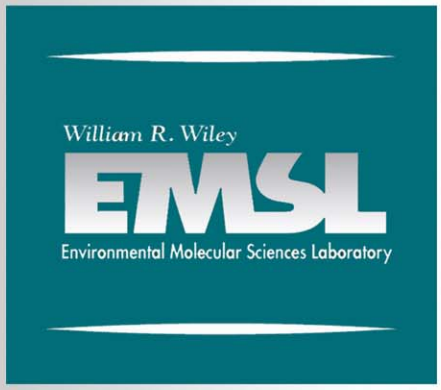


Biomolecular Simulation of Base Excision Repair and Protein Signaling

T. P. Straatsma
J. A. McCammon
J. H. Miller
P. E. Smith
E. R. Vorpagel
C. F. Wong
M. Zacharias

February 2006

This research was performed in part using the Molecular Science Computing Facility (MSCF) in the William R. Wiley Environmental Molecular Sciences Laboratory, a national scientific user facility sponsored by the U.S. Department of Energy's Office of Biological and Environmental Research and located at the Pacific Northwest National Laboratory. Pacific Northwest is operated for the Department of Energy by Battelle.

DISCLAIMER

PACIFIC NORTHWEST NATIONAL LABORATORY
operated by
BATTELLE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC05-76RL01830

PNNL-15687

Biomolecular Simulation of Base Excision Repair and Protein Signaling

T. P. Straatsma
J. A. McCammon
J. H. Miller
P. E. Smith
E. R. Vorpagel
C. CF. Wong
M. Zacharias

February 2006

Published by Pacific Northwest National Laboratory for the
Environmental Molecular Sciences Laboratory

Abstract

The goal of the Biomolecular Simulation of Base Excision Repair and Protein Signaling project is to enhance our understanding of the mechanism of human polymerase-β, one of the key enzymes in base excision repair (BER) and the cell-signaling enzymes cyclic-AMP-dependent protein kinase. This work used molecular modeling and simulation studies to specifically focus on the

- dynamics of DNA and damaged DNA
- dynamics and energetics of base flipping in DNA
- mechanism and fidelity of nucleotide insertion by BER enzyme human polymerase-β
- mechanism and inhibitor design for cyclic-AMP-dependent protein kinase.

Molecular dynamics simulations and electronic structure calculations have been performed using the computer resources at the Molecular Science Computing Facility at the Environmental Molecular Sciences Laboratory.

List of all team members

Team Leader

Name:	T. P. Straatsma
Position:	Associate Division Director, Computational Biology and Bioinformatics Computational Sciences and Mathematics Division
Institution:	Pacific Northwest National Laboratory
Address:	P.O.Box 999, MS K7-90 Richland, WA 99352
Telephone:	(509) 375-2802
Facsimile:	(509) 375-6631
Email:	tps@pnl.gov

Team Member 2

Name:	J. A. McCammon
Position:	Joseph E. Mayer Professor of Theoretical Chemistry
Institution:	University of California at San Diego Department of Chemistry and Biochemistry, and Department of Pharmacology
Address:	9500 Gilman Drive, M/C 0365 La Jolla, CA 92093-0365
Telephone:	(858) 534-2905
Facsimile:	(858) 534-7042
Email:	jmccammon@ucsd.edu

Team Member 3

Name:	J. H. Miller
Position:	Associate Professor

Institution: Washington State University – Tri-Cities
Department of Computer Science
Address: 2710 University Drive
Richland, WA 99352
Telephone: (509) 372-7232
Facsimile: (509) 372-7100
Email: jhmiller@tricity.wsu.edu

Team Member 4

Name: P. E. Smith
Position: Associate Professor
Institution: Kansas State University
Department of Biochemistry
Address: 36 Willard Hall, Kansas State University
Manhattan, KS 66506-3702
Telephone: (785) 532 5109
Facsimile: (785) 532 7278
Email: pesmith@ksu.edu

Team Member 5

Name: E. R. Vorpagel
Position: Chief Scientist
Institution: Pacific Northwest National Laboratory
MSCF Visualization and User Services Group
Address: P.O.Box 999, MS K8-91
Richland, WA 99352
Telephone: (509) 376-0751
Facsimile: (509) 376-0420
Email: erich.vorpagel@pnl.gov

Team Member 6

Name: C. F. Wong
Position: Assistant Professor
Institution: Department of Chemistry and Biochemistry
University of Missouri – St. Louis
Address: 315 Benton Hall – M203
8001 Natural Bridge Road
St. Louis, MO 63121
Telephone: (314) 516-5318
Facsimile: (314) 516-5342
Email: wongch@msx.umsl.edu

Team Member 7

Name: M. Zacharias
Position: Professor of Computational Biology
Institution: School of Engineering and Science
International University of Bremen

Address: P.O.Box 750561
28725 Bremen, Germany
Telephone: 011-49-421-200-3541
Facsimile: 011-49-421-200-3249
Email: m.zacharias@iu-bremen.de

Number of hours allocated for the past three years

FY02: 575,000 node hours
FY03: 500,000 node hours
FY04: 800,000 node hours

Number of hours actually used in the past three years

FY02: 480,000 node hours
FY03: 305,000 node hours
FY04: 606,000 node hours

Overview of the accomplishments and activities

The Biomolecular Simulation of Base Excision Repair and Protein Signaling project has used the massively parallel computing resources in the Environmental Molecular Sciences Laboratory to perform molecular modeling and simulation studies to enhance our understanding of the mechanism of human polymerase-β (pol-β), one of the key enzymes in base excision repair (BER), and the cell-signaling enzymes cyclic-AMP-dependent protein kinase. The broad goal of this project is to study the effect of environmental factors, including ionizing radiation, that contribute to continuous damage of cellular DNA. The damage resulting from oxidative stress and ionizing radiation is primarily in the form of oxidized bases, single-strand breaks, and loss of bases. These are the targets of the BER mechanism enzymes, including pol-β. Failure to correctly and timely repair these damaged DNA sites can result in cell death, carcinogenesis, or genetic diseases. Resulting mutations in cell signal transduction enzymes can lead to uncontrolled cell proliferation or differentiation. For example, mutations in Ras, the molecular switch in several growth-factor signaling pathways, have been found in about 30% of human tumors. These signaling pathways often involve a chain of protein kinases that activate or deactivate each other through phosphorylation reactions, eventually controlling the activation of transcription factors in the cell nucleus. The work on this project is focused on the

- structure and dynamics of DNA and damaged DNA
- dynamics and energetics of base flipping in DNA
- mechanism and fidelity of nucleotide insertion by BER enzyme human pol-β
- mechanism and inhibitor design for cyclic-AMP-dependent protein kinase.

The Structure and Dynamics of DNA and Damaged DNA

The objective of this component is the investigation of the effect of damage induced by oxidative stress on the structure and dynamics of DNA. Damage-induced structural and dynamical changes are very difficult to study experimentally; hence, modeling can play a crucial role to our understanding of the biochemistry of

DNA repair processes. Recent progress in the simulation of nucleic acids due to both force field improvements and a more accurate treatment of long-range electrostatic interactions allows the stable molecular dynamics (MD) simulation of DNA for up to several tens of nanoseconds. Such simulations can complement experimental studies on nucleic acids since they allow extraction of DNA structural and dynamical properties that cannot easily be obtained from experiment.

During the project duration, we finished an MD project on the effect of nicks, extra bases, and gaps in damaged DNA and a project on the effect of modified bases on the dynamics of RNA. In addition, we initiated large-scale simulation studies on a complete nucleosome structure (147 base pairs of DNA in complex with histone proteins).

Modeling the Effect of Nicks, Extra Bases, and Gaps in Damaged DNA Using MD Simulations

Damage of DNA by radiation and chemical reagents is a major cause of cancer and cell aging. Such damage can lead to changes in the DNA structure and dynamics, which in turn influence recognition and repair of damaged DNA. Common forms of DNA damage include the loss of a base (creating an abasic site), a bond break (nick) in the DNA backbone or the loss of one or several nucleotides. Surprisingly, the X-ray structure of a double-stranded nicked DNA shows only modest variation when compared to regular DNA. However, X-ray crystallography provides only a static picture of nucleic acids. It is expected that a nick in DNA or the loss of one or several nucleotides due to radiation or chemical damage (gap) will lead to increased local flexibility or a change in average solution conformation.

During the project duration, multi-nanosecond simulations on an intact 12 base pair DNA, the same sequence with a central nick, and a gap of one nucleotide at the center were performed. For all simulations, NWChem, the massively parallel software for computational chemistry developed at the Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory (PNNL), in combination with the AMBER (parm99) force field, was used. The X-ray structure was used as a starting structure for the simulations of the nicked DNA. Since no experimental structure with exactly the same sequence as the nicked DNA was available, a standard B-DNA structure was used as a reference. The starting structure for the simulations with a single nucleotide gap was obtained by removing a (T) nucleotide at the center of the nicked double-stranded DNA structure. Simulations were performed on the HP/Linux 1960 Itanium-2 processor cluster (MPP2) at the PNNL Molecular Science Computing Facility for approximately 15 nanoseconds. Each simulation included about 5000 water molecules and enough counter ions to neutralize charge.

Although the intact and the nicked DNA oligonucleotide structures showed considerable conformational fluctuations during the MD simulations, some structural features of the experimental structures were well preserved. A stable narrowing of the DNA double helix at the nick was found, which might be a specific structural signature recognized by repair enzymes. No significant changes in twisting or bending, when compared to the regular DNA oligonucleotide, were observed. The average bend angle of the DNA during the simulation was approximately 10 degrees. This result is consistent with experiments investigating global conformational flexibility of nicked and gapped DNA molecules using electric birefringence. In these studies, it was found that a single nick in DNA did not significantly change the global properties of DNA.

In the case of the simulations of gapped DNA, larger conformational changes in the DNA were observed. In the initial DNA structure, the bases that flanked the gap were in an “unstacked” state during the MD simulation and no further unstacking or diffusion of water molecules into the space between the flanking nucleotides was observed. Instead, during the simulation, the gap between the two segments of the second

strand was closed and the two nucleotides adopted a stacked conformation similar to a dinucleotide step conformation in regular DNA. Overall, the DNA adopted a bend conformation with a more accessible major groove than that for the intact or nicked DNA duplex and with a bend angle reaching 30 degrees. This result is, again, consistent with the experimental observation of a significant effect of gaps in DNA on the global conformational properties (bending) of the duplex. The change in major groove size and bending angle may serve as recognition elements for the repair of such DNA damage.

Influence of a Fluorobenzene Nucleobase Analogue on the Conformational Flexibility of RNA Studied by MD Simulations

Chemically modified bases frequently are used to stabilize nucleic acids, to study the driving forces for nucleic acid structure formation, and to tune DNA and RNA hybridization conditions. In particular, fluorobenzene and fluorobenzimidazole base analogues can act as universal bases able to pair with any natural base and to stabilize RNA duplex formation. These modified bases can act as model systems to investigate the chemical base modification on the structure and dynamics of nucleic acids. Although these base analogues are compatible with an A-form RNA geometry, little is known about the influence on the fine structure and conformational dynamics of RNA. Nanosecond MD simulations have been performed to characterize the dynamics of RNA duplexes containing a central 1'-deoxy-1'-(2,4-difluorophenyl)-ß-D-ribofuranose base pair or opposite to an adenine base. For comparison, RNA with a central uridine:adenine pair and a 1'-deoxy-1'-(phenyl)-ß-D-ribofuranose opposite to an adenine was also investigated. The MD simulations indicated a stable overall A-form geometry for the RNAs with base analogues. However, the presence of the base analogues caused a locally enhanced mobility of the central bases, inducing mainly base pair shear and opening motions. No stable "base paired" geometry was found for the base analogue pair or the base analogue:adenine pairs, which explains, in part, the universal base character of these analogues. Instead, the conformational fluctuations of the base analogues led to an enhanced accessibility of the bases in the major and minor grooves of the double-helix compared to a regular base pair. This simulation study suggests a general mechanism for the enhanced local mobility of modified bases in double stranded nucleic acids.

Molecular Dynamics Simulation of the Complete Nucleosome

During the computational grand challenge project an MD simulation of a complete nucleosome has been initiated. The simulation system consists of 147 base pairs of DNA, several histone proteins, ions and several thousand water molecules and is based on the high-resolution structure by Richmond and Davey. A simulation time of 15 ns could already be achieved and allowed the characterization of the local and global mobility of the nucleosome particle on the nanosecond time scale at atomic resolution. The study of the dynamics of packed DNA is also of fundamental importance to understand the effect of DNA damage since DNA in the cell is not free but, to a large degree, associated with proteins to form a compact, only partially accessible structure. Simulation studies on such a large structure are only possible by using supercomputer resources as available at the PNNL supercomputer center. The simulation will be extended in the current grand challenge phase and could also form the basis to study DNA damage in the context of a packed DNA structure.

The Mechanism of BER by Human Polymerase-β

The objective of this component is a computational study of the factors that mediate fidelity of dNTP incorporation by the human enzyme polymerase-β (pol-β). Crystal structures of DNA pol-β suggest that the

"induced fit" model of polymerase fidelity applies to pol-β. In an induced-fit process, conformation changes induced by binding of the correct substrate bring the catalytic groups of the active site into better alignment for catalysis than do incorrect dNTPs. Pol-β consists of two domains, an 8-kDa N-terminal domain with deoxyribosephosphate lyase activity and a 31-kDa C-terminal domain with nucleotidyl transfer activity. Like other polymerases, the 31-kDa domain of pol-β has finger, palm, and thumb subdomains that form a complex with two magnesium ions and template DNA (finger and palm), and dNTPs (following a proposed hinge motion enabling the thumb to close and form the active site). Several dynamic events contribute to the catalytic cycle, including the exchange rate of the two magnesium ions required for catalysis, dynamics of the template DNA strand, and a hinge motion of the thumb subdomain in which main-chain atoms move up to 7Å. Residues of the active site are believed to be affected by these motions such that the transition state for nucleotidyl transfer is stabilized only when the thumb is close to the 8-kDa domain. Hinge motion alone is not sufficient for active site formation. A local switch in aromatic residue packing and hydrogen bonding network contributes to formation of the active site where the incorrect nucleotide cannot simultaneously fit in the binding site created by the template and base-pair with the template. A tyrosine residue detects whether correct Watson-Crick base pairing occurs such that upon correct base pairing, the tyrosine interacts with a phenylalanine, which then displaces a hydrogen bond from the catalytic aspartate, priming it for catalysis.

In particular, the project addresses the question of how magnesium ions, template DNA, subdomain hinge motions, organization of active site aromatics and the hydrogen bond network, and dNTP binding assemble the active site and achieve fidelity. Emphasis has been on the leading hypothesis for fidelity; namely, when the correct Watson-Crick base-pair forms, the thumb subdomain can fully close to assemble the active site. This hypothesis is based on the analysis of models generated from crystal structures of pol-β with and without the correct incoming dNTP, where a 7-Å displacement of the subdomain is seen when these structures are overlaid. We have carried MD simulations of this subdomain to study how such motion would coordinate with other dynamics such as substrate and metal binding. The project has made significant progress in the characterization of the chemistry central to the function of the repair protein pol-β. A Mg^{2+} ion facilitates the organization of the substrates at the active site of pol-β. Extensive simulations and electronic structure calculations have been used to characterize the detailed organization (structure and dynamics) associated with the ligands for the magnesium and their immediate environment. This work is crucial to answering questions concerning the synchronization of substrate binding and product release.

MD simulations were performed on ternary complexes of pol-β bound to gapped DNA and each of the four possible nucleoside triphosphate substrates (dNTP). A simulation was also carried out on a binary complex obtained by removing the substrate and the magnesium ions from the ternary complex. All of the starting structures for the simulations were derived from a crystal structure of a ternary human pol-β – gapped DNA – ddCTP complex (PDB entry: 1BPY), in which the thumb subdomain is positioned in its closed conformation over the ddCTP substrate analog. As a control, a simulation was performed on the crystal structure as is, except for protonation and solvation. Since the crystal structure was intentionally altered from the native complex by removal of the 3' hydroxyl groups from the primer and substrate, a comparison of the ddCTP and dCTP simulations was helpful in determining the effects of the missing 3' hydroxyl groups and estimating the active site structure of the native complex.

The starting structure that differed most radically from the crystal structure was the binary complex, modeled by removing the ddCTP substrate. Therefore, this simulation was extended for the longest duration to search for clear evidence of thumb opening. According to the proposed "induced fit" mechanism of fidelity assurance, the thumb subdomain is expected to reopen if the correct substrate (matching the template base) is

not in place. Both structural and essential dynamics analyses of the 5.2 ns trajectory failed to provide convincing proof of thumb subdomain movement. However, certain specific interactions across the thumb – 8-kDa domain interface showed signs of weakening as the simulation progressed. Kim et al. (2003) estimated that motion of the 8-kDa lyase domain in solution probably occurs on a time scale between 3 and 8 ns. Since opening of the thumb requires the breaking of multiple attractive interactions, a 5.2 ns simulation may well be too short to observe a definite opening movement.

Since we were unable to find any crystal structures of mismatched pol-β complexes, the starting structures used in the simulations of ternary complexes with mismatched substrates opposite the template guanine were created by modeling the mismatched base into the ddCTP structure and restoring the substrate and primer missing the 3' hydroxyl group. Though the essential dynamics analysis revealed no clear evidence of thumb movement in the 1.7 ns simulations, several differences were observed in specific interactions between the substrate and certain nearby residues. The binding of the triphosphate portion of the substrate to the magnesium ions in the active site was found to be very similar among the mismatches and the correct dCTP substrate. However, some peripheral active site differences were noted in details such as the hydrogen bonding partners of the substrate's 3' hydroxyl group. In the case of the purine mismatches, dATP and dGTP, the DNA accommodated the mismatch by moving the template base to provide room for the substrate to assume its normal position. This response may be a result of the lack of flexibility at the other end of the substrate due to the tight binding of the triphosphate moiety to the magnesium ions.

None of the simulations performed provides clear evidence of thumb opening in response to substrate substitution or removal, and thus the proposed "induced fit" fidelity mechanism is not supported by these results. However, this mechanism cannot be rejected on this basis, either, since the energy barrier to thumb opening may be of a magnitude that would require longer simulation times to observe. The rate-limiting step in the pol-β catalytic cycle may be the nucleotidyl transfer reaction rather than a conformational change if nucleotide selection occurs by means of transition state stabilization. If this is the case, then discrimination may depend on specific interactions between the substrate and its surroundings in the transition state structure.

Using the methods of MD simulation and *ab initio* optimization, we have been able to more fully characterize the active conformation of the pol-β active site. Working from the crystal structure closest to the fully formed active site, 1bpy[9] (*i.e.*, missing only the 3'OH groups on the DNA primer and substrate), two 1.7 ns MD simulations were performed; one (DCT) with the missing 3'OH's restored, via modeling, to the primer and substrate, the other (DDC) without restoration of the 3'OH groups. From *ab initio* optimizations performed on simplified *in vacuo* models of the active site, unambiguous conclusions cannot be drawn about ligand protonation states and hydrogen bonding preferences. Taken together, however, the MD simulations and optimized models help determine the probable impact of the missing 3'OH's in the crystal structure and enable us to predict the structure of the complete (native) pol-β active site in its active conformation just prior to the chemical nucleotidyl transfer step.

Besides binding the primer tightly to the catalytic magnesium ion, the DCT simulation revealed a very persistent hydrogen bond between the primer 3'OH and the 5'O (ester) substrate oxygen atom that helps maintain a tight control on the critical distance between the primer 3'OH and the substrate α-phosphate. *Ab initio* geometry optimizations on model complexes suggest that the substrate O2A oxygen should be considered as a potential proton acceptor for primer activation. With the primer 3'OH present, the catalytic ion is hexa-coordinated with a near octahedral geometry. Without the 3'OH, hexa-coordination (counting the

coordinated water molecule) is still achieved by coordination of both Asp 190 carboxylate oxygen atoms, but the ligand geometry about the catalytic ion is far from octahedral. In this case, the OD1 oxygen of Asp 190 is coordinated to both cations. A water molecule coordinates tightly to the catalytic ion in both MD simulations and in all of the optimized model structures, but is not present in the crystal structure, though it has been postulated by the crystallographers. The explanation for this discrepancy lies in the greater freedom of movement evident in the simulations, compared to the water molecule coordinated to the nucleotide-binding ion, which is observed in the crystal structure.

Restoration of the substrate 3'OH in the DCT simulation (*i.e.*, converting ddCTP to dCTP), resulted in a C2' endo to C3' endo change in the ribose pucker conformation, allowing the 3'OH to form a very persistent hydrogen bond to the O1B oxygen of the β phosphate. Besides its effect on the conformation and positioning of the substrate, it is possible that this hydrogen bond assists in the reaction mechanism by helping to stabilize the highly charged triphosphate moiety. Our solute-solvent interaction analysis of the simulation data has identified several tightly bound water molecules that form hydrogen-bonded networks around and between the primer and substrate phosphate groups.

The close agreement among the optimized models, the DCT simulation, and (for the coordination sphere of the nucleotide-binding ion) the 1bpy crystal structure demonstrates the metal cations, rather than the surrounding protein, are the prevailing influence in shaping the pol-β active site. Protonation of the bridging O1A oxygen on the dNTP substrate or either of the bridging aspartates is not supported by optimization tests carried out on the protonated models. Optimization of model QM8 with protonation of the OD2 oxygen of Asp 256 does not rule out this possibility. However, the apparent involvement of Asp 256 in a salt bridge with Arg 254 should favor the basic form.

Based on the combined results of the simulation and model optimization calculations, we expect the structure of the active conformation of the pol-β active site in its native (unaltered) state to closely resemble the structure observed in our DCT simulation. The Mg – O distances should be slightly lengthened and the Mg – Mg distance slightly shortened, in line with the optimized model, QM1, to compensate for some inadequacy of the non-bonded model used for the magnesium ions in the simulations in accounting for ligand-to-metal charge transfer. Of course, electron flow was fully taken into account in the *ab initio* geometry optimizations. The current study provides a detailed analysis of the structure of the active site of pol-β, and serves as an initial step in future investigations of the factors that determine the fidelity of this enzyme, which will require the calculation of free energies of binding and activation free energies of the different nucleotide substrates.

Protein Signaling

To conduct the simulation of protein kinase A (PKA), we started with the crystallographic structure of the catalytic domain of PKA (cPKA) complexed with the inhibitor balanol. cPKA has over 300 residues and, therefore, requires an efficient parallel algorithm such as NWChem to generate a long trajectory of the system within a short time. In this project, we addressed two issues: 1) Identify the determinants of protein-ligand recognition and use this information to guide the design of novel inhibitors for protein kinases and 2) Search for possible new inhibitor binding modes, for which we carried out docking experiments to examine whether known inhibitors of the enzyme bind favorably to dynamical snapshots of the apoenzyme whose conformations are significantly different from those found in existing crystal structures.

Our work on addressing the first issue relied upon continuum electrostatics models. Although these models have already generated many useful insights into designing specific protein kinase inhibitors, these models

have not taken into account conformational fluctuation. We hope to use explicit-solvent MD simulation models, which have a firmer theoretical basis, to validate previous results and generate new guiding hypotheses for designing novel inhibitors for protein kinases.

For identifying the determinants of protein-ligand recognition, we have developed a sensitivity analysis approach similar to those employed in engineering applications. In a sensitivity analysis study, one perturbs the parameters of a model to identify which parameters are truly significant in affecting a system property. A system property that we have looked at in some detail in drug-design applications is the binding free energy between PKA and its inhibitor balanol. Such analysis has already helped to identify the key components of balanol that are important for its recognition of PKA. This information has also generated pharmacophoric models for identifying new drug leads from small-molecule libraries. So far, the calculations on this system have been based largely on the implicit-solvent model implemented in the UHBD program developed in the McCammon group. We hope to repeat these calculations by the MD approach to further evaluate the hypotheses generated from the implicit-solvent model. By accounting for protein flexibility, MD simulations will also allow us to explore the effects of changing the flexibility or size of an inhibitor on binding. Such analysis was not feasible with our previous continuum solvent model using a rigid conformation approximation.

We demonstrated that sensitivity analyses based on MD simulations can be carried out systematically and efficiently by calculating analytical derivatives of the form dO/dli, where O is an observable and li is a parameter describing a certain feature of a functional group. These derivatives describe how small perturbations to the parameters li's affect the property O. For larger perturbations to the parameters, one may use perturbation theory. For example, the Zwanzig formula DA = -<exp(-b(V2-V1)>1/b—where b is the inverse of the product of the gas constant times the absolute temperature, V2 and V1 are, respectively, the potential energies after and before the a parameter is modified—can be used to study the effects of larger parameter perturbations on affecting the free energy of a system. In addition to perturbing the parameters of the ligand, one can also perturb the parameters of the protein to identify the key amino acids that recognize the ligand. By comparing the free energy change resulting from changing the parameters of a protein and ligand functional groups simultaneously with that from changing the parameters of the two functional groups separately, one can also probe the significance of pair interactions in affecting protein-ligand recognition. We have done this using continuum solvent models, and will be generalized to explicit-solvent models in our work. Identification of the key amino acids for recognizing an inhibitor will enable the study of what other amino acids take the place of these residues in other protein kinases. This information will be useful for guiding the design of selective inhibitors targeted toward a specific kinase, as demonstrated from our previous analysis involving almost 400 protein kinases.

Even if the structure of a receptor is determined experimentally, it may not be a conformation to which a ligand would bind when induced-fit effects are significant. In our work, we have evaluated the use of an ensemble of receptor conformations generated from an MD simulation for molecular docking. Two MD simulations were carried out to generate snapshots for PKA: one with the ligand bound, the other without. The ligand, balanol, was then docked to conformations of the receptors presented by these trajectories. The Larmarkian genetic algorithm in AutoDock was used in the docking. Three ligand models were used: rigid, flexible, and flexible with torsional potentials. When the snapshots were taken from the MD simulation of the protein-ligand complex, the correct docking structure was found in all cases. On the other hand, when the snapshots were taken from the simulation of the protein alone, several clusters of structures were found. Out of the ten docking runs for each snapshot, at least one structure was close to the experimental complex structure when the flexible ligand models were used. However, the lowest energy structures, according to

AutoDock, did not always correspond to the correctly docked structure. Rescoring using a more sophisticated Generalized Born electrostatics model did not improve the identification of the correctly docked structure. In fact, we found that the correctly docked structure appeared more frequently as the lowest energy structures with the AutoDock scoring function. This can provide a useful criterion for selecting the correctly docked structure from clusters of structures obtained from molecular docking experiments.

References

Kim S-J, WA Beard, J Harvey, DD Shock, JR Knutson, and SH Wilson. 2003. Rapid segmental and subdomain motions of DNA polymerase-β. *J Biol Chem* 278: 5072-5081.

List of all publications resulting from this work

Barthel, A, D Roccatano, and M Zacharias. 2004. "Molecular dynamics of the nucleosome core particle on the nanosecond time scale." In prep.

Miller, JH, C-C Fan Chiang, TP Straatsma, and MA Kennedy. 2003. "8-oxoguanine Enhances Bending of DNA that Favors Binding to Glycosylases." *J. Am Chem. Soc.* **125**(20):6331-6336.

Rittenhouse, RC, WK Apostoluk, JH Miller, and TP Straatsma. 2003. "Characterization of the Active Site of DNA Polymerase β by Molecular Dynamics and Quantum Chemical Calculation." *Proteins: Structure, Function, and Genetics*, **53**(3):667-682**.**

Wong, CF, J Kua, Y Zhang, TP Straatsma, and JA McCammon. 2005. "Molecular Docking of Balanol to Dynamics Snapshots of Protein Kinase A." *Proteins: Structure Function and Bioinformatics*, 61 (4): 850-858.

Zacharias, M. 2004. "Comparative molecular dynamic simulations of regular, nicked and gapped duplex DNA." *Biopolymers*, submitted.

Zacharias, M and JW Engel. 2004. "Influence of a fluorobenzene nucleobase analogue on the conformational flexibility of RNA studied by molecular dynamics simulations." *Nucleic Acids Res.*, **32:**6304-6311.

List of all presentations resulting from this work

Kennedy, MA, MK Bowman, GW Buchko, PD Ellis, JH Miller, DF Lowry, TJ Straatsma, SS Wallace, and D Wilson, III. 2002. "Damage Recognition, Protein Signaling, and Fidelity in Base Excision Repair." Presented at the 9th Genome Contractor and Grantee Workshop, Oakland, CA, January 27-31, 2002.

Miller, JH, TJ Straatsma, MA Kennedy. 2002. "How does 8-oxoguanine DNA glycosylase (hOGG1) recognize oxoG:C base pairs?" Presented at the Annual Scientific Meeting of the Radiation Research Society, Reno, NA, April 20-24, 2002.

List of all significant methods/routines or codes developed

N/A

List of any awards or special recognitions resulting from research

N/A

Appendix A – Full report of first year activities and accomplishments

Abstract

Environmental factors, such as ionizing radiation, contribute to the continuous damage of cellular DNA, in addition to endogenous sources. The damage resulting from oxidative stress and ionizing radiation is primarily in the form of oxidized bases, single strand breaks, and loss of bases. These are the targets of the base excision repair (BER) mechanism enzymes, including polymerase-β (pol-β). Failure to correctly and timely repair these damaged DNA sites can result in cell death, carcinogenesis, or genetic diseases. Resulting mutations in cell signal transduction enzymes can lead to uncontrolled cell proliferation or differentiation. For example, mutations in Ras, the molecular switch in several growth-factor signaling pathways, have been found in about 30% of human tumors. These signaling pathways often involve a chain of protein kinases that activate or deactivate each other through phosphorylation reactions, eventually controlling the activation of transcription factors in the cell nucleus.

The goal of this Grand Challenge Application Project is to enhance our understanding of the mechanism of human pol-β, one of the key enzymes in BER, and the cell signaling enzymes cyclic-AMP-dependent protein kinase and Ras. This work uses molecular modeling and simulation studies to specifically focus on the

- dynamics of DNA and damaged DNA
- dynamics and energetics of base flipping in DNA
- co-solvent effects on biomolecular structures, including DNA
- mechanism and fidelity of nucleotide insertion by BER enzyme human pol-β
- mechanism and inhibitor design for cyclic-AMP-dependent protein kinase
- dynamics and energetics of Ras and its complex with effector molecule Raf.

Molecular dynamics (MD) simulations and electronic structure calculations are performed using NWChem, the massively parallel software for computational chemistry developed at Environmental Molecular Sciences Laboratory. NWChem is also used for the analysis of the generated MD trajectories, and any additional analysis tools required for the work in this project were developed within the analysis modules of NWChem.

Human Polymerase-β Study (T.P. Straatsma, J.H. Miller)

A detailed understanding of the mechanism by which fidelity is assured by pol-β in the gap-filling DNA repair process is expected to require a complete and accurate description of the pol-β active site in its active conformation just prior to the chemical nucleotidyl transfer step. Thus far, crystal structures obtained for pol-β–gapped DNA–dNTP complexes have been incomplete due to intentional structural alterations necessary to prevent the chemical step in order to obtain a stable structure from the midpoint of the pol-β catalytic cycle. Using the methods of MD and quantum chemical calculation, we have been able to more fully characterize the active conformation of the pol-β active site. Working from the crystal structure that comes the closest to the fully formed active site (*i.e.*, missing only the 3'OH groups on the DNA primer and substrate), two 1.7 ns MD simulations were performed: one with the missing 3'OH's restored, via modeling, to the primer and substrate (converting ddCTP into the true substrate, dCTP); the other without restoration of the 3'OH groups. While the structure of the active site maintained in the first simulation preserved the coordination geometry seen in the crystal structure, the results identified important hydrogen bonds involving the 3'OH's of the

primer and substrate, and confirmed the presence of a previously suspected water molecule tightly coordinated to the catalytic Mg^{2+} ion (Figure A.1). An exhaustive solute-solvent interaction analysis of the simulation trajectory has led to the observation of several hydrogen-bonded networks involving tightly bound water molecules that may assist in the stabilization of the highly charged triphosphate portion of the substrate. The observations made on the basis of MD simulation are supported by *ab initio* geometry optimizations performed on simplified models of the pol-β active site. The optimized models, along with the second MD simulation (without 3'OH's), indicate that the missing 3'OH's in the crystal structure are responsible for certain distortions in the inner coordination sphere of the catalytic ion. A key hydrogen bond formed between the primer 3'OH and the O2A oxygen on the α-phosphate of the substrate may play an important part in the nucleotidyl transfer mechanism by facilitating the activation of the 3' oxygen nucleophile. In addition, the calculations have helped to resolve questions that have been raised regarding the protonation states of certain ligands in the bimetallic coordination environment.

Comparison of crystal structures obtained from different points in the pol-β catalytic cycle have led to the identification of a series of conformational changes involving a large domain "thumb closing" motion and several specific smaller changes. To determine whether the presence or absence of the proper dNTP substrate in the active site of the enzyme is essential for thumb domain movement and assembling of active site residues, simulation of the opening process was performed. The "closed thumb" ternary complex structure was used and the ddCTP substrate and Mg^{2+} ions were removed. Without the substrate, the enzyme should revert to the open thumb conformation. To confirm this, the structure was subjected to a 5.2 ns MD simulation and the trajectory was analyzed with essential dynamics analysis. With this methodology, it is generally possible to separate anharmonic positional fluctuations relevant for the biological function from Gaussian fluctuations. However the analysis results indicate that a simulation of 5.2 ns in length is too short to definitively confirm thumb domain movement. Presently, we are using potential-of-mean-force calculation to force the system through domain opening. In this way we expect to observe the sequence of events, in particular active site residues disassembling, during the opening process.

Since we are particularly interested in the fidelity mechanism employed by pol-β, in addition to simulation of the ternary complexes with correct substrate (dCTP and ddCTP) (Figure A.2), a series of three 1.7 ns MD simulations with incorrect dNTP substrates, *i.e.*, not matching the cytosine template base, were performed. Thus far, analysis of the trajectories has failed to identify significant conformational differences that would support an "induced fit" mechanism for repair fidelity. However, further analysis is needed to confirm this conclusion. We are also exploring an alternate proposed mechanism for pol-β fidelity based on transition state stabilization.

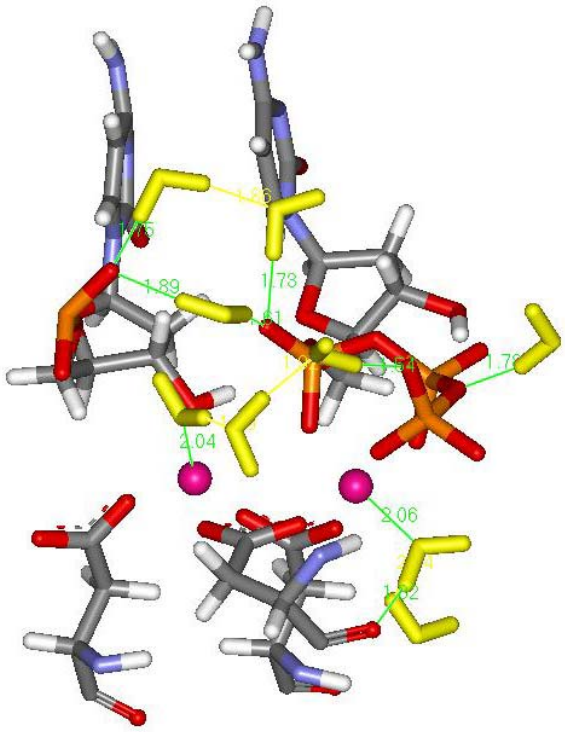


Figure A.1. The pol-β bimetallic active site showing coordinating ligands and tightly bound water molecules (yellow).

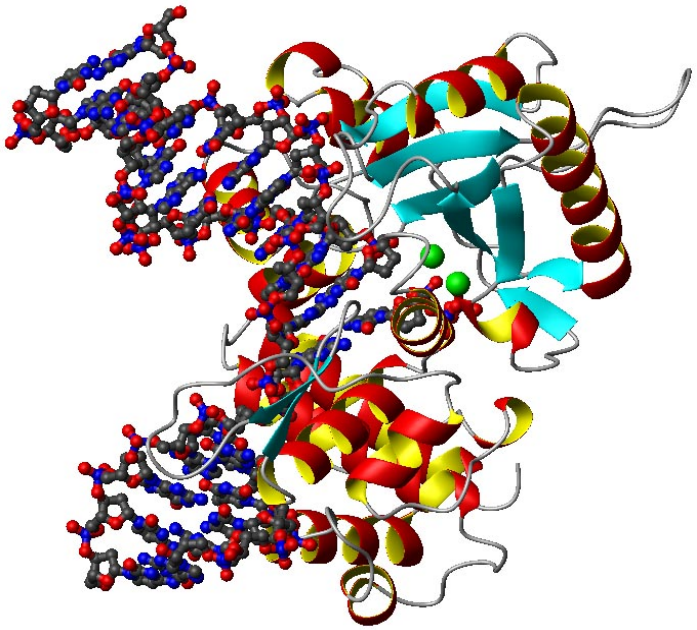


Figure A.2. Closed conformation of the pol-β (as ribbons) – gapped DNA (as balls and sticks) – dCTP ternary complex. dCTP substrate and magnesium ions (in green) present in the active site can be seen in the center of the picture.

Effects of 8-oxoguanine on DNA Bending (J.H. Miller, T.P. Straatsma)

MD simulations were carried out on the fully solvated and cation-neutralized DNA oligonucleotide GGGAA***C***AACTAG:CTAGTT***G***TT. C in its native form and with guanine in the central G_{19}:C_6 base pair replaced by 8-oxoguanine (8oG). The direction and magnitude of global bending were assessed by a technique used previously to analyze simulations of DNA containing a thymine dimer. The presence of 8oG did not greatly affect the magnitude of DNA bending; however, bending directions that compress the major groove and expand the minor groove were significantly more probable when 8oG replaced G_{19}. Crystal structures of glycosylases bound to damaged-DNA substrates consistently showed bending that expands the minor groove. We conclude that changes in bending dynamics that facilitate this expansion are part of the mechanism by which 8oG is recognized by the base excision repair pathway.

Replacing G19 in GGGAACAACTAG:CTAGTTGTT. C by 8oG did not induce any major local perturbations in this oligonucleotide. No tendency for opening of the 8oG19:C6 base pair was detected in MD simulations of 2 ns duration. The most pronounced local changes induced by the presence of 8oG were an attraction for counter ions and a small kink in the helical axis between base pairs 8oG19:C6 and T18:A7. The damaged base had a large effect on the dynamics of global bending. Conformations with the oligomer bent in the direction of minor groove were rarely seen when 8oG was present, even though that was the preferred direction of bending for the native sequence.

The local and global effects of 8oG observed in our simulations are probably related; however, the mechanism of this connection is uncertain. Attraction of counter ions to the damage site did not preferentially neutralize atoms on the major-groove face of the oligomer. Changes in the electrostatic potential of the damaged base will also affect stacking interactions, which may influence bending dynamics. Further investigation of this mechanism is needed.

Since glycosylases recognize 8oG bind in the minor groove, an induced preference for bending into the major groove could be a factor in damage recognition, as has been proposed for recognition of thymine dimers by T4 endonuclease V. This concept is being investigated further by potential-of-mean-force calculations designed to mimic the “pinch-push-pull” mechanism of base flipping.

Simulation of Bulges in DNA and RNA (M. Zacharias)

The aim of the ongoing project is to characterize the structure and dynamics of chemically modified DNA due to oxidative stress or irradiation. During the first phase of the project, MD simulations were performed on a 12 base pair B-DNA structure including water molecules and ions. The simulations were performed for approximately a 2 ns simulation time. The equilibrated simulation will serve as a reference system for simulations that include chemically modified nucleobases. In addition, the equilibrated system will be used to induce structural DNA deformations that are known to occur during DNA repair processes such as minor groove opening and extra-helical base flipping. MD simulation on single unmatched extra bases (bulges) located at the center of otherwise double-stranded DNA and RNA have been initiated. Such non-helical structures occur frequently in RNA and also in DNA as a result of DNA modification and damage. The simulations on a single adenine and a single uridine bulge in RNA have now reached a duration of approximately 1.5 ns. However, due to the increased flexibility of the central bulge nucleotide compared to regular nucleic acids, the simulations are not yet fully equilibrated.

Protein Kinase A (J.A. McCammon, C.F. Wong)

A MD simulation on the apo form of protein kinase A has been carried out, currently to about 3 ns. Snapshots from about 1 ns of this trajectory have been used for ligand docking experiments. We used AutoDock Version 3.05 to dock balanol to these snapshots. The average binding energy is -12± 2 kcal/mol. This value is quite close to the experimental values ranging from -10 to -11 kcal/mol. For the scoring function used in AutoDock 3.05, the electrostatic interactions contribute about -2 kcal/mol to the binding energy. Although these results are consistent with those obtained by using the Poisson-Boltzmann/Solvent-accessible Area (PB/SA) model earlier, they differ quantitatively. The PB/SA model gave significantly more favorable binding energy (-27 kcal/mol), and the electrostatics contributions (-0.9 kcal/mol) accounted for a smaller fraction of the total binding energy. However, the results from AutoDock agree better with a semi-empirical PB/SA model in which a few adjustable parameters were introduced to scale the results of the PB/SA model to fit existing experimental data. The semi-empirical model gave a hydrophobic contribution of -16 kcal/mol, an electrostatics contribution of -1 kcal/mol, and other contributions that accounted to about 2 kcal/mol.

Further study will include other derivatives of balanol and other ligands as planned. Additional analyses will also be carried out to characterize the dynamics of the apo protein and its influence on molecular recognition.

Appendix B – Full report of second year activities and accomplishments

Abstract

Environmental factors, including ionizing radiation, contribute to continuous damage of cellular DNA, in addition to endogenous sources. The damage resulting from oxidative stress and ionizing radiation is primarily in the form of oxidized bases, single strand breaks, and loss of bases. These are the targets of the base excision repair (BER) mechanism enzymes, including polymerase-β (pol-β). Failure to correctly and timely repair these damaged DNA sites can result in cell death, carcinogenesis, or genetic diseases. Resulting mutations in cell signal transduction enzymes can lead to uncontrolled cell proliferation or differentiation. For example, mutations in Ras, the molecular switch in several growth-factor signaling pathways, have been found in about 30% of human tumors. These signaling pathways often involve a chain of protein kinases that activate or deactivate each other through phosphorylation reactions, eventually controlling the activation of transcription factors in the cell nucleus.

This computational grand challenge application project is using the massively parallel computing resources in the Environmental Molecular Sciences Laboratory to perform molecular modeling and simulation studies to enhance our understanding of the mechanism of human pol-β, one of the key enzymes in BER repair, and the cell signaling enzymes cyclic-AMP-dependent protein kinase and Ras.

This document reports on the progress that was made in FY 2003 in the following sub-tasks:

- dynamics of DNA and damaged DNA
- dynamics and energetics of base flipping in DNA
- mechanism and fidelity of nucleotide insertion by BER enzyme human pol-β
- mechanism and inhibitor design for cyclic-AMP-dependent protein kinase.

Modeling the Effect of Nicks, Extra Bases, and Gaps in Damaged DNA Using Molecular Dynamics Simulations

Introduction

Damage of DNA by radiation and chemical reagents is a major cause of cancer and cell aging. Such damage can lead to changes in the DNA structure and dynamics, which in turn influence recognition and repair of damaged DNA. Common forms of DNA damage include the loss of a base (creating an abasic site), a bond break in the DNA backbone, or the loss of one or several nucleotides. Bond breaking or nicking of the DNA backbone may lead to significant changes in the DNA structure (*e.g.*, unstacking and loss of the helical DNA structure). Surprisingly, the crystallographic structure of a double-stranded nicked DNA oligonucleotide (Aymani et al. 1990) shows only a modest variation when compared to a standard intact DNA duplex. One significant difference compared to a standard B-DNA structure is a narrower minor groove around the position of the nick (Aymani et al. 1990). However, X-ray crystallography provides only a static picture of nucleic acids. It is expected that a nick in DNA or the loss of one or several nucleotides due to radiation or chemical damage (gap) can lead to increased local flexibility or a change in average solution conformation. Molecular dynamics (MD) simulations are ideally suited to study such fine structural changes and conformational dynamics of DNA at very high

spatial and time resolution. A better understanding of the structure and dynamics of such DNA modifications is important for an improved understanding of its recognition and repair.

During the project duration, multi-nanosecond simulations on an intact 12 base pair DNA, the same sequence with a central nick, and a gap of one nucleotide at the center were performed. In addition, simulations on single-base bulge structures (in RNA) were performed, which correspond to another common defect in double stranded nucleic acids. For all simulations NWChem, the massively parallel software for computational chemistry developed at Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory (PNNL), in combination with the AMBER (parm99) force field, was used.

An experimental (X-ray) structure was used as a starting structure for the simulations of the nicked DNA (Aymani et al. 1990). Since no experimental structure with exactly the same sequence as the nicked DNA was available, a standard B-DNA structure was used as a reference. The starting structure for the simulations with a single nucleotide gap was obtained by removing a (T) nucleotide at the center of the nicked double-stranded DNA structure. All structures were solvated in cubic boxes with added counter ions and additional ion pairs (approximately 5000 SPCE water molecules per simulation, with an ion concentration of approximately 0.3 M). The simulation systems were heated and equilibrated at 300 K over a period of 2 ns, followed by approximately 12 nanoseconds of data-gathering time. A simulation of each system took about 24 hours using 32 nodes on the HP/Linux 1960 Itanium-2 processor cluster (MPP2) at the PNNL Molecular Science Computing Facility.

Results and Discussion

Although the intact and nicked DNA oligonucleotide structures showed considerable conformational fluctuations during the MD simulations, some structural features of the experimental structures were well preserved. In particular, it has been observed already in the experimental structure that the central nick in the DNA backbone causes a local narrowing of the minor groove (Aymani et al. 1990; see Figure. B.1). This feature was preserved throughout the simulation. Such a stable narrowing of the DNA double strand might be a specific structural signature recognized by repair enzymes. Surprisingly little conformational changes were observed at the central bases around the position of the strand break. The T bases at the position of the strand break remained in a stacked conformation throughout the simulation time. No significant changes in twisting or bending when compared to the regular DNA oligonucleotide were observed. The average bend angle of the DNA during the simulation was approximately 10 degrees. This result is consistent with experiments by Mills et al. (1994, 1999) who investigated the global conformational flexibility of nicked and gapped DNA molecules using gel electrophoresis (Mills et al. 1994) and electric birefringence experiments (Mills et al. 1999). In these studies it was found that a single nick in DNA did not significantly change the global properties of DNA.

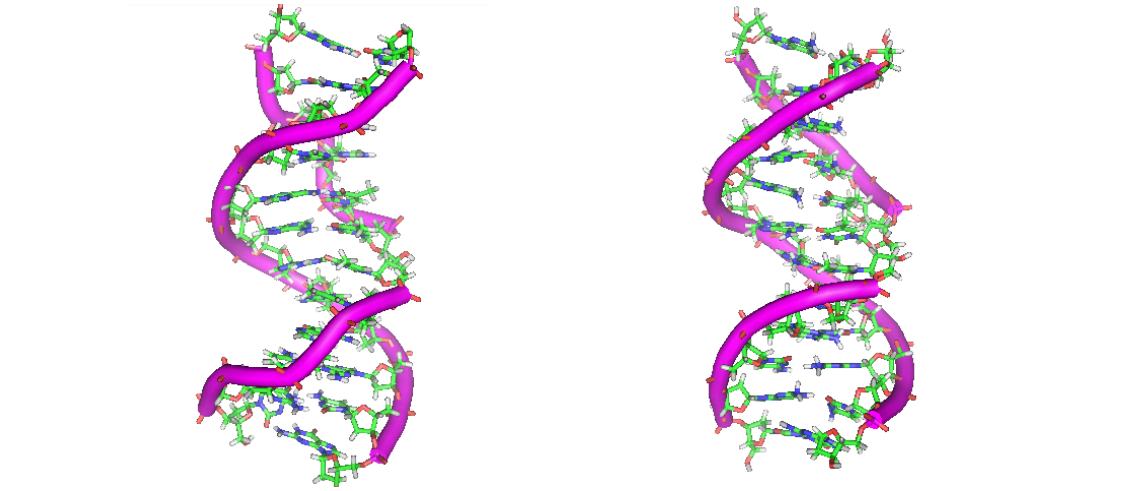


Figure B.1. Comparison of the DNA start structure with a central strand break (nick) at the center (left panel) and a snapshot from the MD simulation after ~10 ns simulation time (right panel). The DNA backbone trace is indicated as a cartoon.

In the case of the simulations of the DNA double strand missing a T nucleotide at the center (gapped DNA), larger conformational changes in the DNA were observed. In the initial DNA structure, the bases that flanked the gap were in an "unstacked" state (Figure B.2, left panel) during the MD simulation and no further unstacking or diffusion of water molecules into the space between the flanking nucleotides was observed. Instead, during the simulation the gap between the two segments of the second strand was closed and the two nucleotides adopted a stacked conformation similar to a dinucleotide step conformation in regular DNA (Figure B.2, right panel). The DNA conformation is actually quite similar to a DNA that contains a single base bulge (*i.e.*, the bulge base corresponds to the base opposite to the gap). Overall, the DNA adopted a bend conformation compared to the intact or nicked DNA duplex with a bend angle reaching 30 degrees. This result is, again, consistent with the experimental observation of Mills et al. (1994, 1999), who found a significant effect of gaps in DNA on the global conformational properties (bending) of the duplex.

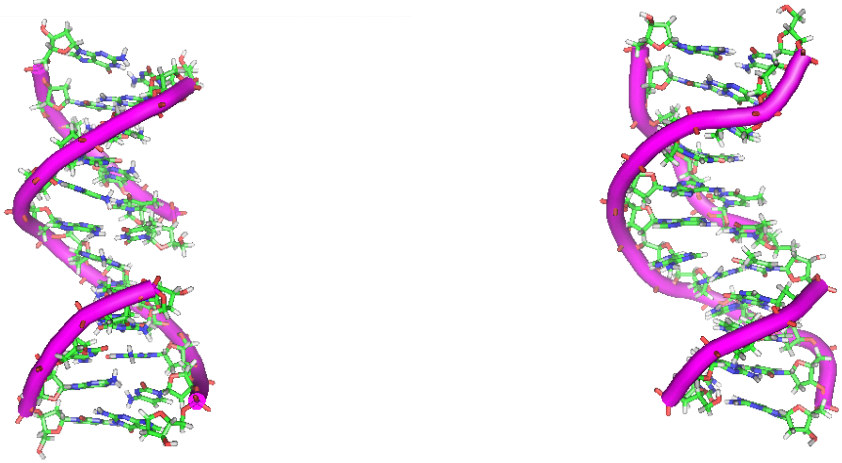


Figure B.2. Comparison of the DNA start structure with a central gap (one missing nucleotide) at the center (left panel) and a snapshot from the MD simulation after ~10 ns simulation time (right panel).

Conclusions

The comparative MD simulations on regular, nicked, and gapped double-stranded DNA indicate some differences in conformational flexibility and conformational preference that might be relevant for the recognition of these modifications by DNA repair enzymes. The present simulation results are consistent with available experimental data on the effect of nicks and gaps on DNA flexibility and allow some further insight into the underlying conformational changes at atomic detail. A publication on this topic is in preparation (Zacharias 2003). Future simulations on damaged DNA will include studies on more extended gap regions, a more detailed characterization of the molecular motions, and the potential of mean force calculations on the deformability of DNA at sites of damage.

References

Aymani J, M Coll, GA van der Marel, JH van Boom, J Wang , and A Rich. 1990. "Molecular structure of nicked DNA: A substrate for DNA repair enzymes." *Proc. Natl. Acad. Sci.*, 87(7): 2526-30.

Mills JB, JP Cooper, and PJ Hagerman. 1994. "Electrophoretic evidence that single-stranded regions of one or more nucleotides dramatically increase the flexibility of DNA." *Biochemistry*. 33(7):1797-1803.

Mills JB, E Vacano, and PJ Hagerman. 1999. "Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA)." *J. Mol. Biol*. 285(1): 245-257.

Zacharias, M. 2003. "Comparative molecular dynamic simulations of regular, nicked and gapped duplex DNA." In prep.

Mechanisms of Damage Recognition in Base Excision of Repair

Aerobic respiration, a stepwise four-electron reduction of O_2 to H_2O, is the energy production mechanism of almost all higher organisms. Partially reduced intermediates, which include O_2-, H_2O_2, and OH radicals, are potent oxidants that can escape from mitochondria and attack vital cellular conponents, such as DNA. The electron-rich bases of DNA are prime targets for interaction with these electrophilic oxidants; hence, higher organisms must have efficient mechanisms to repair base damage induced by endogenous oxidative stress (Lindahl and Wood 1999). The BER system accomplishes this objective by removing the damaged base. Subsequent steps in BER prepare an abasic site that is a good substrate for DNA polymerases that use the undamaged strand as a template for nucleotide gap filling.

The reactive oxygen species induced in cells by exposure to ionizing radiation are similar to endogeous electrophilic oxidants. However, absorption of radiation is generally associated with higher energy states than respiration so that sites of multiple DNA damage from ionizing radiation are considerably more likely than from endogenous ROS (Weinfeld et al. 2001). Whether spatial patterns of radiation-induced DNA damage reduce the efficiency of BER is an important question in assessing the health risks from radiation exposure. Understanding how the DNA glycosylases of the BER system recognizes DNA damage is an essential and, thus far, unrealized part of answering this question.

Osman et al. (2001) proposed that changes in the bending dynamics of DNA induced by thymine dimers (Miaskiewicz et al. 1996) were part of the mechanism by which T4 endonuclease V recognized these lesions. To investigate the effects of 8-oxoguanine (8oG) on the bending dynamics, we performed MD simulations on the DNA oligonucleiotide GGGAACAACTAG:CTAGTTGTT. C in its native form and with guanine in the central C6:G19 base pair replaced by 8oG. The direction and magnitude of global bending were calculated by the technique illustrated in Figure B.3.

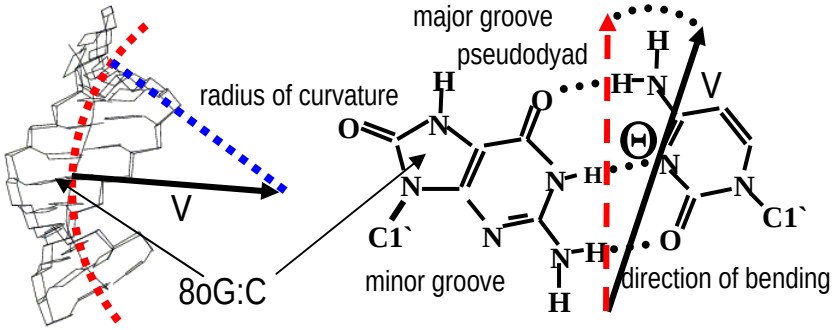


Figure B.3. Schematic representation of global bending analysis.

Helical axis reference points from analysis of DNA structures in Cartesian coordinates by the program CURVES (Lavery and Sklenar 1989) were fit by an arc to obtain the radius of curvature. To quantify the direction of bending, a vector V was defined by connecting the helical axis reference point of the 8oG:C base pair to the center of curvature. The projection of V onto the plane of the 8oG:C base pair was used to define the direction of bending relative to the pseudodyad axis of the base pair, which passes through the Watson-Crick hydrogen bonds and points into the major groove.

Figure B.4 shows the distribution of curvature, reciprocal of the radius of curvature, in 2 ns simulations of the oligomer GGGAACAACTAG:CTAGTTGTT. C with and without 8oG.

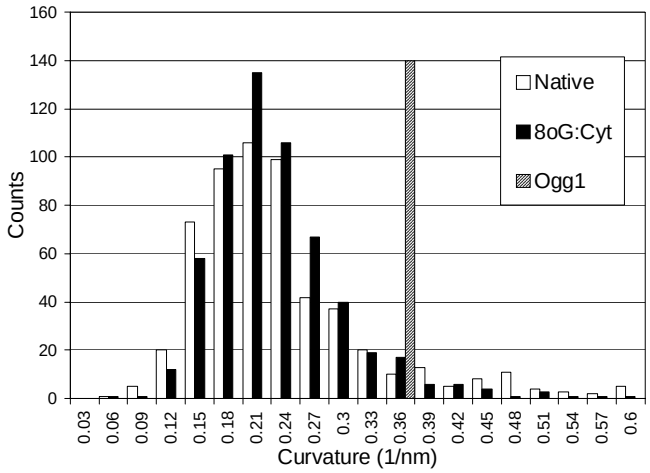


Figure B.4. Distribution of bending magnitude.

Curvature was slightly larger in the presence of 8oG but significantly less than that observed in the crystal structure of DNA containing 8oG bound to the human DNA glycosylase Ogg1 (Bruner et al. 2000).

Figure B.5 shows that the main effect of 8oG was to make bending into the major groove significantly more probable than in the native sequence. In the absence of 8oG, bending into the minor groove is more probable than bending into the major groove. Replacing G19 by 8oG created a strong preference for bending to the major groove. Since crystal structures of glycosylases bound to damaged-DNA substrates consistently show a sharp bend into the major groove at the damage site, we concluded that changes in bending dynamics that assist the formation of this kink are a part of the mechanism by which glycosylases of the base excision repair pathway recognize the presence of 8oG.

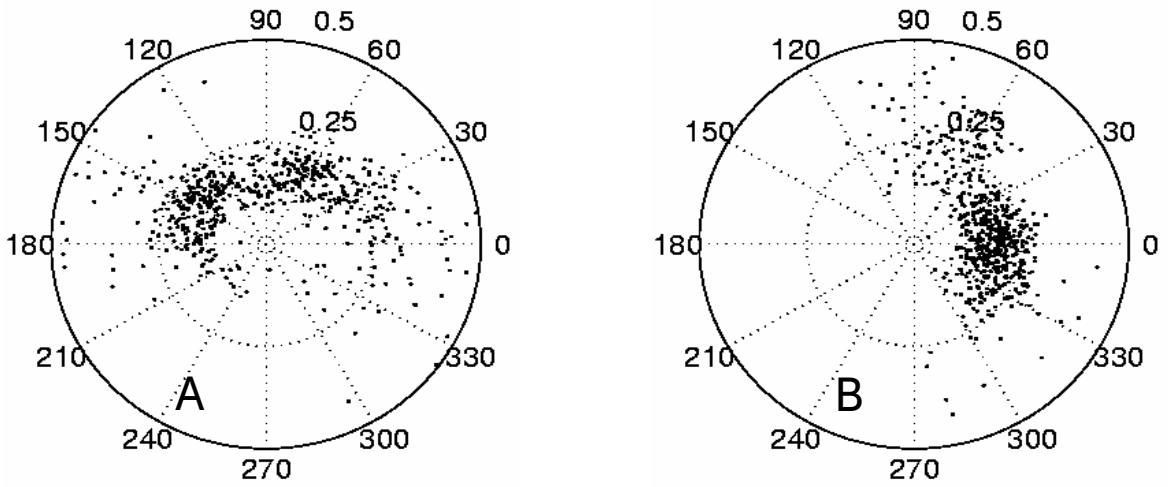


Figure B.5. Curvature (radial coordinate) as a function of bending direction from 2 ns simulations of oligomer GGGAACAACTAG:CTAGTTGTT. C without (A) and with 8oG (B).

Preferential bending of DNA has been associated with asymmetric phosphate neutralization (Strauss and Maher 1994). The presence of 8oG caused counter ions to stay closer to DNA with the major change being the number of counter ions in close proximity to a DNA atom; however, attraction of counter ions to the damage site did not preferentially neutralize atoms on the major-groove face of the oligomer.

Figure B.6 shows that, for the native sequence, bending toward the minor groove correlates with shorter H8-O2P separations. Electrostatic repulsion between O8 and phosphate oxygen atoms is probably one of the reasons why bending into the minor groove is avoided when G19 is replaced by 8oG.

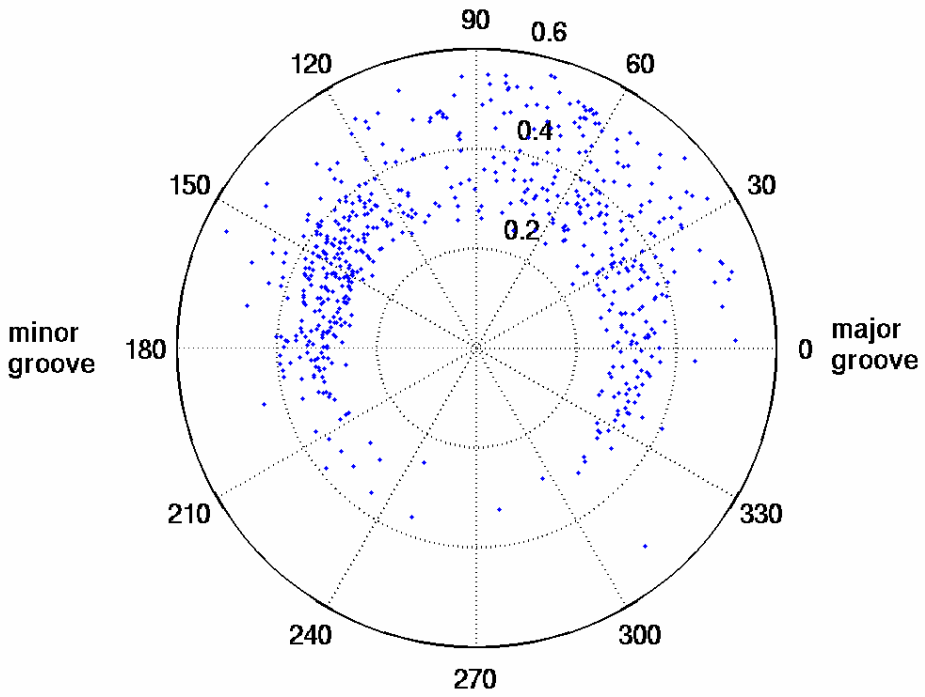


Figure B.6. H8 to O2P distances (nm) in the C6:G19 base pair of the oligomer GGGAACAACTAG:CTAGTTGTT. C as a function of direction of bending.

References

Bruner, SD, DPG Norman, and GL Verdline. 2000. "Structural basis for recognition and repair of the endogenous mutagen 8-oxoguranine in DNA." *Nature* 403(6772): 859-866.

Lavery R and H Sklenar. 1989. "Defining the Structure of Irregular Nucleic Acids: Conventions and Principles." *J Biomol. Struct Dyn*. 6(4), 655-667.

Lindahl, T and RD Wood. 1999. "Quality control by DNA repair." *Science* 286(5446): 1897-1903.

Miaskiewicz, K, J Miller, M Cooney, and R Osman. 1996. "Computational simulation os DNA distortions of a cis,syn-cyclobutane thymine dimer lesion." *J Am Chem Soc,* 118(38): 9156-9163.

Miller, JH, CP Fan-Chiang, TP Straatsma, and MA Kennedy. 2003. "8-Oxoguanine Enhances Bending of DNA that Favors Binding to Glycosylases." *J. Am Chem. Soc.* 125(20):6331-6336.

Osman, R, M Fuxreiter, and N Luo. 2000. "Specificity of damage recognition and catalysis of DNA repair." *Computers and Chemistry* 24(3-4): 331-339.

Rittenhouse, RC, WK Apostoluk, JH Miller, and TP Straatsma. 2003. **"**Characterization of the active site of DNA polymerase β by molecular dynamics and quantum chemical calculation." *Proteins: Structure, Function, and Genetics***,** 53(3):667-682.

Strauss, JK and LJ Maher III. 1994. "DNA Bending by Asymmetric Phosphate Neutralization." *Science* 266(5192): 1829-1834.

Weinfeld, M, A Rasouli-Nia, MA Chaudhry, and RA Britten. 2001. "Response of base excision repair enzymes to complex DNA lesions." *Radiat Res* 156(5), 584-589.

Mechanism and Fidelity of Nucleotide Insertion by BER Enzyme Human Polymerase-β

A key step in the BER process for the repair of damaged cellular DNA is the insertion and attachment of a replacement nucleotide that correctly matches the nucleotide on the opposite (template) strand. This intricate chemical manipulation is catalyzed by the pol-β repair enzyme. Using crystal structures of the pol-β enzyme complexed with a gapped (missing one nucleotide on one strand) DNA fragment and with or without a bound replacement nucleotide (dNTP) substrate in the enzyme's active site, we have investigated details of the catalytic mechanism and the means by which the enzyme ensures the fidelity of the repair process. MD simulations were used to identify important enzyme-substrate interactions and conformational changes that play a role in nucleotide binding and selectivity. In several experiments, the correct substrate, as present in the crystal structure, was replaced with a mismatched substrate (Figure B.7) to examine the dynamic response of the enzyme to the non-Watson-Crick base pair. The larger purine-to-purine mismatches were accommodated by a staggered arrangement of the bases consistent with recent crystallographic observations on complexes with bound mismatched base pairs. It has been proposed that binding of a mismatched substrate prevents a crucial conformational change involving the closing of the "thumb" subdomain of the enzyme over the bound substrate, which is responsible for the proper alignment of the substrate in the active site prior to the chemical reaction that covalently attaches the replacement nucleotide to the 3' side of the gap.

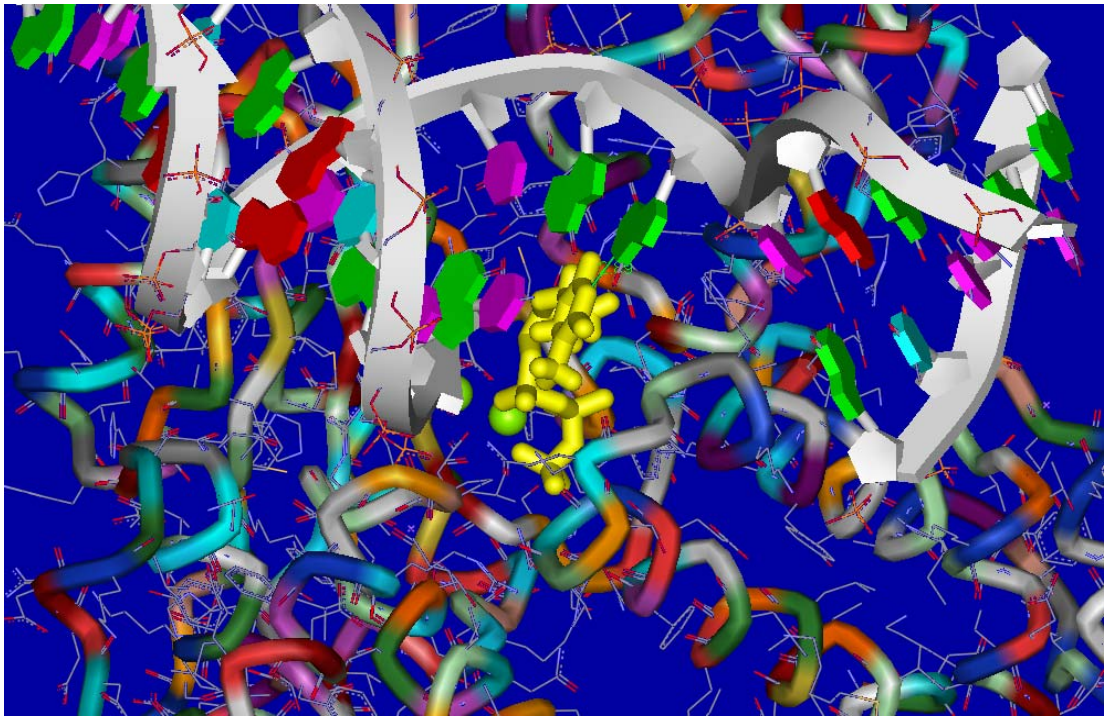


Figure B.7. Thymine (dTTP) mismatched substrate bound in the pol-β active site opposite the template guanine nucleotide.

For the several simulations we performed (with matching template dCTP and ddCTP substrates, and the mismatched dTTP, as well as "close thumb" structure with the removed substrate), we applied the method of essential dynamics analysis, but only in one case (dTTP) were we able to observe persistent domain motions. However, the movement involved the 8-kDa domain rather than the "thumb" (Figure B.8). But since we have not seen any major destabilization of the active site residues enabling release of the incorrect substrate, we have not been able to confirm hypothesis of the "induced fit" mechanism of pol-β action.

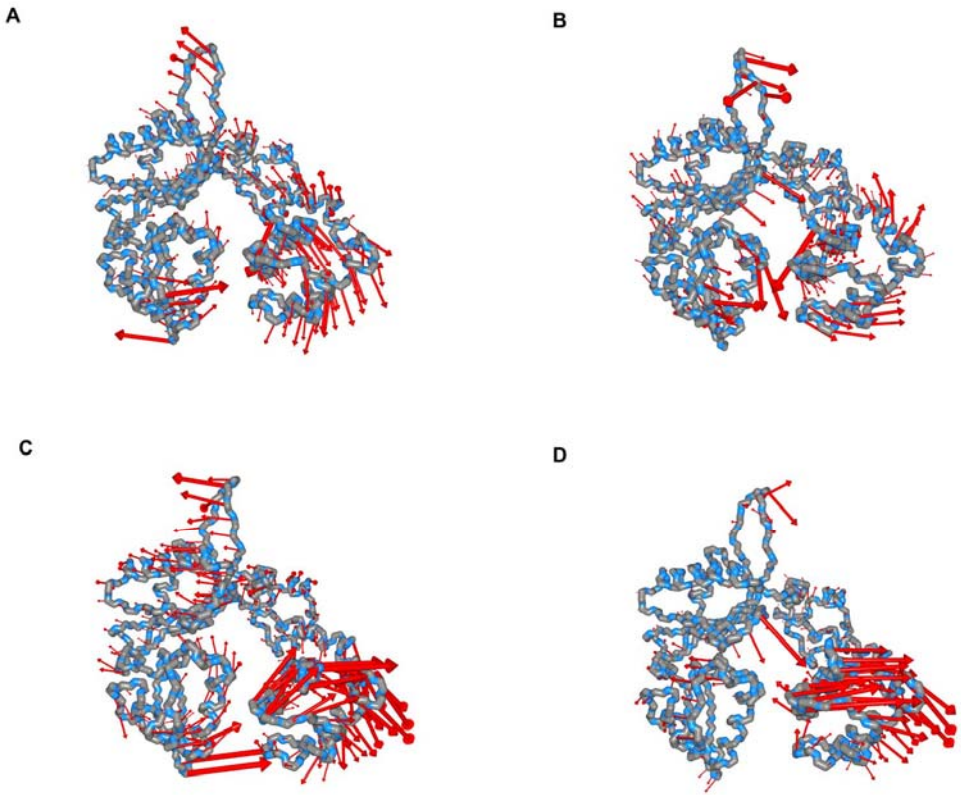


Figure B.8. The plot of the vector with the largest eigenvalue describing domain motion from essential dynamics analysis for Cα atoms of ddCTP complex (A), dCTP complex (B), ddCTP complex with substrate and Mg ions removed (C) and dTTP complex (D).

In addition to the MD simulations of fully solvated pol-β–gapped DNA–dNTP complexes, *ab initio* calculations were performed on models of the enzyme's polymerase active site. The active site contains two magnesium cations that bind the substrate nucleotide in the form of a nucleoside triphosphate (dNTP) and the 3' oxygen of the terminal nucleotide of the primer strand in the relative positions needed to facilitate the phosphorylation reaction via direct nucleophilic attack on the substrate α-phosphate. The structure of the di-magnesium coordination sphere found in the available crystal structures is inaccurate due to intentional removal of the 3' hydroxyl groups from the primer terminus and substrate to prevent complete reaction turnover. *Ab initio* optimizations on various active site models with and without the 3' hydroxyls were helpful, along with the MD simulations, in estimating the structural impact of these groups, particularly the primer 3'OH, which is believed to coordinate tightly to one of the cations. The role of water molecules in the coordination sphere was also explored in depth. Presently, calculations are being performed on a model system to resolve details of the nucleotidyl transfer mechanism, such as the identity of the proton acceptor in the activation of the primer 3'OH nucleophile. A number of plausible intermediates between the "reactant" and "product" coordination complexes have been identified for the model system (Figure B.9).

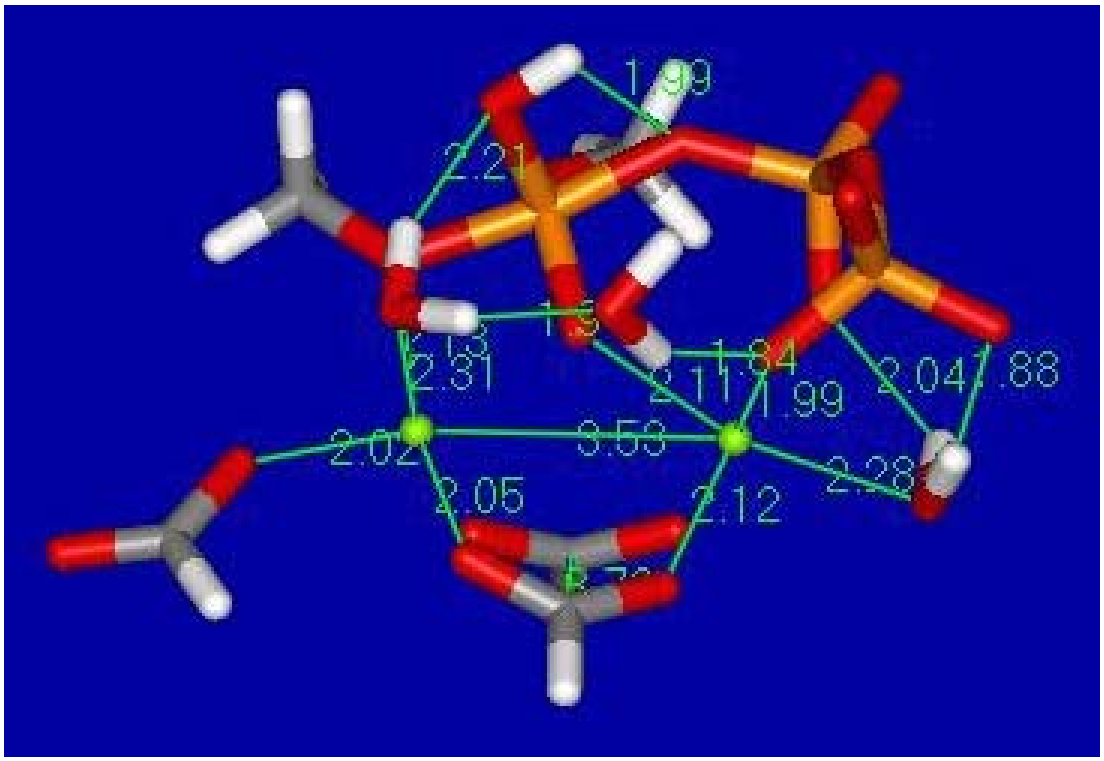


Figure B.9. Structure of an optimized pentacoordinate intermediate in the nucleotidyl transfer reaction for the pol-β active site model system.

Publication

Rittenhouse, RC, WK Apostoluk, JH Miller, and TP Straatsma. 2003**.** "Characterization of the active site of DNA polymerase β by molecular dynamics and quantum chemical calculation." *Proteins: Structure, Function, and Genetics***,** 53(3):667-682.

Protein Kinase A (J.A. McCammon, C.F. Wong)

We carried out more extensive docking studies of balanol to dynamics snapshots of protein kinase A (PKA). Two MD trajectories were used. A control simulation for the complex between PKA and balanol was used to check the reliability of AutoDock 3.05. The other simulation was carried out for the apo enzyme only. This simulation serves to test how well one can use dynamics snapshots of an apo protein to deduce the docking structure of a ligand to the protein. One has to face this situation when no experimental structure for a receptor-ligand structure is available. Three ligand models were used in the docking studies. In one, the ligand was completely rigid and its conformation was taken from the crystal structure of the complex. In another, the ligand was flexible by permitting the torsional angles along the length of the ligand to change. The third model was similar except that dihedral angle energy terms were added to AutoDock to see whether the docking results could be improved. We also used a Generalized Born model as well as surface-area-dependent hydrophobic terms and internal energy terms to rescore results from the flexible docking runs.

The results from the control simulation are provided in Table B.1. The ligand recovered the original structures within 0.27 ± 0.09 Angstrom on average. This may not be surprising because the protein snapshots already had a binding pocket that could accommodate the ligand well. Shape complementarity was sufficient to ensure correct docking. In fact, as seen in Table B.1, one could still recover the original structures after turning off all the charges on the ligand. The root-mean-square deviation (RMSD) only deteriorated slightly, although the binding affinity diminished by about 4 kcal/mol. Turning off charges only on selected functional groups produced similar results.

Table B.1. Docking of balanol to snapshots of PKA obtained from an MD simulation of the complex

Model Variant for Simulation	**Binding Energy (kcal)/mol**	**RMSD[1] (Angstrom)**
Balanol	-20.9 ± 0.6	0.27 ± 0.09
All ligand charge off	-16.9 ± 0.6	0.3 ± 0.1
NH_3^+ charge off	-19.2 ± 0.6	0.3 ± 0.1
CO_2^- charge off	-19.9 ± 0.6	0.26 ± 0.09
OH_4	-20.8 ± 0.6	0.3 ± 0.1
OH_5	-20.5 ± 0.5	0.26 ± 0.09
OH_6	-20.6 ± 0.6	0.28 ± 0.09
1. Root-mean-square deviation between the docked structures and their original structures in the snapshots.		

However, the results were drastically different when snapshots from the apo protein were used. The ligand was not always docked to the right place. For the flexible-ligand models, one can see from Figure B.10b that most of the dihedral angles used to characterize the conformation of the ligand sampled a large range. This suggests that the ligand bound to the receptor with many different conformations. Comparing the dihedral angles sampled in the docking runs with the corresponding angles in the crystal structure (shown in Figure B.10a), one can see that the docking conformation covered all but two of the observed dihedral angles. These two angles characterize the parts of the ligand that would become more exposed if the other end of the ligand docked correctly to the binding pocket. This suggests that the apo structure may not be able to support the observed conformation for this part of the ligand, probably because the two lobes of the kinase domain opened up somewhat in the apo form. The results obtained by using the flexible-ligand model with extra torsional energy terms were similar and are not shown here.

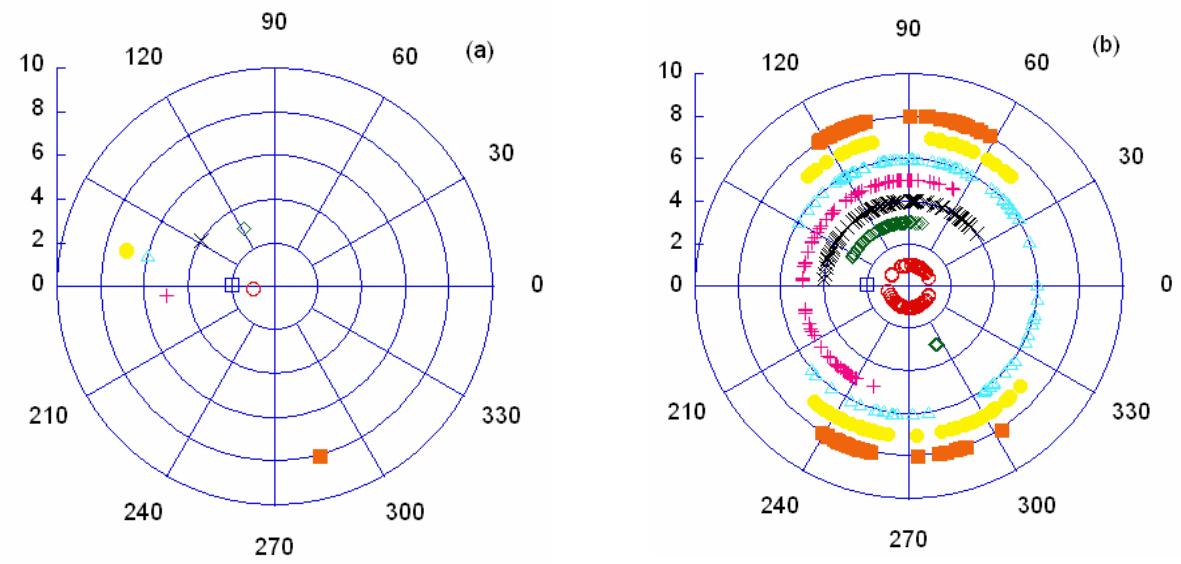


Figure B.10. Dihedral angles of balanol sampled in the docking runs: a) rigid-ligand model b) flexible-ligand model without torsional energy term.

In the crystal structure of the complex, the hydroxyl group of the hydroxylbenzamide moiety forms two hydrogen bonds with the linker region connecting the N-terminal and C-terminal lobes of the kinase domain. If one uses the distance between the hydroxyl oxygen and the peptide carbonyl oxygen of Glu 121 to approximately gauge whether the ligand docked correctly, one can see from Figure B.11 that the ligand could not always dock to its binding site when the rigid-ligand model was used. However, the best docked structure in the flexible-ligand models could always have this "hydrogen bond" distance within 6 Angstrom.

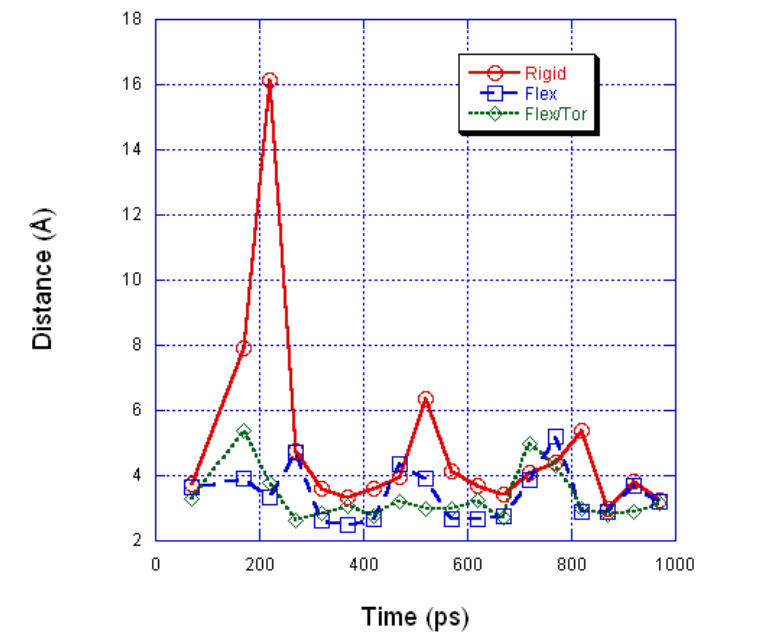


Figure B.11. Best docking structure for each snapshot.

Figure B.12 classifies the docked structures using the same distance criterion as in Figure B.11. For the rigid docking experiment, one can see that there were two major classes of structure; one class that docked correctly, the other that did not. In the flexible docking experiment, there were also two major classes of structures, but the class that did not dock correctly was somewhat different from that in the rigid-docking experiment. From Figure B.12, one can also see that about 50% of the docking runs gave structures that are close to the "correct" structure.

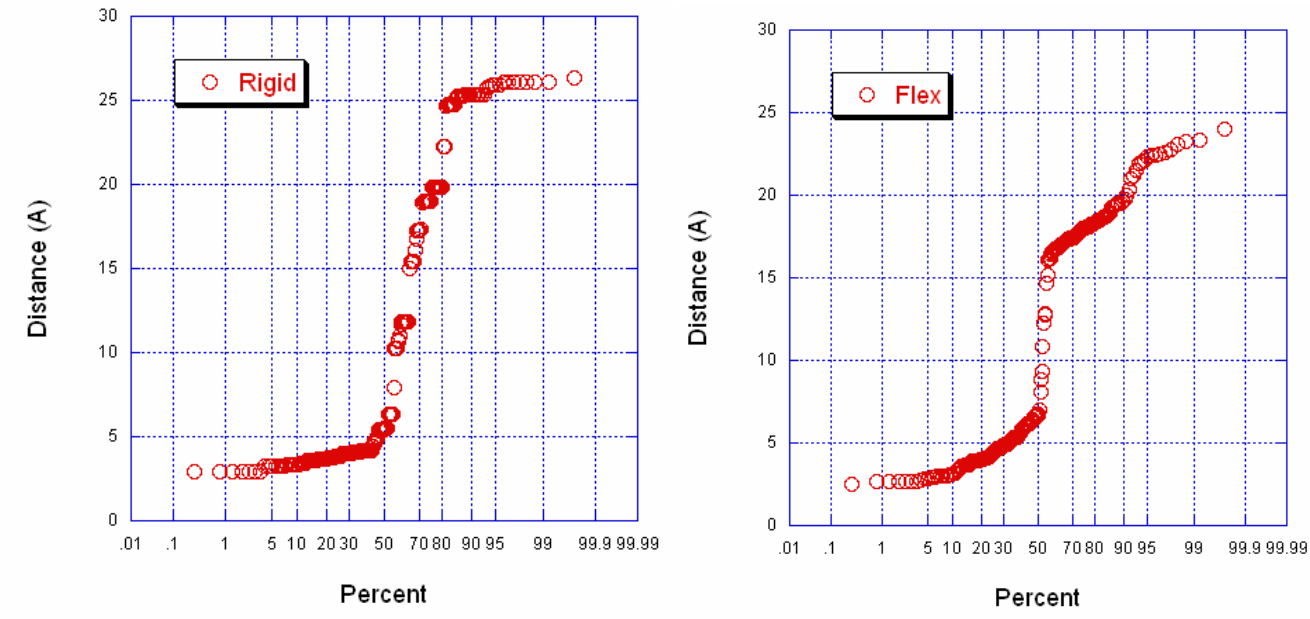


Figure B.12. Classification of docked structures.

As can be seen in Table B.2, BXa, BXb, …BXj have successively higher energy according to AutoDock. An entry with a 0 signifies an "incorrectly" docked structure. A non-zero entry indicates the RMSD of the dihedral angles of the ligand from those observed in the crystal structure of the complex. No torsional potential was used for the ligand in this set of docking runs.

Table B.2. "Successive" docked structures of balanol to snapshots of PKA

ps	BXa	BXb	BXc	BXd	BXe	BXf	BXg	BXh	BXi	BXj
70	84.4	102.5	86.1	102.8	88.5	0	0	0	0	0
120	0	0	0	106.3	0	0	0	0	0	0
170	0	0	86.2	72.8	74.2	0	0	0	0	0
220	62.5	93.2	0	0	0	0	0	95.7	83.4	81.9
270	95.2	84.5	0	74.9	0	58.2	63.9	0	0	0
320	72.3	89.1	83.2	104.1	59.9	62.8	97.4	84.4	87.9	0
370	60.3	67.6	95	84.2	89.3	88.4	0	0	0	60.3
420	0	0	0	0	0	51.2	100.4	0	117.8	0
470	0	66.5	0	0	0	62.8	89.5	0	0	0
520	53.3	0	84.7	0	0	0	0	0	84.5	0
570	65.5	67.1	54.1	0	0	105.1	69.3	101.5	0	0
620	64.1	78.7	0	0	0	81.6	0	0	57	0
670	76.4	91.2	75.4	0	0	0	0	54.4	71	0
720	59.6	54	97.7	0	96	0	56.9	0	0	0
770	0	0	0	0	0	82.4	0	0	97.6	102.8
820	0	0	0	0	0	88.9	61.2	35.9	106.4	0
870	62	95.7	97.9	94.9	72.6	0	76.5	93	111.3	0
920	64.5	0	90.9	78	62.9	0	0	0	0	0
970	85.5	0	0	98.4	0	0	0	65.8	38.6	0

To further characterize the docking structures, Table B.3 presents the characteristics of different docked structures for each snapshot. One can see that out of the 10 runs for each snapshot, there was at least one structure that docked correctly. However, the correctly docked structure did not always have the lowest energy. Also, even for structures that docked roughly to the right place, their conformations deviated quite far from their conformations in the crystal structure of the complex. The table lists AutoDock energy in ascending order within each column. The numbers on the left indicate the "hydrogen bond" distances. A smaller number corresponds to a structure that docks more closely to the true binding site. The numbers on the right were obtained from the Generalized Born model plus surface-area dependent hydrophobic terms (GBSA) model with internal energy terms, with the exception of the Lennard-Jones terms.

Table B.3. Rescoring of AutoDock structures with a GBSA model

70 ps	120 ps	170 ps	220 ps	270 ps
3.7 -7265.9	5.5 -7166.9	11.8 -7279.4	19.9 -7170.0	19.0 -7126.5
3.8 -7265.9	5.5 -7168.5	11.8 -7279.3	19.9 -7170.0	19.0 -7126.7
3.7 -7265.8	5.5 -7167.0	11.8 -7279.3	19.9 -7169.9	19.0 -7126.2
3.8 -7265.7	5.5 -7167.2	11.8 -7279.3	19.9 -7170.0	19.0 -7126.4
3.8 -7265.5	11.7 -7158.3	11.8 -7279.4	19.9 -7169.9	19.0 -7126.4
3.8 -7265.8	25.3 -7142.1	11.8 -7279.2	19.9 -7170.1	19.0 -7126.9
25.3 -7245.5	25.9 -7157.2	11.8 -7278.9	19.8 -7169.9	19.0 -7126.3
25.3 -7245.4	25.9 -7157.1	11.8 -7279.2	19.8 -7169.9	19.0 -7127.5
25.3 -7245.3	25.9 -7156.8	7.9 -7286.7	19.8 -7170.8	4.7 -7142.8
25.4 -7245.4	18.9 -7152.1	25.1 -7292.0	16.1 -7198.8	4.8 -7144.1

320 ps	370 ps	420 ps	470 ps	520 ps
3.6 -7197.9	3.4 -7356.5	3.6 -7320.2	4.0 -7337.3	6.4 -7195.3
3.6 -7197.5	3.4 -7356.4	3.6 -7319.6	4.0 -7337.3	6.4 -7195.3
3.6 -7198.3	3.4 -7356.6	3.6 -7320.4	4.0 -7337.5	6.4 -7195.0
4.1 -7205.6	3.4 -7356.2	3.7 -7320.7	4.0 -7337.2	6.4 -7195.3
4.9 -7213.6	3.4 -7357.2	3.6 -7319.5	4.0 -7336.7	6.4 -7195.1
10.3 -7208.8	3.4 -7356.4	3.7 -7320.2	4.0 -7337.5	22.2 -7189.4
10.2 -7208.4	25.2 -7341.4	3.6 -7319.8	4.0 -7337.2	22.2 -7189.6
10.2 -7208.1	25.2 -7341.5	3.6 -7320.7	4.0 -7337.5	17.3 -7183.6
10.2 -7208.2	25.2 -7341.4	3.7 -7318.8	4.0 -7337.3	17.3 -7183.8
24.7 -7206.6	25.2 -7341.7	11.0 -7318.2	4.0 -7337.0	17.3 -7183.2

570 ps	620 ps	670 ps	720 ps	770 ps
4.1 -7308.8	4.2 -7338.5	3.4 -7444.7	4.1 -7349.2	4.4 -7249.4
4.1 -7308.8	4.2 -7338.5	3.4 -7444.7	4.1 -7349.6	4.4 -7249.6
4.1 -7308.7	4.2 -7338.5	3.4 -7444.6	4.1 -7349.9	16.8 -7250.7
4.1 -7308.7	4.2 -7338.5	3.4 -7444.4	17.3 -7346.1	26.1 -7240.6
4.1 -7308.8	4.2 -7338.5	24.7 -7414.7	17.4 -7346.8	26.1 -7240.7
4.1 -7308.7	4.2 -7338.3	24.7 -7414.7	25.4 -7317.9	26.1 -7240.9
4.1 -7308.7	4.2 -7338.4	24.7 -7414.9	25.4 -7317.9	26.1 -7241.2
4.1 -7308.7	3.7 -7333.3	24.7 -7415.1	25.4 -7317.7	26.1 -7240.7
4.1 -7309.0	3.7 -7333.3	24.7 -7415.8	25.4 -7317.7	26.1 -7241.0
4.1 -7309.1	4.6 -7332.2	24.7 -7415.3	25.4 -7317.9	26.1 -7240.7

820 ps	870 ps	920 ps	970 ps
5.4 -7188.3	3.0 -7373.1	3.8 -7389.1	3.3 -7276.9
5.4 -7188.2	3.0 -7373.1	3.8 -7389.0	3.3 -7277.6
5.4 -7188.6	3.0 -7373.3	3.8 -7389.0	3.3 -7277.3
5.4 -7188.3	3.0 -7373.1	3.8 -7389.0	3.2 -7278.6
5.4 -7188.4	3.0 -7373.1	3.8 -7388.9	3.2 -7277.7
5.4 -7188.2	3.0 -7373.8	3.8 -7389.0	3.2 -7278.6
5.4 -7188.2	3.0 -7373.1	15.4 -7364.6	3.3 -7276.9
6.3 -7179.2	10.8 -7364.4	15.4 -7364.6	5.0 -7259.8
10.6 -7177.6	15.0 -7335.7	15.4 -7364.7	25.8 -7253.2
26.3 -7212.9	15.0 -7343.4	15.4 -7364.4	25.8 -7253.8

To examine whether a more sophisticated GBSA model can identify the correctly docked structure better, we rescored the AutoDock structures using this energy model. When we included all internal energy terms in calculating the energy of the docked structures, the GBSA model gave rankings that were almost completely opposite to those obtained from AutoDock. However, if one removed the Lennard-Jones terms from the energy calculations, the results were much more consistent (Table B.3). Although there were cases that the GBSA model gave the lowest energy for the correctly docked structures when AutoDock did not, this was not always the case. Therefore, even the GBSA model could not identify the correctly docked structure. It is possible that structures that more closely resembled the complex structure may not have the lowest energy when dynamics snapshots of the apo protein were used. If one examines results from all the snapshots, the correctly docked structures appeared more frequently as the lowest energy structures. Thus, one way to identify the correctly docked structure from such docking-to-snapshots experiments may be to identify the structure that has the lowest energy more frequently. At any rate, further refinement of auto docked structures seems necessary as the conformation of the ligand obtained from docking usually deviated a lot from the crystal structure. The refinement will most likely require simulations that treat receptor and ligand flexibility simultaneously.

Appendix C – Full report of third year activities and accomplishments

Human Polymerase-β Study (T.P. Straatsma, J.H. Miller)

Environmental factors, such as ionizing radiation, contribute to continuous damage of cellular DNA, in addition to endogenous sources. The damage resulting from oxidative stress and ionizing radiation is primarily in the form of oxidized bases, single strand breaks, and loss of bases. These are the targets of the base excision repair (BER) mechanism enzymes, including polymerase-β (pol-β). Failure to correctly and timely repair these damaged DNA sites can result in cell death, carcinogenesis, or genetic diseases. Resulting mutations in cell signal transduction enzymes can lead to uncontrolled cell proliferation or differentiation. For example, mutations in Ras, the molecular switch in several growth-factor signaling pathways, have been found in about 30% of human tumors. These signaling pathways often involve a chain of protein kinases that activate or deactivate each other through phosphorylation reactions, eventually controlling the activation of transcription factors in the cell nucleus.

In the third year of this project, results were analyzed of a molecular dynamics (MD) study designed to determine the conformational response of the closed thumb pol-β-gapped DNA complex to the removal of the substrate, and also to the replacement of the correct substrate with the mismatched substrate dTTP. Since the 1BPY crystal structure used as the basis for the simulation starting structures lacks the 3' hydroxyl groups on the primer and substrate nucleotides, preliminary simulations were performed on the complex with the ddCTP substrate analogue and the restored dCTP substrate to obtain a best estimate of the native complex structure. Since there are no crystal structures available for pol-β ternary complexes with a bound mismatched substrate, we believe these simulations shed light on the probable response of the enzyme-DNA complex to the presence of an incorrect nucleotide in the active site, and thus provide a means of testing one proposed mechanism of fidelity assurance.

The starting structures for the simulations in this study were all derived from a crystal structure of a human pol-β ternary complex with gapped DNA, the ddCTP (dideoxyribonucleoside cytosine triphosphate) substrate analogue, and two magnesium ions in the active site (PDB entry 1BPY; Sawaya 1997). This crystal structure represents the mid-point of the polymerase catalytic cycle, just prior to the nucleotidyl transfer step. It differs, however, from the "real" mid-point structure, because of intentional alterations involving removal of the primer and substrate C3' hydroxyls to prevent the chemical step. Recent work has shown that these hydroxyl groups are very important and significantly influence the structure of the active site (Rittenhouse et al. 2003).

Crystal versus Simulated ddCTP and dCTP Ternary Complex Structures

Of all the available crystal structures, the 1BPY structure of polymerase-β complex with gapped DNA and the ddCTP substrate analogue (Sawaya 1997) most accurately represents the structure of the natural complex in mid-catalytic cycle just prior to the chemical nucleotidyl transfer step. The crystallographers were able to prevent the chemical step, and thereby obtain the structure, by eliminating the 3' hydroxyl group on the terminal nucleotide of the primer strand and also the substrate (converting dCTP into ddCTP). In order to provide a baseline for a comparison of simulations where the starting structure was altered significantly from the 1BPY crystal structure, for example, by inserting nucleotide mismatch or removing the substrate entirely, MD simulations were performed on the crystal structure as is, except for protonation and solvation, and on the same structure with the 3' hydroxyl groups restored to the primer terminus and the substrate.

As expected, the differences observed between the ddCTP and dCTP simulations were mostly local to the active site region. The structures obtained at the end of the simulations were found to be closer to each other than to the 1BPY crystal structure. The heavy atom RMSD values for the ddCTP and dCTP simulations with respect to the 1BPY crystal structure are 2.69 and 2.85 Å, respectively. With only protein-heavy atoms considered, the RMSD values are about 0.65 less, well within what typically is observed for MD simulations of proteins starting from crystal structures. In comparison, the heavy atom RMSD value between the two simulated structures is 2.25 Å. This is not surprising since the ddCTP and dCTP complexes differ by only 2 atoms out of more than 3000 heavy atoms. Comparisons between simulated and crystal structures are always complicated by the difference in the physical state represented by the structures, *i.e.,* solution versus solid state, in which the effects of crystal packing forces are difficult to estimate. In addition, differences between the crystal and simulated structures may result from general imperfections of MD methods related to the usage of empirical force fields.

Within the active site region of the enzyme, particularly the coordination sphere of the two magnesium ions, the simulated structures differ from each other as expected, but also from the crystal structure. In both simulations, a water molecule coordinates tightly to each of the magnesium ions (Figure C.1). The water molecule coordinated to the nucleotide-binding cation is present in the crystal structure (#530), but not the water molecule bound to the catalytic ion. However, the crystallographers have postulated the existence of this coordinated water molecule (Sawaya 1997), and when it is included, both magnesium ions are hexacoordinated. In the simulations, the starting structures lack this ligand, but a water molecule migrates into this position early in the simulation protocol.

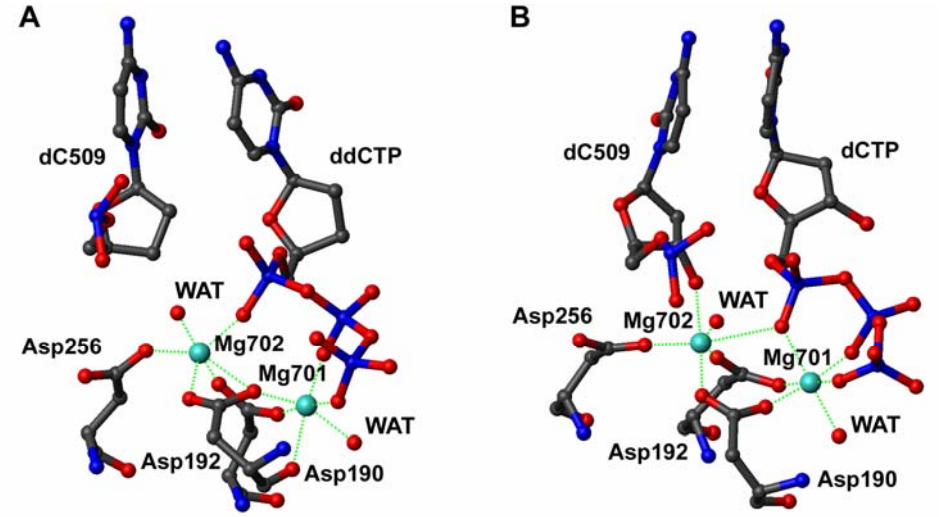


Figure C.1. Coordination spheres of the magnesium atoms (colored in cyan) in ddCTP complex (A) and dCTP complex (B) after a 1.7 ns molecular dynamics simulation. Water molecule coordinating triphosphate binding ion (Mg701) corresponds to crystal water #530.

The coordination sphere of the catalytic magnesium ion observed in the ddCTP simulation is very close to that observed in the crystal structure, assuming the coordinated water molecule. The metal-ligand distances are somewhat shorter in the simulation, but the identity and geometry of the coordinated atoms are the same. In both structures, the catalytic ion coordinates to both carboxyl oxygen atoms of Asp190 with the OD1 oxygen atom acting as a bridging ligand by binding also to the nucleotide-binding cation (Figure C.1A). The coordination sphere of the nucleotide-binding ion, as observed in the ddCTP simulation, is not in complete agreement with the crystal structure. Specifically, the O1A oxygen atom

from the substrate's α-phosphate is a bridging ligand coordinated to both magnesium ions in the crystal structure, but is found in the simulation to be coordinated only to the catalytic ion. Lacking the substrate O1A oxygen atom, the nucleotide-binding ion coordinates to the backbone carbonyl oxygen atom of the Asp190 residue (Figure C.1A).

In the dCTP simulation, the primer terminus coordinates tightly to the catalytic ion through its restored 3' hydroxyl oxygen atom, as expected (Figure C.1B). This causes the primer 3' carbon atom to move about 0.5 Å closer to the substrate α-phosphorus atom. As a result of the added ligand, the catalytic ion releases its hold on the Asp190 OD1 oxygen atom, which then binds only to the nucleotide-binding ion. The coordination sphere of the nucleotide-binding ion, as observed in the dCTP simulation, is identical in ligand identity and geometry to that found in the crystal structure, with only small differences in metal-ligand distances. Thus, surprisingly, the active site structure from the dCTP simulation is in somewhat better overall agreement with the crystal structure than is the active site structure taken from the ddCTP simulation. One plausible explanation for this unexpected result is that the non-bonded model employed in the force field for the magnesium cations slightly overstates the charge by neglecting the ligand-to-metal charge transfer and polarization effects. With the absence of the 3' hydroxyl ligand in the ddCTP simulation, this additional charge may cause the catalytic ion to pull the substrate O1A oxygen atom closer to itself, and thus away from the other cation. The nucleotide-binding ion then, seeking the most available replacement, binds to the Asp190 carbonyl oxygen atom. This explanation is supported by comparisons of the simulated active site structures with *ab initio* optimized models that were performed (Rittenhouse et al. 2003) to confirm the protonation states of the acidic ligands and explore the effects of the restored 3' hydroxyl groups. The calculations suggest that the structure of the active site of the native complex just prior to the chemical step should closely resemble simulated structure of the dCTP complex with slightly longer metal-ligand distances and a somewhat shorter inter-magnesium distance.

Besides the active site, there are some other regions of the dCTP complex where average inter-atomic distances in the simulated structures differ from the corresponding distances in the crystal structure; however, only in a few cases does the difference exceed 0.5 Å. Interestingly, all of these distances involve contacts by atoms of the thumb subdomain. Two of these are with the template nucleotide dC407; eight are with the 8-kDa domain residue Asn37, Arg40, Lys41, Ser44, and Lys48; three are with the palm subdomain residues Thr176, Arg182, and Arg258; and finally one is with the substrate base. Noting that these differences appear in both simulated structures of the ddCTP and dCTP complexes, we attribute them to the force field effects rather then, for instance, the influence of the missing 3' hydroxyl groups in the ddCTP complex. However, there are some distances that differ significantly between the two simulated complexes. The largest difference of 6 Å appears in the distance between the Glu335 OE1 carboxyl oxygen atom and Lys41 NZ nitrogen atom (Figure C.2). It appears from the simulation that contrary to the crystal structure, Lys41 residue might play an important role on the thumb – 8-kDa interface. A comparison of this distance in the simulated native dCTP complex to the corresponding distances in the ddCTP and mismatch simulation, as well as crystal structures of closed and opened complexes, suggests that the native dCTP is much stronger. This implies that this particular inter-domain interaction is somehow influenced and weakened by changes in the active site resulting from the missing 3' hydroxyls in the ddCTP complex or the binding of mismatched nucleotides. The other distances differing between the ddCTP and dCTP simulations are the contacts between the Tyr265 and Thr176 hydroxyl oxygen atoms, and between the Asp276 carboxyl oxygen atom OD2 and the Arg40 nitrogen atom NH1. Since the Tyr265 atom is the first residue of the thumb domain immediately following the hinge region, and the Asp276 – Arg40 and Glu335 – Lys41 pairs form the key interactions on the thumb

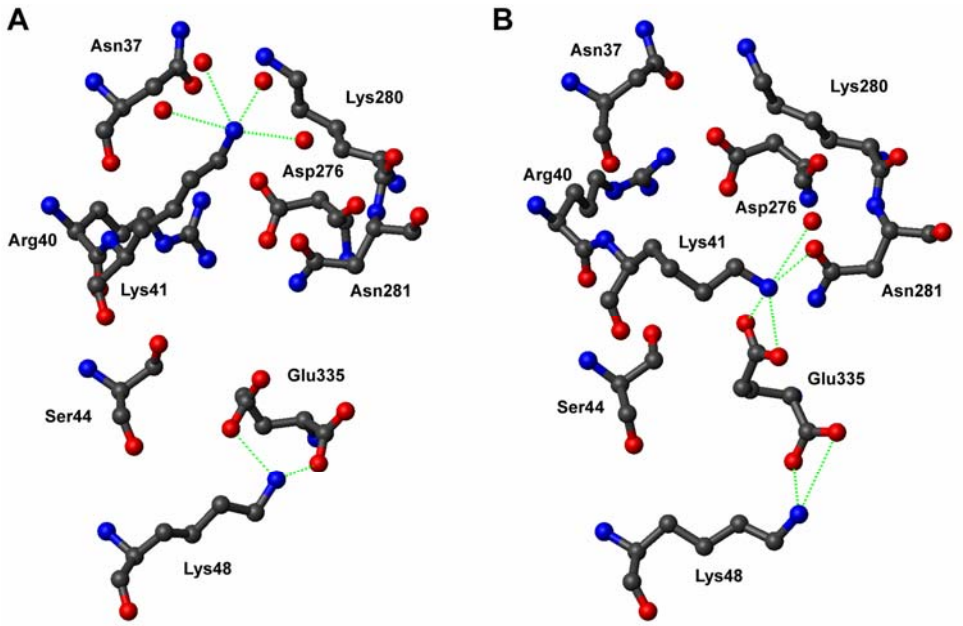


Figure C.2. Thumb – 8-kDa domain interface in the ddCTP complex (A) and dCTP complex (B). Dashed lines point to interaction partners of Lys41 and Lys48.

– 8-kDa domain interface, an increase in these distances may be one of the first symptoms of thumb domain opening.

Ternary Complex with Substrate Removed

To test the proposed "induced fit" mechanism of fidelity assurance in DNA repair performed by pol-β, we performed a simulation of the pol-β closed ternary complex (1BPY) with the ddCTP substrate analogue and both magnesium ions removed from the binding site. Without the substrate, the structure is expected to revert back to an "open thumb" conformation. We assume that, if transitions between open and closed forms do not involve high-energy barriers, the rate-limiting step for the nucleotide transfer reaction would be determined by the stabilization of the transition state during the chemical step. This alternate mechanism of assuring polymerase fidelity has been proposed by Showalter and Tsai (2002). However conformational changes involving large domains can be "very" slow, and, in cases where it takes more than tens of nanoseconds, are beyond the treatment with statistical mechanics methods.

Recent work of Yang et al. (2002a) on the pol-β system has shown that there is no significant thumb subdomain opening within a 5.0-nanosecond simulation starting from the post-chemical, closed structure, *i.e.*, the ternary complex modified by reacting the α-phosphate group of ddCTP with the 3' hydroxyl group of the DNA primer strand. According to their estimations, based on rotation rate constants for particular active site residues, the pol-β conformational change from closed to open state, including disassembly of the active site residues (especially rotation of Arg258), may take up to 10 ms. On the other hand, Kim et al. (2003) estimated that movement of the 8-kDa domain alone may occur on a 3 to 8 ns time scale.

Our analysis of the inter-domain distance, as well as protein radius of gyration, shows no significant domain opening within 5.2 ns of simulation. However, from partial opening of the domain (from ~2.45 nm to ~2.65 nm) occurring between 2 and 3.5 ns, an estimate can be made; *i.e.,* with continued opening motion, the complete opening of the thumb (in which the thumb subdomain to 8-kDa domain center of mass distance increases by almost 5 Å) would take at least 5 nanoseconds.

In order to extract possible thumb subdomain movements from the configurational space explored during the simulation, essential dynamics analysis was applied. This method allows separate anharmonic motions relevant for the protein function from irrelevant local fluctuations. About 90% of the total motions in the system were described by about 70 out of the 978 eigenvectors representing coordinates of the protein C_α atoms and 150 out of the 7,842 eigenvectors representing coordinates of all protein-heavy atoms, which is slightly more than for previously studied systems (Amadei 1993; Stella 1999; Arcangeli 2001). The motions along the two eigenvectors, with the largest corresponding eigenvalues, were obtained from the covariance matrix of the pol-β C_α atoms positional deviations and the corresponding probability distributions. Although atomic displacements obtained from the analysis eigenvectors show a range of fluctuations similar to previously studied systems (Amadei 1993; Stella 1999) they do not transpose in this case into domain movement.

Since our simulation was too short to observe domain rearrangement, our analysis focused on the interfaces between particular domains to determine occurrence of minor changes in conformation prior to the opening of the thumb domain. Inspection of the thumb domain contacts in the crystal structures of closed and open forms revealed that most of the contacts, especially with the DNA, are weakened when in the open form. Interestingly, during the opening process, two new interactions involving the thumb domain apparently are formed. One of these contacts is between the OG1 hydroxyl oxygen atom of Thr273 and the NH2 nitrogen atom of the palm domain Arg182, and the other is between the OE2 carboxyl oxygen atom of Glu335 and the NZ nitrogen atom of 8-kDa Lys48. To further investigate this process, we analyzed the thumb domain distances to DNA as well as to the 8-kDa domain. Inspection of the distances between the thumb domain and DNA revealed that most of the interactions present in the closed form were weakened during the simulation, whereas some of them were fixed or even slightly strengthened (especially the interaction of the Asn294 nitrogen atom ND2 with the dC407 oxygen atom O2P). Based on the contacts with the 8-kDa domain, it appears that, although there is no dramatic domain movement within 5.2 ns of simulation, local rearrangements of the thumb – 8-kDa domain interface do occur. There is a weakening of the interaction between the thumb Asp276 and 8-kDa Arg40 and, although the interaction of at least one of the Glu335 carboxyl oxygen atoms with the OG hydroxyl oxygen atom of Ser44 remains present, a decrease in the distance of both carboxyl oxygen atoms to Lys48 was observed. The removal of the substrate and magnesium ions from the active site resulted in the relaxation of the thumb domain contacts. The contact formed between Glu335 and Lys48 may be a first signal for the thumb domain opening.

A thumb opening motion was observed by Yang (Yang 2002a), but only after use of a "half open" intermediate structure obtained by averaging the open and closed forms. Their findings suggest also that the factor limiting the thumb subdomain opening could be the rotation of active site residue Arg258 and the release of the catalytic Mg^{2+} ion. The rotation of Arg258 is coupled to the flipping of the Phe272 side chain, thereby generating the space that is required for Arg258 to rotate. Although they did not observe rotation of Arg258 in their simulation, they were able to capture movement of the Phe272 ring away from Asp192 (Yang 2002a). In our simulation, we did not observe a change in the Phe272 residue

conformation nor a rotation of the Arg258 side chain toward Asp192. Although residues Arg258 and Asp192 do approach each other, the distance observed is still far from that found in an open thumb structure and therefore provides insufficient evidence of domain opening in the current simulation.

Simulations of Mismatched dNTP Complexes

dTTP

While, there is no steric interference between the guanine base of the template nucleotide dG405 and the thymine replacing the dCTP cytosine in the initial modeled structure of the mismatched dTTP complex, the thymine base is shifted to match a Wobble base pair hydrogen bond pattern. Two hydrogen bonds with the template are formed; one between the N_3 nitrogen atom of thymine and the O_6 oxygen atom of guanine, the other between the O_2 oxygen atom of thymine and the N_1 nitrogen atom of guanine. Similar to the ddCTP and dCTP complexes, the O_2 oxygen atom of the thymine base is hydrogen-bonded to the ND_2 nitrogen atom of Asn279, and there is van der Waals interaction of the thymine base ring with the side chain of Asp276. Also similar to the dCTP complex, the NH_1 nitrogen atom of Arg183 spans the hydroxyl oxygen atoms O3' and O1B of the dTTP substrate.

The value of the root-mean-square deviation (RMSD) of the dTTP complex structure with respect to the initial structure was larger than that observed for the ddCTP and dCTP complexes, but only for the protein-heavy atoms. When the whole complex or just the DNA are considered, the RMSD values are still higher than those for the ddCTP complex, but lower than those for the dCTP complex, suggesting that the DNA in this simulation is less mobile than in the simulation of the dCTP complex. However, when the ddCTP simulated structure is used as a reference, the RMSD values for the dTTP complex and DNA are about the same range as for dCTP.

Just as in the case of the closed ternary complex with the substrate removed, the presence of a mismatched nucleotide in the active site is expected to result in the opening of the thumb domain, in accordance with the proposed "induced fit" mechanism. Although the protein radius of gyration shows the same range of changes as in the ddCTP and dCTP complex simulations, it looks like the inter-domain distance increases more than in other cases. However, we performed an essential dynamics analysis of the protein domains showing a similar atomic displacements range to that in the simulation of ternary complex with substrate removed. This did not reveal any persistent motions of the thumb domain in the dTTP simulation trajectory. A detailed analysis of the simulated dTTP complex structure was also performed and compared with the dCTP complex. Several structural differences were found, within the active site, on the interface of the thumb and 8-kDa domains, and in the conformation of the triphosphate, itself. Unlike the simulation of the dCTP complex, where the O3' hydroxyl oxygen atom alternated between two hydrogen bond acceptors (the O1B and O3A oxygen atoms with minimum distances of 2.6 and 3.0 Å, respectively), the O3' hydroxyl oxygen atom of the dTTP substrate formed a more persistent contact to the O1B oxygen atom with a distance of 2.6 Å.

Since the thymine base in the dTTP complex does not match the template guanine, the distance between the bases, due to repulsion between them, is slightly larger compared to the complex of the correct base pair. However, when compared with the complex with the correct dCTP nucleotide bound in the active site, incorporating the incorrect nucleotide does not appear to affect the structure of the triphosphate group of dTTP, nor most of the distances between the triphosphate atoms and the surrounding atoms of the protein and DNA. The only significant differences are the distance between the O1B oxygen atom of

dTTP and the NH2 nitrogen atom of Arg183, which is 3.59 Å (compared to 3.12 Å for the dCTP complex), and the distance between the O2 oxygen atom of the thymine and the ND2 nitrogen atom of Asn279, which is 3.30 Å (compared to 2.83 Å in the dCTP complex). Other differences caused by the introduction of dTTP in place of dCTP are more distant from the substrate and concern protein – DNA contacts, such as the distance between the OG1 oxygen atom of Thr292 and the backbone O2P oxygen atom of the template residue dG406, as well as contacts between the thumb domain and the 8-kDa domain, including electrostatic interactions of Glu335 with Lys41 and Lys48, and van der Waals interactions between Glu335 and Ser44 (Figure C.3).

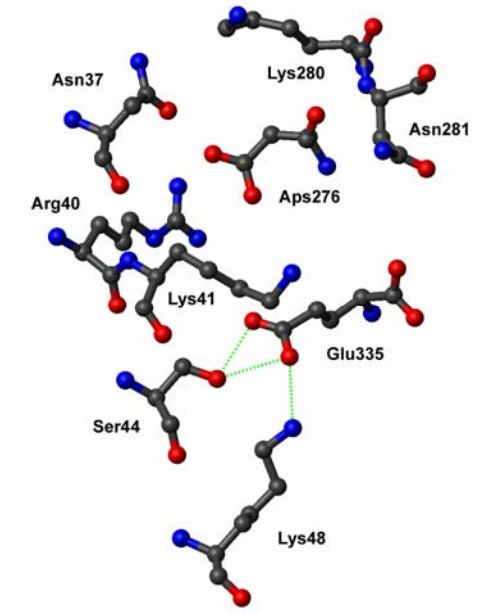


Figure C.3. Thumb – 8-kDa domain interface in the dTTP complex (dashed lines point to interaction partners of Lys41 and Lys48)

Both magnesium ions present in the active site of the dTTP complex are hexacoordinated, with a distance between them comparable to that observed in the dCTP complex. However, contrary to the dCTP complex simulation, the coordination sphere of the catalytic magnesium in the dTTP complex is filled by a non-crystal water molecule. Moreover, the triphosphate-binding magnesium ion stays within about 2.0 Å distance of the O1A oxygen atom of the substrate, whereas in the dCTP complex simulation after 1.2 ns, the magnesium ion starts to fluctuate more and reach even 3.2 Å. Another difference is the much earlier occurrence of the movement of the backbone carbonyl oxygen atom of Asp190 toward the nucleotide-binding ion. In both complexes, this probably results from the rearrangement of one of the water molecules moving away from the hydroxyl oxygen atom of Ser188, allowing a shorter hydrogen bond to the backbone carbonyl oxygen atom of Asp190. This would seem to explain why the average position of the carbonyl oxygen atom of Asp190, unlike the ddCTP case, is shifted so that the distance to the triphosphate-binding magnesium increases slightly and starts to fluctuate more.

dATP and dGTP

We attempted also to model the dATP and dGTP mismatches. The setup of these complexes required manual adjustments to the position of the dATP/dGTP substrate base to avoid steric clashes with the template guanine and nearby Asn279. Since the DNA bound by pol-β is kinked to enable contact

between thumb domain and template base and to make the nucleotide gap accessible for an incoming substrate (Sawaya 1997), there is some relatively vacant space adjacent to the template guanine across from the downstream oligo. We decided to use this available space to position the bases of the dATP and dGTP substrates in the downstream oligo direction and thus avoid steric conflicts.

In the case of the dATP complex, the substrate base during the heating phase of the simulation passed the template guanine and moved back toward the primer strand, forming a stacking interaction with cytosine dC509. This brought the N6 nitrogen atom of dATP into the close vicinity of the hydrogen bonded pair, N4 of dC509 and O6 of dG406. These movements resulted in significant distortions of the DNA, especially the template guanine dG405, in which the purine ring had to move in the direction of the upstream (5' end) template strand.

With respect to the dGTP complex, it appears that there is not enough space for the substrate guanine base to pass by the template nucleotide and move back toward the primer. Instead, the template base moved toward template guanine dG406, and dGTP base basically remained in its originally placed position, with some movement in a plane away from Asn279. Unlike the dGTP mismatch opposite template dGTP studied by the Schlick group (Yang 2002b), which was most distorted among the mismatches they studied, our dGTP complex showed less distortion not only for the DNA, but for the entire complex.

The different behavior of the mismatched guanine pair in both simulations may have resulted from using, in the current work, the pre-chemical step structure instead of the post-chemical step used by the Schlick group. However, it may have come from our procedure of modeling in the mismatched substrates, which fixes it in active site and locks the triphosphate in a reactive conformation. A more appropriate modeling method would be to put the substrate into a binding site of an open form of the enzyme. This will be our future goal.

In our work, MD simulations were performed on ternary complexes of pol-β bound to gapped DNA and each of the four possible nucleoside triphosphate substrates (dNTP). A simulation was also carried out on a binary complex obtained by removing the substrate and magnesium ions from the ternary complex. All of the starting structures for the simulations were derived from a crystal structure of a ternary human pol-β – gapped DNA – ddCTP complex (PDB entry: 1BPY), in which the thumb subdomain is positioned in its closed conformation over the ddCTP substrate analog. As a control, a simulation was performed on the crystal structure as received, except for protonation and solvation. Since the crystal structure was intentionally altered from the native complex by removal of the 3' hydroxyl groups from the primer and substrate, a comparison of the ddCTP and dCTP simulations was helpful in determining the effects of the missing 3' hydroxyl groups and estimating the active site structure of the native complex.

The starting structure that differed most radically from the crystal structure was the binary complex, modeled by removing the ddCTP substrate. Therefore, this simulation was extended for the longest duration to search for clear evidence of thumb opening. According to the proposed "induced fit" mechanism of fidelity assurance, the thumb subdomain is expected to reopen if the correct substrate (matching the template base) is not in place. Both structural and essential dynamics analyses of the 5.2 ns trajectory failed to provide convincing proof of thumb subdomain movement. However, certain specific interactions across the thumb – 8-kDa domain interface showed signs of weakening as the simulation progressed. Kim et al.(2003) estimate that motion of the 8-kDa lyase domain in solution probably occurs on a time scale between 3 and 8 ns. Since opening of the thumb requires the breaking of multiple attractive interactions, a 5.2 ns simulation may well be too short to observe a definite opening movement.

Since we were unable to find any crystal structures of mismatched pol-β complexes, the starting structures used in the simulations of ternary complexes with mismatched substrates opposite the template guanine were created by modeling the mismatched base into the ddCTP structure and restoring the substrate and primer missing the 3' hydroxyl group. Though the essential dynamics analysis revealed no clear evidence of thumb movement in the 1.7 ns simulations, several differences were observed in specific interactions between the substrate and certain nearby residues. The binding of the triphosphate portion of the substrate to the magnesium ions in the active site was found to be very similar among the mismatches and the correct dCTP substrate. However, some peripheral active site differences were noted in details, such as the hydrogen bonding partners of the substrate's 3' hydroxyl group. In the case of the purine mismatches, dATP and dGTP, the DNA accommodated the mismatch by moving the template base to provide room for the substrate to assume its normal position. This response may be a result of the lack of flexibility at the other end of the substrate due to the tight binding of the triphosphate moiety to the magnesium ions. We are exploring other approaches to the insertion of substrate mismatches.

None of the simulations performed provides clear evidence of thumb opening in response to substrate substitution or removal, and thus the proposed "induced fit" fidelity mechanism is not supported by these results. However, this mechanism cannot be rejected on this basis, either, since the energy barrier to thumb opening may be of a magnitude that would require longer simulation times to observe. Recently, Showalter and Tsai (Showalter 2002) reported evidence that the rate-limiting step in the pol-β catalytic cycle is the nucleotidyl transfer reaction rather than a conformational change. They have suggested that nucleotide selection occurs by means of transition state stabilization. If this is the case, then discrimination may depend on specific interactions between the substrate and its surroundings in the transition state structure. We are currently exploring this alternative to the "induced fit" mechanism.

References

Amadei, A, Abm Linssen, Hjc Berendsen. 1993. "Essential Dynamics of Proteins." *Proteins: Structure Function And Genetics*, 17(4): 412-425.

Arcangeli, C, AR Bizzarri, S Cannistraro. 2001. "Molecular Dynamics Simulation And Essential Dynamics Study Of Mutated Plastocyanin: Structural, Dynamical And Functional Effects Of A Disulfide Bridge Insertion At The Protein Surface." *Biophysical Chemistry*, 92 (3):183-199.

Rittenhouse, RC, WK Apostoluk, JH Miller, and TP Straatsma. 2003. "Characterization of the Active Site of DNA Polymerase β by Molecular Dynamics and Quantum Chemical Calculation." *Proteins: Structure, Function, and Genetics*, 53(3):667-682.

Sawaya, MR, R Prasad, SH Wilson. 1997, "Crystal Structures of Human DNA Polymerase Beta Complexed with Gapped and Nicked DNA: Evidence for an Induced Fit Mechanism." *Biochemistry,* 36(37): 11205-11215.

Showalter AK, Tsai MD. 2002. A reexamination of the nucleotide incorporation fidelity of DNA polymerases. *Biochemistry* 41: 10571-10576.

Stella L, Di Iorio EE, Nicotra M, Ricci G. 1999. Molecular dynamics simulations of human glutathione transferase P1-1: Conformational fluctuations of the apo-structure. *Proteins: Struct Funct Genet* 37: 10-19.

Yang LJ, Beard WA, Wilson SH, Broyde S, Schlick T. 2002a. Polymerase beta simulations suggest that Arg258 rotation is a slow step rather than large subdomain motions per se. *J Mol Biol* 317: 651-671.

Yang LJ, Beard W, Wilson S, Roux B, Broyde S, Schlick T. 2002b. Local deformations revealed by dynamics simulations of DNA polymerase Beta with DNA mismatches at the primer terminus. *J Mol Biol* 321: 459-478.

Effects of 8-oxoguanine on DNA Bending (J.H. Miller, T.P. Straatsma)

MD simulations were carried out on the fully solvated and cation-neutralized DNA oligonucleotide GGGAA***C***AACTAG:CTAGTT***G***TT. C in its native form and with guanine in the central G_{19}:C_6 base pair replaced by 8-oxoguanine (8oG). The direction and magnitude of global bending were assessed by a technique used previously to analyze simulations of DNA containing a thymine dimer. The presence of 8oG did not greatly affect the magnitude of DNA bending; however, bending directions that compress the major groove and expand the minor groove were significantly more probable when 8oG replaced G_{19}. Crystal structures of glycosylases bound to damaged-DNA substrates consistently show bending that expands the minor groove. We conclude that changes in bending dynamics that facilitate this expansion are part of the mechanism by which 8oG is recognized by the base excision repair pathway.

Replacing G19 in GGGAACAACTAG:CTAGTTGTT. C by 8oG does not induce any major local perturbations in this oligonucleotide. No tendency for opening of the 8oG19:C6 base pair was detected in MD simulations of 2 ns duration. The most pronounced local changes induced by the presence of 8oG were an attraction for counter ions and a small kink in the helical axis between base pairs 8oG19:C6 and T18:A7. The damaged base had a large effect on the dynamics of global bending. Conformations with the oligomer bent in the direction of the minor groove were rarely seen when 8oG was present even though that was the preferred direction of bending for the native sequence.

The local and global effects of 8oG observed in our simulations are probably related; however, the mechanism of this connection is uncertain. Attraction of counter ions to the damage site did not preferentially neutralize atoms on the major groove face of the oligomer. Changes in the electrostatic potential of the damaged base will also affect stacking interactions, which may influence bending dynamics. Further investigation of this mechanism is needed.

Since glycosylases recognize the 8oG bind in the minor groove, an induced preference for bending into the major groove could be a factor in damage recognition, as has been proposed for recognition of thymine dimers by T4 endonuclease V. This concept is being investigated further by potential-of-mean-force calculations designed to mimic the "pinch-push-pull" mechanism of base flipping.

Structure and Energetics of Clustered Damaged Sites in Duplex DNA (J.H. Miller, T.P. Straatsma)

Introduction

Absorption of ionizing radiation produces DNA damage by direct interaction with charged particles and by a nonhomogeneous distribution of reactive oxygen species. Both of these mechanisms have the potential to produce multiple lesions in DNA on a nanometer scale (Ward 1988). The increasing frequency of multiple damaged sites with increasing linear energy transfer (LET) is widely accepted as a mechanism for the greater biological effectiveness of densely ionizing radiation. The chemical characteristics of "clustered damage" that make it a precursor of biological effectiveness are largely

unknown. As a step toward filling this knowledge gap, several recent experimental studies have investigated the effects of simple, but chemically well characterizied, multiple damaged sites on the DNA structure (Lin and de los Santos 2001), the activity of repair enzymes (Lomax et al. 2004), and mutagenisis (Pearson et al. 2004). Generally, these studies show that pairs of lesions on opposite strands induce larger biological effects than isolated single lesions if they are separated by no more than about 5 base pairs. With the same objective, we have undertaken a theoretical study of the effects of small clusters of base and backbone damage on the dynamics and stability of DNA. In this paper, we report results for single lesions that will be used for comparison with ongoing calculations of multiple damaged sites.

Methods

Starting structures for geometry optimization of DNA trimers by minimization of the energy were obtained by selecting a 3 base pair fragment within a larger oligonucleotide with a standard B-DNA conformation or a conformation determined by a prior classical MD simulation (see Figure C.4). Na+ counter ions were added to neutralize excess negative charge. Lesion-free fragments were modified to obtain damaged fragments with an oxidized purine or an AP site in the central base pair of the trimer. The NWChem computational chemistry package (Harrison et al. 2000) was used to carry out geometry optimization using the density-functional level of quantum electronic-structure theory with the B3LYP functional and 6-31G** basis set. During the optimization, the nucleic acid bases in the top and bottom base pairs of the trimer remained fixed in space to reduce unphysical effects related to the finite size of system.

NWChem was also used to carry out classical MD simulations based on the parm99 version (Cheatham et al. 1999) of the AMBER force field (Cornell et al. 1995). Partial atomic charges for non-standard DNA residues were obtained by fitting the electrostatic potential energy of quantum electronic-structure calculations using the Hartree-Fock level of theory and a 6-31G* basis set. Starting conformations were obtained by modifying the canonical B-DNA structure of the duplex DNA oligonucleotide 3'-TCGCGTTGCGCT-5' to introduce strand breaks between T6 and T7. This sequence was chosen to allow comparison of our results with calculations by Yamaguchi et al. (2002) for the case of a simple strand break obtained by hydrolyzing the P-O bond. Negative charge on the DNA oligonucleotide was neutralized by the addition of Na+ counter ions, and hydration was modeled by placing DNA and counter ions in a box of water molecules.

Results

Figure C.4 shows optimum structures of DNA trimers with the base lesions 8oG and 8-oxoadenine (8oA), and abasic sites in the sequences AGT:AapT, AGC:GapT, and AapC:GTT. The effect of damage on duplex DNA stability was calculated by $\Delta E = \Delta E_{int} + E_{reorg1} + E_{reorg2}$, where ΔE_{int} is change in the interaction energy of DNA strands with and without the lesion present, and E_{reorg1} and E_{reorg2} are energies associated with reorganization of the damaged and undamaged strands, respectively. For the trimers containing 8oG and 8oA, all three components of ΔE are small, which is consistent with experimental data by Plum et al. (1995) on the thermal stability of DNA containing 8oG.

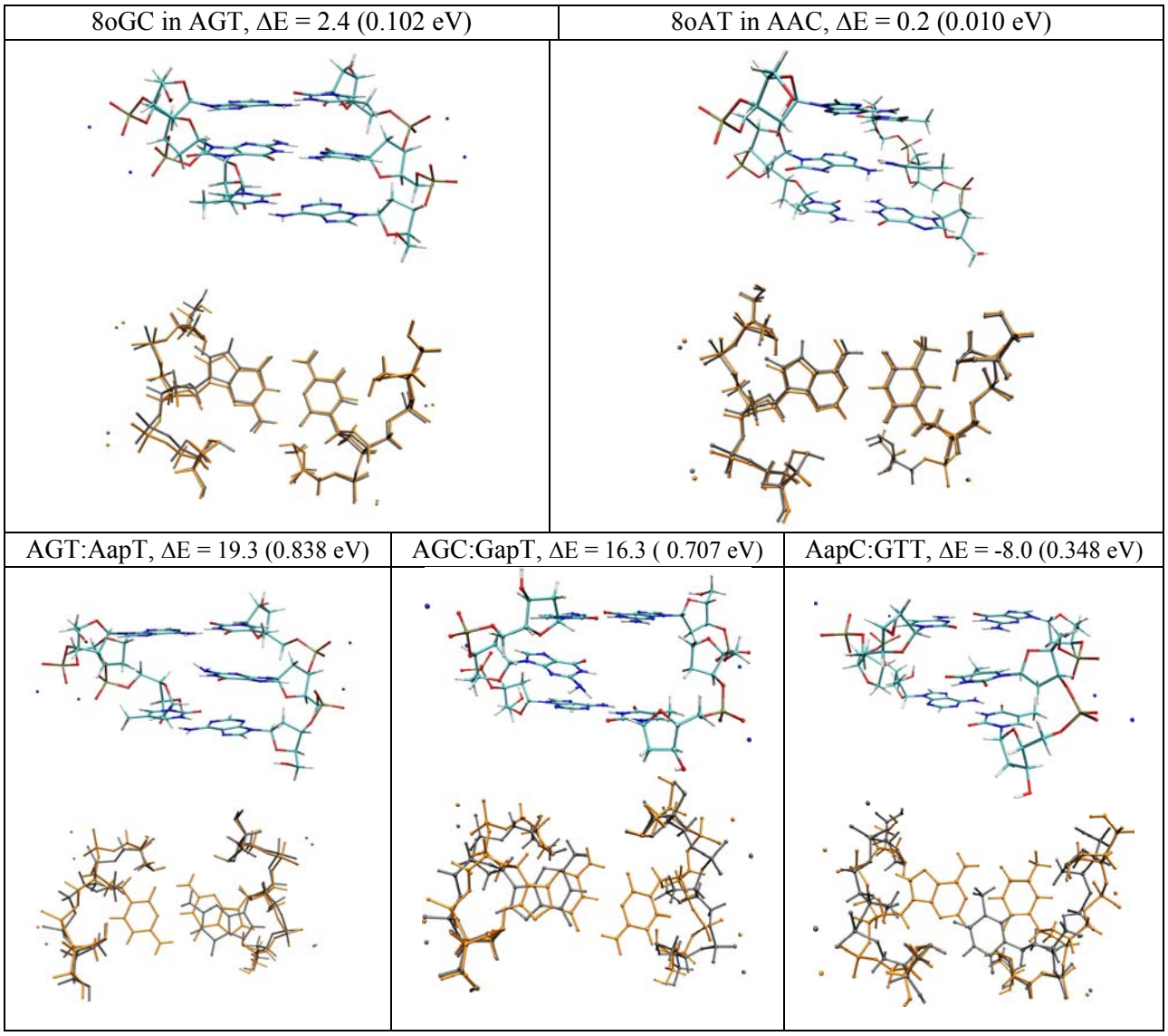


Figure C.4. Conformation and energetics of DNA trimers containing oxidized purine bases and abasic sites. View down the axis of the trimer compares the energy-optimized geometries of damaged and native sequences. The differences in duplex DNA stability (as defined in text) relative to the respective unmodified sequence are in kcal/mol, with equivalent values in eV shown in parentheses.

For DNA trimers containing AP sites, ΔE_{int} is large and positive due to the loss of Watson-Crick hydrogen bonds. In sequences AGT:AapT and AGC:GapT, strand-reorganization energies are small so that ΔE_{int} is the primary factor determining the energy of the abasic sites relative to undamaged DNA. Trimer AapC:GTT has a large negative reorganization energy of the undamaged strand due to a intra-strand hydrogen bond and strong interaction of both phosphate groups with same counter ion. The net effect is that an AP site is more stable than the native sequence by 8.0 kcal/mol (0.348 eV). The energy-optimized conformations in Figure C.4 show that the more stable AP sites exhibit greater conformation change relative to that of their native sequences.

The structures shown in Figure C.5 show the conformation of the 12mer 3'-TCGCGTTGCGCT-5' near 3 types of single-strand scissions after 2-ns classical MD simulations. A simple strand break generated by hydrolyzing the P-O bond between T_6 and T_7 has little effect on DNA conformation, as can be seen in Figure C.5a. Strand breaks induced by ionizing radiation are most often accompanied by loss of the base and sugar residues at the site of strand scission. Figure C.5b shows this combination of damage with complete loss of the sugar moiety, while in Figure C.5c, a 3'phosphoglycolate remains. These are the most common types of end groups for radiation-induced single-strand breaks (Henner et al. 1982).

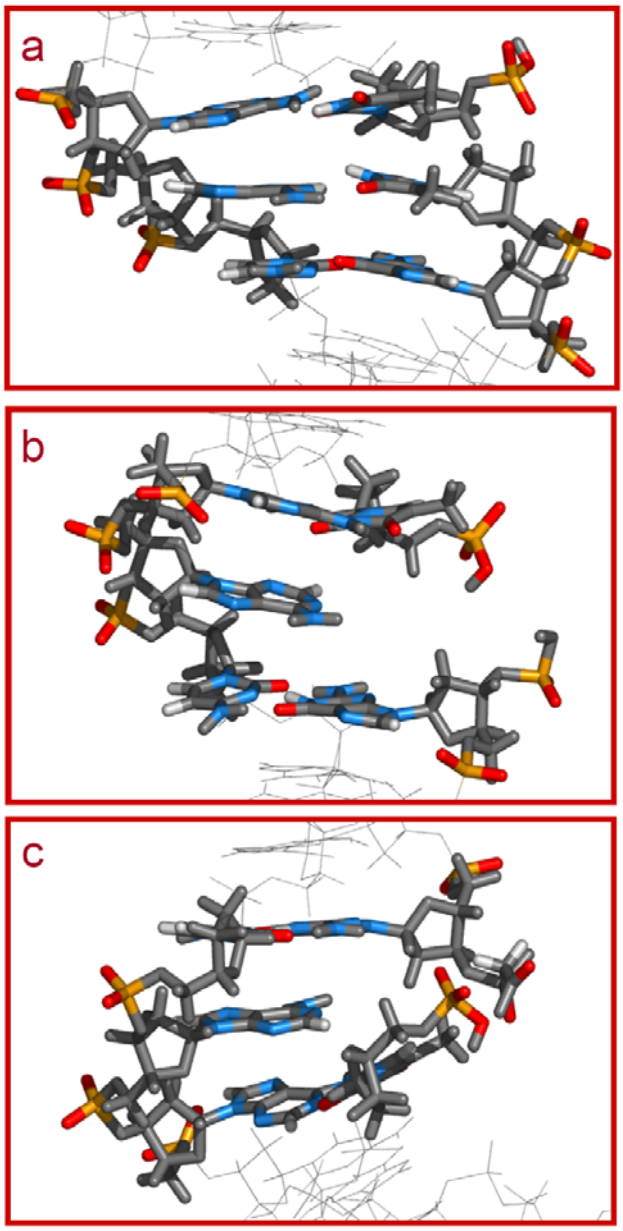


Figure C.5. The effect of different types of strand breaks at T_6T_7 on the conformation of duplex DNA oligonucleotide 3'-TCGCGTTGCGCT-5' during a 2 ns MD simulation. (a) hydrolysis of P-O bond only, (b) missing T_6 base and sugar, (c) missing T_6 base and partially degraded sugar.

For the distance between hydrogen donor and acceptor atoms, we concluded that the Watson-Crick hydrogen bonds of the base pairs on both sides of the strand break remained intact throughout the MD simulation. The main effect of strand cleavage was to allow greater bending flexibility in the helical axis at the site of the strand break. This flexibility enables the base pairs on either side of the break to partially fill the gap when a base is missing, as is illustrated in Figure C.5c for the strand break with a 3'phosphoglycolate end group. A similar distortion of the helical axis frequently occurred in our simulation of a single-strand break with base and sugar completely removed.

References

Cheatham, TE, P Cieplak, and PA Kollman. 1999. "A modified version of the Cornell et al. force field with improved sugar pucker phases and helical repeat." *J. Biomol. Struct. Dyn.* 16(4):845-862.

Cornell, WD, P Cieplak, CI Bayly, IR Gould, KM Merz, DM Ferguson, DC Spellmeyer, T Fox, JW Caldwell, and PA Kollman. 1995. "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules." *J. Am. Chem. Soc*. 117(19):5179-5197.

Harrison, RJ, JA Nichols, TP Straatsma, M Dupuis, EJ Bylaska, GI Fann, T L Windus, E Apra, J Anchell, D Bernholdt, P Borowski, T Clark, D Clerc, H Dachsel, B de Jong, M Deegan, K Dyall, D Elwood, H Fruchtl, E Glendenning, M Gutowski, A Hess, J Jaffe, B Johnson, J Ju, R Kendall, R Kobayashi, R Kutteh, Z Lin, R Littlefield, X Long, B Meng, J Nieplocha, S Niu, M Rosing, G Sandrone, M Stave, H Taylor, G Thomas, J van Lenthe, K Wolinski, A Wong, and Z Zhang. 2000. *NWChem, a computational chemistry package for parallel computers, Version 4.0*. Pacific Northwest National Laboratory, Richland, WA.

Henner, WD, SM Grumberg, and WA Haseltine. 1982. "Sites and structure of γ radiation-induced DNA strand breaks." *J. Biol. Chem*. 257(19), 11750-11754.

Lin, Z and C de los Santos. 2001. "NMR characterization of clustered bistrand abasic site lesions: Effect of orientation on their solution structure." *J. Mol. Biol.* 308(2):341-352.

Lomax ME, S Cunniffe, and P O'Neill. 2004. "8-OxoG retards the activity of the ligase III/XR. 1 complex during the repair of a single-strand break, when present within a clustered DNA damage site." *DNA Repair* 3(3):289-299.

Pearson, CG, N Shikazono, J Thacker, and P O'Neill. 2004. "Enhanced mutagenic potential of 8-oxo-7,8-dihydroguanine when present within a clustered DNA damage site." *Nucleic Acids Res*. 32(1):263-270.

Plum, GE, AP Grollman, F Johnson, and KJ Breslauer. 1995. "Influence of the oxidative damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex." *Biochemistry* 34(49):16148-16160.

Yamaguchi, H, JG Siebers, A Furukawa, N Otagiri, and R Osman. 2002. "Molecular dynamics simulation of a DNA containing a single strand break." *Radiat. Prot. Dosim.* 99(1-4):103-108.

Ward, JF. 1988. "DNA damage produced by ionizing radiation in mammalian cell: Identities, mechanisms of formation, and reparability." *Progress Nucleic Acid Res.* 35, 95-125.

Stability of DNA Containing 8-Oxoguanine:Cytosine Base Pairs (J.H. Miller, T.P. Straatsma)

Introduction

A variety of environmental and endogenous agents, including ionizing radiation and aerobic metabolism, produce oxidative DNA damage. One of the most abundant and extensively studied forms of oxidative DNA damage is 8-oxo-7,8-dihydroguanine (8oG). Its tendency to mispair with adenine during replication makes 8oG a direct source of GC to TA transversion mutations; consequently, elaborate defense systems to protect against the genotoxic effects of 8oG are found in most organisms from bacteria to humans.

The mutagenic potential of 8oG lesions is believed to result from steric repulsion between O8 and deoxyribose that causes 8oG to prefer a syn conformation of the glycosidic torsion angle and forms a Hoogsteen base pair with adenine. Express of this preference occurs with DNA polymerases when 8oG is in the template strand. The anti conformation of the glycosidic torsion angle is maintained in 8oG:C base pairs, and the presence of a 8oG:C base pair destabilizes duplex DNA by at most a few kcal/mole (Plum et al. 1995).

A part of the stability of DNA containing 8oG:C base pairs appears to be related to a change in the bending dynamics that favors bending into the major groove. Miller et al. (2003) demonstrated this effect in MD simulations of the DNA oligonucleotide GGGAACAACTAG:CTAGTTGTT. C with guanine in the central G19:C6 base pair replaced by 8oG. Analysis of the dynamics of the native sequence showed a correlation between bending into the minor groove and shorter distances between the H8 and backbone oxygen atoms of G19, suggesting that bending in this direction will be energetically unfavorable when H8 is replaced by O8.

In this paper, we report on our investigation of stabilization of DNA containing 8oG:C by changes in the hydration of the oxidized base. Ishida (2002) observed water molecules bridging between the atoms O8 and O5' in the 8oG nucleotide during their MD simulation of the 12mer d(CGCGAATTCGCG)2 using the AMBER 95 force field (Cornell et al. 1995). In addition to different sequence and force field, Ishida's simulation was based on somewhat different partial atomic charges in the 8oG nucleotide. Even though the residual charge on O8 was essentially the same as in our model, the charge on the O4' atom was 27% more negative than the standard AMBER charges for deoxyribose, which are what we used. It is of interest to know if the occupation of the bridging site between O8 and O5' is sensitive to these differences in base sequence and simulation methods.

Methods

Results from MD simulations of GGGAACAACTAG:CTAGTTGTT. C were analyzed for the patterns of hydration near the G19:C6 base pair in the presence and absence of 8oG. In the set of simulations (native and modified) for our earlier work (Henner et al. 1982), the oligonucleotide was solvated by 22 Na+ counter ions and about 7000 water molecules. In a repeat of these simulations for this study, we used a larger box of water that contained about 1500 water molecules. In both cases, coordinates of all atoms in the simulation (solute, solvent, and counter ions) were saved every 2 pc, resulting in 1000 frames in each trajectory file that could be analyzed for hydration patterns. Software adapted from code originally written by R.C. Rittenhouse (Rittenhouse et al. 2003) was used to systematically search these frames for sustained contacts between DNA atoms and water molecules. Atoms separated by a distance less than or equal to the sum of their van der Waals radii plus a user-defined distance set to 0.0075 nm

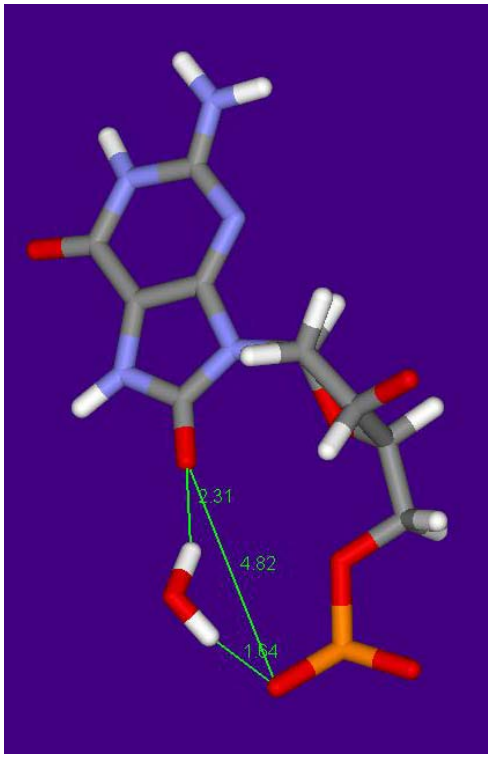
were considered to be in contact. Sustained contacts were defined as those surviving for at least 50 consecutive frames of a trajectory.

For each of the 1000 frames containing coordinates for DNA and water, distances were calculated between all pairings of water oxygen atoms and DNA atoms. If two atoms made contact, the solvent atom was added to a list belonging to the solute atom it contacted. After this first contact, a counter was incremented for that solvent atom each time it again contacted the solute atom. Solvent atoms that had 50 or more contacts with a particular solute atom were listed along with their contact counts and corresponding solute atoms after all frames were processed. Each solute atom that experienced high contact times with one or more solvent atoms had a profile written that showed the distances between itself and those solvent atoms over the course of the 2 ns of simulation. Solvent molecules with long contacts to DNA atoms of special interest were profiled for contacts with other solvent molecules to discover networks of water molecules.

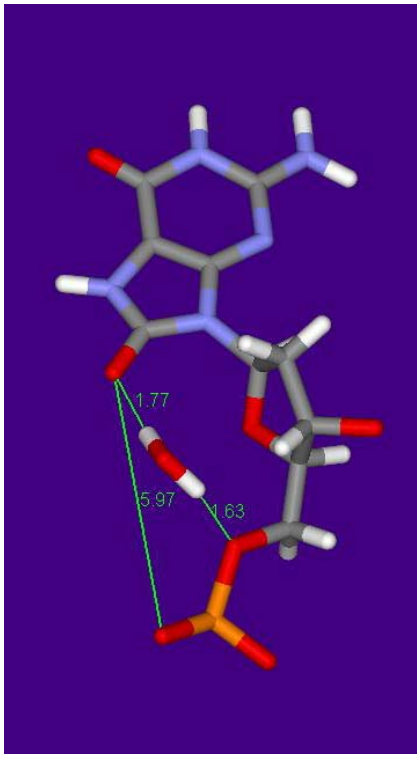
Results

Typical structures within the hydration layer that probably contributed to the stabilization of DNA containing 8oG are shown in Figure C.6. Patterns I and II are bridging configurations of single water molecules connected at O8 to O2P and O5', respectively. In our earlier work (Henner et al. 1982) we noted that the presence of 8oG increased the probability of finding a counter ion near the damaged base pair. Analysis of the effects of 8oG on DNA hydration revealed that these counter ions assist the formations of water networks of the type illustrated by Patterns III and IV in Figure C.6, where the light-blue sphere shows the location of a counter ion.

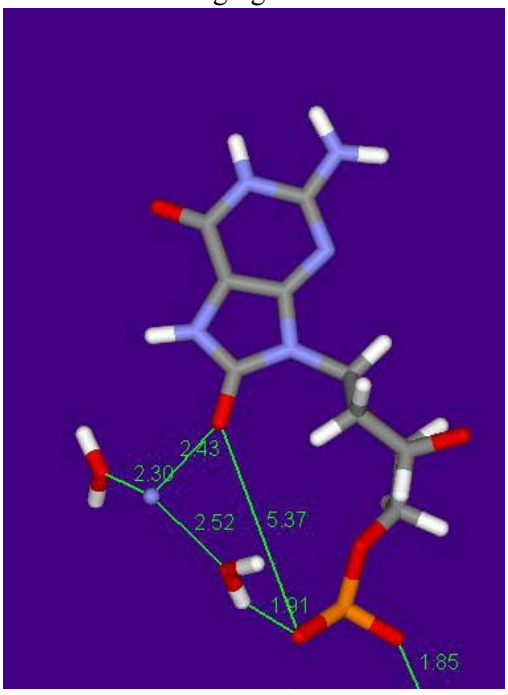
The 1000 frames where water coordinates were saved in the trajectory files of the 2 MD simulations of GGGAA***C***AACTAG:CTAGTT*8o****G***TT. C were exhaustively searched for bridging water molecules similar to Pattern II of Figure C.6. In the simulation with the smaller number of water molecules, 348 instances of Pattern II were found with the combined distance atoms O5* and O8 in $8oG_{19}$ less than 0.6 nm from the oxygen atom of the water molecule. In an exhaustive search of the corresponding simulation of the native sequence with a similar number of water molecules, 158 instances were found of water molecules in a similar orientation as Pattern II with their oxygen atoms less than 0.6 nm (combined distance) from O5* and H8 in G19 native. The same analysis applied to the more recent simulations: with larger numbers of water molecules solvating, the oligomer found 531 examples of Pattern II, with water oxygen atoms less than 0.6 nm combined distance from atoms O5*, and O8 of $8oG_{19,}$ compared with 188 cases of Pattern II in the native simulation, with water oxygen atoms less than 0.6 nm combined distance from atoms O5*, and H8 of G_{19}. Hence we conclude that the more electronegative O8 atom in 8oG makes the occurrence of hydration Pattern II about 2.5 times more probable than a similar hydration of guanine. The effect on DNA hydration probably increases the stability of DNA containing 8oG by reducing the effects of electrostatic repulsion between O8 and backbone oxygen atoms. Future studies will attempt to quantify the magnitude of this stabilization effect.



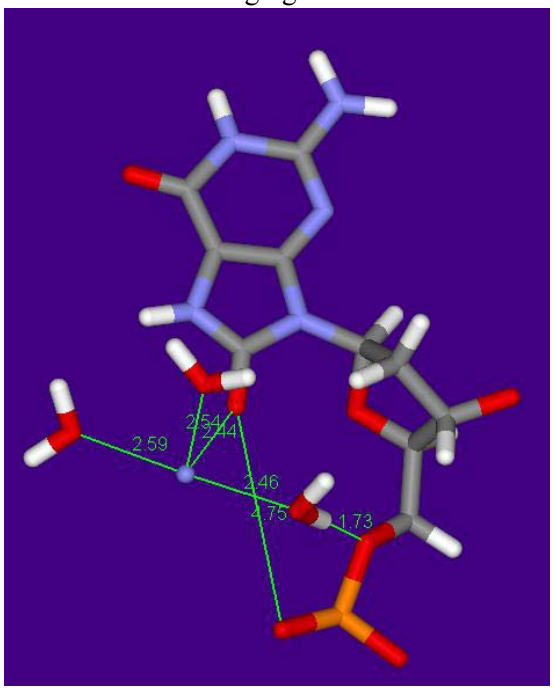
Pattern I: Bridging from O8 to O2P



Pattern II: Bridging from O8 to O5'



Pattern III: O8 – Na^+ - H2O – O2P



Pattern IV: O8 – Na^+ - H2O – O5'

Figure C.6. Typical patterns of hydration in residue 8oG19 of the oligonucleotide GGGAACAACTAG:CTAGTT8oGTT. C that link atom O8 to backbone oxygen atoms.

References

Cornell, WD, P Cieplak, CI Bayly, IR Gould, KM Merz, DM Ferguson, DC Spellmeyer, T Fox, JW Caldwell, and PA Kollman. 1995. "A second generation force field for the simulation of proteins, nucleic acids, and organic molecules." *J. Am. Chem. Soc*. 117(19):5179-5197.

Henner, WD, SM Grumberg, and WA Haseltine. 1982. "Sites and structure of γ radiation-induced DNA strand breaks." *J. Biol. Chem*. 257(19):11750-11754.

Ishida, H. 2002. "Molecular dynamics simulation of 7/8-dihydro-8-oxoguanine DNA." *J. Biomol. Struct. Dyn*. 19(4):839-851.

Miller, JH, CP Fan-Chiang, TP Straatsma, and MA Kennedy. 2003. "8-Oxoguanine Enhances Bending of DNA that Favors Binding to Glycosylases." *J. Am Chem. Soc.* 125:6331-6336.

Plum, GE, AP Grollman, F Johnson, and KJ Breslauer. 1995. "Influence of the oxidative damaged adduct 8-oxodeoxyguanosine on the conformation, energetics, and thermodynamic stability of a DNA duplex." *Biochemistry* 34(49):16148-16160.

Rittenhouse, RC, WK Apostoluk, JH Miller, and TP Straatsma. 2003. "Characterization of the Active Site of DNA Polymerase β by Molecular Dynamics and Quantum Chemical Calculation." *Proteins: Structure, Function, and Genetics*, 53(3):667-682.

Simulation of Bulges in DNA and RNA (M. Zacharias)

During the project duration, we finished an MD project on the effect of nicks, extra bases, and gaps in damaged DNA and a project on the effect of modified bases on the dynamics of RNA. In addition, we initiated large-scale simulation studies on a complete nucleosome structure (147 base pairs of DNA in complex with histone proteins).

Modeling the effect of nicks, extra bases and gaps in damaged DNA using molecular dynamics simulations

Damage of DNA by radiation and chemical reagents is a major cause of cancer and cell aging. It can lead to changes in the DNA structure and dynamics which in turn influence recognition and repair of damaged DNA. Common forms of DNA damage include the loss of a base (creating an abasic site), a bond break (nick) in the DNA backbone, or the loss of one or several nucleotides. Surprisingly, the X-ray structure of a double-stranded nicked DNA shows only modest variation when compared with regular DNA (*e.g.*, narrower minor groove; Aymani et al.1990). However, X-ray crystallography provides only a static picture of nucleic acids. It is expected that a nick in DNA or the loss of one or several nucleotides due to radiation or chemical damage (gap) will lead to increased local flexibility or a change in average solution conformation.

During the project duration, multi-nanosecond simulations on an intact 12 base pair DNA, the same sequence with a central nick, and a gap of one nucleotide at the center were performed. For all simulations, NWChem, the massively parallel software for computational chemistry developed at Environmental Molecular Sciences Laboratory at the Pacific Northwest National Laboratory (PNNL), in combination with the AMBER (parm99) force field, has been used. The X-ray structure was used as

starting structures for the simulations of the nicked DNA (Aymani et al. 1990). Since no experimental structure with exactly the same sequence as the nicked DNA was available, a standard B-DNA structure was used as a reference. The starting structure for the simulations with a single nucleotide gap was obtained by removing a (T) nucleotide at the center of the nicked, double-stranded DNA structure. Simulations were performed on the HP/Linux 1960 Itanium-2 processor cluster (MPP2) at the PNNL Molecular Science Computer Facility for approximately 15 ns. Each simulation included approximately 5000 water molecules and enough counter ions to neutralize the charge.

Although the intact and the nicked DNA oligonucleotide structures showed considerable conformational fluctuations during the MD simulations, some structural features of the experimental structures were well preserved. A stable narrowing of the DNA double helix at the nick was found, which might be a specific structural signature recognized by repair enzymes. No significant changes in twisting or bending compared with the regular DNA oligonucleotide were observed. The average bend angle of the DNA during the simulation was approximately 10 degrees. This result is consistent with experiments by Mills et al. (1999), who investigated the global conformational flexibility of nicked and gapped DNA molecules using electric birefringence experiments. In these studies, it was found that a single nick in DNA did not significantly change the global properties of DNA.

In the case of the simulations of gapped DNA, larger conformational changes in the DNA were observed.. In the initial DNA structure, the bases that flanked the gap were in an "unstacked" state (Figure C.7) during the MD simulation; no further unstacking or diffusion of water molecules into the space between the flanking nucleotides was observed. Instead, during the simulation, the gap between the two segments of the second strand was closed, and the two nucleotides adopted a stacked conformation similar to a dinucleotide step conformation in regular DNA (Figure C.8). Overall, the DNA adopted a bend conformation with a more accessible major groove compared to the intact or nicked DNA duplex and with a bend angle reaching 30 degrees. This result is, again, consistent with the experimental observation of Mills et al. (1999), who found a significant effect of gaps in DNA on the global conformational properties (bending) of the duplex. The change in major groove size and bending angle may serve as recognition elements for the repair of such DNA damage.

Influence of a Fluorobenzene Nucleobase Analogue on the Conformational Flexibility of RNA Studied by MD Simulations

Chemically modified bases are frequently used to stabilize nucleic acids, to study the driving forces for nucleic acid structure formation, and to tune DNA and RNA hybridization conditions. In particular, fluorobenzene and fluorobenzimidazole base analogues can act as universal bases able to pair with any natural base and to stabilize RNA duplex formation (Parsch and Engels 2002; Figure C.9). These modified bases can act as model systems to investigate the chemical base modification on the structure and dynamics of nucleic acids. Although these base analogues are compatible with an A-form RNA geometry (Parsch and Engels 2002), little is known about the influence on the fine structure and conformational dynamics of RNA. Nanosecond MD simulations have been performed to characterize the dynamics of RNA duplexes containing a central 1'-deoxy-1'-(2,4-difluorophenyl)-ß-D-ribofuranose base pair or opposite to an adenine base. For comparison, RNA with a central uridine:adenine pair and a 1'-deoxy-1'-(phenyl)-ß-D-ribofuranose opposite to an adenine was also investigated. The MD simulations indicate a stable overall A-form geometry for the RNAs with base analogues. However, the presence of the base analogues caused a locally enhanced mobility of the central bases inducing mainly base pair shear and opening motions. No stable "base paired" geometry was found for the base analogue pair or the

base analogue:adenine pairs, which explains in part the universal base character of these analogues. Instead, the conformational fluctuations of the base analogues led to an enhanced accessibility of the bases in the major and minor grooves of the helix compared to a regular base pair. This simulation study suggests a general mechanism for the enhanced local mobility of modified bases in double-stranded nucleic acids (Zacharias and Engels 2004).

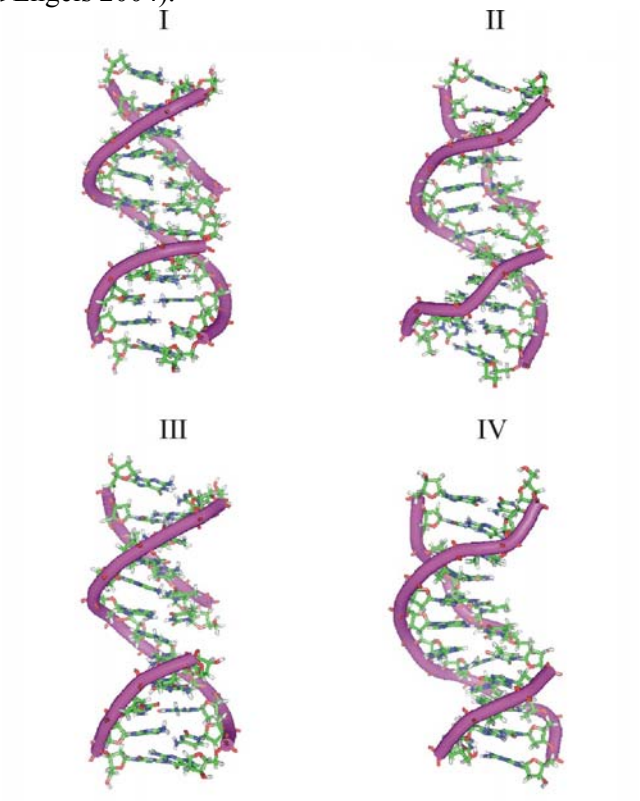


Figure C.7. Comparison of the DNA start structure with a central strand nick (I) or gap (III) and snapshots from the MD simulation after ~10 ns simulation time (II: nick, IV: DNA with gap). The DNA backbone trace is indicated as a cartoon.

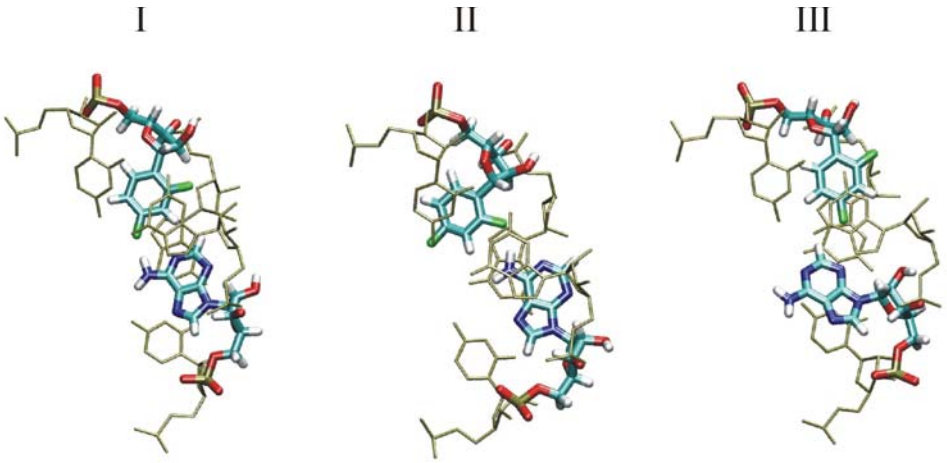


Figure C.8. Conformational snapshots observed during the MD simulation of RNA with a central difluorophenyl base opposite to an adenine. The view is along the helical axis of the RNA. The central base pair (thick lines) and the neighboring base pairs (thin lines) are shown.

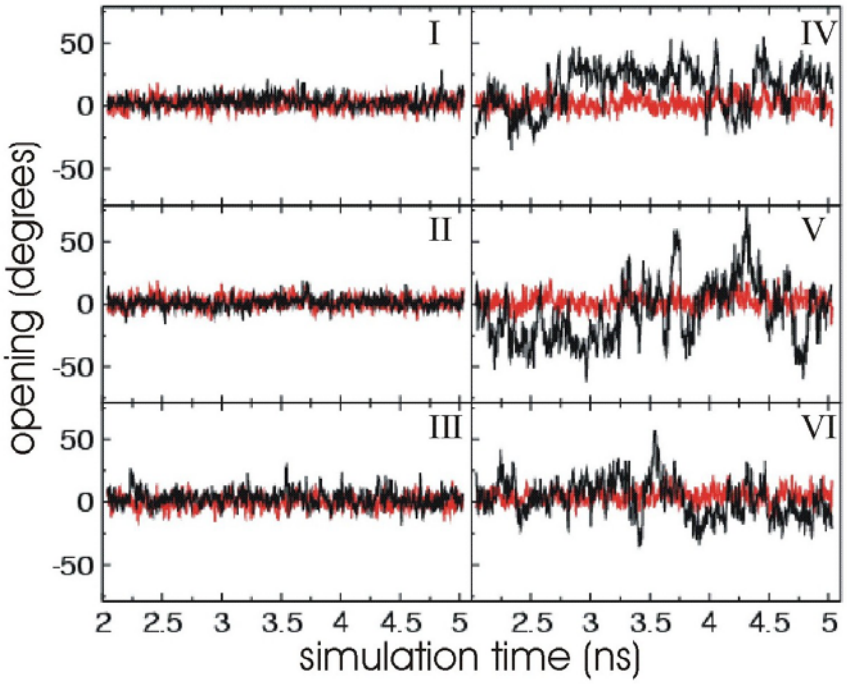


Figure C.9. Base pair opening motions of the central base pairs (black lines in right panels, IV-VI) and the base pair before the central base pair (black lines in left panels, I-III) observed during the MD simulations of a difluorophenyl:adenine containing RNA (I,II and IV, V) and an RNA with a difluorophenyl:difluorophenyl pair (III,VI). The red line corresponds to the reference duplex with only natural base pairs.

MD Simulation of the Complete Nucleosome

During this computational grand challenge project, we initiated an MD simulation of a complete. The simulation system consists of 147 base pairs of DNA, several histone proteins, ions and several thousand water molecules and is based on the high-resolution structure by Richmond and Davey (2003) (Figure C.10). A simulation time of 15 ns could already be achieved and allowed the characterization of the local and global mobility of the nucleosome particle on the nanosecond time scale at atomic resolution. The study of the dynamics of packed DNA is also of fundamental importance to understand the effect of DNA damage since DNA in the cell is not free but, to a large degree, associated with proteins to form a compact, only partially accessible structure. Simulation studies on such a large structure are only possible by using supercomputer resources as available at the PNNL supercomputer center. The simulation will be extended in the current grand challenge phase and could also form the basis to study DNA damage in the context of a packed DNA structure.

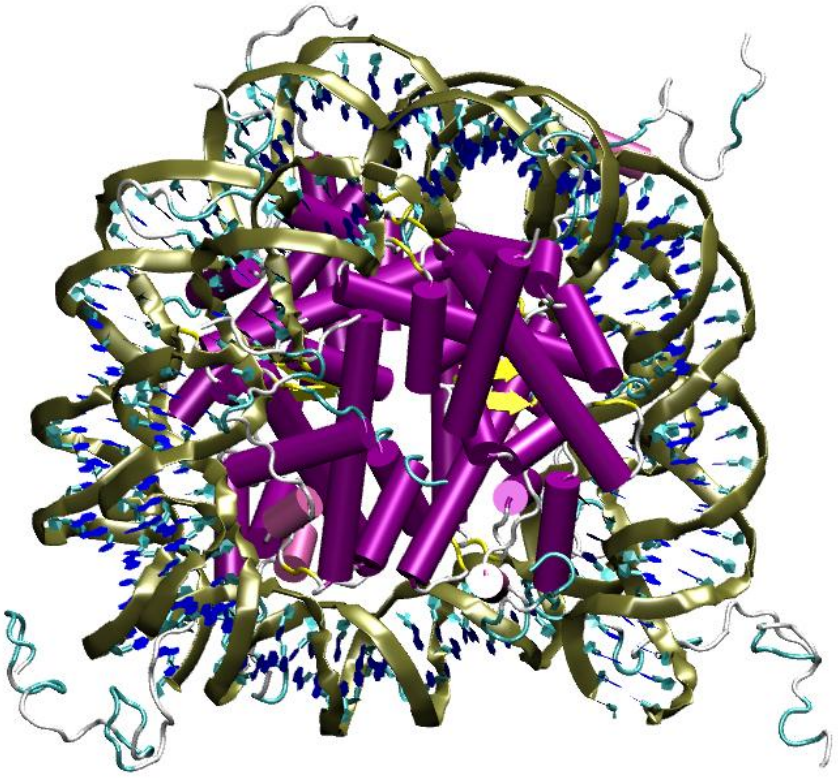


Figure C.10. Simulation start structure of the nucleosome core particle (Richmond and Davey 2003) in cartoon representation.

References

Aymani, J, M Coll, GA van der Marel, JH van Boom, A H-J Wang , and A Rich. 1990. "Molecular structure of nicked DNA: A substrate for DNA repair enzymes." *Proc. Natl. Acad. Sci.* 87(7): 2526-30.

Mills, JB, JP Cooper, and PJ Hagerman. 1994. "Electrophoretic evidence that single-stranded regions of one or more nucleotides dramatically increase the flexibility of DNA." *Biochemistry,* 33(7):1797-1803.

Mills, JB, E Vacano, and PJ Hagerman. 1999. "Flexibility of single-stranded DNA: use of gapped duplex helices to determine the persistence lengths of poly(dT) and poly(dA)." *J. Mol. Biol*., 285(1): 245-257.

Parsch, J and JW Engels. 2002. "C-F···H-C Hydrogen Bonds in Ribonucleic Acids." *J. Am. Chem. Soc*, 124(20):5664-5672.

Richmond, TJ and CA Davey. 2003. "The structure of DNA in the nucleosome core." *Nature*, 423(6936):145-150.

Zacharias, M and JW Engel. 2004. "Influence of a fluorobenzene nucleobase analogue on the conformational flexibility of RNA studied by molecular dynamics simulations." *Nucleic Acids Research*, 32:6304-6311.

Protein Kinase A (J.A. McCammon, C.F. Wong)

Molecular Docking of Balanol to Dynamics Snapshots of Protein Kinase A

Even if the structure of a receptor has been determined experimentally, it may not be a conformation to which a ligand would bind when induced fit effects are significant. In this work, we evaluated the use of an ensemble of receptor conformations generated from an MD simulation for molecular docking. Two MD simulations were carried out to generate snapshots for protein kinase A: one with the ligand bound, the other without. The ligand, balanol, was then docked to conformations of the receptors presented by these trajectories. The Larmarkian genetic algorithm in AutoDock[1,2] was used in the docking. Three ligand models were used: rigid, flexible, and flexible with torsional potentials. When the snapshots were taken from the MD simulation of the protein-ligand complex, the correct docking structure was found in all cases. However, when the snapshots were taken from the simulation of the protein alone, several clusters of structures were found (Figure C.11). Out of the ten docking runs for each snapshot, at least one structure was close to the experimental complex structure when the flexible ligand models were used. However, the lowest energy structures, according to AutoDock[1,2], did not always correspond to the correctly docked structure. Rescoring using a more sophisticated Generalized Born electrostatics model did not improve the identification of the correctly docked structure. In fact, we found that the correctly docked structure appeared more frequently as the lowest energy structures with the Autodock[1,2] scoring function (Figure C.12). This can provide a useful criterion for selecting the correctly docked structure from clusters of structures obtained from molecular docking experiments.

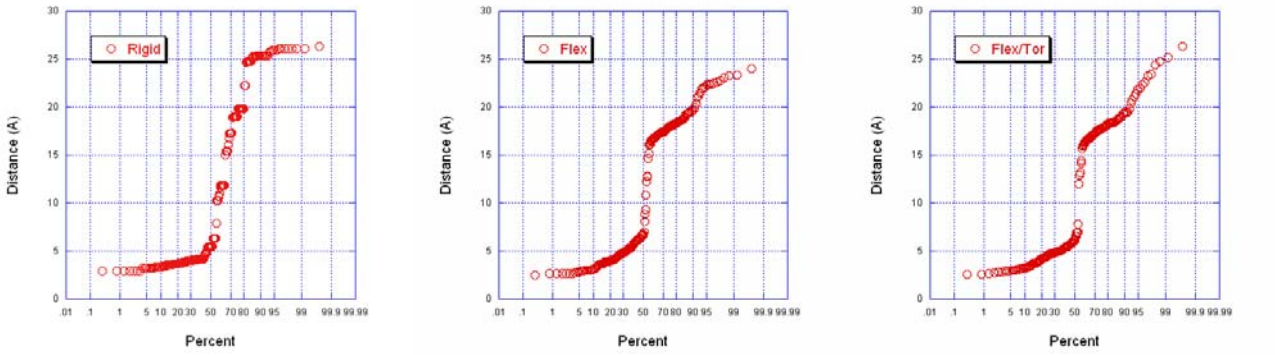


Figure C.11. The use of the distance between a potential hydrogen bond donor on balanol and a potential hydrogen acceptor on the backbone carbonyl oxygen of Glu 121 to classify the docking structures has revealed a few clusters. The clusters with a distance less than 5 Angstrom are those closest to the observed structure of the complex.

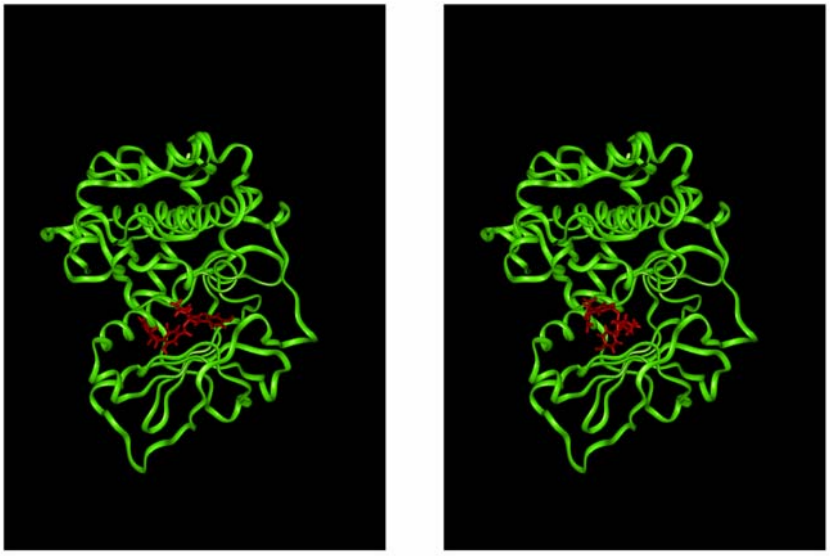


Figure C.12. Two structures of balanol docked to the 270-ps snapshots of protein kinase A. The rank 1 structure, having the lowest docking energy, is shown on the left and is closer to the observed structure than the rank 8 structure, which has a higher docking energy.

References

Goodsell, DS, GM Morris, and AJ Olson. 1996. "Automated docking of flexible ligands: applications of AutoDock." *J Mol Recognit.*, 9(1):1-5.

Morris, GM, DS Goodsell, RS Halliday, R Huey, WE Hart, RK Belew, and AJ Olson. 1998. "Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function." *J Comput Chem.,* 19(14):1639-1662.