Dissertation

submitted to the

Combined Faculties

for the Natural Sciences and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

Presented by

Dipl.-Phys. Andrea C. Vaiana

Born in Cambridge MA (USA)

Oral examination: 02.12.2004

Towards the Understanding of Fluorescence Quenching Mechanisms: Molecular Dynamics Simulations of Dye-Quencher Interactions in Biomolecular Systems.

Referees: Prof. Dr. Jeremy C. Smith

Prof. Dr. Heinz Horner

Zusammenfassung

Der Nachweis von p53-Antikörpern ist ein wichtiges Diagnoseverfahren zur Krebsfrüherkennung und Verlaufskontrolle. Mit Hilfe des Farbstoffes MR121 kann ein solcher Nachweis mittels Fluoreszenspektroskopie geführt werden. Grundlage der Fluoreszenzlöschung (quenching) des Farbstoffes ist dessen bimolekulare Wechselwirkung mit Tryptophanresten. Die molekularen Mechanismen, die dieser Fluoreszenzlöschung zu Grunde liegen, sind weitestgehend unbekannt. In der vorliegenden Arbeit werden MD-Kraftfeldparameter für die Farbstoffe Rhodamine 6G (R6G) und MR121 entwickelt und die Wechselwirkungen zwischen Farbstoff und Quencher (Trp) mittels Molekulardynamik Simulationen untersucht. Zur Bereitstellung eines CHARMM Kraftfelds für beide oben genannten Farbstoffe wurde eine neue, automatisierte Optimisierungsmethode für Kraftfeldparameter von Molekülen kleinerer und mittlerer Größe entwickelt. Als Referenzdaten dienten ab initio quanten-chemische Berechnungen. In einem zweiten Schritt wurden MD-Simulationen zweier Farbstoff/Quencher-Systeme durchgeführt (R6G/Trp und MR121/Trp). Die Untersuchung dieser Modellsysteme ergab wichtige Einsichten in bimolekulare Wechselwirkungen zwischen Farbstoff und Quencher. Basierend auf Abstand und relativer Orientierung des Systems, konnte eine quantitative Interpretation des Quenching-Mechanismus abgeleitet werden. Die an diesen Modellsystemen gewonnenen Resultate dienten dann zur Interpretation der Simulationsdaten zweier, am N-Terminus mit obigen Farbstoffen markierten, Epitope des Tumorsupressor Proteins p53. In allen Phasen der Doktorarbeit wurde gezielt versucht, die theoretischen Resultate mit verfügbaren experimentellen Daten zu vergleichen oder zu kombinieren.

Abstract

The presence of antibodies directed against p53 in human blood serum is a specific and independent marker of cancer. The present thesis is a simulative study of two fluorescent dyes (rhodamine 6G and MR121) used in the development of fluorescence-based immunoassays for the detection of p53 antibodies. The "selective" fluorescence quenching property of tryptophan residues present in dye-conjugated peptide chains enables monitoring of conformational dynamics and antibody binding events by means of fluorescence spectroscopy. The molecular mechanisms of the quenching of fluorescent dyes are mostly unknown. Here MD simulations are used in combination with existing results from ensemble fluorescence experiments in order to obtain predictive theoretical insight into dye/quencher interactions. A new automated refinement method was developed for deriving reliable molecular mechanics force field parameters for small- to medium-sized molecules using reference data from high level *ab initio* quantum chemical calculations. Using this method CHARMM force field parameters for the two dyes were derived. The parameters were then used to perform MD simulations on two simplified, but realistic dye quencher systems: MR121/TRP and R6G/TRP. Results of these simulations have given important insight on the bimolecular interactions between the dyes and the quencher. A quantitative interpretation of the quenching mechanism based on it's dependence on dye/quencher distance and orientation has emerged. These results were then applied to the interpretation of simulation data of an epitope from the tumor suppressor protein p53 which was labelled at the N-terminus first with MR121 and then with R6G. In all phases of the thesis care was taken to confront and/or combine the theoretical results with available experimental data.

"..., so solid from a distance, is riven into sects and clans, which radiate and join, and change their names according to the aspect from which they are approached. Study it for years with the best teachers, and when you raise your head nothing they have told you quite fits."

E.M. Forster on Hinduism
"A Passage to India"

TABLE OF CONTENTS

1. Introduction

This thesis is a simulative study of two fluorescent dyes used in the development of fluorescence-based immunoassays for the detection of p53 antibodies. The presence of antibodies directed against p53 in human blood serum is a specific and independent marker of cancer. In these assays, fluorescently labelled peptides are used to detect p53 antibody binding events in solution. The aim of the thesis is to gain physical insight into the biologically relevant fluorescence properties of the dyes within the physically complex environmental context of a solvated peptide. The thesis is divided into three main parts: in the first part the theoretical model used to describe the dyes, *i.e.* the molecular mechanics force field, is constructed, in the second part the problem of dye/quencher interactions is addressed for the 'simplified but realistic' case of solvated dye/quencher complexes, in the third part the complexity of the simulated system incorporates full-fledged dye-labelled solvated peptides, and again dye/quencher interactions are analysed. At each step care is taken to validate and/or confront the theoretical models with available experimental data on the same systems. This, in turn, allows for a physically meaningful interpretation of experimental results in terms of the structure and dynamics of the molecules investigated.

In the first part of the thesis, a novel automated method is developed for deriving molecular mechanics force field parameters for small to medium sized molecules. The method includes fitting the molecular mechanics potential to both vibrational frequencies and eigenvector projections derived from quantum chemical calculations. The method is benchmarked on a series of aromatic molecules and then applied to derive new parameters for the two fluorescent dyes: rhodamine 6G (R6G) and MR121. The resulting parameter sets are tested, where possible, against independent experimental data.

In the second part the dye parameters are used in molecular dynamics simulations of the bimolecular interaction of the two dyes with the amino acid triptophan (Trp) in aqueous solution. The simulated data are compared with experimental results obtained from steady-state and time-resolved fluorescence quenching experiments of the dyes with Trp. A detailed geometric model of fluorescence quenching of dyes by Trp is obtained.

In the third part simulations are performed on a Trp-containing immuno dominant p53 epitope labelled with MR121 or with R6G, the model derived in part two is used for the interpretation of fluorescence spectroscopy data previously recorded on the same systems.

1.1. Molecular Modelling and Simulation

Computer simulations are nowadays considered to be the "third way" of doing science comparable, in some measure to both experiment and theory. Simulation can provide reliable predictions when experiments are not possible, too difficult or expensive and can supplement pure theory when analytical solutions to a problem are unobtainable. Shortly after the formulation of quantum mechanics Dirac (1929) already had recognized that although the underlying laws and mathematics describing, in his words:

> *"...a large part of physics and the whole of chemistry are thus completely known, and the difficulty is only that the exact application of these laws leads to equations much too complicated to be soluble."*

The complication in solving the equations is often due to the many-body nature of most problems. In disciplines such as physics, chemistry and biology scientists have, over the last half-century, turned to computer simulations as a very powerful investigative tool able to provide detailed and accurate information about many-body systems. Simulations in principle allow, starting from a microscopic description of the system, to extract the macroscopic properties measured in experiments. Statistical mechanics, acts as the physical foundation of simulation *i.e.*, it connects thermodynamic macroscopic variables such as volume, pressure and temperature to microscopic variables such as individual atomic positions, momenta and velocities.

Simulations scale up with the increase in computer power: the earliest simulations involved 32 particles [Alder and Wainwright 1957], today systems with millions of particles are treatable. With persistent progress in computer performance and refined software during the last decades, the use of computer simulations as a theoretical approach for the modelling of the structure, energetics and dynamics of biomolecules has become more and more popular. Molecular modelling as it is intended today, covers a vast range of methodologies and techniques. These span from so called *ab initio* quantum chemical calculations for the

modelling of small, isolated molecules to empirical homology based modelling of the structures of very large biomolecules and complexes with a score of hybrid techniques lying between these two extremes. *Ab initio* methods are mostly based on direct solution of the Schrödinger equation (or equivalent formulations of it) starting from the Hamiltonian of the system. Quantum mechanical methods have been addressed to chemical problems involving structures and reactions of small, isolated molecules. As electrons are explicitly represented in quantum mechanical calculations, properties which depend on the electronic distribution within the molecular system such as bond breaking and bond formation or photophysical properties can, in principle be derived. One recent example of this is the use of time-dependent density functional theory to describe the fluorescence quenching and lifetimes involved in stacking interactions of 2-aminopurine with purines and pyramidines [Jean and Hall 2001].

1.2. Molecular Dynamics Simulations of Biomolecules

In the investigation of biologically relevant systems, one often has to deal with thousands of atoms at a time. Additionally, biopolymers like DNA or proteins form their native, biologically active structures generally only in aqueous solution. Thus, solvent effects have to be taken into account to gain meaningful, experimentally comparable results. For the treatment of these complex systems, quantum mechanical calculations and semi-empirical schemes fail. This often applies even to systems in which the use of quantum methods would in principle be required. For example, fluorescent properties such as quenching are a result of complex interactions between chromophores and the surrounding environment. Although quantum mechanical treatment is required to model the electronic transitions involved in excited state kinetics, the associated task of exploring the configurational space of dye/environment systems is computationally too demanding to be treated with quantum methods. On the other hand, molecular mechanics methods are based on simpler empirical potential energy functions or force fields and calculate the classical energy of the system as a function of the nuclear positions only, electronic effects are partially incorporated in the parameterization of the force field. This allows extensive exploration of the configurational space of much larger systems than what is currently possible using *ab initio* methods. Numerous applications during the past 25 years have proven the validity and accuracy of

these empirical models. These, in many cases, match the accuracy of quantum mechanical calculations and require only a fraction of the computer time. Hence, molecular mechanics is invariably used to perform calculations on systems containing a significant number of atoms.

The two main molecular mechanics methods presently used in the investigation of biological systems are the so-called Monte Carlo (MC) and molecular dynamics (MD) algorithms. There are many modelling software packages capable of performing both MD and MC simulations among the most frequently used for the simulation of biomolecules are the CHARMM, AMBER and GROMACS packages [Brooks, et al. 1983, Pearlman, et al. 1995, Bekker, et al. 1993]. Both MD and MC methods are based on an empirical force field function for evaluating the energy of the system. MC simulations simply impose large, random motions on the system and by comparing the energy of the new configuration to a Boltzman distribution accept or reject the new configuration. In this manner an ensemble of energetically feasible configurations of the system is generated. Because MC simulation samples phase space without a true time variable or a realistic dynamics trajectory, it cannot provide time-dependent quantities. On the contrary MD simulations evaluate forces to determine incremental atomic motions. In MD simulations initial positions and velocities are assigned to each atom, and Newton's laws are applied at the atomic level to propagate the system's motion through phase space. From a sufficiently long MD trajectory *i.e.*, a trajectory in which the system explores a significant portion of phase space, dynamical properties such as time correlation functions, transport coefficients or the dynamics of inter- and intramolecular contact formations can be calculated.

Time-averaged molecular properties directly comparable to experimentally measurable ensemble averages, such as free energy differences, can in principle be extracted from MD trajectories of a single molecule by means of the ergodic hypothesis together with the principles of statistical mechanics. To ensure energy conservation and the stability of the computations of dynamics at atomic detail one has to apply very small integration time-steps, typically of one femtosecond (10^{-15} s). Thus, the simulation of low-frequency biomolecular dynamics and conformational transitions, occuring at pico- to microsecond time scales (10^{-12} to 10^{-6} s) require very long simulations, these are computationally expensive and require high-performance parallel computer systems with large disks for data storage and analysis. The fastest computational resources available to scientists today are parallel clusters of thousands of processors, these reach computational speeds in the teraflop/s range (1 teraflop = 10^{12} floating point operations). The power and speed of these machines is ever increasing, today the number one ranking cluster in the world is the NEC Earth Simulator in Japan, with it's

5120 processors it can reach a maximum of 41.0 teraflop/s. The HELICS cluster of the Interdisziplinäres Zentrum für Wissenschaftliches Rechnen at the University of Heidelberg, on which most of the simulations presented here were performed, ranks 64th in the world and is capable with it's 512 processors of reaching a maximum of 1.4 teraflop/s. An up to date ranking of supercomputers worldwide is available at http://www.top500.org. Even with the fastest modern computer clusters simulation times have only begun to reach the microsecond timescale for small to medium sized biomolecules. A notable example of classical MD on high efficiency parallel computers is a 1-microsecond simulation with explicit representation of water of the folding of a protein [Duan and Kollman 1998]. More recently a 10 nanoseconds simulation study of the dynamics of water penetration through a membrane-protein in a lipid bilayer using an explicit representation of all solvent molecules at atomic detail [de Groot and Grubmuller 2001] has been performed. These and other exciting works demonstrate the possibilities offered by MD simulations in attempting to deal with the protein folding problem and complex biological processes occurring in cells.

On the other hand, short biopolymers have many applications in diagnostic research and biomedical analysis where they are used as tracers to specifically target macromolecular compounds such as antibodies or DNA. For highly sensitive analytics the development of tailor-made fluorescent probes as diagnostic tools is a promising approach. The sensitivity and the specificity of these probes to their target molecules are crucial for the development of reliable assay formats. Small ligands based on peptides or DNA-fragments are frequently labeled with organic fluorescent dyes to serve as tracer molecules for the target. In many cases chemical modification of the ligand, necessary for optical detection, influences its affinity to the target. Also, the spectroscopic properties of the dye can be altered by changes in the microenvironment of the dye upon ligand-binding or recognition of the probe by the target molecule. These matters can be addressed theoretically by means of MD simulations. MD simulations are well suited for describing both intramolecular interaction geometries and conformational dynamics of molecular probes at atomic detail in addition they are capable of explicitly including the effects of solvent. Studies concerning the simulations of organic dye structures and their interactions with biomolecules are rare. Only a few works exist, such as the simulation of the solvation behaviour of coumarin 153 [Cichos, et al. 1999] or an MD simulation study of fluorescein in an antibody binding site [Lim and Herron 1995]. No attempts have been made concerning the simulation of labeled biomolecules as molecular probes. This is most probably due to the difficulty of developing new force field parameters for dye molecules. For most fluorescent probes suitable experimental data (e.g. X-ray crystal

structures, vibrational spectra, NMR measurements) upon which to base the refinement of new parameter sets are scarce.

The automated frequency matching method (AFMM) [Vaiana, et al. 2003a] described in the first part of the thesis is used to derive parameter sets for the CHARMM force field. The method relies solely on reference data from high-level quantum chemical calculations and produces reliable parameter sets for small polycyclic molecules [Vaiana, et al. 2003a, Cournia, et al. 2003, Vaiana, et al. 2003b] in the absence of experimental data. With the new force field parameter sets derived for the commonly used fluorescent dye Rhodamine 6G [Vaiana, et al. 2003a] and for the recently syntesized [Sauer, et al. 1997] oxazine dervative MR121 [Vaiana, et al. 2003b], computer simulations of these dyes have recently become possible. Both MR121 [Neuweiler, et al. 2002] and R6G [H. Neuweiler, unpublished results] have been used as fluorescent labels for p53 epitopes in experiments aimed at developing biosensors based on fluorescence quenching of the dyes by Trp for the detection of p53 antibodies. In these works the dye is covalently linked to the peptide epitope. When the dye-labelled peptide is free in solution it adopts a conformation in which the dye is non-fluorescent (quenched) due to it's interaction with a Trp residue, inherent in the peptide sequence. Upon binding to a p53 antybody this interaction is perturbed and fluorescence of the dye is released. In this manner the indicator of a binding event is a positive fluorescence signal over a negative background of non-fluorescent unbound epitopes. This allows for significantly better signal to noise ratios and therefore an increased sensitivity of the assay in comparison to other commonly used approaches. Conformational dynamics of the peptide conjugates were measured at the single molecule level.

The MD simulations of dye/Trp complexes presented in the second part of the thesis are used, in combination with time-resolved fluorescence experiments, to create a coherent quantitative picture of the dynamic and static mechanisms involved in the quenching of the dyes in the presence of Trp in aqueous solution. Thus, a basis is provided for quantitatively interpreting fluorescence quenching data in terms of the underlying molecular interaction geometries. In the third part of this thesis, MD simulations as well as umbrella sampling studies of one of the peptide conjugates labelled with both R6G and MR121 are performed including explicit solvent representation. The simulations reveal inter- and intramolecular interaction mechanisms between dye and Trp as well as conformational dynamics of the peptide. Umbrella sampling studies were used to calculate free energy differences and to assess relative stabilities of certain peptide conformations and dye-Trp interaction geometries. Results are compared to experimental fluorescence quenching data on the same systems. The

outcome of these theoretical studies help to understand and interpret the experimental results by revealing structures, dynamics, and interaction geometries at atomic detail. In this way the fluorescence quenching mechanism between dye and Trp, derived from the combination of bimolecular simulations and time resolved fluorescence quenching experiments of the second part of the thesis, and the conformational dynamics of the peptide conjugates are elucidated from a theoretical point of view. In turn, the results drawn from MD simulations of the labelled peptide may be used to optimise the fluorescence quenching efficiency of the probes and therefore the detection sensitivity of the experimental assay.

Motivation for this project stems principally from works on the structure and the immunology of human p53, a tumour suppressor protein which is strongly related to human cancer and especially from the recent development of biosensors based on fluorescence quenching of dyes by Trp which allow detection of p53 antibodies at the single molecule level [Neuweiler, et al. 2002]. These works are briefly discussed in chapter 5 of the thesis along with some biological background on p53 antibodies and tumor diagnostics. The theoretical methods used throughout the thesis are described in detail in the next chapter.

References

Alder, B. J. and T. E. Wainwright (1957). "Phase Transition for a Hard Sphere System" J Chem Phys (27): 1208.

Jean, J. M. and K. B. Hall (2001). "2-Aminopurine fluorescence quenching and lifetimes: role of base stacking" Proc Natl Acad Sci U S A **1** (98): 37-41.

Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus (1983). "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations" J Comput Chem **4** 187-217.

Pearlman, D. A., D. A. Case, J. W. Caldwell, W. S. Ross, T. E. Cheatham, I. DeBolt, S. , D. Ferguson, G. Seibel and P. A. Kollman (1995). "AMBER, a package of computer programs for apply-ing molecular mechanics, normal mode analysis, molecular dynamics and free energy cal-culations to simulate the structural and energetic properties of molecules." Comput Phys Commun (91): 1-41.

Bekker, H., H. J. C. Berendsen, E. J. Dijkstra, S. Achterop, R. van Drunen, D. van Der Spoel, A. Sijbers, H. Keegstra, B. Reitsma and M. K. R. Renardus (1993). "GROMACS: A Parallel computer for molecular dynamics simulations" Proc of the 4 Int conf Physics Computing 92 252-6.

Duan, Y. and P. A. Kollman (1998). "Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution" Science **5389** (282): 740-4.

de Groot, B. L. and H. Grubmuller (2001). "Water permeation across biological membranes: mechanism and dynamics of aquaporin-1 and GlpF" Science **5550** (294): 2353-7.

Cichos, F., R. Brown, U. Rempel and C. von Borczyskowski (1999). "Molecular Dynamics Simulations of the Solvation of Coumarine 153 in a Mixture of an Alkane and an Alcohol." J Phys Chem A **15** (103): 2506-12.

Lim, K. and J. N. Herron (1995). "Molecular Dynamics of the Anti-Fluorescein 4-4-20 Antigen- Binding Fragment." Biochemistry **21** (34): 6962-74.

Vaiana, A. C., A. Schulz, J. Wolfrum, M. Sauer and J. C. Smith (2003a). "Molecular Mechanics Force Field Parameterization of the Fluorescent Probe Rhodamine 6G using Automated Frequency Matching" J Comput Chem **5** (24): 632-9.

Cournia, Z., A. C. Vaiana, J. C. Smith and M. Ullmann (2003). "Derivation of a Molecular Mechanics Force Field for Cholesterol." Pure Appl Chem (in press):

Vaiana, A. C., H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer and J. C. Smith (2003b). "Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation" J Am Chem Soc (in press):

Sauer, M., C. Zander, R. Muller, F. Gobel, A. Schulz, S. Siebert, K. H. Drexhage and J. Wolfrum (1997). "Diode laser based time-resolved detection and identification of individual mononucleotide molecules in aqueous solution" Ultrasensitive Biochemical Diagnostics II. Proceedings of SPIE - The International Society for Optical Engineering: 61-8.

Neuweiler, H., A. Schulz, A. C. Vaiana, J. C. Smith, S. Kaul, J. Wolfrum and M. Sauer (2002). "Detection of individual p53-autoantibodies by using quenched peptide-based molecular probes" Angew Chem, Int Ed Engl **24** (41): 4769-73.

2. Theory and Methods

In this chapter the theoretical basis of the methods used in the thesis is presented in detail. The first part of the chapter is aimed at giving some insight into the quantum chemical methods used here. The second part deals with the molecular modelling methods, in particular with MD simulations and empirical force fields. A basic theoretical introduction of fluorescence and quenching of fluorescent dyes is presented in the third part.

2.1. Quantum Chemical Calculations

2.1.1. Hartree-Fock

Hartree-Fock theory is fundamental to much of electronic structure theory. It is the basis of molecular orbital (MO) theory, it assumes that the motion of each electron in a molecule can be described by a single-particle function (orbital) which does not explicitly depend on the instantaneous motions of the other electrons. Hartree-Fock theory was developed to solve the time independent Schrödinger equation within the framework of the Born-Oppenheimer approximation. Although the success of Hartree-Fock MO theory is testified by the ubiquitous use of orbital concepts in chemistry, it is important to bear in mind that molecular orbitals are mathematical constructs which can only approximate reality. The Schrödinger equation cannot be solved exactly for systems with more than two interacting particles (many body problems). Only for the hydrogen atom (and other one-electron systems, such as He^+) do orbitals represent the exact eigenfunctions of the electronic Hamiltonian.

Under the Born-Oppenheimer approximation the motions of the electrons are decoupled from those of the nuclei. The mass of a nucleus is much greater than the mass of an electron (the rest mass of the smallest nucleus, a proton is 1836 times larger than that of an electron). This means that electrons can adjust almost instantaneously to changes in the positions of the nuclei. The electronic wave function is thus a function of the nuclear positions and does not depend on the momenta of the nuclei. Under this approximation the total wavefunction for the molecule can be written as the product of the electronic wavefunction and the nuclear wavefunction:

$$\Psi_{\text{tot}}(\text{nuclei, electrons}) = \Psi(\text{electrons})\Psi(\text{nuclei}) \tag{2.1.1}$$

The total energy equals the sum of the nuclear energy *i.e.*, the electrostatic repulsion between the nuclei and the electronic energy *i.e.*, kinetic and potential energy of the electrons in the field of the nuclei together with electron-electron repulsion: $E_{tot} = E_{electrons} + E_{nuclei}$

The problem is then solving the 'electronic' Schrödinger equation (in atomic units):

$$\left[-\frac{1}{2}\sum_i \nabla_i^2 - \sum_{A,i} \frac{Z_A}{r_{A_i}} + \sum_{A>B} \frac{Z_A Z_B}{R_{AB}} + \sum_{i>j} \frac{1}{r_{ij}} \right] \Psi(\mathbf{r},\mathbf{R}) = E_{el}\Psi(\mathbf{r},\mathbf{R}) \tag{2.1.2}$$

where the electronic and nuclear degrees of freedom (both spacial and spin) are indicated by **r** and by **R**, respectively. Ψ(**r,R**) is the electronic wavefunction. The operator in the brackets is the molecular Hamiltonian, it comprises four terms: the first is electronic kinetic energy, the last three are electrostatic interaction terms for electron-nucleus attraction, nucleus-nucleus repulsion which depends only on the coordinates of the nuclei, and electron-electron repulsion. Recall from the Born-Oppenheimer approximation that E_{el} gives the potential energy experienced by the nuclei. In other words, E_{el}(**R**) gives the potential energy surface as a funtion of the nuclear coordinates, from which properties such as the equilibrium geometry and the vibrational frequencies of the molecule can be derived. Other important molecular properties such as dipole (and multipole) moments, polarizability and the electrostatic field generated by the molecule can be derived from Ψ(**r,R**).

The basic idea of Hartree-Fock theory is to start with the approximation that electrons act as non-interacting particles. In this approximation the last term in the Hamiltonian (electronic repulsion) would be dropped and the Hamiltonian is separable into single electron terms. If the molecule contains N_e electrons, the molecular wavefunction would then reduce to the product of N_e independent hydrogen atom wavefunctions (orbitals): $\Psi(\mathbf{r}_1,\mathbf{r}_2, \ldots, \mathbf{r}_{Ne}) = \Psi(\mathbf{r}_1)\Psi(\mathbf{r}_2) \ldots \Psi(\mathbf{r}_{Ne})$. Although this is a somewhat drastic approximation, it is a good starting point and partial corrections will be introduced later. It is thus plausible to start with a wavefunction of general form:

$$\Psi_{HP}(\mathbf{r}_1,\mathbf{r}_2,\cdots,\mathbf{r}_{N_e}) = \phi_1(\mathbf{r}_1)\phi_2(\mathbf{r}_2)\cdots\phi_{N_e}(\mathbf{r}_{N_e}) \tag{2.1.3}$$

which is known as a Hartree product.

This functional form has one major shortcoming in that it fails to satisfy the antisymmetry principle *i.e.*, a wavefunction describing fermions should be antisymmetric with respect to the interchange of any set of space-spin coordinates. This problem can be overcome by separating

the space dependent part of the wavefunction $\phi(\mathbf{r})$ from the spin dependent part $\alpha(\omega)$ by introducing a spin orbital $\chi(\mathbf{x})=\phi(\mathbf{r})\alpha(\omega)$ where $\mathbf{x}=\{\mathbf{r},\omega\}$ is the set of space-spin coordinates and by representing the wavefunction as a Slater determinant of spin orbitals:

$$\Psi = \frac{1}{\sqrt{N_{el}!}} \begin{vmatrix} \chi_1(\mathbf{x}_1) & \chi_2(\mathbf{x}_1) & \cdots & \chi_{N_{el}}(\mathbf{x}_1) \\ \chi_1(\mathbf{x}_2) & \chi_2(\mathbf{x}_2) & \cdots & \chi_{N_{el}}(\mathbf{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \chi_1(\mathbf{x}_{N_{el}}) & \chi_2(\mathbf{x}_{N_{el}}) & \cdots & \chi_{N_{el}}(\mathbf{x}_{N_{el}}) \end{vmatrix} \qquad (2.1.4)$$

This functional form satisfies the antisymmetry principle and assures indistinguishability of all electrons. Generally, since such a determinant can always be constructed (within a sign) if only the occupied orbitals $\{\chi_i(\mathbf{x}), \chi_j(\mathbf{x}), \ldots, \chi_k(\mathbf{x})\}$ are known, a short hand ket notation is used: $\Psi = |\chi_i(\mathbf{x}), \chi_j(\mathbf{x}), \ldots, \chi_k(\mathbf{x})\rangle$ or more simply $\Psi = |ij\ldots k\rangle$ (the normalisation factor is often also omitted). It can be shown that the assumption that the electrons can be described by the antisymmetrized product above is mathematically equivalent to assuming that each electron moves independently of all others and feels only the average Coulomb repulsion due to the positions of all other electrons and the exchange interaction due to antisymmetrization. Hartree-Fock is therefore a mean field theory.

Once this functional form has been adopted for the wavefunction Ψ, the molecular orbitals can be obtained by making use of the variational theorem which states that the energy calculated as $E_{est}=\langle\Psi_{est}|H_{el}|\Psi_{est}\rangle$, where H_{el} is the electronic Hamiltonian and Ψ_{est} is an estimate of the wavefunction, is always greater than the true energy $E_{el}=\langle\Psi|H_{el}|\Psi\rangle$ calculated from the correct wavefunction Ψ. Better approximate wavefunctions can be obtained by varying their parameters until the energy is minimized within the given functional space. Hence the correct molecular orbitals are those which minimize the estimated electronic energy E_{est}. Usually the molecular orbitals are obtained numerically as a linear combination of a set of given basis functions, the so called 'atomic orbital' basis functions. The atomic orbital basis functions most commonly used in quantum chemical software packages are atom-centred Gaussian type functions.

Quantum chemical, or *ab initio*, methods have the great advantage that a molecule can be treated without the need for any empirically derived parameters. The accuracy attainable is very good and methods exist to reduce the errors caused by the various approximations used.

The greatest drawback of *ab initio* calculations lies in the fact that these computations are quite expensive both in terms of c.p.u. time and memory requirements. The practical limits of these calculations with modern computers is of the order of a few tens of atoms. Due to

constant progress in algorithms and computer power one can certainly expect these limits to extended in the coming years. However these techniques are computationally far too expensive to be applied to be applied to bio-molecules.

In contrast with quantum chemical calculations, the use of molecular mechanics force field methods, based on an empirical representation of the classical potential energy function of the system, is computationally very cheap. Although results obtained with these methods are only approximate they have the great advantage of allowing the introduction of realistic representation of environment in the model. The potential energy function of several conformations of a complex system (ex. including large numbers of water molecules and ions surrounding a protein or nucleic acid strand) can be readily calculated. This facilitates the use of techniques such as molecular dynamics which allow the thermal motions of a system to be explored. Simulations using molecular dynamics and force field based methods are discussed in detail in the following section.

2.2. Molecular Dynamics Simulations

MD simulations are one of the most commonly used tools in the theoretical study of biological molecules. They allow detailed, atomic level description of the fluctuations and conformational changes of proteins and nucleic acids and are also used for the final refinement of molecular structures from X-ray crystallography and NMR experiments. In MD the time evolution of a molecular system through phase space is calculated by integrating Newton's laws of motion. The result of an MD calculation is a trajectory of the time evolution of the positions and velocities of the individual atoms constituting the system.

The first MD simulation was a study of the interactions of 32 hard spheres performed by Alder and Wainwright in the late 1950´s [Alder and Wainwright 1959, 1957]. Many important insights concerning the behaviour of simple liquids emerged from their studies. The next major advance was in 1964, when Rahman carried out the first simulation using a realistic potential for liquid argon [Rahman 1964]. The first MD simulation of a realistic system was performed by Rahman and Stillinger in their simulation of liquid water in 1974 [Stillinger and

Rahman 1974]. This work introduced the first useful force field model for water and together with other theoretical works by Stillinger has led to a deeper understanding of the physics of glass forming liquids using the theory of inherent structures [Weber and Stillinger 1984, Stillinger and Weber 1988]. The first protein simulation appeared in 1977 with the vacuum simulation of the bovine pancreatic enzyme inhibitor (BPTI) [McCammon, et al. 1977]. A new view of proteins emerged from this study substituting the assumption that proteins are essentially rigid structures. Since that time, thanks to the development of computationally inexpensive but precise models for water molecules [Jorgensen, et al. 1983], much progress has been made in the inclusion of solvent effects to study biomolecules under physiological conditions. Other improvements include more realistic treatments of the system boundaries and accurate treatment of long-range electrostatic forces.

Today MD simulations of solvated proteins, protein-DNA complexes as well as membrane-lipid systems are quite common. Issues addressed by means of MD simulations range from the understanding of basic thermodynamic properties of highly simplified model systems to detailed treatment of biological processes such as ligand binding, folding of small proteins, and enzymatic mechanisms.

The protein-folding problem *i.e.*, the problem of how the amino acid sequence of a polypeptide determines its three dimensional native structure, is of special interest in theoretical biophysics. The importance of understanding the protein folding process is evidenced by the Levinthal paradox [Levinthal 1968]: if the folding mechanism were to occur by random attempts of the system to explore all it's accessible configurations (which for an average protein are of the order of 10^{30}), assuming a characteristic time of 10^{-12} sec for the probing of each configuration, the average folding time would be of the order of 10^{10} years. The fact that folding processes in nature occur in biologically relevant timescales *i.e.*, in immensely shorter times, implies that folding cannot be the result of a mere random search of the correct functional form through configurational space. Several attempts have been made to simulate the folding of proteins. For example, a protein-folding pathway has been observed during a 1-microsecond dynamics simulation in explicit water [Duan and Kollman 1998]. The main difficulty in resolving the detailed mechanisms of protein folding lies in effectively exploring the conformational space spanned by the very large number of degrees of freedom of a polypeptide chain. To approach the protein-folding problem, theoreticians are focussing on the simulation of conformational dynamics of small peptides, which serve as a model for protein-folding.

Many attempts have been made to simulate small peptides both with explicit and implicit solvent models, such as in the works of Karplus [Schaefer, et al. 1998] or van Gunsteren [Daura, et al. 1998]. But still, even for small peptide sequences the number of internal degrees of freedom is too large for complete sampling of configurational space with modern computers. The backbone of a 30 amino acid residue peptide contains 60 dihedral angles. In the unphysical approximation that each dihedral angle can assume one of only two possible states, these alone would span a conformational space of approximately 2^{30} ($\sim 10^9$) distinct conformers. Although this incredibly large number of degrees of freedom would seem to rule out any possibility of obtaining significant results from MD simulations, the subset of biologically interesting degrees of freedom turns out to be significantly smaller. Special sampling techniques such as umbrella sampling [Torrie and Valleau 1977] have been developed to explore conformational space of proteins and peptides efficiently [Guo, et al. 1997, Mu and Stock 2002]. The sampling problem is particularly important for the calculation of thermodynamic properties from time trajectories, this issue will be thoroughly discussed later in this chapter.

Prior to performing an MD simulation some necessary steps must be taken. It is necessary to select an initial configuration of the system. This requires some care and it is generally good practice to choose initial states of the system which are close to the state that one wishes to simulate. It is also important to ensure that the starting configurations do not contain any high potential energy interactions as these can lead to instabilities during the simulation. A standard way of relaxing these so called 'hot spots' is to perform an energy minimisation prior to running the simulation. An initial distribution of velocities must also be assigned to the system. The equilibrium distribution of velocities for a thermodynamic system is a Maxwell-Boltzman distribution; in most cases it is more convenient to assign a Gaussian distribution of velocities initially and to run a short equilibration phase of dynamics to allow the kinetic and potential energy of the system to reach their equilibrium distributions spontaneously. A function describing the potential energy of the system must also be selected. Given an initial distribution of velocities and a starting set of coordinate values for the atoms of the system, a trajectory of the time evolution of the molecular system can then be obtained by integrating the potential energy function of the system numerically, using sufficiently small time steps, usually of the order of one femtosecond. Finally, the obtained trajectories must be stored and subsequently analysed. In order to extract desired information and parameters, care must be taken to save coordinates (and velocities if needed) often enough so as not to loose

information. The potential energy function, or force field, describing the forces acting between the atoms as a function of their positions is the most important aspect of the simulation. The accuracy of the applied potential energy function is crucial for obtaining physically reasonable results from a MD simulation.

2.2.1. Force Fields – the Empirical Potential Energy Function

The fact that molecular mechanics works at all is due to the validity of several assumptions. The first to be named is the Born-Oppenheimer approximation already mentioned in section 2.1.1. Without this assumption it would be impossible to contemplate writing the energy as a function of the nuclear coordinates at all. Other, more drastic approximations concern the single energy terms in the force field, these are represented by empirical potential energy functions. The functional form adopted for the potential energy is designed to allow efficient computation of the energy of a system as a function of the coordinates.
The potential energy $V(\mathbf{r})$ can be described as the sum over the bonded (or internal) and non-bonded (or external) energy terms of all the atoms in the system.

$$V(\mathbf{r}) = V_{bonded} + V_{non-bonded} \tag{2.2.1}$$

Bonded Interactions

In most commonly used force fields bonded interactions are described as a sum over four terms: simple harmonic terms describe bond stretching and angle bending. Rotation about single bonds (torsions) is governed by sinusoidal energies. The planarity of groups (e.g., the amide planes of proteins) can also be enforced by harmonic potentials known as improper torsions.

$$V_{bonded} = V_{bonds} + V_{angles} + V_{torsions} + V_{impropers} \tag{2.2.2}$$

Bond Stretching

An empirical potential energy function for a typical bond is the Morse potential which has the form:

$$V_{Morse}(l) = D_e\{1 - \exp[-a(l - l_0)]\}^2 \tag{2.2.3}$$

$$a = \omega\sqrt{\left(\frac{\mu}{2D_e}\right)} \tag{2.2.4}$$

D_e is the depth of the potential minimum, l is the bond length *i.e.*, the distance between the two bonded atoms, l_0 is the equilibrium value of the bond length, μ is the reduced mass, and ω is the frequency of the bond vibration in the approximation of small displacements from the equilibrium position *i.e*, if $l \approx l_0$. In this case the potential function can be approximated by a harmonic well and ω is related to the stretching constant of the bond, k, by:

$$\omega = \sqrt{\frac{k}{\mu}} \tag{2.2.5}$$

Although the Morse potential gives a precise description of the potential energy as a function of the bond length, it is generally not used in molecular mechanics force fields. This is mainly due to the computational expense of evaluating the exponential part of Eq. 2.2.3 and also to the fact that three parameters are required per bond. The Morse curve describes a wide range of behaviour from equilibrium conditions to complete dissociation of the bond. At normal temperatures for biological systems in equilibrium it is rare for bonds to deviate significantly from their equilibrium values, under these assumptions a simple harmonic potential (Hooke´s law) is an adequate approximation for the description of bond stretching energies.

$$V_{bonds}(l) = \frac{k}{2}(l - l_0)^2 \tag{2.2.6}$$

This has the advantage of being computationally inexpensive and requiring only two parameters per bond type. The harmonic approximation to the Morse curve is illustrated in Figure 2.2.1.

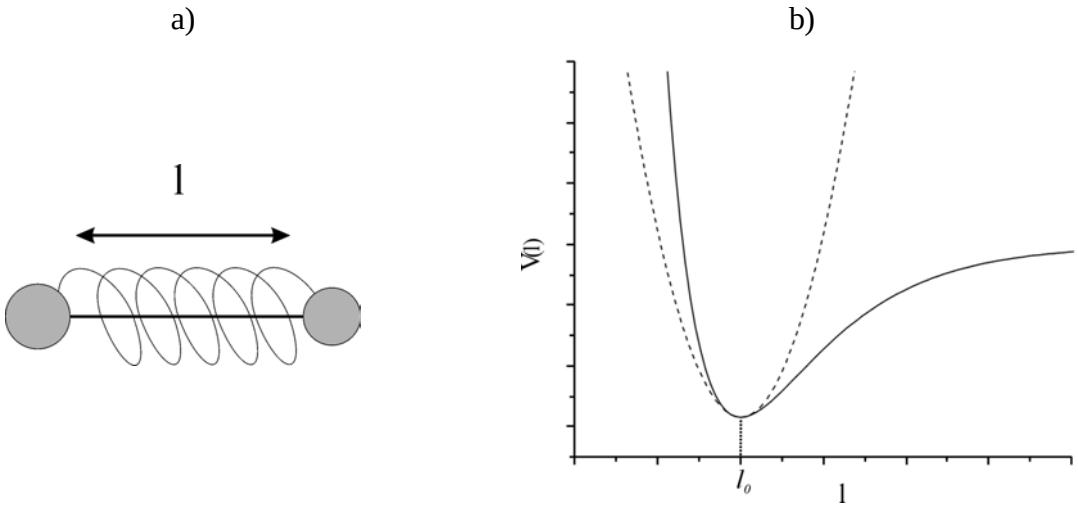

Figure 2.2.1: *a) In most force field implementations bonds between atoms are approximated by harmonic springs. B) In vicinity of the equilibrium value* l_0 *the harmonic potential (dashed line) is a good approximation for the more exact Morse curve (straight line).*

Both, equilibrium bond length l_0 and force constant k are specific for each pair of bonded atoms, *i.e.*, they depend on the chemical type of the constituent atoms. Values for force constants and bond lengths can be evaluated from experimental data such as infrared stretching frequencies high-resolution crystal structures, microwave spectroscopy data or theoretically from quantum mechanical calculations.

Angle Bending

The deviation of angles from their equilibrium values is also described using a harmonic potential:

$$V_{angle}(\theta) = \frac{k}{2}(\theta - \theta_0)^2 \qquad (2.2.7)$$

The contribution of each angle is characterized by a force constant k and an equilibrium value θ_0. Vibrational motions involving angle bending normally occur at lower frequencies than those of typical bond vibrations, less energy is required to distort an angle from its equilibrium value than to stretch a bond. This fact is reflected in the smaller force constants used for angle terms compared to those of bond terms in most force field implementations. The definition of angle bending terms is illustrated in Figure 2.2.2.

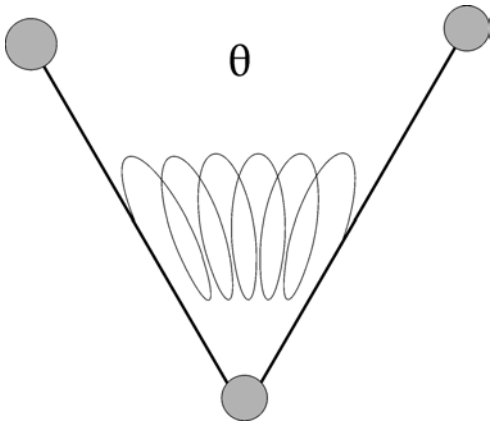

Figure 2.2.2: *Angle bending terms are three atom terms characterized by a harmonic potential dependent on the angle between the three atoms.*

Torsional Terms

The torsional angle terms in the potential energy function model the presence of effective barriers for the rotation around chemical bonds, these are due to steric interactions between atoms or groups of atoms separated by three covalent bonds (1,4-interactions). The proper modelling of barriers of rotation around chemical bonds is fundamental for the reproduction of structural properties of molecules and of conformational transitions. The potential is periodic and often expressed by a cosine function:

$$V_{torsion}(\phi) = K_{\phi}(1 - \cos(n\phi - \phi_0)) \tag{2.2.8}$$

where K_ϕ represents the barrier height, ϕ is the torsion angle between the 1,4-pair, ϕ_0 is an offset which defines the angular position of the first minimum in the potential, and n the multiplicity which gives the number of minima in the function as the bond is rotated through 360°. In most force fields two or more torsion terms may be assigned to the same 1,4-pair with different values of K, n, and ϕ, in the energy evaluation these terms are summed together. This allows the reproduction of very complex shapes for the rotational energy barrier. An example of this is illustrated in Figure 2.2.3 along with the definition of the torsion angle ϕ.

a) b)

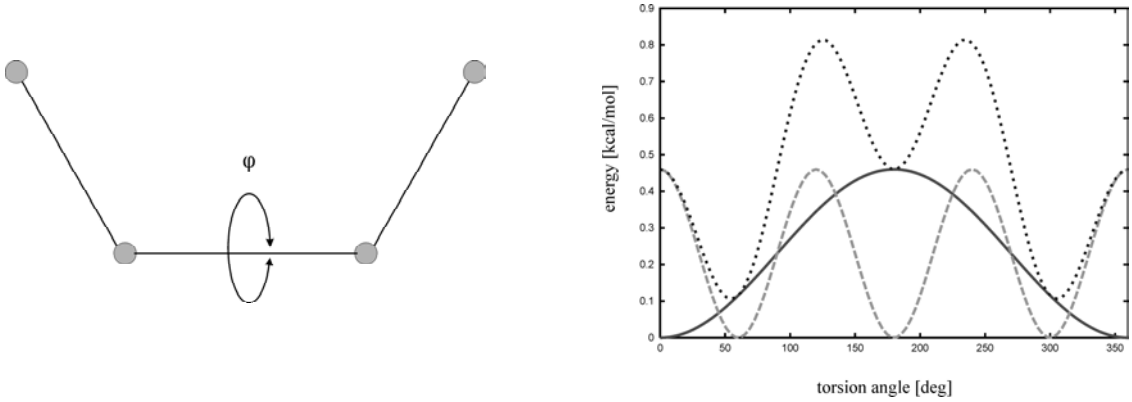

Figure 2.2.3: *a) Definition of the torsion angle ϕ. Torsion terms are four atom terms. b) Two or more torsion terms (continuous and dashed lines) may be applied to the same torsion angle to generate a more complex energy profile (dotted line).*

Improper Torsions

Additionally, so called "improper" torsion energy terms are introduced in the force field to obtain certain special geometries for different hybridisation states of groups of atoms such as the planarity of sp^2-hybridized carbon atoms. These are 4-atom terms as for normal torsions, the potential energy is a harmonic function of the angle ϕ between two planes defined as illustrated in Figure 2.2.4.

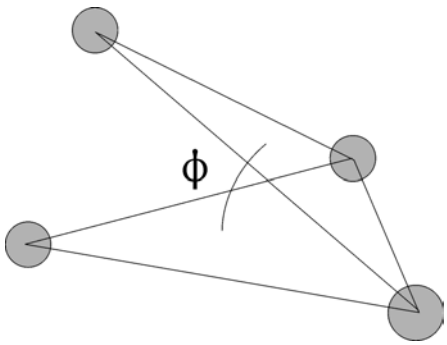

Figure 2.2.4: *Improper torsion angle terms are also four atom terms. The potential is generally defined as a harmonic function of the angle ϕ between two planes as depicted above.*

Non-Bonded Interactions

The energy term representing the contribution of non-bonded interactions has basically two components: the van der Waals interaction energy and the electrostatic interaction energy:

$$V_{non-bonded} = V_{vdWaals} + V_{electrostatic} \tag{2.2.9}$$

The calculation of these contributions in MD is the most time consuming part because they contain long-range interactions of the atoms in the system. The interaction energy of one atom and the rest of the system is calculated as a sum of pair-wise (atom to atom) interactions. The simultaneous interaction of three or more atoms is not calculated (many body problem). Also certain polarization effects are not explicitly included in the force field.

The van der Waals Interactions

The van der Waals interaction between two atoms arises from a balance between repulsive and attractive forces. The attractive forces are long-range forces whereas the repulsive forces act at short distances. The attractive contribution is due to dispersive forces arising from local electron fluctuations, which generate instantaneous dipoles in an atom or a molecule. An instantaneous dipole in a molecule can in turn induce a dipole in neighbouring atoms, giving rise to an attractive inductive effect. The short-range repulsive contributions are due to the electron-electron interactions. These are often referred to as exchange forces. The attractive interactions are longer ranged than the repulsive ones but as the distance becomes shorter the repulsive interaction becomes dominant. The van der Waals interaction is most often modelled using the Lennard-Jones 6-12 potential:

$$V(r) = 4\varepsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \tag{2.2.10}$$

The Lennard-Jones potential contains only two adjustable parameters: The collision diameter σ (the separation for which the energy is zero) and the well depth ε. It is characterized by an attractive part that varies as r^{-6} and a repulsive part that varies as r^{-12}. Although this functional form is used in most force fields, there is plenty of evidence that r^{-12} is a poor representation of the repulsive potential, an exponential form, $e^{-r/r^{\circ}}$ would be a much better approximation. An exponential function is more faithful to the exponential decay of atomic wavefunctions and of their overlap at large distances, which is responsible for repulsion. To use the Lennard-Jones potential in MD simulations one would have to calculate all the pair-wise interactions

of all the atoms in the system because the potential decays to zero only at infinite distance. At a distance $r = 2.5\sigma$ the potential has only 1% of its value at $r = \sigma$. This reflects the r^{-6} distance dependence of the dispersion interaction. In order to reduce the number of interaction terms, and thus the calculation time, the Lennard-Jones potential is often truncated. This can be done by defining an appropriate cut-off distance and calculating the pair-wise interactions only for the atoms lying within this distance. All van der Waals interactions of atoms beyond the cut-off distance are set to zero. The cut-off criterion for the potential energy function can be realized in two different ways. One way is to simply truncate the potential at the cut-off distance: at $r \geq r_c$ the potential is set to zero and interactions beyond that distance are neglected. A problem of this method is the discontinuity introduced in the force at the cut-off distance. At the cut-off distance, the force will have a finite value which drops suddenly to zero just beyond that cut-off. An alternative method is the use of a switching function. A switching function tapers the interaction potential over a predefined range of distances. The potential takes its usual value up to the first cut-off and is then switched to zero between the first and the second cut-off. This model suffers from strong forces in the switching region due to the rapid change in the derivative of the potential energy function. The Lennard-Jones potential and the two cut-off methods are illustrated in Figure 2.2.5.

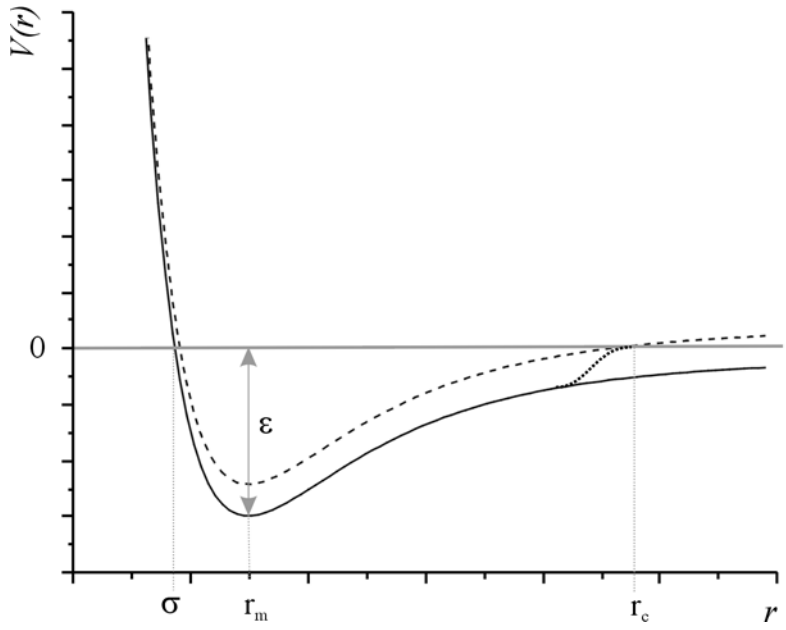

Figure 2.2.5: *The Lennard-Jones potential (straight line). A shifted (dashed line) or a switched (dotted line) function can be used to realize a cut-off criterion: beyond the cut-off distance* r_c *the potential is zero. The parameters* σ *and* ε *denote the collision diameter and the depth of the well of the potential, respectively.* r_m *is the equilibrium distance of the van der Waals interaction.*

Electrostatic Interactions

Accurate reproduction of the electrostatic properties of a molecule is of fundamental importance in force field development. Electrostatic interactions act at longer range than van der Waals interactions. Truncation schemes must be applied carefully, and in some cases should be entirely avoided. A common approach to modelling these interactions is to distribute a number of fractional point charges throughout the molecule. These sets of charges are designed to reproduce the electrostatic properties of the molecule. In most cases the charges are restricted to the nuclear centres, these are referred to as partial atomic charges. The electrostatic interaction between two molecules (or between different parts of the same molecule) is then calculated as a sum of interactions between pairs of point charges, using Coulomb's law:

$$V_{electrostatic} = \sum_i \sum_j \frac{q_i q_j}{4\pi\varepsilon_0 r_{ij}} \tag{2.2.11}$$

Coulomb interactions decay as r^{-1}. Because of this the use of cut-off methods in most cases is not appropriate and can cause significant errors. A variety of methods have been developed to handle long-range electrostatics and make their computation more efficient [Leach 1996]. The method used in this work to calculate the electrostatics is the Particle-Mesh Ewald method[Essmann, et al. 1995].

A major problem in correctly evaluating the electrostatic contribution to the potential energy is that, if one uses periodic boundary conditions a charged particle interacts with all the other charges in the simulation box and with all of their images in an infinite array of periodic cells. The system thus contains an infinite number of charges. The total electrostatic potential energy, $\phi(\mathbf{r}_i)$ at the position $\mathbf{r}_i$ of any charge i of the primary cell is given by the finite difference of two infinite, diverging series:

$$\phi(\mathbf{r}_i) = q_i \sum_{j+}^{\infty} \frac{q_{j+}}{\left|\mathbf{r}_i - \mathbf{r}_{j+}\right|} + q_i \sum_{j-}^{\infty} \frac{q_{j-}}{\left|\mathbf{r}_i - \mathbf{r}_{j-}\right|} \tag{2.2.12}$$

The Ewald summation approach, first proposed by Ewald [Ewald 1921], solves this problem by splitting the potential into two well behaved and rapidly converging parts, the first being represented in real space and the second in reciprocal space. A simple illustration of the method can be made for a one-dimensional ion lattice with a charge distribution as shown in Figure 2.2.6, the extension to three dimensions is trivial. The original lattice is augmented by a set of Gaussian charge distributions of opposite sign to form an auxiliary lattice (lattice 1, in Figure 2.2.6):

$$\rho(\mathbf{r}) = -q_j \left(\frac{\eta^2}{\pi} \right)^{\frac{3}{2}} e^{-\eta^2 (\mathbf{r}-\mathbf{r}_j)^2} \tag{2.2.13}$$

A further lattice (lattice 2, in Figure 2.2.6) is then introduced to compensate the additional Gaussian distributions, such that "lattice 1 + lattice 2 = original lattice". The contributions of the two lattices to the potential energy are computed seperatly. The effect of lattice 1 at large distance, due to the compensating effect of the Gaussian distributions, tends rapidly to zero. The narrower the Gaussian distributions, the more they resemble the delta functions of the original charge distribution and the more the compensating effect of the Gaussian distributions is efficient *i.e.*, the larger the value of η the faster the series will converge. The effect of lattice 1 is therefore best computed in real space, where the series will rapidly converge. For lattice 2 the potential sum is evaluated in k-space. When the Gaussians are broad *i.e.*, when η is small, a smaller number of Fourier components is required. By suitably adjusting the parameter η, optimal convergence of both series can be achieved.

Ewald summations are computationally quite expensive to implement. Nonetheless the Ewald method is the most correct way to accurately include all the effects of long-range forces in computer simulations.

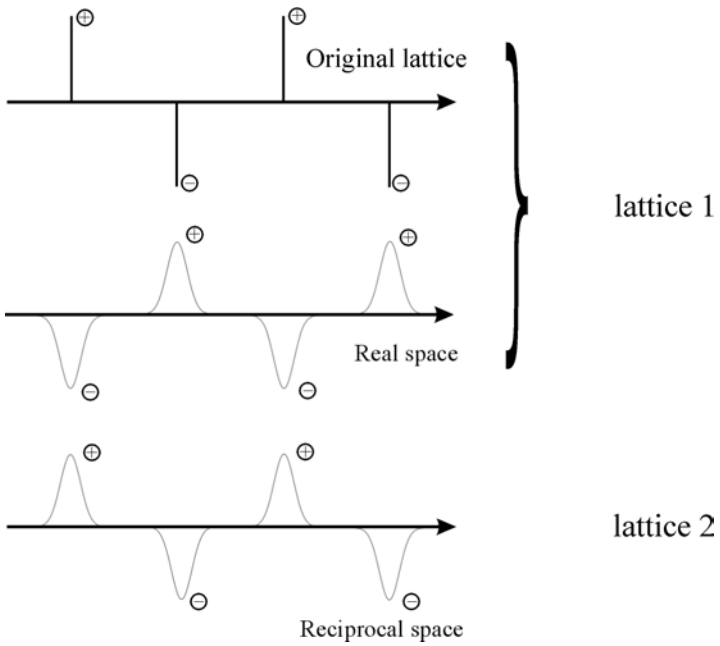

Figure 2.2.6: *The Ewald summation method: the original set of charges is summed to a Gaussian distribution (calculated in real space) and to a cancelling charge distribution (calculated in the reciprocal space).*

In the Ewald procedure, computation of the reciprocal energy from lattice 2 is the most expensive part in the evaluation of long-range electrostatic contributions. The Particle-Mesh Ewald method (PME) significantly speeds up the process. The direct sum in real space is evaluated explicitly using cut-offs while the reciprocal sum is approximated using fast Fourier Transforms (FFT) with convolutions on a grid in which charges are interpolated to the grid points. In addition, PME does not interpolate but rather evaluates the forces by analytically differentiating the energies, thus reducing memory requirements substantially.

In summary, the potential energy in a molecular mechanics force field is calculated as the sum of all the bonded and the non-bonded interactions over all the atoms of the system:

$$V(r) = \sum V_{bond} + \sum V_{angle} + \sum V_{torsion} + \sum V_{vdWaals} + \sum V_{electrostatics} \qquad (2.2.14)$$

Molecular mechanics force fields provide a reasonably good compromise between accuracy and computational efficiency. In some cases molecular mechanics can provide answers that are as accurate as high-level quantum mechanical calculations, in only a fraction of the computer time.

The issue of deriving a good set of force field parameters is of utmost importance. The reliability of the results of molecular mechanics computations is strongly dependent on the parameter sets used in the force field. The CHARMM force field [Foloppe and MacKerell Jr. 2000, MacKerell Jr. and Banavali 2000, MacKerell Jr., et al. 1998] used in this work is equipped with well-tested parameters for the simulation of biomolecules. The matter of force field parameterisation is discussed in detail in the next chapter where a new general method for deriving force field parameters from *ab initio* quantum calculations is presented and parameters for the CHARMM force field are derived for the two fluorescent dyes, rhodamine 6G and MR121.

2.2.2. Energy Minimization

For all except the very simplest systems the potential energy is a complicated, multidimensional function of the 3N Cartesian coordinates of the system. The minima on this multidimensional energy surface are of special interest for both quantum mechanics and molecular mechanics calculations. Indeed, some molecular mechanics force field parameters

can be obtained by fitting molecular properties calculated by quantum chemical methods in the 'ground state' (*i.e.*, in the global potential energy minimum). Minimization algorithms are used to identify geometries of the system corresponding to minima of the potential energy surface. The number of minima can be very large. This is true especially for biomolecular systems consisting of thousands of atoms and a huge number of degrees of freedom. Minimization algorithms are also very important for MD simulations, where it is important to start from well-minimized structures to avoid unwanted high energy interactions.

Given a function f which depends on the variables $\{x_1, x_2, \ldots, x_n\}$, a minimum of f is defined as a point where the first derivative of the function with respect to each of the variables is zero and the second derivatives are all positive:

$$\frac{\partial f}{\partial x_i} = 0\,;\ \frac{\partial^2 f}{\partial x_i^2} > 0 \qquad \{i = 1, \ldots, n\} \tag{2.2.15}$$

For analytical functions, the minima can be found using standard calculus methods. However, this is not possible for systems such as biomolecules where the shape of the potential energy landscape cannot be expressed by a simple analytical formula. In these cases, minima are located using numerical methods that gradually change the coordinates to produce configurations with lower and lower energies until the minimum is reached. The most common minimization algorithms use derivatives of the energy in respect to the coordinates to predict the location of the closest minimum. The two major issues to be considered in the development of a minimization algorithm are the speed of convergence of the method and the memory requirements. No single algorithm has yet been proved to be the best for all problems, rather a combination of different algorithms is commonly used. Most minimization algorithms can only go downhill on the energy surface. Thus, they can only locate the minimum that is nearest (in a downhill sense) to the starting point. To locate more than one minimum or the global energy minimum with these algorithms many minimizations from different starting points are required. For a true exploration of the potential energy landscape, energy minimization alone is not sufficient.

The energy minimization procedures used in this work are a combination of two derivative minimization methods: the steepest descent and the Newton-Raphson method. When discussing derivative methods it is useful to write the potential energy function V as a Taylor series expansion about the point $\mathbf{r_0}$:

$$V(\mathbf{r}) = V(\mathbf{r}_0) + (\mathbf{r} - \mathbf{r}_0)V'(\mathbf{r}_0) + (\mathbf{r} - \mathbf{r}_0)^{\mathrm{T}} \frac{1}{2} V''(\mathbf{r}_0)(\mathbf{r} - \mathbf{r}_0) + \ldots \tag{2.2.16}$$

where $\mathbf{r_0}$ and $\mathbf{r}$ are *3N* dimensional vectors. $V'(\mathbf{r_0})$ is the vector of the partial derivatives of V with respect to each coordinate and $V''(\mathbf{r_0})$ is the Hessian matrix, it contains the partial second derivatives and is of *3N* x *3N* dimensions. In the vicinity of a minimum $V(\mathbf{r})$ can be approximated by truncating the series at the second order. This approximation is used also for normal mode calculations.

The direction of the first derivative of the energy indicates where the minimum lies and the magnitude of the gradient indicates the steepness of the local slope. By moving each atom in response to the force (negative gradient) acting on it the energy of the system is lowered. Second derivatives indicate the curvature of the function, information that can be used to predict where the function will change direction. The steepest descent method is a first-order minimization algorithm. It gradually changes the coordinates of the atoms as it moves the system closer and closer to the minimum point. The direction of movement on the potential energy surface is parallel to the net force, which in geographical analogy corresponds to walking straight downhill along the steepest descent. The minimum is located iteratively using a line search algorithm. The method of steepest descent is very robust, meaning that the minimum is found even when the starting structure is far away from the minimum; so it is a good method to start with. On the other hand, the close-range approach to the minimum may require quite a large number of iterations and, due the inflexible determination of the search direction unwanted oscillations around the minimum may occur.

The Newton-Raphson method is a second-order derivative method. This method makes use of the inverse Hessian matrix for energy minimization. For a purely harmonic function the method finds the minimum in only one step. As the real potential energy function contains anharmonic terms, again an iterative cycle has to be used. But the number of iterative steps required to reach a minimum is generally less in comparison to first-order derivative methods. The efficiency of the method increases the better the potential energy function can be approximated to a harmonic potential well. For most applications a combination of steepest descent and Newton-Raphson minimization is favoured. With the steepest descent method the structure can be efficiently brought close to the minimum and with subsequent Newton-Raphson minimization the minimum is located within a few steps.

2.2.3. Normal Mode Analysis

One of the most important applications of potential energy minimizations is normal mode analysis. In both molecular mechanics and quantum mechanics normal modes are useful

because they provide an orthonormal basis of the configurational space of a system in which collective motions of the atoms can be represented. That is, in a coupled system the normal modes describe independent motions of the whole system that can be individually excited. A non-linear molecule with *N* atoms has *3N-6* vibrational modes and *6* 'trivial' modes representing translation and rotation of the system as a whole. The frequencies of the normal modes together with the displacements of the individual atoms may be calculated from a molecular mechanics force field or from the wavefunction using the Hessian matrix of second derivatives ($V''(\mathbf{r_0})$ in Eq. 2.2.16) calculated in a minimum of the potential energy surface. In fact a normal mode calculation is based on the assumption that the energy surface is quadratic in the vicinity of the energy minimum.

The starting point for a normal mode analysis are Newton's classical equations of motion for a system of coupled harmonic oscillators (harmonic approximation to the potential energy). These can be written in matrix notation as:

$$\mathbf{M}\frac{d^2}{dt^2}\left(\Delta\mathbf{r}\right) = -V''(\mathbf{r}_0)\Delta\mathbf{r} \tag{2.2.17}$$

where **M** is a *3N* diagonal matrix containing the masses of the nuclei and $\Delta\mathbf{r} = \mathbf{r} - \mathbf{r_0}$, is the vector of the displacements of the nuclear coordinates. The solution will be of the form:

$$\Delta\mathbf{r} = \mathbf{Y}e^{i\left(\omega t+\phi\right)} \tag{2.2.18}$$

where ϕ is an arbitrary phase shift and **Y** is a vector and ω is a scalar. Substituting Eq. 2.2.18 in Eq. 2.2.17 it follows that:

$$\left[\mathbf{M}\omega^2 - V''(\mathbf{r}_0)\right]\mathbf{Y} = 0 \tag{2.2.19}$$

This homogeneous system of linear equations must be solved for **Y**. In general the only way to get non-zero solutions is to make the matrix $\left[\mathbf{M}\omega^2 - V''(\mathbf{r}_0)\right]$ singular, *i.e.*:

$$\det\left(\mathbf{M}\omega^2 - V''(\mathbf{r}_0)\right) = \det\left(\omega^2 - \mathbf{M}^{-1/2}V''(\mathbf{r}_0)\mathbf{M}^{-1/2}\right)\det(\mathbf{M}) = 0 \tag{2.2.20}$$

Since $\det(\mathbf{M}) \neq 0$, the second equality is simply the characteristic equation associated with the eigenvalue problem:

$$\overline{\mathbf{K}}\mathbf{u} = \lambda\mathbf{u} \tag{2.2.21}$$

with $\lambda \equiv \omega^2$ giving the square of the frequencies of the normal modes, **u** the respective eigenvectors and $\overline{\mathbf{K}}$ is given by $\overline{\mathbf{K}} = \mathrm{M}^{-1/2} V''(\mathbf{r}_0) \mathrm{M}^{-1/2}$.

The main limit of normal mode analysis lies in the rather drastic approximation of the potential energy using a harmonic function of the coordinates within a local minimum of the potential energy landscape. In spite of the neglect of anharmonic contributions to the potential, normal mode descriptions of protein dynamics and fluctuations have proven to provide useful results concerning the internal motions of proteins[Brooks and Karplus 1983]. In contrast to normal modes, molecular dynamics calculations allow exploration of the full potential energy landscape. Thus MD calculations implicitly include anharmonic motions of the. These are discussed in detail in the following section.

2.2.4. Molecular Dynamics Simulations

The statement of the problem addressed by MD simulations is at a first glance very simple: an MD algorithm simply calculates the classical time evolution of the system. This is done by integrating Newton's laws of motion (or other, equivalent formulations):

$$F_i = m_i a_i \qquad \{i = 1, \ldots, 3N\} \tag{2.2.22}$$

$$\frac{F_{x_i}}{m_i} = \frac{d^2 x_i}{dt^2} \qquad \{i = 1, \ldots, 3N\} \tag{2.2.23}$$

F_i is the force exerted on particle i, m_i is the mass of the particle, and a_i is its acceleration. For a mechanical system the force can also be expressed as the gradient of the potential energy:

$$F_i = -\nabla_i V \qquad \{i = 1, \ldots, 3N\} \tag{2.2.24}$$

The difficulty lies in the fact that for "realistic" systems with a large number of atoms, analytical solutions cannot be derived and the equations have to be solved numerically. The discontinuous nature of computer based numerical calculations, in contrast to the continuous nature of realistic potentials, requires the equations of motion to be integrated by breaking the calculation into a series of very short time intervals (typically 1 or 2 femtoseconds) for dynamics at atomic detail. At each step, the forces on the atoms according to the molecular mechanics force field are computed and combined with the current positions and velocities to generate new positions and velocities at the following time step. The total force on each atom

at a time *t* is calculated as a vector sum of its interaction with other atoms in a pair-wise additive model. During the integration time step the forces are assumed to be constant. The atoms are then moved to the new positions, an updated set of forces is computed, and so on. In this way an MD simulation generates a trajectory that describes how the system moves through phase space as a function of time.

Several computationally efficient algorithms exist for integrating the equations of motion. All of them use Taylor series expansions of the positions and dynamic properties:

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \frac{1}{6}\delta t^3 b(t) + \ldots \tag{2.2.25}$$

where *v* is the velocity (the first derivative of the positions with respect to the time), *a* is the acceleration (the second derivative), *b* is the third derivative, and so on. One of the most commonly used algorithms is the Verlet algorithm [Verlet 1967]. It uses the positions and accelerations at time *t*, and the positions from the previous step to calculate the new positions.

$$r(t+\delta t) = r(t) + \delta t v(t) + \frac{1}{2}\delta t^2 a(t) + \ldots \tag{2.2.26}$$

$$r(t-\delta t) = r(t) - \delta t v(t) + \frac{1}{2}\delta t^2 a(t) - \ldots \tag{2.2.27}$$

Adding these two equations gives:

$$r(t+\delta t) = 2r(t) - r(t-\delta t) + \delta t^2 a(t) \tag{2.2.28}$$

The velocities do not explicitly appear in the Verlet integration but they can be calculated in a variety of ways. A simple approach is to divide the differences in positions at times $t+\delta t$ and $t-\delta t$ by $2\delta t$; or to estimate the velocities at the half-step. A variation on the Verlet algorithm is the leap-frog algorithm [Hockney 1970]. It uses the following relationships:

$$r(t+\delta t) = r(t) + \delta t v(t+\frac{1}{2}\delta t) \tag{2.2.29}$$

$$v(t+\frac{1}{2}\delta t) = v(t-\frac{1}{2}\delta t) + \delta t a(t) \tag{2.2.30}$$

In the leap-frog algorithm, the velocities $v(t+1/2\delta t)$ are first calculated, from the velocities at time $t-1/2\delta t$ and the accelerations at time *t*. The positions $r(t+\delta t)$ are then deduced from the velocities just calculated together with the positions at time *t*. The velocities at time *t* can be calculated as the average:

$$v(t) = \frac{1}{2}\left[v(t + \frac{1}{2}\delta t) + v(t - \frac{1}{2}\delta t) \right] \quad (2.2.31)$$

The advantage over the standard Verlet algorithm is that the velocities are explicitly included and that it does not require the calculation of differences of large numbers.

After running the simulation, a trajectory is obtained containing the changes of the atomic coordinates of the system with time. From this trajectory dynamic parameters can be extracted and evaluated. Furthermore, stable structures and transition states can be analysed and visualized.

2.3. Fluorescence

2.3.1. Absorption and Emission of Light by Molecules

The energy levels of dislocated electrons in molecules fall in the range of 1-5 eV, whereas vibrational energies of the nuclei are at least 10 times lower (0.1 eV) and rotational energies are even lower (0.001 eV). Absorption of a photon by a molecule induces a transition of the molecule to a higher energetic state. Once a molecule has absorbed energy in the form of electromagnetic radiation, there are a number of routes by which it can return to the ground state (the statistically most common energy state for room temperature chemical species). The well-known Bohr frequency condition must be fulfilled for absorption and emission of photons *i.e.*, the energy difference between the ground state and the excited state is given by: $\Delta E = h\nu$, where h is the Plank constant and ν is the frequency of the absorbed/emitted radiation. Using the energy values above, the frequencies corresponding to electronic excitation of molecules fall between the near ultraviolet and the near infrared regions of the spectrum (240-1200 nm). The absorbing molecule can be regarded as an oscillating electric dipole (due to changes in the instantaneous charge distribution of the molecule) interacting with the electrical field component of the electromagnetic wave, which is perpendicular to the wave's propagation direction. The probability of light absorption by a molecule is determined not only by the frequency of the radiation but also by the orientation of the transition dipole moment and by the electronic spin state of the molecule. The transition dipole moment reflects the difference in the molecular charge distribution due to the transition of an electron from the ground state into the excited energy state upon absorption of a photon. Fluorescent

organic dyes are characterized by a strong absorption band in the visible region of the electromagnetic spectrum (in the range between 400 and 800 nm).

2.3.2. Fluorescence and Quenching of Dyes

The transitions involved in the absorption of visible light by fluorescent dyes are those from the electronic ground state S_0 to the first excited singlet state S_1. Relaxation of the system back to the ground state may occur by emission of a fluorescence photon or by non-radiative decay processes in which the excess energy is dissipated in the form of heat *via* vibrational, rotational and translational modes of the system. Another possible relaxation mechanism involves chemical reactions occurring in the excited state, this case will not be discussed here. Electronic transitions of a molecular system upon absorption or emission of photons occur on timescales which are much faster (about 10^{-15} s) than the response of the nuclei (Franck-Condon principle). For this reason, the absorption and emission spectra of most molecules exhibit broad band structures, indeed the rotationally and vibrationally excited states, which depend only on the nuclear degrees of freedom have longer decay times than the electronically excited states, therefore electronic transitions may occur between many different vibrational and rotational states. This fact is illustrated in the absorption and emission spectra of the fluorescent probe R6G (Figure 2.3.1.).

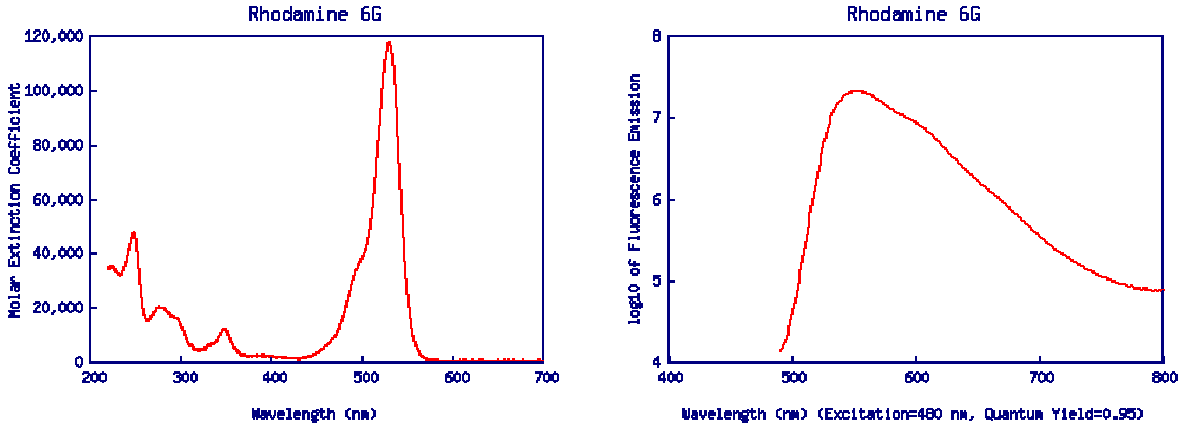

Figure 2.3.1.: *Absorption (left) and emission (right) spectra of R6G. Broad band structures are due to transitions between vibrationally excited states of S_1 and S_0.*

At room temperature, molecules generally occupy the electronic and vibrational ground-state. Upon light absorption the molecule finds itself in a vibrationally excited state of one of the electronically excited levels. This electronic state is generally the first excited singlet state S_1. Excitations into S_2 and higher singlet states also occur, but these generally decay rapidly (10^{-11}-10^{-13} s) in a non-radiative fashion into the S_1 state. After excitation into a vibrationally excited state of S_1, the molecule relaxes to the vibrational ground-state by collisions with surrounding molecules in a time scale of the order of 10^{-12} s. Subsequently, different radiative and non-radiative relaxation processes are possible, these are illustrated in the schematic Jablonski diagram of Figure 2.3.2.

From the vibrational ground state of the S_1 state, radiative relaxation to vibrationally excited S_0 levels can occur by spontaneous emission of a fluorescence photon. As vibrational relaxation of the electronically excited state takes place prior to photon emission, and as the accompanied electronic relaxation goes to vibrationally excited states, the fluorescence photon has a lower energy compared to the absorbed photon, *i.e.* the emission spectrum exhibits a bathochromic shift - the so-called Stokes shift respect to the emission spectrum. The remaining energy in the vibrationally excited S_0 state is again dissipated in the form of heat by collisions with the surrounding medium. The Stokes shift is crucial for the high sensitivity of fluorescence spectroscopy because the fluorescence light can be separated efficiently from Rayleigh and Raman scattering, using optical filtering systems.

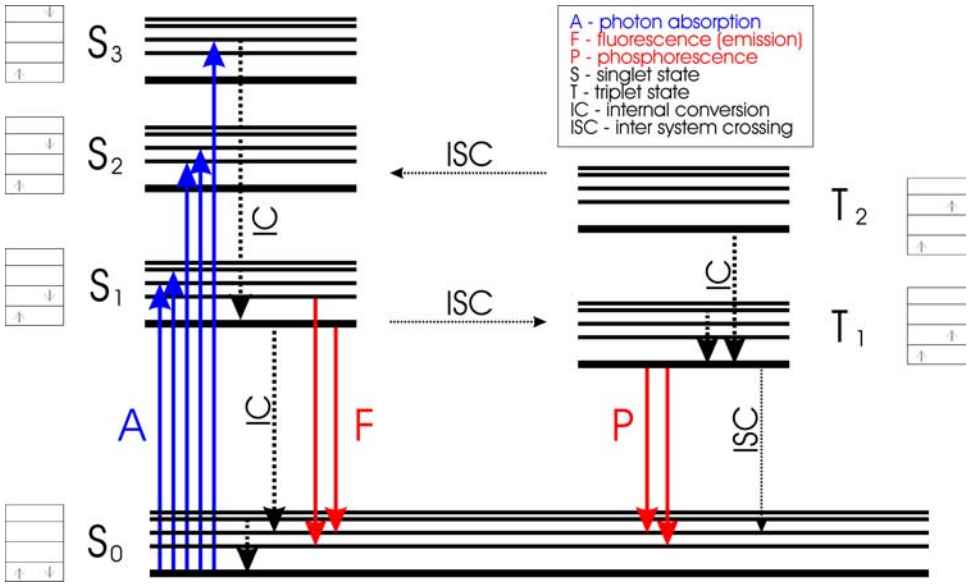

Figure 2.3.2.: *Schematic Jablonski-diagram illustrating different radiative and non-radiative processes. Full lines show processes of light emission (red) or absorption (blue), dotted lines show non-radiative processes. Internal conversion processes (IC) occur between energy levels with the same spin multiplicity, Intersystem crossing (ISC) requires spin flip.*

The non-radiative decay from the lowest vibrational state of electronically excited states to vibrationally and rotationally excited electronic ground-states is called internal conversion (IC) and is mostly responsible for the loss of fluorescence efficiency of organic dyes. Depending on the molecular structure of the dye and the properties of the solvent, the rate of relaxation by IC can vary by many orders of magnitude. Besides the non-radiative decay directly to the ground-state, a molecule in the S_1 state may enter the system of triplet states and relax to the lowest triplet level T_1, this process is called intersystem crossing (ISC). Transitions involving ISC are spin-forbidden, but have a non-zero probability due to spin-orbit coupling. $S_1 \rightarrow T_1$ transitions are energetically always possible because the first excited triplet state always has lower energy compared to the first excited singlet state. From the vibronic ground state of the first excited triplet state T_1 the molecule can relax in a non-radiative fashion by ISC or by the emission of a phosphorescence photon. As this is again a spin forbidden process with low probability, the rate constant for phosphorescence is several orders of magnitude smaller than that of S_1 states and T_1 states have characteristic lifetimes of $\sim 10^{-6}$ - 10^{-3} s.

The preceding considerations concerned absorption and emission characteristics of a single molecule. The measured fluorescence lifetime τ for a statistical ensemble of molecules is related to the fraction of optically excited molecules $F(t)^*$ still present at a time t following excitation. If the excitation is considered to be essentially instantaneous and occurring at time $t = 0$, $F(t)^*$ can be described as follows:

$$\frac{d[F(t)^*]}{dt} = -(k_f + k_{IC} + k_{ISC})F(t)^* \tag{2.3.1}$$

where k_f, k_{IC} and k_{ISC} are the rate constants for fluorescence, internal conversion and intersystem crossing, respectively if one defines a quantity τ as:

$$\tau = \frac{1}{k} = \frac{1}{k_f + k_{IC} + k_{ISC}} \tag{2.3.2}$$

Integration of equation 2.1.4. between time 0 and time t yields:

$$F(t)^* = F(t=0)^* \exp(-t/\tau) \tag{2.3.3}$$

The fraction of excited molecules $F(t)^*$ at time t decays exponentially with time. At any given time $F(t)^*$ is proportional to the measured fluorescence intensity $I(t)$ so that:

$$I(t) = I(t=0)\exp(-t/\tau) \tag{2.3.4}$$

The fluorescence lifetime τ corresponds to the time after which the fluorescence intensity has decayed to 1/*e* times its initial value. The lifetime τ can be determined by time-resolved fluorescence intensity measurements. Fluorescence decays of commonly used dyes generally occur on nanosecond timescales (10^{-9} s).

Often, and especially in complex biological systems, the fluorescence lifetimes cannot be described by a mono-exponential decays. In this case, the decay is fitted as a sum of exponentials:

$$I(t) = \sum_i a_i \exp(-t/\tau_i) \tag{2.3.5}$$

$$\sum_i a_i = 1 \tag{2.3.6}$$

where a_i represents the amplitude of the *i*-th lifetime component in the excited state.

In presence of non-radiative decay processes, the measured fluorescence lifetime τ is smaller than the decay time τ_0^n in absence of non-radiative processes.

The fluorescence quantum yield Φ_f is defined as the ratio between the number of emitted fluorescence photons, N_e, and the number of absorbed photons, N_a:

$$\Phi_f = \frac{N_e}{N_a} \tag{2.3.7}$$

The fluorescence quantum yield is a measure for the fluorescence efficiency of the molecule and reflects the presence of non-radiative processes. Φ_f can also be expressed in terms of lifetimes as:

$$\Phi_f = \frac{\tau}{\tau_0^n} \tag{2.3.8}$$

Other external factors such as interactions with surrounding molecules can contribute to reducing the fluorescence quantum yield of a dye. These are generally referred to as quenching factors.

Dynamic and Static Fluorescence Quenching

Often, after excitation interactions and reactions may occur during the time span in which a fluorescent dye remains in the electronically excited state may cause the system to decay non-radiatively to the ground state *i.e.,* interactions of the dye in the fluorescent state with

surrounding molecules may cause fluorescence quenching. In general, fluorescence quenching processes can be divided into two classes: dynamic and static quenching.

In the dynamic quenching model, a quencher molecule collides by diffusion with the fluorophore during the excited-state lifetime. An encounter complex is formed. Upon formation of the encounter complex, the fluorophore relaxes non-radiatively to the electronic ground-state. This type of dynamic quenching process is characterized by a reduction of both the measured fluorescence lifetime and the fluorescence intensity. Under the conditions that the concentration of quencher molecules [*Q*] is much higher than the concentration of fluorescent dye [*F**], [*Q*] can be assumed to be constant with time and the following expression for the temporal decay of the fluorescence intensity *I(t)* in presence of quencher molecules is valid:

$$I(t,[Q]) = I(t,[Q]=0)\exp\left(-(k_0 + k_q[Q])t\right) \tag{2.3.9}$$

The fluorescence lifetime in this case is given by:

$$\tau = \frac{1}{k_0 + k_q[Q]} \tag{2.3.10}$$

where k_0 is the sum over k_f, k_{IC}, and k_{ISC}. k_q is the rate constant of the bimolecular quenching process. The model of the bimolecular quenching process assumes the formation of an encounter complex with a certain minimum distance between dye and quencher. If the molecular quenching mechanism is fast compared to diffusion, the quenching process is diffusion-limited. In this case, the maximum quenching constant k_d can be calculated by the Smoluchowski equation:

$$k_d = 4\pi r_{DQ} D_{sum} N_A \tag{2.3.11}$$

The radius r_{DQ} is the sum of the radii of the dye molecule and the quencher, D_{sum} is the sum of the diffusion coefficient of the quencher and the dye molecule, and N_A is the Avogadro constant. The diffusion constant D of a molecule depends on the solvent viscosity η, the radius of the molecule r and the temperature T. The diffusion constant can be calculated by the Stokes-Einstein equation:

$$D = \frac{k_B T}{6\pi\eta r} \tag{2.3.12}$$

where k_B is the Boltzmann constant. Typical values of k_d for organic dyes and aromatic quenching molecules in aqueous solution are about 6 x 10^9 $M^{-1}s^{-1}$. In the case of a diffusion-

limited quenching (k_d=k_q) and a typical fluorescence lifetime τ of ~ 4 ns for a rhodamine dye, significant decreases in the fluorescence lifetime can be observed for quencher concentrations of about 10-100 mM. Dynamic fluorescence quenching is described by the Stern-Volmer equation:

$$\frac{I_0}{I} = \frac{\tau_0}{\tau} = 1 + k_q \tau_0 [Q] = 1 + K_D [Q] \tag{2.3.13}$$

where K_D is the dynamic Stern-Volmer constant, I_0 and τ_0 are the measured fluorescence intensity and lifetime in absence of quencher. For determination of the dynamic quenching constant k_q, the fluorescence intensity or the fluorescence lifetime are measured at different quencher concentrations. For a purely dynamic quenching process, the plot of I_0/I or τ_0/τ *vs* the quencher concentration is a straight line with the dynamic Stern-Volmer constant K_D as slope. In many cases, the plot shows a concave or a convex deviation from the straight-line behaviour, signaling an additional static quenching component.

Static quenching is marked by the formation of a ground-state complexes between dye and quencher in equilibrium. This equilibrium between free compounds and complex is also assumed for the fluorophore in the excited state. The ground-state complexes are non- or only weakly fluorescent. Ground state complexes do not influence the measured fluorescence lifetime but only the fluorescence intensity (*i.e.* the relative fluorescence quantum yield), this is due to the fact that the formation of ground-state complexes reduces the concentration of free fluorescent dye. In case of a purely static quenching process with non-fluorescent 1:1 ground-state complexes, the static Stern-Volmer constant K_S corresponds to the thermodynamic equilibrium constant for the complexation:

$$K_S = \frac{[FQ]}{[F][Q]} \tag{2.3.14}$$

$$\frac{I_0}{I} = 1 + K_S [Q] \tag{2.3.15}$$

The combination of steady-state and time-resolved fluorescence measurements can be used to distinguish between dynamic and static quenching.

The Stern-Volmer analysis reveals information about the kinetics and thermodynamics of the interaction between fluorophore and quencher, *i.e.* about the formation of encounter complexes and non-fluorescent ground-state complexes. However, no information about the actual fluorescence quenching mechanism can be extracted. For the interpretation of fluorescence spectroscopic data, especially in bioanalytic applications, it is crucial to

understand how changes in the microenvironment and interactions with surrounding molecular compounds influence the dye in the electronically excited state and thus its fluorescence properties.

References

Alder, B. J. and T. E. Wainwright (1959). "Studies in Molecular Dynamics. I. General Method." J Chem Phys 31): 459-66.

Alder, B. J. and T. E. Wainwright (1957). "Phase Transition for a Hard Sphere System" J Chem Phys 27): 1208.

Rahman, A. (1964). "Correlations in the Motion of Atoms in Liquid Argon" Phys Rev A 136): 405-11.

Stillinger, F. H. and A. Rahman (1974). "Improved Simulation of Liquid Water by Molecular Dyanamics" J Chem Phys 60): 1545-57.

Weber, T. A. and F. H. Stillinger (1984). "The effect of density on the inherent structure in liquids" J Chem Phys **6**(80): 2742-6.

Stillinger, F. H. and T. A. Weber (1988). "Molecular dynamics simulation for chemically reactive substances. Fluorine" J Chem Phys **8**(88): 5123-33.

McCammon, J. A., B. R. Gelin and M. Karplus (1977). "Dynamics of Folded Proteins" Nature 267): 585-90.

Jorgensen, W., J. Chandrasekhar, J. Madura, R. Impey and M. Klein (1983). "Comparison of simple potential functions for simulating liquid water" J Chem Phys 79): 926-35.

Levinthal, C. (1968). "Are There Pathways for Protein Folding?" Journal de Chimie Physique 65): 44–5.

Duan, Y. and P. A. Kollman (1998). "Pathways to a protein folding intermediate observed in a 1-microsecond simulation in aqueous solution" Science **5389**(282): 740-4.

Schaefer, M., C. Bartels and M. Karplus (1998). "Solution conformations and thermodynamics of structured peptides: molecular dynamics simulation with an implicit solvation model" J Mol Biol **3**(284): 835-48.

Daura, X., B. Jaun, D. Seebach, W. F. Van Gunsteren and A. E. Mark (1998). "Reversible peptide folding in solution by molecular dynamics simulation" J Mol Biol **5**(280): 925-32.

Torrie, G. M. and J. P. Valleau (1977). "Nonphysical Sampling Distributions in Monte-Carlo Free Energy Estimation: Umbrella Sampling" J Comput Phys 23): 187-99.

Guo, Z., C. L. Brooks, 3rd and E. M. Boczko (1997). "Exploring the folding free energy surface of a three-helix bundle protein" Proc Natl Acad Sci U S A **19**(94): 10161-6.

Mu, Y. and G. Stock (2002). "Conformational Dynamics of Trialanine in Water: A Molecular Dynamics Study" J Phys Chem B **20**(106): 5294-301.

Leach, A. R. (1996). Molecular Modelling: Principles and Selected Applications. Edimburg Gate, Harlow, Pearson Education Ltd.

Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee and L. G. Pedersen (1995). "A smooth particle mesh Ewald method" J Chem Phys **19**(103): 8577-93.

Ewald, P. (1921). "Die Berechnung optischer und elektrostatischer Gitterpotentiale" Ann Phys 64): 253-87.

Foloppe, N. and A. D. MacKerell Jr. (2000). "All-Atom Empirical Force Field for Nucleic Acids: 1) Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data" J Comput Chem 21): 86-104.

MacKerell Jr., A. D. and N. Banavali (2000). "All-Atom Empirical Force Field for Nucleic Acids: 2) Application to Molecular Dynamics Simulations of DNA and RNA in Solution" J Comput Chem 21): 105-20.

MacKerell Jr., A. D., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. Reiher, W.E, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin

and M. Karplus (1998). "All-atom empirical potential for molecular modeling and dynamics Studies of proteins" J Phys Chem B 102): 3586-16.

Brooks, B. and M. Karplus (1983). "Harmonic dynamics of proteins: normal modes and fluctuations in bovine pancreatic trypsin inhibitor" Proc Natl Acad Sci U S A **21**(80): 6571-5.

Verlet, L. (1967). "Computer "Experiments" on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules" Physical Review 159): 98-103.

Hockney, R. W. (1970). "The Potential Calculation and Some Applications" Methods in Computational Physics 9): 136-211.

3. Molecular Mechanics Force Field Parameterisation of R6G and MR121

The first necessary step for MD simulations of fluorescent probes is to obtain reliable force field parameters of the dyes. Most "all atom" empirical force fields used in common MD packages (such as CHARMM [Brooks, et al. 1983]) are equipped with parameter sets for modelling biomolecules but none have been developed for commonly-used fluorescent labels. Moreover, for many dyes, suitable experimental data against which new parameters can be derived are scarce. Consequently, the present chapter is in two parts. Firstly, a new automated refinement method, the automated frequency matching method (AFMM) [Vaiana, et al. 2003a], is presented. The method is well suited for deriving specifically-tailored sets of molecular mechanics parameters for small- to medium-sized molecules and has been successfully applied to derive force field parameters not only of fluorescent dyes but also of cholesterol [Vaiana, et al. 2003a, Vaiana, et al. 2003b, Cournia, et al. 2003]. The method involves careful choice of an initial parameter set followed by automated refinement against quantum-chemically derived vibrational eigenvector and eigenvalue sets. The method is tested by comparing results obtained using it with existing CHARMM parameters for propene and the five standard nucleic acid bases: adenine, guanine, cytosine, thymine and uracil. In the second part of the chapter the new method is applied to derive parameters for both R6G and MR121. Final testing of the derived R6G parameters involves minimization of the crystal structure (no crystal structure is yet available for MR121). In the case of R6G the experimental geometry and unit cell volume are well reproduced.

3.1 Automated Frequency Matching Method for Force Field Parameterisation

3.1.1 MM Force Field Parameterisation

General sets of procedures for automated parameterisation of MM force fields based on fitting to both experimental and quantum mechanical reference data in order to obtain high quality

force-field parameters for existing MM force-field have been proposed [Norrby and Liljefors 1998]. The main steps involved in these procedures can be summarized as follows:

- Definition of a merit function based on available reference data.
- Choice of atom types to be used and definition of new atom types when needed.
- Choice of initial parameter values.
- Refinement of parameters (optimisation of the merit function).
- Testing and validation of the final parameter set.

The choice of reference data upon which to base the new parameterisation is a critical step in the parameterisation procedure. The reliability and accuracy of the new parameters in reproducing various properties of the molecule will obviously depend upon the quality of the reference data. The accuracy of the parameter set is itself an important issue: parameters must be suited to the task they are designed for. Spectroscopists are generally interested in obtaining precision in frequencies of the order of 3cm-1 [Allinger, et al. 1989]. However, this level of accuracy is not required for most non-spectroscopic aspects. Moreover, with force fields designed for molecular dynamics such precision is often unachievable, and is sacrificed in favour of simpler, computationally cost efficient, potential energy functions (see section 2.2.1). In these cases it is useless to seek spectroscopic accuracy in the parameter set.

A major requirement in MM force fields is portability of the parameter set, that is, the possibility to transfer large groups of parameters from one molecule to another. In this respect, addition of new atom types to the force field should be limited only to specific cases in which existing types cannot be used.

Automated refinement methods are mostly based on optimising the merit function, usually this corresponds to minimising a weighed sum of square deviations from a set of reference values. Refinement of parameter sets always involves exploring a high dimensional space in search of an optimal set. As for any multidimensional search method, in parameter optimisation there is always a substantial risk of the search getting stuck in a high local minimum. Initiating the search procedure from a good initial guess can help in reducing this risk.

Final testing of the parameter set should be done against experimental or theoretical data as much as possible independent of the data used in the optimisation phase.

The method described here has been developed for deriving specifically tailored sets of parameters for small to medium sized molecules within the CHARMM force field. The method can also be applied to similar atom based molecular mechanics force fields, and it follows the commonly used strategy in force-field parameterisation outlined above. Using this method parameter sets are designed to fit reliable quantum-mechanical results. The refinement step is first tested against existing CHARMM parameters for propene and the five standard nucleic acid bases: adenine, guanine, cytosine, thymine and uracil. For the case of novel fluorescent dyes such as MR121, finding suitable spectroscopic data on which to base a parameterisation is often impossible. On the contrary, reliable quantum mechanical data is now more readily available for small to medium sized molecules. R6G was initially used as a test case for parameterisation of a fluorescent dye and subsequently the method was applied to derive MR121 parameters. The main motivations for this choice are that R6G is a representative member of a whole class of widely used xanthene derivative labels and that a crystal structure of the molecule was available in the literature for final, independent testing of the new method [Fun, et al. 1997].

3.1.2 AFMM and the CHARMM Force Field

The AFMM method illustrated here is, in principle, applicable with any atom-based molecular mechanics program, the only prerequisite being the possibility of calculating normal modes and their relative eigenvectors. The CHARMM force field is used throughout the thesis for all MM calculations, attention is therefore confined in the following sections to the CHARMM force field. In the CHARMM model the potential energy is partitioned as follows [Brooks, et al. 1983]:

$$V(R) = \sum_{bonds} K_b (b - b_0)^2 + \sum_{ub} K_{ub} (s - s_0)^2 + \sum_{angles} K_\theta (\theta - \theta_0)^2 + \sum_{dihedrals} K_\chi (1 + \cos(n\chi - \chi_0))$$
$$+ \sum_{impropers} K_\varphi (\varphi - \varphi_0)^2 + \sum_{nonbond} \varepsilon_{ij} \left[\left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{12} - \left(\frac{R_{ij}^{\min}}{r_{ij}} \right)^{6} \right] + \frac{q_i q_j}{D r_{ij}} \qquad (3.1.1)$$

where K_b, K_{ub}, K_θ, K_χ, K_φ are, respectively, the bond, Urey-Bradley[1], angle, dihedral and improper dihedral constants, and b, s, θ, χ, and φ represent, respectively, bond length, Urey-Bradley 1-3 distance, bond angle, dihedral angle and improper torsion angle (a subscript zero is used to represent the corresponding equilibrium value). Non-bonded interactions between pairs of atoms (labelled i and j) at a relative distance r_{ij} consist of Lennard-Jones 6-12 (LJ) terms and Coulomb interactions. $R_{ij}^{\min}$ and ε_{ij} are, respectively, the distance between atoms i and j at which the LJ potential is zero and the depth of the LJ potential well for the same pair of atoms. D is the effective dielectric constant and q_i the partial atomic charge on atom i.

Before any refinement procedure, an initial set of parameters must be determined. The Van der Waals constants ε_{ij} and $R_{ij}^{\min}$ depend mostly on atomic properties and are relatively insensitive to changes in the molecular environment. These are directly transferred from existing CHARMM values and are not modified during refinement. The the partial atomic charges, q_i, are derived from a quantum calculation using a charge fitting method such as the CHELPG method within GAUSSIAN to fit the molecular-mechanics-derived electrostatic potential to that obtained quantum mechanically [Breneman and Wiberg 1990]. This method employs a least-squares fitting procedure to determine the set of atomic partial charges that best reproduces the quantum mechanical electrostatic potential at selected grid points. The fitting is subject to the constraint that the sum of the charges should be equal to the net charge on the molecule. Equilibrium values for bonds b_0, angles θ_0 and dihedrals χ_0 are derived from the quantum chemical ground state structure. An initial guess, based on analogy to other existing CHARMM parameters and on chemical intuition (taking equilibrium values and hybridization into account), should be made for all other missing parameters. The second term in Eq. 3.1.1 (the so-called Urey-Bradley term [MacKerell Jr. and Banavali 2000, MacKerell Jr., et al. 1998, Foloppe and MacKerell Jr. 2000]) is not present in most other force fields, and within the CHARMM model its use is limited to a few special cases. Here K_{ub} is set to zero wherever possible.

The initial parameter set is then used for energy minimization and calculation of normal modes (eigenvalues and eigenvectors) with CHARMM. These can then be directly compared with the results of the 'reference' normal modes calculated with by means of a quantum chemical program. Parameters are thus refined iteratively to fit the results of the quantum chemical normal mode calculations.

[1] The Urey-Bradley term is an additional term, characterisitc of the CHARMM force field and not present in most other MM force fields, it represents an interaction based on the distance between atoms separated by two bonds (so called 1,3 interaction).

The practice of parameterizing against vibrational frequencies is quite old [Schachtschneider and Snyder 1963, Lifson and Warshel 1969, Warshel and Lifson 1970]. One of the major problems of parameterization methods that fit to vibrational frequencies is identifying a calculated mode with the corresponding reference mode [Wang and Kollman 2001]. It is possible to obtain good matching of the frequencies for a molecule while exchanging the corresponding eigenvectors. The resulting model may then well reproduce the vibrational frequencies (and the energy) of the reference molecule but may not reproduce the distribution of energy among the inter-molecular modes, and thus the dynamical properties of the molecule. It is therefore important to use a merit function that takes into account both the frequencies and all the corresponding eigenvectors in order to avoid this problem.

The fitting method proposed here minimizes the above frequency exchange effect. In the 'ideal' case of a perfect molecular mechanics model, the vibrational properties of the molecule, as calculated by molecular mechanics, should perfectly match those resulting from the quantum *ab initio* calculation. For this to occur not only must the frequencies coincide but also the two sets of eigenvectors (resulting from the two different calculations) should coincide. Each eigenvector from the set calculated by molecular mechanics would therefore be orthonormal to all but one (it's corresponding eigenvector) of the vectors from the reference set (calculated using quantum chemical methods).

An efficient way to check simultaneously for both orthonormality and frequency matching is to project each of the CHARMM eigenvectors $\overline{\chi_i^C}$ (where the subscript *i* indicates the normal mode number and the superscript *C* indicates that the modes are calculated with CHARMM) onto the reference set of eigenvectors $\left\{\overline{\chi^G}\right\}$ (the superscript *G* indicates that throughout the thesis these modes are calculated with the GAUSSIAN program [Frisch, et al. 1995]), to find the frequency $\nu_j^{\max}$ corresponding to the highest projection ($j:\overline{\chi_i^C}\cdot\overline{\chi_j^G}=\max$) and to plot this frequency against the corresponding reference frequency, ν_i. In the ideal case mentioned above this plot would give a one-to-one relationship: $\nu_i=\nu_j^{\max}$ and $\overline{\chi_i^C}\cdot\overline{\chi_j^G}=\delta_{ij}$ where δ_{ij} is the Kronecker delta. Points that deviate from the ideal plot may indicate exchanged or mismatched frequencies.

Automated refinement methods are mostly based on minimizing a merit function, usually a weighed sum of square deviations from a set of reference values [Norrby and Liljefors 1998]. The optimisation method proposed here is based on minimization of the weighted sum-of-squares, Y^2 of the deviations from the ideal situation, as follows:

$$Y^2 = \sum_{3N-6} \omega_i^2 \left(\nu_i - \nu_j^{\max} \right)^2$$

$$\omega_i = \frac{1}{\max_j (\overline{\chi_i^C} \cdot \overline{\chi_j^G})} \qquad (3.1.2)$$

where N is the number of atoms in the molecule and there are $3N\text{-}6$ independent vibrational frequencies. The weights ω_i are chosen to be the inverse of the highest eigenvector projection. This has the effect of biasing the penalty function, even in the case of a good frequency assignment, such that minimization of Y^2 leads to an improved eigenvector projection distribution.

Refinement of parameter sets involves exploring a high-dimensional space in search of an optimal set. Consequently, already mentioned, in parameter optimisation there is always a risk of the search becoming trapped in a high local minimum. To reduce this risk it is necessary to generate a physically reasonable set of initial parameters [Wang and Kollman 2001]. To ensure this, the present method also involves checking initial parameter sets by visual inspection of the motions involved in exchanged eigenvector modes, using the Molden program, for example [Schaftenaar and Noordik 2000], and manually adjusting the parameters concerned. This approach is particularly useful for critical torsion parameters. In some cases (of which one is described in detail in the next section for the case of R6G), it is necessary to derive initial parameters from rotational potential energy profiles (single point energy calculations in GAUSSIAN), before achieving a good optimisation.

For the automatic optimisation of the force constants a standard Monte Carlo (MC) scheme [Haeffner, et al. 1997] is used to minimize Y^2. Optimisations are performed simultaneously on bond, angle, and torsion constants. At each step i, all parameters are iteratively varied in the MC algorithm with a uniform distribution within a fixed range, Y_i^2 was evaluated, and, if $Y_i^2 < Y_{i-1}^2$, the new parameter set is used in the next step, $i+1$. The optimisation algorithm is illustrated in Figure 3.1.1.

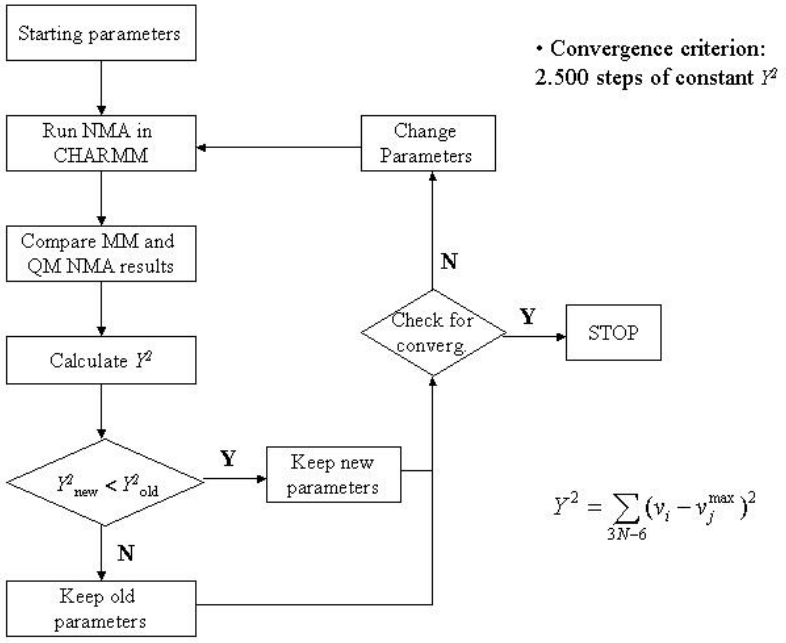

Figure 3.1.1: *Schematic representation of the optimisation algorithm used in the AFMM method. The method iteratively changes the parameters and compares both frequencies and eigenvector projections from the molecular mechanics (CHARMM in this case) normal mode analysis (NMA) with reference quantum mechanical (QM) NMA.*

When comparing results for different molecules, normalization of Y^2 can be rather tedious due to the different weights ω_i. For comparison purposes, then, after minimization of Y^2 the root-mean-square deviation, σ from the reference case is calculated:

$$\sigma = \sqrt{\frac{\sum_{3N-6} \left(\nu_i - \nu_j^{\max}\right)^2}{3N-7}} \qquad (3.1.3)$$

3.1.3 Calculations

Computational details

The computational details described here apply to all calculations presented in the rest of this chapter. All quantum chemical computations were performed with the GAUSSIAN-94 package using the standard 6-31G* basis set [Frisch, et al. 1995]. The restricted Hartree-Fock (RHF) level of theory was used for geometry optimisations and normal mode calculations. Frequencies resulting from the quantum calculations were rescaled by a factor 0.8929 to compensate for the neglect of electron correlation at the Hartree-Fock level [Foresman and Frisch 1993].

All molecular mechanics computations were performed using CHARMM version 27 [Brooks, et al. 1983]. No cutoff was applied to the electrostatic interactions. Except for the new parameters derived here, all force field parameters are given in Refs. [MacKerell Jr. and Banavali 2000, MacKerell Jr., et al. 1998, Foloppe and MacKerell Jr. 2000]. All the molecular mechanics minimizations were carried out using the Steepest Descent algorithm followed by Newton-Raphson minimization with a convergence criterion for the energy gradient of 10^{-8} kcal/mol/Å.

Robustness of the MC refinement Method

The robustness of the MC refinement method with respect to initial parameter choice was tested in calculations using the normal modes and eigenvectors of propene as calculated by CHARMM as reference data. The MC algorithm was used to derive bond force constants, K_b in Eq. 3.1.1, starting from 50 sets of randomly-generated initial values distributed uniformly between 200 and 600 $kcal/mol/Å^2$ (all other parameters were kept fixed at their normal CHARMM values). In the ten lowest-σ parameter sets found the refinement algorithm was able to find the original CHARMM parameters for propene to within a maximum relative error of less than 3%.

Test Calculations and Frequency matching Plots

The Test calculations were performed on molecules for which CHARMM parameters have already been derived by fitting to *ab initio* quantum chemical data. The molecules chosen were propene and the five standard nucleic acid bases. These molecules exhibit chemical similarity to fluorescent dyes. Normal modes calculated with CHARMM were compared with

those calculated with GAUSSIAN, using the comparison method described in methods.

Frequency matching plots (ν_j^{max} vs. ν_i) are shown for the six test molecules in Figure 3.1.2.

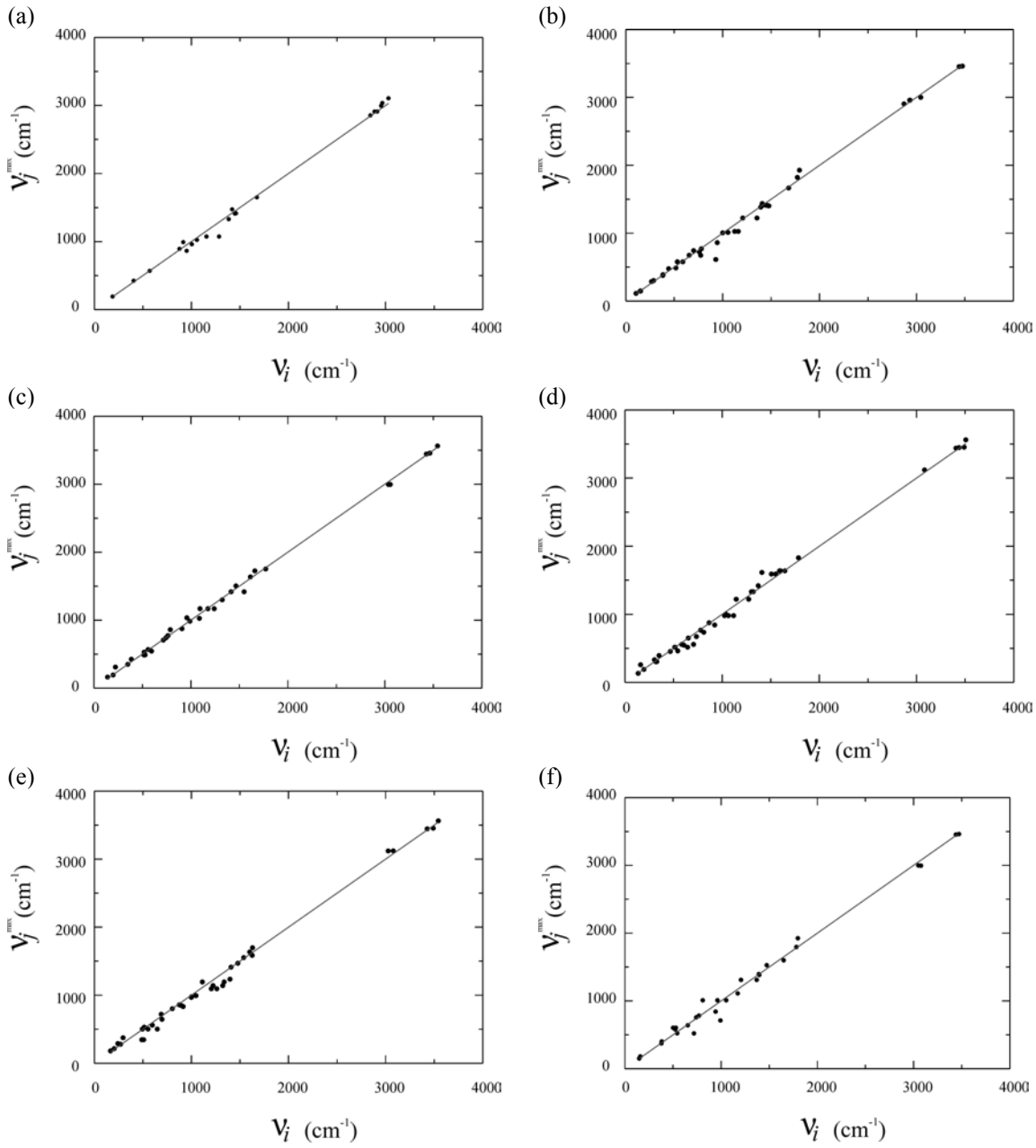

Figure 3.1.2: ν_j^{max} *vs.* ν_i *plots for (a) propene, (b) thymine, (c) cytosine, (d) guanine, (e) adenine and (f) uracil. Lines indicate the ideal case of perfectly matched frequencies and eigenvector projections; points refer to CHARMM parameters taken from Ref.* [Foloppe and MacKerell Jr. 2000]. *Values of σ are: (a) 68 cm^{-1}, (b) 74 cm^{-1}, (c) 47 cm^{-1}, (d) 67 cm^{-1}, (e) 81 cm^{-1}, (f) 92 cm^{-1}.*

The plots in Figure 3.1.2 confirm that the CHARMM and *ab initio* modes agree overall. Better matching of frequencies < 500 cm^{-1} compared to higher frequencies reflect the fact that parameterisation in Ref. [Foloppe and MacKerell Jr. 2000] was focused on the low frequency

modes. Values of σ are 74 cm^{-1} for T, 47 cm^{-1} for C, 67 cm^{-1} for G, 81 cm^{-1} for A and 92 cm^{-1} for U. The values of σ reflect, on average, a better agreement between CHARMM and the *ab initio* data for the pyrimidines than for the purines, again as reported in Ref. [Foloppe and MacKerell Jr. 2000].

3.2 Parameterisation of R6G and MR121

3.2.1 R6G Parameters

The RHF 6-31G* ground state structure for R6G is shown in Figure 3.2.1, and the atom numbering scheme is shown in Figure 3.2.2.

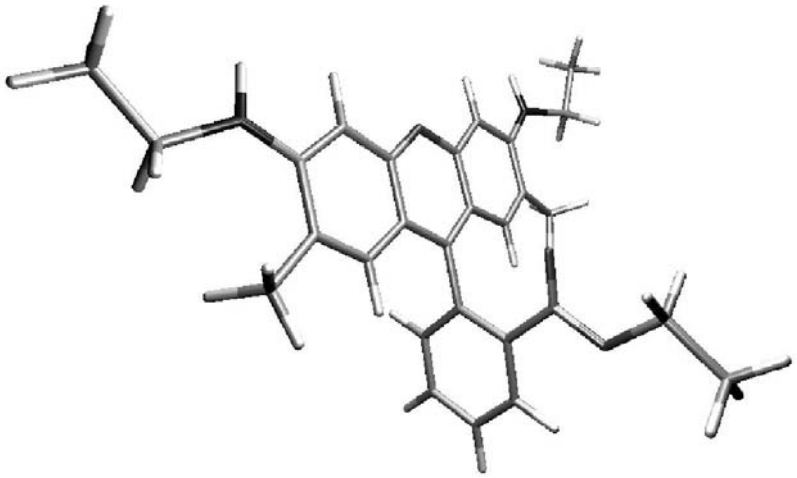

Figure 3.2.1: *Gas phase ground state structure of R6G (RHF 6-31G* optimization).*

In CHARMM, atom type CA describes sp^2 carbons of aromatic rings; to ensure portability this atom type could also be used for the case of the aromatic carbons in R6G. Unfortunately, the use of CA-type carbons for the connection of two aromatic systems in a biphenyl-type configuration results in a structure that is

highly biased toward a planar arrangement of the two rings. It was thus necessary to invent a new atom type (CA1) for the connection between the phenyl ring and the xanthene moiety in R6G (C1 and C45 in Figure 3.2.2). The only changes to CA1 with respect to the standard CA type are the bond and torsion parameters (K_b, b_0, K_χ, χ_0 and n in Eq. 3.1.1). Initial values for these parameters were derived using biphenyl as a model system: equilibrium values (b_0 and

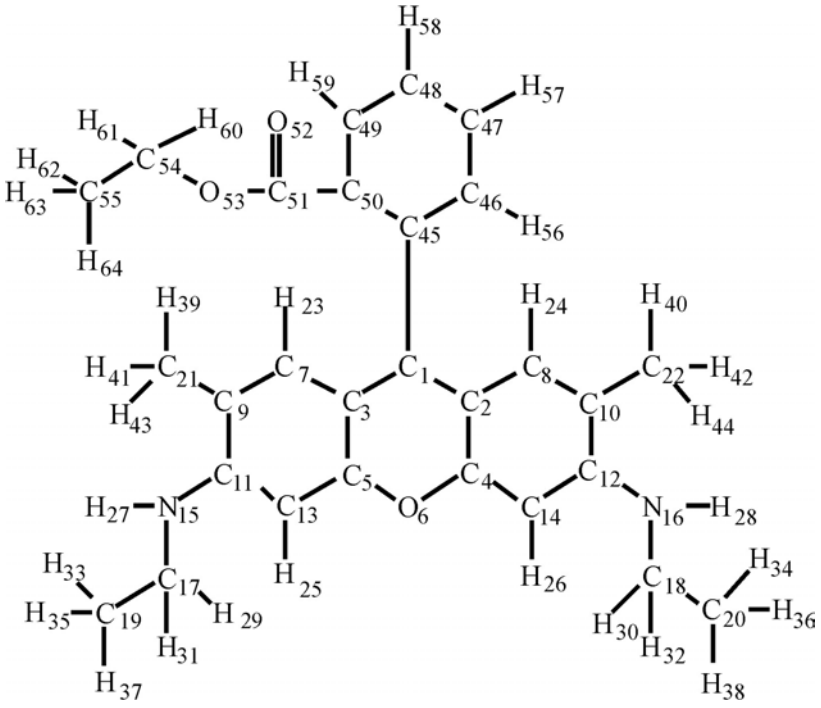

Figure 3.2.2: *Atom numbering scheme for R6G.*

χ_0 in Eq. 3.1.1) were derived from the quantum-chemically optimized structure of biphenyl. The initial value for the bond constant (K_b in Eq. 3.1.1) for the CA1-CA1 bond was taken from existing CHARMM parameters. The torsional force constants (K_χ in Eq. 3.1.1) were derived from the energy barrier for rotation of biphenyl around the central bond as calculated at RHF/6-31G* level (using single point calculations to scan the potential energy landscape). Figure 3.2.3 shows the results of these single point calculations performed with GAUSSIAN and with the CHARMM fitted potential.

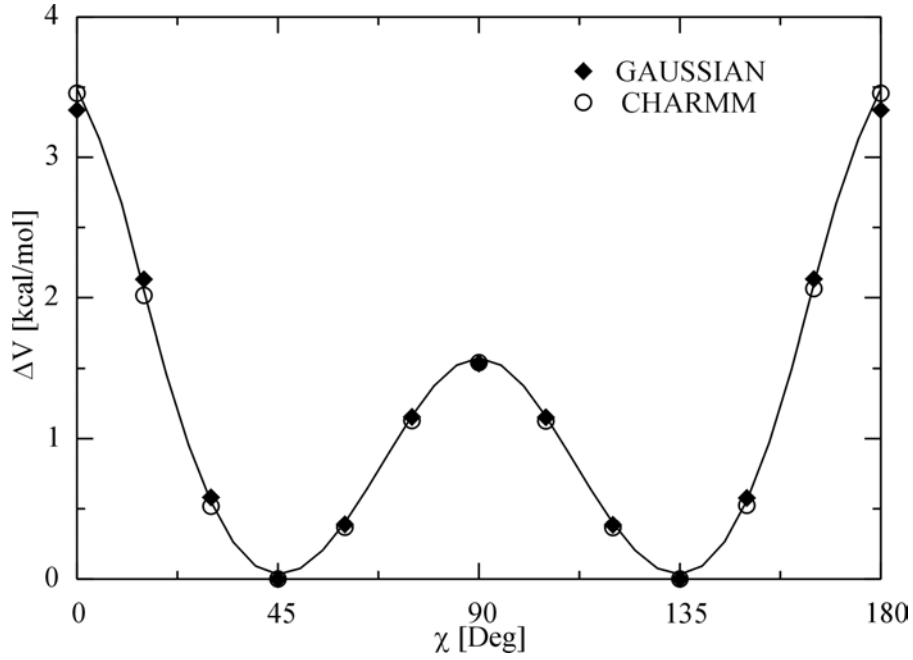

Figure 3.2.3: *Potential energy profile for rotation of the two rings of biphenyl around the central bond. The filled diamonds are the GAUSSIAN relaxed potential energy scan. Circles and the line are the fitted CHARMM potential.*

The OS and NC2 types in CHARMM were used to describe the bridging sp^2 oxygen O6 and the nitrogen atoms N15 and N16 in the partially-charged amino groups in R6G. It was necessary to introduce some new parameters for these atom types to account for the new bonding partners of these atoms in R6G.

Atom type assignments and partial atomic charges for R6G are listed in Tab. A2.1.1. Charges were derived using CHELPG at the RHF/6-31G* level. All new parameters were refined using the MC algorithm as described in section 4.1.2. Final (refined) values of all new parameters for R6G are summarized in Tabs. A1.2.2 to A1.2.4. The resulting $\nu_j^{\max}$ vs. ν_i plot for R6G is shown in Figure 3.2.4. The corresponding value of σ (51.1 cm^{-1}) is within the range seen in the benchmark studies (see Figure 3.1.2).

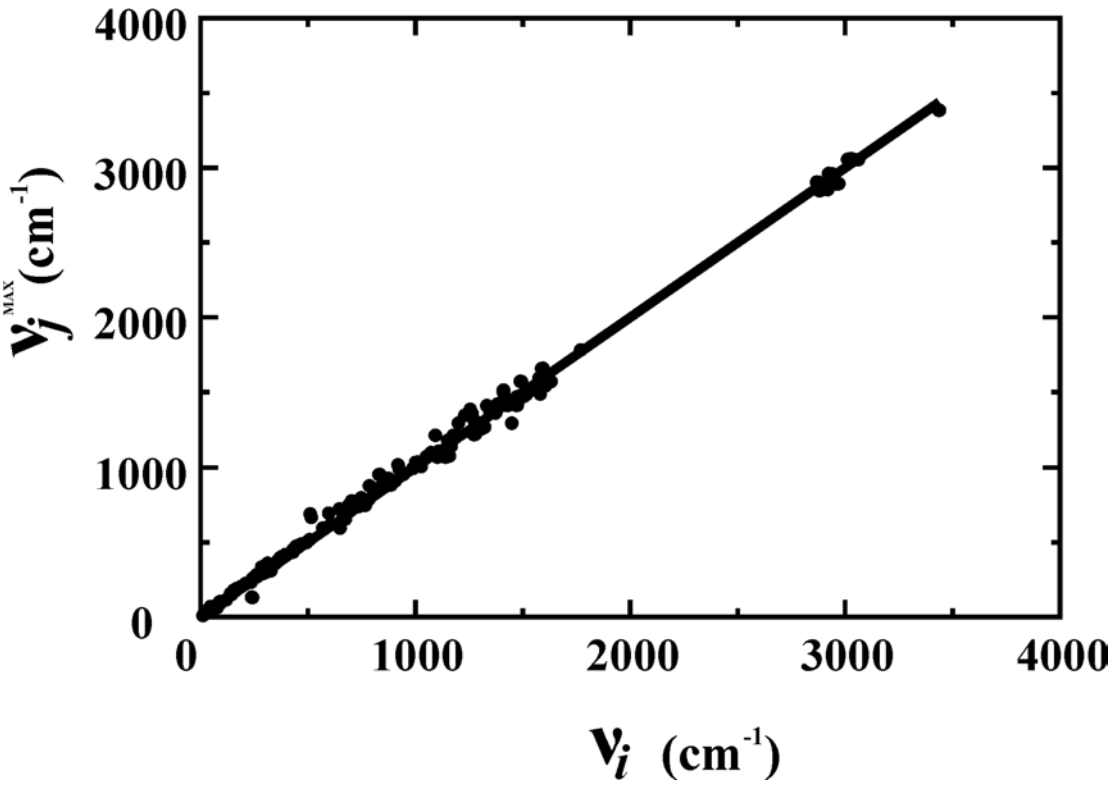

Figure 3.2.4: ν_j^{max} *vs.* ν_i *plot for R6G. The line is the ideal case of perfectly matched frequencies and eigenvector projections; points refer to the present optimized parameters.* σ = *51.1 cm*$^{-1}$.

3.2.2 R6G Crystal Structure

Final testing of a parameter set should be performed against experimental or theoretical data that are as far as possible independent of the data used in the optimisation phase. Here, parameters were tested by performing an energy minimization of R6G in crystalline state using the refined parameter set and comparing the results with the experimental structure by [Fun, et al. 1997]. The calculation was performed including the full crystal environment using periodic boundary conditions. The unit cell has *Pbca* symmetry applied to one R6G molecule, an iodide counterion (R6G is a mono cation) and one water molecule represented here with the TIP3P potential [Jorgensen, et al. 1983]. The resulting unit cell has thus eight R6G molecules, eight iodide ions and eight water molecules. The unit cell dimensions were allowed to vary in the energy minimization. The resulting structure (averaged out over the eight molecules of the unit cell) is compared with experiment in Figure 3.2.5.

a) b)

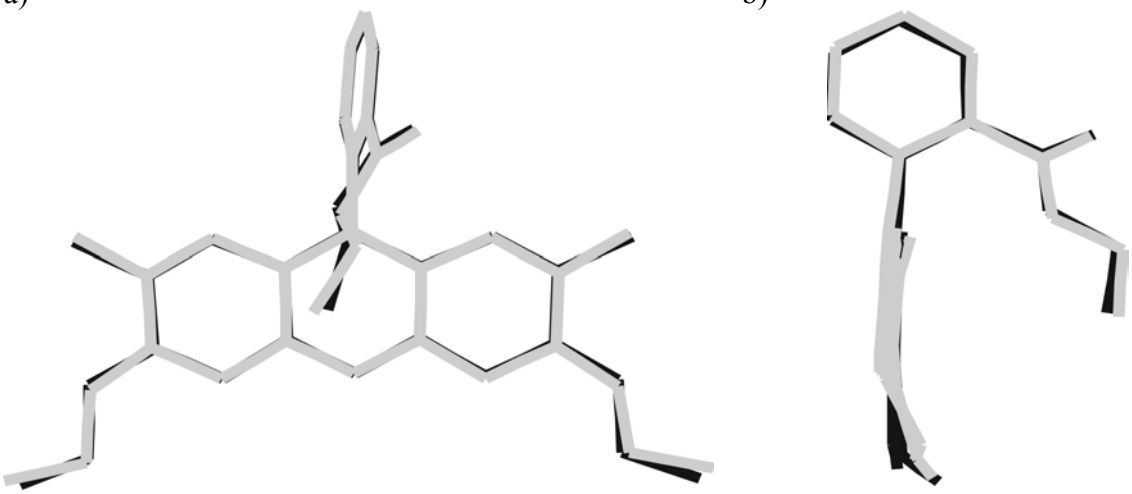

Figure: 3.2.5: *R6G crystal structure, comparison of the simulation result (gray) with the experimental structure by Fun et al. (black): a) front view b) side view. Only heavy atoms are shown. RMSD = 0.083 Å.*

The RMSD between the calculated and experimental molecular structures (not including hydrogen atoms) is 0.083 Å. As shown in Tab. 3.2.1, the molecular mechanics results also match the experimental cell dimensions well. The total unit cell volume difference between the simulated crystal and the experimental data is 26.7 $Å^3$ *i.e.,* 0.5%.

Table 3.2.1: *Experimental unit cell dimensions as reported by* [Fun, et al. 1997] *together with those resulting from the crystal simulation.*

	X	Y	Z
Ref.	14.92 Å	15.16 Å	23.61 Å
Simulation	14.94 Å	15.51 Å	23.15 Å
Δ_i	-0.02 Å	-0.35 Å	0.46 Å
Δ_i/i	0.1%	2.3%	1.9%

These results demonstrate the aptitude of the method to derive force field parameters that reliably reproduce both inter- and intra-molecular properties of R6G. The method can now be

used to derive parameters for MR121, for which no crystal structures on which to base independent testing of the parameters are available.

3.2.3 MR121 Parameters

The method used for deriving the MR121 parameter set follows the same steps as for R6G. An initial set of parameters for MR121 was selected for refinement using AFMM. The partial atomic charges, q_i in Eq. 3.1.1, were derived using the CHELPG methodology with a standard RHF/6-31G* quantum chemical optimization. The ground state structure of MR121 is shown, along with the atom numbering scheme in Figure 3.2.6.

a) b)

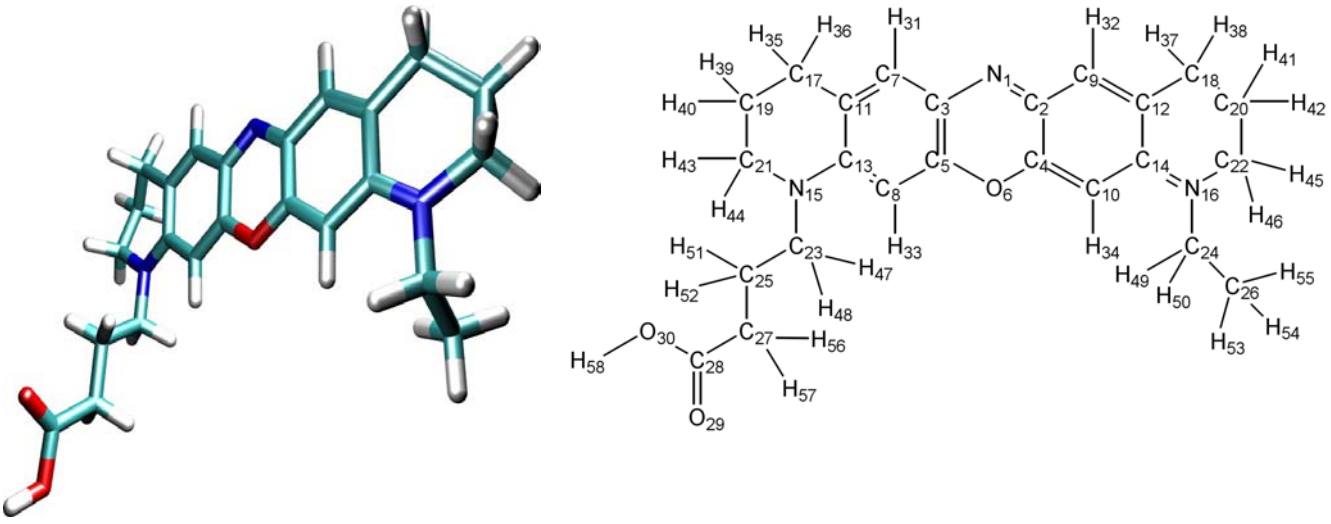

Figure 3.2.6: *a) Gas phase ground state structure of MR12 (RHF 6-31G* optimization). b) Atom numbering scheme for MR121.*

Three new atom types were defined to describe the nitrogens and the oxygen present on the ring system of the dye (see Figure 3.2.6b). The Van der Waals constants ε_{ij} and $R_{ij}^{\min}$ for these were directly transferred from original CHARMM values for NC2, NN3G and OS atom types (see the following references: [MacKerell Jr. and Banavali 2000, MacKerell Jr., et al. 1998, Foloppe and MacKerell Jr. 2000]). Van der Waals constants and partial atomic charges were not modified during refinement. Equilibrium values for bonds b_0, angles θ_0 and dihedrals χ_0 that were not present in the original CHARMM force field parameter set [MacKerell Jr. and Banavali 2000, MacKerell Jr., et al. 1998, Foloppe and MacKerell Jr. 2000] were derived from the structure resulting from the RHF/6-31G* optimization. An initial guess, based on

analogy to similar existing CHARMM parameters and on chemical intuition, was made for all other missing parameters.

The initial parameter set was used for minimization and calculation of normal modes (eigenvalues and eigenvectors) with CHARMM. In the force-field parameter derivation calculation all molecular mechanics minimizations were carried out using the Steepest Descent algorithm followed by Newton-Raphson minimization with a convergence criterion for the energy gradient of 10^{-8} kcal/mol/Å. The normal modes obtained were compared with the normal modes calculated with the quantum chemical method (considered to be the reference). Parameters were then refined iteratively to fit the results of the quantum chemical normal mode eigenvalues and vectors using the MC algorithm described above as for R6G.

Final, refined force field parameters are listed in Tabs. A1.2.1-A1.2.4. The v_i vs. v_j^{max} plot obtained with this parameter set for MR121 is shown in Figure 3.2.7.

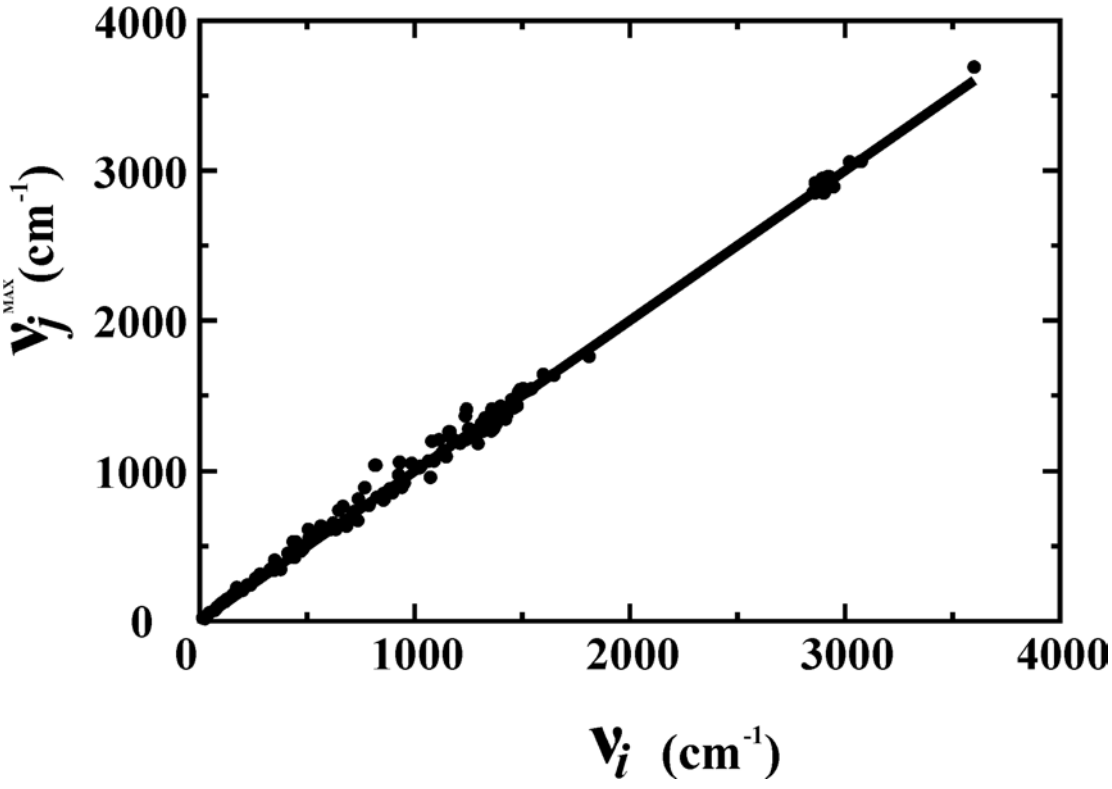

Figure 3.2.7: *ν_j^{max} vs. ν_i plot for MR121. The line is the ideal case of perfectly matched frequencies and eigenvector projections; points refer to the present optimized parameters. σ = 54.8 cm^{-1}.*

The obtained value of the root-mean-square deviation from the reference case, σ in Eq. 3.1.3, is 54.8 cm^{-1} for MR121. This value is within the range seen in the previous benchmark studies on small aromatic compounds described above.

3.3 Conclusions

Results obtained from MD simulations strongly depend on the parameter sets used in the force field. A reliable simulation must use parameter sets that are derived from reference data which are as much as possible independent of the quantities that the simulation is meant to reproduce. The new, general method based on frequency matching (AFMM) which is developed here for obtaining force field parameters for small to medium sized molecules uses results of reliable quantum chemical (*ab initio*) calculations as reference data. The computationally inexpensive penalty function proposed here makes the method very well suited for automation. The use of a relatively simple but efficient MC refinement algorithm has allowed to obtain a frequency matching for R6G and MR121 within the error range found in benchmark studies on test molecules.

In this work the AFMM method is used only as a necessary tool to obtain reliable simulations of fluorescent dyes and their interactions with tryptophan residues. Nonetheless, the importance of the method itself should not be overlooked. The request of molecular mechanics force field parameters for new, non-standard molecules is growing constantly. The AFMM method offers the possibility to develop parameters for new molecules at a relatively low computational cost and requiring, in principle, no experimentally derived reference data. The method also allows checking, and if necessary refining, existing parameter sets.

The CHARMM parameters derived here for R6G and MR121 will be extensively used to simulate the interactions of the dyes with Tryptophan residues in order to gain insight on the quenching of dye fluorescence by Tryptophan. The model systems used are designed to reproduce experimental conditions as closely as possible. Special emphasis will be made in the next chapter on confronting the simulations with time resolved fluorescence measurements.

References

Brooks, B., R. Bruccoleri, B. Olafson, D. States, S. Swaminathan and M. Karplus (1983). "CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations" J Comput Chem **4** 187-217.

Vaiana, A. C., A. Schulz, J. Wolfrum, M. Sauer and J. C. Smith (2003a). "Molecular Mechanics Force Field Parameterization of the Fluorescent Probe Rhodamine 6G using Automated Frequency Matching" J Comput Chem **5** (24): 632-9.

Vaiana, A. C., H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer and J. C. Smith (2003b). "Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation" J Am Chem Soc (in press):

Cournia, Z., A. C. Vaiana, J. C. Smith and M. Ullmann (2003). "Derivation of a Molecular Mechanics Force Field for Cholesterol." Pure Appl Chem (in press):

Norrby, P.-O. and T. Liljefors (1998). "Automated molecular mechanics parameterization with simultaneous utilization of experimental and quantum mechanical data" J Comput Chem **10** (19): 1146-66.

Allinger, N. L., Y. H. Yuh and J. H. Lii (1989). "Molecular mechanics. The MM3 force field for hydrocarbons. 1" J Am Chem Soc **23** (111): 8551-66.

Fun, H. K., K. Chinnakali, K. Sivakumar, C. M. Lu, X. R. G and Y. X (1997). "N-[9-2-Ethoxycarbonylphenyl)-6-(ethylamino)-2,7-dimethyl-3-xanthenylidene]-ethylammonium Iodide Monohydrate" Acta Crystallogr (C53): 1619-20.

Breneman, C. N. and K. B. Wiberg (1990). "Determining Atom-Centered Monopoles from Molecular Electrostatic Potentials. Need for High Sampling Density in Formamide Conformational Analysis" J Comput Chem **3** (11): 361-73.

MacKerell Jr., A. D. and N. Banavali (2000). "All-Atom Empirical Force Field for Nucleic Acids: 2) Application to Molecular Dynamics Simulations of DNA and RNA in Solution" J Comput Chem (21): 105-20.

MacKerell Jr., A. D., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, I. Reiher, W.E, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin and M. Karplus (1998). "All-atom empirical potential for molecular modeling and dynamics Studies of proteins" J Phys Chem B (102): 3586-16.

Foloppe, N. and A. D. MacKerell Jr. (2000). "All-Atom Empirical Force Field for Nucleic Acids: 1) Parameter Optimization Based on Small Molecule and Condensed Phase Macromolecular Target Data" J Comput Chem (21): 86-104.

Schachtschneider, J. H. and R. G. Snyder (1963). "Vibrational analysis of the n-paraffins. II. Normal coordinate calculations" Spectrochim Acta, Part A (19): 117-68.

Lifson, S. and A. Warshel (1969). "Consistent force field for calculations of conformations, vibrational spectra, and enthalpies of cycloalkane and n-alkane molecules" J Chem Phys **11** (49): 5116-29.

Warshel, A. and S. Lifson (1970). "Consistent force field calculations. II. Crystal structures, sublimation energies, molecular and lattice vibrations, molecular conformations, and enthalpies of Alkanes" J Chem Phys **2** (53): 582-94.

Wang, J. and P. A. Kollman (2001). "Automatic parameterization of force field by systematic search and genetic algorithms" J Comput Chem **12** (22): 1219-28.

Frisch, M. J., G. W. Trucks, H. B. Schlegel, P. M. W. Gill, B. G. Johnson, M. A. Robb, J. R. Cheeseman, T. Keith, G. A. Petersson, J. A. Montgomery, K. Raghavachari, M. A. Al-Laham, V. G. Zakrzewski, J. V. Ortiz, J. B. Foresman, J. Cioslowski, B. B. Stefanov, A. Nanayakkara, M. Challacombe, C. Y. Peng, P. Y. Ayala, W. Chen, M. W. Wong, J. L. Andres, E. S. Replogle, R. Gomperts, R. L. Martin, D. J. Fox, J. S. Binkley, D. J. Defrees, J. Baker, J. P. Stewart, M. Head-Gordon, C. Gonzalez and J. A. Pople (1995). Gaussian 94, Revision D.4. Pittsburgh PA, Gaussian, Inc.

Schaftenaar, G. and J. H. Noordik (2000). "Molden: a pre- and post-processing program for molecular and electronic structures" J Comput Aided Mol Des **2** (14): 123-34.

Haeffner, F., T. Brinck, M. Haeberlein and C. Moberg (1997). "Force field parameterization of copper(I)-olefin systems for density functional calculations" THEOCHEM (397): 39-50.

Foresman, J. B. and Æ. Frisch (1993). Exploring chemistry with electronic structure methods, 2nd Ed. Pittsburgh, PA,

Jorgensen, W., J. Chandrasekhar, J. Madura, R. Impey and M. Klein (1983). "Comparison of simple potential functions for simulating liquid water" J Chem Phys (79): 926-35.

4. MD Simulations of Two Simple but Realistic Dye/Quencher Systems: R6G/Trp and MR121/Trp in Explicit Water

The unique property of Trp to serve as specific, efficient fluorescence quencher for certain organic dye structures has enabled direct measurement of intermolecular contact formation rates between dye and Trp in solution and also intramolecular contact rates in peptides. This has allowed detection of antibody binding events and development of highly sensitive assays. In this chapter, fluorescence spectroscopy and molecular dynamics (MD) simulation are combined to characterize the interaction of the two organic fluorescent dyes R6G and MR121, with the amino-acid tryptophan in aqueous solution. Steady-state and time-resolved fluorescence quenching experiments reveal the formation of essentially non-fluorescent ground-state dye/Trp complexes. The MD simulations are used to elucidate the molecular interaction geometries involved. The systems studied are small enough for a thorough investigation of the configurational landscape to be conducted via MD simulations involving full treatment of electrostatics and explicit representation of solvent. The approach presented here provides a useful model case for the treatment of the larger and more biologically relevant systems in the next chapter, *i.e.*, for the structural interpretation of data obtained from one of the two immunodominant p53 epitopes mentioned in section 1.1 labelled first with R6G and then MR121.

4.1 Results of Fluorescence Quenching Experiments

The Fluorescence measurements described in this section were performed by M. Sauer and co-workers at the University of Heidelberg. The main results of these measurements are presented here.

In order to determine the fluorescence quenching efficiency and aspects of the quenching mechanism of the dyes by Trp, steady-state and time-resolved fluorescence intensities of R6G and MR121 were measured as a function of the Trp concentration in phosphate buffered aqueous solution. Plots of the inverse normalized steady-state fluorescence intensities and inverse normalized fluorescence lifetimes of the dyes *vs.* the Trp concentration allow thermodynamic and kinetic quenching data to be determined *via* Stern-Volmer analysis (see

section 2.3.2). The resulting Stern-Volmer plots are shown in Figure 4.1.1. The Stern-Volmer data from both dye/Trp systems are summarized in Table 4.1.1.

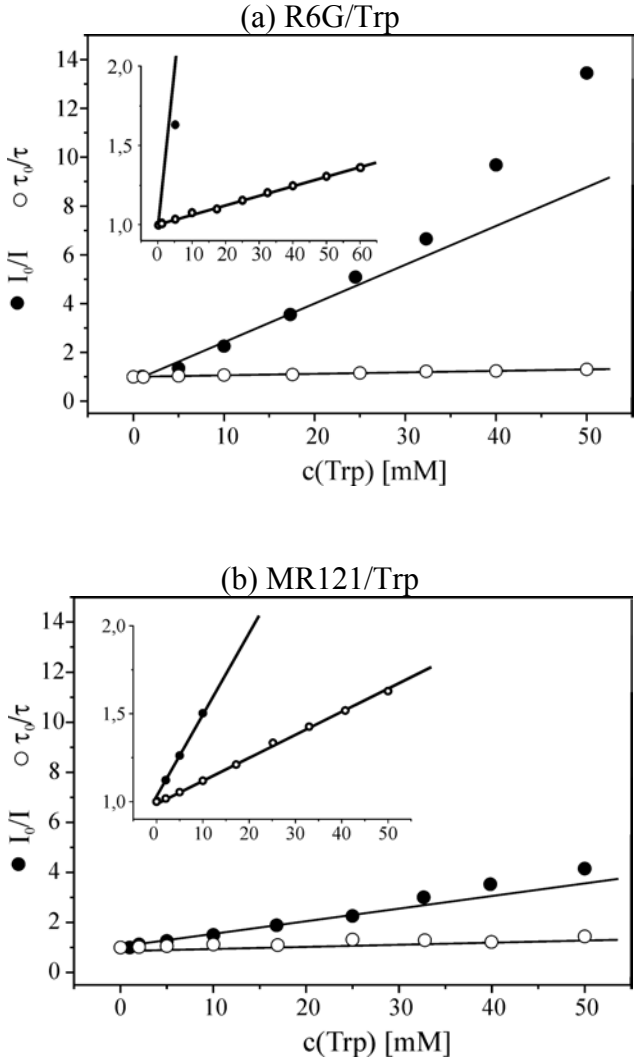

Figure 4.1.1: *a) Static (closed circles) and dynamic (open circles) bimolecular Stern-Volmer plot of R6G/Trp in phosphate buffered saline solution, pH 7.4. Excitation wavelength: 495 nm, detection wavelength: 555 nm. In b) the equivalent plot for MR121 is shown. Insets represent the same data as the main graphs on an expanded y-axis scale.*

The bimolecular dynamic quenching rate, $k_{q,d}$, *i.e.* the rate at which fluorescence-quenched collisional encounter complexes between the dye and Trp are formed, can be extracted from the slope of the fluorescence lifetime data according to the "dynamic" Stern-Volmer equation:

$$\frac{\tau_0}{\tau} = 1 + k_{q,d}\tau_0[Trp] \qquad (4.1.1)$$

where τ denotes the measured fluorescence lifetime, τ_0 the lifetime of the free dye in solution, and [*Trp*] the Trp concentration. This diffusion-controlled process leads to a reduced fluorescence lifetime of the dye in the presence of Trp (see section 2.3.2).

Table 4.1.1: *Experimental fluorescence quenching data for the dyes MR121 and R6G with Trp in phosphate buffered saline solution, pH 7.4.*

	MR121	*R6G*
$k_{q,d}$ / $M^{-1}sec^{-1}$	(3.2±0.3) x 10^9	(3.4±0.3) x 10^9
K_S / M^{-1}	220±20	30.4±3
τ_0 [ns]	1.85±0.06	3.85±0.12

$k_{q,d}$ denotes the bimolecular dynamic quenching rate, K_S the static Stern-Volmer constant and τ_0 the lifetime of the free dye in solution.

The fluorescence photon decay distributions (*i.e.* the time dependence of the fluorescence intensity) of the dyes were found to well fit a mono-exponential model in the presence of Trp with a lifetime which is close to that of the free dye. This indicates that the complexes are essentially non-fluorescent, *i.e.* they exhibit a fluorescence lifetime shorter than the time-resolution of the instrument (~50ps) used to measure the fluorescence decays. Table 4.1.1 shows that $k_{q,d}$ is the same for both systems to within experimental error, whereas τ_0 is significantly longer for R6G.

Complex formation between dye and Trp reduces the amount of free fluorescent dye in solution and therefore also reduces the fluorescence intensity of the measured ensemble. The static Stern-Volmer constant, K_S, *i.e.* the thermodynamic association constant for the formation of non-fluorescent ground state complexes between the dye and Trp, can be extracted from the slope of the linear region of the concentration dependent steady-state fluorescence intensities according to the "static" Stern-Volmer equation:

$$\frac{I_0}{I} = 1 + K_S[Trp] \qquad (4.1.2)$$

where I denotes the measured steady-state fluorescence intensity, and I_0 the fluorescence intensity of the free dye in solution. In the case of a quenching mechanism containing both static and dynamic components the intensity plots are non-linear. This is the case here for both dye/quencher pairs.

K_S is about seven times higher for MR121/Trp than for R6G/Trp. This reveals a stronger tendency of Trp to form non- or only weakly-fluorescent ground state complexes with MR121 than with R6G.

The time-resolved fluorescence quenching experiments are consistent with a model in which the underlying quenching mechanism involves dye/quencher complex formation. This conclusion is supported by the observation that for both systems the static Stern-Volmer constant, K_S, strongly decreases upon addition of detergent or organic solvents like ethanol (data not shown). This suggests that hydrophobic interactions between MR121 or R6G and Trp may play an important role in the formation of the non-fluorescent complexes. In this case the mechanism would involve a PET reaction between the Trp and the dye in stacked interaction geometries, similar to that observed for the dye/quencher pair riboflavin/Trp in the riboflavin-binding protein as described by [Zhong and Zewail 2001].

4.2 MD simulations

4.2.1 Simulation Protocol

MD Simulations were performed on the two model dye/Trp systems: the first (MR121/Trp) consisting of one MR121 dye molecule, one Trp molecule, 1013 TIP3P [Jorgensen, et al. 1983] water molecules and a chloride counter-ion, and the second (R6G/Trp) consisting of one R6G molecule, one Trp molecule, 1002 TIP3P water molecules and a chloride counter-ion.

Five MD simulations were performed on each of the two dye/tryptophan systems. In the rest of the chapter the simulations are referred to as RT1, …, RT5 for the R6G/Trp system and MT1, …, MT5 for the MR121/Trp system. For each system the five simulations were performed with different initial ring separations, so as to improve configurational statistics, as described below. All simulations were performed in the NPT ensemble at 1.0 atm pressure and at 300 K in a truncated octahedral box of (33.6±0.1) Å for MR121/Trp and (33.9±0.1) Å for R6G/Trp using periodic boundary conditions. This corresponds to relative Trp concentrations of 55 mM for the MR121/Trp system and 57 mM for R6G/Trp. Long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) method described in section 2.2.1.

The heating and equilibration procedure was the same for all the MD runs. An arbitrary starting configuration of solvated dye/tryptophan including counter-ions was gradually heated from 0 to 300 K during 5.5ps, this was done by reassigning the velocities every 100 steps according to a Maxwell-Boltzmann distribution at the desired temperature. The integration time step in this heating phase was of 0.1fs. The integration time was then increased to 1fs and a first 100ps equilibration was performed at constant volume and temperature. The

system was then allowed to relax to its equilibrium density at constant temperature and pressure for 100ps. At this point the distance *r*, between the geometrical centers of the main ring systems of the dye and of the tryptophan (the definition of *r* is shown in Figure 4.2.1) was subjected to a harmonic constraint $C(r) = k(r-r_0)^2$ with the force constant *k* set to 10 kcal/mol/Å^2 for a total of 100ps. Values of the equilibrium distance r_0 were: 3.5 Å for RT1 and MT1, 5.0 Å for RT2 and MT2, 7.5 Å for RT3 and MT3, 10.0 Å for RT4 and MT4 and 12.5 Å for RT5 and MT5. Subsequently the system was allowed to relax unconstrained for another 100ps prior to the production phase.

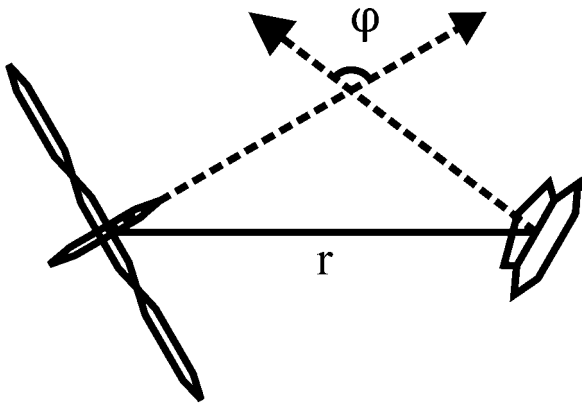

Figure 4.2.1: *definition of the dye/Trp distance, r and the angle, φ between the normals to the dye and Trp planes. The same definition is used for the MR121/Trp system.*

During the production runs no constraints were used. Each production run lasted 10ns *i.e.* 50ns of simulation time was run in total on each dye/Trp system. Altogether the simulations required 41 days (wall time) on 32 CPUs (16 per dye/Trp system) of the HELICS cluster of the Interdisziplinäres Zentrum für Wissenschaftliches Rechnen at the University of Heidelberg. Configurations were downloaded to disk every 0.1ps for subsequent analysis.

4.2.2 Potential of Mean Force: Theoretical Background

The mutual arrangement of the two molecules (Trp and dye) is here described by the distance *r* defined above and the angle φ between the normals to the planes of the main ring systems of the dye and the tryptophan (see Figure 4.2.1). The MD simulations were used to determine the two-dimensional potential of mean force (PMF) landscape, $\omega(r,\varphi)$. This allows the free energy differences between the complexed and noncomplexed states of the dye/Trp system

and the configurational pathways and free energy barriers connecting these states to be determined.

The simulation models a classical thermodynamic system in the NPT ensemble: its microscopic states are described by a set of *3N* coordinates $\{r, \varphi, q_1, \ldots, q_{3N-2}\}$ and their associated momenta $\{p_r, p_\varphi, p_1, \ldots, p_{3N-2}\}$. At equilibrium, the probability $P(r,\varphi)$ of finding the system in a given state of the $\{r, \varphi\}$ subspace *i.e.*, the probability of finding given values of r and φ is:

$$P(r,\varphi) = \frac{\int_V \cdots \int e^{\frac{-U(r,\varphi,q_1,\ldots,q_{3N-2})}{kT}} dq_1 \cdots dq_{3N-2}}{Z} \qquad (4.2.1)$$

where k is the Boltzmann constant, T is the temperature of the system, $U(r, \varphi, q_1, \ldots, q_{3N-2})$ is the potential energy and Z is the partition function. The integrals are extended to the hyper-volume V of phase space spanned by all degrees of freedom except r and φ. $\omega(r,\varphi)$, the associated PMF [Hill 1956], is given by:

$$\omega(r,\varphi) = -kT \ln\left[P(r,\varphi)\right] \qquad (4.2.2)$$

The negative gradient of $\omega(r,\varphi)$ with respect to r and/or φ, $-\nabla_{r,\varphi}\ \omega(r,\varphi)$, is the mean "thermodynamic drive" acting on the system along the coordinates r and/or φ averaged over all possible states of the remaining *3N-2* degrees of freedom. $\omega(r,\varphi)$ thus is a free energy *i.e.*, the reversible work done on the system in moving it from state (r_a,φ_a) to state (r_b,φ_b) is expressed by the difference $\Delta\omega^{a,b} = \omega(r_b,\varphi_b) - \omega(r_a,\varphi_a)$.

$P(r,\varphi)$ can be estimated using MD simulation. If $n(r,\varphi)$ is the number of MD snapshots in which the system is found to be in a volume element $d\Omega = 2\pi r^2 \sin\varphi\ dr d\varphi$ around position (r,φ), then $P(r,\varphi)$ is given by:

$$P(r,\varphi) \approx \frac{n(r,\varphi)}{N_{MD} d\Omega} \qquad (4.2.3)$$

where N_{MD} is the total number of MD snapshots taken during the simulation. If the system is ergodic the equality in Eq. 4.2. is exact in the limit $N_{MD} \rightarrow \infty$.

In the present work, for both systems simulated $n(r,\varphi)$ was evaluated on a 100x100 square grid over 0Å < r < 25Å and 0 < φ < 180º resulting in a bin width of 0.25Å for r and 1.8º for φ.

The statistical error due to incomplete sampling in calculating $\omega(r,\varphi)$ with a finite value of N_{MD} in Eq. 4.2. was estimated in the following manner. The total simulation run was divided into M segments of equal length and for each of these segments the PMF was independently calculated. The average standard deviation from the PMF for $M=1$ is an estimate of the desired statistical error and is evaluated as follows:

$$\sigma_M = \left\langle \left\{ \frac{1}{M} \sum_{i=1}^{M} \left[\omega_i(r,\varphi) - \omega_{tot}(r,\varphi) \right]^2 \right\}^{\frac{1}{2}} \right\rangle_{r,\varphi} \qquad (4.2.4)$$

where $\omega_i(r,\varphi)$ is the PMF, as defined in Eq. 4.2.2, calculated from simulation data from the ith segment and $\omega_{tot}(r,\varphi)$ is calculated from the whole simulation. The average in the angular brackets spans over the whole $\{r, \varphi\}$ space.

4.2.3 Simulation Results

From the fluorescence quenching experiments, $\Phi_{f,rel}$ ($=I/I_0$) is the relative fluorescence quantum yield. The values of $\Phi_{f,rel}$ obtained at Trp concentrations corresponding to those used in the simulations, are ~0.25 for the R6G/Trp system and ~0.10 for the MR121/Trp system, as determined from Figure 4.1.1. This corresponds to a single dye molecule being in a fluorescent state for ~25% of the observation time for R6G and ~10% for MR121. The question arises as to what geometries correspond to the non-fluorescent states for the two systems. To examine this the one-dimensional, normalized probability distribution, $P(r)$ of the dye-quencher distance, r was calculated from the simulations. The results are shown in Figure 4.2.2.

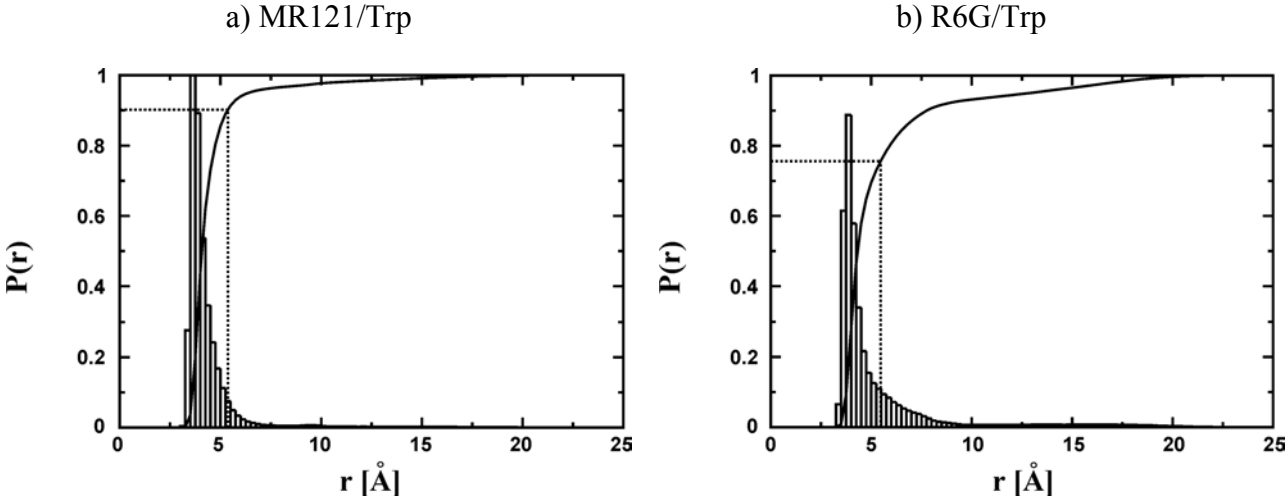

Figure 4.2.2: *Normalized probability distributions P(r) for MR121/Trp (a) and R6G/Trp (b). The continuous line is the integral of P(r), i.e., $F(r^*)$ defined in Eq. 4.2.5. The dotted line indicates the value of r^*=(5.4±0.2)Å for MR121/Trp and r^*=(5.5±0.4)Å for R6G/Trp, corresponding to the experimental quenching efficiencies of 90% and 75%, respectively. The relative difference between the values of $F(r^*)$ for the two dyes is $\Delta F(r^*)$=(15±4)%. The error in r^* was estimated in the same manner as the error σ_5 in the PMF landscape described in Eq.4.2.4, the estimated error in $\Delta F(r^*)$ is back-propagated from the error in r^*. The trajectories were divided into five segments of equal length, for each of them r^* was calculated. The standard deviation from the value of r^* derived from the whole simulation gives the error shown above.*

The probability distributions of the two dye/quencher systems are significantly different, with a higher relative population at r > 5Å for R6G.

The solid line in Figures 4.2.2a and 4.2.2b is the integral of $P(r)$:

$$F(r^*) = \int_0^{r^*} P(r)dr \qquad (4.2.5)$$

$F(r^*)$ is thus the fraction of configurations observed for which $r < r^*$. Using the above-determined experimental probability values of F=75% for R6G/Trp and F=90% for MR121/Trp the corresponding distances r^* can be calculated from Figure 4.2.2. These distances, shown by dotted lines in Figure 4.2.2, are: r^*=(5.5±0.4)Å for R6G/Trp and r^*=(5.4±0.2)Å for MR121/Trp. These two r^* values are remarkably similar given the significantly different form of the probability distributions *P(r)* for the two systems. Configurations occurring at distances >5.5 Å may be considered to be fluorescent. The minimum of the Lennard-Jones potential for the carbon atoms on the ring systems is at ~2.0 Å. Therefore, the values of r^* derived above indicate that quenched configurations occur for both dyes only at distances close to contact with the Trp.

The two-dimensional $\{r, \varphi\}$ PMF maps calculated from the MD simulations are shown for both systems in Fig. 4.2.3, these provide further details on the geometries of the complexes.

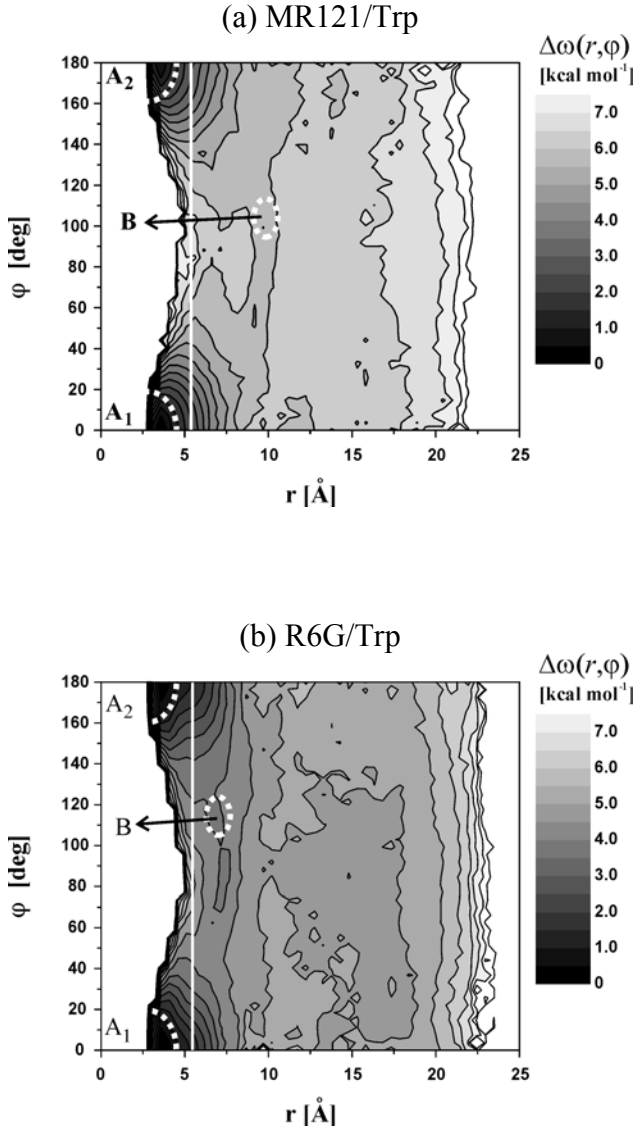

Figure 4.2.3: *Two-dimensional PMF landscapes, Δω(r, φ) derived using Eq. 4.2.2 for MR121/Trp (a) and R6G/Trp (b). For both the PMF value is set to 0 kcal mol^{-1} at the global minimum. Dashed, white lines serve as a guide to the eye and mark respectively the areas A_1, A_2 (locations of the two PMF minima corresponding to close, stacking configurations of the complexes) and B (along the connection pathway between the minima) selected for subpopulation analysis (see Figs. 4.2.5 and 4.2.6). The continuous white lines are placed at the quenching distances r^*=5.5 Å for MR121/Trp and r^*=5.4 Å for R6G/Trp.*

Values of the statistical error σ_M, calculated using Eq. 4.2.4, are reported for M=2 to M=7 in Table 2. These values indicate a statistical error in the PMF landscape in the range 0.21 to 0.43 kcal mol^{-1} for R6G/Trp and 0.16 to 0.38 kcal mol^{-1} for MR121/Trp. These are similar to the values, in the range 0.26 to 0.31 kcal mol^{-1}, obtained in a recent MD analysis of the PMF landscape of aromatic amino-acid complexes in water [Chelli, et al. 2002].

The continuous white line in Figure 4.2.3 is placed at the quenching distance $r=r^*$. To a first approximation the landscapes exhibit similar topologies. Both possess two pronounced minima labeled A_1 and A_2. These minima correspond to close-contact stacking of the ring systems of the dye with the tryptophan, well within the quenching distance. A clearly visible "low-energy" pathway connects the two minima and extends beyond the quenching distance. The PMF barrier along the pathway for the MR121/Trp system is ~5 kcal mol^{-1}, about 1 kcal mol^{-1} higher than that of the R6G/Trp system. Both minima A_1 and A_2 are ~1 kcal mol^{-1} shallower for R6G/Trp than for MR121/Trp. In contrast to the R6G/Trp landscape, the PMF of the MR121/Trp complex is almost symmetrical around φ=90º - this reflects the fact that MR121 has planar symmetry whereas in R6G this symmetry is broken due to the presence of the ester side chain. In both systems, beyond a separation distance of ~10 Å the PMF exhibits no significant dependence on φ.

We now examine more closely subpopulations of the stacked configurations A_1 and A_2 as well as the configurations falling in the areas along the low energy configurational pathway between the minima (the area marked as B in Figures 4.2.3a and b). We define θ as the angle between a fixed vector on the dye plane and a fixed vector on the Trp plane. The definition of θ is illustrated in Figure 4.2.4.

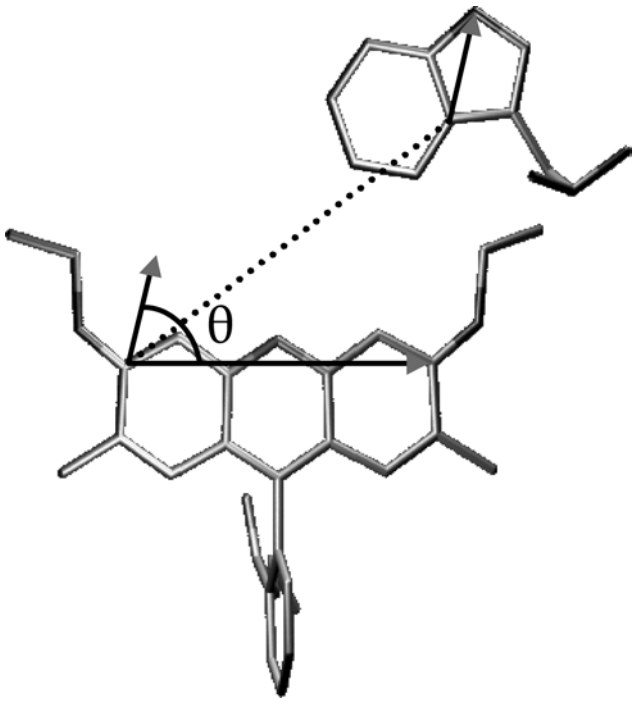

Figure 4.2.4: *Definition of the angle, θ between two fixed vectors lying on the dye and Trp planes. The R6G/Trp system is depicted here. The same definition is used for the MR121/Trp system.*

Subpopulations can be then classified according to the distribution of θ within each region A_1, A_2, and B. The Cartesian coordinates of the most frequently-occurring configurations (*i.e.*, those configurations falling within the peaks of the θ distribution) were averaged in order to obtain a representative structure of each subpopulation - these are shown in Figures 4.2.5 and 4.2.6.

For the R6G/Trp system, stacking of the Trp occurs only on one side of the dye plane whereas stacking on the other side is hindered by the presence of the R6G ester group. In contrast, due to the symmetry of the MR121 dye, in the MR121/Trp system the Trp interacts with both sides of the MR121 plane. The resulting average structures within A_1 and A_2 for MR121 consist of two pairs of almost perfect mirror images (panels marked A_1 and A_2 in Figure 4.2.5). The fact that the free energy minima (A_1 and A_2 in Fig. 4.2.3) are shallower for R6G than for MR121 is also due to the presence of the ester group on R6G, which sterically hinders the dye/Trp interaction. This effect might conceivably be reduced in rhodamine derivatives that have smaller groups at the ester position.

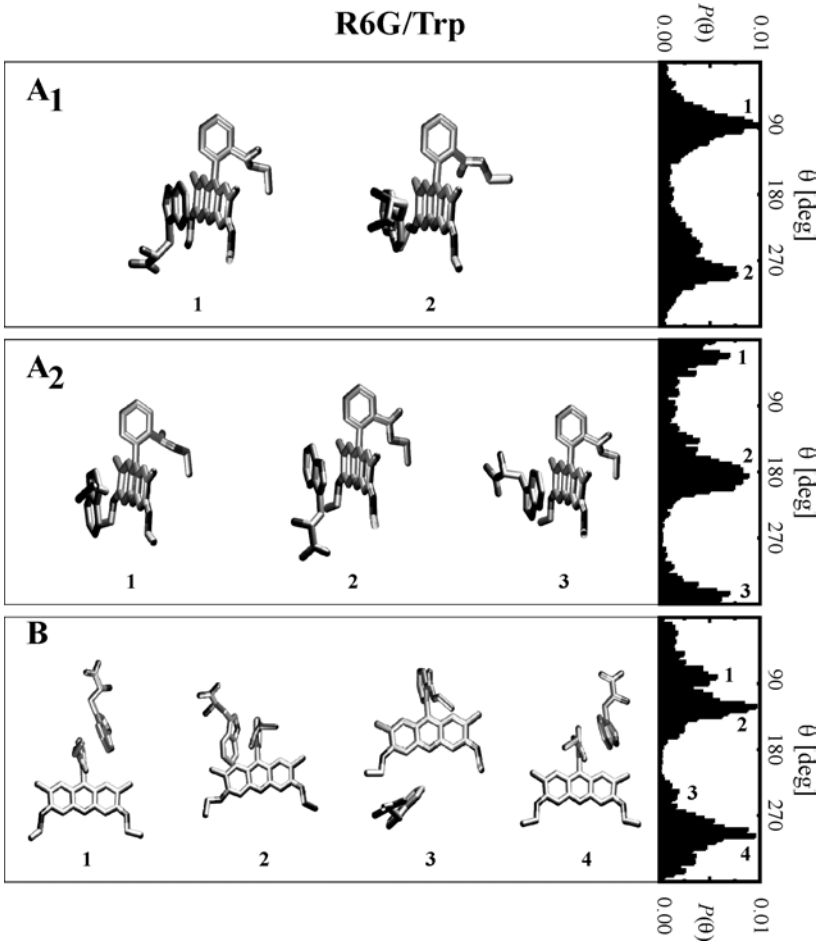

Figure 4.2.5: *Average structures of the most populated subpopulations of the R6G/Trp complex occurring during the simulation runs within the areas defined in Figure 4.2.3 and marked A_1 (top panel), A_2 (middle panel) and B (bottom panel). Corresponding probability distributions of the angle θ are on the right of each panel.*

The average structures of the complexes along the connection/escape pathway between minima are shown in panel B of Figures 4.2.5 and 4.2.6. The most frequently-occurring configurations for the R6G case (Figure 4.2.5, panel B) are those in which the Trp interacts mainly with the phenyl ring of the rhodamine dye (pictures 1, 2 and 4 from the left). Visual inspection of the simulation trajectories showed that these configurations often lead to escape/entrance of the tryptophan to/from regions of the landscape beyond r=10 Å. Interactions between the Trp and the phenyl ring of R6G, which are obviously not present in the case of MR121, are responsible for the fact that in Figure 4.2.2 *P(r)* of R6G is higher than MR121 in the region 5-10Å. This Trp/phenyl ring is a non-quenching interaction as is the Trp/ester group interaction mentioned in the previous paragraph. These two interactions compete with the quenching complexed geometries. For the case of MR121 (Figure 4.2.5, panel B), detailed inspection of the trajectories revealed that the Trp is most often found to slide along the MR121 plane before actually separating from the dye.

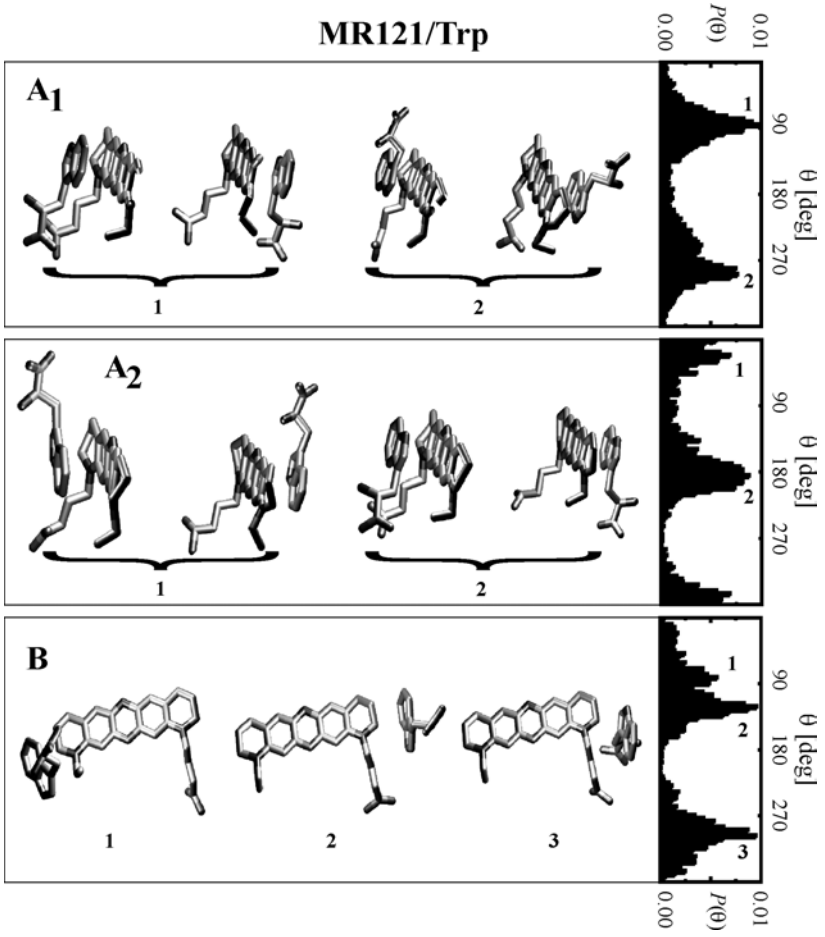

Figure 4.2.6: *Average structures of the most populated subpopulations of the MR121/Trp complex occurring during the simulation runs within the areas defined in Figure 4.2.3: A_1 (top panel), A_2 (middle panel) and B (bottom panel). Corresponding probability distributions of the angle θ are on the right of each panel.*

The bimolecular dynamic quenching rate, $k_{q,d}$ can be derived from the reduction in the average fluorescence lifetime, τ (Eq. 4.1.1), of the dye in the presence of quencher relative to the free dye. This reduction is due to dye/Trp collisional encounters that quench the dye fluorescence earlier than the decay from the excited state of the free dye. τ (or equivalently $k_{q,d}$) is a purely dynamical quantity: it represents the average duration of non-quenched states of the dye, and it is independent of K_S.

Values of τ were obtained from Eq. 4.1.1 using the experimental values of $k_{q,d}$ and τ_0 from Table 4.1.1 and the Trp concentrations used in the simulations *i.e.*, 55 mM for the MR121/Trp system and 57 mM for the R6G/Trp. The resulting values of τ are 2.20±0.07ns for R6G/Trp and 1.40±0.04ns for MR121/Trp.

In principle τ can be estimated from MD simulation given a geometric definition of non-quenched states. The definition assumed here is that used above *i.e.*, the dye is assumed to be quenched if $r < r^*$. With this definition, the duration of the ith non-quenched state occurring along the MD trajectory, $\Delta t_i=(t_{u,i}-t_{q,i})$ is simply the difference between the time $t_{u,i}$ at which the system passes from a quenched ($r < r^*$) to an unquenched ($r > r^*$) state, and the time $t_{q,i}$ at which it returns to a quenched state. τ is calculated from the MD data in a similar manner as for the experiment *i.e.,* by fitting an exponential decay function to the probability distribution for Δt_i. In order to compare with the experimentally-measured values of τ, transitions were excluded for which Δt_i was smaller than the time resolution of the instrument (which is ~50ps).

The time series of r during the MD runs are shown in Figure 4.2.7 for both dye/quencher systems. The total number of transitions found in the simulations is relatively small *i.e.*, 30 for R6G/Trp and 21 for MR121/Trp.

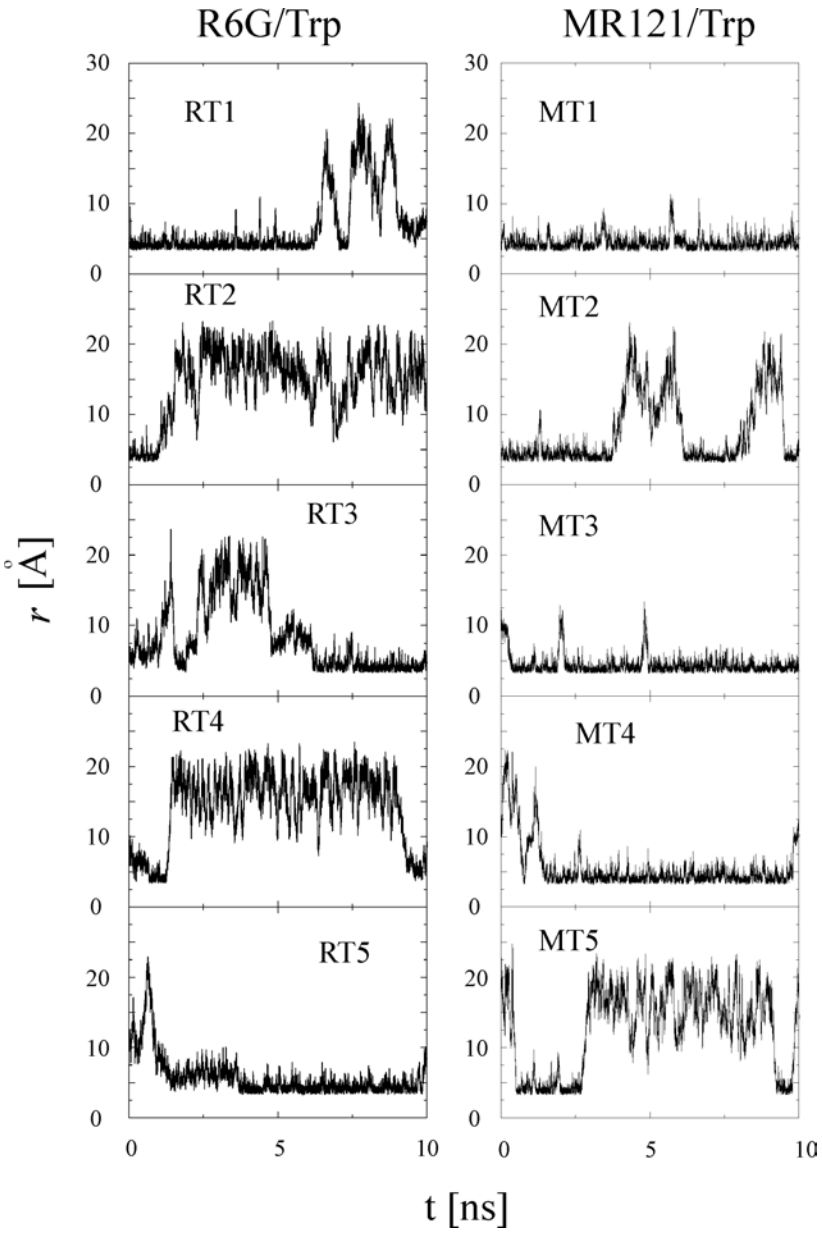

Figure 4.2.7: *Time series of r during the MD runs.*

The resulting values for τ are 1.7±0.2ns for R6G/Trp, (*cf.* the experimental value of 2.20±0.07ns) and 1.5±0.3ns for MR121/Trp compared to the experimental value of 1.40±0.04ns. Given the nanosecond timescale of the fluorescence lifetimes in the present systems, longer MD simulations would be required for an accurate evaluation of this property from the simulation model. Nevertheless, the values of τ obtained from the simulation agree with the experimental values rather well (to within experimental error for MR121/Trp and almost for R6G/Trp). That τ is of the same order as the lifetime of the free species, τ_0, confirms the small contribution of dynamical quenching relative to the contribution arising from stable, non-fluorescent complexes.

4.3 Conclusions

The bimolecular quenching results of this section demonstrate that the combination of MD simulations and time-resolved fluorescence experiments can provide an atomic-detail description of the geometry and kinetics of the quenching interaction of TRP with fluorescent dyes. Interaction geometries resulting from the MD simulations are matched to results from fluorescence quenching experiments.

In both systems studied the quenching mechanism is mostly static i.e., due to the formation of non-fluorescent complexes. At the TRP concentration used in the simulations and experiments, the MD-derived $P(r)$ distribution extends to higher distances for R6G than for MR121, due to the population of fluorescent interaction geometries between TRP and the R6G phenyl ring and ester group. As a consequence of this difference about 75% of the R6G dye is quenched in comparison with 90% of the MR121, as seen experimentally. Combining these observations with the MD-derived $P(r)$ allows determination of the 'quenching distance' r^*. This distance turns out to be physically reasonable and very similar (~5.5Å) for both dye/TRP systems, corresponding to close to van der Waals contact. Finally, the lifetimes of the fluorescent states are on the nanosecond timescale and agree to within experimental error for MR121/TRP and almost for R6G/TRP, further validating the quenching distance derived. To extend understanding of the detailed quenching process beyond the simple 'quenching distance' concept used here will require quantum chemical analysis by using, for example, the geometries derived from MD simulation as input for appropriate molecular orbital calculations.

Quenching interactions similar to those described above for free dyes and tryptophan molecules in solution also occur on an intra-molecular level in dye labelled tryptophan containing peptides [Neuweiler, et al. 2003]. The theoretical description of quenching proposed here for the bimolecular quenching process can, in principle be applied to the intra-molecular quenching processes in order to gain further insight regarding this phenomenon. This is done in the next chapter where the simulated system is a peptide sequence taken from a particularly interesting bio-medical application of intra-molecular quenching interactions: the detection of P53 antibodies, known to be specific tumor markers [Neuweiler, et al. 2002]. A brief introduction on the importance of P53 antibodies as tumor markers is given in the next section.

References

Zhong, D. and A. H. Zewail (2001). "Femtosecond dynamics of flavoproteins: Charge separation and recombination in riboflavin (vitamin B2)-binding protein and in glucose oxidase enzyme" Proc Natl Acad Sci U S A **21** (98): 11867-72.

Jorgensen, W., J. Chandrasekhar, J. Madura, R. Impey and M. Klein (1983). "Comparison of simple potential functions for simulating liquid water" J Chem Phys (79): 926-35.

Hill, T. L. (1956). Chapter 6. Statistical Mechanics, Principles and Selected Applications. New York, Dover Pubblications, Inc.: 179-285.

Chelli, R., F. L. Gervasio, P. Procacci and V. Schettino (2002). "Stacking and T-shape competition in aromatic-aromatic amino acid interactions" J Am Chem Soc **21** (124): 6133-43.

Neuweiler, H., A. Schulz, M. Böhmer, J. Enderlein and M. Sauer (2003). "Measurement of Submicrosecond Intramolecular Contact formation in Peptides at the Single-Molecule Level" J Am Chem Soc **18** (125): 5324-30.

Neuweiler, H., A. Schulz, A. C. Vaiana, J. C. Smith, S. Kaul, J. Wolfrum and M. Sauer (2002). "Detection of individual p53-autoantibodies by using quenched peptide-based molecular probes" Angew Chem, Int Ed Engl **24** (41): 4769-73.

5. MD Simulations of a Fluorescence Labelled p53 Epitope

5.1 p53 Antibodies - a Molecular Marker of Cancer

The tumor suppressor protein p53 is barely detectable in the nucleus of normal cells [Benchimol, et al. 1982]. p53 becomes active upon DNA damage. Some of the known functions of this protein involve preventing the mutation of a normal cell to a malignant cell: p53 can arrest cell cycle progression, allowing the DNA to be repaired, it also influences DNA replication, and it can lead to apoptosis [Soussi 2000]. Due to these key functions the protein has been defined as the "guardian of the genome" [Lane 1992]. The structure of p53 can be, in general, divided into three regions. The N-terminal region contains the transactivation domain, this is the target of several protein partners of p53. The C-terminal region contains the dimerization domain (p53 is a tetramer constituted by association of two dimers). Both N-terminal and C-terminal regions are strongly solvent exposed and therefore easily accessible for binding partners. Finally, the central region of the protein, containing several hydrophobic regions and a small number of charged amino acids. This core region is involved in the specific DNA binding function of the protein. This region has been explored by X-ray crystallography [Cho, et al. 1994].

Gene mutations in p53 are the most frequently found abnormalities in human cancer [Nigro, et al. 1989, Hollstein, et al. 1991]. More than 95% of these alterations in p53 are missense mutations that are scattered in the central part of the gene, corresponding to the central hydrophobic part of the protein, the crucial domain for its tumor suppressive functions. These mutations translate into faulty p53 proteins. They also have dramatic consequences in terms of protein stability. Mutant p53 proteins have a half-life of several hours compared to 20 minutes of the wild-type form. Mutated p53 accumulates in the nucleus of tumor cells and becomes immunologically detectable[Soussi 1996]. It has been shown that circulating antibodies against p53, so called p53 autoantibodies, are present in blood serum or ascites of cancer patients with various cancer types[Soussi 2000]. Furthermore, as shown in Fig. 5.1.1.b, the prescence of p53 antibodies is strongly correlated to p53 gene mutation[Lubin, et al. 1995a].

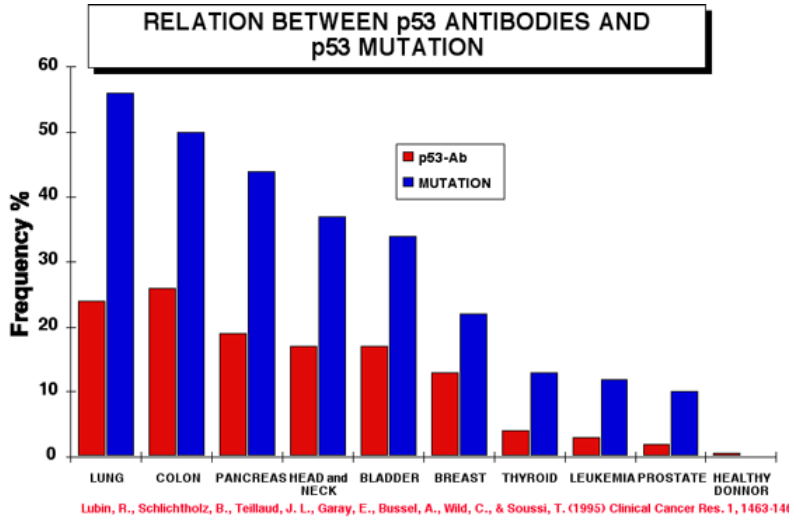

Figure 5.1.1: *Relation between the occurrence of p53 antibodies and p53 mutations taken from [Lubin, et al. 1995a].*

These findings have opened the way for the development of new highly specific, independent cancer diagnostic tools. It has been shown, that the presence of p53 antibodies validates a malignant disease with a specifity of 96%; in contrast, p53 antibodies are very rare in body fluids of healthy donors (less than 0.5%) [Soussi 2000]. It has also been demonstrated that in a few important cancer types, such as lung and liver cancer, the monitoring of p53 autoantibodies has the potential to be used as a standard assay for early-stage diagnosis and follow-up of malignancy [Lubin, et al. 1995b].

For the detailed analysis of the immune response to p53 Lubin and coworkers have determined the precise location of the p53 antigenic sites recognized by serum antibodies of cancer patients with different cancer types [Lubin, et al. 1993]. The results demonstrated that the immune response is restricted to a small subset of peptides localized in the amino and the carboxy terminus of human p53, these regions are among other things, characterized by a total absence of mutations. This finding is in agreement with the observation that p53 antibodies recognize wild-type or mutant p53 in a similar way [Lubin, et al. 1993]. The essential result is the presence of two immunodominant antigenic sites in the amino-terminal sequence of p53 (residues 11 to 35 and 46 to 65), independent on the cancer type. Whereas the epitopes in the carboxy-terminal region are not as well defined and range in general from residue 350 to 393. The structure and the immunodominant antigenic sites in human p53 are illustrated in Figures 5.1.2 and 5.1.3.

Both MR121 and R6G are environmentally sensitive fluorescence dyes and have been shown to be effective probes for the close proximity of Trp residues [Neuweiler, et al. 2003, Vaiana, et al. 2003]. The dyes have the tendency to form complexes with Trp in aqueous solution and intermolecular contacts in dye-labelled Trp-containing peptides. A strong quenching of dye fluorescence acts as a clear signal of dye/Trp contacts. Variations in the environment of the dye, for example the binding of an antibody to the target peptide can disrupt dye/quencher contacts causing dye fluorescence to be restored. This is especially true if the Trp residue is directly involved in the binding process. The two amino-terminal epitopes of p53 mentioned above have been labeled with MR121 and were successfully used for the development of highly sensitive p53 autoantibody immunoassays [Neuweiler, et al. 2002].

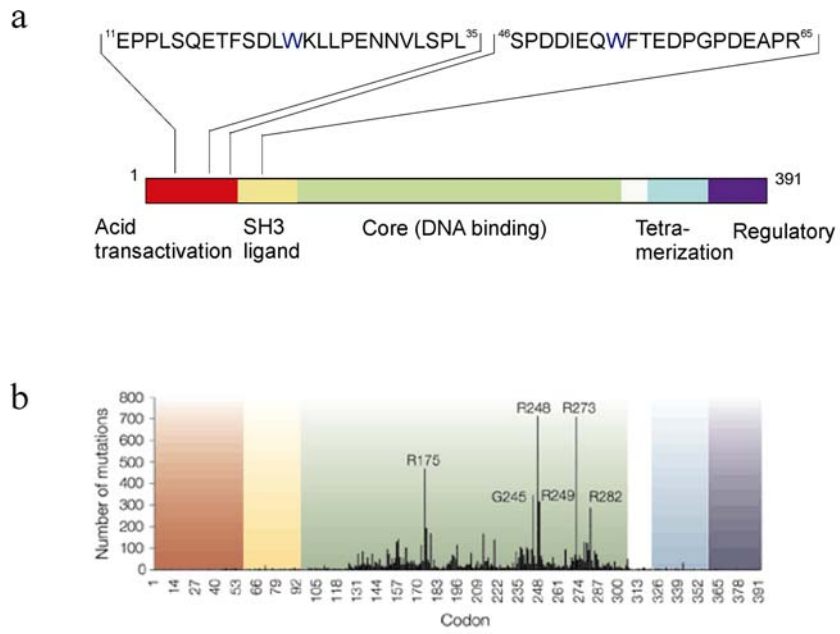

Figure 5.1.2: (a) *Schematic representation of p53 domain structure and immunodominant epitopes. Segment 1 (residues 11-35) lies within the amino-terminal transactivation region, while segment 2 (residues 46-65) lies mostly within the proline-rich SH3 ligand region.* (b)[*] *Histogram showing p53 missense mutations in cancer patients. 95% of mutations occur in the core domain; the six labelled residues are hot spots for mutations.*

[*] *taken from: Bullock A. N. and Fersht A. R., (2001) Rescuing the function of p53. Nat. Rev. Cancer, 1, 68-76.*

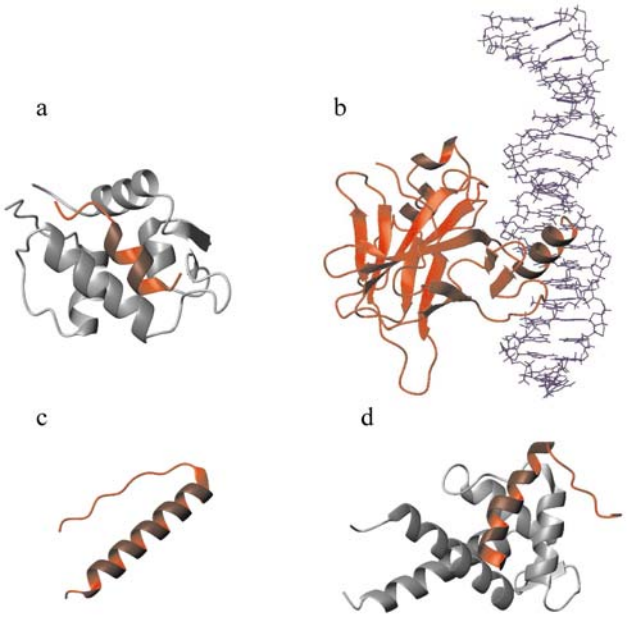

Figure 5.1.3: *Three-dimensional structures of p53 domains.* (a) *Crystal structure of a peptide from the amino-terminal transactivation domain of p53 bound to the N-terminal domain of the Mdm2 oncoprotein[Kussie, et al. 1996].* (b) *Crystal structure of the p53 core domain bound to double-stranded DNA [Cho, et al. 1994].* (c) *NMR structure of the p53 tetramerization domain [Clore, et al. 1995].* (d) *NMR structure of the negative regulator domain of p53 bound to S100B (ββ)[Rustandi, et al. 2000].*

5.2 Selection of the p53 Epitope

The peptide selected for analysis in this thesis is one of the two peptides from the sequence of p53 mentioned above. These peptides were used by Neuweiler et al. [Neuweiler, et al. 2002] for the development of a p53 autoantibody immunoassay. Lubin and co-workers [Lubin, et al. 1993] have showed that the immune response of patients with p53 antibodies is restricted to a small subset of peptides localized in the amino- and carboxy-termini of the human p53 protein, independent of the type of cancer. It could be shown that two short segments in the water-exposed amino-terminal transactivation domain are the most likely to be recognized by antibodies: segment 1 comprising the amino acids 11 to 35, and segment 2, the segment considered in this thesis, the amino acids 46 to 65. The structural properties of segment 1 in complex with a 109-residue fragment of MDM2, a regulatory binding partner of p53 have been studied by x-ray crystallography by [Kussie, et al. 1996]. The crystal structure reveals, that MDM2 has a deep hydrophobic cleft on which the p53 peptide binds as an amphipathic

α-helix. The interface shows steric complementarities between the MDM2 cleft and the hydrophobic face of the α-helix and, in particular, a triad of p53 amino acids - F19, W23, and L26 - which insert deep into the MDM2 cleft see Figure 5.2.1.

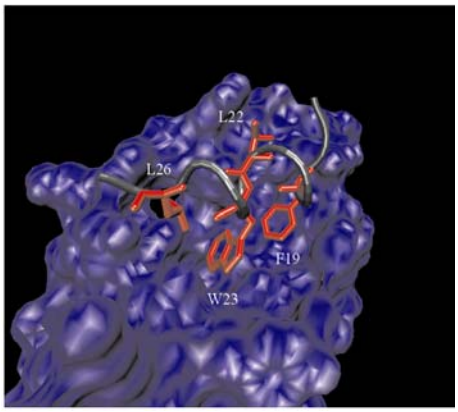

Figure 5.2.1: *Crystal structure of segment 1 in complex with a 109-residue fragment of MDM2. Amino acids - F19, W23, and L26 are shown.*

In developing the immunoassay mentioned above, one of the basic assumptions was that the MDM2 interface is similar to the antibody binding site. Although no similar crystal structure exists for the second immunodominant segment, it is interesting to note that segment 2 also contains a tryptophan and a phenylalanine residue (residues 53 and 54) which might be important interaction partners in the binding to the corresponding antibody. Indeed both segments, when labelled with MR121, have proven to work as highly sensitive probes for antibody binding. The peptide sequence for segment 2 is given in Figure 5.1.2.

The success of the approach illustrated in the preceding chapter in elucidating intermolecular quenching of both R6G and MR121 by TRP raises the question as to whether a similar approach may be used for the treatment of the more bio-medically relevant issue of intramolecular quenching of dyes covalently linked to TRP-containing peptide sequences. Although there are several similarities the case of intramolecular quenching by a TRP residue inherent in the peptide sequence presents some fundamental differences respect to the bimolecular quenching case. The first difficulty is due to the fact that the line of reasoning in which an unquenched free dye in solution is used as a reference for the experimental relative fluorescence quantum yield is no longer valid. In fact, it is clear that in the unquenched case both the dynamical behaviour and the local environment of the dye attached to the peptide will be very different from that of a free dye molecule in solution. A reference can be obtained by synthesizing a point-mutated peptide conjugate in which the TRP residue is substituted by another amino acid. In confronting experimental results which use this type of reference with a theoretical approach based on the probability distribution $P(r)$ of the distance

r between the TRP and dye ring systems for the interpretation of intramolecular quenching, one must make the assumption that the dynamical properties of the dye are not significantly changed by the mutation. A second theoretical difficulty arises from the larger size of the systems to be simulated and the longer time scales needed to explore the conformational landscape of a solvated peptide. This can be partially overcome by using a biased sampling technique such as umbrella sampling to explore areas of configurational space that would otherwise be inaccessible on the time scales which can be simulated. However, this has the drawback, compared to unbiased MD simulation, of allowing only the calculation of time averaged quantities.

5.3 Umbrella Sampling and the WHAM Equations, Theoretical Background

5.3.1 Umbrella Sampling

Calculation of free energies and PMF from simulation data requires a complete sampling of the possible configurations of the investigated system. Often, especially for biomolecular systems, the free energy surface is rugged and the barriers are too high to be overcome within reasonable computer time and the simulation 'gets stuck' in these free energy wells. Currently, MD trajectories on the nanosecond timescale can be achieved but the dynamics of interest often takes place in a time ranges of hundreds of nanoseconds or even microseconds. For these cases special sampling techniques have been developed to calculate free energy differences by means of MD simulations. One of these approaches is the umbrella sampling technique [Torrie and Valleau 1977]. Umbrella sampling attempts to overcome the sampling problem by adding a series of biasing potentials to the energy function of the system. The biasing potentials are used to enhance the sampling of otherwise unreachable areas of configurational space. This allows the system to explore larger regions of phase space during the time scale of normal MD simulations, the effect of the biasing potentials is then analytically compensated for in calculating the PMF. This is illustrated here for the case of a one-dimensional PMF calculation where the reaction coordinate is the distance *r* between the dye and the TRP as defined in chapter 4.

In umbrella sampling *M* simulations are performed with the biasing potentials $V_i(r)$, $i = 1, \ldots, M$. The total simulated potential can thus be expressed as a sum:

$$H(r,\mathbf{R}) = \sum_0^M V_i(r,\mathbf{R}) \qquad (5.3.1)$$

where $V_0(r, \mathbf{R})$ is the 'true', unbiased potential energy function and **R** is the 3*N*-1 dimensional vector of all space coordinates of the system except the distance *r*. The biased probability $P(r)$ of finding a given value of *r* is:

$$P(r) = \frac{\int e^{-\beta \sum_0^M V_i(r,\mathbf{R})} d\mathbf{R}}{\int e^{-\beta \sum_0^M V_i(r,\mathbf{R})} d\mathbf{R} dr} = \frac{\int e^{-\beta \sum_0^M V_i(r,\mathbf{R})} d\mathbf{R}}{Z} \qquad (5.3.2)$$

where $\beta = 1/kT$ and *Z* is the 'biased' partition function of the system. Whereas the desired unbiased probability $P_0(r)$ is:

$$P_0(r) = \frac{\int e^{-\beta V_0(r,\mathbf{R})} d\mathbf{R}}{\int e^{-\beta V_0(r,\mathbf{R})} d\mathbf{R} dr} = \frac{\int e^{-\beta V_0(r,\mathbf{R})} d\mathbf{R}}{Z_0} \qquad (5.3.3)$$

where Z_0 is the partition function for the true (unbiased) system. Combining Eqs. 5.3.3 and 5.3.2 one obtains an expression of the unbiased probability as a function of the biased probability and the biasing potential:

$$P_0(r) = \frac{Z}{\Omega Z_0} P(r) e^{\beta \sum_1^M V_i(r)} \qquad (5.3.4)$$

here Ω is simply the volume integral over all of **R**. From Eq. 5.3.4 it is possible to calculate the PMF $\omega_0(r)$, this can be simply expressed using Eq. 4.2.2, as:

$$\omega_0(r) = -\sum_1^M V_i(r) + \omega(r) + C(\beta) \qquad (5.3.5)$$

where $\omega(r)$ is the 'biased' PMF *i.e.*, as calculated from the biased probability distribution. From a theoretical point of view one can, in this manner, run a series of MD simulations with added biasing potentials to enhance the sampling and then *a posteriori* reconstruct the true probability distribution and PMF.

5.3.2 The WHAM Method

From a practical point of view it is more efficient to calculate the probability distribution and the PMF using an extension of the umbrella sampling method called weighted histogram analysis method (WHAM) [Kumar, et al. 1992]. In this case each biasing potential, typically a harmonic function of the reaction coordinate, is used to enhance sampling in a small region, referred to as a window. A total of *M* MD simulations are performed, one for each biasing potential. After compensating as above for the bias introduced by the potentials $V_i(r)$, an

unbiased probability distribution is obtained for each window and the M distributions are stitched together. For this purpose it is important that the distributions from adjacent windows overlap so that all points in r are well sampled.

In order to construct the probability distribution:

$$P_0(r) = \frac{1}{Z_0} e^{-\beta V_0(r)} \tag{5.3.6}$$

WHAM combines the M measured probability distributions $P_{0i}(r)$, corresponding to potentials $V_{0i} = V_0 + V_i$, in a weighted sum such that the biasing potentials V_i are discounted. For this purpose one employs the expression:

$$P_0(r) = \sum_1^M \kappa_i(r) e^{\beta V_i(r)} \frac{Z_0}{Z_{0i}} P_{0i}(r) \tag{5.3.7}$$

Here Z_0/Z_{0i} is the ratio of the configuration integral of the unbiased system to that of the system with the additional biasing potential V_i. The weighting factors $\kappa_i(r)$ are subject to the normalization constraint:

$$\sum_1^M \kappa_i(r) = 1 \tag{5.3.8}$$

and are chosen such that the variance of $P_0(r)$, $\left\langle P_0^2(r) \right\rangle - \left\langle P_0(r) \right\rangle^2$ is minimized. The assumption that the M simulations are statistically independent of one another yields an expression for the weights [Gullingsrud, et al. 1999],

$$\kappa_i(r) = \lambda(r) N_i \frac{Z_{0i}}{Z_0} e^{-\beta V_i(r)} \tag{5.3.9}$$

where N_i is the number of data points taken to construct the probability distribution in simulation i and where $\lambda(r)$ is the Lagrange multiplier which serves to enforce the normalization condition 5.3.8. Substituting 5.3.9 into 5.3.7 one obtains

$$P_0(r) = \frac{\sum_1^M P_{0i}(r) N_i}{\sum_1^M \left(Z_0 / Z_{0i}\right) N_i e^{-\beta V_i(r)}} \tag{5.3.10}$$

with

$$\frac{Z_0}{Z_{0i}} = \int_{r_0}^{r_f} dr P_0(r) e^{-\beta V_i(r)} \tag{5.3.11}$$

Eqs. 5.3.10 and 5.3.11 can be used to obtain $P_0(r)$ and, hence, $\omega_0(r)$. In practice, 5.3.10 and 5.3.11 are solved iteratively until they converge.

5.4 Results of Fluorescence Quenching Experiments

As for the bimolecular case of the previous chapter, the experiments on the labelled native and mutated peptides described here were performed by M. Sauer and co-workers at the University of Heidelberg. The main results of these measurements are summarized here. The Trp-containing peptide epitope (segment 2) was labeled at the amino-terminus with R6G (denoted here as PEP/R6G) and MR121 (PEP/MR121). Additionally, point-mutated labeled peptide conjugates were synthesized, where the Trp residue (W^{53}) was replaced by a single point mutation with a phenylalanine (F) residue (these peptides are denoted here as PEP*/R6G and PEP*/MR121).

The absorption and emission characteristics of the synthetic wild-type and mutated peptide conjugates were measured by steady-state fluorescence and absorption spectroscopy and are summarized in Table 5.4.1. The relative fluorescence quantum yields $\Phi_{f,rel}$ were determined with respect to the point mutated conjugates (*i.e.* PEP*/R6G and PEP*/MR121).

Table 5.4.1.: *Ensemble spectroscopic characteristics of the labeled peptide conjugates. Measurements were performed in phosphate buffer (pH 7.7) at room temperature. λ_{abs} and λ_{em} are the absorption and the emission maxima of the measured probes, $\Phi_{f,rel}$ denotes the relative fluorescence quantum yield respect to the point mutated conjugates, and τ_i the fluorescence lifetimes with corresponding amplitudes a_i.*

	λ_{abs} (nm)	λ_{em} (nm)	$\Phi_{f,rel}$	τ_1 (ns)	a_1	τ_2 (ns)	a_2	χ^2
PEP/MR121 [(1)]	667	683	0.17	0.51	0.09	2.00	0.91	1.138
PEP*/MR121 [(1)]	664	677	1.00	0.61	0.12	2.26	0.88	1.071
PEP/R6G [(2)]	531	558	0.23	0.64	0.07	3.73	0.93	1.175
PEP*/R6G [(2)]	529	561	1.00	0.10	0.01	4.37	0.99	1.097

[(1)] Data from [Neuweiler, et al. 2003]
[(2)] H. Neuweiler (unpublished results)

The fluorescence quantum yields of the conjugates reflect the expected selective fluorescence quenching of MR121 and R6G by Trp residues. There is a dramatic decrease in fluorescence quantum yield for peptides containing Trp (83% and 77% for PEP/MR121 and PEP/R6G, respectively), which can be attributed to intramolecular complex formation between dye and Trp.

In contrast to the steady-state fluorescence intensities, the fluorescence lifetimes of the Trp containing peptides remain nearly unchanged compared to the mutated ones. PEP/MR121 and PEP*/MR121 show similar fluorescence decays with approximately 10% of a shorter fluorescence lifetime of ~ 0.5-0.6 ns, and a main component with a lifetime slightly longer than the lifetime of the free dye of ~ 2 ns. As already observed in the intermolecular quenching experiments, complexes are formed between Trp and MR121 that exhibit very efficient fluorescence quenching. Such ultrafast processes fall beyond the detection limit of the experimental set-up used. These complexes can effectively be considered as non-fluorescent. In the case of the R6G labeled peptides the short lifetime component is rather different between the TRP containing peptide and the mutated one (0.6 ns compared to 0.1 ns). The amplitudes of the short lifetime components are in any case rather small. All these lifetime components of 100-600 ps could be ascribed to slight quenching of the other amino acids.

5.5 Simulations

5.5.1 MD Simulations of a Fluorescently Labelled Peptide

Simulation Protocol

A first set of 6 MD simulations was performed on the MR121 labeled PEP/MR121 system. Each simulation was performed with different initial TRP/dye separations, as for the bimolecular simulations of chapter 4, so as to improve configurational statistics. All simulations were performed in the NPT ensemble at 1.0 atm pressure and at 300 K in an orthorhombic box containing ~2150 TIP3P water molecules and 6 sodium counter-ions to neutralize the system. Long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) method described in section 2.2.1.

The heating and equilibration procedure was the same for all the MD runs. Three arbitrary starting configurations of PEP/MR121 were selected from a high temperature (800 K) vacuum unconstrained dynamics simulation by taking one snapshot every 1 ns. The remaining three initial structures were obtained by continuing the high temperature dynamics and applying a harmonic constraint $C(r) = k(r-r_0)^2$ on the distance r, between the geometrical centres of the main ring systems of the dye and of the tryptophan (the definition of r is shown in Figure 4.2.1) for a total of 1 ns. Values of the equilibrium distance r_0 used were: 15 Å, 20 Å, and 25 Å respectively. All constraints were removed at this point. Each starting configuration was solvated in TIP3P and counter-ions, waters overlapping with the peptide were removed and a steepest descent minimization was performed in order to remove additional stress due to close contact interactions. Each solvated system was then subjected to 6 ps of heating (CVT and 0.1 fs integration time) and 100 ps (CPT and 1 fs integration time) equilibration during which velocities were reassigned every 100 steps. Subsequently the systems were allowed to relax for another 100ps prior to the production phase. The production runs were all 5 ns long for a total of 30 ns simulation time.

Analysis of the Trajectories

As for the previous bimolecular case, in order to monitor interaction geometries between MR121 and Trp, the distance and the angle between the chromophore and indole derivative (as described in chapter 4) were extracted from the simultation trajectories. In addition, the trajectories were analysed for structural and conformational aspects of the peptide.

No secondary structural elements were detected during conformational analysis of the peptide backbone. In Figure 5.5.1 the time series of the distance r are plotted for each run.

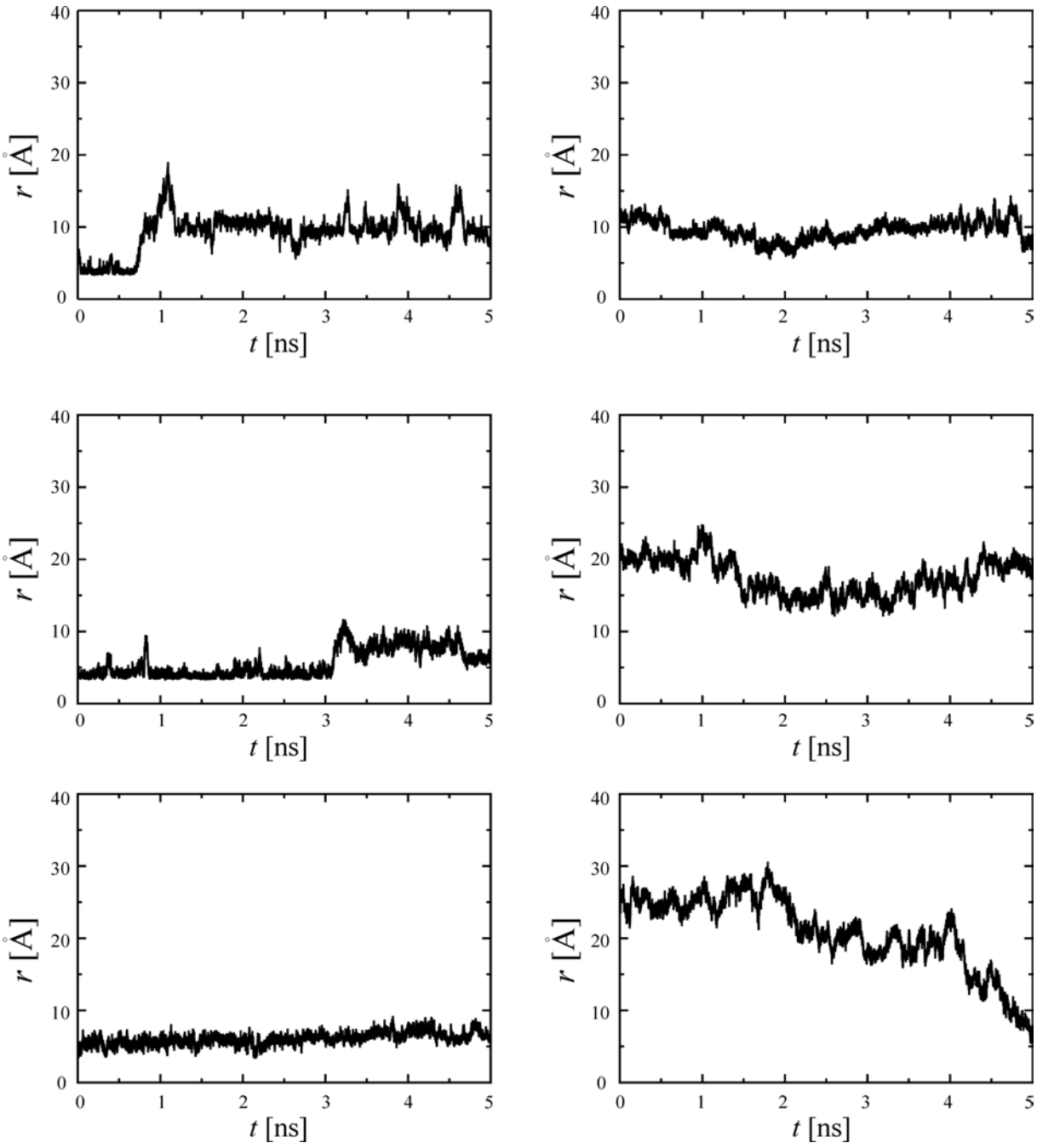

Figure 5.5.1: *Time series of the distance r(t) between the dye and the TRP during the six 5ns MD runs of the PEP/MR121 system, transitions between quenched and unquenched configurations are even less frequent than in the bimolecular case of Fig. 4.2.7.*

The main feature noticeable in Figure 5.5.1 is that the transitions within each simulation between quenched configurations *ie.*, values of r corresponding to close contact interactions of the dye and the TRP residue, are extremely rare (less than 7 for quenching distances of $r \leq 5.4$ Å).

As already mentioned, in order to have a good estimate of the PMF from MD simulations the configurational space must be well sampled. From both the time series $r(t)$ and the probability

distributions $P(r)$ (Figure 5.5.2), referring to each of the simulations performed on the PEP/MR121 system, it is clear that the system is stuck within a few minima of the free energy landscape and sampling of the r space is insufficient for a correct statistical analysis of this system. Due to the higher degree of complexity of this system compared to the simple bimolecular systems of the previous chapter, a sufficient sampling of configurational space would require much longer simulations (at least in the hundreds of ns range).

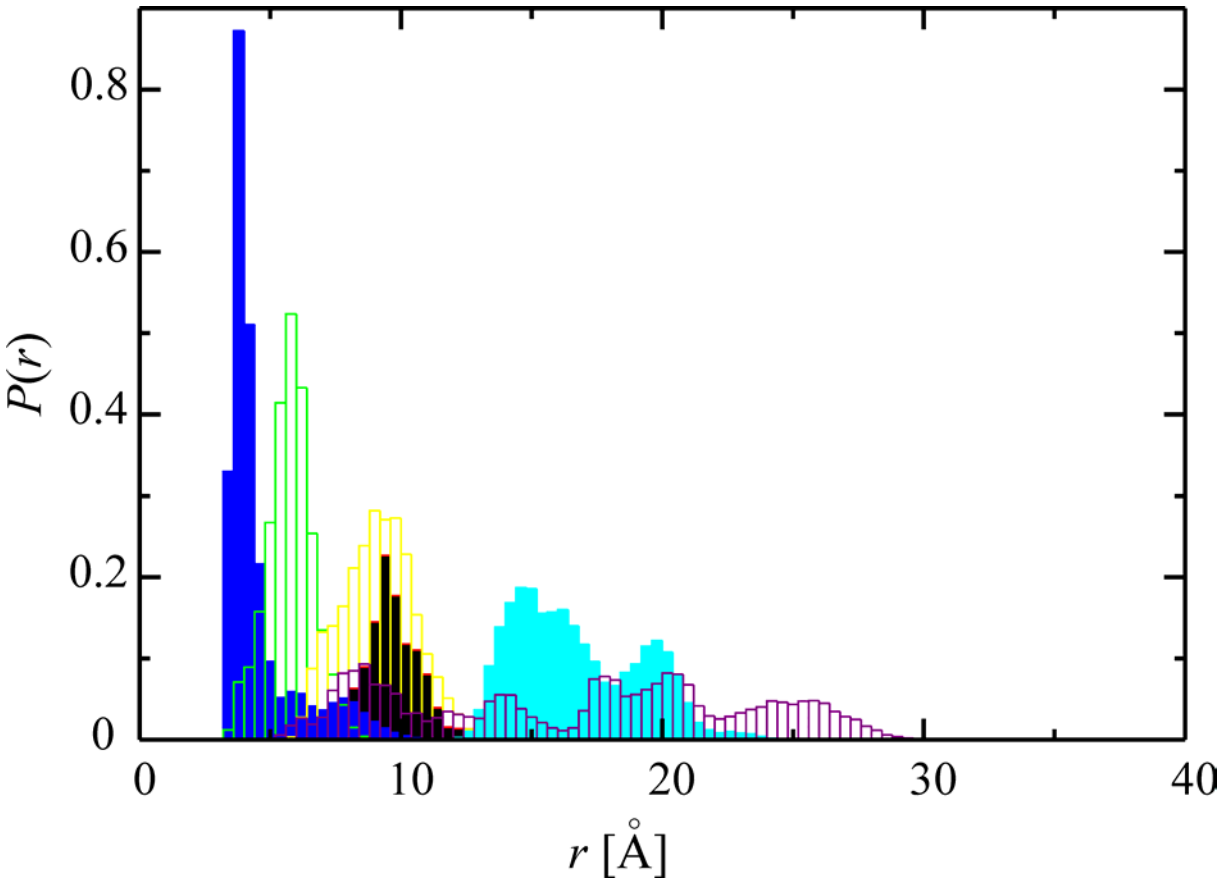

Figure 5.5.2: *Normalised probability distributions of the distance r between the dye and the TRP for six 5ns MD runs of the PEP/MR121 system, distributions lack sufficient overlap for statistical analysis even of the one-dimensional PMF landscape.*

The simulations mentioned above performed on the PEP/MR121 system required a total of ~120 CPU days of computer time on 8 Dual Athlon 1.2 GHz processors connected *via* Fast Ethernet cards *i.e.*, 15 days of real time. Under these same conditions, simulating a total of 100 ns on both PEP/MR121 and PEP/R6G would require 800 CPU days *i.e.*, 100 days of real time. Using the umbrella sampling method, the one dimensional PMF and associated probability distribution can be obtained, with reasonable accuracy, in a fraction of the computer time. One must however, in this case, renounce obtaining time dependent quantities such as lifetime information.

5.5.2. Umbrella Sampling Simulations of PEP/R6G and PEP/MR121

Umbrella sampling was performed on both systems solvated in explicit water. The distance between the centres of geometry of the chromophore and the indole ring system of Trp was used to monitor the process of unstacking. The reaction coordinate for both systems ranged from r_0 = 3.0 to r_0 = 8.0 Å. 15 sampling windows were used, the umbrella potentials were of the type $V_i(r)=k/2(r-r_{oi})^2$. The force constant k was set to 5.0 kcal/(mol Å) values of r_{oi} were chosen starting from 4.0 Å and incremented by 0.25 Å. In each sampling window a first 10 ps of equilibration was performed followed by 100 ps of production dynamics yielding a total of 1.5 ns dynamics for each system. For the calculation of distance distributions from the time series, the reaction coordinate was binned into 0.1 Å distance intervals. The unbiased probability distributions and their relative one dimensional PMF profiles were calculated using the WHAM algorithm [Kumar, et al. 1992].

Figures 5.5.3 and 5.5.4 show the distance time series under the applied harmonic potentials for the intramolecular separation distance r between MR121 and Trp in PEP/MR121 and between R6G and TRP in PEP/R6G, respectively. Additionally, characteristic structures of the systems during the transitions between the "closed" and the "open" configurations are shown. The $r(t)$ plots demonstrate that the balance chosen between the force constant of the umbrella potentials and the distance increments between the sampling windows was correctly chosen: the distance distributions of adjacent sampling windows are completely overlapping for both dye/peptide systems.

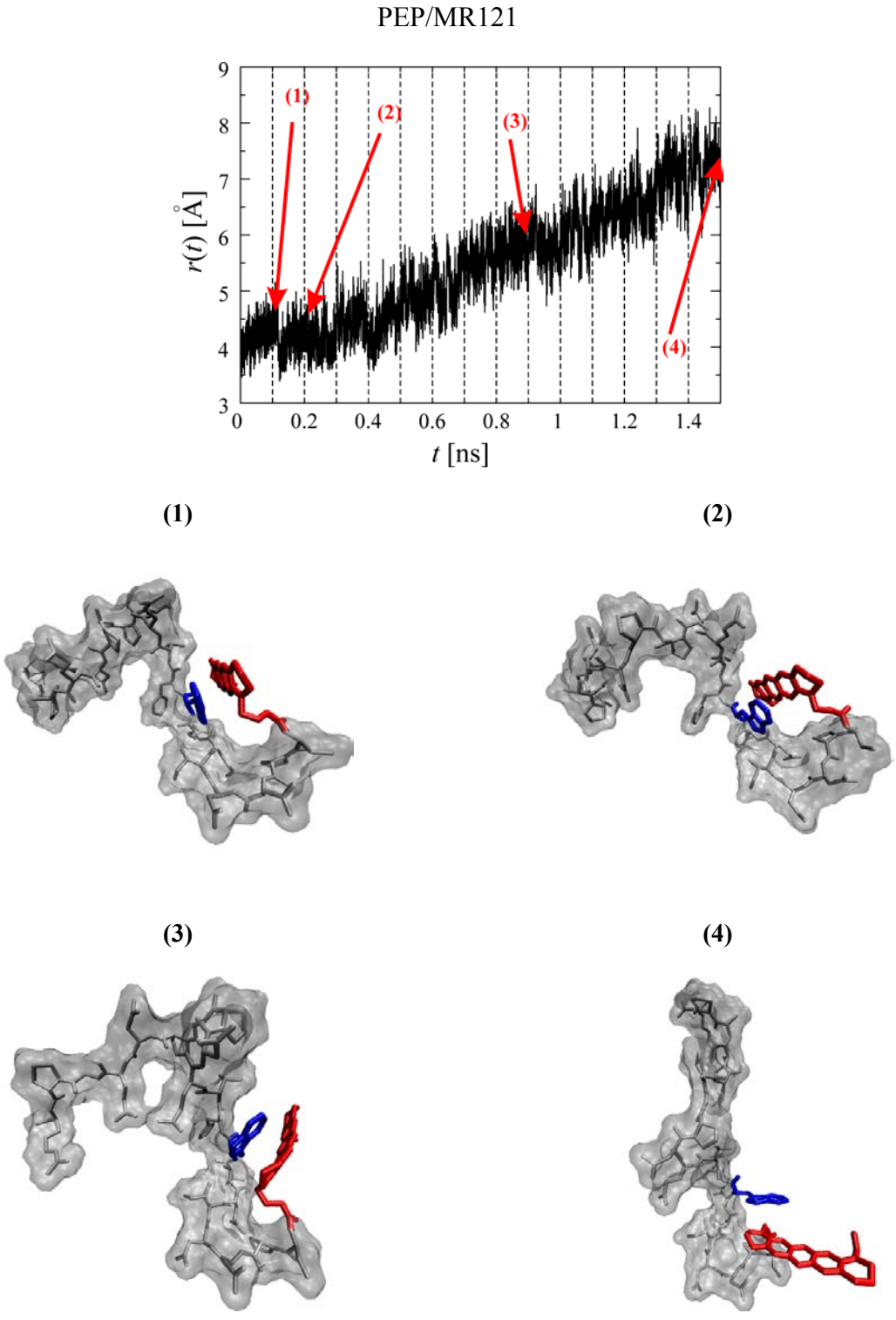

Figure 5.5.3: *The time series r(t) of the PEP/MR121 system during the umbrella sampling process (top). Adjacent sampling windows are evidenced by vertical dashed lines. Overlap between adjacent windows is very good. Characteristic structures taken at sample points 1-4 are shown. TRP (blue) and MR121 (red) relative positions are evidenced. Water accessible surface of the remaining residues is also shown.*

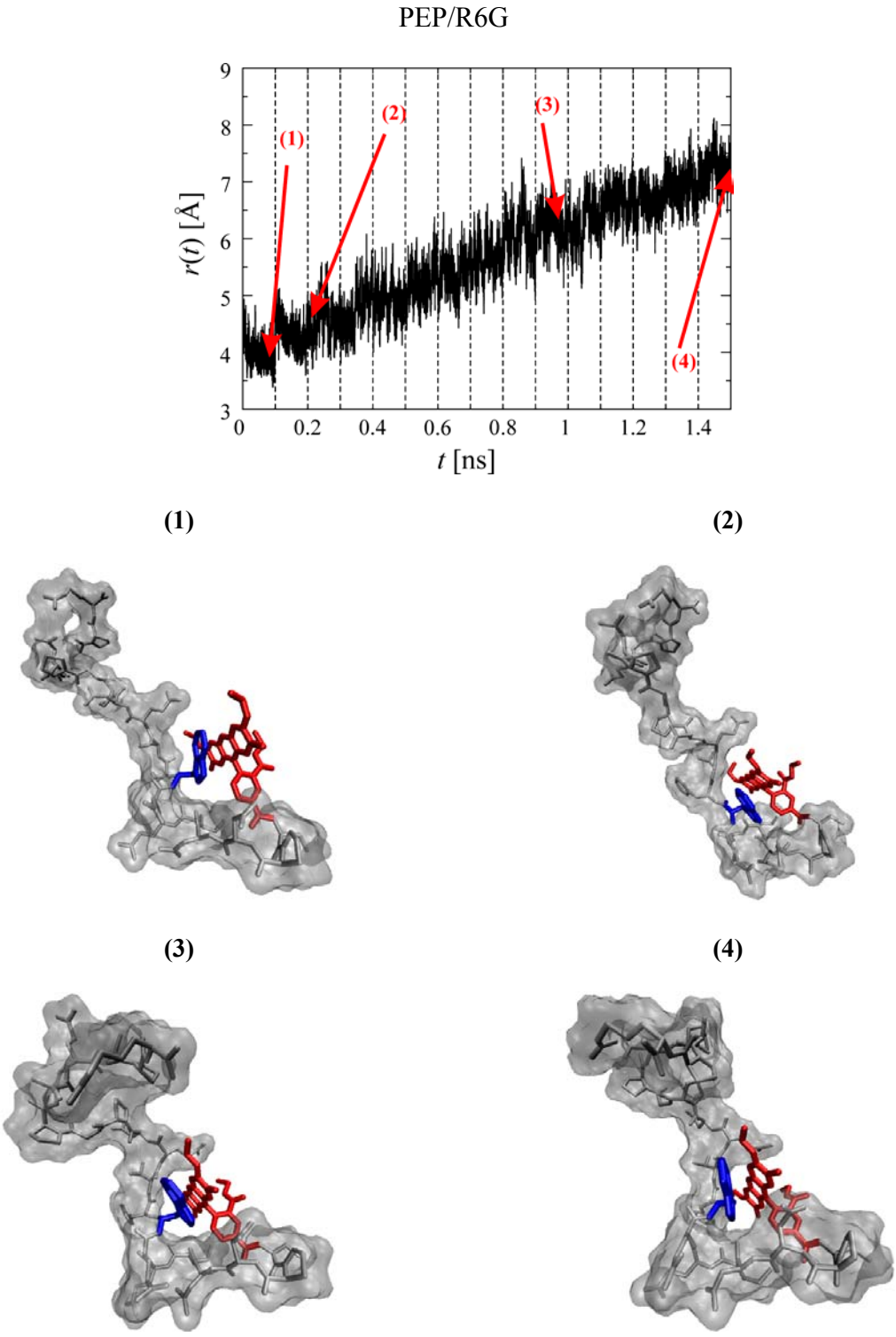

Figure 5.5.4: *The time series r(t) of the PEP/R6G system during the umbrella sampling process (top). Adjacent sampling windows are evidenced by vertical dashed lines. Overlap between adjacent windows is very good. Characteristic structures taken at sample points 1-4 are shown. TRP (blue) and R6G (red) relative positions are evidenced. Water accessible surface of the remaining residues is also shown.*

Figure 5.5.5 shows the resulting one dimensional PMF profiles for both PEP/MR121 and PEP/R6G and also, for comparison, the profile resulting from the 30 ns of unconstrained MD runs. Both PMF profiles from the umbrella sampling runs are characterized by a deep minimum corresponding to coplanar stacking interaction geometries between Trp and MR121 at a distance of ~3.7 Å. The depth of the minimum for the case of PEP/MR121 demonstrates that these configurations are very stable. A comparison between the two profiles for PEP/MR121 reveals insufficient sampling of the minimum PMF region resulting from the MD simulations.

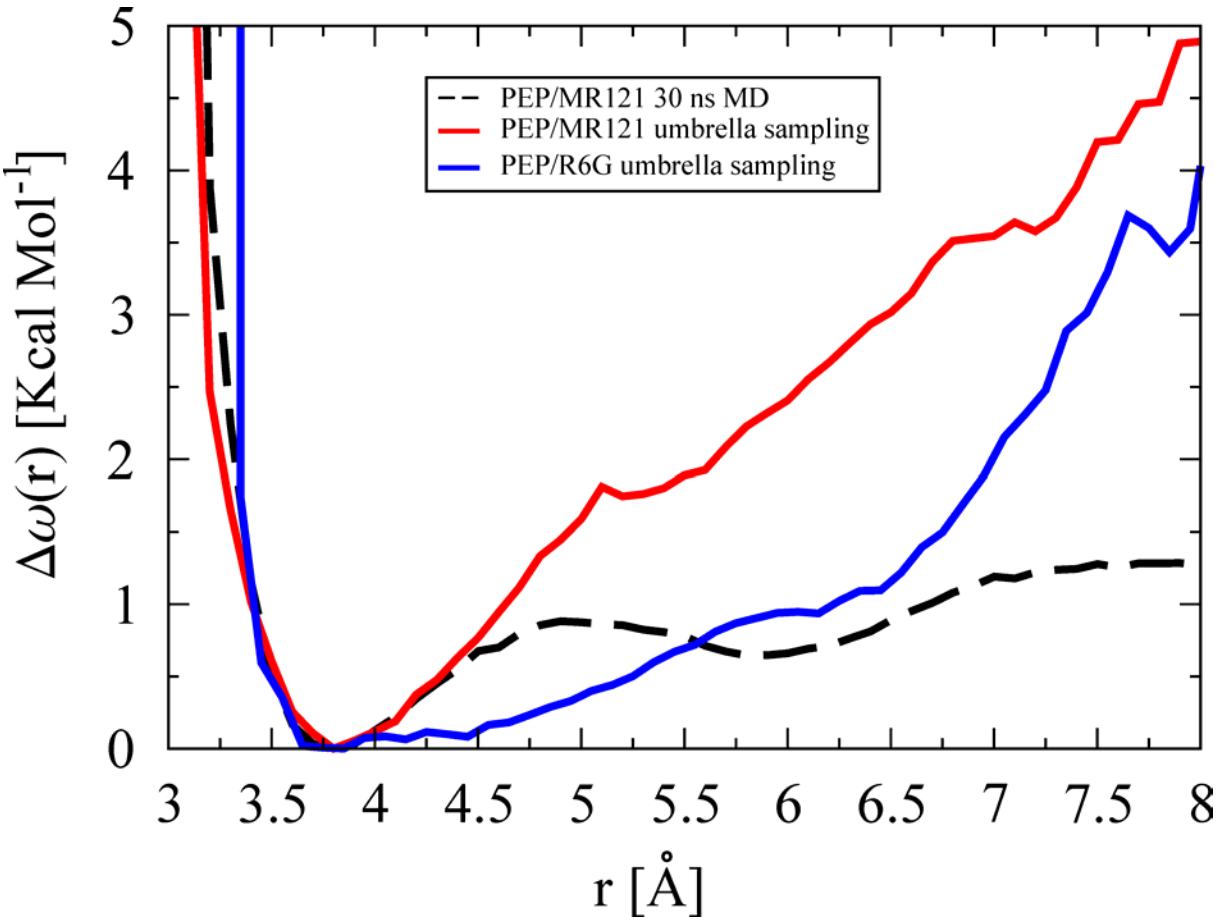

Figure 5.5.5: *One dimensional umbrella sampling PMF profiles Δω(r) for PEP/MR121 (red) and PEP/R6G (blue). The dashed black line refers to the PMF profile for PEP/MR121 as calculated from the MD simulations of section 5.5.1.*

The PMF profiles for PEP/MR121 and PEP/R6G show significant differences. As was the case for the bimolecular studies, the PMF minimum for the MR121 labelled system (red line in Figure 5.5.5) is more pronounced than for the R6G system (blue line in Figure 5.5.5).

In Figure 5.5.6 the unbiased probability distributions $P(r)$ for both PEP/MR121 and PEP/R6G are shown and the quenching distance r^* is derived as for the bimolecular case (section 4.2).

The experimental values of the relative fluorescence quantum yield $\Phi_{f,rel}$, *i.e.*, the probability of the dye being unquenched were taken from Tab. 5.4.1 ($\Phi_{f,rel}$ = 0.17 for the PEP/MR121 system and $\Phi_{f,rel}$ = 0.23 for the PEP/R6G system). It is clear from Figure 5.5.6 that both the probability distributions and their integrals $F(r)$ (see section 4.2) are very different. The tendency observed for the bimolecular case of a higher relative population for the R6G/TRP system is here even more evident. This fact cannot be ascribed, as was shown to be for the bimolecular quenching case, simply to the presence of the ester side chain of R6G (not present in MR121). In fact from the structures shown in Figures 5.5.3 and 5.5.4 it is clear that for both dye labelled systems, the configuration adopted by the peptide forces the TRP to interact only with one side of the dye, and for the case of PEP/R6G this side is the one which does not contain the hindering ester tail. The ester tail adopts a solvent exposed configuration.

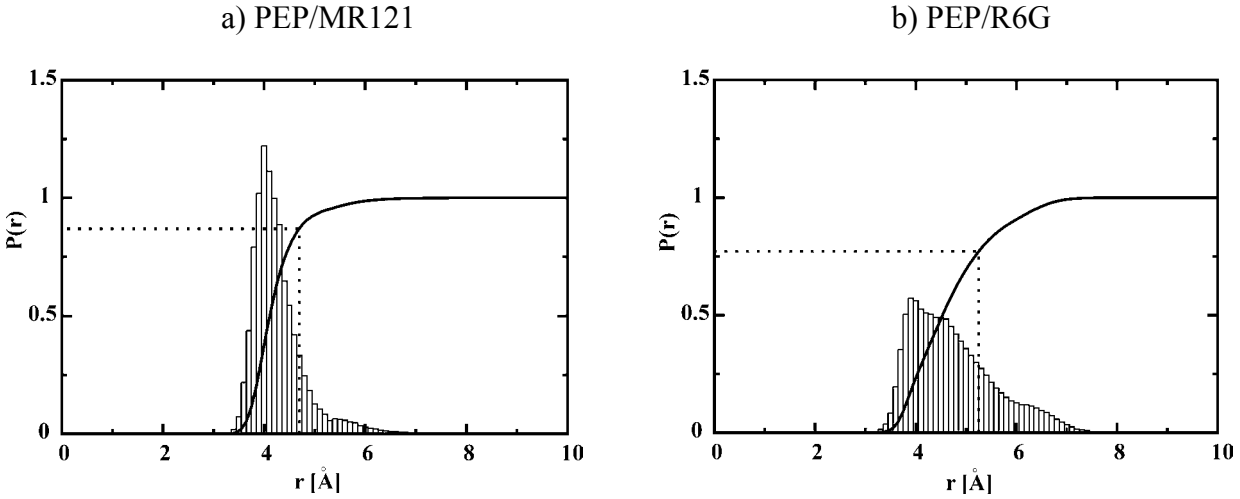

Figure 5.5.6: *Normalized probability distributions P(r) for PEP/MR121 (a) and PEP/R6G (b). The continuous line is the integral of P(r), i.e., F(r*) defined in Eq. 4.2.5. The dotted line indicates the value of r*=(4.7±0.2)Å for PEP/MR121 and r*=(5.3±0.2)Å for PEP/R6G, corresponding to the experimental quenching efficiencies of 83% and 77%, respectively.*

The shape of the probability distributions $P(r)$ is most probably a result of complex interplay between solute-solute and solute-solvent effects. In the PEP/MR121 case the peptide maintains a rather loose configuration which allows solvent exposure of the dye and the TRP residue from all sides and the configurations adopted by the two residues are very similar to the most favourable ones seen in the deep PMF minima of section 4.2 (Figure 4.2.6 configurations A1-1 and A2-2), these configurations in the bimolecular case referred to distances up to ~5 Å. The case of PEP/R6G is somewhat different, the peptide adopts a more closed structure forming a solvent shielded pocket in which the TRP is inserted and the rhodamine 'closes' the pocket exposing mainly the ester tail to the solvent. The configurations

adopted by the backbone of the peptide force the TRP to interact with the R6G in a spacial arrangement resembling that of Figure 4.2.5 configuration A2-1, whereas configurations similar to those of minimum A1 in section 4.2 are not present. Therefore the relative probability of closed configurations for the PEP/R6G system is almost half as large as that for PEP/MR121. On the other hand, the broader probability distribution for the PEP/R6G system (between 4 and 5 Å) partially compensates the lower peak and this is reflected in a fluorescence quenching efficiency of 77%, only slightly lower than that of the PEP/MR121 system *i.e.*, 83%.

In contrast to the bimolecular case, here also the quenching distances r^* are slightly different: for the PEP/R6G the value of $r^* = (5.3\pm0.2)$Å is very similar (practically the same if error bars are considered) to the value obtained for the bimolecular quenching of R6G by TRP *i.e.*, $r^*=(5.5\pm0.4)$Å. The result for the PEP/MR121 system is $r^* = (4.7\pm0.2)$Å. Although slightly smaller than the value obtained in the bimolecular case i.e., $r^* = (5.4\pm0.2)$Å, this value is still within an acceptable range for van der Waals contact.

The fact that the "closed" configurations (Figure 5.5.3, frame 1) are structurally very different from the "open" ones (Figure 5.5.3, frame 4) for the PEP/MR121 system may explain why this peptide experimentally behaves so well as a molecular probe of p53 antibodies. In solution, the free peptide adopts the low free energy closed conformations and dye fluorescence is mostly quenched. Experimentally it is known that upon selective binding to the antibody the dye releases its fluorescence. Using the distance dependent interpretation of quenching proposed here this would mean that upon binding (or maybe shortly before the binding event) the peptide finds itself in one of the "open" structures above. The binding process must somehow freeze this "open" structure by imposing a similar "open" conformation to the peptide in the bound state (or the part of the peptide containing the TRP and dye). This same reasoning would suggest that, instead the PEP/R6G system, although efficiently quenched in the "closed" state, would probably not behave as well i.e., in terms of the release of fluorescence, in the "open" state, where the structure of the peptide is not significantly different from the "closed" one. Unfortunately no binding experiments have yet been performed with this system.

A full fledged MD treatment including very long simulations of the peptide systems described here would be necessary for a complete mapping of the PMF and related probability landscapes and also for obtaining lifetime information. Nonetheless, the umbrella sampling approach used here is still able to provide valuable picture of the intramolecular quenching mechanism of the two dyes in terms of the $P(r)$ distributions and the structural properties of

peptide when labelled with MR121 or R6G. This is done by using only a fraction of the computational resources that would be required for an, unbiased exploration of configurational space. This is very important if this approach is to be used for *in silico* testing either of new dyes or of point mutations within the peptide with the goal of optimising the "closed" quenched states (to favour quenched conformations of the free peptide in solution) while maintaining flexible "open" states (to favour binding to the antibody and simultaneous fluorescence release).

5.6 Conclusions

30 ns of MD simulation were not sufficient to obtain an adequate sampling of conformational space for the peptide systems studied here. The sampling problem is overcome by means of "biased" sampling methods (umbrella sampling). This allows sampling the statistically relevant configurations of the system but the possibility of obtaining time dependent quantities must be abandoned.

As in the bimolecular quenching case, the umbrella sampling yields a $P(r)$ distribution that extends to higher distances for the R6G labelled system than for PEP/MR121. The same type of geometrical interpretation of the data as used for the bimolecular case yields quenching distances which are still physically reasonable and similar to the ones found for the bimolecular case.

The relative configurations of the dye and the TRP obtained from the umbrella sampling runs reflect only three of the more populated configurations found in the bimolecular quenching case. These three geometries are probably the best candidates for future molecular orbital calculations aimed at understanding the details of the quenching process at the level of electronic transitions.

References

Benchimol, S., D. Pim and L. Crawford (1982). "Radioimmunoassay of the cellular protein p53 in mouse and human cell lines" EMBO J **9** (1): 1055-62.

Soussi, T. (2000). "p53 Antibodies in the sera of patients with various types of cancer: a review" Cancer Res **7** (60): 1777-88.

Lane, D. P. (1992). "Cancer. p53, guardian of the genome" Nature **6381** (358): 15-6.

Cho, Y., S. Gorina, P. D. Jeffrey and N. P. Pavletich (1994). "Crystal structure of a p53 tumor suppressor-DNA complex: understanding tumorigenic mutations" Science **5170** (265): 346-55.

Nigro, J. M., S. J. Baker, A. C. Preisinger, J. M. Jessup, R. Hostetter, K. Cleary, S. H. Bigner, N. Davidson, S. Baylin, P. Devilee and et al. (1989). "Mutations in the p53 gene occur in diverse human tumour types" Nature **6250** (342): 705-8.

Hollstein, M., D. Sidransky, B. Vogelstein and C. C. Harris (1991). "p53 mutations in human cancers" Science **5015** (253): 49-53.

Soussi, T. (1996). "The humoral response to the tumor-suppressor gene-product p53 in human cancer: implications for diagnosis and therapy" Immunol Today **8** (17): 354-6.

Lubin, R., B. Schlichtholz, J. L. Teillaud, E. Garay, A. Bussel and C. P. Wild (1995a). "p53 antibodies in patients with various types of cancer: assay, identification, and characterization" Clin Cancer Res **12** (1): 1463-9.

Lubin, R., G. Zalcman, L. Bouchet, J. Tredanel, Y. Legros, D. Cazals, A. Hirsch and T. Soussi (1995b). "Serum p53 antibodies as early markers of lung cancer" Nat Med **7** (1): 701-2.

Lubin, R., B. Schlichtholz, D. Bengoufa, G. Zalcman, J. Tredaniel, A. Hirsch, C. C. de Fromentel, C. Preudhomme, P. Fenaux, G. Fournier and et al. (1993). "Analysis of p53

antibodies in patients with various cancers define B-cell epitopes of human p53: distribution on primary structure and exposure on protein surface" Cancer Res **24** (53): 5872-6.

Neuweiler, H., A. Schulz, M. Böhmer, J. Enderlein and M. Sauer (2003). "Measurement of Submicrosecond Intramolecular Contact formation in Peptides at the Single-Molecule Level" J Am Chem Soc **18** (125): 5324-30.

Vaiana, A. C., H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer and J. C. Smith (2003). "Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation" J Am Chem Soc (in press):

Neuweiler, H., A. Schulz, A. C. Vaiana, J. C. Smith, S. Kaul, J. Wolfrum and M. Sauer (2002). "Detection of individual p53-autoantibodies by using quenched peptide-based molecular probes" Angew Chem, Int Ed Engl **24** (41): 4769-73.

Kussie, P. H., S. Gorina, V. Marechal, B. Elenbaas, J. Moreau, A. J. Levine and N. P. Pavletich (1996). "Structure of the MDM2 oncoprotein bound to the p53 tumor suppressor transactivation domain" Science **5289** (274): 948-53.

Clore, G. M., J. Ernst, R. Clubb, J. G. Omichinski, W. M. Kennedy, K. Sakaguchi, E. Appella and A. M. Gronenborn (1995). "Refined solution structure of the oligomerization domain of the tumour suppressor p53" Nat Struct Biol **4** (2): 321-33.

Rustandi, R. R., D. M. Baldisseri and D. J. Weber (2000). "Structure of the negative regulatory domain of p53 bound to S100B(betabeta)" Nat Struct Biol **7** (7): 570-4.

Torrie, G. M. and J. P. Valleau (1977). "Nonphysical Sampling Distributions in Monte-Carlo Free Energy Estimation: Umbrella Sampling" J Comput Phys (23): 187-99.

Kumar, S., D. Bouzida, R. H. Swendsen, P. A. Kollman and J. M. Rosenberg (1992). "The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method" J Comput Chem **8** (13): 1011-21.

Gullingsrud, J. R., R. Braun and K. Schulten (1999). "Reconstructing Potentials of Mean Force through Time Series Analysis of Steered Molecular Dynamics Simulations" J Comput Phys **1** (151): 190-211.

6. Concluding Remarks

The "selective" fluorescence quenching property of TRP in dye-conjugated peptide chains enables monitoring of conformational dynamics and antibody binding events by means of fluorescence spectroscopy. Moreover, most dyes quenched by TRP are also quenched by the nucleic acid base guanosine. TRP residues are almost omnipresent in proteins. In order to fully exploit the possibilities offered by modern labelling techniques along with the remarkable sensitivity achievable in single molecule fluorescence spectroscopy a detailed theoretical understanding of the dye/quencher systems involved is necessary. This understanding cannot be limited to simplified models as is the case for most quantum chemical calculations. A full treatment of environmental effects with a full, explicit representation of system details at the atomic level is necessary.

6.1 Summary of Current Work and Conclusions

The present thesis shows how MD simulations, in combination with ensemble fluorescence experiments, may be efficiently used to obtain predictive insight into dye/quencher interactions, not only for simplified model systems but also in presence of a complex, realistic environment as is the case for biomolecules in solution. In this work the quenching problem has been studied “from scratch” i.e., starting from an *ab initio* parameterisation of the two dyes. The obtained parameters are subsequently used to perform MD simulations on two simplified, but realistic dye quencher systems: MR121/TRP and R6G/TRP. Results of these simulations provide important insight on the bimolecular interactions between dye and quencher. A quantitative interpretation of the quenching mechanism based on it’s dependence on dye/quencher distance and orientation has emerged. These results are then applied to the interpretation of simulation data of an epitope from the tumor suppressor protein p53 which was labelled at the N-terminus first with MR121 and then with R6G. In all phases of the thesis care was taken to confront and/or combine the theoretical results with available experimental data.

6.1.1 Development of the AFMM Method

The reliability of MD simulations strongly depends on the parameters used in the force field. In the course of the thesis work, for the purpose of obtaining force field parameters of the two

dyes, a general method (AFMM) was developed. AFMM allows parameterisation of small to medium sized molecules using results of reliable quantum chemical calculations as reference data. Due to the computationally inexpensive penalty function proposed here, the method is also well suited for automation. The MC refinement algorithm is relatively simple, although it could be, in the future, substituted by more elaborate methods such as a genetic algorithm, as suggested by [Wang and Kollman 2001]. Nonetheless, with the present algorithm it was possible to obtain a frequency matching for R6G and MR121 within the error range found in benchmark studies on test molecules. The refined parameters for R6G reproduce well available experimental crystal structure data that are independent of the data used in the parameterisation procedure.

6.1.2 Bimolecular Quenching of MR121 and R6G by TRP

Simulation of the bimolecular quenching of R6G/TRP and MR121/TRP provides an atomic-detail description of the geometry and kinetics of the quenching interaction of TRP with fluorescent dyes. The results illustrate the validity of the approach used, which combines MD simulation with time-resolved fluorescence experiments. This approach involves use of MD simulations to determine molecular interaction geometries along with thermodynamic properties which can then be matched with results from appropriate fluorescence experiments. The most striking feature emerging from the bimolecular interactions is the existence of a well defined “quenching distance” almost at Van der Waals contact distance. The quenching distance is the same for both dye/quencher systems analysed. Detailed analysis of subpopulations of the PMF landscape reveals few, well defined “states” or geometries corresponding to quenched or fluorescent interactions between the dye and the quencher.

6.1.3 Conformational Dynamics of a Dye Labelled p53 Epitope

The umbrella sampling studies of the p53 epitope presented here show that, using the knowledge acquired from simulation of simpler model systems, it is possible to obtain important information on the conformation dependent quenching of dye-labelled peptides in solution. Intramolecular dye/TRP fluorescence quenching experiments are widely applied to a variety of biomolecular systems of fundamental, medical and technological interest. Many of the interactions present in systems such as those employing peptide conjugates involve a considerably larger number of intramolecular degrees of freedom than in the model systems studied here, thus complicating the obtention of converged interaction probability

distributions from computer simulation. For this reason, even with the aide of the most advanced computational resources, biased sampling methods are required to study dye quencher interactions in biologically relevant systems.

Results from the p53 epitope umbrella sampling in solution show very different behaviour of the two dye labelled peptides, especially at larger values of the ring separation distance i.e., the unquenched (open) configurations. The presence of the peptide reduces the number of possible quenched (closed) conformations accessible to the system in the case of R6G. Nonetheless the experimental quenching efficiency is only slightly lower for the R6G labelled system than for the MR121 system this is due to the compensating effect of a wider probability distribution at contact distance.

6.2 Future Perspectives

One of the most exciting aspects of scientific research work is that, when a project is finally brought to conclusion, for each question addressed in the course of the project many more new questions will have emerged giving rise to a host of "possible" spin-off projects. This project is no exception: a number of promising directions in which future work may evolve from the work presented here are imaginable. Some of these directions are briefly traced in the following paragraphs and where possible, preliminary data illustrating the feasibility of such projects are presented.

6.2.1 AFMM

Although development of the AFMM method was actually a necessary "side effect" of the parameterisation of MR121 and R6G in the course of the thesis, because of its general applicability, its importance for future simulative studies should not be underestimated. This applies not only to fluorescent dyes but also to other small to medium sized molecules interacting with biological systems and for which force field parameters have not yet been developed. This is reflected in the fact that the method has been already applied for obtaining parameters of cholesterol [Cournia, et al. 2003], and it is currently being applied for the parameterisation of small, commonly used antibiotic molecules which bind to RNA (work in progress).

The AFMM method as it is today, derives parameters using reference data respect to one minimum energy conformation of the molecule to be parameterised. In the case of more

flexible molecules, it is desirable to have parameter sets able to describe the internal motions in more than one minimum. In principle, this can be achieved without major changes to the method by simply performing more quantum chemical and classical energy minimisations and normal mode analysis (one for each minimum) and by extending the penalty function to take into account new sets of eigenvector projections. This method would not be much more expensive than the original one but would still involve using a harmonic approximation of the potential energy function within each minimum. A better approximation of the potential energy surface, including anharmonic terms, could be achieved at a much higher computational cost by optimising parameters to fit the complete Hessian matrix calculated in several points of configurational space. This would involve performing several single point quantum chemical calculations far from the minimum i.e., in various different conformations of the molecule.

Apart from the obvious application of the AFMM method to derive new force field parameters, another not less important one is verification of existing force fields. The AFMM method offers the possibility of confronting different parameter sets and energy functions. It is imaginable to perform a "force field screening" in which each molecule parameterised within the force field is tested using AFMM as was illustrated here for the CHARMM benchmark studies.

6.2.2 Future QM Studies of Quenching

As seen in chapter 4, the quenching of the two dyes by TRP is due to the formation of non-fluorescent complexes in solution. It was also mentioned that the detailed mechanism would most probably resemble the ultra-fast photo-induced electron transfer reaction between the TRP and the dye in the stacked interaction geometries. This has already been observed for the riboflavin/TRP dye/quencher pair [Zhong and Zewail 2001]. To extend understanding of the detailed quenching process beyond the simple 'quenching distance' concept used here will require quantum chemical analysis of these complexes. This may be done by using, for example, the geometries derived from MD simulation as input for appropriate molecular orbital calculations. In fact, the electronic transitions involved in the emission of a fluorescence photon by molecules, schematically shown in Fig. 2.3.2, can be described in terms of transitions between an "excited" state which usually corresponds to the lowest unoccupied molecular orbital (LUMO) and the highest occupied molecular orbital (HOMO). The concepts of HOMO and LUMO are very useful in that they can be directly calculated

with quantum chemical methods. Calculation of excited states is not a simple matter, methods based on HF theory usually give results that are at best, only qualitatively correct and computationally very expensive. Other methods based on density functional theory (DFT) are computationally more feasible and achieve a somewhat better precision than their HF equivalents. Fig. 6.2.1 shows some preliminary results of DFT calculations performed separately on TRP and R6G. These already seem to qualitatively confirm the hypothesis of a quenching mechanism involving a photo-induced electron transfer reaction.

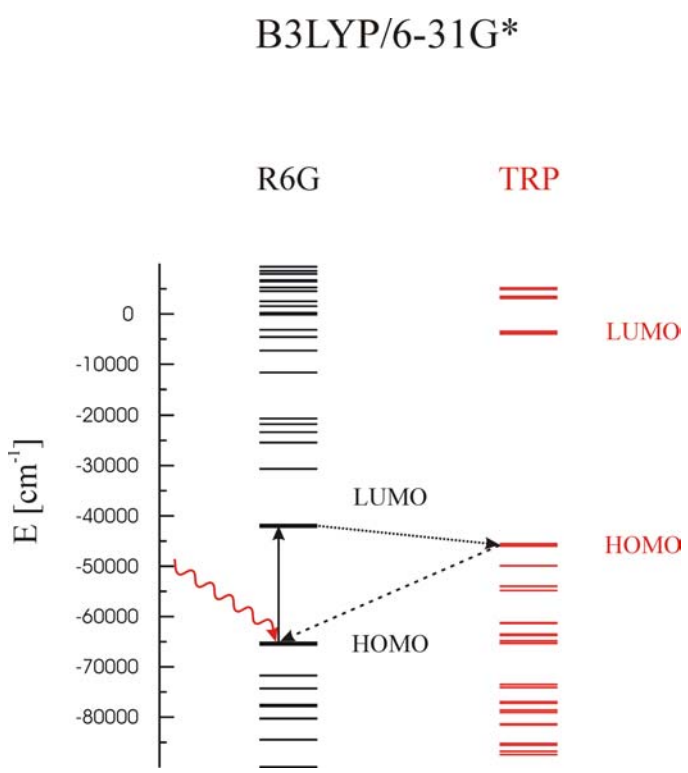

Figure 6.2.1: HOMO, LUMO and neighbouring *energy levels for isolated R6G and TRP molecules resulting from a B3LYP/6-31G* DFT calculation. A schematic representation of a possible photo-induced electron transfer reaction is displayed.*

Among DFT based methods, a particularly good method for calculating excited state transitions in molecular complexes similar to those described in this thesis, is the method of time dependent density functional theory (TDDFT). TDDFT has been successfully used by [Jean and Hall 2001] to give a rigorous electronic description of the context dependent quenching and lifetimes of 2-Aminopurine (a fluorescent analog of guanosine) in stacked interactions with nucleotide bases. Starting geometries for these calculations were easily obtainable from canonical x-ray structures of nucleic acids. This same method can now be

applied to the dye/quencher systems described in this work. The quenched and unquenched interaction geometries resulting from this thesis can serve as starting points for TDDFT calculations. The main difference being that while the interaction geometries in the case of the 2-Aminopurine study were well characterized experimentally, in the case of the MR121/TRP and R6G/TRP complexes the interaction geometries to be studied using TDDFT have been characterized using reliable MD simulations and umbrella sampling. These calculations would represent the "other half" of this present work, closing the circle and returning to a molecular orbital characterisation of the fluorescence quenching properties of the two dyes.

6.2.3 Studies of the Peptide Binding Mechanism

A minimal interpretation of the binding mechanism of the peptide to its antibody partner is proposed in Chapter 9. This interpretation is at the moment only speculative. An important possible direction of future work would be to clarify the issue of how the peptide interacts with its binding partners. This could be done by simulating the labelled peptide in the presence of possible binding partners (a p53 auto-antibody in this case). Unfortunately, there are no x-ray or NMR studies available today which could provide structural information on specific antibodies directed against p53. A homology based modelling of the antibody would be necessary before engaging in any MD study. The matter is further complicated by the fact that any simulation of the binding complex would have to fully incorporate environmental effects such as solvent interactions. Thus the system to be simulated becomes quite large and the simulations would require large computational resources. However it is likely that with, for example, the use of coarse-grained cluster computing, useful information could be obtained using the present approach for these systems. One could envisage, given a reliable dynamic model of the labelled-peptide/antibody complex, to perform *"in silico"* mutational studies aimed at making the assay more sensitive and more specific. This would involve finding an optimal compromise between strong binding, efficient quenching of the fluorescence of the isolated labelled peptide in solution, and high fluorescence intensity of the bound complex.

References

Wang, J. and P. A. Kollman (2001). "Automatic parameterization of force field by systematic search and genetic algorithms" J Comput Chem **12** (22): 1219-28.

Cournia, Z., A. C. Vaiana, J. C. Smith and M. Ullmann (2003). "Derivation of a Molecular Mechanics Force Field for Cholesterol." Pure Appl Chem (in press):

Zhong, D. and A. H. Zewail (2001). "Femtosecond dynamics of flavoproteins: Charge separation and recombination in riboflavin (vitamin B2)-binding protein and in glucose oxidase enzyme" Proc Natl Acad Sci U S A **21** (98): 11867-72.

Jean, J. M. and K. B. Hall (2001). "2-Aminopurine fluorescence quenching and lifetimes: role of base stacking" Proc Natl Acad Sci U S A **1** (98): 37-41.

Appendix A1

A1.1 Molecular Mechanics Force Field Parameters for the Oxazine Dye MR121

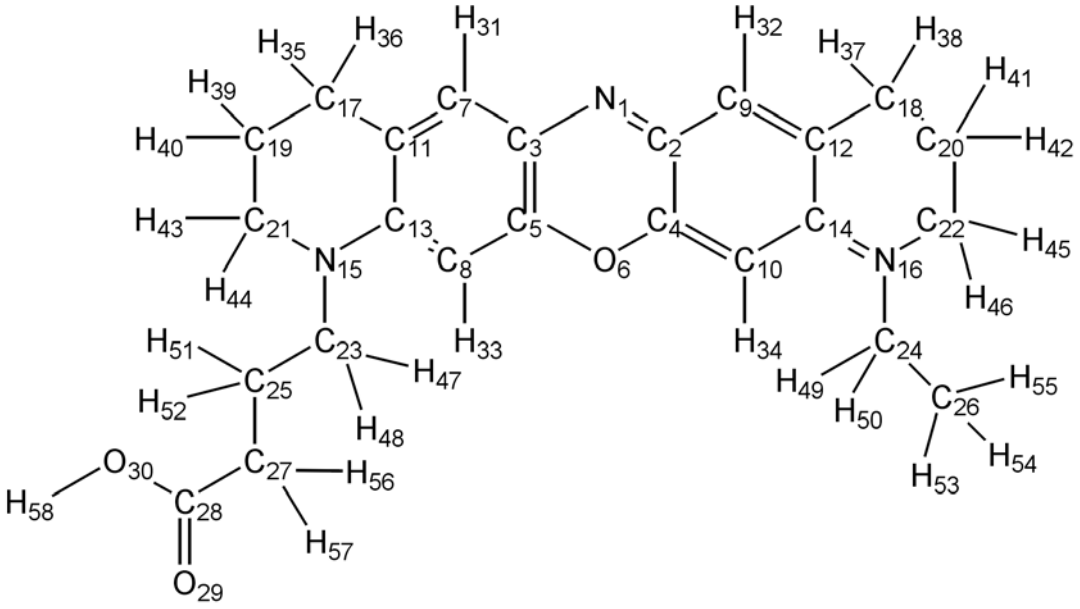

Figure A1.1.1: *Atom numbering scheme for MR121.*

Number	Atom type	Charge	Number	Atom type	Charge
N1	NP2	-0.74	C27	CT2	-0.21
C2, C3	CA	0.37	C28	CD	0.75
C4, C5	CA	0.30	O29	OB	-0.55
O6	OP1	-0.60	O30	OH1	-0.61
C7-C10	CA	-0.115	H31-H34	HP	0.115
C11, C12	CA	0.00	H35-H57	HA	0.09
C13, C14	CA	0.38	H58	H	0.44
N15, N16	NP1	-0.80			
C17-C20	CT2	-0.18			
C21-C24	CT2	0.28			
C25	CT2	-0.18			
C26	CT3	-0.27			

Table A1.1.1: *Atom type assignments and partial atomic charges for MR121.*

Atom types	$K_b \left[\frac{kcal}{mol Å^2}\right]$	$b_0 [Å]$	Atom Types	$K_b \left[\frac{kcal}{mol Å^2}\right]$	$b_0 [Å]$
NP2-CA	320.0	1.32	CA-OP1	300.0	1.335
CT2-NP1	261.0	1.49	NP1-CA	463.0	1.365

Table A1.1.2: *Bond parameters for MR121 (only parameters not already published in Refs. [MacKerell 1995, MacKerell 1998] are given).*

Atom types	$\mathbf{K_\theta}\left[\frac{kcal}{mol\ \mathrm{rad}^2}\right]$	$\boldsymbol{\theta_0}\,[\mathrm{deg}]$	**Atom types**	$\mathbf{K_\theta}\left[\frac{kcal}{mol\ \mathrm{rad}^2}\right]$	$\boldsymbol{\theta_0}\,[\mathrm{deg}]$
CA-NP2-CA	90.0	109.4	NP2-CA-CA	70.0	129.9
CA1-CA1-OP1	45.0	120.0	CA-OP1-CA	45.0	120.0
CA-CT2-CT2	70.0	113.0	CA-CA-NP1*	40.0	120.0
CA-NP1-CT2	45.8	122.3	CT2-NP1-CT2	40.4	115.0
CT2-CT2-NP1	67.7	107.5	NP1-CT2-CT3	67.7	107.5
NP1-CT2-HA	51.5	107.5			

* For this angle parameter also a Urey-Bradley term is used (K_{ub} = 40.0 *kcal mol*$^{-1}$ *Å*$^{-2}$, S_0 = 2.4162 *Å*)

Table A1.1.3: *Angle parameters for MR121 (only parameters not already published in Refs. [MacKerell 1995, MacKerell 1998] are given).*

Atom types	$\mathbf{K_\gamma}\left[\frac{kcal}{mol}\right]$	**n**	$\boldsymbol{\chi_0}\,[\mathrm{deg}]$	**Atom types**	$\mathbf{K_\gamma}\left[\frac{kcal}{mol}\right]$	**n**	$\boldsymbol{\chi_0}\,[\mathrm{deg}]$
NP2-CA-CA-OP1	3.1	2	180.0	NP2-CA-CA-CA	3.1	2	180.0
NP2-CA-CA-HP	4.2	2	180.0	NP2-CA-CA-OP1	3.1	2	180.0
CA-NP2-CA-CA	3.1	2	180.0	CA-CA-OP1-CA	3.1	2	180.0
OP1-CA-CA-CA	3.1	2	180.0	OP1-CA-CA-HP	4.2	2	180.0
CA-CA-CA-NP1	1.55	2	180.0	CA-CA-NP1-CT2	1.55	2	170.0
HP-CA-CA-NP1	4.2	2	180.0	CT2-CA-CA-NP1	1.55	2	180.0
CT2-CT2-NP1-CA	0.23	2	180.0	CT2-CT2-NP1-CT2	0.15	1	0.0
CA-NP1-CT2-HA	0.0	3	0.0	CA-NP1-CT2-CT3	0.23	2	180.0
CT2-NP1-CT2-HA	0.0	3	0.0	CT2-NP1-CT2-CT3	0.15	1	0.0
NP1-X-X-CT2*	40.0	0	0.0				

*This is an improper dihedral term which is used to achieve planarity of the groups containing the sp^2 nitrogens N_{15} and N_{16}.

Table A1.1.4: *Dihedral parameters for MR121 (only parameters not already published in Refs. [MacKerell 1995, MacKerell 1998] are given).*

A1.2 Molecular Mechanics Force Field Parameters for R6G

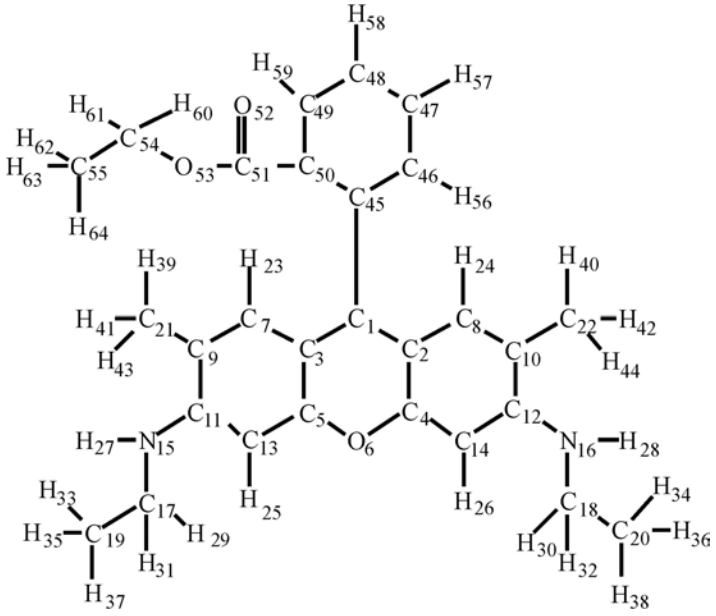

Figure A1.1.1: *Atom numbering scheme for R6G.*

Number	Atom type	Charge	Number	Atom type	Charge
C1	CA1	0.00	H27, H28	H	0.46
C2, C3	CA	0.00	H29-H44	HA	0.09
C4, C5	CA	0.17	C45	CA1	0.00
O6	OS	-0.34	C46-C49	CA	-0.115
C7, C8	CA	-0.115	C50	CA	0.00
C9, C10	CA	0.00	C51	C	0.73
C11, C12	CA	0.64	O52	OB	-0.52
C13, C14	CA	-0.115	O53	OS	-0.34
N15, N16	NC2	-0.60	C54	CT2	-0.05
C17, C18	CT2	-0.18	C55	CT3	-0.27
C19-C22	CT3	-0.27	H56-H59	HP	0.115
H23-H26	HP	0.115	H60-H64	HA	0.09

Table A1.2.1: *Atom type assignments and partial atomic charges for R6G.*

Atom types	$K_b \left[\frac{kcal}{mol Å^2}\right]$	b_0 [Å]	Atom Types	$K_b \left[\frac{kcal}{mol Å^2}\right]$	b_0 [Å]
CA – OS	300.0	1.335	CD – CA	302.0	1.480
CA – NC2	463.0	1.365	CT2 – OS	340.0	1.430
CA1- CA1	221.0	1.490	CA1 - CA	305.0	1.3750

Table A1.2.2: *Bond parameters for R6G (only parameters not already published in Refs. [MacKerell 1995, MacKerell 1998] are given).*

Atom types	$K_\theta \left[\frac{kcal}{mol\ rad^2}\right]$	θ_0 [deg]	Atom types	$K_\theta \left[\frac{kcal}{mol\ rad^2}\right]$	θ_0 [deg]
CA1 - CA1 - CA	48.9	122.3	CA - NC2 - CT2	53.0	129.9
CA – OS - CA	52.5	121.6	NC2 - CT2 - CT3	69.1	107.5
CA – CA - NC2	39.4	120.0	CA - CD – OB	37.1	123.0
OS – CA – CA	76.3	119.0	CA - CD - OS	38.5	113.0
CA - NC2 – H	39.0	113.4	OS – CT2 – CT3	30.9	107.5

Table A1.2.3a: *Angle parameters for R6G (only parameters not already published in Refs. [MacKerell 1995, MacKerell 1998] are given).*

Atom types	$K_\theta \left[\frac{kcal}{mol\ \mathrm{rad}^2}\right]$	θ_0 [deg]	$K_{ub} \left[\frac{kcal}{mol\,Å^2}\right]$	S_0 [Å]
CA – CA - CA1	40.0	120.0	35.0	2.4162
CA - CA1 - CA	40.0	120.0	35.0	2.4162
HP - CA - CA1	30.0	120.0	22.0	2.1525
CA – CA - CD	89.6	119.5	33.5	2.4162
CA1 – CA - CD	89.6	119.5	33.5	2.4162

Table A1.1.3b: *Angle parameters with Urey-Bradley terms forR6G (only parameters not already published in Refs. [MacKerell 1995, MacKerell 1998] are given).*

Atom types	$K_\chi \left[\frac{kcal}{mol}\right]$	n	χ_0 [deg]	Atom types	$K_\chi \left[\frac{kcal}{mol}\right]$	n	χ_0 [deg]
CA – CA - CA - CA1	3.1	2	180.0	CA - CA - CD - OB	1.0	2	180.0
CA – CA - CA1 – CA	3.1	2	180.0	CA1 - CA - CD – OB	1.0	2	180.0
CA1 – CA – CA - HP	4.2	2	180.0	CA - C - OS - CT2	2.5	1	0.0
CA – CA1 – CA - HP	4.2	2	180.0	CA - CA – CD – OS	1.0	2	180.0
OS –CA –CA - CA	1.7	2	180.0	CA1 - CA - CD – OS	1.0	2	180.0
OS – CA - CA - CA1	3.1	2	180.0	NC2 - CA - CA – HP	4.2	2	180.0
CA - CA - CA - CD	4.2	2	180.0	CA - CA1 - CA1 - CA	-0.6	2	0.0
CA – CA1 - CA - CD	4.2	2	180.0	CA - CA1 - CA1 - CA	0.1	4	0.0
CA1 – CA1 - CA - CD	4.2	2	180.0	CT2 - NC2 - CA – CA	0.0	2	140.0
CA – CA - CA1 - CA1	2.1	2	180.0	CA - CA - NC2 - HC	2.5	2	180.0
OSR - CA - CA - HP	1.3	2	180.0	CD – OS – CT2 – CT3	0.0	3	180.0
CA - OS - CA - CA	4.0	2	180.0	CD - OS – CT2- HA	0.0	3	180.0
CA - CA - CA - NC2	3.5	2	180.0	CA – NC2 – CT2 - CT3	1.0	2	-160.0
CA1 - CA1 - CA - HP	3.2	2	180.0	CA – NC2 – CT2 – HA	0.0	6	180.0
HP - CA - CA - CD	3.2	2	180.0	CT3 – CT2 – NC2- HC	0.0	3	180.0
NC2 -CA -CA -CT2	1.1	2	180.0	HC – NC2 – CT2 - HA	0.0	6	180.0
NC2 - CA - CA - CT3	1.1	2	180.0				

Table A1.2.4: *Dihedral parameters for R6G (only parameters not already published in Refs. [MacKerell 1995, MacKerell 1998] are given).*

Appendix A2 - Publications Arising from this Thesis

1) Neuweiler, H., A. Schulz, A. C. Vaiana, J. C. Smith, S. Kaul, J. Wolfrum and M. Sauer (2002). "Detection of individual p53-autoantibodies by using quenched peptide-based molecular probes" Angew Chem, Int Ed Engl **24** (41): 4769-73.

2) Vaiana, A. C., A. Schulz, J. Wolfrum, M. Sauer and J. C. Smith (2003). "Molecular Mechanics Force Field Parameterization of the Fluorescent Probe Rhodamine 6G using Automated Frequency Matching" J Comput Chem **5** (24): 632-9.

3) Cournia, Z., A. C. Vaiana, J. C. Smith and M. Ullmann (2003). "Derivation of a Molecular Mechanics Force Field for Cholesterol." Pure Appl Chem (in press).

4) Vaiana, A. C., H. Neuweiler, A. Schulz, J. Wolfrum, M. Sauer and J. C. Smith (2003). "Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation" J Am Chem Soc (in press).

Molecular Mechanics Force Field Parameterization of the Fluorescent Probe Rhodamine 6G Using Automated Frequency Matching

ANDREA C. VAIANA,[1] ANDREAS SCHULZ,[2] JÜRGEN WOLFRUM,[2] MARKUS SAUER,[2] JEREMY C. SMITH[1]

[1]*IWR-Biocomputing, Universität Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany*

[2]*Physikalisch-Chemisches Institut, Universität Heidelberg, Im Neuenheimer Feld 253, D-69120 Heidelberg, Germany*

Received 27 May 2002; Accepted 14 August 2002

Abstract: Novel single-molecule fluorescence experimental techniques have prompted a growing need to develop refined computational models of dye-tagged biomolecules. As a necessary first step towards useful molecular simulations of fluorescence-labeled biomolecules, we have derived a force field for the commonly used dye, rhodamine 6G (R6G). A novel automated method is used that includes fitting the molecular mechanics potential to both vibrational frequencies and eigenvector projections derived from quantum chemical calculations. The method is benchmarked on a series of aromatic molecules then applied to derive new parameters for R6G. The force field derived reproduces well the crystal structure of R6G.

 J Comput Chem 24: 632–639, 2003

Key words: computer simulation; molecular modeling; force field parameters; fluorescent dyes; normal modes

Introduction

Recent advances in single-molecule detection techniques have stimulated interest in the application of computer simulation, in particular molecular dynamics (MD), as a means of interpreting experimental results in terms of the structure and dynamics of the molecules investigated.[1–4] In single-molecule fluorescence experiments biomolecules are tagged with specifically tailored fluorescent dyes.[5] There is therefore a growing need to develop refined computational models of dye-tagged biomolecules to fully exploit the possibilities offered by the new experimental methods.

Fluorescence properties such as quenching are a result of complex interactions between chromophores and the surrounding environment. Although quantum mechanical treatment of fluorescent systems is required to model the electronic transitions involved in excited-state kinetics,[6] the associated task of exploring the configurational space of dye/environment systems is computationally too demanding to be treated with quantum methods. However, this configurational space can be extensively explored using MD methods, leading to detailed information on interaction geometries, free energies, configurational energy landscapes, and effects of solvent on single dye/biomolecule systems.

Most "all-atom" empirical force fields used in common MD packages (such as CHARMM[7]) are equipped with parameter sets for modeling biomolecules but, to our knowledge, none have been developed for commonly used fluorescent labels. Moreover, for many dyes, suitable experimental data against which new parameters can be derived are scarce. Consequently, the present article is in two parts. First, an automated refinement method is presented for deriving specifically tailored sets of molecular mechanics parameters for small- to medium-sized molecules. The method involves careful choice of an initial parameter set followed by automated refinement against quantum-chemically derived vibrational eigenvector and eigenvalue sets. The method is tested by comparing results obtained using it with existing CHARMM parameters for propene and the five standard nucleic acid bases: adenine, guanine, cytosine, thymine, and uracil.

In the second part of the article the new method is applied to derive parameters for a common fluorescent dye, rhodamine 6G (R6G), a representative member of a class of widely used xanthene derivative labels. Final testing of the derived R6G parameters

Correspondence to: J.C. Smith; e-mail: biocomputing@iwr.uni-heidelberg.de

Contract/grant sponsor: the Forschungsschwerpunktprogramm of the Land Baden-Württemberg

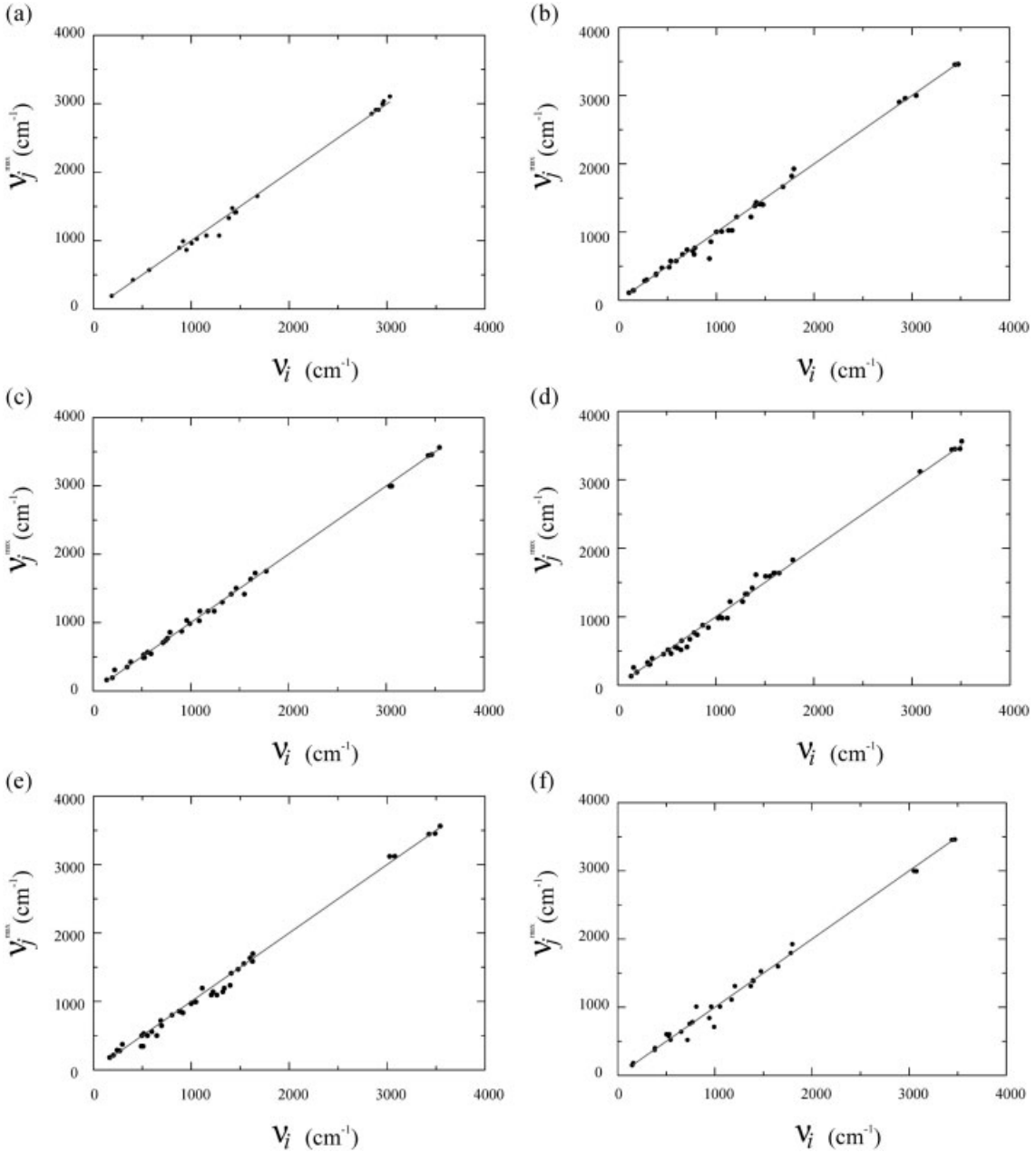

Figure 1. ν_j^{max} vs. ν_i plots for (a) propene, (b) thymine, (c) cytosine, (d) guanine, (e) adenine, and (f) uracil. Lines indicate the ideal case of perfectly matched frequencies and eigenvector projections; points refer to CHARMM parameters taken from ref. 12. Values of σ are: (a) 68 cm^{-1}, (b) 74 cm^{-1}, (c) 47 cm^{-1}, (d) 67 cm^{-1}, (e) 81 cm^{-1}, (f) 92 cm^{-1}.

involves minimization of the crystal structure. The experimental geometry and unit cell volume are well reproduced.

Methods

Computational Details

All quantum chemical computations were performed with the GAUSSIAN-94 package using the standard 6-31G* basis set.[8] The restricted Hartree–Fock (RHF) level of theory was used for geometry optimizations and normal mode calculations. Frequencies resulting from the quantum calculations were rescaled by a factor 0.8929 to compensate for the neglect of electron correlation at the Hartree–Fock level.[9]

All molecular mechanics computations were performed using CHARMM version 27.[7] No cutoff was applied to the electrostatic interactions. Except for the new parameters derived here, all force field parameters are given in refs. 10–12. All the molecular me-

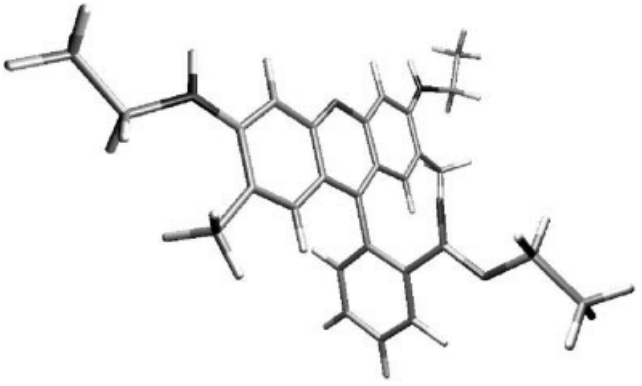

Figure 2. Gas-phase ground-state structure of R6G (RHF 6-31G* optimization).

chanics minimizations were carried out using the Steepest Descent algorithm followed by Newton–Raphson minimization with a convergence criterion for the energy gradient of 10^{-8} kcal/mol/Å.

Parameter Refinement

The refinement method proposed here is, in principle, applicable with any atom-based molecular mechanics program, the only prerequisite being the possibility of calculating normal modes and their relative eigenvectors. We will, however, in the following, confine our attention to the CHARMM force field. In the CHARMM model the potential energy is partitioned as follows:[7]

$$V(R) = \sum_{\text{bonds}} K_b(b - b_0)^2 + \sum_{ub} K_{ub}(s - s_0)^2 + \sum_{\text{angles}} K_\theta(\theta - \theta_0)^2 + \sum_{\text{dihedrals}} K_\chi(1 + \cos(n\chi - \chi_0)) + \sum_{\text{impropers}} K_\varphi(\varphi - \varphi_0)^2 + \sum_{\text{nonbond}} \varepsilon_{ij}\left[\left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{12} + \left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{D r_{ij}} \quad (1)$$

where K_b, K_{ub}, K_θ, K_χ, K_ϕ are, respectively, the bond, Urey–Bradley, angle, dihedral and improper dihedral constants, and b, s, θ, χ, and ϕ represent, respectively, bond length, Urey–Bradley 1–3 distance, bond angle, dihedral angle, and improper torsion angle (a subscript zero is used to represent the corresponding equilibrium value). Nonbonded interactions between pairs of atoms (labeled i and j) at a relative distance r_{ij} consist of Lennard–Jones 6–12 (LJ) terms and Coulomb interactions. $R_{ij}^{\min}$ and ε_{ij} are, respectively, the distance between atoms i and j at which the LJ potential is minimum and the depth of the LJ potential well for the same pair of atoms. D is the effective dielectric constant and q_i the partial atomic charge on atom i.

Before any refinement procedure, an initial set of parameters must be determined. The Van der Waals constants ε_{ij} and $R_{ij}^{\min}$ depend mostly on atomic properties and are relatively insensitive to changes in the molecular environment. These were directly transferred from existing CHARMM values and were not modified during refinement. To derive the partial atomic charges, q_i, the

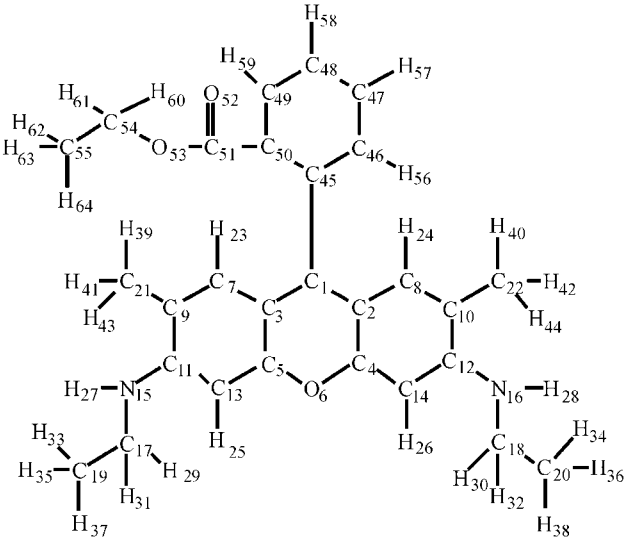

Figure 3. Atom numbering scheme for R6G.

CHELPG method within GAUSSIAN was used to fit the molecular mechanics-derived electrostatic potential to that obtained quantum mechanically.[13] Equilibrium values for bonds b_0, angles θ_0, and dihedrals χ_0 were derived from the RHF 6-31G* ground-state structure. An initial guess, based on analogy to other existing CHARMM parameters and on chemical intuition (taking equilibrium values and hybridization into account), was made for all other missing parameters. The second term in eq. (1) (the so-called Urey–Bradley term[10–12]) is not present in most other force fields, and within the CHARMM model its use is limited to a few special cases. Here K_{ub} was set to zero wherever possible.

The initial parameter set was used for energy minimization and calculation of normal modes (eigenvalues and eigenvectors) with

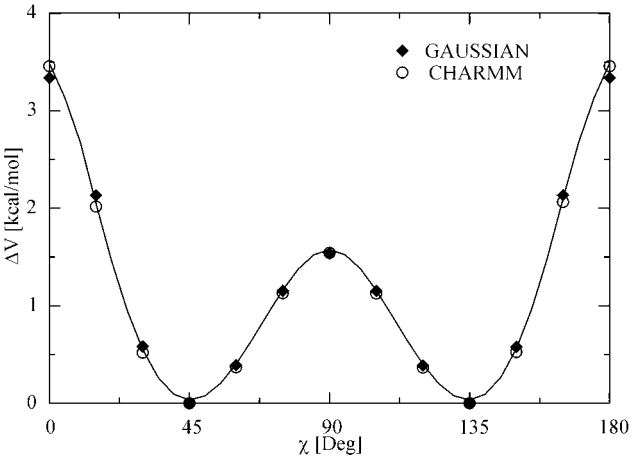

Figure 4. Potential energy profile for rotation of the two rings of biphenyl around the central bond. The filled diamonds are the GAUSSIAN relaxed potential energy scan. Circles and the line are the fitted CHARMM potential.

Table 1. Atom Type Assignments and Partial Charges for R6G.

Number	Atom type	Charge	Number	Atom type	Charge
C1	CA1	0.00	H27, H28	H	0.46
C2, C3	CA	0.00	H29–H44	HA	0.09
C4, C5	CA	0.17	C45	CA1	0.00
O6	OS	−0.34	C46–C49	CA	−0.115
C7, C8	CA	−0.115	C50	CA	0.00
C9, C10	CA	0.00	C51	C	0.73
C11, C12	CA	0.64	O52	OB	−0.52
C13, C14	CA	−0.115	O53	OS	−0.34
N15, N16	NC2	−0.60	C54	CT2	−0.05
C17, C18	CT2	−0.18	C55	CT3	−0.27
C19–C22	CT3	−0.27	H56–H59	HP	0.115
H23–H26	HP	0.115	H60–H64	HA	0.09

CHARMM. These were then directly compared with the results of the "reference" normal modes calculated with GAUSSIAN. Parameters were thus refined iteratively to fit the results of the quantum chemical normal mode calculations.

The practice of parameterizing against vibrational frequencies is quite old.[14–16] One of the major problems of parameterization methods that fit to vibrational frequencies is identifying a calculated mode with the corresponding reference mode.[17] It is possible to obtain good matching of the frequencies for a molecule while exchanging the corresponding eigenvectors. The resulting model may well then reproduce the vibrational frequencies (and the energy) of the reference molecule but may not reproduce the distribution of energy among the intermolecular modes, and thus the dynamical properties of the molecule. It is therefore important to use a penalty function that takes into account both the frequencies and all the corresponding eigenvectors to avoid this problem. Methods based on fitting the whole Cartesian Hessian matrix[18,19] would, in principle, solve this problem. A numerical implementation with the inclusion of off-diagonal terms in the fit did not significantly improve the results.[20] Analytical implementations have been described.[21] An excellent review on force field parameterization can be found in ref. 22.

The fitting method proposed here minimizes the above frequency exchange effect. In the "ideal" case of a perfect molecular mechanics model, the vibrational properties of the molecule, as calculated by molecular mechanics, should perfectly match those resulting from the quantum *ab initio* calculation. For this to occur not only must the frequencies coincide but also the two sets of eigenvectors (resulting from the two different calculations) should be orthonormal to each other.

An efficient way to check simultaneously for both orthonormality and frequency matching is to project each of the CHARMM eigenvectors $\overline{\chi_i^C}$ (where the subscript i indicates the normal mode number and the superscript C indicates that the modes are calculated with CHARMM) onto the reference set of eigenvectors $\{\overline{\chi^G}\}$ (the superscript G indicates that these modes are calculated with GAUSSIAN), to find the frequency $\nu_i^{\max}$ corresponding to the highest projection (j:$\overline{\chi_i^C} \cdot \overline{\chi_j^G} = \max$) and to plot this frequency against the corresponding reference frequency, ν_i. In the ideal case mentioned above this plot would give a one-to-one relationship: $\nu_i = \nu_i^{\max}$ and $\overline{\chi_i^C} \cdot \overline{\chi_j^G} = \delta_{ij}$. Points that deviate from the ideal plot may indicate exchanged or mismatched frequencies.

Automated refinement methods are mostly based on minimizing a penalty function, usually a weighed sum of square deviations from a set of reference values.[20] The optimization method proposed here is based on minimization of the weighted sum of squares, Y^2 of the deviations from the ideal situation, as follows:

$$Y^2 = \sum_{3N-6} \omega_i^2(\nu_i - \nu_j^{\max})^2$$

$$\omega_i = \frac{1}{\max_j(\overline{\chi_i^C} \cdot \overline{\chi_j^G})} \tag{2}$$

where N is the number of atoms in the molecule and there are $3N - 6$ independent vibrational frequencies. The weights ω_i are chosen to be the inverse of the highest eigenvector projection. This has the effect of biasing the penalty function, even in the case of a good frequency assignment, such that minimization of Y^2 leads to an improved eigenvector projection distribution.

Refinement of parameter sets involves exploring a high-dimensional space in search of an optimal set. Consequently, as for any multidimensional search method, in parameter optimization there is always a risk of the search becoming trapped in a high local minimum. To reduce this risk it is necessary to generate a physically reasonable set of initial parameters.[17] To ensure this, the present method involves checking initial parameter sets by visual

Table 2. Bond Parameters for R6G (Only Parameters Not Already Published in Refs. 10–12 Are Shown).

Atom types	$K_b\left[\frac{\text{kcal}}{\text{mol Å}^2}\right]$	b_0[Å]	Atom types	$K_b\left[\frac{\text{kcal}}{\text{mol Å}^2}\right]$	b_0[Å]
CA—OS	300.0	1.335	CD—CA	302.0	1.480
CA—NC2	463.0	1.365	CT2—OS	340.0	1.430
CA1—CA1	221.0	1.490	CA1—CA	305.0	1.3750

Table 3. Angle Parameters for R6G (Only Parameters Not Already Published in Refs. 10–12 Are Shown).

Atom types	$K_\theta\left[\frac{\text{kcal}}{\text{mol rad}^2}\right]$	θ_0[deg]	Atom types	$K_\theta\left[\frac{\text{kcal}}{\text{mol rad}^2}\right]$	θ_0[deg]
CA1—CA1—CA	48.9	122.3	CA—NC2—CT2	53.0	129.9
CA—OS—CA	52.5	121.6	NC2—CT2—CT3	69.1	107.5
CA—CA—NC2	39.4	120.0	CA—CD—OB	37.1	123.0
OS—CA—CA	76.3	119.0	CA—CD—OS	38.5	113.0
CA—NC2—H	39.0	113.4	OS—CT2—CT3	30.9	107.5

inspection of the motions involved in exchanged eigenvector modes (using the Molden program, for example[23]) and manually adjusting the parameters concerned. This approach is particularly useful for critical torsion parameters. In some cases (of which one is described in detail in the next section for the case of R6G), it is necessary to derive initial parameters from rotational potential energy profiles (single point calculations in GAUSSIAN), before achieving a good optimization.

For the automatic optimization of the force constants a standard Monte Carlo (MC) scheme[24] was used to minimize Y.[2] Optimizations were performed simultaneously on bond, angle, and torsion constants. At each step i, all parameters were iteratively varied in the MC algorithm with a uniform distribution within a fixed range, Y_i^2 was evaluated, and, if $Y_i^2 < Y_{i-1}$,[2] the new parameter set was used in the next step, $i+1$.

When comparing results for different molecules, normalization of Y^2 can be rather tedious due to the different weights ω_i. For comparison purposes, then, after minimization of Y^2 the root-mean-square deviation, σ from the reference case was calculated:

$$\sigma = \sqrt{\frac{\sum_{3N-6} (\nu_i - \nu_j^{\max})^2}{3N-7}} \quad (3)$$

The robustness of the MC refinement method with respect to initial parameter choice was tested in calculations using the normal modes and eigenvectors of propene as calculated by CHARMM as reference data. The MC algorithm was used to derive bond force constants, K_b in eq. (1), starting from 50 sets of randomly generated initial values distributed uniformly between 200 and 600 kcal/mol/Å^2 (all other parameters were kept fixed at their normal CHARMM values). In the 10 lowest-σ parameter sets found the refinement algorithm was able to find the original CHARMM parameters for propene to within a maximum relative error of less than 3%.

A related method for comparing modes of different molecules has been proposed in ref. 25.

Results

Test Calculations

Test calculations were performed on molecules for which CHARMM parameters have already been derived by fitting to *ab initio* quantum chemical data. The molecules chosen were propene and the five standard nucleic acid bases. These molecules exhibit chemical similarity to fluorescent dyes. Normal modes calculated with CHARMM were compared with those calculated with GAUSSIAN, using the comparison method described in methods.

Frequency matching plots ($\nu_j^{\max}$ vs. ν_i) are shown for the six test molecules in Figure 1. The plots in Figure 1 confirm that the CHARMM and *ab initio* modes agree overall. Better matching of frequencies <500 cm^{-1} compared to higher frequencies reflect the fact that parameterization in ref. 12 was focused on the low frequency modes. Values of σ are 74 cm^{-1} for T, 47 cm^{-1} for C, 67 cm^{-1} for G, 81 cm^{-1} for A and 92 cm^{-1} for U. The values of σ reflect, on average, a better agreement between CHARMM and the *ab initio* data for the pyrimidines than for the purines, again as reported in ref. 12.

Parameterization of R6G

The RHF 6-31G* ground state structure for R6G is shown in Figure 2, and the atom numbering scheme is shown in Figure 3. In CHARMM, atom type CA describes sp^2 carbons of aromatic rings; to ensure portability this atom type could also be used for the case of the aromatic carbons in R6G. Unfortunately, the use of CA-type carbons for the connection of two aromatic systems in a biphenyl-type configuration results in a structure that is highly biased toward a planar arrangement of the two rings. It was thus necessary to invent a new atom type (CA1) for the connection between the phenyl ring and the xanthene moiety in R6G (C1 and C45 in Fig. 3). The only changes to CA1 with respect to the standard CA type are the bond and torsion parameters [K_b, b_0, K_χ, χ_0 and n in eq. (1)]. Initial values for these parameters were derived using biphenyl as a model system: equilibrium values [b_0 and χ_0 in eq. (1)] were derived from the quantum-chemically optimized structure of

Table 4. Angle Parameters with Urey–Bradley Term for R6G (Only Parameters Not Already Published in Refs. 10–12 Are Shown).

Atom types	$K_\theta\left[\frac{\text{kcal}}{\text{mol rad}^2}\right]$	θ_0[deg]	$K_{ub}\left[\frac{\text{kcal}}{\text{mol Å}^2}\right]$	S_0[Å]
CA—CA—CA1	40.0	120.0	35.0	2.4162
CA—CA1—CA	40.0	120.0	35.0	2.4162
HP—CA—CA1	30.0	120.0	22.0	2.1525
CA—CA—CD	89.6	119.5	33.5	2.4162
CA1—CA—CD	89.6	119.5	33.5	2.4162

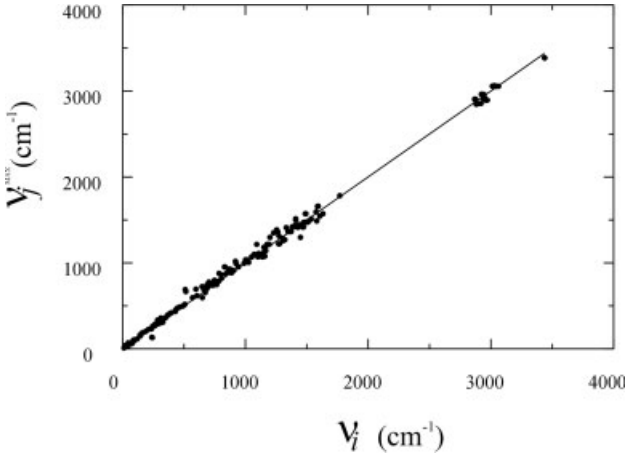

Figure 5. ν_i^{max} vs. ν_i plot for R6G. The line is the ideal case of perfectly matched frequencies and eigenvector projections; points refer to the present optimized parameters. σ = 51.1 cm^{-1}.

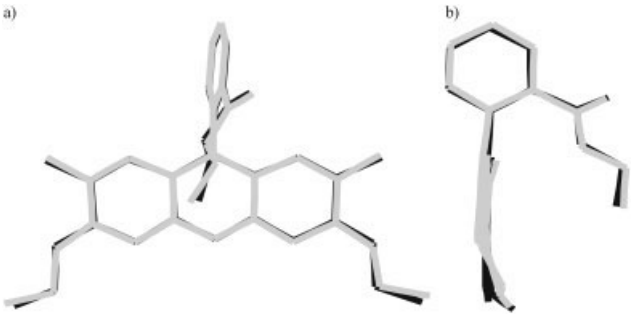

Figure 6. R6G crystal structure, comparison of the simulation result (gray) with the experimental structure by Fun et al. (black): (a) front view, (b) side view. Only heavy atoms are shown. RMSD = 0.083 Å.

biphenyl. The initial value for the bond constant [K_b in eq. (1)] for the CA1–CA1 bond was taken from existing CHARMM parameters. The torsional force constants [K_χ in eq. (1)] were derived from the energy barrier for rotation of biphenyl around the central bond as calculated at RHF/6-31G* level (using single-point calculations to scan the potential energy landscape). Figure 4 shows the results of these single-point calculations performed with GAUSSIAN and with the CHARMM fitted potential. The OS and NC2 types in CHARMM were used to describe the bridging sp^2 oxygen O6 and the nitrogen atoms N15 and N16 in the partially charged amino groups in R6G. It was necessary to introduce some new parameters for these atom types to account for the new bonding partners of these atoms in R6G.

Atom-type assignments and partial atomic charges for R6G are listed in Table 1. Charges were derived using CHELPG at the RHF/6-31G* level. Charges resulting from an MP2/6-31G*//6-31G* calculation did not differ significantly from the RHF/6-31G* charges: the average relative difference being ~1%. All new parameters were refined using the MC algorithm as described in Methods. Final (refined) values of all new parameters for R6G are summarized in Tables 2 to 4. The resulting ν_j^{max} vs. ν_i plot for R6G is shown in Figure 5. The corresponding value of σ (51.1 cm^{-1}) is within the range seen in the benchmark studies (see Fig. 2).

Final testing of a parameter set should be performed against experimental or theoretical data that are as far as possible independent of the data used in the optimization phase. Here, parameters were tested by performing an energy minimization of R6G in crystalline state using the refined parameter set and comparing the results with the experimental structure from ref. 26. The calculation was performed including the full crystal environment using

Table 5. Dihedral Parameters for R6G (Only Parameters Not Already Published in Refs. 10–12 Are Shown).

Atom types	$K_\chi\left[\frac{\text{kcal}}{\text{mol}}\right]$	n	χ_0[deg]	Atom types	$K_\chi\left[\frac{\text{kcal}}{\text{mol}}\right]$	n	χ_0[deg]
CA—CA—CA—CA1	3.1	2	180.0	CA—CA—CD—OB	1.0	2	180.0
CA—CA—CA1—CA	3.1	2	180.0	CA1—CA—CD—OB	1.0	2	180.0
CA1—CA—CA—HP	4.2	2	180.0	CA—C—OS—CT2	2.5	1	0.0
CA—CA1—CA—HP	4.2	2	180.0	CA—CA—CD—OS	1.0	2	180.0
OS—CA—CA—CA	1.7	2	180.0	CA1—CA—CD—OS	1.0	2	180.0
OS—CA—CA—CA1	3.1	2	180.0	NC2—CA—CA—HP	4.2	2	180.0
CA—CA—CA—CD	4.2	2	180.0	CA—CA1—CA1—CA	−0.6	2	0.0
CA—CA1—CA—CD	4.2	2	180.0	CA—CA1—CA1—CA	0.1	4	0.0
CA1—CA1—CA—CD	4.2	2	180.0	CT2—NC2—CA—CA	0.0	2	140.0
CA—CA—CA1—CA1	2.1	2	180.0	CA—CA—NC2—HC	2.5	2	180.0
OSR—CA—CA—HP	1.3	2	180.0	CD—OS—CT2—CT3	0.0	3	180.0
CA—OS—CA—CA	4.0	2	180.0	CD—OS—CT2—HA	0.0	3	180.0
CA—CA—CA—NC2	3.5	2	180.0	CA—NC2—CT2—CT3	1.0	2	−160.0
CA1—CA1—CA—HP	3.2	2	180.0	CA—NC2—CT2—HA	0.0	6	180.0
HP—CA—CA—CD	3.2	2	180.0	CT3—CT2—NC2—HC	0.0	3	180.0
NC2—CA—CA—CT2	1.1	2	180.0	HC—NC2—CT2—HA	0.0	6	180.0
NC2—CA—CA—CT3	1.1	2	180.0				

Table 6. Experimental Unit Cell Dimensions as Reported in Ref. 26 Together with Those Resulting from the Crystal Simulation.

	X	Y	Z
Ref.	14.92 Å	15.16 Å	23.61 Å
Simulation	14.94 Å	15.51 Å	23.15 Å
Δ_i	−0.02 Å	−0.35 Å	0.46 Å
Δ_i/i	0.1%	2.3%	1.9%

periodic boundary conditions. The unit cell has *Pbca* symmetry, with one R6G molecule, an iodide counterion (R6G is a mono cation) and one water molecule represented here with the TIP3P potential.[27] The unit cell dimensions were allowed to vary in the energy minimization. The resulting structure is compared with experiment in Figure 6. The RMSD between the calculated and experimental molecular structures (not including hydrogen atoms) is 0.083 Å. As shown in Table 5, the molecular mechanics results also match the experimental cell dimensions well. The total unit cell volume difference between the simulated crystal and the experimental data is 26.7 Å,[3] i.e., 0.5%.

Conclusions

The method proposed here for force field determination is of particular use in deriving parameters for molecules for which appropriate experimental data are scarce, as is the case for most novel fluorescent dyes. The method uses quantum mechanical calculations to provide reference data. In contrast to other parameterization procedures, the present method uses not only the vibrational frequencies provided by quantum calculations but also information on the forms of the internal motions of the molecule, i.e., the eigenvectors. This is done in a computationally cost-efficient manner. The method involves correction of errors resulting from modes with exchanged frequencies, thus easing the task of producing a first reasonable guess of parameters prior to automated refinement.

Methods based on fitting the whole Hessian matrix (or equivalently, the whole set of eigenvectors) would yield a slightly more accurate parameterization than the present method. However, the present method is well adapted for the accuracy required for MD simulation of fluorescent systems and has the considerable practical advantage of simple geometrical interpretation of mismatched modes in terms of motions along the normal coordinates. This allows for manual fine-tuning of the parameters. There is a risk of underdetermination of the data set due to the use of only $3N - 6$ terms in the penalty function as opposed to $0.5((3N)^2+3N)$ if the whole Hessian were to be used. However, this risk is low when extending an existing parameter set by transferring existing parameters to a new molecule for which only a few new parameters vary significantly during the refinement procedure.

Due to the computationally inexpensive penalty function proposed here, the method is also well suited for automation. The MC refinement algorithm is relatively simple, although it could be, in the future, substituted by more elaborate methods such as a genetic algorithm, such as that suggested in ref. 17. Nonetheless, with the present algorithm it was possible to obtain a frequency matching for R6G within the error range found in benchmark studies on test molecules. The refined parameters for R6G reproduce well available experimental crystal structure data that are independent of the data used in the parameterization procedure.

Deriving force field parameters for dyes is a first essential step towards accurate MD simulation of dye-tagged biomolecule interactions. Use of MD simulation, in combination with quantum chemical calculations of excited state properties, can be expected to provide a detailed picture of fluorescent properties of dyes and dye–biomolecule–solvent interactions. The parameters derived here for R6G are currently being used in MD simulations aimed at gaining insight into the quenching of R6G fluorescence in a number of dye–biomolecule systems (work in progress).

Acknowledgments

The authors thank G. M. Ullmann, D. Mihailescu, and S. Fischer for fruitful discussions.

References

1. Grubmüller, H.; Heymann, B.; Tavan, P. Science 1996, 271, 997.
2. Israelewitz, B.; Gao, M.; Schulten, K. Curr Opin Struct Biol 2001, 11, 224.
3. Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. Science 1997, 275, 1295.
4. Heymann, B.; Grubmüller, H. Phys Rev Lett 2000, 84, 6126.
5. Sauer, M.; Angerer, B.; Ankenbauer, W.; Foldes–Papp, Z.; Gobel, F.; Han, K. T.; Rigler, R.; Schulz, A.; Wolfrum, J.; Zander, C. J Biotechnol 2001, 86, 181.
6. Frank, I.; Hutter, J.; Marx, D.; Parrinello, M. J Chem Phys 1998, 108, 4060.
7. Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. J Comput Chem 1983, 4, 187.
8. Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head–Gordon, M.; Gonzalez, C.; Pople, J. A. Gaussian 94, Revision D.4; Gaussian, Inc.: Pittsburgh, PA, 1995.
9. Foresman, J. B.; Frisch, Æ. Exploring Chemistry with Electronic Structure Methods; Gaussian, Inc.: Pittsburgh, PA, 1993; 2nd ed.
10. MacKerell, A. D., Jr.; Banavali, N. J Comput Chem 2000, 21, 105.
11. MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph–McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I., W.E; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz–Kuczera, J.; Yin, D.; Karplus, M. J Phys Chem B 1998, 102, 3586.
12. Foloppe, N.; MacKerell, A. D., Jr. J Comput Chem 2000, 21, 86.
13. Breneman, C. N.; Wiberg, K. B. J Comput Chem 1990, 11, 361.
14. Schachtschneider, J. H.; Snyder, R. G. Spectrochim Acta 1963, 19, 117.

15. Lifson, S.; Warshel, A. J Chem Phys 1969, 49, 5116.
16. Warshel, A.; Lifson, S. J Chem Phys 1970, 53, 582.
17. Wang, J.; Kollman, P. A. J Comput Chem 2001, 22, 1219.
18. Dasgupta, S.; Yamasaki, T.; Goddard, W. A., III. J Chem Phys 1996, 104, 2898.
19. Maple, J. R.; Hwang, M. J.; Stockfisch, T. P.; Dinur, U.; Waldman, M.; Ewig, C. S.; Hagler, A. T. J Comput Chem 1994, 15, 162.
20. Norrby, P.-O.; Liljefors, T. J Comput Chem 1998, 19, 1146.
21. Halgren, T. A. J Comput Chem 1996, 17, 490.
22. Dinur, U.; Hagler, A. T. Rev Comput Chem 1991, 2, 99.
23. Schaftenaar, G.; Noordik, J. H. J Comput Aided Mol Des 2000, 14, 123.
24. Haeffner, F.; Brinck, T.; Haeberlein, M.; Moberg, C. Theochem 1997, 397, 39.
25. Grafton, A. K.; Wheeler, R. A. J Comput Chem 1998, 19, 1663.
26. Fun, H. K.; Chinnakali, K.; Sivakumar, K.; Lu, C. M.; Xiong, R. G.; You, X. Acta Crystallogr 1997, C53, 1619.
27. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. J Chem Phys 1983, 79, 926.

J|A|C|S
ARTICLES
Published on Web 00/00/0000

Fluorescence Quenching of Dyes by Tryptophan: Interactions at Atomic Detail from Combination of Experiment and Computer Simulation

Andrea C. Vaiana,[†] Hannes Neuweiler,[§] Andreas Schulz,[§] Jürgen Wolfrum,[§] Markus Sauer,[§] and Jeremy C. Smith*[,†]

Contribution from the IWR-Computational Molecular Biophysics, Universität Heidelberg, Im Neuenheimer Feld 368, D-69120 Heidelberg, Germany, and Physikalisch-Chemisches Institut, Universität Heidelberg, Im Neuenheimer Feld 253, D-69120 Heidelberg, Germany

Received May 12, 2003; E-mail: biocomputing@iwr.uni-heidelberg.de

Abstract: Fluorescence spectroscopy and molecular dynamics (MD) simulation are combined to characterize the interaction of two organic fluorescent dyes, rhodamine 6G (R6G) and an oxazine derivative (MR121), with the amino acid tryptophan in aqueous solution. Steady-state and time-resolved fluorescence quenching experiments reveal the formation of essentially nonfluorescent ground-state dye/Trp complexes. The MD simulations are used to elucidate the molecular interaction geometries involved. The MD-derived probability distribution of the distance r between the centers of geometry of the dye and quencher ring systems, $P(r)$, extends to higher distances for R6G than for MR121 due to population in the R6G/Trp system of fluorescent interaction geometries between Trp and the phenyl ring and ester group of the dye. The consequence of this is the experimental finding that under the conditions used in the simulations about 25% of the R6G dye is fluorescent in comparison with 10% of the MR121. Combining the above findings allows determination of the "quenching distance", r^*, above which no quenching occurs. r^* is found to be very similar (~5.5 Å) for both dye/Trp systems, corresponding to close to van der Waals contact. Both experimental dynamic Stern−Volmer analysis and the MD trajectories demonstrate that the main determinant of the fluorescence intensity is static quenching. The approach presented is likely to be useful in the structural interpretation of data obtained from fluorescent conjugates commonly used for monitoring the binding and dynamics of biomolecular systems.

1. Introduction

The development of innovative fluorescence-based techniques for probing molecular recognition and conformational dynamics of biopolymers is of major interest in biophysical chemistry. The naturally occurring amino acid tryptophan (Trp) is of particular interest in fluorescence-based work on peptides and proteins. For example, the quenching of the triplet state of Trp by cysteine has been used to measure the rate of formation of specific intramolecular contacts in disordered peptides.[1] Moreover, Trp can serve as an efficient electron donor in photoinduced electron transfer (PET) reactions with certain dye molecules, a property conferred by the Trp indole side chain which is the most readily oxidized functional group among all naturally occurring amino acids.[2] In an example of the use of PET, it has been shown that efficient PET between Trp and riboflavin in the binding pocket of the riboflavin-binding protein can be used to detect different conformational states of the protein.[3] In the crystal structure of this complex, Trp and riboflavin are oriented in a stacked face-to-face geometry enabling ultrafast charge transfer between their aromatic moieties.

In further promising applications, PET-based biosensors have been developed that use conformationally induced alterations in PET efficiency upon binding for the specific detection of DNA or RNA sequences or antibodies at the single-molecule level.[4,5] These biosensors take advantage of specific properties of naturally occurring DNA nucleotides and amino acids, in particular, the low oxidation potential of Trp and the DNA base guanosine and the tendency of many fluorophores to aggregate in aqueous environment to decrease their water-accessible surface area. In contrast to electronic energy transfer-based systems, in which long-range dipole−dipole interactions occur, the above sensors require contact formation between the fluorophore and the guanosine or tryptophan residue. Depending on the reduction potential of the fluorophore used, efficient fluorescence quenching via PET can then occur. With careful design of conformationally flexible molecules and the use of

[†] IWR-Computational Molecular Biophysics.
[§] Physikalisch-Chemisches Institut.

(1) Lapidus, L. J.; Steinbach, P. J.; Eaton, W. A.; Szabo, A.; Hofrichter, J. *J. Phys. Chem. B* **2002**, *106*, 11628−11640.
(2) Jones, G.; Lu, L. N.; Vullev, V.; Gosztola, D. J.; Greenfield, S. R.; Wasielewski, M. R. *Bioorg. Med. Chem. Lett.* **1995**, *5*, 2385−2390.
(3) Zhong, D.; Zewail, A. H. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 11867−11872.
(4) Neuweiler, H.; Schulz, A.; Vaiana, A. C.; Smith, J. C.; Kaul, S.; Wolfrum, J.; Sauer, M. *Angew. Chem., Int. Ed.* **2002**, *41*, 4769−4773.
(5) Knemeyer, J. P.; Marme, N.; Sauer, M. *Anal. Chem.* **2000**, *72*, 3717−3724.

10.1021/ja036082j CCC: $25.00 © xxxx American Chemical Society

appropriate fluorophores (rhodamine and oxazine dyes are well-suited), efficient single-molecule sensitive PET sensors can be produced. For example, if quenching interactions between the fluorophore and the guanosine or tryptophan residue are hindered upon specific binding to the target (for example, due to binding of a complementary DNA sequence or antibody or due to cleavage by an endonuclease or protease enzyme) the fluorescence of the sensor is restored. In addition to detecting the presence of a target, the contact-induced quenching process can also be used to monitor the rate of intramolecular contact formation in peptides or the opening/closing dynamics of DNA hairpins at the single-molecule level.[6] Therefore, PET-based molecules are ideally suited for single-molecule studies and offer an elegant alternative to conventional biosensors based on electronic energy transfer processes.

Contact formation between Trp and the dye is signaled by a strong quenching of dye fluorescence. To correlate fluorescence spectroscopic data recorded from dye-labeled Trp-containing peptides or proteins with conformational properties, a detailed understanding of the relation between fluorescence quenching by Trp and the underlying interaction geometries is necessary. A complete theoretical understanding of PET reactions and fluorescence quenching in fluorophore/quencher complexes would in principle require accurate quantum chemical calculations of ground and excited-state properties of the complexes involved, including the effects of the surrounding environment. Although the speed and accuracy of quantum chemical methods is ever increasing, these calculations are still computationally too expensive to permit incorporation of complex environmental effects, and they are presently mostly limited to the treatment of relatively small, isolated molecules. One recent example of this is the use of time-dependent density functional theory to describe the fluorescence quenching and lifetimes involved in stacking interactions of 2-aminopurine with purines and pyrimidines.[7] Molecular dynamics (MD) simulations have the advantage over quantum chemical methods of computational speed. Thus, although MD is unable to model the electronic transitions involved in excited-state kinetics, it does allow the incorporation of environmental effects such as explicit solvent molecules together with extensive configurational averaging. Since pioneering work in which fluorescence anisotropy decays of the six Trp residues of lysozyme were examined by MD simulation,[8] a number of MD studies of fluorophores in complex environments have been reported, including a MD study relating the fluorescence quenching and the dynamical behavior of a flavin adenine dinucleotide molecule,[9] several studies of the quenching of tryptophan fluorescence in proteins and peptides,[10−14] and a mixed quantum chemical/molecular mechanics study of the variability of the fluorescence quantum yield of Trp in proteins.[15] Studies concerning the simulations of organic dye structures and their interactions with biomolecules are rare; this is most probably due to the difficulty of developing new force field parameters for dye molecules. Only a few of these studies exist, such as the simulation of the solvation behavior of coumarin 153[16] and of the dynamics of fluorescein in an antibody binding site.[17]

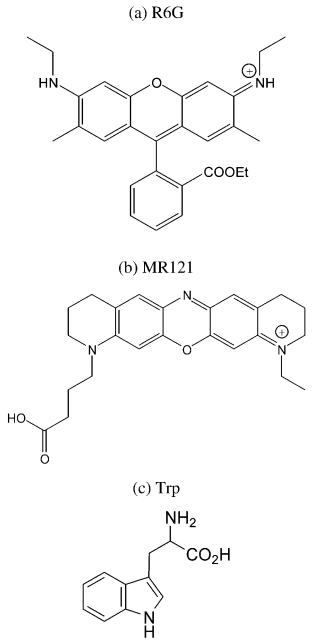

Figure 1. Chemical structures of the two dyes and of the amino acid Trp: (a) R6G, (b) MR121, and (c) Trp.

Here we report a combined experimental fluorescence spectroscopy and MD study of the bimolecular interaction of two organic dyes, Rhodamine 6G (Figure 1a) and the oxazine derivative MR121 (Figure 1b), with Trp (Figure 1c) in aqueous solution. Steady-state and time-resolved fluorescence intensities were measured as a function of Trp concentration, allowing extraction of thermodynamic and kinetic quenching parameters via Stern−Volmer analysis. These parameters are interpreted using MD simulation of the two systems in conditions designed to match the experiments as closely as possible. To enable the simulations to be carried out, it was necessary to derive force-field parameters for the dyes. This was performed using the Automated Frequency Matching Method.[18] The MD simulations, in explicit solvent, enabled two-dimensional potentials of mean force, $\omega(r,\varphi)$ (where r is the distance between the centers of geometry of the dye and quencher ring systems and φ the angle between the normals to the ring planes), to be determined with good statistical accuracy. Significant differences are found both in the experimental effective binding constants and in the simulation-derived potentials of mean force. Interestingly, combining these two quantities, both of which yield information on the probability distributions of different states of the dye/quencher system, leads to the same ring separation cutoff distance, ∼5.5 Å, for fluorescence quenching. A physical explanation of the interactions involved is made possible via

(6) Neuweiler, H.; Schulz, A.; Böhmer, M.; Enderlein, J.; Sauer, M. *J. Am. Chem. Soc.* **2003**, *125*, 5324−5330.
(7) Jean, J. M.; Hall, K. B. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 37−41.
(8) Ichiye, T.; Karplus, M. *Biochemistry* **1983**, *22*, 2884−2893.
(9) van den Berg, P. A. W.; Feenstra, K. A.; Mark, A. E.; Berendsen, H. J. C.; Visser, A. J. W. G. *J. Phys. Chem. B* **2002**, *106*, 8858−8869.
(10) Bismuto, E.; Martelli, P. L.; Casadio, R.; Irace, G. *Protein Sci.* **2000**, *9*, 1730−1742.
(11) Laboulais, C.; Deprez, E.; Leh, H.; Mouscadet, J.-F.; Brochon, J.-C.; Le Bret, M. *Biophys. J.* **2001**, *81*, 473−489.
(12) Yeh, I.-C.; Hummer, G. *J. Am. Chem. Soc.* **2002**, *124*, 6563−6568.
(13) Muino, P. L.; Harris, D.; Berryhill, J.; Hudson, B.; Callis, P. R. *Proc. SPIE-Int. Soc. Opt. Eng.* **1992**, *1640*, 240−251.
(14) Ludescher, R. D.; Peting, L.; Hudson, S.; Hudson, B. *Biophys. Chem.* **1987**, *28*, 59−75.
(15) Callis, P. R.; Vivian, J. T. *Chem. Phys. Lett.* **2003**, *369*, 409−414.
(16) Cichos, F.; Brown, R.; Rempel, U.; von Borczyskowski, C. *J. Phys. Chem. A* **1999**, *103*, 2506−2512.
(17) Lim, K.; Herron, J. N. *Biochemistry* **1995**, *34*, 6962−6974.
(18) Vaiana, A. C.; Schulz, A.; Wolfrum, J.; Sauer, M.; Smith, J. C. *J. Comput. Chem.* **2003**, *24*, 632−639.

the $\omega(r,\varphi)$ plots. Finally, the experimental kinetic parameters are compared with quenching kinetics determined from MD time series analysis: both indicate nanosecond time scale lifetimes of the fluorescent states. The combination of simulation and experiment presented here allows the derivation of a detailed geometric and dynamic model of fluorescence quenching of the dyes by Trp, an approach that will be widely applicable in the interpretation of spectroscopic data recorded from dye-labeled Trp-containing biopolymers, in studies aimed at determining biomolecular structure and dynamics from fluorescence measurements.

2. Methods

2.1. Experimental Procedure. Fluorescence measurements of R6G and MR121 in the presence of different Trp concentrations were performed using standard fluorescence spectrometers. All measurements were performed at room temperature (25 °C) in aqueous solution of phosphate-buffered saline (PBS, pH 7.4). Steady-state fluorescence intensities were recorded with a fluorescence spectrometer LS100 from Photon Technology Int. (Wedel, Germany). Corrected fluorescence spectra were obtained using a high-pressure xenon flash lamp as the excitation source. Ensemble fluorescence lifetimes, τ, were measured with a standard spectrometer from IBH (model 5000MC; Glasgow, U.K.) for time-correlated single-photon counting (TCSPC) using a pulsed diode laser (635 nm) as the excitation source (4096 channels, 12.5 ps/channel, 5000 photons in the maximum channel). To exclude polarization effects, fluorescence was observed under the magic angle (54.7°). The decay parameters were determined by least-squares deconvolution, and their quality was judged by the reduced χ^2 values and the randomness of the weighted residuals. In the present systems the decays were well fitted by a monoexponential function:

$$I(t) = I(0)\exp(-t/\tau) \tag{1}$$

where τ is the lifetime of the excited species and $I(t)$ is the measured fluorescence intensity at time t after excitation. Because of the time resolution limit of the apparatus, strongly quenched populations with decay times shorter than ~50 ps were not detectable.

2.2. Simulations. 2.2.1. Force Field. All molecular mechanics and dynamics computations were performed using CHARMM version 27.[19] The CHARMM potential energy function is given by:

$$V(R) = \sum_{\text{bonds}} K_b(b-b_0)^2 + \sum_{\text{ub}} K_{ub}(s-s_0)^2 + \sum_{\text{angles}} K_\theta(\theta-\theta_0)^2 + \sum_{\text{dihedrals}} K_\chi(1+\cos(n\chi-\chi_0)) + \sum_{\text{impropers}} K_\varphi(\varphi-\varphi_0)^2 + \sum_{\text{nonbond}} \left\{ \epsilon_{ij}\left[\left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{12} - \left(\frac{R_{ij}^{\min}}{r_{ij}}\right)^{6}\right] + \frac{q_i q_j}{r_{ij}} \right\} \tag{2}$$

where K_b, K_{ub}, K_θ, K_χ, K_φ are the bond, Urey−Bradley, angle, dihedral, and improper dihedral force constants, respectively, and b, s, θ, χ, and φ represent the bond lengths, Urey−Bradley 1−3 distances, bond angles, dihedral angles, and improper torsion angles, respectively. The subscript zero, where present, is used to represent the corresponding equilibrium value. Nonbonded interactions between pairs of atoms (labeled i and j) at a relative distance r_{ij} are described by the Lennard-Jones 6−12 and Coulombic interaction terms; $R_{ij}^{\min}$ is the distance between atoms i and j at which the Lennard-Jones potential is zero, and ϵ_{ij} is related to the depth of the Lennard-Jones potential well for the same pair of atoms. q_i is the partial atomic charge on atom i.

Parameters for R6G and MR121 were derived using the Automated Frequency Matching Method (AFMM) described in ref 18. The AFMM method involves refinement of an initial parameter set by fitting to match vibrational eigenvectors and eigenvalues derived from quantum-chemical normal mode calculations. AFMM has already been successfully used to derive CHARMM force field parameters for polycyclic molecules.[18,20] The force field parameter derivation for R6G is described in ref 18. Here we apply the method to derive force-field parameters for MR121, a recently synthesized oxazine dye.[21] No force field parameter set has yet been developed for this molecule. Furthermore, suitable experimental data on which to base a parametrization for MR121 do not exist. Hence, the parametrization scheme used here relies solely on reference data from high-level quantum chemical calculations. All quantum chemical calculations required for parametrization of MR121 were performed with the GAUSSIAN-94 package using the standard 6-31G* basis set[22] and the restricted Hartree−Fock (RHF) level of theory for both geometry optimizations and normal mode calculations.

An initial set of parameters for MR121 was selected for refinement using AFMM. The partial atomic charges, q_i in eq 2, were derived using the CHELPG methodology[23] with a standard RHF/6-31G* quantum chemical optimization. The CHELPG method employs a least-squares fitting procedure to determine the set of atomic partial charges that best reproduces the quantum mechanical electrostatic potential at selected grid points around the molecule. Three new atom types were defined to describe the nitrogens and the oxygen present on the ring system of the dye (see Figure 1b). The van der Waals constants ϵ_{ij} and $R_{ij}^{\min}$ for these were directly transferred from original CHARMM values for NC2, NN3G, and OS atom types (see refs 24−26). van der Waals constants and partial atomic charges were not modified during refinement. Equilibrium values for bonds b_0, angles θ_0, and dihedrals χ_0 that were not present in the original CHARMM force field parameter set[24−26] were derived from the structure resulting from the RHF/6-31G* optimization. An initial guess, based on analogy to similar existing CHARMM parameters and on chemical intuition, was made for all other missing parameters. All force field parameters not given in refs 24−26 are available as Supporting Information.

The initial parameter set was used for minimization and calculation of normal modes (eigenvalues and eigenvectors) with CHARMM.[27] In the force field parameter derivation calculation, all molecular mechanics minimizations were carried out using the Steepest Descent algorithm followed by Newton−Raphson minimization with a convergence criterion for the energy gradient of 10^{-8} kcal/mol/Å. The normal modes obtained were compared with the normal modes calculated with the quantum chemical method, which are considered to be the reference. Parameters were then refined iteratively to fit the results of the quantum chemical normal mode eigenvalues and vectors.

(19) Brooks, B.; Bruccoleri, R.; Olafson, B.; States, D.; Swaminathan, S.; Karplus, M. *J. Comput. Chem.* **1983**, 187−217.
(20) Cournia, Z.; Vaiana, A. C.; Smith, J. C.; Ullmann, M. *Pure Appl. Chem.* **2003**.
(21) Sauer, M.; Zander, C.; Muller, R.; Gobel, F.; Schulz, A.; Siebert, S.; Drexhage, K. H.; Wolfrum, J. *Proc. SPIE-Int. Soc. Opt. Eng.* **1997**, *2985*, 61−68.
(22) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Gill, P. M. W.; Johnson, B. G.; Robb, M. A.; Cheeseman, J. R.; Keith, T.; Petersson, G. A.; Montgomery, J. A.; Raghavachari, K.; Al-Laham, M. A.; Zakrzewski, V. G.; Ortiz, J. V.; Foresman, J. B.; Cioslowski, J.; Stefanov, B. B.; Nanayakkara, A.; Challacombe, M.; Peng, C. Y.; Ayala, P. Y.; Chen, W.; Wong, M. W.; Andres, J. L.; Replogle, E. S.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Binkley, J. S.; Defrees, D. J.; Baker, J.; Stewart, J. P.; Head-Gordon, M.; Gonzalez, C.; Pople, J. A. *Gaussian 94*, revision D.4; Gaussian, Inc.: Pittsburgh, PA, 1995.
(23) Breneman, C. N.; Wiberg, K. B. *J. Comput. Chem.* **1990**, *11*, 361−373.
(24) Foloppe, N.; MacKerell, A. D., Jr. *J. Comput. Chem.* **2000**, *21*, 86−104.
(25) MacKerell, A. D., Jr.; Banavali, N. *J. Comput. Chem.* **2000**, *21*, 105−120.
(26) MacKerell, A. D., Jr.; Bashford, D.; Bellott, M.; Dunbrack, R. L., Jr.; Evanseck, J. D.; Field, M. J.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F. T. K.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D. T.; Prodhom, B.; Reiher, I. W. E.; Roux, B.; Schlenkrich, M.; Smith, J. C.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. *J. Phys. Chem. B* **1998**, *102*, 3586−3516.
(27) Brooks, B. R.; Janezic, D.; Karplus, M. *J. Comput. Chem.* **1995**, *16*, 1522−1542.

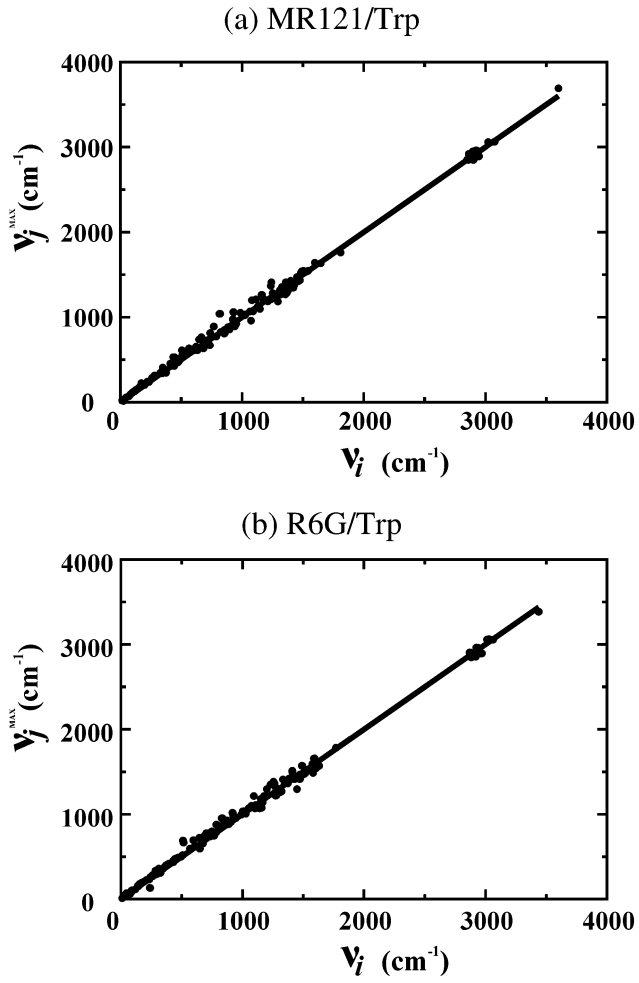

Figure 2. ν_i^{max} vs ν_i plot for MR121 (a) and R6G (b). Lines refer to the ideal case of perfectly matched frequencies and eigenvector projections; points refer to the parameters used in this work.

The AFMM method optimizes frequency matching by using a penalty function that takes into account both frequencies and all the corresponding eigenvectors. We denote the eigenvectors calculated from the MM force field as $\overline{\chi_C^i}$ and the eigenvectors calculated from the quantum chemical method as $\overline{\chi_Q^j}$, where the superscripts i and j represent the normal mode number. An efficient way to check simultaneously for both orthonormality and frequency matching is to project each of the CHARMM eigenvectors, $\overline{\chi_C^i}$, onto the reference set of eigenvectors, $\overline{\chi_Q^j}$, and to find the frequency, ν_j^{max}, corresponding to the highest projection (j: $\overline{\chi_C^i}\cdot\overline{\chi_Q^j}$ = max). In the ideal case of a perfect fit, $\nu_i = \nu_j^{max}$ and $\overline{\chi_C^i}\cdot\overline{\chi_Q^j} = \delta_{ij}$, where δ_{ij} is the Kronecker delta. The ν_i vs ν_j^{max} plot obtained with the parameter set for MR121 used here is shown in Figure 2a, along with the same plot for R6G (Figure 2b, data taken from ref 18). Values of the root-mean-square deviation from the reference case are 54.8 cm^{-1} for MR121 and 51.1 cm^{-1} for R6G. These values are within the range seen in previous benchmark studies on small aromatic compounds.[18]

The AFMM method optimizes intramolecular vibrations to reproduce the ground-state properties of a molecule in a vacuum. To model intermolecular interactions AFMM uses, as do most other MM parametrization methods, the correct reproduction by the partial atomic charges of the electrostatic potential as calculated quantum mechanically, together with the Lennard-Jones parameters. The Lennard-Jones parameters used here were transferred from existing CHARMM parameters that were derived from experimental solvation free energies and quantum chemical calculations on small model compounds.[25] The combination of AFMM, CHELPG for the atomic charges, and standard CHARMM Lennard-Jones parameters has proven to successfully reproduce sensitive intermolecular properties such as the crystal structures of both R6G[18] and cholesterol (Z. Cournia, personal communication).

2.2.2. Molecular Dynamics Simulations. MD simulations were performed on two model dye/Trp systems: the first (MR121/Trp) consisting of one MR121 dye molecule, one Trp molecule, 1013 TIP3P[28] water molecules, and a chloride counterion and the second (R6G/Trp) consisting of one R6G molecule, one Trp molecule, 1002 TIP3P[28] water molecules, and a chloride counterion.

In a recent MD study of intrapeptide interactions it was noted that the TIP3P model underestimates the viscosity of water, and the calculated time constants were thus scaled.[12] However, whereas the dynamics involved in ref 12 occurs over a long time scale and has large amplitude, that studied here is faster and more local, resulting from close-range, atomic-level interactions between the dyes and the Trp molecule that mostly involve small numbers of water molecules. Thus, viscosity plays no clear role in determining the relaxation times examined here, and no results were scaled.

Five MD simulations were performed on each of the two dye/tryptophan systems. In what follows, the simulations are referred to as RT1, ..., RT5 for the R6G/Trp system and MT1, ..., MT5 for the MR121/Trp system. For each system, the five simulations were performed with different initial ring separations so as to improve configurational statistics, as described below. All simulations were performed in the NPT ensemble at 1.0 atm pressure and at 300 K in a truncated octahedral box of (33.6 ± 0.1) Å for MR121/Trp and (33.9 ± 0.1) Å for R6G/Trp using periodic boundary conditions. This corresponds to relative Trp concentrations of 55 mM for the MR121/Trp system and 57 mM for R6G/Trp. Long-range electrostatic interactions were computed using the Particle Mesh Ewald (PME) method.[29]

The heating and equilibration procedure was the same for all the MD runs. An arbitrary starting configuration of solvated dye/tryptophan including counterions was gradually heated from 0 to 300 K during 5.5 ps, reassigning velocities every 100 steps using an integration time step of 0.1 fs. The integration time was then increased to 1 fs, and a first 100-ps equilibration was performed at constant volume and temperature. The system was then allowed to relax to its equilibrium density at constant temperature and pressure for 100 ps. At this point the distance r, between the geometrical centers of the main ring systems of the dye and of the tryptophan (the definition of r is shown in Figure 3a), was subjected to a harmonic constraint $C(r) = k(r - r_0)^2$ with the force constant k set to 10 kcal/mol/Å^2 for a total of 100 ps. Values of the equilibrium distance r_0 were 3.5 Å for RT1 and MT1, 5.0 Å for RT2 and MT2, 7.5 Å for RT3 and MT3, 10.0 Å for RT4 and MT4, and 12.5 Å for RT5 and MT5. Subsequently, the system was allowed to relax unconstrained for another 100 ps prior to the production phase.

During the production runs no constraints were used. Each production run lasted 10 ns, i.e., 50 ns of simulation time was run in total on each dye/Trp system. Altogether, the simulations required 41 days (wall time) on 32 CPUs (16 per dye/Trp system) of the HELICS cluster of the Interdisziplinäres Zentrum für Wissenschaftliches Rechnen at the University of Heidelberg. Configurations were downloaded to disk every 0.1 ps for subsequent analysis.

2.2.3. Potential of Mean Force: Theoretical Background. The mutual arrangement of the two molecules (Trp and dye) is here described by the distance r defined above and the angle φ between the normals to the planes of the main ring systems of the dye and the tryptophan (see Figure 3a). The MD simulations were used to determine the two-dimensional potential of mean force (PMF) landscape, $\omega(r,\varphi)$. This allows the free-energy differences between the complexed and

(28) Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. *J. Chem. Phys.* **1983**, *79*, 926−935.
(29) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577−8593.

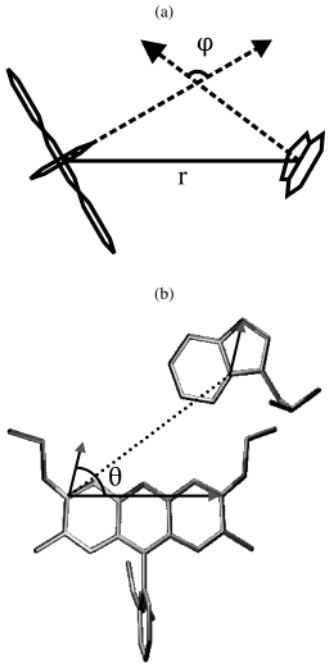

Figure 3. (a) Definitions of the dye/Trp distance, r, and the angle, φ, between the normals to the dye and Trp planes. (b) Definition of the angle, θ, between two fixed vectors lying on the dye and Trp planes. The R6G/Trp system is depicted here. The same definitions are used for the MR121/Trp system.

noncomplexed states of the dye/Trp system and the configurational pathways and free energy barriers connecting these states to be determined.

The simulation models a classical thermodynamic system in the NPT ensemble: its microscopic states are described by a set of $3N$ coordinates $\{r, \varphi, q_1, ..., q_{3N-2}\}$ and their associated momenta $\{p_r, p_\varphi, p_1, ..., p_{3N-2}\}$. At equilibrium, the probability $P(r, \varphi)$ of finding the system in a given state of the $\{r, \varphi\}$ subspace, i.e., the probability of finding given values of r and φ is:

$$P(r,\varphi) = \frac{\int_V \cdots \int \exp[-U(r,\varphi,q^1,...,q^{3N-2})/kT]\, \mathrm{d}q_1 ... \mathrm{d}q_{3N-2}}{Z} \quad (3)$$

where k is the Boltzmann constant, T is the temperature of the system, $U(r, \varphi, q_1, ..., q_{3N-2})$ is the potential energy, and Z is the reduced partition function. The integrals are extended to the hypervolume V of phase space spanned by all degrees of freedom except r and φ. $\omega(r,\varphi)$, the associated PMF,[30] is given by:

$$\omega(r,\varphi) = -kT \ln[P(r,\varphi)] \quad (4)$$

The negative gradient of $\omega(r,\varphi)$ with respect to r and/or φ, $-\nabla_{r,\varphi}\, \omega(r,\varphi)$, is the mean "thermodynamic drive" acting on the system along the coordinates r, φ, or both averaged over all possible states of the remaining $3N - 2$ degrees of freedom. Thus, $\omega(r,\varphi)$ is a free energy i.e., the reversible work done on the system in moving it from state (r_a,φ_a) to state (r_b,φ_b) is expressed by the difference $\Delta\omega^{a,b} = \omega(r_b,\varphi_b) - \omega(r_a,\varphi_a)$.

$P(r,\varphi)$ can be estimated using MD simulation. If $n(r,\varphi)$ is the number of MD snapshots in which the system is found to be in a volume element $\mathrm{d}\Omega = 2\pi r^2 \sin\varphi\, \mathrm{d}r\, \mathrm{d}\varphi$ around position (r,φ), then $P(r,\varphi)$ is given by:

$$P(r,\varphi) \approx \frac{n(r,\varphi)}{N_{\mathrm{MD}}\, \mathrm{d}\Omega} \quad (5)$$

where N_{MD} is the total number of MD snapshots taken during the simulation. If the system is ergodic, the equality in eq 5 is exact in the limit $N_{\mathrm{MD}} \rightarrow \infty$.

In the present work, for both systems simulated $n(r,\varphi)$ was evaluated on a 100 × 100 square grid over 0 Å < r < 25 Å and 0 < φ < π, resulting in a bin width of 0.25 Å for r and 1.8° for φ.

The statistical error due to incomplete sampling in calculating $\omega(r,\varphi)$ with a finite value of N_{MD} in eq 5 was estimated in the following manner. The total simulation run was divided into M segments of equal length, and for each of these segments the PMF was independently calculated. The average standard deviation from the PMF for $M = 1$ is an estimate of the desired statistical error and is evaluated as follows:

$$\sigma_M = \left\langle \left\{ \frac{1}{M} \sum_{i=1}^{M} [\omega_i(r,\varphi) - \omega_{\mathrm{tot}}(r,\varphi)]^2 \right\}^{1/2} \right\rangle_{r,\varphi} \quad (6)$$

where $\omega_i(r,\varphi)$ is the PMF, as defined in eq 4, calculated from simulation data from the ith segment and $\omega_{\mathrm{tot}}(r,\varphi)$ is calculated from the whole simulation. The average in the angular brackets spans over the whole $\{r, \varphi\}$ space.

3. Results

3.1. Experimental. To determine the fluorescence quenching efficiency and aspects of the quenching mechanism of the dyes by Trp, steady-state and time-resolved fluorescence intensities of R6G and MR121 were measured as a function of the Trp concentration in phosphate-buffered aqueous solution. Plots of the inverse-normalized steady-state fluorescence intensities and inverse-normalized fluorescence lifetimes of the dyes vs the Trp concentration allow thermodynamic and kinetic quenching data to be determined via Stern−Volmer analysis (Figure 4). The corresponding plots for the dye/quencher pair R6G/Trp are shown in Figure 4a along with those for MR121/Trp (Figure 4b: data taken from ref 6) for comparison. The Stern−Volmer data from both dye/Trp systems are summarized in Table 1.

The bimolecular dynamic quenching rate, $k_{\mathrm{q,d}}$, i.e., the rate at which fluorescence-quenched collisional encounter complexes between the dye and Trp are formed, can be extracted from the fluorescence lifetime data according to the "dynamic" Stern−Volmer equation:

$$\frac{\tau_0}{\tau} = 1 + k_{\mathrm{q,d}} \tau_0 [\mathrm{Trp}] \quad (7)$$

where τ denotes the measured fluorescence lifetime, τ_0 the lifetime of the free dye in solution, and [Trp] the Trp concentration. This diffusion-controlled process leads to a reduced fluorescence lifetime of the dye in the presence of Trp.

The fluorescence photon decay distributions (i.e., the time dependence of the fluorescence intensity) of the dyes were found to fit well a monoexponential model in the presence of Trp with a lifetime which is close to that of the free dye. This indicates that the complexes are essentially nonfluorescent, i.e., they exhibit a fluorescence lifetime shorter than the time-resolution of the instrument (~50 ps) used to measure the fluorescence decays. Table 1 shows that $k_{\mathrm{q,d}}$ is the same for both systems to within experimental error, whereas τ_0 is significantly longer for R6G.

Complex formation between dye and Trp reduces the amount of free fluorescent dye in solution and therefore also reduces the fluorescence intensity of the measured ensemble. The static

(30) Hill, T. L. In *Statistical Mechanics, Principles and Selected Applications*; Dover Publications, Inc.: New York, 1956; pp 179−285.

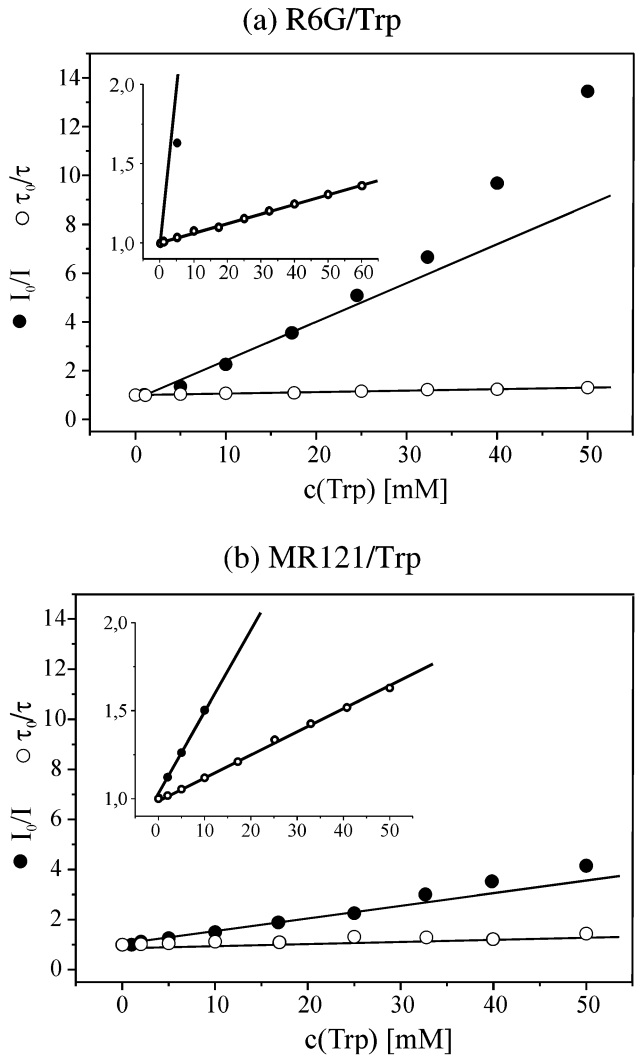

Figure 4. (a) Static (closed circles) and dynamic (open circles) bimolecular Stern−Volmer plot of R6G/Trp in phosphate-buffered saline solution, pH 7.4. Excitation wavelength: 495 nm, detection wavelength: 555 nm. (b) An equivalent plot for MR121 is shown (data taken from ref 6). Insets represent the same data as the main graphs on an expanded *y*-axis scale.

Table 1. Experimental Fluorescence Quenching Data for the Dyes MR121 and R6G with Trp in Phosphate-Buffered Saline Solution, pH 7.4[a]

	MR121	R6G
$k_{q,d}/M^{-1}\,sec^{-1}$	$(3.2 \pm 0.3) \times 10^9$	$(3.4 \pm 0.3) \times 10^9$
K_S/M^{-1}	220 ± 20	30.4 ± 3
τ_0 [ns]	1.85 ± 0.06	3.85 ± 0.12

[a] $k_{q,d}$ denotes the bimolecular dynamic quenching rate, K_S the static Stern−Volmer constant, and τ_0 the lifetime of the free dye in solution.

Stern−Volmer constant, K_S, i.e., the thermodynamic association constant for the formation of nonfluorescent ground-state complexes between the dye and Trp, can be extracted from the slope of the linear region of the concentration-dependent steady-state fluorescence intensities according to the "static" Stern−Volmer equation:

$$\frac{I_0}{I} = 1 + K_S[\mathrm{Trp}] \tag{8}$$

where I denotes the measured steady-state fluorescence intensity and I_0 denotes the fluorescence intensity of the free dye in solution. In the case of a quenching mechanism containing both static and dynamic components, the intensity plots are nonlinear as the dependence of I_0/I becomes second order in [Trp]. This is the case here for both dye/quencher pairs.

K_S is about seven times higher for MR121/Trp than for R6G/Trp. This reveals a stronger tendency of Trp to form non- or only weakly fluorescent ground-state complexes with MR121 than with R6G.

The time-resolved fluorescence quenching experiments are consistent with a model in which the underlying quenching mechanism involves dye/quencher complex formation. This conclusion is supported by the observation that for both systems the static Stern−Volmer constant, K_S, strongly decreases upon addition of detergent or organic solvents such as ethanol (data not shown). This suggests that hydrophobic interactions between MR121 or R6G and Trp may play an important role in the formation of the nonfluorescent complexes. In this case, the mechanism would involve a PET reaction between the Trp and the dye in stacked interaction geometries, similar to that observed for the dye/quencher pair riboflavin/Trp in the riboflavin-binding protein.[3]

3.2. MD Simulation Results. To obtain an atomic-detail description of the geometries involving fluorescence quenching of the organic dyes by Trp, MD simulations were performed on the systems MR121/Trp and R6G/Trp in aqueous solution.

$\Phi_{f,rel}$ ($= I/I_0$) is the relative fluorescence quantum yield. The values of $\Phi_{f,rel}$ obtained at Trp concentrations corresponding to those used in the simulations are ~0.25 for the R6G/Trp system and ~0.10 for the MR121/Trp system, as determined from Figure 4. This corresponds to a single dye molecule being in a fluorescent state for ~25% of the observation time for R6G and ~10% for MR121. The question arises as to what geometries correspond to the nonfluorescent states for the two systems. To examine this, the one-dimensional, normalized probability distribution, $P(r)$ of the dye-quencher distance r, was calculated from the simulations. The results are shown in Figure 5. The probability distributions of the two dye/quencher systems are significantly different, with a higher relative population at $r > 5$ Å for R6G.

The solid line in Figure 5a,b is the integral of $P(r)$:

$$F(r^*) = \int_0^{r^*} P(r)\,\mathrm{d}r \tag{9}$$

$F(r^*)$ is thus the fraction of configurations observed for which $r < r^*$. Using the above-determined experimental probability values of $F = 75\%$ for R6G/Trp and $F = 90\%$ for MR121/Trp, we can calculate the corresponding distances r^* from Figure 5. These distances, shown by dotted lines in Figure 5, are $r^* = (5.5 \pm 0.4)$ Å for R6G/Trp and $r^* = (5.4 \pm 0.2)$ Å for MR121/Trp. The error in r^* was estimated in the same manner as the error σ_5 in the PMF landscape described in eq 6. These two r^* values are remarkably similar given the significantly different form of the probability distributions $P(r)$ for the two systems. Configurations occurring at distances >5.5 Å may be considered to be fluorescent. The minimum of the Lennard-Jones potential for the carbon atoms on the ring systems is at ~2.0 Å. Therefore, the values of r^* derived above indicate that quenched configurations occur for both dyes only at distances close to contact with the Trp.

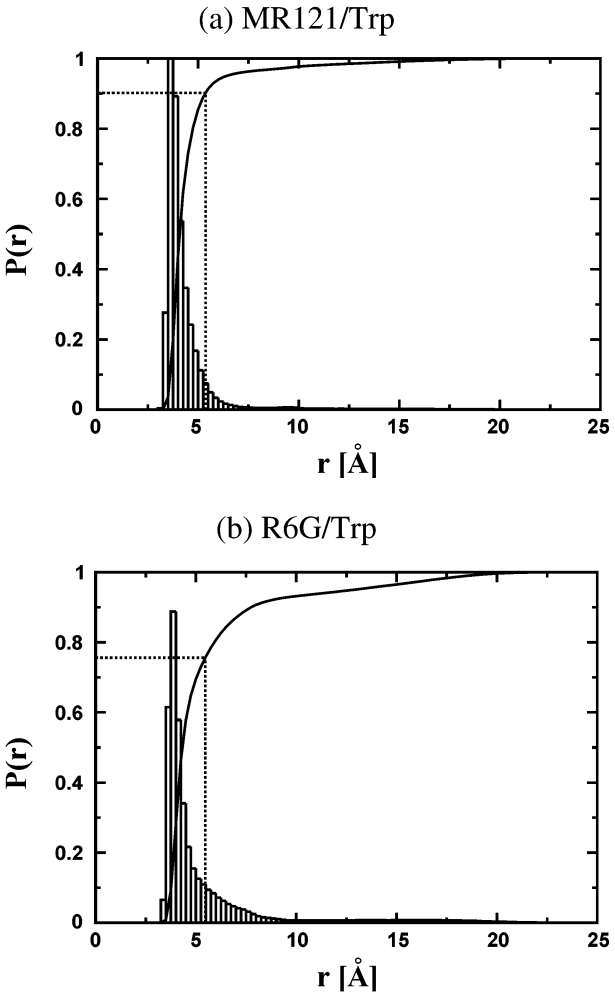

Figure 5. Normalized probability distributions $P(r)$ for MR121/Trp (a) and R6G/Trp (b). The continuous line is the integral of $P(r)$, i.e., $F(r^*)$ defined in eq 9. The dotted line indicates the value of $r^* = (5.4 \pm 0.2)$ Å for MR121/Trp and $r^* = (5.5 \pm 0.4)$ Å for R6G/Trp, corresponding to the experimental quenching efficiencies of 90 and 75%, respectively. The relative difference between the values of $F(r^*)$ for the two dyes is $\Delta F(r^*) = (15 \pm 4)\%$. The error in r^* was estimated in the same manner as the error σ_5 in the PMF landscape described in eq 6; the estimated error in $\Delta F(r^*)$ is back-propagated from the error in r^*. The trajectories were divided into five segments of equal length, and for each of them r^* was calculated. The standard deviation from the value of r^* derived from the whole simulation gives the error shown above.

To provide further details on the geometries of the complexes, two-dimensional $\{r, \varphi\}$ PMF maps calculated from the MD simulations are shown for both systems in Figure 6. Values of the statistical error σ_M, calculated using eq 6, are reported for $M = 2$ to $M = 7$ in Table 2. These values indicate a statistical error in the PMF landscape in the range 0.21−0.43 kcal mol^{-1} for R6G/Trp and 0.16−0.38 kcal mol^{-1} for MR121/Trp. These are similar to the values in the range 0.26−0.31 kcal mol^{-1} obtained in a recent MD analysis of the PMF landscape of aromatic amino acid complexes in water.[31]

The continuous white line in Figure 6 is placed at the quenching distance $r = r^*$. To a first approximation the landscapes exhibit similar topologies. Both possess two pronounced minima labeled A_1 and A_2. These minima correspond to close-contact stacking of the ring systems of the dye with

(31) Chelli, R.; Gervasio, F. L.; Procacci, P.; Schettino, V. *J. Am. Chem. Soc.* **2002**, *124*, 6133−6143.

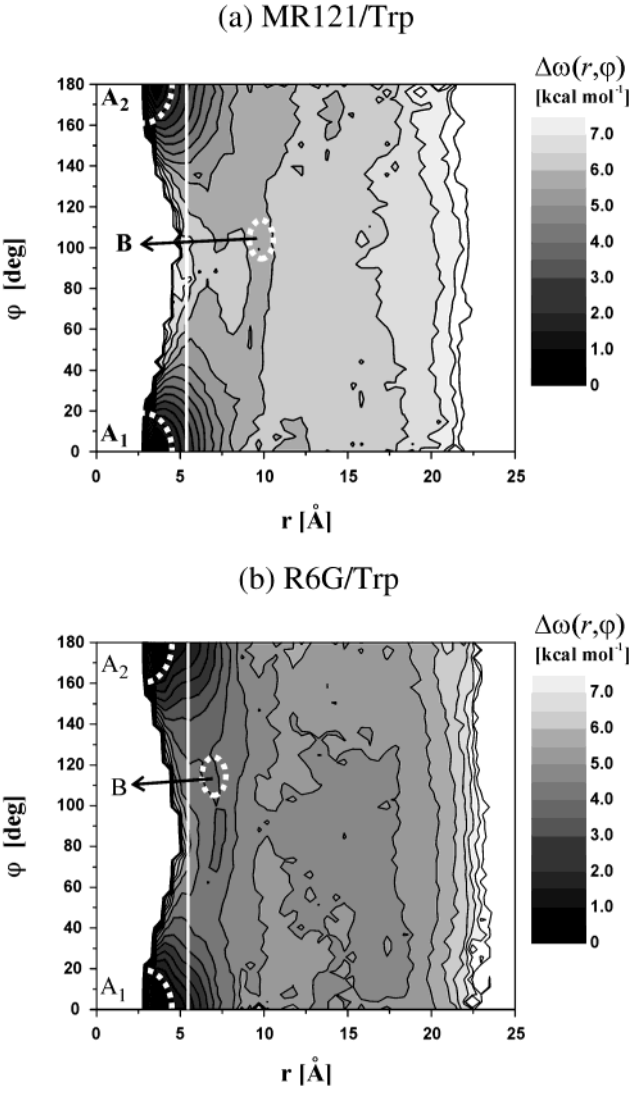

Figure 6. Two-dimensional PMF landscapes, $\Delta\omega(r,\varphi)$ derived using eq 4 for MR121/Trp (a) and R6G/Trp (b). For both, the PMF value is set to 0 kcal mol^{-1} at the global minimum. Dashed, white lines serve as a guide to the eye and mark respectively the areas A_1, A_2 (locations of the two PMF minima corresponding to close, stacking configurations of the complexes), and B (along the connection pathway between the minima) selected for subpopulation analysis (see Figures 7 and 8). The continuous white lines are placed at the quenching distances $r^* = 5.5$ Å for MR121/Trp and $r^* = 5.4$ Å for R6G/Trp.

Table 2. Values of σ_M, the Average Standard Deviation of the PMF for M Segments of Simulation, as Calculated from Eq 6 for Both Dye/Trp Systems

M	σ_M (MR121) [kcal mol^{-1}]	σ_M (R6G) [kcal mol^{-1}]
2	0.16	0.21
3	0.29	0.29
4	0.31	0.33
5	0.34	0.36
6	0.37	0.39
7	0.38	0.43

the tryptophan, well within the quenching distance. A clearly visible "low-energy" pathway connects the two minima and extends beyond the quenching distance. The PMF barrier along the pathway for the MR121/Trp system is ∼5 kcal mol^{-1}, about 1 kcal mol^{-1} higher than that of the R6G/Trp system. Both minima A_1 and A_2 are ∼1 kcal mol^{-1} shallower for R6G/Trp than for MR121/Trp. In contrast to the R6G/Trp landscape, the

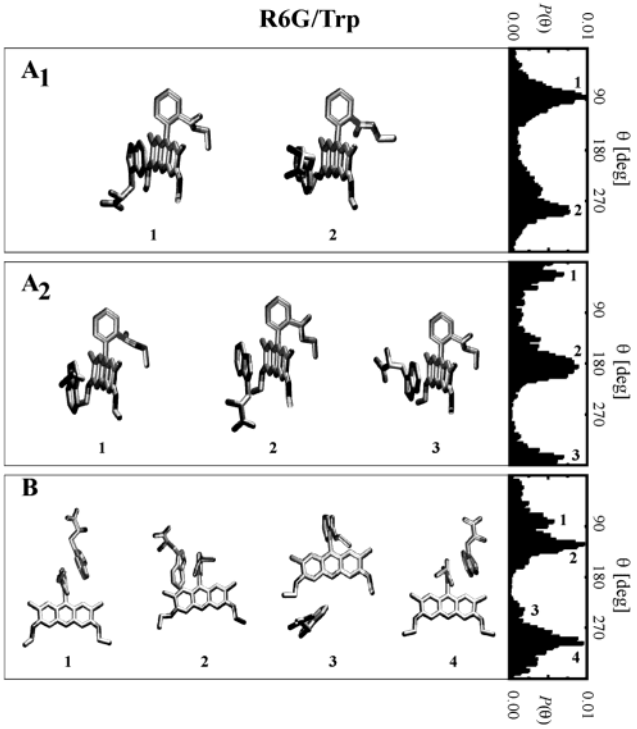

Figure 7. Average structures of the most populated subpopulations of the R6G/Trp complex occurring during the simulation runs within the areas defined in Figure 4 and marked A_1 (top panel), A_2 (middle panel), and B (bottom panel). Corresponding probability distributions of the angle θ are on the right of each panel.

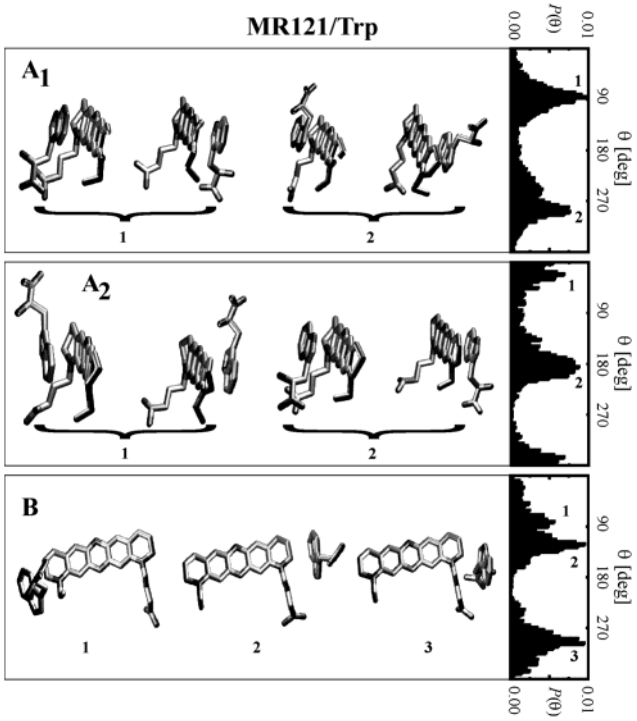

Figure 8. Average structures of the most populated subpopulations of the MR121/Trp complex occurring during the simulation runs within the areas defined in Figure 6: A_1 (top panel), A_2 (middle panel), and B (bottom panel). Corresponding probability distributions of the angle θ are on the right of each panel.

PMF of the MR121/Trp complex is almost symmetrical around $\varphi = 90°$; this reflects the fact that MR121 has planar symmetry, whereas in R6G this symmetry is broken due to the presence of the ester side chain. In both systems, beyond a separation distance of ~10 Å the PMF exhibits no significant dependence on φ.

We now examine more closely subpopulations of the stacked configurations A_1 and A_2 as well as the configurations falling in the areas along the low-energy configurational pathway between the minima (the area marked as B in Figure 6a,b). We define θ as the angle between a fixed vector on the dye plane and a fixed vector on the Trp plane (see Figure 3b). Subpopulations can be then classified according to the distribution of θ within each region A_1, A_2, and B. The Cartesian coordinates of the most frequently occurring configurations (i.e., those configurations falling within the peaks of the θ distribution) were averaged to obtain a representative structure of each subpopulation; these are shown in Figures 7 and 8.

For the R6G/Trp system, stacking of the Trp occurs only on one side of the dye plane, whereas stacking on the other side is hindered by the presence of the R6G ester group. In contrast, due to the symmetry of the MR121 dye, in the MR121/Trp system the Trp interacts with both sides of the MR121 plane. The resulting average structures within A_1 and A_2 for MR121 consist of two pairs of almost perfect mirror images (panels marked A_1 and A_2 in Figures 7 and 8). The fact that the free-energy minima (A_1 and A_2 in Figure 6) are shallower for R6G than for MR121 is also due to the presence of the ester group on R6G, which sterically hinders the dye/Trp interaction. This effect might conceivably be reduced in rhodamine derivatives that have smaller groups at the ester position.

The average structures of the complexes along the connection/escape pathway between minima are shown in panel B of Figures 7 and 8. The most frequently occurring configurations for the R6G case (Figure 7, panel B) are those in which the Trp interacts mainly with the phenyl ring of the rhodamine dye (pictures 1, 2, and 4 from the left). Visual inspection of the simulation trajectories showed that these configurations often lead to escape/entrance of the tryptophan to/from regions of the landscape beyond $r = 10$ Å. Interactions between the Trp and the phenyl ring of R6G, which are obviously not present in the case of MR121, are responsible for the fact that in Figure 5 $P(r)$ of R6G is higher than MR121 in the region 5−10 Å. This Trp/phenyl ring is a nonquenching interaction as is the Trp/ester group interaction mentioned in the previous paragraph. These two interactions compete with the quenching complexed geometries. For the case of MR121 (Figure 8, panel B), detailed inspection of the trajectories revealed that the Trp is most often found to slide along the MR121 plane before actually separating from the dye.

The bimolecular dynamic quenching rate, $k_{q,d}$ can be derived from the reduction in the average fluorescence lifetime, τ (eq 7), of the dye in the presence of quencher relative to the free dye. This reduction is due to dye/Trp collisional encounters that quench the dye fluorescence earlier than the decay from the excited state of the free dye. τ (or equivalently $k_{q,d}$) is a purely dynamical quantity: it represents the average duration of nonquenched states of the dye, and it is independent of K_S.

Values of τ were obtained from eq 7 using the experimental values of $k_{q,d}$ and τ_0 from Table 1 and the Trp concentrations used in the simulations, i.e., 55 mM for the MR121/Trp system and 57 mM for the R6G/Trp. The resulting values of τ are 2.20 ± 0.07 ns for R6G/Trp and 1.40 ± 0.04 ns for MR121/Trp.

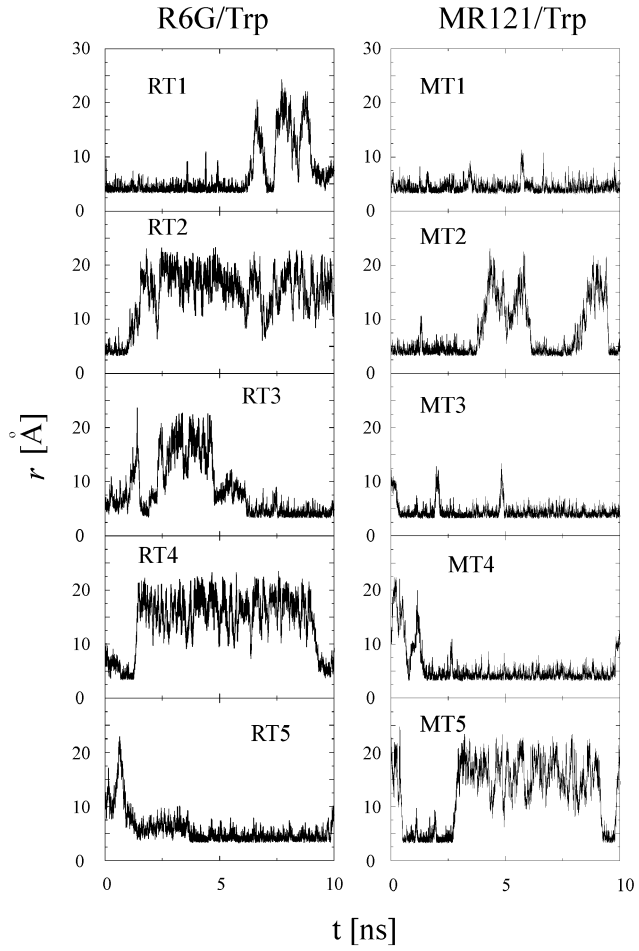

Figure 9. Time series of r during the MD runs.

In principle, τ can be estimated from MD simulation given a geometric definition of nonquenched states. The definition assumed here is as that used above: the dye is assumed to be quenched if $r < r^*$. With this definition, the duration of the ith nonquenched state occurring along the MD trajectory, $\Delta t_i = (t_{u,i} - t_{q,i})$ is simply the difference between the time $t_{u,i}$ at which the system passes from a quenched ($r < r^*$) to an unquenched ($r > r^*$) state and the time $t_{q,i}$ at which it returns to a quenched state. τ is calculated from the MD data in a similar manner as for the experiment, i.e., by fitting an exponential decay function to the probability distribution for Δt_i. To compare with the experimentally measured values of τ, transitions were excluded for which Δt_i was smaller than the time resolution of the instrument (which is ~50 ps).

The time series of r during the MD runs are shown in Figure 9 for both dye/quencher systems. The total number of transitions found in the simulations is relatively small, i.e., 30 for R6G/Trp and 21 for MR121/Trp. The resulting values for τ are 1.7 ± 0.2 ns for R6G/Trp, (cf. the experimental value of 2.20 ± 0.07 ns) and 1.5 ± 0.3 ns for MR121/Trp compared to the experimental value of 1.40 ± 0.04 ns. Given the nanosecond time scale of the fluorescence lifetimes in the present systems, longer MD simulations would be required for an accurate evaluation of this property from the simulation model. Nevertheless, the values of τ obtained from the simulation agree with the experimental values rather well (to within experimental error for MR121/Trp and almost within experimental error for R6G/Trp). That τ is of the same order as the lifetime of the free species, τ_0, confirms the small contribution of dynamical quenching relative to the contribution arising from stable, nonfluorescent complexes. The predominance of static quenching is also consistent with the fact that the time series in Figure 9 evidently exhibit confined, rather than free, diffusional characteristics.

4. Conclusions

The present work demonstrates that the combination of MD simulation with time-resolved fluorescence experiments can provide an atomic detail description of the geometry and kinetics of the quenching interaction of Trp with fluorescent dyes. The MD simulations determine molecular interaction geometries which can then be matched with results from appropriate fluorescence experiments.

In both systems studied here, the quenching mechanism is mostly static and due to formation of nonfluorescent complexes. At the Trp concentration used in the simulations and experiments, the MD-derived $P(r)$ distribution extends to higher distances for R6G than for MR121 due to the population of fluorescent interaction geometries between Trp and the R6G phenyl ring and ester group. As a consequence of this difference, about 75% of the R6G dye is quenched in comparison with 90% of the MR121, as seen experimentally. Combining these observations with the MD-derived $P(r)$ allows determination of the "quenching distance" r^*. This distance turns out to be physically reasonable and very similar (~5.5 Å) for both dye/Trp systems, corresponding to close to van der Waals contact. Finally, the lifetimes of the fluorescent states are on the nanosecond time scale and agree to within experimental error for MR121/Trp and almost within experimental error for R6G/Trp, further validating the quenching distance derived. To extend our understanding of the detailed quenching process beyond the simple "quenching distance" concept used here will require quantum chemical analysis by using, for example, the geometries derived from MD simulation as input for appropriate molecular orbital calculations.

Intramolecular dye/Trp fluoresence quenching experiments are widely applied to a variety of biomolecular systems of fundamental, medical, and technological interest. Many of the interactions involved, such as those employing peptide conjugates, involve a considerably larger number of intramolecular degrees of freedom than in the systems studied here, thus complicating the obtention of converged interaction probability distributions from computer simulation. However, it is likely that with the use of, for example, coarse-grained parallel cluster computing, useful information can also be obtained using the present approach for these systems. Thus, a basis will be provided for quantitatively interpreting fluorescence quenching data in terms of the underlying molecular interaction geometries.

Acknowledgment. The authors thank G. M. Ullmann, T. Becker, and A. Tournier for fruitful discussions, Z. Cournia for access to unpublished results, K. H. Drexhage for providing the oxazine derivative MR121, and B. Costescu for skillful technical assistance.

Supporting Information Available: MR121 force field parameters (PDF). This material is available free of charge via the Internet at http://pubs.acs.org.

JA036082J

Acknowledgments

At this point, I wish to thank all the people who have supported me during my thesis work and without whom this work would not have been possible.

In particular I wish to thank Prof. Wolfrum and Prof. Smith for allowing me to work on this very interesting subject and for the great interest they showed in my work.

A very special thanks goes to Andreas Schulz for bearing with me for more than two years, with his innate chemical intuition he greatly contributed to the parameterisation of the dyes, and to Markus Sauer and Hannes Neuweiler for their experimental expertise and for having given me access to unpublished results.

I wish to thank Bogdan Costescu for his skilful technical assistance, and Dan Michailescu and Stefan Fischer for their help with CHARMM and fruitful discussions.

A great thanks goes to Matthias Ullmann, Torsten Becker and Alexander Tournier for the very interesting, illuminating discussions and for creating a stimulating working atmosphere.

I wish to thank all members of the Computational Molecular Biophysics group in the IWR and everyone in the Biophysical Chemistry group for discussions and the great working atmosphere.

I am also grateful to Prof. Eric Westhof for his patience while I was finishing the writing of this thesis, and to Prof. M.U. Palma and Prof. M.B. Palma-Vittorelli who taught me much more than “just” physics.