Simulations and Experiments: How close can we get?

Martin Höfling

Munich 2011

Simulations and Experiments: How close can we get?

Martin Höfling

Dissertation
Faculty of Physics
Ludwig–Maximilians–University
of Munich

by
Martin Höfling
born in Frankfurt a.M.

Munich, December 12, 2011

First referee: Prof. Hermann E. Gaub
Second referee: Prof. Helmut Grubmüller
Date of the defence: December 8, 2011

Reprint permissions

- **Publication P1, pp 79**: Reprinted with the permission from John Wiley and Sons, License # 2735951293498.

- **Publication P1a, pp 83**: Reprinted with the permission from John Wiley and Sons, License # 2735951465854.

- **Publication P2, pp 87**: Reprinted with the permission from John Wiley and Sons, License #2735961136425.

- **Publication P3, pp 91**: Reprinted with the permission from Elsevier, License # 2735960840290.

- **Publication P4, pp 103**: Reprinted with permission from Journal of Chemical Theory and Computation. Copyright 2010 American Chemical Society.

- **Publication P5, pp 119**: Reprinted with the permission of John Wiley and Sons, License # 2735960547983.

- **Publication P6, pp 125**: Reprinted with the permission of John Wiley and Sons, License # 2735960074367.

- **Publication P7, pp 131**: Reprinted with the permission of Langmuir. Copyright 2010 American Chemical Society.

- **Publication P8, pp 137**: No Permission Required, Redistributed under the Creative Commons Attribution License (CCAL).

- **Publication P9, pp 157**: No Permission Required, Redistributed under the Creative Commons Attribution License (CCAL).

Contents

Contents

List of Abbreviations

AFM   atomic force microscopy 50, 58
AMBER   assisted model building with energy refinement 64
APL   area per lipid 64

BD   Brownian dynamics 4, 15, 48, 52, 54, 56, 57, 69, 70
BSE   bovine spongiform encephalopathy 58

CG   coarse grained 28, 29
CMT   collagen mineralized tissue 49
COM   center of mass 12, 16, 55, 69
CPMw   clique percolation method weighted 70
cryo-EM   cryo electron microscopy 26

DFT   density functional theory 48, 53, 57
DNA   deoxyribonucleic acid 5, 28, 42, 50
DOPC   1,2-Dioleoyl-sn-glycero-3-phosphocholine 63, 64

ECM   extracellular matrix 49, 62
EET   electronic energy transfer 17
EM   energy minimization 64

FRET   fluorescence resonance energy transfer 1–3, 7, 11, 14, 17–19, 21, 25–34, 37–41, 45, 195, 196

GaAs   gallium arsenide 50
GCMC   grand canonical Monte Carlo 14, 71, 72
GdHCl   guanidine hydrochloride 41
GolP   gold-protein force field 10, 53, 54
GROMACS Groningen machine for chemical simulations 5, 11, 13, 14, 53, 59, 71, 72

HLRB-II   Höchstleistungsrechner Bayern II 14
HTH   helix-turn-helix 41

IDA   ideal dipole approximation 44, 45

LD	Langevin dynamics 28, 29
LHC	light harvesting complex 17
LRZ	Leibnitz Rechenzentrum 13
MC	Monte Carlo 1–3, 11, 14, 25, 28, 34
MD	molecular dynamics 1–5, 9–14, 16, 29, 32, 33, 38, 48, 51–54, 56–58, 64, 69–71
mRNA	messenger ribonucleic acid 50
NAMD	"Not just Another Molecular Dynamics program" 53
NMR	nuclear magnetic resonance 26, 30, 33, 62, 63
OPLS/AA	optimized potentials for liquid simulations, all-atom 7, 12, 53, 64
PBC	periodic boundary conditions 12
PD	phage display 50, 51, 56
PME	particle mesh Ewald 9
PMF	potential of mean force 16, 48, 54, 67, 69, 70
RET	resonance energy transfer 2, 17–19, 26, 27, 45
RMSD	root mean square deviation 14
RNA	ribonucleic acid 5, 42, 68
RNase	ribonuclease 67, 68
SAS	solvent accessible surface 14, 63
SMD	steered molecular dynamics 65
smFRET	single molecule fluorescence resonance energy transfer 2, 23, 27, 28, 30–32, 34–36, 41
SPC	simple point charge 12
SPR	surface plasmon resonance 50–52, 57
STED	stimulated emission depletion 26
SWIG	simplified wrapper and interface generator 72
TDC	transition density cube 43–45
TEM	transmission electron microscopy 26
TF	transfer function 25, 38, 39, 41, 42
THP	thermal hysteresis protein 49

TIP4P	transferable intermolecular potential with 4 points 12
TM	transmembrane 2, 59, 62–65
UHBD	University of Houston Brownian dynamics 48, 56
VCAM	vascular cell adhesion molecule 62
VMD	visual molecular dynamics 14

List of Figures

List of Tables

Summary

The interactions between biomolecules and their environment can be studied by experiments and simulations. Results from experiments and simulations are often interpretations based on the raw data. For an accurate comparison of both approaches, the interpretation of the raw data from experiments and simulation have to be in compliance.

The design of such simulations and interpretation of raw data is demonstrated in this thesis for two examples; fluorescence resonance energy transfer (FRET) experiments and surface adsorption of biomolecules on inorganic surfaces like gold.

FRET experiments allow to probe molecular distances via the distance-dependent energy transfer efficiency from an excited donor dye to its acceptor counterpart. In single molecule settings, not only average distances, but also distance distributions or even fluctuations can be probed, providing a powerful tool to study flexibilities and structural changes in biomolecules.

However, the measured energy transfer efficiency does not only depend on the distance between the two dyes, but also on their mutual orientation, which is typically inaccessible to experiments. Thus, assumptions on the orientation distributions and averages have to be employed, which severely limit the accuracy of the distance distributions extracted from FRET experiments alone.

In this work, I combined efficiency distributions from FRET experiments with dye orientation statistics from molecular dynamics (MD) simulations to calculate improved estimates of the distance distributions. From the time-dependent mutual dye orientations, the FRET efficiency was calculated and the statistics of individual photo-absorption, FRET, and photo-emission events were determined from subsequent Monte Carlo (MC) simulations. All recorded emission events were then collected to bursts from which efficiencies were calculated in close resemblance to the actual FRET experiment. The feasibility of this approach has been tested by direct comparison to experimental data.

As my test system, I chose a poly-proline chain with Alexa 488 and Alexa 594 dyes attached. Quantitative agreement of calculated efficiency distributions from simulations with the experimental ones was obtained. In addition, the presence of cis-isomers and specific dye conformations were identified as the sources of the experimentally observed heterogeneity.

This agreement of *in silico* FRET with experiments allows employment of the dye orientation dynamics from simulations in the distance reconstruction. For multiple levels of approximation, the dye orientation dynamics was used in dye orientation models. At each level, fewer assumptions were applied to the dye orientation model.

Each model was then used to reconstruct distance distributions from experimental efficiency distributions.

Comparison of reconstructed distance distributions with those from simulations revealed a systematically improved accuracy of the reconstruction in conjunction with a reduction of model assumptions.

This result demonstrates that dye orientations from MD simulations, combined with MC photon generation, can indeed be used to improve the accuracy of distance distribution reconstruction from experimental FRET efficiencies.

A second example of simulations and interpretation in compliance with experiments are the studies of protein adsorption on gold surfaces. Interactions between biomolecules and inorganic surfaces, e.g. during the biomineralization of bone, are fundamental for multicellular organisms. Moreover, understanding these interactions is the basis for biotechnological applications such as biochips or nano-sensing.

In the framework of the PROSURF project, a multi-scale approach for the simulation of biomolecular adsorption was implemented. First, parameters for MD simulations were derived from *ab initio* calculations. These parameters were then applied to simulate the adsorption of single amino acids and to calculate their adsorption free energy profiles. For the screening of adsorbed protein conformations, rigid body Brownian dynamics (BD) docking on surfaces was benchmarked with the free energy profiles from the MD simulations.

Comparison of the protein adsorption rate from surface plasmon resonance experiments and BD docking yielded good agreement and therefore justifies the multi-scale approach. Additionally, MD simulations of protein adsorption on gold surfaces revealed an unexpected importance of positively charged residues on the surface for the initial adsorption steps.

The multi-scale approach presented here allows the study of biomolecular interactions with inorganic surfaces consistently at multiple levels of theory: Atomistic details of the adsorption process can be studied by MD simulations whereas BD allows the extensive screening of protein libraries or adsorption geometries.

In summary, compliance of simulation and experimental setup allows benchmarking of the simulation accuracy by comparison to experiments. In contrast to employing experiments alone, the combination of experiments and simulations enhances the accuracy of interpreted results from experimental raw data.

Zusammenfassung

Wechselwirkungen zwischen Biomolekülen und ihrer Umgebung können sowohl in Experimenten als auch in Simulationen untersucht werden. Die Ergebnisse von beiden Methoden beruhen häufig auf der Auswertung von Rohdaten. Um die beiden Ansätze genau miteinander vergleichen zu können, muss die Auswertung der Rohdaten aus Experiment und Simulation übereinstimmen.

In dieser Arbeit wird der Entwurf von Simulationen sowie die gemeinsame Interpretation von deren Rohdaten zusammen mit dem Experiment demonstriert. Gezeigt wird dies zum einen am Beispiel von Fluoreszenz-Resonanzenergietransfer (FRET), zum anderen an der Adsorption von Biomolekülen auf anorganischen Oberflächen wie Gold.

FRET Experimente bestimmen molekulare Abstände durch die Abstandsabhängigkeit der Energietransfereffizienz von einem angeregten Donorfarbstoff zum Akzeptorfarbstoff. Neben den mittleren Abständen können in Einzelmolekülexperimenten auch Abstandsverteilungen und sogar Fluktuationen des Abstandes bestimmt werden. Dies ist eine wichtige Voraussetzung für die Untersuchung von Flexibilität sowie Strukturänderungen in Biomolekülen.

Die gemessene Effizienz des Energietransfers hängt nicht nur vom Abstand, sondern auch von der relativen Orientierung der Farbstoffe ab. Da diese normalerweise nicht im Experiment bestimmt werden kann, werden Modellannahmen über die Orientierungsverteilung der Farbstoffe gemacht und Mittelwerte verwendet. Diese Annahmen begrenzen jedoch die Genauigkeit der rekonstruierten Abstände.

Um genauere Abstandsverteilungen zu erzielen, wurden in dieser Arbeit Effizienzverteilungen aus FRET Experimenten mit Statistik und Dynamik der Farbstofforientierung aus Molekulardynamik (MD) Simulationen kombiniert. Dabei wurde zuerst aus der zeitabhängigen Farbstofforientierung die zugehörige FRET Effizienz berechnet. Diese sind die Grundlage für Monte Carlo (MC) Simulationen mit dem Ziel die Kinetik von Photo-Absorption, FRET und Photo-Emission zu modellieren. In Anlehnung an Einzelmolekül-FRET-Experimente wurden anschließend Photonenbündel erzeugt, aus denen dann die jeweilige Effizienz bestimmt wurde. Durch direkten Vergleich zu experimentellen Daten wurde die Genauigkeit dieses Ansatzes überprüft.

Als Modellsystem dienten Poly-Prolinketten mit gebundenen Alexa 488 und 594 Farbstoffen. Dabei wurde eine quantitative Übereinstimmung mit dem Experiment beobachtet. Darüber hinaus wurden Cis-Isomere sowie spezielle Farbstoffkonformationen als Ursache für die experimentell beobachtete Heterogenität identifiziert.

Die Übereinstimmung von *in silico* FRET mit Experimenten ermöglichte die Abstandsrekonstruktion mit Hilfe der simulierten Farbstofforientierung und Dynamik.

Auf Grundlage der Farbstoffdynamik wurden Farbstofforientierungsmodellen erstellt. Eine Reduktion der Modellannahmen erfolgte dabei in mehreren Schritten. Anschließend wurde an jedem dieser Modelle die Rekonstruktion von Abstandsverteilungen aus den experimentell bestimmten Effizienzverteilungen getestet.

Beim Vergleich von den rekonstruierten Abständen konnte eine systematische Verbesserung der Genauigkeit bei gleichzeitiger Reduktion von Modellannahmen beobachtet werden.

Dieses Ergebnis zeigt, dass eine höhere Genauigkeit der rekonstruierten Abstände erzielt werden kann, wenn experimentelle FRET Effizienzen mit der Farbstofforientierung aus MD Simulationen kombiniert werden.

Als zweites Beispiel wurde die Adsorption von Proteinen auf Goldoberflächen untersucht. Für multizelluläre Organismen sind Wechselwirkungen zwischen Biomolekülen und anorganischen Oberflächen von hoher Wichtigkeit, wie etwa bei der Biomineralisierung von Knochen. Des Weiteren ist das Verständnis dieser Wechselwirkungen grundlegend für biotechnologische Anwendungen wie in Biochips oder beim Nano-Sensing.

Im Rahmen des PROSURF Projekts wurden daher Ansätze für Simulationen von biomolekularer Adsorption auf mehreren Skalen entwickelt. Zu Beginn wurden mit *ab initio* Berechnungen Kraftfeldparameter für MD Simulationen bestimmt. Mit diesen wurde anschließend die Adsorption von einzelnen Aminosäuren simuliert, sowie deren freie Energielandschaft berechnet. Mit den Energielandschaften wurde Brownsche Dynamik (BD) mit starren Körpern getestet, mit welcher im Anschluss adsorbierte Proteinkonformationen erzeugt wurden.

Es konnte gezeigt werden, dass die Adsorptionsraten von Oberflächenplasmonenresonanzspektroskopie Experimenten mit denen von BD Simulationen übereinstimmen. Dies zeigt die Genauigkeit von verschiedenen, aufeinander aufbauenden Simulationsskalen. In MD Simulationen von Proteinadsorption auf Goldoberflächen wurde eine überraschende Schlüsselrolle von positiv geladenen Seitenketten für die anfängliche Adsorption beobachtet.

Der hier gezeigte Ansatz ermöglicht Simulationen von biomolekularen Wechselwirkungen mit anorganischen Oberflächen auf verschiedenen Skalen. So können mit Hilfe von MD Simulationen atomare Details der Adsorption untersucht werden. Mit Hilfe von BD Simulationen wird das Auswählen aus Proteinbibliotheken und von Adsorptionsgeometrien ermöglicht.

Zusammenfassend kann eine gemeinsame Rohdatenauswertung von Experiment und Simulation zum Testen der Genauigkeit von Simulationen herangezogen werden. Darüber hinaus ermöglicht die Kombination mit Simulationen auch die Auswertung von experimentellen Rohdaten mit verbesserter Genauigkeit.

1

Overview

In many cases, results from experiments and simulations show only qualitative agreement or are even in contradiction [12]. Prominent examples are the simulations of water or lipid bilayers, where the experimentally determined properties are not accurately reproduced in simulations without additional surface tension [13]. For complex biomolecular systems, it is often challenging to mimic the experiment and only indirect comparisons between experiments and simulations are made.

The reasons for such discrepancies are manifold. The interpretation of experimental raw data can be wrong, e.g. when crystal structures are derived from the measured electron densities. For the simulations, the model may not accurately describe the system and therefore lead to different results than in the experiments. But even if experiment and simulation are accurate and correct on their own, discrepancies can arise from the differences in the setups and different interpretation of the raw data.

Therefore, one of the guidelines in this thesis is to interpret raw data from simulations in compliance with experiments. The foundation for this is accurate modeling of the experimental conditions. Simulation of interactions between biomolecules often requires an atomistic description, and therefore atomistic molecular dynamics (MD) simulations are employed in my work. Further, additional techniques are required to accurately model the processes in the experiment, since not all interactions are modeled in MD simulations.

Examples in my work are the combination of MD with Monte Carlo (MC) techniques to simulate fluorescence resonance energy transfer (FRET) or *ab initio* parameterization of dyes and surface interactions. During the design of each simulation system, tight cooperation with experimental scientists is required to reflect the experimental conditions in the theoretical model. Finally, direct comparison of raw data from experiment and simulation will serve as benchmark for the accuracy of simulations.

Instead of simulating a particular biomolecular process, I developed simulations reflecting the entire experimental setup. As I will show, the combination of such

simulations with experiments enables higher accuracy of measurements and a better throughput in contrast to stand alone experiments or simulations. This principle will be demonstrated on two major examples. First, I performed *in silico* single molecule fluorescence resonance energy transfer (smFRET) experiments by combining MD simulations with MC techniques (Chapter 3). In addition, the combination of FRET experiments and their *in silico* counterpart resulted in a better accuracy of FRET distance measurements. Second, the assembly of a computational toolbox for protein-gold interactions and the validation by experiments will serve as a prototype for characterization of protein interactions with arbitrary inorganic surfaces *in silico* (Chapter 4). This *in silico* characterization yields a higher protein screening throughput by reducing the number of proteins that have to be tested experimentally.

Both projects constitute the main part of this thesis and will be introduced in more detail in Chapter 3 and 4. Three further small projects are summarized in Chapter 5-7. In Chapter 5, the first steps towards simulations of the transmembrane (TM) signaling of integrins are described whereas protein complex formation is studied computationally in Chapter 6. Finally, the technical details of my GromPy analysis library and their usage in other projects are summarized in Chapter 7.

1.1 Distance Measurement in Molecular Biology via FRET

Distance measurements on the nanometer scale are fundamental to understand functional processes in biomolecular systems. Interaction, relative orientation and conformational changes of biomolecules are revealed by distance measurements. Preferably, the measurement should be performed in conditions similar to the *in vivo* process and the measurement itself should not interfere with the analyzed process.

FRET is an excellent method to measure distances even *in vivo*. In FRET experiments, two dyes are covalently attached to the sites between which the distance should be measured. After excitation of one dye, the efficiency of the resonance energy transfer (RET) is measured and the distance dependency of the efficiency is exploited to determine distances.

Unfortunately, the RET does not only depend on the distance but — similar to antennas — also on the mutual orientation of the two dyes. Moreover, the dyes move during the measurement process and therefore their mutual orientation changes during time. The instantaneous dye orientation during RET cannot be measured in experiments.

Instead of the inaccessible mutual dye orientation, assumptions are commonly applied in experiments. Assuming an isotropic dye orientation results in the well known mean orientation factor $\kappa^2 = 2/3$. To which extend this and further assumptions limit the accuracy of distance measurements in FRET experiments is difficult to assess [14].

In contrast to experiments, the instantaneous dye orientation is accessible by MD simulations. Therefore, my major goal in Chapter 3 is to replace the assumptions of the dye orientation made in experiments by orientational information calculated from MD simulations. Thereby, I will demonstrate that a systematic reduction of model assumptions improves the accuracy of experimental distance reconstruction.

As a first validation step, I verified that the simulations performed are an accurate model of the experimental setup. To achieve this, I developed an *in silico* FRET protocol based on MD simulations and MC photon statistics. This method was successfully tested on a poly-proline model system.

After successful validation, the dye orientation statistics from simulations was employed to calculate distance distributions from the experimental efficiency distributions. A systematic improvement in distance accuracy was found upon reduction of the dye orientation model assumptions.

In summary, the virtue of this approach is to use the results of MD simulations for the analysis of experimental data — in contrast to only compare them — to improve the accuracy of distances measurement in FRET experiments.

1.2 Biomolecular Adsorption on Inorganic Surfaces

The second topic are interactions of biomolecules with inorganic surfaces. Studying these interactions is not only important for biotechnological applications such as biosensing [15, 16] but also for the design of biocompatible implants and to assess toxicity of nano crystals [17, 18]. Moreover, biomolecules interact with surfaces during the mineralization of dentin and bones *in vivo*. In all cases, the understanding of the biomolecular interactions with inorganic surfaces is key to specifically design or avoid adsorption [19].

Recently, a variety of peptides with a high affinity for gold surfaces were found by screening with combinatorial experiments [20–22]. However, these experiments require extensive library scans. Specialized interactions, e.g. for enzyme adsorption not interfering with the enzymatic activity, are difficult to create [21]. Furthermore, the exact adsorption mechanisms are not understood by combinatorial experiments. For these reasons, the specific design of surface interactions is currently challenging.

For the biotechnological screening of protein-surface interactions, a reduction of the required library scan experiments is desirable. Therefore, *in silico* docking of proteins to surfaces is expected to aid the design of protein-surface interactions similar to *in silico* docking of compounds to proteins which revolutionized drug discovery during the last decade [23–25].

The principle interactions with surfaces (electrostatics, van-der-Waals) are known. However, they often compensate each other such that the overall adsorption behavior is difficult to predict in purely qualitative terms. In addition, the interactions of the biomolecule with the surface compete with desolvation of the biomolecule and the surface. Force fields in MD simulations accurately model the complex interactions between biomolecules in solution. Thus, MD simulations of biomolecules and surfaces in explicit solvent are promising in the study of the adsorption process in atomistic detail. To enable adsorption simulations, additional force field parameters of biomolecules with the inorganic surface had to be developed and tested in a first step (Chapter 4) since the interactions are not described by standard force fields.

In Chapter 4, computational protein-surface docking is demonstrated for gold surfaces. In the development process, multiple levels of theory are required and my MD simulations are pivotal for connecting the highest level of theory, *ab inito* calculations of amino acid fragments [26–28], with rigid body Brownian dynamics (BD) docking of proteins to the surface.

The *in silico* docking of proteins was validated experimentally by comparison of calculated and experimental adsorption rates [29] such that the first step towards efficient computational screening of protein-surface interactions is achieved. In this second topic, BD simulations close to experiments are employed to verify the multi-scale approach at the end.

2

Theoretical Background and Methods

2.1 Approximations in Molecular Dynamics Simulation

The time-dependent Schrödinger equation of nuclei and electrons describes the atomic motions of a biological system. The solution of this equation computationally faces two main problems, the size of biological systems and the typical time scales of biological processes.

At high levels of detail, taking electronic degrees of freedom explicitly into account, the time-dependent Schrödinger equation can only be solved in the attosecond range, whereas typical biological processes on the molecular level occur in the range of picoseconds to hours. In addition, this solution is only possible for small systems consisting of few atoms (≈ 10).

Even smallest proteins in explicit solvent easily reach system sizes in the order of 10,000 atoms, e.g. the small protein complex without solvent in Publication P7 already consists of $\approx 3,200$ atoms. Therefore, several approximations will be outlined in this section, which will allow a sufficient description of systems, consisting of more than a few atoms. They will lead to the electron interaction being taken into account by a force-field in which the nuclei move according to classical equations of motion.

These approximations allow MD simulations of systems consisting of proteins, solvent, membranes, deoxyribonucleic acid (DNA), ribonucleic acid (RNA) and inorganic surfaces. Simulation of large systems consisting of millions of atoms [30] and simulations up to milliseconds are possible with specialized hardware [31]. Detailed theoretical background, algorithms and MD simulation protocols are found in textbooks [32–36] and in the Groningen machine for chemical simulations (GROMACS) simulation software manual [37].

2.1.1 Born Oppenheimer Approximation

Aim: *Separation of nuclear and electronic degrees of freedom*

As stated before, the goal is to (approximatively) solve the time-dependent Schrödinger equation

$$\mathcal{H}\Psi(\boldsymbol{R},\boldsymbol{r},t) = i\hbar\frac{\partial\Psi(\boldsymbol{R},\boldsymbol{r},t)}{\partial t}, \tag{2.1}$$

which describes the time evolution of a system consisting of nuclei and electrons. Here, $\mathcal{H}$ denotes the Hamilton-operator, Ψ the time dependent wave-function and $\hbar$ the reduced Planck constant. The wave-function depends on the position of all electrons $\boldsymbol{r}$ and nuclei $\boldsymbol{R}$ at time t.

Since the electron mass is much smaller than a nuclei mass, the electron configuration can be assumed to adapt instantaneously to slow changes of the nuclei positions. This adiabatic approximation is known as "Born Oppenheimer approximation" [38]. As a consequence, the molecular wave-function Ψ is expressed by a product

$$\Psi(\boldsymbol{R},\boldsymbol{r},t) = \Psi_{\mathrm{nuc}}(\boldsymbol{R},t)\cdot\Psi_{\mathrm{el}}(\boldsymbol{r};\boldsymbol{R}) \tag{2.2}$$

of nucleic and electronic wave-functions Ψ_{nuc} and Ψ_{el}.

The electronic wave-function depends on the nuclei coordinates $\boldsymbol{R}$ only parametrically and therefore, applying the Hamilton-Operator $\mathcal{H}$ results in a separate expression for the electronic part

$$\mathcal{H}_{\mathrm{el}}(\boldsymbol{R})\Psi_{\mathrm{el}}(\boldsymbol{r};\boldsymbol{R}) = E_{\mathrm{el}}(\boldsymbol{R})\Psi_{\mathrm{el}}(\boldsymbol{r};\boldsymbol{R}) \tag{2.3}$$

in any given nuclei configuration $\boldsymbol{R}$. Here the electronic Hamiltonian $\mathcal{H}_{\mathrm{el}}$ is defined as $\mathcal{H}_{\mathrm{el}} = \mathcal{H} - T_{\mathrm{nuc}}$, the complete Hamiltonian $\mathcal{H}$ without the kinetic energy operator T_{nuc} of the nuclei (see Ref. 35, p. 21).

The system of interest is commonly in the electronic ground state. Therefore, the smallest eigenvalue $E_{\mathrm{el}}^0(\boldsymbol{R})$ of Eq. 2.3 is used in the remaining Schrödinger equation for the nuclei dynamics:

$$(E_{\mathrm{el}}^0(\boldsymbol{R}) + T_{\mathrm{nuc}})\Psi_{\mathrm{nuc}}(\boldsymbol{R},t) = i\hbar\frac{\partial\Psi_{\mathrm{nuc}}(\boldsymbol{R},t)}{\partial t}. \tag{2.4}$$

Alternatively, $\mathcal{H}_{\mathrm{el}}$ can be defined without containing the Coulomb interaction between the nuclei (e.g. as in Ref. 32, p. 107) which then leads to an additional term $\frac{q_i q_j}{r_{ij}}$ for each pair of nuclei ij.

2.1.2 Approximation of the Electronic Potential by a Force Field

Aim: *Avoid expensive solving of the electronic Schrödinger Eq. 2.3 for each set of nuclei coordinates R by approximation through an empirical force field.*

The separation into electronic and nuclear parts of the Schrödinger equation still requires separated solving of the time-independent electronic part (Eq. 2.3) for a given nuclei configuration. Solving this equation computationally is not feasible for systems consisting of more than a few atoms. Thus, the interactions are described by a potential function

$$V = \underbrace{\sum_{\text{bond } i} V_{\text{B}}^{i} + \sum_{\text{bond angle } j} V_{\alpha}^{j} + \sum_{\text{dihedral } k} V_{\text{dih}}^{k} + \sum_{\text{imp.dih. } l} V_{\text{imp}}^{l}}_{\text{bonded interactions}} + \underbrace{\sum_{\text{pairs } m,n} (V_{\text{vdW}}^{mn} + V_{\text{coul}}^{mn})}_{\text{non-bonded interactions}} \quad (2.5)$$

whose derivative is a conservative force field which only depends on the nuclei positions.

As shown by Eq. 2.5, typical force fields for bio-molecular applications consist of bond, bond-angle, dihedral and improper dihedral potentials describing the covalent bonds (see Tab. 2.1) between atoms as well as "non-bonded" potentials (Tab. 2.2) for dispersion, Pauli-repulsion and Coulomb interactions. The non-bonded interaction potentials are pair additive.

During most chemical reactions, molecules are in the ground state, which is therefore commonly selected as the electronic state for which force-field parameters are derived. In Publication P8, the donor-dye is in the excited state before FRET occurs, thus the first excited state was used for parameterization here (see Section "Force Field", page 139, Publication P8).

Force fields are parameterized using quantum chemical calculations [39, 40] and/or by calibrating the free parameters to reproduce experimentally known thermodynamic properties in simulations [41].

In my studies, the optimized potentials for liquid simulations, all-atom (OPLS/AA) force field [42, 43], designed for proteins, was used. In Publications P5, P7 and P9, additional parameters [26–28] modeling the interaction between biomolecules and gold surfaces were employed. The dyes used in Publication P8 are not covered by the standard force field parameters of OPLS/AA and therefore required calculation of additional parameters.

Table 2.1: Bonded Interactions Between Atoms

Bond potentials

Harmonic: $V_{\mathrm{B_{harm}}}(\boldsymbol{r}) = \frac{1}{2}k_{\mathrm{B}}\left(\boldsymbol{r} - \boldsymbol{r_0}\right)^2$

standard potential for covalent bond-stretching.

Morse: $V_{\mathrm{B_{morse}}}(\boldsymbol{r}) = k_{\mathrm{B}}\left[1 - \exp(\beta(\boldsymbol{r} - \boldsymbol{r_0}))\right]^2$

allows anharmonic bond-stretching and breaking of covalent bonds.
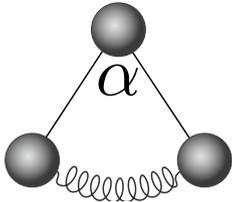
Angle potential

Harmonic: $V_{\mathrm{B_{harm}}}(\alpha) = \frac{1}{2}k_{\alpha}\left(\alpha - \alpha_0\right)^2$

standard angle potential, e.g. modeling tetrahedral bond angles from sp^3 hybridization.
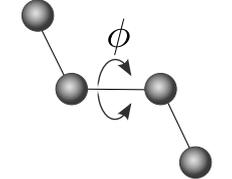
Dihedral potentials

Periodic: $V_{\mathrm{dih}}(\phi) = k_{\phi}\left[1 - \cos\left(n \cdot \phi - \phi_0\right)\right]$

Ryckaert-Bellemans: $V_{\mathrm{RBdih}}(\phi) = \sum\limits_{n=0}^{5} C_n(\cos(\phi - \pi))^n$

cosine power series potential.

dihedral potentials for sterical restriction, e.g. due to Pauli-repulsion of orbitals from first and fourth atom.
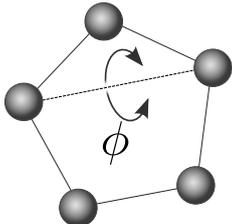
Improper dihedral potential

Harmonic: $V_{\mathrm{imp}}(\phi) = \frac{1}{2}k_{\phi}\left(\phi - \phi_0\right)^2$

Improper dihedral potentials atoms restrain atoms to a plane, e.g. in conjugated planar ring systems with delocalized $\pi-$electron systems.

Table 2.2: Non-bonded Interactions Between Atoms

	Combined repulsion and dispersion potentials
	Lennard Jones: $\quad V_{\mathrm{LJ}}(\boldsymbol{r}) = \frac{C_{(12)}}{\boldsymbol{r}^{12}} - \frac{C_{(6)}}{\boldsymbol{r}^{6}}$
	standard potential for repulsion and dispersion
	Buckingham: $\quad V_{\mathrm{B}}(\boldsymbol{r}) = A \cdot \exp(-B \cdot |\boldsymbol{r}|) - \frac{C}{\boldsymbol{r}^{6}}$
	more realistic repulsion (3 free parameters), computationally expensive due to expansion of the exponential function.
	repulsion and dispersion usually cut-off at distances of $\approx 1\,\mathrm{nm}$
	Coulomb potential
	Standard Coulomb: $\quad V_{\mathrm{c}}(\boldsymbol{r}) = \frac{1}{4\pi\epsilon_0} \frac{q_i \cdot q_j}{\epsilon_{\mathrm{r}} \cdot |\boldsymbol{r}|}$
	Commonly, the standard Coulomb interaction is cut-off at distances of $\approx 1\,\mathrm{nm}$ for computational efficiency. This however can cause severe artifacts [44] in simulations. The method I used in my simulations to accurately treat long-range electrostatic interactions is particle mesh Ewald (PME) [45].

2.1.3 The Nuclei are Treated as Classical Particles

Aim: *Avoid expensive solving of the nuclei Schrödinger Eq. 2.4 by classical description of nuclei dynamics.*

In the last approximation, is is assumed that the nuclei dynamics (Eq. 2.4) are well described by classical particle dynamics

$$m_i \frac{d^2 \boldsymbol{r}_i(t)}{dt^2} = -\nabla_i V(\boldsymbol{r}_1, \dots, \boldsymbol{r}_n) = \boldsymbol{F}_i, \qquad (2.6)$$

where m_i are the masses and $\boldsymbol{r}_i$ the coordinates of the i-th particle (nucleus).

Several justifications are made such as Ehrenfest's theorem [46] or the parametric dependency of Eq. 2.3 (Ref. 35, p. 22-25), however they are not a rigorous derivation of this approximation. MD was applied successfully for many systems in the past so that the classical approximation is considered to be valid for bio-molecules within a certain range (for limitations see the following Section).

2.1.4 Known Limitations

In the next paragraph, I will discuss some limitations of these approximations which might affect results obtained from MD simulations carried out in this work.

Force fields are simplified and approximate representations of the electronic potential. Therefore describing the reconfigurations of the electronic system due to bond formation would require additional parameterization and treatment with non-standard (computationally expensive) potential functions, e.g. Morse instead of harmonic potentials for bonds (Tab. 2.1).

In Publications P5, P7 and P9 Lennard-Jones instead of Morse potentials (Tab. 2.2) were used to efficiently model the surface-peptide interaction. How does this affect the results and the design of my adsorption studies on gold surfaces? While most amino acids only physisorb on gold surfaces [26], cysteine is known to form covalent bonds with gold atoms on the surface. For this reason, the cysteine adsorption in Publications P7 and P5 only models the physisorption step. This is also reflected in the low interaction free energy of 37.7 kJ/mol compared to the covalent bonding energy of at least 200 kJ/mol [47]. In summary, adsorption of proteins or peptides with exposed cysteines at the surface is not correctly modeled by the current approach. Details are found in the "Results and Discussion" section of Publication P7.

A second limitation arises from the fact that an important property of gold as a metal is its polarizability. Current force fields used for biomolecules and water, however, do not explicitly include atomic polarizability of the electronic system. The gold-protein force field (GolP) [26] applied in Publications P5, P7 and P9 (see Section 4 and Fig. 4.2 for the implementation in my simulations) includes the polarization of the gold surface, since the polarizability of gold is expected to be much stronger than the bio-molecular one. Through this treatment, the induced atomic polarization at the interface is only partially described — polarization of the gold surface is included while polarization in the peptides or water molecules close to the surface is neglected. In summary, e.g. the structure of the un-polarizable water molecules near the surface might not be modeled correctly.

In Chapter 3, the lacking of inducible atomic polarization in current force fields might also influence the dynamics of the dyes in Publication P8, which form a highly delocalized electron system with inducible polarizability. This would influence the sampling of dye conformations. However, since fluorescence anisotropy decay experiments agree with anisotropy decay in our simulations, the dye orientations are most likely accurately sampled. In Section 3.3.2, future directions towards an improved parameterization and quantification of error are pointed out.

A further issue is the force field parameterization itself. MD simulates trajectories of (classical) nuclei, the electron contribution is incorporated by a mean field in which the nuclei move and commonly described by a force field. As stated in Section 2.1.2, force fields are parameterized for a distinct electronic state of the molecule, usually the ground state, and therefore effects like photo-absorption, fluorescence, electron transfer and FRET processes are not described by MD simulations.

For this reason, I developed a hybrid approach in Publication P8. First simulations of the molecule are performed choosing the donor parameters in the excited state and the acceptor parameters in the ground state, as this is the situation after donor excitation and before FRET (see Section 2.4). The resulting trajectories are then used in a Markov chain MC process to discriminate between FRET and donor relaxation, as described later in detail in Section 2.4.

2.2 Simulation of Biological Systems via Molecular Dynamics

2.2.1 Integrating the Equations of Motion

For the integration of Newton's equations of motion, the leap-frog integrator [48]

$$\boldsymbol{v}(t + \tfrac{1}{2}\Delta t) \;=\; \boldsymbol{v}(t - \tfrac{1}{2}\Delta t) + \frac{\Delta t}{m}\boldsymbol{F}(t) \tag{2.7}$$

$$\boldsymbol{r}(t + \Delta t) \;=\; \boldsymbol{r}(t) + \Delta t \cdot \boldsymbol{v}(t + \tfrac{1}{2}\Delta t)\,, \tag{2.8}$$

implemented in GROMACS [37], was used for all simulations. As Eqs. 2.7 and 2.8 show, the name originates from calculation of velocity $\boldsymbol{v}$ and positions $\boldsymbol{r}$ at different times, interleaved by $\frac{1}{2}\Delta t$. Energy is conserved accurately, if the integration time step is small compared to the fastest degrees of freedom in the system.

The leap-frog algorithm was chosen, because it has two advantages over classical $\mathcal{O}_4$ Runge-Kutta [49] algorithms. First, it requires calculation of forces only once per full iteration step and therefore is computationally more efficient. Second and in contrast to Runge-Kutta schemes, leap-frog is a symplectic integrator (see 32, p. 492) with the property that during numerical integration, the system Hamiltonian is conserved up to a small perturbation. Symplectic integrators preserve the phase space area [48] which is also fulfilled by the time evolution of Hamilton's equations [35, 50, p. 48].

In comparison to standard Verlet [51] integration, velocities e.g. required to calculate temperature or the kinetic energy are explicitly present. The velocity-Verlet integrator also explicitly contains the velocities [52] and has the same numerical stability and computational efficiency as the leap-frog algorithm. Velocity-Verlet, however, has a slight disadvantage in parallel communication [37].

Some methods, e.g. rigorously corrected pressure control [53, 54] require v and r at the same instant t. In those cases, leap-frog integration is only applicable at twice of the computational cost compared to velocity-Verlet integration. No methods with the requirement of v and r at the same instant were used in my studies and therefore the leap-frog algorithm was chosen for all simulations.

2.2.2 Solvent Environment and Boundary Conditions

In the biological systems I studied, the proteins or poly-peptides are dissolved in water. The structure of the solvent at biomolecular interfaces strongly influences biological function [55]. Thus, simulations in explicit solvent are desirable [56].

In my studies, the simple point charge (SPC) [57] water model was used for Publications P3, P5, P7 and P9, because this water model was combined with the OPLS/AA force field for parameterization of water gold interactions [26]. In Publication P8 the transferable intermolecular potential with 4 points (TIP4P) [58] water model was applied, since this is the water model, OPLS/AA was parameterized with [42].

Simulations of molecules require a box to constrict the space in which the molecules move. Exceptions are simulations in the gas phase, in vacuum, or when using implicit solvent models, where the center of mass (COM) translation is removed. All simulations in my work were performed in explicit solvent. Therefore, to avoid unphysical interfaces for the solvent molecules at the microscopic scale, periodic boundary conditions (PBC) were employed in MD simulations.

2.2.3 Temperature and Pressure

MD simulations in my work were performed in the canonical NVT or NPT ensembles [59, p. 12]. In both cases, the number of particles and the temperature are kept constant. In the NVT ensemble, the volume, and thus the (periodic) box size, is fixed while in the isothermal-isobaric (NPT) ensemble, the box size is adjusted to obtain constant pressure.

The simplest algorithm to maintain temperature is the Berendsen thermostat [60], which rescales atomic velocities to a certain temperature within a decay time τ. This thermostat shows no temperature oscillations when the starting temperature strongly

differs from the target temperature. Berendsen thermostats however do not generate proper canonical ensembles [37, 61].

The Nosé-Hoover thermostat uses an extended-ensemble approach in which the system Hamiltonian is extended by a thermal reservoir and a friction term is used in the equations of motion (more details can be found in Ref. 37). Nosé-Hoover thermostats produce correct canonical ensembles but show temperature oscillations when far from the target temperature.

In Publications P3, P5, P7 and P9, Berendsen temperature coupling was used during the equilibration due to the non-oscillatory behavior and the Nosé-Hoover [62, 63] thermostat was employed for the production runs.

Recently, the Berendsen thermostat was improved by including a stochastic term to generate a proper canonical ensemble, leading to the v-rescale thermostat [64]. This approach, however, was not implemented in GROMACS at the beginning of my studies, and thus was used in Publication P8 only.

In particular, during the equilibration of the simulation system, it might be desirable to establish or maintain (atmospheric) pressure in the system. Throughout my studies, I used the Parrinello-Rahman [65] barostat, which is conceptually similar to the Nosé-Hoover thermostat [37] by also using an extended ensemble.

In Publications P3 and P8, the protein or poly-peptide is dissolved in water only, and no membrane or surface is present. Therefore, the pressure is isotropically applied to the system (equally to all three dimensions). However, this is not applicable in simulations involving an incompressible gold surface as in Publications P5, P7 and P9. In those cases, the pressure was applied by adjusting the box vector perpendicular to the surface only.

2.2.4 Software and Hardware

In the following, an overview of the applied software packages, the computational hardware and programming languages is given. In addition, the computational workload and simulation performance is exemplified for one of the projects.

For the MD simulations in this work, the GROMACS software package [66–70] was employed. GROMACS has an outstanding single core performance [70] and also shows an excellent parallel scaling since version 4 [67, 68, 71]. Publications P5 and P7 use the version 3.3.3 and 4.0.1 of the GROMACS package, version 4.0.1 was used for Publication P3 and version 4.0.7 for Publications P8 and P9.

Different cluster installations have been used for my studies. The work in Publications P3, P4, P7 and P5 was performed mainly on the Linux cluster of the Leibnitz Rechenzentrum (LRZ) in Munich whereas the computations in Publications P6, P8 and

P9 were conducted on the cluster of the Computational and Theoretical Biophysics group at the Max Planck Institute for Biophysical Chemistry in Göttingen.

The computationally largest project in my studies was the extensive sampling of dye-conformations in Publication P8 with $\approx 1,000,000$ CPUh, which is comparable to a medium sized project on a national supercomputer such as the Höchstleistungsrechner Bayern II (HLRB-II)[1]. Each of the 230 simulations in Publication P8 covers 100 ns of dynamics. On compute nodes with 16 CPU cores, the typical performance of a simulation was 25 ns/day.

For analysis, the g_tools, provided with the GROMACS software package were used, and additional analysis tools were written in C [72] and Python (cPython, NumPy, SciPy, Matplotlib) [73, 74]. During my thesis, I wrote a python library interfacing the GROMACS MD package, which was the basis for the grand canonical Monte Carlo (GCMC) package in Manuscript M1. The FRET MC algorithm [P8] was implemented in Python using optimized C-code for performance critical parts of the algorithm. Molecules were visualized using visual molecular dynamics (VMD) [75] and PyMol [76].

2.2.5 Trajectory Analysis

Typically, the result of MD simulations are trajectories containing the atomic coordinates and velocities, (constraint) forces as well as the (decomposed) potential energy. This information is then used to calculate observables such as the root mean square deviation (RMSD) or the solvent accessible surface (SAS) in Publication P3 as well as mutual orientations or distances of (parts of) molecules in Publications P8 and P9.

Often, one is interested in ensemble averages of observables. For my work in Publication P8, e.g. the ensemble average of the orientation factor $\langle \kappa^2 \rangle$ was an important property to study. When using the ergodic hypothesis

$$\langle A \rangle_{\text{ens}} = \lim_{t \to \infty} \int_t A(t')dt',$$

(2.9)

time averages can be employed instead of averages from multiple systems to obtain ensemble averages of an observable $\langle A \rangle_{\text{ens}}$. In many cases, e.g. for the $\langle \kappa^2 \rangle$ calculation in Publication P8, a hybrid approach is taken using multiple systems (simulations) combined with averaging over time. The advantage is that sampling of multiple systems in parallel shows a perfect scaling with the available computational power — it is easy to simulate 100 systems for 100 ns but not a single system for $10\,\mu$s.

---

[1] `https://www.lrz.de/services/compute/hlrb/projectproposal/`

2.3 Free Energy Calculations via Thermodynamic Integration

Free energy is an important property to understand biophysical processes, such as transient-protein complex formation. In Publication P3, I studied the dependence of the free energy landscape on mutations and the salt concentration. In the adsorption studies on gold surfaces (Publications P3 and P7), the free energy landscape of amino acid adsorption was calculated and later used to benchmark potentials for BD simulations in Publication P4. Therefore, the calculation of the exact free energy landscape or the projection on an appropriate reaction coordinate for the process involved was essential for my work.

The Helmholtz free energy $F = U - TS$, with internal energy U, temperature T and entropy S, is the thermodynamic potential of the natural variables temperature T, volume V and particle count N. F can also be expressed by the partition function Q as

$$F = k_{\mathrm{B}} T \ln Q(N, V, T) \,. \tag{2.10}$$

This equation connects thermodynamics and statistical mechanics in the canonical ensemble by stating that calculation of F allows estimation of Q. The probability ratio $P_{\mathrm{A}}/P_{\mathrm{B}}$ to find a system in state A or B (e.g. adsorbed and in solution) is given by the free energy difference ΔF between those states,

$$\frac{P_{\mathrm{B}}}{P_{\mathrm{A}}} = \exp(-\Delta F \beta) \text{ with } \beta = (k_{\mathrm{B}} T)^{-1}, \tag{2.11}$$

k_{B} as the Boltzmann constant and T as temperature.

Transition rates k are governed by barriers ΔF in the free energy landscape as

$$k = \omega \exp(-\Delta F \beta), \tag{2.12}$$

where ω is the frequency pre-factor or attempt frequency. Two models involving barriers in the free energy profile that limit the association rate of barnase-barstar (see Chapter 6) were suggested in Ref. 77. In Publication P3, I examined which of the models applies and what are the interactions that create these barriers.

2.3.1 Calculating Free Energies from Molecular Dynamics Simulations

Many methods exist to estimate free energy differences from MD simulations. Some examples are the usage of thermodynamic integration, perturbation theory, probability distributions and histograms, non-equilibrium methods or transition path sampling. An extensive overview of the individual methods is found in literature [78]. For recent developments see Refs. 79, 80. In Publications P3 and P7, thermodynamic integration was used as one of the well tested methods with excellent convergence behavior [81].

The key idea of thermodynamic integration is that free energy differences are calculated by integrating forces [82], the derivative of the free energy. These forces constitute a conservative force field and therefore the integration is independent of the integration path. Statistical averages

$$\frac{dF}{d\lambda} = \left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda \tag{2.13}$$

are calculated along a reaction path described by a reaction coordinate λ. The average on the right side of Eq. 2.13 is computed for a fixed λ and is interpreted as constraint force or generalized force along λ [78, p. 121]. Simulating a constrained system at multiple points along λ from state A to B and thereby recording the constraint force allows numerical integration of Eq. 2.13 to calculate free energy differences.

The right side of Eq. 2.13 is also written as

$$\left\langle \frac{\partial \mathcal{H}}{\partial \lambda} \right\rangle_\lambda = \left\langle \frac{\partial U}{\partial \lambda} - k_B T \frac{\partial \ln |J|}{\partial \lambda} \right\rangle_\lambda , \tag{2.14}$$

assuming that λ is a function of atomic coordinates [78, 82, p. 121]. Here, U is the potential energy and $|J|$ the determinant of the Jacobian matrix, which measures the change in volume when switching from Cartesian to general coordinates. The last term leads to an entropic correction for the increase in phase space volume [81], e.g. when using radial symmetric reaction coordinates as in Publication P3.

Although free energy differences between two states are independent from the choice of the reaction coordinate, the free energy profile (also referred to as potential of mean force (PMF)) itself depends on the integration path. In many cases, the most probable pathway is the one of interest and therefore the path with the lowest barriers, and thus fastest kinetics. Therefore, the careful choice of the reaction coordinate is important and non-trivial in many cases. The distance between the COM of Barnase and Barstar was used as reaction coordinate in Publication P3 and the COM distance from the surface atom layer was chosen in Publication P7.

2.4 Fluorescence Resonance Energy Transfer

In biotechnology, RET in combination with fluorescence (FRET) is a widely used method to measure distances. In addition, RET plays a key role in photosynthesis as energy transfer mechanism in the light harvesting complex (LHC). The theoretical description of RET has been pioneered by Theodor Förster in the late 1940's [83], therefore, FRET is also referred to as Förster Resonance Energy Transfer. First, the principal mechanism and limits of RET will be explained (Section 2.4.1) and RET measurements are sketched (Section 2.4.2). This will be followed by a summary of the theory and of the approximations used in experiments (Section 2.4.3). Finally, the orientational dependency of the RET mechanism and the approximations commonly applied in experiments are highlighted (Section 2.4.4), leading to the combination of experiments and simulations demonstrated in the following Chapter 3.

2.4.1 Resonance Energy Transfer and the Limitations

During RET, also referred to as electronic energy transfer (EET), the excitation, and thus the energy localized on (part of) a donor molecule is transferred to the corresponding acceptor molecule. Thereby, the electronic system of the donor molecule relaxes to the ground state with the excitation transfered to the acceptor counterpart (Fig. 2.1, step 2b).

RET is the dominant energy transfer process for molecular separations of $1 - 20\,\mathrm{nm}$, where the orbital overlap of donor and acceptor can be neglected (short range limit). The long range limits of this radiationless energy transfer are distances, where the energy is transferred via radiation (photon emission and absorption). The classical analogon for the RET regime is the near field zone of an antenna, where the distance of the receiver to the emitting antenna is small compared to the wavelength.

2.4.2 Measurement of FRET

Despite the term "fluorescence" in the common abbreviation of FRET, the energy transfer mechanism itself is unrelated to molecular fluorescence. However, fluorescence or fluorescence quenching is often used to measure the efficiency of the energy transfer [84]. In the following, determination of the FRET efficiency based on measurement of fluorescence intensities is sketched [85]. An overview of the different FRET detection methods is found in Ref. 86.
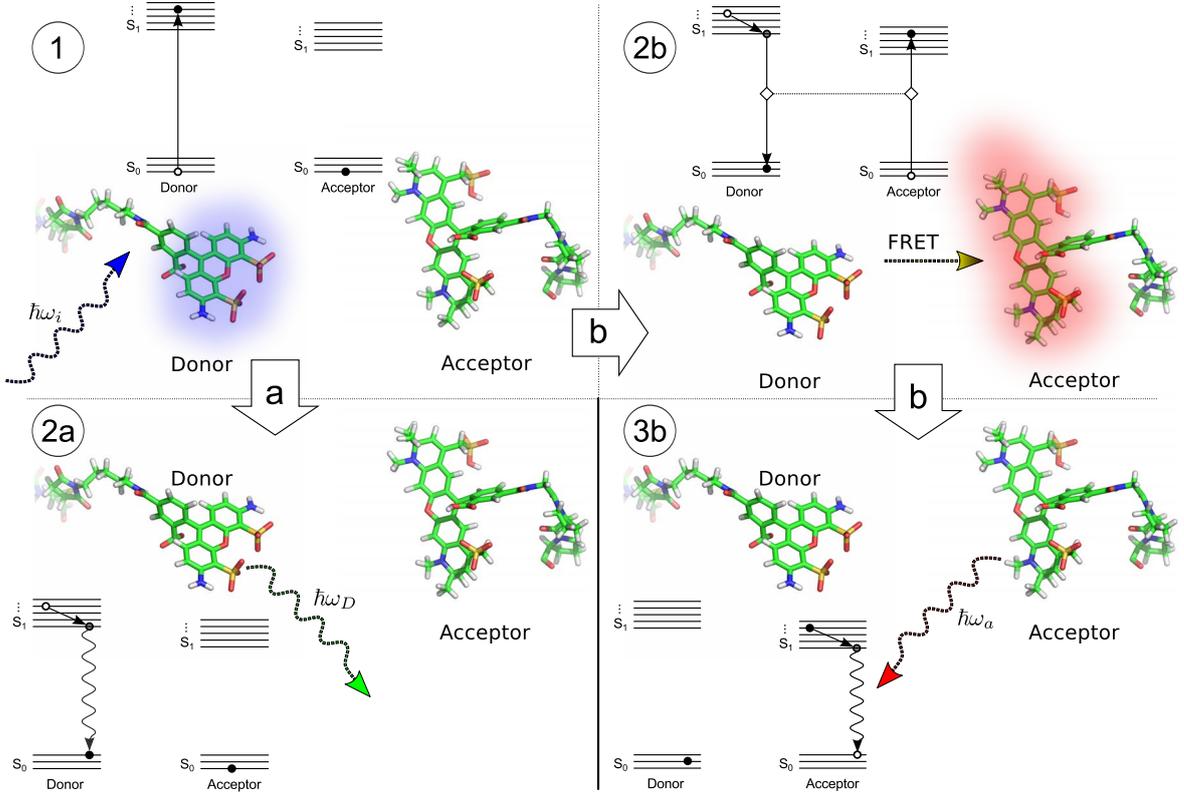
Figure 2.1: **Two Competing Pathways During FRET:** Normal fluorescence, the donor is excited by an absorbed photon (1) and the energy is released via photo-emission (2a). After excitation via photo-absorption (1), the excitation is transferred by RET to the acceptor (2b) and then again released via photo-emission, in this case from the acceptor (3b). The number 1–3 denote the first, second and third step while a/b discriminates between the two pathways

Figure 2.1 shows the two pathways in the FRET process. By recording the donor (1 → 2a) and acceptor (1 → 2b → 3b) fluorescence decay intensities after donor excitation, the efficiency E_{FRET} is calculated via

$$E_{\text{FRET}} = \frac{I_{\text{acc}}}{I_{\text{acc}} + I_{\text{don}}}. \tag{2.15}$$

I_{acc} and I_{don} are the measured acceptor and donor photon fluorescence intensities (photon counts per time-interval), respectively. Distances are obtained from the efficiency measurement, when the Förster formula,

$$E_{\text{FRET}} = \frac{1}{1 + (R/R_0)^6},\qquad(2.16)$$

is applied. R is the distance between the two dyes and R_0 the Förster radius, the distance at which the excitation is transferred with 50% probability. The Förster radius is assumed to be constant for a given dye pair.

Combination of Eq. 2.15 and Eq. 2.16 together with measurement of donor and acceptor fluorescence intensities allows measurement of distances. The accuracy, however, strongly depends on assumptions "hidden" in the constant R_0 which will be discussed in the next section.

2.4.3 Derivation of the Förster Formula

In the following paragraph, the derivation of the Förster formula, Eq. 2.16, and the Förster radius R_0, commonly used in FRET experiments, are sketched and the underlying assumptions are explained. For extensive discussion of the individual approximations the reader is referred to Refs. 87–89. Reference 90 summarizes determination of the individual parameters in the Förster radius R_0 by experiments.

The rate (or probability) of RET is described by an equilibrium Fermi's golden rule approach [91, 92] of initial and final state, D*A and DA* (the asterisk denotes the excitation), in a time-dependent perturbing Hamiltonian H',

$$k_{\text{RET}} = \frac{2\pi}{\hbar} \left| \langle \text{D}^*\text{A} | H' | \text{DA}^* \rangle \right|^2 J = \frac{2\pi}{\hbar} \left| V_{\text{cpl}} \right|^2 J_{\text{DA}},\qquad(2.17)$$

with V_{cpl} as coupling between donor and acceptor between state D*A and DA*. J is the density of final states or Franck-Condon overlap [93, 94]. Derivation of Fermi's golden rule based on second order time-dependent perturbation theory is found in Ref. 95.

In the following, it is assumed that the transition density (see Ref. 96, page 128ff),

$$P_{\text{g}\to\text{e}}(\boldsymbol{x}) = \int \Psi_{\text{g}}(x_1, \ldots, x_{\text{N}}) \Psi_{\text{e}}^*(x_1, \ldots, x_{\text{N}}) dx_2 \ldots dx_{\text{N}} \;,\qquad(2.18)$$

of each dye from ground to excited state ($g \to e$) is not affected by the presence of surrounding molecules, in particular not by the presence of the second dye (no orbital overlap). $\Psi_{\text{g}}(x_1, \ldots, x_{\text{N}})$ and $\Psi_{\text{e}}(x_1, \ldots, x_{\text{N}})$ are the wave-functions of N electrons in

ground and excited state, respectively. x_i are the spatial coordinates and spin of the electron N_i. The electronic coupling is independent of the spin, therefore the spatial transition density is

$$\rho_{g \to e}(\boldsymbol{r}) = \int\limits_s P_{g \to e}(\boldsymbol{x})ds \ . \tag{2.19}$$

With this assumption, the coupling,

$$V_{\text{cpl}} \approx V_{\text{coul}} = \frac{1}{4\pi\epsilon_0} \int\limits_{\boldsymbol{r}} \int\limits_{\boldsymbol{r'}} \frac{\rho_{\text{D}}(\boldsymbol{r})\rho_{\text{A}}(\boldsymbol{r'})}{R_{rr'}} d\boldsymbol{r}d\boldsymbol{r'} \ , \tag{2.20}$$

is described by the Coulombic coupling V_{coul} between the transition densities of ρ_{D} and ρ_{A} for donor and acceptor, respectively. $R_{rr'}$ is the distance between the two volume elements.

In the next step, the transition density coupling is approximated by dipole-dipole coupling,

$$V_{\text{coul}} \approx V_{\text{dip}-\text{dip}} = \frac{1}{4\pi\epsilon_0 n^2} \frac{\kappa|d_{\text{D}}||d_{\text{A}}|}{R_{\text{DA}}^3} \tag{2.21}$$

neglecting higher coupling moments. d_{D} and d_{A} are the dipole moments of the transition densities ρ_{D} and ρ_{A} for donor and acceptor, respectively, and R_{DA} is the separation between the two dyes.

Starting from the dipole-dipole coupling approximation, the refractive index n of the media at optical frequencies has to be considered [97], commonly assumed to be homogeneous and isotropic. The orientation factor κ represents the orientational dependency of the dipole-dipole coupling,

$$\kappa = [\cos\theta_{\text{DA}} - 3\cos\theta_{\text{D}}\cos\theta_{\text{A}}] \ . \tag{2.22}$$

The overlap J (Eq. 2.17) of the Franck-Condon envelopes[1] is calculated by the spectral overlap integral I of normalized donor emission and acceptor absorption spectra [88, 89].

With these approximations and assumptions, the Förster radius R_0 is

$$R_0^6 = \frac{9(\ln 10)Q_{\text{D}}I\kappa^2}{128\pi^5 n^4 N_{\text{A}}} \ , \tag{2.23}$$

with N_{A} as Avogadro number, and Q_{D} as the quantum yield of the donor dye.

---

[1]Franck-Condon envelopes are the intensity distribution shape in the adsorption or emission bands.

The Förster radius is commonly considered as constant in experiments, assuming that all parameters do not change for different conformations. Q_D, n and I are considered independent of the local environment. In the following, the approximations to obtain a constant orientation factor κ^2 are summarized.

2.4.4 The $\langle \kappa^2 \rangle = 2/3$ Approximation

The competing process to FRET is donor decay, e.g. via fluorescence. Therefore, donor fluorescence decay limits the time scale on which FRET can occur to the lifetime of the donor excitation τ_D in absence of the acceptor dye.

To understand the basis for the $\kappa^2 = 2/3$ approximation, two limiting cases are considered first[1]:

- **The dye re-orientation is slow compared to the donor lifetime τ_D.** The probability for FRET, which is equal to the FRET efficiency, depends on the static dye orientation while the donor is excited. $\longrightarrow \kappa^2(t) \approx \kappa^2_{\text{static}}$

- **The dye re-orientation is fast compared to the donor lifetime τ_D.** Therefore, the mean orientation factor determines the FRET efficiency. $\longrightarrow \kappa^2(t) \approx \langle \kappa^2 \rangle_{\tau_D}$

Assuming that the second case applies for the system of interest, the average of the orientation factor within τ_D is close to the ensemble average. In addition, it is assumed that the orientation distributions of both dyes are isotropic and both dye orientations are uncorrelated. With this, Eq. 2.22 integrates to

$$\langle \kappa^2 \rangle_{\tau_D} \overset{\text{isotropic}}{=} \frac{2}{3} \,. \tag{2.24}$$

The focus of the following Chapter 3 are the limits of this approximation.

---

[1]The cases are discussed for dye distances close to or larger than the Förster radius R_0, where donor fluorescence and FRET have comparable probabilities. Therefore the donor lifetime with acceptor dye present is still comparable to the lifetime of the donor τ_D in absence of the acceptor dye.

Main Projects
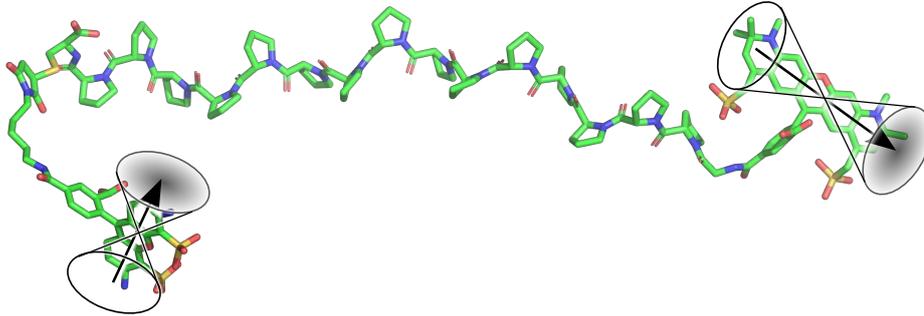
Simulation-Aided Distance Reconstruction in FRET Experiments

Aims:

- Development of a smFRET simulation protocol and test against experimental data.

- Assess the impact of the isotropic $\kappa^2 = 2/3$ approximation on the accuracy of distance reconstruction in experiments.

- Improve accuracy of distance reconstruction in experiments through combination with simulated orientational dye dynamics.
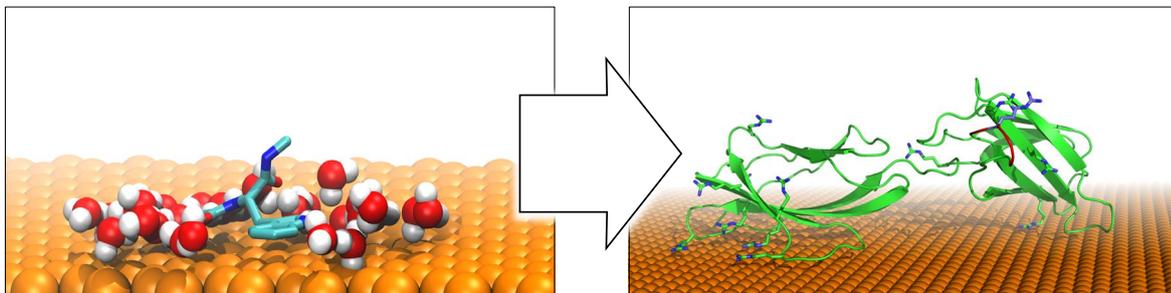
Simulated Adsorption of Biomolecules on Gold Surfaces

Aims:

- Development of experimentally validated surface docking toolkit for high throughput screening.

- Find design principles for surface adsorption of proteins and biomolecules.

3

Simulation-Aided Distance Reconstruction in FRET Experiments

- M. Hoefling, N. Lima, D. Hänni, B. Schuler, C. A. M. Seidel and H. Grubmüller. Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines Through a Hybrid Atomistic Simulation and Monte Carlo Approach. *PLoS One*, 6(5):e19791, **2011**. (Publication P8).

Individual Contributions

Parameterization and testing of the dye parameters in simulations was performed by Nicola Lima. Simulations of poly-prolines with dyes attached, trajectory analysis and the photon generation MC were done by myself. Helmut Grubmüller and myself established the theoretical framework. The transfer function (TF) formalism for the different levels of approximation was introduced by Helmut Grubmüller and myself. Experiments were performed by Benjamin Schuler and Dominik Haenni. The manuscript of Publication P8 was written by Nicola Lima, Helmut Grubmüller and myself. All authors of Publication P8 aided in the interpretation of the results from simulation and experiments.

3.1 Introduction

After considering the theoretical foundations of FRET in the previous chapter, I will first compare FRET to other techniques for the measurement of distances on the nanometer scale (Section 3.1.1) and thereby highlight the advantages of FRET. The development of FRET over the last decades shows, why the dye orientation models employed in the Förster formula limit the accuracy in particular in recent single molecule setups (Section 3.1.2). The current knowledge on how much the dye orientation influences the measured distances is summarized in Section 3.1.3.

3.1.1 Measurement Techniques in the Nanometer Range

Several techniques allow the measurement of distances in the nanometer range. Most prominent examples are X-ray crystallography, cryo electron microscopy (cryo-EM) and transmission electron microscopy (TEM). These techniques are successfully applied in structural biology to resolve the structure of large proteins. However, they are not applicable *in vivo*; X-ray requires crystallized structures of the analyte, TEM and cryo-EM require staining with heavy metals or frozen structures, respectively.

Nuclear magnetic resonance (NMR) spectroscopy is a further technique routinely employed in structural biology to resolve protein structures in solution. No crystal is required and therefore NMR is particularly useful to study intrinsically disordered proteins or membrane proteins. However, it allows the measurement of distances in small proteins only. Further, NMR can only detect changes in the ensemble and — like X-Ray crystallography — is not suitable to study single molecules.

Since the 17th century, optical microscopy has been used for imaging of cells and cell compartments. The resolution of optical microscopy is diffraction-limited and therefore only structures larger than $\approx 200\,\mathrm{nm}$ can be resolved. Recent advances in optical microscopy like stimulated emission depletion (STED) microscopy broke this limit [98, 99]. STED typically improves the resolution of optical microscopy by one order of magnitude [100].

The FRET technique combines the optical readout and thus the *in vivo* applicability with the measurement of the distance-dependent RET process. The distance is measured between two sites, labeled by donor and acceptor dyes, respectively. FRET is applied to measure distances in a single molecule or in an ensemble.

FRET can also easily be combined with other techniques. *In vivo* force imaging in the cell [101] is a beautiful example of such a combination: Vinculin, a protein connecting integrins (see Chapter 5) to the actin cytoskeleton in the cell is separated into two parts, connected by an elastic linker with the ends labeled by a pair of FRET dyes. The FRET signal of the elastic linker is calibrated by force spectroscopy measurements with optical tweezers on single dye-linker constructs. Recording the FRET signal of the calibrated linker *in vivo* allows force imaging in a living cell.

FRET is not limited to the measurement of single distances. Monitoring the transfer efficiency of multiple FRET pairs and thus distances at the same time allows trilateration, e.g. of domain positions in biomolecules [102]. The measurement of multiple distance via FRET has recently been improved by using switchable dyes [103] instead of multiple dye pairs. In summary, FRET is a versatile method to measure distances.

To understand why the dependency of the FRET efficiency on dye orientation limits the accuracy in particular in recent experiments, the history of FRET experiments is discussed in the following Section 3.1.2.

3.1.2 Historical Perspective on the Dye Orientation Problem

Förster developed his theory in 1948 [83] to describe RET of dyes in solution. In Förster's setup, the dyes are assumed to freely diffuse in the solvent. Therefore assuming isotropic and uncorrelated movement of the dyes is accurate at low concentrations.

In 1967, Stryer and Haughland published FRET measurements of different polyproline chain lengths and demonstrated that FRET can be used as "spectroscopic ruler" [104]. Stryer and Haughland measured the FRET efficiency in the ensemble, as Förster did before. In the ensemble, the efficiency is measured as the average of intensities from multiple dye-pairs and thereby the derived distances also represent an ensemble average.

Advancements in several fields were required to allow FRET experiments of single molecules. Most notably in this context are the improvements in detection efficiency of single photons reducing the undetected signal from the dyes [105]. On the other side, the signal itself has been improved by the development of dyes with large quantum yields [106] such as quantum dots [107].

Through these advancements, FRET became a standard tool to determine interactions and to measure distances *in vivo* [108]. SmFRET was used for example to monitor the folding process in chaperons [109, 110], or to estimate free energy surfaces of protein folding in solution from measured distance distributions [111].

The assumed isotropic orientation, described by an orientation factor of $\kappa^2 = 2/3$, was recognized as accuracy-limiting quite early [112]. With FRET being increasingly used to determine exact distances [113] and not only the presence of the second dye, the unknown orientation became a major uncertainty [114, 115]: "*In almost all previous single molecule FRET studies [...], it was assumed that $\kappa^2 = 2/3$, not because this assumption is strictly valid, but for lack of further information. In fact, most fluorophores [...] orientations relative to the host molecules are not entirely randomized within their fluorescence lifetimes. Therefore, the best we can hope for is an approximation to a true distance.*" [116].

Particularly, two factors render the orientation dependency of FRET as accuracy-limiting in recent setups. First, and in contrast to Förster's experiments, the dyes are employed to label biomolecules and therefore interact with them. When dyes are attached to the protein surface for example, the surface restricts the available

conformational space for the attached dyes and therefore the dye orientation sampling is non-isotropic. As a consequence, the averaged orientation factor κ^2 is not 2/3.

The second issue affects in particular smFRET experiments. Förster and also Stryer and Haughland performed ensemble experiments and thereby measured the average efficiency of the dye conformation ensemble. In single molecule experiments, the measurement time can be too short to obtain a converged dye conformation ensemble and therefore the ergodic hypothesis (Section 2.2.5) is not fulfilled.

The following section summarizes previous approaches to assess the unknown dye orientation in experiments and by theoretical considerations.

3.1.3 The Unknown Orientation Factor κ^2

The common experimental procedure to estimate the deviation of the isotropic dye orientation is to perform anisotropy measurements. Thereby, the dye is excited by a polarized laser pulse, and the fluorescence relaxation intensities parallel and perpendicular to the excitation beam are recorded. By this, the rotation of the dyes between excitation and decay is estimated and for anisotropy values is below 0.2, the isotropic assumption $\kappa^2 = 2/3$ is considered sufficiently accurate [108].

However, anisotropy measurements detect only orientational dynamics perpendicular to the transition dipole axis. Also, anisotropy decay measurements are typically ensemble measurements and therefore the specific dynamics of individual dyes might be hidden in the ensemble decay. Moreover, the decay is measured for each dye of the FRET pair separately and therefore anisotropy decay does not detect correlations between the mutual orientations of both dyes.

One question arising from these measurements is how to handle those cases, in which the anisotropy is large and thus the dye orientation is not accurately described by the isotropic κ^2.

Improved accuracy was achieved by estimating the mean κ^2 for the measured system through models and simulations. In some models, geometries in the confined dye orientation space are generated by a MC approach [117, 118]. Other models are based on a combination of intensity measurements and geometrical considerations [119]. For rigid periodic structures like DNA, the orientation factor κ^2 has implicitly been determined by fitting to a structural model for the Förster radius R_0 [120, 121] .

For many dye environments like complex protein surfaces, simple treatment of the interactions between dye with its environment as geometrical constraint is not an accurate model. Simulations are therefore a promising approach to obtain accurate dye orientation dynamics. From coarse grained (CG) Langevin dynamics (LD) simulations

of immunity protein Im7 in implicit solvent it was concluded, that the error in κ^2 is smaller than the experimental one [122].

In MD simulations of lipid systems it has been shown, that κ^2 strongly depends on the local environment of the dyes [123, 124]. The simulations were performed at atomistic detail with explicit solvent, and the findings are in contrast to the influence of the dye environment smaller than the experimental error observed in CG/LD simulations in implicit solvent. As a result, fully atomistic simulations in explicit solvent might be required to accurately model interactions of dye and its environment.

In summary, the dye motion is often strongly restricted due to interactions with the biomolecule attached to, thus also restricting the conformational sampling of orientational space [125]. This anisotropic sampling is difficult to detect and quantify in experiments [126].

Besides the non-isotropicity of the individual dye orientation, the correlations between the two dyes might also limit the accuracy (see Section 2.4.4). First, the dye orientations not necessarily have to be uncorrelated. Second, and more importantly, each dye orientation can be correlated with the distance between the dyes.

Recent findings suggest that assuming a distance independent, ensemble averaged orientation factor κ^2 limits the accuracy in FRET measurements due to the averaging over the distance. Van Beek *et al.* [14] found that the averaging of orientation factor κ^2 and the distance R, $\langle \kappa^2 R^{-6} \rangle$, differed by a factor of 1.6 from to the product of the individual averages of κ^2 and R, $\langle \kappa^2 \rangle \langle R^{-6} \rangle$. Therefore, the correlation between κ^2 and R is not negligible in experiments. Since this correlation is difficult to measure, I included a distance-dependent orientation factor $\kappa^2(R)$ in the distance reconstruction as shown further below.

In summary, the effect of the employed approximations in the orientation factor κ^2 on the accuracy of distances from FRET experiments is not clear. For these reasons, I systematically studied different levels of approximation. Each level distance distributions, e.g. κ^2 from a refined model or distance-dependent $\kappa^2(R)$ were compared to distances from the isotropic approximation $\kappa^2 = 2/3$. This comparison allows to probe the impact on the accuracy of reconstructed distances. The aim was twofold: First, the error in measurements — which is different for each system — was quantified. Second, I replaced the isotropic κ^2 assumption by dye orientation dynamics and statistics to improve the accuracy of distance reconstruction from experimental raw data .

3.1.4 The Poly-Proline Model System

Poly-proline forms a type II helix in water [127] with the peptide bonds in the trans-configuration. Since the structure and thus also length of the straight type II helix was known, Stryer and Haughland used poly-proline as a spacer between two dyes in their FRET experiments [104]. The poly-proline chain was considered being in the all-trans configuration by Stryer *et al.* and therefore suitable as "spectroscopic ruler" [104].

Schuler *et al.* recently re-investigated poly-proline chains of different lengths in smFRET experiments and found unexpectedly high efficiencies [128], hinting at additional dynamics in the system. These dynamics were attributed to the formation of cis-isomers as later confirmed according to NMR measurements and simulations of poly-proline chains [129].

Cis-conformations result in a kink in the poly-proline helix. These kinks decrease the distance and thus result in higher FRET efficiencies as shown by Fig P8.9 on page 146. Besides the change in distance through isomers, a second source of heterogeneity is the anisotropicity of the dye orientation sampling. Interactions of the dye with the poly-proline chain alter isotropic orientation sampling and thus also affect the FRET efficiency.

Simulations of small poly-proline chains in implicit solvent show κ^2 deviations from 2/3 for [129], although the steric restrictions of the poly-proline chain on the dye conformations are small. This finding suggests that the dynamics of the dyes is influenced by the interactions with the poly-proline chain. However the exact mechanism of this interaction and the influence on the FRET efficiencies are unknown.

There are two main reasons, why poly-proline is an excellent model system to study FRET *in silico* and the impact of dye orientation on reconstructed distances. First the cis/trans isomer probabilities are known from NMR experiments [129]. This allows studying the effect of this heterogeneity by a simulated ensemble as I will show below. Second, accurate smFRET efficiency distributions [128] are available as reference to test the accuracy of my *in silico* FRET approach.

For these reasons, I studied a system consisting of poly-proline chains with different lengths (15, 20 and 30). Alexa 488 and Alexa 594 dyes (Fig. P8.1 on page 139) were attached via linkers at the termini of the poly-proline chain respectively as shown in Fig. P8.2 on page 139.

$$E = \frac{I_A}{I_D + I_A}$$

$$R_{DA} = R_0 \cdot \sqrt[6]{1/E - 1}$$

$$R_{XY} \approx R_{DA}$$
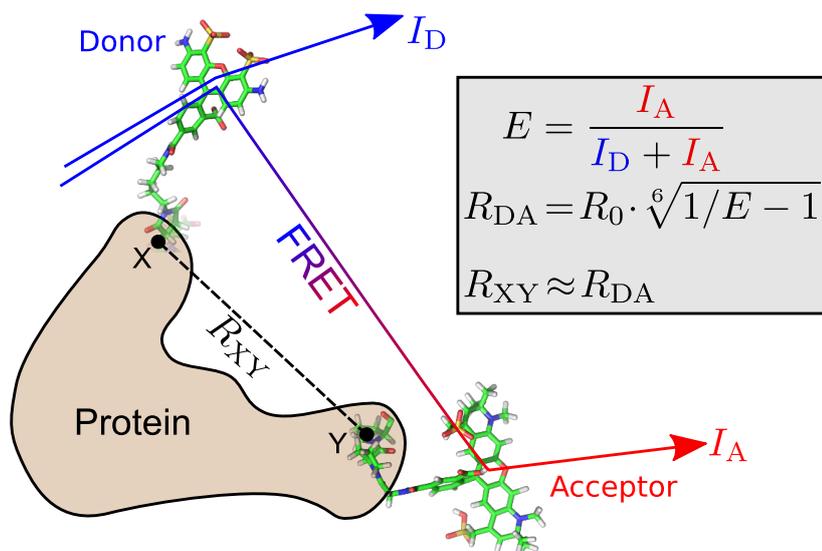
Figure 3.1: **FRET Distance Measurement in Biomolecules.** To measure the distance R_{XY} between X and Y in a protein, the donor dye is excited by incident photons. After the excitation, the donor dye relaxes to the ground state either emitting a photon or by transferring the excitation to the acceptor via FRET which also relaxes to the ground state after emitting a photon. By measurement of the donor and acceptor fluorescence intensities I_D and I_A, the efficiency E of the FRET process is determined. Assuming isotropic dye orientation sampling, the distance between donor and acceptor R_{DA} can be calculated from the efficiency E and the Förster radius R_0 (Eq. 2.16). Since the dyes are attached to X and Y via linkers, the distance R_{DA} is a measure for R_{XY}.

3.2 Results and Discussion

As outlined already in Chapter 1, I proceeded in two steps:

- First, I modeled smFRET experiments *in silico* as close as possible to experiments. Agreement of *in silico* FRET with experiments served as a test if the *in silico* model accurately describes the experiment (Section 3.2.1).

- Second, the impact of approximations on the distance reconstruction was systematically studied (Section 3.2.2).

To better understand the steps of my *in silico* FRET approach, a typical ensemble FRET measurement as shown in Figure 3.1 will be considered first. In ensemble experiments, intensities I_D and I_A of donor and acceptor fluorescence are measured

from an ensemble consisting of many FRET pairs. The obtained efficiency is therefore a measure for the ensemble average of the distance between the dyes.

To monitor the dynamics of individual molecules and the heterogeneity in the ensemble, smFRET experiments have to be performed which are for these reasons also the experimental reference [128] in this work.

In smFRET experiments a single molecule diffuses through the detection volume of a confocal laser spot. As sketched in Fig. 3.3 (a), the donor is excited multiple times during this diffusion, and single photons from donor and acceptor are measured, in analogy to the measurement of intensity in the ensemble. The collected photons from a single diffusion event are referred to as a burst.

Each burst has its own FRET efficiency, which is calculated from the donor and acceptor photon counts. This efficiency is an average of the single molecule and dye dynamics in the burst in contrast to ensemble experiments, where the efficiency is an average of the entire ensemble dynamics. From multiple bursts, an efficiency distribution is obtained.

The diffusion through the detection volume is on the time scale of milliseconds whereas the cis/trans isomerization occurs on much longer times scales of minutes to hours. For this reason each poly-proline chain diffusing through the laser spot is in a distinct isomeric state and smFRET experiments allow the detection of the isomeric ensemble heterogeneity [128].

3.2.1 *In silico* FRET Experiments

In the following, the three steps of my developed *in silico* smFRET approach are explained and the obtained efficiency distributions are compared with the experiment.

Step 1: *Calculation of dye orientation and dynamics in MD simulations*

At the beginning of my *in silico* approach the poly-proline and dye dynamics was sampled by MD simulations. The interaction balance between dye, solvent and the poly-proline chain determines the dye dynamics as highlighted in Section 3.1.3. For this reason, my simulations were performed in explicit solvent to accurately model interactions between all three components. To assess the required sampling by simulations, three time scales are considered here.

First, continuous dye dynamics from photo-adsorption to photo-emission is required to simulate a single FRET event. This dynamics is on the time scale of the donor lifetime ($\approx 4\,\mathrm{ns}$ for Alexa 488) and thus covered by my simulations covering 100 ns of dynamics each.
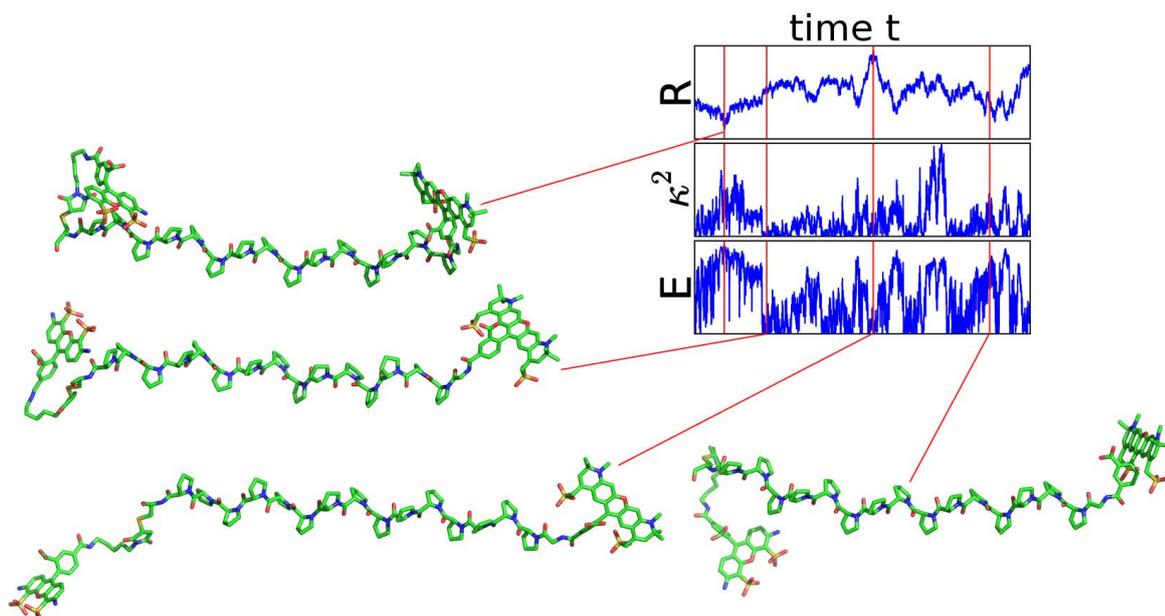
Figure 3.2: **Orientational Dynamics of Poly-Proline and the Corresponding Instantaneous FRET Efficiency.** The time dependent dye-to-dye distance $R(t)$ and orientation factor $\kappa^2(t)$ as well as the resulting efficiency trajectory $E_{\mathrm{FRET}}(t)$ are shown as plot. The red bar shows the time t of the four poly-proline snapshots.

Second, the total dynamics of one burst covers a few milliseconds. This is difficult to simulate by MD and for this reason, I will assume that the dye dynamics of each recorded photon in the burst is uncorrelated to the dynamics of the other recorded photons. With this assumption, multiple simulations were conducted to sample the dye dynamics during a burst. The convergence behavior of multiple simulations is shown in Fig. P8.5 and P8.6 on page 143 for the distance and orientation factor distributions.

The third time scale is the cis-trans isomerization which results in the isomeric heterogeneity of the poly-proline ensemble detected by NMR experiments. The cis-trans transition times from minutes to hours are far beyond the time scale accessible to MD simulations. Therefore, to obtain the proper isomer ensemble, simulations starting from cis and all-trans isomers were performed and weighted according to the measured probabilities from NMR spectroscopy. The number of trajectories of the simulation ensemble is summarized in Tab. P8.1 on page 140 and covers a total of $22\ \mu s$ of dynamics of the poly-proline ensemble.

The simulation trajectories were then employed to calculate FRET efficiency trajectories. As shown in Fig. 3.2, each trajectory snapshots contains distance and orientation of the dyes. From the distance and orientation, the instantaneous FRET efficiency $E_{\text{FRET}}(t)$ or FRET rate $k_{\text{FRET}}(t)$ was calculated for each trajectory snapshot using Eq. 2.16.

Step 2: *Generation of photons using a Monte Carlo process*

In the next step, photons are generated using the FRET rate $k_{\text{FRET}}(t)$ in a kinetic model:

$$
\begin{array}{ccccc}
\text{D} + \text{A} + h\nu & \longrightarrow & \text{D}^* + \text{A} & \xrightarrow{k_{\text{FRET}}(t)} & \text{D} + \text{A}^* \\
& & \downarrow k_{\text{D}} & & \downarrow k_{\text{A}} \\
& & \text{D} + \text{A} + h\nu_{\text{D}} & & \text{D} + \text{A} + h\nu_A
\end{array}
\tag{3.1}
$$
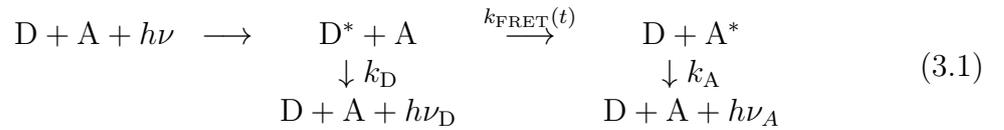
D and A are donor and acceptor, respectively; the asterisk denotes the excited state. The initial point of the kinetics is the excitation of the donor by an incident photon with energy $h\nu$. In this excited state $\text{D}^* + \text{A}$, the donor relaxes to the ground state either via fluorescence k_{D} or via FRET $k_{\text{FRET}}(t)$ followed by acceptor fluorescence k_{A}.

As summarized in in Fig. P8.4 on page 141, the kinetic Scheme 3.1 was evaluated as Markov Chain MC process starting at a random excitation time t in the simulation trajectory ensemble. When donor or acceptor decay is reached in the kinetic model, the corresponding photon was recorded.

It is useful here to point out that for ensemble experiments as reference, the FRET efficiency can be calculated from the obtained donor and acceptor photon counts. As explained before, the experimental reference in my work are smFRET experiments where photon bursts are measured. Therefore, in the next step photon bursts were generated from the recorded photons.

Step 3: *Calculation photon bursts from multiple trajectories*

As pointed out in Section 3.2, the dynamics of the dye and the poly-proline chain is averaged in a photon burst (Fig. 3.3 (a)) when the efficiency of the burst is calculated[1]. The photons in my approach are not generated from continuous dynamics, i.e. a single simulation trajectory, and thus burst formation using multiple simulations has to take the dynamics into account.

---

[1]Dynamics on shorter time scales can be recovered by maximum likelihood techniques [130, 131]
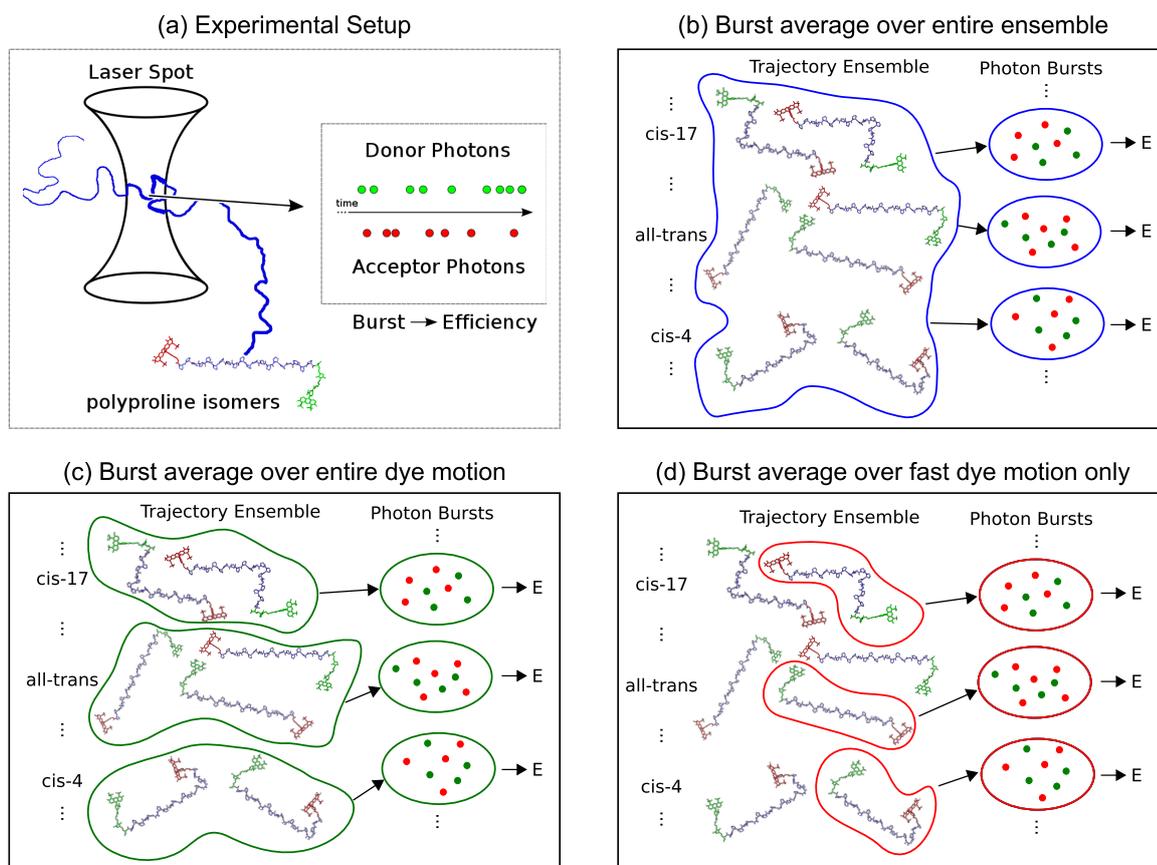
Figure 3.3: **Combining Simulation Trajectories to Mimic Experimental Bursts.** (a) smFRET experiments measure photon bursts from a single molecule, diffusing through the laser spot. (b-d) describes the bursts formation from simulations. In (b), photons of each burst originate from the entire trajectory ensemble. All trajectories of the same isomer are used for each burst in (c) whereas in (d) each burst originates from a single trajectory from the ensemble.

The poly-proline chain dynamics has two main components: The first one are changes in length, which are much shorter than the simulation length as shown by the chain length convergence within simulation time in Fig. P8.5 on page 143.

The second component of the dynamics is the isomerization which is slower than the burst recording time. To test the impact of the isomerization, each bursts was formed from photons of the entire ensemble first (Fig. 3.3 (b)). Comparison of this burst formation method (blue) with the measured efficiency distributions (black) from

smFRET in Fig. P8.10 on page 147 shows a clear discrepancy due to the averaging over the entire ensemble dynamics. Thus, each bursts will consist of photons from the same isomer in the following.

In contrast to the poly-proline chain, the dye dynamics in each burst can only be estimated indirectly, e.g. by anisotropy decay measurements as discussed previously. To assess the impact of the dye dynamics on the efficiency distributions two burst formation schemes shown in Fig. 3.3 (c-d) were tested.

In the first (c), the barriers between different dye conformations are low and the dyes sample the conformational space within each burst. To reflect this in the smFRET approach, photons from multiple trajectories are combined.

In the second approach (d), it is assumed that large barriers between different dye conformations are present preventing the transition within burst recording time. For this reason, the sampling of a single trajectory is used for each burst to represent the photon burst heterogeneity. The results from these two approaches (c-d) are compared to the measured efficiency distribution in the following.

Comparison of efficiency distributions from simulations and experiment

Comparison of the efficiency distribution from the methods referred to as (c) and (d) (Fig. 3.3) shows agreement with the experiment at different parts of the burst efficiency histogram (Fig. 3.4). The burst efficiencies from all isomer trajectories (c) match the experimental efficiency on the low efficiency side while the burst efficiencies from single trajectories (d) are in agreement with the high efficiency side.

This difference in agreement has its origin in different dye conformations leading to the high and low efficiencies. While the open conformation, shown in Fig. 3.4, has an increased distance between both dyes and therefore results in low burst efficiencies, the opposite is found for the closed conformation.

The two conformations not only have different distances between the dyes but also different dynamics. In the open conformation (Fig. 3.4, right side), only weak interactions with the poly-proline chain are present. Therefore multiple simulations are required to accurately sample the conformation space as in the much longer bursts. In the closed conformation, the interactions of dye and poly-proline backbone are strong such that an interconversion of different closed conformations is inhibited.

In summary, the simulations revealed that the slow dye dynamics is an additional source of the heterogeneity observed in the smFRET experiments apart from the isomeric heterogeneity. The agreement between experiment and simulation suggests that the *in silico* approach is a realistic description of smFRET measurements. Thus, distance reconstruction based on the recorded dye statistics is feasible and will be demonstrated in the following.
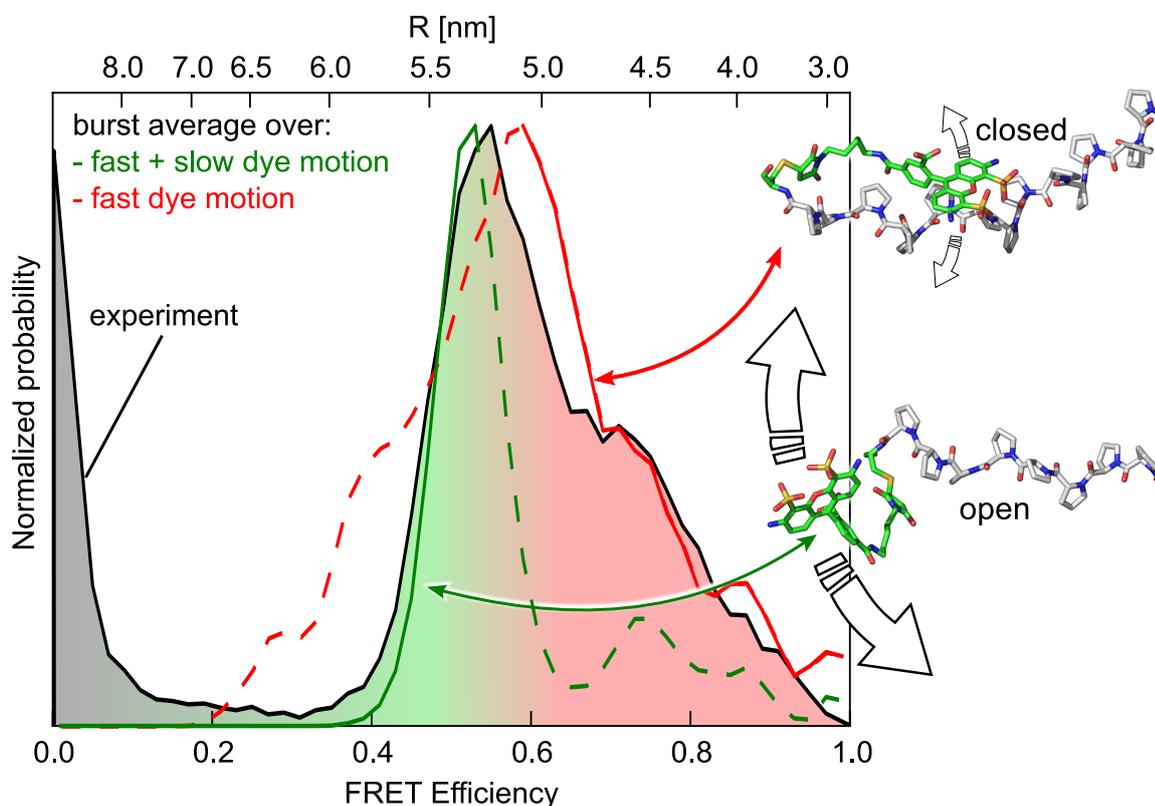
Figure 3.4: **Comparison of *in silico* FRET With Experiments.** Normalized FRET efficiency histograms of bursts from: isomer trajectories (green), single trajectories (red), experiment (black, with fill). The filled area indicates the agreement of experiment and simulation with isomer trajectories (green) and single trajectories (red) for the burst formation. The black area around the zero efficiency are poly-prolines with only the donor dye attached to in the experiment. Agreement on the high and low efficiency side is explained by the difference in dynamics of open and closed dye conformation shown on the right side.

Figure 3.5: **Combination of Experiment and Simulation to Reconstruct Distances.** Experiments measure FRET efficiency distributions or time-resolved FRET efficiency. MD simulations yield the dye orientation statistics and dynamics. Simulation and experiments are then combined to reconstruct the underlying distance distributions.
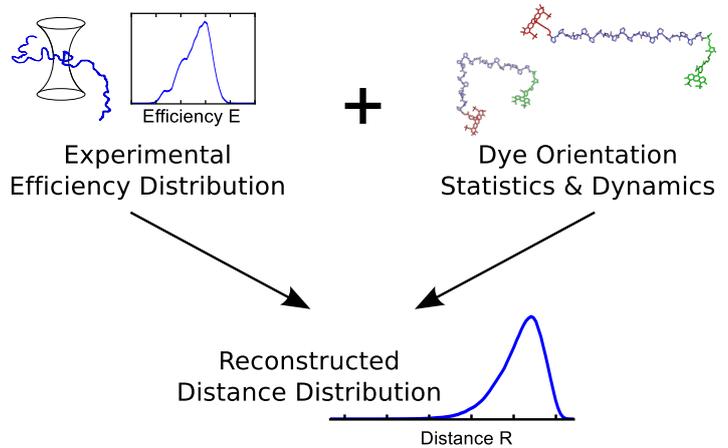
3.2.2 Combination of Simulations and Experiments for Distance Reconstruction

So far, it has been demonstrated that the *in silico* approach is suitable to calculate accurate efficiency distributions from simulation dye distances and orientations. However, the opposite direction — determining accurate distances from efficiencies — is after all the purpose of FRET experiments.

As in conventional FRET experiments, the efficiency distribution or time-resolved efficiency is measured, but instead of employing models like the isotropic $\kappa^2 = 2/3$ approximation in Eq. 2.24, the dye orientations from simulations will be used. Further, not only distance-independent averages of the dye orientation will be considered. In multiple steps, the dye orientation assumptions will be replaced by orientation statistics from simulations in the reconstruction process.

To assess the quality of reconstruction, the distance reconstruction is not only tested with the efficiency distribution from experiments, but also with the one from *in silico* FRET. In the latter case, the distance distribution is known from the simulations and therefore the comparison of distances serves as benchmark for the obtained accuracy in each step.

The Förster formula, Eq. 2.16, transfers an efficiency distribution $q(E)$ to a distance distribution $p(R)$ and vice versa and is therefore interpreted as transfer function (TF) $g(E, R)$. An exemplary TF is shown in Fig. 3.6. The reconstructed distance distribution with the isotropic model for the orientation is plotted in green in Fig. 3.6 and shows an error $\Delta x_{\mathrm{iso}} = 0.62\,\mathrm{nm}$ in peak location compared to the reference peak from simulations.
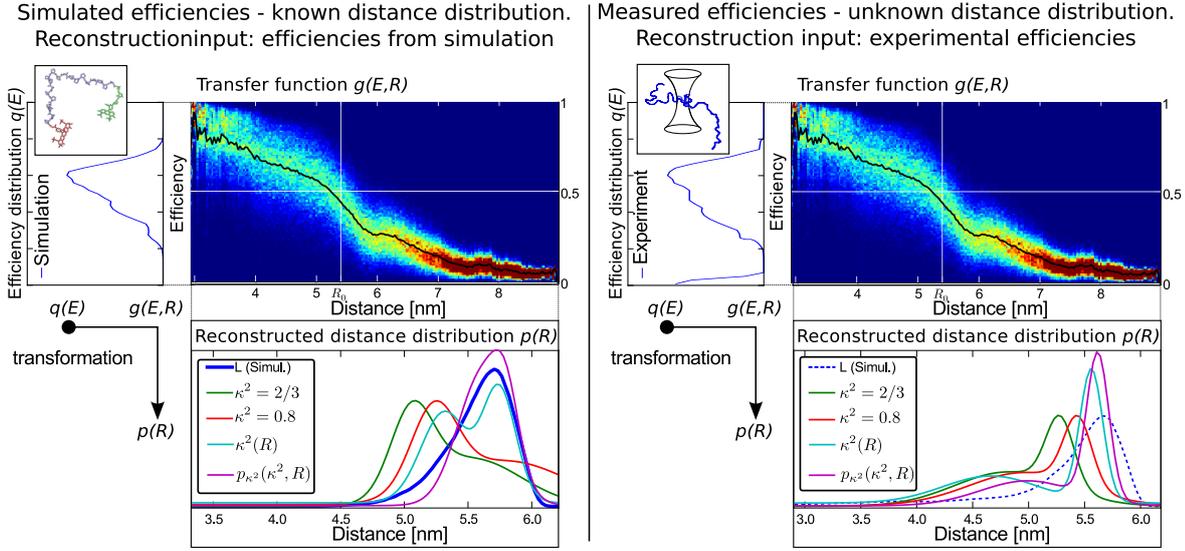
Figure 3.6: **Distance Reconstruction from Experiment and Simulations.** The distance distribution $p(R)$ is reconstructed by transformation of the reference efficiency distribution $q(E)$ via TFs $g(E, R)$ from simulations. Panel on each side: $q(E)$: top left, $g(E, R)$: top right, $p(R)$: bottom right, transformation path: bottom left. Left panel: Reconstruction with reference efficiencies from simulation and comparison to the known distance distribution. Right panel: Application of distance reconstruction using reference efficiencies from experiments. The reconstructed distance distributions $p(R)$ for different levels of approximation are color-coded as follows: blue is the distance distribution from simulations, green, red and cyan are reconstructed distributions from $\kappa^2 = 2/3$, improved κ^2 and distance-dependent $\kappa^2(R)$, respectively. Pink shows the distance reconstruction using TFs based on photo emission probability normalized (R, κ) samples (see Publication P8, Eq. P8.23 on page 152).

In the following steps, the dye orientation dependency in R_0 of the TF is modified by replacing assumptions on the dye orientation by information from my simulations. For each step, the key idea and replacement is given. The resulting TFs of each step are shown in Fig P8.12 on page 150 and the mathematical description of the steps can be found in section "Reconstructing Distance Distributions from FRET Efficiencies" on page 148 (Publication P8).

- **The dye orientations are not necessarily isotropic.** Therefore, the isotropic model assumption was dropped and $\langle \kappa^2 \rangle \neq 2/3$ in the Förster radius (Eq. 2.23) was replaced by the average from simulations $\langle \kappa^2 \rangle = 0.80$. This reduced the

error in peak location to $\Delta x_{\mathrm{noniso}} = 0.42\,\mathrm{nm}$ as seen by the shift (red in Fig. 3.6) towards the reference distance peak from simulations .

- **The dye orientation distribution is distance-dependent.** Therefore, correlation between κ^2 and R has to be taken into account. In this step, the assumption of distance independent orientation averages was dropped and the mean $\langle\kappa^2\rangle$ was replaced by a distance-dependent $\langle\kappa^2\rangle_R(R)$. As a result, also the Förster radius R_0 is distance-dependent. In the *in silico* efficiency reconstruction, the primary peak (cyan in Fig. 3.6) is close to the distances in the simulations.

- **Orientation averages of the dyes mask the conformational heterogeneity.** Therefore, the assumption is dropped that the dye dynamics averaged over the donor lifetime is converged to the ensemble average. Instead of averages, the distributions $p(R, \kappa^2)$ at each distance R were employed.

  The probability to measure a photon is not constant for each sample from $\{R, \kappa^2\}$ (see Fig. P8.14 on page 152) and for this reason, the samples are weighted according to their photon emission probability (Eq. P8.23 on page 152). Distributions of κ^2 (magenta in Fig. 3.6) showed the best agreement with the reference distances from the simulations. The peak position of the simulation distance distribution is accurately reproduced and compared to shape deviations of isotropic, non-isotropic and distance dependent averages of κ^2, only a slight underestimation of low distances is seen.

In summary, the reconstructed distances from *in silico* FRET efficiency distributions yield better agreement with the distances from simulation for decreasing approximation levels (Fig. 3.6, left side). When using the isotropic dye orientation model, the peak location has an error of $0.62\,\mathrm{nm}$ in contrast to an error which is within the histogram bin width of $0.035\,\mathrm{nm}$ for the reconstruction with distributions of κ^2. This shows that distances with improved accuracy are obtained by combination with simulations and reduced levels of dye orientation assumptions.

The approximation levels used for the reconstruction were also employed to calculate distances from experimental efficiencies (Fig. 3.6, right panel). Here, the true distance distribution is unknown and only the relative differences are observable.

In the experimental reconstruction, the location of the peak differs by up to $0.65\,\mathrm{nm}$ using the isotropic dye orientation model in contrast to the usage of κ^2 distributions. In the latter case, the peak positions differ less than $0.1\,\mathrm{nm}$ from the simulation distance peak. This corresponds to an error[1] reduction from $\approx 12\%$ using experiments alone

---

[1]relative error of the peak location

to 2% obtained by the combination of experiments and simulation. However, this is only a strong indication for an improved distance reconstruction since the true distance distribution in experiments is unknown.

In contrast to the poly-proline, many systems have even stronger interactions with the dyes or spatially restrict the conformational space of the dyes. For these cases, distance reconstruction from FRET experiments with higher accuracy is reached through the combination with simulations as demonstrated here.

3.3 Outlook

3.3.1 Application to Further Systems

The foundations laid in this thesis have established an approach to simulate FRET *in silico* and a method to combine experiments and simulations for accurate distance reconstruction. From this stage, a number of routes appear promising.

To that aim, one possibility are folding studies of small proteins. In fact, we have already established a collaboration with the group of Jörg Enderlein[1] who will study the folding of small proteins or protein domains like the helix-turn-helix (HTH) motif [132] with smFRET experiments.

For the measurement, a single protein labeled with two dyes will be encapsulated in a polymerosome nanocontainer to enable continuous measurements on a single molecule. The protein will then be unfolded by changing the guanidine hydrochloride (GdHCl) or urea concentration. Simultaneously, donor and acceptor photons are recorded. The dye distance measured by the FRET efficiencies will serve as a measure of the folding degree and folding dynamics.

While folding time scales itself are too long for simulations, they will provide the dye dynamics for accurate distance reconstruction. In simulations, the protein will be unfolded in urea [133] or using high temperatures. The system will be sampled at different stages of folding to calculate TFs from the dye conformations. The TFs will be employed in the distance reconstruction and thus will allow monitor the folding process of HTH molecules with high accuracy.

Further the computer-aided distance reconstructions will be employed in FRET experiments with extreme deviations from $\kappa^2 = 2/3$. Here, our focus are FRET measurements in the ribosome which will be performed in the group of Marina Rodnina[2].

---

[1]Universität Göttingen
[2]Max Planck Institute for Biophysical Chemistry, Göttingen

Dyes labeling substrates or sites in the ribosome are severely restricted in their orientational freedom. Therefore, the orientational sampling will be highly anisotropic. This anisotropicity will be estimated by simulations, and TFs from simulations will enable an accurate distance reconstruction as in the previous example.

3.3.2 Improving the Dye Parameterization

A general problem in simulations are the required force field parameters. Force fields not only contain the topology of a molecule but also reflect its hydrophobic or hydrophilic behavior. The balance between these two is crucial to correctly model the tendency of dyes to adsorb on the protein surface or to expose themselves to the solvent. This tendency is crucial for accurate simulation of the dye dynamics. While well tested force fields exist for proteins, DNA or RNA, dye parameterization is challenging for reasons explained in the following.

The fluorescent centers of dyes are often ring systems with a strong hydrophobicity. To increase solubility in experiments, highly charged groups (e.g. SO^3 groups in Fig. P8.1 on page 139) are attached to the fluorescent center of the dye. A related function of the soluble groups is to increase the probability of the dye and its linker to expose to the solvent instead of sticking to the protein surface. When in solution, more dye orientations are sampled and thus the dye orientation factor is closer to the isotropic case.

Ab initio parameterization of the partial charges (Coulomb interaction, Tab. 2.2) is commonly performed in vacuo [40]. However, for highly charged molecules and groups, the shielding of the solvent, not included in the parameterization in vacuo, is a problem leading to inaccurate partial charges. As a result, hydrophobicity of dyes might not not be reflected accurately in the simulations.

A further problem affecting the dye hydrophobicity is the delocalized electron system of dye, which is easily polarized in contrast to the localized charges of the soluble groups. This results in inducible dipoles which are not included in current force fields. Due to the uncertain hydrophobicity caused by the static and inducible charges, the dye mobility might be inaccurate in simulations.

To test the mobility of the dyes, the calculated and measured anisotropy decay was compared and found to be in agreement in Publication P8 (see page 144). Anisotropy decay is however an indirect measure of the dye hydrophobicity and ideally, the parameterization should be tested before extensive simulations are performed.

To this end, a direct comparison of the hydrophobicity in experiments and simulations is desirable. A test for the dye hydrophobicity is the difference in the free energy of solvation between two solvents. This difference can be measured and calculated.

Experimentally, partition coefficients of the dye between two non-mixing solvents measure the probability and therefore the free energy difference. In simulations, free energies will be calculated in a thermodynamic cycle between the two identical solvents. The comparison of the free energies will allow an assessment of the partial charge parameter quality. Further, multiple simulation cycles with different sets of partial charges can be performed for a systematic optimization.

Table 3.1: **Transition Density Coupling Methods**

Multipole coupling	TDC coupling

With the convoluted multipole moment [134]

$$Q_{LM}^{\mathrm{CP}} = \sum_{l+l'=L} (-1)^l \frac{4\pi(2L+1)}{(2l+1)(2l'+1)}$$

$$\times \sum_{m+m'=M} \frac{\xi_{lm}\xi_{l'm'}}{\xi_{LM}} q_{lm}^{\mathrm{C}} q_{l'm'}^{\mathrm{P}} \qquad (3.2)$$

and $\xi_{lm} = \sqrt{\frac{2l+1}{4\pi(l-m)!(l+m)!}}$, the potential can be written as

$$V_{\mathrm{mp}} = \sum_{L,M} \frac{4\pi}{2L+1} Q_{LM}^{\mathrm{CP}} \frac{Y_{LM}(\hat{\mathbf{d}})}{|d|^{L+1}} \; . \qquad (3.3)$$

Y_{LM} are spherical harmonics, q_{lm}^{C} and $q_{l'm'}^{\mathrm{P}}$ the multipole moments of the two transition densities respectively. For an order $L_{\mathrm{max}} = 2$, dipole-dipole coupling as in the Förster formula Eq. 2.16 is obtained.
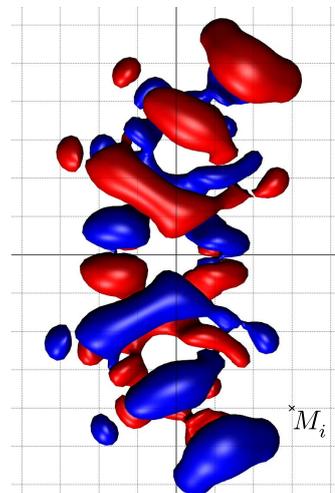
The transition density cube (TDC) method calculates the direct coupling between the two transition densities. The coupling potential is expressed as

$$V_{\mathrm{coul}} = \sum_{i,j} \frac{M_i \cdot M_j}{|d_{ij}|} \; , \qquad (3.4)$$

with M_i and M_j as the transition density of each dye, discretized on a grid.
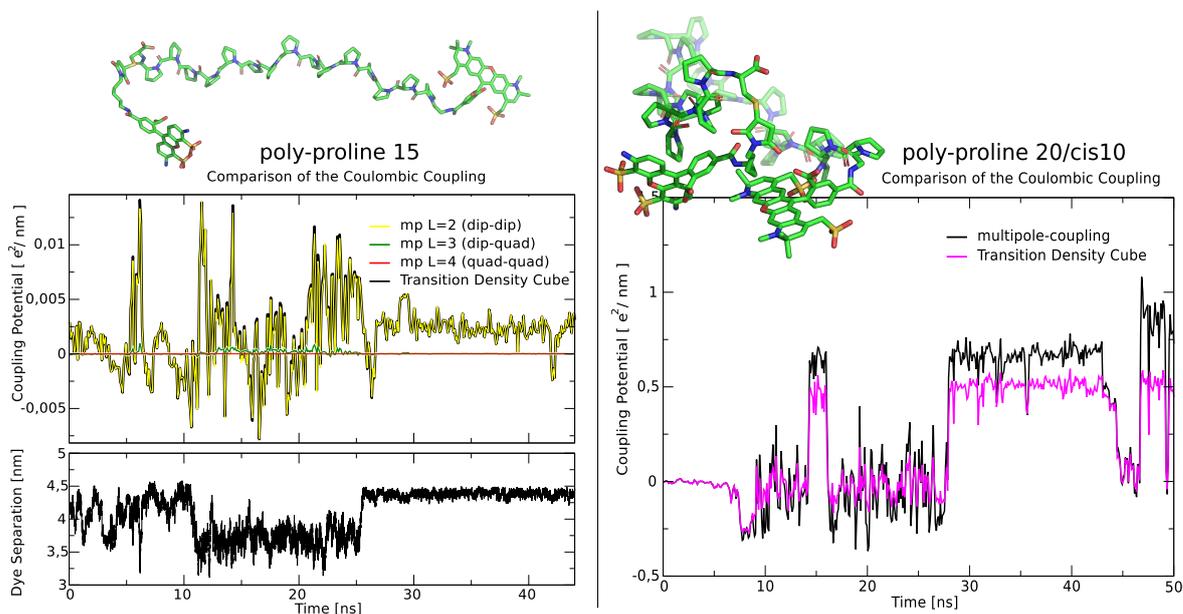
Figure 3.7: **Comparison of Multipole Coupling and Transition Density Cube Coupling.** On the left side, coupling potentials for an all-trans poly-proline 15 system are shown. Multipole contributions of dipole-dipole (yellow), dipole-quadrupole (green), quadrupole-quadrupole (red) and TDC (black) coupling are plotted. Below the dye-to-dye distance is shown. The right side shows poly-proline 20 with a central cis-bond leading to a small dye-to-dye distance. Multipole coupling up to fourth order (quadrupole-quadrupole, dipole-octopole) and TDC coupling is shown.

3.3.3 Transition Density Coupling Beyond the Ideal Dipole Approximation

A central assumption of the Förster approximation (Eq. 2.21) is the sufficient description of the electronic coupling between the two dyes by dipole-dipole interactions. This approximation, commonly referred to as ideal dipole approximation (IDA), is valid for dyes with a strong dipolar transition density and for large distances compared to the extension of the two electronic systems. When measuring small distances or using dyes with non negligible higher transition density moments, the IDA limits the accuracy.

In order to take higher moments into account, the dipole-dipole coupling has to be replaced by a multipole expansion including higher coupling moments. Table 3.1

(left column) summarizes the coupling of two transition densities using convoluted multipole moments.

Preliminary tests of the influence were performed for the system in Publication P8. Figure 3.7 (left side) shows the distance and the Coulombic coupling potential between both dyes for the system used in Publication P8. For the two employed Alexa dyes, the effect of higher coupling moments (green and red) than dipole-dipole coupling (yellow) is small. Thus the dipole-dipole coupling is close to the reference calculation with the TDC method which will be discussed below. This dominating behavior of the dipole-dipole coupling term is not surprising for the two Alexa dyes since both exhibit a strong dipolar transition density. However, e.g. for FRET between nano crystals, higher coupling moments like the dipole-quadrupole coupling is not negligible [135].

For small distances, the extension of the electronic system is large compared to the distance. As a result, higher moments L of the multipole expansion have to be included due to the higher inverse power in the distance dependency of the potential (see Eq. 3.3 in Table 3.1). Evaluation of the sum in Eq. 3.3 is however not feasible for large coupling moments L. Thus, instead of using a multipole expansion of the transition densities, the direct coupling can be calculated via the TDC method [87].

Table 3.1 (right column) summarizes the key idea of the TDC method: The transition density of both dyes is discretized on a grid and the coupling between the grid points is calculated. For first tests, the poly-proline 20 simulations with central cis-bond were used (see Fig. 3.7, right side). Due to the central cis, the inter dye distances are below 2 nm. In Fig. 3.7, a fourth order multipole coupling (black) is compared with the TDC method (pink). The error of the multipole coupling potential at these short distances is up to 40% for the test trajectory. Since the squared coupling potential is proportional to the FRET rate, this error is not negligible.

In addition, RET is not accurately described by Försters theory at short distances, not only due to the IDA but also due to the assumption of weak coupling between the dye orbitals. Therefore, e.g. coupling by Dexter electron transfer should be examined as well in the future.

Finally, also the surrounding medium influences the coupling between the transition densities as demonstrated by Curutchet *et al.* [136]. Therefore, the effect of the dielectric constant at optical frequencies $\epsilon(\lambda)$ should also be investigated in future studies.

4

Simulated Adsorption of Biomolecules on Gold Surfaces

- O. Cohavi, S. Corni, F. D. Rienzo, R. D. Felice, K.-E. Gottschalk, M. Hoefling, D. Kokh, E. Molinari, G. Schreiber, A. Vaskevich and R. C. Wade. Protein-Surface Interactions: Challenging Experiments and Computations. *J. Mol. Recognit.*, 23(3):259–262, **2010**. (Publication P2).

- D. B. Kokh, S. Corni, P. J. Winn, M. Hoefling, K.-E. Gottschalk and R. C. Wade. ProMetCS: an Atomistic Force Field for Modeling Protein-Metal Surface Interactions in a Continuum Aqueous Solvent. *J. Chem. Theory Comput.*, 6(5):1753–1768, **2010**. (Publication P4).

- M. Hoefling, F. Iori, S. Corni and K.-E. Gottschalk. The Conformations of Amino Acids on a Gold(111) Surface. *ChemPhysChem*, 11(8):1763–1767, **2010**. (Publication P5).

- M. Hoefling, F. Iori, S. Corni and K.-E. Gottschalk. Interaction of Amino Acids With the Au(111) Surface: Adsorption Free Energies From Molecular Dynamics Simulations. *Langmuir*, 26(11):8347–8351, **2010**. (Publication P7).

- M. Hoefling, S. Monti, S. Corni and K.-E. Gottschalk. Interaction of β-Sheet Folds With a Gold Surface. *PLoS One*, 6(6):e20925, **2011**. (Publication P9).

Individual Contributions

- **Publication P2:** All authors contributed to this workshop review publication according to their field. Kay E Gottschalk and myself gave the overview of atomistic simulations of biomolecules near inorganic surfaces.

- **Publication P4:** Daria B Kokh, Peter J Winn and Rebecca C Wade established the BD protocol and the additional potentials for the surface interactions of amino acids, which were then implemented by Daria B Kokh and Peter J Winn in the University of Houston Brownian dynamics (UHBD) program. The MD simulation of water and the analysis of the PMF for individual water molecules were done by Stefano Corni. MD simulations and the analysis of the PMF and conformations of the amino acids near the surface were performed by myself. Kay E Gottschalk aided in the system setup and interpretation of the amino acid simulations. All authors interpreted the results and contributed to the manuscript.

- **Publication P5:** MD simulations of amino acids and density functional theory (DFT) fragments, used for parameterization, on gold surfaces were done by myself. Kay E Gottschalk and myself analyzed the resulting trajectories and performed the cluster analysis of conformations on the surface. The DFT conformations were calculated by Francesco Iori and Stefano Corni. The manuscript was written by Kay E Gottschalk and myself with the aid of the other authors.

- **Publication P7:** Kay E Gottschalk and myself designed the simulation setup and the free energy calculations. Simulations of amino acids and the analysis of the adsorption free energy were performed by myself. Kay E Gottschalk analyzed the correlation of the adsorption free energies with known physiochemical properties. The analysis of the adsorption pathway was done by myself and interpreted by all authors. The manuscript was written by Kay E Gottschalk, Stefano Corni and myself.

- **Publication P9:** Simulations of the oligopeptide were suggested by Susanna Monti and performed and analyzed by Stefano Corni. The MD simulations of fibronectin adsorption, the analysis of the adsorption pattern and the analysis of structural changes during adsorption were done by myself. Kay E Gottschalk interpreted the physiological impact of adsorbed conformations and aided in the design of the fibronectin studies. The manuscript was written by all authors which also interpreted the results.

4.1 Introduction

The following Section 4.1.1 summarizes examples for the importance of biomolecular interactions with inorganic surfaces *in vivo* as highlighted in Chapter 1.2. Biocombinatorial approaches to select molecules with high affinities and the drawbacks of biocombinatorial techniques are discussed in Section 4.1.2. Computational approaches to aid in the design of biomolecular surface interactions are proposed in Section 4.1.3.

4.1.1 Inorganic Surfaces in Biomolecular Systems

The interactions between proteins and inorganic surfaces are of increasing importance to understand processes in many natural systems. One example is hard-tissue growth, e.g. of dentin or bone, where collagen mineralized tissue (CMT) is formed. The biomineralization in CMTs relies on the interactions of the protein collagen with inorganic surfaces to control deposition of apatite crystals [137].

A further example is cell-adhesion on inorganic surfaces [138] which depends on the surface properties apart from the surface topology. Cell adhesion molecules like integrins (see Chapter 5) link the cell to the surface via extracellular matrix (ECM) proteins adsorbed on the surface.

Some cells contain thermal hysteresis proteins (THPs) to prevent freezing. Thereby, the THPs adsorb to the surface of ice crystal nuclei leading to a hysteresis between the reduced freezing temperature and the unchanged melting temperature of the solution [139]. The nature of freezing inhibition and the responsible interactions of THPs with the ice crystal surface are still under debate [139].

Interactions of biomolecules with inorganic surfaces are also pivotal for biotechnological applications. The design of bio-sensing devices and bio-compatible implants requires a fine tuning of biomolecule-surface interactions [15, 16]. A further example is nano-toxicity, which not only depends on the size and shape of nano-particles but also on the type of surface exhibited [17, 18].

For all given examples, it is fundamental to understand the design principles of biomolecular adsorption on inorganic surfaces.

What is the optimal surface to avoid protein binding or to facilitate adsorption? How can we design proteins specifically binding to distinct surfaces? Experimentally, biocombinatorial techniques allow the screening for interactions and are therefore discussed in the following section.

4.1.2 Design of Interactions Between Biomolecules and Inorganic Surfaces

Biocombinatorial approaches like phage display (PD) selected small peptides with a specific binding affinity, e.g. to gallium arsenide (GaAs) [140] or gold surfaces [22]. In PD, a recombinant library of peptide fragments is displayed on the surface of bacteriophages. After binding to the surface and removal of the non-binding species, the phages binding to the surface are amplified by bacteria. By repetition of these steps, an enrichment of phages with the high affinity peptides on the surface is achieved.

Despite its success in selecting peptides binding specifically to the target surface [20], the PD technique has several drawbacks:

- PD is limited to small proteins and peptide libraries. Design of custom libraries, e.g. specific protein mutations, is expensive or not feasible at all.

- Amino acids sequences are encoded as codons (nucleotide triplets) on the messenger ribonucleic acid (mRNA). To generate peptide libraries, the nucleotide sequence of the mRNA is randomized. The number of codons encoding the same amino acid varies, e.g. six different codons encode leucine while tryptophane is encoded by only one nucleotide triplet. Therefore, peptide libraries are biased in their amino acid composition [21].

- The influence of the phage on the surface binding affinity of peptides and proteins is elusive. All phage-encoded proteins are fusion proteins, which might limit the activity or accessibility for protein binding.

- Reasons for high or low affinity and the mechanism of binding are not revealed by PD experiments [19].

In summary, biocombinatorial approaches allow the testing of available libraries for components with high affinity. However, they do not explain the nature of binding. Moreover, affinity tests of biomolecules not available in libraries is often very expensive if feasible at all.

Further examples to experimentally study the adsorption of single biomolecules on surfaces are atomic force microscopy (AFM) and magnetic tweezers while adsorption rates in the ensemble can be monitored in surface plasmon resonance (SPR) experiments.

Single molecule techniques are not suitable for library screening but excel in the study of the responsible interactions. As an example, electrically controlled adsorption of DNA to gold surfaces has been studied by AFM experiments recently [141, 142].

4.1.3 Computational Methods to Probe Interactions Between Biomolecules and Inorganic Surfaces

Monitoring the adsorption process in experiments is challenging (Publication P2) since the adsorption to the surface is often very strong. As a consequence, only on-rates are measurable in SPR experiments. Therefore, the combination with simulations not only provides an atomistic model of the adsorption process but in principle also allows calculation e.g. of desorption rates, difficult to access in experiments [143].

For an accurate modeling of adsorption, the balance of interactions between solvent, solute and surface is crucial [144]. In addition, the flexibility of the solute plays an important role during adsorption or desorption [145]. Both aspects are accurately described by atomistic MD simulations in explicit solvent, which I therefore employed to contribute to the understanding of biomolecular surface adsorption.

Besides understanding the details of the interaction balance, methods allowing to select promising species with high or low affinity for surfaces are key to biotechnological applications. Due to the disadvantages of the currently available combinatorial techniques described in the previous Section 4.1.2, alternatives were developed in this work. In the following, I will show that computational methods may contribute in analogy to the role of *in silico* docking in the field of drug discovery.

In silico docking selects promising compounds with desired properties from a library. This docking of compounds to protein surfaces led to a revolution in drug discovery [23–25]. The main advantage in comparison to PD is that arbitrary libraries can be tested. Unbiased peptide libraries are easily produced and protein modifications, e.g. only at distinct sites, are screenable. A further advantage is that the computational power of supercomputers is still growing exponentially from which this technique will automatically benefit in the future.

For these reasons, the availability of a computational docking toolkit to estimate interactions between proteins and inorganic surfaces would be an optimal starting point for technological applications. Design, implementation and experimental verification of such a docking toolkit was one of the goals of the PROSURF[1] project, which will be described below.

An overview of this project and related computational studies from other groups concerning interactions of peptides with inorganic surfaces can be found in Ref. 146.

---

[1]PROSURF, European Union FP6 NEST project no. 028331 `http://www.s3.infm.it/prosurf/`

Figure 4.1: **Work Flow of the PROSURF Project.** The contributions of different computational groups are marked by color:

- Modena[a], *ab-initio* parameterization: blue

- Munich, MD simulations: red

- Heidelberg[b], BD docking: yellow.

Steps with experimental validation (Rehovot[c], SPR) are shown with dashed green frame.

_____

[a]Group of Stefano Corni, Theoretical Nanoscience, University of Modena, Italy

[b]Group of Rebecca Wade, Heidelberg Institute for Theoretical Studies (HITS), Germany

[c]Groups of Gideon Schreiber and Israel Rubinstein, Weizmann Institute of Science, Rehovot, Israel

Derive force field parameters (based on ab-initio methods)

Binding free-energy calculations with MD for different peptides

validate results vs. experiment

Fit empirical free-energy scoring function in docking to MD/experimental results

Implement new scoring function and search algorithm in docking code

Validate docking results vs. MD/experiment

Validated protein-surface docking tool
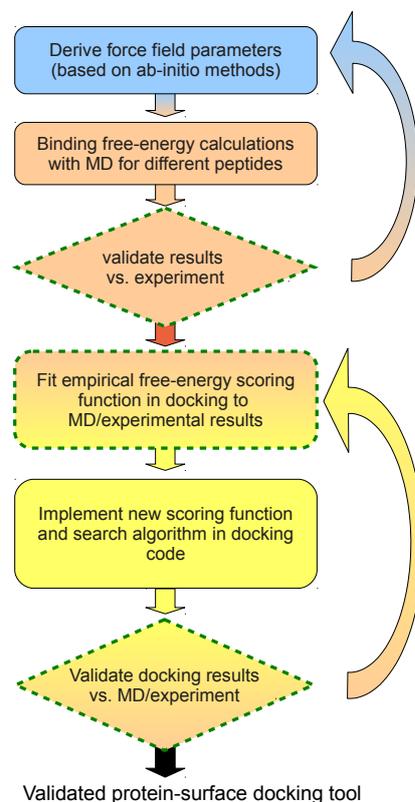
4.2 Results and Discussion

4.2.1 The PROSURF Project

The work presented in Publications P2, P4, P5, P7 and P9 is part of the PROSURF project. Aim of PROSURF was the development of a computational toolkit for docking of proteins on gold surfaces.

Figure 4.1 sketches the work flow towards a validated surface docking toolkit and the contribution of each group. The required steps and the results will be summarized in the following paragraphs.
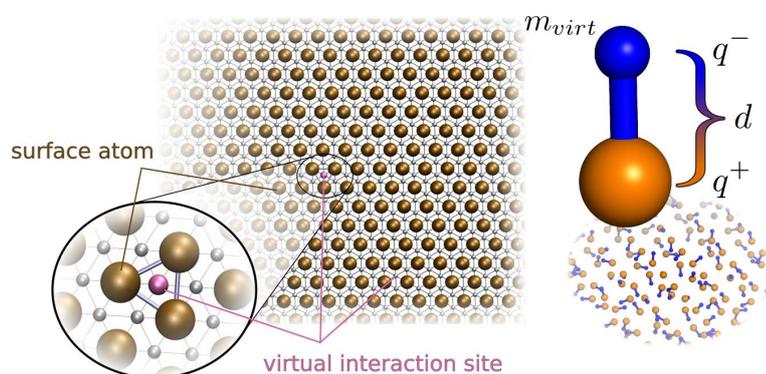
Figure 4.2: **Parameterization of Gold(111) Surfaces.** On the left side, the location of the virtual interaction sites, responsible for the van-der-Waals interactions of the surface, is shown. The right side shows the modeling of the metal polarizability through freely rotatable dipoles.

4.2.2 Ab inito Force Field Parameterization

Before adsorption simulations with MD were performed, a force field modeling the interactions of proteins with the inorganic gold surface had to be developed. This task was performed by our partners in Modena[1] using a combination of *ab-initio* DFT quantum chemistry calculations and experimentally known properties. The additional GolP force field parameters [26, 27] extend the OPLS/AA force field by interactions with gold surface and were the basis for my MD simulations.

One of the advantages of the developed GolP force field is that all interactions are described by standard non-bonded MD potentials (see Tab. 2.2 in Chapter 2). Thereby, employing standard MD packages like GROMACS or "Not just Another Molecular Dynamics program" (NAMD) takes advantage of available optimizations for these potentials and the interactions are easy to implement.

Technically, each of the two long-range potentials — Coulomb and Lennard-Jones from Tab. 2.2 — is used to mimic parts of the interaction:

- As shown in Fig. 4.2 (left), chemisorption and van-der-Waals interactions with the surface are modeled by Lennard-Jones potentials on virtual interaction sites. This special treatment is required, since atomic adsorption on gold is centered on the hollow sites rather than on top of the gold atom. Each surface gold atom is represented by two virtual interaction sites on the hollow sites.

- The polarization of the metal surface is described by static dipoles which are allowed to freely rotate around the gold atoms (Fig. 4.2, right side) [28]. For the electrostatic interaction, standard Coulomb potentials are employed.

---

[1]Group of Stefano Corni, Theoretical Nanoscience, University of Modena, Italy

For efficiency reasons no additional potentials to describe bond formation are included. As a consequence, only the first step of chemisorption is described by the force field. Bond formation, e.g. of cysteines with the surface is therefore not described. This is a general limitation of force fields which is not solved easily without additional computational cost (see Section 2.1.4). In the following step, the GolP force field was applied in MD simulations.

4.2.3 Molecular Dynamics of Peptide Adsorption

Screening of large protein libraries or sampling of the adsorbed conformations of a single protein are not feasible with MD. Therefore, BD simulations were used for the actual docking procedure (see Fig. 4.1). While BD interactions for protein-protein interactions exist, parameters for the interactions with surfaces had to be developed first.

To test the implemented BD interactions, the adsorption of amino acids — the building blocks of proteins — was compared to MD simulations. The accuracy of the simplified treatment in BD was verified by comparison to the corresponding adsorption free energy landscape from MD simulations. Beyond verification of the BD simulations, the correlation of the calculated free energy of adsorption from MD simulations with known physiochemical properties or available experimental data was tested.

In the following, the calculation of the amino acid adsorption free energy from my MD simulations is summarized. The surface distance was chosen as reaction coordinate λ (Section 2.3.1). Each amino acid was simulated with constraints at different distances.

The obtained mean constraint forces from multiple simulations (over 100 per amino acid) were integrated to obtain the free energy profile according to Eq. 2.14 in Section 2.3.1[1]. The PMF from infinite distance to adsorption was reconstructed by performing simulations at increasing distances until no mean force and thus interaction was recorded. The obtained PMF was interpreted as one-dimensional adsorption free energy landscape, perpendicular to the surface.

The adsorption free energies for all 20 natural amino acids are summarized in Tab. P7.1 on page 133 and range from 17.5 kJ/mol for glutamate to 44.2 kJ/mol for tyrosine. In general, the highest interaction free energies are observed for aromatic residues. Sulfur-containing, positively charged, aliphatic and negatively charged amino acids show reduced interaction energies in this order (Fig. P7.1 (c) on page 132).

---

[1]The principal idea of constraint biased force analysis [81] to calculate free energies of adsorption is shown in Fig. P7.1 (a) and (b) on page 132.
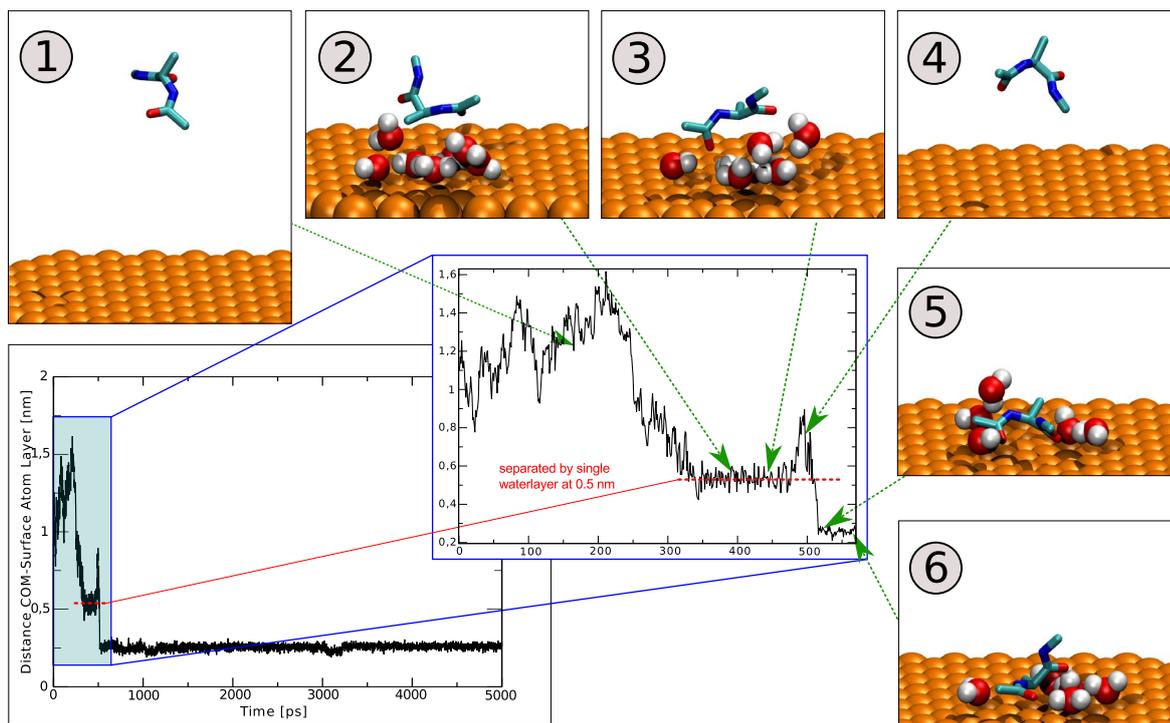
Figure 4.3: **Alanine Adsorption on a Gold(111) Surface.** The plot shows the COM distance of alanine to the surface atom layer. The insets show the different stages connected by arrows to the corresponding trajectory point. 1 & 4: diffusion, 2 & 3: association, 5 & 6: adsorption. No desorption was observed in the remaining simulation time after step 5.

The adsorption free energy profiles also allowed the identification of different stages in the adsorption process. As shown in Fig. P7.3 on page 134, diffusion dominates until penetration of the second water layer occurs. This stage is followed by the association step in which parts of the amino acid have penetrated the first water layer and are in contact with the surface. The steepest change in free energy was observed upon surface binding, which maximizes the contact of amino acid and surface. Figure 4.3 shows an example trajectory with associated snapshots of each stage.

The free energy of adsorption is one order of magnitude larger than the thermal energy at room temperature. Therefore desorption events are very unlikely within simulation time. In fact, no spontaneous desorption was observed in my simulations.

Further, the correlation of the interaction free energies with available experimental results and other known chemical properties was investigated (Fig. P7.2 on page 134).

Most notably is the correlation of the adsorption free energies from simulations with the surface coverage in PD experiments [147].

In addition, a strong correlation of the adsorption free energies with the propensity of each amino acid type to form β-sheets was found. The propensity to form α-helices was uncorrelated with the adsorption free energies. This observed adsorption preference of β-sheets was one of the reasons to continue with the adsorption study of β-sheet folds as described later in Section 4.2.5 and in Publication P9.

4.2.4 Brownian Dynamics Docking

The free energy profiles (Publication P7) and geometries (Publication P3) of single amino acids were used in cooperation with our partners in Heidelberg[1] to implement BD protein-surface interactions.

The interactions in BD simulations of the protein and amino acid with the gold surface are decomposed into three components:

- Lennard-Jones interaction in analogy to the interactions in MD simulations (see Fig. 4.2, left).

- Protein metal electrostatic interactions for the charged amino acids modeled by image charges.

- The metal desolvation term required for implicit solvent simulations was described by modulation of the effective dielectric constant ϵ, parameterized by MD simulations in explicit solvent.

The implementation in the existing UHBD simulation software package is described in Publication P4 and was the focus of our partners in Heidelberg[1].

The input conformations were chosen from MD simulations, since the protein in BD is treated as a rigid body. For each amino acid, representative structures of the cluster(s) with the highest population(s) in the MD (see Publication P5) were selected for the BD simulation ensemble.

In a next step, the free energy of adsorption was calculated for each conformation. Comparison with the free energies from MD simulations shows a good agreement (Fig. P4.10 on page 114).

The influence of the rigid treatment in BD on the free energy was probed by testing multiple conformations of the same amino acid. In most cases, the differences between individual conformations were smaller than 5 kJ/mol, whereas for glycine,

---

[1]Group of Rebecca Wade, Heidelberg Institute for Theoretical Studies (HITS), Germany

isoleucine, cysteine and histidine, differences up to 16 kJ/mol were observed (Fig. P4.10 on page 114).

To verify this computational multi scale approach from *ab initio* DFT calculations via MD to BD docking, the agreement of BD simulations and experiment was tested. The adsorption of amino acids is difficult to measure. Therefore, SPR experiments with entire proteins were performed by an experimental collaboration[1]and the comparison to BD simulations was performed in a follow up study of the group in Heidelberg[2] and Rehovot[1].

Gold(111) surfaces and the identical proteins were used in experiments and BD simulations. The experimental adsorption strength show a good agreement with the adsorption strength revealed in simulations [29].

In summary, this agreement with experimental results suggests that our simulation approach is suitable to study protein surface adsorption. Moreover, it renders BD docking with the implemented surface interactions a valuable tool to rapidly screen for the surface binding potential of proteins and other biomolecules.

4.2.5 Molecular Dynamics Simulations of Protein Adsorption on Gold Surfaces

The correlation of free energies in simulations with the β-sheet propensity of each amino acid studied in Publication P7 suggests that these structural elements support adsorption to surfaces. Therefore, the adsorption of β-sheet rich RAD16II fibers and the fibronectin domains 9 & 10 on the surface was simulated in Publication P9 as a first step towards simulations of larger molecules.

In the protein, amino acids are restricted in their conformations and cooperativity is required during adsorption in contrast to single amino acids. Therefore, the side chains of individual amino acids on the protein surface might contribute to differently to each adsorption step, and not just act as the sum of individual interactions.

The study of β-sheet rich proteins was focused in particular on the formation of initial contacts during the adsorption process. Interestingly, I observed a special role of the positively charged amino acids with long side chains, arginine and lysine, in the initial anchoring of the protein to the surface. The anchored protein conformations then drive further adsorption to the surface.

---

[1]Groups of Gideon Schreiber and Israel Rubinstein, Weizmann Institute of Science, Rehovot, Israel

[2]Group of Rebecca Wade, Heidelberg Institute for Theoretical Studies (HITS), Germany

The special behavior of positively charged amino acids reveals an interesting starting point to electrically control adsorption kinetics of proteins with arginine and lysine residues on the surface. In the simulations performed, the surface is neutral but in principle the surface charge is adjustable in simulations. Experimentally, the surface charge in metals can be controlled as well, e.g. in AFM setups [141, 142]. Therefore, single molecule studies combining experiments and simulations may lead to a consensus picture of the adsorption process in the future.

4.3 Outlook

With a tested force field available, adsorption studies of large molecules such as fibronectin (Publication P9) are now possible. These studies will provide insights into the stability of proteins or protein subunits close to surfaces.

The stabilization of certain conformations is important for diseases like Alzheimer or bovine spongiform encephalopathy (BSE), where fibrillous β-sheet forming structures play a key role. Surfaces in the organism might be responsible for the initial formation of β-sheet like structures and thereby related to the diseases. Further, the interactions with surfaces might aid in research of diseases like BSE by inducing certain protein folds *in vitro*.

For the design of protein surface interactions, simulations now allow the calculation of preferred adsorption geometries and conformational changes on the surface. These properties will help answering the question if an enzyme is still active on the surface. Moreover, mutations preventing the inactivation of an enzyme after the adsorption are testable.

Although an MD simulation protocol for interactions between proteins and gold surfaces was tested, further steps are required to obtain the same accuracy in protein-surface adsorption simulations as in protein-protein simulations. A crucial point is the lack of experiments which can be accurately modeled by MD simulations to benchmark force fields [146]. The availability of such experimental data not only would aid in the parameterization but could also be applied to benchmark the available force fields for surface interaction studies. In this context single molecules studies combining AFM experiments and MD simulations are promising.

Further Projects

Mechanical Signal Transduction through Transmembrane Proteins
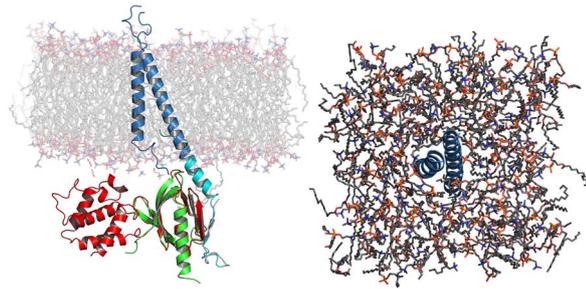
Aims:

- Classify available structures and models of the integrin TM domain for future atomistic simulations.

- Development of a simulation protocol for TM protein insertion into previously equilibrated bilayers with minimal disturbance of the bilayer equilibrium.

Association Mechanism of Transient Protein-Protein Complexes
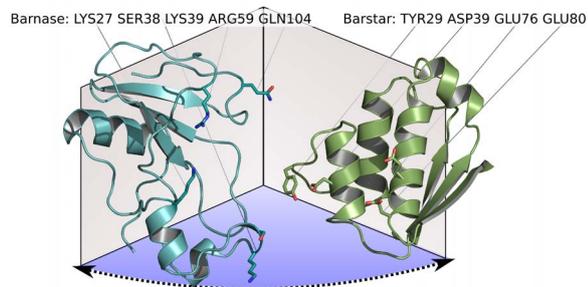
Aims:

- Calculate the free energy landscape of the complex formation and understand the impact on kinetics.

- Test the effect of mutations and salt concentration on the association pathways.

GromPy: Python Interface for GROMACS

Aim:

- Development of a GROMACS simulation package interface for the rapid prototyping programming language Python.

5

Mechanical Signal Transduction through Transmembrane Proteins

- M. Hoefling, H. Kessler and K.-E. Gottschalk. The Transmembrane Structure of Integrin αIIbβ3: Significance for Signal Transduction. *Angew. Chem., Int. Ed.*, 48(36):6590–6593, **2009**. (Publication P1).

- M. G. Wolf, M. Hoefling, C. Aponte-Santamaría, H. Grubmüller and G. Groenhof. g_membed: Efficient Insertion of a Membrane Protein Into an Equilibrated Lipid Bilayer With Minimal Perturbation. *J. Comput. Chem.*, 31(11):2169–2174, **2010**. (Publication P6).

Individual Contributions

- **Publications P1, P1a:** The manuscript was written by Kay E Gottschalk, Horst Kessler and myself. Kay E Gottschalk and myself proposed the force induced straightening model for integrin activation.

- **Publication P6:** A first prototype of the membrane insertion procedure was written by myself in Python using GromPy (see Manuscript M1) while the g_membed tool was implemented and tested by Maarten G Wolf. The manuscript was mainly written by Maarten G Wolf with the aid of all other authors. The Integrin αIIbβ3 system setup was contributed from myself, the yeast Aquaporin from Camilo Aponte-Santamaria and the β-Barrel Platform and Reaction Center from Maarten G Wolf. All tests with these systems were performed and analyzed by Maarten G Wolf.

5.1 Introduction

In multicellular organisms, controlled adhesion of cells to surfaces or to the ECM is pivotal. For this task, a specialized class of heterodimeric adhesion glycoprotein receptors, known as integrins [148], is found in cells.

Integrins provide a stable linkage between the ECM and the intracellular actin cytoskeleton of the cell. The affinity of the integrin ectodomain, e.g. to vascular cell adhesion molecule (VCAM) 1 is controlled from the cytoskeleton [149] (inside-out signal). The reverse signaling pathway through the membrane is found when binding to the ECM changes shape and arrangement of the cytoskeleton [150] (outside-in signal). Both signals are transduced through the TM domains of the integrin heterodimer (Fig. P1.1 on page 79).

Signal cascades starting from integrins are involved in a multitude of physiological processes like apoptosis, cell adhesion and migration or cancer metastasis [151]. Therefore, an understanding of the signal transduction on the atomistic level would not only lead to additional insight into the physiology of multicellular organisms but also aid in the discovery of integrin targeting drugs.

In signal transduction, the TM region of integrins plays a key role by relaying the signal from the inside to the outside of the cell and vice versa [152, 153]. It is known for example, that integrins are permanently activated after mutation of the GFFKR[1] motif in the TM α-domain [154]. How does this mutation affect the stability of the TM domain complex? Here, I aim at atomistic simulations to aid in answering the question.

In the following, I addressed two important problems to improve the accuracy and feasibility of atomistic simulations of the integrin TM domains. First, an accurate model or structure of the TM domains is required. Therefore, I evaluated available computational models and structural models from NMR. Second, the TM structure was inserted into an equilibrated bilayer.

TM structures are difficult to resolve by crystallographic techniques. It is known that the TM part of the integrin αIIbβ3 complex consists of two helices. On this basis and using experimentally known properties, e.g. from cross-link studies, computational models of the structure [155] and the clustering of multiple integrins [156] were proposed.
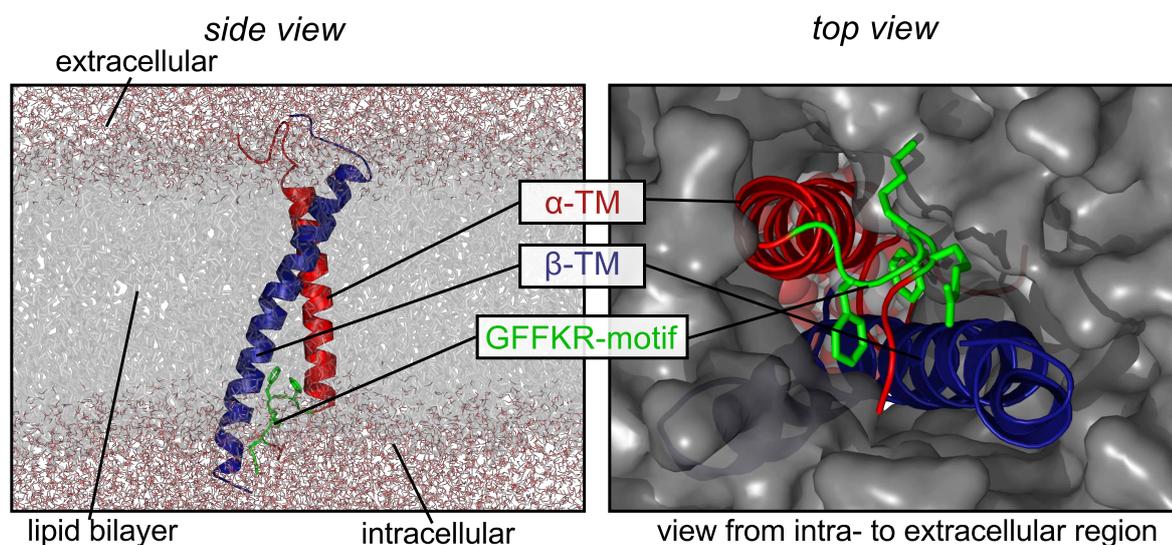
Figure 5.1: **Integrin Embedded in a DOPC Bilayer via g_membed.** Left panel: side view. Right panel: view from intra to extracellular. The DOPC bilayer is shown as SAS. Extracellular waters are shown in sphere representation in the background of the hole.

5.2 Results and Discussion

Recently, the first structures of the integrin αIIbβ3 TM domain were resolved by Lau *et al.* Refs. 157–159 using NMR experiments. The proposed structure from computational models from Gottschalk *et al.* [155] are in good agreement with the structures from NMR (see Fig. P1.4 on page 81).

A remarkable feature is the structure of the GFFKR motif which binds the helix of the α-subunit with two salt bridges to the β-subunit. This structure might act as a force sensor inducing a conformational change of the two helices upon external force (Fig. P1.2 on page 80). In summary, the new NMR structures are a big step forward to test and propose new models for the TM signal transduction. Moreover, these structures are the structural basis for atomistic simulations of integrin activation.

Once the structure of TM integrin domains is chosen, the two helices had to be inserted into an equilibrated lipid bilayer to perform atomistic simulations including the membrane and water. Thereby, the distortion of the bilayer itself and the bilayer

---

[1]amino acid single letter code

water interface should be minimized to avoid long and expensive equilibration runs after the insertion.

In Publication P6, a method for efficient membrane insertion of TM proteins was tested and implemented. The key idea of this insertion protocol is to laterally shrink the protein (see Fig. P6.1 on page 126) and only to remove lipids, whose van-der-Waals radii are in contact with the shrunken protein. The protein is then slowly grown to its original lateral extension performing MD or energy minimization (EM) steps during the growth process. Figure 5.1 shows the resulting integrin TM domain embedded in a DOPC bilayer from the side and top view.

It was also examined how fast the initial area per lipid (APL) is re-established in the simulation after insertion of the protein with this protocol. Relaxation times shorter than 1 ns are found for all tested systems (see Publication P6). By carefully choosing the initial protein scaling factor, the APL is not disturbed after insertion at all (see Tab. P6.2 on page 128).

With the structure of the integrin TM domain available and an established efficient insertion protocol into bilayers (Publication P6), two key problems towards atomistic simulations of the integrin TM domain system are solved.

5.3 Outlook

One remaining key issue towards accurate TM simulations of integrins are lipid force fields and their combination with the protein force field. Here, we first have to distinguish between all-atom force fields with parameters for every atom type including hydrogens and united-atom force fields where methyl and methylene groups are represented by a single pseudo atom. For all-atom simulations, two approaches exist: Consistent description of protein and lipid interactions with a single all-atom force field or the combination of an all-atom force field for the protein with a united-atom lipid force field.

For all-atom force fields, e.g. OPLS/AA or assisted model building with energy refinement (AMBER), simulations of lipid bilayers only reproduce a subset of properties experimentally determined [13]. Such simulations require for example additional surface tension potentials to exert lateral pressure on the lipids such that the APL of experiments is reproduced [160]. This additional lateral pressure might affect the integrin conformations in the simulations, therefore artificial potentials should be avoided.

United-atom lipid force fields [161] in contrast allow simulations without application of additional lateral pressure [161]. Protocols for the combination with all-atom force fields exist [162], rendering it the more promising approach at the moment. However,
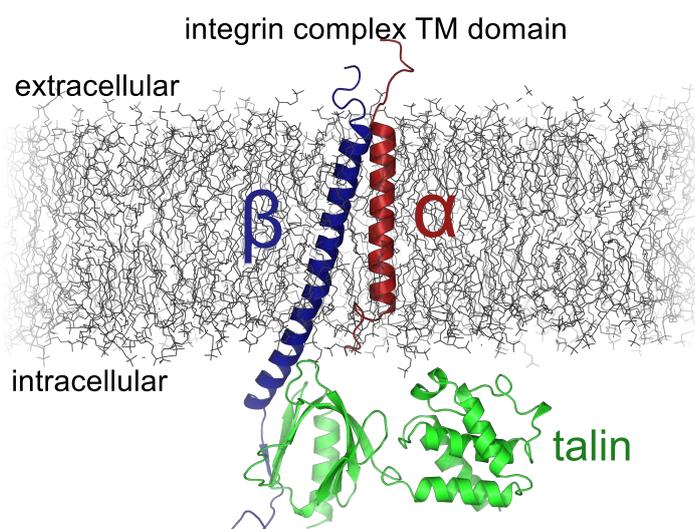
Figure 5.2: **Integrin Transmembrane Domain With Talin Bound.** The α and β subunits are shown in red and blue, talin is shown in green.

when performing simulations with such a combination, the effect of mixing different force fields is not negligible [162]. So far, this problem has no general solution so that the effect of mixed force fields has to be carefully investigated for each system separately.

After selection and testing of a force field, multiple *in silico* research directions are possible. So far, it is known that the signal transduction is achieved by a conformational change of the helical alignment of the integrin TM domains.

Recent results suggest that externally applied forces dramatically enhance the signal transduction, rendering integrins as force sensors. Most likely, the external forces therefore also act on the TM domain. A computational method to mimic this setup are steered molecular dynamics (SMD) simulations. In these simulations, forces are applied to individual TM domains to probe how force direction, magnitude and the point of application change the conformation of the TM domain.

In one step further, calculation of the free energy landscape for the TM helix alignment in the membrane will give a detailed insight into the dimerization pathways and barriers. Considering the two helices as rigid, the free energy landscape has five or seven dimensions[1], depending if rotation around the helix axes is included. This multidimensional free energy landscape contains information about the activation mechanism. The required dimensionality of the free energy landscape however renders this calculation computationally very expensive. Therefore, SMD simulations are the first step.

---

[1]Helix-helix distance, 2 helix tilt angles, relative helix orientation, relative helix location

Different helix tilt angles are observed in experiments (see Ref. 153 Fig. 2). Is the origin of the different angles the membrane composition or a change in tilt angle induced by the binding of intracellular proteins like talin? Atomistic simulations including the intracellular part (see Fig. 5.2) are suitable to probe the effect of talin binding. Moreover, simulations are able to probe the influence of different lipid compositions and therefore could contribute in answering this question.

6

Association Mechanism of Transient Protein-Protein Complexes

- M. Hoefling and K.-E. Gottschalk. Barnase-Barstar: From First Encounter to Final Complex. *J. Struct. Biol.*, 171:52–63, **2010**. (Publication P3).

Individual Contributions

The simulations and the setup of the barnase-barstar complex and mutants at different ion concentrations were done by myself. Kay E Gottschalk suggested the mutations, aided in the system preparation and discussed the steering behavior in detail. I performed the analysis of the PMF, the adsorption path graphs and the contacts. The manuscript was written by Kay E Gottschalk and myself.

6.1 Introduction

Proteins in the cell associate to transient protein complexes which exchange between complexed and free form. Cellular life heavily relies on these protein complexes. They are involved in the regulation of signal transduction in cells of multicellular organisms [163] or act as natural antibiotics of unicellular organisms [164].

To obtain a distinct binding between two or more proteins, the complex partners have to recognize each other reliably, and binding to other proteins is minimized. In nature, binding sites have been optimized during evolution. In order to design or modify specific protein-protein interactions, the understanding of the underlying association mechanism is crucial.

In my studies, the well known barnase-barstar model system [164] shown in Fig. 6.1 was investigated, consisting of a RNase and its inhibitor. The RNase is used by the

Figure 6.1: **Barnase (blue) Barstar (red) in Complex.** Residues responsible for the enzymatic catalysis (RNA cleavage) are shown in green. The binding helix of barstar is responsible for the inhibition of the barnase RNase in complex.
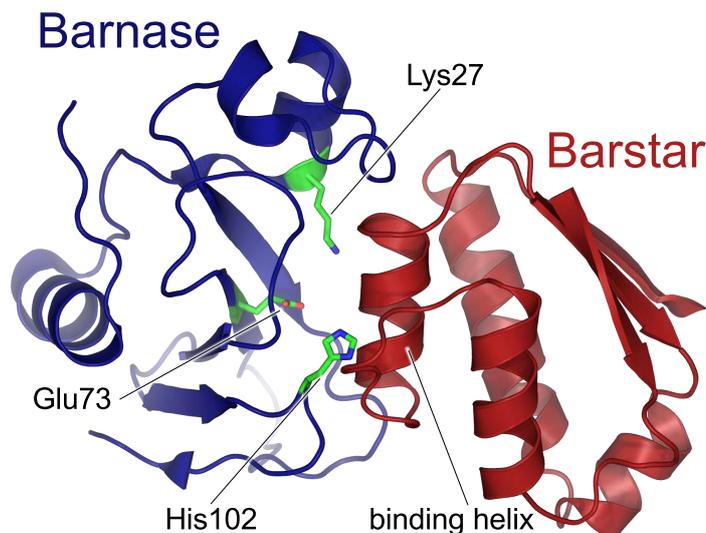
bacterium *Bacillus amyloliquefaciens* as an antibiotic and thus is deadly to itself. Therefore, the RNase is inhibited by barstar in the bacteria and barstar is removed after secretion. Since survival of the bacteria critically depends on the reliability of this inhibition, the binding kinetics became fast and reliable during evolution. Barstar shows inhibitory behavior even when only the binding helix (Fig. 6.1) is folded [165].

The kinetics of the complex formation, responsible for the fast binding, are determined by the shape of the free energy landscape. In this landscape, the magnitude of free energy barriers modulate the frequency factor or attempt frequency. If the barriers are small compared to the available thermal energy, the complex formation is purely driven by the attempt frequency, which depends on the concentration as well as the lateral and rotational diffusion behavior (see Section 2.3). When large barriers are present on the association pathway, the barrier height determines the probability of barrier crossing. Therefore, calculation of the free energy landscape is important to understand the kinetics and was one of the goals in my studies.

The fast association of barnase-barstar is mainly driven by electrostatic interactions [166]. Despite the strong interaction, no disulfide bonds or cations are involved in the barnase-barstar complex. Due to its moderate size, the absence of disulfide bridges in complex, and the biotechnological importance [167], barnase-barstar is an excellent target for *in silico* studies of protein complex formation [168, 169] and was therefore used in this work.

In experiments, a large number of interacting residue pairs of barnase and barstar were identified using double-mutant cycles [170, 171]. In double mutant cycles, relative

free energies of a mutant A, mutant B and double mutant AB are compared. When the mutated residues interact with each other during complex formation, the difference in free energy is not the sum of the single mutations. Only relative free energy values between mutants are obtained in such studies, however.

The exact mechanism of association, the shape of the underlying free energy landscape and the influence of mutations and salt concentration are still under debate. MD simulations were suggested to contribute to the atomistic understanding of the complex formation [172].

The experimental interaction characterization of the mutants provided a starting point to select mutations for my study. The selected mutations will then serve to analyze the impact of mutations on the free energy landscape of binding and thus on the binding kinetics.

So far, the diffusion and the free energy landscape of the barnase-barstar complex were extensively studied by BD simulations [173–175]. Due to the rigidity of the proteins in BD simulations, only the first encounter and not the full complex formation has been simulated. Moreover, diffusional studies are performed using implicit solvent models and thus do not accurately recover the effect of desolvation.

However, accurate description of the solvent is required to simulate the final complex formation steps since it has been demonstrated that the structural modulation of the solvent structure has a strong influence on the binding [55, 139, 176–178]. Therefore I will employ MD simulations in explicit solvent to assess the importance of desolvation and structural rearrangement for the final complex formation steps.

6.2 Results and Discussion

In my study, I compared the wild type complex and three mutants at two different salt concentrations and thus eight different simulation setups in total (Tab. P3.2 on page 93). For each setup, I conducted over 100 simulations constraining the distance between the COM of barnase and COM of barstar to different lengths. From the constraint force in these simulations, the free energy landscape of association (PMF) was calculated using Eq. 2.14 as discussed in Section 2.3.1[1].

Comparison of our calculated free energy landscape with previously suggested ones in Ref. 77 implies that the only barrier present in the complex formation is the desolvation transition (Fig. P3.9 (b) on page 101). No significant barriers at larger

---

[1]In contrast to PMF calculations of amino-acid adsorption (Publication P7), the available phase space increases with increasing COM-COM distances λ and therefore the correction term with the Jacobian Matrix $k_{\mathrm{B}}T\frac{\partial \ln|J|}{\partial \lambda}$ in Eq. 2.14 does not vanish.

separation distances were found. With weak electrostatic shielding present — low salt concentration in our case — the desolvation barrier was compensated by the stronger electrostatic attraction forces.

The shape of the free energy landscape suggests two different interaction regions (Fig. P3.9 (a) on page 101). First, the complex formation is steered by the long ranging electrostatics, and at separation distances below 0.4 nm a short ranged funneling region driven by desolvation is reached. In summary, the final complex formation is not the rate limiting step for barnase-barstar; complexes with weaker electrostatics might behave differently, however.

Structural association modes emerge from clustering of representative structures in our trajectories. To test if different modes are present, I created k-clique community graphs of our representative structures with the clique percolation method weighted (CPMw) algorithm [179, 180]. From these graphs, up to two association modes were found. The equilibrium between both modes depends on the mutations and salt concentrations (Fig. P3.5 on page 96). Two distinct association patterns were identified from the structures which constitute the two paths (Fig. P3.6 on page 97).

My results underline the importance of atomistic simulations to study the effect of solvent structure on the binding behavior and to identify highly populated intermediate conformations during the adsorption process.

6.3 Outlook

The resulting free energy landscapes highlights the importance of the final desolvation barrier, which has been addressed for the same system by MD simulations recently [55]. As a next step, further systems — less rigid and with weaker electrostatic interactions than barnase-barstar — should be investigated to test the generality of the findings.

Technically, sampling of the free energy at large separations is inefficient via MD simulations while BD simulations are not able to capture the final association due to the rigidity of proteins. Therefore, it would be interesting to develop and test a hybrid approach in the future, combining MD simulations at low and BD simulations at large separation distances to obtain an overall PMF at minimal computational cost.

7

GromPy: Python Interface for GROMACS

- R. Pool, A. Feenstra, M. Hoefling, R. Schulz, J. C. Smith and J. Heringa. Enabling Grand-Canonical Monte Carlo: Extending the Flexibility of GROMACS Through the GromPy Python Interface Module. *submitted.* (Manuscript M1).

Individual Contributions

The initial version of the GromPy library was written by Roland Schulz and myself and was made available to the public in 2009[1]. Starting from this version, René Pools, Anton Feenstra, Jeremy C Smith and Jaap Heringa extended the GromPy library by an interface to the main MD routines and performed simulations and analysis in the grand canonical ensemble via GCMC as described in Manuscript M1. The manuscript was mainly written by René Pools, Anton Feenstra, Jeremy C Smith and Jaap Heringa with minor additions from Roland Schulz and myself.

7.1 Summary and Outlook

The MD package GROMACS [70] provides a wide variety of library functions. These libraries implement the MD simulation algorithms and provide functions for the reading, writing and analysis of simulation trajectories.

The simulation core of GROMACS is implemented in C with generated assembler code for the parts critical for the computational performance. GROMACS analysis tools and the analysis library are implemented in C as well. For rapid prototyping of

---

[1]`http://lists.gromacs.org/pipermail/gmx-developers/2009-March/003183.html`

applications, analysis tools and new algorithms, dynamic programming languages like Python are more suitable. Therefore an interface for the available GROMACS routines for Python is desirable.

The GromPy interface was implemented using the ctypes Python module[1], which allows direct function calls to shared libraries from Python code. Shared libraries contain compiled machine code, e.g. implemented in C. The calls to shared libraries from the Python interpreter require no previous compilation of a Python interface as required for other interfacing techniques like the simplified wrapper and interface generator (SWIG).

As the name indicates, ctypes in addition provides wrappers for the fundamental data types and structures in C. Thereby, shared library functions can be directly executed without computationally expensive conversion of variable types (typecasting) rendering the overhead of the Python interpreter small. In summary, the advantage of this approach is the usage of Python as "glue", to quickly combine and extend available and tested library functions optimized for computational speed.

The first implementation of GromPy was written for my trajectory analysis in Publications P3, P6, P8 and P9 and was made available to the public in 2009[2]. Similar interfaces exist for structure manipulation [181]. René Pools and coworkers extended the GromPy library by an interface for the GROMACS simulation core to perform GCMC simulations described in Manuscript M1.

In the future, extension of the GromPy interface to support additional GROMACS library functions is planned. Moreover, the library will be ported to future GROMACS releases which will be based on C++.

---

[1]`http://docs.python.org/library/ctypes.html`
[2]`http://lists.gromacs.org/pipermail/gmx-developers/2009-March/003183.html`

Appendix

A

Publication List

1. M. Hoefling, H. Kessler and K.-E. Gottschalk. The Transmembrane Structure of Integrin αIIbβ3: Significance for Signal Transduction. *Angew. Chem., Int. Ed.*, 48(36):6590–6593, **2009**. (see page 79).

2. M. Hoefling, H. Kessler and K.-E. Gottschalk. Transmembranstruktur von Integrin αIIbβ3 - Bedeutung für die Signalübertragung. *Angew. Chem.*, 121:6714–6717, **2009**. (see page 83).

3. O. Cohavi, S. Corni, F. D. Rienzo, R. D. Felice, K.-E. Gottschalk, M. Hoefling, D. Kokh, E. Molinari, G. Schreiber, A. Vaskevich and R. C. Wade. Protein-Surface Interactions: Challenging Experiments and Computations. *J. Mol. Recognit.*, 23(3):259–262, **2010**. (see page 87).

4. M. Hoefling and K.-E. Gottschalk. Barnase-Barstar: From First Encounter to Final Complex. *J. Struct. Biol.*, 171:52–63, **2010**. (see page 91).

5. D. B. Kokh, S. Corni, P. J. Winn, M. Hoefling, K.-E. Gottschalk and R. C. Wade. ProMetCS: an Atomistic Force Field for Modeling Protein-Metal Surface Interactions in a Continuum Aqueous Solvent. *J. Chem. Theory Comput.*, 6(5):1753–1768, **2010**. (see page 103).

6. M. Hoefling, F. Iori, S. Corni and K.-E. Gottschalk. The Conformations of Amino Acids on a Gold(111) Surface. *ChemPhysChem*, 11(8):1763–1767, **2010**. (see page 119).

7. M. G. Wolf, M. Hoefling, C. Aponte-Santamaría, H. Grubmüller and G. Groenhof. g_membed: Efficient Insertion of a Membrane Protein Into an Equilibrated Lipid Bilayer With Minimal Perturbation. *J. Comput. Chem.*, 31(11):2169–2174, **2010**. (see page 125).

8. M. Hoefling, F. Iori, S. Corni and K.-E. Gottschalk. Interaction of Amino Acids With the Au(111) Surface: Adsorption Free Energies From Molecular Dynamics

Simulations. *Langmuir*, 26(11):8347–8351, **2010**. (see page 131).

9. M. Hoefling, N. Lima, D. Hänni, B. Schuler, C. A. M. Seidel and H. Grubmüller. Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines Through a Hybrid Atomistic Simulation and Monte Carlo Approach. *PLoS One*, 6(5):e19791, **2011**. (see page 137).

10. M. Hoefling, S. Monti, S. Corni and K.-E. Gottschalk. Interaction of β-Sheet Folds With a Gold Surface. *PLoS One*, 6(6):e20925, **2011**. (see page 157).

B

Unpublished Manuscripts

1. R. Pool, A. Feenstra, M. Hoefling, R. Schulz, J. C. Smith and J. Heringa. Enabling Grand-Canonical Monte Carlo: Extending the Flexibility of GROMACS Through the GromPy Python Interface Module. *submitted*. (see page 171).

The Transmembrane Structure of Integrin αIIbβ3: Significance for Signal Transduction**

*Martin Hoefling, Horst Kessler, and Kay-Eberhard Gottschalk**

cell adhesion · computational prediction · membrane proteins · receptors · signal transduction

*T*he formation of multicellular organisms requires concerted action by cells, which alter their adhesive and migratory behaviors. Cell adhesion and migration are tightly regulated by intra- and extracellular signals, which are conveyed through the cellular membrane by specialized receptors known as integrins.[1] Integrins are the starting point of a variety of signaling cascades and are involved in a multitude of physiological events important to multicellular organism morphogenesis, ranging from cell adhesion to migration, apoptosis, and angiogenesis as well as pathophysiological behaviors such as those found in cancer metastasis.[2] The transmembrane (TM) domains of integrins are at the center of integrin signaling.[3] Recently, a structure of the TM domains of the αIIbβ3 integrin has been reported that sheds light on the signal transduction mechanism of integrins.[4–7]

Integrins are essential TM proteins that couple the extracellular matrix to the cytoskeleton. They consist of noncovalently bound heterodimers, in which each subunit (α and β) contains one TM helix.[8] With eighteen α and eight β subunit types identified, there are 24 known distinct heterodimer combinations with partially overlapping yet specific function.[9]

Cells regulate their integrin-mediated adhesion through a variety of mechanisms on different time scales. On the slower time scale, expression patterns are altered by external signals, such as growth factors.[10] On faster time scales, integrins can be redistributed (by clustering or recycling) on the cellular surface,[11] change their intracellular attachment state to the cytoskeleton[12] as exemplified by different lateral mobilities of integrins on a single cell,[13] or alter the affinity state for their extracellular ligand (integrin activation).[14] Changes in the affinity state are correlated with integrin conformational

changes,[14] which act as one potential mechanism to relay signals either from the extracellular to the intracellular space (outside-in signaling) or vice versa (inside-out signaling). Integrins are bidirectional signaling molecules, as both signaling directions take place along an allosteric pathway.[15] The two TM helices of the integrin heterodimer are pivotal in signaling events as linkers between the extracellular and intracellular domains. Hence, a variety of groups have pursued different strategies to understand the role of the TM domains in signaling.[16–23]

It was postulated that the TM helices were not merely connectors between the extra- and the intracellular space but active structures forming specific heterodimers.[24,25] However, the structural details of this TM heterodimer have remained elusive until recently. In their recent report, Lau et al. have solved the structure of the αIIbβ3 integrin heterodimer in its resting state using highly sophisticated, well-designed NMR spectroscopy experiments on the TM domains in bicelles.[7] This experimentally determined structure allows a solid, structural basis for the prior experimental biochemical results by providing a foundation for an atomistic understanding of TM integrin signaling.

The experimental structure shows a right-handed helical dimer (Figure 1). Given that the majority of soluble helix dimers form left-handed structures,[26] this unusual right-

[*] M. Hoefling, Dr. K.-E. Gottschalk
Angewandte Physik und Biophysik & Center for NanoScience
Ludwig-Maximilians Universität
Amalienstrasse 54, 80799 München (Germany)
Fax: (+49) 89-2180-2050
E-mail: kay.gottschalk@physik.uni-muenchen.de

Prof. Dr. H. Kessler
Institute for Advanced Study, Department Chemie
Technische Universität München
Lichtenbergstrasse 4, 85747 München (Germany)

Figure 1. Overview of the αIIbβ3 integrin TM structure recently solved by Lau et al. The α subunit is shown in red, the GFFKR motif in green, and the β subunit in blue. Several important residues described in the text are highlighted.

6590

WILEY InterScience®
DISCOVER SOMETHING GREAT

© 2009 Wiley-VCH Verlag GmbH & Co. KGaA, Weinheim

Angew. Chem. Int. Ed. **2009**, *48*, 6590–6593

handed conformation and the resulting lower thermodynamic stability may play an important role in signaling. It has been demonstrated that the TM domains separate in the course of signaling.[27] Therefore, a fine-tuned energy balance favoring specific interactions versus the propensity to separate needs to be encoded in the interaction between these helices. A right-handed helical dimer is particularly suited by offering the needed structural flexibility. This propensity to separate, which complicated the structure determination of this important complex,[28] is also reflected by the low affinity of the helices for each other.[29]

The TM structure can be divided into two different interaction regions. Adjacent to the extracellular domains, the helices form canonical helix–helix interactions in tight contact. Proximal to the intracellular face of the membrane, an unusual loop comprising the highly conserved GFFKR[25,30] motif of the α subunit makes important interactions between the subunits, while the helices are already well-separated.

Near the extracellular domain, the conserved ^{972}Gxxx^{976}G motif in the α subunit, a well-known interaction motif of TM helices (x is a nonconserved amino acid),[31] is located at the interface of the dimer to allow tight interactions with the β subunit (Figure 1). A similar but less canonical motif of the β subunit (^{699}Sxxx^{703}A) points out from the interface and does not participate in the interaction. The relevance of this motif remains to be determined, but it may be important for lateral association with other TM proteins such as tetraspanins or integrins in integrin clustering, or it might be involved in transient TM conformations. It would be instructive to mutate these residues in animal models to test their significance.

Towards the intracellular interface, the C-terminal end of the α subunit contains an unusual loop (initiated by ^{991}G in the α subunit). Gly is a well-known helix breaker, given its highly flexible nature resulting from its lack of side-chain atoms. The main contact between the subunits is formed in this loop region by the ^{992}F^{993}F motif and the subsequent salt bridge (Figure 1). These interactions have previously been shown to be indispensable in keeping the integrin in its resting state.[25,30]

Integrins are subjected to force during the mechanical processes of adhesion and migration. In fact, force has been identified as a facilitator for integrin activation.[32] Interestingly, the deviation from helicity in the α subunit may be significant for integrin function: if force is acting on the α subunit tail, the nonhelical loop would facilitate activation under applied force, as less of an energetic penalty is paid for straightening a less-ordered loop than a hydrogen-bond-stabilized helix. Hence, this loop is an attractive candidate for a trigger in integrin activation by force (Figure 2). Surprisingly, no such trigger is structurally obvious in the canonical helix of the β subunit, given that the main intracellular adaptor protein talin, which connects the integrin to the cytoskeleton and is likely the force-transducing protein, binds to the β and not the α subunit. It would be interesting to identify adaptor proteins that bind to the α subunit and facilitate activation by mechanical structural changes in the loop region of the α subunit. Recent evidence suggests that, at least in the case of the integrin α4β1, paxillin might be a candidate for such a mechanism.[12]
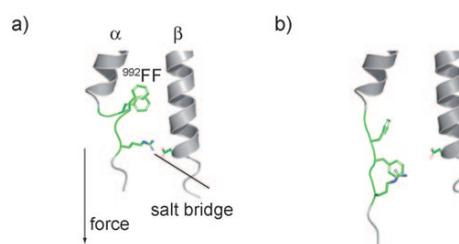
Figure 2. Potential impact of force on the α subunit. a) Experimental structure. GFFKR motif is shown in green. b) Force-induced straightening of the GFFKR motif (putative model).

The formation of a complex between the integrin β tail and talin has been shown to be the common final step in integrin activation.[33] During complex formation, talin displaces the α subunit, thus inducing TM separation and integrin activation.[34,35] The new structure of the TM domains in combination with recently solved structures of the integrin–talin complex now enables a structural test of this hypothesis. Surprisingly, talin can bind to the integrin β subunit without structural constraints imposed by the α subunit (Figure 3).
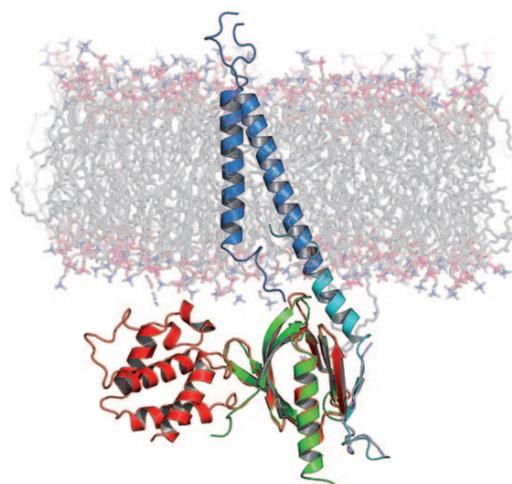
Figure 3. Concatenation of TM structure (blue) and two integrin–talin complexes (red and green). Background: approximate position of the membrane.

Hence, a possible mechanism of a purely sterical displacement of the α subunit needs to be refined. It must be remembered, however, that the intracellular tail of the α subunit is longer than the portion that was resolved by NMR spectroscopy.

In the past, models of the TM domains have been generated by a variety of approaches.[19,24,36–42] The publication of the structure of the TM domains allows the predictive power of TM structure versus computer modeling to be tested. We published in 2002 a first computational model of the core of the TM domains of the integrin αIIbβ3 calculated in the absence of experimental data on the TM domains.[37] Despite this lack of experimental support at the time of

calculation, the predicted conformation and the experimental structure are identical with a root-mean-square deviation (RMSD) of less than 1 Å across the Cα atoms over 34 residues (Figure 4). Later computations by the groups of
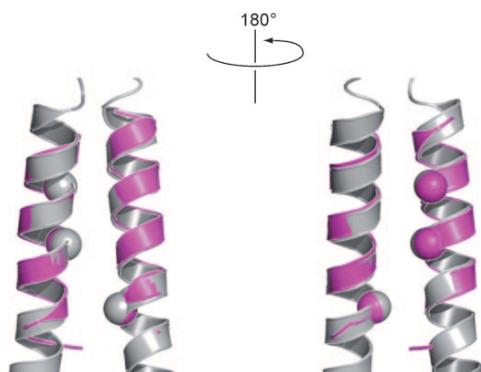
Figure 4. Superposition of experimental (gray) and predicted (magenta) TM domain conformation. Two Gly residues on the α chain and one Gly on the β chain are shown as spheres to demonstrate rotational orientation for a front and rear view. The predicted conformation was superimposed on Model 1 of the NMR ensemble.

DeGrado,[39] Torres,[41] and most recently Springer,[43] obtained with different methods, led to very similar models. The Springer group even correctly predicted the unusual C-terminal loop of the α subunit using a large number of experimental restraints together with an ab initio structure-prediction tool. This close agreement between experiment and computation impressively underlines the power of computational approaches for these kinds of systems.

Obtaining the structure of integrin TM domains provides the foundations for a structural understanding of TM integrin signaling. However, open questions remain. These include the impact of force on the TM domains, the possible existence of structural intermediates between complexed and uncomplexed subunits, and the mechanism of TM activation through talin binding. Other important issues are the possibility of formation of homooligomers and their role in formation of focal adhesion. The molecular nature of cell migration along concentration gradients as well as the spatiotemporal demands of chemotactic proteins in cooperation with integrin ligands also require further study.[44,45] Furthermore, the details of transducing information of ligand binding from the headgroup to the transmembrane region or vice versa are still not well-understood. The now available structure of the transmembrane complex is a tremendous step forward in understanding the function of these important molecules. Future computational and experimental efforts will address the questions on the basis of this structure.

[1] T. A. Springer, *Nature* **1990**, *346*, 425.
[2] K. R. Legate, S. A. Wickstrom, R. Fassler, *Genes Dev.* **2009**, *23*, 397.
[3] K. E. Gottschalk, H. Kessler, *Angew. Chem.* **2002**, *114*, 3919; *Angew. Chem. Int. Ed.* **2002**, *41*, 3767.
[4] C. Kim, T. Lau, T. S. Ulmer, M. H. Ginsberg, *Blood* **2009**, *113*, 4747.
[5] T. Lau, A. W. Partridge, M. H. Ginsberg, T. S. Ulmer, *Biochemistry* **2008**, *47*, 4008.
[6] T. L. Lau, V. Dua, T. S. Ulmer, *J. Biol. Chem.* **2008**, *283*, 16162.
[7] T. L. Lau, C. Kim, M. H. Ginsberg, T. S. Ulmer, *EMBO J.* **2009**, *28*, 1351.
[8] M. J. Humphries, *Biochem. Soc. Trans.* **2000**, *28*, 311.
[9] J. D. Humphries, A. Byron, M. J. Humphries, *J. Cell Sci.* **2006**, *119*, 3901.
[10] S. M. Wahl, J. B. Allen, B. S. Weeks, H. L. Wong, P. E. Klotman, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4577.
[11] C. Cluzel, F. Saltel, J. Lussi, F. Paulhe, B. A. Imhof, B. Wehrle-Haller, *J. Cell Biol.* **2005**, *171*, 383.
[12] R. Alon, S. W. Feigelson, E. Manevich, D. M. Rose, J. Schmitz, D. R. Overby, E. Winter, V. Grabovsky, V. Shinder, B. D. Matthews, M. Sokolovsky-Eisenberg, D. E. Ingber, M. Benoit, M. H. Ginsberg, *J. Cell Biol.* **2005**, *171*, 1073.
[13] O. Lieleg, M. Lopez-Garcia, C. Semmrich, J. Auerheimer, H. Kessler, A. R. Bausch, *Small* **2007**, *3*, 1560.
[14] J. Takagi, B. M. Petre, T. Walz, T. A. Springer, *Cell* **2002**, *110*, 599.
[15] M. A. Arnaout, B. Mahalingam, J. P. Xiong, *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 381.
[16] H. Yin, J. S. Slusky, B. W. Berger, R. S. Walters, G. Vilaire, R. I. Litvinov, J. D. Lear, G. A. Caputo, J. S. Bennett, W. F. DeGrado, *Science* **2007**, *315*, 1817.
[17] J. Q. Zhu, C. V. Carman, M. Kim, M. Shimaoka, T. A. Springer, B. H. Luo, *Blood* **2007**, *110*, 2475.
[18] B. H. Luo, C. V. Carman, J. Takagi, T. A. Springer, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3679.
[19] K. E. Gottschalk, H. Kessler, *Structure* **2004**, *12*, 1109.
[20] B. H. Luo, T. A. Springer, J. Takagi, *PLoS Biol.* **2004**, *2*, 776.
[21] A. Stefansson, A. Armulik, I. M. Nilsson, G. von Heijne, S. Johansson, *J. Biol. Chem.* **2004**, *279*, 21200.
[22] R. H. Li, N. Mitra, H. Gratkowski, G. Vilaire, R. Litvinov, C. Nagasami, J. W. Weisel, J. D. Lear, W. F. DeGrado, J. S. Bennett, *Science* **2003**, *300*, 795.
[23] R. H. Li, C. R. Babu, J. D. Lear, A. J. Wand, J. S. Bennett, W. F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12462.
[24] B. D. Adair, M. Yeager, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14059.
[25] P. E. Hughes, F. DiazGonzalez, L. Leong, C. Y. Wu, J. A. McDonald, S. J. Shattil, M. H. Ginsberg, *J. Biol. Chem.* **1996**, *271*, 6571.
[26] A. N. Lupas, M. Gruber, *Adv. Protein Chem.* **2005**, *70*, 37.
[27] M. Kim, C. V. Carman, T. A. Springer, *Science* **2003**, *301*, 1720.
[28] T. S. Ulmer, B. Yaspan, M. H. Ginsberg, I. D. Campbell, *Biochemistry* **2001**, *40*, 7498.
[29] D. Schneider, D. M. Engelman, *J. Biol. Chem.* **2004**, *279*, 9840.
[30] T. E. O'Toole, Y. Katagiri, R. J. Faull, K. Peter, R. Tamura, V. Quaranta, J. C. Loftus, S. J. Shattil, M. H. Ginsberg, *J. Cell Biol.* **1994**, *124*, 1047.
[31] A. Senes, D. E. Engle, W. F. DeGrado, *Curr. Opin. Struct. Biol.* **2004**, *14*, 465.
[32] R. Alon, M. L. Dustin, *Immunity* **2007**, *26*, 17.
[33] S. Tadokoro, S. J. Shattil, K. Eto, V. Tai, R. C. Liddington, J. M. de Pereda, M. H. Ginsberg, D. A. Calderwood, *Science* **2003**, *302*, 103.
[34] K. L. Wegener, A. W. Partridge, J. Han, A. R. Pickford, R. C. Liddington, M. H. Ginsberg, I. D. Campbell, *Cell* **2007**, *128*, 171.
[35] K. L. Wegener, I. D. Campbell, *Mol. Membr. Biol.* **2008**, *25*, 376.
[36] K. E. Gottschalk, *Structure* **2005**, *13*, 703.
[37] K. E. Gottschalk, P. D. Adams, A. T. Brunger, H. Kessler, *Protein Sci.* **2002**, *11*, 1800.

[38] K. E. Gottschalk, H. Kessler, FEBS Lett. 2004, 557, 253.

[39] W. Li, D. G. Metcalf, R. Gorelik, R. H. Li, N. Mitra, V. Nanda, P. B. Law, J. D. Lear, W. F. DeGrado, J. S. Bennett, Proc. Natl. Acad. Sci. USA 2005, 102, 1424.

[40] X. Lin, S. M. Tan, S. K. A. Law, J. Torres, Proteins Struct. Funct. Bioinf. 2006, 63, 16.

[41] X. Lin, S. M. Tan, S. K. A. Law, J. Torres, Proteins Struct. Funct. Bioinf. 2006, 65, 274.

[42] A. W. Partridge, S. C. Liu, S. Kim, J. U. Bowie, M. H. Ginsberg, J. Biol. Chem. 2005, 280, 7294.

[43] V. C. Hirschfeld-Warneken, M. Arnold, A. Cavalcanti-Adam, M. Lopez-Garcia, H. Kessler, J. P. Spatz, Eur. J. Cell. Biol. 2008, 87, 743.

[44] V. C. Hirschfeld-Warneken, M. Arnold, A. Cavalcanti-Adam, M. Lopez-Garcia, H. Kessler, J. P. Spatz, Eur. J. Cell Biol. 2008, 87, 743.

[45] M. Arnold, V. C. Hirschfeld-Warneken, T. Lohmuller, P. Heil, J. Blummel, E. A. Cavalcanti-Adam, M. Lopez-Garcia, P. Walther, H. Kessler, B. Geiger, J. P. Spatz, Nano Lett. 2008, 8, 2063.

Signalübertragung

Transmembranstruktur von Integrin αIIbβ3 – Bedeutung für die Signalübertragung**

Martin Hoefling, Horst Kessler und Kay-Eberhard Gottschalk*

Computervorhersage · Rezeptoren · Zelladhäsion · Signaltransduktion · Membranproteine

Mehrzeller benötigen ein eng aufeinander abgestimmtes Handeln von Zellen, die ihre Adhäsions- und Migrationseigenschaften verändern können. Zelladhäsion und -migration werden stark durch intra- und extrazelluläre Signale reguliert, die durch die Zellmembran mithilfe von spezialisierten Rezeptoren, bekannt als Integrine, transportiert werden.[1] Integrine sind der Ausgangspunkt für eine Vielzahl von Signalkaskaden. Außerdem sind sie an vielen physiologischen Prozessen der Morphogenese von Mehrzellern beteiligt, von der Zelladhäsion über die Migration bis hin zur Apoptose und Angiogenese. Außerdem spielen sie bei pathophysiologischem Verhalten, z. B. bei der Metastasenbildung, eine Rolle.[2] Das Zentrum der Signalübertragung des Integrins sind die Transmembran(TM)-Domänen.[3] Erst vor kurzem wurde eine Struktur der TM-Domänen des αIIbβ3-Integrins vorgestellt, die ein neues Licht auf den Signalübertragungsmechanismus von Integrinen wirft.[4–7]

Integrine sind essenzielle TM-Proteine, die die extrazelluläre Matrix mit dem Zytoskelett verbinden. Sie bestehen aus nichtkovalent gebundenen Heterodimeren, von denen jede Untereinheit – α und β – jeweils eine TM-Helix enthält.[8] Es sind 18 verschiedene α- und 8 verschiedene β-Untereinheiten bekannt, aus denen sich 24 bekannte Heterodimerkombinationen, teils mit spezifischer, teils mit überlappender Funktion, bilden.[9]

Zellen regulieren ihre Integrin-vermittelte Adhäsion durch eine Vielzahl von Mechanismen auf unterschiedlichen Zeitskalen. Auf größeren Zeitskalen werden Expressionsmuster durch externe Signale wie Wachstumsfaktoren verändert.[10] Auf kleineren Zeitskalen können sich Integrine z. B. auf Zelloberflächen durch Clustering und Recycling neu anordnen[11] oder ihren intrazellulären Bindungszustand zum Zytoskelett verändern,[12] beispielhaft verdeutlicht durch unterschiedliche laterale Mobilität von Integrinen auf einer Zelle.[13] Außerdem spielt sich auf dieser Zeitskala auch die Änderung ihrer Affinität zu einem extrazellulären Liganden ab (Integrinaktivierung).[14] Veränderungen in der Affinität sind stark mit Konformationsänderungen von Integrinen korreliert,[14] was ein potenzieller Mechanismus zur Signalvermittlung ist, entweder vom extra- zum intrazellulären Raum oder umgekehrt. Integrine können Signale bidirektional übertragen, da beide Signalübertragungen entlang eines allosterischen Pfades verlaufen.[15] Die beiden TM-Helices des Integrinheterodimers sind der Dreh- und Angelpunkt dieser Signalereignisse, da sie die extra- und intrazellulären Domänen miteinander verbinden. Aus diesem Grund wurden von mehreren Forschergruppen unterschiedliche Strategien entwickelt, um die Rolle der TM-Domänen in diesem Signalprozess besser zu verstehen.[16–23]

Es wurde postuliert, dass die TM-Helices nicht nur als Verbindung zwischen extra- und intrazellulärem Raum wirken, sondern auch als aktive Strukturen, die spezifische Heterodimerstrukturen bilden können.[24,25] Allerdings waren die Strukturdetails dieser TM-Heterodimere bis vor kurzem noch spekulativ. In ihrer kürzlich erschienenen Publikation haben Lau et al. die Struktur des αIIbβ3-Integrinheterodimers in seinem Ruhezustand mithilfe anspruchsvoller NMR-Experimente an TM-Domänen in Bizellen gelöst.[7] Diese experimentell bestimmte Struktur bildet eine solide Grundlage für das Verständnis bisheriger experimenteller Untersuchungen sowie der TM-Signalübertragung auf atomarer Ebene.

Die experimentelle Struktur zeigt ein rechtshändig helicales Dimer (Abbildung 1). In Anbetracht der Tatsache, dass die meisten löslichen Helixdimere linkshändige Strukturen bilden,[26] könnten diese ungewöhnliche rechtshändige Konformation und die daraus resultierende niedrigere thermodynamische Stabilität eine wichtige Funktion bei der Signalübertragung innehaben. Die TM-Domänen trennen sich nachweislich im Verlauf der Signalweiterleitung.[27] Daher ist eine fein abgestimmte Energiebalance nötig, um bestimmte Wechselwirkungen gegenüber der Auftrennung zu begünstigen, die in den Helix-Helix-Wechselwirkungen kodiert sein muss. Ein rechtshändig helicales Dimer kann die dafür benötigte besondere Strukturflexibilität bereitstellen. Die Tendenz zur Trennung, die die Bestimmung der Struktur dieses wichtigen Komplexes in der Vergangenheit erschwerte,[28] spiegelt sich

[*] M. Hoefling, Dr. K.-E. Gottschalk
Angewandte Physik und Biophysik & Center for NanoScience
Ludwig-Maximilians-Universität
Amalienstraße 54, 80799 München (Deutschland)
Fax: (+49) 89-2180-2050
E-Mail: kay.gottschalk@physik.uni-muenchen.de
Prof. Dr. H. Kessler
Institute for Advanced Study, Department Chemie
Technische Universität München
Lichtenbergstraße 4, 85747 München (Deutschland)
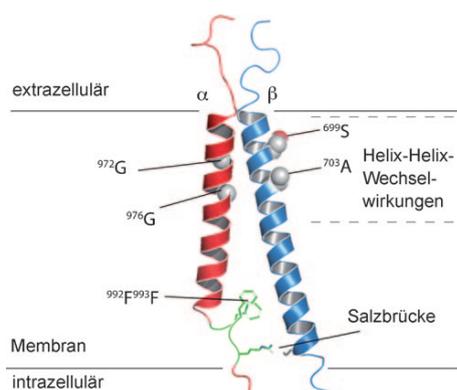
Abbildung 1. Kürzlich von Lau et al. gelöste αIIbβ3-Integrin-TM-Struktur. Die α-Untereinheit ist in Rot gezeigt, das GFFKR-Motiv in Grün und die β-Untereinheit in Blau. Einige im Text beschriebene, wichtige Reste wurden hervorgehoben.

auch in der geringen Affinität der Helices zueinander wider.[29]

Die TM-Struktur kann in zwei unterschiedliche Wechselwirkungsregionen aufgeteilt werden: Nahe den extrazellulären Domänen gehen die Helices enge, kanonische Helix-Helix-Wechselwirkungen ein, und nahe der intrazellulären Seite zeigt sich eine ungewöhnliche Schleife, die ein hochkonserviertes GFFKR-Motiv[25,30] auf der α-Untereinheit umfasst und wichtige Wechselwirkungen zwischen den Untereinheiten vermittelt, wenn die Helices schon weiter getrennt sind.

Nahe den extrazellulären Domänen befindet sich das konservierte und in TM-Helices bekannte[31] ^{972}Gxxx^{976}G-Motiv (x: nichtkonservierte Aminosäure) des Dimers, das eine enge Wechselwirkung mit der β-Untereinheit (Abbildung 1) ermöglicht. Ein ähnliches, aber weniger kanonisches Motiv der β-Untereinheit (^{699}Sxxx^{703}A) zeigt von den Berührungsflächen weg und ist nicht an der Wechselwirkung beteiligt. Dieses Motiv, dessen Bedeutung noch unklar ist, könnte wichtig für die laterale Assoziation mit anderen TM-Proteinen, wie Tetraspaninen, oder beim Integrin-Integrin-Clustering sein; außerdem könnte es eine Rolle bei der Bildung von Zwischenzuständen der TM-Helices spielen. Es wäre aufschlussreich, die Auswirkungen von Mutationen in diesem Motiv im tierischen Modellversuch zu testen.

Zur intrazellulären Grenzschicht orientiert findet sich das C-terminale Ende der α-Untereinheit mit einer ungewöhnlichen Schleife (eingeleitet durch ^{991}G in der α-Untereinheit). Glycin wirkt wegen seiner hohen Flexibilität und dem Fehlen von Seitenkettenatomen als Helixunterbrecher. Der Hauptkontakt zwischen den Untereinheiten wird in dieser Schleifenregion durch das ^{992}F^{993}F-Motiv und die nachfolgende Salzbrücke (Abbildung 1) hergestellt. Diese Wechselwirkungen sind für die Stabilität von Integrinen im Ruhezustand unentbehrlich.[25,30]

Bei Adhäsion und Migration unterliegen Integrine angelegten Kräften, und tatsächlich konnte die Kraft als ein Moderator der Aktivierung von Integrinen identifiziert werden.[32] Interessanterweise könnte in diesem Kontext die Abweichung von der Helizität eine Bedeutung für die

Funktion des Integrins haben: Wenn eine Kraft auf das Ende der α-Untereinheit wirkt, könnte das nichthelicale Ende eine Aktivierung erleichtern, da eine geringere Energie für das Strecken der weniger geordneten Schleife nötig ist als für die H-Brücken-stabilisierte Helix. Daher ist diese Schleife ein vielversprechender Kandidat für eine kraftinduzierte Aktivierung (Abbildung 2). Interessanterweise findet sich in der Struktur der β-Untereinheit kein klarer Hinweis auf einen
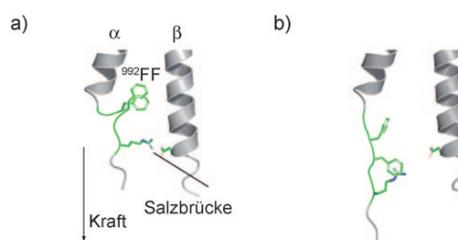
Abbildung 2. Mögliche Auswirkungen einer an die α-Untereinheit angelegten Kraft. a) Experimentelle Struktur; das GFFKR-Motiv ist in Grün gezeigt. b) Kraftinduziertes Strecken des GFFKR-Motivs (putatives Modell).

solchen Auslösemechanismus, obwohl das prominenteste intrazelluläre Adapterprotein Talin, welches das Integrin mit dem Zytoskelett verbindet und daher vermutlich das kraftübertragende Protein ist, an die β- und nicht an die α-Untereinheit bindet. Daher wäre es interessant, weitere Adapterproteine zu identifizieren, die an die α-Untereinheit binden und daher eine Aktivierung durch mechanische Strukturveränderung in der Schleifenregion der α-Untereinheit induzieren können. Neueste Indizien stützen die Vermutung, dass zumindest im Fall von α4β1 das Paxillin ein Kandidat für einen solchen Mechanismus ist.[12]

Die Bildung eines Komplexes zwischen dem Ende der β-Untereinheit und Talin ist der letzte gemeinsame Schritt bei der Aktivierung von Integrinen.[33] Talin verdrängt dabei die α-Untereinheit und bewirkt damit eine Trennung der TM-Domänen, die letztlich zur Aktivierung führt.[34,35] Die neue Struktur der TM-Domänen ermöglicht nun in Kombination mit vor kurzem gelösten Strukturen des Integrin-Talin-Komplexes eine Überprüfung dieser Hypothese. Überraschenderweise kann Talin an die β-Untereinheit binden, ohne strukturell durch die α-Untereinheit eingeschränkt zu sein (Abbildung 3). Daher muss der mögliche Mechanismus einer rein sterischen Verschiebung der α-Untereinheit revidiert werden. Außerdem ist zu beachten, dass der intrazelluläre Teil der α-Untereinheit länger ist als der Teil, der mithilfe der NMR-Spektroskopie gelöst wurde.

In der Vergangenheit wurden Modelle von TM-Domänen mit einer Vielzahl von Methoden erstellt.[19,24,36–42] Die Publikation der Struktur dieser TM-Domänen ermöglicht es nun, die Vorhersagekraft von am Computer berechneten Modellen zu testen. 2002 veröffentlichten wir ein erstes Modell des Kerns der TM-Domänen von Integrin αIIbβ3, das berechnet worden war, ohne dass experimentelle Daten zu den TM-Domänen selbst verfügbar waren.[37] Trotz des Fehlens der experimentellen Verifikation während der Rechnung
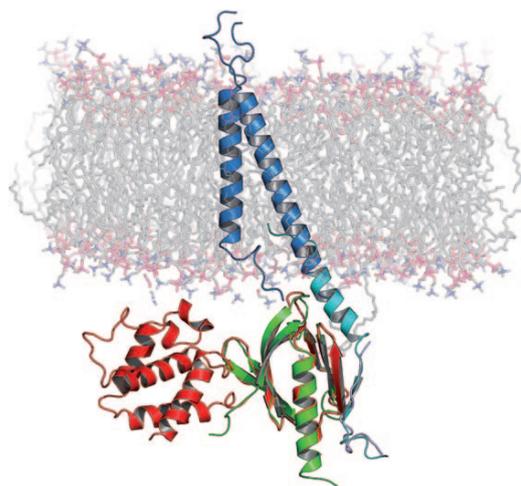
Abbildung 3. Verknüpfung der TM-Struktur (blau) und zweier Integrin-Talin-Komplexe (rot und grün). Hintergrund: ungefähre Position der Membran.

sind die vorhergesagte Konformation und die experimentelle Struktur innerhalb einer mittleren quadratischen Standard-abweichung von weniger als 1 Å über die Cα-Atome der 34 Aminosäurereste identisch (Abbildung 4). Spätere Rech-nungen mit anderen Methoden durch DeGrado,[39] Torres[41] und neuerdings auch Springer et al.[43] führten zu sehr ähnli-
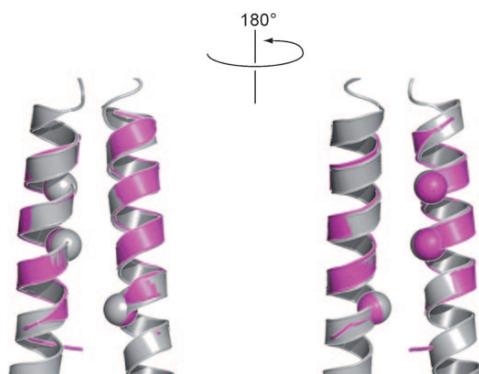
Abbildung 4. Überlagerung der experimentellen (grau) und vorherge-sagten TM-Domänenkonformation (magenta). Zwei Gly-Reste auf der α-Untereinheit und ein Gly-Rest auf der β-Untereinheit sind zur besse-ren Demonstration der Rotationsorientierung in einer Vorder- und Rückansicht als Kugeln dargestellt. Die vorhergesagte Konformation wurde mit Modell 1 des NMR-Ensembles überlagert.

chen Ergebnissen. Springer und Mitarbeiter gelang es sogar, die ungewöhnliche C-terminale Schleife der α-Untereinheit vorherzusagen, indem sie eine Reihe von experimentellen Beschränkungen mit einem Programm zur Ab-initio-Struk-turvorhersage kombinierten. Diese enge Übereinstimmung zwischen Experiment und Rechnung demonstriert die Leis-tungsfähigkeit von Rechenverfahren bei dieser Art von Sys-temen.

Strukturen von Integrin-TM-Domänen bilden die Basis für ein Verständnis der Signalweiterleitung von Integrin-TM-Domänen. Allerdings sind noch einige offene Fragen zu klä-ren, z. B. zur Auswirkung einer angelegten Kraft auf die TM-Domänen, zur Existenz von Übergangsstrukturen zwischen komplexierten und nichtkomplexierten Untereinheiten und zum Mechanismus der Integrinaktivierung durch das Binden von Talin. Weitere wichtige offene Punkte sind die mögliche Bildung von Homooligomeren und ihre Rolle bei der Bildung einer fokalen Adhäsion. Auch die molekulare Natur der Zellmigration entlang von Konzentrationsgradienten sowie die räumlichen und zeitlichen Anforderungen an die Protei-ne, die an der Chemotaxis beteiligt sind, sowie deren Ko-operation mit Integrinliganden müssen noch weiter unter-sucht werden.[44,45] Auch die Details der Informationsweiter-leitung nach dem Binden eines Liganden von der Kopfgruppe hin zur Transmembranregion sind noch immer nicht gut ver-standen. Die nun verfügbare Struktur des Transmembran-komplexes ist ein enormer Schritt nach vorne für das Ver-ständnis der Funktion dieser Moleküle und bietet eine Grundlage, um die offenen Fragen durch weitere Modell-rechnungen und experimentelle Untersuchungen zu beant-worten.

[1] T. A. Springer, *Nature* **1990**, *346*, 425.
[2] K. R. Legate, S. A. Wickstrom, R. Fassler, *Genes Dev.* **2009**, *23*, 397.
[3] K. E. Gottschalk, H. Kessler, *Angew. Chem.* **2002**, *114*, 3919; *Angew. Chem. Int. Ed.* **2002**, *41*, 3767.
[4] C. Kim, T. Lau, T. S. Ulmer, M. H. Ginsberg, *Blood* **2009**, *113*, 4747.
[5] T. Lau, A. W. Partridge, M. H. Ginsberg, T. S. Ulmer, *Bio-chemistry* **2008**, *47*, 4008.
[6] T. L. Lau, V. Dua, T. S. Ulmer, *J. Biol. Chem.* **2008**, *283*, 16162.
[7] T. L. Lau, C. Kim, M. H. Ginsberg, T. S. Ulmer, *EMBO J.* **2009**, *28*, 1351.
[8] M. J. Humphries, *Biochem. Soc. Trans.* **2000**, *28*, 311.
[9] J. D. Humphries, A. Byron, M. J. Humphries, *J. Cell Sci.* **2006**, *119*, 3901.
[10] S. M. Wahl, J. B. Allen, B. S. Weeks, H. L. Wong, P. E. Klotman, *Proc. Natl. Acad. Sci. USA* **1993**, *90*, 4577.
[11] C. Cluzel, F. Saltel, J. Lussi, F. Paulhe, B. A. Imhof, B. Wehrle-Haller, *J. Cell Biol.* **2005**, *171*, 383.
[12] R. Alon, S. W. Feigelson, E. Manevich, D. M. Rose, J. Schmitz, D. R. Overby, E. Winter, V. Grabovsky, V. Shinder, B. D. Matt-hews, M. Sokolovsky-Eisenberg, D. E. Ingber, M. Benoit, M. H. Ginsberg, *J. Cell Biol.* **2005**, *171*, 1073.
[13] O. Lieleg, M. Lopez-Garcia, C. Semmrich, J. Auerheimer, H. Kessler, A. R. Bausch, *Small* **2007**, *3*, 1560.
[14] J. Takagi, B. M. Petre, T. Walz, T. A. Springer, *Cell* **2002**, *110*, 599.
[15] M. A. Arnaout, B. Mahalingam, J. P. Xiong, *Annu. Rev. Cell Dev. Biol.* **2005**, *21*, 381.
[16] H. Yin, J. S. Slusky, B. W. Berger, R. S. Walters, G. Vilaire, R. I. Litvinov, J. D. Lear, G. A. Caputo, J. S. Bennett, W. F. DeGrado, *Science* **2007**, *315*, 1817.
[17] J. Q. Zhu, C. V. Carman, M. Kim, M. Shimaoka, T. A. Springer, B. H. Luo, *Blood* **2007**, *110*, 2475.
[18] B. H. Luo, C. V. Carman, J. Takagi, T. A. Springer, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 3679.

[19] K. E. Gottschalk, H. Kessler, *Structure* **2004**, *12*, 1109.

[20] B. H. Luo, T. A. Springer, J. Takagi, *PLoS Biol.* **2004**, *2*, 776.

[21] A. Stefansson, A. Armulik, I. M. Nilsson, G. von Heijne, S. Johansson, *J. Biol. Chem.* **2004**, *279*, 21200.

[22] R. H. Li, N. Mitra, H. Gratkowski, G. Vilaire, R. Litvinov, C. Nagasami, J. W. Weisel, J. D. Lear, W. F. DeGrado, J. S. Bennett, *Science* **2003**, *300*, 795.

[23] R. H. Li, C. R. Babu, J. D. Lear, A. J. Wand, J. S. Bennett, W. F. DeGrado, *Proc. Natl. Acad. Sci. USA* **2001**, *98*, 12462.

[24] B. D. Adair, M. Yeager, *Proc. Natl. Acad. Sci. USA* **2002**, *99*, 14059.

[25] P. E. Hughes, F. DiazGonzalez, L. Leong, C. Y. Wu, J. A. McDonald, S. J. Shattil, M. H. Ginsberg, *J. Biol. Chem.* **1996**, *271*, 6571.

[26] A. N. Lupas, M. Gruber, *Adv. Protein Chem.* **2005**, *70*, 37.

[27] M. Kim, C. V. Carman, T. A. Springer, *Science* **2003**, *301*, 1720.

[28] T. S. Ulmer, B. Yaspan, M. H. Ginsberg, I. D. Campbell, *Biochemistry* **2001**, *40*, 7498.

[29] D. Schneider, D. M. Engelman, *J. Biol. Chem.* **2004**, *279*, 9840.

[30] T. E. O'Toole, Y. Katagiri, R. J. Faull, K. Peter, R. Tamura, V. Quaranta, J. C. Loftus, S. J. Shattil, M. H. Ginsberg, *J. Cell Biol.* **1994**, *124*, 1047.

[31] A. Senes, D. E. Engle, W. F. DeGrado, *Curr. Opin. Struct. Biol.* **2004**, *14*, 465.

[32] R. Alon, M. L. Dustin, *Immunity* **2007**, *26*, 17.

[33] S. Tadokoro, S. J. Shattil, K. Eto, V. Tai, R. C. Liddington, J. M. de Pereda, M. H. Ginsberg, D. A. Calderwood, *Science* **2003**, *302*, 103.

[34] K. L. Wegener, A. W. Partridge, J. Han, A. R. Pickford, R. C. Liddington, M. H. Ginsberg, I. D. Campbell, *Cell* **2007**, *128*, 171.

[35] K. L. Wegener, I. D. Campbell, *Mol. Membr. Biol.* **2008**, *25*, 376.

[36] K. E. Gottschalk, *Structure* **2005**, *13*, 703.

[37] K. E. Gottschalk, P. D. Adams, A. T. Brunger, H. Kessler, *Protein Sci.* **2002**, *11*, 1800.

[38] K. E. Gottschalk, H. Kessler, *FEBS Lett.* **2004**, *557*, 253.

[39] W. Li, D. G. Metcalf, R. Gorelik, R. H. Li, N. Mitra, V. Nanda, P. B. Law, J. D. Lear, W. F. DeGrado, J. S. Bennett, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 1424.

[40] X. Lin, S. M. Tan, S. K. A. Law, J. Torres, *Proteins Struct. Funct. Bioinf.* **2006**, *63*, 16.

[41] X. Lin, S. M. Tan, S. K. A. Law, J. Torres, *Proteins Struct. Funct. Bioinf.* **2006**, *65*, 274.

[42] A. W. Partridge, S. C. Liu, S. Kim, J. U. Bowie, M. H. Ginsberg, *J. Biol. Chem.* **2005**, *280*, 7294.

[43] V. C. Hirschfeld-Warneken, M. Arnold, A. Cavalcanti-Adam, M. Lopez-Garcia, H. Kessler, J. P. Spatz, *Eur. J. Cell. Biol.* **2008**, *87*, 743.

[44] V. C. Hirschfeld-Warneken, M. Arnold, A. Cavalcanti-Adam, M. Lopez-Garcia, H. Kessler, J. P. Spatz, *Eur. J. Cell Biol.* **2008**, *87*, 743.

[45] M. Arnold, V. C. Hirschfeld-Warneken, T. Lohmuller, P. Heil, J. Blummel, E. A. Cavalcanti-Adam, M. Lopez-Garcia, P. Walther, H. Kessler, B. Geiger, J. P. Spatz, *Nano Lett.* **2008**, *8*, 2063.

Journal of **Molecular Recognition**

Protein–surface interactions: challenging experiments and computations

Ori Cohavi[a], Stefano Corni[b]*, Francesca De Rienzo[b,c], Rosa Di Felice[b], Kay E. Gottschalk[d], Martin Hoefling[d], Daria Kokh[e], Elisa Molinari[b,f], Gideon Schreiber[a], Alexander Vaskevich[a] and Rebecca C. Wade[e]

Protein–surface interactions are fundamental in natural processes, and have great potential for applications ranging from nanotechnology to medicine. A recent workshop highlighted the current achievements and the main challenges in the field. Copyright © 2009 John Wiley & Sons, Ltd.

INTRODUCTION

Several natural processes involve the interaction between polypeptides and solid surfaces (Gray, 2004). To cite a few, the growth of hard tissue is regulated by protein–mineral interactions, the adhesion of organisms to solid supports (e.g. mussels to rocks) depends on protein–surface interactions and the anti-freezing capability of animals living in cold environments relies on proteins binding to ice.

In recent years, the scope of polypeptide–surface interactions has enlarged. Proteins inside the body interact with the artificial materials of implants, an interaction that may induce unwanted reactions or may also be exploited to perform useful tasks (bioactive materials). In addition, the natural recognition capabilities of proteins have inspired new bottom-up approaches to the self-assembly of nanostructures (Sarikaya *et al.*, 2003). Moreover, the increasing distribution of man-made nanoparticles (NPs) naturally raises concerns about possible toxic effects when they accidentally contact living organisms (Fubini *et al.*, 2006; Nel *et al.*, 2006). Such effects will likely involve the interactions of nanomaterials with the protein arsenal of the body.

Protein–surface interactions were the focus of a recent workshop held in Sestri Levante (Italy) on 4–6 June 2009 (www.s3.infm.it/prosurf_meeting) organized in the context of the FP6 EU project Prosurf (www.s3.infm.it/prosurf). The workshop brought together about 60 scientists from around the world working on this topic from different perspectives, ranging from basic science to applications, and from experiments to computation. Here we report on the main themes that emerged at the workshop, and discuss some of the important new results presented there.

Selectivity of proteins/peptides for surfaces

Despite the potential of surface recognition by proteins for applications and its role in natural processes, the basic mechanisms determining affinity and specificity are still poorly understood. Experimental results on the selection of peptides that possess specific binding properties to different types of surfaces were described by Candan Tamerler, Istanbul Tech. Univ.,

Turkey (Tamerler and Sarikaya, 2009). To select an initial pool of binding peptides, her team proceeds by combinatorial biology methods such as phage and cell display. According to their work, obtaining consensus sequence is usually not the case. They apply post-selection engineering to improve the peptide affinity: the binding of peptides is characterized experimentally (e.g. by quartz crystal microbalance (QCM) and surface plasmon resonance (SPR)), then evolutionary engineering approaches and bioinformatics tools are combined to obtain the next generation peptides with tailored properties. Their knowledge-based approach results in peptide sequences that can bind selectively to a desired material (Au, Pt, SiO$_2$). Tamerler stressed that high affinity and specificity of peptides to various surfaces can be achieved, and that proper *characterizing of the peptides is crucial* (Tamerler *et al.*, 2006). The surface specific peptides can be used as linkers (e.g. for immobilization of alkaline phosphatase on gold), to control material assembly, form films and scaffolds or as catalysers in the controlled formation of inorganic NPs. Furthermore, the specific recognition can be used to guide

* Correspondence to: S. Corni, CNR-INFM National Research Center S3, Modena, Italy.
E-mail: stefano.corni@unimore.it

a   O. Cohavi, G. Schreiber, A. Vaskevich
    Weizmann Institute of Science, Rehovot, Israel

b   S. Corni, F. De Rienzo, R. Di Felice, E. Molinari
    CNR-INFM National Research Center S3, Modena, Italy

c   F. De Rienzo
    Department of Chemistry, University of Modena and Reggio Emilia, Modena, Italy

d   K. E. Gottschalk, M. Hoefling
    Ludwig Maximilian University, Munich, Germany

e   D. Kokh, R. C. Wade
    EML Research, Heidelberg, Germany

f   E. Molinari
    Department of Physics, University of Modena and Reggio Emilia, Modena, Italy

proteins towards specific surfaces, including in living organisms (like teeth or bone).

A different approach to selectivity towards surfaces was presented in the work by Arbel Artzy-Schnirman *et al.*, Technion, Israel. Arguing that short peptides are too flexible to be able to recognize surfaces selectively, in particular different facets of the same material, these scientists focused on methods to attain selectivity by using libraries of antibodies with a phage display approach. They demonstrated that *antibodies can indeed select different facets of the same material* (GaAs), a task that was not fulfiled by peptides (Artzy-Schnirman *et al.*, 2006). The antibody library was also used to select an antibody able to discriminate between the two states of an 'electronic antigen' comprised of a hydroquinone monolayer assembled on a gold electrode (Artzy-Schnirman *et al.*, 2008). An oxidative pulse switches the monolayer to the benzoquinone state which is recognized by the specific antibody. A subsequent opposite pulse reduces the monolayer back to its hydroquinone state and releases the antibody. These developments show the potential use of protein–surface interactions as an *electronic switch*, which provides a means for *external control of biological pathways*.

Quantifying the affinity of proteins to surfaces

The need for quantifying the binding affinity of proteins/peptides to surfaces was mentioned in several presentations. The often irreversible character of the adsorption was remarked (see also the next paragraph). Besides the above-mentioned QCM and SPR studies by Tamerler, SPR measurements of peptide adsorption on functionalized gold surfaces were reported by Robert Latour, Clemson Univ., USA (Wei and Latour, 2009). A workflow for measuring the binding of peptides and proteins to gold NPs was described by Ori Cohavi, Weizmann Inst., Israel. This method utilizes high-throughput, real-time SPR measurements. To characterize the binding of amino acids to the NPs, a set of fusions of short homopeptides to the beta-lactamase inhibitor protein (BLIP) were prepared. Wild-type BLIP binds poorly to the Au NPs, allowing for the measurements of the binding strength of each of the fused polypeptides to gold. Synthesized 3-mer homopeptides of some amino acids confirmed the fusion protein results. All the fusions bound irreversibly to the Au NPs, suggesting a rather different mechanism of binding from that observed between proteins. The adsorption of wild-type BLIP and BLIP-His fusion protein to gold was also measured in a different system with well-defined, atomically flat (111) surfaces and found to give the same results (Alexander Vaskevitch, Weizmann Inst., Israel). These data indicate that *Au NPs can be used as a model for the determination of quantitative binding constants of proteins to Au surfaces*.

In addition to SPR, high-throughput fluorescent-based detection (Marcy *et al.*, 2008) can be adapted to study protein adsorption (Claude Weisbuch, Ecole Polytechnique, France and UCSB, USA). Biochips with spatially selective adsorption of different proteins (Weisbuch; François Rossi, IHCP Joint Research Centre, EU) or surface spots with variable chemistry/surface morphology can be used in this format (Rossi; Paolo Milani, Univ. of Milan) (Giorgetti *et al.*, 2008; Ceriotti *et al.*, 2009).

The structure of proteins at surfaces

The question of whether proteins unfold on a surface or maintain their native conformation is not only fundamental to interpreting

the adsorption data, but it also determines how the adsorbed proteins interact with other agents (e.g. how blood plasma proteins adsorbed on NPs elicit immune response). Latour presented measurements of the activity of lysozyme bound to surfaces. He showed that the reduction in activity was greater when the enzyme was bound to hydrophilic than to hydrophobic surfaces, although changes in the protein secondary structure were the same. The difference in activity may be the consequence of different orientations of lysozyme on the two kinds of surfaces.

Ornella Cavalleri, Univ. of Genoa, Italy, remarked the key role of surface hydrophobicity and crystalline structure in the protein adsorption onto highly ordered pyrolytic graphite, HOPG (Svaldo-Lanero *et al.*, 2008). Proteins (e.g. lysozyme) create ordered striped domains on HOPG, losing their native folding. The stripe periodicity is the same irrespectively of the protein structural properties. Interestingly, such patterned interface works as template to align amyloid protofilaments.

Electron Spin Resonance (ESR) of spin-labelled proteins was applied, in association to other techniques, to determine the orientation of adsorbed species and for the assessment of proteins adsorbed on silica NPs (Ivana Fenoglio, Univ. of Turin, Italy). Interesting results were found in studies of peptide adsorption on GaAs (Karsten Goede, Univ. of Leipzig, Germany). Changes in the molecular structure upon adsorption were detected by Fourier transform infrared spectroscopy (FTIR). It was found that adsorbed peptides contain only small amounts of water, enabling FTIR spectroscopy without the use of D_2O. This finding is also important for *experimental verification of the role of water molecules* present at the surface–protein interface (see next paragraph). Rossi presented an NMR characterization of a model protein, ubiquitin, on a Au NP. Ubiquitin remained folded upon binding and these measurements permitted the protein residues affected by adsorption to be identified.

Heiko Seeger, CNR-INFM Natl. Ctr. S3, Italy, presented Atomic Force Microscopy results on the lateral redistribution of the KcsA ion channel in supported lipid bilayers upon the formation of lipid domains. Finally, the capability of hydrophobin to form layers on surfaces (Houmadi *et al.*, 2008) was discussed by Michele Giocondo, CNR-INFM Licryl Lab., Italy, while the use of surfaces to obtain protein crystals (Tosi *et al.*, 2008) was presented by Giuseppe Falini, Univ. of Bologna, Italy.

The importance of the environment: the role of water and ions

A recurring observation was the importance of water for protein–surface binding. Water depletion is observed near hydrophobic surfaces due to the orientation of water molecules in the first layer to hydrogen bond to each other rather than to the surface (Sedlmeier *et al.*, 2008). This water is more mobile than on a hydrophilic surface, affecting protein diffusion as demonstrated by MD simulations (Roland Netz, Tech. Univ. of Munich, Germany). Different computational treatments of the solvent-surface–biomolecule interface effects were presented. Existing implicit solvent models can be successfully used for hydrophobic surfaces, where there is no water between the surface and the biomolecule: hydrogenated Si was discussed by Michael Bachmann, FZ Jülich, Germany (Mitternacht *et al.*, 2007); carbon nanotubes and graphite were presented by Tiffany Walsh, Univ. of Warwick, UK (Tomasio and Walsh, 2009). In other cases, implicit-solvent methods were found to be unsatisfactory (Latour,

2008) or should be designed specifically for the surface considered (as discussed by Rebecca Wade, EML Research, Heidelberg, Germany). Some MD simulations with explicit water show that, *for strongly polar surfaces, binding is not to the surface itself, but rather to a structured water layer on the surface* (Walsh; Netz; Lucio Colombi Ciacchi; Univ. of Bremen, Germany; Susanna Monti, CNR-IPCF, Italy; Monti, 2007; Skelton *et al.*, 2009). The importance of hydration at metal surfaces was clearly demonstrated by Luigi Delle Site, MPI Mainz, Germany, who showed computationally how amino acids with similar affinities for Pt *in vacuo* behave quite differently in water (Ghiringhelli *et al.*, 2008). Not only water, but also *ions play a fundamental role in controlling interactions at interfaces*. Alkali ion specificity was shown by Uri Sivan, Technion, Israel, for the forces between two silica surfaces (Dishon *et al.*, 2009). As the ion concentration was varied, even inversion of the force, from repulsive to attractive, was observed. This behaviour was related to the different hydration properties of ions (strength of the hydration shell and the consequent effective hydrated ion radius). From the computational side, Netz showed how existing force field (ff) parameter sets for ions are not reliable and contradict each other. Netz presented instead a natural-ionic ff parameter set (Horinek *et al.*, 2009) constructed from the experimental solvation data (comparing computed vs. experimental solvation-free energies). They found that hydrophobicity of the ion, rather than its polarizability, is the key to explaining ion behaviour at the water/vapour interface.

The challenges for computations

Experiments can be effectively complemented by theoretical and computational approaches based on coarse-grained simulations, classical molecular dynamics and density functional theory, as demonstrated by several presentations. Protein–surface systems cover a huge diversity which poses *a challenge for the derivation of accurate, consistent force fields*. One needs to consider the properties of the surface material and the protein, as well as the aqueous solution in which they are immersed (Cole *et al.*, 2007). The level of details necessary to obtain reliable results depends on the properties under study. While most computational studies presented in the workshop were based on atomistic molecular mechanics models, there were also quantum mechanical (QM) studies (Susan Köppen, Univ. of Bremen, Germany, Köppen *et al.*, 2008; Stefano Corni, CNR-INFM Natl. Ctr. S3, Italy, Delle Site) and models with continuum solvent (Walsh, Bachmann) or further coarse graining (Delle Site). Most MD simulations rely on an established ff for proteins, with tailored additions aimed at describing the interaction with the surface. For polymer surfaces, a hybrid ff is being developed by Latour: it combines the class-I ff, CHARMM, for proteins and the class-II ff, PCFF, for polymers;

further terms that describe the interface are derived to fit experimental data (Latour, 2008). Another approach is to develop parameters for the protein–surface and water–surface interactions based on QM calculations and the available experimental data (Iori *et al.*, 2009). MD simulations in explicit water (Martin Hoefling, Ludwig Maximilian Univ., Germany) based on these parameters can then form the basis for developing implicit solvent models and coarser protein descriptions able to explore the adsorption preferences of entire proteins (Wade and Kokh, EML Research, Heidelberg, Germany). *Force field evaluation or modification needs experimental benchmark data sets*, such as the peptide–surface interaction set reported by Latour, or the peptide–gold binding studies presented by Cohavi.

Extensive sampling of conformations is another fundamental problem to be addressed for gaining insights into the binding process. Latour highlighted the importance of both a proper combination of ffs and the improvement of sampling techniques. He described temperature intervals with global exchange of replicas (TIGER2), a sampling algorithm that has a higher efficiency than conventional replica-exchange molecular dynamics, REMD (Li *et al.*, 2009). Bachmann exploited Monte Carlo based sampling techniques to study peptides known to bind to hydrogenated Si. He reproduced the experimental relative affinities of the peptides and predicted the relative binding order of new peptides, which was experimentally confirmed afterwards.

A key issue in the context of simulations is *bridging time and length scales*. An adaptive simulation scheme was proposed by Delle Site (Praprotnik *et al.*, 2008). Here, the level of coarse graining is spatially controlled in a consistent way during the simulation, thus allowing the treatment of interactions with the appropriate detail. Although this approach has not yet been applied to the protein–surface problem, it would significantly improve the range of accessible sizes and times.

In conclusion, the Prosurf 'Protein Surface Interactions' workshop provided a comprehensive overview of the current achievements and the main challenges of the protein–surface interaction research field. The interdisciplinarity of this research field is one of its intrinsic strengths, but it is also a potential hindrance for developments. Occasions such as the Prosurf workshop are needed to overcome the boundaries between different disciplines and to develop a common language and shared concepts in the emerging scientific community working on protein–surface interactions.

REFERENCES

Artzy-Schnirman A, Zahavi E, Yeger H, Rosenfeld R, Benhar I, Reiter Y, Sivan U. 2006. Antibody molecules discriminate between crystalline facets of a gallium arsenide semiconductor. *Nano Lett.* **6**: 1870–1874.

Artzy-Schnirman A, Brod E, Epel M, Dines M, Hammer T, Benhar I, Reiter Y, Sivan U. 2008. A two-state electronic antigen and an antibody selected to discriminate between these states. *Nano Lett.* **8**: 3398–3403.

Ceriotti L, Buzanska L, Rauscher H, Mannelli I, Sirghi L, Gilliland D, Hasiwa M, Bretagnol F, Zychowicz M, Ruiz A, Bremer S, Coecke S, Colpo P,

Rossi F. 2009. Fabrication and characterization of protein arrays for stem cell patterning. *Soft Matter* **5**: 1406–1416.

Cole DJ, Csányi G, Payne MC, Spearing SM, Colombi Ciacchi L. 2007. Development of a classical force field for the oxidised Si surface: application to hydrophilic wafer bonding. *J. Chem. Phys.* **127**: 204704.

Dishon M, Zohar O, Sivan U. 2009. From repulsion to attraction and back to repulsion: the effect of NaCl, KCl, and CsCl on the force between silica surfaces in aqueous solution. *Langmuir* **25**: 2831–2836.

Fubini B, Fenoglio I, Martra G, Ceschino R, Tomatis M, Cavalli R, Trotta M. 2006. An overview on the toxicity of inhaled nanoparticles. In Surface

Chemistry in Biomedical and Environmental Science, Blitz JP, Gun'ko VM (eds). Springer: Amsterdam; 241–252.

Ghiringhelli LM, Hess B, van der Vegt NFA, Delle Site L. 2008. Competing adsorption between hydrated peptides and water onto metal surfaces: from electronic to conformational properties. *J. Am. Chem. Soc.* **130**: 13460–13464.

Giorgetti L, Bongiorno G, Podesta A, Berlanda G, Scopelliti PE, Carbone R, Milani P. 2008. Adsorption and stability of streptavidin on cluster-assembled nanostructured TiO$_x$ films. *Langmuir* **24**: 11637–11644.

Gray JJ. 2004. The interaction of proteins with solid surfaces. *Curr. Opin. Struct. Biol.* **14**: 110–115.

Horinek D, Mamatkulov SI, Netz RR. 2009. Rational design of ion force fields based on thermodynamic solvation properties. *J. Chem. Phys.* **130**: 124507.

Houmadi S, Ciuchi F, De Santo MP, De Stefano L, Rea I, Giardina P, Armenante A, Lacaze E, Giocondo M. 2008. Langmuir-Blodgett film of hydrophobin protein from *Pleurotus ostreatus* at the air–water interface. *Langmuir* **24**: 12953–12957.

Iori F, Di Felice R, Molinari E, Corni S. 2009. GolP: an atomistic force-field to describe the interaction of proteins with Au(111) surfaces in water. *J. Comp. Chem.* **30**: 1465–1476.

Köppen S, Bronkalla O, Langel W. 2008. Adsorption configurations and energies of amino acids on anatase and rutile surfaces. *J. Phys. Chem. C* **112**: 13600–13606.

Latour RA. 2008. Molecular simulation of protein–surface interactions: benefits, problems, solutions, and future directions. *Biointerphases* **3**: FC2–FC9.

Li XF, Latour RA, Stuart SJ. 2009. TIGER2: an improved algorithm for temperature intervals with global exchange of replicas. *J. Chem. Phys.* **130**: 174106.

Marcy Y, Cousin PY, Rattier M, Cerovic G, Escalier G, Bena G, Gueron M, McDonagh L, le Boulaire F, Benisty H, Weisbuch C, Avarre JC. 2008. Innovative integrated system for real-time measurement of hybridization and melting on standard format microarrays. *Biotechnology* **44**: 913–920.

Mitternacht S, Schnabel S, Bachmann M, Janke W, Irbäck A. 2007. Differences in solution behavior among four semiconductor-binding peptides. *J. Phys. Chem. B* **111**: 4355–4360.

Monti S. 2007. RAD16II beta-sheet filaments onto titanium dioxide: dynamics and adsorption properties. *J. Phys. Chem. C* **111**: 16962–16973.

Nel A, Xia T, Madler L, Li N. 2006. Toxic potential of materials at the nanolevel. *Science* **311**: 622–627.

Praprotnik M, Delle Site L, Kremer K. 2008. Multiscale simulation of soft matter: from scale bridging to adaptive resolution. *Annu. Rev. Phys. Chem.* **59**: 545–571.

Sarikaya M, Tamerler C, Jen AK-Y, Schulten K, Baneyx F. 2003. Molecular biomimetics: nanotechnology through biology. *Nat. Mater.* **2**: 577–585.

Sedlmeier F, Janecek J, Sendner C, Bocquet L, Netz RR, Horinek D. 2008. Water at polar and nonpolar solid walls. *Biointerphases* **3**: FC23–FC39.

Skelton AA, Liang T, Walsh TR. 2009. Interplay of sequence, conformation and binding at the titania–peptide interface as mediated by water ACS. *Appl. Mater. Interfaces* **1**: 1482–1491.

Svaldo-Lanero T, Penco A, Prato M, Canepa M, Rolandi R, Cavalleri O. 2008. Nanopatterning by protein unfolding. *Soft Matter* **4**: 965–967.

Tamerler C, Sarikaya M. 2009. Molecular biomimetics: nanotechnology and molecular medicine utilizing genetically engineered peptides. *Philos. Transact. A* **367**: 1705–1726.

Tamerler C, Duman M, Oren EE, Gungormus M, Xiong X, Kacar T, Parviz BA, Sarikaya M. 2006. Materials specificity and directed assembly of a gold-binding peptide. *Small* **2**: 1372–1378.

Tomasio SM, Walsh TR. 2009. Modeling the binding affinity of peptides for graphitic surfaces. Influences of aromatic content and interfacial shape. *J. Phys. Chem. C* **113**: 8778–8785.

Tosi G, Fermani S, Falini G, Gallardo JAG, Ruiz JMG. 2008. Crystallization of proteins on functionalized surfaces. *Acta Crystallogr. D* **64**: 1054–1061.

Wei Y, Latour RA. 2009. Benchmark experimental data set and assessment of adsorption free energy for peptide–surface interactions. *Langmuir* **25**: 5637–5646.

262

90   *– P2 –*

Barnase–Barstar: From first encounter to final complex

Martin Hoefling, Kay E. Gottschalk *

Chair for Applied Physics, Biophysics and Molecular Materials & CeNS, Ludwig-Maximilians University, Amalienstr. 54, 80799 Munich, Germany

ARTICLE INFO

ABSTRACT

Formation of transient protein complexes is an important process in cells. Details of the association process as well as the energy landscapes of association are not well understood. In particular, the nature, height and position of the energy barriers during complexation are debated. Computational studies are well suited for atomistically investigating protein association processes. The Barnase–Barstar complex constitutes a well-studied target for computational studies as a small system with fast association rates. Here, we performed constraint biased Molecular Dynamics simulations along the reaction coordinate reaching from the diffusion regime to the bound complex. We simulated the wild-type and different mutants at different salt concentrations. A structural analysis of our simulation trajectories revealed not a single, but two distinct association patterns dominated by an interplay between two charged contact points near the binding site. Electrostatics and/or mutations influence the relative population of these patterns. Further, we computed the energy landscape of association as PMF (Potential of Mean Force) profiles within a reasonable agreement to experiment. We find a single energy barrier at a distance of ∼0.3 nm, which corresponds to the final desolvation transition. Electrostatics has a profound influence on the height of this energy barrier, but not on its position.

© 2010 Elsevier Inc. All rights reserved.

1. Introduction

Transient protein–protein complexes regulate the cellular life. Therefore, the complexation and the recognition of the complex partners need to be fine-tuned, specific and reliable. Investigations of the fundamentals of protein–protein complexation require suitable model systems. One well established model system is the ribonuclease Barnase and its inhibitor Barstar. Barnase–Barstar is one of the best-studied systems by experiments and in theory (Hartley, 1989). This complex has also found applications in bioengineering (Strittmatter et al., 1995). Barnase is deadly to the hosting organism and is excreted. For protection, it is complexed with Barstar prior to export. This interaction needs to be very fast and strong to prevent cell death. The complexation is mainly driven by electrostatics leading to the fast association. With an on-rate of $10^8 s^{-1} M^{-1}$ (Buckle et al., 1994), it is one of the fastest known in biological systems. Together with its moderate size, this system provides an excellent target for a detailed examination of complexation processes in silico (Janin, 1997; Gottschalk et al., 2004).

We aim here at analyzing the association behavior of this complex. Protein–protein association can be divided into three steps: free diffusion, steered diffusion and desolvation (Janin, 1997). Some proteins reach the diffusion limit for complexation (North-rup and Erickson, 1992; Schreiber and Fersht, 1996), despite the fact that the rotational degrees of freedom hamper the protein recognition process. Successful complexation is strongly facilitated if the proteins are pre-aligned before contact (Janin, 1997). This implies that during complexation certain relative orientations of the two proteins are overrepresented, forming specific patterns of complexation. In order to detect these complexation patterns, the focus of this study is the structural and energetical analysis of the last complexation step starting from encounter complexes. Factors influencing the association process in this regime include electrostatic complementary regions of the complex, desolvation and in some cases structural reorganization in the interface region (Dong and Zhou, 2006).

Computational methods have been shown to be well suited for analyzing protein–protein structures and predicting conformations (Gottschalk and Kessler, 2004) and interactions (Gottschalk et al., 2002; Grater et al., 2005). Several previous studies on Barnase–Barstar and similar systems focused on the steps towards forming an encounter complex (Janin, 1997; Gabdoulline and Wade, 1998; Camacho et al., 1999; Frisch et al., 2001). Due to computational restrictions, these studies are typically performed using implicit water models, which is a good approximation for well-separated proteins. However, recent studies have shown that a structural modulation of water by solvated specimen can greatly influence the physical properties of the solvent (Abseher et al., 1996; Gallagher and Sharp, 2003; Despa et al., 2004; Yang and Sharp, 2005) and thus affect the modeling of the short-range interactions. Since

---

* Corresponding author. Fax: +49 89 2180 2050.
E-mail addresses: kay.gottschalk@physik.uni-muenchen.de, kay.gottschalk@gmail.com (K.E. Gottschalk).

we are interested in the last steps of association including short-range interactions between the complex partners, we explicitly treated the surrounding environment, namely water and ions. This is required to correctly model the barriers in the association process (Schröder et al., 2006).

Although it is well established that electrostatics plays a major role for complexation in this system (Schreiber and Fersht, 1996), the exact effects of changes in the electrostatics of the system such as salt concentrations or mutations involving charged residues on the system are still under debate (Sheinerman et al., 2000; Lee and Tidor, 2001; Novotny and Sharp, 1992). Molecular Dynamics has been suggested as a good source of information about the role of electrostatics in protein association by Dong et al. (2003).

To obtain information about both structural and energetic aspects of the association of Barnase–Barstar and the influence of electrostatics, we performed constraint-biased simulations along the reaction coordinate, the distance between the Centers of Masses (COM). To this end, we separated the proteins along the vector connecting the COMs (Fig. 1) and simulated the system with constrained distance of the COMs, while all other degrees of freedom were left unconstrained. We have simulated the wild-type and differently charged mutations under varying salt conditions. This yields not only information about free energy changes during association (Trzesniak et al., 2007; Schlitter et al., 2001) under a variety of electrostatic conditions, the overall small perturbance of our system along the reaction coordinate furthermore allows detailed insight into structural processes accompanying the complexation.

The here computed interaction energy of −46.9 kJ/mol of the wild-type complex of Barnase with its inhibitor Barstar at 150 mM NaCl is in reasonable agreement with experimental studies reporting an interaction energy of −79.8 kJ/mol at ionic conditions of 50 mMol Tris–HCl buffer (Schreiber and Ferscht, 1995). In our simulations, we find two different overrepresented relative orientations during association for the wild-type complex under physiological ion concentrations, arguing for the existence of multiple pathways of complexation. The energy landscape of association has a single barrier at a separation distance of ∼0.3–0.4 nm. Electrostatics strongly influences the height of this barrier, however hardly its location.

2. Material and methods

2.1. Simulations

The initial point of our studies is the crystal structure of the Barnase–Barstar complex (1BRS) (Buckle et al., 1994). A single complex (chain A and E) from the crystal structure was used for the wild-type and mutations. From the obtained initial complex structure, Barstar was simulated in 111 steps along the vector connecting the COMs up to a final separation distance of 5.6 nm (Fig. 1). Barnase and Barstar additionally were rotated separately in all three directions randomly up to 15°. The system was solvated with SPC water in truncated octahedron boxes with a distance of 1.1 nm to the boundaries avoiding direct protein–protein interactions from van-der-Waals and Coulomb interactions by separating the protein by at least twice the cutoff radius from its image. Studies on two-ion model systems showed that the effect of the periodicity together with long-range electrostatic treatment via PME has only a small effect on the computed Potential of Mean Force (Bergdorf et al., 2003), thus the overall effect of periodicity of our system on the computed free energy differences should be small. Ions were added to neutralize the net charge to avoid artifacts from Particle Mesh Ewald background charges when calculating free energy differences (Donnini et al., 2005). Additionally, we achieved physiological ion concentrations of 150 mMol NaCl by adding additional ions in a subset of simulations. Altogether, eight different setups were studies as summarized in Table 2. The OPLS-AA force field was used for all simulations.

The box was scaled in a separate 500 ps simulation coupled to Parinello-Rahman barostat while temperature coupling to 300 K with Nose–Hoover thermostat to protein and non-protein parts was used in all simulations. On the result of these runs a distance constraint was established between the COMs of Barnase and Barstar via the SHAKE algorithm (Ryckaert et al., 1977). Hereby, in every step, the atoms of the pull group are translated according to the new COM position to fulfil the constraint. Then, the constraint force is calculated by $F = m_{tot} \, dr_{COM} \, dt^{-2}$, where m_{tot} is the total mass of the constrained group, dr_{COM} the displacement vector to fulfill the constraint and dt the integration time. Simulations of 5 ns were performed with temperature coupling to 300 K. All simulations were performed with 2 fs integration steps and PME with a Fourier grid of 0.12 nm above 1.1 nm to avoid artifacts from a plain coulomb cutoff on the protein and water (Schreiber and Steinhauser, 1992; van der Spoel and van Maaren, 2006). Switch-cutoff for van-der-Waals interactions with a switch radius of 0.9 nm and cutoff at 1.0 and the Gromacs 4.0 package was used for all simulations (Hess et al., 2008; Kutzner et al., 2007; van der Spoel et al., 2005).

2.2. Potential of Mean Force

The obtained average constraint forces from the simulations of the last 3 ns of the simulations where integrated and corrected for their entropic part to obtain the PMF corresponding to $f_R = \frac{\partial U}{\partial R} - \frac{2k_B T}{R}$ and $\Delta F_C(R_{\alpha\beta}) = \int_0^{R_{\alpha\beta}} \langle f_R \rangle dR'$ (Swegat et al., 2003). The standard error
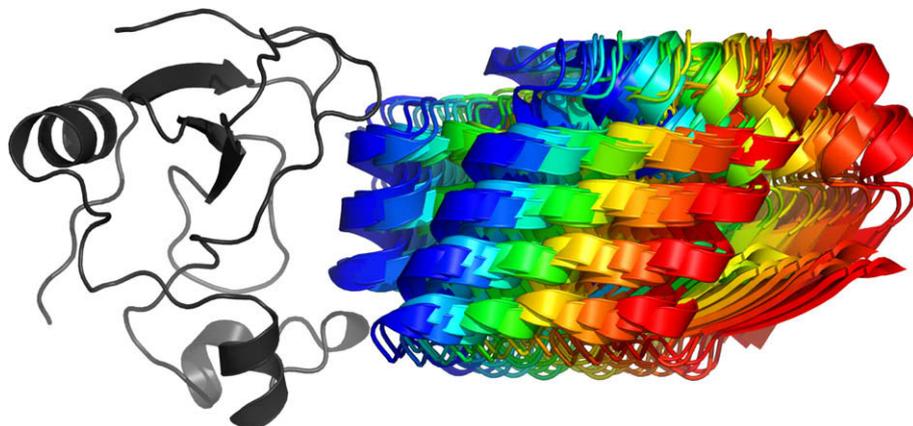
Fig. 1. Initial structures for constraint simulations. We generated a multitude of initial orientations with increasing distance by displacing the proteins along the COM–COM vector.

Table 1
Energy comparison, Molecular Dynamics simulations vs. experiment.

	ΔG_{exp} [kJ/mol] (50 mM Tris–HCl)	ΔG_{PMF} [kJ/mol] (150 mM NaCl)	ΔG_{PMF} [kJ/mol] (counterions only)
Wild-type	−79.8	−46.9 ± 2.1	−69.1 ± 2.0
Bn27 K and Bs39D	−43.2	−27.1 ± 1.7	−36.7 ± 1.8
Bn 59R and Bs76E	−59.2	−33.3 ± 2.8	−28.8 ± 2.2
Bn27 K, Bn59R, Bs39D and Bs76E	n/a	−15.1 ± 1.4	−11.4 ± 1.9

Table 2
Constraint-biased simulations summary.

Setup	Ionic strength	Simulation #	Separation range
wt	150 mMol NaCl	111	2.360–4.966 nm
Mutant 1	150 mMol NaCl	111	2.414–5.435 nm
Mutant 2	150 mMol NaCl	111	2.366–4.964 nm
Mutant 12	150 mMol NaCl	111	2.416–4.886 nm
wt	Na + counterions (neutral box)	111	2.338–5.134 nm
Mutant 1	Na + counterions (neutral box)	111	2.414–4.818 nm
Mutant 2	Na + counterions (neutral box)	111	2.338–5.563 nm
Mutant 12	Na + counterions (neutral box)	110	2.353–5.416 nm

was calculated by first estimating the errors in the individual simulation via block averaging of the force (Hess, 2002) and then propagating the errors in the integration. The obtained standard errors for the potential minimum were between 1.4 kJ/mol and 2.2 kJ/mol, except for mutant 2 in 150 mMol ionic strength with a calculated standard error 2.8 kJ/mol.

2.3. Mutants

Experimental studied mutations revealed a strong effect on binding kinetics (Frisch et al., 1997), in particular when charged residues were mutated. We have chosen two sets of interacting residues from experiments, which showed strong (Bn27K, Bs39D) and moderate (Bn59R–Bs76E) impact on binding energy (Frisch et al., 2001; Schreiber and Ferscht, 1995) when mutated. Each of our double mutations to Alanine (Table 3 and Fig. 2) change the net charge complementary by 2, and lead to Barnase charged with 1*e* in case of mutant 1 and mutant 2 and neutral Barnase for mutant 12. Mutations were created with the mutagenesis tool from PyMol (DeLano, 2009).

2.4. Clustering

We tried to find common structures for analyzing association modes. To this end, the last 3 ns of the trajectories were reduced to 30 snapshots per simulation (100 ps steps). These where then preprocessed to find representative structures for the trajectories. For all snapshots of a trajectory normalized contact maps of all residues of Barnase vs. those of Barstar were calculated with g_mdmat (Lindahl et al., 2001). From the obtained matrices the absolute dif-

Table 3
Setup and system description.

wt	Wild-type Barnase(Bn) – Barstar(Bs) complex (chain A and E from 1BRSBuckle et al. (1994)
Mutant 1	Bn27 K → Bn27A, Bs39D → Bs39A
Mutant 2	Bn59R → Bn59A, Bs76E → Bs76A
Mutant 12	Combined mutations from mutant 1 and mutant 2
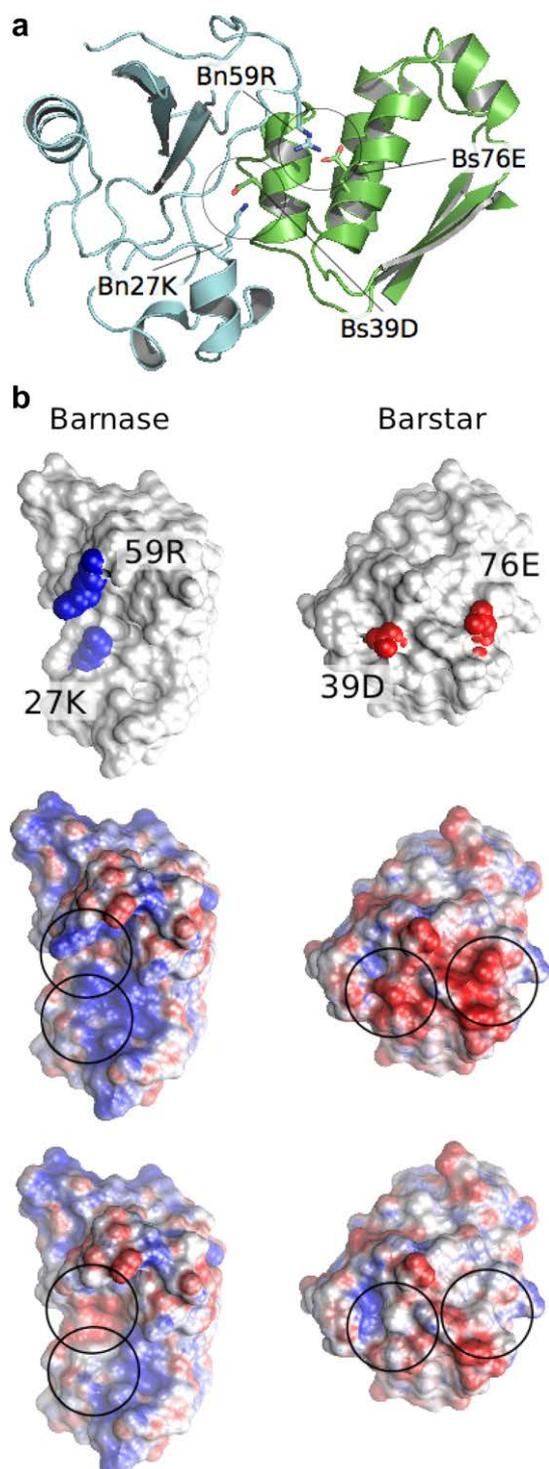
Fig. 2. Mutations performed on Barnase–Barstar. (a) Mutations to alanine performed on 1BRS (Buckle et al., 1994) complex structure. Chain A (Barnase) and E (Barstar) were extracted as input structure for the simulations. The side chains pairs were mutated to Alanine with PyMol (DeLano, 2009) as shown in Table 3. Bn59R is located in the Guanidine binding-loop while Bn27 K is oriented towards the pocket for the binding helix of Barstar. (b) The impact of the mutations on the electrostatics of the complex. The complex is opened with the binding site facing the spectator to show the impact on the interface region. The electrostatic maps were calculated with APBS (Baker et al., 2001) and mapped on surfaces generated by msms (Sanner et al., 1996) with VMD (Humphrey et al., 1996).

ference was calculated for each pair and used as input for preclustering step with CFinder (Palla et al., 2005) and the CPMw (Clique Percolation Method weighted) algorithm. A value of 1.2

was used as intensity threshold. CFinder locates and visualizes overlapping dense groups of nodes in networks, based on the clique percolation method (Palla et al., 2007). For the obtained communities (total count typically in the magnitude of 1–3) and their nodes, a representative structure was selected based on minimal RMSD to all other community members.

The obtained representative frames from all simulations and communities of a setup were then used for clustering the association patterns. In the second clustering step the RMSD between the structures was used as cluster distance metric and an intensity threshold of 15.0 were used in CFinder. The community graphs were calculated from all connected communities with k-cliques of 3. To further describe specific regions with their average COM distance clusters were split up in their higher k-clique constituents for visualization if possible.

3. Results

3.1. Association pattern

First, we were interested in structural variations of the individual proteins during complexation. To investigate this, we calculated the average Root-Mean-Square-Deviation (RMSD) values of our sampled structures of the individual proteins in our simulations compared to the crystal structure of the individual proteins in the complex. For Barnase, a significant increase with growing separation distance can be observed (Fig. 3). For Barstar, a small initial increase is observed. However, the RMSD decreases again for larger separation distances. The most significant changes of Barnase occur in the range up to separations of 1 nm. In general, the RMSD is higher for Barstar compared to Barnase.

We went onto analyze the relative orientation of the proteins to each other during complexation. To initially analyze the individual contribution of the residues to the early complexation steps, we plotted the 10 residue pairs with the smallest distance to the complex partner at each distance from both complexation partners (Fig. 4). This reveals that in early stages of complexation at larger distances, Barnase contacts the approaching partner with three distinct residue regions around the residues K39, R59 and Q104. Mutations on the complexation partners significantly change the role of the residues in early complexation of Barnase. In case of Barstar, Y29 and E80 maintain their key-role in early complexation while residue D39 and E76 (mutant 1 and 2) reveal changes only in shorter separation distances upon mutation. No change in preference of contacts can be observed upon changes in ion concentration for Barnase (Fig. 4) and Barstar (data not shown).

We then analyzed the relative orientations of the proteins in structural terms. To this end, we clustered the structures according to their similarity using the clique percolation method as described in Section 2. During clustering, the RMSD distance of two structures was used to establish a simple and effective metric in configurational space. Interestingly, we find distinct communities in a distance dependent manner (Fig. 5), arguing for multiple overrepresented relative orientations during complexation. Under physiological ion conditions, we find a splitup into two specific association patterns, which merge at a separation distance of approximately 0.4 nm. This is not observed for mutant 2. In the absence of ionic environment, only mutant 1 shows this splitup. In general, the splitup length decreases when mutations are applied.

In the case of the wild-type at physiological ion concentration, we created an exemplary reconstruction of the overrepresented orientations corresponding to the two clusters by ordering the representative frames in our cluster communities (Fig. 6). Here, the hinge region between both patterns is near residue S38-K39 on Barnase, residues previously found to play an important role in early complexation. While the first orientations (0.55 nm/
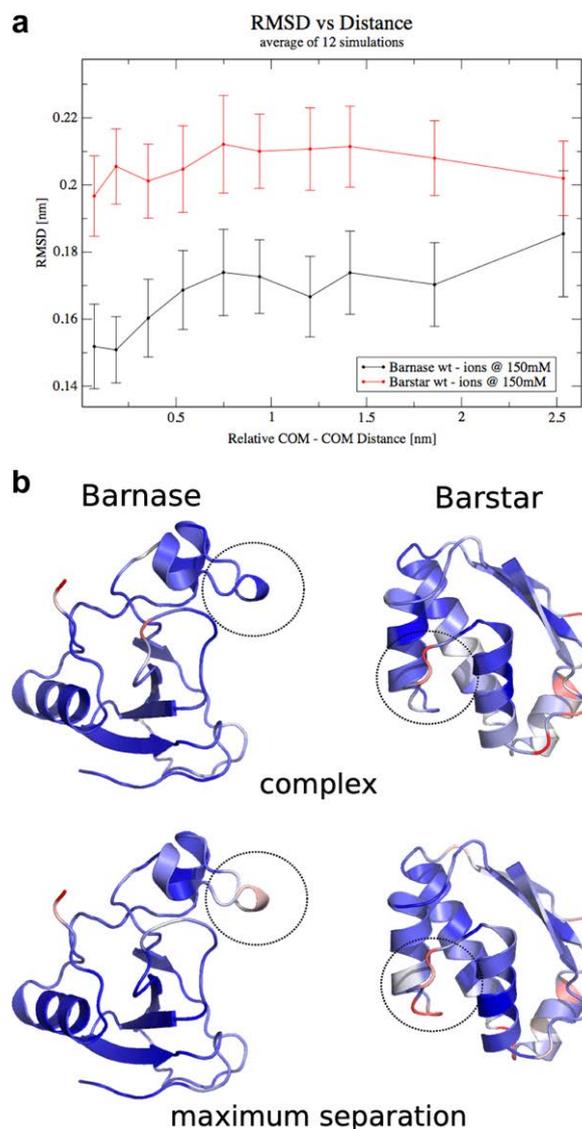
Fig. 3. RMSDs (a) RMSD vs. Distance of Barnase and Barstar. The RMSD of Barnase increases with the distance, in particular in the first nm, indicating a structural reorganization in the close distance. Barstar displays an overall higher flexibility than Barnase with a less pronounced RMSD increase. (b) Map of residue-wise Cα RMSD onto structures, showing the increased structural flexibility at larger separation distances. Red: high RMSD; Blue: low RMSD. In Barnase, the loop region 29–33 shows significant deviations. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

0.74 nm, each corresponding to the COM separation distance in the two major pathways) only show a tilt around the hinge region after splitup, the last two states of our overlapping communities (distances 0.70 nm/0.82 nm and 0.72 nm/0.84 nm) additionally reveal rotations around the connection vector and not only tilted orientations. Both pathways differ in their contact regions along the path. While in the first pathway (red in Fig. 6) Barstar maintains contact via residue R59 in the Guanidine binding-loop of Barnase, this can only be observed for the first distance set of the second pathway.

In order to obtain structural information of the association process at greater distances, we mapped the mutual orientation of the two constituents in our simulation on spherical shells. For small separation distances, the sampling is concentrates on a small spot (Fig. 7). By increasing the distance, the sampled regions increase in area. Additionally, an increase of area can be observed comparing wild-type and the double mutant 12, while changes in ionic environment do not affect the sampled region size.

Barnase wt (150 mMol NaCl)
Residues with minimal distance to Barstar

Barstar wt (150mMol NaCl)
Residues with minimal distance to Barnase

Barnase: LYS27 SER38 LYS39 ARG59 GLN104   Barstar: TYR29 ASP39 GLU76 GLU80

Barnase mutant1 (150 mMol NaCl)
Residues with minimal distance to Barstar

Barnase mutant2 (150 mMol NaCl)
Residues with minimal distance to Barstar

Barnase mutant12 (150 mMol NaCl)
Residues with minimal distance to Barstar

Barnase wt (Counterions only)
Residues with minimal distance to Barstar

Fig. 4. Most frequent contacts vs. separation distance. In these graphs, for each distance the 10 smallest contact distances are shown with symbol size reproducing the order. Each residue can have multiple contacts in this list together with its counterparts. Residues with contacts in early complexation are highlighted as sticks in the opened complex structure below. In terms of contacts – virtually no difference can be observed at different ion concentrations, while mutation completely extinct the early interactions of the corresponding residue region.
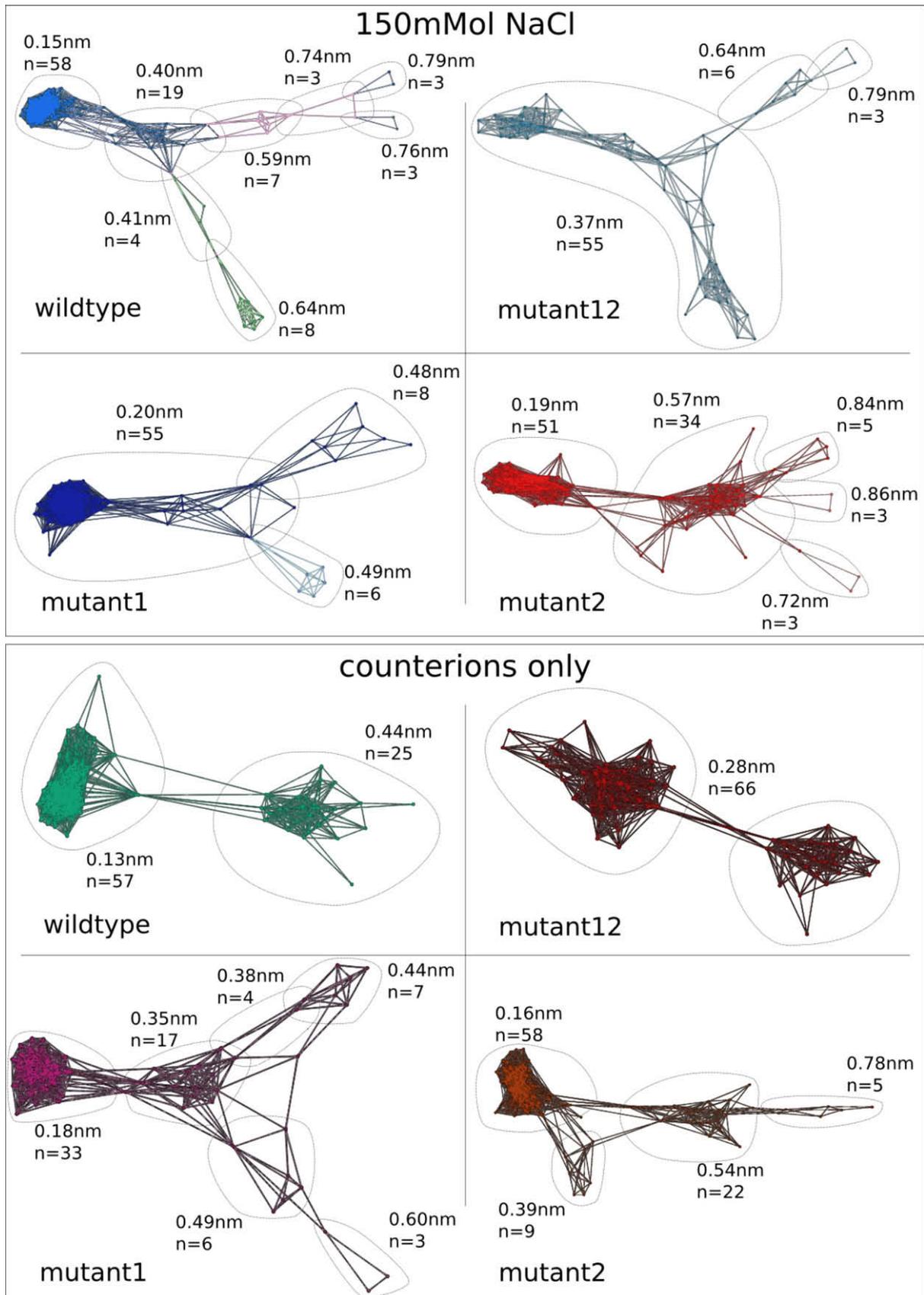
Fig. 5. k-clique community graphs of all setups based on the clique percolation method. Each data point corresponds to a representative frame from simulation trajectories. The metric is defined by the RMSD between the frames. For all complexes, except mutant 2, a splitting up into two domains can observed at a separation of about 0.4 nm at physiological ion concentration. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
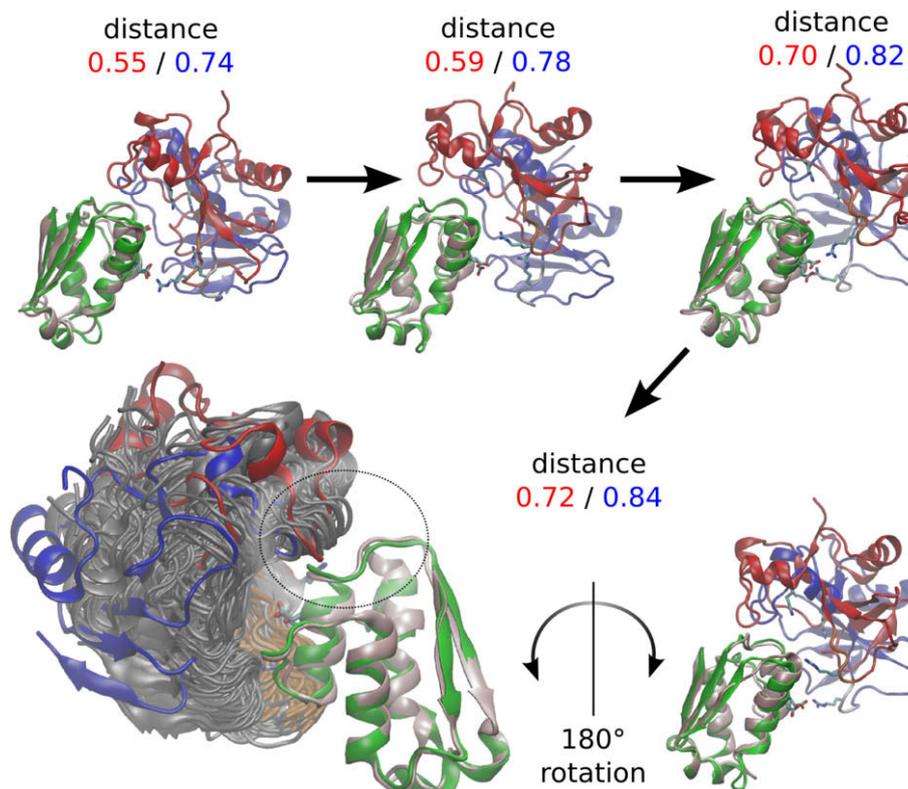
Fig. 6. Association pattern of the wt simulations in physiological ion concentration. The longer reaching pattern (see Fig. 5) is colored in blue, the shorter one in red. Residues used for our mutants are displayed as sticks. All units if not denoted otherwise are nm. The contact region on the opposite site of the Guanidine binding-loop is highlighted in the rotated view of the maximum split distance. Common clusters at shorter separation distances are shown in gray. Patterns differ in their tilt of the binding helix of Barstar as well as the rotation around the COM – connection vector. This also explains the limitation of cluster splitting. For higher distances it has been found that this orientation is only of minor importance for the association (Selzer and Schreiber, 2001). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3.2. Potentials during complex formation

Apart from structural considerations, the mean forces in the constrained simulations (Fig. 8) allow the calculation of the Potential of Mean Force along the reaction coordinate and hence an energetically detailed picture of the association process (Fig. 8). The absolute free energy of association of wt and mutants in different ionic conditions can be found in Table 1. The shape of the free energy profile reveals interesting details. For physiological ion concentration, our profiles reveal a pronounced barrier around 0.25 nm separation. The barrier height is 6.0 kJ/mol in case of the wild-type, 3.8 kJ/mol for mutant 1 as well as 4.6 kJ/mol for the double mutant 12. Exception is mutant 2, where the barrier vanishes and only a plateau is observable at this separation (Fig. 8). For simulations with increased electrostatic attraction due to reduced ion content in the simulated cell (only counterions were added), no barriers can be observed. Another observation is a shift in the location of the absolute minima, which corresponds to deviations from the equilibrium distance in the crystal structure. Here, only mutation 1 and double mutation 12 show a shift in their minimum location of about 0.12 nm. The wild-type complex has the longest reaching interactions compared to the mutants with a significant decrease in energy already at 2 nm separation, arguing for a far-reaching pre-orientation.

4. Discussion

4.1. Conformations of individual proteins

The crystal structures of the Barnase and Barstar individually are very similar to the structures of the two proteins in the com-

plex, showing only minor conformational changes. For Barstar, conformational changes are known to occur in the loop region 29–33 which mainly involves rearrangement of side chains (Buckle et al., 1994; Selzer and Schreiber, 1999) (Fig. 3b). Additionally, Barstar as the inhibitor displays a higher overall flexibility. This is does not hamper complexation as long as the binding helix preserves the conformation still allowing successful inhibition (SolerGonzalez and Fersht, 1997). In our simulations, also no major conformational change is observed. However, an increase of the RMSD can be observed for Barnase. For Barstar, we indeed found an overall higher RMSD comparing to Barnase. Steered Molecular Dynamics simulations on this complex results suggest a higher liability of this protein, with the stable inhibiting binding helix as the strongest element in the inhibitor protein (Neumann and Gottschalk, 2009), in agreement with our results here. The overall change in solvent accessible surface (SAS) during complexation computed here (ΔSAS $\approx$ 20 nm^2, data not shown) is slightly larger than predicted by other calculations (17 nm^2) (Sheinerman and Honig, 2002) and has its origin in the structural relaxation during de-complexation observed in our simulations. Additionally, we did not observe an effect of the ion concentration on the RMSD (data not shown), underlining that electrostatics does not significantly stabilize/destabilize the individual proteins. Hence, effects observed when altering the ionic environment are due to altered interactions between the proteins rather than structural modifications within the proteins.

4.2. Choice of mutations

A wide variety of Barnase–Barstar mutants have been studied experimentally under various conditions (Frisch et al., 2001,
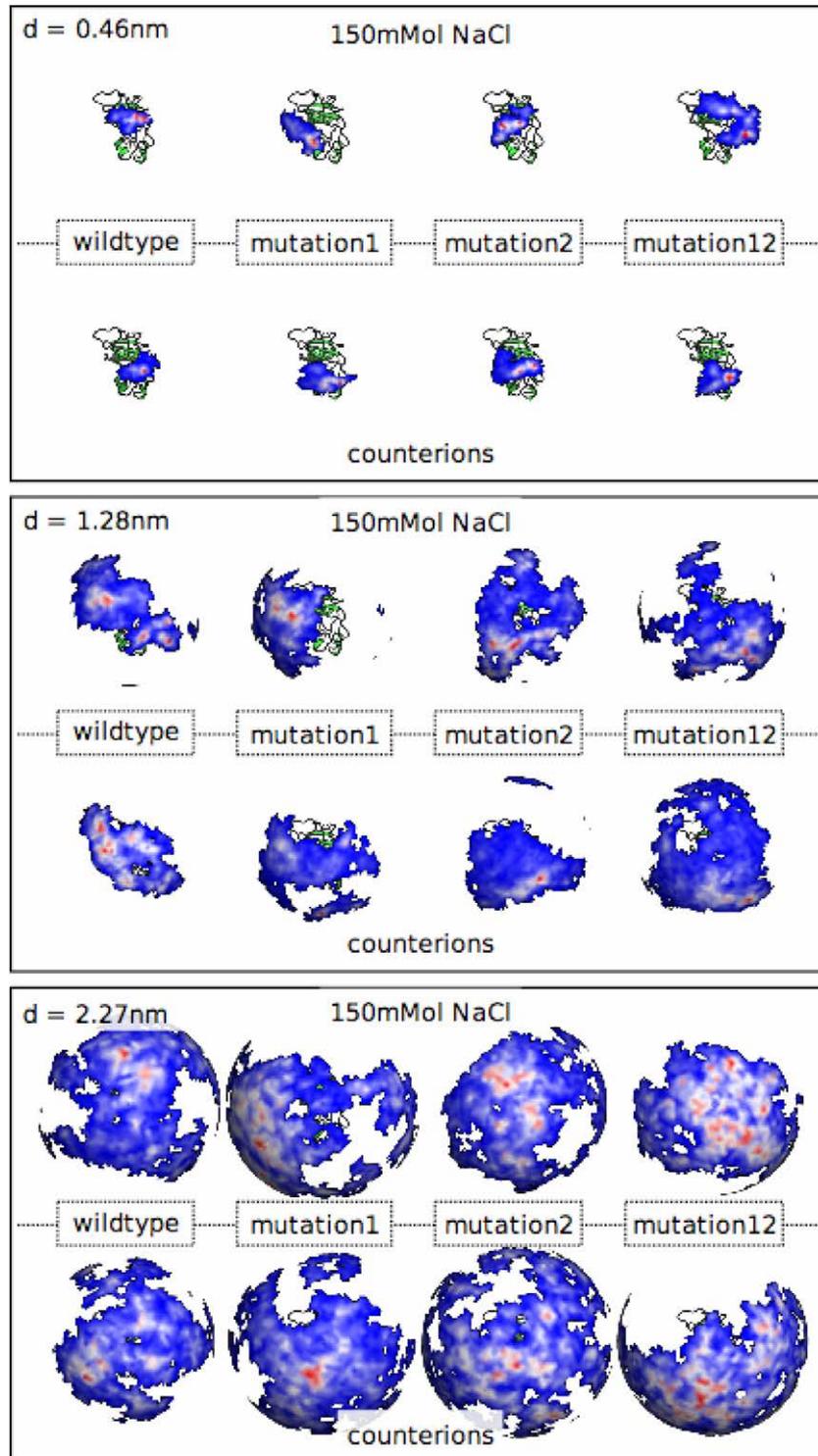
Fig. 7. Probability densities at different distances. In this figure, the normalized probability density of Barstar position from Barnase is calculated from our simulations at three different average distances. Distances are shown with the corresponding average sphere radii of the 10 used simulations. Blue color depicts a low probability while red corresponds to areas with a high probability. The simulations did not sample transparent areas. When mutated, the area increases independent of the size, reducing the probability of a distinct pattern. In the low distance regime, no difference between the ion concentrations is observable while for greater distances – the physiological ion concentration differences in the highly occupied regions are observable. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

1997; Schreiber and Ferscht, 1995) in order to gain insight into contributions of individual amino acids to the association. We decided therefore to not only study the wild-type protein, but also mutants with an altered electrostatic profile. Several consider-ations lead us to the choice of Bn27–Bs39 and Bn59–Bs76 as our mutation set. Barnase and Barstar form a tight complex and most potential mutation targets can interact with multiple residues. Thus the idea is to maximize the impact by mutating both partners
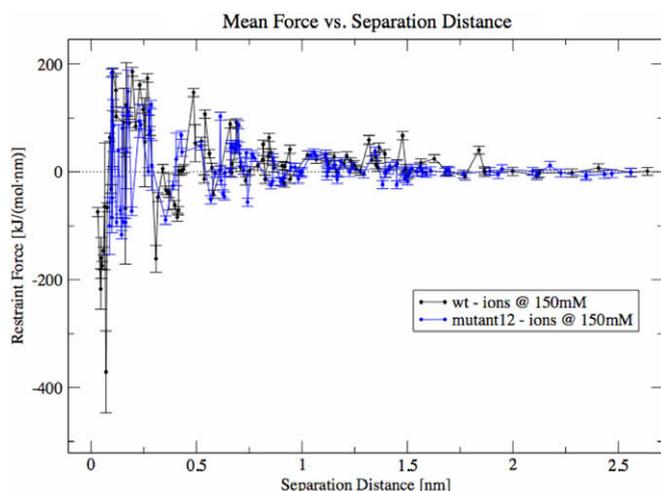
Fig. 8. Computed mean force from a set of distance restrained simulations. Here, wt and mutant 12 setups are shown exemplarily. For each setup – 111 simulations constrained to a certain distance were performed and the constraint force was averaged after equilibration. Errorbars indicate the individual error estimate of the force based on block average (Hess, 2002) of each simulation.

of strongly interacting pairs. The pair Bn27–Bs39 shows the highest coupling energies in experiments (Schreiber and Ferscht, 1995) while Bn59–Bs76 is located in the guanine binding-loop (residues 57–60 of Barstar), known to play an important role in the recognition process (Camacho et al., 1999; Spaar and Helms, 2005). Both are known to still interact at high salt concentrations despite their different location and exposure to solvent (Frisch et al., 2001). Finally, the mutation of both residues allows us to determine the additive content of these interactions as well as the examination of electrostatic interactions with no net charge remaining on Barnase. In this case, the monopole character and thus also electrostatic attraction of Barnase is reduced and higher momentums can contribute more to electrostatic interactions resulting in a complex associating at the minimum rate of $k_a \cong 10^5 \, M^{-1} \, s^{-1}$ (Janin, 1997).

4.3. Structural motifs during complexation

As a first step towards an identification of structural motifs during complexation and the influence of charged residues on association, we analyzed contact maps at different separation distances (data not shown). Here, no clear preference of a single distinct contact pair was observable at higher distances. This underlines a rather dynamic ensemble of structures. For lower distances, the contacts as found in the crystal structure are preserved, demonstrating a pre-orientation leading to the final complex.

For a better analysis of the contribution of individual amino acids in the complex, we identified residues being close to the binding partner without specifying the exact contacting amino acid in the partner protein at varying distances (Fig. 4). For Barnase, the identified residues being close to Barstar are mainly K39, R59 and Q104. K39 and R59 are on opposite sites of the binding pocket, located in two well-exposed loops. R59 is located in the RNA binding-loop. Q104 is located in a neighboring loop near the RNA binding-loop and thus also accounts for binding in this direction. These three state the most important initial early contact residues on Barnase and suggest two different regions for the initial contact located at the two loops adjacent to the binding interface. The location of residue K27 inside the binding pocket (mutation 1) precludes contact formation in the early complexation stadium and thus a significant contribution at this early stage. This behavior

shows that although both double mutations form tight contacts in the crystal structure and strongly affect the binding energy, their individual role in early complexation differs significantly.

The initial contacts are not equally well defined for Barstar. Two main regions are found: one involves Glu76 and Glu80, which are both located at the outer rim of the binding site, while a second, broad region is centered around Asp39, which is in the center of the binding helix of Barstar. For both complexation partners, a change in ion concentration does not affect the importance of the residues for the first contact.

Interestingly, exclusively charged and polar residues participate in establishing the first contact independent of the ion concentration. This suggests that the ions do not individually shield charges on the surface but instead have a global shielding effect on the complex. Additionally, charged residues are exposed to the solvent and thus are more likely involved in the first contact.

Mutations lead to a loss of interaction with the selected residue while having only a limited effect on neighboring residues. This underlines that the binding interface has a modular architecture forming well-defined, but separated, independent binding patches as described before (Reichmann et al., 2005). This finding further suggests that the overall binding modes are only slightly affected by individual mutations, indicating that at least for the Barnase–Barstar complex, information about the correct relative orientations of the proteins during complexation is redundant. The redundancy may be a safety measure minimizing the potential deadly effect of random point mutations.

Particularly in the short-range domain, single or multiple overrepresented relative orientations, constituting something resembling an association pathway, can ensure a proper pre-orientation for final complexation to overcome desolvation penalties and to steer the complex into the correct final configuration (Lee and Tidor, 2001; Sheinerman and Honig, 2002). While the encounter complex can be seen as a loosely bound structure, displaying high orientational flexibility, steric mismatch in the final complexation steps can lead to long living intermediates with incorrect alignment, which can be prevented by the existence of a strong bias towards certain, 'correct' relative orientations. To analyze this potential bias, we further analyzed the structures of complexation using a sophisticated clustering scheme. The obtained graphs with connected community clusters provide us a picture of structures along the complexation pathways. Preferred patterns reaching up to 0.8 nm separation could be observed. At physiological ion concentration, the structures split up into two branches of connected clusters, merging at a separation of ∼0.4 nm. Mutation and/or changes on the electrostatics of the complex by mutation can alter the occurrence of the splitting. A reconstruction of the wild-type pathway revealed that the structural difference is the tilt of Barstar towards the contact region on Barnase. Both motives have been previously recognized: The first motif we found has been suggested previously by Gabdoulline and Wade (Gabdoulline and Wade, 1998) as well as Camacho et al. (1999). The second path corresponds to the suggestion of a second energy minimum opposite to the RNA binding-loop location important to steer the complex into the encounter complex region (Spaar et al., 2006). Here, Barstar residue Y27 maintains close contact only with the Barnase pocket region opposite to the Guanidine binding-loop. Due to the heterogeneity of the structures at larger separation distances, a result of the weaker steering forces, the range of this method is limited. Rotations around the connection vector have a decreasing impact on the complex formation at higher separation distances (Selzer and Schreiber, 1999) and thus limit the range of observable patterns with this method.

For larger separations, we mapped the complex orientations on spherical surfaces providing a probability distribution in a certain distance range. In general, spots with a high probability together

with a high orientational restraint can be only observed for small distances. Upon increasing distance, the sampled area increases as a result of the weaker steering. The same effect can be obtained by mutation. The covered area from our simulations increases with separation and at an average distance of 2 nm sampling shows only slight preference for some spots (Fig. 6). At this separation, the interactions are weak at some orientations and nearly vanishing after only slight orientational changes. These regions of weak interaction are overrepresented. Surprisingly, electrostatic changes caused by the investigated mutations and in particular the change of ion concentration do not strongly affect the sampled area. However, mainly the occupancy of highly populated spots changes. Hence, the pre-orientation is influenced by electrostatics.

4.4. Sampling performance

Furthermore, these maps provide us with information on our sampling at a certain distance. The sampled region strongly increases with distance. Hence, the small regions sampled at low distance are caused by protein–protein interactions and not by limited simulation time. However, at large distances sampling may be insufficient. This is a significant difference to Brownian Dynamics based approaches (Fig. 6). While sufficient sampling at low distances is achievable with MD including the flexibility of the complex constituents and explicit solvent molecules, computational restrictions may compromise the sampling performance for increasing distances. Not only the phase space to sample increases with increasing distance additionally, more explicit water molecules $\propto r^3$ are required, in the best case leading to a linear increase of computation time with the number of atoms. At lower distances REMD combined with free energy calculation techniques such as umbrella sampling may improve sampling in some cases (Wolf et al., 2008), at the price of higher computational costs. We choose initial random rotations of up to 15°, increasing the sampling but still allowing successful encounter as previously suggested in theoretical studies of Janin and co-workers (1997). In BD simulations, a free energy contribution of roughly −13 kJ/mol has been suggested from infinity up to a separation of 2 nm, based on average distance of the contact surfaces (Spaar and Helms, 2005). Here, the loss of rotational entropy ΔG_{rot} of both complexation partners after diffusion is about 16.5 kJ/mol (Janin, 1997). These contributions are only partially included in our free energy profiles.

4.5. Free energy profiles

The absolute free energy values computed here can further be affected by the choice of a specific force field. In order to accurately model the protein stability and structure, we applied the OPLS/AA (Jorgensen and TiradoRives, 1988; Jorgensen et al., 1996) force field, which is known to perform very good for unbiased simulations involving protein interactions. However, Spaar et al. (2006) observed a significant improvement in the agreement of absolute free energy values with experimental data upon the usage of the AMBER 95 (Cornell et al., 1996) force field together with a rigid protein model in implicit solvent using BD simulations. Hence, force field inaccuracies may play a role in the absolute computed values. However, the general shape of the energy landscape is not expected to be strongly influenced by the choice of a specific force field.

Several free energy profiles have been suggested for transient protein complexes with fast associations such as Barnase–Barstar (Selzer and Schreiber, 1999; Spaar and Helms, 2005). In particular, two different scenarios have been discussed previously (Selzer and Schreiber, 2001). In these model landscapes, two barriers have been proposed: one before the encounter-complex formation and one before the final complexation. The two landscapes differ in the relative height of these barriers: in one landscape, the encounter-complex formation is the rate determining step, while in the second landscape, the final desolvation transition has the higher barrier. Our results stress the importance of the desolvation transition (Fig. 9). We further find that first significant interaction takes place at a separation distance of approx. 2 nm, which leads to a strong steering effect at closer distances. This is in agreement with earlier findings (Selzer and Schreiber, 2001). Our simulations indicate the potential existence of an initial small barrier with a barrier height of 2 kJ/mol at this separation distance of ∼2 nm. The size of the barrier is very small compared to our standard error in the MF (Fig. 8) and at a region where sampling may be sub-optimal. However, nearly all setups display similar barriers at this region, arguing that it may be no artifact of our simulation. The barrier has a height on the order of one $k_B T$ and therefore may have a noticeable effect on the overall rate. The barrier vanishes at lower ion concentrations for all systems but the double mutant. Notably, the wild-type at a physiological ion concentration does not show this desolvation barrier. This stresses the importance of electrostatics for the reduction of this first barrier.

In particular at lower separation distances, sampling with MD simulations allows an accurate picture of the energy landscape during association. Here, we find a pronounced barrier (see Section 3) for wt, mutant 1 and mutant 12 at a separation distance of ∼0.3 nm at physiological ion concentrations. This distance correlates with the thickness of the first hydration shell. Interestingly, for all of the three setups displaying this barrier, two different overrepresented relative orientations of the proteins at larger distances can be observed (see Fig. 4). The location of the barrier (around 0.3–0.4 nm COM separation) is close to the junction position observed in clustering. This indicates that at this distance a structural re-orientation might have to take place. Additionally, the distance suggests the involvement of bound water being displaced, known to play an important role in protein–protein interactions (Camacho et al., 2000; Bhattacharyya and Bagchi, 2000; Perutz, 1978). Hence, an interplay between desolvation and structural transitions appears to be responsible for this barrier. Single molecular force experiments, which measure the distance between ground state and rate determining barrier, are in good agreement with the barrier position computed here and often report values in the range of 0.2–0.5 nm(Cai and Yang, 2003; Kada et al., 2001; Yuan et al., 2000; Chen and Moy, 2000).

From our simulations, we could not reliably observe the suggested barriers (Selzer and Schreiber, 1999) of the encounter-complex formation. There are two possible explanations: (1) The proposed barriers (in the free energy) are small, at least in the Barnase–Barstar system. (2) The barrier is beyond our sampled range. The latter might play a role at long distances, where we face the previously discussed sampling limitations. On the other hand, for lower distances, we also could only observe a relatively small barrier on the order of 6 kJ/mol. This is supported by other computational results, where a significant local minimum along the reaction coordinate could not be observed (Spaar and Helms, 2005; Selzer and Schreiber, 2001). Furthermore, the very fast association rate of this complex argues against high barriers in the energy landscape. For the complex of the beta lactamase, TEM, and its natural inhibitor, BLIP, it has been suggested based on thorough biochemical and biophysical investigations that the encounter-complex formation is not hindered by a significant barrier, while the final complexation step is the rate determining process (Selzer and Schreiber, 2001). Although TEM–BLIP represents a different protein system, these findings are consistent with our results here.

We can directly compare our obtained values of the free energy of binding with other calculations and experimental values. The energies of wt (−46.9 vs. 79.8 kJ/mol), mutant 1 (−27.1 vs. 43.2 kJ/mol) and mutant 2 (−33.3 vs. 59.2 kJ/mol) in 150 mMol
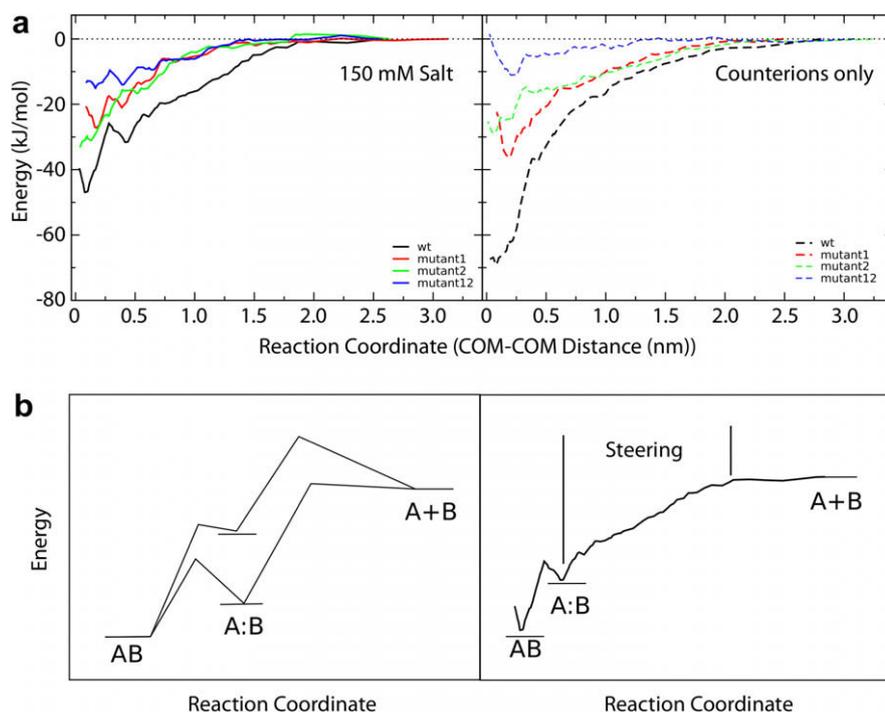
Fig. 9. Potential of mean force of Barnase–Barstar complexation. Potentials calculated by integration of the corresponding mean force profile (shown in Fig. 7). As a general trend, lack of ions increases the potential profile depth between wt and mutants. A local minimum around 0.4 nm is observable at physiological ion concentration for wt, mutant 1 and the double mutant 12. The lower part is a direct comparison between suggested profiles (Selzer and Schreiber, 2001) and our wild-type PMF. Our results stress the importance of a desolvation transition after initial complex formation.

NaCl correlate with a factor of 2 with experimental and computational values in 50 mMol Tris (Frisch et al., 1997; Wang et al., 2004). For our double mutant 12 we found that the absolute free energy difference in presence of physiological ion concentration (−15.1 kJ/mol) is even larger than in absence of ionic environment (−11.4 kJ/mol). This suggests that the presence of stronger electrostatic interactions here might even destabilizes the complex. However, the observed difference is close to the standard error of the mean in the minimum for mutant 12 in absence and presence of ionic environment (±1.9 kJ/mol and ±1.4 kJ/mol).

Another interesting observation from our simulation is that also the order of absolute free energy difference of the mutants 1 and 2 switches (mutant 1: −36.7 kJ/mol, mutant 2: −28.8 kJ/mol) depending on the ion concentration, while the overall effect on the absolute values is small. A significant impact was only observable for the wild-type complex (from −46.9 to −69.1 kJ/mol).

5. Conclusions

The details of the complexation process of Barnase and Barstar in particular and other systems in general are still under debate. Our simulations suggest a steep decrease in free energy starting from a distance of 2 nm with small barriers only for this complex with strong electrostatics. The final barrier is caused by both structural reorganizations and displacement of water. It affects the rates up to one order of magnitude. This suggests that the final complexation of this complex is not the overall rate-limiting step and that complexation for such strongly charged system is predominantly a downhill process. However, less-charged systems may well have differing energy landscapes.

Our findings suggest the existing of two main interaction regions each with a short ranging funneling region. The terminology path might be misleading, because orientational coherence is lost very quickly at larger distance. Structurally, two association modes

could be identified, both interacting with different interaction regions. The equilibrium between both patterns can be shifted by mutation in different ionic environment.

References

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

Baker, N.A., Sept, D., Joseph, S., Holst, M., McCammon, J.A., 2001. Electrostatics of nanosystems: application to microtubules and the ribosome. Proceedings of the National Academy of Sciences of the United States of America 98, 10037–10041.

Bergdorf, M., Peter, C., Hunenberger, P.H., 2003. Influence of cut-off truncation and artificial periodicity of electrostatic interactions in molecular simulations of solvated ions: a continuum electrostatics study. The Journal of Chemical Physics 119, 9129–9144.

Bhattacharyya, K., Bagchi, B., 2000. Slow dynamics of constrained water in complex geometries. Journal of Physical Chemistry A 104, 10603–10613.

Buckle, A., Schreiber, G., Fersht, A., 1994. Protein–protein recognition – crystal structural-analysis of a Barnase–Barstar complex at 2.0-angstrom resolution. Biochemistry 33, 8878–8889.

Cai, X., Yang, J., 2003. The binding potential between the cholera toxin B-oligomer and its receptor. Biochemistry 42, 4028–4034.

Camacho, C., Weng, Z., Vajda, S., DeLisi, C., 1999. Free energy landscapes of encounter complexes in protein–protein association. Biophysical Journal 76, 1166–1178.

Camacho, C., Kimura, S., DeLisi, C., Vajda, S., 2000. Kinetics of desolvation-mediated protein–protein binding. Biophysical Journal 78, 1094–1105.

Chen, A., Moy, V., 2000. Cross-linking of cell surface receptors enhances cooperativity of molecular adhesion. Biophysical Journal 78, 2814–2820.

Cornell, W., Cieplak, P., Bayly, C., Gould, I., Merz, K., Ferguson, D., Spellmeyer, D., Fox, T., Caldwell, J., Kollman, P., 1996. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules (vol. 117, p. 5179, 1995) (117). Journal of the American Chemical Society 118, 2309.

W. DeLano, 2009. The PyMOL User's Manual, Manual.

Despa, F., Fernandez, A., Berry, R.S., 2004. Dielectric modulation of biological water. Physical Review Letters 93, 228104.

Dong, F., Zhou, H.-X., 2006. Electrostatic contribution to the binding stability of protein–protein complexes. Proteins-Structure Function and Bioinformatics 65, 87–102.

Dong, F., Vijayakumar, M., Zhou, H., 2003. Comparison of calculation and experiment implicates significant electrostatic contributions to the binding stability of Barnase and Barstar. Biophysical Journal 85, 49–60.

Donnini, S., Mark, A., Juffer, A., Villa, A., 2005. Incorporating the effect of ionic strength in free energy calculations using explicit ions. Journal of Computational Chemistry 26, 115–122.

Frisch, C., Schreiber, G., Johnson, C., Fersht, A., 1997. Thermodynamics of the interaction of Barnase and Barstar: changes in free energy versus changes in enthalpy on mutation. Journal of Molecular Biology 267, 696–706.

Frisch, C., Fersht, A., Schreiber, G., 2001. Experimental assignment of the structure of the transition state for the association of Barnase and Barstar. Journal of Molecular Biology 308, 69–77.

Gabdoulline, R.R., Wade, R.C., 1998. Brownian dynamics simulation of protein–protein diffusional encounter. Methods 14, 329–341.

Gallagher, K., Sharp, K., 2003. A new angle on heat capacity changes in hydrophobic solvation. Journal of the American Chemical Society 125, 9853–9860.

Gottschalk, K.-E., Kessler, H., 2004. Evidence for hetero-association of transmembrane helices of integrins. FEBS Letters 557, 253–258.

Gottschalk, K.-E., Günther, R., Kessler, H., 2002. A three-state mechanism of integrin activation and signal transduction for integrin alpha(v)beta(3). ChemBioChem 3, 470–473.

Gottschalk, K.-E., Neuvirth, H., Schreiber, G., 2004. A novel method for scoring of docked protein complexes using predicted protein–protein binding sites. Protein Engineering, Design & Selection 17, 183–189.

Grater, F., Shen, J., Jiang, H., Grubmueller, H., 2005. Mechanically induced titin kinase activation studied by force probe molecular dynamics simulations. Biophysical Journal 88, 405A.

Hartley, R.W., 1989. Barnase and Barstar: two small proteins to fold and fit together. Trends in Biochemical Sciences 14, 450–454.

Hess, B., 2002. Determining the shear viscosity of model liquids from molecular dynamics simulations. Journal of Chemical Physics 116, 209–217.

Hess, B., Kutzner, C., van der Spoel, D., Lindahl, E., 2008. GROMACS 4: algorithms for highly efficient, load-balanced, and scalable molecular simulation. Journal of Chemical Theory and Computation 4, 435–447.

Humphrey, W., Dalke, A., Schulten, K., 1996. VMD – visual molecular dynamics. Journal of Molecular Graphics 14, 33–38.

Janin, J., 1997. The kinetics of protein–protein recognition. Proteins-Structure Function and Genetics 28, 153–161.

Jorgensen, W., TiradoRives, J., 1988. The OPLS potential functions for proteins – energy minimizations for crystals of cyclic-peptides and crambin. Journal of the American Chemical Society 110, 1657–1666.

Jorgensen, W., Maxwell, D., TiradoRives, J., 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. Journal of the American Chemical Society 118, 11225–11236.

Kada, G., Blayney, L., Jeyakumar, L., Kienberger, F., Pastushenko, V., Fleischer, S., Schindler, H., Lai, F., Hinterdorfer, P., 2001. Recognition force microscopy/spectroscopy of ion channels: applications to the skeletal muscle Ca^{2+} release channel (RYR1). Ultramicroscopy 86, 129–137.

Kutzner, C., van der Spoel, D., Fechner, M., Lindahl, E., Schmitt, U.W., Groot, B.L.D., Grubmueller, H., 2007. Software news and update – speeding up parallel GROMACS on high-latency networks. Journal of Computational Chemistry 28, 2075–2084.

Lee, L., Tidor, B., 2001. Optimization of binding electrostatics: charge complementarity in the Barnase–Barstar protein complex. Protein Science 10, 362–377.

Lindahl, E., Hess, B., van der Spoel, D., 2001. GROMACS 3.0: a package for molecular simulation and trajectory analysis. Journal of Molecular Modeling 7, 306–317.

Neumann, J., Gottschalk, K.E., 2009. The effect of different force applications on the protein–protein complex Barnase–Barstar. Biophysical Journal 97, 1687–1699.

Northrup, S.H., Erickson, H.P., 1992. Kinetics of protein–protein association explained by Brownian dynamics computer-simulation. Proceedings of the National Academy of Sciences of the United States of America 89, 3338–3342.

Novotny, J., Sharp, K., 1992. Electrostatic fields in antibodies and antibody antigen complex. Progress in Biophysics and Molecular Biology 58, 203–224.

Palla, G., Derényi, I., Farkas, I., Vicsek, T., 2005. Uncovering the overlapping community structure of complex networks in nature and society. Nature 435, 814–818.

Palla, G., Barabási, A.-L., Vicsek, T., 2007. Quantifying social group evolution. Nature 446, 664–667.

Perutz, M., 1978. Electrostatic effects in proteins. Science 201, 1187–1191.

Reichmann, D., Rahat, O., Albeck, S., Meged, R., Dym, O., Schreiber, G., 2005. The modular architecture of protein–protein binding interfaces. Proceedings of the National Academy of Sciences of the United States of America 102, 57–62.

Ryckaert, J., Ciccotti, G., Berendsen, H.J.C., 1977. Numerical integration of the cartesian equations of motion of a system with constraints; molecular dynamics of *n*-alkanes. Journal of Computational Physics 23, 327–341.

Sanner, M.F., Olson, A.J., Spehner, J.C., 1996. Reduced surface. An efficient way to compute molecular surfaces. Biopolymers 38, 305–320.

Schlitter, J., Swegat, W., Mulders, T., 2001. Distance-type reaction coordinates for modelling activated processes. Journal of Molecular Modeling 7, 171–177.

Schreiber, G., Fersht, A.R., 1995. Energetics of protein–protein interactions – analysis of the Barnase–Barstar interface by single mutations and double mutant cycles. Journal of Molecular Biology 248, 478–486.

Schreiber, G., Fersht, A.R., 1996. Rapid, electrostatically assisted association of proteins. Nature Structural Biology 3, 427–431.

Schreiber, H., Steinhauser, O., 1992. Cutoff size does strongly influence molecular-dynamics results on solvated polypeptides. Biochemistry 31, 5856–5860.

Schröder, C., Rudas, T., Boresch, S., Steinhauser, O., 2006. Simulation studies of the protein–water interface. I. Properties at the molecular resolution. The Journal of Chemical Physics 124, 234907.

Selzer, T., Schreiber, G., 1999. Predicting the rate enhancement of protein complex formation from the electrostatic energy of interaction. Journal of Molecular Biology 287, 409–419.

Selzer, T., Schreiber, G., 2001. New insights into the mechanism of protein–protein association. Proteins-Structure Function and Genetics 45, 190–198.

Sheinerman, F., Honig, B., 2002. On the role of electrostatic interactions in the design of protein–protein interfaces. Journal of Molecular Biology 318, 161–177.

Sheinerman, F.B., Norel, R., Honig, B., 2000. Electrostatic aspects of protein–protein interactions. Current Opinion in Structural Biology 10, 153–159.

SolerGonzalez, A., Fersht, A., 1997. Helix stability in Barstar peptides. European Journal of Biochemistry 249, 724–732.

Spaar, A., Helms, V., 2005. Free energy landscape of protein–protein encounter resulting from Brownian dynamics simulations of Barnase:Barstar. Journal of Chemical Theory and Computation 1, 723–736.

Spaar, A., Dammer, C., Gabdoulline, R., Wade, R., Helms, V., 2006. Diffusional encounter of Barnase and Barstar. Biophysical Journal 90, 1913–1924.

Strittmatter, G., Janssens, J., Opsomer, C., Batterman, J., 1995. Inhibition of fungal disease development in plants by engineering controlled cell-death. Bio-Technology 13, 1085–1089.

Swegat, W., Schlitter, J., Kruger, P., Wollmer, A., 2003. MD simulation of protein–ligand interaction: formation and dissociation of an insulin–phenol complex. Biophysical Journal 84, 1493–1506.

Trzesniak, D., Kunz, A.-P.E., Gunsteren, W.F.V., 2007. A comparison of methods to compute the potential of mean force. Chemphyschem: a European Journal of Chemical Physics and Physical Chemistry 8, 162–169.

van der Spoel, D., van Maaren, P., 2006. The origin of layer structure artifacts in simulations of liquid water. Journal of Chemical Theory and Computation 2, 1–11.

van der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A., Berendsen, H.J.C., 2005. GROMACS: fast, flexible, and free. Journal of Computational Chemistry 26, 1701–1718.

Wang, T., Tomic, S., Gabdoulline, R.R., Wade, R.C., 2004. How optimal are the binding energetics of Barnase and Barstar? Biophysical Journal 87, 1618–1630.

Wolf, M.G., Jongejan, J.A., Laman, J.D., de Leeuw, S.W., 2008. Rapid free energy calculation of peptide self-assembly by REMD umbrella sampling. Journal of Physical Chemistry B 112, 13493–13498.

Yang, C., Sharp, K.A., 2005. Hydrophobic tendency of polar group hydration as a major force in type I antifreeze protein recognition. Proteins 59, 266–274.

Yuan, C., Chen, A., Kolb, P., Moy, V., 2000. Energy landscape of streptavidin–biotin complexes measured by atomic force microscopy. Biochemistry 39, 10219–10223.

JCTC Journal of Chemical Theory and Computation

ProMetCS: An Atomistic Force Field for Modeling Protein−Metal Surface Interactions in a Continuum Aqueous Solvent

Daria B. Kokh,*[,†] Stefano Corni,[‡] Peter J. Winn,[§] Martin Hoefling,[||]
Kay E. Gottschalk,[||] and Rebecca C. Wade*[,†]

*Molecular and Cellular Modeling Group, Heidelberg Institute for Theoretical Studies
(HITS gGmbH), Schloss-Wolfsbrunnenweg 35, D-69118 Heidelberg, Germany,
INFM-CNR National Research Center on nanoStructures and BioSystems at Surface
(S3), Modena, Italy, Centre for Systems Biology, School of Biosciences, The University
of Birmingham, Edgbaston, Birmingham, B15 2TT, United Kingdom, and Ludwig
Maximilians University, Munich, German*

Abstract: In order to study protein−inorganic surface association processes, we have developed a physics-based energy model, the ProMetCS model, which describes protein−surface interactions at the atomistic level while treating the solvent as a continuum. Here, we present an approach to modeling the interaction of a protein with an atomically flat Au(111) surface in an aqueous solvent. Protein−gold interactions are modeled as the sum of van der Waals, weak chemisorption, and electrostatic interactions, as well as the change in free energy due to partial desolvation of the protein and the metal surface upon association. This desolvation energy includes the effects of water−protein, water−surface, and water−water interactions and has been parametrized using molecular dynamics (MD) simulations of water molecules and a test atom at a gold−water interface. The proposed procedure for computing the energy terms is mostly grid-based and is therefore efficient for application to long-time simulations of protein binding processes. The approach was tested for capped amino acid residues whose potentials of mean force for binding to a gold surface were computed and compared with those obtained previously in MD simulations with water treated explicitly. Calculations show good quantitative agreement with the results from MD simulations for all but one amino acid (Trp), as well as correspondence with available experimental data on the adhesion properties of amino acids.

1. Introduction

Protein−surface binding events are of great importance in many bioengineering, biomedical and nanotechnology applications. For example, protein adsorption properties are crucial for the integration of medical implants with tissue, and for the assembly of interfacial protein constructs in

nanotechnology, such as sensors, activators, and other functional components at the biological/electronic junction. Over the past decades, extensive experimental investigations on the molecular recognition, binding, and self-assembly of proteins, peptides, and amino acids on inorganic surfaces have been reported (for gold, see refs 1−7), and even combinatorially selected peptides with affinity for specific inorganic materials have been successfully synthesized.[8−10] For some examples of protein adsorption studies, particularly in connection with possible applications, see the reviews in refs 11−14 and references therein.

Because of the high complexity of protein adsorption phenomena and the scarcity of experimental data at the

* Corresponding author e-mail: daria.kokh@h-its.org (D.B.K.) and rebecca.wade@h-its.org (R.C.W.).

† Heidelberg Institute for Theoretical Studies.

‡ INFM-CNR National Research Center.

§ The University of Birmingham.

|| Ludwig Maximilians University.

atomistic level, however, the mechanisms by which biomolecules interact with inorganic surfaces are still poorly understood,[14] and until very recently, investigations of protein adsorption properties either had a rather qualitative character or were done on the macroscopic scale. This is why, in recent years, great efforts have been applied to adapt computational methods that are usually employed for molecular modeling in solution to the protein−surface problem. In particular, all-atom empirical force field methods, treating water molecules and the internal coordinates of an adsorbate explicitly, are now widely used to investigate biomolecule−surface binding behavior at the atomistic level[13−24] and have been shown to be able to provide qualitative agreement with experimentally observed adsorption tendencies for some small peptides.[19] However, all-atom molecular modeling methods, with explicit inclusion of water molecules, are extremely computationally demanding and therefore restricted to short time (typically of 10−100 ns) and length scales, while most experimental studies give an averaged behavior of large biomolecules over milliseconds or longer. Most of the atoms in molecular dynamics (MD) simulations with explicit water molecules come from the solvent itself. Furthermore, the presence of explicit water molecules slows protein motions. These two factors can make the computational time needed for the convergence of calculated properties extremely long. Therefore, a possible way to reduce computational time is to use an implicit solvent model that, in combination with an all-atom force field representation of a protein, may provide a reasonable compromise between accuracy and computational cost.

Existing implicit solvent models have primarily been developed for simulation of protein or peptide behavior in solution alone[25] and are generally not appropriate for protein interactions with inorganic interfaces.[13] This was demonstrated, in particular, by Sun and Latour[26] in their comparative analysis of commonly used empirical force-field-based implicit solvent models. It was found that the adsorption free energy of a peptide on a self-assembled monolayer (SAM) may change by up to several tens of kilocalories per mole, depending on which solvent model was used for the calculations. Furthermore, it has been recognized recently that the microscopic properties of the hydration shell vary for different solid surfaces, thereby altering the mechanism of adsorbate−surface interaction. For example, on metal surfaces, the desolvation energy may cause a transition barrier to adsorption due to the energetically unfavorable displacement of the water layer,[24] whereas, for some polar surfaces, peptides may be bound to the structured water layer rather than to the surface itself.[23] Hence, to provide a reliable description of protein−metal association in aqueous solvent, the solvent model should include a microscopic characterization of processes at the protein−surface interface.

In the present paper, we propose an approach for computation of the adsorption free energy of a biomolecule to a gold surface with an implicit solvent model that accounts for the short- and long-range effects of the protein−solvent−metal interactions. We employed the Au(111) surface for modeling because of its importance in the field of protein−surface interactions, both for fundamental studies (well-characterized, stable surface in both air and water) and for potential applications (e.g., contacts in nanobioelectronics and optical detection systems). Moreover, extensive theoretical investigations of small organic molecules adsorbed on gold,[1,7,20,27−29] as well as experimental data on protein and peptide adsorption,[3,5,7] are available and can be used for model optimization and validation. The present energy function is designed for use in Brownian Dynamics (BD) simulations of protein adsorption to surfaces but is not limited to this application.

BD methods in which solute molecules are treated as rigid bodies diffusing in a continuum solvent are commonly applied to simulate diffusion-influenced reactions and have been shown to be successful for computing protein−protein,[30] protein−small molecule,[31] and protein−membrane[32] association kinetics. Similarly, BD methods can be directly applied to a large group of proteins with high internal stability that can adsorb onto inorganic surfaces without appreciable changes in conformation or can form a transient complex before conformational changes occur. This method may also open the way for simulation of protein−protein interactions mediated by solid surfaces or protein self-assembly on inorganic substrates.

Despite the apparent similarity of the protein−protein and protein−solid surface association reactions, they have intrinsic differences in kinetics and in the driving forces for the binding processes. Indeed, the leading interaction in the case of protein−protein association to a diffusional encounter complex often arises from the long-range electrostatic forces, while the short-range effects can be described simply by prohibiting overlap of the exclusion volumes of the proteins. The influence of electrostatics on the interaction between a protein and an uncharged metal surface is much weaker since it arises solely from polarization effects. For a neutral solute molecule without a well-pronounced dipole moment, the image-charge potential must rapidly converge to zero as the distance from the metal increases due to the cancellation of contributions from opposite charges. On the other hand, at small distances from the surface, short-range interactions such as van der Waals forces and small metal−solute molecule charge transfers (that may also involve π electrons), along with the desolvation free energy, dominate over the electrostatic interaction. The construction of such an energy function is facilitated by the fact that the van der Waals interaction between organic molecules and a solid state surface, in particular gold, has recently been parametrized with a set of force field parameters[27] which can be directly implemented at an atomistic level in the energy function. A continuum solvent model able to provide a reliable description of the solvent−protein−solid interface, especially hydrophobic effects, needs to be developed and parametrized. This task is complicated by the fact that there is no well-established microscopic model of the protein−water−metal interface even though the behavior of aqueous solvent itself on the metal (in particular, gold) surfaces has been intensively studied both theoretically[33−40] and experimentally,[33,41−45] and some solution-driven effects in MD simulations of peptide adsorption on metal surfaces have been reported.[19,23,24]

ProMetCS: An Atomistic Force Field

J. Chem. Theory Comput., Vol. 6, No. 5, 2010 **1755**

In experimental studies of metal wetting properties, gold surfaces have been described both as hydrophilic (on the basis of contact angle measurements)[43] and rather hydrophobic (from the sublimation kinetics of ice layers).[45] On the other hand, experimental and theoretical evidence indicates that water−Au(111) interactions are weak relative to the hydrogen bonds between water molecules.[33,34,40] To minimize the intermolecular interaction energy in the interfacial region at an uncharged surface, water forms hydrogen-bonded clusters (in some studies described as having an ice-like structure[40,42,44]) in which the water dipoles in contact with the metal surface are oriented in the surface plane or slightly tilted with hydrogen oriented toward the bulk water (oxygen points to the surface).[33,42] MD simulations and *ab initio* calculations show that the water density has a maximum in the vicinity of the Au(111) surface (the first water layer) and some density fluctuations at larger distances caused by screening effects (the hydration shell).[34,37−40] This effect is responsible for some energy penalty due to displacement of the hydration shell upon adsorption, which makes binding to gold in water less favorable than in a vacuum as observed in MD simulations.[24] Furthermore, oriented water molecules on the metal surface, acting as dipoles, induce an electrostatic field, which may affect the behavior of charged and polar molecules.[39] Taking all these data into account, it is reasonable to include in the continuum solvent model both the hydrophobic and the electrostatic effects of the interaction with the hydration shell of the metal, and to employ MD simulations of water molecules to compute their hydrophobic and electrostatic contributions to the desolvation energy, in order to parametrize them.

Finally, to check the designed continuum solvent model, we need a well-characterized test system that not only allows us to verify the reliability of our representation of solvation effects and the derived energy parameters but also helps us to understand the contributions of different interaction mechanisms to the total protein−metal binding free energy. It is reasonable therefore to start with validation on small systems, whose interaction can be studied accurately either by experimental or by theoretical methods or both, and then make use of parameter transferability to apply the method to larger ones. The natural choice of such small systems is a set of amino acids whose binding to metal surfaces has been analyzed using MD simulations[28,29] and can also be related to available experimental studies.[1,5,7,9] For the sake of consistency, we used the same gold surface representation, amino acid structures, and force field parameters as used previously in the MD simulations.[28,29] We also employed the same image-charge model and water force field in the MD simulations with explicit water molecules performed to support the development of the continuum approach presented here. Thus, we note that the ProMetCS model developed is based on the force field employed in MD simulations (the GolP model[27]) and, therefore, inherits the limitations of the latter.

The paper is organized as follows. In the next section (Computational Methods), we describe the procedure used for calculating adsorption free energy. We show how the effects of solvent−metal−solute interactions can be ap-
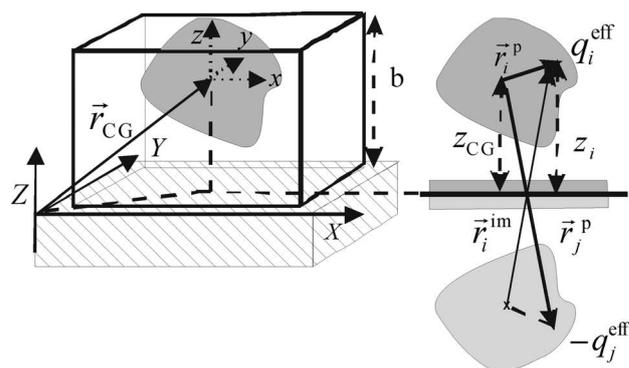
Figure 1. Illustration of the simulation box used for the calculations (left panel) and of the protein-image system employed for computing metal polarization effects (right panel). The low dielectric cavities of the protein and surface are shaded dark gray, and their images are light gray. The gold cluster is shown by a hatched block. Vectors are defined as $\vec{r}_j^{\,p} \equiv (x_i^p, y_i^p, -z_i^p - 2z_{CG})$ and $\vec{r}_i^{\,im} \equiv (x_i^p, y_i^p, z_i^p + 2z_{CG})$ from the geometric centers of the real and image protein, respectively. z_{CG} is the distance between the geometric center of the protein and the surface, and $z_i \equiv z_{CG} + z_i^p$ is the distance between a protein effective charge i and the surface.

proximated by physics-based energy terms, parametrized using explicit solvent simulations, and how they are implemented in the ProMetCS model. We give details of the MD simulations of the behavior of water molecules in the metal hydration shell, which we have used for the design and parametrization of the desolvation energy term. Finally, we show how the adsorption free energy and the potential of mean force (PMF) obtained from MD simulations can be calculated with the ProMetCS model. In the following section, we present the results of the application of the ProMetCS model to amino acid residues and compare the computed PMF binding energies with those from MD simulations as well as with available experimental data. In Appendix I, we give details of the MD simulations of the behavior of water molecules and test atoms in the surface hydration shell that were used for the design and parametrization of the desolvation energy. In Appendix II, we estimate the influence of the intrinsic electrostatic field of the hydration shell on the adsorption of charged molecules.

2. Computational Methods

2.1. Description of System (Setup). The Au(111) surface is described by a gold cluster with atomic layers. A minimum of three layers is necessary (and sufficient) for the accurate calculation of protein−gold van der Waals interactions, as will be shown below. During the calculations, the position of the cluster is fixed with the centers of the atoms in the surface layer at $z = 0$, i.e., in the xy plane of the simulation box, as illustrated in Figure 1. The surface area of the cluster must be larger than the size of the adsorbate in order to account for interatomic interaction effects up to the cutoff employed in calculations (see below). In the present study, a gold cluster with surface dimensions of 100 Å × 100 Å was employed. Since we used the force field parameters for the biomolecule−gold interaction derived in ref 27, the cluster was constructed accordingly (see details below). In

calculations of electrostatic and desolvation effects, the gold surface was considered to be a plane.

A distance b from the surface defines the limit of the simulation box where the protein−surface interaction energy is negligible and serves as a reference state for the calculation of the protein adsorption free energy and PMFs. The $z = b$ plane is also used for the generation of the starting positions of the adsorbate for the computation of BD trajectories.

Amino acid residues capped with an acetyl group at the N terminus and a methylamide group at the C terminus, corresponding to those studied in refs 28 and 29, were employed as test adsorbates. Calculations were performed for all 20 natural amino acids with their side chains assigned the standard protonation state at pH 7. Cysteine is known to form a strong bond with the gold surface which cannot be described by the Lennard-Jones-based force field parameters and a rigid gold surface. Therefore, we considered only the protonated form (denoted CysH), which cannot form a strong bond to gold, for the present simulations. To evaluate the effect of conformational variability of the capped amino acids upon binding to gold, we compared the binding properties of several of the most populated binding conformations obtained in MD simulations.[28]

2.2. Implicit Solvent Model: Interaction Energy Function. The protein−metal interaction energy function, U, which implicitly includes solvent effects, is expressed in the ProMetCS model as a sum of three separate contributions:

$$U = E_{LJ} + U_{EP} + U_{desolv} \qquad (1)$$

The E_{LJ} energy term describes nonpolar, van der Waals, and weak chemical interactions between a protein and a metal surface. It is parametrized to reproduce experimental binding properties of small organic molecules on gold.[27] It is a sum of classical Lennard-Jones 12−6 terms and will hereafter be denoted as a Lennard-Jones (LJ) term.

U_{EP} is the protein−metal electrostatic interaction free energy in aqueous solvent. (In the general case, it also includes the energy due to the electrostatic interaction between the charges in the protein binding site and the interfacial water potential on the metal surface, see Appendix II; this latter term is neglected in the implementation described in this work.)

The last term in eq 1, U_{desolv}, describes desolvation effects, i.e., the free energy change arising from protein−water, solid surface−water, and water−water interactions. Desolvation effects can be further split into two separate components: the desolvation energy of the protein, U_{desolv}^{p}, and the desolvation energy of the metal surface, U_{desolv}^{m}:

$$U_{desolv} = U_{desolv}^{p} + U_{desolv}^{m} \qquad (2)$$

The first term, the nonpolar (or hydrophobic) protein desolvation energy, is the free energy change of the protein−water system that arises from the replacement of the protein−water interface in the region of the adsorption site by a protein−vacuum interface. The second term in eq 2 represents effects arising from the partial replacement of the metal hydration shell by a protein adsorption site and is given by the free energy change due to insertion of a hydrophobic cavity (which mimics the binding site of the protein) into the hydration shell of the metal surface (note that the change of the protein−metal electrostatic interaction due to surface desolvation is instead included in the electrostatic energy term, U_{EP}).

It should be noted that the entropy contribution to the function represented by eq 1 [specifically, the second and third terms] is limited to the entropy change upon binding of the solvent only. The entropy change due to the restriction of protein motion upon binding the metal surface must be calculated separately. Hence, although U in eq 1 includes some entropic effects, it does not correspond to the complete adsorption free energy. The procedure for calculation of the entire adsorption free energy will be considered at the end of the present section. The three terms contributing to U in eq 1 are now described in more detail.

2.2.1. Lennard-Jones Term: E_{LJ}. van der Waals and weak chemical interactions between the biomolecule and the gold surface are described by the sum of 12−6 Lennard-Jones atom−atom pair potentials corresponding to interactions between each atom i of the biomolecule and each atom j of the gold cluster

$$E_{LJ} = \sum_{j} \sum_{i} 4\varepsilon_{ij}[(\sigma_{ij}/R_{ij})^{12} - (\sigma_{ij}/R_{ij})^{6}] \qquad (3)$$

where R_{ij} is the interatomic distance and

$$\varepsilon_{ij} = \sqrt{\varepsilon_{ii}\varepsilon_{jj}} \text{ and } \sigma_{ij} = \sqrt{\sigma_{ii}\sigma_{jj}}$$

are the (OPLS/AA-like) gold force field (GolP) parameters optimized by Iori et al.[27] for the interaction between organic molecules and a Au(111) surface.

The most important additions introduced in the GolP force field with respect to the standard OPLS/AA force-field[46] can be briefly summarized as follows: (i) The physical position of each Au atom in the upper layer of the gold cluster was replaced by two virtual atoms that occupy hollow sites. This particular representation of the structure of the surface layer has been proposed[27] to reproduce the correct binding position of the adsorbed molecules on the Au(111) surface. (ii) A new generic atom type for the Au atom was introduced, with generic ε_{AuAu} and σ_{AuAu} LJ parameters to be used for calculating E_{LJ} for gold−water and gold−protein atom pairs. (iii) Specific LJ parameters for the interaction between Au and the unprotonated N atom in His and the S atoms in CysH/Met were optimized to introduce N−Au and S−Au chemical bonding, respectively. (iv) The ε_{ij} value of carbon atoms in π rings was fitted to reproduce the rather strong interaction between the π electrons of aromatic molecules and the metal surface observed experimentally (if the π ring is oriented parallel to the surface plane). (v) A shell type model describes polarization effects of the gold surface,[47] although the latter feature is not used in the calculations presented here. Details on the derivation of the GolP parameters and a comparison between the adsorption energies calculated with GolP and experimental results for different molecules (typical deviations of less than 5−10% or a few kJ/mol), can be found in ref 27.

ProMetCS: An Atomistic Force Field

J. Chem. Theory Comput., Vol. 6, No. 5, 2010 **1757**

The direct pairwise calculation of the E_{LJ} energy between all atoms of the protein and of the metal cluster is too expensive for the large biological molecules usually studied by BD methods. Therefore, a grid-based procedure was implemented in which the LJ interaction energy between the protein and a gold atom is saved on the nodes of a three-dimensional grid with the origin placed at the protein center. The grid size is chosen so that the long-range limit of the protein atom—Au interactions with a cutoff of ~10 Å is inside the LJ grid. Then, the E_{LJ} interaction energy between the gold surface and a protein can be obtained by summation over all Au atoms of the gold cluster.

The balance between the repulsive and attractive parts of the LJ potentials arising from neighboring protein atoms is extremely important at small protein—surface distances. The binding energy is thus very sensitive to the grid spacing. Our test calculations of amino acid adsorption in a vacuum showed, for example, that a spacing of 0.2 Å may lead to an error in binding energy of up to about 3 kJ mol^{-1} compared to the binding energy of amino acids obtained directly by summation of the pairwise terms in eq 3. For comparison, a grid spacing of 0.5 Å results in an error in the binding energy of up to 12 kJ mol^{-1}. Therefore, a grid spacing of 0.2 Å has been used throughout the present study. Both the accuracy of the computed energy and the calculation speed depend on the number of gold layers employed. The optimal number of layers is three, since the energy correction due to adding a fourth layer is smaller than the uncertainty due to the grid discretization.

At short interaction distances (less than the sum of the atomic van der Waals radii in the OPLS force field), a constant positive energy of 100 kJ mol^{-1} was assigned to avoid very strong repulsion and therefore excessively high forces in BD simulations.

2.2.2. Protein—Metal Electrostatic Free Energy in an Aqueous Solution: U_{EP}. The interaction of a fixed set of partial point charges with a flat infinite uncharged metal surface is represented in classical electrostatics by the interaction between real charges q_i and their image charges, $q_i^{img} = -q_i$, placed symmetrically with respect to the metal surface plane. This approximation was shown to give good agreement with density functional calculations at a surface-charge distance of >2.5 Å.[48] Likewise, the electrostatic field of a fixed charge density in a nonuniform dielectric medium in the presence of the uncharged metal surface can be simulated by introducing an oppositely charged mirror image of the charge system instead of the metal surface. It is important that, to satisfy the boundary conditions (zero surface potential of the metal), electrostatic potentials of the protein and its opposite-charged image should exactly cancel each other at the surface plane. Therefore, not only the spatial distribution but also the dielectric surroundings of the real/image charges should be symmetrical with respect to the metal surface plane. Practically, in the implicit solvent model, a protein interacting with its image system consists of two charge distributions (one distribution for the real protein and one for its image), each immersed in low dielectric cavities surrounded by a high dielectric solvent and separated by a low dielectric cavity that surrounds the metal surface. The

latter cavity is introduced since the centers of the surface layer of metal atoms (defining the metal surface plane) are separated from the solvent by the LJ radius for the metal—water interaction.

The image potential is defined as $\Phi^{im}(\vec{r}_i^{im}) \equiv -\Phi(\vec{r}_j^p)$, where $\vec{r}_i^{im} \equiv (x_i^p, y_i^p, z_i^p + 2z_{CG})$ and $\vec{r}_j^p \equiv (x_i^p, y_i^p, -z_i^p - 2z_{CG})$ are vectors from the geometric centers of the image and real protein, respectively, and z_{CG} is the distance between the protein center and the metal surface as illustrated in Figure 1. Hence, we replace the protein—metal electrostatic interaction by a protein-image interaction with an additional low-dielectric cavity between the protein and the image as illustrated in Figure 1.

The electrostatic interaction free energy of two macromolecules (including solvent-related entropic effects only) can be calculated by numerical solution of the Poisson—Boltzmann equation. This however requires considerable computational resources and cannot be done at each time step of a BD simulation. Alternatively, the problem can be quite accurately solved by using the effective charge approximation for macromolecules (ECM) developed for protein—protein interactions.[49] Following this work, we describe the electrostatic interaction free energy between a protein and its image in the presence of the metal cavity as

$$U_{EP} = U_p/2 + U_{im}/2 + U_{p-c} + U_{im-c} \quad (4)$$

where U_p (U_{im}) corresponds to the energy of interaction of the protein (image) charges with the image (protein) electrostatic potential computed in the presence of both protein and image cavities as well as the metal cavity; U_{p-c} (U_{im-c}) describes perturbation of the protein (image) electrostatic potential by the low-dielectric cavity of the image (protein). The latter term decreases rapidly with the protein-image distance (i.e., with the distance from the metal surface) and will be referred to hereafter as the electrostatic desolvation energy.

The first and the second terms in eq 4 are equal, and so are the third and fourth terms. Thus,

$$U_{EP} = U_p + 2U_{p-c} \quad (4a)$$

For the real protein, the effective charges, q_i^{eff}, in a uniform high dielectric medium give the same electrostatic potential outside the protein surface as that computed for the real protein treated as a low dielectric cavity immersed in high dielectric solvent.[49] The electrostatic energy, U_p, can then be approximated by the interaction energy of the real protein effective charges q_i^{eff},[49] immersed in a uniform solvent medium, with the electrostatic potential of the protein image

$$U_p = \sum_i \Phi^{im}(\vec{r}_i^{im}) q_i^{eff} \quad (\text{where } \Phi^{im}(\vec{r}_i^{im}) \equiv -\Phi(\vec{r}_j^p)) \quad (5)$$

The electrostatic potential, $\Phi(\vec{r}_j^p)$, of a protein in water was calculated by numerically solving the linearized Poisson—Boltzmann equation using the UHBD (University of Houston Brownian Dynamics) program.[50] The relative dielectric constant of the protein was assigned as 4 and that of the solvent as 78, and the dielectric boundary was defined by the van der Waals radii of the protein atoms. The protein

atoms were assigned partial charges from the OPLS force field.[46] The electrostatic potential was computed on a three-dimensional grid centered on the geometric center of the protein. Since the electrostatic potential changes smoothly with $\vec{r}$, it does not require as accurate a representation as the LJ potential, and we have therefore used a grid with a spacing of 0.5 Å in the present calculations. The effective charges q_i^{eff} were positioned on selected atoms of the charged residues (the carboxylate oxygen atoms of Asp and Glu residues, as well as the amine nitrogen atoms of Lys, Arg, and protonated His residues), and their values were derived by fitting the protein electrostatic potential in a 3-Å-thick layer extending outward from the protein's accessible surface computed with a probe of radius 4 Å.[49,51]

The last term in eq 4a is the electrostatic desolvation term of the protein with effective charges q_i^{eff} due to presence of the image protein cavity and the metal surface cavity. This can be accounted for by the introduction of a positive energy term analogous to that proposed in ref 49 as

$$U_{\text{p-c}} = \sum_i (\Phi_{\text{ed}}^{\text{met}}(\vec{r}_i) + \Phi_{\text{ed}}^{\text{im}}(\vec{r}_i)) \times (q_i^{\text{eff}})^2 \qquad (6)$$

A general equation for the electrostatic desolvation potentials, $\Phi_{\text{ed}}(\vec{r})$ [$\Phi_{\text{ed}}^{\text{met}}(\vec{r})$ or $\Phi_{\text{ed}}^{\text{im}}(\vec{r})$], due to a set of spherical low dielectric cavities is given in the dipole approximation in ref 52 as

$$\Phi_{\text{ed}}(\vec{r}_i) = \alpha \frac{\varepsilon_s - \varepsilon_p}{\varepsilon_s(2\varepsilon_s + \varepsilon_p)} \sum_j (1 + k\vec{r}_{ij})^2 \exp(-2kr_{ij} \frac{a_j^3}{r_{ij}^4}) \qquad (7)$$

k is the Debye−Hückel parameter, ε_p is the protein dielectric constant, ε_s is the solvent dielectric constant, a_j is the van der Waals radius of the jth atom of the protein image (or atoms of the metal surface), and r_{ij} is the distance from the jth atom to the effective charge of the protein q_i^{eff}. The scaling factor α was estimated[52] for protein−protein association as $\alpha = 1.67$. Since $\Phi_{\text{ed}}^{\text{met}}(\vec{r}_i)$ is at least 2^4 times larger than $\Phi_{\text{ed}}^{\text{im}}(\vec{r}_i)$ (eq 7), we can omit the effect of the image cavity on the electrostatic field of the protein and keep only the metal cavity terms:

$$U_{\text{EP}}(\vec{r}) = \sum_i \Phi^{\text{im}}(\vec{r}_i^{\text{im}})q_i^{\text{eff}} + 2\sum_i \Phi_{\text{ed}}^{\text{met}}(\vec{r}_i) \times (q_i^{\text{eff}})^2 \quad (8)$$

The electrostatic energy given by eq 8 has been derived for the case of nonoverlapping cavities of the protein and the metal surface. To complete this model, we have to consider the case in which the adsorbed molecule penetrates the first hydration layer of the surface, which in the context of the implicit solvent model means that the low-dielectric cavities of the protein and the metal merge. In general, this has two effects: (i) The change in Born solvation energy should be taken into account; this, however, rapidly vanishes with increasing adsorbate size[53] and can be neglected in the case of molecular adsorption (the case of ions will be discussed at the end of the present section). (ii) The metal-charge interaction energy must be scaled appropriately for the transition from high to low dielectric surroundings.
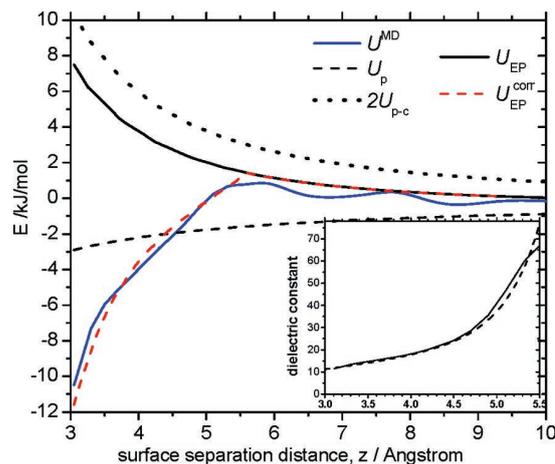
Figure 2. Total electrostatic energy for a test charge atom as a function of distance from the gold surface with explicit (U^{MD}) and implicit ($U_{\text{EP}}^{\text{corr}}$) water models ($U_p$, U_{EP}, and $U_{\text{p-c}}$ are separate contributions to the electrostatic energy, see text for details). Insert: Plot of effective dielectric constant derived from the image-charge potential computed from explicit water simulations (solid line) and approximated by an analytical function (dashed line).

Effect ii can be estimated from the electrostatic energy of an ion in the presence of the metal surface obtained in an MD model in which the solvent is treated explicitly. To this end, we have computed the image-charge energy, U^{MD}, of a test charge atom (with unit charge and $\sigma_{ii} = 2.87$ Å) as the difference in ion energy in explicit-water simulations with and without image-charge effects, see Figure 2. One can see in Figure 2 that, at surface separation distances smaller than $z \sim 5.5$ Å, the electrostatic ion-metal energy computed in the explicit water model, U^{MD}, is much lower than that obtained in the present implicit solvent approximation, U_{EP}. This z value can be considered as the approximate ion-surface distance at which the ion (or an effective charge in a molecule) and surface cavities start to merge. Indeed, this agrees with the Au−water and test charge−water LJ radii (∼3 Å and ∼2.5 Å, respectively).

The simplest way to account for this effect at small charge-surface distances in the ProMetCS model is to introduce a variable dielectric constant that increases as an effective charge moves away from the surface and reaches the value of $\varepsilon = \varepsilon_s$ when the cavities of the charge and the surface are separated and water molecules are able, at least partially, to screen the charge−metal interaction. Keeping the electrostatic desolvation energy $U_{\text{p-c}}$ unchanged, we fitted the dependence of the dielectric constant on the unit charge-surface separation distance, z, to reproduce the explicit-water electrostatic energy:

$$U_{\text{EP}}^{\text{corr}} \equiv \frac{-1}{2z(4\pi\varepsilon_0)\,\varepsilon(z)} + 2U_{\text{p-c}} \approx U^{\text{MD}} \qquad (9)$$

and therefore,

$$\varepsilon(z) = \frac{1}{2z(4\pi\varepsilon_0)(2U_{\text{p-c}} - U^{\text{MD}})}$$

ProMetCS: An Atomistic Force Field

J. Chem. Theory Comput., Vol. 6, No. 5, 2010 **1759**

The computed variable dielectric constant can be approximated by an analytical function $\varepsilon(z) = 4.0 + 0.8z^2 + \exp(z/0.385 - 10.4)$, where z is in Å, for $z < 5.5$ Å (see the insert in Figure 2).

In the case of a set of effective charges, the corrected electrostatic energy, U_{EP}^{corr}, can be directly applied to the diagonal terms, which correspond to the interaction of an effective charge i with its own image (for charge-surface distances of $z_i < 5.5$ Å):

$$U_{EP}^{corr} = U_{EP} + \sum_i \frac{(q_i^{eff})^2}{2z_i(4\pi\varepsilon_0)}\left(\frac{1}{\varepsilon_s} - \frac{1}{\varepsilon(z_i)}\right) \quad (10)$$

This approximation is valid for the case of amino acids that are described by one effective charge in the ECM model,[49] but for a system of many effective charges, cross-terms that correspond to the interaction between a charge and the image of another charge should also be taken into account.

The minimum value of the relative dielectric constant is ~10 (see Figure 2), which is consistent with typical values used in modeling of the electrochemical interface.[54] This value can lead to an up to 4-times larger Coulomb energy for a monatomic ion in pure water.

2.2.3. Protein Nonpolar (Hydrophobic) Desolvation Energy. The free energy change of a protein due to its partial desolvation by the gold surface can be described by a nonpolar desolvation energy that is proportional to the solvent accessible surface area (SASA) of a protein and an energy coefficient (Φ_{pd}):[55]

$$U_{desolv}^p = \sum_m \Phi_{pd}SASA_m \quad (11)$$

where the energy potential Φ_{pd} is computed on a three-dimensional grid and is defined as a function of the distance r from the van der Waals surface of a protein:[55]

$$\Phi_{pd}(r) = \beta c \begin{cases} 1 & \text{if} \quad r < a \\ \frac{b - r}{b - a} & \text{if} \quad a < r < b \\ 0 & \text{if} \quad r > b \end{cases}$$

The parameters a and b have been optimized[55] by using a standard method for SASA calculations (NACCESS) and are set to 3.1 Å and 4.35 Å, respectively; $c = 0.5$; the coefficient β was set to ~-0.021 kJ mol^{-1} Å^{-2} in the present calculations. It should be noted that the regions of nonzero desolvation energy and LJ binding energy strongly overlap, and this may lead to the relatively smaller hydrophobic desolvation term being dominated by the larger LJ attraction.

2.2.4. Metal Desolvation Energy for Nonpolar Adsorption Sites. To understand the nature of the solvation effects arising from the partial replacement of the metal hydration shell by a biomolecule, we considered the properties of the water in the vicinity of the Au(111) surface that can be derived from MD simulations. We first computed the partial water density as a function of surface water separation distance from a simulation of bulk water in the presence of an Au(111) surface, see Figure 3. The hydration shell consists of two water layers (at 3 Å and 6 Å) with a high partial
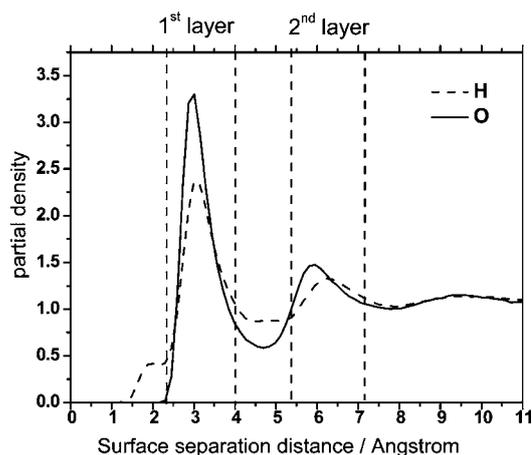
Figure 3. Dependence of the partial density of the water oxygen atoms (solid line) and hydrogen atoms (dashed line) on the distance from the gold surface computed from MD simulations of water in the presence of a gold surface. Densities are normalized to the bulk values; details of calculations are given in Appendix I.
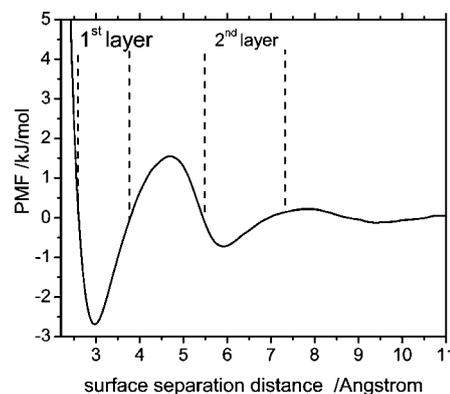
Figure 4. PMF of a water molecule as a function of the distance from the surface computed from MD simulations. Details of calculations are given in Appendix I.

density of water molecules. The comparison of the density of oxygen and hydrogen inside the first and second layers is higher than that of hydrogen, which indicates a nonuniform orientation of the water molecules, in agreement with other studies.[38,40] We then computed the PMF for one water molecule as a function of the surface water separation distance, see Figure 4. From the PMF, the computed binding free energy for a water molecule is ~-2.8 kJ mol^{-1} and ~-0.6 kJ mol^{-1} for the first and second hydration layers, respectively. The bound water in the first hydration layer is separated by a free energy barrier of ~-4.4 kJ mol^{-1} from the bulk water. The PMF shows that there will be an unfavorable positive energy change of the solvent-metal system when a water molecule is removed from the hydration shell to the bulk.

The metal desolvation energy is the free energy change caused by the replacement of the hydration shell of the metal surface by the protein adsorption site. It is, therefore, reasonable to assume that the desolvation energy is proportional to the desolvated area of the metal so that we can use an expression similar to eq 11:

$$U_{\text{desolv}}^m = \sum_i \Phi_{\text{metd}} S_i^{\text{desolv}} \tag{12}$$

where the coefficient Φ_{metd} is a free energy change for desolvation of a unit surface area of the metal. A proper modeling of the metal desolvation requires Φ_{metd} to depend on the distance of the protein surface atom i from the atoms of the metal surface. At large separations, when the distance between a protein atom i and the metal surface is greater than the LJ cutoff value, Z_{max}, Φ_{metd} must converge to zero. The summation in eq 12 must be carried out over the protein surface atoms, and S_i^{desolv} defines a desolvated area of the metal surface associated with the contacting protein atom i.

In the ProMetCS model, the desolvation energy describing replacement by a protein atom i of the first ($z_i < Z_{\text{adw}}$) and the second and higher ($z_i > Z_{\text{adw}}$) hydration layers is given by

$$\Phi_{\text{metd}} = \begin{cases} \Phi_{\text{metd}}^0 & z_i \leq Z_{\text{adw}} \\ \Phi_{\text{metd}}^0 \exp(-(z_i - Z_{\text{adw}})/\gamma)) & Z_{\text{adw}} < z_i < Z_{\text{max}} \\ 0 & z_i > Z_{\text{max}} \end{cases} \tag{13}$$

where z_i is the distance between the center of the protein surface atom i and the metal surface, Φ_{metd}^0 is the desolvation energy per unit area of the first hydration layer, γ describes the decrease in magnitude of the desolvation energy when the second and higher hydration layers are replaced, Z_{adw} corresponds to the position of the first hydration layer as defined above (~ 3 Å) plus the average LJ radius for the protein−metal atom interaction, which gives $Z_{\text{adw}} \sim 5$ Å, and $Z_{\text{max}} \sim 10$ Å is the cutoff for computing the desolvation term. Using the binding free energies per water molecule derived from the PMF in Figure 4, and assuming that the surface area occupied by one water molecule is ~ 9 Å^2, we estimate $\Phi_{\text{metd}} \equiv \Phi_{\text{metd}}^0 \sim 0.31$ kJ mol^{-1} Å^{-2} and $\Phi_{\text{metd}} \sim 0.07$ kJ mol^{-1} Å^{-2} for the first and the second hydration layers, respectively, which leads to an assignment of $\gamma \sim 1.51$ Å.

It should be noted that the desolvation energy U_{desolv}^m of eq 12 represents only the part of the free energy change of the metal hydration shell due to replacement of parts of the hydration shell by a noninteracting cavity. Electrostatic effects caused by the interaction of the charges at the adsorption site with oriented water dipoles on the metal surface (see Appendix II) are neglected here. Hence, the value of Φ_{metd}^0 estimated from the water adsorption energy is only a first approximation that may need further correction.

In order to calculate the desolvation area due to binding of a protein, we placed a two-dimensional grid on the surface plane, centered on the protein. Then, the positions of all protein atom−metal contacts with $z_i < Z_{\text{max}}$ were stored on the grid, and the area defined by the distance around the contact points R_{adw} (defined below) was considered as the desolvation area (illustrated in Figure 5). The total contact areas for atoms with $z_i < Z_{\text{adw}}$ and with $Z_{\text{adw}} < z_i < Z_{\text{max}}$ were calculated separately (they are shown in Figure 5 by the bold solid and dashed lines, respectively). These areas were then multiplied by the corresponding energy coefficients given by eq 13.
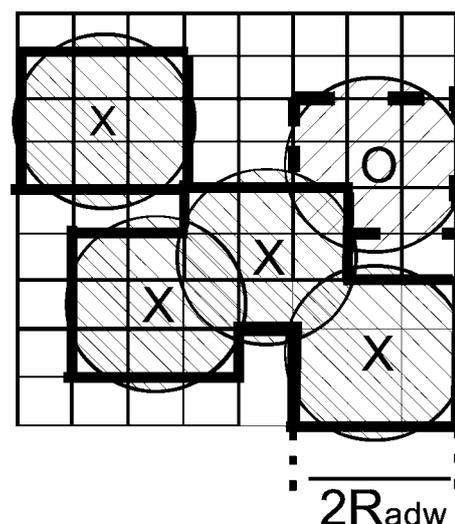
Figure 5. Illustration of the method employed for the calculation of the metal surface area in which water molecules are replaced by the adsorption site of the protein. The crosses and zeros show the positions of the centers of the protein atoms, with $z < Z_{\text{adw}}$ and $Z_{\text{adw}} < z < Z_{\text{max}}$, respectively. The hatched circles with radius R_{adw} show the area each atom is assumed to desolvate. The computed desolvation area is shown by bold lines, the solid and dashed lines corresponding to water desorption from the first and second hydration layers, respectively. See the text for details.
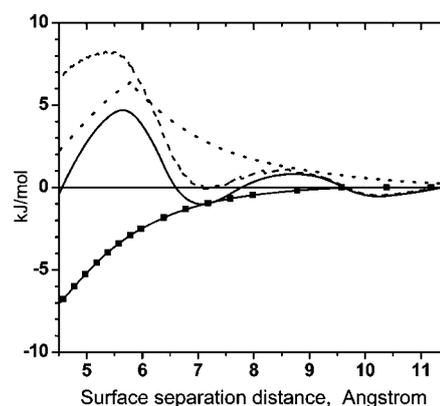
Figure 6. PMF obtained from MD simulations for the test atom (solid line), corresponding LJ potential (squares), and their difference (dashed lines) associated with the desolvation energy. Dotted line, PMF energy computed using the present model (includes both LJ and metal desolvation energies).

The value of R_{adw} was estimated by considering the desolvation energy, S^{desolv}, of a single test atom with "iodine-like" force-field parameters ($\sigma_{ii} = 5.4$ Å, $\varepsilon_{ii} = 0.293$ kJ mol^{-1}) that mimics a small nonpolar functional group of a protein. The PMF of the test atom obtained from MD simulations using a harmonic restraint potential applied along the z axis (the x and y coordinates were fixed during the simulations) is shown in Figure 6, along with the corresponding LJ potential. Since the translational entropy change along the PMF is zero for the present case, the difference between the PMF and LJ energies (dashed line in Figure 6) corresponds to the metal desolvation energy. It shows maxima at the first and second hydration layers at surface

ProMetCS: An Atomistic Force Field

J. Chem. Theory Comput., Vol. 6, No. 5, 2010 **1761**

separation distances of $\sim$5.5 Å and $\sim$8.5 Å, respectively. From the magnitude of the energy maxima, we estimate the desolvation energy change associated with partial replacement of the first hydration layer by an adsorbed atom as $U_{\text{desolv}}^m = 8.4$ kJ mol^{-1}. With $\Phi_{\text{md}}^0 = 0.31$ kJ mol^{-1} Å^2, $S_i^{\text{desolv}} \sim 27$ Å^2, which can be described by an effective desolvation radius, R_{adw}, of $\sim$3 Å. The total adsorption energy for the test atom computed as a function of the separation distance with the ProMetCS model with the parameters derived above is shown by the dotted line in Figure 6. The energy function given by eq 13 is by definition not able to reproduce the energy fluctuation at $z_i > Z_{\text{adw}}$, but in the case of a protein, the contribution of this effect is expected to be relatively small. The value of $R_{\text{adw}} = 3$ Å was used throughout all the calculations and appeared to be a good first approximation, as will be shown below.

2.3. Calculation of the Adsorption Free Energy and Potential of Mean Force. We consider an adsorbate as a rigid molecule moving relative to the solid surface. The geometry of the simulation box is the same as described above and shown in Figure 1. An adsorbate energy, defined by eq 1, is generally a six-dimensional function of the protein position and orientation. Three translational degrees of freedom define the position of the molecular center of geometry, x_{CG}, y_{CG}, and z_{CG}, where $\vec{r}_{\text{CG}} = (x_{\text{CG}}, y_{\text{CG}}, z_{\text{CG}})$, and the three rotational coordinates, $\Omega = (\Omega_1, \Omega_2, \Omega_3)$, are represented by Euler angles of the coordinate frame centered at the protein.

The solid surface is usually characterized by a periodic structure, and one can, therefore, expect a periodic variation of interaction energy as the protein position is shifted in the xy plane. Without any loss of generality, the molecule motion in the xy plane can, therefore, be considered in the area of $\Delta S = \Delta x_{\text{CG}} \Delta y_{\text{CG}}$, where Δx_{CG} and Δy_{CG} define a period of energy variation along the corresponding coordinate. The unit volume of configurational space of the protein−surface system is defined as $dS\, d\Omega\, dz$ ($dS = dx_{\text{CG}}\, dy_{\text{CG}}$, $d\Omega = \sin \Omega_1\, d\Omega_1\, d\Omega_2\, d\Omega_3$), and the total simulation volume of the configurational space is $8\pi^2 \Delta S b$.

The free energy change upon protein adsorption is then given by[56]

$$\Delta G = -k_B T \ln \left[\frac{Q_b}{Q_f} \right]$$
$$= -k_B T \ln \left[\frac{\int_b dz \int_{\Delta S,\Omega} d\Omega\, dS \exp(-U(\vec{r}_{\text{CG}}, \Omega, S)/k_B T)}{b 8\pi^2 \Delta S} \right] \tag{14}$$

where Q_b and Q_f denote configurational partition functions of an adsorbate in the bound and free states per unit volume, $S = (x,y)$, and

$$Q_b = \frac{\int_{\text{bound}} dz \int_{\Delta S,\Omega} d\Omega\, dS \exp(-U(\vec{r}_{\text{CG}}, \Omega, S)/k_B T)}{V_b} \tag{15}$$

The value $Q_f = 8\pi^2$ represents a uniform distribution of an unbound protein over configurational space, and $V_b = \Delta S b$ is the simulation volume.

Direct calculation of the complete 6-dimensional free energy landscape is difficult, and we have therefore used the system symmetry to reduce the dimensions of the energy matrix. First, due to the periodicity of the interaction potential along the x_{CG} and y_{CG} coordinates, only a small area, ΔS, must be explored. A period of the interaction potential is about the dimension of the Au metal cell. In fact, preliminary calculations showed that the greatest variations in potential occur within an area of 6×6 Å with a grid spacing of 0.5 Å; i.e., 13×13 grid nodes should be computed for the x_{GC} and y_{GC} coordinates. Variations in potential in the xy plane as well as with respect to rotation around z (Ω_3 angle) arise only from the short-range energy terms (i.e., LJ and desolvation energy terms) and can therefore be neglected if the smallest separation between protein surface atoms and the metal surface, z_{min}, is larger than the LJ cutoff, $Z_{\text{max}} \sim 10$ Å. Thus, at large distances, only the electrostatic component is important, and only three coordinates, Ω_1, Ω_2, and z_{GC}, must be explored. Moreover, since the electrostatic potential is quite smooth, the grid spacing over z_{GC} can be notably increased at $z_{\text{min}} > Z_{\text{max}}$. After some test calculations, we chose a grid spacing $d\Omega_1 = d\Omega_2 = 3°$, $d\Omega_3 = 6°$, and $dz_{\text{GC}} = 0.2$ Å at $z_{\text{min}} < Z_{\text{max}}$, and $d\Omega_1 = d\Omega_2 = 6°$, $d\Omega_3 = 12°$, $dz_{\text{CG}} = 2$ Å at $z_{\text{min}} > Z_{\text{max}}$.

Computation of the free energy change upon binding using standard molecular dynamics simulations is not feasible since it requires very extensive sampling to reach and cross high-energy regions of the underlying energy landscape. To overcome this problem, enhanced sampling techniques, such as those based on the umbrella-sampling concept,[57,58] can be used. To explore a reaction coordinate z, a series of simulations can be performed with a biasing harmonic restraint potential defined at each point of interest, z_{CG}^0, along the reaction coordinate: $V(z_{\text{CG}} - z_{\text{CG}}^0) = -1/2k(z_{\text{CG}} - z_{\text{CG}}^0)^2$.[58] The biased energy distribution function is given by $\exp(-(U(\vec{r}_{\text{CG}}, S, \Omega) + V(z_{\text{CG}} - z_{\text{CG}}^0))/kT)$, and, for a very sharp harmonic potential, the energy distribution function can be approximately described as

$$\exp(-(U(\vec{r}_{\text{CG}}, S, \Omega) + V(z_{\text{CG}} - z_{\text{CG}}^0))/kT) \approx$$
$$\delta(z_{\text{CG}} - z_{\text{CG}}^0) \exp(-U(\vec{r}_{\text{CG}}^0, S, \Omega))/kT)$$

where $\delta(z_{\text{CG}} - z_0)$ is the Dirac delta function and $\vec{r}_{\text{CG}}^0 = (x_{\text{CG}}, y_{\text{CG}}, z_{\text{CG}}^0)$. In this case, we are concerned with a local function at fixed z_{CG}^0 that describes the Boltzmann distribution over the adsorbate positions in the $x_{\text{CG}} y_{\text{CG}}$ plane and over the adsorbate orientation. Instead of an adsorption free energy given by eq 14, we have a PMF along the reaction coordinate, z_{GC}, with

$$G_{\text{PMF}}(z_{\text{CG}}^0) = -k_B T \ln \left[\frac{\int_{\Delta S \Omega} d\Omega\, dS \exp(-U(\vec{r}_{\text{CG}}^0, \Omega, S)/k_B T)}{8\pi^2 \Delta S} \right] \tag{16}$$

It is important to note that the $G_{\text{PMF}}(z_{\text{CG}})$ given by eq 16 includes the same kinds of entropy contributions as the MD simulations and can, therefore, be directly compared with the MD results.
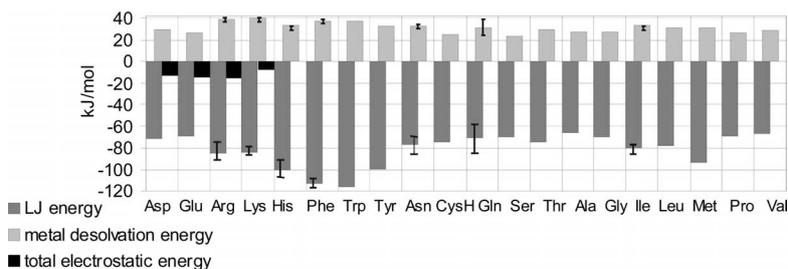
Figure 7. Contribution of the LJ, metal desolvation, and electrostatic (U_{EP}^{corr}) terms to the binding energy of capped amino acids on gold as calculated with the ProMetCS energy function. Error bars show energy deviation for different binding conformations used in simulations.

3. Results and Discussion: Testing of the Model for Adsorption of Capped Amino Acids on the Au Surface

To evaluate the accuracy of the energy model described, we computed the adsorption free energies and the PMFs of capped amino acids and compared the results with those obtained in MD simulations reported recently.[28,29] The main aim of this comparison was the testing of the proposed implicit solvent model against results with an explicit representation of water molecules. However, the solvent representation is obviously not the only difference between the ProMetCS energy function and that used in the MD simulations. To minimize the differences in physical characteristics of the amino acid−gold−water system employed in the two models, we used the same structure of the gold cluster and the same force-field parameters for the LJ energy as used by Hoefling et al.[28,29] Furthermore, in both models, an image-charge approximation was employed for calculation of the electrostatic effects. Finally, we used the most populated binding conformation of each capped amino acid obtained in MD simulations[28,29] in the present simulations. If several conformations with comparable populations were reported, we carried out simulations for all of them separately. However, we did not take into account the change in internal energy of the molecules upon conformational transition during the adsorption process, and this may cause some uncertainty in binding energy as will be discussed below in more detail.

Before presenting the results of the PMF simulations, let us consider the relative contributions of the energy terms of eq 1 to the binding energy of the amino acids. The largest contribution to the binding energy for almost all the amino acids arises from the LJ term, see Figure 7. As can be expected, the LJ energy increases with the number of atoms contacting the surface and, therefore, tends to increase with the size of the amino acid side chain. Due to the empirical design of the GolP force field parameters used in the present study, the binding to gold of Cys, Met, and His is favored if sulfur or nitrogen, respectively, comes close to the surface. Similarly, molecules with π rings (His, Phe, Trp, Tyr) are rather strongly bound to gold if their rings are parallel to the plane of the surface. Indeed, the absolute value of the LJ term shown in Figure 7 demonstrates the largest magnitudes for His, Met, Phe, Tyr, and Trp. In His, we found that binding through the π ring is stronger than attraction via an unprotonated nitrogen atom, and in our calculations, the

conformation with the ring parallel to the plane of the surface is more preferable than the tilted one.

The image-charge interaction is quite weak because of the charge-image distances (>6 Å) and the high dielectric constant aqueous medium between them. Indeed, the image-charge energy (first term in eq 8, U_p) is ∼−1.5 kJ mol^{-1} for all charged amino acids. Moreover, as an effective charge approaches the metal surface, induced solvent polarization around the low-dielectric cavities makes the electrostatic interaction effectively repulsive. This effect is described by the positive electrostatic desolvation penalty ($2U_{p-c} \sim +5−7$ kJ mol^{-1} at an ionic strength of 150 mM used in the MD simulations). In fact, the total electrostatic energy becomes negative only when an effective charge penetrates the hydration shell of the metal and its field is not screened by the water molecules any more. This effect, simulated by the variable dielectric constant, leads to an electrostatic energy of about −7 to −14 kJ mol^{-1} for charged residues, which is, however, still notably smaller than the $|E_{LJ}|$ binding energy of up to ca. 115 kJ mol^{-1}.

Whereas the electrostatic contribution to binding to a neutral gold surface is small for capped amino acids, as is the favorable hydrophobic protein desolvation energy ($|U_{desolv}^p|$ < 3 kJ mol^{-1}), the positive metal desolvation penalty varies from +20 kJ mol^{-1} to +40 kJ mol^{-1} and provides the largest compensation to the LJ term, see Figure 7. Since, U_{desolv}^m is proportional to the capped amino acid−metal contact area, the larger residues are in general characterized by a larger desolvation penalty as well as larger LJ binding energies. From Figure 7, one can also see that residues with long side chains, such as Arg and Lys, have a larger desolvation penalty than more compact residues. On the other hand, comparing all amino acids, the difference in the U_{desolv}^m value does not exceed 15 kJ mol^{-1}, while the LJ energy differs by up to ∼50 kJ mol^{-1}.

Finally, the binding energy is additionally compensated by the loss in translational and rotational entropy of the molecule upon binding to the surface. The entropy contribution has quite a small dependence on the amino acid type since all the capped amino acids have a well-defined binding position that corresponds to a rather sharp energy minimum. The entropic part due to restriction of rotation and of translation in the *xy* plane (which is included in the PMF) is about 25 kJ mol^{-1}, whereas the entropy difference due to translation along the *z* coordinate is ∼10 kJ mol^{-1}.

ProMetCS: An Atomistic Force Field

*J. Chem. Theory Comput., Vol. 6, No. 5, 2010* **1763**
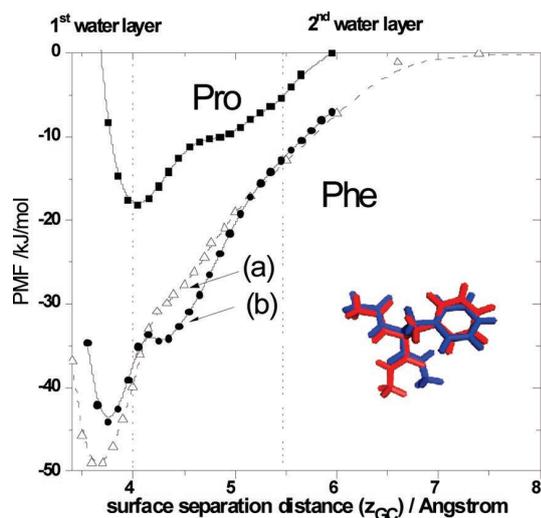
Figure 8. Representative PMF profiles computed using the ProMetCS model for two capped amino acids shown as a function of surface the separation distance: squares, Pro; triangles, circles, Phe. For Phe, PMFs computed for two slightly different conformations are shown: the aromatic ring is in the backbone plane (a, red conformation) and slightly tilted (b, blue conformation). Phe conformations are shown in projection onto the Au surface plane.

Representative PMF profiles for weakly (Pro) and strongly (Phe) bound residues are shown in Figure 8. The shape of the PMF, and the value of its global minimum depend on the conformation of the molecule used in the simulations. For example, the Phe conformation with the aromatic ring and backbone oriented in the same plane is more strongly bound and has only one minimum, whereas tilting the aromatic ring with respect to the backbone plane leads to two minima in the PMF (at ~3.7 Å and ~4.4 Å, see Figure 8), corresponding to the orientation of either the side chain or the backbone in the surface plane, respectively.

From the above analysis, we conclude that the relative binding strength of the amino acids in the present model is mainly driven by the LJ energy term. Since the LJ binding energy is very sensitive to the positions of the interacting atoms, conformational effects can be very important in the adsorption energy calculations. However, with the rigid-body approximation, the conformation is not adjusted at each simulation step, and therefore, changes in the internal energy of adsorbed molecules are not included in the present simulations. Therefore, to account for the existence of multiple conformations and minimize the uncertainty in the computed energies due to neglecting the change in internal molecular energy upon binding, calculations were carried out for some of the capped amino acids for several of the most populated binding conformations obtained in the MD simulations.[28] For most of the residues, there was only one dominant binding conformation in the MD simulations. For Asn, Arg, Cys, Gln, Glu, Leu, Lys, and Tyr, however, there were two bound conformations with similar populations, and for Ile, there were at least three conformations. The variation in binding energies obtained in the ProMetCS model for the different conformations is less than within ~5 kJ mol^{-1} for most of these amino acids, except for Gln, Ile, and Cys, whose
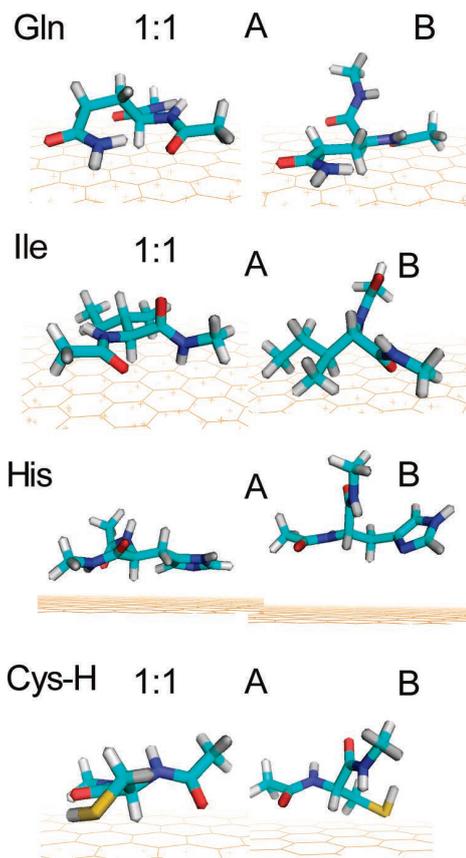
Figure 9. The most populated binding conformations of capped amino acids[28] for which significant dependence of computed PMF binding energy on conformation was observed, shown with their relative populations in the MD simulations. The corresponding PMF binding energies computed with the ProMetCS model are as follows: Gln, −40 (A) and −27 (B) kJ mol^{-1}; Ile, −24.3(A), −16.3 (B) kJ mol^{-1}; His, −48 (A), −31.9 (B) kJ mol^{-1}, where configurations A and B correspond to HIE and HID, respectively; CysH, −43 (A), −33(B) kJ mol^{-1}.

binding energy variation reaches ~10−15 kJ mol^{-1}. The most populated binding conformations for the latter residues are shown in Figure 9 along with their relative populations; the corresponding computed PMF binding energies are given in the figure caption. For all of these residues, the most strongly bound conformation has a nearly "flat" geometry (denoted as A in Figure 9) with the side chain as well as part of the backbone oriented parallel to the plane of the surface so that the LJ energy is optimized for the most atoms. In Figure 9, two binding conformations of His, "flat" and "tilted", which were observed in MD simulations[28] for two His forms corresponding to protonation of different nitrogen atoms in the aromatic ring (HIE, HID), are also shown. These have a binding energy difference of ~16 kJ mol^{-1} in the present calculations with the ProMetCS model, but almost equal binding free energies were computed from the MD simulations.[29] Here, the energy difference may be caused by underestimation of the desolvation penalty of the aromatic ring if it is placed in the surface plane since, as noted above, the U_{desolv}^{m} value is relatively small for aromatic residues.
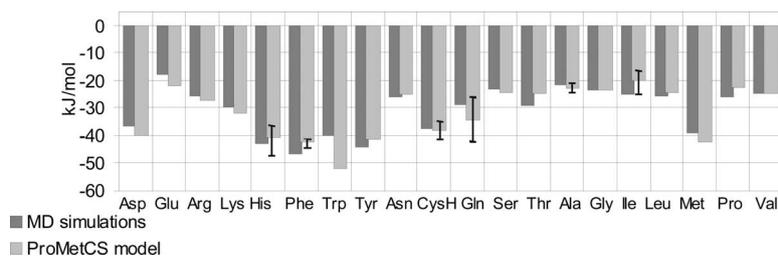
Figure 10. PMF binding free energies of the capped amino acids on the gold surface obtained using the ProMetCS model and MD calculations[29] as calculated with the ProMetCS energy function. Error bars show energy deviation for different binding conformations used in simulations.

Figure 10 shows a comparison of the PMF binding energies (the minimum of the PMF along the z coordinate) computed from the MD simulations[29] and with the ProMetCS model. For the residues mentioned above, characterized by several binding conformations in MD simulations that have notably different binding energies in ProMetCS, the PMF energies for the different conformations were averaged with equal weights. With the exception of Trp, the deviation of the PMF binding energies computed with the ProMetCS model from those from MD simulations does not exceed 5 kJ mol^{-1}. The binding energy of Trp is overestimated by $\sim$13 kJ mol^{-1}, which may be due to two reasons: (i) underestimation of the metal desolvation energy for aromatic residues, which was discussed above and might be especially pronounced in the case of Trp (the notable overestimation of the binding energy for conformation A of His is consistent with this suggestion, see Figure 9); (ii) a conformational energy change upon binding that is not taken into account in these calculations.

Taken into account the uncertainties in the present calculations, we can select capped amino acids with adsorption energies below -40 kJ mol^{-1} and above -25 kJ mol^{-1} and assign them to groups of strongly and weakly bound amino acids, respectively. Thus, His, Met, Phe, Trp, and Tyr belong to the group of strong binders, whereas Ala, Glu, Gly, Ile, Leu, Pro, Ser, and Val can be described as weak binders. The list of "strong binders" agrees with the experimental study of Peelle et al.,[5] in which notable binding was observed only for homopeptides of Cys, His, Met, and Trp. Furthermore, high affinity to the gold lattice has also been suggested for Trp and Tyr by Hnilova et al.[9] However, Phe was not mentioned among the amino acids with high affinity to gold reported in these experimental studies.[5,9]

In general, some overestimation of computed binding energy relative to that observed experimentally might be expected because, in particular, there should be some conformational restrictions on the amino acids in the peptides studied in experiments (i.e., the optimal binding conformation of an amino acid on the surface may be greatly unfavorable in the peptide). Furthermore, the binding to gold of the capping residues used in the calculations may additionally contribute to the computed affinity, leading to some overestimation of binding strength, especially for weakly bound residues. For example, the adsorption free energy of L-phenylalanine derived from electrochemical measurements[7] was characterized as typical for weak chemisorption of small aromatic molecules (from -18 to -37 kJ mol^{-1}, where the larger value is associated with electrostatic binding of the carboxylic group to a positively charged electrode). Considering that, in the present simulations, about $20-25$ kJ mol^{-1} of the adsorption energy of the capped Phe molecule come from the capping residues (see adsorption geometry of Phe shown in Figure 8), the binding energy of the Phe residue can be estimated as $\sim-25-30$ kJ mol^{-1}, which is in the range of experimental values.

4. Summary and Future Directions

In the present paper, we propose an approximation for the calculation of the binding free energy of biomolecules on an atomically flat uncharged Au(111) surface in a continuum aqueous solvent. The interfacial interaction energy is based on an atomistic representation of short-range interactions (van der Waals, weak charge transfer, π orbital interactions) that are approximated by a set of Lennard-Jones potentials, electrostatic interactions described by the image charge method combining with the effective charge approximation, and adsorbate desolvation and metal desolvation free energies. The latter term simulates the solvation free energy change due to the replacement of part of the gold hydration shell by the uncharged binding region of an adsorbate and has been parametrized by using the results of MD simulations of water molecules on gold. MD simulations were also the basis for parametrizing a model of the desolvation effects based on the electrostatic energy. The case when the adsorption site of the biomolecule is charged and interacts with the induced electrostatic field of the oriented water dipoles on the gold surface was also considered. When parametrized using the PMFs for surface binding of negatively and positively charged ions obtained from MD simulations, this effect was found to generally lead to slightly stronger binding of positively charged adsorbates than negatively charged ones (see Appendix II).

The proposed energy model, ProMetCS, has been verified against the recently reported PMFs of capped amino acids obtained from MD simulations.[29] We computed the binding energy of 1$-$3 of the most populated binding conformations observed in the MD simulations for each amino acid.[28,29] When averaged over these conformations, the computed PMF minimum values (i.e., PMF binding energies) reproduce the results of MD simulations with an error of less than 5 kJ mol^{-1} for all residues except Trp. The trends in amino acid

ProMetCS: An Atomistic Force Field

J. Chem. Theory Comput., Vol. 6, No. 5, 2010 **1765**

binding to gold are mostly in agreement with available experimental observations of the binding of homopeptides to gold despite the different conditions of the experiments.[5,7,9]

Analysis of the computed binding energies, in particular in comparison with experimental data, gives strong evidence that short-range van der Waals interactions (described by LJ potentials) are a driving force for adsorption of amino acids to a neutral metal surface. The change of the solvation free energy upon adsorption species is positive due to unfavorable distortion of the structure of the water layer on the gold surface. The balance of the short-range LJ attraction and the surface desolvation penalty makes the adsorption energy very sensitive to conformational variations of the adsorbed species and the orientation of the molecule on the gold surface. Therefore, it is advisable to explore a range of adsorbate conformations that are energetically accessible in aqueous solution.

As can be expected, the image-charge electrostatic effects on amino acid−gold interactions are quite small in comparison with the LJ term and the metal desolvation penalty, except in cases where a charged residue penetrates the hydration shell of the metal surface. On the other hand, in adsorption kinetics of large molecules, the electrostatic effects may gain more importance due to their long-range character.

The next step in validation of the ProMetCS model will be to apply it to a set of proteins and compare it with experimental adsorption data on the relative binding properties. Furthermore, due to the time-saving grid-based technique employed in the present model, it can be extended to the simulation of coadsorption and adsorption kinetics.

Appendix I. MD Simulations with Explicit Water Molecules

Three MD simulations with explicit water have been performed for this work:

(i) MD simulation of the water−Au(111) interface.

(ii) Calculation of the PMF of "fluorine-like" ions, i.e., an ion/atom with the LJ parameters assigned in the OPLS/AA force field to F^-, but with a +1, 0, and −1 e charge.

(iii) Calculation of the PMF of an "iodine-like" neutral atom, i.e., a neutral atom with the OPLS/AA[46] LJ parameters corresponding to those for I^-.

All these calculations were performed with the GROMACS (version 3.3.3 and 4.0.1) software.[59] The simple point charge, SPC, model was used for water (with rigid internal geometry constrained by the RATTLE algorithm), while the force field used for the water−gold, ion−gold, and atom−gold interactions is GolP.[27] The Au surface was simulated by a 5-layer gold slab, using a $7 \times 4\sqrt{3}$ supercell. 3D periodic boundary conditions were used. A second Au slab was placed at ∼3.5 nm from the first in the direction perpendicular to the surface (z), to confine water in a fraction of the periodic

box along z (10 nm). In this way, possible spurious effects due to fictitious periodicity along z were minimized. The interslab space is large enough to have a ∼1-nm-thick region of water behaving like bulk SPC water in the middle of the slab (as verified by density profile and oxygen−oxygen correlation functions). The precise value of the interslab space was adjusted for each simulation to yield the bulk density of SPC water at 1 bar of pressure and 300 K at the center of the water layer.

The PME electrostatic model was employed in all the simulations, using GROMACS defaults for the PME parameters. For neutral systems, we performed some tests employing the Yeh and Berkowitz[60] electrostatic corrections proposed for 2D periodic systems treated with 3D periodicity. No significant variation in the reported result was found. For charged systems, no counterions were inserted to neutralize the simulation box to avoid sampling issues.[61] All the simulations were performed in the NVT ensemble, with T = 300 K. LJ interactions were cut off at 1.0 nm.

For simulation i, an initial equilibration of 100 ps was followed by a 5 ns production run. The density profiles in Figure 3 were obtained from the resulting trajectory. The PMF of water in Figure 4 was obtained as

$$\text{PMF}(z) = -RT \ln\left(\frac{d(z)}{d_{\text{bulk}}}\right)$$

where $d(z)$ is the water density in the slab centered at z and d_{bulk} is the bulk density.

PMF calculations ii and iii were performed by integration of the average force along the ion-surface separation coordinate,[62] using either umbrella-biased simulations, also called umbrella integration[63] (calculation ii), or a constraint-biased simulation, with a LINCS constraint.[64] In both cases, an initial simulation was performed in which the ion/atom was pulled through the box, along the direction perpendicular to the surface, in 1 ns. From this simulation, 30 snapshots corresponding to 30 different ion-surface distances ranging from 0.25 to 1.2 nm were extracted. From each of these snapshots, a 5-ns-long simulation was started, keeping the ion/atom at the initial distance by using a tight harmonic restraint for ii or a constraint for iii. The thermodynamic restraint/constraint forces were collected during the last 3 ns of the dynamics and averaged to get the opposite of the PMF derivative with respect to the ion-surface distance. By numerically integrating such PMF derivatives, the PMF profiles in Figures 6 and 11 were obtained.

The ability of the GolP force field to reproduce experimental adsorption energies on gold for small molecules has been verified in ref 27, and the soundness of the calculated adsorption free energies of amino acids in solution is shown in ref 29. As a further test of the force field underlying ProMetCS, we calculated the central quantity to characterize the liquid water−gold interaction, i.e., the wetting coefficient k as defined by the relation $k = (\gamma_{\text{sv}} - \gamma_{\text{sl}})/\gamma_{\text{lv}}$ where γ_{sv}, γ_{sl}, and γ_{lv} are the solid−vapor, solid−liquid, and liquid−vapor interface tensions, respectively.[65] The difference $\gamma_{\text{sv}} - \gamma_{\text{sl}}$ was calculated from the 5 ns simulation described above in two ways: by the virial-based expression[65,66] as implemented in GROMACS 4.0.1 and by the energy-based method proposed in ref 20, using the entropic term correction
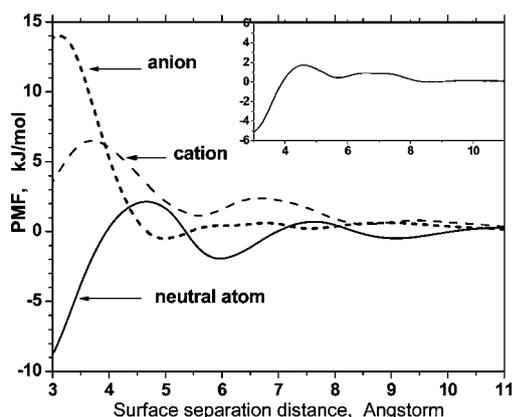
Figure 11. PMF of a test neutral atom (solid curve) and test positively and negatively charged ions (dashed and dotted curves, respectively) as a function of the surface separation distance. Insert depicts half of the difference between the PMF of a cation and of an anion.

proposed there. For the latter calculation, separate simulations of the gold slabs and the water slab were needed, and we performed simulations of 5 ns for each. In both cases, k was then calculated by using the value for γ_{lv} obtained by the virial method on the 5 ns water slab simulation. The virial-based expression yielded $k = 0.95 \pm 0.1$, while the energy-based expression yielded $k = 1.35 \pm 0.03$. For polycrystalline gold, it is known that $k \geq 1$ (i.e., the water contact angle = 0°), and for the Au(111) surface, a value close to or higher than 1 is also expected.[43] Therefore, the calculated k values are compatible with the experimental results. This discussion should not be considered as a detailed study of the water−gold surface tension obtained with GolP, which would require tests with respect to the cell size, the duration of the simulation, and the effects of the LJ cutoff. While such a detailed treatment is outside the scope of this article, it remains that the approximate k value computed here supports the use of GolP results in ProMetCS.

Appendix II. Electrostatic Interaction of an Ion with an Interfacial Water Potential

The electrostatic energy terms described above are derived for a uniform CS and do not take into account the effect related to the ordered water layer that directly contacts the metal surface. To estimate the contribution of this effect to the surface-binding energy of an ion, we first considered adsorption of test ions onto the metal surface. Computed PMFs for the positively and negatively charged test fluorine-like ion s, as well as a corresponding neutral atom, are shown in Figure 11.

The PMF function of an ion can be decomposed into four separate energy terms: (i) the LJ and the image-charge electrostatic interaction energies between the ion and the metal surface; (ii) the positive Born solvation energy given by $q/(8\pi\varepsilon_0 a)(1/\varepsilon_S - 1/\varepsilon_I)^{53}$ in the case of a charged atom of radius a that is transferred from a high dielectric ε_s to a low dielectric medium with dielectric constant ε_I; (iii) the free energy change of the solvent arising from distortion of the hydration shell of the metal (discussed above); and (iv) the interaction energy of the ion with the electrostatic field of the interfacial water.

Only the last term depends on the sign of the ion's charge and, therefore, can be computed as half of the difference between the PMF of the positively and negatively charged ions plotted in Figure 11. As expected, the resultant function (see insertion in Figure 11) shows fluctuations that roughly correlate with the variation of the oxygen partial density, i.e., with the negative partial charge variation within the hydration shell of the metal surface. One can also see from this plot, that term iv is relatively small (less than ~ 5 kJ mol^{-1}) and attractive for positively charged ions that are localized at an ion-surface distance of 3−4 Å (i.e., when the ion is inserted into the first hydration layer). On the other hand, the electrostatic field of the surface water layer is preferentially attractive for an anion when it is placed slightly beyond the first hydration layer, at 4−5 Å, which corresponds to the maximum of the hydrogen partial density.

The Born solvation energy is dominant at small distances ($z < 4$ Å) for ions but should be negligible for molecules. A charged fragment of a protein adsorption site would be surrounded by neighboring neutral atoms of the protein. This would mean that the electrostatic effect caused by the hydration shell would be less pronounced than for a bare ion because the charge-water distance would be too large to make the magnitude of the effect significant. Thus, the desolvation effect may be represented solely by the metal desolvation term iii due to the distortion of the hydration shell of the metal. Taking into account all these uncertainties and a modest contribution to the free binding energy, we did not implement the term accounting for this effect in the free energy calculations.

References

(1) Slojkowska, R.; Palys, B.; Jurkiewicz-Herbich, M. *Electro-chem. Acta* **2004**, *49*, 4109–4118. Prado, C.; Prieto, F.; Rueda, M.; Feliu, J.; Aldaz, A. *Electrochim. Acta* **2007**, *52*, 3168–3180.

(2) Brown, S. *Nat. Biotechnol.* **1997**, *15*, 269–272.

(3) Willett, R. L.; Baldwin, K. W.; West, K. W.; Pfeiffer, L. N. *Proc. Nat. Acad. Sci. U.S.A.* **2005**, *102*, 7817–7822.

(4) Wei, Y.; Latour, R. A. *Langmuir* **2008**, *24*, 6721–6729.

(5) Peelle, B. R.; Krauland, E. M.; Wittrup, K. D.; Belcher, A. M. *Langmuir* **2005**, *21*, 6929–6933.

(6) Krauland, E. M.; Peelle, B. R.; Wittrup, K. D.; Belcher, A. M. *Biotechnol. Bioeng.* **2007**, *97*, 1009–1020.

(7) Li, H.-Q.; Chen, A.; Roscoe, Sh. G.; Lipkowski, J. *J. Electroanal. Chem.* **2001**, *500*, 299–310.

(8) Tamerler, C.; Duman, M.; Oren, E. E.; Gungormus, M.; Xiong, X.; Kacar, T.; Parviz, B. A.; Sarikaya, M. *Small* **2006**, *11*, 1372–1378.

(9) Hnilova, M.; Oren, E. E.; Seker, U. O.; Wilson, B. R.; Collono, S.; Evans, J. S.; Tamerler, C.; Sarikaya, M. *Langmuir* **2008**, *24*, 12440–12445.

(10) Tamerler, C.; Oren, E. E.; Duman, M.; Venkatasubramanian, E.; Sarikaya, M. *Langmuir* **2006**, *22*, 7712–7718.

(11) Gray, J. J. *Curr. Opin. Struct. Biol.* **2004**, *14*, 110–115.

(12) Harding, J. H.; Duffy, D. M.; Sushko, M. L.; Rodger, P. M.; Quigley, D.; Elliot, J. A. *Chem. Rev.* **2008**, *108*, 4823–4854.

ProMetCS: An Atomistic Force Field

J. Chem. Theory Comput., Vol. 6, No. 5, 2010 **1767**

(13) Latour, R. A. *Biointerphases* **2008**, *3*, FC2–FC12.

(14) Cohavi, O.; Corni, S.; De Rienzo, F.; Di Felice, R.; Gottschalk, K. E.; Höfling, M.; Kokh, D.; Molinari, E.; Schreiber, G.; Vaskevich, A.; Wade, R. C. *J. Mol. Recog.* **2010**, *23*, 259–262.

(15) Nakanishi, K.; Sakiyama, T.; Imamura, K. *Biosci. Bioing.* **2001**, *91*, 233–244.

(16) Horinek, D.; Serr, A.; Geisler, M.; Pirzer, T.; Slotta, U.; Lud, S. Q.; Garrido, J. A.; Scheibel, T.; Hugel, T.; Netz, R. R. *Proc. Nat. Acad. Sci. U.S.A.* **2008**, *105*, 2842–2847.

(17) Raut, V. P.; Agashe, M. A.; Stuart, S. J.; Latour, R. A. *Langmuir* **2005**, *21*, 1629–1639.

(18) Raffaini, G.; Ganazzoli, F. *Langmuir* **2004**, *20*, 3371–3378. Makrodimitris, K.,; Masica, D. L.; Kim, E. T.; Gray, J. J. *J. Am. Chem. Soc.* **2007**, *129*, 13713–13722.

(19) Heinz, H.; Farmer, B. L.; Pandey, R. B.; Slocik, J. M.; Patnaik, S. S.; Pachter, R.; Naik, R. R. *J. Am. Chem. Soc.* **2009**, *131*, 9704–9714.

(20) Heinz, H.; Vaia, R. A.; Farmer, B. L.; Naik, R. R. *J. Phys. Chem. C* **2008**, *112*, 17281–17290.

(21) Braun, R.; Sarikaya, M.; Schulten, K. J. *Biomater. Sci. Polymer Ed.* **2002**, *13*, 747–757.

(22) Kantarei, N.; Tamerler, C.; Sarikaya, M.; Halilogla, T.; Doruker, P. *Polymer* **2005**, *46*, 4307–4313.

(23) Ghiringhelli, L. M.; Hess, B.; van der Vegt, N. F. A.; Delle Site, L. *J. Am. Chem. Soc.* **2008**, *130*, 13460–13464.

(24) Verde, A. V.; Acres, J. M.; Maranas, J. K. *Biomacromolecules* **2009**, *10*, 2118–2128.

(25) Feig, M.; Brooks, C. L., III. *Curr. Opin. Struct. Biol.* **2004**, *14*, 217–224.

(26) Sun, Y.; Latour, R. *J. Comput. Chem.* **2006**, *27*, 1908–1922.

(27) Iori, F.; Di Felice, R.; Molinari, E.; Corni, S. *J. Comput. Chem.* **2009**, *30*, 1465–1476.

(28) Hoefling, M.; Iori F.; Corni, S.; Gottschalk, K. E. *ChemPhysChem* **2010**, [Online] DOI: 10.1002/cphc.200900981.

(29) Hoefling, M.; Iori, F.; Corni, S.; Gottschalk, K. E. *Langmuir* **2010**, in press.

(30) Gabdoulline, R. R.; Wade, R. C. *Biophys. J.* **1997**, *72*, 1917–1929.

(31) Shkel, I. A.; Kim, S. Two Dimensional Reaction of Biological Molecules Studied by Weighted Ensemble Brownian Dynamics *Proceedings of the 4th Pacific Symposium on Biocomputing*, (PSB '99) Hawaii, January 4–9, 1999; Altman, R. B., Dunker, A. K., Hunter, L., Klein, T. E., Eds. http://psb.stanford.edu/psb-online/ (accessed Apr 20, 2010); pp 468–479.

(32) Goba, C.; Geyer, T.; Helms, V. *J. Chem. Phys.* **2004**, *121*, 457–464.

(33) Tiel, P. A.; Madey, Th. E. *Surf. Sci. Rep.* **1987**, *7*, 211–285. Henderson, M. A. *Surf. Sci. Rep.* **2002**, *46*, 1–308.

(34) Meng, Sh.; Wang, E. G.; Gao, Sh. *J. Chem. Phys.* **2003**, *119*, 7617–7620. Meng, Sh.; Wang, E. G.; Gao, Sh. *Phys. Rev. B* **2004**, *69*, 195404–17.

(35) Michaelides, A.; Ranea, V. A.; de Andres, P. L.; King, D. A. *Phys. Rev. Lett.* **2003**, *90*, 216102–216106.

(36) Criscenti, L. J.; Cygan, R. T.; Kooser, A. S.; Moffat, H. K. *Chem. Mater.* **2008**, *20*, 4682–4693.

(37) Ju, Sh.-P. J. *Chem. Phys.* **2005** *122*, 094718–094718–6.

(38) Chang, Ch.-I.; Lee, W.-J.; Young, T.-F.; Ju, Sh.-P.; Chang, Ch.-W.; Chen, H.-L.; Chang, J.-G. *J. Chem. Phys.* **2008**, *128*, 154703–154703–9.

(39) Torrie, G. M.; Kusalik, P. G.; Patey, G. N. *J. Chem. Phys.* **1988**, *88*, 7826–7840. Bérard, D. R.; Kinoshita, M.; Ye, X.; Patey, G. N. *J. Chem. Phys.* **1994**, *101*, 6271–6280. Bérard, D. R.; Kinoshita, M.; Ye, X.; Patey, G. N. *J. Chem. Phys.* **1995**, *102*, 1024–1033.

(40) Guidelli, R.; Schmickler, W. *Electrocheim. Acta.* **2000**, *45*, 2317–2338.

(41) Barten, D.; Kleijn, J. M.; Duval, J.; v. Leeuwen, H. P.; Lyklema, J.; Cohen Stuart, M. A. *Langmuir* **2003**, *19*, 1133–1139. Hillier, A. C.; Kim, S.; Bard, A. J. *J. Phys. Chem.* **1996**, *100*, 18808–18817.

(42) Ataka, K.; Yotsuyanagi, T.; Osawa, M. *J. Phys. Chem.* **1996**, *100*, 10664–10672. Nihonyanagi, S.; Ye, Sh.; Uosaki, K.; Dreesen, L.; Humbert, Ch.; Thiry, P.; Peremans, A. *Surf. Sci.* **2004**, *573*, 11–16.

(43) Schrader, M E. *J. Colloid Interface Sci.* **1984**, *100*, 372–380.

(44) Sendner, Ch.; Horinek, D.; Bocquet, L.; Netz, R. R. *Langmuir* **2009**, *25*, 10768–10781.

(45) Löfgren, P.; Ahlström, P.; Chakarov, D. V.; Lausmaa, J.; Kasemo, B. *Surf. Sci.* **1996**, *367*, L19–L25. Smith, R. S.; Huang, C.; Wong, E. K. L.; Kay, B. D. *Surf. Sci.* **1996**, *367*, L13–L18.

(46) Jorgensen, W. L.; Tirado-Rivers, J. *J. Am. Chem. Soc.* **1988**, *110*, 1657–1666.

(47) Iori, F.; Corni, S. *J. Comput. Chem.* **2008**, *29*, 1656–1666.

(48) Finnis, M. W. *Surf. Sci.* **1991**, *241*, 61–72. Sahni, V.; Bohnen, K.-P. *Phys. Rev. B* **1985**, *31*, 7651–7661.

(49) Gabdoulline, R. R.; Wade, R. C. *J. Phys. Chem.* **1996**, *100*, 3868–3878.

(50) Madura, J. D.; Briggs, J. M.; Wade, R. C.; Davis, M. E.; Luty, B. A.; Ilin, A.; Antosiewicz, J.; Gilson, M. K.; Bagheri, B.; Scott, L. R.; McCammon, J. A. *Comput. Phys. Commun.* **1995**, *9*, 57–95.

(51) Gabdoulline, R. R.; Wade, R. C. *Methods* **1998**, *14*, 329–341.

(52) Gabdoulline, R. R.; Wade, R. C. *J. Mol. Biol.* **1999**, *291*, 149–162.

(53) Davis, M. E. *J. Chem. Phys.* **1994**, *100*, 5149–5159.

(54) Hamann, C. H.; Hamnet, A.; Vielstich, W. *Electrochemistry*; Wiley: New York, 1998.

(55) Gabdoulline, R. R.; Wade, R. C. *J. Am. Chem. Soc.* **2009**, *131*, 9230–9238.

(56) Ben-Tal, N.; Honig, B.; Bagdassarian, C. K.; Ben-Shaul, A. *Biophys. J.* **2000**, *79*, 1180–1187.

(57) Harvey, S. C.; Prabhakaran, M. *J. Phys. Chem.* **1987**, *91*, 4799–4801.

(58) Beutler, T. C.; van Gunsteren, W. F. *J. Phys. Chem.* **1994**, *100*, 1492–1497.

(59) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(60) Yeh, I. C.; Berkowitz, M. L. *J. Chem. Phys.* **1999**, *111*, 3155–3162.

(61) Donnini, S.; Mark, A. E.; Juffer, A. H.; Villa, A. *J. Comput. Chem.* **2005**, *26*, 115–122.

(62) Trzesniak, D.; Kunz, A.-P. E.; van Gunsteren, W. F. *ChemPhysChem* **2007**, *8*, 162–169.

(63) Kaestner, J.; Thiel, W. *J. Chem. Phys.* **2005**, *123*, 144104. Van Eerden, J.; Briels, W. J.; Harkema, S.; Feil, D. *Chem. Phys. Lett.* **1989**, *164*, 370–376.

(64) Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. *J. Comput. Chem.* **1997**, *18*, 1463–1472.

(65) Sedlmeier, F.; Janecek, J.; Sedner, C.; Bocquet, L.; Netz, R. R.; Horinek, D. D. *Biointerphases* **2008**, *3*, FC23–FC39.

(66) Harris, J. G. *J. Phys. Chem.* **1992**, *96*, 5077–5086.

The Conformations of Amino Acids on a Gold(111) Surface

Martin Hoefling,[a] Francesco Iori,[b, c] Stefano Corni,[b] and Kay-Eberhard Gottschalk*[a]

The interactions of amino acids with inorganic surfaces are of interest for biologists and biotechnologists alike. However, the structural determinants of peptide–surface interactions have remained elusive, but are important for a structural understanding of the interactions of biomolecules with gold surfaces. Molecular dynamics simulations are a tool to analyze structures of amino acids on surfaces. However, such an approach is challenging due to lacking parameterization for many surfaces and the polarizability of metal surfaces. Herein, we report

DFT calculations of amino acid fragments in vacuo and molecular dynamics simulations of the interaction of all amino acids with a gold(111) surface in explicit solvent, using the recently introduced polarizable gold force field GolP. We describe preferred orientations of the amino acids on the metal surface. We find that all amino acids preferably interact with the gold surface at least partially with their backbone, underlining an unfolding propensity of gold surfaces.

1. Introduction

The assembly of proteins and peptides on gold surfaces plays an important role in biotechnology ranging from organic/inorganic hybrid materials for protein detection to cancer therapy. Biocombinatorial techniques allow obtaining proteins and peptides with high affinities for gold.[2–5] Some of these also behave as selective binders, that is, they are able to discriminate between gold and other inorganic surfaces by their affinity. These findings open up the way to new strategies to guide the self-assembling on nanosystems containing gold,[4,6–8] and have the potential for many other applications, such as protein-oriented immobilization on gold, or the use of gold nanoparticles as multifunctional nanocarriers.

These biocombinatorial techniques (phage display or cell surface display) use large compound libraries generated by the genetic manipulation of viruses or bacteria, each exposing a single protein of the library. Performing repeated nature-selection like cycles, the library is progressively enriched with the best binders for the target surface.[3,6] While these methods deliver peptides with the desired properties, they do not provide any direct information on the underlying physicochemical foundations. This is not only unsatisfactory from the point of view of a basic understanding of the affinity and specificity, but it also hinders the rational design of surface-binding biomolecules or mimetics. To fully exploit the potential of biomolecule/gold assemblies, the structural foundations of biomolecular adsorption on gold need to be known. While structural proteomic tools have greatly improved our understanding of biomolecular interactions, they cannot be easily extended to biomolecule/gold interactions. Computational methods, on the other hand, can give direct access to the structural properties of interacting partners, and are therefore invaluable in the understanding of protein-surface interactions. Unfortunately, full quantum mechanical approaches for even the single amino acids on surfaces in vacuo are computationally demanding, and have been performed only for a few systems.[9–12] At present, a systematic quantum mechanical study of natural amino

acid adsorption on gold, including the effects of solvation, is not feasible due to the prohibitive computational cost of such complex systems. Atomistic classical molecular dynamics, on the other hand, are feasible and offer at present the best compromise between accuracy and speed of computation. However, reliable MD simulations of biomolecules on gold or other metals are still rare,[10,13] the main problem being the lack in accuracy of the metal surface description. One of the core elements of classical MD is the force field. In fact, force fields tailored to describe protein–metal interactions in water appeared only recently.[14–16] In particular, the force field GolP, based on quantum mechanical calculations and experimental data, also taking into account metal polarizability,[17] has recently been introduced for gold surfaces. Here, we will use a combined approach of DFT computations and MD simulations with the GolP force field to compute preferred conformations of amino acids on the gold surface.

We find that both solvation and backbone have a profound effect on the preferred sidechain conformations on gold. Interestingly, the backbone of amino acids tends to strongly bind

[a] M. Hoefling, Dr. K.-E. Gottschalk+
Chair for Applied Physics, Biophysics and Molecular Materials
Ludwig-Maximilians Universität
Amalienstr. 54, 80799 Munich (Germany)
Fax: (+49) 89 2180 2050
E-mail: kay.gottschalk@physik.uni-muenchen.de

[b] Dr. F. Iori, Dr. S. Corni
S3, CNR-Istituto di Nanoscienze
via Campi 213/A, 41125 Modena (Italy)

[c] Dr. F. Iori
Dept. of Physics
University of Modena and Reggio Emilia
via Campi 213/A, 41125 Modena (Italy)

[+] Current address:
ZIK HIKE
Ernst Moritz Arndt University Greifswald
Fleischmannstraße 42–44, 17475 Greifswald (Germany)
E-mail: kay.gottschalk@uni-greifswald.de

the gold surface. This suggests beta-sheet forming peptides as good gold binders and underlines the unfolding propensity of gold surfaces for helical peptides.

2. Results and Discussion

The conformations of solvated amino acids adsorbed on gold are the results of the competition between the gold-affinity, solvation propensity and the geometric preferences of the various chemical groups (the side chain, the N and C termini) that compose each amino acid. These geometric preferences are determined by basic physical interactions between the chemical groups, the surface and the solvent.

In order to analyze this basic physics, we employed the following strategy: first we have defined a set of small molecules that mimic the corresponding chemical functionality found in amino acids (e.g. methanol to mimic the serine side chain, formamide for the protected N- and C- termini, ammonium for the unprotected N-terminus). Then, we have studied the geometric preferences of such amino acid fragments on Au(111) in vacuo, and finally we have explored how these preferences change by hydrating the system at 300 K. Finally, we simulated whole solvated amino acids absorbed on Au(111).

The relative small size of the amino acid fragments renders the use of an ab initio description of such fragments in the absence of water feasible, allowing a high-level computation of preferred orientations in comparison to force-field simulations.

2.1. DFT Calculations In Vacuo

For a high-level description of the orientation of these fragments on the gold surface, we have performed plane-waves DFT calculations for the chosen molecules on a Au(111) slab in the periodic supercell approach. The minimum energy structure of each molecule on gold has been determined. Further details can be found in refs. [15, 18].

The conformations of the unsolvated fragments found in our DFT calculations are shown in Figure 1. For chemisorbed molecules (imidazole, dimethylsulfide, methanthiol), the arrangement is dominated by the directionality of the chemical bond. Thus, the sp2 N of imidazole induces an upright orientation, while the sp3 S of dimethylsulfide and

methanthiol induces a pyramidal-like orientation, where one of the vertex of the pyramid is a surface gold atom.

The arrangement of the other, physisorbed, molecules on Au(111) is dictated by the interplay of two different molecule–surface interactions: 1) a short-range term composed of an attractive part (that collects the share of dispersion interaction mimicked by the employed DFT functional, plus tiny charge-transfer effects), and a repulsive Pauli exchange term; 2) the purely attractive image interaction.

(1) tends to maximize the number of molecular atoms contacting the surface, in the on-top sites; the molecule–gold distance is of typical non-bonding character ($\sim$ 0.3 nm). (2) favors an adsorption arrangement where the highest charge portion of the molecule is in contact with the surface, and orientations that maximize the electric field acting on the metal (e.g. perpendicular molecular dipoles vs parallel ones). (2) poses the molecule as close as possible to the gold surface. The resulting molecular arrangement is a non-obvious compromise between the intrinsic preferences of (1) and (2).

The upright position of formate, methylguanidinium and formamide is dictated by the different equilibrium distance of each atom in the molecule from the surface: due to image in-
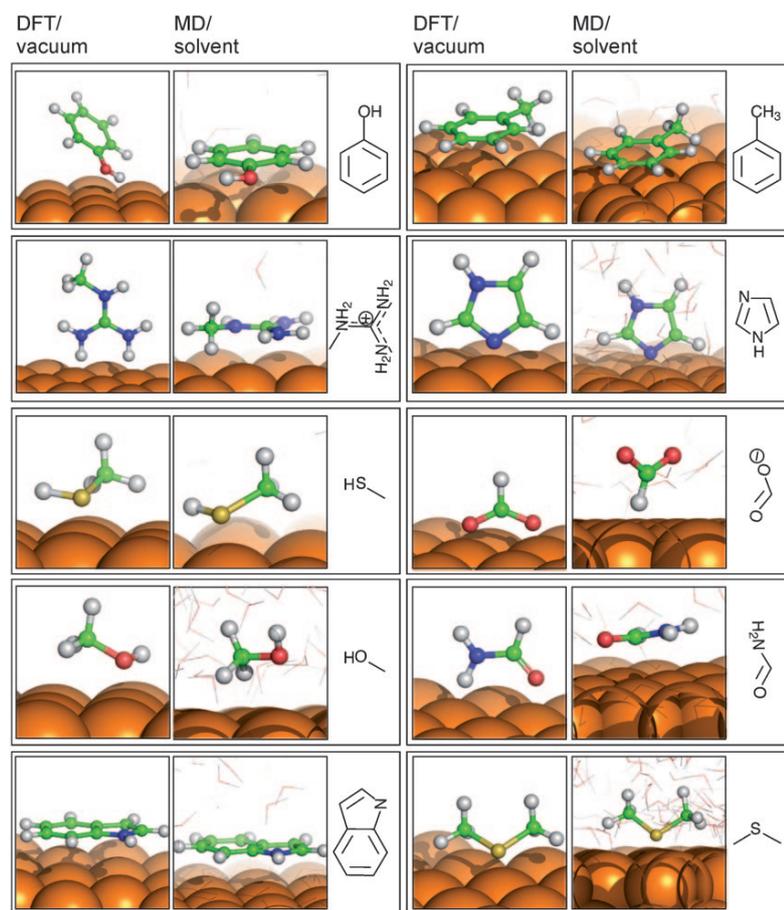

Figure 1. Conformations of amino acid building blocks from DFT calculations and the preferred clusters in fully solvated MD simulations. Figure rendered with PyMol.[1]

teraction, highly charged, exposed atoms (e.g. the O atoms of HCOO- or the H and N atoms of methylguanidinium) have a higher propensity to closely interact with the surface than the other atoms of the molecules. When the molecule is standing up, both kinds of atoms are close to their preferences. Notably, a flat orientation may be a possible arrangement (i.e. a local energy minimum) also for upright molecules, as we accidentally observed for methylguanidinium. However, our simulations show that it is not the most stable in vacuo.

In contrast, non-polar molecules (such as indole and toluene) with a neglectable image contribution prefer a flat conformation. Molecules such as methanol or phenol show an intermediate behavior, in line with their intermediate nature.

2.2. Solvated Molecular Dynamics

DFT calculations were previously compared to minimized conformations using the classical GolP forcefield with agreement in the general conformation and slight differences in the distances,[15] stressing the validity of the force field. However, the situation changes substantially for hydrated molecules. The interaction (1) is still at work (although it is now in competition with the analogous water–Au interaction), but the high dielectric water environment now screens effects of image interactions. In particular, highly charged, exposed atoms have now the possibility to interact with water molecules, with energies comparable to the image charge one. Therefore, the balance that led to standing up arrangement for the in vacuo case is compromised in water, and the standing-up orientation looses stability either with respect to a flat conformation or even an upside-down conformation, where the charges are exposed to water. The ammonium ion as a small molecule with a high charge is an extreme case demonstrating these considerations. Indeed, when solvated, it completely detaches from the gold surface and does not significantly interact with it, despite the strong interaction energy of this ion with gold observed in the DFT calculations. Methylguanidine, on the other hand, preferring an upright position in vacuo, is absorbing flatly to the gold surface in water. The differences between the two positively charged entities, the ammonium ion and methylguanidine, are the available surface area and the charge density. Methylguanidine has a significantly higher surface area allowing for significant van der Waals interaction with the gold surface in a flat conformation, which combined with a reduced charge density tips the balance of the described effects in favor of a flat absorption on gold. Interestingly, the preferred conformation of formate reverses in water: not only does it only transiently absorb onto the surface, the high dielectric environment of the solvating water interacts better with the oxygens than the gold surface. The positive charge of methylguanidine, as opposed to the negative charge of formate, appears to be better suited for interacting with the gold surface as exemplified by the flat conformation of the former and the upright conformation with water-exposed charges of the latter. Somewhat surprisingly, imidazole is hardly affected by the solvation. The DFT-conformation in vacuo and the MD-conformation in solvent are virtually indistinguishable. This underlines

the subtleties of the effects of charge screening and van der Waals interactions. As expected, more hydrophobic fragments prefer a flat orientation: phenol, indole and toluene maximize their contact area with gold in water. This arrangement minimizes the solvent accessible surface of the test molecule.

2.3. Molecular Dynamics Simulations of Solvated Amino Acids

How does the backbone of the amino acids influence the absorption behavior? To investigate this, we simulated as a next step all amino acids on the gold surface without restraints for 50 ns per amino acid. Here, we capped the N- and C-termini to avoid artifacts from charges. During this simulation, no amino acid detached from the surface. On the surface, each amino acid shows preferential conformations (Figure 2).

The analysis of Gly yields the preferred orientation of the backbone on the surface, since no sterical or electronic sidechain properties influence its absorption behavior. Our simulations show that the backbone of Gly prefers a flat conformation with maximum contact between backbone and surface. The carbonyl oxygen is pointing towards the surface, while the NH-moiety is slightly pointing upwards. In fact, not only Gly, but all of the amino acids prefer to contact the gold surface with at least part of their backbone. However, the interaction of the sidechains with the gold surface forces part of the backbone away from the surface for some of the amino acids. Arg with its guanidine side chain is a prominent example of those. As already seen in the simulations of the fragment, the sidechain is lying flat on the surface, forcing the C-terminus away from the surface. The situation of the other positively-charged amino acid, Lys, is completely different. As seen in the fragment simulations, the terminal nitrogen has no particular propensity to interact with the gold surface. Hence, one highly occupied cluster exposes the whole sidechain to the solvent. However, the methylen groups of the sidechain have a tendency to contact the gold surface in order to minimize their solvent accessible surface. This is reflected by a second, approximately equally populated cluster with a flat sidechain orientation. Differences are also seen between Asp and Glu: Glu forms two clusters on the surface with a relative population density of approximately 1:3. The lower occupied cluster has an adsorbed geometry comparable to Asp with fully solvent exposed sidechain. The higher occupied cluster, on the other hand, forms sidechain contacts of the additional methylen moiety of Glu. These two clusters demonstrate the energetic balance leading to the preferred conformations: the amino acids need to compromise between maximum backbone contact, exposure of charges to the solvent and minimizing the solvent exposure of hydrophobic moieties. This trend is also observed when going from Ala to Val. While in Ala, the one methyl group is not sufficient to disturb the backbone, the additional methylen moiety in Val tips the balance and enforces sidechain contact with accompanying backbone distortion. Interestingly, also the aromatic amino acids do not behave identically: while for Tyr and Trp, the sidechains have full contact with the gold surface with compromised backbone contact,
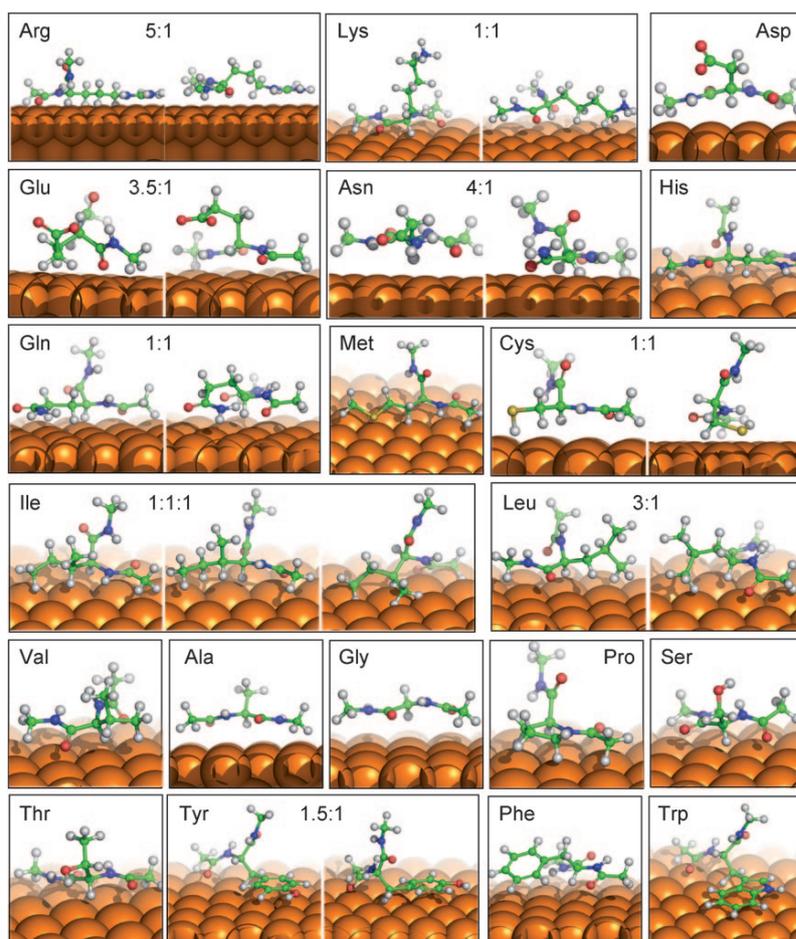
Figure 2. Conformation clusters with high population and their ratio from amino acids in free simulations on Au(111) surface. Figure rendered with PyMol.[1]

amino acid fragments studied both by DFT and MD simulations and then describing the adsorption properties of all amino acids. We find an intriguing interplay between steric and electrostatic factors. In passing from molecules adsorbed on gold in vacuo to molecule adsorbed in water, microscopic effects tends to favor orientations, where the charges are exposed to the solvent. The backbone plays an important role in the adhesion process of amino acids, indication an unfolding propensity of the gold surface. These studies are the first step for an in-depth understanding of biomolecular adsorption on gold surfaces.

Computational Methods

The DFT calculations of the amino acids fragments on the gold surface are described in details in refs. [15, 18]. They have been performed with the PWscf code included in the Quantum Espresso package. To represent the Au(111) surface, we used a slab of 4 gold layers in a $2\sqrt{3}\times3$ supercell. The PBE gradient corrected exchange correlation functional has been exploited for all the calculations, with ultrasoft pseudopotentials to represent the electron-ion interac-

the sidechain of Phe is only partially contacting the surface with optimal backbone contact. The underlying, fine balance in energetic contributions is not obvious from the chemical composition. Intuitively, one would expect full sidechain contact also for Phe. Indeed, one minor cluster shows such a conformation. However, the conformation with tilted aromatic ring is approximately tenfold overrepresented.

The observed propensity to maximize backbone contact with gold indicates 1) a certain unfolding propensity of a gold surface and 2) shows that naturally unfolded proteins or proteins with a high beta-sheet content, which potentially allows for backbone contact as opposed to alpha-helices, may be preferred gold binders. The unfolding propensity of gold surfaces has recently been demonstrated experimentally: certain proteins, when exposed to gold nanoparticles, form aggregates with unspecific interactions, indicative of unfolded proteins.[19]

4. Conclusions

We have analyzed the absorption of amino acids onto a gold surface in great detail, starting from the conformations of

tions. The basis set was composed of plane waves basis with a kinetic energy up to $E_{cut} = 25$ Ry. Brillouin zone sums have been performed with a Monkhorst–Pack k-point mesh of $4\times4\times1$. Geometry relaxations were started by different initial positions of the fragment with respect to gold, and the most stable final geometry was used for comparison with classical MD results.

For the MD simulations, we used the recently developed GolP force field[15, 18] which extends the OPLS/AA by interactions of aminoacid building blocks with Au(111), and SPC water. All simulations were performed with Nose-Hoover temperature coupling at 300 K, integration timesteps of 2 fs, Particle-Mesh-Ewald for electrostatics above 1.1 nm as well as a switch-cutoff for van der Waals interactions (0.9–1.0 nm). The Gromacs 3.3.3[20] simulation package was used for all simulations of aminoacids while Gromacs 4.0.3[21] was used for the DFT fragments. We choose to simulate in physiological ion concentration (150 mM NaCl) with a neutralized box to avoid effects from particle mesh ewald background charge on the PMF calculations.[22] Each setup was prepared by energy minimization with steepest descent algorithm followed by a position restrained run with a Parrinello–Rahman barostat applied in the z-direction for 1 ns to scale the box size after solvation.

The fragments used in the DFT calculations were placed in contact with the surface. During the relaxation with pressure coupling, no

restraints were applied. Each fragment was then simulated for 20 ns.

All aminoacids were capped with acetyl groups to prevent interactions of charged ends with the surface. The obtained structures were placed in a 4.5 nm x 4.3 nm x 4.7 nm box containing a gold slab with 5 layers. For the simulation of the adsorbed amino acids, the amino acid was put into contact with the gold surface and solvated. No restraints or constraints were applied for a simulation time of 50 ns. The amino acids were clustered with the Gromacs tool g_cluster, single linkage algorithm and a cutoff criterion of 0.5 Å RMSD.

Acknowledgements

[1] W. DeLano, *Pymol Manual* **2002**.
[2] R. Braun, M. Sarikaya, K. Schulten, *J. Biomater. Sci. Polym. Ed.* **2002**, *13*, 747.
[3] S. Brown, *Nat. Biotechnol.* **1997**, *15*, 269.
[4] B. R. Peelle, E. M. Krauland, K. D. Wittrup, A. M. Belcher, *Langmuir* **2005**, *21*, 6929.
[5] C. Tamerler, M. Duman, E. Oren, M. Gungormus, X. Xiong, T. Kacar, B. A. Parviz, M. Sarikaya, *Small* **2006**, *2*, 1372.
[6] M. Sarikaya, C. Tamerler, A. K. Jen, K. Schulten, F. Baneyx, *Nat. Mater.* **2003**, *2*, 577.
[7] M. Hnilova, E. Oren, U. Seker, B. Wilson, S. Collino, J. Evans, C. Tamerler, M. Sarikaya, *Langmuir* **2008**, *24*, 12440.
[8] K. Ishikawa, K. Yamada, S. Kumagai, K. Sano, K. Shiba, I. Yamashita, M. Kobayashi, *Appl. Phys. Express* **2008**, *1*, 034006.
[9] P. Schravendijk, N. van der Vegt, L. Delle Site, K. Kremer, *ChemPhysChem* **2005**, *6*, 1866.
[10] L. M. Ghiringhelli, L. Delle Site, *J. Am. Chem. Soc.* **2008**, *130*, 2634.
[11] A. Rimola, M. Corno, C. M. Zicovich-Wilson, P. Ugliengo, *J. Am. Chem. Soc.* **2008**, *130*, 16181.
[12] G. Hong, H. Heinz, R. R. Naik, B. L. Farmer, R. Pachter, *ACS Appl. Mater. Interfaces* **2009**, *1*, 388.
[13] P. Schravendijk, L. M. Ghiringhelli, L. Delle Site, N. Van Der Vegt, *J. Phys. Chem. C* **2007**, *111*, 2631.
[14] L. M. Ghiringhelli, B. Hess, N. F. van der Vegt, L. Delle Site, *J. Am. Chem. Soc.* **2008**, *130*, 13460.
[15] F. Iori, R. Di Felice, E. Molinari, S. Corni, *J. Comput. Chem.* **2009**, *30*, 1465.
[16] H. Heinz, R. Vaia, B. Farmer, R. Naik, *J. Phys. Chem. C* **2008**, *112*, 17281.
[17] F. Iori, S. Corni, *J. Comput. Chem.* **2008**, *29*, 1656.
[18] F. Iori, S. Corni, R. Di Felice, *J. Phys. Chem. C* **2008**, *112*, 13540.
[19] D. Zhang, O. Neumann, H. Wang, V. Yuwono, A. Barhoumi, M. Perham, J. Hartgerink, P. Wittung-Stafshede, N. Halas, *Nano Lett.* **2009**, *9*, 666.
[20] D. Van der Spoel, E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, H. J. C. Berendsen, *J. Comput. Chem.* **2005**, *26*, 1701.
[21] B. Hess, C. Kutzner, D. Van Der Spoel, E. Lindahl, *J. Chem. Theory Comput.* **2008**, *4*, 435.
[22] A. Serr, D. Horinek, R. R. Netz, *J. Am. Chem. Soc.* **2008**, *130*, 12408.

g_membed: Efficient Insertion of a Membrane Protein into an Equilibrated Lipid Bilayer with Minimal Perturbation

MAARTEN G. WOLF,[1] MARTIN HOEFLING,[2] CAMILO APONTE-SANTAMARÍA,[2]
HELMUT GRUBMÜLLER,[2] GERRIT GROENHOF[1]

[1]*Computational Biomolecular Chemistry Group, Max-Planck-Institute for Biophysical Chemistry, Göttingen D-37077, Germany*
[2]*Department for Theoretical and Computational Biophysics, Max-Planck-Institute for Biophysical Chemistry, Göttingen D-37077, Germany*

Abstract: To efficiently insert a protein into an equilibrated and fully hydrated membrane with minimal membrane perturbation we present a computational tool, called g_membed, which is part of the Gromacs suite of programs. The input consists of an equilibrated membrane system, either flat or curved, and a protein structure in the right position and orientation with respect to the lipid bilayer. g_membed first decreases the width of the protein in the *xy*-plane and removes all molecules (generally lipids and waters) that overlap with the narrowed protein. Then the protein is grown back to its full size in a short molecular dynamics simulation (typically 1000 steps), thereby pushing the lipids away to optimally accommodate the protein in the membrane. After embedding the protein in the membrane, both the lipid properties and the hydration layer are still close to equilibrium. Thus, only a short equilibration run (less then 1 ns in the cases tested) is required to re-equilibrate the membrane. Its simplicity makes g_membed very practical for use in scripting and high-throughput molecular dynamics simulations.

© 2010 Wiley Periodicals, Inc.   J Comput Chem 31: 2169–2174, 2010

Key words: computer simulations; membrane; Gromacs; molecular dynamics; protein insertion

Introduction

The number of computer simulations to probe the molecular details of membrane proteins is rapidly increasing.[1] Proteins embedded in or associated with membranes are of interest because they perform various crucial tasks, including (selective) transport, signal transduction and energy conversion. It is estimated that in most organisms 20–30% of the proteins encoded in the genome are membrane proteins[2] and that many drug targets are located at or near the cell surface (60% in 2006).[3] Nevertheless, these percentages are not reflected in the ratio of studies between soluble and membrane proteins. Because of the requirement of a lipid bilayer, membrane proteins are more complicated to crystallize or study with NMR. Consequently less than 1% of the structurally resolved proteins are membrane proteins.[4] However, the number of membrane protein structures in the pdb-database is rising quickly (see http://blanco.biomol.uci.edu/Membrane_Proteins_xtal.html for a maintained list) and thus their availability for computational studies.

Also on the simulation side, setting up a membrane protein system requires additional effor*t*s compared to soluble proteins. First, an appropriate membrane model has to be selected for which parameters are available in the desired force field. Setting

up plain membrane simulations can now be considered standard procedure[1] and hence will not be further discussed here. Furthermore, many equilibrated membrane patches and the corresponding force field parameters can be downloaded from webservers (URLs given in[1]) maintained by research groups specialized in membrane simulations. Second, and crucially, once a suitable membrane is selected the protein has to be embedded in the membrane before the production runs can be performed.

Several protocols have been suggested to insert a protein into a lipid bilayer, but up to now no standard has been established, mainly for reasons discussed later. The simplest way to insert a protein is to remove lipids and solvent molecules that overlap with the protein after combining the coordinate sets (X, Y, Z). However, the highly disordered nature of lipid tails results in an irregular and oversized hole and hence a poor lipid-protein packing. Long simulation runs are then needed to equilibrate the system. To obtain a good lipid-protein packing one can build a lipid bilayer around a protein from preequilibrated and prehydrated single lipids.[5] Unfortunately, assembling the bilayer from uncorrelated lipids will result

---

***Correspondence to:*** M. G. Wolf; e-mail: mwolf@gwdg.de
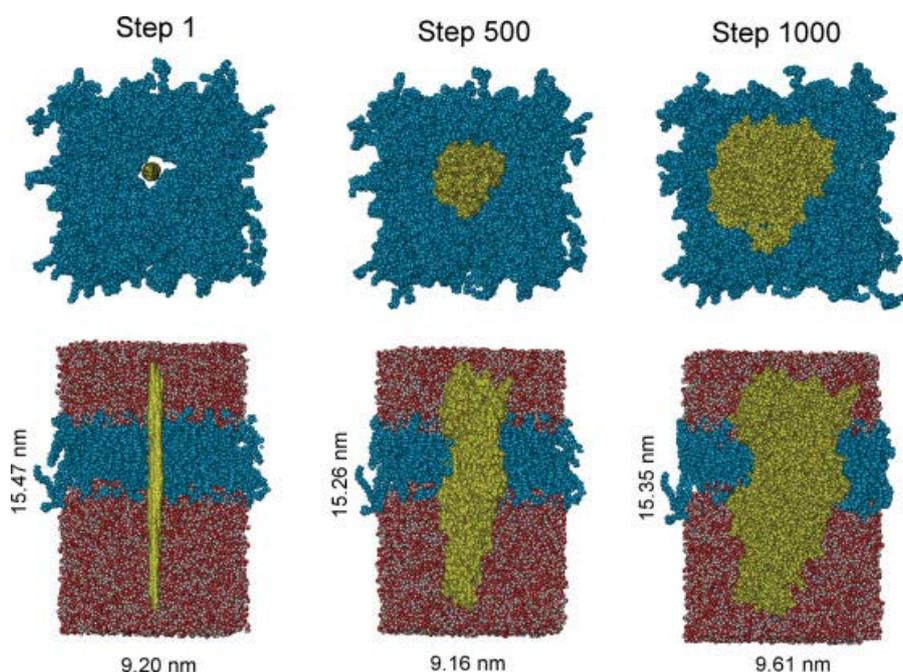
Figure 1. The embedding proces. Three snapshot of a protein (Reaction Center, see Table 1) grown within the membrane using g_membed are shown in top view (upper) and side view (lower). The protein, membrane and water are displayed as Van der Waals spheres and colored yellow, blue, and red/white, respectively. The width of the protein in the membrane plane is 0.1, 0.55, and 1.0 for step 1, step 500 and step 1000, respectively. Also the length of the box-vectors are shown.

in bad contacts (overlapping atoms), which require extensive optimisation by, for instance, rigid body translation and rotations.[5] A more sophisticated method uses repulsive forces to create a hole at the intended protein position before inserting the protein itself.[6, 7] This will result in a proper lipid-protein packing in an equilibrated membrane, but needs considerable tuning by the user to find the optimal hole generating parameters.[8] Recently, it has also been proposed to inflate the lipid bilayer, insert the protein and then, alternatingly deflate the lipid bilayer in small steps and perform an energy minimization, until the area per lipid reaches the reference equilibrium value.[8] However, this method cannot be applied with a solvated membrane and thus requires resolvation. Furthermore, it is not clear if the membrane remains in equilibrium during the deflation procedure. The lipid tails that were entangled in the equilibrated membrane are likely to clash when going from a very diluted (inflated) to a fully packed membrane, resulting in a different packing of the lipid tails. Finally, with coarse grained simulations it is possible to let the lipids self-assemble around the membrane protein.[9] Unfortunately, the level of detail of coarse grained simulation might not be sufficient to assess the relevant interactions in membrane proteins adequately. Also the inverse mapping problem, i.e., obtaining the atom positions from the coarse grained coordinates, is highly nontrivial.

To address the issue more systematically, we propose that an insertion method needs to fulfil four requirements: (1) It should be easy to use, without the requirement of parameter tuning or manual intervention during or after the embedding process. (2) It should yield a structure that is close to equilibrium, to reduce the equilibration simulation time before production runs. (3) It should be easily automated, to allow for large scale high-throughput simulations. (4) It should be distributed as part of a popular MD package, to maintain availability, functionality and retraceability. None of the aforementioned methods satisfies these four criteria and therefore we developed a new Gromacs[10] tool, g_membed, that grows a protein into a lipid bilayer during a short md simulation (1000 steps typically). By using an already hydrated and equilibrated lipid bilayer the output system (protein embedded in the solvated membrane) only requires a short equilibration run (maximal 1 ns in our test systems) to reequilibrate the membrane part of the system. Finally, by making g_membed part of Gromacs availability, functionality, and retraceability is guaranteed.

Method

The basic idea is to slowly grow the protein into an already equilibrated membrane, thereby pushing away the lipids and waters and fit the protein nicely into the lipid bilayer. Starting point of the growth process is the narrowed protein, where the narrowing of the protein is performed within the membrane plane, but not in the direction of the membrane normal. In the limiting case, all protein atoms start on a line perpendicular to the membrane plane. Shrinking the protein in 2D instead of 3D is crucial, as it avoids severe perturbation of the lipids, especially when approaching the limiting cases (a line instead of a point). The growth phase itself is a short md run. After each md step the size of the protein is slightly increased until the protein has reached its initial size (see Fig. 1 for a typical example). During this process all protein–protein interactions, including bonded interactions, are switched off.

Table 1. Systems Used to Test g_membed.

	pdb-Code	Atoms (nr)	A_{prot} (nm^2)	Membrane
Integrin αIIbβ3[11]	2K9J	1413	5.78	popc
β-Barrel Platform[12,a]	2JMM	2372	9.52	popc
Reaction Center[13]	1PRC	22,196	29.59	dopc
Yeast Aquaporin[14]	2W2E	15,868	39.10	dopc
β-Barrel Platform[12,a]	2JMM	1584	7.25	pope (vesicle)

[a]The discrepancies are due to the use of different force fields.

The method, as implemented into the Gromacs tool g_membed (*embed*ding a protein within a *mem*brane) requires a file with the protein structure overlapping the lipid bilayer spanning the *xy*-plane at the desired position and orientation as input (we assume the user already has this knowledge about the system). Optimal placement of the protein with respect to the membrane will avoid needlessly-long equilibration runs after embedding. In the case of a curved bilayer (e.g., a vesicle), it is important that the protein is overlapping a membrane part that has its normal directed along the z-axis. g_membed then performs the following steps to embed the protein (or any other desired group of atoms defined as a Gromacs index group) within a lipid bilayer.

1. Protein narrowing—The coordinates of the atoms in the protein are scaled with respect to the geometrical center of the transmembrane part of the protein by a user specified scaling factor of original (input) coordinates in the *xy*-plane and, if applicable, in the *z*-direction. Normally the protein should not be scaled in the *z*-direction. However in special cases, such as a protein that has the same height as the bilayer, increasing the size of the protein in the *z*-direction prevents lipids to envelop the protein during the growth phase.
2. Remove overlapping molecules—Every molecule not part of the protein for which at least one atom is within a user defined radius of a protein atom will be removed. If the difference between the number of lipids removed from the lower (n_{lower}) and the upper (n_{upper}) membrane leaflet is not equal to a user defined number (n_{diff}), i.e., $n_{lower} - n_{upper} \neq n_{diff}$, additional lipids will be removed, such that this equality will be obtained. This option is usefull when inserting an asymetrically shaped protein.
3. Growth phase—Iterate steps 3a and 3b $nxy + nz$ times to grow the narrowed protein to its original size.
   a. md step—Do a normal md step.
   b. Protein resizing—Change the atom coordinates of the protein by linear interpolation between the coordinates of the narrowed protein (step 1) and those of the input configuration by

$$\mathbf{r}_i = \mathbf{r}_{geom} + \mathbf{s}_i \cdot (\mathbf{r}_0 - \mathbf{r}_{geom}) \qquad (1)$$

$$\mathbf{s}_i = \begin{cases} s_{i,x} = s_{i,y} = s_{0,xy} + \frac{i \cdot (1 - s_{0,xy})}{n_{xy}} \text{ and } s_{i,z} = s_{0,z} \\ \qquad\qquad\qquad i \leq n_{xy} \\ s_{i,x} = s_{i,y} = 1 \text{ and } s_{i,z} = s_{0,z} + \frac{(i - n_{xy}) \cdot (1 - s_{0,z})}{n_z} \\ \qquad\qquad\qquad n_{xy} < i \leq n_{xy} + n_z \end{cases}$$
$$(2)$$

with $\mathbf{r}_i$ the protein atom coordinates at step i, $\mathbf{r}_0$ the input atom coordinates of the protein, $\mathbf{r}_{geom}$ the coordinates of the

geometrical center of the transmembrane region, $\mathbf{s}_i$ the scaling factor at step i, $s_{0,xy}$, and $s_{0,z}$ the initial scaling factors and n_{xy} and n_z the number of embedding steps in the *xy*-plane and the *z*-direction, respectively. This way the atom coordinates of the proteins reach the input configuration in n_{xy} steps in the *xy*-plane and, if applicable, after the *xy*-dimension is completely expanded in n_z steps in the *z*-dimension.

The output will be a file containing the protein structure properly embedded in the membrane. All atom coordinates of the protein in the output file will be equal to the coordinates provided in the input (see the Appendix for a user manual).

Case Studies

As a test we have applied g_membed to insert four proteins of various shape and size: Integrin αIIbβ3 2K9J,[11] β-Barrel Platform 2JMM,[12] Reaction Center 1PRC,[13] and Yeast Aquaporin 2W2E[14] (see Table 1). We have embedded the first two into a united-atom ($POPC$[15, 16]) and the latter two into an all-atom ($DOPC$[17]) membrane. In addition, we have embedded the β-Barrel Platform within the lipid bilayer of a vesicle (POPE), see Figure 2. The crystal waters available in the Reaction Centers pdb entry were preserved and considered as part of the protein. Initially, the width of the protein in these tests was scaled to 10% of the full-size protein ($s_{0,xy} = 0.1$). All proteins listed in Table 1 were inserted successfully within the bilayer within 1000 md steps.

To assess the time required to reach equilibrium after embedding the protein within the membrane, a 10 ns standard md simulation was performed after the embedding (insertion within the vesicle not
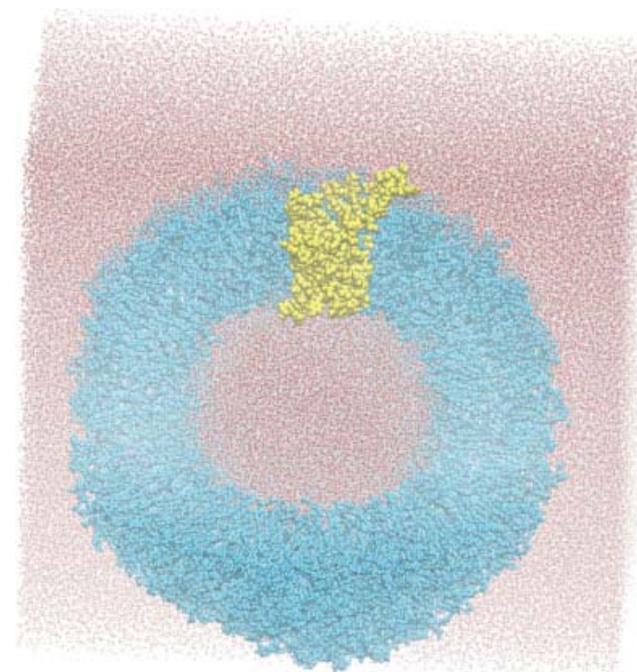

Figure 2. A protein embedded in a vesicle bilayer. The β-Barrel platform (yellow) has been embedded within a POPE vesicle (blue) in water (red/white) with g_membed.
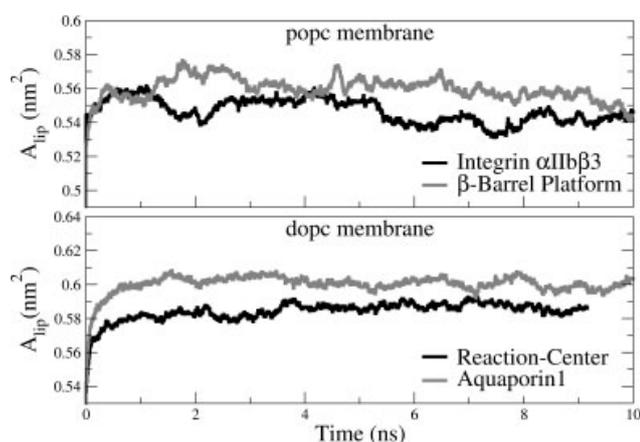
Figure 3. Equilibrating a system after embedding a protein with g_membed into a membrane starting with an initial scaling factor of 0.1. The area per lipid (A_{lip}, eq. 3) as a function of simulation time shows that within 1 ns the equilibrium area per lipid is reached.

included). Figure 3 shows the area per lipid (A_{lip}) as a function of time, where A_{lip} is obtained through

$$A_{lip} = 2 * (A_{xy} - A_{prot})/n_{lip} \tag{3}$$

with n_{lip} the number of lipids in the membrane, A_{xy} the area spanned by the x- and y-box vectors (membrane plane) and A_{prot} the protein area estimated by projecting all the atoms of the protein that overlap the membrane slab on a two-dimensional square grid in the xy-plane and summing the area of all grid elements that contain a protein atom. From Figure 3 it is clear that already within 1 ns the system reaches the equilibrium area per lipid. For a pure popc and dopc membrane system we found an area per lipid of 0.59 and 0.62, respectively, which is slightly higher than the values we found for the protein-membrane system. This discrepancy can partly be explained by the fact that the protein area is overestimated, resulting in an underestimation of the area per lipid. Furthermore, since the influence of proteins on lipid properties is still not completely understood, it is not clear whether the area per lipid should actually be unchanged after protein insertion.[18, 19]

Reaching the equilibrium area per lipid does not necessarily imply that the membrane is in equilibrium. As a more rigorous test, we therefore also analyzed the deuterium order parameter, S_{CD}, a measure for the amount of disorder in the lipid tails. Comparing the deuterium order parameter before and after embedding the protein into the membrane did not show any significant differences. In addition, we calculated the density profile of water normal to the membrane plane to compare the hydration of the membrane. Also here we found no significant difference before and after embedding the protein. Monitoring the deuterium order parameter and the water density profile normal to the membrane over time indicate that the membrane is still in equilibrium.

To determine the time required to reach area equilibrium, we calculated the relaxation time τ associated with the area per lipid by fitting

$$A_{lip} = \langle A_{lip} \rangle - A_0 * exp(-t/\tau) \tag{4}$$

to the area per lipid as a function of time (Fig. 3). In eq. 4, $\langle A_{lip} \rangle$ is the ensemble average of the area per lipid, where we averaged over the 2–10 ns time interval and A_0 the fitted difference with $\langle A_{lip} \rangle$ at $t = 0$. The resulting relaxation times are listed in Table 2, column 0.1. The average relaxation time within the scaling factor range 0.1–0.3 (including data not listed in Table 2) is 180 ± 190 ps. From this value and the values listed in Table 2 we expect that a 1 ns equilibration run suffices after embedding a protein in an equilibrated membrane using g_membed. Nevertheless, as the number of cases studied here is limited, the required equilibration time might differ for some specific systems.

Finally, we examined the influence of the protein scaling factor $s_{0,xy}$ used in the first step of the embedding process. We expect that a larger initial scaling factor implies slower equilibration, since more lipids have to be moved to accomodate the protein. We embedded a Yeast Aquaporin tetramer and the β-Barrel Platform starting with initial scaling factors ranging from 0.1 to 0.7, followed by a 10 ns standard simulation. The root mean square deviation (rmsd) of the Yeast Aquaporin tetramer with respect to the X-ray structure approaches the same value, 0.15 nm, for all initial scaling factors, whereas the rmsd for the β-Barrel Platform ranges from 0.3 to 0.6 nm without any correlation to the initial scaling factor. These findings indicate that the protein structure is unperturbed by the initial scaling factor. Table 2 shows the relaxation times of the area per lipid for different initial scaling factors. These values show that indeed our hypothesis that a smaller scaling factor implies slower equilibration is true. When the scaling factor approaches a value, for which the removed lipid area is only slightly smaller than the estimated protein area (0.6 and 0.7 for the β-Barrel Platform and Yeast Aquaporin, respectively) the perturbation to the membrane as measured by the area per lipid, deuterium order parameter and water density profile is minimal in the cases tested. The length of the simulation to equilibrate the system is then mainly determined by the equilibration of the protein.

Conclusions

We have presented a program, g_membed, for efficient insertion of a protein structure into a hydrated and equilibrated membrane. g_membed can handle both flat and curved membranes, thereby for instance also allowing embedding of protein structures within vesicle bilayers. The program only minimally perturbs the properties and the hydration of the system during protein insertion. As a consequence short subsequent equilibration runs suffice, enabling drastically increased throughput. In addition, one set of parameters (set as default) can be aplied to insert proteins of various shape and

Table 2. Relaxation Time τ (**ps**) of the Area per Lipid After Embedding the Protein in a Membrane with Different Initial Protein Sizes.

		Initial protein scaling factor $s_{0,xy}$						
		0.1	0.2	0.3	0.4	0.5	0.6	0.7
Protein (membrane)	Integrin αIIbβ3 (popc)	2						
	β-Barrel Platform (popc)	113	115	622	2		2	0
	Reaction Center (dopc)	306						
	Yeast Aquaporin (dopc)	153		66		49		0

size into an arbitrary lipid bilayer, suggesting this set to be useful for most proteins.

We consider the following two setups most useful. (1) The protein is embedded within a membrane slab that is relatively small compared to the protein size, so that a small number of lipids should be removed. Consequently, the initial scaling factor $s_{0,xy}$ needs to be set to a very small value (e.g., 0.1 of the full protein size). After embedding, a 1 ns equilibration is recommended to reequilibrate the membrane, but longer equilibration times may be required for the protein. This setup is the default for g_membed. (2) Alternatively, if the protein is to be embedded within a membrane slab that is large with respect to the proteins size, the protein scaling factor $s_{0,xy}$ can be set to a value, such that g_membed removes a lipid portion that is slightly smaller than the estimated protein area (depending on the protein size a scaling factor between 0.5 and 0.7). In this case, the perturbation of the system is even smaller, and the length of the equilibration run is determined by the time it takes to equilibrate the protein structure.

Computational Details

g_membed

As g_membed uses the molecular dynamics code implemented in Gromacs,[10] all system options available to a standard md simulation also apply to g_membed. The Integrin $\alpha IIb\beta 3$ and the β-Barrel Platform were simulated using the OPLS-AA force field[20] and embedded in the popc lipid bilayer[15, 16] combined with simple point charge (SPC) water.[21] For the Reaction-Centre and Yeast Aquaporin we used the Amber03 parameter set[22] with the dopc membrane[17] and TIP3P water.[23] During embedding of the β-Barrel Platform within the pope vesicle we applied the Gromos-87 force field[24] with SPC water.

To insert a protein in the membrane by g_membed, 1000 time steps were performed. A Van der Waals cut-off of 1.4 nm was used

and the electrostatic interactions were treated using PME[25] with a real space cut-off of 1.0 nm. All bonds were constraint using the LINCS algorithm,[26] allowing a time step of 2 fs. The temperature and pressure were kept constant at 300 K and 1 bar using velocity rescaling[27] and Berendsen semi-isotropic pressure coupling,[28] respectively. Protein–protein interactions were excluded and the atoms of the protein were set to the position specified in step 3 of the g_membed protocol.

Equilibration Runs

After embedding the protein in the membrane a 10 ns equilibration run was performed. The same system parameters were used as during embedding, with the exception that the temperature of the popc simulations was set to 323 K to prevent ordering of the bilayer. Also, the protein–protein interactions are not excluded and the atom positions of the protein are no longer fixed.

Acknowledgement

Appendix: Manual

The required input for g_membed is a tpr-file containing the group to embed (further called protein) and the (solvated) membrane. The protein should be placed at the desired position and orientation overlapping the lipid bilayer by using for instance pymol.[29] If the position and orientation of the protein is satisfactory, merge the protein and (solvated) membrane structure files into one (e.g., `merged.gro`) and create a matching topology file.

Download `sample.mdp` from our website (http://wwwuser.gwdg.de/~ggroenh/membed.html) or set the following options in an mdp file.

```
integrator     = md
energygrp      = Protein (or other group that you want to insert)
freezegrps     = Protein
freezedim      = Y Y Y
energygrp_excl = Protein Protein
```

Generate the input file for g_membed, input.tpr, with the gromacs preprocessor.

```
grompp -f sample.mdp -c merged.gro -p merged.top -o input.tpr
```

Then run

```
g_membed -f input.tpr -p merged.top -xyinit 0.1 -xyend 1.0 -nxy 1000 or
g_membed -f input.tpr -p merged.top -xyinit 0.1 -xyend 1.0 -nxy 1000 -zinit 1.1 -zend 1.0
-nz 100
```

and select the group that has to be inserted. The options of g_membed are -f the required input tpr-file, -p to update the number of molecules in the topology file, -xyinit and -zinit the scaling factor for the width and height, respectively, of the protein in the membrane plane at the start of the embedding and -nxy and -nz the number of steps to reset -xyinit and -zinit to

-xyend and -zend, respectively. If the protein is (very) asymmetric in shape use -ndiff to remove more lipids from the lower (negative integer) or upper (positive integer) leaflet.

Finally, a 1 ns equilibration run should be run before performing any production runs.

```
grompp -f equi.mdp -c membedded.gro -p merged.top -o equilibrate.tpr
mdrun -deffnm equilibrate
```

When the group to embed is not a default group, such as a protein and its crystal water, an ndx file should also be provided to g_membed. Make sure all the molecule types in the group to embed are unique, e.g., the molecule type of the crystal waters should be different from that of the solvent. Also the freeze and energy exclusion parameters in the mdp file should be changed to match the name of the group to embed.

Note that the program will also issue a warning for most common mistakes we have encountered (will be updated).

References

1. Biggin, P.; Bond, P. Methods Mol Biol 2008, 443, 147.
2. Wallin, E.; von Heijne, G. Protein Sci 1998, 7, 1029.
3. Overington, J.; Al-Lazikani, B.; Hopkins, A. Nat Rev Drug Discov 2006, 5, 993.
4. Carpenter, E.; Beis, K.; Cameron, A.; Iwata, S. Curr Opin Struct Biol 2008, 18, 581.
5. Woolf, T. B.; Roux, B. Proteins 1996, 24, 92.
6. Faraldo-Gómez, J.; Smith, G.; Sansom, M. Eur Biophys J 2002, 31, 217.
7. Shen, L.; Bassolino, D.; Stouch, T. Biophys J 1997, 73, 3.
8. Kandt, C.; Ash, W.; Tieleman, D. Methods 2007, 41, 475.
9. Scott, K.; Bond, P.; Ivetac, A.; Chetwynd, A.; Khalid, S.; Sansom, M. Structure 2008, 16, 621.
10. Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. J Chem Theor Comp 2008, 4, 435.
11. Lau, T.-L.; Kim, C.; Ginsberg, M.; Ulmer, T. EMBO J 2009, 28, 1351.
12. Johansson, M.; Alioth, S.; Hu, K.; Walser, R.; Koebnik, R.; Pervushin, K. Biochemistry 2007, 46, 1128.
13. Deisenhofer, J.; Epp, O.; Sinning, I.; Michel, H. J Mol Biol 1995, 246, 429.
14. Fischer, G.; Kosinska-Eriksson, U.; Aponte-Santamaría, C.; Palmgren, M.; Geijer, C.; Hedfalk, K.; Hohmann, S.; de Groot, B. L.; Neutze, R.; Lindkvist-Petersson, K. PLoS Biol 2009, 7, e1000130.
15. Berger, O.; Edholm, O.; Jühnig, F. Biophys J 1997, 72, 2002.
16. Tieleman, P.; MacCallum, J.; Ash, W.; Kandt, C.; Xu, Z.; Monticelli, L. J Phys: Condens Matter 2006, 18, S1221.
17. Siu, S. W. I.; Vácha, R.; Jungwirth, P.; Böchmann, R. A. J Chem Phys 2008, 128, 125103.
18. Periole, X.; Huber, T.; Marrink, S.-J.; Sakmar, T. P. J Am Chem Soc 2007, 129, 10126.
19. Ollila, O. H. S.; Risselada, H. J.; Louhivuori, M.; Lindahl, E.; Vattulainen, I.; Marrink, S. J Phys Rev Lett 2009, 102, 078101.
20. Kaminski, G.; Friesner, R.; Tirado-Rives, J.; Jorgensen, W. J Phys Chem B 2001, 105, 6474.
21. Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; Hermans, J. In Intermolecular Forces; Pullman, B., Eds.; D. Reidel Publishing Company: Dordrecht, 1981; pp. 331–342.
22. Sorin, E. J.; Pande, V. S. Biophys J 2005, 88, 2472.
23. Jorgensen, W.; Chandrasekhar, J.; Madura, J.; Impey, R.; Klein, M. J Chem Phys 1983, 79, 926.
24. van Gunsteren, W. F.; Berendsen, H. J. C. Gromos-87 Manual Biomos BV Nijenborgh 4, 1947 AG Groningen, The Netherlands, 1987.
25. Darden, T.; York, D.; Pedersen, L. J Chem Phys 1993, 98, 10089.
26. Hess, B.; Bekker, H.; Berendsen, H. J. C.; Fraaije, J. G. E. M. J Comput Chem 1997, 18, 1463.
27. Bussi, G.; Donadio, D.; Parrinello, M. J Chem Phys 2007, 126, 014101.
28. Berendsen, H.; Postma, J.; DiNola, A.; Haak, J. J Chem Phys 1984, 81, 3684.
29. DeLano, W. The Pymol Molecular Graphics System; DeLano Scientific LLC: Palo Alto, CA, USA, 2008. Available at: http://www.pymol.org.

Interaction of Amino Acids with the Au(111) Surface: Adsorption Free Energies from Molecular Dynamics Simulations

Martin Hoefling,[†] Francesco Iori,[‡] Stefano Corni,[‡] and Kay-Eberhard Gottschalk*,[†,§]

[†]*Chair for Applied Physics, Biophysics and Molecular Materials, Ludwig-Maximilians University, Amalienstr. 54, 80799 Munich, Germany, and* [‡]*Center S3, CNR Institute of Nanoscience and Department of Physics, University of Modena and Reggio Emilia, Via Campi 213/A, I-41125 Modena, Italy.* [§]*Present address: Center for Autoimmunity Diseases, University of Greifswald, Fleischmannstraße 42-44, 17475 Greifswald, Germany*

Interactions of proteins with inorganic surfaces are of high importance in biological events and in modern biotechnological applications. Therefore, peptides have been engineered to recognize inorganic surfaces with high specificity. However, the underlying interactions are still not well understood. Here, we investigated the adsorption of amino acids as protein building blocks onto a Au(111) surface. In particular, using molecular dynamics simulations, we calculated the potential of mean force between all the 20 amino acids and the gold surface. We found a strong dependence of the binding affinities on the chemical character of the amino acids. Additionally, the interaction free energy is correlated with the propensity of amino acids to form β-sheets, hinting at design principles for gold binding peptides and induction of β-sheet formation near surfaces.

Introduction

Interactions of proteins with inorganic surfaces are involved in several biological processes[1] such as hard-tissue growth,[2] cell–surface adhesion,[3] and inhibition of ice formation inside cells.[4] The physiologically important event of protein interaction with inorganic surfaces recently obtained new attention emerging from the ability of engineered peptides[5] and antibodies[6] to recognize specific target inorganic surfaces, which bears promise for applications in fields such as nanobiotechnology and biomedicine.[7]

In this framework, gold is a particularly relevant material for different reasons: Combinatorial techniques yielded peptides with high affinities for gold,[5,8,9,10] and self-assembled nanosystems have been suggested as leveraging such optimized gold-protein interactions.[9,7,11,12] In addition, gold is important for technological applications (e.g., contacts in bioelectronics) due to its biocompatibility and well-behaved properties.

Despite the importance of protein–surface interactions, the underlying physicochemical principles are still not fully understood. The crucial first step in a bottom-up approach aimed at understanding the interactions of proteins with gold surfaces is to focus on the building blocks of proteins, the natural amino acids. At this level, several important questions can be addressed: What is the relative strength of the interaction between amino acids and gold surfaces? How do these correlate with the intrinsic properties of amino acids, such as hydrophobicity and secondary structure propensity? Computational methods can contribute to answering these questions and provide deepened insight into the atomistic foundations of protein–surface interactions.[13,14] However, full quantum mechanical approaches for even the single amino acids on surfaces in vacuo are computationally demanding, and have been performed only for a few systems.[15–18] A computationally more feasible approach is using molecular dynamics simulations as described for some peptides or amino acids on gold surfaces.[8,19–22] Problematic here is the polarizability of metal surfaces, which is difficult to include in classical force fields. However, new force fields tailored to describe the interactions of biomolecules with metal surfaces appeared recently.[19,23] Based on one of these, developed for the gold (111) surface, we have performed extensive molecular dynamics (MD) studies on the free energy of adsorption, answering the question of relative

*To whom correspondence should be addressed. E-mail: kay.gottschalk@physik.uni-muenchen.de. Telephone: +49-89-2180 3436. Fax: +49-89-2180 2050.

(1) Gray, J. J. *Curr. Opin. Struct. Biol.* **2004**, *14*, 110–115.
(2) Addadi, L.; Weiner, S. *Angew. Chem., Int. Ed. Engl.* **1992**, *31*, 153–169.
(3) Iuliano, D. J.; Saavedera, S. S.; Truskey, G. A. *J. Biomed. Mater. Res.* **1993**, *27*, 1103–1113.
(4) Yeh, Y.; Feeney, R. E. *Chem. Rev.* **1996**, *96*, 601–617.
(5) Brown, S. *Nat. Biotechnol.* **1997**, *15*, 269–272.
(6) Schnirman, A. A.; Zahavi, E.; Yeger, H.; Rosenfeld, R.; Benhar, I.; Reiter, Y.; Sivan, U. *Nano Lett.* **2006**, *6*, 1870–1874.
(7) Sarikaya, M.; Tamerler, C.; Jen, A. K. Y.; Schulten, K.; Baneyx, F. *Nat. Mater.* **2003**, *2*, 577–585.
(8) Braun, R.; Sarikaya, M.; Schulten, K. *J. Biomater. Sci.* **2002**, *13*, 747–757.
(9) Peelle, B. R.; Krauland, E. M.; Wittrup, K. D.; Belcher, A. M. *Langmuir* **2005**, *21*, 6929–6933.
(10) Tamerler, C.; Duman, M.; Oren, E. E.; Gungormus, M.; Xiong, X. R.; Kacar, T.; Parviz, B. A.; Sarikaya, M. *Small* **2006**, *2*, 1372–1378.
(11) Hnilova, M.; Oren, E. E.; Seker, U. O. S.; Wilson, B. R.; Collino, S.; Evans, J. S.; Tamerler, C.; Sarikaya, M. *Langmuir* **2008**, *24*, 12440–12445.
(12) Ishikawa, K.; Yamada, K.; Kumagai, S.; Sano, K. I.; Shiba, K.; Yamashita, I.; Kobayashi, M. *Appl. Phys. Express* **2008**, *1*, 034006-3–.

(13) Harding, J. H.; Duffy, D. M.; Sushko, M. L.; Rodger, P. M.; Quigley, D.; Elliott, J. A. *Chem. Rev.* **2008**, *108*, 4823–4854.
(14) Latour, R. A. *Biointerphases* **2008**, *3*, FC2–FC12.
(15) Ghiringhelli, L. M.; Delle Site, L. *J. Am. Chem. Soc.* **2008**, *130*, 2634–2638.
(16) Hong, G. Y.; Heinz, H.; Naik, R. R.; Farmer, B. L.; Pachter, R. *ACS Appl. Mater. Interfaces* **2009**, *1*, 388–392.
(17) Rimola, A.; Corno, M.; Zicovich-Wilson, C. M.; Ugliengo, P. *J. Am. Chem. Soc.* **2008**, *130*, 16181–16183.
(18) Rimola, A.; Corno, M.; Zicovich-Wilson, C. M.; Ugliengo, P. *Phys. Chem. Chem. Phys.* **2009**, *11*, 11662–11662.
(19) Ghiringhelli, L. M.; Hess, B.; van der Vegt, N. F. A.; Delle Site, L. *J. Am. Chem. Soc.* **2008**, *130*, 13460–13464.
(20) Heinz, H.; Farmer, B. L.; Pandey, R. B.; Slocik, J. M.; Patnaik, S. S.; Pachter, R.; Naik, R. R. *J. Am. Chem. Soc.* **2009**, *131*, 9704–9714.
(21) Schravendijk, P.; Ghiringhelli, L. M.; Delle Site, L.; van der Vegt, N. F. A. *J. Phys. Chem. C* **2007**, *111*, 2631–2642.
(22) Verde, A. V.; Acres, J. M.; Maranas, J. K. *Biomacromolecules* **2009**, *10*, 2118–2128.
(23) Iori, F.; Di Felice, R.; Molinari, E.; Corni, S. *J. Comput. Chem.* **2009**, *30*, 1465–1476.

interaction strengths between the amino acids and gold surfaces. Our results correlate with available experimental data and show a clear dependency of the interaction free energy depending on the chemical nature of the amino acid. We also analyzed the correlation between the obtained free energies of adsorption and other intrinsic physiochemical properties of amino acids. Here we found a correlation of the β-sheet propensity of amino acids with the free energy of binding, hinting at a valuable design principle for gold binding peptides.

Materials and Methods

For the MD simulations, we used the recently developed GolP force field[23,24] which extends the OPLS/AA[25] force field by interactions of amino acid building blocks with Au(111) and includes a term describing metal polarizability. All simulations were performed in explicit (SPC) solvent, Nose−Hoover temperature coupling at 300 K, integration time steps of 2 fs, particle mesh Ewald for electrostatics above 1.1 nm, as well as a switch cutoff for van der Waals interactions (0.9−1.0 nm). The gromacs 3.3.3[26] simulation package was used for all constrained and free simulations of amino acids. We chose to simulate in physiological ion concentration (150 mM NaCl) with a neutralized box to avoid effects from particle mesh Ewald background charge on the potential of mean force (PMF) calculations.[27] Each setup was prepared by energy minimization with a steepest descent algorithm followed by a position restrained run with a Parrinello-Rahman barostat coupled to the z-direction for 1 ns to scale the box size after solvation.

All amino acid termini were capped to mimic a peptide chain and to prevent interactions of charged amino or carboxyl moieties with the surface. The N-terminus was extended with an acetyl group; for the C-terminus, an N-methyl moiety replaced the OH group of the carboxylic acid. The obtained capped amino acids were placed in a 4.5 nm × 4.3 nm × 4.7 nm box containing a gold slab with five layers. The side chain topologies were set according to the protonation state suitable for neutral pH.

Free adsorption simulations of amino acids were performed with starting structures with the center of mass (COM) placed > 1 nm away from the surface. The time per simulation was 5 ns.

In unrestrained simulations, the COM of an amino acid can move freely within the simulation box. The average force that one has to apply to the COM of an amino acid to restrain it at a given distance from the surface corresponds to the negative thermodynamic force acting on the COM at the given distance. Simulating the amino acid at different distances and integrating over the distances of restraining forces yields the free energy of association as given by the PMF. We used an initial pull simulation starting from the middle between the slab surfaces to generate input structures at different distances from the surface for the constraint-biased simulations. From the pull simulations, 4 sets of input conformations each consisting of frames in 27 different distances from the surface were set up. The first 0.4 nm were sampled in 0.025 nm steps, the following 0.2 nm were sampled by 0.05 nm steps, while for the remaining 0.6 nm steps of 0.1 nm were applied. In each simulation, a harmonic spring with a spring constant of 5000 kJ/mol/nm^2 in the z-direction was used to constrain the COM of the amino acid (stiff spring approximation). The xy-translations as well as rotations were not constrained. Only the last 0.5 ns of the 5 ns simulations were used for analysis. The restraint forces during the simulations were monitored for each amino acid and distance. An average of these forces was then mathematically integrated in order to obtain
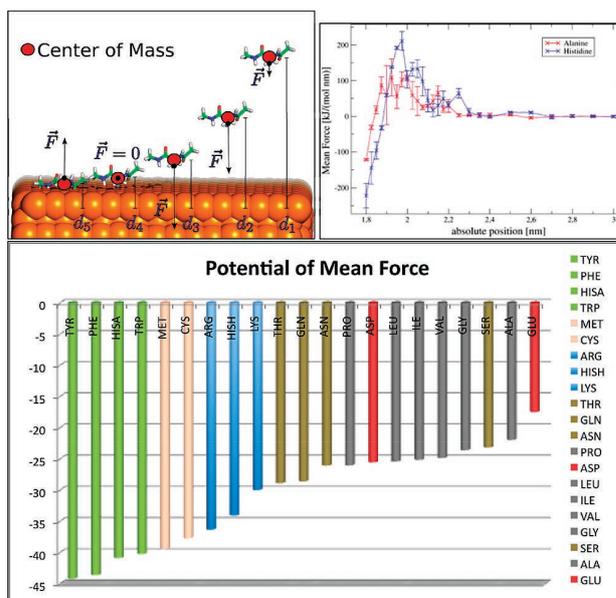

Figure 1. (a) PMF calculation based on constraint biased simulations at different distances. (b) Mean force (MF) plot which can be integrated. (c) Free energy differences for each amino acid obtained from the PMF in kJ/mol.

relative free energy differences. The boundary condition of free energy $A = 0$ kJ/mol at our maximum separation distance of 2 nm yielded the integration constant. The standard error of the adsorption free energy was calculated by first taking the block average of the force[28] and then propagating the errors in the integration. These standard errors were between 1.6 and 3.5 kJ/mol.

Results and Discussion

Adsorption Free Energies. A crucial question in understanding the interaction of amino acids with gold surfaces is the free energy of interaction. Different strategies can be employed for this task. Recently, it has been demonstrated that the potential of mean force calculated from constrained biased simulations is a computationally effective and within the limits of the applied force field accurate way to compute free energies of interacting particles.[29] Using constrained biased simulations, a reaction coordinate between the two states of interest needs to be defined. For the interaction of amino acids with a surface, the reaction coordinate is the distance between the COM of each amino acid and the gold layer (Figure 1). Using this, we computed the PMF for all amino acids by monitoring and averaging the constraint force at a given distance (Figure 1). Integrating this force versus distance yields the free energy of association (Figure 1 and Table 1). Cys, containing a thiol group, is known to chemisorb onto gold surfaces. Scanning probe microscopy experiments showed that Au−S bond formation can induce surface defects upon adsorption.[30] Later on, combined crystallographic and theoretical studies revealed the substructure of the vacancies.[31] The description of bond breakage and bond formation is beyond the capabilities of our method. Therefore, in our simulations, we

(24) Iori, F.; Corni, S. *J. Comput. Chem.* **2008**, *29*, 1656–1666.

(25) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. *J. Am. Chem. Soc.* **1996**, *118*, 11225–11236.

(26) Van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. *J. Comput. Chem.* **2005**, *26*, 1701–1718.

(27) Donnini, S.; Mark, A. E.; Juffer, A. H.; Villa, A. *J. Comput. Chem.* **2005**, *26*, 115–122.

(28) Hess, B. *J. Chem. Phys.* **2002**, *116*, 209–217.

(29) Trzesniak, D.; Kunz, A. P. E.; van Gunsteren, W. F. *ChemPhysChem* **2007**, *8*, 162–169.

(30) McCarley, R. L.; Dunaway, D. J.; Willicut, R. J. *Langmuir* **1993**, *9*, 2775–2777.

(31) Cossaro, A.; Mazzarello, R.; Rousseau, R.; Casalis, L.; Verdini, A.; Kohlmeyer, A.; Floreano, L.; Scandolo, S.; Morgante, A.; Klein, M. L.; Scoles, G. *Science* **2008**, *321*, 943–946.

Table 1. Absolute Value of the Interaction Free Energy in kJ/mol of the Noncovalent Association of the Amino Acids with a Gold (111) Surface

Ala	Arg	Asn	Asp	Cys	Gln	Glu	Gly	His	Ile	Leu	Lys	Met	Phe	Pro	Ser	Thr	Trp	Tyr	Val
21.9	36.3	26.1	25.5	37.7	28.6	17.5	23.6	34.0	25.1	25.4	30.0	39.3	43.6	26.0	23.1	28.9	40.2	44.2	24.8

do not calculate chemisorption, but only the initial phase of association. Our strongest obtained interaction energies are much smaller than the adsorption energy predicted for the Au−S bond (∼200 kJ/mol),[31] and therefore, less disruptive effects on the gold surface are expected. While our interaction energies may not be directly comparable with this Au−S value (we are calculating free energies, in solutions), the directly comparable interaction energies of neutral chemical groups composing the natural amino acids are also much smaller than the Au−S value.[19,23] However, the energy needed to form a defect on the Au surface is not very high (around 50 kJ/mol)[31] and is within the range of our here computed energies. Therefore, our modeling assuming an atomistically flat and rigid surface might underestimate effects stemming from structural reorganization upon the adsorption process, in particular in the dense covered regime. However, our approach using a flat surface for weakly bound molecules is supported by ab initio modeling of water,[32] phenylalanine,[15] and imidazole,[33] where only minimal rearrangements of the Au(111) surfaces were found. Yet, more computational and theoretical work is needed to gain deeper insight into the effects of surface irregularities on the adsorption.

From our simulations, we observe a general trend following aromatic < sulfur < positive < polar < aliphatic ∼ negative. Several factors (such as electrostatic, dispersion, metal polarization, direct chemical interactions, hydrophobic effect) may contribute to determine this order. From previous computational studies,[23,33,34] we know that His and Met side chains can create interactions approaching covalent bonds, although the resulting interaction is much weaker than that for deprotonated thiols and Cys. This explains their high affinity for gold. For aromatic amino acids, stronger-than-expected interactions were also observed.[23] These strong interactions may be caused by π-electron mediated effects. Moreover, the planarity of aromatic amino acids allows maximizing the dispersion interaction with the equally planar surface. Remarkably, although the interaction is so favorable with aromatic amino acids, the surface is not behaving hydrophobically, since apolar aliphatic amino acids have poor binding affinities. The different behavior of positive and negative amino acids is surprising, considering that the gold surface is overall neutral. We are performing more work on this result, but the origin of this difference (that may also be related to the different structures of Arg and Lys with respect to Glu and Asp, and not only to their different charges) is not well understood yet.

Comparison with Experimental Data. Most of the experimental literature mainly focuses the on the strong binding of sulfur containing amino acids to gold surfaces. Moreover, even if several gold-binding peptides have been reported so far (e.g., refs 5 and 8−10), experimental data for amino acids binding affinity to gold, that would be directly comparable with our results, are lacking. Thus, the here reported values for adsorption free energies are computational predictions waiting for experimental verification. Nevertheless, the soundness of these predictions can be estimated by comparison to experiments on related systems in the weak binding regime. Recently, the coverage of different inorganic surfaces with homopeptides of natural amino acids has been reported. Unfortunately, the sensitivity for measurements on gold surfaces was too low to allow quantitative comparison to our data.[35] In a peptide library screened for gold binding peptides, Arg, Trp, Tyr, and Cys were overrepresented in strong binding peptides, in agreements with our results.[11] However, the method used to discriminate between strong, medium, and weak binders in ref 11 (based on adhered cell counting) does not directly measure the gold-binding affinity of the peptides, a fact that further decreases the comparability with our calculated data. Reference 11 indeed reports absorption free energy data, but for some specific peptides and not for single amino acids.

Our computed energies correlate with experimentally determined affinities of peptides displayed on phage viruses.[9] In this experiment, the strongest interacting, nonsulfur containing homohexamers were Trp and His.

Both residues are also strongly interacting in our simulations. Interestingly, other aromatic residues like Phe did not interact strongly in the experimental system. This may be caused by intramolecular hydrophobic or π-stacking interactions in these homomeric peptides, which prevent the side chains from interacting with gold in the experiments. The authors of the experimental study also designed histidine-containing heptapeptides (XHXHXHX, where X is a variable amino acid). Our energies correlate with the gold coverage of phage viruses expressing these peptides (Figure 2A), although the correlation is not very strong. The experimental system with small, heterogeneous peptides expressed on phage viruses is different from the system simulated here in many regards. In the experiment, possible cooperative effects between the histidine and the amino acids are likely, in particular since the experimental system was dependent on the exact sequence of X and His, and not only on the relative composition. This underlines the difficulties encountered when comparing the experimental results with our simulations. Despite these considerable differences in the systems, the correlation is encouraging.

Relation with Intrinsic Amino Acid Properties. Interestingly, the association energy is correlated with the energy of amino acids to be brought from a water/lipid interface to bulk water as determined by Wimley and White[36] (Figure 2B). This experimental system is conceptually similar to our simulations, since it measures the distribution of amino acids in a heterogeneous system containing a two-dimensional surface. The observed correlation is an indication for a rather hydrophilic nature of the gold surface in our simulation, in agreement with the low contact angle of water on gold.[37]

As expected, the interaction free energy is also correlated with the size of the amino acid (Figure 2C). The importance of sterical contact between gold and interacting biomolecules has been recognized before. In one of the early studies on this subject, a strong correlation between protein size and the protein's affinity to gold nanoparticles has been described, in accordance with our simulations here.[38]

(32) Michaelides, A.; Ranea, V. A.; de Andres, P. L.; King, D. A. *Phys. Rev. Lett.* **2003**, *90*, 216102.
(33) Iori, F.; Corni, S.; Di Felice, R. *J. Phys. Chem. C* **2008**, *112*, 13540–13545.
(34) Hoefling, M.; Iori, F.; Corni, S.; Gottschalk, K. E. *ChemPhysChem*, in press; DOI: 10.1002/cphc.200900990.

(35) Willett, R. L.; Baldwin, K. W.; West, K. W.; Pfeiffer, L. N. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7817–7822.
(36) White, S. H.; Wimley, W. C. *Annu. Rev. Biophys. Biomol. Struct.* **1999**, *28*, 319–365.
(37) Schrader, M. E. *J. Colloid Interface Sci.* **1984**, *100*, 372–380.
(38) Deroe, C.; Courtoy, P. J.; Baudhuin, P. *J. Histochem. Cytochem.* **1987**, *35*, 1191–1198.
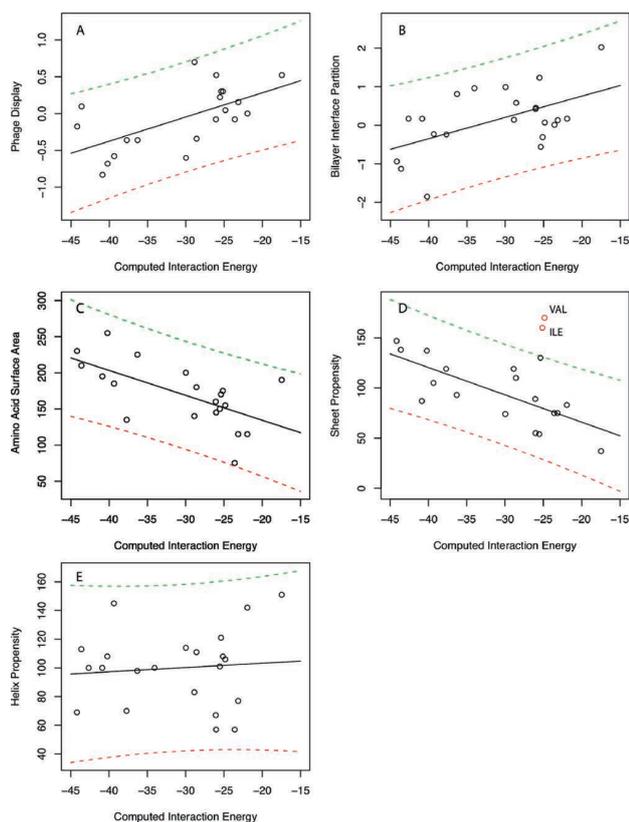
Figure 2. Correlation of computed energies with experimental data. Y-axis corresponds to (A) phage display (here, the log of the surface coverage as estimated from ref 9 is taken as *y*-axis); (B) bilayer partition; (C) amino acid surface area; (D) sheet propensity; (E) helix propensity. Solid line, linear fit; dashed lines, 95% prediction interval. Correlation coefficients are shown in Table 2. Please note that, in direct comparison of energies (A and B), a positive slope corresponds to correlation. Due to the sign in the energy, this is reversed in the comparison with propensities and size (C−E), and thus, a positive slope corresponds to anticorrelation.

Table 2. Correlation coefficients *R* corresponding to Figure 2[a]

	R	*p*
bilayer interface	0.57	0.01
surface area	0.63	< 0.01
phage display	0.62	< 0.01
helix propensity	0.08	0.69
sheet propensity	0.70	< 0.01

[a] *p* is the probability to obtain this correlation by random numbers.

Surprisingly, the interaction energies are rather strongly correlated with the propensity of a certain amino acid to form a β-sheet (when excluding Ile and Val), but not with the propensity to form a helix (Figure 2D and E). The gold surface favors conformations that allow contact of the backbone of the amino acid. These conformations are not available in helices, pointing at an unfolding propensity of gold surfaces for folded, helical proteins. In fact, amino acids having high β-sheet propensity have intrinsically preferred Φ,Ψ angles[39] that are ideally suited to interact with the surface (with β-sheet-like angles, the entire backbone including the capping can contact the surface). On the contrary, the other amino acids have to distort from their preferred conformation to

(39) Swindells, M. B.; MacArthur, M. W.; Thornton, J. M. *Nat. Struct. Biol.* **1995**, *2*, 596–603.
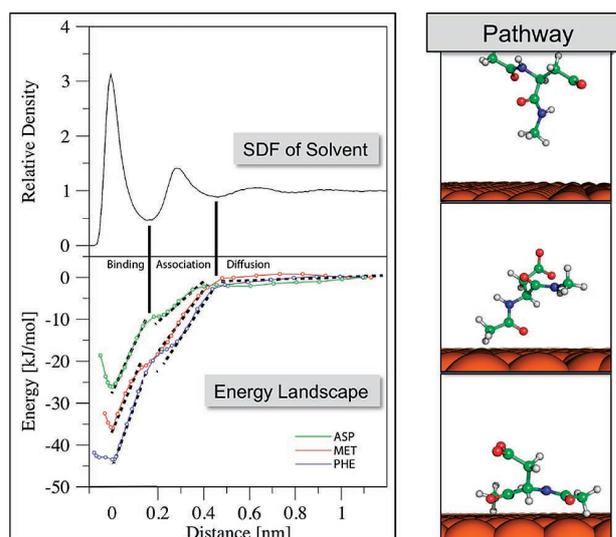
Figure 3. Changes in energy landscape during association of three exemplary amino acids correlate with low solvent density in the spatial distribution function (SDF). Different steps during association are depicted with snapshots from Asp MD simulations.

maximize the interaction with the surface, an effect that will cost energy. Therefore, our results are an indication that the gold surface may induce β-sheet-like conformations, a finding that points to a future design principle for gold-binding peptides. In line with our observation here, it has been suggested that other surfaces such as graphite induce β-sheet-like conformations.[40,41] This is structurally comprehensible, since β-sheet-like conformations allow maximizing the contact between the polypeptide and the surface.

Adsorption Pathways. The potential of mean force computed here provides detailed insight into the energy landscape of the association pathways. The association of the amino acids with the gold surface has a triphasic behavior (Figure 3). In an initial diffusive phase, the potential of the gold is hardly influencing the amino acid. At a distance of ∼5 Å from the bound state, the slope of the energy potential (equaling the force experienced by the amino acids) increases significantly, and the amino acid associates with the gold. In the final binding phase, the potential becomes even steeper.

As seen from an overlay of three representative PMFs onto the vertical distribution function of water over the gold surface, the phases correspond to the dewetting transitions. In the associative phase, the center of mass moves through the second water layer, and first contact is made with the gold surface. During the binding phase, the center of mass moves through the final water layer, and full contact with the surface is established. The constrained biased simulations also allow a structural insight into the different steps of association. As shown for the example of Asp, the amino acids predominantly touch the gold surface first with their capped backbone. However, in one of the four runs of Arg, first contact is made with the side chain. Due to the length of this side chain, this is one of the first contacts observed for all amino acids. This might indicate that Arg may be of importance for the adsorption of proteins on gold, since for proteins the backbone is in general not available for the initial association. In this context, experimental findings suggest Arg as a mediator for nanoparticle assembly.[42]

(40) Kowalewski, T.; Holtzman, D. M. *Proc. Natl. Acad. Sci. U.S.A.* **1999**, *96*, 3688–3693.
(41) Raffaini, G.; Ganazzoli, F. *Langmuir* **2004**, *20*, 3371–3378.
(42) Sethi, M.; Knecht, M. R. *ACS Appl. Mater. Interfaces* **2009**, *1*, 1270–1278.
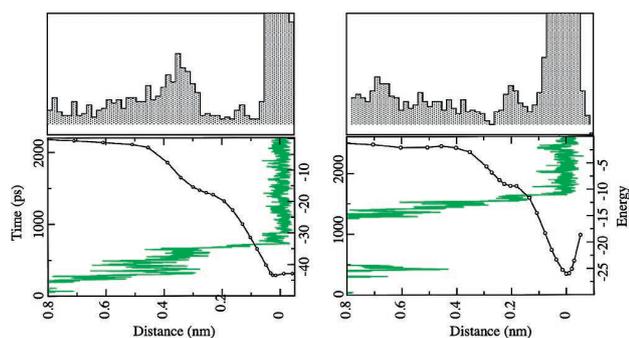
Figure 4. Energy landscape and free association. Left, Tyr; right, Asp. Bottom: Superposition of a free adsorption trajectory (green) with the energy landscape (black). Left axis, time of trajectory; right axis, energy of energy landscape; *x*-axis, distance of COM of amino acid from top layer of gold. Top: Spatial probability density of amino acid at a given distance during unconstrained simulation. The density was obtained from a distance histogram of the free simulation. A low probability is observed at steep areas of the potential.

The adsorption pathway is analogous to the pathways postulated for protein−protein interactions.[43] In both cases, something such as an encounter complex, where the two interacting partners first meet, is formed. The last stage of complexation in both cases is determined by the dewetting transition, allowing for maximal contact.

The constrained simulations use an additional harmonic potential acting on the amino acid that may restrict the conformational space. Therefore, we performed additional free, unconstrained simulations of the solvated amino acids positioned more than 1 nm away from the surface to reveal the unbiased adsorption pathway. All of the amino acids adsorbed within 5 ns of simulation. In the majority of trajectories, the initial contact is made by one of the backbone caps also in the unconstrained simulations, indicating that the constrained simulations are in equilibrium. The backbone caps are exposed, hydrophobic moieties, which tend to minimize their solvent exposed surface. Future studies of polypeptides, which do not require capping,

need to show whether also in a more proteinlike context backbone atoms form the initial contact. After initial contact, the equilibrium conformation is reached by further conformational rearrangements. The trajectories can be superimposed on the energy landscape previously obtained by the PMF simulations (Figure 4). The superimposition reveals that once the amino acid passed the last barrier, it adsorbs ballistically despite the overdamped environment in agreement with theoretical models.[44] Time reversal of molecular dynamics simulations does imply that the adsorption trajectory is also relevant for desorption. The barriers observed in the energy landscape are reflected in increased spatial probability density at the local minima. The agreement between the distance distribution of the amino acid relative to the gold surface in the free simulations and the free-energy landscape obtained from the biased simulations gives further confidence in the computed potentials of mean force.

Conclusions

In this Article, we described molecular dynamics simulations for the natural amino acids on the gold (111) surface. The obtained free energies of adsorption of the amino acids are dependent on their chemical character. They reasonably correlate with previous experimental findings in related systems. Additionally, we found that gold surfaces may induce *β*-sheet-like conformations, hinting at a future design principle for gold binding peptides. Simulations of individual amino acids provided are a crucial first step in describing the interactions of more complex biomolecular systems with inorganic surfaces. An important next step is the investigation of potential cooperative effects between amino acids[9,45] and of whole peptides and proteins. However, already this basic level revealed insight important for future rational design efforts of gold-binding polypeptides.

(43) Selzer, T.; Schreiber, G. *Proteins: Struct., Funct., Genet.* **2001**, *45*, 190–198.
(44) Bier, M.; Derenyi, I.; Kostur, M.; Astumian, R. D. *Phys. Rev. E* **1999**, *59*, 6422–6432.

(45) Serr, A.; Horinek, D.; Netz, R. R. *J. Am. Chem. Soc.* **2008**, *130*, 12408–12413.

Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines through a Hybrid Atomistic Simulation and Monte Carlo Approach

Martin Hoefling[1], Nicola Lima[1], Dominik Haenni[2], Claus A. M. Seidel[3], Benjamin Schuler[2], Helmut Grubmüller[1]*

1 Theoretical and Computational Biophysics Department, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, 2 Department of Biochemistry, University of Zurich, Zurich, Switzerland, 3 Institute of Molecular Physical Chemistry (MPC), Heinrich Heine University, Düsseldorf, Germany

Abstract

Förster Resonance Energy Transfer (FRET) experiments probe molecular distances via distance dependent energy transfer from an excited donor dye to an acceptor dye. Single molecule experiments not only probe average distances, but also distance distributions or even fluctuations, and thus provide a powerful tool to study biomolecular structure and dynamics. However, the measured energy transfer efficiency depends not only on the distance between the dyes, but also on their mutual orientation, which is typically inaccessible to experiments. Thus, assumptions on the orientation distributions and averages are usually made, limiting the accuracy of the distance distributions extracted from FRET experiments. Here, we demonstrate that by combining single molecule FRET experiments with the mutual dye orientation statistics obtained from Molecular Dynamics (MD) simulations, improved estimates of distances and distributions are obtained. From the simulated time-dependent mutual orientations, FRET efficiencies are calculated and the full statistics of individual photon absorption, energy transfer, and photon emission events is obtained from subsequent Monte Carlo (MC) simulations of the FRET kinetics. All recorded emission events are collected to bursts from which efficiency distributions are calculated in close resemblance to the actual FRET experiment, taking shot noise fully into account. Using polyproline chains with attached Alexa 488 and Alexa 594 dyes as a test system, we demonstrate the feasibility of this approach by direct comparison to experimental data. We identified *cis*-isomers and different static local environments as sources of the experimentally observed heterogeneity. Reconstructions of distance distributions from experimental data at different levels of theory demonstrate how the respective underlying assumptions and approximations affect the obtained accuracy. Our results show that dye fluctuations obtained from MD simulations, combined with MC single photon kinetics, provide a versatile tool to improve the accuracy of distance distributions that can be extracted from measured single molecule FRET efficiencies.

Introduction

Since the development of the Resonance Energy Transfer theory by Förster (FRET) in the late forties [1], and the definition of this technique as a "spectroscopic ruler" in biological systems by Stryer and Haugland [2], single molecule detection [3–5] and time-resolved experiments [6] have opened up a new window to probe inter- and intramolecular distances and motions. In a typical experiment, donor molecules are excited by a laser pulse, and part of the excitation energy is transferred to nearby acceptor molecules. The transfer efficiency

$$E = \frac{I_A}{I_D + I_A} \qquad (1)$$

is measured via the donor fluorescence intensity I_D and the acceptor fluorescence intensity I_A. Among other factors, E

depends on the distance R between the donor and the acceptor fluorophores, as well as on the mutual orientation of their respective transition dipole moments. After orientational averaging, the distance dependency is described by Förster's approximation,

$$E = \frac{1}{1 + \left(\dfrac{R}{R_0}\right)^6} \quad , \qquad (2)$$

where R_0 is the so-called Förster radius which denotes the distance at which 50% of the donor excitation is transferred to the acceptor molecule.

This relation is widely used to monitor structural changes in biomolecules via FRET efficiency measurements [2,7]. To that aim, donor and acceptor fluorophores are covalently attached to specific sites of the macromolecule of interest. Taking into account

the flexibility of the fluorophores and their linkers, the measured intensities provide information on the mutual distance of these specific sites [8–11]. The use of multiple dye pairs allows for triangulation of biomolecules, which provides three-dimensional structural information [10,12–16].

In single molecule setups, distributions and distance fluctuations of individual molecules are accessible [4,17–19]. If the scatter of the observed efficiency distributions in these experiments is broader than the expected shot noise, distance distributions can be estimated [20]. For distance changes in the biomolecule, which are slow compared to the burst duration, time resolved information is then accessible [21,22]. By recording millisecond fluorescence bursts while the molecules diffuses through a confocal laser volume, conformational motions in the same time scale have been resolved [21,23,24].

FRET spectroscopy has proven particularly successful in situations where the mutual orientation distribution of the transition dipole moments can be considered isotropic and uncorrelated. Examples are freely diffusing dyes, or dyes attached to flexible and solvent-exposed parts of a protein [18] or nucleic acids [10,11]. In this case, orientational averaging gives rise to the well-known orientation factor $\kappa^2 = 2/3$, which is by convention included within the Förster radius R_0 [7]. In contrast to this average κ^2, the instantaneous orientation factor $\kappa^2(t)$ can assume values in the range of 0 to 4.

Particularly when triangulating biomolecules, however, the dye motion is often far from isotropic due to steric restrictions set by the biomolecule, as well as due to electrostatic or hydrophobic interactions between the dye and the protein surface [25–30]. Since the mutual dye orientation is typically inaccessible to experiments, the $\kappa^2 = 2/3$ approximation provides only qualitative insights, unless the free and rapid reorientation of the dyes is commonly verified by fluorescence anisotropy measurements [31]. For this reason, efficiency distributions rather than distances are often reported.

The orientational dynamics uncertainty of fluorophores has been addressed via several routes. Empirical, semi-empirical, and theoretical models [32–36] for the orientational factor have been developed, assuming that the dynamics of the dyes can indeed be described by a time average. Recent computer simulations [37,38] have suggested that the mutual dye orientation can be highly anisotropic, with κ^2-values deviating markedly from 2/3 (0.24–1.02 [38]; 0.71–2.81 [37]). R_0 has been refined through fluorescence quenching measurements of multiple fluorophores [39].

Despite these efforts, three main problems remain. First, the assumption of an isotropic dye orientation distribution is invalid or difficult to establish in most cases [40,41]. Second, possible correlations between the distance and dye orientation distribution are neglected in the above treatments [38]. Third, the orientational sampling during individual bursts may be incomplete, in which case the dye distribution relevant for the observed efficiency depends on the duration of the bursts. In all three cases, applying an average κ^2 – as opposed to the κ^2 of instantaneous and time-dependent Förster transfer rate coefficients – leads to an additional broadening of the efficiency distribution [25], and biased distance distributions are obtained.

To overcome these limitations, we have developed an approach that combines molecular dynamics (MD) simulations of a dye-labeled biomolecule in solution with Monte Carlo (MC) simulations of dye excitation, FRET transfer, and fluorescence decay events. This approach involves four steps.

First, extended and fully atomistic MD simulations of the solvated biomolecule, labeled with a FRET dye pair, serve to cover the biomolecular dynamics at the fluorescence decay time scales of the system. To capture structural motions that are slower than the nanoseconds time scale accessible to MD simulation, several MD trajectories are recorded starting from different isomers and combined into a comprehensive ensemble using appropriate Boltzmann weights.

In the second step, time-dependent mutual dye orientations extracted from these trajectories are recorded. These orientations are then used to derive time-dependent instantaneous resonance energy transfer rate coefficients $k_T(t)$. Within a short time interval Δt, these rate coefficients specify the probability $p_T(t) = \Delta t \cdot k_T(t)$ that a FRET transfer event takes place, for each instant of time.

In the third step, using $p_T(t)$, a large number of MC runs is carried out to simulate and collect many individual photon absorption and excitation, FRET transfer, and emission events. For each photon absorption event, an instant of the trajectories is chosen randomly, and the probabilities are propagated appropriately until a photon emission or radiationless decay event occurs. After averaging over sufficiently many events, fluorescence intensities I_D and I_A are calculated. The numbers of photos recorded from the donor and the acceptor dyes, respectively, finally determine an average FRET efficiency value E. Similar approaches using dye conformations from simulations have been proposed recently [42–45].

To mimic single molecule FRET (smFRET) experiments, in a fourth step the emitted photons are collected into bursts according to the experimental photon burst size distribution (BSD). The efficiency in each burst is then calculated, and efficiency histograms are obtained, similar to single molecule experiments. By construction, this procedure takes shot noise accurately into account.

This hybrid simulation approach will enable one to calculate efficiency distributions that can be directly compared to measured efficiency distributions. Vice versa, we will develop a systematic approach to reconstruct distance distributions by combining the dye orientation and photon statistics at hand with measured efficiency distributions.

Here we apply this approach to a polyproline 15, 20, and 30-mer [46] with two FRET dyes (Alexa 488 and 594, Fig. 1) attached to both termini [2,31,45](Fig. 2A). As dye-labeled polyproline chains have been widely used as "rigid rods" to test the validity of the approximations underlying Förster's theory, and to gauge the Förster radius of several of FRET pairs in different environments [2,31,45], much of the current understanding relies on the particular properties of these systems. Initially assumed to be quite rigid, all-trans polyproline helices were used in the definition of FRET as a "spectroscopic ruler" [2]. This assumption was challenged quite early [47,48], suggesting that polyproline chains exhibit a substantial degree of flexibility [49]. The issue is still not fully resolved.

For these reasons, polyproline flexibility has been revisited recently by performing single molecule FRET recordings [31,50] and simulations [45] on these molecules. Indeed, unexpectedly broad efficiency distributions were seen, suggesting substantial structural heterogeneity. A detailed analysis of single molecule data showed the heterogeneity that persists on time scales greater than $10\,\mathrm{ns}$ [50]. Recent NMR experiments [45] pointed to a considerable population of cis-isomers within all-trans polyproline helices, which might contribute further to the structural flexibility and heterogeneity of polyprolines. These findings put the suitability of these molecules as "rigid rods" in question, and the unexpected complexity of their dynamics requires a detailed study of the structural ensemble in solution at room temperature.

Here we attempt a comprehensive characterization of the polyproline structural heterogeneity by combining atomistic
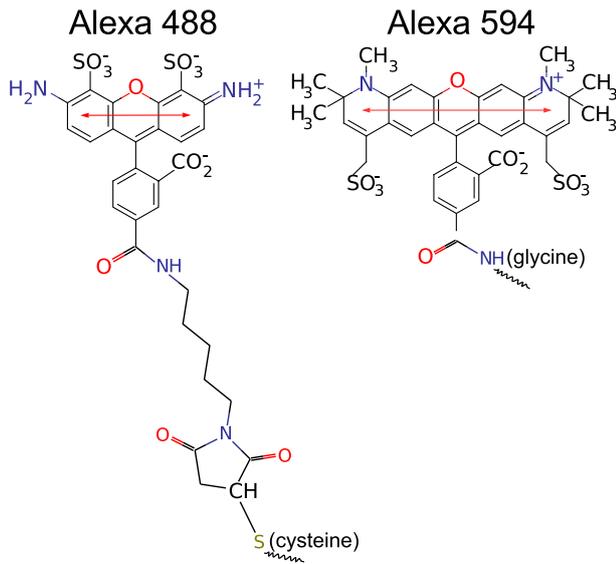
Figure 1. Dye and Linker Structures. Structure and transition dipole moments of Alexa 488 and Alexa 594. The red arrows show the orientation of the transition dipole moments. MarvinSketch was used to draw the chemical structures, Marvin 5.3.0.2 , 2010, ChemAxon (http://www.chemaxon.com).
doi:10.1371/journal.pone.0019791.g001

Figure 2. System Setup. (A) All-*trans* polyproline-20 molecular structure including Alexa 488 (green) and Alexa 594 (red) dyes attached by their corresponding linkers. The simulation box is shown in blue, terminal prolines used to restrain the position are depicted in black. (B) Fully solvated system is shown including Na^+ (blue) and Cl^- (yellow) ions.
doi:10.1371/journal.pone.0019791.g002

simulations with single molecule FRET data. Resting on a direct comparison of single burst efficiencies collected over many bursts, our approach is based on much fewer assumptions than the standard interpretation of FRET experiments. In particular, this approach includes κ^2 averages, on the basis of the detailed molecular dynamics of the system, and cases where the motion of the dyes is slower than the donor fluorescence decay time are readily handled. Moreover, all possible correlations between the dye movement and the distances are included, such that accurate mutual orientation distributions are obtained. Finally, the approach fully accounts for the photon count shot noise. Vice versa, comparison with experiments will enable us to test our approach. As we will demonstrate, our approach serves to combine dye orientational dynamics from MD with experimental FRET efficiency distributions at increasingly refined approximation levels.

The good agreement of distance distributions of polyproline obtained by this approach with the reference distribution suggests that this combination allows extraction of improved quantitative geometrical information from single molecule FRET experiments. By comparison with synthetic FRET data, the validity of the reconstruction will be established.

Methods

System Setup

The studied system comprises a polyproline peptide of 15, 20 or 30 proline residues [46], an amino-terminal glycine and a carboxyl-terminal cysteine residue, to which a succinimide ester and maleimide derivatives of Alexa 594 and Alexa 488 dyes [52](Fig. 1), respectively, are attached. Figure 2A shows the simulation system for the polyproline-20 [53] within a rectangular simulation box. Figure 2B depicts the box filled with explicit water molecules and 300 mM NaCl, corresponding to the ionic strength of 50 mM sodium phosphate buffer used in the experiment [31]. The number of Na^+ and Cl^- ions was chosen such as to obtain a neutral system.

In aqueous solution the most stable configuration for polyproline chains is the polyproline II (PPII) helix [53,54], characterized by dihedral angle values Φ, Ψ and Ω of $-75°$, $150°$, and $180°$, respectively [47], with the *trans*-isomer as the most favorable configuration. Nevertheless, in water a marked fraction of *cis* peptide bonds the PPII helices is observed. By NMR experiments a fraction of approximately 10% for proline at the C-terminus of the chain and 2% within the chain was measured [45], with *trans* to *cis* transition times of 10^3 to 10^4 seconds [50,55]. As this is far beyond MD time scales, separate simulations were performed for all relevant isomers, for subsequent weighted averaging. To this end, all possible isomers containing one single *cis* peptide bond were considered, i.e., 20 *cis*-trajectories for the polyproline-20 with dyes attached. Additionally, for polyproline-30, a subset of 61 isomers with *cis*-bonds at two positions was simulated.

Force Field

For water molecules, the TIP4P model was employed [56]. Force field parameters for the peptide were taken from a modified OPLS-AA force field [57] including custom parameters for the two dyes and their corresponding linkers. Alexa 488 and Alexa 594 are highly conjugated systems whose parameters are not included within the standard OPLS-AA force field. Figure 1 depicts the atomic structure of the two dyes together with the orientation of the transition dipole moments. All dye parameters (bonded and Lennard-Jones) – except for the partial charges –

were assigned via an analogy approach from similar OPLS-AA groups [58].

Because FRET occurs when the donor dye is in the excited state and the acceptor in the ground state, partial charges of these corresponding states were used in all our simulations for the dyes. The fact that the partial charges calculated for the ground and excited states differed only by a small amount suggests that the effect of this simplified treatment on the dynamics of the dyes is small. All partial charges were calculated by fitting to the electrostatic potential surfaces (EPS approach [59]) obtained from *ab-initio* B3LYP Density Functional Theory (DFT) calculations with the 6-31G* basis set. All *ab-initio* calculations were performed with the GAUSSIAN 03 program package [60]. First, for reference, the point charges for the 20 natural amino acids were calculated with B3LYP/6-31G* CHelpG population analysis to assure compatibility of the derived charges with OPLS-AA. A mean scaling factor of 0.9 was calculated by averaging the multiplicative factors of each amino-acid, which minimizes the mean square deviation between OPLS-AA and DFT charges (amino-acid scaling factors shown in Suppl. Table S1).

For the ground state of the two dyes, the same protocol was used. For the excited state, we determined charge differences with respect to the ground state for each atom in two steps. First, point charges were determined from Configuration Interaction Singlets (CIS) calculations for the first excited state using the STO-3G basis set. From these values, in a second step, point charges were subtracted, that were obtained from Hartree Fock (HF) calculations with the same STO-3G basis.

For both, ground and excited state, the charges were averaged to reflect the internal symmetry of the molecule, and scaled with the previously calculated scaling factor of 0.9. Finally, a small offset was added to all partial charges to re-establish the correct total charge of the system.

Molecular Dynamics Simulations

All MD simulations were carried out with the GROMACS 4.0.7 simulation software package [61–63]. Each proline system was energy-minimized by steepest descent to convergence. Periodic boundary conditions were applied in all three dimensions. V-Sites on hydrogens [64] were used allowing 4 fs integration time steps. After minimization, 10 ns equilibration simulations were performed. From the last 5 ns of these simulations, starting conformations for all subsequent production runs were selected at random instances (Table 1). Solvent and ions as well as the solute were separately coupled to an external temperature bath with a time constant of 0.1 ps applying the v-rescale algorithm [65,66]. The system was coupled to an isotropic pressure bath of 1 atm using the Parinello-Rahman algorithm [67] and a time constant of 1 ps. Bond lengths were constrained to their equilibrium lengths with LINCS [68]. The cut-off for Lennard-Jones interactions was set to 1 nm. Electrostatic interactions between charged groups at distances below 1 nm were calculated in direct space, while for the long-range interactions the particle-mesh-Ewald method [69] with a grid spacing of 0.12 nm and fourth order spline interpolation was used. All simulations were performed with random Maxwell-distributed starting velocities at 293 K, 303 K, and 313 K (Table 1).

Soft restraints were imposed to suppress rotation of the entire molecule in the box and thus to allow the use of a small simulation box, adapted to the shape of the molecule. To this end, the component of the difference vector perpendicular to the x-axis (Fig. 2A) between the centers of mass of the two terminal prolines was restrained to zero with a weak harmonic potential ($k = 9.744 \, \text{kJ mol}^{-1} \, \text{nm}^{-2}$, corresponding to a Boltzmann distri-

**Table 1.** Performed molecular dynamics simulations.

Proline Length	Isomer	Temperature	Number of simulations
		K	
pro15	all-*trans*	293	10
pro15	single-*cis*	293	30
pro20	all-*trans*	293	20
pro20	all-*trans*	303	10
pro20	all-*trans*	313	10
pro20	single-*cis*	293	40
pro30	all-*trans*	293	10
pro30	single-*cis*	293	30
pro30	double-*cis*	293	61

Simulations are listed according to isomer and applied temperature. Single-*cis* simulation were carried out for all possible *cis*-isomer positions. For polyproline-30, in addition, a representative set of 61 isomers, randomly picked from the 870 possible isomers with two *cis* bonds, was simulated. All simulation lengths are 100 ns summing up to a total sampling of 22.1 μs.
doi:10.1371/journal.pone.0019791.t001

bution of width $\sigma = 0.5 \, \text{nm}$). We assume that these soft restraints leave the internal dynamics of the molecule unperturbed.

Resonance Energy Transfer Rates

All FRET efficiencies were calculated from the MD simulations using following kinetics,

$$
\begin{array}{ccc}
& \text{D+A} & \text{D+A} \\
& \uparrow k_{\text{D}_i} & \uparrow k_{\text{A}_i} \\
\text{D+A+}hv \;\rightarrow\; & \text{D}^*\text{+A} \;\overset{k_{\text{T}(t)}}{\rightarrow}\; & \text{D+A}^* \\
& \downarrow k_{\text{D}} & \downarrow k_{\text{A}} \\
& \text{D+A+}hv_{\text{D}} & \text{D+A+}hv_A
\end{array}
\tag{3}
$$
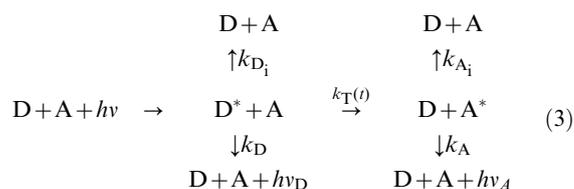
Starting after a photon adsorption event by the donor dye, this kinetics is described by

$$
\dot{p}(\text{D}^*\text{+A}) = -\left(k_{\text{D}_i} + k_{\text{D}} + k_{\text{T}}(t)\right) \cdot p(\text{D}^*\text{+A}) \text{ and} \tag{4}
$$

$$
\dot{p}(\text{D+A}^*) = k_{\text{T}}(t) \cdot p(\text{D}^*\text{+A}) - \left(k_{\text{A}_i} + k_{\text{A}}\right) \cdot p(\text{D+A}^*). \tag{5}
$$

In Eq. 3, D/D^* is the donor (Alexa 488) and A/A^* is the acceptor (Alexa 594) dye in their ground and the excited state, respectively. hv, hv_{D} and hv_{A} denote the exciting photon and photons emitted by the donor and the acceptor dye. The rate coefficients refer to FRET (k_T), fluorescence and internal conversion of the donor ($k_{\text{D}}, k_{\text{D}_i}$), and fluorescence as well as internal conversion of the acceptor dye ($k_{\text{A}}, k_{\text{A}_i}$).

The rate coefficients were calculated from the lifetimes τ of the dyes and their respective quantum yields Q,

$$
k_{\text{D}} = \frac{Q_{\text{D}}}{\tau_{\text{D}}}, \; k_{\text{A}} = \frac{Q_{\text{A}}}{\tau_{\text{A}}}, \tag{6}
$$

$$k_{D_i} = \frac{1 - Q_D}{\tau_D}, \, k_{A_i} = \frac{1 - Q_A}{\tau_A}. \tag{7}$$

For the Alexa 488 and 594 dyes attached to polyproline peptides, we used the measured lifetimes τ_D of $4.0\,\mathrm{ns}$ and $\tau_A = 3.9\,\mathrm{ns}$. To obtain photon statistics directly comparable to the experiment, the quantum yields were combined with the detector efficiencies into (relative) effective quantum yields using the correction matrix defined in Ref. [70]. In this framework, Q_A and Q_D correspond to the diagonal correction matrix elements. For the simulations, we averaged the two detector channels used in the experiment, yielding 0.77 and 1.0 for donor and acceptor effective quantum yields, respectively. Crosstalk, direct acceptor excitation, and background were found to change the photon statistics only by a small amount and thus are neglected in our MC approach.

For the time-dependent FRET rate coefficient $k_T(t)$, which depends on the electronic coupling between the two dyes and thus also on their mutual orientation at each instant, we used Förster's dipole approximation for the electronic coupling,

$$k_T = \left(k_D + k_{D_i} \right) \left(\frac{R_0}{R} \right)^6. \tag{8}$$

In Eq. 8, R is the distance between the geometric center of the ring system of the acceptor and the donor dyes, and R_0 is the Förster radius (the distance of 50% excitation transfer), which is proportional to the time-dependent orientation factor κ^2,

$$R_0^6 = \frac{9(\ln 10) Q_D J \kappa^2}{128 \pi^5 n^4 N_A} = R_{\mathrm{const}}^6 \cdot \kappa^2, \tag{9}$$

where Q_D is the quantum yield of the donor in the absence of the acceptor, J the spectral overlap integral (Franck Condon factor), N_A Avogadro's number, n the index of refraction of the solvent, and κ^2 is the time-averaged orientation factor [3,70,71]. For the pair Alexa 488 – Alexa 594, a Förster radius R_0 of $5.4\,\mathrm{nm}$ has been determined [7,72], based on the assumption of isotropic dye orientations i.e., $\kappa^2 = 2/3$. To describe time-dependent Förster transfer, R_0^6 in Eq. 8 is therefore replaced by $R_{\mathrm{const}}^6 \kappa^2(t)$, with

$$R_{\mathrm{const}} = 5.4\,\mathrm{nm} / \sqrt[6]{\frac{2}{3}} = 5.78\,\mathrm{nm}.$$

The orientation factor

$$\kappa^2(t) = [\cos\theta_{DA}(t) - 3\cos\theta_D(t)\cos\theta_A(t)]^2 \tag{10}$$

depends on the three relevant angles defined in Fig. 3. The transition dipole moment orientations within the molecular frame of the dyes were chosen parallel to the ring system plane, and connecting the terminal rings of each dye (Fig. 1) [73].

Using the above framework, for all MD trajectories orientation factors $\kappa^2(t)$ and distances $R(t)$ were calculated and stored for each time step, thus obtaining time-dependent FRET rate coefficients $k_T(t)$, which will be used below. Supplementary Video S1 shows distance, orientation factor and transfer efficiency for an exemplary trajectory.

We note that for small inter-dye distances ($<2\,\mathrm{nm}$), when terms of higher order than the dipolar are not negligible, Eq. 8 can be replaced by multipole expansion of the coupling potential or the transition density cube method [42,74] in a straightforward
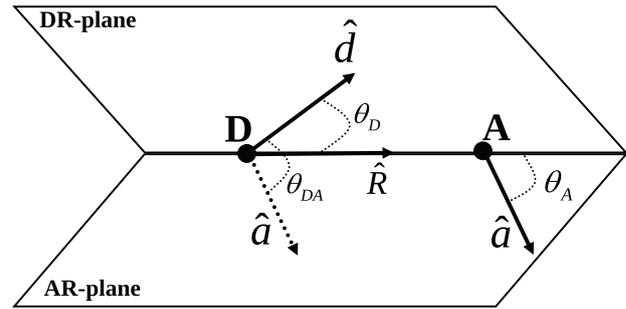

Figure 3. Geometry of dye orientations. Three angles define the orientation factor κ^2, the angle θ_{DA} between $\hat{d}$ and $\hat{a}$, and the angles θ_D and θ_A between $\hat{d}$ and $\hat{a}$, respectively, and $\hat{R}$. The DR and DA plane are defined by $\hat{R}$ and $\hat{d}$ as well as $\hat{R}$ and $\hat{a}$.
doi:10.1371/journal.pone.0019791.g003

manner, such that accurate FRET rate coefficients are also obtained in these cases. In the present work, the dipolar coupling potential was used.

Single Photon Generation

For direct comparison with smFRET burst counts, we developed a Monte Carlo (MC) procedure to calculate single burst FRET efficiencies from $k_T(t)$. In the experiments, the arrival times of individual photons from single molecules were recorded. Accordingly, and following the kinetics scheme Eq. 3, multiple individual photons were generated in a Monte Carlo process (Fig. 4). For each photon, we proceeded as follows.
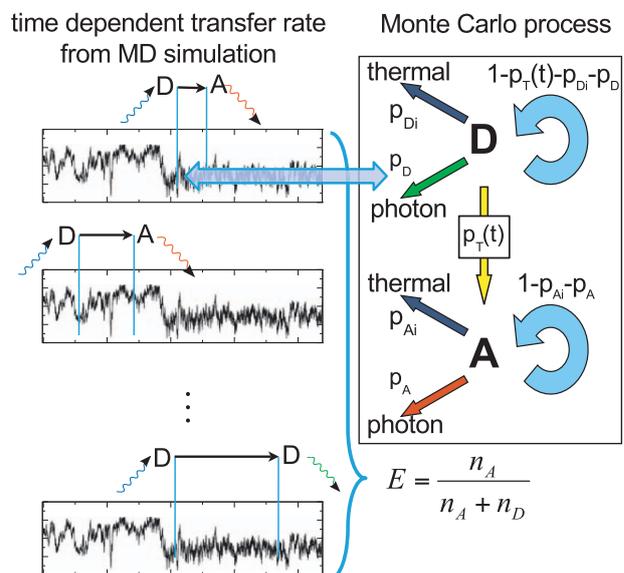

Figure 4. Photon generation by Monte Carlo. FRET transfer rate coefficient vs. time, calculated from a molecular dynamics simulation (box, left part). A random starting point on the trajectory is chosen at which the donor dye is assumed to be excited by a photon (blue). Then, for each time step the MC process on the right side is evaluated according to the corresponding probabilities until de-excitation occurs. Four de-excitation pathways are considered, thermal de-excitation of donor or acceptor (dark-blue) and respective photon emissions (donor: green; acceptor: red). The ratio of the collected donor and acceptor photons is used to calculate a FRET transfer efficiency.
doi:10.1371/journal.pone.0019791.g004

First, a random donor excitation instance was chosen from a randomly chosen trajectory (Fig. 4 left). Next, the Markov scheme in Fig. 4 (right) was iterated in time steps of Δt until either photon emission or radiationless decay occurred (see Suppl. Video S2). In the latter case, the MC run was discarded, in the former, the photon (donor or acceptor) was recorded. During each MC cycle and using an integration time step $\Delta t = 1\,\mathrm{ps}$, transitions were randomly selected according to probabilities $p_{\mathrm{Di}} = k_{\mathrm{Di}} \cdot \Delta t$ for thermal de-excitation, $p_{\mathrm{D}} = k_{\mathrm{D}} \cdot \Delta t$ for donor photon emission, $p_{\mathrm{T}} = k_{\mathrm{T}}(t) \cdot \Delta t$ for FRET transfer and $1 - p_{\mathrm{Di}} - p_{\mathrm{D}} - p_{\mathrm{T}}(t)$ for no state change. Acceptor de-excitation probabilities were calculated in the same way, but with consistent transition probabilities p_{Ai} and p_{A}, which allowed to skip the remaining Monte Carlo step and to record the emitted photon right away. All random numbers were generated with an SIMD-optimized Mersenne Twister algorithm [75,76].

In the experiment, no FRET is seen for dyes in or close to van-der-Waals contact, presumably due to quenching by electron transfer [77]. The effect of quenching at low inter-dye distances is not described with Förster theory, and therefore also not in our MC process. To correct for this, photons are rejected if the inter-dye distance is below $1\,\mathrm{nm}$ during the photon generation when comparing to experiments.

FRET Efficiency Calculation

Averaged over many MC runs, the collected de-excitation events $n_{\mathrm{Atot}} = n_{\mathrm{Ai}} + n_{\mathrm{A}}$ and $n_{\mathrm{Dtot}} = n_{\mathrm{D}} + n_{\mathrm{Di}}$ from donor and acceptor, respectively, were used to determine the average efficiency

$$E = \frac{n_{\mathrm{Atot}}}{n_{\mathrm{Atot}} + n_{\mathrm{Dtot}}}. \tag{11}$$

In experiments, only radiative de-excitation events $(n_{\mathrm{A}}, n_{\mathrm{D}})$ can be recorded. We therefore followed the same way in reconstructing the total number of de-excitation events using the respective fluorescence quantum yields,

$$n_{\mathrm{Atot}} = \frac{n_{\mathrm{A}}}{Q_{\mathrm{A}}}, \tag{12}$$

and analogously for n_{Dtot}.

To directly relate efficiency distributions from MC sampling to single molecule FRET measurements, the effect of shot noise and burst size distribution has to be taken into account properly [31,78]. Here, a sufficiently large number (>50000) of bursts has been measured, which provided sufficient statistics such that the experimental burst size distribution was used for combining the MC generated photons into bursts. After correction for quantum yield and detector efficiency, for each burst a single FRET efficiency value was calculated using Eq. 11. Collecting FRET efficiencies from many bursts yielded efficiency distributions that can be directly compared to the measured ones. As in the experiment, only bursts larger than 100 photons, after correction for the effective quantum yield, were used.

Inclusion of *cis/trans* isomer heterogeneity

So far, we have considered only one isomeric state of the proline polymer, e.g., the all-*trans* state. As has been found by NMR, however, each peptide bond undergoes isomerizations, with a small but non-negligible population in the *cis*-isomer, and with a larger *cis*-population for the terminal peptide bond at the C-terminus [45]. Because the isomerization times of minutes to hours are much longer than all other relevant time scales, we considered

a weighted ensemble of all possible relevant isomerization states and performed the above MD and MC simulations with efficiency calculations separately for each isomer. Subsequently, employing $p_{\mathrm{ter}}^{\mathrm{cis}}$ and $p_{\mathrm{int}}^{\mathrm{cis}}$ from NMR experiments [45] as probabilities for the occurrence of *cis*-isomers for C-terminal and internal peptide bonds, receptively, a weighted average was obtained (Table 2).

Single-Molecule Experiments

Peptide samples were prepared as described previously [31]. Single-molecule fluorescence experiments were performed with a MicroTime 200 confocal microscope (PicoQuant, Berlin, Germany) equipped with a pulsed $485\,\mathrm{nm}$ diode laser (LDH-P-C-485B, PicoQuant) and an Olympus UplanApo 60 x/1.20 W objective. After passing through a $100\,\mathrm{mm}$ pinhole, sample fluorescence was separated by a polarizing beam splitter cube into components parallel and perpendicularly polarized with respect to the excitation light. Subsequently, both components were further divided into donor and acceptor photons by means of dichroic mirrors (585DCXR, Chroma), filtered (donor emission filters: Chroma ET525/50 M, acceptor emission filters: Chroma HQ650/100), focused on avalanche photodiodes (PerkinElmer Optoelectronics SPCM-AQR-15), and the arrival times of all detected photons were recorded using suitable counting electronics (Hydra Harp, PicoQuant, Berlin, Germany).

Results and Discussion

Time-dependent conformations of the two dyes and their mutual orientations for the three polyproline systems considered here (Fig. 2) were obtained from multiple $100\,\mathrm{ns}$ MD trajectories of the all-*trans* and *cis*-isomers. MD simulations totaling $22.1\,\mathrm{ms}$ were carried out for the different isomers, chain lengths, and temperatures (Table 1). We first focus on polyproline-20 in the all-*trans* isomer as the most stable configuration in water and analyzed two main factors relevant for the FRET efficiencies, the distance R between the two fluorophores and the orientation factor κ^2.

Distance Distributions

Fig. 5A shows the distributions of dye-to-dye distances (defined by the geometric center of the ring system) from individual $100\,\mathrm{ns}$ simulations. The fact that the distributions differ from each other shows that the individual simulations are not fully converged to represent the full all-*trans* ensemble. To improve convergence, multiple simulations were combined. The apparent differences between the individual distance distributions are mainly due to

Table 2. Isomer weights.

Isomer	Probability
all-*trans*	$p^{\mathrm{trans}} = \left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-1}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
#1 cis, other trans	$p^{\mathrm{cis}_1} = p_{\mathrm{int}}^{\mathrm{cis}}\left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-2}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
#2 cis, other trans	$p^{\mathrm{cis}_2} = p_{\mathrm{int}}^{\mathrm{cis}}\left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-2}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
⋮	⋮
#n-1 cis, other trans	$p^{\mathrm{cis}_{n-1}} = p_{\mathrm{int}}^{\mathrm{cis}}\left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-2}\left(1 - p_{\mathrm{ter}}^{\mathrm{cis}}\right)$
#n cis, other trans	$p^{\mathrm{cis}_n} = \left(1 - p_{\mathrm{int}}^{\mathrm{cis}}\right)^{n-1} p_{\mathrm{ter}}^{\mathrm{cis}}$

Here, cis₁,cis₂,…,cisₙ indicate the position 1,2,…,*n* of the *cis* peptide bond in the chain, starting from the amino terminus.
doi:10.1371/journal.pone.0019791.t002
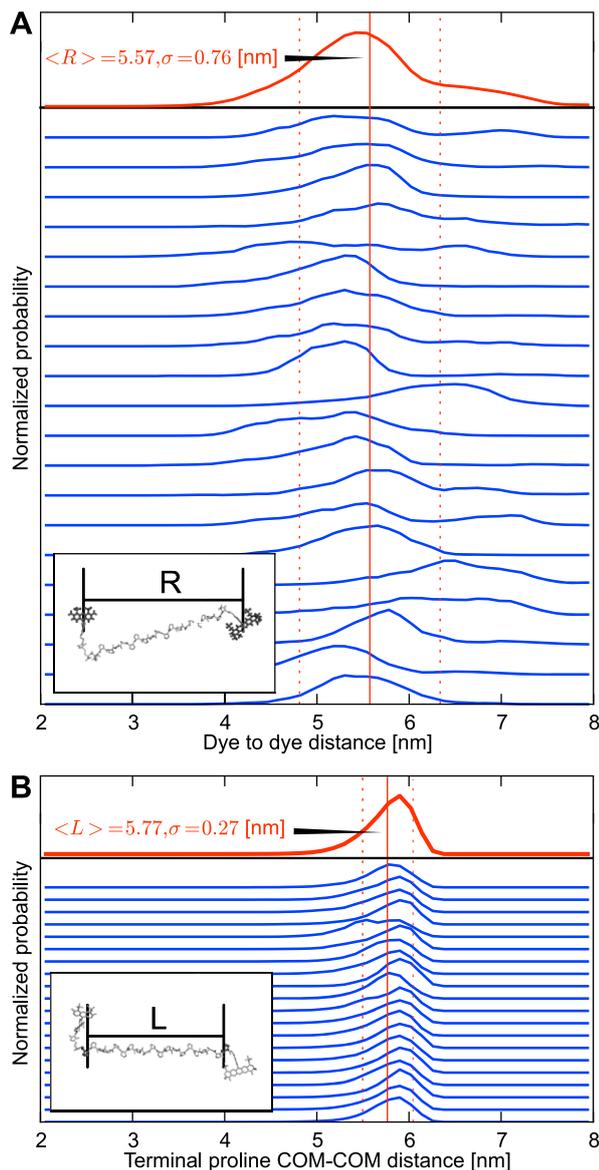
Figure 5. Distance distributions. (A) Histograms of the distances between the geometric centers of the ring systems of the two dyes for 20 all-*trans* MD simulations. (B) The distance histogram between the COM of terminal prolines from the polyproline-20 chain, for the same simulations. The insets visualize the measured distance in each plot. Respective averages are shown in red; vertical lines denote the mean and standard deviation.
doi:10.1371/journal.pone.0019791.g005

slow transitions between subpopulations of dye-conformations (for more details see section "Preferred Dye Conformations").

To better characterize the subpopulations and how they differ between the individual trajectories, the distance between two terminal proline residues center of mass (COM) was analyzed. As shown in Fig. 5B, these length fluctuations are much smaller compared to the dye-to-dye distances. In addition, the mean length of individual simulations shows only small variations.

These small length fluctuations point to considerable rigidity of the polyproline peptide, which indeed originally motivated its use as a molecular ruler. From the angular fluctuation θ of selected segment pairs, separated by length L, a persistence length

$P = 18.3 \pm 0.3$ nm was obtained via

$$P = -\frac{L}{\ln \langle \cos \theta \rangle}. \qquad (13)$$

Here, 3 proline residues ($\equiv 1$ PPII helix turn) defined a segment and its tangent with a segment length of 0.93 nm. The all-*trans* chains are indeed quite rigid and do not strongly deviate from the type II helix structure model.

Because of the stiffness of the polyproline, the observed broader distribution between the dyes mainly originate from the flexible dye linkers rather than from the flexibility of the polyproline chain.

Orientational Dye Dynamics and Orientation Factor κ^2

Figure 6 shows the κ^2 distributions derived from 20 all-*trans* simulations (gray) as well as their average (red). For comparison, an isotropic κ^2 distribution is shown (black). As shown, the individual simulations scatter considerably, with respective mean κ^2 values between $0.58 - 1.06$. Averaging over all 20 simulations, the mean κ^2 of all-*trans* simulations was 0.83 ± 0.03, and 0.80 ± 0.02 for the complete ensemble including all *cis*-isomers (Fig. 6). Both values agree within statistical error and significantly deviate from the isotropic κ^2 value of $2/3$.

As seen from the κ^2 histograms of individual MD simulations, the sampled dye geometries differ for each simulation, which underscores the importance of averaging multiple simulations. The obtained more realistic κ^2-value shifts the effective Förster radius from 5.4 nm to 5.6 nm.

Next, we determined the correlation between R and κ^2 for the 20 all-*trans* simulations and found a mean Pearson correlation coefficient of -0.13 ± 0.02. Because R and κ^2 are assumed to be uncorrelated in Försters RET theory, this finding suggests that using a distance-dependent $\langle \kappa^2 \rangle (R)$ might further improve the distance reconstruction, as will be discussed below.

Table 3 shows mean auto-correlation times of different variables from the simulations (exemplary autocorrelation plot shown in
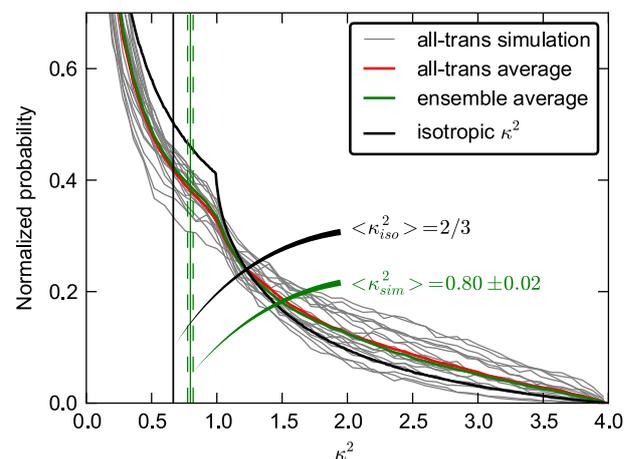

Figure 6. Distributions of the orientation factor κ^2. Each gray line shows to an orientation factor histogram from one of the 20 all-*trans* simulations at 293 K, with the average shown in red. The green curve (full ensemble) additionally includes the *cis*-isomers with appropriate weights, the green vertical line shows the corresponding average and its statistical error (dashed). The black curve shows the κ^2 histogram for an isotropic dye orientation distribution, with the well known mean value of $2/3$ (vertical black line).
doi:10.1371/journal.pone.0019791.g006

Table 3. Time scales of motions.

	mean	SEM	min	max
	[ns]	[ns]	[ns]	[ns]
R (dye-to-dye)	2.96	0.52	0.71	8.68
κ^2 (orientation factor)	0.34	0.04	0.15	0.86
L (terminal prolines)	0.80	0.20	0.30	4.12
Ω (terminal orientation)	4.96	0.86	1.06	14.55
Anisotropy decay (Alexa 488)	0.90	0.08	0.42	1.66

Autocorrelation times of all-*trans* polyproline-20 with their respective standard error of the mean (SEM), minimum and maximum. Terminal orientation Ω denotes the autocorrelation times of the cosine of the angle between the terminal proline tangent vectors.
doi:10.1371/journal.pone.0019791.t003

Suppl. Fig. S1). The orientation factor κ^2 shows the fastest decay ($0.3\,\text{ns}$), whereas the terminal orientation and the dye to dye distance are in the ns regime (Table 3) and thus comparable to the donor fluorescence decay times. Calculated fluorescence anisotropy decay timescales [34,35] of $0.9\,\text{ns}$ in our simulations agree with experimentally measured decay times of $0.3-0.8\,\text{ns}$ [31] within the accuracy of the simulation [34] and thus indicate a correct modeling of the dye dynamics by our force field.

These autocorrelation times determine the correlation of the dye conformations and distances as probed by successive photons and, therefore also, how many structures probed by each burst are effectively statistically independent. Further, this autocorrelation time may determines the size of the sub-ensemble of conformations that is actually probed by FRET, because the fluorescence intensities of the two dyes also depend on past transfer efficiencies. We will therefore examine the influence of these effects on the quality that can be achieved for the distance reconstruction described further below.

Preferred Dye Conformations

What is the structural origin of the orientation factor κ^2 deviation from its isotropic value of $2/3$? A closer inspection of the MD simulations revealed that hydrophobic interactions of the dye linker with the proline chain enhanced the population of certain conformational sub-states, similar to previous reports[45]. This effect is more pronounced for Alexa 488 due to the longer linker. For Alexa 488, two distinct conformation sub-states (*open* and *closed*) were seen (Fig. 7).

To test the stability of these confomer ensembles, we analyzed distances and the orientation factors of the all-*trans* polyproline-20 system at elevated temperatures (303 and $313\,\text{K}$, Table 1). No significant impact on the values for κ^2 was found ($293\,\text{K}$: $0.82 \pm 0.03, 303\,\text{K} : 0.83 \pm 0.05, 313\,\text{K} : 0.86 \pm 0.04$). Also the dye-to-dye distance R showed no systematic trend towards open or closed conformations ($293\,\text{K} : 5.57 \pm 0.08\,\text{nm}, 303\,\text{K} : 5.22 \pm 0.08\,\text{nm}, 313\,\text{K} : 5.37 \pm 0.06\,\text{nm}$). For the polyproline-20 chain length L ($293\,\text{K} : 5.77 \pm 0.01\,\text{nm}, 303\,\text{K} : 5.73 \pm 0.01\,\text{nm}, 313\,\text{K} : 5.70 \pm 0.02\,\text{nm}$), a small decrease with increasing temperature was seen. In summary, the applied temperature changes neither seem to significantly influence the population ratios of the two conformations, nor the relative dye-to-dye orientations. However, due to the limited sampling, we cannot fully exclude small effects, which may arise at larger temperature changes. It will be an interesting
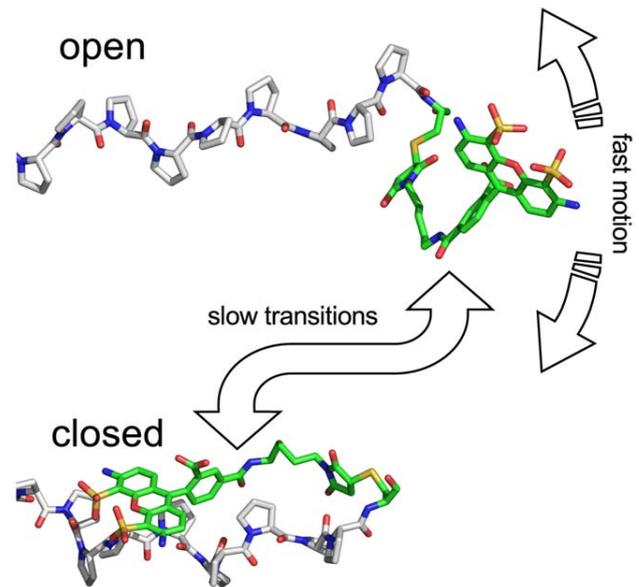

Figure 7. Conformational heterogeneity of Alexa 488. Several conformations of the Alexa 488 dye and its linker attached to the proline chain during MD simulations are seen in the simulations. For the *open* conformation, fast large amplitude motions are seen for the dye whereas hydrophobic interactions restrict the dye mobility in the *closed* conformations (one representative example is shown). Additionally slow transition between the open and closed conformations are seen.
doi:10.1371/journal.pone.0019791.g007

challenge for future experimental work to directly identify the presence of the dye conformations observed here, e.g. from a broadening of fluorescence anisotropy distributions in single molecule experiments [79], or from the effect of measurements under conditions that increase the solubility of the fluorophores on the transfer efficiency histograms.

Efficiency Distributions from Individual Simulations

Figure 8 shows FRET efficiencies calculated separately from all 20 all-*trans* MD simulations. As already expected from the dye-to-dye distance distributions, also the mean FRET efficiencies cover a broad range from 0.27 to 0.66 with $\sigma = 0.037$ to 0.043. These standard deviations σ were compared to the expected shot noise $\sigma^2 = E(1-E)/n_{\text{tot}_\text{L}}$ [80,81] for each simulation mean efficiency E using the lower experimental BSD limit ($n_{\text{tot}_\text{L}} = 100$) resulting in a width $\sigma = 0.033$ to 0.036. Thus, the efficiency peak observed in the individual traces of our simulations is mainly broadened due to the photon shot noise.

Comparison of the distance distributions (Fig. 5A) with the efficiencies (Fig. 8) illustrates the effect of signal averaging over an entire fluorescence burst, subsequently referred to as "burst averaging". To see this, consider naive transformation from distances to efficiencies using Eq. 2, which would result in much broader efficiency distributions than those observed in Fig. 8. This narrowing is due to the combination of multiple photons, and thus also of distances, into one burst, such that each efficiency value represents a corresponding average [11]. It is this averaging, which markedly narrows obtained efficiency distributions and also obscures much of the structure seen in the distance distribution.

Isomeric Heterogeneity

To account for the isomeric heterogeneity due to the presence of *cis*-isomers, which reduce the average distance between the two
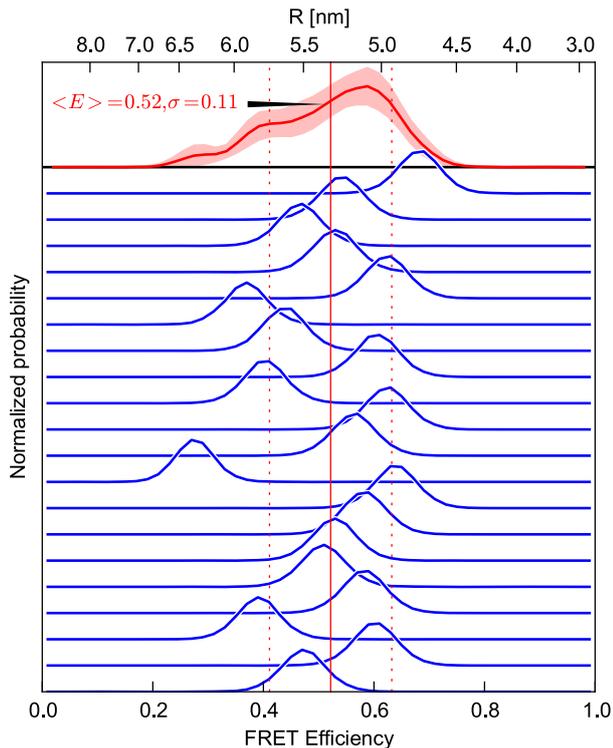
Figure 8. Spread of the efficiency amongst individual simulations. Transfer efficiency histograms (blue) obtained via MC sampling from 20 all-*trans* MD simulations of polyproline-20 at 293 K. The red curve at the top depicts the efficiency combined from all 20 trajectories, where each burst is still combined from photons of one trajectory; the bootstrapping standard error, calculated from 100 random samples, is indicated by the shaded area. The vertical lines indicate the mean efficiency and its standard deviation.
doi:10.1371/journal.pone.0019791.g008

dyes [45], additional MD simulations for all possible isomers were performed (Table 1). Using the population estimate of Table 2, the full ensemble includes 5%, 8%, and 15% of isomers with more than one *cis* bond for polyprolines of length 15, 20, and 30, respectively. Thus, for proline 15 and 20, we included only the single-*cis* conformers within the ensemble. For polyproline-30, estimating the impact of multi-*cis* isomers, additionally a subset of double-*cis* isomers was considered (Table 1). In the isomer simulations, all the other bonds were kept in the *trans* configuration, and the same MD parameters and protocol as for the all-*trans* isomer were used. FRET efficiencies were then calculated as explained before.

Figure 9 shows FRET efficiency distributions and averages for the all-*trans* and *cis* polyproline-20 chains in comparison with experiment. As expected, the average efficiencies of the *cis*-chains are larger than that of the all-*trans* isomer, due to the reduced distance of the terminal prolines. The largest reduction is seen for *cis*-bonds in central positions, thus attributing measured high efficiencies to those isomers. This behavior can be captured in a simple model (Fig. 9, top), in which the *cis*-isomer is described by a kink angle α between the two stiff parts of the molecule, with distances R_A and R_D between the *cis*-bond and the respective termini, and $R = R_A + R_D$ being the all-*trans* distance between the two termini. R, was determined from the all-*trans* mean efficiency using Eq. 2 and split up on R_A and R_D for each *cis* isomer according to the *cis*-bond position. To account for the distance

change due to the linker and the observed dye conformations (Fig. 7), an offset $R_D + x_0$ and $R_A - x_0$ was allowed for as an additional fit parameter. After fitting to the model to the average *cis*-efficiencies using Eq. 2, an angle of 95.5°, and and offset $x_0 = 0.27$ nm was obtained. The resulting model is shown as green line in Fig. 9 and has to be compared to the mean efficiency values (red dots). The dashed line shows an offset of ± 1 nm in efficiency space as error estimate. The offset towards Alexa 594 $x_0 = 0.27$ nm agrees with the deviation of the average dye-to-dye distance from the proline length $(6.01 \text{ nm} - 5.57 \text{ nm} = 0.44 \text{ nm})$ within the accuracy of this simple model.

Next, ensemble efficiency distributions were calculated by combining *cis* and *trans* isomers according to their population in solution. Using the population of individual isomers as determined by NMR [45], p_{int}^{cis} and p_{ter}^{cis}, weights were determined as listed in Table 2. For polyproline-20, these weights are $p^{trans} = 0.6131$, $p^{cis[1-19]} = 0.0125$, and $p^{cis20} = 0.0681$.

For poly-15 and polyproline-30, the same p_{int}^{cis} and p_{ter}^{cis} measured on polyproline-20, were applied assuming that they are not strongly influenced by the proline chain length. Because the *cis*-content is larger in polyproline-30, an error in p_{int}^{cis} and p_{ter}^{cis} has a larger impact on the accuracy of the ensemble composition. For example, if polyproline-30 has a p_{int}^{cis} value of 4% instead of 2%, the multi-*cis* isomer ensemble content increases from 15% to 37%, whereas the all-*trans* isomer contribution drops from 50% to 28%. As a result, the obtained ensemble efficiency histograms sensitively depend on the value of p_{int}^{cis} and p_{ter}^{cis}, particularly for the longer polyproline-30 chain.

As seen before, the *cis* ensemble content and thus the content of isomers with double-*cis* bonds increases with the chain length. For polyproline-30, this contribution is about 15%. To estimate the impact of double-*cis* species on the efficiency histogram, we simulated a subset of double-*cis* isomers (Table 1). The obtained weights for each chain length and isomer were used in the next step, to calculate efficiency distributions of the entire ensemble.

Combining Photons into Bursts

So far, we calculated efficiency distributions of single simulations (Fig. 8) and their accumulated histograms (Fig. 9). To calculate burst efficiencies in closer resemblance to single molecule experiments, we need to define how the recorded photons are combined, e.g. from multiple trajectories. The specific approach depends on the relative time scales of the relevant processes in the experiment and the simulation. In single molecule experiments on freely diffusing molecules, ten to hundreds of photons are recorded in each burst of several ms duration. On the simulation side, in contrast, multiple 100 ns trajectories are available. We consider three different ways of combining photons into bursts and compare the resulting efficiencies to experiment.

The relevant time scales are the two autocorrelation times for the dye dynamics, namely those of the orientation factor κ^2 and of the distance R fluctuation, from hundreds of picoseconds (κ^2) to ns (R) (Table 3); further the polyproline chain dynamics of a specific isomer with the slowest motions in the 5 ns range (Table 3, L and Ω), the *cis* to *trans* isomerization time ranging from minutes to hours for polyproline [82], the experimental burst recording duration of several ms and the respective inter photon times [83], as well as the simulation trajectory length of 100 ns.

In the first case (burst average over fast and slow dye motions as well as the polyproline isomerization), the burst duration is assumed to be longer than all other time scales mentioned above. Accordingly, in this case, each measured burst consists of photons from the entire isomeric ensemble. To achieve a most comprehensive sampling, therefore, photons from all available trajectories
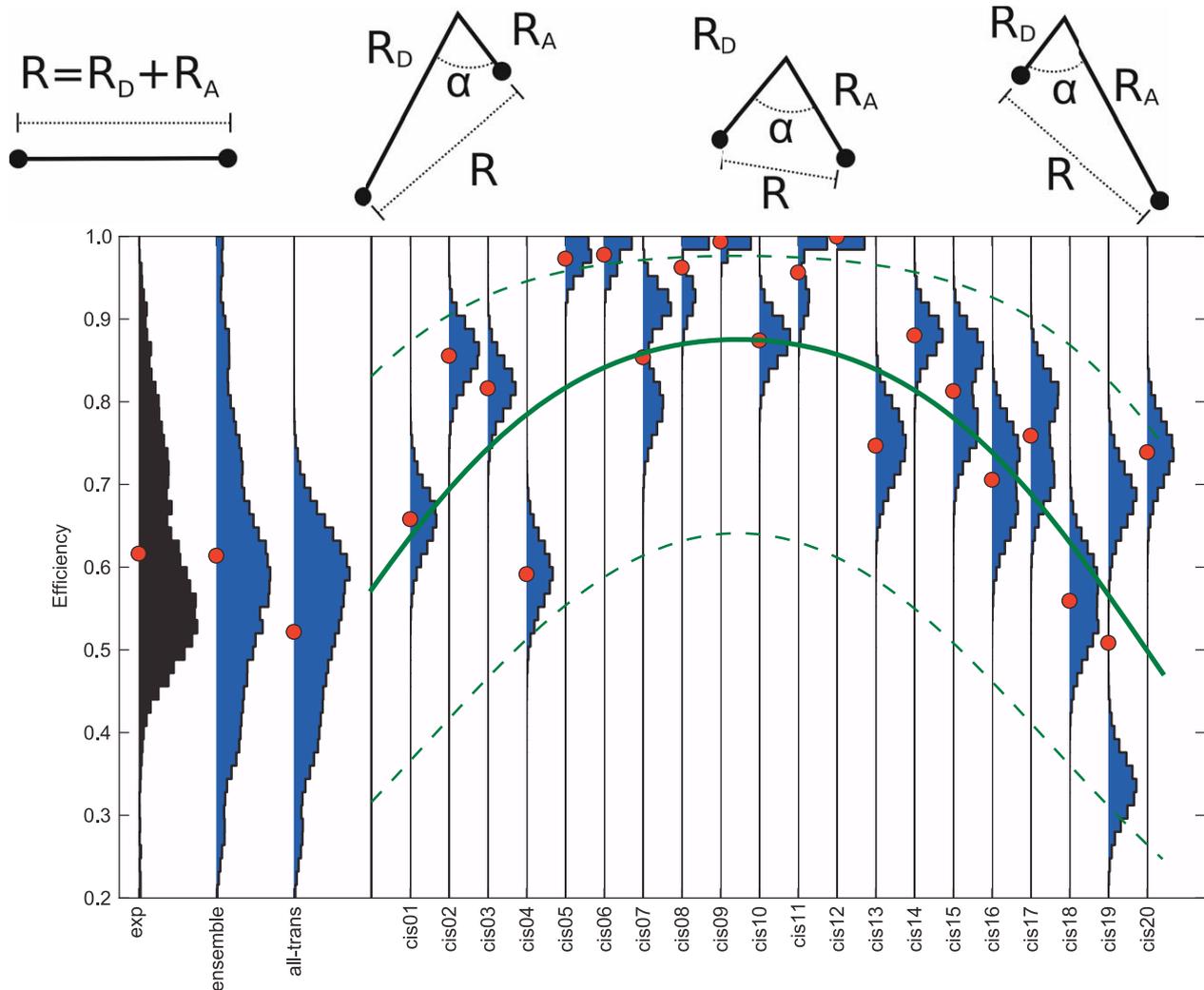
Figure 9. FRET efficiency of *trans* and *cis* isomers. Comparison between measured FRET efficiency histograms (black) and histograms computed from the simulations (blue: ensemble, all-trans and cis01 ... cis20). Red dots denote the respective mean values. The simple model sketched on top and defined in the text describes the general trend (green line) that isomers with a *cis*-bond close to the termini show lower efficiencies, whereas those with *cis*-bonds close to the polymer center tend to yield higher efficiencies. The dashed green lines estimate the spread of the average efficiencies of the *cis* simulations mirroring the spread found for the all-*trans* simulations ($\Delta R = \pm 1.0$ nm). For illustration purposes, the photons of the individual *cis* were not discarded when generated below 1 nm as described in the Methods Section. The high efficiencies observed for *cis*-6 to *cis*-12 result from dyes in contact and are quenched in the experiment.
doi:10.1371/journal.pone.0019791.g009

with their appropriate ensemble weight are combined. The blue line in Fig. 10 shows the resulting efficiency distribution as a single peak whose width is solely determined by the shot-noise. Experiments measuring ensemble efficiencies (e.g., CW in bulk) correspond to this case, except that in ensemble measurements an effectively infinite number of photons is gathered, and therefore the shot noise vanishes. For the polyproline system at hand, however, the isomerization times are long compared to the burst duration, and thus this case is not expected to apply here. Indeed, the measured efficiency distribution (Fig. 10, black) is much broader.

Accordingly, for the second case (burst average over fast and slow dye motion), we assume that the isomerization time is longer than the average burst duration, with the remaining dye and chain dynamics still being fast compared to the burst duration. In this case, all photons from a measured burst originate from one particular

isomer. Because the trajectory length is much shorter than the burst duration, each burst is generated from all trajectories of a particular isomer. Figure 10, green line, shows the resulting efficiency distribution. Because in contrast to the previous case, averaging is not done over multiple isomers within each burst, as assumed above, the individual *cis* isomers contribute high efficiencies (>0.7) to the efficiency distribution (Figure 10). As shown in Fig. 10 (dashed green line), these high efficiencies are also observable in the experiment (black line). In addition, Fig. 10 reveals that the low efficiency side agrees with the experimental distribution (solid green line). However, when comparing the region around 0.7, a gap between the all-*trans* peak and the high efficiency *cis* region is present, not found in the experiment. In analogy to the comparison of this case and the above case, which averages over the polyproline isomerization, this hints at additional dynamics slower than the burst duration, averaged out in the current case.
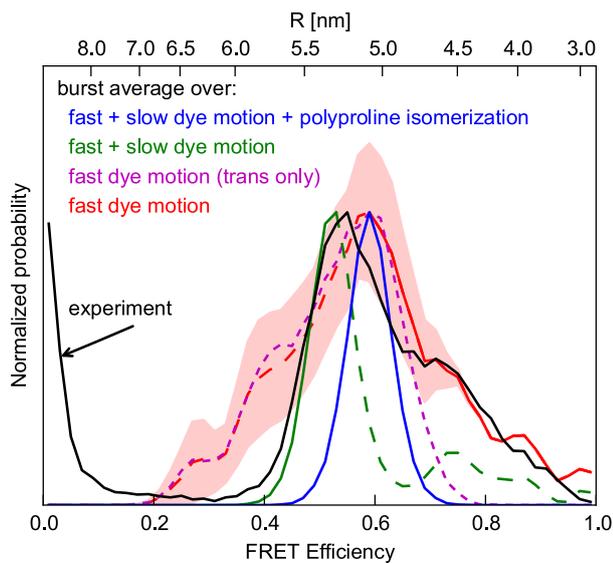
Figure 10. Combining photons into bursts. Comparison of different photon accumulation methods for a full polyproline-20 ensemble at 293 K with the experiment (black). Three different accumulation methods (colors) were considered. First, efficiencies were calculated from the full ensemble (blue), for which each photon burst has been combined from photons of all *cis* and *trans* simulations, and which therefore average over all motions and heterogeneities covered by the simulations. Second, each efficiency value was calculated from photons of all simulations of a randomly chosen isomer (green), thereby averaging over all dye motions but not over different isomers. Third, each efficiency value is derived from photons of one single trajectory, and weighted by the appropriate ensemble probability (red). The impact of the *cis*-isomers is demonstrated by comparison to the all-*trans* only efficiency histogram (magenta). The bootstrapping standard error (Fig. 8) of the all-*trans* isomers is drawn as light red area. Efficiency histograms were normalized to their maxima.
doi:10.1371/journal.pone.0019791.g010

If this is true, one would expect a better agreement for the third case considered here. In this case (burst averaging over fast dye motion only), we now assume that the dye dynamics contains additional components that are slow compared to the burst duration. An example of such a component is the transition between different conformations of the dye, e.g. the ones shown in Fig. 7. Therefore, all photons in a burst originate from a distinct dye conformation with an interconversion time larger than 100 ns. In resemblance to this, each burst is generated from one distinct simulation trajectory. The previous assumption of slow isomerization times compared to the burst recording duration is automatically contained in this case, since each trajectory contains a single distinct isomer. Figure 10 (red line) shows the resulting efficiency distribution. In contrast to the burst average over fast *and* slow dye motion, where all-*trans* and *cis*-isomers were resolvable (Fig. 10, green line), the conformational heterogeneity on time scales beyond 100 ns and thus of different simulations, is now visible as already observed in Fig. 8. As shown in Fig. 10 (red line), this heterogeneity is particularly pronounced for the all-*trans* simulations due to the largest number of simulations (Table 1) and the all-*trans* isomer being the largest fraction of the ensemble. The small numbers of simulations result in a considerable statistical error, shown as red area in Fig. 10 and calculated from the all-*trans* isomer. When comparing this result to the experiment, the high efficiency side (solid red line) with *cis*-efficiencies agrees with the experiment (black). The discrepancy (gap around 0.7) previously observed (burst average over fast *and* slow dye motion) vanishes.

However, an additional low efficiency shoulder is visible not present in the experiment (dashed red line).

This deviation is not within statistical uncertainty (Fig. 10, red area) and may be due to several reasons. First, because all simulations have been started from the *open* conformation (Fig. 7), this conformation may have been oversampled. Second, although the dye dynamics described by the fluorescence anisotropy decay times agrees with the experiment, we cannot fully exclude over- or underestimation of the dye-hydrophobicity with our choice of partial charges. Third, this discrepancy can be explained by the presence of two different dye dynamics in the experiment as described below.

Overall, the low-efficiency side (<0.5 in Fig. 10) in case of burst averaging over fast *and* slow dye motions agrees well with the experiment, whereas on the high efficiency side (>0.5), better agreement is seen for burst averaging over fast dye motions only (Fig. 10, solid green and blue vs. black). From the above discussion of time scales, this finding would imply that the low efficiency side (i.e., large distances) is governed by fast dynamics, whereas parts of the slow dynamics govern the high efficiency (i.e., shorter distances) side only. Close inspection of our simulations suggests a possible structural explanation for this finding. In particular, the hydrophobic interactions between the polyproline and the Alexa 488, which give rise to the structural heterogeneity shown in Fig. 7, with very slow transitions between the *open* and *closed* conformation. In the *open* conformation, the dye-reorientation is fast compared to the burst duration and thus sampled within a single burst, in agreement with the low efficiency side (Fig. 10). In the *closed* conformations, the dye dynamics is largely restricted, with the high FRET efficiency therefore being governed by the slow transitions between these sub-states, in agreement with the observed burst averaging over fast and slow dye motions.

Next, we compare efficiency distributions for different dye-labeled proline lengths. Figure 11 shows the calculated efficiency distributions (burst averaging over fast dye motions only) from simulations with proline lengths 15, 20, and 30 (solid lines) as well as measured efficiencies for lengths of 14, 20, 27, and 33 (dashed). The general length effect, increase in efficiency for shorter prolines and vice versa, is observed.

For polyproline-15, the calculated distribution has the same narrow shape as found in the experiment, however with the simulated efficiency distribution shifted towards higher efficiencies. Purely from the length difference between polyproline-14 (experiment) and polyproline-15 (simulation), an opposite shift is expected. A similar slight discrepancy is seen for polyproline-30, where the peak should be located between the experimental peaks of polyproline-27 and -33, but is seen in Fig. 11 somewhat below polyproline-33.

While the overall agreement between simulation and experiment is good, this observed systematic deviation is striking. Apparently, compared to our simulation results, the experimental efficiencies tend to be shifted slightly towards 0.5 within both the high as well as the low efficiency regime. Overall, such behavior cannot be explained by an uncertainty in the measured R_0, which would lead to a uniform shift in one direction. With the same argument, also force field inaccuracies, which might, e.g., overestimate the hydrophobicity of the dyes and thus also the population of the *closed* conformation, are incompatible with the observed deviation. As a possible explanation one might consider a modified Förster law with, e.g., an effective power smaller than 6 in Eq. 2 (e.g. a power of ≈ 4 yields the best agreement of the simulated and experimental peak positions). Such effects have been observed previously [31] and may originate from inter-dye quenching or the breakdown of the point dipole approximation
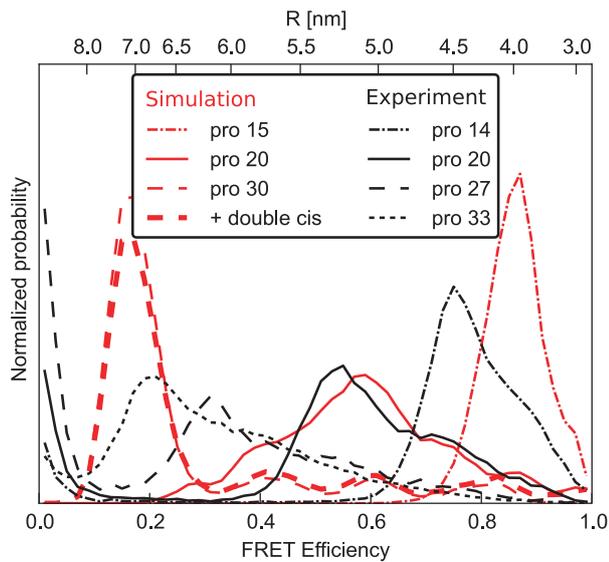
Figure 11. Comparison between proline 15, 20 and 30 and experiment. Efficiency histograms averaging over fast dye motions only (Fig. 10) are shown for three different polyproline-lengths (red, dashed-dotted, solid, dashed), and corresponding measured efficiency distributions (black). For polyproline-30, inclusion of double-*cis* isomers (see Text S1) only slightly changes the efficiency histogram (bold dashed line).
doi:10.1371/journal.pone.0019791.g011

[41,84]. As a third possible cause, decreased fluorescence lifetimes at high efficiencies leading to a stronger deviation from $\kappa^2 = 2/3$ has been discussed [31], but is already included within our simulation approach and thus unlikely to explain the deviation.

Comparing the shapes of the polyproline-30 curves, both the calculated as well as the measured efficiency distributions share shoulders reaching into the high efficiency regime. However, this shoulder is much more pronounced in the experiment than in the simulation. Closer inspection shows that the shoulder originates exclusively from *cis*-isomers. To interpret this discrepancy it is thus helpful to ask what fraction of the *cis*-population, according to the NMR results, is expected to fall into this high efficiency range. Interestingly, with the 2% *cis*-population (per bond) from NMR, and considering the fact that only about 2/3 of the cis-population contributes to the high efficiency shoulder (whereas about 1/3 contributes to the main, all-*trans* peak, see Suppl. Fig. S2), the NMR results are incompatible with the high (ca. 50%) population seen by FRET. Accordingly, a small correction of the NMR values towards higher populations of the *cis*-isomers would resolve both the discrepancy between NMR and FRET as well as that between FRET and our calculated efficiency histogram. In contrast, our neglect of multiple *cis*-conformers is unlikely to explain the discrepancy, as seen from the small effect when including the double-*cis*-species (Fig. 11, bold dashed red line) as the dominant multiple *cis*-population.

In experiments, a peak around zero efficiency is seen for all proline species. This peak originates from polyproline molecules lacking an active acceptor dye, either because of imperfect labeling or because of photobleaching of the acceptor dye during the measurement [31]. In our simulations, all molecules carry a donor and an acceptor dye, and photobleaching is not considered; the zero efficiency peak is thus absent. The clear separation of the zero efficiency peak from the rest of the signal allows us to compare only the signal from the "intact" molecules with the simulated data.

Reconstructing Distance Distributions from FRET Efficiencies

We have shown above that accurate efficiency histograms can be calculated from a combination of atomistic MD simulations and Monte Carlo photon sampling. Now we will ask the inverse question: Can the dye orientation distributions obtained from the simulations be combined with *measured* FRET efficiency histograms in such a way as to enable reconstruction of more accurate distances and, possibly, also distance distributions, than by the established $\kappa^2 = 2/3$ approximation? And if so, which accuracy can be expected at the different conceivable levels of approximations that were mentioned in the Introduction?

To address these questions, the efficiency histogram calculated from the hybrid MD/MC approach (where the distance distribution is known) as well as the single molecule FRET efficiency histogram from the experiment (where the distance distribution is unknown) were used as input for the backward calculations. The thus reconstructed distance distribution, both from the synthetic and the experimental FRET data, were then compared to the known distribution from the simulation. For each level of approximations, thereby, the impact on accuracy of the respective assumptions is quantified.

As a common framework for a proper definition of the applied approximations, we consider the most general (linear) transformation from a distance distribution $p(R)$ to an efficiency distribution $q(E)$ in terms of transfer functions $g(E,R)$,

$$q(E) = \int_0^{R_{max}} g(E,R)p(R)dR. \tag{14}$$

Each level of approximation, will be defined through an approximately specified transfer function $g(E,R)$ – or, after discretization, transfer matrix. In all cases, the all-*trans* polyproline-20 structural ensembles were used for the calculation of the transfer function as well as to generate the synthetic efficiency distribution $q(E)$; to reconstruct $p(R)$ from the experimental efficiency distribution, which involves an isomer mixture, the full structural ensemble with appropriate weights was used to calculate the transfer function (except for transfer functions g_4 and g_5). At each approximation level, $p(R)$ was then reconstructed from $q(E)$ and $g(E,R)$ by inverting a discretized version of Eq. 14.

As, generally, such inversion is numerically highly unstable, regularization assumptions are required. Here, motivated from the observation of two structural conformers (*open* and *closed* conformation, cf. Fig. 7), we assumed that $p(R)$ can be described sufficiently accurate by the sum of $n = 2$ Gaussian functions centered at R_i of width σ_i,

$$p(R) = \sum_{i=1}^{n} a_i \exp\left[-\frac{(R - R_i)^2}{2\sigma_i^2} \right]. \tag{15}$$

With this description, the χ^2 between the calculated and the reference efficiency distribution was minimized by variation of R_i and σ_i using the two array differential evolution algorithm [85]. Extension of this method to more Gaussian functions or to a more sophisticated model [86] is straightforward.

At the lowest level of refinement, the usually assumed isotropic dye orientation distribution is considered, implying $\kappa^2 = 2/3$, independent of the mutual distance between the two dyes. The efficiency distribution $q(E)$ was obtained from the donor-acceptor

distance distribution $p(R)$ via the usual Förster formula, Eq. 2,

$$q(E) = p(R)\frac{dR}{dE} = -\frac{1}{6}p(R)\frac{R_0}{E^2(1/E-1)^{5/6}}. \quad (16)$$

In the more general transfer function formalism used further below (Eq. 14), the above result (Eq. 16) is readily recovered from the transfer function

$$g_1(E,R) = \delta\left(E - \frac{1}{1+(R/R_0)^6}\right) \quad (17)$$

shown in Fig. 12A.

Figures 13A and B show how well the respective transfer functions capture the relation between $p(R)$ and $q(E)$ as obtained from the simulations. At this first level of refinement, using the above $\kappa^2 = 2/3$ transfer function for both the all-trans ensemble (A) as well as for the full ensemble, containing all isomers (B), quite narrow efficiency distributions (green curves) are obtained, which are also shifted towards lower efficiencies with respect to the reference efficiency distributions (blue, black). As expected, the reconstructed distance distributions, Fig. 13E and F (same color scheme), are also shifted towards smaller distances, with the maximum being off by more than $0.5\,\text{nm}$. Further, the reconstructed distance distribution has a shoulder that is not seen in the reference distribution. Overall, the reconstruction is not satisfactory at this level of refinement. Figure 13C and D show, that for the all-trans MC and the full ensemble experimental efficiencies, respectively, adjusting of parameters in Eq. 15 led to convergence.

To quantify to which extent the assumption of an isotropic dye orientation distribution causes this discrepancy, at a second level of refinement the correct $\langle\kappa^2\rangle$ value was used, as obtained from the respective MD simulation ensemble (cf. Fig. 6). Still, this value was assumed independent of the distance between the two dyes. This approximation is described by the transfer function

$$g_2(E,R) = \delta\left(E - \frac{1}{1+(R/R_{\text{adj}})^6}\right), \quad (18)$$

with $R_{\text{adj}}^6 = \frac{3}{2}\langle\kappa^2\rangle R_0^6$, and $\langle\kappa^2\rangle = 0.80$. As seen in Fig. 12B, this refinement results in a slight shift of the Förster curve with respect to the isotropic dye orientation approximation (Fig. 12A).

At this improved level of refinement, a slight shift of the calculated efficiency distributions towards the reference distributions is observed (red curves in Figs. 13 A, B). As a result, correspondingly improved reconstructed distance distributions are obtained (Fig. 13 E, F). However, the shapes of the efficiency curves are still too narrow, and the shoulder in the reconstructed distance distribution is still present. Apparently, these artifacts are mainly caused by further approximations not investigated so far.

Therefore, at the third level of refinement, we drop the previous assumption that the dye orientation distribution is independent of the donor-acceptor distance. Accordingly, the MD structure ensemble was split into groups according to mutual dye distance, and an average orientation factor $\langle\kappa^2\rangle_R$ was calculated separately for every group, i.e., as a function of R. Note that this distance dependent orientation factor

$$\langle\kappa^2\rangle(R) = \langle(\cos\theta_{\text{DA}} - 3\cos\theta_{\text{D}}\cos\theta_{\text{A}})^2\rangle(R) \quad (19)$$

differs from the previous ones in that it captures correlations between the dye orientation distribution and the donor-acceptor distances. This can be used to construct the transfer function

$$g_3(E,R) = \delta\left(E - \frac{1}{1+(R/R_{\text{red}})^6\cdot(1/\langle\kappa^2\rangle(R))}\right), \quad (20)$$

defining $R_{\text{red}}^6 = R_0^6\cdot 3/2$.

As seen in Fig. 12C, the resulting transfer function is not strictly monotonic any more, such that the inverse transformation to $R(E)$ is not straightforward and, the above regularization techniques need to be applied.

This refinement step yields a marked improvement of both peak position and shape of the obtained efficiency distributions (Figs. 13 A and B, cyan). Only a slight peak shift towards lower efficiencies remains for the all-trans ensemble (Fig. 13 A), as is also seen for the experimental efficiencies in Fig. 13B. Also for the distance reconstruction, the dominant peak is now at the correct position in both cases (Fig. 13 E, F), although the second peak in the synthetic distance reconstruction using the all-trans ensemble still remains and leads to an overestimate of the distribution for smaller distances.

So far, our transfer functions uniquely defined the efficiency E for each distance R. Before continuing with further refinement steps, we demonstrate how the experimental shot noise impacts the reconstruction of distances. Two fundamental approaches have been used so far to calculate the shot noise contribution via numeric solution [78,80] or via simulation [87,88]. Mathematically, the shot noise free efficiency distribution $\tilde{q}(E)$ is convoluted with a shot noise kernel S resulting in an efficiency distribution $q(E)$ including the shot noise. This convolution

$$q(E) = \int \tilde{q}(E')S(E-E')dE', \quad (21)$$

broadens the underlying efficiency distribution $\tilde{q}(E)$ to $q(E)$. Because of this broadening, the reconstructed distance distribution $p(R)$ is narrowed when shot noise is taken into account. In analogy to image reconstruction from a de-focused image by inversion of the convolution with the appropriate image transfer function, the achieved accuracy and the ability to recover finer details of the original distance distribution are limited by the information loss due to convolution of the shot noise kernel S with the transfer function $g(E,R)$, Eq. 14.

Since determining the shot noise kernel S of an experimental BSD is non-trivial, the experimental shot noise (bursts ≥ 100) was included in the transfer function as follows. Each distance bin of the transfer matrix (columns in Fig. 12) was randomly sampled by 1200 bursts from the experimental BSD. The target efficiency for each burst was directly calculated from the transfer function (g_1, g_2, g_3) or randomly picked from the efficiency distribution (for the following refinement steps). According to the target efficiency, donor and acceptor photons were randomly generated, and the obtained burst efficiency was then recorded in the transfer function. Figure 12D and H illustrate the impact of an experimental shot noise (bursts ≥ 20) on transfer functions (Fig. 12C and G). Comparison of C and D illustrates, that after the inclusion of shot noise, the transfer function not uniquely defines an efficiency E value for each distance R, but instead an efficiency distribution. The here observed effect of the BSD on the transfer function is purely of stochastic origin, whereas a similar but independent effect will be seen in the following refinement level.
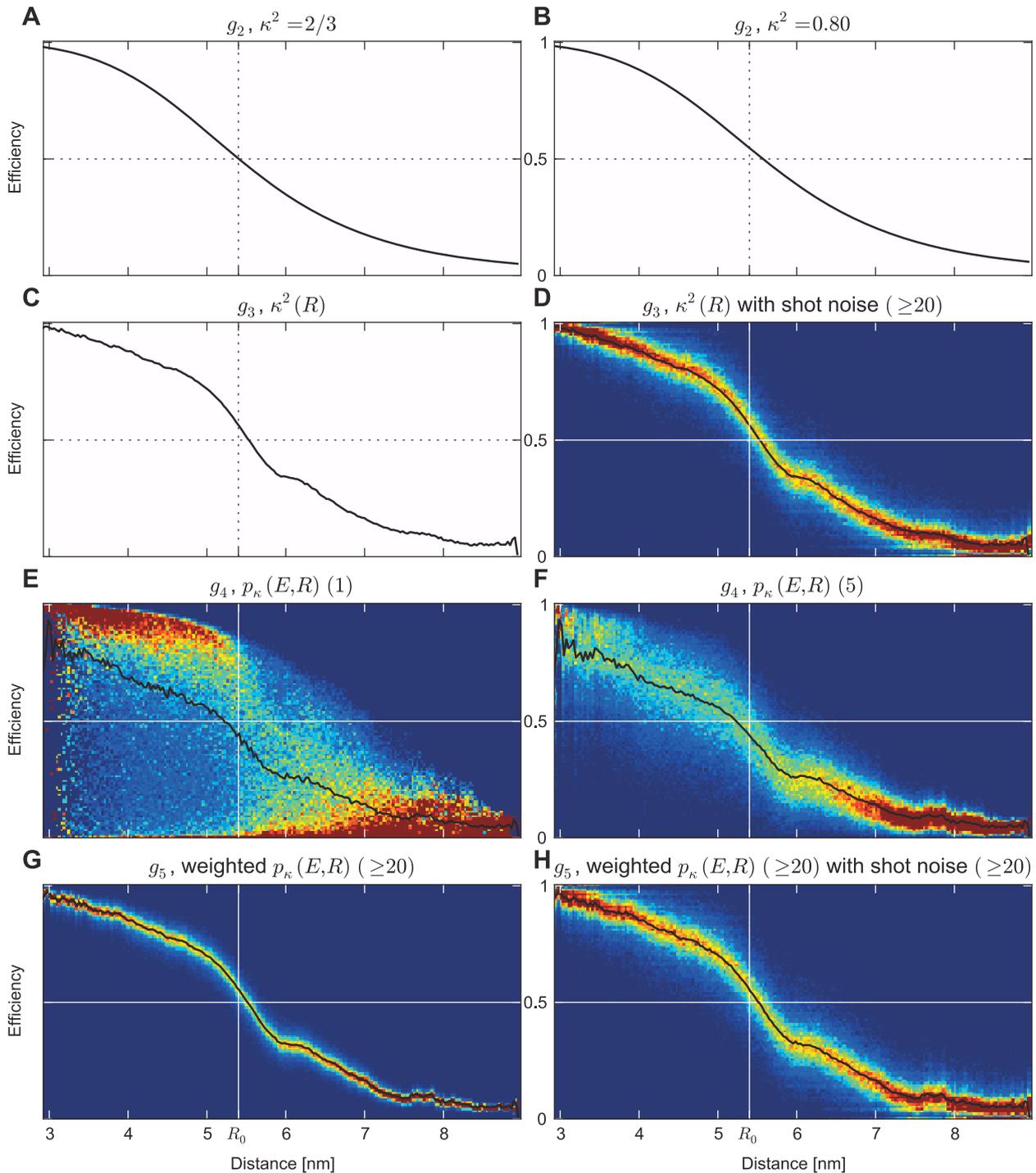
Figure 12. Transfer functions g_i at increasingly refined approximation levels. Transfer functions A, B, and C are shown as black curves; the remaining transfer functions are shown color-coded, with averages highlighted as black curves. Transfer function g_1 was calculated using the assumption of $\kappa^2 = 2/3$ (A). For g_2, κ^2 was adjusted to represent the ensemble average in the simulations (B). g_3 includes the distance dependency of κ^2 without (C) and with (D) shot noise derived from the experimental BSD (burst size or lower burst size cutoff given in brackets). In contrast to a distance dependent averaged κ^2, g_4 includes the κ^2 distributions at each distance without (E) and including averaging within a burst (F). In g_5, the time dependent photon emission (Fig. 14) is included, shown without (G) and with experimental shot noise (H).
doi:10.1371/journal.pone.0019791.g012

To motivate this level of refinement, recall that in all levels of refinement considered so far the full structure ensemble has been used to calculate appropriate averages for the orientation factor κ^2.

This approach implies the salient assumption that each single burst samples the same dye orientation distribution – which, however, holds true only if all components of the dye motion are much faster
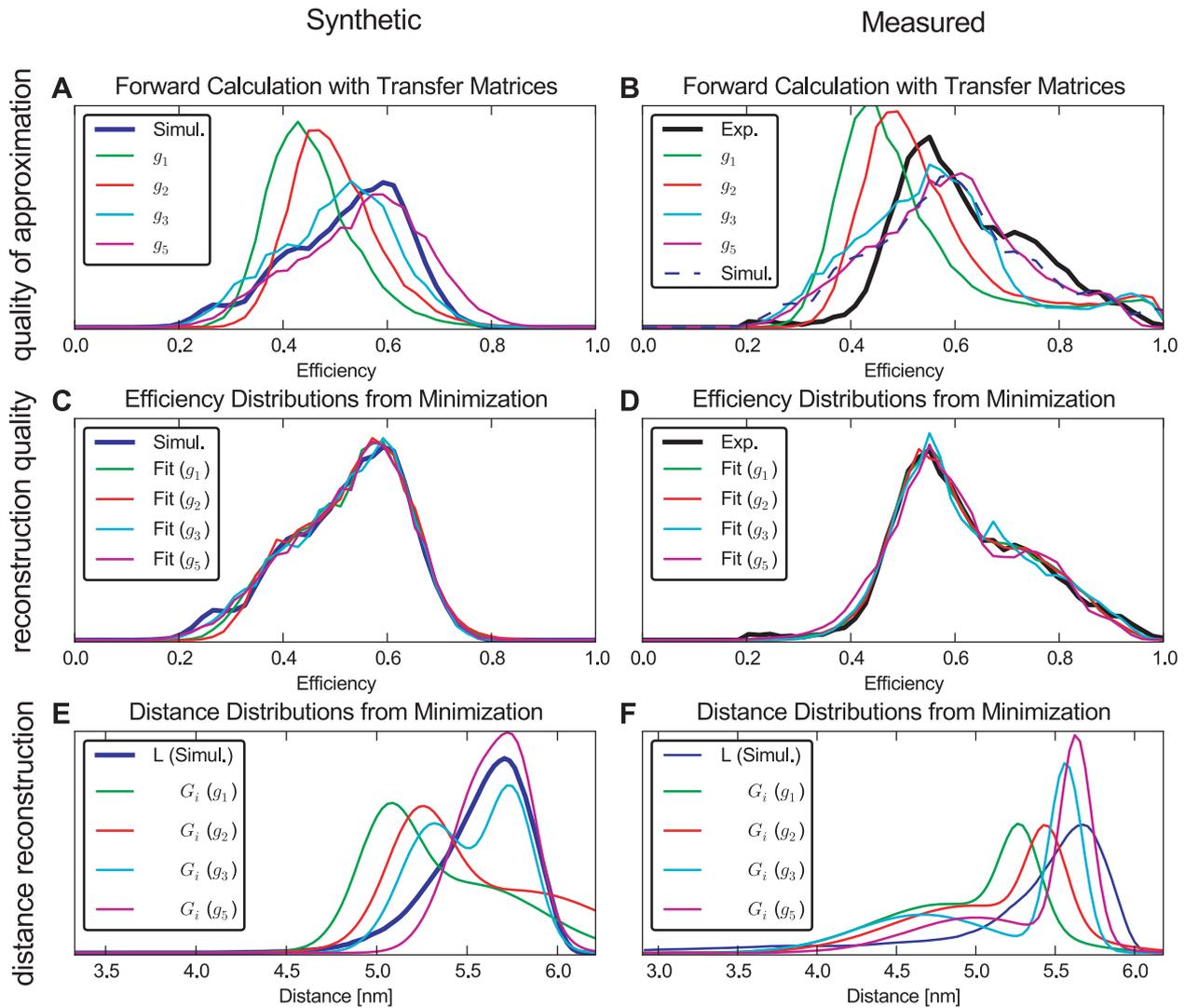
Figure 13. Distance reconstruction from efficiencies. Reconstruction of distance distributions from synthetic efficiencies with known distance distributions from simulations, and measured efficiencies with unknown distance distribution. Two ensembles were considered: The left column consists of all-*trans* polyproline-20 at $293\,\mathrm{K}$. The right column also includes the *cis*-isomers with appropriate weights and uses experimental efficiencies as reference. In the first row (A and B), the efficiency distributions obtained from multiplying the transfer matrices (discretized transfer function) with the distance distributions obtained from simulations are shown together with efficiencies derived from our simulations and experiment as reference. The second row (C and D) depicts the efficiency obtained by optimizing the parameters of two Gaussian distance distributions as a measure for the reconstruction quality. The efficiencies were calculated by multiplying the transfer matrix with the Gaussian distance distributions with optimized parameters (see Text S1). The distance distributions obtained from reconstruction and simulation are shown in the third row (E and F). In all graphs, the reference is plotted with a bold line. Notably the experimental reference distance is inaccessible in F. The employed transfer matrices include experimental shot noise.
doi:10.1371/journal.pone.0019791.g013

than the burst duration. As has been shown from the above comparison between measured efficiency distributions and those obtained from three different structure ensembles, there are slow components of the dye motion, which may render this salient assumption questionable. Our last level of refinement attempts to include the dominant effect of this limited dye orientation sampling within the transfer function. Note, however, that a rigorous treatment of this effect would require to go beyond the limits of the transfer function framework, and here is only achieved conceptually by our explicit hybrid MD/MC simulation approach.

At this refinement level, accordingly, the columns of the transfer matrix are formed from distance dependent transfer efficiency distributions $p_\kappa(\kappa^2,R)$ rather than single valued R-dependent

averages $\langle\kappa^2\rangle(R)$, from which the transfer function $g_4(E,R)$ is derived as

$$g_4(E,R)=\frac{p_\kappa(\kappa^2,R)}{\int d\kappa^2 p_\kappa(\kappa^2,R)}\cdot\frac{d\kappa}{dE}. \tag{22}$$

Here, the integral over κ^2 in the denominator of g_4 normalizes the probability distribution on the distances and $d\kappa/dE$ transforms $p_\kappa(\kappa^2,R)$ to $p_\kappa(E,R)$. The normalized transfer function is obtained from orientation factor histograms for different distances from the MD trajectory ensemble, applying Eq. 2.
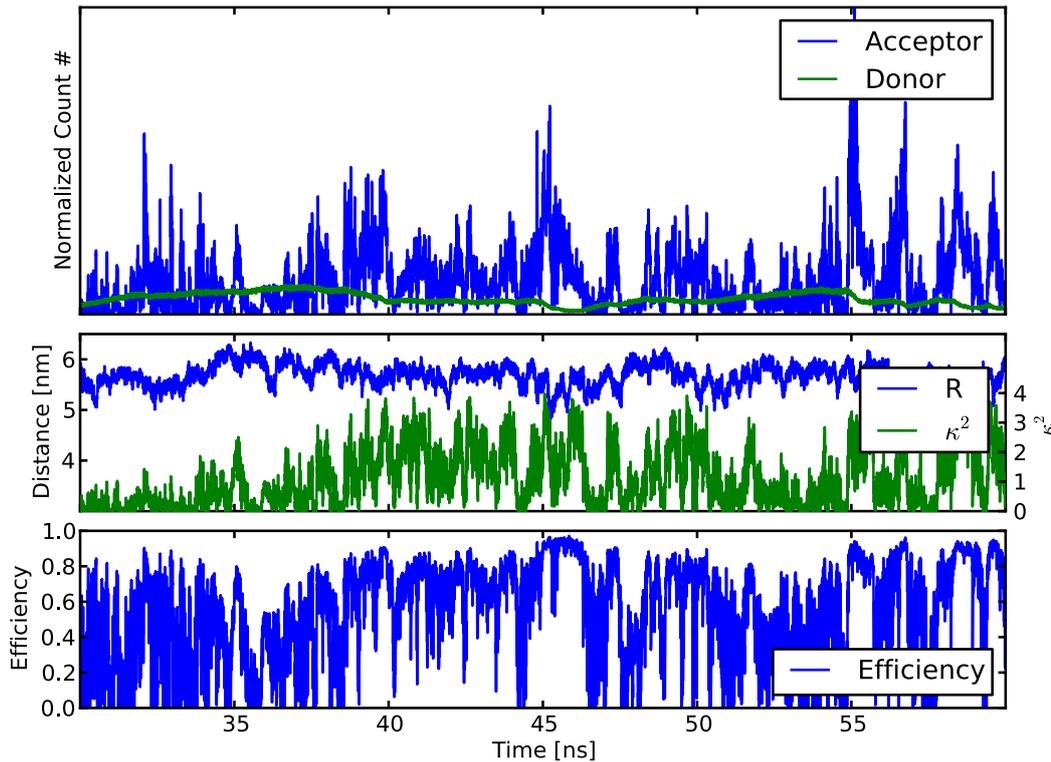
Figure 14. Time dependent photon emission along a single trajectory. Top: normalized acceptor (blue) and donor (green) photon count for time independent excitation probability. Mid: corresponding distance $R(t)$ and orientation factor $\kappa^2(t)$. Bottom: resulting time dependent instantaneous efficiency.
doi:10.1371/journal.pone.0019791.g014

Figure 12E shows the transfer function resulting from the samples of our simulations. Notably, there is a distance dependent maximum efficiency due to the κ^2 range from 0 to 4. The samples of $p_\kappa(\kappa^2, R)$ from our simulations each determine the efficiency samples of $p_\kappa(E, R)$. In experiments, however, efficiencies are determined using multiple photons. Thus, the efficiencies in the transfer function need to be averaged over multiple (κ^2, R) samples according to the BSD. Figure 12F shows this effect for a constant burst size of 5. As seen, this burst size dependent averaging introduces a narrowing of the transfer function, independent of the photon shot noise.

Figure 14 motivates and illustrates the next level of refinement. Shown are the donor and acceptor photon counts from a trajectory, extensively sampled with photons created in our MC process. In high efficiency regions with efficiency values in the lower plot close to one (e.g., around 45 and 55 ns), a marked depletion of donor and acceptor photon counts (dips in the green curve in the upper plot) is observed. As a result, also the mean intensity I_{avg} depends on κ^2 and R and thus affects the probability of obtaining a photon from a distinct (κ^2, R) conformation. Thus, for each instant t, the orientation $(\kappa^2(t), R(t))$ depends on the history of orientations within the fluorescence lifetime. Because in the construction of g_4, the probability distribution of dye orientations $p_\kappa(\kappa^2, R)$ is only normalized at each distance, this memory is not described. By applying appropriate intensity weights, our last, most realistic transfer function g_5 includes also this effect.

In our transfer function construction, the intensity of each (κ^2, R) sample was determined by extensive photon sampling of our trajectories. Thereby, the adsorption events were equally distributed over the whole trajectory and the emission times of the photons was recorded. The (κ^2, R) samples were then weighted according to their total emitted photon count. Notably, the samples are implicitly weighted according to their efficiency history in experiments. In Fig. 12G, a shift towards higher efficiencies as an effect of this weighting is seen. To reduce computationally expensive photon sampling of the trajectories, g_4 and g_5 were calculated from 20 all-*trans* simulations only.

Applying this transfer function g_5 to our known distance distributions to asses the quality of approximations results in efficiency distributions only slightly different from the ones for g_3 (Fig. 13 A, B). Nevertheless, as seen in Fig. 13, the high efficiencies in the experimental ensemble were reproduced better than for g_3. When using g_5 for the reconstruction of distances using the synthetic efficiencies, the best agreement with the distance distribution from the simulations was found (Fig. 13 E). Also, the reconstructed peak location using the experimental efficiencies is slightly closer to the peak from the simulations (Fig. 13 F).

Overall, in the experimental reconstruction, all distance distributions of different refinement levels are shifted towards lower distances in comparison to the simulation distance distribution. This agrees with the observation of low efficiency overestimation shown in Fig. 10.

These tests demonstrate, that a markedly improved reconstruction over the established approaches is achieved by inclusion of dye motion and photon statistics obtained by our hybrid simulation approach of simulated data. Further, by using step-

$$g_5(E, R) = \frac{p_\kappa(\kappa^2, R) \cdot I_{avg}(\kappa^2, R)}{\int d\kappa^2 p_\kappa(\kappa^2, R) \cdot I_{avg}(\kappa^2, R)} \cdot \frac{d\kappa}{dE}. \qquad (23)$$

by-step refined approximation levels for the transfer functions, a systematic improvement of the inverse distance reconstruction is achieved for the polyproline system.

Conclusions

We have demonstrated that structural information on the dynamics of FRET dye pairs from MD simulations improves the reconstruction of distances and distance distributions from experimental FRET efficiency distributions over the usual $\kappa^2 = 2/3$ approximation, which assumes isotropic and uncorrelated distributions of the dye transition dipole orientations. A hybrid MC/MD method was developed and tested, which uses this structural information in combination with a Monte Carlo description of photon absorption, FRET-transfer, and emission, to calculate quantitative efficiency distributions. Based on the obtained good agreement with measured efficiency distributions of polyproline constructs, we have investigated several levels of approximation, resting on the particular relation of the different relevant time scales of the experiment and of the simulations. For the system at hand, this analysis revealed a previously unknown slow component of the dye movement. Our analysis further highlights that careful consideration of the time scales of the involved processes is crucial, and offers a framework that is flexible enough to capture the different time scale relationships expected for a broad range of systems. Unexpectedly, already for the simple polyproline system at hand, where the dyes are usually assumed to be sufficiently flexible to justify the established $\kappa^2 = 2/3$ approximation, severe deviations were seen. Our results suggest that for FRET dye pairs attached to proteins or DNA/RNA complexes, the orientational dynamics are typically more restricted due to steric hindrance and electrostatic interactions, a simulation approach like the one developed here is essential.

Supporting Information

Figure S1 Autocorrelation decay times of multiple parameters. R is the inter-dye distance, L the chain end-to-end distance and κ^2 the orientation factor. The 3D, 2D and 2nd Legendre Polynomial

of 2D (Anisotropy decay) was determined from Alexa 488. The decay here is from an all-*trans* polyproline-20 simulation.
(EPS)

Figure S2 Polyproline-30 *cis*-isomer efficiencies. For each isomer, the normalized probability is shown.
(EPS)

Video S1 Distance and Orientation factor from simulations. For illustration, a fragment of 10 ns simulation time from polyproline-15 with the two dyes attached is shown as example for the dynamics. The box in the bottom shows the time dependent orientation and distance as well as the resulting FRET efficiency. The position in the time trace is shown as moving red bar in the box.
(M4V)

Video S2 FRET from Monte Carlo and simulation trajectories. In addition to Video S1, exemplary excitation and de-excitation events are shown. Both competing pathways, de-excitation of donor, as well as the alternate pathway via FRET followed by acceptor de-excitation are displayed.
(M4V)

Text S1
(PDF)

Table S1
(PDF)

Author Contributions

Conceived and designed the experiments: MH NL BS CAMS HG. Performed the experiments: MH NL BS HG DH. Analyzed the data: MH NL BS. Contributed reagents/materials/analysis tools: MH NL BS CAMS HG DH. Wrote the paper: MH NL BS CAMS HG.

References

1. Förster T (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz. Annalen der Physik 2: 55–75.
2. Stryer L, Haughland RP (1967) Energy transfer - a spectroscopic ruler. Proc Natl Acad Sci U S A 58: 719–726.
3. Sisamakis E, Valeri A, Kalinin S, Rothwell PJ, Seidel CAM (2010) Accurate single-molecule FRET studies using multiparameter uorescence detection. Methods Enzymol 475: 455–514.
4. Ha T, Enderle T, Ogletree DF, Chemla DS, Selvin PR, et al. (1996) Probing the interaction between two single molecules: uorescence resonance energy transfer between a single donor and a single acceptor. Proc Natl Acad Sci U S A 93: 6264–6268.
5. Shera EB, Seitzinger NK, Davis LM, Keller RA, Soper SA (1990) Detection of single uorescent molecules. Chem Phys Lett 174: 553 - 557.
6. Berglund AJ, Doherty AC, Mabuchi H (2002) Photon statistics and dynamics of uorescence resonance energy transfer. Phys Rev Lett 89: 068101.
7. Meer BWVD, III GC, Chen SYS (1994) Resonance Energy Transfer: Theory and Data. John Wiley & Sons, 1 edition.
8. Kalinin S, Felekyan S, Valeri A, Seidel CAM (2008) Characterizing multiple molecular states in single-molecule multiparameter uorescence detection by probability distribution analysis. J Phys Chem B 112: 8361–8374.
9. Neubauer H, Gaiko N, Berger S, Schaffer J, Eggeling C, et al. (2007) Orientational and dynamical heterogeneity of rhodamine 6G terminally attached to a DNA helix revealed by NMR and singlemolecule uorescence spectroscopy. J Am Chem Soc 129: 12746–12755.
10. Wozniak A, Nottrott S, Kuhn-Holsken E, Schröder G, Grubmüller H, et al. (2005) Detecting protein-induced folding of the U4 snRNA kink-turn by single-molecule multiparameter FRET measurements. RNA-A Publication of the RNA Society 11: 1545–1554.
11. Sindbert S, Kalinin S, Nguyen H, Kienzler A, Clima L, et al. (2011) Accurate distance determination of nucleic acids via FRET: Implications of dye linker length and rigidity. J Am Chem Soc 133: 2463–2480.
12. Wozniak AK, Schröder GF, Grubmüller H, Seidel CAM, Oesterhelt F (2008) Single-molecule FRET measures bends and kinks in DNA. Proc Natl Acad Sci U S A 105: 18337–18342.
13. Brunger AT, Strop P, Vrljic M, Chu S, Weninger KR (2010) Three-dimensional molecular modeling with single molecule FRET. J Struct Biol 173: 497–505.
14. Düser MG, Bi Y, Zarrabi N, Dunn SD, Börsch M (2008) The proton-translocating a subunit of F0F1-ATP synthase is allocated asymmetrically to the peripheral stalk. J Biol Chem 283: 33602–33610.
15. Schröder G, Grubmüller H (2004) FRETsg: biomolecular structure model building from multiple FRET experiments. Comput Phys Commun 158: 150–157.
16. Uphoff S, Holden SJ, Reste LL, Periz J, van de Linde S, et al. (2010) Monitoring multiple distances within a single molecule using switchable FRET. Nat Methods 7: 831–836.
17. Ha T (2001) Single-molecule uorescence resonance energy transfer. Methods 25: 78–86.
18. Schuler B, Eaton WA (2008) Protein folding studied by single-molecule FRET. Curr Opin Struct Biol 18: 16–26.
19. Widengren J, Kudryavtsev V, Antonik M, Berger S, Gerken M, et al. (2006) Single-molecule detection and identification of multiple species by multiparameter uorescence detection. Anal Chem 78: 2039–2050.
20. Gopich IV, Szabo A (2003) Single-macromolecule uorescence resonance energy transfer and freeenergy profiles. J Phys Chem B 107: 5058–5063.
21. Margittai M, Widengren J, Schweinberger E, Schröder G, Felekyan S, et al. (2003) Single-molecule uorescence resonance energy transfer reveals a dynamic equilibrium between closed and open conformations of syntaxin 1. Proc Natl Acad Sci U S A 100: 15516–15521.
22. Hofmann H, Hillger F, Pfeil SH, Hoffmann A, Streich D, et al. (2010) Single-molecule spectroscopy of protein folding in a chaperonin cage. Proc Natl Acad Sci U S A 107: 11793–11798.

– P8 –

153

23. Schröder GF, Grubmüller H (2003) Maximum likelihood trajectories from single molecule uorescence resonance energy transfer experiments. J Chem Phys 119: 9920–9924.

24. Santoso Y, Torella JP, Kapanidis AN (2010) Characterizing single-molecule FRET dynamics with probability distribution analysis. Chemphyschem 11: 2209–2219.

25. Hillger F, Hänni D, Nettels D, Geister S, Grandin M, et al. (2008) Probing protein-chaperone interactions with single-molecule uorescence spectroscopy. Angew Chem Int Ed Engl 47: 6184–6188.

26. Van Oort B, Murali S, Wientjes E, Koehorst RBM, Spruijt RB, et al. (2009) Ultrafast resonance energy transfer from a site-specifically attached uorescent chromophore reveals the folding of the N-terminal domain of CP29. Chem Phys 357: 113–119.

27. Yefimova SL, Gural'chuk GY, Sorokin AV, Malyukin YV, Borovoy IA, et al. (2008) Hydrophobicity effect on interactions between organic molecules in nanocages of surfactant micelle. J Appl Spectrosc 75: 658–663.

28. Sanborn ME, Connolly BK, Gurunathan K, Levitus M (2007) Fluorescence properties and photophysics of the sulfoindocyanine Cy3 linked covalently to DNA. J Phys Chem B 111: 11064–11074.

29. Ranjit S, Gurunathan K, Levitus M (2009) Photophysics of backbone uorescent DNA modifications: reducing uncertainties in FRET. J Phys Chem B 113: 7861–7866.

30. Kalinin S, Sisamakis E, Magennis SW, Felekyan S, Seidel CAM (2010) On the origin of broadening of single-molecule FRET effciency distributions beyond shot noise limits. J Phys Chem B 114: 6197–6206.

31. Schuler B, Lipman EA, Steinbach PJ, Kumke M, Eaton WA (2005) Polyproline and the spectroscopic ruler revisited with single-molecule uorescence. Proc Natl Acad Sci U S A 102: 2754–2759.

32. Lakowicz CDGJR, ed (2006) Reviews in Fluorescence / Annual volumes 2006 (Reviews in Fluorescence). Springer, 1 edition.

33. Rothwell PJ, Berger S, Kensch O, Felekyan S, Antonik M, et al. (2003) Multiparameter single-molecule uorescence spectroscopy reveals heterogeneity of HIV-1 reverse transcriptase: primer/template complexes. Proc Natl Acad Sci U S A 100: 1655–1660.

34. Schröder GF, Alexiev U, Grubmüller H (2005) Simulation of uorescence anisotropy experiments: Probing protein dynamics. Biophys J 89: 3757 - 3770.

35. Henry ER, Hochstrasser RM (1987) Molecular dynamics simulations of uorescence polarization of tryptophans in myoglobin. Proc Natl Acad Sci U S A 84: 6142–6146.

36. Rindermann JJ, Akhtman Y, Richardson J, Brown T, Lagoudakis PG (2011) Gauging the exibility of uorescent markers for the interpretation of uorescence resonance energy transfer. J Am Chem Soc 13: 1857–1871.

37. Corry B, Jayatilaka D, Martinac B, Rigby P (2006) Determination of the orientational distribution and orientation factor for transfer between membrane-bound uorophores using a confocal microscope. Biophys J 91: 1032–1045.

38. VanBeek DB, Zwier MC, Shorb JM, Krueger BP (2007) Fretting about FRET: correlation between kappa and R. Biophys J 92: 4168–78.

39. Di Fiori N, Meller A (2010) The effect of dye-dye interactions on the spatial resolution of singlemolecule FRET measurements in nucleic acids. Biophys J 98: 2265–2272.

40. Corry B, Jayatilaka D (2008) Simulation of structure, orientation, and energy transfer between AlexaFluor molecules attached to MscL. Biophys J 95: 2711–2721.

41. Dolghih E, Ortiz W, Kim S, Krueger BP, Krause JL, et al. (2009) Theoretical studies of short polyproline systems: Recalibration of a molecular ruler. J Phys Chem A 113: 4639–4646.

42. Speelman AL, Munosz-Losa A, Hinkle KL, VanBeek DB, Mennucci B, et al. (2011) Using molecular dynamics and quantum mechanics calculations to model uorescence observables. J Phys Chem A in press.

43. Unruh JR, Kuczera K, Johnson CK (2009) Conformational heterogeneity of a leucine enkephalin analogue in aqueous solution and sodium dodecyl sulfate micelles: comparison of time-resolved FRET and molecular dynamics simulations. J Phys Chem B 113: 14381–14392.

44. Schuetz P, Wuttke R, Schuler B, Caisch A (2010) Free energy surfaces from single-distance information. J Phys Chem B 114: 15227–15235.

45. Best RB, Merchant KA, Gopich IV, Schuler B, Bax A, et al. (2007) Effect of exibility and cis residues in single-molecule FRET studies of polyproline. Proc Natl Acad Sci U S A 104: 18964–18969.

46. Berger A, Kurtz J, Katchalski E (1954) Poly-L-Proline. J Am Chem Soc 76: 5552–5554.

47. Harrington WF, Sela M (1958) Studies on the structure of poly-L-proline in solution. Biochim Biophys Acta 27: 24–41.

48. Dorman DE, Torchia DA, Bovey FA (1973) Carbon-13 and proton nuclear magnetic resonance observations of the conformation of poly(L-proline) in aqueous salt solutions. Macromolecules 6: 80–82.

49. Zagrovic B, Lipfert J, Sorin EJ, Millett IS, van Gunsteren WF, et al. (2005) Unusual compactness of a polyproline type II structure. Proc Natl Acad Sci U S A 102: 11698–11703.

50. Doose S, Neuweiler H, Barsch H, Sauer M (2007) Probing polyproline structure and dynamics by photoinduced electron transfer provides evidence for deviations from a regular polyproline type II helix. Proc Natl Acad Sci U S A 104: 17400–17405.

51. Hoffmann A, Nettels D, Clark J, Borgia A, Radford S, et al. Quantifying heterogeneity and conformational dynamics from single molecule FRET of diffusing molecules: Recurrence analysis of single particles (RASP). Phys Chem Chem Phys, (in press).

52. Panchuk-Voloshina N, Haugland RP, Bishop~nStewart J, Bhalgat MK, Millard PJ, et al. (1999) Alexa dyes, a series of new uorescent dyes that yield exceptionally bright, photostable conjugates. J Histochem Cytochem 47: 1179–1188.

53. Cownan PM, McGavin S (1955) Structure of poly-L-proline. Nature 176: 501–503.

54. Moradi M, Babin V, Roland C, Darden TA, Sagui C (2009) Conformations and free energy landscapes of polyproline peptides. Proc Natl Acad Sci U S A 106: 20746–20751.

55. Grathwohl C, Wüthrich K (1981) NMR studies of the rates of proline cis-trans isomerization in oligopeptides. Biopolymers 20: 2623–2633.

56. Jorgensen WL, Madura JD (1985) Temperature and size dependence of Monte Carlo simulations of TIP4P water. Mol Phys 56: 1381–1392.

57. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118: 11225–11236.

58. Carlson HA, Nguyen TB, Orozco M, Jorgensen WL (1993) Accuracy of free-energies of hydration for organic molecules from 6-31G-asterisk-derived partial charges. J Comput Chem 14: 1240–1249.

59. Singh UC, Kollman PA (1984) An approach to computing electrostatic charges for molecules. J Comput Chem 5: 129–145.

60. Frisch MJ, Trucks GW, Schlegel HB, Scuseria GE, Robb MA, et al. Gaussian 03, B.04. Gaussian, Inc.,Wallingford, CT, 2004.

61. Berendsen HJC, van der Spoel D, van Drunen R (1995) GROMACS - a message-passing parallel molecular-dynamics implementation. Comput Phys Commun 91: 43–56.

62. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for highly effcient, load-balanced, and scalable molecular simulation. J Chem Theory Comput 4: 435–447.

63. Van der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, et al. (2005) GROMACS: fast, exible, and free. J Comput Chem 26: 1701–1718.

64. Feenstra KA, Hess B, Berendsen HJC (1999) Improving effciency of large time-scale molecular dynamics simulations of hydrogen-rich systems. J Comput Chem 20: 786–798.

65. Berendsen HJC, Postma JPM, van Gunsteren WF, Dinola A, Haak JR (1984) Molecular-dynamics with coupling to an external bath. J Chem Phys 81: 3684–3690.

66. Bussi G, Donadio D, Parrinello M (2007) Canonical sampling through velocity rescaling. J Chem Phys 126: 014101.

67. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 52: 7182–7190.

68. Hess B, Bekker H, Berendsen HJC, Fraaije JGEM (1997) LINCS: A linear constraint solver for molecular simulations. J Comput Chem 18: 1463–1472.

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

70. Schuler B (2007) Application of single molecule förster resonance energy transfer to protein folding. Methods Mol Biol 350: 115–138.

71. Braslavsky SE, Fron E, Rodrguez HB, Romn ES, Scholes GD, et al. (2008) Pitfalls and limitations in the practical use of förster's theory of resonance energy transfer. Photochem Photobiol Sci 7: 1444–1448.

72. Schuler B, Lipman EA, Eaton WA (2002) Probing the free-energy surface for protein folding with single-molecule uorescence spectroscopy. Nature 419: 743–747.

73. Rusinova E, Tretyachenko-Ladokhina V, Vele OE, Senear DF, Ross JBA (2002) Alexa and Oregon green dyes as uorescence anisotropy probes for measuring protein-protein and protein-nucleic acid interactions. Anal Biochem 308: 18–25.

74. Krueger BP, Scholes GD, Fleming GR (1998) Calculation of couplings and energy-transfer pathways between the pigments of LH2 by the ab initio transition density cube method. J Phys Chem B 102: 9603–9604.

75. Matsumoto M, Nishimura T (1998) Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. ACM Trans Model Comput Simul 8: 3–30.

76. Saito M, Matsumoto M (2008) SIMD-oriented fast mersenne twister: a 128-bit pseudorandom number generator. In: Keller A, Heinrich S, Niederreiter H, eds (2008) Monte Carlo and Quasi-Monte Carlo Methods 2006, Springer Berlin Heidelberg. pp 607–622.

77. Brune R, Doose S, Sauer M (2007) Analyzing the inuence of contact-induced quenching processes on Förster resonance energy transfer (Proceedings Paper). In: Popp J, von Bally G, eds. Biophotonics 2007: Optics in Life Science volume 6633 of Proceedings of SPIE;doi:10.1117/12.727853.

78. Antonik M, Felekyan S, Gaiduk A, Seidel CAM (2006) Separating structural heterogeneities from stochastic variations in uorescence resonance energy transfer distributions via photon distribution analysis. J Phys Chem B 110: 6970–6978.

79. Kalinin S, Felekyan S, Antonik M, Seidel CAM (2007) Probability distribution analysis of singlemolecule uorescence anisotropy and resonance energy transfer. J Phys Chem B 111: 10253–10262.

80. Gopich I, Szabo A (2005) Theory of photon statistics in single-molecule forster resonance energy transfer. J Chem Phys 122: 14707.

81. Gopich IV, Szabo A (2007) Single-molecule FRET with diffusion and conformational dynamics. J Phys Chem B 111: 12925–12932.

82. Venkatachalam CM, Price BJ, Krimm S (1975) A theoretical estimate of the energy barriers between stable conformations of the proline dimer. Biopolymers 14: 1121–1132.
83. Gopich IV, Szabo A (2009) Decoding the pattern of photon colors in single-molecule FRET. J Phys Chem B 113: 10965–10973.
84. Scholes GD (2003) Long-range resonance energy transfer in molecular systems. Annu Rev Phys Chem 54: 57–87.
85. Feoktistov V (2006) Differential Evolution: In Search of Solutions. Springer Science+Business Media.
86. Gopich IV, Szabo A (2010) FRET effciency distributions of multistate single molecules. J Phys Chem B 0.
87. Nir E, Michalet X, Hamadani KM, Laurence TA, Neuhauser D, et al. (2006) Shot-noise limited single-molecule FRET histograms: comparison between theory and experiments. J Phys Chem B 110: 22103–22124.
88. Torella JP, Santoso Y, Holden SJ, Hohlbein J, Joyce CM, et al. (2010) Separating static and dynamic heterogeneity in single-molecule FRET experiments with burst variance analysis (BVA). Biophys J 98: 591.

– P8 – *155*

Interaction of β-Sheet Folds with a Gold Surface

Martin Hoefling[1], Susanna Monti[2], Stefano Corni[3], Kay Eberhard Gottschalk[4]*

1 Theoretical and Computational Biophysics, Max Planck Institute for Biophysical Chemistry, Göttingen, Germany, 2 CNR Istituto di Chimica dei Composti OrganoMetallici (ICCOM), Pisa, Italy, 3 Centro S3, CNR Istituto Nanoscienze, Modena, Italy, 4 ZIK HIKE Centre for Humoral Immune Reactions in Cardiovascular Disease, Universität Greifswald, Greifswald, Germany

Abstract

The adsorption of proteins on inorganic surfaces is of fundamental biological importance. Further, biomedical and nanotechnological applications increasingly use interfaces between inorganic material and polypeptides. Yet, the underlying adsorption mechanism of polypeptides on surfaces is not well understood and experimentally difficult to analyze. Therefore, we investigate here the interactions of polypeptides with a gold(111) surface using computational molecular dynamics (MD) simulations with a polarizable gold model in explicit water. Our focus in this paper is the investigation of the interaction of polypeptides with β-sheet folds. First, we concentrate on a β-sheet forming model peptide. Second, we investigate the interactions of two domains with high β-sheet content of the biologically important extracellular matrix protein fibronectin (FN). We find that adsorption occurs in a stepwise mechanism both for the model peptide and the protein. The positively charged amino acid *Arg* facilitates the initial contact formation between protein and gold surface. Our results suggest that an effective gold-binding surface patch is overall uncharged, but contains *Arg* for contact initiation. The polypeptides do not unfold on the gold surface within the simulation time. However, for the two FN domains, the relative domain-domain orientation changes. The observation of a very fast and strong adsorption indicates that in a biological matrix, no bare gold surfaces will be present. Hence, the bioactivity of gold surfaces (like bare gold nanoparticles) will critically depend on the history of particle administration and the proteins present during initial contact between gold and biological material. Further, gold particles may act as seeds for protein aggregation. Structural re-organization and protein aggregation are potentially of immunological importance.

Introduction

The interaction of inorganic surfaces with biomolecules like peptides and proteins is central to biotechnology [1], being involved, e.g., in biosensors, biomaterials [2] or the biological use of nanoparticles [3]. Further, aspects of nanotoxicity of inorganic particles may be based on particle-protein interactions [4,5]. Despite such importance, the microscopic understanding of how proteins interact with inorganic surfaces is still limited. Besides relatively well-characterized bonds such as *Cys* on gold [6], little is known about the exact nature of the interaction. However, the interaction between proteins and inorganic surfaces is based on well-defined principles as demonstrated by the ability to genetically engineer proteins specific for and discriminating between given surfaces [7]. Yet, technical limitations hamper the detailed structural and dynamical experimental characterization of biomolecular interactions with surfaces [8]. Computational methods, on the other hand, are well-suited to atomistically describe the effect of surfaces on biomolecules [9,10,11]. One important biocompatible metal widely used in biotechnological application is gold [12]. Being a metal, gold polarization effects are important and realistic simulations need to include them. Only recently, force fields suitable for this task became available [13]. Using such a force

field in earlier work, we analyzed the adsorption of single amino acids (the 20 natural ones) onto gold surfaces [14,15,16]. However, the complex cooperative behavior of a polypeptide [17,18,19,20,21,22] may significantly alter the adsorption properties. Therefore, our focus in this work is to analyze the association of a complex biomolecule with a gold(111) surface. Our earlier results indicated that amino acids with an intrinsic propensity to form β-sheets are predisposed to interact with gold surfaces [15]. We therefore concentrate here on simulating the adsorption of two polypeptides with high β-sheet content using a polarizable gold(111) surface as parameterized in the force field GolP [13]. First, we simulated by molecular dynamics an oligopeptide (RAD16II) forming a fibrillious structure. It is a spontaneous self-assembling amphiphilic peptide, known to form hydrogel-like matrices that support cell attachment, proliferation and differentiation [23,24]. The structure of the elastic properties of filaments formed by RAD16II have been investigated by atomistic simulations [25]. This peptide has been suggested as a coating to make inorganic surfaces biocompatible, and the interaction of RAD16II fibers with surfaces other than gold (TiO₂) has been studied by classical MD simulations [24]. These studies showed that the fibrous structure is maintained in typical MD simulation times Fibrillious structures are important for many protein-folding

related diseases like Alzheimer or bovine spongiforme encephalopathy (BSE). The peptide absorbs quickly in our simulations, is stable on the gold surface and shows no tendency of unfolding. For this peptide, which is rich in *Arg*, we characterize how the *Arg* side chain can penetrate the water easily, creating a stable initial anchoring points.

Of particular importance, in terms of nanotechnological applications, is the interaction of a gold surface with members of the extracellular matrix (ECM). Not only are ECM proteins used as anchors for cells on gold electrodes, these proteins will also be the first encountered by gold nanoparticles administered to organisms [26,27]. One of the most important members of the ECM proteins is Fibronectin. This protein is involved in cell migration, adhesion, and metastasis [28]. The large size of Fibronectin and the lack of a full-length atomistic structure of this protein prevent the simulation of the adsorption of the whole protein on the gold surface. However, a crucial part of this protein is composed of the FibronectinIII modules 9 and 10 (FNIII9–10). FNIII9–10 contains the RGD binding site in FNIII10 as well as a synergy site in FNIII9, important for integrin-mediated cell adhesion and force-initiated signalling [29]. The structure of the two domains is known and the modest size of this fragment allows atomistic simulations in a reasonable time frame. Due to the biological importance of FNIII9–10, appreciating its interaction with gold not only fosters our basic understanding of biomolecule-metal interactions in general, but has technological and health-care implications. Therefore, we report here long atomistic molecular dynamics simulations of the type III fibronectin domains 9 and 10 (FNIII9–10) on the polarizable gold(111) surface in explicit water.

We show that gold strongly interacts with FNIII9–10, but does not lead to unfolding of this β-sheet rich protein within the simulation time. Also in this case, *Arg* plays an important role for penetrating the dense water layer covering the hydrophilic gold surface. In *in silico* mutational studies, we show that an *Arg→Lys* mutation does not alter the general adsorption pattern, while an *Arg→Val* significantly changes the proteins behavior, underlining the importance of charges for the adsorption process. Our results have implications for assessing possible health risks associated with gold particles and may serve as a starting point for the design of gold binding proteins.

Results and Discussion

Model peptide

First, we have performed a simulation of a β-sheet like structure formed by the oligopeptide RAD16II. This is a 16-mer peptide with a RARADADARARADADA sequence that spontaneously self-assembles in anti-parallel β-sheet fibers [30,31]. Hence, it is a relevant model system for other β-sheet forming peptides like the Alzheimer precursor protein. Further, RAD16II represents an ideal system to study the basics of β-sheet forming peptide adsorption. With *Arg* and *Asp* in proximity, it offers two competing interaction partners with opposite charge. However, it is overall uncharged. Its rigid, planar structure facilitates the analysis, in particular in comparison to the complex surface topology of a multi-domain protein. With its β-sheet structure, it is also a good simplified representation of the β-sheet fold of Fibronectin.

The *Arg*-rich β-sheet quickly replaces water molecules, thus forming strong peptide-surface contacts. However, water replace-
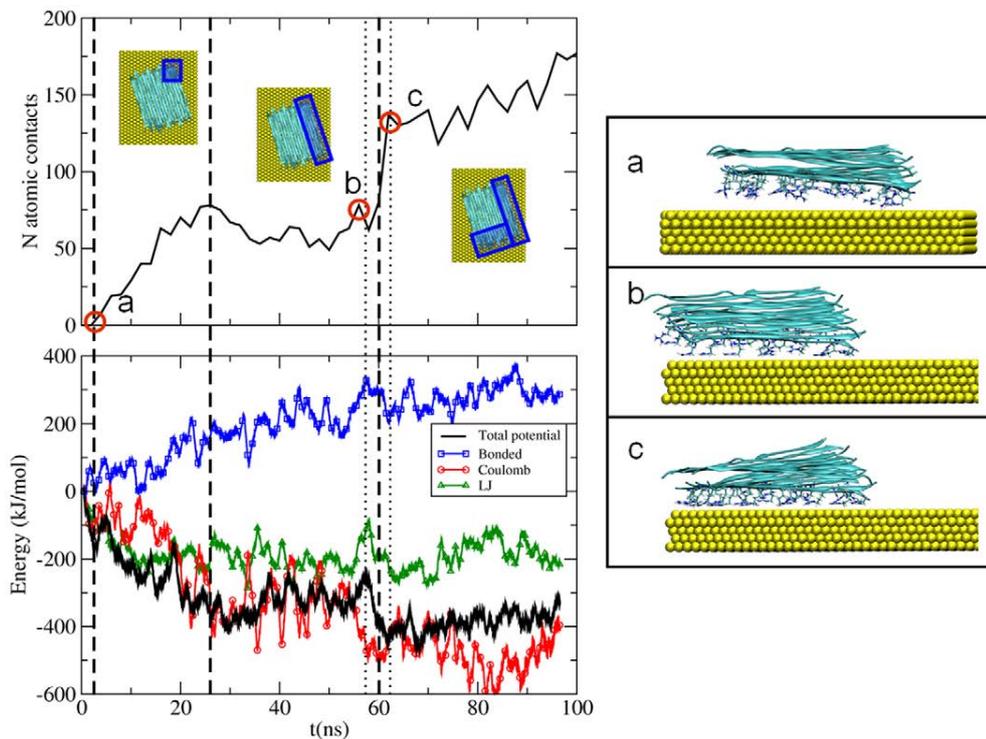

Figure 1. Evolution of the atomic contact number between the fiber and the Au surface, and of the corresponding adsorption energy. In the right panels, relevant snapshots from the dynamics are shown. The blue box in the top view indicates the adsorbed parts with progressing simulation time.
doi:10.1371/journal.pone.0020925.g001

ment and peptide adsorption are not a continuous process despite the symmetric and regular structure of the peptide. The adsorption process can be divided into four different phases (Fig. 1). The first contact is formed after ~3 ns. Then, a gradual increase in the number of contacts takes place during the subsequent 20 ns leading to the adsorption of all the *Arg* residues of the lateral oligopeptide molecule. After a ~35 ns plateau, the sharp increase in the number of contacts in the interval 58–60 ns corresponds to the adhesion of the edge of the β-sheet perpendicular to the adsorbed portion of the molecule (Fig. 1B and 1C). The sharpness of the transition indicates that the whole edge is cooperatively getting into contact with the surface. Cooperativity is enforced by the H-bond network between the chains, which fosters a zipping-like mechanism. After this jump, the number of contacts gradually increases till the end of the simulation. No sign of structural changes in the β-sheet is seen during the dynamics.

The evolution of the adsorption energy for the peptides approaching the surface is also shown in Fig. 1, so as to be compared with the number of atomic contacts. The various energy components (bonding, electrostatics and Lennard Jones, LJ) are depicted as well. Since the contact of the peptide with the surface is steadily increasing with time (besides local fluctuations), the energy curve is representative of the adsorption energy landscape of the peptide while it settles on the surface, neglecting entropic contributions. The profile of the adsorption energy, although fluctuacting, is clearly correlated with the number of atomic contacts. The increase in the number of contacts during the 0~25 ns time interval is accompanied by a steady decrease in the adsorption energy. This was initially due to the Lennard Jones term (~20 ns) and then also to electrostatic interactions. From 25 to 58 ns (i.e., the dotted line) the number of contacts has a plateau and the potential energy is almost constant. Just before the number of contacts ramps up (i.e., before the dotted line at 58 ns), there is a decrease in the electrostatic energy and an increase in both bonding and LJ terms, suggesting that during this period the system behavior is guided by the electrostatic interactions, which
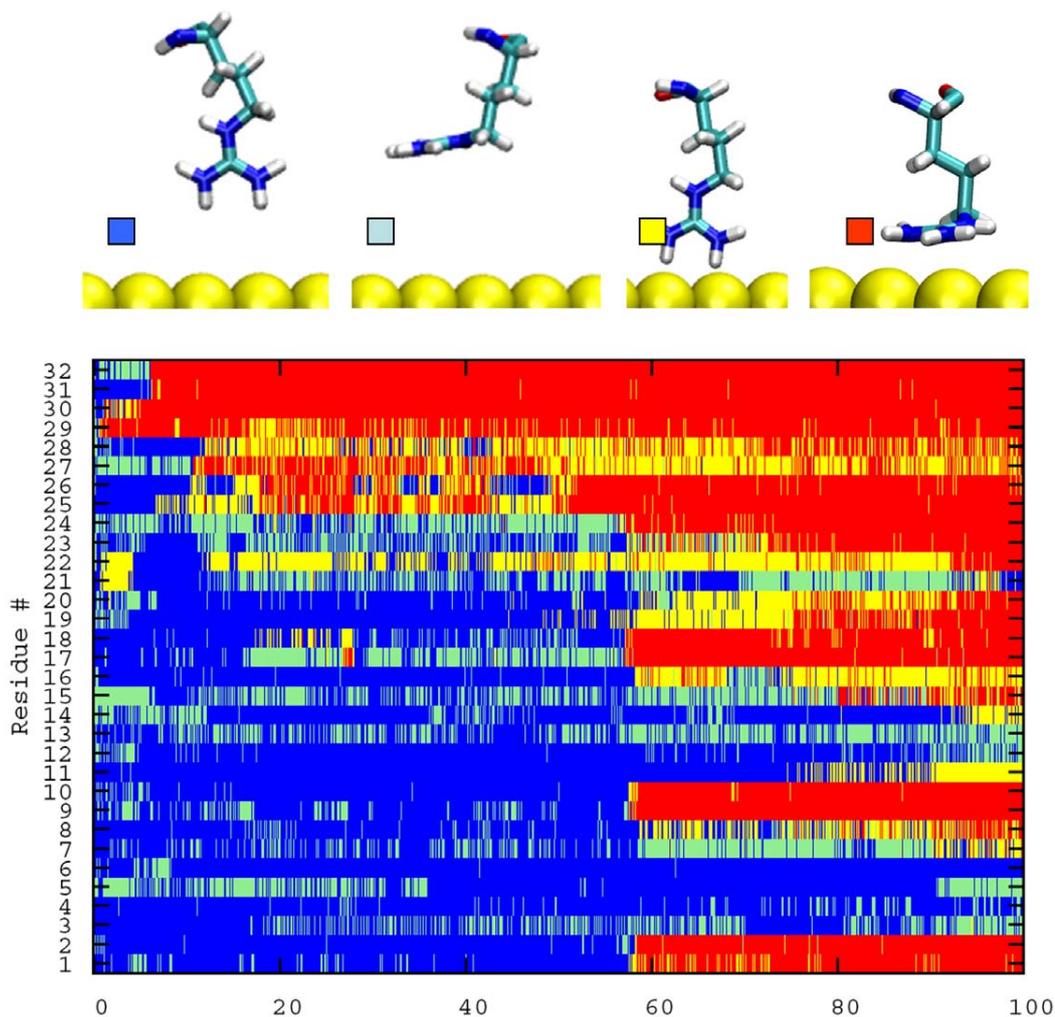

Figure 2. Contact and orientation map for the 32 gold-exposed *Arg* of the RAD16II peptide. The color code (defined in the upper panels) indicates whether *Arg* is in contact with the surface or not, and whether its side chain orientation is perpendicular or parallel to the surface. *Arg* is defined to be in contact when any of its atoms is within 0.35 nm of the surface. Side chain is defined to be parallel to the surface when all the guanidinium C-N bonds have angles with the surface smaller than 30°, otherwise orientation is called perpendicular. If a looser definition of parallel orientation is used (angles smaller than 45° instead of 30°), the number of first contacts taking place in the perpendicular orientation is still >70% of the total.
doi:10.1371/journal.pone.0020925.g002

temporarily induce the formation of an higher-energy structure (the total potential energy has a positive fluctuation). However, this conformation rapidly relaxes (region between the dotted lines) via a rearrangement of the sidechains that corresponds to the sharp increase in the number of atomic contacts described above. After this step, further stabilization of the system configuration is achieved through a steady increase in the atomic contacts which is clearly evidenced after point c. Notably, the bonding term (blue) is increasing throughout the dynamics, in order to maximize the interaction with the surface. Indeed, the internal geometry of the peptide gradually departs from the optimal conformation adopted by the isolated molecules.

Strikingly, the interactions of *Arg* side chains with the gold layer are the driving force of peptide-surface the association. The electrostatic energy minimum that accompanies the sharp increase of atomic contacts at 58–62 ns corresponds to several *Arg* finding their way to the surface (Fig. 2). In all the described adsorption steps, the *Arg* capability to penetrate the water layer is fundamental, forming the initial peptide-gold contact. In >80% of first contact events, the *Arg* side chain is perpendicular to the surface (Fig. 2), minimizing the number of water molecules to be displaced (Fig. 3). Afterwards, the side chain changes its conformation becoming parallel to the gold surface. Such an adsorption geometry has already been observed for an isolated *Arg*, competing with a second geometry with an adsorbed backbone. The observed orientation is the preferred orientation found for methylguanidinium (i.e., the molecule representing the *Arg* side-chain) in water [16]. Our simulations strongly indicate that *Arg* may play a dominant role in the adsorption of polypeptides.

Fibronectin Close to Surface

As the next level of complexity, we simulated the last steps of adsorption of FNIII9–10. To this end, we placed the protein in eight different orientations over the gold surface with approximately one water layer between FN9–10 and the metal surface (Fig. 4) and monitored the protein-gold distances over a simulation of 100 ns per orientation. With these simulations, we save computational time due to the reduced need to simulate the

diffusional encounter of the protein with the surface. This allows us to sample many different protein-gold orientations. Effects of diffusion are considered later in this work. Well-defined contact areas are formed quickly (Fig. 5), being stable over the whole length of the simulation. Within the 100 ns simulation, no significant re-arrangement of the initial adsorption geometry occurs. The contacts between the protein and the metal surface do not break once formed (Fig. 5). This demonstrates the strong interaction of the protein with the surface. In a variety of Atomic Force Microscopy experiments, it has been demonstrated that adsorption of proteins on surface leads to interactions that are so strong that the individual protein domains unfold rather than detach from the surface when force is applied, in line with our results here [32,33]. Our results furthermore indicate that once adsorbed, the proteins are in a kinetically trapped state (translational and rotational), which prevents the reorientation to a potentially more optimized adsorption geometry. However, although our simulations of altogether more than one µs approach experimental time frames, slow re-orientation processes on longer timescales may occur [34,35].

Despite the strong interaction with the gold surface, the protein does not unfold on the gold surface within the simulation time in either of the simulations. The β-sheet content is virtually unchanged (Fig. 6 [36]), and no major conformational changes within each domain occur. The Cα RMSD of the individual domains remains below 3 Å (Fig. 7 and 8). However, the total root mean square deviation (RMSD) of the two domains together shows drastic changes, indicative of the gold-induced remodeling of the domain-domain orientation. The stepwise increase in the total RMSD indicates collective structural transitions between different energy minima, and the resulting relative domain orientation is rigid.

Although the individual domains remain rather stable, the protein adapts to the gold surface. The originally roundish surface of the protein becomes planar at the site of adsorption, constituting something like an induced-fit mechanism.

Like the model peptide, FNIII9–10 has a very high β-sheet content. Based on simulations of individual amino acids on gold surfaces [15], we proposed earlier that gold surface might in fact
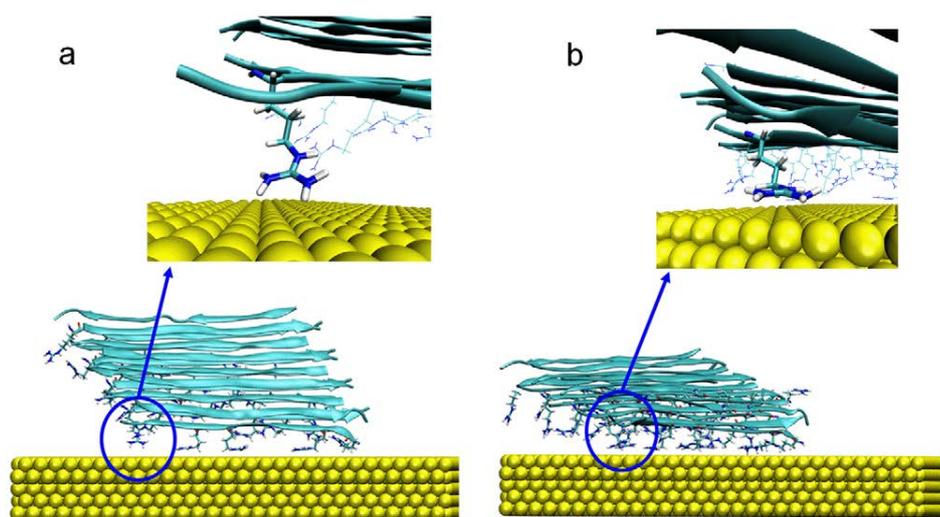

Figure 3. Switching of *Arg* side-chain conformation during the association process. *Arg* contacts the surface having the plane of the side chain perpendicular to it (Fig. 3a). After a few ns, the plane of the methylguanidinum moiety has changed its orientation becoming parallel to the surface.
doi:10.1371/journal.pone.0020925.g003
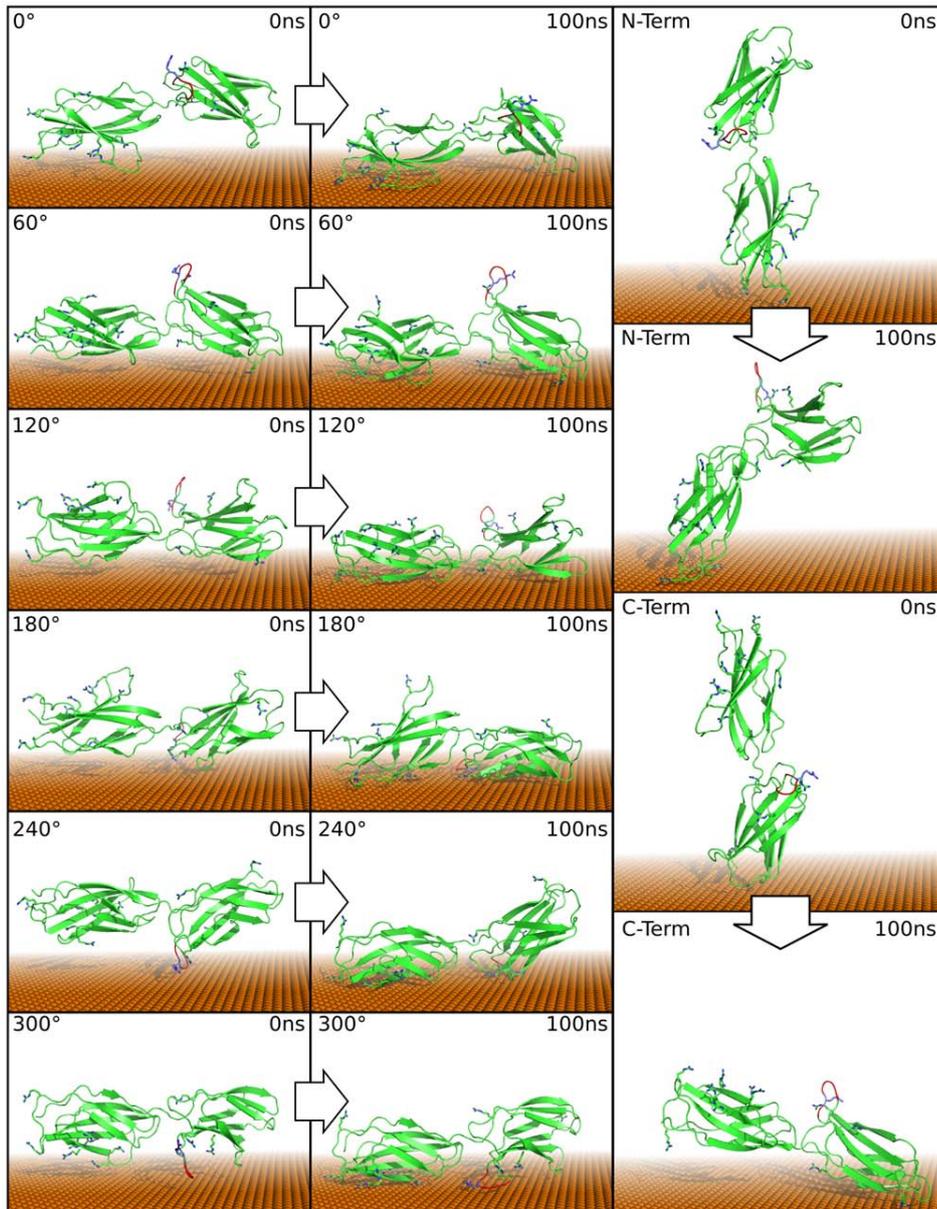
Figure 4. Starting and final conformations of FN. On the left side, the six rotations of the conformation parallel to the surface are shown while the right side depicts the perpendicular conformations with N- or C-Terminus close to the surface. The integrin-binding RGD-loop is indicated in red. Domain 9&10 from 1FNF were used for the simulations. First the principal axis was aligned parallel to the surface. Then, stepwise rotations of 60° around the principal axis produced the initial conformations. The obtained structures were separated by one water layer from the surface.
doi:10.1371/journal.pone.0020925.g004

stabilize β-sheets, in agreement with the stable individual domain folds observed here.

The behavior of α-helical domains on gold may be different and will be the focus of future studies.

Inclusion of Diffusion

As a next step, we were also interested in the diffusional states of FN further away from the protein. To investigate this, we placed FNIII(9–10) between two gold surfaces with a separation of 8 Å to one of the gold surfaces resulting in a distance of 23 Å to the other gold surface. This allows the protein to freely choose on which side to adsorb and to change the relative orientation of the protein

relative to the gold surface during the diffusional phase. We simulated the protein starting from identical conformations but from two different sets of randomly Maxwell distributed atomic velocities. Due to the chaotic behavior of many particle systems, simulation trajectories with different initial atomic velocities quickly diverge in the sub nanosecond regime.

Fast Adsorption

In both cases, FNIII(9–10) quickly adsorbed on the gold surface. However, despite the identical starting conformation, the association times and geometries vary significantly, demonstrating that the protein is still in a diffusional state at the start of the

0	10	20	30	40	50	60
GLDSPTGIDF	SDITANSFTV	HWIAPRATIT	GYRIRHHPEH	FSGRPREDRV	PHSRNSITLT	NLTPGTEYaVV

70	80	90	100	110	120	130
SIVALNGREE	SPLLIGQQST	VSDVPRDLEV	VAATPTSLLI	SWDAPAVTVR	YYRITYGETG	GNSPVQEFTV

140	150	160	170	180		
PGSKSTATIS	GLKPGVDYTI	TVYAVTGRGD	SPASSKPISI	NYRT		
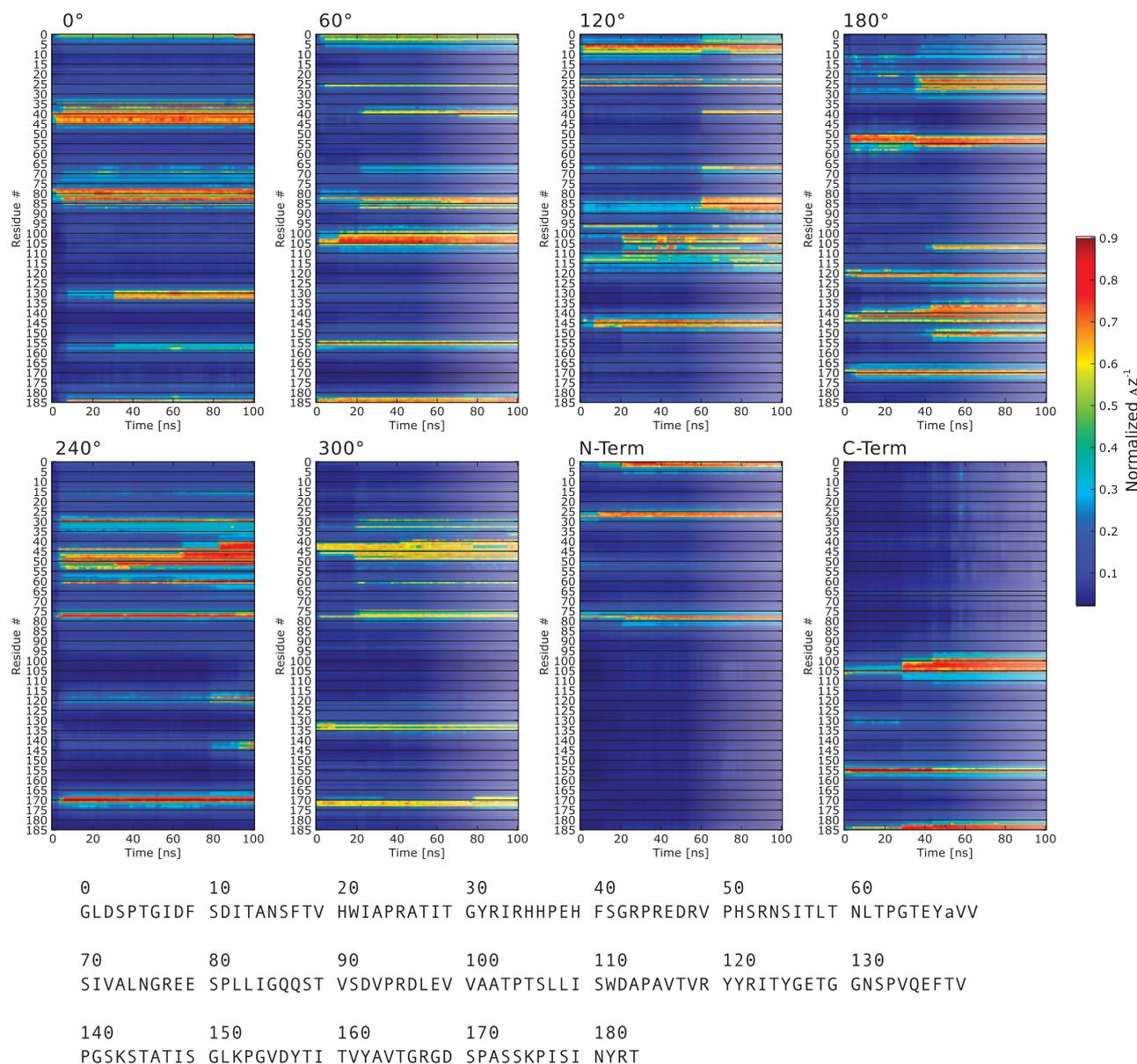
Figure 5. Time resolved adsorption of FN residues. The distance from closest atom of each residue from the surface is shown during each 100 ns simulation color coded according to Δz^{-1} where 1.0 corresponds to the minimum distance and 0 to ∞. Residue # refers to the protein sequence reported below.
doi:10.1371/journal.pone.0020925.g005

simulation. This behavior is consistent with free energy profiles calculated for the adsorption of single amino acids to the gold surface. There, strong attraction was only observed at distances below 8 Å [15,16].

The fast association reaches the diffusional limit. From simulations in water, we found that the 3D diffusion constant is D = 0.1031 (+/− 0.0215) 1e-5 cm^2/s. This translates to a mean travelled length of L = 0.786 nm within 1 ns. In the simulations with a separation of 0.8 nm from the gold surface, the first contact times were 0.4 ns and 21 ns. Thus, in one of the two simulations, the adsorption time is in fact faster than expected for random diffusion. It indicates that no major energy barriers exist on the adsorption trajectory. Earlier results on the adsorption of **individual** amino acids support that the adsorption of

biomolecules on gold surfaces is energetically predominantly a downhill process with only a small dewetting barrier. The comparable behavior of the complex biomolecule shows that no energetically costly conformational changes need to occur for the adsorption. The fast adsorption is further indicative of a rather unspecific binding interface. Once adsorbed, no desorption is observed within our simulation time. Further, also in this case we observe no major rearrangements of the relative orientation of the gold and the protein once adsorbed.

Physicochemical Properties of the Protein Patch Forming the Initial Contact

After formation of a first stable contact, the protein does not diffuse away from the gold surface. Hence, the initial contact
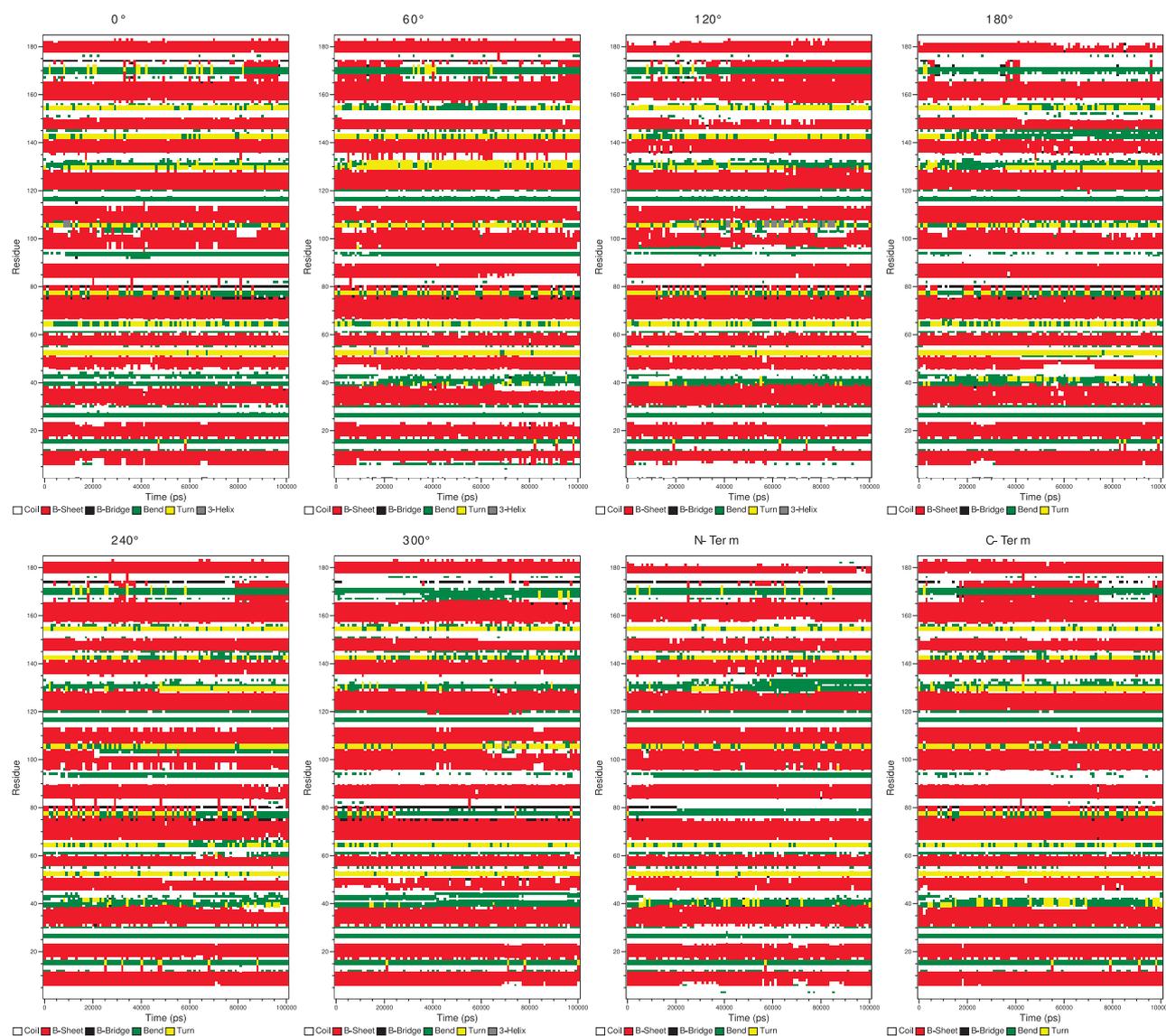
Figure 6. Secondary structures changes during simulation time. No significant conversions could be observed. The secondary structure maps were created with do_dssp, the Gromacs interface to dssp.
doi:10.1371/journal.pone.0020925.g006

between protein and surface is the determining step in the adsorption process. Understanding the physicochemical properties of this patch may allow identifying putative gold binding sites in other proteins without reverting to time-consuming atomistic MD simulations. Two different surface patches form this initial contact in our simulations. This underlines the rather unspecific nature of the adsorption process. To analyze the physicochemical properties of the contacting surface patches, we mapped the electrostatic potential of proteins as calculated with an adaptive Poisson-Boltzman approach [37] on the solvent accessible surface. Patches with electrostatic potential close to zero, surrounded by area with potential values deviating from zero, form this initial contact (Fig. 9). Earlier results on the free energy of adsorption of individual amino acids showed a different trend: there, positively charged amino acids formed stronger interactions than apolar amino acids [15]. Different reasons may contribute to reconcile these apparently contradicting findings. The first and more

fundamental is that the adsorption free energy of the amino acid is a thermodynamic quantity, while the present results for the entire proteins refer to the dynamic behavior (i.e., to the kinetics) of the system. To bring a protein patch in contact with the surface, it is necessary to desolvate such patch. Therefore, a dewetting barrier should be overcome. Such a barrier, which can be seen even for single amino acids [15], will be larger for protein patches that bound water more tightly, e.g, patches at potentials far from zero. As a result, patches with a potential far from zero will dewett more slowly. Hence, the here observed first contacts may be determined by the low dewetting barrier of patches with a potential close to zero.

Of course similar effects related to desolvation may modify the adsorption thermodynamics for the whole protein, not just the kinetics. In fact, in the earlier thermodynamics studies on single amino acids, the charged side chain in proximity to the surface is still exposed to solvent. Hence, no major desolvation penalty is
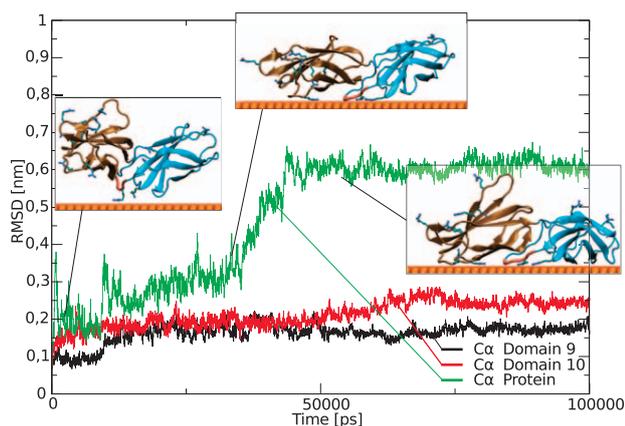
Figure 7. Cα RMSD of the two domains and the entire protein. The first inlet shows the initial adsorption while the others depict the twisting of the domains in the region between 35 ns and 44 ns. *Arg* are shown in licorice representation. See Fig. 8 for all simulations.
doi:10.1371/journal.pone.0020925.g007

paid for individual charged amino acids, while simultaneously attractive interactions between the amino acid charge and the polarizable gold surface are possible. In our studies of a whole protein reported here, the sidechains are shielded from the solvent by the neighboring amino acids once adsorbed. This shielding effect may tip the balance so that the interactions between polarizable gold surface and the charged amino acids together with the entropic gain obtained from freeing the water molecules bound to the gold surface and the charged surface patches cannot compensate for the loss of solvation enthalpy.

The physicochemical properties observed here resemble the architecture of protein-protein interaction sites. However, protein-protein interactions are highly specific and the results of long evolutionary process, while the here observed interactions are not physiological and unspecific. Hence, the interactions observed here can be regarded as a model system for protein-protein interactions before evolutionary optimization. This supports the earlier notion that protein-protein interaction sites evolved from unspecific interactions, formed by patches with an intrinsic propensity to aggregate. These unspecific interaction sites were then optimized by mechanisms inducing shape and charge complementarity. This is facilitated by the plasticity of both interaction partners and the complexity of a protein surface, allowing for a high variety in shapes. The challenge for designing specific protein-to-surface interaction lies in the fact that the surface is flat, isotropic and not adaptive, so that only one interaction partner is available for evolutionary optimization.

Adsorption is a step-wise process

Analyzing the contact times of individual amino acids of the protein with the gold surface shows a step-wise process. First, a contact is initiated by a few amino acids. This is then followed by a very fast cooperative adsorption of a complete surface patch. Despite the more complex nature of the protein, this adsorption sequence is very similar to the adsorption of the model peptide. This suggests that the basic features observed in the model peptide are well preserved in complex systems. The surface patches of FNIII9–10 behave like independent binding units. This hints at a modular architecture of the binding interface. Modular architectures of binding interfaces have previously been

proposed for protein protein interactions, demonstrating a conceptual similarity. This raises the question whether a certain modularity is an intrinsic property of interactions involving proteins, stemming from the self-organization of the polypeptide chain during folding.

Also here *Arg* seems to be overrepresented in the early initiation phase of the interaction, underlining the particular ability of *Arg* to penetrate the final water layer, establishing a first contact and facilitating the dewetting transition.

Binding-Site Reorientation

Binding of cells to fibronectin induces cellular signaling cascades. It has been shown that two binding sites in FNIII(9–10) are of particular importance for cellular force sensing: the main RGD binding site located in FNIII(10) and the synergy-site located in FNIII(9). Since biological recognition processes rely on the fine-tuned relative orientation of different binding partners, it is expected that changes in the relative orientation impact binding and signaling properties. Therefore, we analyzed the relative orientation of the synergy site and the main RGD binding loop as indicated by the dihedral angle distribution between the two binding sites (Fig. 10). Adsorption on the gold surface leads to a narrowing of this distribution compared to simulations in water, indicating a stabilizing effect of the adsorption. Furthermore, adsorption of both domains leads to a relative orientation of the two binding sites only rarely observed in free simulations. The observed binding-site reorientation raises interesting questions about the physiological activity of surface adsorbed multi-domain proteins, even if the individual fold of each domain may structurally not be affected by adsorption. For the here simulated domains FN9–10, it has been suggested that the structural stability of FN9–10 is important for the proper ligation of the protein with the cell-surface receptor integrin A5B1 [38], and that the relative orientation plays an important role in mechanosensing [39,40]. Hence, surface adsorption of adhesion molecules might potentially alter their signaling properties, as already suggested earlier [41]. Further, these reorientations can expose hidden epitopes, potentially causing an immunological threat.

Impact of Mutations

Since our results of both the model peptide and the protein imply an important role for *Arg* in early steps of the adsorption process, we mutated all *Arg* in silico to *Val* or to *Lys*. The *Arg→Lys* mutation retains the charge distribution on the FN surface, while the *Arg→Val* mutation leads to an overall highly negatively charged protein. For both mutants, we performed 100 ns adsorption simulation each. Our results underline the importance of electrostatic properties for the adsorption: while the *Arg→Lys* mutant adsorbs very similar to the WT, the *Arg→Val* mutant needs longer time for adsorption, only adsorbs at the rather uncharged ends of the protein and bridges the two gold layers. Hence, the desolvation penalty of the negatively charged protein is too high to allow fast and strong adsorption. This has strong implications for the design of specifically adsorbing proteins: it is very desirable to create proteins that adsorb in a specific orientation onto the gold surface. For such an effort, comparable to drug design or to the design of protein-protein interaction, two aspects need to be regarded. First, one has to create a surface patch that easily adsorbs onto the surface (design-in). But second and equally important, one has to design the remaining of the surface such that it is unlikely to adsorb (design-out). Our results give first hints for both strategies: overall neutral patches containing *Arg* are preferred
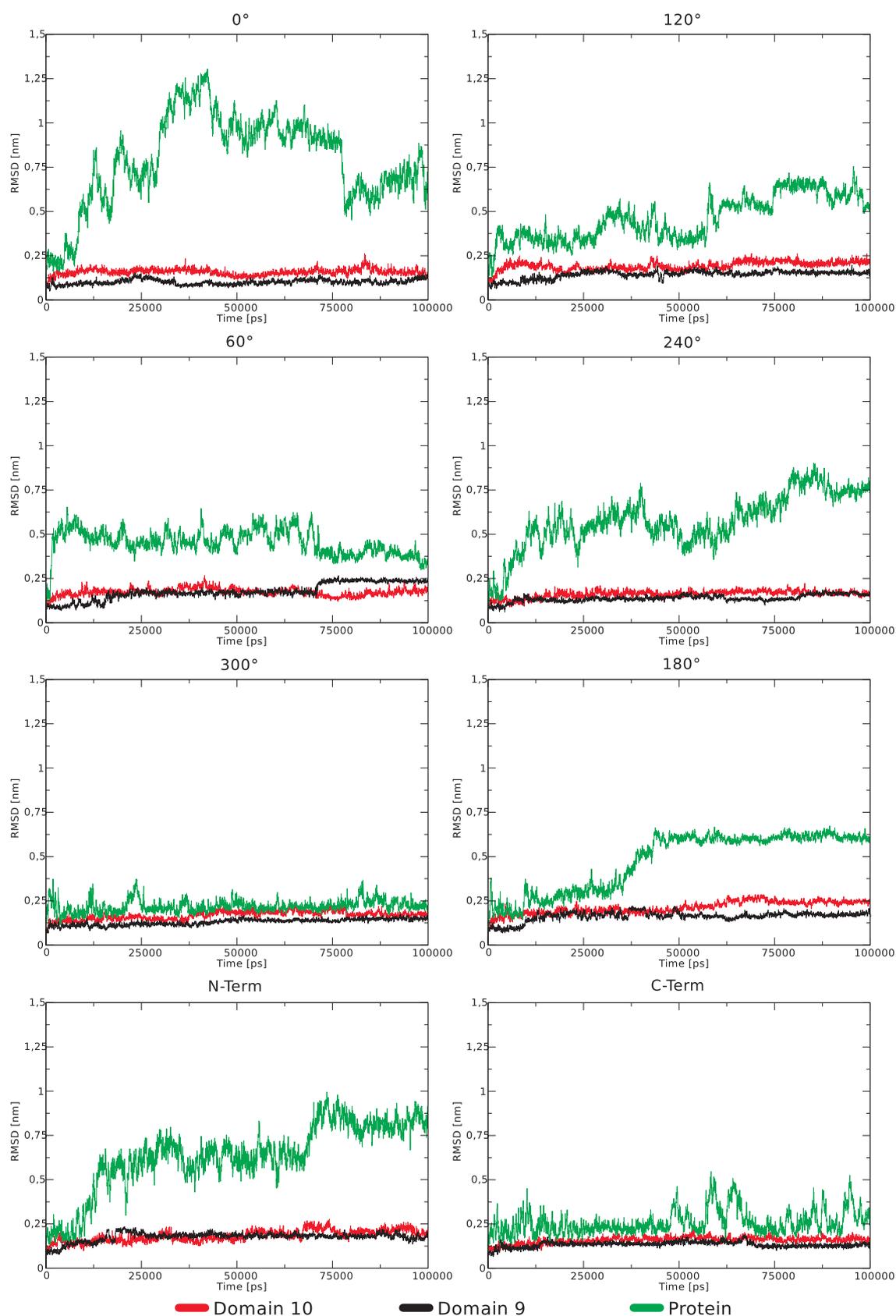
Figure 8. RMSD of Domain 9 & 10 and the entire protein from all simulations. Only Cα - atoms are considered. Major changes are observable only in the RMSD of the entire protein.
doi:10.1371/journal.pone.0020925.g008
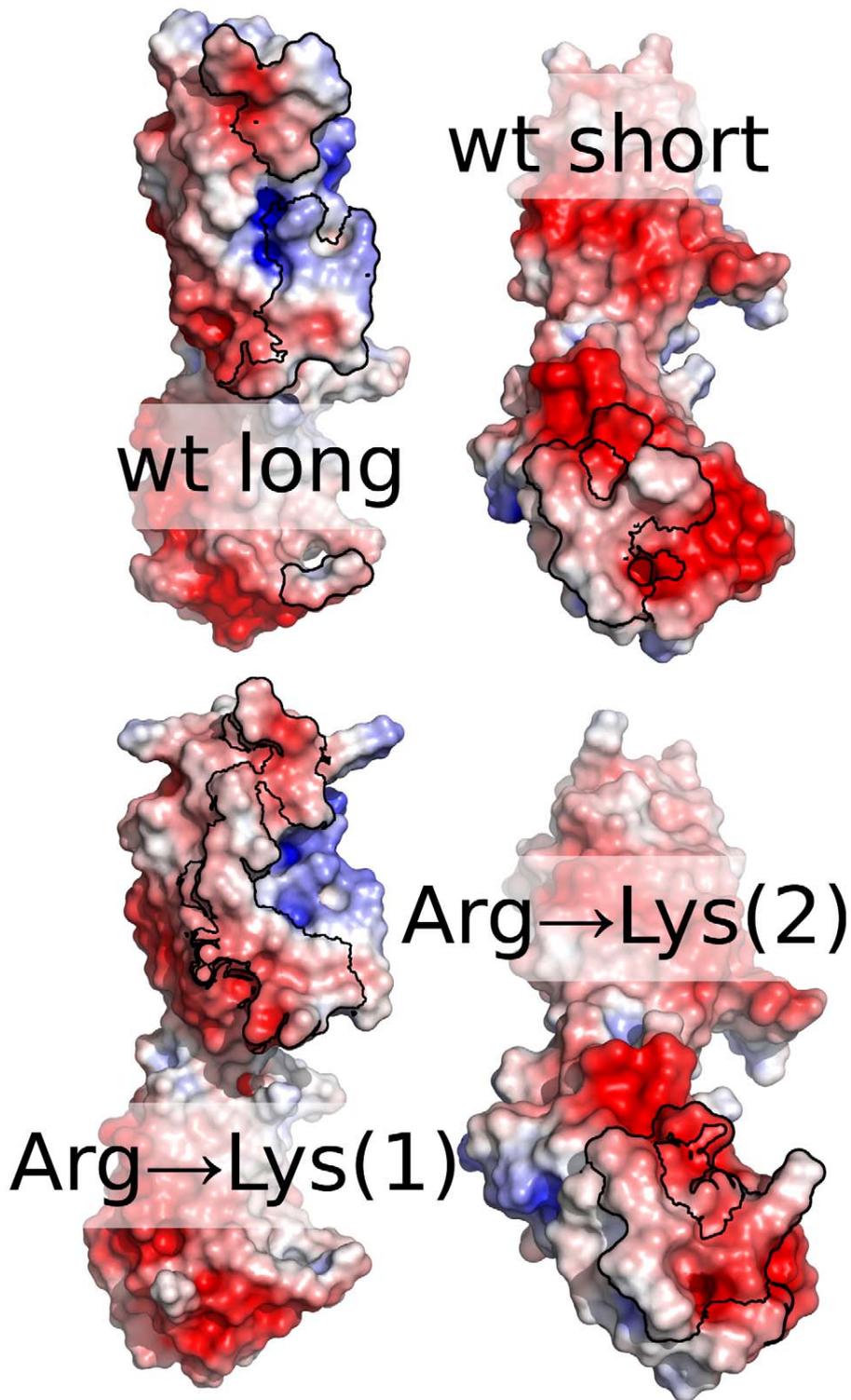
Figure 9. Electrostatic potential on the initial adsorption sites. The electrostatic surface potential on the solvent accessible surface is mapped on the van-der-Waals surface of FN. The first contact region is surrounded by a black line. Most parts of the contact regions exhibit a low electrostatic surface potential.
doi:10.1371/journal.pone.0020925.g009
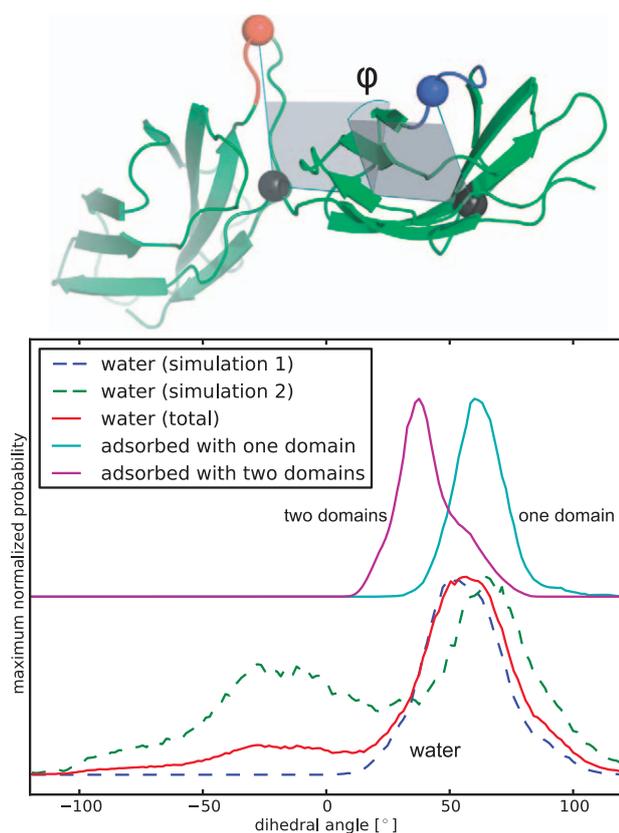
Figure 10. RGD and synergy site orientation during adsorption. Dihedral angle between the RGD (blue), two representative residues (black) and the synergy site (red) in FN is shown top. Below the angular distribution of this dihedral is shown for two simulations of FN in water and for two adsorption simulations. The individual water angle distributions are shown as dashed lines, the cumulative distribution as solid line. A narrowing of the distribution is observed after adsorption. Further, for the simulation with both domains adsorbed, an additional shift in the distribution is seen.
doi:10.1371/journal.pone.0020925.g010

binding sites (desgn-in). Negatively charged areas hinder adsorption (design-out).

Summary

Our results of the two weakly related systems, which mainly share a common secondary structure, reveal certain similarities, which may be regarded as common topics in protein adsorption on gold: the adsorption in both cases is a step-wise and cooperative process. When closely analyzing the adsorption both for the model peptide and the protein, it becomes apparent that *Arg* plays a crucial role. *Arg* residues are either providing the first contact or involved within the following ns after the first contact with the surface (Table 1). No other amino acid has the same behavior. The availability on the protein surface, the high affinity for gold [17,42] and its length predestine *Arg* as an early anchor for protein-gold interactions. Biotechnological applications may not only require surface-specific interactions, but additionally preferred orientations of proteins on the surface, best without sophisticated covalent attachment schemes. Our simulations indicate that intelligent placements of *Arg* could fulfill such a role. Furthermore, β-sheet rich proteins are stable on the gold surface, but domains may reorient impacting biological effects.

Conclusions

We have presented long atomistic molecular dynamics simulations of complex polypeptides on a polarizable gold(111) surface in water. On the time scale of the simulation, no unfolding has been observed. However, we did observe significant domain-domain re-orientation, which may promote the exposure of cryptic epitopes [41]. The simulations have also demonstrated the crucial role of *Arg* in promoting early on in the adhesion process a direct contact with the gold surface, similar to what has recently been shown for the adsorption of Lysozyme on a charged model surface [43]. Leveraging these special properties of *Arg* may help to design proteins adsorbing on a gold surface in a specific orientation without the need for sophisticated covalent coupling schemes. Taking into account that *Arg* is cationic at neutral pH, our result hint at the possibility to electrochemically control the protein-gold interactions, since the gold surface can be made positively charged (and thus *Arg*-repelling) by an applied potential bias. Hence, our simulations offer a first in-depth view into the possibilities to rationally control the binding of proteins on gold.

Methods

We used fibronectin domains FN9–10 from crystal structure 1FNF [44] as input (starting from residue 1326). The structure was placed in 6 rotations in $60°$ steps parallel and 2 conformations perpendicular to the gold surface with a water layer separating the protein from the gold surface (Fig. 4). In the solvated box, Na and Cl ions were placed in order to obtain physiological conditions of 150 mMol and an overall neutral box. The obtained systems were energy minimized and then subject to 500 ps relaxation with restraints on all heavy atoms followed by a separate 50 ps simulation with semi-isotropic Parrinello-Rahman pressure coupling [45] in z-direction. Compressibility was set to $0\ m^2N^{-1}$ in xy-direction. The obtained structures were then simulated for 100 ns. All simulations were performed with the Gromacs 4.0.7 package [46], integration time steps of 2 fs and a Nose-Hoover thermostat [47] at 300 K. Particle Mesh Ewald [48] for long range electrostatics above 1.1 nm and switch cutoff 0.9–1.0 nm for van der Waals interactions was used. GolP [13] and OPLS/AA parameters [49] were used for the surface and the protein and the SPC water model were used. For distance and secondary structure analysis, we used the gromacs tools g_traj and do_dssp, which is a trajectory interface to the dssp program [50]. The simulation of RAD was performed by placing the equilibrated RAD β-sheet [31] parallel to the gold surface at a distance that do not prevent the formation of a complete water monolayer between the peptide and the surface. The β-sheet fragment is composed of 16 RAD16II peptide molecules. After minimization, the NVT dynamics was started and run for 100 ns. Values of the adsorption energies and of its components (see Fig. 1) were collected every 0.2 ps during the entire dynamics. From them we subracted the energy averaged over the first ns of the production dynamics (when the peptide is not in contact with the surface) to define the time-dependent adsorption energy reported in Fig. 1. The bonded term (blue) also includes the Lennard-Jones and Coulomb interactions between the peptide atoms separated by three bonds (1–4 interactions), as their parameterization and meaning is entangled with those of the bonding terms. To smooth out statistical fluctuations in the energy profiles, shown in Fig. 1, we present values obtained by averaging the energy on a 1 ns window centered on each collected energy. The use of smaller averaging windows yield similar trends but noiser profiles (see Fig. S1). We also recall that the image charge procedure used here provides correct statistical

Table 1. For each simulation and residue, the first contact time of a specific residue type is given.

AA/Simulation	0°	60°	120°	180°	240°	300°	N-Term	C-Term	Average first contact time
ACE	*0.275*	1.45	20.2	42.375			0.0		12.86
ALA		2.275	20.8	37.45		84.4	9.975	28.825	30.62
ARG	0.75	*0.025*	0.1	*0.475*	*0.05*	0.9	*0.025*	*0.175*	*0.31*
ASN	30.7			18.55	5.425	0.3		52.95	21.59
ASP	59.275	4.875	1.65	4.7	0.175	0.05	20.775		13.07
GLN	23.95	21.725	17.575	36.6					24.96
GLU	1.15	11.525	60.4	43.25	4.6	18.875	22.775	43.525	25.76
GLY	2.475	1.225	1.65	5.725	4.55	0.85	10.1	0.2	3.35
HIS	0.025	21.625	60.425	4.075	4.625	46.65			22.90
ILE	5.8	21.325	*0.075*	36.9				58.25	24.47
LEU	5.45	7.975	20.75			21.975	20.775		15.39
LYS			7.15	2.6					4.88
NAC	54.25	1.225						0.4	18.63
PHE	0.875	71.2	73.975	31.375	81.7	41.6			50.12
PRO	2.875	1.75	60.5	0.7	4.625	*0.05*		0.0	10.07
SER	1.55	30.55	2.1	0.025	6.8	0.125			6.86
THR	30.75	1.95	1.65	1.15	4.575	22.025	0.1	*0.175*	7.80
TYR		2.05		4.4				28.9	11.78
VAL	19.975	11.3	20.525	40.9	69.875	60.9		19.275	34.68

Criterium for a contact is a distance Δz from the surface <0.3 nm, as measured from the atom closest to the surface. Time unit is ns. The last columns provide averages over the first contact times in the simulations. Caps are listed as ACE (acetyl) and NAC (n-acetyl). The lowest time for each column is highlighted in bold (for N-Term and C-Term simulations, contact times of ACE and NAC, respectively, are 0.0 ns by construction). *Arg* covers ~13% from the total Solvent Accessible Surface (SAS) of the studied portion of Fibronectin (Domain 9 & 10).
doi:10.1371/journal.pone.0020925.t001

sampling and free energies [51] but may overestimate the image contribution to the interaction enthalpy. No entropic term is included in Fig. 1.

Supporting Information

Figure S1 Impact of averaging of the potential energy of the system. Potential energy of the system as obtained directly from the dynamics (Potential E) and averaged over time windows of different lengths.
(TIF)

References

1. Zhdanov VP, Kasemo B (2002) Bistability in catalytic reactions on the nm scale. Surf Sci 496: 251–263.
2. Ratner BD, Bryant SJ (2004) BIOMATERIALS: Where We Have Been and Where We are Going. Annu Rev Biomed Eng 6: 41–75.
3. Gobin AM, Lee MH, Halas NJ, James WD, Drezek RA, et al. (2007) Near-Infrared Resonant Nanoshells for Combined Optical Imaging and Photothermal Cancer Therapy. Nano Lett 7: 1929–1934.
4. Nel AE, Madler L, Velegol D, Xia T, Hoek EMV, et al. (2009) Understanding biophysicochemical interactions at the nano-bio interface. Nat Mater 8: 543–557.
5. Alberola AP, Rädler JO (2009) The defined presentation of nanoparticles to cells and their surface controlled uptake. Biomaterials 30: 3766–3770.
6. Di Felice R, Selloni A, Molinari E (2003) DFT Study of Cysteine Adsorption on Au(111). J Phys Chem B 107: 1151–1156.
7. Sarikaya M, Tamerler C, Jen AKY, Schulten K, Baneyx F (2003) Molecular biomimetics: nanotechnology through biology. Nat Mater 2: 577–585.
8. Cohavi O, Corni S, Rienzo FD, Felice RD, Gottschalk KE, et al. (2010) Protein-surface interactions: challenging experiments and computations. J Mol Recognit 23: 259–262.
9. Harding JH, Duffy DM, Sushko ML, Rodger PM, Quigley D, et al. (2008) Computational Techniques at the Organic-Inorganic Interface in Biomineralization. Chem Rev 108: 4823–4854.
10. Notman R, Walsh TR (2009) Molecular dynamics studies of the interactions of water and amino acid analogues with quartz surfaces. Langmuir 25: 1638–1644.
11. Oren EE, Notman R, Kim IW, Evans JS, Walsh TR, et al. (2010) Probing the molecular mechanisms of quartz-binding peptides. Langmuir 26: 11003–11009.
12. Shukla R, Bansal V, Chaudhary M, Basu A, Bhonde RR, et al. (2005) Biocompatibility of gold nanoparticles and their endocytotic fate inside the cellular compartment: A microscopic overview. Langmuir 21: 10644–10654.
13. Iori F, Di Felice R, Molinari E, Corni S (2009) GoIP: An Atomistic Force-Field to Describe the Interaction of Proteins With Au(111) Surfaces in Water. J Comput Chem 30: 1465–1476.
14. Kokh DB, Corni S, Winn PJ, Hoefling M, Gottschalk KE, et al. (2010) ProMetCS: An Atomistic Force Field for Modeling Protein - Metal Surface Interactions in a Continuum Aqueous Solvent. J Chem Theory Comput 6: 1753–1768.
15. Hoefling M, Iori F, Corni S, Gottschalk K (2010) Interaction of amino acids with the Au(111) surface: adsorption free energies from molecular dynamics simulations. Langmuir 26: 8347–8351.
16. Hoefling M, Iori F, Corni S, Gottschalk K (2010) The conformations of amino acids on a gold(111) surface. Chemphyschem 11: 1763–1767.
17. Peelle BR, Krauland EM, Wittrup KD, Belcher AM (2005) Design criteria for engineering inorganic material-specific peptides. Langmuir 21: 6929–6933.

18. Reichmann D, Rahat O, Cohen M, Neuvirth H, Schreiber G (2007) The molecular architecture of protein-protein binding sites. Curr Opin Struct Biol 17: 67–76.

19. Matthes D, de Groot BL (2009) Secondary structure propensities in peptide folding simulations: a systematic comparison of molecular mechanics interaction schemes. Biophys J 97: 599–608.

20. Dölker N, Zachariae U, Grubmüller H (2010) Hydrophilic linkers and polar contacts affect aggregation of FG repeat peptides. Biophys J 98: 2653–2661.

21. Stumpe MC, Grubmüller H (2009) Urea impedes the hydrophobic collapse of partially unfolded proteins. Biophys J 96: 3744–3752.

22. Snow CD, Nguyen H, Pande VS, Gruebele M (2002) Absolute comparison of simulated and experimental protein-folding dynamics. Nature 420: 102–106.

23. Holmes TC, de Lacalle S, Su X, Liu G, Rich A, et al. (2000) Extensive neurite outgrowth and active synapse formation on self-assembling peptide scaffolds. Proc Natl Acad Sci U S A 97: 6728–6733.

24. Narmoneva DA, Oni O, Sieminski AL, Zhang S, Gertler JP, et al. (2005) Self-assembling short oligopeptides and the promotion of angiogenesis. Biomaterials 26: 4837–4846.

25. Park J, Kahng B, Kamm RD, Hwang W (2006) Atomistic simulation approach to a continuum description of self-assembled beta-sheet filaments. Biophys J 90: 2510–2524.

26. Lacerda SHDP, Park JJ, Meuse C, Pristinski D, Becker ML, et al. (2010) Interaction of gold nanoparticles with common human blood proteins. ACS Nano 4: 365–379.

27. Lundqvist M, Stigler J, Elia G, Lynch I, Cedervall T, et al. (2008) Nanoparticle size and surface properties determine the protein corona with possible implications for biological impacts. Proc Natl Acad Sci U S A 105: 14265–14270.

28. Leiss M, Beckmann K, Giros A, Costell M, Fässler R (2008) The role of integrin binding sites in fibronectin matrix assembly in vivo. Curr Opin Cell Biol 20: 502–507.

29. Mardon HJ, Grant KE (1994) The role of the ninth and tenth type III domains of human fibronectin in cell adhesion. FEBS Lett 340: 197–201.

30. Ellis-Behnke RG, Liang Y, You S, Tay DKC, Zhang S, et al. (2006) Nano neuro knitting: Peptide nanofiber scaffold for brain repair and axon regeneration with functional return of vision. Proc Natl Acad Sci U S A 103: 5054–5059.

31. Monti S (2007) RAD16II beta-Sheet Filaments onto Titanium Dioxide: Dynamics and Adsorption Properties. J Phys Chem B 111: 16962–16973.

32. Rief M, Gautel M, Schemmel A, Gaub HE (1998) The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy. Biophys J 75: 3008–3014.

33. Meadows PY, Walker GC (2005) Force microscopy studies of fibronectin adsorption and subsequent cellular adhesion to substrates with well-defined surface chemistries. Langmuir 21: 4096–4107.

34. Antia M, Islas LD, Boness DA, Baneyx G, Vogel V (2006) Single molecule fluorescence studies of surface-adsorbed fibronectin. Biomaterials 27: 679–690.

35. Baugh L, Vogel V (2004) Structural changes of fibronectin adsorbed to model surfaces probed by fluorescence resonance energy transfer. J Biomed Mater Res A 69: 525–534.

36. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–637.

37. Baker NA, Sept D, Joseph S, Holst MJ, McCammon JA (2001) Electrostatics of nanosystems: Application to microtubules and the ribosome. Proc Natl Acad Sci U S A 98: 10037–10041.

38. Altroff H, Choulier L, Mardon HJ (2003) Synergistic activity of the ninth and tenth FIII domains of human fibronectin depends upon structural stability. J Biol Chem 278: 491–497.

39. Friedland JC, Lee MH, Boettiger D (2009) Mechanically Activated Integrin Switch Controls alpha5beta1 Function. Science 323: 642–644.

40. Krammer A, Craig D, Thomas WE, Schulten K, Vogel V (2002) A structural model for force regulated integrin binding to fibronectin's RGD-synergy site. Matrix Biol 21: 139–147.

41. Lynch I, Dawson KA, Linse S (2006) Detecting Cryptic Epitopes Created by Nanoparticles. Science Signaling 2006: pe14–.

42. Hnilova M, Oren EE, Seker UOS, Wilson BR, Collino S, et al. (2008) Effect of Molecular Conformations on the Adsorption Behavior of Gold-Binding Peptides. Langmuir 24: 12440–12445.

43. Kubiak-Ossowska K, Mulheran PA (2010) What governs protein adsorption and immobilization at a charged solid surface?. Langmuir 26: 7690–7694.

44. Leahy DJ, Aukhil I, Erickson HP (1996) 2.0 Å Crystal Structure of a Four-Domain Segment of Human Fibronectin Encompassing the RGD Loop and Synergy Region. Cell 84: 155–164.

45. Parrinello M, Rahman A (1981) Polymorphic transitions in single crystals: A new molecular dynamics method. J Appl Phys 52: 7182–7190.

46. Hess B, Kutzner C, van der Spoel D, Lindahl E (2008) GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. J Chem Theory Comput 4: 435–447.

47. Hoover WG (1985) Canonical dynamics: Equilibrium phase-space distributions. Phys Rev A 31: 1695–1697.

48. Darden T, York D, Pedersen L (1993) Particle Mesh Ewald - An n.log(N) Method for Ewald Sums in large Systems. J Chem Phys 98: 10089–10092.

49. Jorgensen WL, Maxwell DS, TiradoRives J (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J Am Chem Soc 118: 11225–11236.

50. Kabsch W, Sander C (1983) Dictionary of protein secondary structure: pattern recognition of hydrogen-bonded and geometrical features. Biopolymers 22: 2577–2637.

51. Iori F, Corni S (2008) Including Image Charge Effects in the Molecular Dynamics Simulations of Molecules on Metal Surfaces. J Comp Chem 29: 1656–1666.

– P9 –

169

Journal of Computational Chemistry

Enabling Grand-Canonical Monte Carlo: Extending the Flexibility of GROMACS Through the GromPy Python Interface Module

SCHOLARONE™
Manuscripts

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Enabling Grand-Canonical Monte Carlo: Extending the Flexibility of GROMACS Through the GromPy Python Interface Module

René Pool[1,2,*], Jaap Heringa[1,2], Martin Hoefling[3], Roland Schulz[4], Jeremy C. Smith[4], and K. Anton Feenstra[1,2]

[1]Centre for Integrative Bioinformatics Vrije Universiteit (IBIVU), Vrije Universiteit Amsterdam, De Boelelaan 1081a, 1081HV Amsterdam, the Netherlands
[2]Netherlands Bioinformatics Centre, Geert Grooteplein 28, 6525GA Nijmegen, the Netherlands
[3]Max-Planck-Institute for Biophysical Chemistry Dept. for Theoretical & Computational Biophysics, Am Fassberg 11, 37077 Goettingen, Germany
[4]University of Tennessee/Oak Ridge National Laboratory, Center for Molecular Biophysics, 1 Bethel Valley, Oak Ridge, Tennessee, USA
[*]Corresponding author. Email: r.pool@vu.nl

July 12, 2011

Abstract

We report on a python interface to the GROMACS molecular simulation package, GromPy (available at https://github.com/GromPy). This interface uses the ctypes python module that allows function calls to shared libraries e.g. written in C. To the best of our knowledge, this is the first reported interface to the GROMACS library that uses direct library calls. GromPy can be used for extending the current GROMACS simulation and analysis modes. In this work we demonstrate that the interface enables hybrid Monte-Carlo/Molecular Dynamics simulations in the grand-canonical ensemble, a simulation mode that is currently not implemented in GROMACS. For this application, the interplay between GromPy and GROMACS requires only minor modifications of the GROMACS source code, not affecting the operation, efficiency and performance of the GROMACS applications. We validate the grand-canonical application against molecular dynamics in the canonical ensemble by comparison of equations of state. The results of the grand-canonical simulations are in complete agreement with molecular dynamics in the canonical ensemble. The python overhead of the grand-canonical scheme is only minimal.

1

1 Introduction

The `GROMACS` molecular simulation package [1, 2, 3, 4] is widely used in the field of (bio)molecular simulation. The most common setup of a simulation system in `GROMACS`, as in most other major molecular simulation software, assumes a system with a fixed composition of molecules. In addition, for each type of molecule, a fixed, predefined chemical connectivity is assumed. This setup does not readily allow the breaking of chemical bonds, nor the addition or removal of atoms or molecules to or from the system, e.g. simulations in the grand-canonical ensemble. A common workaround implementation involves shell scripting to apply modifications to the simulated system setup, followed by (re)starting of the simulation engine. Needless to say, this adds significant overhead and subtracts from the overall efficiency. Moreover, such approaches show unfavorable scaling behavior with respect to computational efforts, e.g. when increasing the amount of insertions/removals or when increasing the system size.

The grand-canonical ensemble can be used for studying systems where one is interested in the average number of molecules as a function of the external chemical potential and temperature. This renders it a suitable ensemble e.g. for exploring adsorption behavior of a given molecular species to the system of interest [5]. In a grand-canonical Monte Carlo (GCMC) simulation, one imposes the chemical potential μ_i of species i, the system volume V and the temperature T. During simulation, particles of type i are removed or inserted as a result of the imposed chemical potential. At equilibrium, the amount of removals is equal to the amount of insertions, and one can sample the average number of molecules i. The main computational advantage with respect to MD and NVT Monte Carlo (MC) is that equilibration times are drastically reduced as well as the sizes needed for the molecular systems [5]. It is also possible to combine MD with grand-canonical MC (GCMC) [5, 6]. The result is a 'hybrid' scheme that alternates short MD trajectories for particle translations of a system containing N_i particles of type i, with trial particle removals ($N_i \leftarrow N_i - 1$) or insertions ($N_i \leftarrow N_i + 1$).

The main application of `GROMACS` is as an engine to perform Molecular Dynamics (MD) simulations. Based on such simulations, dynamic system properties of interest can be determined. In addition, the MD trajectories also contain (non-)equilibrium thermodynamic properties of molecular systems. For analyzing the simulation outcomes, `GROMACS` comes with a range of applications that facilitate this process. With respect to specific simulation options (e.g. GCMC) or with respect to data analysis, it would however be useful to have the `GROMACS` data structures accessible to the user via interpreted, high-level programming languages such as `python`.

In contrast to `C` or Fortran, `python` is suitable for rapid-prototyping and is easy to read and learn. Moreover the `python` user community is active and growing [7] and several `python` packages such as BioPython [8] and PyCogent [9] have become standards. A `python` interface would therefore extend the scope of users that can contribute to and use the flexibility of the `GROMACS` simulation package.

In this work we describe an approach that makes the `GROMACS` data structures available to the user via the `python` module `GromPy` acting as an interface to the `GROMACS` C-library. The module allows access to the desired `GROMACS` data structures in memory from the `python` interpreter that can then be used to implement analysis tools and new simulation schemes. Here, we illustrate the use of the `GromPy` interface by implementing a

2

grand-canonical Monte Carlo simulation scheme [5] for which we utilize `GROMACS` C-library functions to perform energy calculations.

2 Methods

2.1 The `GromPy` Python Interface

The `GROMACS` package is written in the `C` programming language. We base our development tree on `GROMACS` version 4.0.5 that will be ported to the latest development branch in the near future.

To implement the interface, we choose the `python` programming language. `Python` is a high-level, interpreted, object-oriented and multi-platform programming language. It provides a large standard library and is easy to code. We use the free and open source CPython implementation of `python` [10]. Apart from the standard library, `python` has excellent extensions for numerical data analysis and data display [11, 12, 13, 14, 15, 16]. CPython is written in `C` and compiles `python` programs into intermediate code that can be executed by a virtual machine. The CPython implementation also allows the implementation of modules in `C` and the interfacing of (precompiled) libraries.

In our setup, we use the `ctypes` module [17] as interface between `python` and the `GROMACS` C-library. The `ctypes` module contains `python` equivalents for all basic `C` data types and allows the mapping of compound structures in `C` to `python` classes. As soon as the `GROMACS` data structures are accessible via `ctypes`, we can pass them to external `GROMACS` functions and access the result from the `python` interpreter during the execution with the `GromPy` module.

The initial `GromPy` implementation can be used for the analysis of trajectories, e.g. using `GROMACS`' periodic boundary condition removal and structure fitting routines [18]. `GromPy` can also read in index groups and topologies and was applied in the prototyping of `GROMACS` tools which were later implemented in `C` [19]. Recently, `GromPy` was applied to design a combined MD/MC approach to simulate FRET experiments and aid in the distance reconstruction [20]. This work involves extending `GromPy` by a GCMC simulation mode. The `GromPy` source code is publicly available at `https://github.com/GromPy`.

2.2 Hybrid Grand-Canonical Monte Carlo / Molecular Dynamics Simulations

In grand-canonical Monte Carlo (GCMC) the simulation box is in chemical equilibrium with an external bath. Hence, the chemical potential μ of both systems is equal. One therefore imposes the chemical potential of a particular molecular species upon which molecules are exchanged between the external reservoir and the simulation box [5]. In practice, this means that molecules are inserted into or removed from the simulation box during simulation. The MC acceptance rule for insertion of a molecule reads

$$P_{\text{acc}}(N \to N+1) = \min\left[1, \frac{V}{\Lambda^3(N+1)} \exp\left(\beta[\mu - \Delta U]\right)\right], \tag{1}$$

where N is the number of molecules, V is the box volume, $\Lambda = \sqrt{h^2/(2\pi m k_B T)}$ is the thermal De Broglie wavelength (h denotes Planck's constant, m the molecular mass, k_B

3

Boltzmann's constant and T the temperature), $\beta = 1/(k_B T)$ is the inverse temperature and $\Delta U = U(N+1) - U(N)$ is the energy difference of adding one molecule at a random position in the simulation box. For removal of a molecule, we use the following acceptance rule

$$P_{\text{acc}}(N \to N-1) = \min\left[1, \frac{\Lambda^3 N}{V}\exp\left(-\beta[\mu + \Delta U]\right)\right], \tag{2}$$

where $\Delta U = U(N-1) - U(N)$ is the potential energy difference associated with the removal of a randomly selected molecule.

To simulate thermal motion, we apply several MD steps at constant NVT using the velocity rescale thermostat [21], that generates a canonical ensemble, in between the GCMC moves. The nature of the MC move (i.e. a trial insertion/removal or an MD move) during a MC cycle is chosen at random based on a user-defined list of probabilities for each type of MC move.

2.3 Extending `GromPy` and Modifying the `GROMACS` Source Code

This work involves an extension of `GromPy`, enabling GCMC using the `GROMACS` C-library. The general setup is shown in Fig. 1. When used in GCMC mode, `GromPy` needs a starting configuration with a number of molecules $N_{i,\text{start}}$ of type i in the form of a `GROMACS` `tpr` file stored on disk. Such a `tpr` file serves as input for a `GROMACS` simulation and contains all simulation parameters and a configuration of a system. The `tpr` file range $N_i \in [N_{i,\text{min}}, N_{i,\text{max}}]$ is generated in the preprocessing stage, where $N_{i,\text{min}} \leqslant N_{i,\text{start}}$ and $N_{i,\text{max}} \geqslant N_{i,\text{start}}$ are the extrema of the N_i sampling range. By imposing a chemical potential μ_i of this molecule type, `GromPy` samples the N_i range via the hybrid GCMC/MD algorithm.

All MC moves in our hybrid MD/grand-canonical MC module require having the current state $s_c : \left[N_{i,c}, \mathbf{r}^{N_{i,c}}, \mathbf{v}^{N_{i,c}}\right]$ and associated total potential energy U_c in memory, c.f. Fig. 2. This state is a member of the grand-canonical ensemble and thus comprises the current number of molecules $N_{i,c}$ of type i, the coordinates $\mathbf{r}^{N_{i,c}}$ and the velocities $\mathbf{v}^{N_{i,c}}$ (we use the $\mathbf{r}^N$ and $\mathbf{v}^N$ short hand notation for the coordinate and velocity arrays consisting of N elements). The GCMC module uses two MC move types: one that performs several MD steps on s_c to simulate thermal motion of the molecular system and one that performs a grand-canonical MC move that tries to modify s_c by inserting or removing a molecule. For computational efficiency, the MD move is always accepted since the resulting configuration is already part of the correct statistical mechanical ensemble. After the MD move, we update the coordinates, velocities and total potential energy. Inside the GCMC move we select either the removal or the insertion of a molecule with a probability of $P_{\text{insert}} = P_{\text{remove}} = \frac{1}{2}$. For insertion we generate a trial state s_t that has $N_{i,t} = N_{i,c} + 1$ molecules. The first $N_{i,c}$ elements of the coordinate and velocity arrays are copied from s_c. The last element is filled by a random molecular position $\mathbf{r}'$ inside the box and by a molecular velocity $\mathbf{v}'$ chosen at random from the Maxwell-Boltzmann velocity distribution associated with the imposed temperature T, respectively. This step requires having s_t in memory. If this is not the case, we first read a `tpr` file with $N_i = N_{i,t}$ from disk. A molecular removal involves generating a trial state s_t that has $N_{i,t} = N_{i,c} - 1$ molecules. We randomly select a molecule (k) from the list and copy the $N_{i,c}$ elements of the coordinate and velocity arrays from s_c to s_t, while excluding the kth

4

Figure 1: The GCMC simulation setup used in this work. The preprocessing stage (1) involves generating `tpr` files for each configuration $N_i \in [N_{i,\min}, N_{i,\max}]$ using `grompp`. The input (2) for `GromPy` comprises the molecule type i for which the chemical potential μ_i is imposed, a range of numbers of molecules $[N_{i,\min}, N_{i,\max}]$ of type i that can be sampled, a starting configuration $N_{i,\mathrm{start}}$, the `tpr` path on disk, and the output path on disk. The preprocessing step requires prior knowledge of the input parameters (a). `GromPy` (3) reads the input parameters (b). Molecular insertions/removals requires `tpr` reads from disk (c). Once read, the associated data structures are kept in memory. The necessary energy evaluations are performed by the `GROMACS` library (4) with which `GromPy` communicates (d). This shared object library is compiled (e) from a slightly modified version (5) of `GROMACS` 4.0.5. The generated output can be further analyzed by the native `GROMACS` analysis suite, `GromPy`, or other software.

5

element. Again, we require having s_t in memory and read from disk otherwise. Trial insertions or removals with associated U_t are accepted according to Eq. 1 or Eq. 2 (where $\Delta U = U_t - U_c$), respectively. If accepted, we update s_t to s_c and the associated potential energy U_t becomes U_c. Otherwise, we keep s_c and U_c. After each MC move we update the averages and increment the MC loop iterator.

As described above, the GCMC module uses the current and trial states (s_c and s_t) to sample the grand-canonical ensemble. For this, energy evaluations are needed to obtain U_c and U_t that serve as input for the acceptance rules for insertion (Eq. 1) and removal (Eq. 2). At run time, the states are stored in memory by interfacing with specific GROMACS library functions. The associated energies U_c and U_t are computed by calls to the GROMACS library. Both operations are performed using the python ctypes module. To achieve the interfacing, we modified the GROMACS 4.0.5 source code as shown in Fig. 3. Although the modifications were performed for the serial implementation of GROMACS, we intend to make the modifications compatible with the parallel parts of the code. We expect that this will require relatively little effort. The GROMACS function mdrunner() loads a tpr file and can perform an MD simulation on a given system. This function is called by the GROMACS mdrun executable. Since ctypes can load only shared object libraries, we compile the mdrun executable as a shared object library: libmdrun.so. During a GCMC run we generate trial states s_t by copying the current state s_c to s_t and adding a trial position (and velocity) for insertion or excluding a randomly selected molecule for removal. To achieve this flexibility, we have split up the mdrunner() function into three parts: mdr_init(cs), mdr_int(cs) and mdr_fin(cs). We added a new data structure cs for the current state that enables communication between the sub-functions. For our purposes the most important member of cs is the state s. By subsequently calling the three separate functions (and without modifying cs in between), the behavior of the original mdrunner() function is reproduced exactly. Function mdr_init(cs) reads a tpr file from disk and stores the state s in cs. Function mdr_int(cs) performs an MD calculation of N_{MD} steps. N_{MD} is also a member of cs and can be set from within GromPy. For an MD move the number of MD steps is set to $N_{MD} > 0$ and for energy evaluations in a GCMC move it is set to $N_{MD} = 0$ (which results in a single point energy calculation). Computational performance of the simulation is calculated by function mdr_fin(cs). The gain in total computational time is realized by keeping cs in python memory once initialized by a disk read. In this way cs can be (re)used efficiently for MD or GCMC moves.

Note that the $N_{i,start}$ configuration should be an equilibrated one. However, this is not a precondition for all other $N_i \neq N_{i,start}$ tpr files that the user wishes to use for sampling, since this tpr file is merely used to fill the coordinate and velocity arrays in a trial move. During simulation, s_c will always be part of the correct ensemble.

To summarize, once in memory cs can be manipulated for whatever intended purpose and can serve as input for mdr_int(cs). Our purpose is GCMC and we therefore need to manipulate the cs members s and N_{MD}. Obviously, the same behavior can be achieved by executing a shell script that calls the necessary GROMACS executables, i.e. grompp and mdrun. The downside of such an approach is that most of the time the GCMC shell will perform file I/O and/or system calls. Having the relevant GROMACS data structures in memory, combined with the modified GROMACS source code drastically reduces the time spent on file I/O and renders GromPy an efficient GCMC application, with less than 6%

6

Figure 2: Flowchart of GromPy in GCMC mode. Each MC move is based on the current state $s_c : \left[N_{i,c}, \mathbf{r}^{N_{i,c}}, \mathbf{v}^{N_{i,c}}\right]$ and associated total potential energy U_c, kept in memory. State s_c is defined by the number of molecules $N_{i,c}$ of type i, their coordinates $\mathbf{r}^{N_{i,c}}$ and velocities $\mathbf{v}^{N_{i,c}}$. GromPy uses two MC move types: an MD move of several MD steps and a grand-canonical MC move. After the MD move, the coordinates, velocities and total potential energy are updated. The GCMC move involves removal or insertion of a molecule selected with a probability $P_{\text{insert}} = P_{\text{remove}} = \frac{1}{2}$. Insertion or removal requires having s_t in memory. If this is not the case, we first read a tpr file with $N_i = N_{i,t}$ from disk. Insertions or removals are accepted with the probabilities in Eq. 1 and Eq. 2, respectively. If accepted, s_t becomes s_c and U_t becomes U_c. Otherwise, we keep s_c and U_c. After each MC move we update the averages and increment the MC loop iterator j.

7

Figure 3: Modification of the source code of `GROMACS` version 4.0.5. Left: default compilation yields the `mdrun` executable (amongst others). This program calls function `mdrunner()` that is the calculation engine for MD simulations. Right: compilation of the `mdrun` executable as the shared object library `libmdrun.so` and splitting up function `mdrunner()` into an initialization stage an integration stage and a finalization stage. Communication between the stages is achieved through the new `cs` data structure. Library `libmdrun.so` is loaded into the `GromPy` module where `mdr_init(cs)`, `mdr_int(cs)` and are called for performing MD moves and GCMC trial moves by manipulating `cs` before each MC move.

of run time spent in overhead. This overhead involves logging to disk, reading of `tpr` files from disk, iterating over the MC loop, replacing the $\mathbf{r}^N$ and $\mathbf{v}^N$ arrays for trial insertions/removals and associated evaluations of Eqs. 1 and 2.

2.4   Validation of the Grand-Canonical Monte Carlo Module

We aim to validate the `GROMACS`-GCMC scheme by comparing equations of state determined by GCMC and *NVT* MD. For this, it is necessary to simulate a single phase. We therefore choose to simulate supercritical fluids. The validation is performed for two model systems. The first system consists of single Lennard-Jones (LJ) particles of the same type. For this, we use water particles of the MARTINI coarse-grained force field [22] that are modeled as single LJ particles. For this system type we approximate the critical properties by Gibbs ensemble simulation results [5]. For the second system, with polar SPC water [23], we also need to account for charges and insertions/removals of multi-atomic molecules, rendering it a more complicated and challenging test case. The critical properties for the SPC model are taken from the literature [24]: $T_{c,\text{SPC}} = 587$ K and

8

Table I: MD parameters used in this work for the LJ and SPC models. The MD time step is denoted by Δt, the total simulation time for each NVT simulation by t_{NVT}, the simulation time per MD move in each μVT simulation by $t_{\mathrm{MD},\ \mu VT}$ and the 'simulation time' for a single point energy calculation needed for a GCMC trial move by $t_{\mathrm{GCMC},\ \mu VT}$. We apply the velocity rescale thermostat [21] that ensures a canonical ensemble. The associated coupling frequency is represented by $\tau_{\mathrm{thermostat}}^{-1}$.

	Model	
	LJ	SPC
$N_{\mathrm{molecules}}$ $[-]$	400	500
Δt [ps]	2×10^{-2}	2×10^{-3}
t_{NVT} [ps]	2×10^{3}	2×10^{3}
$t_{\mathrm{MD},\ \mu VT}$ [ps]	2	0.2
$t_{\mathrm{GCMC},\ \mu VT}$ [ps]	0	0
$\tau_{\mathrm{thermostat}}^{-1}$ [ps^{-1}]	0.1	0.1

$\rho_{c,\mathrm{SPC}} = 15$ mol/l.

In the LJ simulations, we use a shift potential for the non-bonded interactions with a switch radius of $r_s = 0.9$ nm. The non-bonded interactions were truncated at $r_c = 1.2$ nm [22]. In the SPC simulations all non-bonded interactions were calculated up to a cut-off distance of $r_c = 0.9$ nm (corrections to the total energy and pressure due to truncation are taken into account) and he Coulombic interactions are calculated by the particle mesh Ewald method [25] with a spacing of the Fourier grid of 0.12 nm.

The NVT equations of state (EOS) for both systems are determined at $T = 773$ K and $T = 900$ K. The simulation parameters are summarized in Table I. For each density ρ, we perform a separate simulation of which the ranges are the x-values in Figs. 4a and 4b for the LJ and SPC models respectively. These density ranges are obtained by changing the box volume, while keeping the amount of molecules constant. A pilot experiment showed that NVT results are consistent when varying the box volume V at constant N or varying N at constant V. We average the total pressure p and hence obtain a pressure profile as a function of density ρ.

The μVT EOSs at $T = 773$ K and $T = 900$ K are obtained by imposing a range of chemical potentials μ to fixed volume systems of either LJ particles or SPC water molecules. The simulation parameters of the μVT simulations can be found in Table II. The MD parameters used in MD moves are listed in Table I. For the density ranges studied, c.f. the x-values in Figs. 4a and 4b for the LJ and SPC models respectively. For this type of simulation, we obtain a density profile as a function of μ.

The Gibbs-Duhem equation is used to validate the μVT results

$$\rho \left[\frac{\partial \mu}{\partial \rho} \right]_T = \left[\frac{\partial p}{\partial \rho} \right]_T, \tag{3}$$

from which the pressure profile

$$p(\rho) = \rho k_B T + \int_{\rho'=0}^{\rho'=\rho} d\rho' \left(\rho' \left[\frac{\partial \mu_{\mathrm{ex}}}{\partial \rho'} \right]_T \right) \tag{4}$$

9

Table II: GCMC parameters used in this work for the LJ and SPC models. The length of the cubic simulation box is denoted by b and the number of MC cycles by N_{cycles}. Each MC cycle consists of N_{moves} trial MC moves where the MC move type is chosen randomly with probabilities P_{MD} and P_{GCMC} for an MD move and GCMC move respectively.

	Model	
	LJ	SPC
b [nm]	3.64	2.70
N_{cycles} $[-]$	2500	2500
N_{moves} $[-]$	42	42
P_{MD} $[-]$	0.05	0.05
P_{GCMC} $[-]$	0.95	0.95

is derived. The excess part of the chemical potential μ_{ex} is calculated as

$$
\begin{aligned}
\mu_{\text{ex}} &= \mu - \mu_{\text{id}} \\
&= \mu - k_B T \ln\left(\Lambda^3 \rho\right),
\end{aligned}
\tag{5}
$$

where μ_{id} is the ideal part of the chemical potential. The pressure as a function of density in Eq. 4 is determined from a numerical least-squares fit of a 6$^{\text{th}}$ degree polynomial to the $\mu V T$ data of $N_{\text{dat}} = 1000$ data points.

3   Results and Discussion

3.1   A Supercritical Lennard-Jones System

For the supercritical Lennard-Jones (LJ) system, we used a system of single particle MARTINI [22] water (W) molecules. A system consisting of just this molecule type, involves non-bonded LJ interactions only and therefore renders it a relatively simple test system. We calculated the critical temperature of this system as $T_{c,\text{W}} = 647.2$ K and its associated critical density as $\rho_{c,\text{W}} = 4.99$ mol/l by Gibbs ensemble simulations [5].

The NVT results are shown in Fig. 4a (left and bottom axes). We examined if the NVT EOS is different when varying the number of molecules compared to varying the simulation box volume. This was found not to be the case. The results of the LJ $\mu V T$ simulations are shown in Fig. 4a (right and bottom axes). The NVT and $\mu V T$ equations of state are completely equivalent.

3.2   A Supercritical SPC Water System

Apart from non-bonded LJ interactions between the water oxygen atoms, the SPC model [23] involves Coulomb interactions between the partially charged hydrogen and oxygen atoms. The relative orientation of the hydrogen and oxygen atoms within a water molecule is assumed constant, i.e. bond stretching and bond bending is constrained during simulation using the SETTLE algorithm [26].

10
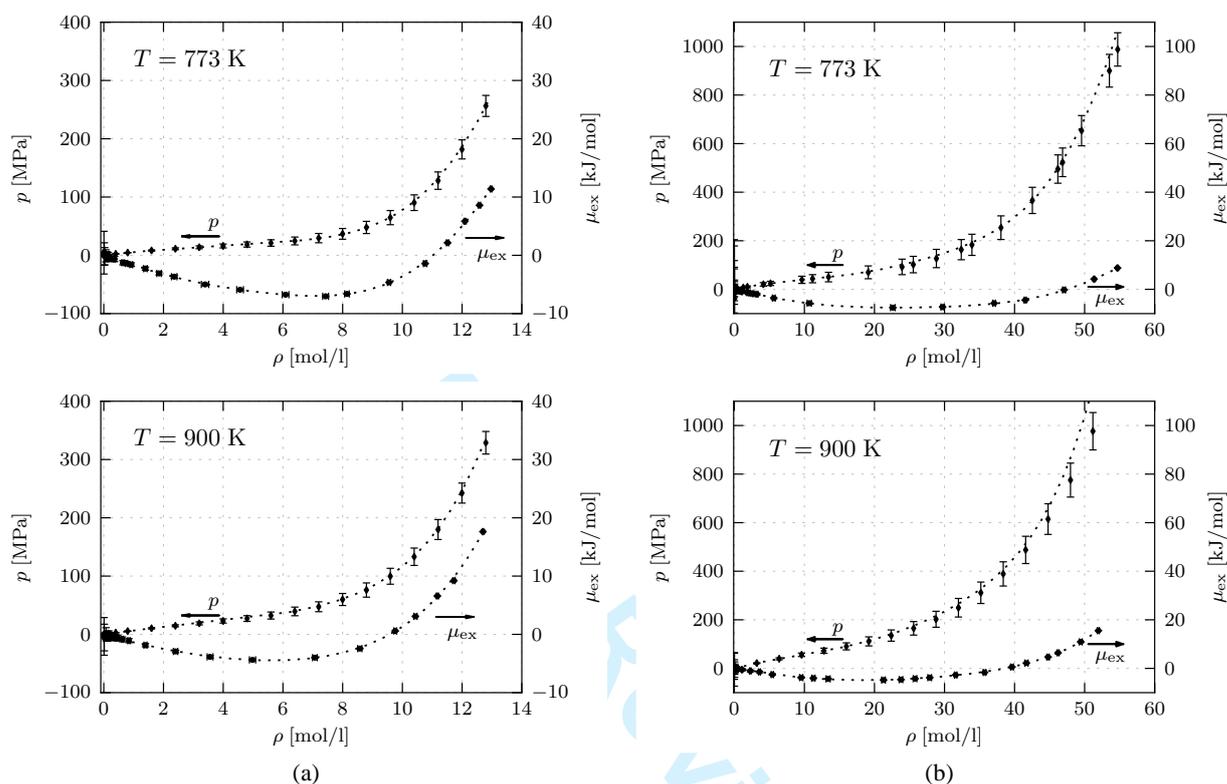
Figure 4: Equations of state for the LJ model (a) and the SPC water model (b)) at $T = 773$ K (top) and $T = 900$ K (bottom). The data points on the $p(\rho)$ line are determined by MD at NVT. The points on the $\mu_{\mathrm{ex}}(\rho)$ line are determined by grand-canonical MC using `GromPy` in GCMC mode. The standard deviations of μ_{ex} and of ρ for the μVT data were calculated by conventional error propagation rules. The least-squares fit of a 6th degree polynomial to these μ_{ex} points was transformed into a $p(\rho)$ curve using Eq. 4. Both results are shown as dotted lines in each plot.

11

In Fig. 4b (left and bottom axes), we show the NVT results. We again validated the NVT EOS when varying the number of molecules compared against the NVT EOS when varying the simulation box volume. The results of the μVT simulations are shown in Fig. 4b (right and bottom axes).

As we can see from Fig. 4b, the μVT data at $T = 773$ K are in agreement with the NVT results. The μVT data at $T = 900$ K agrees with the NVT data up to a density of $\rho \approx 48$ mol/l within the NVT error bars. The overestimation of the $p(\rho)$ profile at higher particle densities derived from the grand-canonical results can be explained by GCMC sampling difficulties. At $T = 773$ K the deviation is only very slight, while at $T = 900$ K, the deviation is higher due to more extreme simulation conditions. SPC molecule insertions are performed by generating a random position in the simulation box for the oxygen atom, followed by randomly orienting the hydrogens while meeting the bond angular and bond length constraints. A more efficient sampling at higher densities could be achieved by applying the configurational bias Monte Carlo (CBMC) technique [5]. In CBMC one selects the most favorable insertion configuration from a set of trial configurations and appropriately corrects for this bias. It should be kept in mind that for both temperatures, the conditions at higher densities can be considered extreme, e.g. pressures of over 5×10^8 Pa.

3.3   Computational Performance and Accuracy

To get an impression of the computational performance of `GromPy` in GCMC mode, we again determined the EOS for the LJ system at $T = 773$ K. For the GCMC case, simulation parameters are the same as above. For each data point in Fig. 5, the number of particles in the NVT simulation were taken the same as the average number of particles, calculated from the μVT simulation series. In this way a fair comparison can be made between the two simulation modes. Per μVT or NVT simulation, we used a total of 749700 integration steps of which the first 16.7% was used for equilibration.

Both EOSs were determined on a 32-bit Linux machine with the applications running on a single CPU. The total simulation time for the NVT EOS is 2400 s (8 s spent on system calls and 2392 s spent on 'real' CPU time). The total simulation time for the μVT EOS is 2547 s (149 s spent on system calls and 2397 s spent on 'real' CPU time). The ~ 150 s difference between the two simulation modes comes from the limited amount of time spent on system calls and can be considered as 'python overhead' as described in Sec. 2.3. Note that this also involves the evaluations of Eqs. 1 and 2 in `python`.

The μVT and NVT EOSs are completely equivalent, c.f. Fig. 5. The uncertainty bandwidth in the pressure profile, based on the standard deviations in the μVT data is the area between the thin solid lines in Fig. 5. The μVT EOS uncertainty is well within the error bars of the pressure sampled in the NVT ensemble.

To illustrate the file I/O overhead problem in a shell approach that does not use direct calls to the `GROMACS` library, we implemented such a shell that can also sample the grand-canonical ensemble but uses the `GROMACS` executables to perform the necessay MC moves. We simulated the LJ system at $T = 900$ K at a chemical potential yielding an average number of $\langle N \rangle \approx 377$ using both GCMC approaches. For both simulations, the parameters are listed in Table II (and Table I for the parameters of the MD moves). The 'shell' GCMC module requires 10800 s for 2500 MC cycles, whereas `GromPy` in GCMC

12

Figure 5: Equation of state for the LJ model at $T = 773$ K. The data points on the $p(\rho)$ line are determined by MD at NVT. The points on the $\mu_{ex}(\rho)$ line are determined by grand-canonical MC using `GromPy` in GCMC mode. The standard deviations of μ_{ex} and of ρ for the μVT data were calculated by conventional error propagation rules. The least-squares fit of a 6th degree polynomial to these μ_{ex} points was transformed into a $p(\rho)$ curve using Eq. 4. Both results are shown as dotted lines. Note that the data points for both simulation types are obtained by an equal number of integration steps. The error band width in the pressure profile based on the uncertainty in the μVT data is the area between the thin full lines.

mode does the same 14 times faster (771 s). It thus turns out that the 'shell' approach spends over 90% of the computation time on system calls and disk operations.

4   Conclusions

We have successfully implemented and extended the `GromPy` module (available at https://github.com/GromPy) and enabled simulations in grand-canonical ensemble using the `GROMACS` C-library. To this end, only minor modifications to the `GROMACS` source code needed to be applied, and these do not in any way affect the operation, efficiency and/or performance of the `GROMACS` applications built with the `GROMACS` source. To the best of our knowledge, `GromPy` is the first reported interface to the `GROMACS` library and MD engine that uses direct library calls. It can be used for further extending the current `GROMACS` simulation and analysis modes.

We validated our grand-canonical scheme for two system types. For the simplest one that involves only Lennard-Jones interactions, the μVT results are in complete agreement with those of NVT MD simulations performed with `GROMACS`. For a second, more complicated, system that also involves Coulombic interactions and insertions of multi-atomic molecules, the μVT results agree with the NVT results at $T = 773$ K, but overestimate the high density region of the NVT equation of state at $T = 900$ K. The deviation is explained by sampling difficulties at these high densities. Sampling efficiency might be enhanced by implementing configurational bias Monte Carlo for multi-atomic molecules.

The computational performance of `GromPy` in GCMC mode is comparable to the `GROMACS` `mdrun` executable. The accuracy of the μVT data is well within that of conven-

13

tional MD in the NVT ensemble.

Our work is compatible with the 4.0.7 version of `GROMACS`, and only minor modifications are needed for the 4.5 version and higher versions. For the near future, we plan to merge the necessary changes on the code to the main development tree which will make our GCMC compatible with the latest `GROMACS` releases, of course in consultation with the `GROMACS` developers community. In addition our minor modifications to the serial implementation of the `GROMACS` source code should be made compatible with the parallel implementation. We expect that grand-canonical MC and hybrid MD/MC is of interest to the `GROMACS` users community. Our modifications to the source code are only minor and do not stand in the way of 'normal' use of the MD engine. Additionally, a `python` interface to `GROMACS` will contribute significantly to the flexibility of the package.

Acknowledgments

References

[1] Berendsen, H. J. C.; van der Spoel, D. and Van Drunen, R., Comput Phys Commun, 1995, 91, 43–56.

[2] Lindahl, E.; Hess, B. and van der Spoel, D., J Mol Mod, 2001, 7, 306–317.

[3] van der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E. and Berendsen, H., J Comput Chem, 2005, 26, 1701–1718.

[4] Hess, B.; Kutzner, C.; van der Spoel, D. and Lindahl, E., J Chem Theory Comput, 2008, 4, 435–447.

[5] Frenkel, D. and Smit, B., *Understanding molecular simulation: from algorithms to applications*, Academic Press, San Diego, 2nd ed., 2002.

[6] Pool, R. and Bolhuis, P. G., J Phys Chem B, 2005, 109, 6650–6657.

[7] Pérez, F.; Granger, B. E. and Hunter, J. D., Comput Sci Eng, 2011, 13, 13–21.

[8] Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B. and de Hoon, M. J. L., Bioinformatics , 2009, 25, 1422–1423.

14

[9] Knight, R.; Maxwell, P.; Birmingham, A.; Carnes, J.; Caporaso, J. G.; Easton, B. C.; Eaton, M.; Hamady, M.; Lindsay, H.; Liu, Z.; Lozupone, C.; McDonald, D.; Robeson, M.; Sammut, R.; Smit, S.; Wakefield, M. J.; Widmann, J.; Wikman, S.; Wilson, S.; Ying, H. and Huttley, G. A., Genome Biol , 2007, 8, R171.

[10] The Python Software Foundation, Python Programming Language – Official Website, `http://www.python.org/`.

[11] Dubois, P. F.; Hinsen, K. and Hugunin, J., Comput Phys, 1996, 10, 262–267.

[12] Ascher, D.; Dubois, P. F.; Hinsen, K.; Hugunin, J. and Oliphant, T., *Numerical Python*, ucrl-ma-12 ed., 1999.

[13] Jones, E.; Oliphant, T. and Peterson, P., SciPy: Open Source Scientific Tools for Python, `http://www.scipy.org`, 2001.

[14] Oliphant, T. E., *Guide to NumPy*, 2006.

[15] Oliphant, T. E., Comput Sci Eng, 2007, 9, 10–20.

[16] Hunter, J. D., Comput Sci Eng, 2007, 9, 90–95.

[17] Heller, T., The ctypes package, `http://pypi.python.org/pypi/ctypes`.

[18] Hoefling, M. and Gottschalk, K. E., J Struct Biol , 2010, 171, 52–63.

[19] Wolf, M. G.; Hoefling, M.; Aponte-Santamaría, C.; Grubmüller, H. and Groenhof, G., J Comput Chem, 2010, 31, 2169–2174.

[20] Hoefling, M.; Lima, N.; Haenni, D.; Seidel, C. A. M.; Schuler, B. and Grubmüller, H., PloS One , 2011, 6, e19791.

[21] Bussi, G.; Donadio, D. and Parrinello, M., J Chem Phys , 2007, 126, 014101.

[22] Marrink, S. J.; Risselada, H. J.; Yefimov, S.; Tieleman, D. P. and de Vries, A. H., J Phys Chem B , 2007, 111, 7812–7824.

[23] Berendsen, H. J. C.; Postma, J. P. M.; Van Gunsteren, W. F. and Hermans, J., In *Intermolecular Forces*, Edited by Pullman, B., D. Reidel Publishing Company, vol. 11, pp. 313–338, 1981.

[24] de Pablo, J. J.; Prausnitz, J. M.; Strauch, H. J. and Cummings, P. T., J Chem Phys, 1990, 93, 7355–7359.

[25] Essmann, U.; Perera, L. and Berkowitz, M. L., J Chem Phys, 1995, 103, 8577–8593.

[26] Miyamoto, S. and Kollman, P. A., J Comput Chem, 1992, 13, 952–962.

15

Table I: MD parameters used in this work for the LJ and SPC models. The MD time step is denoted by Δt, the total simulation time for each NVT simulation by t_{NVT}, the simulation time per MD move in each μVT simulation by $t_{\text{MD, }\mu VT}$ and the 'simulation time' for a single point energy calculation needed for a GCMC trial move by $t_{\text{GCMC, }\mu VT}$. We apply the velocity rescale thermostat [21] that ensures a canonical ensemble. The associated coupling frequency is represented by $\tau_{\text{thermostat}}^{-1}$.

	Model	
	LJ	SPC
$N_{\text{molecules}} \; [-]$	400	500
$\Delta t \; [\text{ps}]$	2×10^{-2}	2×10^{-3}
$t_{NVT} \; [\text{ps}]$	2×10^{3}	2×10^{3}
$t_{\text{MD, }\mu VT} \; [\text{ps}]$	2	0.2
$t_{\text{GCMC, }\mu VT} \; [\text{ps}]$	0	0
$\tau_{\text{thermostat}}^{-1} \; [\text{ps}^{-1}]$	0.1	0.1

16

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60

Table II: GCMC parameters used in this work for the LJ and SPC models. The length of the cubic simulation box is denoted by b and the number of MC cycles by N_{cycles}. Each MC cycle consists of N_{moves} trial MC moves where the MC move type is chosen randomly with probabilities P_{MD} and P_{GCMC} for an MD move and GCMC move respectively.

|	Model	
	LJ	SPC
b [nm]	3.64	2.70
N_{cycles} $[-]$	2500	2500
N_{moves} $[-]$	42	42
P_{MD} $[-]$	0.05	0.05
P_{GCMC} $[-]$	0.95	0.95

17

Figure 1:

The GCMC simulation setup used in this work. The preprocessing stage (1) involves generating `tpr` files for each configuration $N_i \in [N_{i,\min}, N_{i,\max}]$ using `grompp`. The input (2) for `GromPy` comprises the molecule type i for which the chemical potential μ_i is imposed, a range of numbers of molecules $[N_{i,\min}, N_{i,\max}]$ of type i that can be sampled, a starting configuration $N_{i,\text{start}}$, the `tpr` path on disk, and the output path on disk. The preprocessing step requires prior knowledge of the input parameters (a). `GromPy` (3) reads the input parameters (b). Molecular insertions/removals requires `tpr` reads from disk (c). Once read, the associated data structures are kept in memory. The necessary energy evaluations are performed by the `GROMACS` library (4) with which `GromPy` communicates (d). This shared object library is compiled (e) from a slightly modified version (5) of `GROMACS` 4.0.5. The generated output can be further analyzed by the native `GROMACS` analysis suite, `GromPy`, or other software.

18

Figure 2:

Flowchart of `GromPy` in GCMC mode. Each MC move is based on the current state $s_c : \left[N_{i,c}, \mathbf{r}^{N_{i,c}}, \mathbf{v}^{N_{i,c}}\right]$ and associated total potential energy U_c, kept in memory. State s_c is defined by the number of molecules $N_{i,c}$ of type i, their coordinates $\mathbf{r}^{N_{i,c}}$ and velocities $\mathbf{v}^{N_{i,c}}$. `GromPy` uses two MC move types: an MD move of several MD steps and a grand-canonical MC move. After the MD move, the coordinates, velocities and total potential energy are updated. The GCMC move involves removal or insertion of a molecule selected with a probability $P_{\text{insert}} = P_{\text{remove}} = \frac{1}{2}$. Insertion or removal requires having s_t in memory. If this is not the case, we first read a `tpr` file with $N_i = N_{i,t}$ from disk. Insertions or removals are accepted with the probabilities in Eq. 1 and Eq. 2, respectively. If accepted, s_t becomes s_c and U_t becomes U_c. Otherwise, we keep s_c and U_c. After each MC move we update the averages and increment the MC loop iterator j.

19

Figure 3:

Modification of the source code of GROMACS version 4.0.5. Left: default compilation yields the mdrun executable (amongst others). This program calls function mdrunner() that is the calculation engine for MD simulations. Right: compilation of the mdrun executable as the shared object library libmdrun.so and splitting up function mdrunner() into an initialization stage an integration stage and a finalization stage. Communication between the stages is achieved through the new cs data structure. Library libmdrun.so is loaded into the GromPy module where mdr_init(cs), mdr_int(cs) and are called for performing MD moves and GCMC trial moves by manipulating cs before each MC move.

20

Figure 4:

(a)

(b)

Equations of state for the LJ model (a) and the SPC water model (b)) at $T = 773$ K (top) and $T = 900$ K (bottom). The data points on the $p(\rho)$ line are determined by MD at NVT. The points on the $\mu_{ex}(\rho)$ line are determined by grand-canonical MC using `GromPy` in GCMC mode. The standard deviations of μ_{ex} and of ρ for the μVT data were calculated by conventional error propagation rules. The least-squares fit of a 6th degree polynomial to these μ_{ex} points was transformed into a $p(\rho)$ curve using Eq. 4. Both results are shown as dotted lines in each plot.

21

Figure 5:

Equation of state for the LJ model at $T = 773$ K. The data points on the $p(\rho)$ line are determined by MD at NVT. The points on the $\mu_{\text{ex}}(\rho)$ line are determined by grand-canonical MC using `GromPy` in GCMC mode. The standard deviations of μ_{ex} and of ρ for the μVT data were calculated by conventional error propagation rules. The least-squares fit of a 6th degree polynomial to these μ_{ex} points was transformed into a $p(\rho)$ curve using Eq. 4. Both results are shown as dotted lines. Note that the data points for both simulation types are obtained by an equal number of integration steps. The error band width in the pressure profile based on the uncertainty in the μVT data is the area between the thin full lines.

22

C
References

1. M. Hoefling, H. Kessler and K.-E. Gottschalk. The Transmembrane Structure of Integrin αIIbβ3: Significance for Signal Transduction. *Angew. Chem., Int. Ed.*, 48(36):6590–6593, **2009**. (see page 79).

2. M. Hoefling, H. Kessler and K.-E. Gottschalk. Transmembranstruktur von Integrin αIIbβ3 - Bedeutung für die Signalübertragung. *Angew. Chem.*, 121:6714–6717, **2009**. (see page 83).

3. O. Cohavi, S. Corni, F. D. Rienzo et al. Protein-Surface Interactions: Challenging Experiments and Computations. *J. Mol. Recognit.*, 23(3):259–262, **2010**. (see page 87).

4. M. Hoefling and K.-E. Gottschalk. Barnase-Barstar: From First Encounter to Final Complex. *J. Struct. Biol.*, 171:52–63, **2010**. (see page 91).

5. D. B. Kokh, S. Corni, P. J. Winn et al. ProMetCS: an Atomistic Force Field for Modeling Protein-Metal Surface Interactions in a Continuum Aqueous Solvent. *J. Chem. Theory Comput.*, 6(5):1753–1768, **2010**. (see page 103).

6. M. Hoefling, F. Iori, S. Corni et al. The Conformations of Amino Acids on a Gold(111) Surface. *ChemPhysChem*, 11(8):1763–1767, **2010**. (see page 119).

7. M. G. Wolf, M. Hoefling, C. Aponte-Santamaría et al. g_membed: Efficient Insertion of a Membrane Protein Into an Equilibrated Lipid Bilayer With Minimal Perturbation. *J. Comput. Chem.*, 31(11):2169–2174, **2010**. (see page 125).

8. M. Hoefling, F. Iori, S. Corni et al. Interaction of Amino Acids With the Au(111) Surface: Adsorption Free Energies From Molecular Dynamics Simulations. *Langmuir*, 26(11):8347–8351, **2010**. (see page 131).

9. M. Hoefling, N. Lima, D. Hänni et al. Structural Heterogeneity and Quantitative FRET Efficiency Distributions of Polyprolines Through a Hybrid Atomistic

Simulation and Monte Carlo Approach. *PLoS One*, 6(5):e19791, **2011**. (see page 137).

10. M. Hoefling, S. Monti, S. Corni et al. Interaction of β-Sheet Folds With a Gold Surface. *PLoS One*, 6(6):e20925, **2011**. (see page 157).

11. R. Pool, A. Feenstra, M. Hoefling et al. Enabling Grand-Canonical Monte Carlo: Extending the Flexibility of GROMACS Through the GromPy Python Interface Module. *submitted*. (see page 171).

12. W. F. van Gunsteren, D. Bakowies, R. Baron et al. Biomolecular Modeling: Goals, Problems, Perspectives. *Angew. Chem., Int. Ed.*, 45(25):4064–4092, **2006**.

13. S. W. I. Siu, R. Vácha, P. Jungwirth et al. Biomolecular Simulations of Membranes: Physical Properties From Different Force Fields. *J. Chem. Phys.*, 128(12):125103, **2008**.

14. D. B. VanBeek, M. C. Zwier, J. M. Shorb et al. Fretting About FRET. Correlation Between κ and R. *Biophys. J.*, 92(12):4168–4178, **2007**.

15. J. J. Gray. The Interaction of Proteins With Solid Surfaces. *Curr. Opin. Struct. Biol.*, 14(1):110–115, **2004**.

16. M. Sarikaya, C. Tamerler, A. K.-Y. Jen et al. Molecular Biomimetics: Nanotechnology Through Biology. *Nat. Mater.*, 2(9):577–585, **2003**.

17. B. Fubini, I. Fenoglio, G. Martra et al. An Overview on the Toxicity of Inhaled Nanoparticles. In J. Blitz and V. Gun'ko, editors, *Surface Chemistry in Biomedical and Environmental Science*, pages 241–252 (Springer), **2006**.

18. A. Nel, T. Xia, L. Mädler et al. Toxic Potential of Materials at the Nanolevel. *Science*, 311(5761):622–627, **2006**.

19. R. Braun, M. Sarikaya and K. Schulten. Genetically Engineered Gold-Binding Polypeptides: Structure Prediction and Molecular Dynamics. *J. Biomater. Sci., Polym. Ed.*, 13(7):747–757, **2002**.

20. B. Peelle, E. Krauland, K. Wittrup et al. Design Criteria for Engineering Inorganic Material-Specific Peptides. *Langmuir*, 21(15):6929–6933, **2005**.

21. M. Sarikaya, C. Tamerler, D. Schwartz et al. Materials Assembly and Formation Using Engineered Polypeptides. *Annu. Rev. Mater. Res.*, 34:373–408, **2004**.

22. C. Tamerler, M. Duman, E. E. Oren et al. Materials Specificity and Directed Assembly of a Gold-Binding Peptide. *Small*, 2(11):1372–1378, **2006**.

23. H. Bohm. The Computer-Program Ludi - a New Method for the DeNovo Design of Enzyme-Inhibitors. *J. Comput.-Aided Mol. Des.*, 6(1):61–78, **1992**.

24. A. R. Ortiz, M. T. Pisabarro, F. Gago et al. Prediction of Drug Binding Affinities by Comparative Binding Energy Analysis. *J. Med. Chem.*, 38(14):2681–2691, **1995**.

25. J. C. Alvarez. High-Throughput Docking as a Source of Novel Drug Leads. *Curr. Opin. Chem. Biol.*, 8(4):365–370, **2004**.

26. F. Iori, R. D. Felice, E. Molinari et al. GolP: An Atomistic Force-Field to Describe the Interaction of Proteins With Au(111) Surfaces in Water. *J. Comput. Chem.*, 30(9):1465–1476, **2009**.

27. F. Iori, S. Corni and R. D. Felice. Unraveling the Interaction Between Histidine Side Chain and the Au(111) Surface: A DFT Study. *J. Phys. Chem. C*, 112(35):13540–13545, **2008**.

28. F. Iori and S. Corni. Including Image Charge Effects in the Molecular Dynamics Simulations of Molecules on Metal Surfaces. *J. Comput. Chem.*, 29(10):1656–1666, **2008**.

29. O. Cohavi, D. Reichmann, R. Abramovich et al. A Quantitative, Real-Time Assessment of Binding of Peptides and Proteins to Gold Surfaces. *Chem.–Eur. J.*, 17(4):1327–1336, **2011**.

30. M. Zink and H. Grubmüller. Mechanical Properties of the Icosahedral Shell of Southern Bean Mosaic Virus: A Molecular Dynamics Study. *Biophys. J.*, 96(4):1350–1363, **2009**.

31. D. E. Shaw, P. Maragakis, K. Lindorff-Larsen et al. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science*, 330(6002):341–346, **2010**.

32. H. J. C. Berendsen. *Simulating the Physical World: Hierarchical Modeling From Quantum Mechanics to Fluid Dynamics* (Cambridge University Press), 1st edition, **2007**.

33. C. L. Brooks, M. Karplus and B. M. Pettitt. *Advances in Chemical Physics, Proteins: A Theoretical Perspective of Dynamics, Structure, and Thermodynamics*, volume 71 (Wiley-Interscience), 1st edition, **1990**.

34. D. Frenkel and B. Smit. *Understanding Molecular Simulation, Second Edition: From Algorithms to Applications* (Academic Press), 2nd edition, **2001**.

35. M. Griebel, S. Knapek, G. Zumbusch et al. *Numerische Simulation in der Moleküldynamik. Numerik, Algorithmen, Parallelisierung, Anwendungen* (Springer), 1st edition, **2003**.

36. W. van Gunsteren, P. Weiner and A. Wilkinson, editors. *Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications*, volume 3 (Springer), 1st edition, **1997**.

37. B. Hess, D. van der Spoel and E. Lindahl. *GROMACS Manual 4.5.4*, **2011**.

38. M. Born and R. Oppenheimer. Zur Quantentheorie der Molekeln. *Ann. Phys.*, 389(20):457–484, **1927**.

39. Y. Duan, C. Wu, S. Chowdhury et al. A Point-Charge Force Field for Molecular Mechanics Simulations of Proteins Based on Condensed-Phase Quantum Mechanical Calculations. *J. Comput. Chem.*, 24(16):1999–2012, **2003**.

40. J. Wang, R. M. Wolf, J. W. Caldwell et al. Development and Testing of a General Amber Force Field. *J. Comput. Chem.*, 25(9):1157–1174, **2004**.

41. L. Schuler, X. Daura and W. F. van Gunsteren. An Improved Gromos96 Force Field for Aliphatic Hydrocarbons in the Condensed Phase. *J. Comput. Chem.*, 22(11):1205–1218, **2001**.

42. W. L. Jorgensen, D. Maxwell and J. Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *J. Am. Chem. Soc.*, 118(45):11225–11236, **1996**.

43. G. Kaminski, R. Friesner, J. Tirado-Rives et al. Evaluation and Reparametrization of the OPLS-AA Force Field for Proteins via Comparison With Accurate Quantum Chemical Calculations on Peptides. *J. Phys. Chem. B*, 105(28):6474–6487, **2001**.

44. M. Patra, M. Karttunen, M. T. Hyvonen et al. Molecular Dynamics Simulations of Lipid Bilayers: Major Artifacts Due to Truncating Electrostatic Interactions. *Biophys. J.*, 84(6):3636–3645, **2003**.

45. T. A. Darden, D. York and L. Pedersen. Particle Mesh Ewald: An N log (N) Method for Ewald Sums in Large Systems. *J. Chem. Phys.*, 98:10089, **1993**.

46. P. Ehrenfest. Bemerkung über die Angenäherte Gültigkeit der Klassischen Mechanik Innerhalb der Quantenmechanik. *Z. Phys. A: Hadrons Nucl.*, 45(7):455–457, **1927**.

47. A. Cossaro, R. Mazzarello, R. Rousseau et al. X-Ray Diffraction and Computation Yield the Structure of Alkanethiols on Gold(111). *Science*, 321(5891):943–946, **2008**.

48. E. Hairer, C. Lubich and G. Wanner. Geometric Numerical Integration Illustrated by the Störmer–Verlet Method. *Acta Num.*, 12:399–450, **2003**.

49. E. Hairer, S. P. Norsett and G. Wanner. *Solving Ordinary Differential Equations I: Nonstiff Problems* (Springer), 2nd edition 1993. corr. 3rd printing edition, **2010**.

50. T. Fliessbach. *Mechanik: Lehrbuch Zur Theoretischen Physik I* (Spektrum Akademischer Verlag), 6th edition, **2009**.

51. L. Verlet. Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules. *Phys. Rev.*, 159(1):98, **1967**.

52. W. Swope, H. Andersen, P. Berens et al. A Computer-Simulation Method for the Calculation of Equilibrium-Constants for the Formation of Physical Clusters of Molecules - Application to Small Water Clusters. *J. Chem. Phys.*, 76(1):637–649, **1982**.

53. M. E. Tuckerman, J. Alejandre, R. Lopez-Rendon et al. A Liouville-Operator Derived. Measure-Preserving Integrator for Molecular Dynamics Simulations in the Isothermal-Isobaric Ensemble. *J. Phys. A: Math. Gen.*, 39(19):5629–5651, **2006**.

54. T.-Q. Yu, J. Alejandre, R. Lopez-Rendon et al. Measure-Preserving Integrators for Molecular Dynamics in the Isothermal-Isobaric Ensemble Derived From the Liouville Operator. *Chem. Phys.*, 370(1-3):294–305, **2010**.

55. M. Ahmad, W. Gu, T. Geyer et al. Adhesive Water Networks Facilitate Binding of Protein Interfaces. *Nat. Commun.*, 2:261, **2011**.

56. A. Baumketner and J.-E. Shea. The Influence of Different Treatments of Electrostatic Interactions on the Thermodynamics of Folding of Peptides. *J. Phys. Chem. B*, 109(45):21322–21328, **2005**.

57. H. J. C. Berendsen, J. Postma, W. van Gunsteren et al. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*, pages 331–342 (Kluwer Academic Publishers), **1981**.

58. W. L. Jorgensen and J. Madura. Temperature and Size Dependence for Monte Carlo Simulations of TIP4P Water. *Mol. Phys.*, 56(6):1381–1392, **1985**.

59. T. L. Hill. *Statistical Mechanics: Principles and Selected Applications* (Dover Publications), **1987**.

60. H. J. C. Berendsen. Transport Properties Computed by Linear Response Through Weak Coupling to a Bath. In M. Meyer and V. Pontikis, editors, *Computer Simulations in Material Science*, pages 139–155 (Kluwer Academic Publishers), **1991**.

61. T. Morishita. Fluctuation Formulas in Molecular-Dynamics Simulations With the Weak Coupling Heat Bath. *J. Chem. Phys.*, 113(8):2976–2982, **2000**.

62. W. Hoover. Canonical Dynamics - Equilibrium Phase-Space Distributions. *Phys. Rev. A*, 31(3):1695–1697, **1985**.

63. S. Nose. A Molecular-Dynamics Method for Simulations in the Canonical Ensemble. *Mol. Phys.*, 52(2):255–268, **1984**.

64. G. Bussi, D. Donadio and M. Parrinello. Canonical Sampling Through Velocity Rescaling. *J. Chem. Phys.*, 126(1):014101, **2007**.

65. M. Parrinello and A. Rahman. Polymorphic Transitions in Single-Crystals - a New Molecular-Dynamics Method. *J. Appl. Phys.*, 52(12):7182–7190, **1981**.

66. H. J. C. Berendsen, D. van der Spoel and R. VanDrunen. GROMACS - a Message-Passing Parallel Molecular-Dynamics Implementation. *Comput. Phys. Commun.*, 91(1-3):43–56, **1995**.

67. B. Hess, C. Kutzner, D. van der Spoel et al. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.*, 4(3):435–447, **2008**.

68. C. Kutzner, D. van der Spoel, M. Fechner et al. Improved GROMACS Scaling on Ethernet Switched Clusters. In *Recent Advances in Parallel Virtual Machine and Message Passing Interface*, volume 4192, pages 404–405 (Springer), **2006**.

69. E. Lindahl, B. Hess and D. van der Spoel. GROMACS 3.0: A Package for Molecular Simulation and Trajectory Analysis. *J. Mol. Model.*, 7(8):306–317, **2001**.

70. D. van der Spoel, E. Lindahl, B. Hess et al. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.*, 26(16):1701–1718, **2005**.

71. C. Kutzner, D. van der Spoel, M. Fechner et al. Speeding Up Parallel GROMACS on High-Latency Networks. *J. Comput. Chem.*, 28(12):2075–2084, **2007**.

72. B. W. Kernighan and D. M. Ritchie. *C Programming Language* (Prentice Hall), 2nd edition, **1988**.

73. H. P. Langtangen. *Python Scripting for Computational Science.* Texts in Computational Science and Engineering (Springer), **2010**.

74. H. P. Langtangen. *A Primer on Scientific Programming With Python.* Texts in Computational Science and Engineering (Springer), 2nd edition, **2011**.

75. W. Humphrey, A. Dalke and K. Schulten. VMD – Visual Molecular Dynamics. *J. Mol. Graphics.*, 14:33–38, **1996**.

76. W. Delano. *The Pymol User's Manual*, **2002**.

77. T. Selzer and G. Schreiber. New Insights Into the Mechanism of Protein-Protein Association. *Proteins: Struct., Funct., Genet.*, 45(3):190–198, **2001**.

78. C. Chipot and A. Pohorille, editors. *Free Energy Calculations: Theory and Applications in Chemistry and Biology.* Springer Series in Chemical Physics (Springer), **2007**.

79. M. Goette and H. Grubmüller. Accuracy and Convergence of Free Energy Differences Calculated From Nonequilibrium Switching Processes. *J. Comput. Chem.*, 30(3):447–456, **2009**.

80. M. G. Wolf, J. A. Jongejan, J. D. Laman et al. Rapid Free Energy Calculation of Peptide Self-Assembly by Remd Umbrella Sampling. *J. Phys. Chem. B*, 112(43):13493–13498, **2008**.

81. D. Trzesniak, A.-P. E. Kunz and W. F. van Gunsteren. A Comparison of Methods to Compute the Potential of Mean Force. *ChemPhysChem*, 8(1):162–169, **2007**.

82. J. Kirkwood. Statistical Mechanics of Fluid Mixtures. *J. Chem. Phys.*, 3:300–314, **1935**.

83. T. Förster. Zwischenmolekulare Energiewanderung und Fluoreszens. *Ann. Phys.*, 2(1-2):55–75, **1948**.

84. T. W. J. Gadella, editor. *FRET and FLIM Techniques, Volume 33.* Laboratory Techniques in Biochemistry and Molecular Biology (Elsevier Science), 1st edition, **2009**.

85. J. R. Lakowicz. *Principles of Fluorescence Spectroscopy* (Springer), 2nd edition, **1999**.

86. E. A. Jares-Erijman and T. M. Jovin. FRET Imaging. *Nat. Biotechnol.*, 21(11):1387–1395, **2003**.

87. B. P. Krueger, G. D. Scholes and G. R. Fleming. Calculation of Couplings and Energy-Transfer Pathways Between the Pigments of LH2 by the Ab Initio Transition Density Cube Method. *J. Phys. Chem. B*, 102(27):5378–5386, **1998**.

88. A. Munosz-Losa, C. Curutchet, B. P. Krueger et al. Fretting About FRET. Failure of the Ideal Dipole Approximation. *Biophys. J.*, 96(12):4779–4788, **2009**.

89. G. D. Scholes. Long-Range Resonance Energy Transfer in Molecular Systems. *Annu. Rev. Phys. Chem.*, 54:57–87, **2003**.

90. Z. K. Majumdar, R. Hickerson, H. F. Noller et al. Measurements of Internal Distance Changes of the 30S Ribosome Using FRET With Multiple Donor-Acceptor Pairs: Quantitative Spectroscopic Methods. *J. Mol. Biol.*, 351(5):1123–1145, **2005**.

91. P. Dirac. The Quantum Theory of the Emission and Absorption of Radiation. *Proc. R. Soc. London, A*, 114(767):243–265, **1927**.

92. E. Fermi. *Nuclear Physics: A Course Given by Enrico Fermi at the University of Chicago* (University Of Chicago Press), revised edition, **1974**.

93. E. Condon. A Theory of Intensity Distribution in Band Systems. *Phys. Rev.*, 28:1182–1201, **1926**.

94. J. Franck and E. Dymond. Elementary Processes of Photochemical Reactions. *Trans. Faraday Soc.*, 21(2):536–542, **1926**.

95. D. J. Griffiths. *Introduction to Electrodynamics* (Benjamin Cummings), 3rd edition, **1999**.

96. R. McWeeny. *Methods of Molecular Quantum Mechanics (Theoretical Chemistry)* (Academic Press), 2nd edition, **1969**.

97. R. Knox and H. van Amerongen. Refractive Index Dependence of the Forster Resonance Excitation Transfer Rate. *J. Phys. Chem. B*, 106(20):5289–5293, **2002**.

98. S. W. Hell and J. Wichmann. Breaking the Diffraction Resolution Limit by Stimulated-Emission - Stimulated-Emission-Depletion Fluorescence Microscopy. *Opt. Lett.*, 19(11):780–782, **1994**.

99. S. W. Hell. Far-Field Optical Nanoscopy. *Science*, 316(5828):1153–1158, **2007**.

100. G. Donnert, J. Keller, R. Medda et al. Macromolecular-Scale Resolution in Biological Fluorescence Microscopy. *Proc. Natl. Acad. Sci. U. S. A.*, 103(31):11440–11445, **2006**.

101. C. Grashoff, B. D. Hoffman, M. D. Brenner et al. Measuring Mechanical Tension Across Vinculin Reveals Regulation of Focal Adhesion Dynamics. *Nature*, 466(7303):263–266, **2010**.

102. A. T. Brunger, P. Strop, M. Vrljic et al. Three-Dimensional Molecular Modeling With Single Molecule FRET. *J. Struct. Biol.*, 173(3):497–505, **2011**.

103. S. Uphoff, S. J. Holden, L. L. Reste et al. Monitoring Multiple Distances Within a Single Molecule Using Switchable FRET. *Nat. Methods*, 7(10):831–836, **2010**.

104. L. Stryer and R. P. Haughland. Energy Transfer - a Spectroscopic Ruler. *Proc. Natl. Acad. Sci. U. S. A.*, 58(2):719–726, **1967**.

105. S. Weiss. Fluorescence Spectroscopy of Single Biomolecules. *Science*, 283(5408):1676–1683, **1999**.

106. N. Panchuk-Voloshina, R. P. Haughland, J. Bishop-Stewart et al. Alexa Dyes, a Series of New Fluorescent Dyes That Yield Exceptionally Bright, Photostable Conjugates. *J. Histochem. Cytochem.*, 47(9):1179–1188, **1999**.

107. I. L. Medintz, A. R. Clapp, H. Mattoussi et al. Self-Assembled Nanoscale Biosensors Based on Quantum Dot FRET Donors. *Nat. Mater.*, 2(9):630–638, **2003**.

108. R. Roy, S. Hohng and T. Ha. A Practical Guide to Single-Molecule FRET. *Nat. Methods*, 5(6):507–516, **2008**.

109. H. Hofmann, F. Hillger, S. H. Pfeil et al. Single-Molecule Spectroscopy of Protein Folding in a Chaperonin Cage. *Proc. Natl. Acad. Sci. U. S. A.*, 107(26):11793–11798, **2010**.

110. F. Hillger, D. Haenni, D. Nettels et al. Probing Protein-Chaperone Interactions With Single-Molecule Fluorescence Spectroscopy. *Angew. Chem., Int. Ed.*, 47(33):6184–6188, **2008**.

111. B. Schuler, E. A. Lipman and W. A. Eaton. Probing the Free-Energy Surface for Protein Folding With Single-Molecule Fluorescence Spectroscopy. *Nature*, 419(6908):743–747, **2002**.

112. R. E. Dale, J. Eisinger and W. E. Blumberg. Orientational Freedom of Molecular Probes - Orientation Factor in Intra-Molecular Energy-Transfer. *Biophys. J.*, 26(2):161–193, **1979**.

113. A. Hillisch, M. Lorenz and S. Diekmann. Recent Advances in FRET. Distance Determination in Protein-DNA Complexes. *Curr. Opin. Struct. Biol.*, 11(2):201–207, **2001**.

114. D. Kajihara, R. Abe, I. Iijima et al. FRET Analysis of Protein Conformational Change Through Position-Specific Incorporation of Fluorescent Amino Acids. *Nat. Methods*, 3(11):923–929, **2006**.

115. E. Haas. The Study of Protein Folding and Dynamics by Determination of Intramolecular Distance Distributions and Their Fluctuations Using Ensemble and Single-Molecule FRET Measurements. *ChemPhysChem*, 6(5):858–870, **2005**.

116. I. Rasnik, S. A. McKinney and T. Ha. Surfaces and Orientations: Much to FRET About? *Acc. Chem. Res.*, 38(7):542–548, **2005**.

117. J. Yang and M. A. Winnik. The Orientation Parameter for Energy Transfer in Restricted Geometries Including Block Copolymer Interfaces: A Monte Carlo Study. *J. Phys. Chem. B*, 109(39):18408–18417, **2005**.

118. D. Badali and C. C. Gradinaru. The Effect of Brownian Motion of Fluorescent Probes on Measuring Nanoscale Distances by Förster Resonance Energy Transfer. *J. Chem. Phys.*, 134(22):225102, **2011**.

119. B. Corry, D. Jayatilaka, B. Martinac et al. Determination of the Orientational Distribution and Orientation Factor for Transfer Between Membrane-Bound Fluorophores Using a Confocal Microscope. *Biophys. J.*, 91(3):1032–1045, **2006**.

120. R. M. Clegg, A. Murchie, A. Zechel et al. Observing the Helical Geometry of Double-Stranded DNA in Solution by Fluorescence Resonance Energy Transfer. *Proc. Natl. Acad. Sci. U. S. A.*, 90(7):2994–2998, **1993**.

121. I. H. Stein, V. Schüller, P. Böhm et al. Single-Molecule FRET Ruler Based on Rigid DNA Origami Blocks. *ChemPhysChem*, 12(3):689–695, **2011**.

122. L. R. Allen and E. Paci. Orientational Averaging of Dye Molecules Attached to Proteins in Förster Resonance Energy Transfer Measurements: Insights From a Simulation Study. *J. Chem. Phys.*, 131(6):065101, **2009**.

123. B. Corry and D. Jayatilaka. Simulation of Structure, Orientation, and Energy Transfer Between Alexafluor Molecules Attached to MscL. *Biophys. J.*, 95(6):2711–2721, **2008**.

124. L. Loura, A. Carvalho and J. Ramalho. Direct Calculation of Förster Orientation Factor of Membrane Probes by Molecular Simulation. *J. Mol. Struct. Theochem*,

946(1-3):107–112, **2010**.

125. E. Dolghih, A. E. Roitberg and J. L. Krause. Fluorescence Resonance Energy Transfer in Dye-Labeled DNA. *J. Photochem. Photobiol., A*, 190(2-3):321–327, **2007**.

126. S. Kalinin, S. Felekyan, M. Antonik et al. Probability Distribution Analysis of Single-Molecule Fluorescence Anisotropy and Resonance Energy Transfer. *J. Phys. Chem. B*, 111(34):10253–10262, **2007**.

127. W. F. Harrington and M. Sela. Studies on the Structure of Poly-L-Proline in Solution. *Biochim. Biophys. Acta*, 27(1):24–41, **1958**.

128. B. Schuler, E. A. Lipman, P. J. Steinbach et al. Polyproline and the "Spectroscopic Ruler" Revisited With Single-Molecule Fluorescence. *Proc. Natl. Acad. Sci. U. S. A.*, 102(8):2754–2759, **2005**.

129. R. B. Best, K. A. Merchant, I. V. Gopich et al. Effect of Flexibility and Cis Residues in Single-Molecule FRET Studies of Polyproline. *Proc. Natl. Acad. Sci. U. S. A.*, 104(48):18964–18969, **2007**.

130. G. F. Schröder and H. Grubmüller. Maximum Likelihood Trajectories From Single Molecule Fluorescence Resonance Energy Transfer Experiments. *J. Chem. Phys.*, 119(18):9920–9924, **2003**.

131. I. V. Gopich and A. Szabo. Decoding the Pattern of Photon Colors in Single-Molecule FRET. *J. Phys. Chem. B*, 113(31):10965–10973, **2009**.

132. R. G. Brennan and B. W. Matthews. The Helix-Turn-Helix DNA Binding Motif. *J. Biol. Chem.*, 264(4):1903–1906, **1989**.

133. M. C. Stumpe and H. Grubmüller. Urea Impedes the Hydrophobic Collapse of Partially Unfolded Proteins. *Biophys. J.*, 96(9):3744–3752, **2009**.

134. W. Schulz. Note on the Computation of the Multipole Tensor in Spherical Harmonics. *Chem. Phys. Lett.*, 254(5-6):337–340, **1996**.

135. R. Baer and E. Rabani. Theory of Resonance Energy Transfer Involving Nanocrystals: The Role of High Multipoles. *J. Chem. Phys.*, 128(18):184710, **2008**.

136. C. Curutchet, G. D. Scholes, B. Mennucci et al. How Solvent Controls Electronic Energy Transfer and Light Harvesting: Toward a Quantum-Mechanical Description of Reaction Field and Screening Effects. *J. Phys. Chem. B*, 111(46):13253–13265, **2007**.

137. A. S. Deshpande and E. Beniash. Bioinspired Synthesis of Mineralized Collagen Fibrils. *Cryst. Growth Des.*, 8(8):3084–3090, **2008**.

138. M. A. Arnaout, S. L. Goodman and J.-P. Xiong. Structure and Mechanics of Integrin-Based Cell Adhesion. *Curr. Opin. Cell Biol.*, 19(5):495–507, **2007**.

139. C. Yang and K. A. Sharp. Hydrophobic Tendency of Polar Group Hydration as a Major Force in Type I Antifreeze Protein Recognition. *Proteins*, 59(2):266–274, **2005**.

140. S. R. Whaley, D. S. English, E. L. Hu et al. Selection of Peptides With Semiconductor Binding Specificity for Directed Nanocrystal Assembly. *Nature*, 405(6787):665–668, **2000**.

141. M. Erdmann, R. David, A. R. Fornof et al. Electrically Controlled DNA Adhesion. *Nat. Nanotechnol.*, 5(2):154–159, **2010**.

142. A. R. Fornof, M. Erdmann, R. David et al. Electric Glue: Electrically Controlled Polymer-Surface Adhesion. *Nano Lett.*, 11(5):1993–1996, **2011**.

143. R. A. Latour. Thermodynamic Perspectives on the Molecular Mechanisms Providing Protein Adsorption Resistance That Include Protein-Surface Interactions. *J Biomed Mater Res A*, 78A(4):843–854, **2006**.

144. D. Horinek, A. Serr, D. J. Bonthuis et al. Molecular Hydrophobic Attraction and Ion-Specific Effects Studied by Molecular Dynamics. *Langmuir*, 24(4):1271–1283, **2008**.

145. D. Horinek, A. Serr, M. Geisler et al. Peptide Adsorption on a Hydrophobic Surface Results From an Interplay of Solvation, Surface, and Intrapeptide Forces. *Proc. Natl. Acad. Sci. U. S. A.*, 105(8):2842–2847, **2008**.

146. R. D. Felice and S. Corni. Simulation of Peptide–Surface Recognition. *J. Phys. Chem. Lett.*, 2:1510–1519, **2011**.

147. R. Willett, K. Baldwin, K. West et al. Differential Adhesion of Amino Acids to Inorganic Surfaces. *Proc. Natl. Acad. Sci. U. S. A.*, 102(22):7817–7822, **2005**.

148. T. A. Springer. Adhesion Receptors of the Immune System. *Nature*, 346(6283):425–434, **1990**.

149. R. Alon, S. W. Feigelson, E. Manevich et al. α4β1-Dependent Adhesion Strengthening Under Mechanical Strain Is Regulated by Paxillin Association With the α4-Cytoplasmic Domain. *J. Cell Biol.*, 171(6):1073–1084, **2005**.

150. C. Kim, T.-L. Lau, T. S. Ulmer et al. Interactions of Platelet Integrin αIIb and β3 Transmembrane Domains in Mammalian Cell Membranes and Their Role in Integrin Activation. *Blood*, 113(19):4747–4753, **2009**.

151. K. R. Legate, S. A. Wickstrom and R. Fassler. Genetic and Cell Biological Analysis of Integrin Outside-In Signaling. *Genes Dev.*, 23(4):397–418, **2009**.

152. D. Schneider and D. M. Engelman. Involvement of Transmembrane Domain Interactions in Signal Transduction by α/β Integrins. *J. Biol. Chem.*, 279(11):9840–9846, **2004**.

153. C. Kim, F. Ye and M. H. Ginsberg. Regulation of Integrin Activation. *Annu. Rev. Cell Dev. Biol.*, **2011**.

154. P. Hughes, F. Diazgonzalez, L. Leong et al. Breaking the Integrin Hinge - a Defined Structural Constraint Regulates Integrin Signaling. *J. Biol. Chem.*, 271(12):6571–6574, **1996**.

155. K.-E. Gottschalk. A Coiled-Coil Structure of the αIIβ3 Integrin Transmembrane and Cytoplasmic Domains in Its Resting State. *Structure*, 13(5):703–712, **2005**.

156. K.-E. Gottschalk and H. Kessler. A Computational Model of Transmembrane Integrin Clustering. *Structure*, 12(6):1109–1116, **2004**.

157. T.-L. Lau, C. Kim, M. H. Ginsberg et al. The Structure of the Integrin αIIβ3 Transmembrane Complex Explains Integrin Transmembrane Signalling. *EMBO J.*, 28(9):1351–1361, **2009**.

158. T.-L. Lau, V. Dua and T. S. Ulmer. Structure of the Integrin αIIb Transmembrane Segment. *J. Biol. Chem.*, 283(23):16162–16168, **2008**.

159. T.-L. Lau, A. W. Partridge, M. H. Ginsberg et al. Structure of the Integrin β3 Transmembrane Segment in Phospholipid Bicelles and Detergent Micelles. *Biochemistry*, 47(13):4008–4016, **2008**.

160. B. Jójárt and T. A. Martinek. Performance of the General Amber Force Field in Modeling Aqueous POPC Membrane Bilayers. *J. Comput. Chem.*, 28(12):2051–2058, **2007**.

161. O. Berger, O. Edholm and F. Jahnig. Molecular Dynamics Simulations of a Fluid Bilayer of Dipalmitoylphosphatidylcholine at Full Hydration, Constant Pressure, and Constant Temperature. *Biophys. J.*, 72(5):2002–2013, **1997**.

162. D. P. Tieleman, J. Maccallum, W. L. Ash et al. Membrane Protein Simulations With a United-Atom Lipid and All-Atom Protein Model: Lipid-Protein Interac-

tions, Side Chain Transfer Free Energies and Model Proteins. *J. Phys.: Condens. Matter*, 18(28):S1221–S1234, **2006**.

163. I. M. A. Nooren and J. M. Thornton. Structural Characterisation and Functional Significance of Transient Protein-Protein Interactions. *J. Mol. Biol.*, 325(5):991–1018, **2003**.

164. R. W. Hartley. Barnase and Barstar: Two Small Proteins to Fold and Fit Together. *Trends Biochem. Sci.*, 14(11):450–454, **1989**.

165. J. Neumann and K.-E. Gottschalk. The Effect of Different Force Applications on the Protein-Protein Complex Barnase-Barstar. *Biophys. J.*, 97(6):1687–1699, **2009**.

166. G. Schreiber and A. R. Fersht. Interaction of Barnase With Its Polypeptide Inhibitor Barstar Studied by Protein Engineering. *Biochemistry*, 32(19):5145–5150, **1993**.

167. G. Strittmatter, J. Janssens, C. Opsomer et al. Inhibition of Fungal Disease Development in Plants by Engineering Controlled Cell-Death. *Nat. Biotechnol.*, 13(10):1085–1089, **1995**.

168. J. Janin. The Kinetics of Protein-Protein Recognition. *Proteins: Struct., Funct., Genet.*, 28(2):153–161, **1997**.

169. K.-E. Gottschalk, H. Neuvirth and G. Schreiber. A Novel Method for Scoring of Docked Protein Complexes Using Predicted Protein-Protein Binding Sites. *Protein Eng., Des. Sel.*, 17(2):183–189, **2004**.

170. G. Schreiber, C. Frisch and A. R. Fersht. The Role of Glu73 of Barnase in Catalysis and the Binding of Barstar. *J. Mol. Biol.*, 270(1):111–122, **1997**.

171. G. Schreiber and A. R. Fersht. Energetics of Protein-Protein Interactions - Analysis of the Barnase-Barstar Interface by Single Mutations and Double Mutant Cycles. *J. Mol. Biol.*, 248(2):478–486, **1995**.

172. F. Dong, M. Vijayakumar and H.-X. Zhou. Comparison of Calculation and Experiment Implicates Significant Electrostatic Contributions to the Binding Stability of Barnase and Barstar. *Biophys. J.*, 85(1):49–60, **2003**.

173. C. Camacho, Z. Weng, S. Vajda et al. Free Energy Landscapes of Encounter Complexes in Protein-Protein Association. *Biophys. J.*, 76(3):1166–1178, **1999**.

174. R. R. Gabdoulline and R. C. Wade. Brownian Dynamics Simulation of Protein-Protein Diffusional Encounter. *Methods*, 14(3):329–341, **1998**.

175. A. Spaar and V. Helms. Free Energy Landscape of Protein-Protein Encounter Resulting From Brownian Dynamics Simulations of Barnase : Barstar. *J. Chem. Theory Comput.*, 1(4):723–736, **2005**.

176. R. Abseher, H. Schreiber and O. Steinhauser. The Influence of a Protein on Water Dynamics in Its Vicinity Investigated by Molecular Dynamics Simulation. *Proteins: Struct., Funct., Genet.*, 25(3):366–378, **1996**.

177. K. Gallagher and K. A. Sharp. A New Angle on Heat Capacity Changes in Hydrophobic Solvation. *J. Am. Chem. Soc.*, 125(32):9853–9860, **2003**.

178. F. Despa, A. Fernandez and R. S. Berry. Dielectric Modulation of Biological Water. *Phys. Rev. Lett.*, 93(22):228104, **2004**.

179. I. Derenyi, G. Palla and T. Vicsek. Clique Percolation in Random Networks. *Phys. Rev. Lett.*, 94(16):160202, **2005**.

180. G. Palla, I. Derenyi, I. Farkas et al. Uncovering the Overlapping Community Structure of Complex Networks in Nature and Society. *Nature*, 435(7043):814–818, **2005**.

181. D. Seeliger and B. L. D. Groot. Protein Thermostability Calculations Using Alchemical Free Energy Simulations. *Biophys. J.*, 98(10):2309–2316, **2010**.

D
Acknowledgements

It is also a great pleasure to thank everyone from the IDK-NBT doctorate program, my colleagues but foremost Marie-Christine Blüm, Julia Zimmermann and Marilena Pinto who always lent an ear to us.

A big "thank you!" also to Stefano Corni, Francesco Iori and Elisa Molinari for the great times in Modena. I would also like to show my gratitude to Daria Kokh, Gideon Schreiber, Alexander Vaskjevich, Rebecca Wade and all the others for working together on the PROSURF project. I would also like to show my gratitude to Ben Schuler for the cooperation in the FRET project.

I am also obliged to many of my colleagues who supported me: Carl Burmeister, Nicole Dölker and Gerrit Groenhof for finding the right number of empty lines in the Gaussian input. Martin Fechner and Ansgar Esztermann realized my special software and hardware requirements on clusters and workstation and Eveline Heinemann and Ulrike Gerischer for making administrative work as painless as possible. I also thank Jan Neumann, Stephanus Fengler, Christian Blau and Maarten Wolf for the countless scientific discussions and both groups — Gambicrew in Munich and Grubigroup in Göttingen — for the pleasant atmosphere during the last years.

It is a great pleasure to thank everyone who helped me writing my dissertation successfully by making corrections: Camilo Aponte, Christian Blau, Carl Burmeister, Ulrike Daub, Stefanie Daub-Höfling, Thomas Eichner, Konstantin Eisler, Stephanus Fengler, Ulrike Gerischer, Helmut Grubmüller, Kay Gottschalk, Georg and Stephanie Höfling, Cornelius Kaschinski, Hadas Leonov, Jan Neumann and Maarten Wolf.

This thesis would not exist without my family and friends, who supported and accompanied me during the last years. I am indebted to my parents for letting me follow my interests and for their continuous support.

Finally, my greatest thanks belong to my wife for her boundless patience, support and trust through all the years we had together.

E

curriculum vitæ — Martin Höfling

Education

since 2007   PhD student at the Chair for Applied Physics, LMU[1], Munich and the MPI[2] for Biophysical Chemistry, Göttingen.

Jul 2007   Diploma in Physics, 1.1 (very good), Certificate of Special Studies in Biophysics.

2006-2007   Diploma thesis, Chair for Applied Physics, LMU, Munich, Supervisor: Dr. Kay E. Gottschalk. *Computer Based Analysis of Proteins and their Interactions with the Solvent and Surfaces.*

2004   Advanced practical: Chair for Applied Physics, LMU, Munich, Supervisor: Ferdinand Kühner. *Force spectroscopy with gold coated small cantilevers.*

2001-2007   Diploma studies in Physics at the LMU.

Research & Professional Experience

since 2009   PhD student in the group of Prof. Helmut Grubmüller, Dept. for Theoretical and Computational Biophysics, MPI for Biophysical Chemistry, Göttingen, *Combined Molecular Dynamics and Monte Carlo to simulate FRET in dye-labeled polyproline molecules.*

since 2007   PhD student in the group of Dr. Kay E. Gottschalk, Chair for Applied Physics, LMU, Munich, *Development and validation of an all-atom forcefield modeling interactions of biomolecules with gold.*

---

[1]Ludwig-Maximilians-University
[2]Max-Planck-Institute

Jul-Oct 2005	Student trainee in the Group of Prof. Vincent Moy, Dept. of Physiology and Biophysics, University of Miami, Miller School of Medicine, *Design and implementation of data retrieval and analysis software for Atomic Force Microscopy experiments.*
2004-2005	Student trainee, Chair for Applied Physics, LMU, Mentor: Ferdinand Kühner, *Gold coating of small cantilevers and implementation of peak detection in force spectroscopy analysis software.*

Honors & Prizes

Apr 2010	Poster Award, Computer Simulation and Theory of Macromolecules, Hünfeld, *Improving the accuracy of the Coulombic coupling treatment in FRET systems.*
Dec 2008	Apple Research & Technology Support (ARTS) Laureate, *iMD: interactive Molecular Dynamics as a human/machine interface for nanomedical applications* (Dr. Kay E. Gottschalk & Martin Höfling).
Oct 2008	International Doctorate Program - Nano Bio Technology (IDK-NBT) scholarship funded by the Elite Network of Bavaria (ENB).