M. J.; Ford, G. D.; Lentz, P. J.; Smiley, I. E.; Steindel, S. J.; Rossmann, M. G. *J. Mol. Biol.* **1976**, *102*, 759.
28. Schar, H. P.; Zuber, H. *Hoppe-Seyler's Z. Physiol. Chem.* **1979**, *360*, 795.
29. Creighton, T. E. *Proteins: Structures and Molecular Properties*; W.H. Freeman Co.: New York, 1984.
30. Pearlman, D. A.; Case, D. A.; Caldwell, J. W.; Ross, W. S.; Cheatham, T. E.; Ferguson, D. M.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. (1995), AMBER 4.1, University of California, San Francisco.
31. Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
32. Jorgensen, W. L. *J. Am. Chem. Soc.* **1981**, *103*, 335.
33. Cummins, P. L.; Ramnarayan, K.; Singh, U. C.; Gready, J. E. *J. Am. Chem. Soc.* **1991**, *113*, 8247.
34. Caves, L. S. D.; Evanseck, J. D.; Karplus, M. *Prot. Sci.* **1998**, *7*, 649.

35. Fox, T.; Kollman, P. A. *Proteins: Struct. Funct. Genet.* **1996**, *25*, 315.
36. Guenot, J.; Kollman, P. A. *Prot. Sci.* **1992**, *1*, 1185.
37. Clarke, A. R.; Wigley, D. B.; Chia, W. N.; Barstow, D. A.; Atkinson, T.; Holbrook, J. J. *Nature* **1986**, *324*, 699.
38. Foley, C. K.; Pedersen, L. G.; Charifson, P. S.; Darden, T. A.; Wittinghofer, A.; Pai, E. F.; Anderson, M. W. *Biochemistry* **1992**, *31*, 4951.
39. Guenot, J.; Kollman, P. A. *J. Comput. Chem.* **1993**, *14*, 295.
40. Kraulis, P. J. *J. Appl. Cryst.* **1991**, *24*, 946.