Structure and Hydration of the DNA-Human Topoisomerase I Covalent Complex

Giovanni Chillemi,* Tiziana Castrignanò,* and Alessandro Desideri†

*CASPUR, c/o University of Rome "La Sapienza," 00185; and †INFM and Department of Biology, University of Rome "Tor Vergata," 00133 Rome, Italy

ABSTRACT The structure and hydration of reconstituted human topoisomerase I comprising the core and the carboxyl-terminal domains in covalent complex with 22-basepair DNA duplex has been investigated by molecular dynamics simulation. The structure and the intermolecular interactions were found to be well maintained over the simulation. The complex displays a high degree of flexibility of the contact area, confirmed by the presence of numerous water-mediated protein-DNA hydrogen bonds comparable in quantity and distribution to the direct ones. The interaction between the enzyme and the solvent also provides the key for interpreting the experimental reduction of activity or affinity observed upon single residue mutation. Finally, four long lasting water molecules are observed in the proximity of the active site, one of which in the appropriate position to accept a proton from the active Tyr723.

INTRODUCTION

Topoisomerase enzymes control the level of supercoiled DNA in cells by transiently breaking one or two DNA strands. Their role is vital for cellular processes like replication, transcription, and recombination. Human topoisomerase I is a type IB enzyme composed of four major domains (N-terminal, core, linker, and C-terminal domains) and is involved in the relaxation of both positive and negative supercoiled DNA, reversibly nicking one DNA strand. The crystal structures of a reconstituted topoisomerase comprising the core and the C-terminal domain, in both covalent and non-covalent complexes with a 22-basepair duplex DNA, have been reported (Redinbo et al., 1998; Stewart et al., 1998). The reconstituted enzyme is active in vitro and in the crystallographic structure with DNA; the authors described a large number of direct protein to DNA interactions that are unevenly distributed, being more frequent in the intact and in the upper region of the scissile strand but not in the lower region (Redinbo et al., 1998). This observation, together with the presence of a positively charged surface downstream of the cleavage site that makes minimal contact with DNA, has led to the proposal of a model which, without underlining the role of water, describes a mechanism for DNA relaxation called controlled rotation (Stewart et al., 1998).

Recent surveys of structural data on protein-protein and protein-DNA recognition sites has indicated that water is present in abundance at the interface and the analysis of different high resolution structures allowed the identification of at least as many water-mediated interactions as direct hydrogen bonds or salt bridges (Janin, 1999). In particular, water can play an important role in mediating protein-DNA interaction due to the polarity of the surface characterized by the presence of phosphate groups on the DNA side and by the frequent abundance of positively charged groups on the protein side. The major function of water in promoting macromolecular association was first suggested by the structure of the trp repressor-DNA complex (Otwinoski et al., 1988). Since then, many examples have been found, suggesting that the description of protein-DNA interactions is incomplete without an atomic description of water molecules at the interface (Schwabe, 1997; Janin, 1999). The understanding of macromolecular hydration can be provided by high resolution x-ray and neutron diffraction studies of crystals (Kossiakoff et al., 1992; Burling et al., 1996), which, however, are not able to provide a dynamical description. Dynamical information on water behavior is provided by inelastic neutron scattering and NMR spectroscopy (Otting et al., 1991; Zanotti et al., 1999; Melacini et al., 2000), or by accurate molecular dynamics (MD) simulation procedures. The latter approach is a powerful tool for examining the macromolecule solvation because it can provide a description at atomic level and at the appropriate time scale, of the macromolecule-water interaction mechanism and it has been successfully applied to describe the hydration properties of proteins (Abseher et al., 1996; Luise et al., 2000) and DNA (Bonvin et al., 1998; Castrignanò et al., 2000). In the last years, MD simulation has been used to investigate the molecular basis of protein-DNA or RNA recognition (Billeter et al., 1996; Sen and Nilsson, 1999; Reyes and Kollman, 1999; Tang and Nilsson, 1999; Tsui et al., 2000).

Here we present a state of the art 1.2 nanosecond MD simulation of the covalent topoisomerase-DNA complex carried out in periodic boundary condition, starting from the crystallographic structure (Redinbo et al., 1998). The system, fully solvated and neutralized, consists of nearly 79,000 atoms and was simulated using the Amber force field (Cornell et al., 1995) and the particle mesh method (Darden et al., 1993; et al., 1995) on a cluster of Compaq

Received for publication 1 February 2001 and in final form 5 April 2001.

Address reprint requests to A. Desideri, Department of Biology, University of Rome Tor Vergata, via della ricerca scientifica 00133, Rome, Italy. Tel.: 39-06-72594376; Fax: 39-0672594326; E-mail: desideri@uniroma2.it.

0006-3495/01/07/490/11 $2.00

ES40. The protein-DNA interface was found to be highly hydrated and the number of water-mediated hydrogen bonds is comparable to the direct ones, suggesting that they provide a comparable contribution to the binding affinity and specificity. Moreover, the active site region was found to trap for a long time four water molecules supporting the recently proposed active site mechanism, in which a water molecule plays a direct role in the deprotonation process of the active site residue Tyr723 (Redinbo et al., 2000).

METHODS

System setup and simulation protocol

The starting coordinates of the DNA-topoisomerase complex were obtained from x-ray diffraction (PDB entry 1a31) (Redinbo et al., 1998). The system topology was built using the AMBER leap module and the AMBER95 all atom force-field (Cornell et al., 1995), including the parameter for the 723 phosphotyrosine residue covalently bound to the −1 thymine scissile base. The partial atomic charges and intra-residue force parameters of the tyrosine residue and the phosphate group forming the phosphotyrosine residue were let identical to those found in the AMBER95 force field but the partial charge of the phosphorous atom which was imposed to have a value of 1.2593. This small change permits to have a charge for the phosphotyrosine residue equal to −0.6921, identical to the charge of a DNA 3′ base in the AMBER95 force field. Crystallographic water molecules were not retained and the DNA-protein complex was immersed in a rectangular box ($105 \times 92 \times 99$ Å^3) filled with TIP3P (Jorgensen, 1981) water molecules, imposing a minimum solute-wall box distance of 10 Å. The system was neutralized with Na^+ cations using the AMBER leap module. The total system is composed of 7679 protein atoms, 1401 DNA atoms, 27 Na^+ counterions and 23216 water molecules, giving a total of 78755 atoms. The system was simulated in periodic boundary conditions, using a cutoff radius of 9 Å for the non-bonded interactions, and updating the neighbor pair list every 10 steps. The electrostatic interactions were calculated with the Particle Mesh Ewald method (Darden et al., 1993; Cheatham et al., 1995). The SHAKE algorithm (Ryckaert et al., 1977) was used to constrain all bond lengths involving hydrogen atoms. Optimization and relaxation of solvent and ions were initially performed, while keeping the solute atoms constrained to their initial position with decreasing force constants of 500, 25, 15 and 5 Kcal/(mol Å). Thereafter the system was minimized without any constraint, simulated for 1 ps at 100 K and then warmed up to 300 K and equilibrated over 20 ps. A 1.2-ns simulation was carried out at constant temperature of 300 K using Berendsen's method (Berendsen et al., 1984) and at a constant pressure of one bar with a 2-fs time step, on a cluster of Compaq ES40. Pressure and temperature coupling constants were 0.5 ps. The atomic positions of the last nanosecond of trajectory were saved every 250 steps (0.5 ps) for analysis.

Analysis

The analysis reported in this manuscript refers to the last nanosecond of the trajectory (i.e., from 200 to 1200 ps), because the trend of the root mean square deviations (RMSD) of the protein-DNA complex indicates that the system was well stabilised after 200 ps. RMSD were calculated removing the overall translational and rotational motions upon superimposition to the initial x-ray structure. The AMBER carnal module was used to calculate some structural properties (root mean square deviations and fluctuations; hydrogen bonds). The hydrogen bond criterion was a maximum donor-acceptor distance of 3.5 Å and a minimum donor-proton-acceptor angle of 120°, the default for the carnal AMBER module. Protein and DNA are considered to form a stable direct hydrogen bond if the criterion is maintained for more than 70% of the simulation time. In the case of water-mediated hydrogen bonds we do not impose any restriction on the lifetime of individual water molecules but only on the contact itself. Therefore protein-DNA are considered to form a water-mediated contact when at least one water molecule simultaneously bridges through an hydrogen bonds both the protein and DNA, and when this configuration is present for more than 70% of simulation time, regardless the lifetime of individual water hydrogen bonds. Crystallographic water molecules, present in the PDB structure 1a31, were considered to be in interaction with the protein or DNA when the water oxygen-macromolecule atom distance was lower than 3.5 Å. The solvent accessible surface area was evaluated with the program NACCESS (Hubbard and Thornton, 1993).

RESULTS AND DISCUSSION

Root mean square deviations

Protein and DNA root mean square deviation (RMSD), calculated after a mass weighted superposition on the starting structure, are shown in Fig. 1 *a*. The enzyme RMSD reach a plateau at around 2 Å after about 200 ps, whereas the DNA RMSD show a fast increment in the first 50 ps, after which it stabilizes at around 2 Å displaying some oscillations which are comparable or even lower to those observed in simulations of isolated DNA strands (Castrignanò et al., 2000; Young et al., 1997). A closer inspection of the DNA RMSD indicates that its variation is mainly due to the contribution of the DNA bases at both ends of the duplex (data not shown) because the RMSD of the central bases (−4 + 6) oscillate around 1 Å. On the other hand the central bases shows a relatively high deviation from the canonical B structure (RMSD, ~2.5 Å) suggesting that the

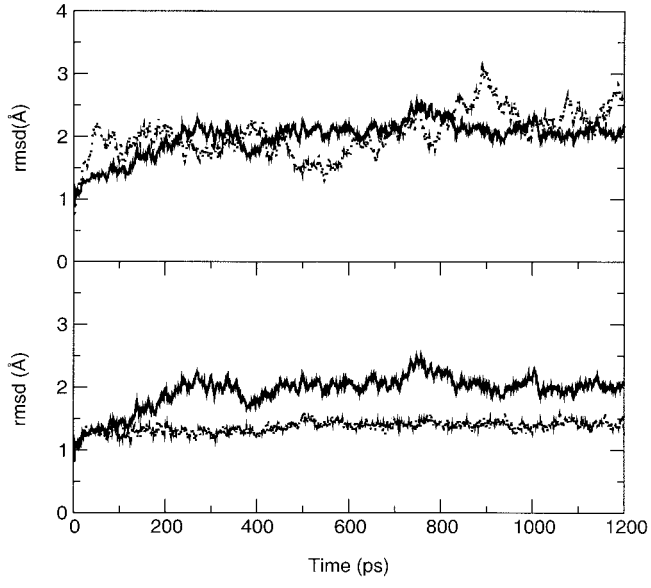

FIGURE 1 Root-mean-square atom positional deviations as a function of simulation time from the starting x-ray structure. (*a*) Protein (*full line*) and DNA (*dotted line*); (*b*) protein core domain (*full line*) and C-terminal domain (*dotted line*).

deviations observed in the x-ray diffraction study were not simply due to crystal contacts (Redinbo et al., 1998).

Analysis of topoisomerase core and C-terminal domain RMSD (Fig. 1 *b*) indicate that the latter, which contains the essential catalytic Tyr 723, show an high stability reaching a plateau of 1.5 E already after 20 ps. The core domain shows a higher deviation indicating that the overall enzyme deviation observed in Fig. 1a is mainly due to its contribution.

Direct protein-DNA hydrogen bonds

Fig. 2 *a* shows a schematic representation of the direct protein-DNA hydrogen bonds present for more than 70% of the simulation time. Twenty-two direct hydrogen bonds are formed between the DNA intact strand and the enzyme; 15 between the DNA scissile strand upstream of the cleavage site, and only 4 downstream. These results confirm the trend observed in the x-ray diffraction study showing a tight interaction between protein and both DNA intact strand and the upstream portion of the scissile strand, and indicating the free or controlled rotation hypothesis (Stewart et al., 1998; Pommier et al., 1998) as the most probable mechanism for the DNA supercoiling relaxation event. The number of interactions detected by MD simulations is higher than those observed in the crystal structure and some of them, absent in the covalent DNA-topoisomerase complex,

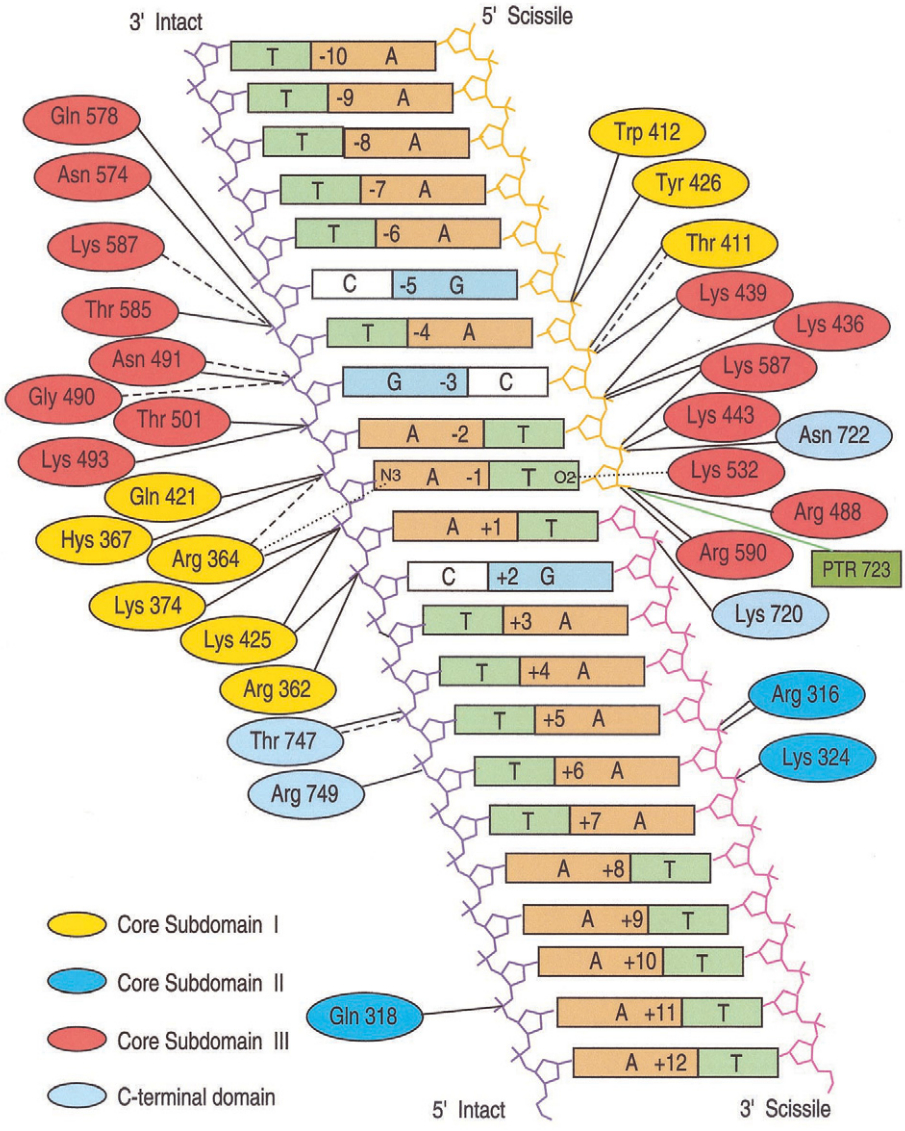

FIGURE 2A Schematic representation of the protein-DNA direct (*a*) and water-mediated (*b*) interactions present for more than 70% of the simulation time. *Full line*, protein side chain- DNA phosphate group; *dashed line*, protein main chain–DNA phosphate group; *dotted line*, protein side chain-DNA base atoms; *dash-dotted line*, protein main chain-DNA base atoms. The phosphotyrosine bond is represented by a full green line. Different colors refer to different protein domains. (*b*) Thick and thin lines represent long life and fast exchanging waters, respectively.

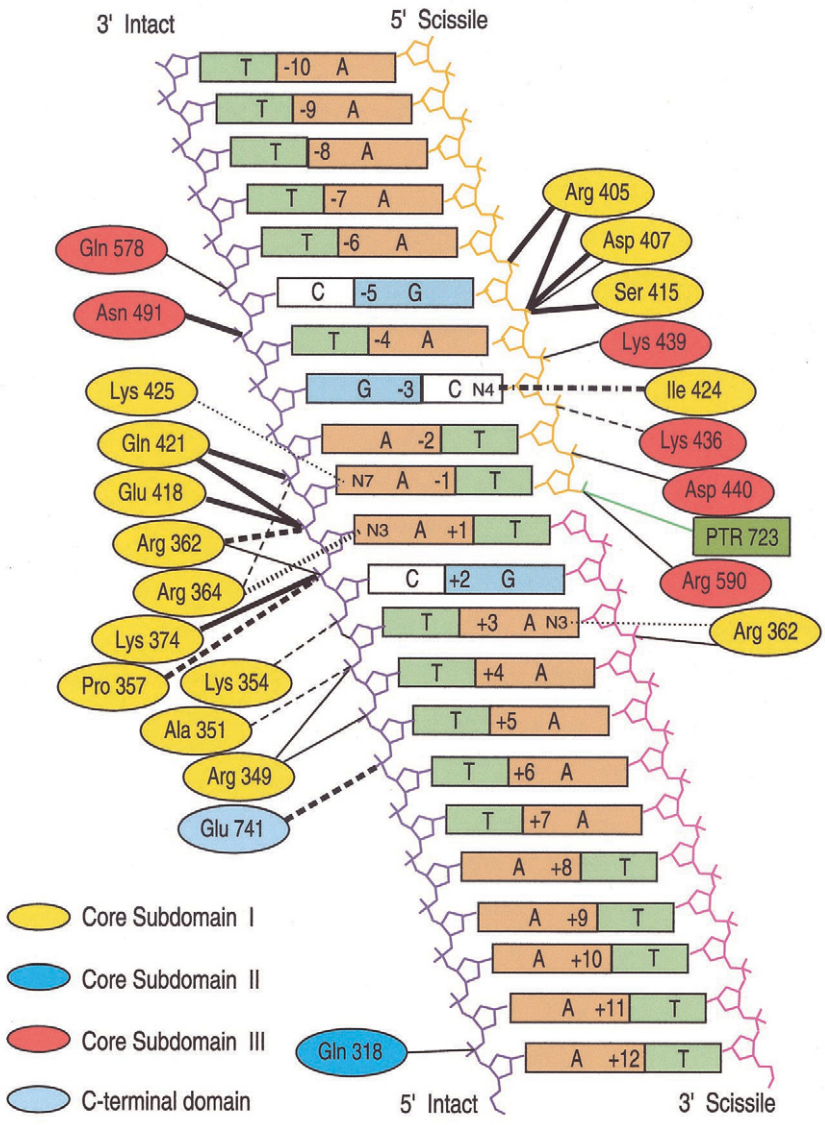

FIGURE 2B

are present in the non-covalent one, indicating that MD is sampling configurations that have been observed in similar but different experimental conditions.

In the upstream strand all of the direct contacts observed in the covalent crystal structure are conserved in the MD simulation, including the two stable hydrogen bonds connecting Arg488 and Arg590 with the phosphate group of the catalytic Tyr723 (named PTR723 in the phosphate derivative of the covalent complex). Five protein side chain-DNA direct contacts, not detected in the x-ray diffraction study of the covalent complex, are observed in the MD trajectory. One of them involves a residue (Lys587) in which the MD simulation is contacting both the −1 and −2 thymine bases whereas only the contact with the −2 thymine base was described in the diffraction study; two have been described to occur in the crystal structure of the non-covalent complex (Lys436 and Asn722 with the −2 and −1 thymine bases, respectively) whereas the remaining two contacts involve a residue (Lys439) not contacting DNA in the crystal structures.

In the downstream strand, three contacts not observed in the crystal structure (involving Arg316 and Lys324 of the sub domain II) are observed in the MD simulation whereas the contact between His266 and the +4 adenine DNA base, detected in the crystal, is lost. In actuality, this last residue is still forming a hydrogen bond with DNA but only for ~60% of the simulation time, i.e., for a time shorter than the chosen cutoff.

In the intact strand, 10 new contacts were found during the simulation. Particularly noteworthy are Asn491 and Thr747, in which the x-ray structures contact the DNA

through their side chains, and also interact with DNA with their main chains in the MD simulation. This is opposed to Lys425, in which the x-ray structure contacts DNA through a single interaction with the −1 adenine, but interacts with both the −1 adenine and the +1 adenine bases in the MD. Arg364, in which the x-ray structure has two DNA contacts (one of which is base-specific with the −1 adenine base), displays a third interaction between its lateral chain and the −1 adenine phosphate group, confirming its crucial role in mediating the DNA-topoisomerase interaction. The DNA contacts involving Asn491 main chain and Gln578 side chain, present in the MD simulation, are observed in the x-ray diffraction of the non-covalent complex. The remaining five contacts regard the side chains of residues (Lys493, Gln421, Arg362, Arg749, and Gln318) not contacting DNA in the crystal structures. On the contrary, four protein residues contacting DNA in the crystal (Asn745, Asp533, Ala489, and Ala586) do not form stable hydrogen bonds in the simulation, although two of them (Asp533 and Ala489) are still interacting with DNA, but the connecting hydrogen bonds are present for a low percentage of time (50 and 28%, respectively).

Many of the contacts observed in the MD simulation involve positively charged residues in which the crystals are seen to face, but not to directly interact with, the negatively charged phosphates. This result, together with the occurrence in the MD trajectory of contacts observed in the crystal structure of the non-covalent complex, suggests that the MD simulation is sampling different configurations close in energy to the starting one, indicating that the DNA-protein contacts are made by a relatively flexible surface. In line, some of the groups involved in the protein-DNA contacts maintain rotational freedom, i.e., several lateral chains rotate quickly, continuously changing their hydrogen atom involved in a non-covalent interaction with DNA. The same behavior is shown by the DNA phosphate oxygen atoms that quickly rotate around their backbone, giving rise to an alternate interaction between the O1P and O2P atoms and the protein donor atom. Such a property contributes to confer to the complex the high degree of flexibility already inferred by the crystallisation of the complex in multiple non-isomorphous structures (Redinbo et al., 1999).

Water-mediated protein-DNA hydrogen bonds

Fig. 2 *b* shows the protein-DNA water-mediated contacts present for >70% of the simulation time, either mediated by a single long lasting water molecule (long life waters) or through many transient water molecules characterised by short residence times (fast exchanging waters). MD simulation identifies 18 water-mediated hydrogen bonds between the protein and the DNA intact strand, 10 water-mediated hydrogen bonds between the DNA scissile strand upstream of the cleavage site, and only two downstream. They are nearly as numerous as the direct protein-DNA bonds, in line with the crucial role attributed to water in mediating sequence specific recognition or enhancing the affinity for DNA (Janin, 1999; Jones et al., 1999). Analysis of PDB structure 1a31 permitted us to identify a number of water-mediated contacts similar to that obtained by the MD analysis, being 20, 12, and 1 for the intact strand and the upstream and downstream portion of the scissile DNA strand, respectively. The large number of water molecule-mediated contacts observed in the upstream and intact strand may contribute to furnish the enzyme with an adaptable contact surface, indispensable for the binding of different DNA sequences. Human topoisomerase is, in fact, known to display only slight preference for particular DNA sequences (Capranico and Binaschi, 1998), as confirmed by the presence of only two base-specific direct topoisomerase-DNA hydrogen bonds observed both in x-ray diffraction (Redinbo et al., 1998) and in our simulation (Fig. 2 *a*). In agreement the MD trajectory indicates the occurrence of only four water-mediated specific contacts of 30, the remaining ones occurring between the enzyme and phosphate backbone DNA atoms.

In detail, the number of water-mediated hydrogen bonds (10 contacts) in the upstream portion of the scissile strand is comparable to the direct bonds (15 contacts) previously described, and they include both long life and fast exchanging water molecules. In particular, Lys439, Lys436, Arg590, Asp440, and Asp407 interact with DNA through fast exchanging water molecules, whereas Arg405, Asp407, Ser415, and Ile424 interact through long life water molecules. Among these latter contacts, the one between Ile424 and the N4 atom of the −3 cytosine base is base-specific. This is the only base-specific contact involving a protein main chain atom. Moreover, some water molecules are involved in more than one interaction, such as the long lasting interaction shared by Arg405 and Asp407, bridging the −4 adenine phosphate group. It is interesting that the identical base-specific water contact made by Ile424 can be detected analyzing the 1a31 PDB structure, as it is the contact mediated by the water shared by Arg405 and Asp407 that bridges the −4 adenine phosphate group. In actuality 5 of the 8 residues contacting DNA through water molecules in the crystal display a contact also in the MD simulation indicating that this technique provides a clear and reliable atomic description of the structure and dynamics of water molecules around a macromolecule.

In the intact strand, MD simulation indicates that only two contacts over the 18 water-mediated contacts are base-specific and again the interactions include both long life and fast exchanging water molecules. In details, Asn491, Glu418, Gln421, Arg362, Lys374, Arg364, Pro357, and Glu741 display long life water-mediated contacts, whereas the other amino acid residues represented in Fig. 2 *b* interact with DNA through different short living water molecules which, however, provide an interaction for >700 ps. Some

water molecules are involved in more than one interaction. For example, the side chains of Glu418 and Gln421 share a long lasting water molecule (>800 ps) interacting with the −1 adenine phosphate group. The same occurs for the lateral chain of Lys374 and the main chain of Pro357, both of which share a water molecule contacting the +1 adenine phosphate group. An interesting role is played by Arg362, which shows two water-mediated contacts with the intact strand and two with the scissile strand. Many of these water-mediated contacts are also observed in the x-ray structure. Indeed of the 16 residues identified by x-ray diffraction to contact DNA through water molecules, only three are not observed by MD simulation, again confirming the power of this method in describing the water-macromolecule interaction at the atomic level. Moreover, MD simulation adds a dynamical description to the static view provided by x-ray diffraction. In fact, crystallographic waters are observed either in correspondence with the long life water-mediated contacts made by Asp491, Gln421, Arg362, Lys374, and Pro357, or for the base-specific contact between Lys425 and the N7 atom of the −1 guanine base and the contacts between Arg364, Lys354, Ala351, and the DNA phosphate groups, which MD simulation describes as fast exchanging water-mediated contacts. These results indicate that the sites where water molecules are observed by x-ray diffraction correspond to sites where water molecules have a high probability to be found but they can be in fast exchange with the solvent, a property that can be detected by MD simulation.

The downstream portion of the scissile strand is characterized by only two water-mediated contacts made by Arg362 (one of which base-specific). In actuality, Arg362 is able to bridge, through water molecules, either the N3 atom of the +3 adenine base and the +4 adenine phosphate group of the downstream scissile strand (this latter interaction being also present in the crystallographic structure), or the +1 and −1 adenine phosphate groups of the intact strand (Fig. 2 *b*), besides having a direct hydrogen bond with the +1 adenine phosphate group (Fig. 2 *a*). An example of these interactions, taken from a MD frame, is shown in Fig. 3. The contacts between the +3 and +4 adenine bases of the scissile strand are mediated during the simulation by 13 and 27 different water molecules, respectively, which bridge the DNA for >800 ps, although each one is fast exchanging with the bulk. This fast exchange rate confers a potential flexibility to the DNA-protein surface, suggesting that Arg362 may participate either in the protein-DNA recognition or in the modulation of the rotation process. These findings provide structural evidence of the mutagenesis experiments that indicate an essential role for Arg362 in the cleavage/ligation phase (Li et al., 1997). Moreover it has been shown that mutation of the residues next to Arg362 (Phe361 and Gly363) produce enzymes resistant to the antitumor drug camptothecin (CPT) (Benedetti et al., 1993; Rubin et al., 1994; Li et al., 1997; Fiorani et al., 1999; Bailly et al., 1999). It has been also proposed that the interactions between Arg362, Arg364, and DNA could play a role in the correct positioning of the catalytic tyrosine in the active site (Fiorani et al., 1999). Our simulation indicates that Arg364 (core subdomain I) lies in the center of a complex network of interactions shown in Fig. 4, composed of three direct hydrogen bonds with DNA intact strand, two water-mediated hydrogen bonds with the +1 adenine N3 base atom and the −2 adenine phosphate group, and two stable hydrogen bonds with the carboxyl lateral chain of Asp533 (core subdomain III), the mutation of which gives rise to a CPT resistant enzyme (Tamura et al., 1991). This latter residue is next to Lys532, which mediates the other specific contact with the thymine scissile base covalently bound to the active Tyr723 residue. Moreover, it must be noticed that Arg362 and Arg364 belong to the protein region constituting one of the two protein lips (lip1, Core subdomain I 361–369; lip2, core subdomain III 491–501) that close the clamp around the DNA. The two lips directly interact through a salt bridge, as in the crystal structure, between Lys369 and Glu497 present for 89% of the simulation time and a direct hydrogen bond between Asp366 and Asn500, present for 82% of the simulation time. It is likely that this complex network of hydrogen bonds, which connects residues belonging to different domains, is needed to orient correctly the DNA-topoisomerase complex and that the modulation

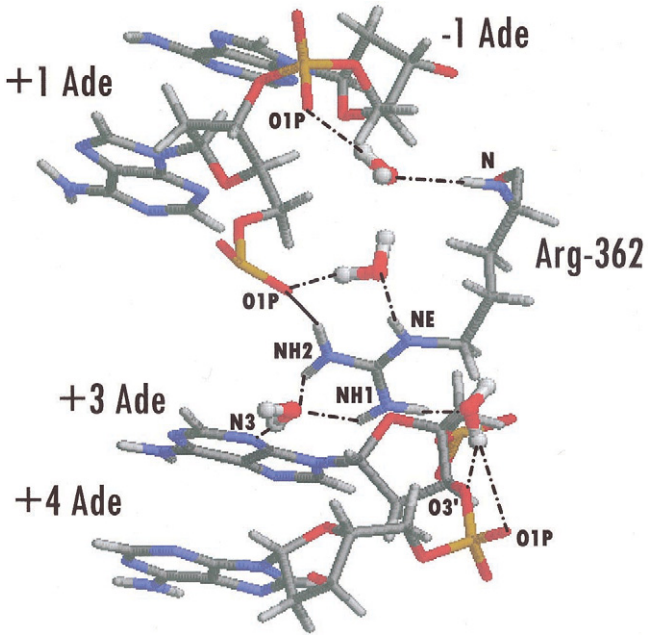

FIGURE 3 Network of direct and water-mediated interactions between Arg362 and DNA. The direct hydrogen bond with the +1 adenine phosphate group is represented by a *full line*; the water-mediated contacts occurring with the downstream scissile strand (N3 atom of the +3 adenine base and the phosphate group of the +4 adenine base) and the intact strand (phosphate groups of the +1 and −1 adenine bases) are represented by *dash-dotted lines*. Water molecules are shown in a ball and stick representation.

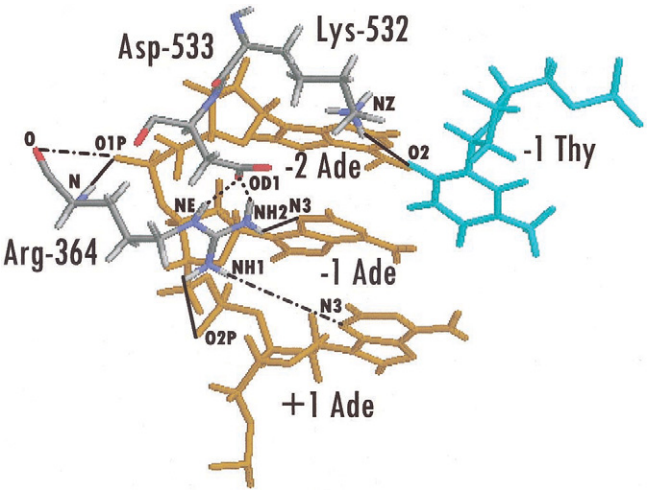

FIGURE 4 Network of direct and water-mediated interactions between Arg364 and DNA. The three direct hydrogen bonds with the DNA-intact strand (N3 atom of the −1 adenine base and the phosphate groups of the −1 and −2 adenine bases) are represented by *full lines*; the two water-mediated hydrogen bonds with the DNA-intact strand (N3 atom of the +1 adenine base and phosphate groups of −2 adenine base) are represented by *dash-dotted lines*; the two direct hydrogen bonds with Asp533 (OD1 atom) are represented by *dotted lines*. In the figure is also represented the direct hydrogen bond between Lys532 and the −1 thymine base (O2 atom) by a *full line*. The three bases of the DNA intact strand and the scissile base are represented in *yellow* and *cyan*, respectively.

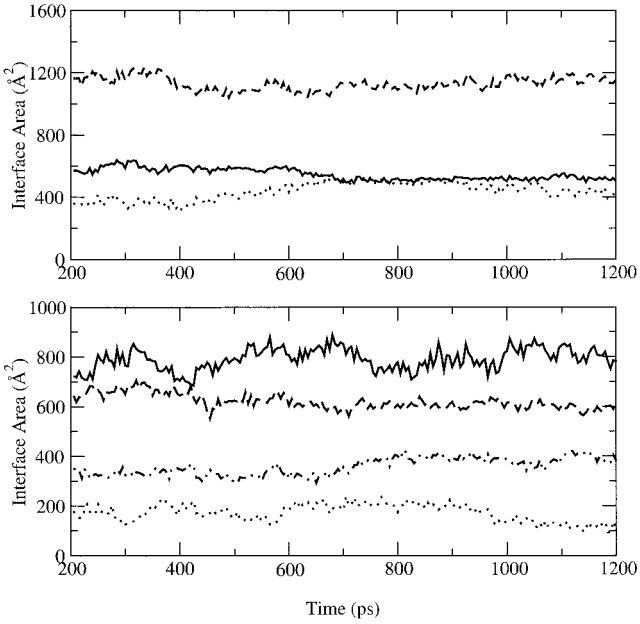

FIGURE 5 Time history of the accessible surface area buried upon complex formation (interface area). (*a*) Intact DNA strand (*dashed line*), DNA upstream strand (*full line*) and DNA downstream strand (*dotted line*). (*b*) Core subdomain I (*full line*), core subdomain II (*dotted line*), core subdomain III (*dashed line*), and C-terminal domain (*dashed-dotted line*).

of the strength of these interactions may play a role in the correct binding, cleavage, rotation, religation, and release of the enzyme.

Protein-DNA interface

Analysis of the accessible surface area buried upon complex formation (interface area) is shown in Fig. 5 *a* as a function of time for upstream, downstream and intact DNA strand. The intact strand has the widest interface area, which is fairly constant over the trajectory, displaying an average value of 1129 Å^2. The interface area of the upstream portion of scissile DNA strand is also quite constant over the trajectory, having an average value of 547 Å^2. On the other hand the interface area of the downstream portion of scissile DNA strand shows a large variation during the simulation, going from a minimum of 308 to a maximum of 541 Å^2. This result confirms that the intact strand and the upstream portion of the scissile strand are tightly bound to the protein whereas the downstream portion shows a lower and more variable interface area and displays a higher flexibility being able to sample a variable number of configurations. This finding is in agreement with the uneven distribution of both direct and water-mediated DNA-protein contacts and with the proposal of a controlled rotation mechanism.

The protein interface area as a function of time is represented in Fig. 5 *b* for the three core subdomains and for the C-terminal domain. The average interface area is 790, 172, 620, and 359 Å^2 for subdomain I, II, III, and C-terminal domains, respectively. These values are all greater than those obtained from x-ray diffraction (654 Å^2 core subdomain I; 40 Å^2 core subdomain II; 592 Å^2 core subdomain III; 221 Å^2 C-terminal domain) in line with the evidence that during the simulation polar and charged residues, not directly contacting DNA in the crystal structure, are now observed to reach a contact distance.

Fluctuations

The RMS fluctuations (RMSF) of each residue belonging to the core and C-terminal domain are reported in Fig. 6 *a* and *b*, respectively. The core domain displays regions characterized by large differences in fluctuation, their values ranging from 1 to 6 Å. The regions having the lowest fluctuations display a large number of contact with DNA, such as the two lip regions. Another minimum is found for the β_{11} region (418–427), which interacts with the major groove of the DNA through numerous direct and water-mediated hydrogen bonds (Fig. 2). On the other hand, residues 434–453 and 455–466 in core subdomain III, belonging to α_8 and α_9 respectively, which were proposed to play a role in the mechanism of opening and closure of the enzyme around DNA, show high fluctuations with average values around 4.5 and 3.5 Å for α_8 and α_9, respectively. The C-terminal domain has an average fluctuation higher than the core domain (Fig. 6 *b*), but also in this case the minimum

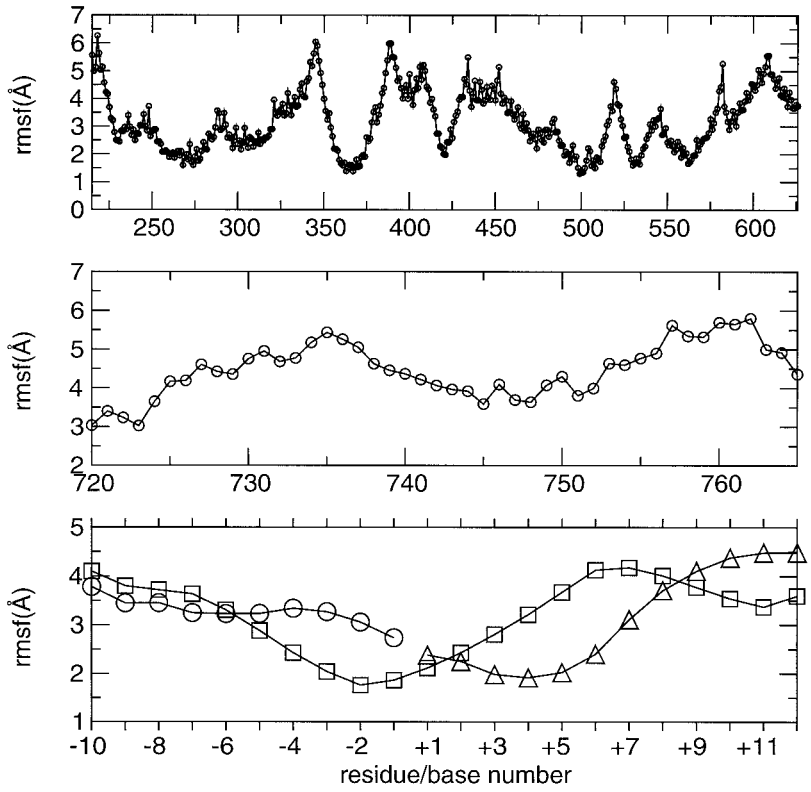

FIGURE 6 Root-mean-square per residue or per base fluctuations of core domain (*a*), C-terminal domain (*b*), and DNA double strand (*c*). In panel c the bases of the intact, scissile upstream, and downstream DNA strands are represented by *squares*, *circles* and *triangles*, respectively.

corresponds to the only regions (720–723; 741–751) that contact DNA.

Fig. 6 *c* shows the DNA per base average fluctuations. The region with lower fluctuations lies in the intact strand between the +1 and the −2 adenine bases, in which numerous direct or water-mediated protein-DNA contacts are present (Fig. 2). More surprisingly, also the +3 and +4 adenine bases in the scissile strand show low fluctuations even though they do not have any direct stable protein-DNA hydrogen bonds. As we have already described, these bases have water-mediated hydrogen bonds with Arg362, which has water-mediated hydrogen bonds also with the +1 and −1 adenine bases of the intact strand. The observation that the DNA bases having the lowest fluctuations are bridged by Arg362 confirms that this residue plays a crucial role and leads us to the hypothesis that it can represent a sensor that transduces the information between the active site and the downstream portion of the scissile strand. This hypothesis is supported by the network of covalent and non-covalent interactions between Arg362, Arg364, the −1 adenine and thymine bases, Asp533, Lys532, and Tyr723 (Figs. 3 and 4) and it explains the strict conservation of Arg362 and Arg364 over the evolution of this enzyme. Moreover, the occurrence of CPT-resistant enzymes upon mutation of Phe361 and Gly363 suggests that Arg362 may play also an important role in mediating the drug interaction (Benedetti et al., 1993; Rubin et al., 1994; Li et al., 1997; Fiorani et al., 1999; Bailly et al., 1999).

Long life water molecules around the active site

Restriction of our analysis to water molecules having a maximum residence time >800 ps. identifies only 17 long life water molecules in the topoisomerase-DNA system, and four of them are localized in the proximity of the active site indicating that this region behaves as a water attractor. In this region two water molecules form an hydrogen bond with Arg488, one with Arg590, and the fourth one is hydrogen bonded with the phosphate group belonging to the Tyr723 derivative (Fig. 7). This last water molecule is in the appropriate position to accept a proton from the active Tyr723 residue supporting the hypothesis that a water close to the active site could function as a catalytic water (Redinbo et al., 2000).

Correlation between MD results and site-directed mutagenesis experiments

Analysis of the interactions observed in the MD trajectory helps us in interpreting some experimental results observed upon single residue mutations which cause a reduction of

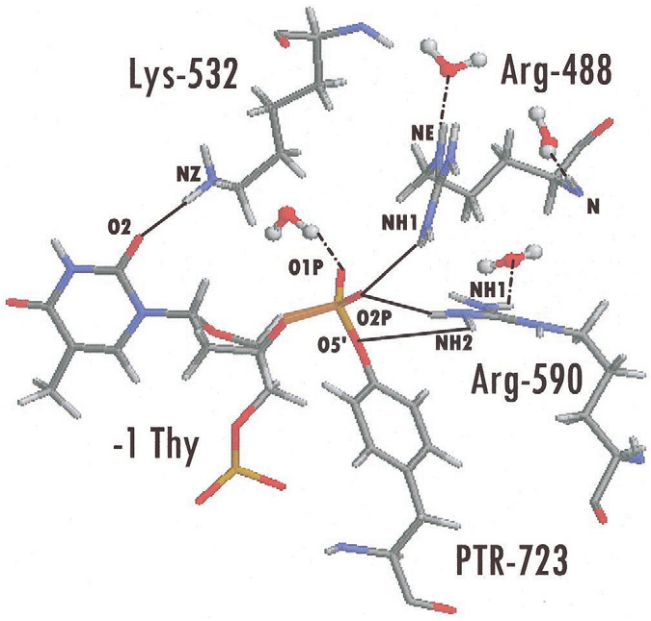

FIGURE 7 Long life water molecules present in the active site region. The four water molecules which are making a hydrogen bond with atoms close to the active site for >80% of the total simulation time are shown in stick and ball representation. The hydrogen bonds occurring between the water molecules and the protein atoms are represented by *dash-dotted lines*. *Full lines* represent the direct hydrogen bond between Lys532 and the −1 thymine base (O_2 atom), the direct hydrogen bond between Arg488 and the phosphotyrosine PTR723 (phosphate group) and the two direct hydrogen bonds between Arg590 and the phosphotyrosine PTR723 (phosphate group).

binding affinity and activity or the production of drug resistant enzymes.

In the C-terminal domain, besides the catalytic Tyr723 residue, there are other residues that are conserved and whose mutation is able to alter the enzymatic activity or sensitivity to different drugs (Pommier et al., 1998). Mutation of yeast Asn726 (corresponding to Asn722 in human topoisomerase) in Leu results in a decrease of enzyme activity, whereas its mutation in Ala does not modify topoisomerase activity, but strongly enhances its resistance to CPT and other topoisomerase poisons (Knab et al., 1993, 1995; Hann et al., 1998b). Mutations of this residue in Ser, Asp or His interfere with both enzyme activity and drug sensitivity (Leteurtre et al., 1994, Fertala et al., 2000). Crystallographic data indicate that in the upstream DNA portion of the covalent complex, Asn722 forms a direct hydrogen bond with DNA, whereas Lys720 makes a direct hydrogen bond with DNA in the downstream portion of the non-covalent complex. In our MD trajectory, we find stable direct hydrogen bonds between DNA and both Lys720 and Asn722, confirming that this region has a fundamental role in the interaction with DNA, most likely in the orientation of the active site residue Tyr723, thus explaining the modulation of the drug binding. Mutation of Thr729 to Ile or Ala gives rise to CPT resistance (Kubota et al., 1992; Wang et al., 1997) and in the latter case also to reduction of enzyme activity, however x-ray diffraction does not display any direct interaction between Thr729 and DNA. MD simulation allows one to hypothesize a structural mechanism for such an effect because a long life water molecule, present for >84% of the simulation time, strongly bridges this residue to Asp725 via three hydrogen bonds with the main chain (NH group) and the lateral chain of Thr729 and with the main chain of Asp725 (CO group). In turn, Asp725 makes a direct hydrogen bond with its main chain (NH group) and the lateral chain of Ser595 for >95% of the simulation time. The lateral chain of Ser595 also interacts, via direct hydrogen bond, with the main chain (CO group) of the catalytic Tyr723 for >960 ps, designating a hydrogen bond connection between Thr729 and Tyr723.

In core subdomain III, mutation in yeast of Gly369, corresponding to human topoisomerase Gly437, to Asp diminishes DNA binding and catalytic activity (Hann et al., 1998a). Gly437 is located in the upper portion of α_8 and is conserved in this position in 13 of 14 cellular topoisomerase I's with known sequences (Redinbo et al., 2000). Analysis of MD trajectory indicates that Lys436 and Lys439 can have direct contact with the upstream portion of DNA, providing a structural explanation for the mutational effect observed at the level of residue 437. Mutation of Asp533 to Gly gives rise to a CPT-resistant enzyme (Tamura et al., 1991). Our trajectory indicates that Asp533 is involved in a complex hydrogen bond network that includes Arg364, Arg362, and Lys532 (Fig. 4). In particular, Asp533 makes two very stable hydrogen bonds through its carboxylic lateral chain with the lateral chain of Arg364, which, in its turn, makes direct hydrogen bonds with the DNA intact strand. Mutation to Gly of Arg590, which constitutes the catalytic active site with Arg488 and Tyr723, produces a 1000-fold less active enzyme (Megonigal et al., 1997). In line MD results show that this residue is involved in numerous direct and water-mediated contacts (Figs. 2 and 7).

In core subdomain I the region 361–365 is highly conserved among eukaryotic species and is homologous to sequences found in other DNA-binding proteins (Li et al., 1997). In human topoisomerase, mutation of Phe361 to Ser gives rise to a CPT and rebeccamycin-derivatives resistant enzyme (Rubin et al., 1994; Li et al., 1997; Bailly et al., 1999), whereas mutation of Gly363 to Cys, Ser or Val directly affects the formation of the CPT binding site and enzyme catalysis (Benedetti et al., 1993; Fiorani et al., 1999). Moreover, mutations of Arg362 and Arg364 gave convincing evidence that the interactions of 361–364 region with DNA are required for DNA cleavage/ligation (Li et al., 1997). These results may be explained by our MD simulation which attributes a crucial role to Arg362, involved in four water-mediated hydrogen bonds, two with the intact strand and two with the downstream region (Figs. 2 and 3), and to Arg364, involved in various direct and water-medi-

ated hydrogen bond contacts (Figs. 2 and 4). For this latter residue a crucial role has also been indicated by x-ray diffraction (Redinbo et al., 1998).

CONCLUSION

This study makes use of MD simulation to extend the analysis of the interaction of the DNA-topoisomerase complex obtained through x-ray diffraction to a description which takes into consideration the pattern of hydration around the topoisomerase-DNA interface. Water is now recognized to play a crucial role in mediating protein-DNA interaction, and MD simulation has been successfully used to investigate the hydration pattern both in protein-DNA complexes (Billeter et al., 1996; Tsui et al., 2000) and in single proteins or DNA sequences (Luise et al., 2000; Castrignanò et al., 2000). This work identifies a number of water-mediated hydrogen bonds comparable to the direct ones, as found in many high resolution structures (Schwabe, 1997; Janin, 1999), indicating that water plays a crucial role in mediating affinity and specificity. The identification of water-mediated interaction allows us to interpret experimental data from site directed mutagenesis reporting reduced protein-DNA affinity or the production of a drug resistant enzyme. These results can not be easily explained taking into consideration only the direct interaction described in the x-ray diffraction studies (Redinbo et al., 1998; Stewart et al., 1998; Redinbo et al., 2000). Moreover, the simulation identifies the presence of four long life water molecules confined in the proximity of the active site, indicating that this region acts as a water-attractor. Such a finding is in agreement with the results of a recent x-ray diffraction study which identifies a single water molecule proximal to the active site residue Tyr723 and gives support to the hypothesis made by the crystallographers that this water is able to activate the catalytic Tyr723 (Redinbo et al., 2000).

We acknowledge the contribution of O. Incani in the early stages of this project. We thank P. Benedetti, G. Capranico, P. D'Angelo, M. Ceruso and N. Sanna for helpful suggestions. We also acknowledge the CASPUR Computational Center for the computer architecture used in this work. This work was partly supported by a Ministero dell'Università e Ricerca Scientifica e Tecnologica COFIN2000 project.

REFERENCES

Abseher, R., H. Schreiber, and O. Steinhauser. 1996. The influence of a protein on water dynamics in its vicinity investigated by molecular dynamics simulation. *Proteins.* 25:366–378.

Bailly, C., C. Carrasco, F. Hamy, H. Vezin, M. Prudhomme, A. Saleem, and E. Rubin. 1999. The camptothecin-resistant topoisomerase I mutant F361S is cross-resistant to antitumor rebeccamycin derivatives: a model for topoisomerase I inhibition by indolocarbazoles. *Biochemistry.* 38: 8605–8611.

Benedetti, P., P. Fiorani, L. Capuani, and J. C. Wang. 1993. Camptothecin resistance from a single mutation changing glycine 363 of human DNA topoisomerase I to cysteine. *Cancer Res.* 53:4343–4348.

Berendsen, H. J. C., J. P. M. Postma, W. F. van Gusteren, A. Di Nola, and J. R. Haak. 1984. Molecular dynamics with coupling to an external bath. *J. Comput. Phys.* 81:3684–3690.

Billeter, M., P. Guntert, P. Luginbuhl, and K. Wuthrich. 1996. Hydration and DNA recognition by homeodomains. *Cell.* 85:1057–1065.

Bonvin, A. M., M. Sunnerhagen, G. Otting, and W. F. van Gunsteren. 1998. Water molecules in DNA recognition: a molecular dynamics view of the structure and hydration of the trp operator. *J. Mol. Biol.* 282: 859–873.

Burling, F. T., W. I. Weis, K. M. Flaherty, and A. T. Brunger. 1996. Direct observation of protein solvation and discrete disorder with experimental crystallographic phases. *Science.* 271:72–77.

Capranico, G., and M. Binaschi. 1998. DNA sequence selectivity of topoisomerases and topoisomerase poisons. *Biochim. Biophys. Acta.* 1400:185–194.

Castrignanò. , T., G. *Chillemi and A. Desideri.* 2000: Structure and hydration of BamHI DNA recognition site: a molecular dynamics investigation *Biophys. J.* 79:1263–1272.

Cheatham, T. E., J. L. Miller, T. Fox, T. A. Darden, and P. A. Kollman. 1995. Molecular dynamics simulation on solvated biomolecular systems: the particle mesh Ewald method leads to stable trajectories of DNA, RNA and proteins. *J. Am. Chem. Soc.* 117:4193–4194.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, M. Kenneth, J. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kolman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald-an N.log(n) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.

Fertala, J., J. R. Vance, P. Pourquier, Y. Pommier, and M-A. Bjornsti. 2000. Substitutions of Asn-726 in the active site of yeast DNA topoisomerase I define novel mechanisms of stabilizing the covalent enzyme-DNA intermediate. *J. Biol. Chem.* 275:15246–15253.

Fiorani, P., J. F. Amatruda, A. Silvestri, R. H. Butler, M-A. Bjornsti, and P. Benedetti. 1999. Domain interactions affecting human DNA topoisomerase I catalysis and camptothecin sensitivity. *Mol. Pharmacol.* 56:1105–1115.

Hann, C., A. L. Carlberg, and M-A. Bjornsti. 1998a. Intragenic suppressors of mutant DNA topoisomerase I-induced lethality diminish enzyme binding of DNA. *J. Biol. Chem.* 273:31519–31527.

Hann, C., D. L. Evans, J. Fertala, P. Benedetti, M-A. Bjornsti, and D. J. Hall. 1998b. Increased camptothecin toxicity induced in mammalian cells expressing Saccharomyces cerevisiae DNA topoisomerase I. *J. Biol. Chem.* 273:8425–8433.

Hubbard, S. J., and J. M. Thornton. 1993. NACCESS, Computer Program. Department of Biochemistry and Molecular Biology, University College, London.

Janin, J. 1999. Wet and dry interfaces: the role of solvent in protein-protein and protein-DNA recognition. *Structure.* 7:277–279.

Jones, S., P. van Heyningen, H. M. Berman, and J. M. Thornton. 1999. Protein-DNA interactions: a structural analysis. *J. Mol. Biol.* 287: 877–896.

Jorgensen, W. L. 1981. Transferable intermolecular potential functions for water alcohols and ethers: application to liquid waters. *J. Am. Chem. Soc.* 103:335–340.

Knab, A. M., J. Fertala, and M-A. Bjornsti. 1993. Mechanisms of camptothecin resistance in yeast DNA topoisomerase I mutants *J. Biol. Chem.* 268:22322.

Knab, A. M., J. Fertala, and M-A. Bjornsti. 1995. A camptothecin-resistant dna topoisomerase I mutant exhibits alteres sensitivies to other DNA topoisomerase poisons. *J. Biol. Chem.* 270:6141–6148.

Kossiakoff, A. A., M. D. Stintchak, J. Shpunginand, and L. G. Presta. 1992. Analysis of solvent structure in proteins using neutron D2O H_2O solvent

maps: pattern of primary and secondary hydration of tripsin. *Proteins.* 12:223–236.

Kubota, N., F. Kanzawa, K. Nishio, Y. Takeda, T. Ohmori, Y. Fujiwara, Y. Terashimaand, and N. Saijo. 1992. Detection of topoisomerase I gene point mutation in CPT-11 resistant lung cancer cell line. *Biochem. Biophys. Res. Commun.* 188::571.

Leteurtre, F., A. Fujimori, A. Tanizawa, A. Chhabra, A. Mazumder, G. Kohlhagen, H. Nakano, and Y. Pommier. 1994. Saintopin, a dual inhibitor of DNA topoisomerase I and II, as a probe for drug-enzyme interactions. *J. Biol. Chem.* 269:28702–28707.

Li, X. G., P. Haluska, Y-H. Hsiang, A. K. Bharti, D. W. Kufe, L. F. Liu, and E. H. Rubin. 1997. Involvement of amino acids 361 to 364 of human topoisomerase I in camptothecin resistance and enzyme catalysis. *Biochem. Pharmacol.* 53:1019.

Luise, A., M. Falconi, and A. Desideri. 2000. Molecular dynamics simulation of solvated azurin: correlation between surface solvent accessibility and water residence times. *Proteins.* 39:56–67.

Megonigal, M. D., J. Fertala, and M. A. Bjornsti. 1997. Alterations in the catalytic activity of yeast DNA topoisomerase I result in cell cycle arrest and cell death. *J. Biol. Chem.* 272:12801–12808.

Melacini, G., A. M. Bonvin, M. Goodman, R. Boelens, and R. Kaptein. 2000. Hydration dynamics of the collagen triple helix by NMR. *J. Mol. Biol.* 300:1041–1048.

Otting, G., E. Liepinish, and K. Wutrick. 1991. Protein hydration in aqueous solution. *Science.* 254:974–980.

Otwinoski, Z., R. W. Schevitz, R. G. Zhang, C. L. Lawson, R. Q. Marmorstein, B. F. Luisi, and P. B. Sigler. 1988. Crystal structure of TRP repressor/operator complex at atomic resolution. *Nature.* 335:321–329.

Pommier, Y., P. Pourquier, Y. Fan, and D. Strumberg. 1998. Mechanism of action of eukaryotic DNA topoisomerase I and drugs targeted to the enzyme. *Biochim. Biophys. Acta.* 1400:83–106.

Redinbo, M. R., J. J. Champoux, and W. G. J. Hol. 2000. Novel insight into catalytic mechanism from a crystal structure of human topoisomerase I in complex with DNA. *Biochemistry.* 39:6832–6840.

Redinbo, M. R., L. Stewart, P. Kuhn, J. J. Champoux, and W. G. J. Hol. 1998. Crystal structures of human topoisomerase I in covalent and noncovalent complexes with DNA. *Science.* 279:1504–1513.

Redinbo, M. R., L. Stewart, J. J. Champoux, and W. G. J. Hol. 1999. Structural flexibility in human topoisomerase I revealed in multiple non-isomorphous crystal structures. *J. Mol. Biol.* 292:685–696.

Reyes, C. M., and P. A. Kollman. 1999. Molecular dynamics study of U1A-RNA complexes. *RNA.* 5:235–244. .

Rubin, E., P. Pantazis, A. Bharti, D. Toppmeyer, B. Giovannella, and D. Kufe. 1994. Identification of a mutant topoisomerase I with intact catalytic activity and resistance to 9-nitro-camptothecin. *J. Biol. Chem.* 269:2433.

Ryckaert, J.-P., G. Ciccotti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327–341.

Schwabe, J. W. R. 1997. The role of water in protein-DNA interactions. *Curr. Opin. Struct. Biol.* 7:126–134.

Sen, S., and L. Nilsson. 1999. Structure, interaction, dynamics and solvent effects on the DNA-*Eco*RI complex in aqueous solution from molecular dynamics simulation. *Biophys. J.* 77:1782–1800.

Stewart, L., M. R. Redinbo, X. Qiu, W. G. J. Hol, and J. J. Champoux. 1998. A model for the mechanism of human topoisomerase I. *Science.* 279:1534–1541.

Tamura, H., C. Kohchi, R. Yamada, T. Ikeda, O. Koiwai, E. Patterson, J. D. Keene, K. Okada, E. Kjeldsen, K. Nishikawa et al. 1991. Molecular cloing of a cDNA of a camptothecin-resistant human DNA topoisomerase I and identification of mutation sites. *Nucleic Acids Res.* 19:69.

Tang, Y., and L. Nilsson. 1999. Molecular dynamics simulations of the complex between human U1A protein and hairpin II of U1 small nuclear RNA and of free RNA in solution. *Biophys. J.* 77:1284–1305.

Tsui, V., I. Radhakrishnan, P. E. Wright, and D. A. Case. 2000. NMR and molecular dynamics studies of the hydration of a zinc finger-DNA complex. *J. Mol. Biol.* 302:1101–1117.

Wang, L. F., C. Y. Ting, C. K. Lo, J. S. Su, L. A. Mickley, and A. T. Fojo. 1997. Identification of mutations at DNA topoisomerase I responsible for camptothecin resistance. *Cancer Res.* 57:1516–1522.

Young, A. Y., G. Ravishanker, and D. L. Beveridge. 1997. A 5-nanosecond molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *Biophys. J.* 73:2313–2336.

Zanotti, J. M., M. C. Bellisent-Funel, and J. Parello. 1999. Hydration-coupled dynamics in proteins studied by neutron scattering and NMR: decays of the typical EF-hand calcium-binding parvalbumin. *Biophys. J.* 76:2390–2411.