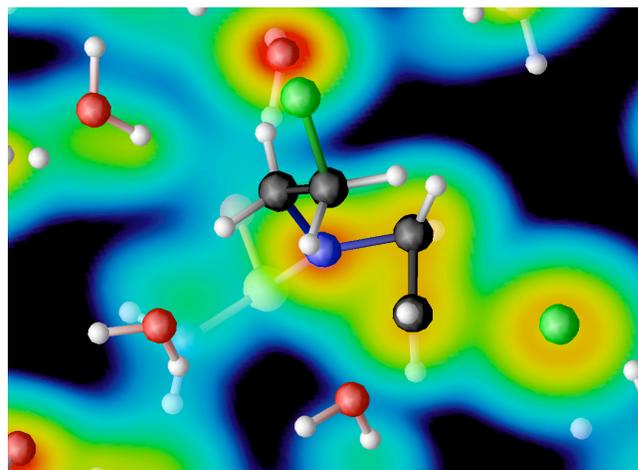
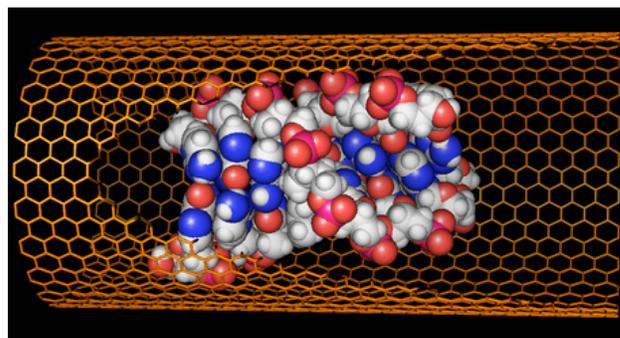
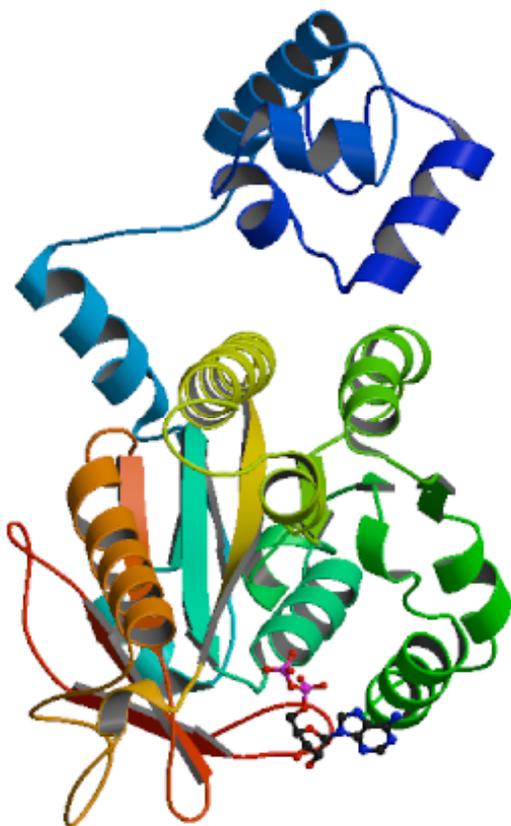


UCRL-TR-202418

A Strategic Initiative in Applied Biological Simulations



01-SI-012

Final Report for FY01 - FY03

February 17, 2004

Michael E. Colvin

Biology and Biotechnology Research Program

Co-Investigators: Giulia Galli (Physics), Francois Gygi (Computations), Felice Lightstone (BBRP), Eric Schwegler (Physics), Daniel Barsky (BBRP), Ceslovas Venclovas (BBRP), Chris Mundy (BBRP), Edmond Lau (BBRP), Brian Bennion (BBRP), Dorota Sawicka (BBRP)

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This work was performed under the auspices of the U.S. Department of Energy by University of California, Lawrence Livermore National Laboratory under Contract W-7405-Eng-48.

Table of Contents

1 Introduction	3
1.1 Relevance to the laboratory	3
1.2 Research Overview	5
2 Accomplishments FY01-FY03	6
2.1 Computational Biochemistry	6
2.1.1 Radioimmunotherapy	6
2.1.1.1 Molecular modeling of HLA-DR10	8
2.1.1.2 Identification of potential binding sites on HLA-DR 10	11
2.1.1.3 Computational Screening of Virtual Library.....	12
2.1.1.4 Molecular modeling of anti-MUC-1 single-chain antibody structures.....	14
2.1.1.5 Sequence analysis of single-chain Fv antibodies raised against DOTA.....	17
2.1.1.4 Simulations on the DOTA-DOTA linker configurations.....	17
2.1.1.4.1 Quantum chemical calculations of DOTA-DOTA linker.	18
2.1.1.4.2 Classical molecular dynamics simulations of ⁹⁰ Y-DOTA-LINKER-DOTA- ⁹⁰ Y.....	19
2.1.1.5 Quantum chemical calculations of ⁹⁰ Y-DOTA and ¹¹¹ In-DOTA.....	21
2.1.1.6 Quantum chemical calculations of Y-DO3AM and In-DO3AM.	22
2.1.2 DNA-Alkylating Anticancer Drugs	23
2.1.2.1 First Principles Molecular Dynamics Simulations on the Activation of Phosphoramidate.....	24
2.1.2.2 Classical Molecular Dynamics Simulations of DNA Crosslinking.....	25
2.1.2.3 Acrolein-Guanine Reactions	27
2.1.3 Abasic DNA Repair	28
2.1.4 Computational development of carboxylesterase inhibitors	31
2.1.5 Estrogen Receptor Modeling, Ligand Docking, and Molecular Dynamics Simulations	33
2.1.6 Dynamics of Protamine based peptides:	36
2.2 Predicted Interactions of Biological Materials with Nanostructures	38
2.3 Advanced Computational Chemical Methods	39
2.3.1 Applications of First Principles Molecular Dynamics	39
2.3.1.1 Liquid water simulations.....	39
2.3.1.2 Ion Solvation	41
2.3.1.3 Reactions of hydroxyl radical with guanine:.....	43
2.3.2.4 Reactions of hydroxyl radical with thymine:	44
2.3.2 Algorithm Development of First Principles Molecular Dynamics	46
2.4 Protein Structure Prediction	47
2.4.1 Development of methods for homology-based protein modeling	47
2.4.2 Application of Molecular Modeling	48
3 Appendix	50
3.1 Publications — In Print, In Press or Submitted	50
3.2 Talks and posters (FY01-FY03):	52

1 Introduction

The goal of this Strategic Initiative in Applied Computational Biology has been to apply LLNL's expertise in computational simulation to forge a new laboratory core competency in biological simulation. By every measure, this SI has been very successful in this goal. Based on a strong publication record and large number of conference presentations and invited talks, we have built a recognized niche for LLNL in the burgeoning field of computational biology. Further, many of the projects that were previously part of this LDRD are now externally funded based on the research results and expertise developed under this SI. We have created successful collaborations with a number of outside research groups including several joint projects with the new UC Davis/LLNL Comprehensive Cancer Center. In addition to these scientific collaborations, the staff developed on this SI is involved in computational biology program development and advisory roles with other DOE laboratories and DOE Headquarters. Moreover, a number of capabilities and expertise created by this SI are finding use in LLNL programmatic applications. Finally, and most importantly, this SI project has brought to LLNL the human talent on who will be the ensuring the further success of computational biology at this laboratory.

1.1 Relevance to the laboratory

Lawrence Livermore Laboratory has a long history of excellence in computational simulation, going back to the earliest days of electronic computers. Notable past achievements at the lab include the computational discovery of two dimensional phase transitions and the development of the first widely used non-linear structural dynamics software. LLNL is presently the world leader in high performance computing and related computational sciences, and is the site for the world's fastest supercomputers. The role of simulation is also a central theme in the LLNL Strategic Plan, which identifies computational simulation as a "critical technology" for 19 of the 22 "major activities" at the lab.

The biological sciences, including molecular biology, genomics, and biotechnology, have also have a growing role at LLNL. Indeed, expanding the laboratory's biology and biotechnology research efforts was one of the primary recommendations of the LLNL *Long Range Strategy Planning Project*. Several LLNL directorates are now involved in biology-related research and the biosciences are expected to play a role in an increasing number of laboratory programs. With the broadening national security missions at the laboratory,

especially since the biological attacks in October, 2001, the areas of biological weapons non-proliferation and defense are also emerging as very significant programs at the lab.

Given the laboratory's expertise in the computational sciences and growing involvement in biology, computational biology combines natural strengths of the laboratory in this rapidly emerging new field of biosciences. No field offers a greater combination of challenge and promise for computer simulation than biology. Biochemical processes involve subtle, low-energy reactions and complex interactions of large macromolecules; yet predictive biochemical simulations could lead to breakthroughs with profound impact on human health, environmental protection, engineered biomaterials, and, of great current relevance, to defense against biological threats.

For example our extensive work on the heterocyclic amine food mutagens is now funded by a large NIH Program Project grant. The computational docking is being applied to the development of biological warfare (BW) agent sensors, and the protein structure prediction methods are being applied to predict the function of newly discovered proteins in BW agents. Five of the post-docs originally supported under this SI have been promoted to staff positions (including three who were hired specifically for this project).

In the past year of this project a major focus has been the application of first principles molecular dynamics (FPMD) running on the ASCI parallel computers to simulate complex aqueous phase reactions. We have completed a first series of FPMD simulations of the activation reaction of DNA-binding anticancer drugs. Additionally, we have applied FPMD to study the complex behavior in water of small metal ions that are critical to the function of many enzymes. This year we also began a collaboration with Professors Gerald and Sally DeNardo at the UC Davis Medical Center to develop a new generation of radioimmunotherapy compounds. This collaboration will involve the entire range of computational biology methods that have been developed under this SI to develop new therapies of immediate clinical relevance. In addition, we continued to make progress on our investigation of how the structural characteristics of damaged DNA correlate with the efficiency of their repair by a DNA repair enzyme and have opened a new dimension in this research project by modeling the effects of single amino acid mutations on these repair enzymes. Finally, we have completed and submitted for publication several collaborative projects, including a computational study of inhibitors of carboxylesterase enzymes and simulations of the reactions of guanine with the hydroxyl radical.

During the final year of this project we will finish the remaining projects, with the specific goal of achieving external funding in each ongoing research area. We will apply first principles molecular dynamics simulations to understand the structural properties of the first solvation shell of the calcium ion (in collaboration with researchers at PNNL). As part of the development of new radioimmunotherapy compounds we will use FPMD to simulate the interactions of two radioisotopes, ^{90}Y and ^{111}In with the chelator compound that is used to bind them to tumor targeting compounds. In another part of this collaboration, homology-based protein structure prediction methods will be applied to the antibody proteins that actually target the radioisotopes to the cancer cells. We will also build on our earlier work on the protein complexes that load circular “clamps” onto DNA but studying alternative clamp-loading complexes. Such modeling will also be applied to the Rad51 protein responsible for repairing DNA double-strand breaks. We will continue to use computational docking methods to develop compounds that specifically bind toxin proteins by studying the light chain of botulinum neurotoxin, the most potent toxin known. Finally classical molecular dynamics simulations will be applied to evaluate the conformational flexibility of damaged DNA and correlate this with observed repair rates.

This project has yielded many benefits to the laboratory. Firstly, the research in this project has earned recognition and visibility for the new LLNL capabilities in computational biology, both within the DOE system and in the larger research community. Secondly, the research performed under this project provided the foundation and preliminary data for several successful research grant applications to the DOE and to the National Institutes of Health. Thirdly, the expertise in computational docking and protein structure prediction developed under this project has found application in the development of bioweapon detection technologies and in elucidating the virulence mechanisms of potential biowarfare pathogens. Finally, this project has brought new scientific talent to the laboratory, including three LLNL staff scientists originally hired as post-doctoral fellows for this project.

1.2 Research Overview

The scientific goal of this project was to apply a wide range of computational biology methods to help solve significant biological problems. To this end, during this Strategic Initiative we were involved in a wide range of collaborative biological research projects. The goals of these collaborative studies have included: understanding the mechanisms of enzymes

that repair damaged DNA, elucidating the mechanisms of how food-borne mutagens damage DNA, developing anticancer drugs with improved therapeutic effectiveness, determining the effect of a nanotube surrounding a piece of DNA, and determining the structures of proteins that bind to DNA to modulate a cell's life cycle. These research efforts have produced a large number of refereed publications and established the necessary expertise and collaborations to provide continued funding for these projects.

In summary, this Strategic Initiative LDRD Project has been successful in creating a core competency at the laboratory in computational biology. This project has used diverse human talent and applied laboratory resources to advance simulation methods to solve biology questions. The publications, conference presentations and community visibility arising from this project have lead to a sustained LLNL program in the important new discipline of computational biology.

2 Accomplishments FY01-FY03

2.1 Computational Biochemistry

The goal of the computational biochemistry has been to collaborate with experimental biologists on projects where advanced chemical simulations could have a significant impact on biological understanding. Over the course of the SI we expanded the number of collaborations and published a large number of papers in life sciences journals, thus demonstrating the effectiveness of joint experimental/theoretical approaches to elucidating biological questions. These projects are described in the following numbered sections. For each project at least one paper has been published or submitted for publication in the relevant biological literature.

2.1.1 Radioimmunotherapy

In a new collaboration with Professors Gerald and Sally DeNardo at the UC Davis Medical Center, our group has begun studying radioimmunotherapy (RIT) compounds. The goal is to develop novel compounds that target specific antigens on cancerous cells. Our group focused on two projects: 1) development of synthetic high affinity ligands (SHALs) that target HLADR10 on lymphoma and 2) development of bispecific multivalent single-chain antibodies (scFv) to be used to pretarget prostate cancer and to be followed by a small multivalent DOTA chelator of ⁹⁰Y. There are many aspects of this effort for computational biochemistry. The work of this SI

was used as preliminary data for a program project grant that was submitted to NIH and will be funded starting April 2004.

Radioimmunotherapy (RIT) with ^{90}Y -labeled immunoconjugates have shown promise in treating cancer because of its high-energy beta emission and a physical half-life of 2.67 days. The macrocyclic chelating agent 1,4,7,10-tetraazacyclododecane- $\text{N},\text{N}',\text{N}'',\text{N}'''$ -tetraacetic acid (DOTA) binds ^{90}Y with extraordinary stability. However, the ^{90}Y -DOTA immunoconjugate product yields have been typically only <50%. Improved yields are needed for RIT with ^{90}Y -DOTA immunoconjugates to be practical.

For Project 1, our goal was to aid in the design of new SHALs that target HLADR10 on lymphoma. This effort involved homology modeling, molecular docking, and molecular dynamics simulations. Figure 1 is a cartoon of the components of this effort.

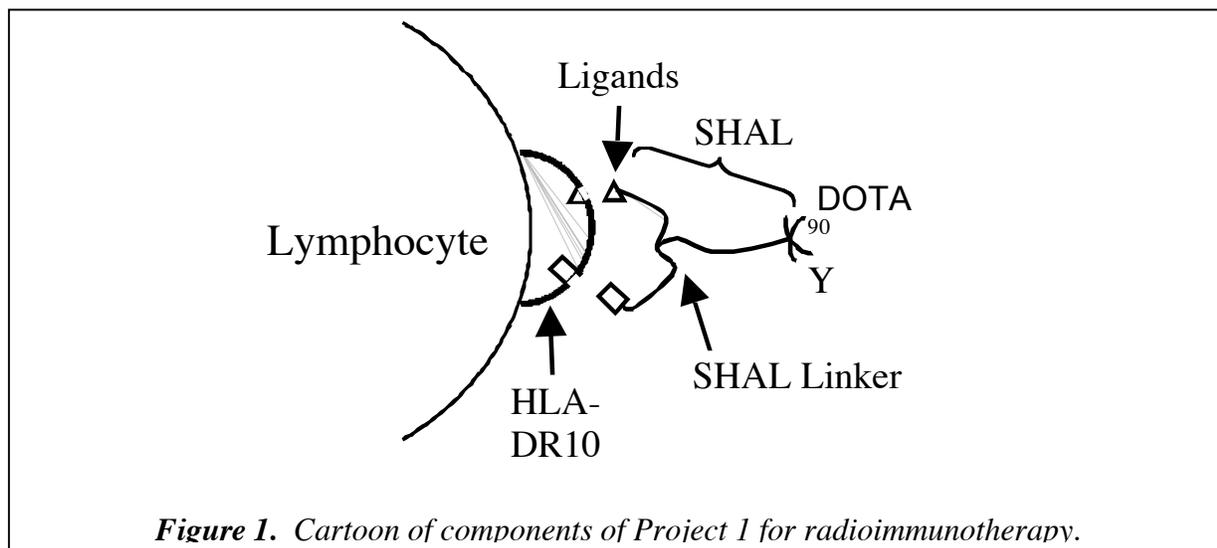
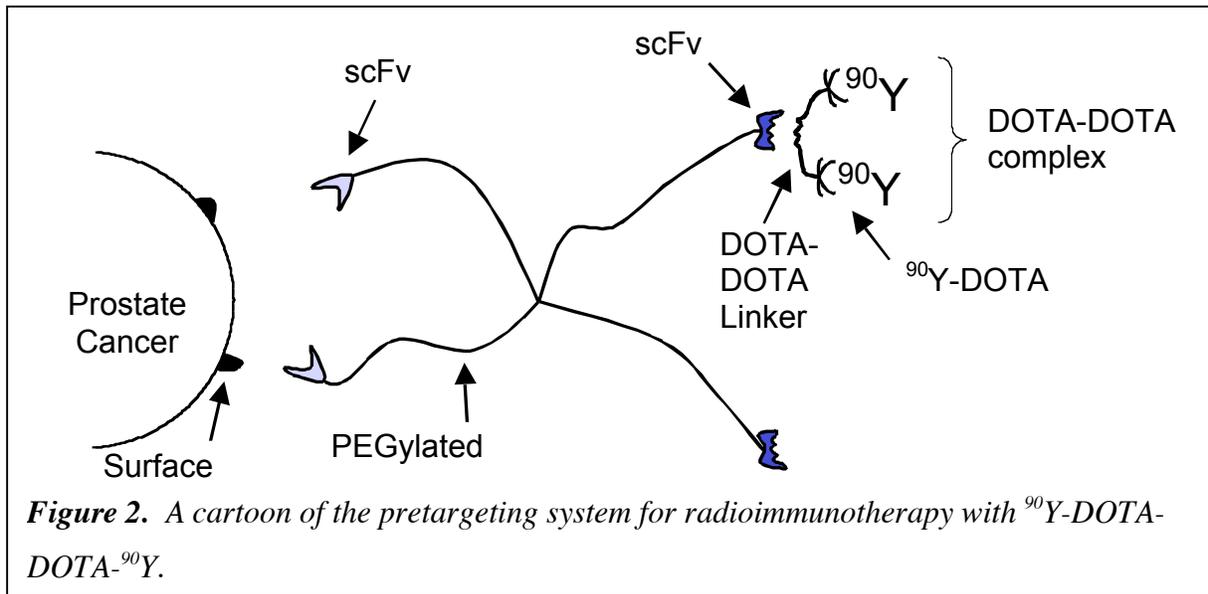


Figure 1. Cartoon of components of Project 1 for radioimmunotherapy.

For Project 2, single chain variable fragment (scFv) monoclonal antibodies have been raised against ^{90}Y -DOTA and with ^{111}In -DOTA with high binding affinities. These monoclonal antibodies are then used in a pretargeting system to target prostate cancer cells (Figure 2).



2.1.1.1 Molecular modeling of HLA-DR10

Crystal structures for four different closely related HLA-DR molecules (HLA-DR 1-4) have been determined previously and deposited in the PDB structure database by others. Protein sequences for these four proteins, HLA-DR1, HLA-DR2, HLA-DR3 and HLA-DR4, were aligned with the HLA-DR10 sequence and compared to identify both the locations of the variable amino acids and those regions of the HLA-DR10 molecule containing the amino acid residues that had been identified as the critical epitope of the Lym-1 antibody. This alignment revealed that all five proteins exhibit such a high degree of sequence similarity (Figure 3) that we should be able to create a sufficiently accurate 3-D model of the HLA-DR10 beta subunit by homology modeling and use the coordinates of the model to screen for ligand binding using the program DOCK.

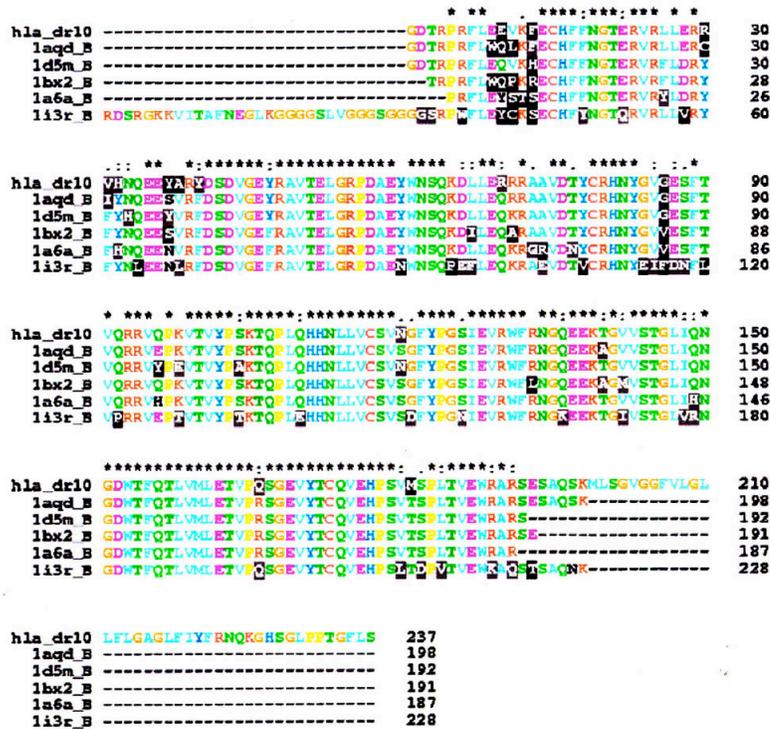


Figure 3. Amino acid sequence alignment of HLA-DR molecules with known crystal structures. PDB codes identify the HLA-DR1(1aqd), HLA-DR2(1bx2), HLA-DR3(1a6a) and HLA-DR4 molecules (1dm5) sequences. li3r is a homologous MHC fusion protein.

Two different approaches were used to create models of the HLA-DR10 beta subunit for use in ligand docking. The first approach used the coordinates of the entire structure of HLA-DR3 as the template for creating the homology model, and the nineteen amino acids that differed between HLA-DR3 and HLA-DR10 were mutated (changed) in the HLA-DR3 sequence. The coordinates of the amino-terminal four amino acids, which are present in HLA-DR10, HLA-DR1 and HLA-DR2 but absent in HLA-DR3, were obtained from the HLA-DR1 structure and used to complete the model. In the second approach, a hybrid model was generated using the atomic coordinates obtained from different segments of the HLA-DR 1, HLA-DR2 and HLA-DR4 crystal structures. The particular segments of the three HLA crystal structures used in the model were selected based on similarities in their secondary structural elements. Sequence-structure alignments were generated using the Smith-Waterman, FASTA, BLAST and PSI-BLAST algorithms, and the backbone of the model was created automatically using the AS2TS system (<http://sb9.llnl.gov/adamz/LGA/AL2TS/as2ts.html> see website). The coordinates for the amino terminal four residues of the structure were taken from the 1seB crystal structure of HLA-DR1,

residues 5-122 were obtained from the 1aqd structure of HLA-DR1, residues 123-170 were taken from the 1d5m HLA DR4 structure, and the remaining residues (aa171-193) were obtained from the 1bx2 structure of HLA-DR2. The construction of the terminal regions and loops, amino acid insertions and deletions, and template-model structure comparisons were performed using the LGA program developed at LLNL (see website <http://predictioncenter.llnl.gov/local/lga/lga.html>). The majority of the side chain atom's coordinates were incorporated from the four structural templates (listed above) due to their high level of homology. The side chains in selected regions of the protein model were built using the SCWRL program. Energy minimization was performed on both structures to eliminate inappropriate side chain contacts and the resulting structures were "optimized" using molecular dynamics.

Analyses of the resulting models revealed the two approaches yielded structures that were remarkably similar. Extended molecular dynamics runs appeared to provide little additional improvements. The results of the modeling revealed that the structure of the HLA-DR10 molecule is comprised of two domains linked by a hinge with one of the Lym-1 reactive residues, V85, positioned directly adjacent to the hinge (Figure 4).

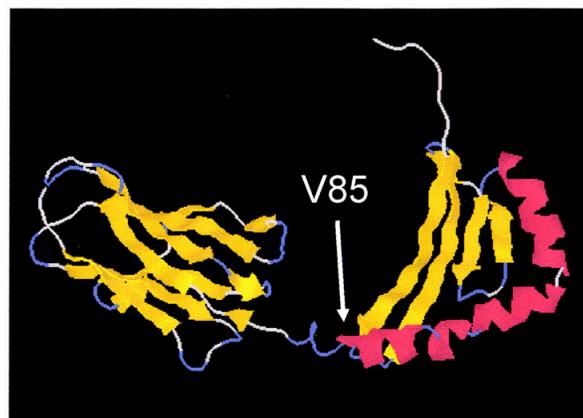


Figure 4. Homology model of beta subunit of HLA-DR10 showing the two structural domains. The majority of the core of the relaxed structure of HLA-DR10, when compared with the HLA-DR3 crystal structure, was found to be essentially identical (Figure 7). The other three amino acids that were observed to play a role in Lym-1 binding, R70, R71 and A74 (A or E at this position appears important for Lym-1 binding), are all located on the exposed surface of a long alpha helix (Figure 4 and 5) located immediately adjacent to the hinge.

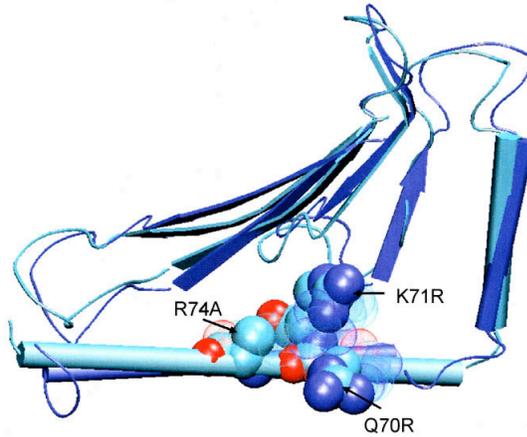


Figure 5. Superposition of HLA-DR3 crystal structure (light blue, transparent atoms) and homology model of HLA-DR10 (dark blue, solid atoms) showing the structural similarity in the region of the beta subunit that comprises the Lym-1 epitope. The amino acid residues critical for Lym-1 binding are shown as space filling atoms.

This effort was part of the Program Project grant submitted and funded by NIH.

2.1.1.2 Identification of potential binding sites on HLA-DR 10

The solvent accessible surface of the HLA-DR 10 protein and the crystal structures of HLA-DR 1, 2, 3, and 4 were calculated by using the generated model of HLA-DR 10 and the atomic coordinates of HLA-DR 1, 2, 3, and 4 obtained from the Protein Data Bank. Three key amino acids in the HLA-DR 10 sequence change both the charge distribution and the topology of the protein's solvent accessible surface in the Lym-1 epitope when compared to the homologous HLA-DR 1, 2, 3, and 4 protein surfaces. The solvent accessible surface of

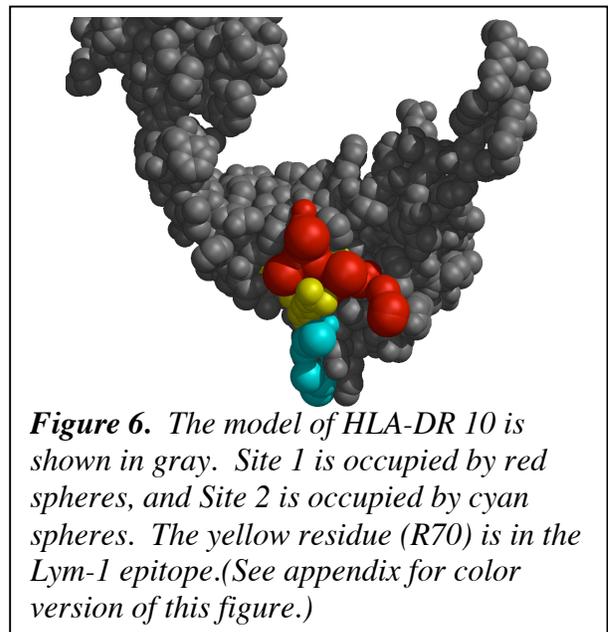


Figure 6. The model of HLA-DR 10 is shown in gray. Site 1 is occupied by red spheres, and Site 2 is occupied by cyan spheres. The yellow residue (R70) is in the Lym-1 epitope. (See appendix for color version of this figure.)

HLA-DR 10 was used by the program SPHGEN (part of the DOCK 4.0 suite of programs, UCSF) to determine potential binding site pockets on the surface of the protein. Two adjacent binding sites in the modeled HLA-DR 10 surface were selected as appropriate sites for ligand binding based on their proximity to each other, the three essential Lym-1 residues, and the

uniqueness of the amino acids lining the pocket. These sites, identified as Site 1 and Site 2 (Figure 6), flank both sides of the most important amino acid in the Lym-1 epitope, R70. This effort was part of the Program Project grant submitted and funded by NIH.

2.1.1.3 Computational Screening of Virtual Library

The program DOCK 4.0 was used to computationally screen the Available Chemical Directory (~300,000) of small molecules to identify the top ranked 2,500 molecules predicted to bind to the identified Site 1 and Site 2. In brief, to simulate a form of “flexible” docking, each ligand had 10 unique conformations generated, representing the range of possible conformations that could be adopted by the ligand. Each of these conformations was then docked into Site 1, scored energetically, and ranked. The top ranked 2,500 molecules were then visually inspected to select down to thirty-five molecules for experiment binding assays (Table 1). This final selection process was based on chemical properties including hydrophobic interactions, hydrogen bonding, and molecular size, and well as practical criteria including commercial availability and cost, ease of synthetic linkage, the overall structural diversity of the set of molecules.

Table 1. Ligands predicted to bind to Site 1 on the beta subunit of HLA-DR10 by computational docking

1. 7-Amino-4-chloromethylcoumarin, glycyl-l-proline amide, hydrochloride	14. '7-Amino-4-chloromethylcoumarin, l-alanyl-l-proline amide, hydrochloride
2. 5-([4,6-Dichlorotriazin-2-yl]amino)-fluorescein hydrochloride	15. '5-(N'-[2-aminoethyl]thioureidofluorescein)
3. 2-[2-[3-Chloro-5-(trifluoromethyl)-2-pyridinyl]carbohydrazonyl]benzene carboxylic acid	16. Achatin I, Ammonium salt
4. 4-[[2-(4-Cyano-3-phenyl-5-isoxazolyl)vinyl]amino]benzenecarboxylic acid	17. Fmoc-Asp(OBzl)-OH
5. 4-[2-(2,4-Dichlorophenyl)hydrazino]-4-oxo-2-phenyl-2-butenic acid	18. Fmoc-Bip-OH
6. 3-(2-[[3-Chloro-5-(trifluoromethyl)-2-pyridinyl]oxy]anilino)-3-oxopropanoic acid	19. Menai H535
7. 3-[(4-Chlorobenzyl)thio]imidazo[1,5-a]pyridine-1-carboxylic acid	20. 2'-Methoxy-5'-methyl-3,4,5,6-tetrachlorophthalanilic acid
8. 2-[[[1,1'-Biphenyl]-4-ylamino]carbonyl]benzoic acid	21. 4-Dimethylaminoazobenzene-4'-sulfonyl-l-valine
9. Bis[4-(3-aminophenoxy)phenyl]sulfone	22. Bigchap
10. 4,4'-Bis(4-aminophenoxy)biphenyl	23. Arg-gly-asp-thr
11. 5(6)-Carboxytetramethylrhodamine n-succinimidyl ester	24. n-Allyl-2-[(1-benzyl-2-oxo-1,2-dihydro-3-pyridinyl)carbonyl]-1-hydrazinecarbothioamide
12. 1,4-Phenylenebis[[4-(4-aminophenoxy)phenyl]methanone]	25. n'-Methoxy-n-[7-(4-phenoxyphenyl)[1,2,4]triazolo[1,5-a]pyrimidin-2-yl]iminoforamide
13. H-Tyr(Br-Z)-OEt	26. n-[[[6-(4-Chlorophenoxy)-3-pyridyl]carbonyl]-n'-[3-(trifluoromethyl)phenyl]urea
	27. n-[[[6-(4-Chlorophenoxy)-3-pyridyl]carbonyl]-n'-(4-chlorophenyl)urea
	28. n,n'-Diphenylbenzidine
	29. Rcl s16,963-3
	30. N,N'-bis-(4-amino-2-chloro-phenyl)-terephthalamide4-[[5-(trifluoromethyl)pyridin-2-yl]oxy]phenyl N-phenylcarbamate

- | | |
|--|---|
| <p>31. Methidiumpropyl ethylenediamine tetraacetic acid</p> <p>32. N-(4-[[3-Chloro-5-(trifluoromethyl)-2-pyridinyl]methyl]phenyl)-4-iodobenzenecarboxamide</p> | <p>33. 2-(4-Chlorophenyl)-2-[6-[(4-chlorophenyl)sulfanyl]-3-pyridazinyl]acetamide</p> <p>34. 6-Chloro-n4-(4-phenoxyphenyl)-2,4-pyrimidinediamine</p> <p>35. 4-Amino-2-anilino-5-benzoyl-3-thiophenecarbonitrile</p> |
|--|---|

This procedure was repeated for identification of thirty-five potential binders for Site 2.

Table 2 are those ligands predicted by molecular docking to bind to Site. 2

Table 2. Ligands predicted to bind to Site 2 on the beta subunit of HLA-DR10 by computational docking

1. Leu-enkephalin
2. 4-[4-(4-Chlorobenzyl)piperazino]-3-nitrobenzenecarboxylic acid
3. Beta-casomorphin (1-2)
4. L-Aspartic acid, alpha-(4,5-dimethoxy-2-nitrobenzyl) ester, hydrochloride
5. Cefadroxil
6. 3,3',5-Triiodo-dl-thyronine
7. H-Glu(anilide)-OH
8. H-Trp-phe-OH
9. Glycylglycyl-D,L-phenylalanine
10. Thymopoietin II (33-36)
11. Thiophosphoric acid S-(3-(3-amino-propylamino)-propyl) ester, di-hydrate
12. Dynorphin A (13-17), Porcine
13. l-Alanyl-l-alanyl-l-tryptophan
14. Asp-arg-val-tyr
15. A-VI-5
16. Glu-thr-pro NH₂
17. (D-ala2)-Beta-casomorphin (1-5) (bovine)
18. Tyr-D-ala-gly-phe-D-met acetate salt
19. Arg-gly-asp-thr
20. N-Alpha,n-omega-di-cbz-l-arginine
21. Asp-lys acetate salt
22. (+)-Allo-octopine
23. Sodium 7-[(2-amino-2-phenylacetyl)amino]-3-methyl-8-oxo-5-thia-1-azabicyclo[4.2.0]oct-2-ene-2-carboxylate
24. Val-ile-his-asn
25. 2-Amino-8-(diphenylphosphinyl)-octanoic acid
26. Glu-his-pro NH₂
27. Bap-arg(m)aba
28. H-glu(lys)-OH
29. Bis-boc-l-arg
30. Htr)-OH
31. 4-Aminomethyl-L-phenylalanine boc
32. H-met-met-OH
33. H-trp-ile-OH
34. N-alpha benzoyl-arginine-4-amino benzoic acid
35. (Thr46)-osteocalcin (45-49) (human)

As a result of the computational docking, 30 compounds from both the Site 1 and Site 2 lists were experimentally screened by NMR. Eleven compounds were found to bind to HLA-DR10, giving a successful hit rate of 37%. These ligands are listed in Table 3.

Table 3. NMR screened ligands that were found to bind to HLA-DR10. Ligands are separated by computationally predicted docking sites.

Site 1:	Site 2:
5(6) carboxytetramethylrhodamine-n-succinimidyl ester	N-alpha benzoyl-arginine-4-amino benzoic acid
Methidiumpropyl EDTA	5-leu-enkephalin (YAGFM)
Deoxycholic acid	N alpha N omega dicarbobenzoxyarginine
Fmoc-aspartic acid(O-benzyl)-OH	Angiotensin II (DRVY)
4-dimethylaminoazobenzene-4'-sulfonyl-L-valine	Bis-BOC-L-arginine
4-[[5-(trifluoromethyl)pyridin-2-yl]oxy]phenyl N-phenylcarbamate	

From Table 3, 5 synthetic high affinity ligands (SHALs) have been synthesized, and 2 SHALs have been tested for binding. By surface plasmon resonance, SHAL JP459 has a K_d of 65 nM, and SHAL JP459B has a K_d of 23 nM. SHAL JP459B has also been shown to compete for binding against Lym-1 on HLA-DR10 in Raji cell lysates and will bind intact Raji human lymphoma cells. This effort was part of the Program Project grant submitted and funded by NIH. We have also submitted an ROI for the work.

2.1.1.4 Molecular modeling of anti-MUC-1 single-chain antibody structures

We have applied homology-based protein structure prediction to model two of the human single-chain antibody constructs, named clone 12E and clone 3D, (named for their coordinates in a multiwell plate) that showed promise in targeting cancer cells. For modeling both of these constructs we used as a template the known structure of the murine single-chain F_v antibody MFE-23 (PDB code: 1QOK).

A three-dimensional model for the clone 12E is given in Figure 7. The model includes two structural domains corresponding to the variable heavy-chain and light-chain domains. The linker between these domains was not explicitly modeled, since the crystallographic study of single-chain F_v antibody MFE-23 suggests that this linker is flexible and most likely does not assume a unique conformation. Short sequence fragments at the N- and C-terminus of clone 12E, extending beyond the structure of V_H and V_L domains are also expected to be flexible and were therefore not modeled. The modeled structure of clone 12E is very similar to that of MFE-23

including four out of six hypervariable antigen-binding loops. Two loops (H1 and H3) display marked differences from the MFE-23 template, and H3 loop has the most dissimilar conformation.

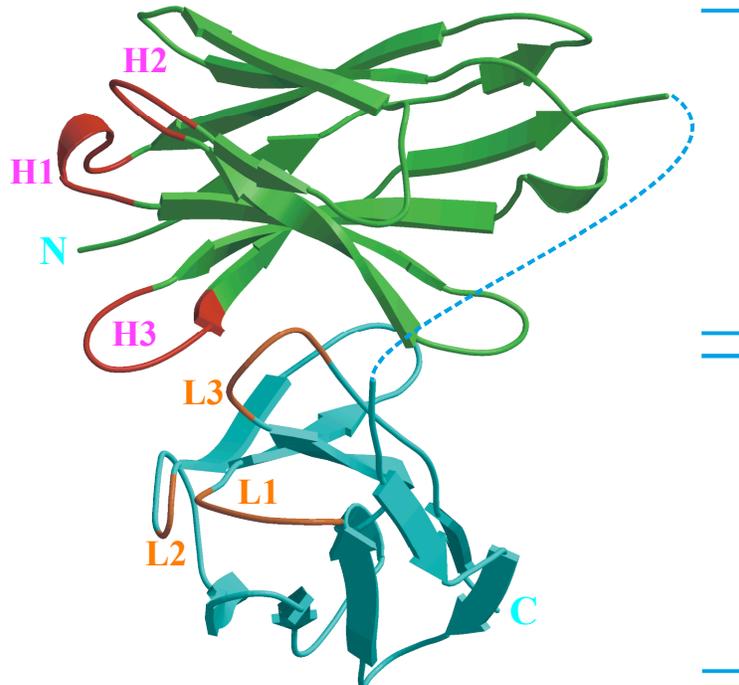


Figure 7. A three-dimensional model for the clone 12E. Variable heavy-chain (V_H) and light-chain (V_L) domains are colored green and cyan respectively. Colored red (H1-H3) and orange (L1-L3) are the hypervariable antigen-binding loops of respectively V_H and V_L domains. The approximate path of Gly/Ser-rich connector between domains is indicated with a dashed line. (See appendix for color version of this figure.)

The clone 3D sequence includes a variable heavy-chain region and a Gly/Ser-rich linker, identical to those in clone 12E (Figure 8). However, instead of complete variable light-chain domain, the sequence of clone 3D following the linker is a composition of short subsequences of the N-terminal and C-terminal regions of light-chain domain. Since this short composite region is flanked by Gly/Ser-rich linker and the E-tag that are not expected to have rigid conformations, this region is unlikely to form stable three-dimensional structure. Secondary structure prediction for the clone 3D sequence suggests that at most this composite C-terminal region could form a β -hairpin.

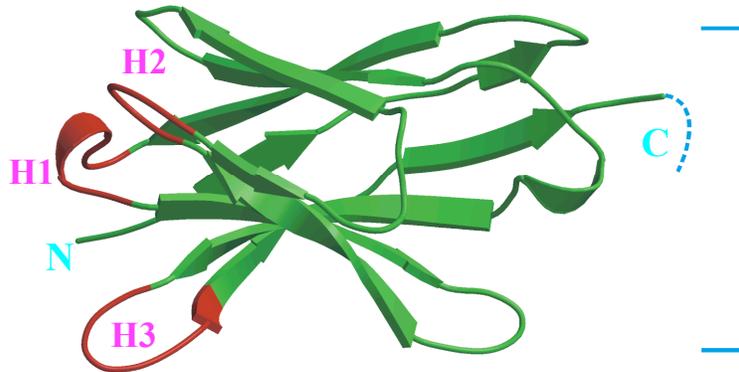


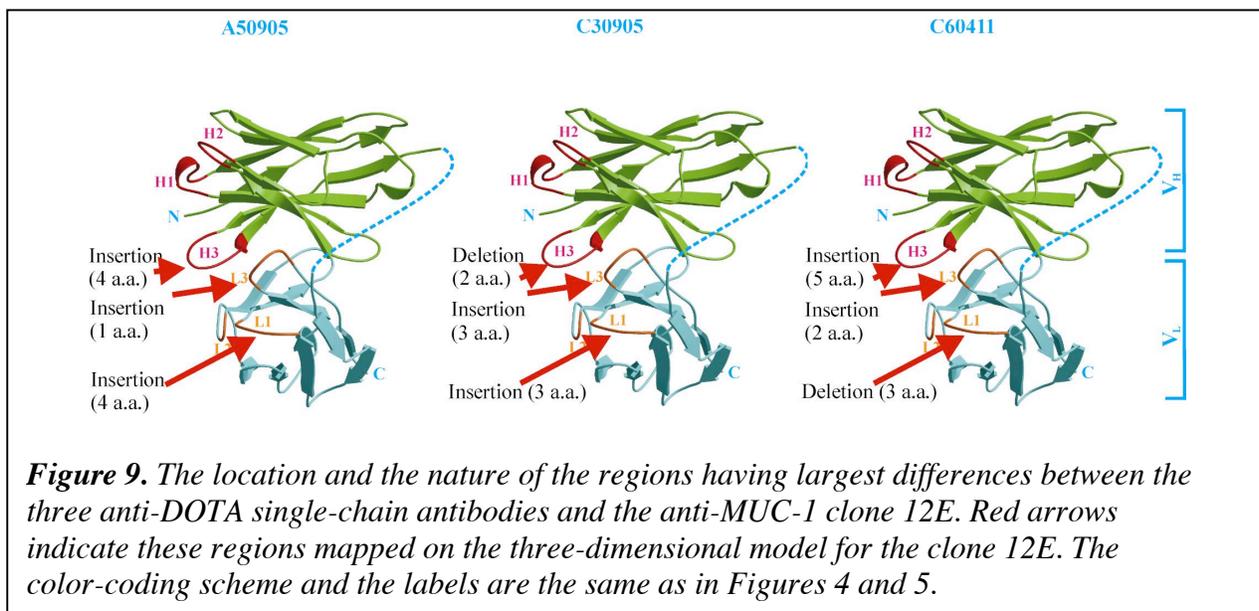
Figure 8. A three-dimensional model for the clone 3D. Colored and labeled in red are antigen-binding loops of the variable heavy-chain (V_H) domain region. (See appendix for color version of this figure.)

Both the 12E and 3D clones bind the same antigen with almost identical affinity. At the same time, the C-terminal part of clone 3D, following Gly/Ser-rich linker, does not include any of the light-chain antigen-binding loops, and, as discussed above, also is not expected to form a stable three-dimensional structure. As can be seen from the structural model, the Gly/Ser-rich linker is on the other side of the heavy-chain domain hypervariable loops, making it unlikely that the C-terminal part of clone 3D would affect the interaction of these loops with the antigen. Thus, these modeled structures, combined with the binding affinity measurements, strongly suggest that in both clones the antigen-binding is mediated only (or mostly) by heavy-chain antigen-binding loops (H1-H3). This is in good agreement with the antigen-binding mode suggested for MFE-23, based on the intermolecular packing in the crystal.

Since the selected template structure has high sequence homology (73% identical residues) as well as an identical linker sequence connecting the variable heavy-chain and light-chain domains, we expect the model to be comparable in quality to a structure that could be obtained using X-ray crystallography. We expect that the comparative modeling of antibodies should yield high-quality predicted structures because the database of known structures (PDB) currently holds close to 1000 antibody structures. This available structural information allows comparative modeling to be effective not only in modeling conserved antibody core but also variable antigen recognition loops. This work has been published in *Clinical Cancer Research* and was part of the Program Project grant submitted to and funded by NIH.

2.1.1.5 Sequence analysis of single-chain Fv antibodies raised against DOTA.

We have performed preliminary computational analysis of the recently determined sequences of the three clones of single-chain antibody constructs (A50905, C30905, C60411) selected by their ability to bind DOTA. The analysis has been done using standard pairwise sequence comparison methods. We have looked at the differences between the sequences of these clones and the sequence of anti-MUC-1 single-chain antibody (clone 12E), selected to bind cancer cells. In the course of the analysis we identified regions in the anti-DOTA antibodies that displayed the largest differences compared with the anti-MUC-1 antibody 12E (see Figure 9).



Surprisingly, in all three anti-DOTA antibodies these regions were localized within three (L1, L3 and H3) of the six variable loops responsible for antigen recognition. This observation suggests that the specificity for DOTA binding by these antibodies resides in these particular loops. The next step will be to develop the all-atom molecular models needed for detailed structural analysis of these regions. This effort was part of the Program Project grant submitted to and funded by NIH.

2.1.1.4 Simulations on the DOTA-DOTA linker configurations

Two ⁹⁰Y-chelating DOTA groups are linked together by a chemical chain. Information about the structural conformation of this DOTA-DOTA linker, both in solution and in possible scFv binding sites is necessary to understand the factors leading to efficient and selective capture of the ⁹⁰Y-DOTA-DOTA by the “anti-DOTA” single-chain antibody.

2.1.1.4.1 Quantum chemical calculations of DOTA-DOTA linker.

As a first step in this structural study, we have performed quantum chemical simulations of the DOTA-DOTA linker currently in use, 2IT-BAD. The DOTA moieties were not included in these initial simulations, but were simply replaced by methyl groups (Figure 10, indicated by arrows in structure (a)). Classical MD simulations of the 2IT-BAD linker are described in Section 2.1.1.4.2.

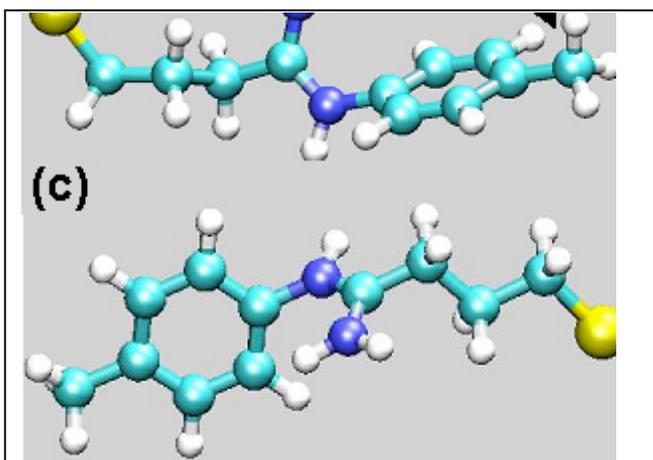


Figure 10. Different calculated conformations of the DOTA-DOTA linker, 2IT-BAD. (a) Configurations 1-3, (b) Configuration 4 (c) Configuration 5. (See appendix for color)

An initial set of linker configurations was generated by a simulated annealing process implemented in the molecular modeling software package AMPAC (using the AM1 semiempirical method to calculate the conformational energies). This procedure produced a list of different molecular configurations of the linker, ordered by energy. The five lowest energy conformations generated by the simulated annealing had a relative energy range of 16 kcal/mole and the next lowest energy configuration was 71 kcal/mole higher in energy. We took the five lowest energy conformations and performed a full gas-phase structural optimization using *ab initio* density functional theory with B3LYP functional and 6-31G(d) basis set. These optimizations led to no significant change in the linker conformations from that generated by the simulated annealing process. The B3LYP/6-31G(d) relative energies are listed below in Table 4 and the optimized structures are shown in Figure 7. (Note that the first three conformations differ only in the rotations of the terminal methyl and phenyl groups and are not shown separately.) Configurations 1-3 are “closed” structures stabilized by an intramolecular hydrogen bond between an amine and an amide carbonyl. Configuration 4 is a “semi-closed” structure lacking an intramolecular hydrogen bond and configuration 5 is a fully extended conformation. Inclusion of solvent effects using the COSMO dielectric continuum model dramatically changes the ordering of the energies and indicates that the fully extended chain is favored in water. This is because the polar amine, amide, and thiol groups are most available for favorable interactions with water in this extended

conformation. These results demonstrate the subtle mix of intramolecular and solvent effects that will determine the structure and rigidity of the DOTA-DOTA linker.

Table 4. Relative energies of five configurations of the DOTA-DOTA linker molecule as calculated by *ab initio* quantum chemistry. Energies in kcal/mole.

Config #	Gas-phase Energy	Aqueous-phase Energy
1	0.0	0.0
2	0.0	-0.2
3	0.0	1.1
4	9.7	3.5
5	15.8	-2.8

This effort was part of the Program Project grant that was submitted to and funded by NIH.

2.1.1.4.2 Classical molecular dynamics simulations of ⁹⁰Y-DOTA-LINKER-DOTA-⁹⁰Y.

To effectively crosslink the scFv's on the scaffolding, the DOTA-DOTA linker must satisfy a number of structural and chemical criteria; it must: 1) not interfere with the DOTA-scFv binding; 2) not weaken the DOTA-⁹⁰Y interaction; and 3) adopt a conformation with a DOTA-DOTA distance that maximizes crosslinking. Since the PEGylated scaffolding is highly flexible, the optimal linker length corresponds to the most probable scFv-scFv distance. As an initial estimate of the most probable scFv-scFv distance, we used the valence angle model that predicts a PEG polymer of N monomers the most probable end-to-end distance is given by a constant times the square root of N. For the specific PEGylated scaffolding currently being evaluated in Project 2, this corresponds to a scFv-scFv distance of ~100Å. Designing a DOTA-DOTA linker that will optimally span this distance requires consideration of the contour length of the linker and the conformations adopted by a particular linker type (e.g. PEG vs peptide).

We have used classical MD to study the conformational distributions for several DOTA-DOTA linkers. The program CHARMM was used to perform classical molecular dynamics simulations on ⁹⁰Y-DOTA-PEG-DOTA-⁹⁰Y and ⁹⁰Y-DOTA-2IT-BAD-DOTA-⁹⁰Y in solution to determine the distribution of ⁹⁰Y -⁹⁰Y distances sampled by these molecules. The simulations performed with PEG linkers had a different number of PEG units (4, 6, and 8) in the molecules. The starting structures for all four simulations have the molecules in a fully extended conformation. These extended molecules were solvated in water boxes and two sodium ions

were added to each water box to neutralize the systems. These solvated systems were heated to 300 K and allowed to equilibrate for 200 picoseconds. All simulations were run at constant temperature (NVT ensemble) and electrostatic interactions were treated by particle mesh Ewald (PME) summation. Each simulation was performed for a total of 2.0 nanoseconds. During the simulations, all molecules formed more compact structures in solution. Plots of the ^{90}Y - ^{90}Y distance distributions are shown in Figure 11 with representative structures of the molecules from the simulations. Molecules having a PEG linker had a single major conformer. Although the PEG linkers formed kinks during the simulations, the two DOTA units remain separated and solvent exposed. The simulation of ^{90}Y -DOTA-2IT-BAD-DOTA- ^{90}Y shows two primary conformers for this molecule in solution (Figure 8D). The structure corresponding to the peak at ~ 26 Å is fully extended. At the second maximum (~ 17 Å), the 2IT-BAD linker adopts a closed conformation similar to the structures obtained from *ab initio* calculations. This closed conformation causes the two DOTAs to be in close proximity.

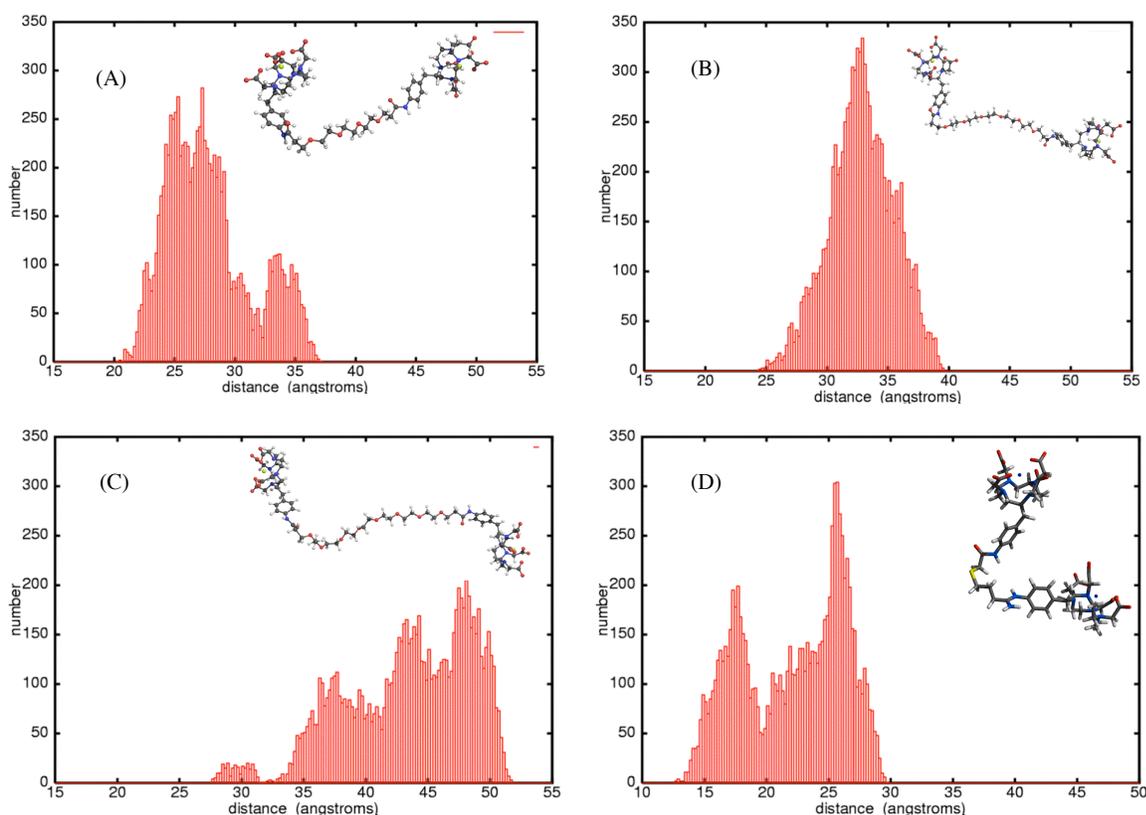
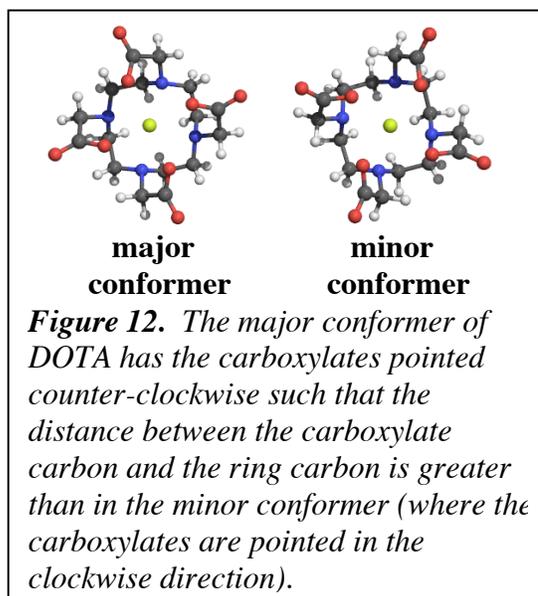


Figure 11. Distributions of yttrium-yttrium distances during the molecular dynamics simulations for the two ^{90}Y -DOTA linked by 4 PEG (A), 6 PEG (B), 8 PEG (C), and 2IT-BAD (D).

2.1.1.5 Quantum chemical calculations of ^{90}Y -DOTA and ^{111}In -DOTA

Li and Meares have shown experimentally, that the rate of loss of In(III) is 5 times greater than Y(III) when chelated to a peptide linked DOTA in human serum. To determine possible differences between ^{90}Y -DOTA and ^{111}In -DOTA in solution, quantum chemical calculations were performed with these molecules complexed with a water molecule. The program Gaussian 98 was used to perform the quantum chemical calculations. Hartree-Fock calculations were performed using the LANL2DZ effective core potential (ECP) for yttrium and indium and the 6-31G(d) basis set for all other atoms. There is a major and minor conformation for DOTA containing either an yttrium or an indium ion (Figure 12). A water molecule coordinating to the yttrium in ^{90}Y -DOTA causes small changes in the structure of both the major and minor conformers relative to the noncoordinated

structures. In the major conformer of ^{90}Y -DOTA, the yttrium is 1.7060 Å above the plane formed by the nitrogens in DOTA. When a water molecule coordinates to the yttrium, the metal ion rises to 1.8070 Å from the plane. A similar effect is seen in the minor conformer of ^{90}Y -DOTA (1.7770 Å without a water molecule and 1.9063 Å with a water molecule). When a water molecule coordinates to the indium of ^{111}In -DOTA, there is a considerable structural change in this complex. Without the water molecule, the indium is coordinated to the



four nitrogens in DOTA. When the water molecule is coordinated to the indium, the interaction between the nitrogens and indium is disrupted. The coordination changes from 8 to 5 and the indium rises in the DOTA ring relative to the structure without the water. This change in coordination occurs for both the major and minor conformers of ^{111}In -DOTA. In the major conformer, the indium is 1.707 Å above the plane formed by the nitrogens in DOTA (Figure 13A). Coordination of a water molecule causes the indium to be 1.979 Å above the plane (Figure 13B). An even larger effect is seen in the minor conformer of ^{111}In -DOTA. The indium is 2.270 Å above the plane when coordinated to a water molecule and 1.827 Å when not coordinated to a water molecule. Because a single water molecule can have such a large

structural change in the chelation of indium, first principles molecular dynamics simulations will be used to explore the solvation effects on the chelation of indium and yttrium by DOTA. This effort was part of the Program Project grant that was submitted to and funded by NIH.

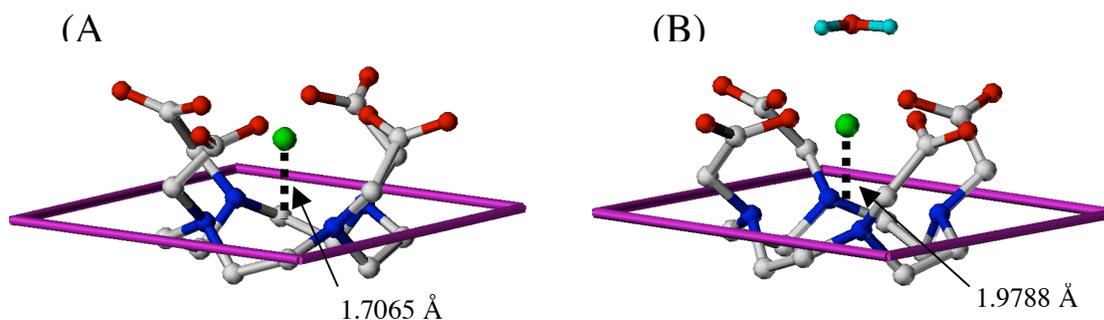
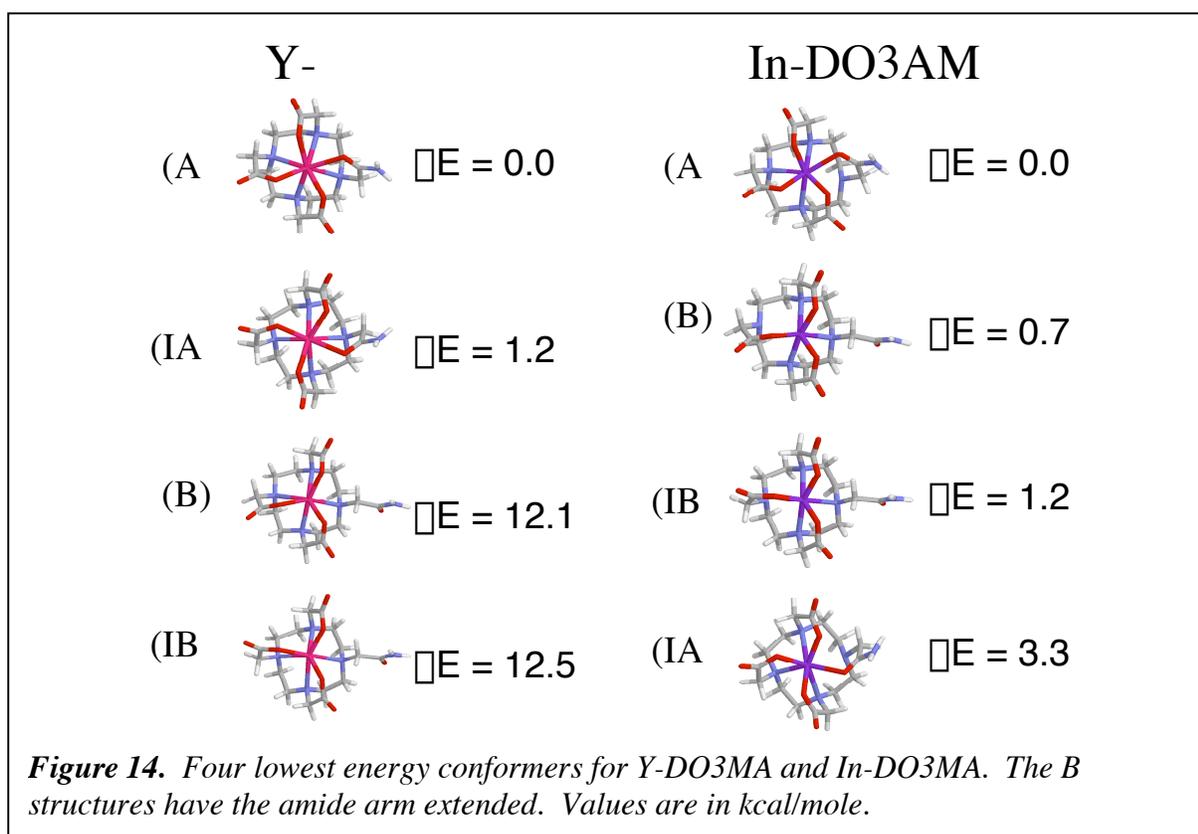


Figure 13. Major conformers of indium-DOTA without water (A) and with water (B) from quantum chemical calculations. The box represents the plane formed by the nitrogens in DOTA. The values are the indium distances from the plane. Hydrogen atoms in DOTA have been deleted for ease of viewing.

2.1.1.6 Quantum chemical calculations of Y-DO3AM and In-DO3AM.

DOTA is commonly attached to other molecules by reacting one of the carboxylate arms to form an amide linkage. A recent NMR study of yttrium and indium coordinated by modified DOTAs (1,4,7,10-tetraaza-4,7,10-tris(carboxymethyl)-1-cyclododecylacetyl-benzylamine) has shown differences in these complexes. The yttrium complex has one predominant conformer. The indium complex showed significant line broadening indicating multiple conformers in solution. Quantum chemical calculations were performed on yttrium and indium chelated by DO3AM (DOTA with one carboxylate modified to an amide) to determine these different conformers. Hartree-Fock calculations were performed using the LANL2DZ effective core potential (ECP) for yttrium and indium and the 6-31G(d) basis set for all other atoms. The energies were calculated using MP2/6-31+G(d)//HF/6-31G(d). The zero-point vibration energy (ZPE) for each complex was scaled by 0.893. The four lowest energy conformers for the metal complexes are shown in Figure 14 with their relative energies. The lowest energy structures (A) for both metal complexes have the same geometry but are in equilibrium with different conformers. Similar to Y-DOTA (see Figure 12), the two lowest energy conformers of Y-

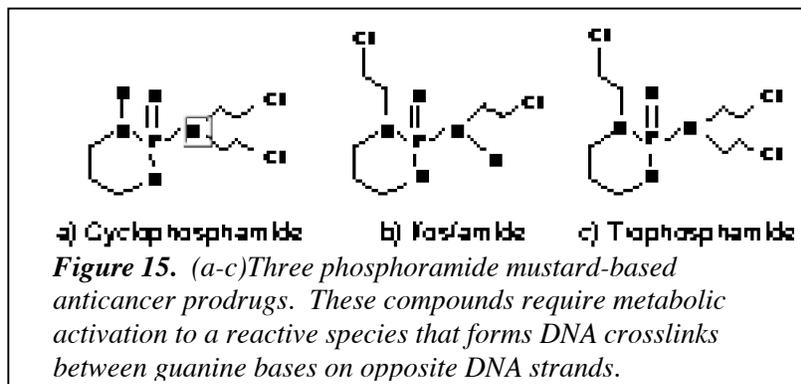


DO3MA (A and IA) differ only in the orientation of the arms coordinating the metal ion. For In-DO3MA, the major conformer is in equilibrium with structures which do not have the amide coordinated to the metal (B and IB). The three carboxylates coordinate the indium and the amide arm is extended away from the ion. The physical properties of the complex differ when the amide is not coordinated to the metal. The indium moves $\sim 0.5 \text{ \AA}$ relative to its position in the major conformer (A) and a water molecule does not coordinate to the metal ion (work in progress). These differences can have significant consequences if an antibody uses metal coordination in recognizing or binding this complex. This effort was part of the Program Project grant that was submitted to and funded by NIH.

2.1.2 DNA-Alkylating Anticancer Drugs

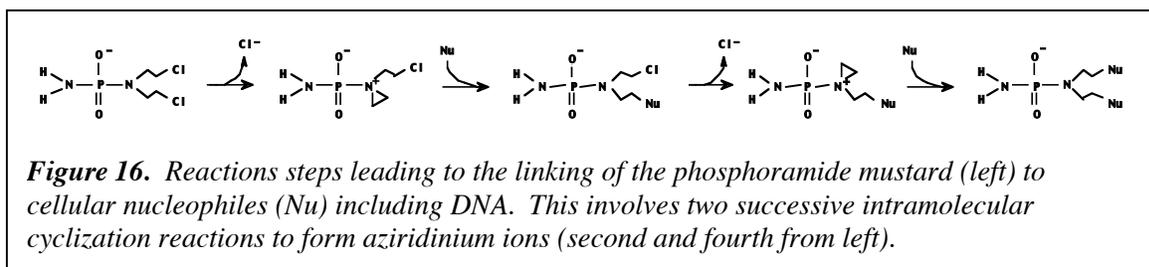
The nitrogen mustard based DNA alkylating agents were the first effective anticancer drugs and remain important drugs in the fight against many forms of cancer. In particular, cyclophosphamide and ifosfamide (see Figure 15) are members of the nitrogen mustard prodrug class that are currently widely used in cancer chemotherapy. More than fifty years of research on nitrogen mustards has yielded a broad range of therapeutically useful compounds and a detailed

knowledge of the biochemical mechanism of these drugs. Nevertheless, there is much ongoing research on the phosphoramidic mustards to increase their potency and reduce their toxic and mutagenic side effects.



Because of the existing detailed knowledge of these drugs, it is hoped that atomic-level chemical modeling will play an important role in developing improved analogs.

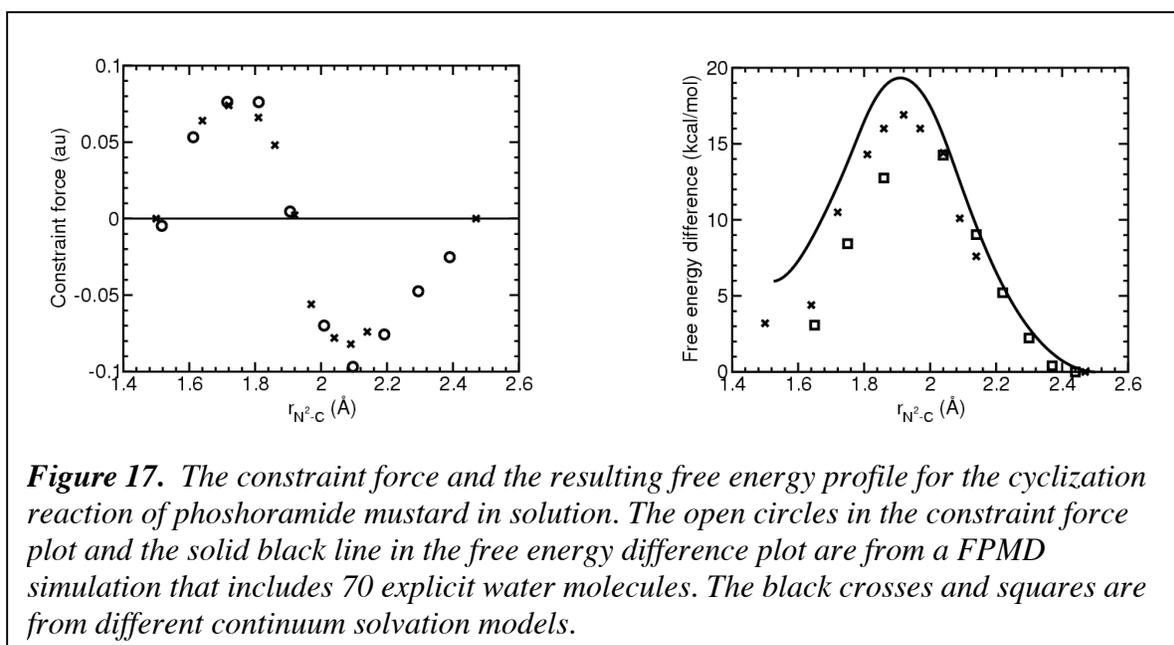
The phosphoramidic mustard prodrugs shown in Figure 15, are known to require metabolic activation in order to liberate an active phosphoramidic mustard. Overwhelming experimental evidence indicates that the phosphoramidic mustards alkylate DNA via a highly electrophilic aziridinium species formed from an intramolecular SN₂ reaction. In this activation process (see Figure 16), the mustard nitrogen sequentially attacks each of the carbons in the chloroethyl moieties, displacing the chloride to form the aziridinium ion that is subsequently reopened by nucleophilic attack at one of the carbon atoms.



2.1.2.1 First Principles Molecular Dynamics Simulations on the Activation of Phosphoramidic

In order to examine the effects of an aqueous environment on the activation of phosphoramidic mustard, we have performed a series of molecular dynamics simulations. In particular, we have used FPMD simulations with explicit solvation and with a continuum solvation model to examine important features of the cyclization reaction. It has been shown in several studies that de-esterification or deprotonation of the phosphate to form its conjugate base is necessary for aziridinium ion formation. The chemical origins of the strong effect of phosphate esterification on phosphoramidic mustard alkylating ability are not yet known. We

have used FPMD simulations to investigate the several possible chemical factors underlying this effect, including: increased nucleophilicity of the mustard nitrogen, better conformational accessibility of the transition state, or electrostatic stabilization for loss of the chloride. The long-term outcome of these studies of the phosphoramidate mustard activation reactions will be the capability to use computational simulations to design new mustard analogs with modulated activation rates. In our initial investigation, we used first principles molecular dynamics to evaluate the effect of an aqueous environment on the free energy profile for the cyclization reaction on phosphoramidate mustard (Figure 16). We have used constrained molecular dynamics



simulations at ten points along the cyclization reaction path. At each of these points, the force associated with maintaining a fixed N-C distance was monitored and used to determine a free energy difference profile for the reaction. In Figure 17, the free energy profile from our FPMD simulation is shown along with two implicit solvation model calculations that were obtained in FY01. Although the implicit solvation model results are in agreement with the FPMD simulation for the initial phase of the cyclization reaction, there are significant differences near the top of the reaction barrier. A manuscript describing this work is currently being prepared for publication.

2.1.2.2 Classical Molecular Dynamics Simulations of DNA Crosslinking

The phosphoramidic mustard anticancer drugs, including cyclophosphamide and ifosfamide, are widely used for treating many forms of cancer and are finding new uses as

immunosuppressive agents. However, a number of toxic side effects, including secondary tumors, limit their clinical utility. The long-term goal of our research in this area is to develop new versions of these drugs that are effective at lower doses and have fewer side effects. There is strong evidence that the effectiveness for killing cancer cells (cytotoxicity) of these drugs is directly related to the ability of their metabolites to form interstrand DNA crosslinks. These crosslinks are formed by the sequential alkylation of two bases on complementary DNA strands (see Figure 18). Other products, including monoalkylated adducts and intrastrand crosslinks are much less cytotoxic, and can mutate the DNA in normal cells leading to the formation of secondary tumors. Our objective is to develop analogs of phosphoramidic mustard metabolites that are designed to increase the fraction of interstrand DNA crosslinks resulting from the initially monoalkylated DNA. Once developed, these analogs could be incorporated into new versions of the anticancer drugs that would have therapeutically useful cytotoxicity at lower doses. To this end, we collaborated with Dr. Dat Nguyen (now at Harvard University Medical School) and Prof. Bill Fink of UC Davis to perform molecular dynamics simulations of final DNA crosslinking step to determine how variations in the chemical structure of these drugs are related to the efficiency of crosslink formation.

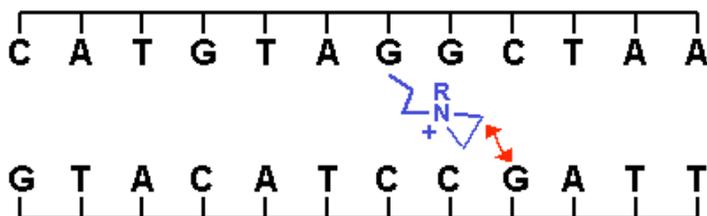


Figure 18. This diagram shows alkylating agent (shown in blue) attached to a single strand of the DNA. The next step is the reaction of the agent to the other DNA strand. The rate of this reaction is dictated by the distance between the guanine N7 and the reactive site on the alkylating agent. (Distance shown by red arrow.) One set of results is shown in the figure below. These graphs plot the N7 to reactive carbon distance (shown as red arrow in Figure 18) versus simulation time for two possible versions of the alkylating agents. These results indicate that the phosphoramidic mustard (Figure 19, upper plot) should be more efficient at forming chemical crosslinks than a larger analog (Figure 19, lower plot). These results are currently being prepared for submission to the *Journal of Medicinal Chemistry*.

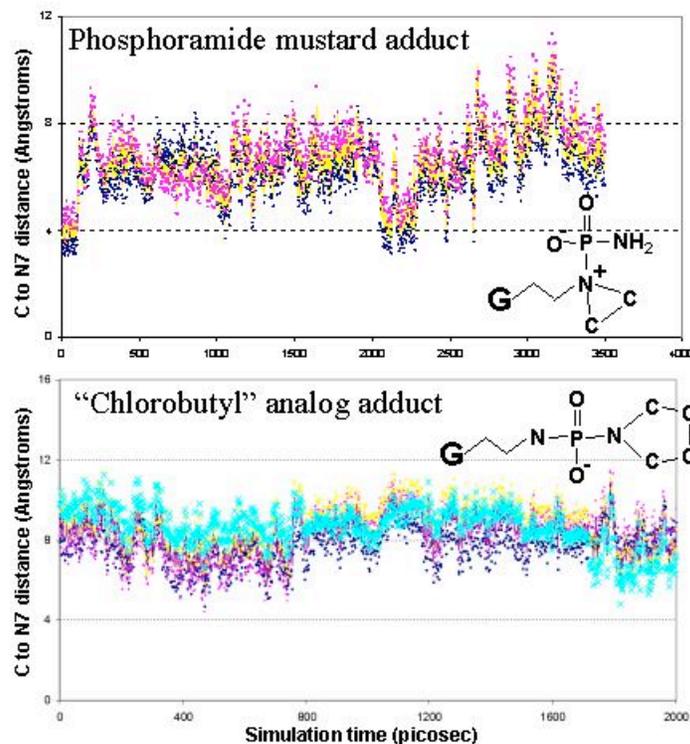


Figure 19. Plot of distances between the guanine N7 and the reactive carbon in the alkylating agents calculated from classical molecular dynamics simulations. The overall shorter distance observed for the phosphoramidate mustard (top) compared to the “chlorobutyl” analog (bottom) indicates that the former will be a more effective crosslinking agent with fewer mutagenic side effects.

2.1.2.3 Acrolein-Guanine Reactions

Acrolein is a chemical byproduct produced in the metabolic activation of the anticancer drugs cyclophosphamide and ifosfamide. Acrolein has long been known to cause kidney damage, one of the dose-limiting side effects of this class of drugs. However, there is now some evidence that acrolein can also react with DNA bases and through causes DNA mutations may contribute to treatment-induced secondary cancers. We collaborated with Prof. Susan Ludeman and her colleagues at the Duke University Cancer Center to investigate the nature of acrolein reactions with the DNA base guanine. In particular, we used first principles quantum chemical methods to model the initial reaction of acrolein at four different sites on guanine (primary reaction) as well as all possible subsequent ring-closures at other sites on guanine (secondary reactions). Figure 20 below summarizes the results for the primary and secondary reactions of acrolein with guanine. These results indicate that the “N2-N1” tricyclic structure should be the dominant final

product of acrolein-guanine reactions. This work has been published in *Chemical Research in Toxicology*.

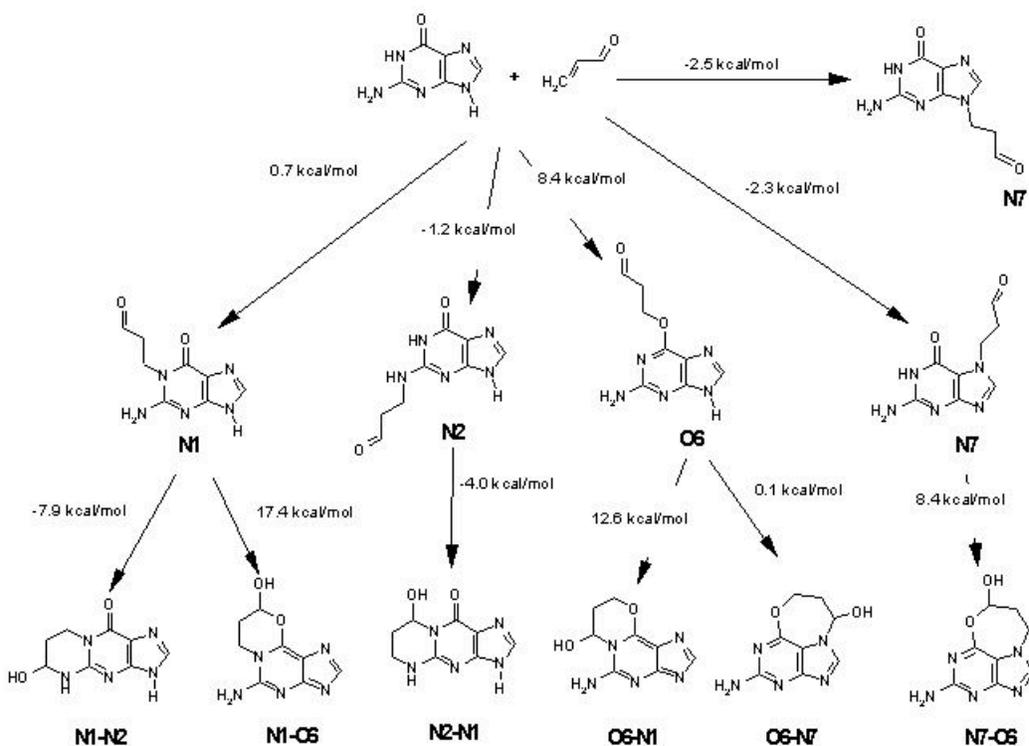


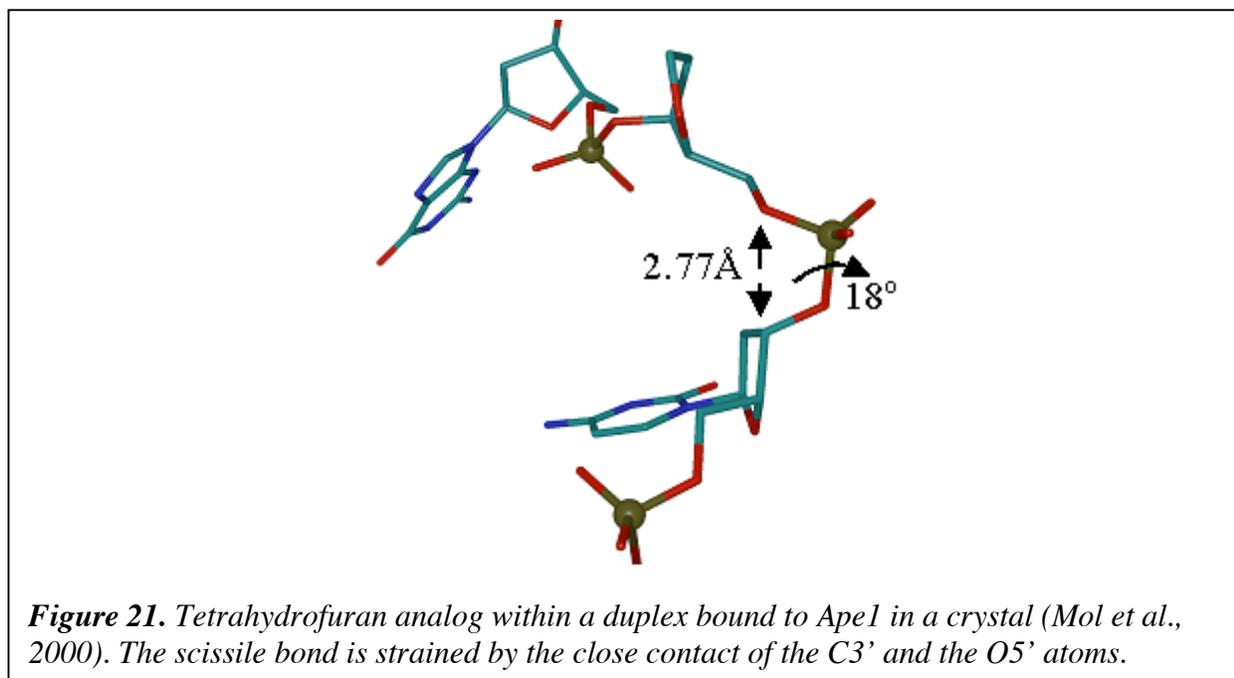
Figure 20. Aqueous-phase reaction free energies of guanine and acrolein in kcal/mole. Energies calculated at the B3LYP/6-31+G**//B3LYP/6-31+G** level of theory; solvation effects included using the COSMO solvation model.

2.1.3 Abasic DNA Repair

A mechanistic understanding of how the body recognizes DNA damage is key to understanding the mutagenic potency of chemicals and radiation and to developing therapeutics that modify DNA repair efficiency. To avoid mutations DNA damage must be repaired in advance of cellular replication, and efficient DNA repair depends on DNA damage recognition factors, such as altered DNA structure or structural flexibility. Baseless sites in DNA—a prevalent type of DNA damage also known as abasic (AP) sites—are bound by the DNA repair protein, human apurinic/apyrimidinic endonuclease (Ape1), with a selectivity one billion times greater than that of non-AP sites. Through data gleaned from a series of previous computational and experimental studies in our laboratory (Erzberger et al., 1998; Barsky et al., 2000) and X-ray crystallography (Mol et al., 2000), it appears that the Ape1 repair protein binds its DNA targets

with high specificity due to a unique flexibility in DNA at AP sites that is not present in undamaged DNA.

The arrival of the first X-ray crystal structure of AP-DNA bound to Ape1 in January 2000 allowed us to immediately ask the question: To what extent do the AP sites take on the contortions observed in the co-complex crystal structure, and which contortions in particular are the hallmarks of specificity. Figure 21 shows that the C3'•••O5' distance is very short, and the



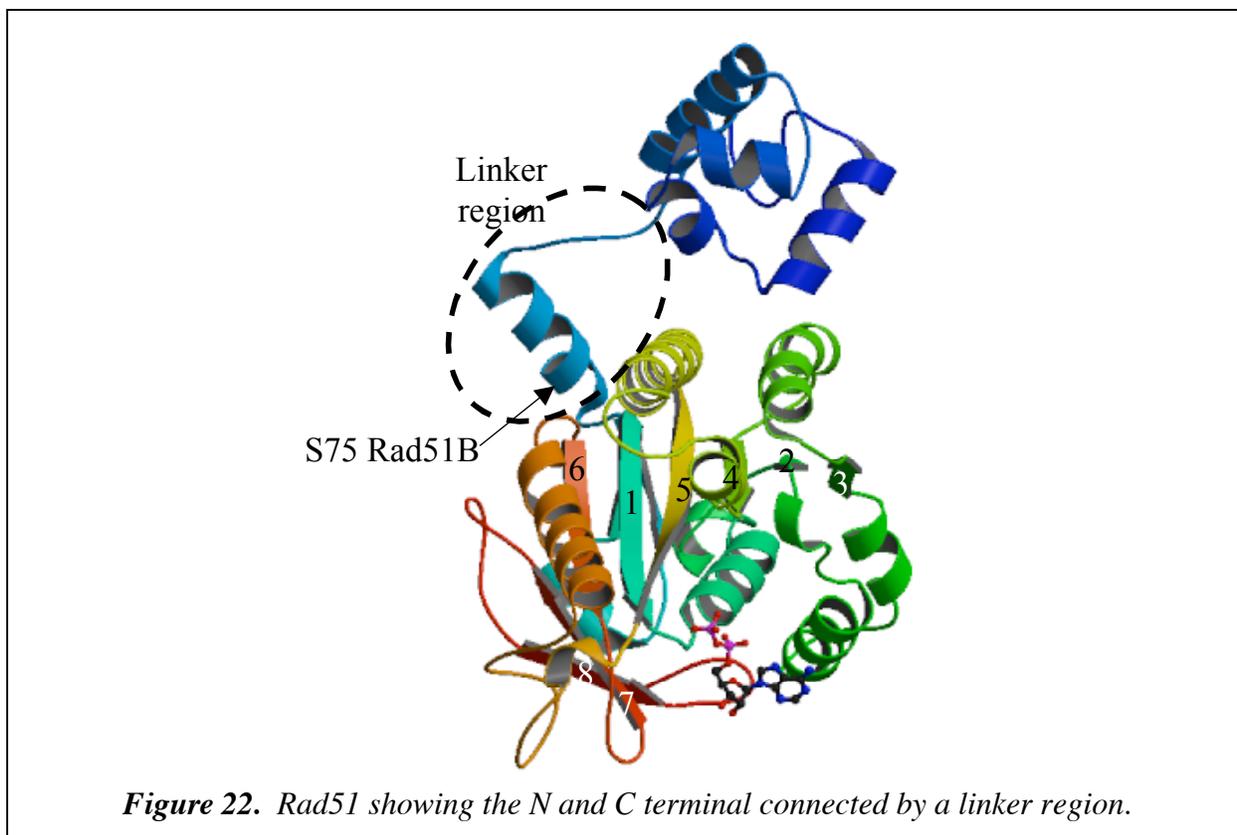
C3'-O3'-P-O5' dihedral angle is close to 0°. Our earlier work suggests that substrate flexibility seems essential for Ape1 activity. It is only by sifting through the trajectories of various DNA substrates, however, that we can begin to understand what specifically we mean by flexibility.

We have also performed simulations of the major apurinic/aprimidinic endonuclease (Ape1) enzyme itself, essential for life because it recognizes DNA damage and begins the repair process. Several mutations have been identified in the human population and are associated with an increased risk for cancer. We have used molecular dynamics simulations to analyze the way that these mutations might affect the function of this protein. Our results suggest that, for example the L104R mutation affects the way this protein works with other proteins, making the overall repair pathway less effective. Another mutation, R237C, is not only destabilizing to the protein, but our simulations suggest that the protein might actually be less selective, potentially

leading to occasional mistakes in its repair. This hypothesis is currently under investigation in David Wilson's lab.

In human DNA repair the Rad51 family of proteins is absolutely essential for cell viability, by eliminating DNA double strand breaks that occur naturally. Implicated in the origins of breast cancer, these proteins are currently the focus of many researchers, including three experimental groups in the BBRP of LLNL. These proteins have been found to form two distinct complexes *in vivo*, Rad51B-Rad51C-Rad51D-Xrcc2 (BCDX2) and Rad51C-Xrcc3 (CX3). By molecular modeling we have provided structural predictions of several rad51 proteins (rad51, rad51B, rad51C, rad51D, and xrcc3).

Furthermore, by combining homology modeling with experiments, we have identified the binding domains within the BCDX2 and CX3 complexes. Based the recent *Pyrococcus furiosus* Rad51 structure, we have used homology modeling to design truncation mutants of five Rad51 proteins: Rad51B, hsRad51C, Rad51D, and hsXrcc3 proteins. The models reveal distinct N-terminal and C-terminal domains connected by a linker region (see Figure 22). The C-terminal domains contain intertwined beta-strands, suggesting why certain deletion mutants exhibit no binding activity in yeast two-hybrid assays and enabling subsequent rational design of active deletion mutants. Yeast two-hybrid and co-immunoprecipitation techniques verify the predictions based on our models, namely that N-terminal domain of one paralog will interact with the C-terminal domain of the adjacent paralogs. The experiments demonstrate that a deletion construct of hsRad51B containing residues 1-75 of the N-terminus interacts with the linker and C-terminus of hsRad51C, residues 79-376, and this region of hsRad51C also interacts with mmRad51D and hsXrcc3. Using our models to identify key protein-protein interactions, additional site-directed mutagenesis experiments were designed and will be carried out to determine the details of how these proteins form a complex.



2.1.4 Computational development of carboxylesterase inhibitors

We have collaborated with Dr. Craig Wheelock and Prof. Bruce Hammock at UC Davis to identify chemical properties that correlate well with a compound's ability to inhibit carboxylase enzymes. Carboxylesterases are important enzymes responsible for the hydrolysis and metabolism of numerous pharmaceuticals and xenobiotics. These enzymes are potently inhibited by trifluoromethylketone-containing inhibitors. These ketones are known to undergo hydration to form *gem*-diols as shown in the following figure:

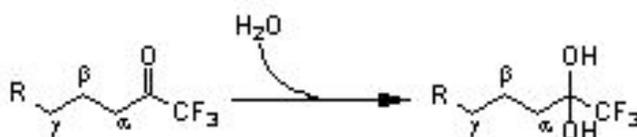


Figure 23. Hydration reaction of trifluoromethyl ketone from the ketone to the *gem*-diol. We found that the energy of this reaction was a good predictor of the inhibitory potency of the compound.

Ab initio quantum chemical calculations were performed on a series of chemically analogous ketones to determine the energy of hydration of the ketone and these values were correlated with esterase inhibition data for a series of carboxylesterase inhibitors.

The ketone compounds that were synthesized were experimentally tested for their ability to inhibit esterase activity in several different contexts. We used three different mammalian enzyme systems: human liver microsomes, murine liver microsomes, and commercial porcine liver esterase, as well as an insect enzyme, juvenile hormone esterase. The regression results are shown in Table 5. We demonstrated that the ketone hydration state was affected by the surrounding chemical moieties and was related to inhibitor potency, with inhibitors that favored the *gem*-diol conformation exhibiting greater potency. In all cases, the energy of ketone hydration was strongly correlated with biological potency. Our results showed a very strong correlation with the hydration extent accounting for 94% of activity for human liver microsomal esterase inhibition ($p < 0.01$). These results were published in the *Journal of Medicinal Chemistry*.

Table 5. Summary regression analysis of quantum chemical hydration energies against the experimentally measured inhibitory potencies in four different enzyme systems. For the mammalian enzyme systems six compounds were tested; for the juvenile hormone esterase, eleven compounds were tested. R^2 is the least squares regression coefficient; ΔE_{gas}^d : electronic energy values in the gas phase (kcal/mole). ΔG_{gas}^e : free energy values in the gas phase (kcal/mole). ΔE_{aq}^f : electronic energy values in the aqueous phase (kcal/mole). ΔG_{aq}^g : Free energy of hydration values as performed by the conductor-like screening solvation model (COSMO) method.

	Equation	R ²
Human Carboxylesterase^c		
ΔE_{gas}^d	y=0.39x-0.23	0.94
ΔG_{gas}^e	y=0.40x-6.27	0.92
ΔE_{aq}^f	y=0.36x-1.48	0.92
$\Delta G_{\text{aq}}(\text{COSMO})^g$	y=0.38x-7.21	0.89
Murine Carboxylesterase^c		
ΔE_{gas}^d	y=0.43x+1.28	0.90
ΔG_{gas}^e	y=0.45x-5.53	0.88
ΔE_{aq}^f	y=0.41x-0.14	0.87
$\Delta G_{\text{aq}}(\text{COSMO})^g$	y=0.42x-6.59	0.85
Porcine Carboxylesterase^c		
ΔE_{gas}^d	y=0.42x+0.78	0.84
ΔG_{gas}^e	y=0.42x-5.74	0.84
ΔE_{aq}^f	y=0.39x-0.57	0.81
$\Delta G_{\text{aq}}(\text{COSMO})^g$	y=0.41x-6.75	0.81
Juvenile Hormone Esteraseⁱ		
ΔE_{gas}^d	y=0.34x-0.43	0.82
ΔG_{gas}^e	y=0.37x-5.71	0.85
ΔE_{aq}^f	y=0.39x-0.48	0.87
$\Delta G_{\text{aq}}(\text{COSMO})^g$	y=0.42x-6.52	0.87

2.1.5 Estrogen Receptor Modeling, Ligand Docking, and Molecular Dynamics Simulations

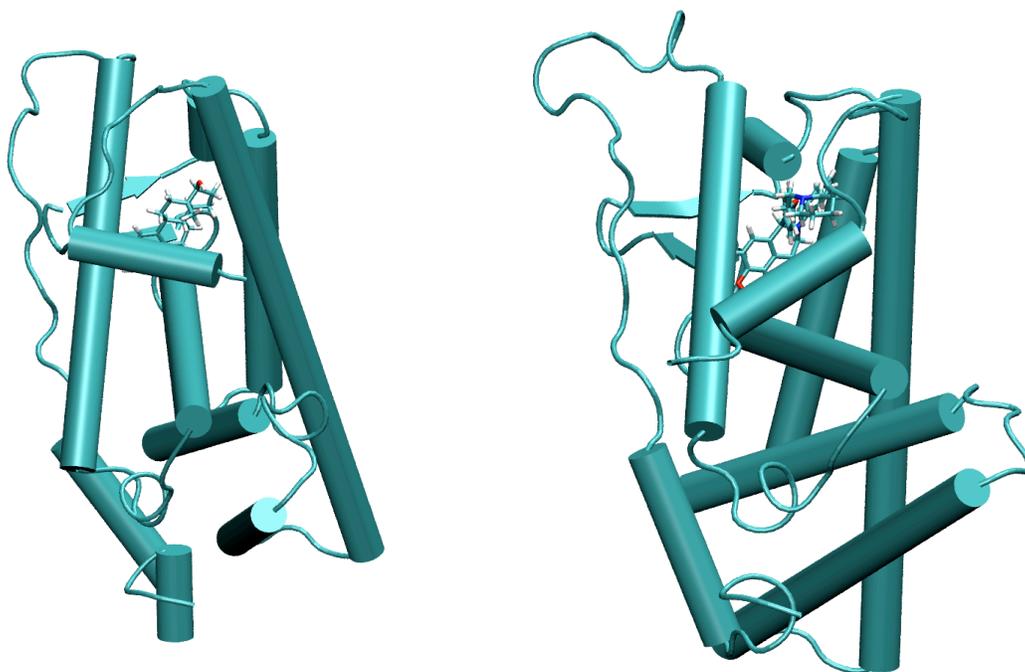


Figure 24 a) GWR

Figure 24 b) IUOM

Estrogen receptors (ER) are proteins that are a part of a superfamily of nuclear transcription factors that are important in growth, development and disease. These receptors are composed of several structurally independent but functionally dependent domains that; bind the hormone (C-terminal ligand-binding domain, LBD), bind DNA (DNA-binding domain, DBD), and a N-terminal activation domain (AP). Two major types of ER have been observed in mammals, Era and Erb. The AP and DBD domains are composed of nearly identical amino acid residues, however the LBD of each receptor are significantly different. Recently it has been shown by others that these two ER subtypes may oppose each other in their mechanism of action *in vivo*. Both receptors bind the hormone estradiol as well as a numerous other chemicals found in food (phytoestrogens) and the environment. These other chemicals are classified as xenoestrogens because of their ability to stimulate estrogen receptor mediated processes inside the cell, i.e. cell proliferation in breast tumors. Xenoestrogens are structurally similar to estradiol, they are generally narrow, planar, and aromatic that are about 10Å long with a hydrogen bond donor/receptor moiety on one end (e.g. a hydroxyl group) of the molecule. Because these chemicals have the potential for disrupting normal estrogen signaling, a significant effort is underway to study and predict their effects on humans and other animals.

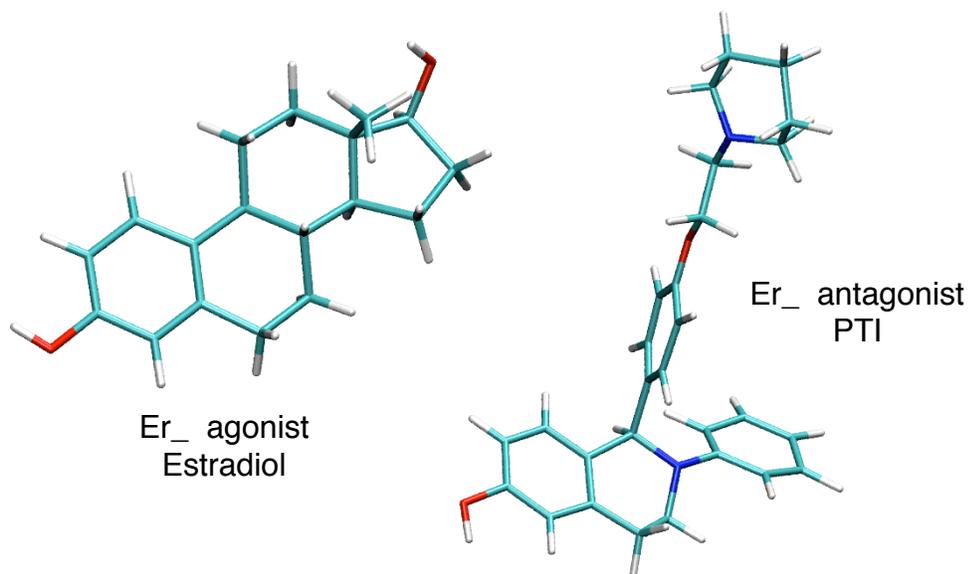


Figure 25. *Estradiol and PTI.*

The Era receptor has been targeted in the treatment of breast cancers that are estradiol-sensitive. Normally estradiol would bind to the receptor and cause a conformational change in the protein. This change promotes dimerization of the receptor which can then interact with other transcription factors and eventually bind to DNA and activate gene transcription. Tamoxifen has been shown to be a potent antiestrogen that competes with estradiol for the ligand binding site in the receptor and after binding promotes dimerization but prevents interaction with the other required transcription activators and transcription is prevented. Several crystal structures of Era with Tamoxifen other antiestrogens bound in the estradiol binding pocket, show a major change in the protein structure when compared to estradiol bound to Era. Specifically, the peptide chain containing the last helix (H12) and loop of the ligand binding domain is displaced and reoriented by 10Å. Amino acid residues within H12 do not make contact with bound estradiol in the native binding conformation; however, when potent antiestrogens are bound, several residues in the helix and loop make polar and hydrophobic contacts with the ligand. Moreover, the displaced H12 is blocking the binding site of a co-activator protein and therefore is the true cause of the loss in ER mediated transcription when Tamoxifen and like antiestrogens are present.

There are few detailed studies on the conformational dynamics of Era, even though significant changes in conformation have been proposed for many years. Recently, studies using EPR spin labels have been published and do suggest that dynamics of the protein can be modulated by ligand binding. However, MD simulations of Era of sufficient length to catch significant backbone conformational changes have not been reported. Consequently, we started our studies on two x-ray crystallographic determined structures of the Era protein. The first structure contains estradiol bound in the appropriate cavity of Era (Figure 24, PDB ID 1gwr). The second structure has an antagonist bound in the estradiol cavity (Figure 24, PDB ID 1uom). It was proposed that the different conformations of the same protein would converge over time in MD simulations.

To complete the docking and MD studies the structures needed to be 'cleaned up'. Small sections of residues were missing in both structures which corresponded mainly to loop regions too flexible to be seen the x-ray experiments. One of these missing sections in 1uom was the loop preceding H12. Comparative modeling tools developed here at the Lab were used to create complete sequences and 3D structures of the ER proteins.

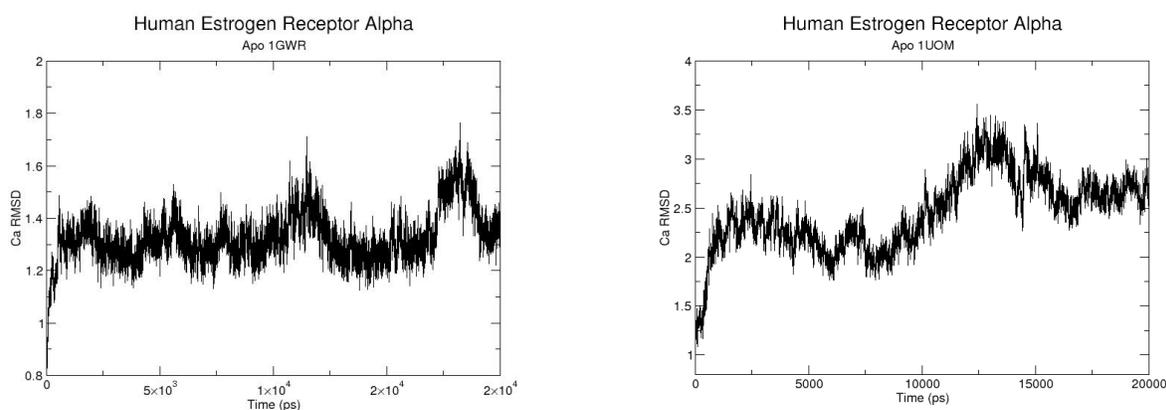


Figure 26. RMS deviation for ERa from native substrate. **Figure 27.** RMS deviation for ERa from antagonist structure.

Once the two Era structures were prepared (no ligands were present), MD simulations at 298 K were performed using the software package NAMD (University of Illinois Champagne-Urbana). This massively parallelizable code was compiled and installed at the Lab on LCs MCR, a 1024 node Linux cluster. A total of 50 nanoseconds of simulations were performed.

Analysis of the two trajectories showed that the 1GWR structure was very stable for the entire 20ns (see Figure 26). Small, periodic fluctuations were also observed for defined areas in the 3D structure. Fluctuations in the structure may be a part of the receptors resting state and that

estradiol binding is dependent on these periodic changes in conformation. The analysis of 1UOM gave different results. The large C α RMSD at the end of the 20 ns simulation was a consequence of binding cavity opening up completely to the solvent (see Figure 27). This may be a result of the modeled H12 loop and further simulations on the unliganded protein are planned.

To model ligand-protein interactions, estradiol and TETRAHYDRO-ISOQUINOLIN (PTI) were docked (Autodock, The Scripps Research Institute) into the binding cavities of the 1GWR and 1UOM structures respectively. The docking procedure recovered the crystallographic binding conformation for both ligands. The ligands were then parameterized for MD simulations of bound Era. These MD simulations are now in progress.

This work has been used in the submission of a DOD Concept Award and will be used in a grant submission to the NIH.

2.1.6 Dynamics of Protamine based peptides:

Protamine molecules bind to and condense DNA in the sperm of most vertebrates, packaging the sperm genome in an inactive state until it can be reactivated following fertilization. Molecular dynamics runs were performed on the R₆GGR₆ and R₆FGR₆ peptides to examine how the presence of the aromatic F residue located between two anchoring domains might affect the conformation adopted by the adjacent peptide segments and ultimately impact the binding of these sequences to DNA. Clustering analysis performed on 200 structures selected at 10 ps intervals during the last 2.0 ns of the simulations are shown in Figures 28A,B. Comparisons of these structures suggest that both peptides form a significant amount of structure along the backbone of the entire sequence. In the simulation performed with the R₆GGR₆ sequence, this backbone remains reasonably extended as a result of the extension and solvation of the R sidechains. The mobility of the carboxy-terminal region of this peptide is constrained by the formation of a salt bridge between the terminal carboxyl group and the sidechain of residue R11. The most interesting structure, however, is that developed in the R₆FGR₆ peptide during the last 1.1 ns of the simulation. In this peptide, the phenyl ring stacks against the peptide backbone and appears to form a stable core that is surprisingly tight (Figure 28B). In these structures, the aromatic ring of F interacts with the backbone amides of G8 and R9. One consequence of this stacking is that the backbone is bent between the two anchoring (R₆) domains.

The distances measured from the amide hydrogen to the center of mass of the F ring during the last nanosecond of the simulation are 3.88 ± 0.63 and 3.52 ± 0.42 Å for G8 and R9, respectively. One effect of the interaction between the aromatic ring and backbone is that the amount of motion sampled by the ring is reduced significantly. The positional fluctuation of the F side-chain is 4.54 Å for the first nanosecond of dynamics but reduces to 2.72 Å when stacked against the peptide backbone. The average solvent accessible surface area of R₆FGR₆ (2766.5 ± 87.4 Å²) is lower than for R₆GGR₆ (2850.4 ± 79.4 Å²) even though R₆FGR₆ has the greater intrinsic surface area. The results of this study were published in *Journal Biological Chemistry*.

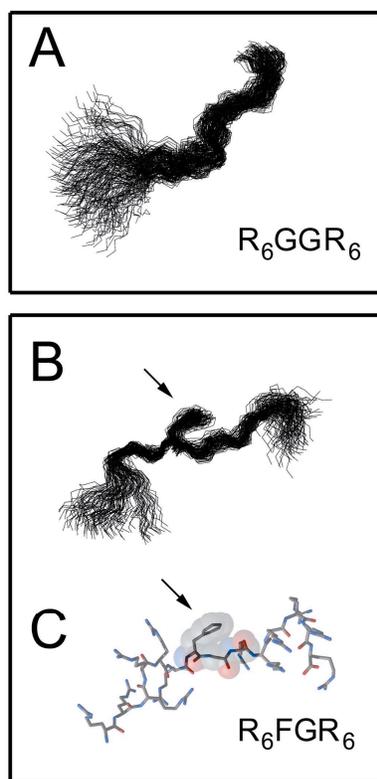


Figure. 28A The superposition of the backbone atoms of R₆GGR₆ from molecular dynamics calculations. Only the N, CA, and C atoms of the peptide are shown. A total of 200 structures, taken from 10 ps intervals, are shown. **28B** Overlay of the backbone atoms for R₆FGR₆ from the last 1.16 ns of molecular dynamics calculations (116 structures). Only the N, CA, and C atoms of the peptide are shown for clarity. **28C** The last of the 116 structures shown in **28B** and **28C** point to the aromatic ring of Phe7 that stacks against the backbone amides of G8 and R9.

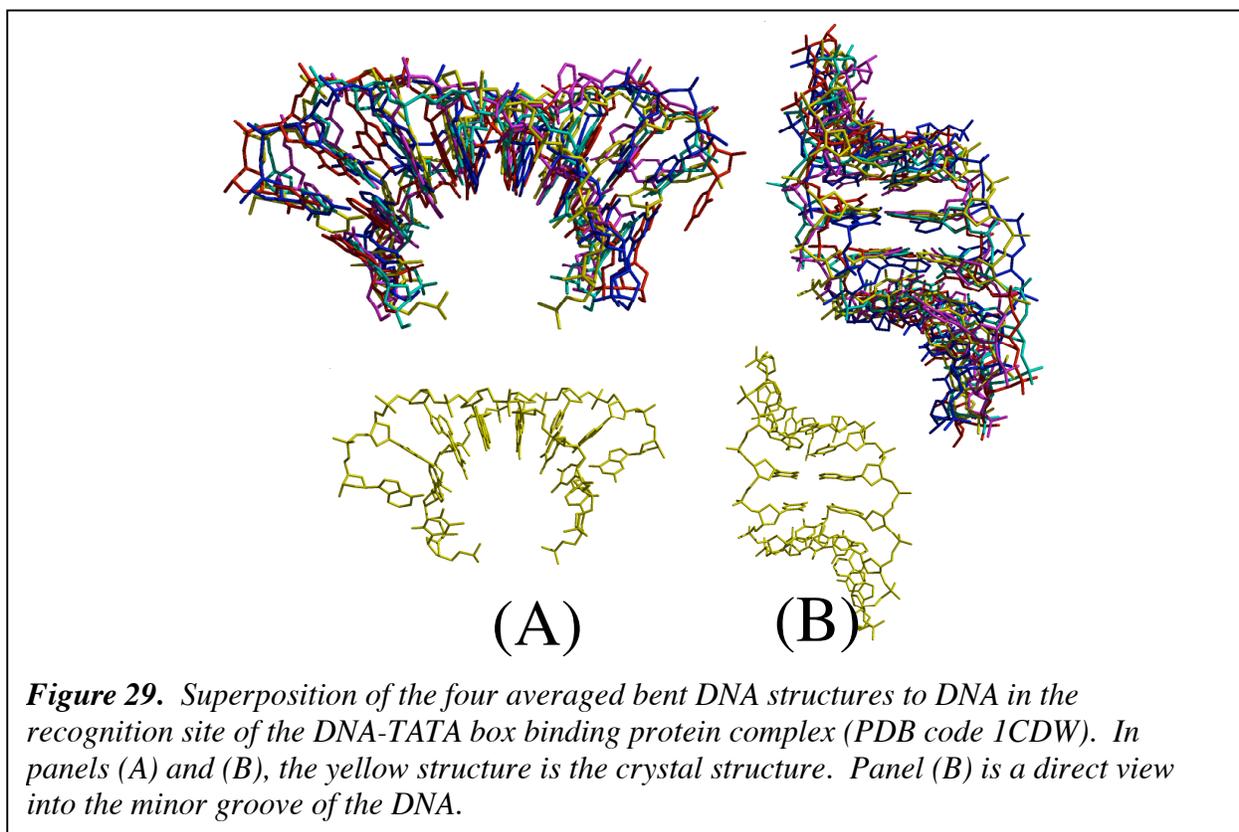
2.2 Predicted Interactions of Biological Materials with Nanostructures

We have used molecular dynamics to simulate the structure and counter-ion interactions of double stranded DNA confined inside a single-walled carbon nanotube (NT). We studied five systems: an 8-bp DNA oligomer inside 30 Å and 40 Å diameter nanotubes with the counter ions either inside or outside the nanotube, as well as the unenclosed DNA as a control. Each of these systems was simulated four times with different initial velocity distributions and the system with DNA in a 30 Å NT with the ions on the outside was simulated for four additional runs.

The simulations of the DNA-NT systems with the ions excluded from the NT yielded the most significant perturbations to the DNA structure. Interestingly, in 4 out of 8 of the simulations of the DNA in the 30 Å NT with the ions excluded the DNA adopted a characteristic bend that were very nearly superimposeable (The DNA had bend angles of 84° to 92°). The DNA bases remain paired even though the overall structure is severely bent. In the solvated DNA simulations the DNA bends from 17° to 26°. The characteristics of the DNA bending were consistent in the simulations. The minor groove of the DNA widens to expose the bases, the major groove narrows, and DNA bends towards the major groove. The bending of the DNA is very rapid. In each simulation where bending occurs there is an initial transition from B-form to A-form DNA followed by the structures bending by 80° before 300 ps and reaching its maximum by 500 ps.

This result is intriguing due to previous experimental and theoretical evidence for structurally similar bending of DNA near hydrophobic surfaces. Crystallographic structures of DNA bound to the TATA box binding protein (TBP) that acts to initiate eukaryotic RNA polymerization, showed that the DNA had an 80° bend in the seven bases within the TBP recognition site and also underwent a 117° unwinding in these seven bases. Figure 29 shows a superposition of the four bent structures observed for the DNA in the 30 Å NT with the ions excluded on the crystallographic structure of the TBP-bound DNA.

This work has been submitted as a communication to the *Journal of the American Chemical Society*.



2.3 Advanced Computational Chemical Methods

2.3.1 Applications of First Principles Molecular Dynamics

During this Strategic Initiative LDRD Project, we performed first principles molecular dynamics (FPMD) simulations of liquid water, extending these to higher pressures and temperatures. Additionally, we performed FPMD simulations of solvated calcium ion.

2.3.1.1 Liquid water simulations

Water is an important liquid in many different areas of science. In addition to its ubiquitous, life-supporting presence on earth, water happens to be one of the most anomalous liquids known. As such, an enormous amount of effort by both experimentalists and theoreticians has been devoted to unraveling its properties. However, many questions about water have so far remained unanswered, and it is only very recently that different experimental approaches for probing the structure of water have come into agreement on, e.g., the oxygen-oxygen radial distribution function.

Important theoretical improvements in the description of water have occurred over the past thirty years. Molecular dynamics simulations have increased in complexity from simple pair-wise interaction potentials to fully quantum mechanical descriptions. In particular, the Car-Parrinello (CP) method has enabled simulations of liquid water entirely from first-principles. These simulations, which treat electronic degrees of freedom quantum mechanically within density functional theory (DFT) and ionic motion classically, have allowed for important insight into the properties of water at a variety of temperature, pressure and density conditions. CP techniques combined with significant increases in available computer power have pushed the computational simulation of water to an exciting stage. Quantum simulations can now be used both to make genuine predictions directly comparable with experiment as well as to establish the accuracy of newly developed classical potentials, which are critical for accessing much longer time scales and much larger systems. In this light, it is of great importance to understand in detail the level of accuracy and predictive power of first principles molecular dynamics (FPMD) simulations. In order to do so, it is useful to make a broad range of comparisons between simulations carried out using different theoretical and technical approximations (e.g., density functionals, total simulation time), by systematically testing one specific approximation at a time.

We have carried out a series of ten FPMD simulations of water under ambient conditions that are aimed at accessing the accuracy of density functional theory in describing the dynamical and structural properties of water. Our simulations have addressed a number of issues that include simulation timescales, density functionals, size effects, and computational parameters that are specific to the CP method. We have found negligible differences in structural properties obtained using different density functionals that employ the generalized gradient approximation and also have found that, although not completely negligible, size effects for systems larger than 32 water molecules are rather small.

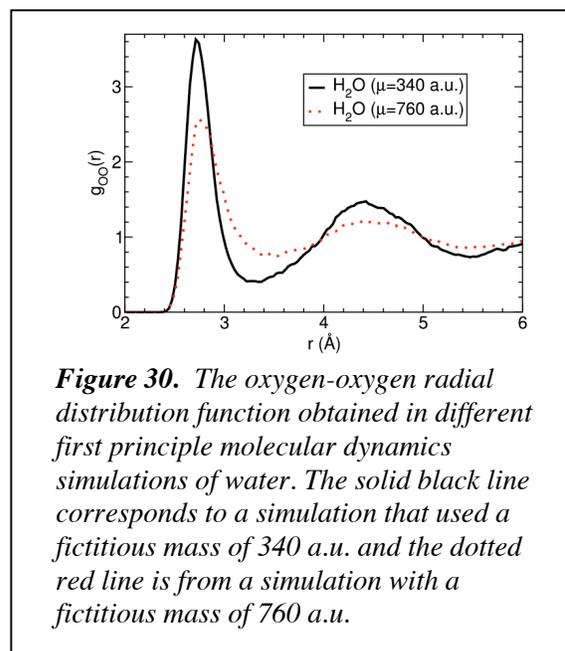


Figure 30. The oxygen-oxygen radial distribution function obtained in different first principle molecular dynamics simulations of water. The solid black line corresponds to a simulation that used a fictitious mass of 340 a.u. and the dotted red line is from a simulation with a fictitious mass of 760 a.u.

In addition, we have found a wide range of values for the fictitious mass parameter that is used in the CP Lagrangian for which the electronic ground state is accurately described. However, as shown in Figure 30, large values of the fictitious mass can lead to substantial differences in the structural description of liquid water. A paper describing this work has been published in the *Journal of Chemical Physics* **120**, 300 (2004).

Since relatively small values of the fictitious mass parameter must be used in CP simulations of water, this in turn necessitates the use of small integration time steps used the molecular dynamics algorithm. The reason for this restriction is due to the high frequency vibrational modes that are characteristic of water i.e. O-H stretch (3200 to 3600 cm^{-1}) and H-O-H bending modes ($\sim 1600 \text{ cm}^{-1}$), which can lead to a coupling between the ionic and electronic degrees of freedom when large values of the fictitious mass are used. In fact, the time step may need to be as small as 0.08 fs, which is approximately ten times smaller than what is often used in classical MD simulations of water with empirical inter-atomic potentials. This poses a severe restriction on the time scales that can be accessed in CP simulations of water. In order to address this time scale limitation we have investigated the use of a rigid water model within a FPMD context. By completely removing the high frequency vibrations, a rigid water model allows for much larger values of the fictitious mass, as well as time steps in the CP method. Our results indicate that the rigid water model enables the use of time steps as large as 0.24 fs within CP simulations of water. This represents an important advantage for first-principle simulations of aqueous solutions where chemical reactions do not occur, and opens up the possibility of investigating phenomena that occur on a long timescale. A paper describing this work has been accepted for publication in the *Journal of Chemical Physics*.

2.3.1.2 Ion Solvation

The solvation of ions is a fundamental process that is found in a wide range of biological and chemical systems. In particular, the manner in which water solvates alkali and alkaline earth cations is relevant to systems such as enzymatic catalysis and the structural stability of DNA and RNA. We have performed FPMD simulations of calcium (II) in water. By comparison with the available experimental data on these ions in solution, we are able to directly evaluate how well solute-solvent interactions are reproduced in the simulations. In addition, comparisons are made

to classical MD simulation models, which clearly demonstrate the advantages of using a highly accurate first principles simulation approach over more traditional simulation models.

We have performed a FPMD simulation of a calcium cation (Ca^{2+}) in liquid water. Because calcium ion has the capacity to coordinate multiple ligands, it is difficult to determine its hydration shell and preferred coordination number experimentally. From various experimental methods, calcium ion has been shown to coordinate any number of water molecules from 6 to 10. In this study, two different types of water molecules were used, fully flexible and rigid water molecules. In all our simulations, we observe that the calcium ion has a flexible coordination shell and is able to coordinate 6, 7, and 8 water molecules in the first solvation shell. For the fully flexible water with calcium ion simulations, we determined an average coordination number of 6.2 while for the rigid water simulations the average coordination number is 7.0. These results are within the experimental values and compare nicely with a recent X-ray absorption fine structure of a 6 m CaCl_2 , having an average coordination number of 7.2 ± 1.2 . Because several coordination states are observed, several exchange processes are also observed for the rigid water simulation.

A careful analysis of the first solvation shell around Ca^{2+} reveals that the water molecules tend to coordinate to Ca^{2+} in an asymmetric orientation. This orientational property is not reproduced by classical MD simulations or in QM calculations of Mg^{2+} /water clusters. The water molecules in the first solvation shell are primarily oriented with the oxygen atoms coordinating the ion and the hydrogen atoms pointing away into the bulk water. The asymmetric orientation is observed with a 39.7° tilt angle.

In addition to the structural properties of the first solvation shell around Ca^{2+} , we have investigated the difference between the electronic structure of the first solvation shell waters and the “bulk” waters (the water molecules outside of the first solvation shell). This analysis was performed by computing maximally localized Wannier functions in a manner similar to the Boys localization procedure that is commonly used in quantum chemistry. The localized Wannier functions can be associated with features such as distributions of lone pairs and covalent O-H bonds and can be used to define local dipole moments for each of the water molecules in the simulation. A probability distribution of the water dipole moments in our FPMD simulation is shown in Figure 31. As can be seen, the dipole moments of the bulk water molecules form a broad distribution from 2 to 4.5 Debye, with an average value of 3.19 Debye. For the waters in

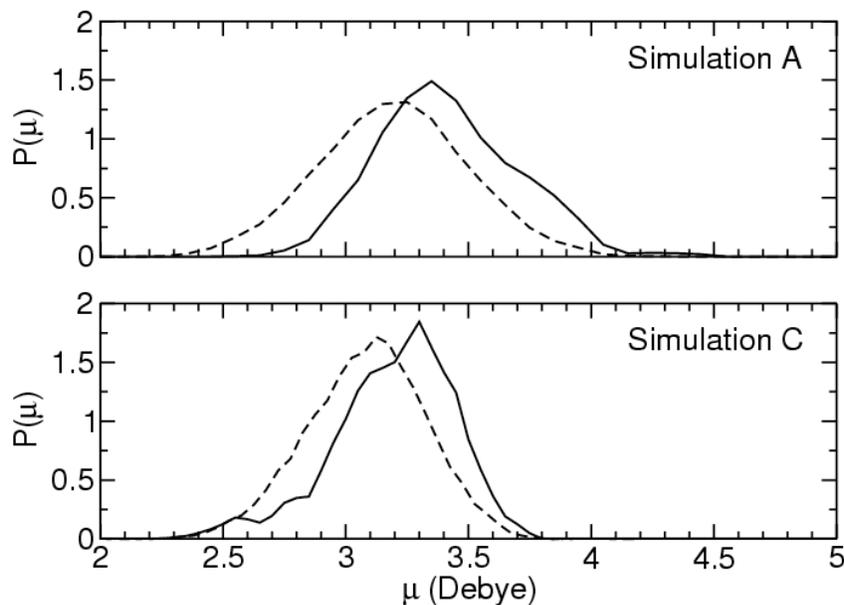


Figure 31. The probability distribution of water dipole moments averaged over the flexible water simulation A (top panel) and the rigid water simulation C (bottom panel). The solid line corresponds to the dipole moment distribution of water molecules within the first solvation shell around calcium and the dashed line to the water molecules outside of the first solvation shell.

the first solvation shell around Ca^{2+} , the distribution is shifted 0.2 Debye to an average value of 3.39 Debye. These subtle properties of Ca^{2+} may prove to be essential in the accurate simulation of chemical reactions where Ca^{2+} is known to assist in enzymatic reactions

2.3.1.3 Reactions of hydroxyl radical with guanine:

In collaboration with Drs. Chris Mundy and Andrew Quong in the Chemistry and Material Science directorate, we studied the reactions of the hydroxyl radical with guanine. Hydroxide radical is one of the primary cellular products of ionizing radiation; therefore the damage that this reactive species can induce in DNA is a likely cause of radiation-induced DNA damage experimentally observed. In this study we used both Car-Parrinello molecular dynamics and *ab initio* quantum chemical methods were used to investigate possible reactions induced by hydroxyl radicals. In this paper we focused on reactions with guanine, the DNA base known to be most susceptible to damage due from radiation-induced radicals. Figure 32 below shows the results for possible reactions leading to loss of a hydrogen atom from the guanine N2 position (one of the primary sites of guanine damage) in the gas-phase ($\square\text{G}_{\text{gas}}$) and aqueous phase ($\square\text{G}_{\text{aq}}$). This data shows that the over all hydrogen abstraction (reaction 1), can efficiently occur by a two

step process (reactions 3 and 4) in which the hydroxyl radical first abstracts an electron from the guanine and subsequently removes a proton from the N2 position. (Note that the large difference in the gas- and aqueous-phase energies, show the importance of including aqueous solvation effects in modeling reactions involving ionic intermediates.) These results were published in the *Journal of Physical Chemistry*.

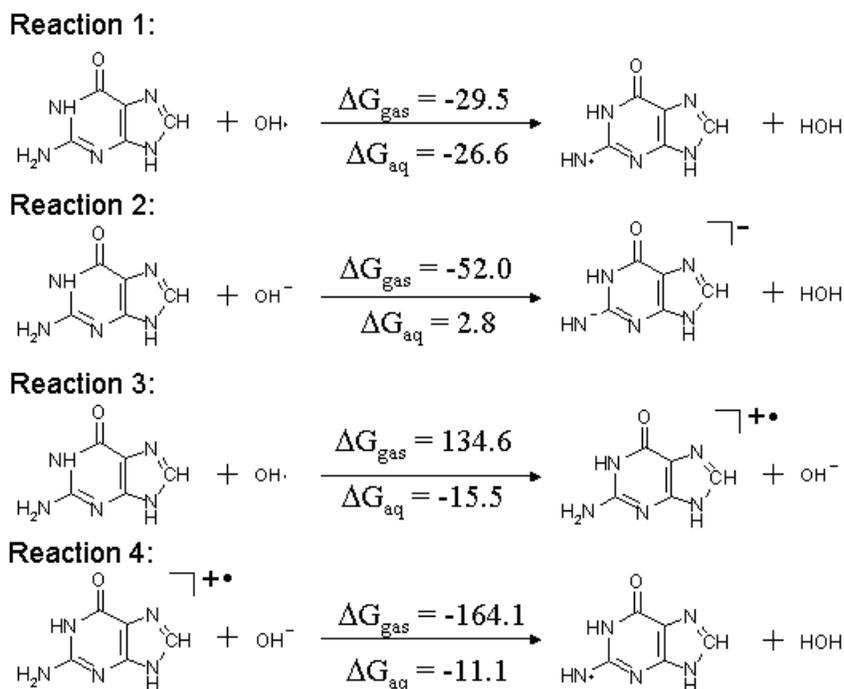


Figure 32. Gas- and aqueous-phase reaction energies of guanine and hydroxide radicals in kcal/mole. Energies calculated at the B3LYP/6-31++G(3df,3pd)//B3LYP/6-31++G(3df,3pd) level of theory; solvation effects included using the COSMO solvation model.

2.3.2.4 Reactions of hydroxyl radical with thymine:

In collaboration with Drs. Chris Mundy in the Chemistry and Material Science directorate and Dr. Yu Dong Wu and Roberto Car of Princeton University, we studied the reactions of the hydroxyl radical with thymine. As with the study described above, both Car-Parrinello molecular dynamics and *ab initio* quantum chemical methods were used to study the hydroxylation and dehydrogenation reactions induced by hydroxyl radical. Figures 33 and 34 below summarize the predicted reaction energies for both the dehydrogenation and hydroxylation reactions. Since none of these reactions involve an ionic intermediate, the gas- and aqueous phase energetics are in fairly good agreement. This data indicates that

dehydrogenation is favored at two sites on thymine, the exocyclic methyl group and one of the endocyclic amines (N1) and that hydroxylation is favored at the two non-carbonyl endocyclic carbons (C5 and C6). These results have been accepted for publication in the *Journal of Physical Chemistry*.

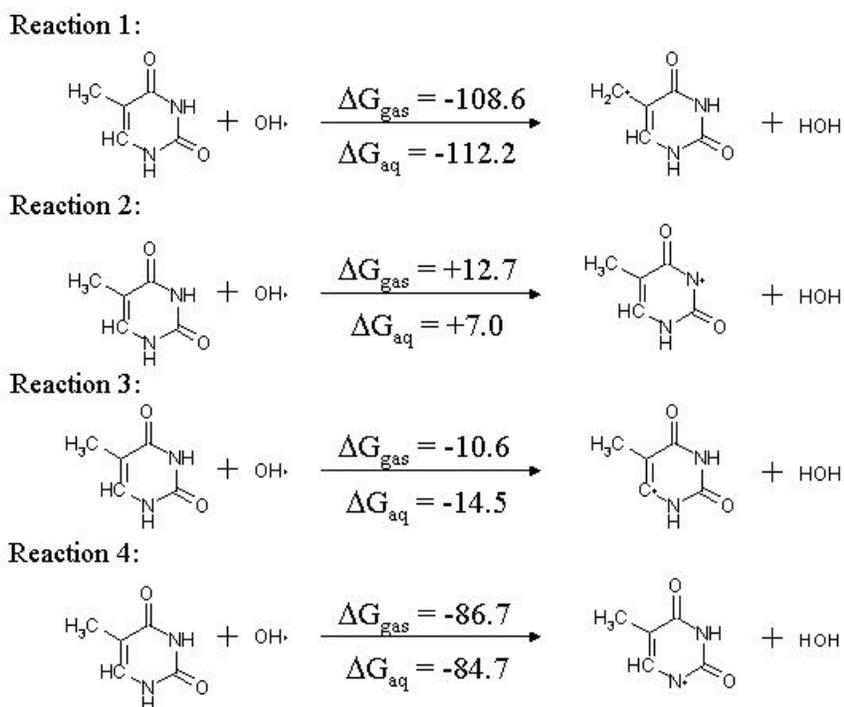


Figure 33. Gas- and aqueous-phase reaction energies for dehydrogenation of thymine by hydroxide radicals in kJ/mole. Energies calculated at the B3LYP/6-31G**//B3LYP/6-31G** level of theory; solvation effects included using the COSMO solvation model.

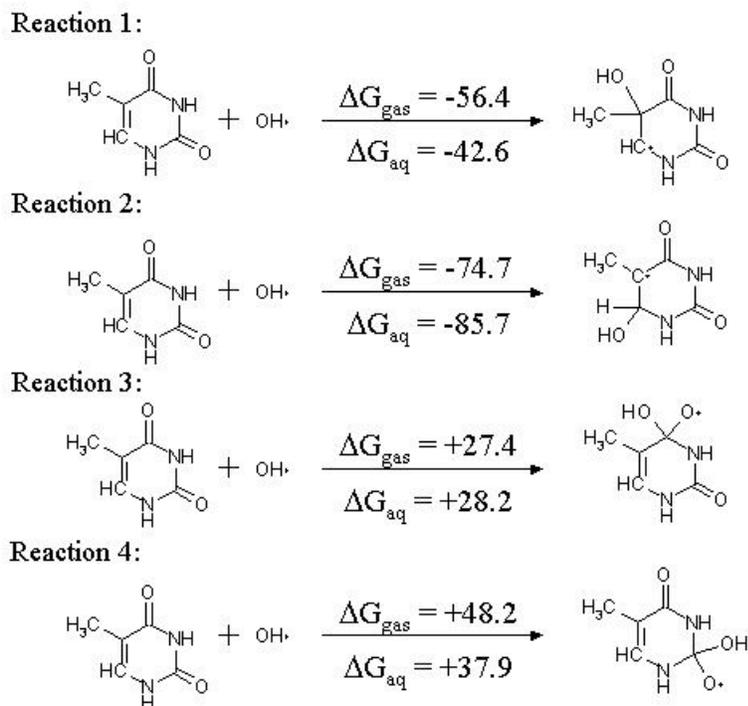


Figure 34. Gas- and aqueous-phase hydroxylation energies of thymine and hydroxide radicals in kJ/mole. Energies calculated at the B3LYP/6-31G**//B3LYP/6-31G** level of theory; solvation effects included using the COSMO solvation model.

2.3.2 Algorithm Development of First Principles Molecular Dynamics

In order to examine various chemical reactions within a molecular dynamics simulation it is often convenient to force the reaction of interest to occur with the use of bond length constraints. Although simple bond length constraints have allowed us to examine a wide range of chemical reactions, in many cases the reaction coordinates of the system are complex and can not be effectively examined with a single constraint. One approach for mapping out complex reaction paths involves the use of multiple holonomic distance constraints. We have implemented the use of multiple bond distance constraints in the GP FPMD simulation code. This new computational capability is relevant to many ongoing projects at the lab. For example, this new feature was recently used to examine various functionalization reactions on the surface of a silicon quantum dot.

Software Developed:

VACF: Computes the velocity autocorrelation function from a molecular dynamics trajectory with a multi-windowing technique to improve the statistical sampling. (E. Schwegler)

Spectrum: Computes the power spectrum from the output of VACF, by direct Fourier transform or by the maximum entropy method. (E. Schwegler)

SDF: Computes and visualizes the spatial distribution function around a given solute in a molecular dynamics simulation. (E. Schwegler)

Lifetime: Determines the lifetime of proton transfer events in liquid water simulations, including the possibility of “proton shuttling” mechanisms. (E. Schwegler)

RingStats: Counts the number and size distribution of non-short-circuited hydrogen bonded rings of water molecules around a solute. (E. Schwegler)

SMOR: Computes the single molecule orientational relaxation function from a molecular dynamics trajectory. (E. Schwegler)

2.4 Protein Structure Prediction

2.4.1 Development of methods for homology-based protein modeling

The basic concept of homology-based protein structure prediction relies on the observation that structural features of proteins are conserved during evolution to a much higher degree than their sequences, and therefore proteins related even by distant sequence similarity can be expected to have similar 3D structures. Thus once a 3D structure is determined for at least one representative of a protein family, models for the other family members can be derived using the known structure as a template.

During the report period we continued the development of a comparative modeling method focusing specifically on low level homology regime, where the correct sequence-structure alignment cannot be obtained trivially. The approach is based on the idea of the “consensus prediction”. More specifically, multiple models for the same protein molecule are obtained based on alternative sequence-structure alignments and the best three-dimensional model is then selected from these models using consensus evaluation results. The method with subsequent improvements has been rigorously tested during international “blind” mode protein structure prediction experiments (CASPs). During CASP4 we tested the method together with over 80 other groups and during the latest CASP round (CASP5, in 2002) with more than 100

research groups from all over the world. In both experiments the results obtained with this method were exceptionally good. More specifically, the assessment by independent experts ranked the results obtained with this modeling method as the best in CASP4. In CASP5 the refined method again produced results that together with those produced by two other methods stood out in the homology-based protein structure prediction category. The top performance of the method during both CASP4 and CASP5 resulted in the invited talks during the corresponding international CASP conferences as well as invited publications in the special issues of “Proteins” devoted entirely to CASP experiments.

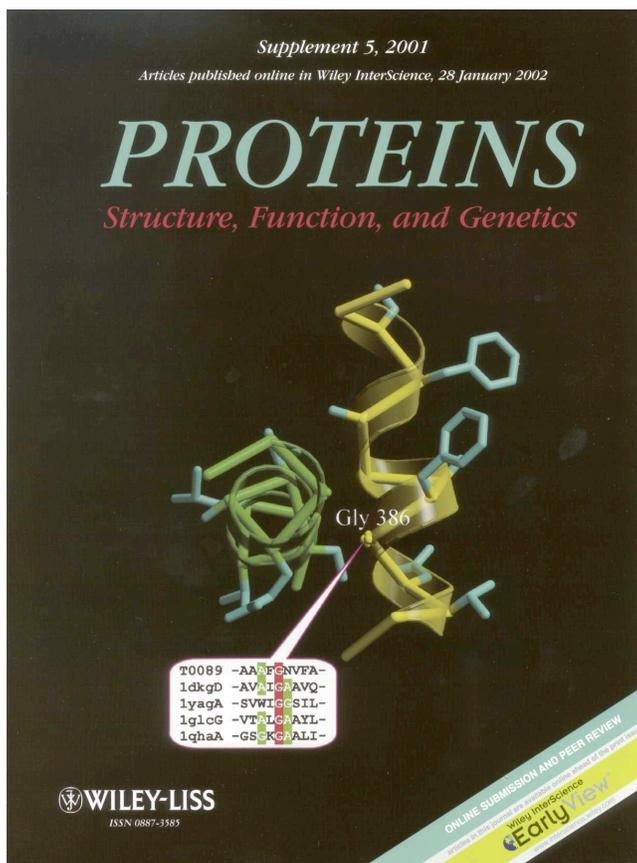


Figure 35. Some of our prediction results were featured as a cover story in the special issue of “Proteins”.

2.4.2 Application of Molecular Modeling

In many cases the structural conservation is characteristic not only of individual proteins, but also of the higher-order structure of protein complexes. We have shown that this evolutionary feature can be successfully exploited to characterize protein complexes with the help of molecular modeling.

Prediction of the structure, intermolecular interactions and the functional mechanism for the Replication Factor C, a clamp loading protein “machine”.

We have extended comparative modeling-based approach to structurally characterize fairly large protein complex, Replication Factor C (RFC), composed of five different proteins. This modeling study provided a number of novel findings, including detailed models of RFC interaction with its molecular partner – the PCNA complex, and topographical model for how the modified version of RFC interacts with the “9-1-1” complex. This study has provided a structural framework for interpretation of existing experimental data and for many new experimental tests.

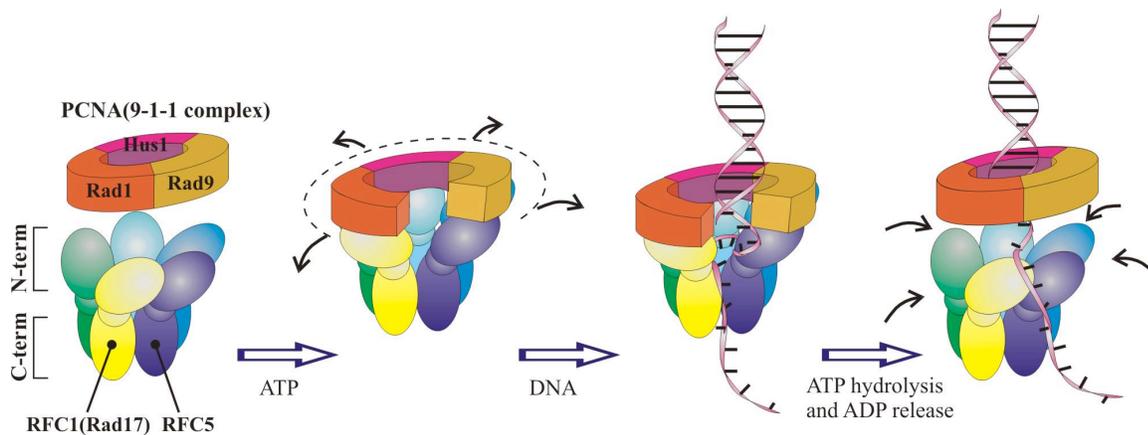


Figure 36. Results of modeling-based study of protein-protein interactions: the proposed molecular structure and functional steps of the clamp loader/clamp complexes.

The demonstrated effectiveness of modeling-based studies of protein complexes has been used as a basis to obtain funding for further exploration of the modeling capabilities in this important research area. Specifically, modeling of protein complexes is a central theme in the funded Lab wide LDRD proposal (Venclovas, PI).

3 Appendix

3.1 Publications — In Print, In Press or Submitted

1. Hatch, F.T., M.G. Knize, M.E. Colvin (2001) Extended QSAR for 80 Aromatic and Heterocyclic Amines: Structural, Electronic and Hydrophobic Factors Affecting Mutagenic Potency *Environmental and Molecular Mutagenesis* **38**: 268-291.
2. Felice Lightstone, Eric Schwegler, Giulia Galli, and François Gygi “A first principles molecular dynamics simulation of the hydrated magnesium ion”, *Chemical Physics Letters*, **343**, 549 (2001)
3. Long, S. A., C. Quan, J. van de Water, M. H. Nantz, M. J. Kurth, D. Barsky, M. E. Colvin, K. S. Lam, R L. Coppel, A. Ansari, and M. E. Gershwin (2001) Synthetic Organic/Peptide Conjugate Mimetopes of the Major Autoantigen of Primary Biliary Cirrhosis. *J. Immunol.* **167**: 2956–2963.
4. Eric Schwegler, Giulia Galli, and François Gygi “Conformational dynamics of the dimethyl phosphate anion in solution”, *Chemical Physics Letters*, **342**, 434 (2001)
5. Eric Schwegler, Giulia Galli, François Gygi and Randolph Hood “Dissociation of water under pressure”, *Physical Review Letters*, **87** 265501 (2001)
6. Wilson, III, D.M. and D. Barsky (2001) The Major Human Abasic Endonuclease Ape1. Formation, Consequences, and Repair of Abasic Lesions in DNA. [Review] *Mutation Res.* **485**: 283–307.
7. Venclovas, C. (2001) Comparative modeling of CASP4 target proteins: combining results of sequence search with three-dimensional structure assessment. *Proteins*, Suppl. **5**: 47-54.
8. Venclovas, C., Zemla, A., Fidelis, K. and Moulton, J. (2001) Comparison of performance in successive CASP experiments. *Proteins*, Suppl. **5**: 163-170.
9. Zemla, A., Venclovas, C., Moulton, J. and Fidelis, K. (2001) Processing and evaluation of predictions in CASP4. *Proteins*, Suppl. **5**: 13-21.
10. Tran, N.L., C.P. Salmon, M.G. Knize, M.E. Colvin (2002) Experimental and Simulation Studies of Heat Flow and Heterocyclic Amine Mutagen/Carcinogen Formation in Pan-Fried Meat Patties. *Food and Chemical Toxicology* **40**: 673-684.
11. Balu, N., M.P. Gamcsik, M.E. Colvin, O.M. Colvin, M.E. Dolan, S.M. Ludeman (2002) Modified Guanines Representing O⁶-Alkylation by the Cyclophosphamide Metabolites Acrolein and Chloroacetaldehyde: Synthesis, Stability, and *Ab initio* Studies *Chemical Research in Toxicology* **15**: 380-387.
12. M. C. Cosman, F. C. Lightstone, V. V. Krishnan, L. Zeller, M. C. Prieto, D. C. Roe, and R. Balhorn, (2002) Identification of Novel Small Molecules That Bind to Two Different Sites on the Surface of Tetanus Toxin C Fragment. *Chem. Res. Toxicol.*, **15**, 1218.
13. B. Lee, L. H. Nguyen, D. Barsky, M. Fernandes, and D. M. Wilson III (2002) Molecular interactions of human Exo1 with DNA *Nucleic Acids Res.* **30**(4): 942–949.

14. Leininger, M.L., I.M.B. Nielsen, M.E. Colvin, C.L. Janssen (2002) Complete Basis Set MP2 Binding Energies for Stacked Uracil Dimers *Journal of Physical Chemistry A* **106**: 3850-3854.
15. Grossman, J.C., M.E. Colvin, N.L. Tran, S.G. Louie, M.L. Cohen (2002) Aromaticity and Hydrogenation Patterns in Highly Strained Fullerenes. *Chemical Physics Letters* **356**: 247-253.
16. Nguyen, D.H., M.E. Colvin, Y. Yen, R.E. Feeney, W.H. Fink, (2002) The Dynamics, Structure, and Free Energy Profile of Proline-Containing Antifreeze Glycoprotein *Biophysical Journal* **82**: 2892-2905
17. Sasaki, J.C., R.S. Fellers, M.E. Colvin (2002) Metabolic Oxidation of Carcinogenic Arylamines by P-450 Monooxygenases: Theoretical Support for the One Electron Transfer Mechanism *Mutation Research* **506-507**: 79-89.
18. Eric Schwegler, Giulia Galli, François Gygi and Randolph Hood “Reply: Dissociation of under pressure”, *Physical Review Letters*, **89** 199602 (2002)
19. Venclovas, C., M.E. Colvin, M.P. Thelen, (2002) Molecular modeling-based analysis of interactions in the RFC-dependent clamp-loading process *Protein Science* **11**: 2403-2416.
20. Colvin, M.E., J.N. Quong (2002) DNA-Alkylating Events Associated with Nitrogen Mustard Based Anticancer Drugs and Metabolic Byproduct Acrolein *Advances in DNA Sequence-Specific Agents Vol. 4*; G.B. Jones, ed., Elsevier, New York, 29-46.
21. Mundy, C.J., M.E., Colvin, A.A. Quong, (2002) Irradiated Guanine: A Car-Parrinello molecular dynamics study of dehydrogenation in the presence of an OH radical *Journal Physical Chemistry A* **106**: 10063-10071.
22. Wheelock, C.E., M.E. Colvin, I. Uemura, M.M. Olmstead, J.R. Sanborn, Y. Nakagawa, B.D. Hammock (2002) Use of *ab initio* calculations to predict the biological potency of carboxylesterase inhibitors. *Journal of Medicinal Chemistry* **45**: 5576-5593.
23. Tran, N.L., M. Colvin, S. Gronert, and W. Wu (2003) Catalysis of decarboxylation by an adjacent negative charge: a theoretical approach. *Bioorganic Chemistry* **31**: 271-277.
24. M.D. Winthrop, S.J. DeNardo, H. Albrecht, G.R. Mirick, L.A. Kroger, K.R. Lamborn, C. Venclovas, M.E. Colvin, P.A. Burke, G.L. DeNardo (2003) Selection and Characterization of Anti-MUC-1 scFvs Intended for Targeted Therapy *Clinical Cancer Research* **9**: 3845s-3853s.
25. S. J. Shields, O. Oyeyemi, F. C. Lightstone, and R. Balhorn, (2003) Mass Spectrometry and Non-covalent Protein-Ligand Complexes: Confirmation of Binding Sites and Changes in Tertiary Structure. *J. Am. Soc. Mass Spectrom.*, **14**, 460.
26. J.R. Keefe, S. Gronert, M.E. Colvin, N.L. Tran (2003) Identity Proton Transfer Reactions from C-H, N-H, and O-H Acids. An *ab initio*, DFT, and CPCM-B3LYP Aqueous Solvent Model Study *Journal of the American Chemical Society* **125**: 11730-11745.
27. D.H. Nguyen, T. Dieckmann, M.E. Colvin and W.H. Fink (2003) Dynamics Studies of a Malachite Green-RNA Complex Revealing the Origin of the Red-Shift and Energetic Contributions of Stacking Interactions *Journal of Physical Chemistry B* **108**: 1279-1286.

28. L. Brewer, M. Corzett, E. Y. Lau, and R. Balhorn (2003) Dynamics of Protamine 1 Binding to Single DNA Molecules *J. Biol. Chem.*, **278**: 42403 - 42408.
29. Fernando A. Reboredo, Eric Schwegler, and Giulia Galli “Optically Activated Functionalization Reactions in Si Quantum Dots” *The Journal of the American Chemical Society*, **125**, 15243 (2003)
30. François Gygi, Jean-Luc Fattebert, and Eric Schwegler “Computation of Maximally Localized Wannier Functions Using a Simultaneous Diagonalization Algorithm”, *Computer Physics Communications*, **155**, 1 (2003)
31. Venclovas, C. (2003) Comparative modeling in CASP5: progress is evident, but alignment errors remain a significant hindrance. *Proteins*, Suppl. **6**: 380-388.
32. Markus Allesch, Eric Schwegler, Giulia Galli, and François Gygi “First-principle simulations of rigid water” *Journal of Chemical Physics*, **120** (2004)
33. Jeffrey C. Grossman, Eric Schwegler, Erik W. Draeger, François Gygi, and Giulia Galli “Towards an assessment of the accuracy of density functional theory for first principles simulations of water” *Journal of Chemical Physics*, **120**, 300 (2004)
34. K. A. Miller, D. Sawicka, D. Barsky, and J. S. Alcala. Domain mapping of the Rad51 paralog protein complexes. *Nucleic Acids Res.*, (In Press).
35. Y. Wu, C.J. Mundy, M.E. Colvin and R. Car (2003) On the mechanisms of OH radical induced DNA base damage: A comparative quantum chemical and Car-Parrinello molecular dynamics study *Journal of Physical Chemistry B* (In press)

3.2 Talks and posters (FY01-FY03):

1. D. Barsky, “Molecular Dynamics Simulations of Damaged DNA” Poster Presentation at Gordon Conference on Mammalian DNA Repair, Ventura, CA, January 24, 2001.
2. “Computer Modeling Studies of DNA-binding Food Mutagens and Anticancer Drugs”, Dept. of Nutritional Sciences and Toxicology, UC Berkeley, March 14, 2001, Berkeley, CA.
3. “Computational Biology: Principles and Practice”, Invited lecture, Dept. of Chemistry, San Francisco State University, CA, March 16, 2001.
4. “Conformational dynamics of the dimethyl phosphate anion in solution” International Conference on Computational Nanoscience, March 2001 (invited talk)
5. “Efficient computation of Hartree-Fock exchange matrices” American Chemical Society, San Diego, April 2001 (invited talk)
6. “A first principles molecular dynamics simulation of hydrated magnesium ion” American Chemical Society, San Diego, April 2001 (poster)
7. “Computational Chemical Studies of DNA-Binding Compounds”, Chemistry Department Organic Chemistry Colloquium, UCLA, Los Angeles, CA, May 2, 2001.
8. D. Barsky, “DNA Damage Probed by Molecular Dynamics Simulations“ *Albany Conversations in Biomolecular Stereodynamics*, Albany, NY, June 2001. (Both a talk and poster presented at this meeting.)

9. "Computational Chemical Studies of DNA-Damaging Drugs and Food Mutagens" Keynote speaker: DOE Computational Science Graduate Fellowship Conference, Washington, DC, July 20, 2001.
10. "The dissociation of water under pressure" American Chemical Society, Division of Nuclear Chemistry and Technology, Chicago, August 2001 (invited talk)
11. "Efficient computation of Hartree-Fock exchange matrices" Martin Head-Gordon group seminar, University of California, Berkeley, October 2001 (invited talk)
12. "Computer Simulations of the Activation of Heterocyclic and Aromatic Amine Mutagens", Eighth International Conference on Carcinogenic/Mutagenic N-Substituted Aryl Compounds, Washington DC, November 12, 2001.
13. Venclovas, C., Colvin, M and Thelen, M.P. Structural model of the RFC clamp loading mechanism. – A poster presented at the Keystone Symposia meeting: Molecular Mechanisms of DNA Replication and Recombination. Snowbird, Utah, USA, January 7 – 13, 2002.
14. "Prospects for Computational Biochemistry" DOE/NASA/DOD Mission Computing Conference, February 5, 2002. "Computational Biochemistry" UC Davis Biotechnology Seminar Series, Davis, California, February 15, 2002. "Dissociation of water under pressure" American Physical Society, Indianapolis, Indiana, March 2002 (contributed talk)
17. Computational Biochemistry & Roles for Nanotechnology in the Biosciences" Briefing for JASON Study on Opportunities At The Intersection Of Nanoscience, Biology And Computation, La Jolla, CA, June 25, 2002. "First principles simulations of ion solvation" 85th Canadian Society for Chemistry Conference, Vancouver, Canada June 2002 (invited talk)
19. "Computational Biochemistry: Principles and Practice", Georgetown University Georgetown University Medical Center, Washington DC July 29, 2002
20. "Prospects for Computational Biochemistry" Supercomputing 2002 "Masterworks Speaker", Baltimore, MD, November 20, 2002.
21. "Comparative modeling based on a combination of sequence comparison and assessment of structural fitness". Fifth Meeting on the Critical Assessment of Techniques for Protein Structure Prediction., Asilomar, CA, USA. December 1 - 5, 2002.
22. "Computational Biology at Lawrence Livermore National Laboratory". University of California, Santa Barbara, December 4, 2002.
23. "Bringing the Genome to Life: the Role of Computing" AAAS National Meeting, Denver, CO, February 18, 2003.
24. "First-principles simulations of rigid water" Theresa Head-Gordon Group seminar, University of California, Berkeley, September 2003 (invited talk)
25. "Computational Biochemistry" AIChE National Meeting, San Francisco, CA, November 17, 2003.