1. General Information (Algemeine Angaben)

New proposal for DGF Emmy Noether program

1.1 Applicant (Antragsteller)

Name:	Gerrit Groenhof, Ph. D.
Contract:	TVöD until 30 juni 2010
Date of birth:	18 November 1976
DFG reference number:	-
Institution and department:	Max-Planck-Institut für biophysikalische Chemie Abteilung Theoretische und Computergestützte Biophysik
Work address:	Am Fassberg 11 D-37077, Göttingen
Telephone number:	+49(0)551-201-2321
Fax number:	+49(0)551-201-2302
Email address:	ggroenh@gwdg.de
Home address:	Schmiedestrasse 9 D-37077, Göttingen
Home telephone number:	+49(0)551-309-7157

1.2 Topic (Thema)

Bio-inspired design of molecular motors

1.3 Research area and field of work (Fach und Arbeitsrichtung)

Discipline:	Physical and theoretical chemistry
Specialisation:	Biomolecular modeling, nanotechnology

1.4 Anticipated total duration (Voraussichtlicher Gesamtduaer)

Duration:	5 years

1.5 Application period (Antragszeitraum)

Duration:	60 months
Starting date:	1 juli 2010

1.6 Summary (Zusammenfassung)

1.6.1 summary

Manipulating matter at the atomic level by means of molecular machines is one of the major challenges in nanotechnology. An essential element of such machine is the motor that performs mechanical work under the control of external stimuli. Because the conversion of energy into motion is a common process in many biological systems, the development of artificial motors can benefit from understanding the principles by which these biological systems operate. To exploit these princinples it is necessary to understand how the protein controls the motion at a molecular level. As the relevant time and spacial resolution are notoriously difficult to access experimantally, we will make use of computer simulations to reveal how the protein enviroment mediates the photochemical process. With these insights, we will then create artificial systems in which we control the photochemistry by means of chemical substituents that mimic the protein. In particular, we will design rotating molecular motors that generate mechanical force under the influence of light, and light-driven proton pumps that transport protons between media.

1.6.2 Zusammenfassung

Die Manipulation von materie auf atomares ebene mit molekulaire machinen ist einer von die groeste herausfoerderungen in nanotechnologie. Ein essentieles element fuer so eine machine ist das motor, welche mechanische arbeit leistet unter die einfluss von externe stimulanzen. Weil die umsetzung von energy in bewegung ein algemeines proces ist in viele biologische systeme, kann es die entwicklung von artifiziele motoren stark nutzen um die prinzipien woauf die operation von diese biologische systeme basiert ist zu begreifen. Um diese prinzipien an zu wenden ist es notwendig zu verstehen wie das protein die bewegungen auf moleculares ebene controliert. Da die dabei relevanten Zeit- und Raumskalen mit experimentellen Methoden sehr schwierig zu erreichen sind, ist hier die Computersimulation die Methode der Wahl um auszufinden wie die Protein umgebung die photochemische Reaktionen ansteuert. Wir werden diese Erkentnisse für die Entwicklung von artifiziele systeme anwenden in denem wir die photochemische reaktionen controlieren durch chemische substituenten die die protein umgebebung imitieren. Ins besonderes sind wir vornemens zwei typen molekulare motoren zu entwicklen: rotierende molekulare motoren die mechanische kraft generieren unter einfluss von licht, und licht-gesteuerte protonon pumpe die protonen transportieren koennen zwischen zwie medien.

2. State of the art, preliminary work (Stand der Forschung, eigene Vorarbeiten)

2.1 State of the art (Stand der Forschung)

Manipulating matter at the atomic level by means of molecular machines is one of the major challenges in nanotechnology. The essential element of a machine is the engine that can perform mechanical work under the control of external stimuli. Despite tremendous progress in the elucidation of the atomic details of biomolecular motors like ATPsynthase[1], kinesin, myosine[2], DNA polymerase[3], the design of artificial molecular motors has not been very much inspired by these biological examples.

The first succesful light-driven rotational motors were developed by Feringa and co-workers, and are based on chiral overcrowed alkenes[4,5] (Fig. 1). Photo-isomerization around the central double bond places the achiral methyl substituents in an energetically unfavourable position, so that thermal relaxation at elevated temperatures completes the 360° rotation. Based on the same principle the Feringa group has recently also developed a rotational motor powered by chemical energy[6].

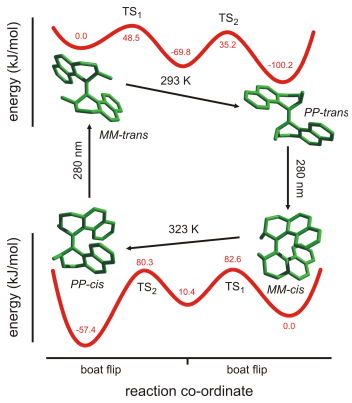

Figure 1: In the overcrowded alkene motor unidirectional 360° rotation, indicated by arrows, is achieved by photo-isomerizations followed by thermal relaxation. Minima and transition states were optimized a the AM1 level.

The linear motors pioneered by Stoddard and co-workers are based on rotaxanes[7]. Typically a rotaxane consists of a molecular ring that can shuttle reversibly between different binding sites along an extended axle molecule. The movement of the ring between these sites is controlled by external stimuli, such as light[7], pH[8] or redox chemistry[9].

None of the artificial motors developed to date can be efficiently incorporated into nanomachines. The most important limitations are (i) slow ground-state recovery dynamics, (ii) photo-chemical degradation in the case of alkenes[10], (iii) low quantum yields of the photochemical steps, or (iv) lack of functionalization possibilities. Furthermore, since the currently available motors require human interventions (*i.e.* heating, adding reactants, applying voltage) to complete their cycles, they cannot be operated autonomously. Thus, despite initial success in the development of molecular motors, major challenges remain.

A deeper mechanistic understanding of the chemical events that take place during the operation of both artificial and biological motors could help to improve the artificial motors. As the relevant time and spatial resolution are notoriously hard to access experimentally, computer simulations are the methods of choice to provide a better understanding of how these motors function at an atomic level.

2.2 Preliminary work (Eigene Vorarbeiten)

We have developed a simulation protocol for simulating photochemical reactions in biomolecular systems. In our scheme, a multi-configurational quantum mechanical description is used to model the electronic rearrangement for those parts of the system that are involved in the absorption, typically the chromophore. For the remainder, normally consisting of the apoprotein and the solvent, a simple forcefield model suffices. The interactions in the systems are thus computed within a hybrid quantum/classical framework. Forces are calculated on-thefly, and a diabatic surface hopping procedure is used to model the excited state decay.

We have demonstrated the validity of this hybrid QM/MM approach for photobiological reactions in recent applications on photoactivation of photoreceptor proteins[11,12], on photo-switching of fluorescent proteins[13,14] and on benign and malign photochemical reactions in DNA[15,16]. In addition to providing quantities that are experimentally accessible, such as structural intermediates, fluorescence lifetimes, quantum yields and spectra, the simulations have also provided information that is much more difficult to measure experimentally, such as reaction mechanisms and the influence of individual aminoacid residues.

We have recently gone one step further and use our protocol to engineer photoactive proteins. We first use QM/MM simulations to understand how these proteins have evolved to mediate the photochemical process. In the second step, we use the insights obtained from these simulations to design mutants with desired photochemical properties (*e.g.* high fluorescence quantum yield, different absorption maximum, alternative photoproducts). After testing the mutants by QMMM simulations, the most promising mutants will be selected and validated experimentally by our co-workers. The outcome of these experiments will be the most critical test for our simulation protocol thus far.

In the project we propose here, we want to keep going in the direction of molecular design, but now try to replace the protein altogether by synthetic substituents on the chromophore that mimic the interactions with the protein environment.

3.1 Objectives (Ziele)

The aim of this proposal is to design a new generation of molecular motors, strictly based on biological principles by means of atomistic computer simulations. To design a motor that overcomes limitations of the present molecular motors, we will derive our inspiration mainly from the biological processes of vision and proton pumping, as these are prime examples of converting light into directed motion. To reach our objectives we proceed in three steps:

1. use molecular dynamics simulations to get atomistic insight into the principles at the basis of these biological processes;

2. use these principles to design synthetic molecular motors;

3. perform simulations of the motors to test and improve their design.

After demonstrating that this design protocol works in selected proof-of-principle applications, we will focus on developing two types of molecular motors that in our opinion will be important components in the nanomachines of the future:

- rotating motors that generate mechanical force;

- proton conductors that transport protons.

3.2 Work schedule (Arbeitsprogramm)

3.2.1 Introduction

In this project we focus on light-driven molecular motors. Because these motors are powered and stimulated by photons, their action is easy to control from outside (*i.e.* the macroscopic world). Ignoring for the moment the ease to synthese, there are four criteria we can use to measure the succes of our designs: quantum yield, efficiency, directionality and autonomous operation. Creating a system that fulfils these criteria is possible if we can control the underlying molecular dynamics of the system by means of chemical substituents. In particular, these substituents have to

- catalyze the dynamics in the excited state (S_1), so that the energy of the absorbed photon is used to generate a power stroke. This can be achieved if the absorption induces a structural change in the molecule, such as an isomerization of a conjugated bond.

- catalyze the recovery dynamics in the ground state (S_0), so that subsequent photon absorption does not reverse the previous powerstroke, but generates a new powerstroke in the forward direction. This can be achieved if the recovery happens on a shorter timescale than the interval between two photon absorptions.

Although we are convinced that we can devise systems, which achieve efficient directed molecular motion with a high quantum yield in the computer simulations, the ultimate proof can only be obtained by experiment. Therefore, we are in close contact with synthetic chemists, whose experience concerning the practical feasibility of molecular systems will help us to improve the designs.

In what follows we first define our concept of a molecular motor. We will then show that the principles on which this concept is based are also found in

various photobiological systems. To transfer these principles from the biological systems to the artificial systems, we proceed in three steps:

- First, will use molecular dynamics simulations to reveal how the protein environment in these systems catalyzes the photochemical processes.

- Second, we will create model systems in which the protein environment is replaced and mimiced by simpler chemical substituents. Molecular dynamics simulations will be used to demonstrate that these simplified model systems undergo the same photochemistry as the protein systems.

- Third, after we have shown that the protein can be mimiced by artificial systems, we will focus on realizing different types of molecular motors by redesiging the artificial systems to operate according to the desired specifications.

After a concise description of the methodogical framework, which necessarly goes beyond the established simulation techniques, we conclude this section with a discussion of the risks and challenges we expect to encounter during the project and how these will be addressed.

3.2.2 Concept of a molecular motor

Our basic idea to create uni-directional motors is to combine the concepts of double bond isomerization in the excited-state, with proton transfer on the ground state. As shown in Fig. 2, the powerstroke is provided by photo-isomerization around a conjugated bond, while ground-state recovery is mediated by a proton transfer reaction. Using proton transfer to avoid unproductive re-photoisomerization, rather than thermal relaxation (Fig. 1), has two important advantages. First, proton transfer is a very fast process. This is not only because protons can tunnel through barriers on the potential energy surface, but also because proton transfer does not require large changes in the motor configuration. The second advantage is that, because the two photo-isomerization reactions take place in different charge states (Fig. 2), the direction of the rotation can be controlled by the local polarity. We will elaborate on this aspect later (3.2.4).

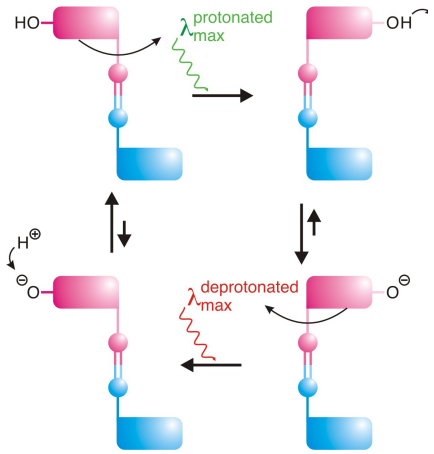

Figure 2: concept of a molecular motor. The two powerstrokes are stimulated at different wavelengths due to the pK_a difference between the two isomers.

Thus, to build a motor that operates according to these specifications, we need to design a molecule that (i) undergoes photo-isomerization around a specific chemical bond, and (ii) has an titrating site that shifts pK_a upon

isomerization. The principle of bond-specific photo-isomerization coupled to pK_a changes is found in many photobiological systems (Fig. 3). In bacteriorhodopsin, for example, photo-isomerization of a covalently bound retinal chromophore is used to translocate protons across the membrane[17]; In the photoactive yellow protein photo-isomerization of the *p*-coumaric chromophore induces a proton transfer,[11] and in so-called switchable fluorescent proteins, the protonation state of the chromophore changes upon photo-isomerization[13,18].

Understanding how the protein environment mediates the photochemical reactions in these systems is the crucial step towards the rational design of molecular motors. By means of computer simulations, in which the influence of the protein environment can be probed very precisely by selectively altering the interactions of specific aminoacids, it is possible to identify the interactions between the chromophore and the protein that catalyze the photochemical processes.

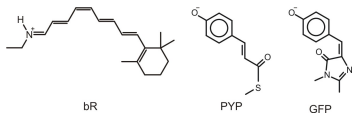

Figure 3: chromophores of the biological systems bacteriorhodopsin (bR), photoactive yellow protein (PYP), and green fluorescent protein (GFP).

3.2.3. Controlling photochemistry

To proof that it is possible to use organic substituents to replace the effect of the protein environment on the reactivity of a biological chromophore, we will design a simple system that mimics the photochemistry of the photoactive yellow protein (PYP). We estimate this phase to cover the first year of the project.

PYP is the primary photoreceptor for the photo-avoidance response of the salt-tolerant bacterium *Halorhodospira halophila*. PYP contains a deprotonated 4-hydroxy-cinnamic acid (or *p*-coumaric acid) chromophore linked covalently to the γ-sulphur atom of a cysteine residue via a thioester bond (Fig. 4). Upon absorption of a blue-light photon, PYP enters a fully reversible photocycle involving several intermediates on a timescale spanning from a few hundred femtoseconds to seconds.[19]

By using QM/MM molecular dynamics simulations, we have identified the key amino acids and the mechanism by which they control the primary events in the photocycle of PYP.[11] These are (i) double-bond photoisomerization (torsion *b*, Fig. 4), and (ii) the break of a hydrogen bond between the chromophore and the protein backbone. These events trigger a proton transfer from the protein to the chromophore, which ultimately leads to the signalling state of PYP.

The positive guanidinium moiety of Arg52 located just above the chromophore ring, was identified as the 'catalytic' residue that enforces double bond isomerization (torsion *b*, Fig. 4).[11,12] Replacing this arginine by a neutral residue results in photo-isomerization of a single bond (torsion *a*, Fig. 4))[12], which we recently found to be also the predominant decay channel in the

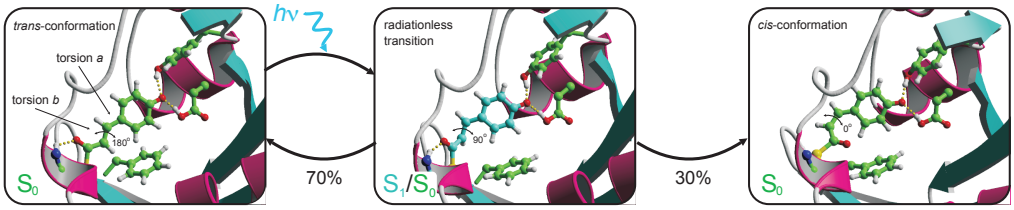

Figure 4: Snapshots from excited state trajectories of wild-type PYP, showing the chromophore in the active site pocket. The guanidium group of Arg52 that lies stacked on the chromophore ring is not shown. The first snapshot is at the excitation. The second shows the configuration at the radiationless transition from S_1 to S_0. The third snapshot shows the photoproduct, in which the carbonyl oxygen of the thioester linkage has flipped and is no longer hydrogen bonded to the backbone of Cys69.

isolated chromophore[20]. Thus, preferential electrostatic stabilization of the double bond twisted S_1 minimum by the positive Arg52 strongly favours double bond isomerization over single bond isomerization. Subsequent excited state decay from the double-bond twisted S_1 minimum is mediated by two essential hydrogen bonding interactions between the chromophore and the protein.[20]

We will create and test *p*-coumaric acid analogues with subsituents that mimic the positive charge on top of the ring and the two hydrogen bond donors to the phenolate oxygen (Fig. 5). Based on our previous findings, we predict that, like in the protein, the positive imidazole ring catalyzes a rotation around the double-bond in the excited state, while the two built-in alcohol groups provide the hydrogen bonds necessary for efficient radiationless decay from the S_1 minimum. Thus, we expect to find in our QM/MM simulations that the predominant excited-state decay channel will involve rotation of the double bond, identical to wild-type PYP. We will also test variations of this chromophore derivative. For instance, instead of imidazole we can introduce a guanidinium, or ammonium group. These groups are also positively charged and can stack onto the chromophore ring.

The second half of this proof-of-princinple project will be dedicated to control the proton transfer between the hydrogen bond donors and the chromophore. In PYP the double bond isomerization of the chromophopre increases the pK_a of the phenolate ring.[11] In the initial model, aimed at demonstrating that the photoisomerization can be controlled, the pK_a's of the alcohol moieties are probably too high to observe proton transfer to the phenol moiety. Therefore, we will also perform simulations on analogues with the alcohols replaced by more acidic hydrogen bond donors. Alternatively, a proton could come the solution. To estimate the pK_a shift, we will make use of thermodynamic integration, which is an established method to calculte the

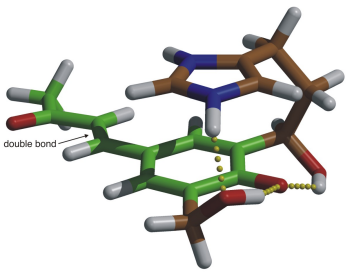

Figure 5: The intramolecular interactions in this *p*-coumaric acid derivative mimic the interactions of the chromophore and the protein environment. Therefore, we predict that the photoisomerization will take place around the double bond, as in PYP.

free energy difference between two thermodynamic states. In our case, the first state would be a protonoted hydrogen bond donor and a deprotonated chromophore, and the second state a deprotonated hydrogen bond donor and a protonated chromophore. It is also possible to include the proton dynamics explicitly into our simulations, using one of the protocols mentioned in methodology section (3.2.6).

A definite proof of concept can be obtained if our chromophore analogue is synthesised and measured experimentally. Such experimental data will provide important feedback not only on the quality of our design, but also on the accuracy of our simulation methodology, which is often difficult to assess in a biological applications due to their higher complexity. We expect that the outcome of our simulations will provide a strong incentive to perform the experiments and that we are ultimately able to compare to experimental data.

After we have demonstrated that we are able to imitate the effect of the protein environment on the chemistry of the chromophore, we will focus on achieving unidirectional motion in photochemical model systems.

3.2.4 Bio-inspired rotational motors

The next step, designing systems that convert light into rotational motion, will cover the second and third year of the project. In addition to the substituents required to achieve double bond photoisomerization and intramolecular proton transfer, we need to incorporate functional groups to achieve unidirectionality as well. The sources of inspiration are photoactive yellow protein and green fluorescent protein (GFP). In particular, we will let our design efforts be guided by the chemistry observed in switchable variants of this protein, (Padron[18], Dronpa[21], and asFP595[13,22]). The chromophore in these reversibly switchable fluorescent proteins (RSFP) can undergo photo-isomerize between a fluorescent and non-fluorescent isomer. Since the isomerization induces a pK_a shift of the chromophore, the two isomers exist in different protonation states.[13,18]

Initially, we will investigate the mechanism of selected RFSPs in atomic detail. By means of QM/MM simulations we will identify the interactions responsible for the isomerization and pK_a shift. Since the chromophore is larger than the one of PYP and we might have to include sidechains into the QM subsystem, part of our efforts will be dedicated to employing and validating more efficient quantum chemistry approaches.

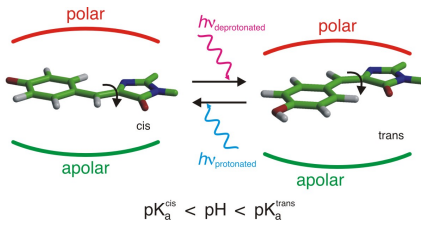

Figure 6: GFP-based motor. The pK_a difference between the two isomers in combination with a different polarity of the environment above and below the chromophore is used to achieve unidirectional rotation

Although the details of the design will depend on the outcome of the protein simulations, we anticipate that it will be similar to the concept shown in Fig. 6. Optical excitation of the cis conformation induces a photo-isomerization of the methine bridge, providing the first power stroke. After decay the system

relaxes into the trans conformation. Intra- or intermolecular interactions raise the pK_a of the trans isomer with respect to the pK_a of the cis conformation. Provided the pH of the solution is in between these two pK_a's, the chromophore takes up a proton from solution spontanously. Because protonation alters the absorption wavelength, unproductive back isomerization is avoided, and the first half of the 360° rotation cycle is completed.

Unidirectional rotation can be achieved by creating a local environment that has a different polarity below and above the plane of the chromophore. Due to the proton transfer, the phenol ring of cis and trans isomers exists in different charge states. The preferred rotation direction of this ring in the negatively charged cis chromophore will be along the more polar environment, whereas the rotation direction in the neutral trans chromophore will be along the more apolar environment. The major challenge in this part of the project is to design a suitable asymmetric environment that does not interfere with the pK_a shift, or open up different excited state decay channels. Since this is the problem we have to solve for the PYP-motors as well, we will work on both motortypes simultaneously. We expect to obtain hints on how to construct the environment from the QM/MM simulations of the protein systems.

3.2.5 bio-inspired proton pumps

The final part of project, which is expected to cover the remaining two years, is dedicated to devise a light-driven proton pump. Pumping protons against concentration gradients is an essential step in energy conversion in all kingdoms of life. Probably the best studied biological example of a light-driven proton pump is the membrane protein bacteriorhodopsin (bR). Conformational strain induced by photoisomerization of a covalently bound retinal is used to pump protons across the cell membrane. Despite structural studies of kinetically or cryogenically trapped intermediates and computational studies, the complete mechanism of proton pumping is still debated.

We therefore choose the GFP chromophore instead of retinal, as the basic element in our pumps. In analogy to bR and our previous designs, we will

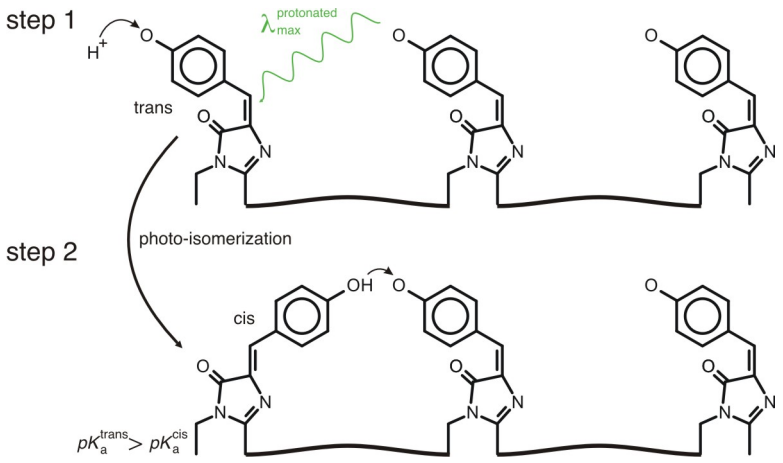

Figure 7: Principle of a GFP-based proton pump. Photon absorption should lead to proton transport from left to right.

exploit photo-isomerization in combination with a pK_a shift to achieve motion of protons in a unique direction. Our current idea is to place a series of photo-isomerizing chromophores next to eachother (Fig. 7). By incorporating such wire inside a hydrophobic pore (*e.g.* a nanotube), it should be possible to move protons over larger distances by means of light.

If the pH of the medium is above the pK_a of trans isomer, a proton is picked up by the phenolate ring exposed to the medium (left, Fig. 7). Photo-isomerization of the methine bridge brings the protonated phenol ring in contact with the phenolate of the second chromophore. Because the pK_a of the phenol ring is lower in the cis conformation than the trans conformation, the proton spontaneously hops from the first to the second chromophore. Photo-isomerization of the second chromophore brings the proton to the third chromophore, and so on until the proton reaches the medium on the other end of the wire. Although, the pump can transport protons from a medium with a lower pH to a medium with a higher pH, the pH difference between the media on the left of the wire and on the right can not be larger than the pK_a difference between the cis and trans conformation. Thus, to overcome large pH differences, we need to introduce interactions that increases this pK_a difference.

Protonation alters the absorption spectrum of the chromophore. Therefore, unidirectional proton transport could in principle be achieved by irrariation only at a wavelength that matches the absorption maximum of the protonated chromophore. However, this would not allow more than one operation cycle. To reset the units (*i.e.* bring them back to the deprotonated trans conformation), we need to simultaneously irradiate at wavelengths corresponding to the spectrum of the deprotonated chromophore. However, since both forward and backward photo-isomerizations will then occur, it is not clear if we can achieve a nett proton flux under these conditions.

To find out under what conditions our wires would function, we propose to perform large scale classical simulations at various illumination intensities. Because we need to include very many chromophores, QM/MM simulations are not affordable. Instead, we will build an excited-state forcefield based on our QM/MM simulations and higher level QM calculations. Since at this stage of the project we already have peformed extensive simulations of the GFP chromophore in a wide variety of systems, we should have enough data to fit a reliable forcefield. Photon absorption will be described as a stochastic process. At randomly selected timesteps, a randomly selected chromophore will be brought into the excited state. The interval between such timesteps will depend on the light intensity. In the simulations, we will use one of the approaches to model the proton dynamics, so that the proton transfers can be described realistically as well.

An alternative mechanism for light-driven proton transport is excited-state proton transfer (ESPT). Although we have studied this process in DNA base pairs,[15] we currently understand too little of this process to design a machines based on this princinple. Furthermore, an EPST process is much harder to simulate than a photoisomerization. First, both donor and acceptor(s) need to

be included in the QM subsystem. Second, the CASSCF method significantly overestimates barriers for proton transfer in the excited state, as it does not include sufficient dynamic electron correlation.[23] Going beyond CASSCF, however, is not possible due to the large QM subsystem size. Therefore, we leave this theme optional. Depending on how the main project evolves, we might have the opportunity to learn more about this process by looking into the ESPT process in GFP and its mutants. To perform the simulations on these systems, we would then combine enhanced sampling techniques, proton hopping and new approaches for electronic excited states.

3.2.6 Methodology

The size and complexity of a typical photobiological system, together with the timescales that must be reached necessitate the use of classical molecular dynamics for the nuclear degrees of freedom. In molecular dynamics simulations Newton's equations of motion are solved numerically to obtain a trajectory of the dynamics of a molecule over a period of time. To model the electronic rearrangement upon excitation a quantum mechanical description (QM) is required for those parts of the system that are involved in photon absorption. For the remainder, a simple molecular mechanics forcefield model suffices. The interactions in the systems are thus computed within a hybrid QM/MM framework.

Theoretical background

The central mechanistic feature in a photochemical reaction is the intersection seam between the potential energy surfaces of the excited and ground states (Fig. 8). Any point on the seam provides a funnel for efficient radiationless decay to the ground state. Just as a transition state separates reactants and products in ground state chemistry, the seam separates the excited state branch from the ground state branch in a photochemical reaction. The crucial difference, however, is that while a transition state connects a reactant to a *single* product via a *single* reaction path, the seam connects the excited state react to *several* products on the ground state via *several* paths. Just as ground state reactivity is enhanced by a stabilization of the transition state, photoreactivity can be enhanced too by stabilization of the seam. For example, in the photoactive yellow protein[11,12,20], electrostatic interaction between the chromophore and the active site pocket lowers the seam in energy, facilitating radiationless decay and enhancing the quantum yield of the photoisomerization process.

Excited state quantum chemistry

The established computational approach to photochemical problems is to model the electronic structure of isolated chromophores at the highest possible level of ab initio theory and characterize reactant, product and intersection geometries. Our approach is different in two main respects. First, in contrast to the conventional quantum chemistry approach, molecular dynamics trajectories at room temperature are computed rather than stationary points (essentially 0 K), which is not only more realistic, but also avoids the choice of an *a priori* reaction coordinate. Second, while most theoretical studies on photoreactivity still concentrate on the isolated chromophores, we include the protein environments explicitly in our simulations.

For reliable excited state calculations multi-configurational methods are required. In contrast to the more popular single-configuration techniques, such as time-dependent density functional theory (TDDFT) or equation of motion coupled cluster (EOM-CCSD), only multi-configurational methods provide wavefunctions that are sufficiently flexible to describe bond rearrangements, electronic state mixing and electronic reorganisations. In addition analytical energy gradients (forces) are necessary for computing molecular dynamics trajectories. Since the complete active space self-consistent field (CASSCF) method fulfills all of these requirements, it has often been used in the framework of excited state dynamics.

In CASSCF, a judicious set of occupied and virtual orbitals is chosen, the so-called active space orbitals[24]. In this active space, a full configuration interaction calculation is performed, while the other orbitals are being kept doubly occupied or empty in all configurations. The active orbitals are optimized such that the electronic energy of the state considered is minimal. The CASSCF method captures to a large extent so-called electron correlation. However, due to the necessary truncation of the active space, it does not recover dynamic electron correlation completely. Dynamic correlation is known to play a key role in the quantitative description of barrier heights and excitation energies. Dynamic correlation is accounted for in multi-reference perturbation theory approaches such as CASPT2[y]. However, since CASSCF usually provides a qualitatively correct describtion of the excited and ground state surface topologies, it has been widely used for mechanistic studies of photochemical reactions[x].

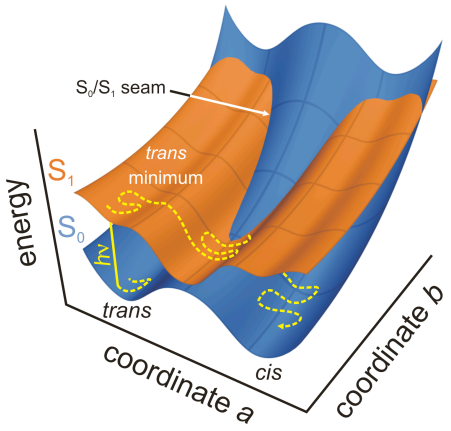

Figure 8: Photochemical reaction pathway (dashed line). After photon absorption, evolution takes place on the excited-state potential energy surface until the system hits the S_1/S_0 intersection seam. There a radiationless transition takes place to the ground-state. After the decay, the system continues evolving on the ground-state surface.

Excited state molecular dynamics

To model the dynamics of a photochemical reaction, the ground state and excited state potential energy surfaces must be described accurately. After light absorption, the reaction starts in the excited state (S_1), but ends in the ground state (S_0, Fig. 8). Therefore, it is essential to model the radiationless transition between the excited and ground state surfaces in a manner that is consistent with a quantum mechanical treatment of the complete system. Because we use Newton's equation of motion to compute molecular dynamics trajectories, the quantum mechanical character of the nuclei is ignored. As a consequence, population transfer from S_1 to S_0 cannot occur and the classical trajectory is restricted to a single potential energy surface. Thus, in contrast to a full quantum mechanical approach, radiationless transitions do not take place spontaneously. Instead, a binary decission to jump to a different electronic surface must be made at every timestep in a single trajectory. The criterion for switching between electronic states must result in a distribution of state populations, whose average can be compared to observable quantities quantities, such as quantum yield, lifetimes, etc.

In our simulations we allow hopping only at the intersection seam[x]. In principle, this strict diabatic hopping criterium could lead to an underestimation of the population transfer probability. However, because of the high dimensionality of the seam, most trajectories can usually encounter such regions of high propobability. The diababtic hopping model is clearly an approximation, but helps one to keep a proper physical insight, which is crucial in understanding complex systems.

Classical dynamics has the advantage that the potential energy surface can be computed on-the-fly. Forces are evaluated for the geometry at timestep t and used to compute the geometry at the next timestep $t+\Delta t$. Thus, only at configurations sampled by the classical trajectory electronic structure calculations are required. For high-dimensional system, for which computing potential energy surfaces beforehand is impossible, the on-the-fly strategy is the only way to perform a molecular dynamics simulation.

However, the on-the-fly calculation of CASSCF energies and gradients places severe constraints on the sophistication of the QM subsystem and we are forced to use minimal active spaces. Therefore, we need to calibrate our computations on the isolated QM subsystem against results obtained with higher levels of theory before performing the molecular dynamics simulations. Thus, CASPT2 results are used as benchmarks to validate reduced active space calculations that are sufficiently efficient on modern computer hardware.

Summarizing, we use classical molecular dynamics, with forces derived on-the-fly from a QM/MM hamiltonian. The QM subsystem consists of the chromophore and is described at the CASSCF level of theory, while the remainder is modelled with a forcefield. The active space used in the simulation is validated beforehand, by comparing the excited state properties of the isolated QM subsystem calculated at the CASSCF level to those computed at much higher levels of theory.

Method developments I: accelerating the simulation

The on-the-fly calculation of the multi-reference wavefunctions and gradients is the main computational bottleneck. Therefore part of our research effort will be dedicated to further develop the methodology with the goal of increasing the efficiency of the simulations.

First, the computation cost per MD step can be reduced by making use of more approximate levels of theory to describe the QM subsystem, such as semi-empirical and valence bond based methods, that are computationally much more efficient than CASSCF. These methods, however, need to be carefully calibrated against the results from higher level ab initio calculations before they can be used in the simulations. Prof. Walter Thiel and co-workers from the Max Plank Institut für Kohlenforschung have pioneered the use of semi-empirical methods for excited states, and therefore his department is an optimal place to carry out the research project.

Second, the number of MD steps required to simulate a photochemical reaction can be reduced by employing enhanced sampling techniques, such as chemical flooding[x], and metadynamics[x]. The trick of these methods is to locally perturb the potential energy surface during the simulation so that barriers are overcome more rapidly and the dynamics is accelerated. In chemical flooding, a Gaussian-shaped flooding potential is constructed from a principal component analysis of short excited-state QM/MM simulations (Fig. 9). By acting only on the internal degrees of the chromophore, the escape from local S_1 minima is accelerated in an unbiased manner. As the development and application of enhanced sampling methods are important research themes of Prof. Helmut Grubmüller, his department at the Max Planck Institut für Biophysikalische Chemie would also provide an excellent environment to perform the proposed reseach.

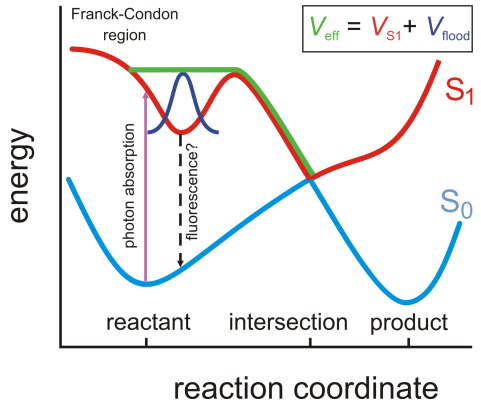

Figure 9: To overcome a high barrier on S_1 in a short QM/MM MD simulation, a flooding potential is applied (V_{flood},), applied diminishing the transition state barrier on S_1. From the strenght of the flooding potential, the time it would require to pass the barrier without flooding, can be estimated.

Method developments II: proton transfer

In addition to enhanced sampling techniques, we also require efficient simulation strategies for modeling proton dynamics. In many photo-biological processes, proton transfer plays an important role. Since photon absorption induces changes in the electron distribution, the initial response is often a proton transfer reaction, which can transiently stabilize the new charge distribution[x]. Because we want to use this principle in our artifical motors, we need to understand photobiological proton transfer in

atomic detail. However, simulating such processes is difficult, due to the inherent quantum mechanical character of the proton. Rather than moving classically, a proton 'hops' between donors to acceptors in a quantum mechanical fashion[x].

The model proton hops correctly in our MD simulations, we should use quantum mechanics for both electronic and nuclear degrees of freedom. Since for the systems we are interested in, this is far beyond the capabilities of the most modern computer hardware, we have to rely on semi-classical alternatives. One possibility is to use Q-hop, which combines classical molecular dynamics with a stochastic proton hopping.x The acceptance criterium for the hops is based on empirical hopping rates, parametrized on a small quantum mechanical donor-acceptor model system. We will not only implement this method in our MD package gromacs, but also improve its efficiency and accuracy.

Due to the stochastic hops, the underlying potential energy function is not continous, and energy is not conserved in Q-hop runs. In addition, because the hopping probabilities are derived from the potential energy surface of a small isolated model system, rather than the surface of the simulation system, detailed balance is not maintained. This has serious consequences for the thermodynamics.

To describe proton dynamics in a classical simulation that is both thermodynamically and kinetically correct, we propose to develop a model based on the λ-dynamics of Brooks and co-workers.x,y The proton transfer is described a fictious particle λ, with mass mλ, moving along a one dimensional coordinate that interpolates between the reactant and product states (Fig. 10). The λ-particle moves under the influence of the local free energy gradient, which, in contrast the free energy difference between the endstates, is a strictly local. Because the λ-particle samples the underlying free energy surface, the approach would not only conserve energy but also fulfil detailed balance. The transfer rate can be influenced by applying a symmetric biasing potential in the middle of the λ-coordinate. In order to obtain the correct kinetics, the height of this barrier will be parametrized against experimentally determined proton diffusion constants.

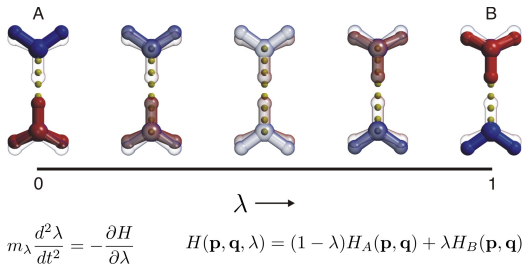

Figure 10: λ-hop proton transfer. The proton is not transferred, but a proton donor morphs into an acceptor, and vice versa, under the influence of the local thermodynamic gradient. This force is the gradient of the hamiltonian, linearly interpolated between the donor-acceptor and accpetor-donor situation.

A third alternative, if the above methods fail, is the multi-state empirical valence bond (MS-EVB) method by Schmidt and Voth.y The MS-EVB methods provides quantum chemical accuracy for the proton energy landscape at reasonable computational cost. Although he MS-EVB method is computationally far more expensive the semi-classical algorithms, it greatly

outperforms semi-empirical and ab initio approaches. Therefore, whenever the semi-classical approaches fail to describe the proton dynamics correctly, we will implement and use the MS-EVB method.

Finally, the new simulation approaches that we develop here will be made generally available in the Gromacs molecular dynamics package.

3.2.7 Risks and Challenges

Designing an artificial system capable of carrying out a task that has been specified beforehand is the ultimate goal of the nanosciences. Here we try to reach this goal by controlling the dynamics of the system by means of chemical substituents that mimic the catalytic power of photoactive proteins. We are aware that this project goes further than understanding and predicting the properties of chemical systems, which is already difficult to do by means of computation. Therefore, we might not be able to achieve all of our objectives within the time of the project.

The largest threat is that manipulating photochemistry is not as easy as our earlier work on PYP suggests. Although these results do not conflict with experimental data, the comparison to experiment is obscured by the complexity of the biogical systems. However, by performing the simulations of the less complex model systems we will learn about the basic photochemistry in these systems and the ways to control it. Thus, performing atomistic QM/MM simulations not only on photoactive proteins, but also on artifical systems designed to mimic the protein, could be a very promising approach to understand how the protein environment controls the photochemistry.

There is also the risk that the motors we design cannot be used in practice. Even if we can demonstrate that our designs achieve efficient unidirectional motion, the impact will depend on how easy these systems can be incorporated in molecular devices. To perform useful mechanical work, the rotating elements of our motors must be connected to other components of the nanomachine. For instance, to transmit the rotation of the GFP motor (Fig. 8), we would have to attach an axle to the phenolring. Thus, in addition to the substituents required for efficient rotation, we also need substituents that are functionalized to allow incorporation into more complex systems. In this project, we are in contact with synthetic chemists, and together we should be able to propose suitable substitents that functionalize our motor, but do not interfere with the rotation. The latter will be demonstrated by QM/MM simulations of these functionalized motors. It is even possible to include other components (*e.g.* axles, propellors) in these simulations to study their effect on the motor.

Closely related to the previous issue is photofatigueness. Side reactions, such as photocyclization[24] will compete with photoisomerization. Because sidereactions form a minor excited state decay channel, they will be difficult to identify in the simulations. However, in practice the molecular motors have to perform many operation cycles and the risk of sidereactions can become significant. Some of the sidereactions can be anticipated by chemical intuition.

Therefore, we could attempt to compute minimum energy paths that connect the Franck-Condon region with the ground-state product of a sidereaction (Fig. 7). This is an established approach in quantum chemistry. From the height of potential energy barriers encountered along these paths, we can estimate the likelyhood of the sidereaction.

Finally, one could argue that CASSCF with truncated active spaces might lack the accuracy required for our applications. We would like to stress again that we employ a hierarchical calibration strategy. All static properties are compared to higher levels of theory, *e.g.* CASPT2 or RASSCF. Not only do we validate the ordering of excited states, but we also reoptimize geometries at both levels. In addition, we characterize important configurations sampled in the dynamics with higher accuracy methods. Furthermore, as our previous results on PYP have been confirmed in several studies, both theoretical and experimental, we are confident that our approach of combining high level electronic structure and molecular dynamics is valid and reliable.

3.2.8 References

1. D. Stock, A. G.W. Leslie, J. E. Walker, *Science*, **286**, 700 (1999)
2. R.D. Vale, R.A. Milligan, *Science*, **288**, 88 (2000)
3. E.C. Friedberg, R. Wagner, M. Radman, *Science*, **296**, 1627 (2002)
4. N. Koumura *et al*., *Nature*, **401**, 152 (1999)
5. M.K.J. ter Wiel *et al*., *J. Am. Chem. Soc*., **125**, 15076 (2003)
6. S.P. Fletcher *et al*., *Science*, **310**, 80 (2005)
7. J.F. Stoddart, *Acc. Chem. Res*. **34**, 410 (2001)
8. A.M. Brouwer *et al*., *Science*, **291**, 2124 (2001)
9. J.N. Lowe *et al*., *Chem. Eur. J.* **10**, 1926 (2004)
10. R.A. Bissel *et al*., *Nature*, **369**, 133 (1994)
11. G. Groenhof *et al*., *J. Am. Chem. Soc.* **124**, 4228 (2004)
12. G. Groenhof *et al*., *J. Am. Chem. Soc.* **130**, 3250 (2008)
13. L.V. Schäfer *et al*., *Angew. Chem. Int. Ed.* **119**, 530 (2007)
14. L.V. Schäfer *et al*., *PLoS comp. biol*. **4**, e1000034 (2008)
15. G. Groenhof *et al*., *J. Am. Chem. Soc*. **129**, 6812 (2007)
16. M. Boggio-Pasqua *et al*., *J. Am. Chem. Soc*. **129**, 10996 (2007)
17. J. Tittor, D. Oesterhelt, *FEBS Lett*. **263**, 269 (1990)
18. T. Brakemann *et al*., Submitted to *J. Biol. Chem.* (2009)
19. K.J. Hellingwerf, J. Hendriks, T. Gensch, *J. Phys. Chem. A.* **107**, 1082 (2003)
20. M. Boggio-Pasqua, M.A. Robb, G. Groenhof, *J. Am. Chem. Soc*. **131**, 13580 (2009)
21. M. Andresen *et al*., *Proc. Natl. Acad. Sci.* **104**, 13005 (2007)
22. M. Andresen *et al*., *Proc. Natl. Acad. Sci.* **102**, 13070 (2005)
23. J.D. Coe, B.G. Levine, T.J. Martinez, *J. Phys. Chem. A.* **111**, 11302 (2007)
24. H. Meyer, *Angew. Chem. Int. Ed*. **31**, 1399 (1992)
25. B.O. Roos, P.R. Taylor, *Chem. Phys*. **48**, 157 (1980)
26. B.O. Roos, *Acc. Chem. Res.* **32**, 137 (1999)
27. E. Fabiano, G. Groenhof, W. Thiel, *Chem. Phys.* **351**, 111 (2008)
28. H. Grubmüller, *Phys. Rev. E*. **52**, 2893 (1995).
29. A. Laio, M. Parrinello, *Proc. Natl. Acad. Sci.* **99**, 12562 (2002)
30. A.B. Wöhri *et al*., submitted to *Science* (2009)
31. M.A. Lill, V. Helms, *J. Chem. Phys.* **115**, 7993 (2001)
32. J.L. Knight, C.L. Brooks III, *J. Comp. Chem.* **30**, 1692 (2009)
33. U.W. Schmitt, G.A. Voth, *J. Chem. Phys.* **111**, 9361 (1999)

4. Funds requested (Beantragte Mittel)

4.1 Staff costs (Personalkosten)

An overview of project is shown schematically in Fig. 11. For myslef I request a full-time TVöD position for the duration of the project In addition to supervising the project, I will implement and validate alternative QM schemes and the proton hopping algorithms. I will also be actively involved in analyzing the data together with my students.

For the the first Ph. D. student I request a 0.65 TVöD position for the first three years of the project. This Ph. D. student will carry out the simulations of the biological systems and build and test the PYP model systems. Later this Ph. D. student will also get involved in building motor systems based on GFP and PYP.

For the second Ph. D. student I also request a 0.65 TVöD position, for the last three years of the project. This Ph. D. student will develop the excited-state forcefields and build and test the proton pumps.

4.2 Scientific instrumentation (Wissenschaftliche Geräte)

The computations will be performed on the computational resources of the host institute. However, in order to contribute to these facilities, and equip the computers with the hardware that is most suitable for the project, I request **50.000 €** to purchase a small computer cluster.

4.3 Consumables (Verbrauchsmaterial)

To maintain collaborations with groups at different institutes and universities, I request 9,000 € per year for travel. The travel budget will also be used to cover the costs of workshops, summerschools and conferences. Publication costs I estimate at 2,000 € per year, especially since I wish to publish the results within the open access framework. Thus, in total, I request **11,000 €** per year for consumables.